



ROYAUME DU MAROC  
UNIVERSITE MOHAMMED V  
FACULTE DE MEDECINE ET DE PHARMACIE  
RABAT



Année : 2021

THESE N° : 02/20 CSVS

Centre d'Étude Doctorale des Sciences de la vie et de la Santé

Formation Doctorale : Biologie médicale, Pathologie Humaine et  
Expérimentale et Environnementale  
Option : Bio-Informatique / Biotechnologie

THÈSE DE DOCTORAT

Intitulé de thèse

**Analyses bio-informatiques des données génomiques,  
transcriptomiques et épigénétiques des gliomes grade II et III**

**Présentée et soutenue par Mohammed Amine Bendahou**

**le 22-01-2021**

**Jury**

<b>Pr. IBRAHIMI Azeddine</b>	<b>Faculté de Médecine et Pharmacie de Rabat, UMV</b>	<b>Président/Co- directeur de thèse</b>
<b>Pr. BADOU Abdallah</b>	<b>Faculté de Médecine et Pharmacie de Casablanca, UHII</b>	<b>Rapporteur</b>
<b>Pr. OUADGHIRI Mouna</b>	<b>Faculté de Médecine et Pharmacie de Rabat, UMV</b>	<b>Rapporteur</b>
<b>Pr. SBABOU Laila</b>	<b>Faculté des Sciences de Rabat, UMV</b>	<b>Rapporteur</b>
<b>Pr. BENKABBOU Amine</b>	<b>Faculté de Médecine et Pharmacie de Rabat, UMV</b>	<b>Examineur</b>
<b>Pr. BOUTARBOUH Mahjouba</b>	<b>Faculté de Médecine et Pharmacie de Rabat, UMV</b>	<b>Directeur de thèse</b>

## AVANT-PROPOS

Cette thèse a été réalisée au sein du Laboratoire de Biotechnologie Médicale (Medbiotech), Centre d'Étude Doctoral Science de la vie et de la Santé, Faculté de Médecine et Pharmacie, Université Mohammed V Rabat.

J'ai réalisé ma thèse sous la direction du Professeur BOUTARBOUCH Mahjouba, Professeur de l'enseignement supérieur à la Faculté de Médecine et Pharmacie de Rabat et neurochirurgienne à l'Hôpital des Spécialités, CHU Ibn Sina de Rabat, et du Professeur IBRAHIMI Azeddine, Professeur de l'enseignement supérieur et Directeur du Laboratoire de Biotechnologie Médicale (Médbiotech) à la Faculté de Médecine et de Pharmacie de Rabat.

Au cours de ces trois années quatre études ont été effectuées, la première était l'étude de l'impact des mutations IDH1, IDH2 et TP53 dans les gliomes de bas grade sur les protéines générées en utilisant des algorithmes génomiques et des outils de modélisation moléculaire, la 2ème étude était l'identification des biomarqueurs potentiels pour les gliomes de bas grade en analysant le profil génomique et le phénotype CpG Island Methylator, la 3ème étude était l'analyse bio-informatique des gènes et des miRNAs différentiellement exprimés dans les gliomes de bas grade et la 4<sup>ème</sup> étude était l'identification des modifications épigénétiques impliquées dans les gliomes de bas grade.

## Dédicaces

*A ma très chère mère*

*Affable, honorable, aimable : Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi.*

*Ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études.*

*Aucune dédicace ne saurait être assez éloquente pour exprimer ce que tu mérites pour tous les sacrifices que tu n'as cessé de me donner depuis ma naissance, durant mon enfance et même à l'âge adulte.*

*Tu as fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études.*

*Je te dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur.*

## *A mon cher Père*

*Tu as toujours été pour moi un exemple du père respectueux, honnête, de la personne méticuleuse, je tiens à honorer l'homme que tu es.*

*Grâce à toi papa j'ai appris le sens du travail et de la responsabilité. Je voudrais te remercier pour ton amour, ta générosité.*

*Ton soutien fut une lumière dans tout mon parcours. Aucune dédicace ne saurait exprimer l'amour l'estime et le respect que j'ai toujours eu pour toi.*

*Ce modeste travail est le fruit de tous les sacrifices que tu as déployés pour mon éducation et ma formation.*

*A mon très cher frère Oussama*

*Mon cher petit frère présent dans tous mes moments d'examens par son soutien moral et ses*

*belles surprises sucrées.*

*Je te souhaite un avenir plein de joie, de bonheur, de réussite et de sérénité.*

*Je t'exprime à travers ce travail mes sentiments de fraternité et d'amour.*

## Remerciement

Cette thèse est le fruit de la collaboration entre le Laboratoire de Biotechnologie Médicale (MedBiotech) de la Faculté de Médecine et de Pharmacie de Rabat de l'Université Mohammed V, et le service de Neurochirurgie à l'Hôpital des Spécialités, CHU Ibn Sina de Rabat. De plus, elle est le résultat d'un effort constant qui n'aurait pu aboutir sans la contribution d'un nombre de personnes. Ainsi se présente l'occasion de leur exprimer mes remerciements:

A ma Directrice de thèse Pr. Mahjouba BOUTARBOUCH, Professeur de l'enseignement supérieur à la Faculté de Médecine et Pharmacie de Rabat et neurochirurgienne à l'Hôpital des Spécialités, CHU Ibn Sina de Rabat, pour votre confiance dès le départ, pour votre encadrement, pour votre disponibilité sans faille malgré un emploi du temps plus que chargé et pour votre patience. Je tiens également à vous remercier pour la liberté d'action que vous m'avez donnée à chaque étape, vos précieux conseils, vos connaissances scientifiques et vos encouragements tout au long de ces trois années, afin d'aboutir à cette thèse. Je vous remercie également pour la qualité de vos relations humaine, votre rigueur scientifique, votre œil critique et pour tout le temps que vous m'avez consacré lors des discussions constructives que nous avons échangés. J'espère sincèrement que ce travail sera l'occasion de pouvoir continuer à travailler avec vous. Veuillez trouver ici l'expression de mes sincères remerciements et ma profonde reconnaissance.

A mon Co-encadrant de thèse Pr. Azeddine IBRAHIMI, Professeur d'Enseignement Supérieur et Directeur du Laboratoire de Biotechnologie Médicale (Médbiotech) à la Faculté de Médecine et de Pharmacie de Rabat, Université Mohammed V, pour m'avoir accueilli dans votre équipe de recherche, pour m'avoir initié dans ce domaine de Bio-Informatique et pour m'avoir guidé

avec compétence. Je tiens également à vous remercier pour votre patience et soutien tout au long de mon stage de master et de ces années de thèse. Son dynamisme et son enthousiasme ont également été une source de motivation et de persévérance. Je lui suis reconnaissant de ses conseils, son soutien et les opportunités qui m'a donné pour faire des cours de « cytologie et techniques cellulaires » aux étudiants de la première année du Master de Biotechnologie Médicale et du Master de Bio-informatique ainsi que la chance de co-encadrer des projets de fin d'études de Master. Son sens de la rigueur et son souci de perfection ont contribué l'accomplissement de ce travail.

A tous les membres de jury, malgré vos multiples occupations vous me faites l'honneur d'évaluer mon travail en acceptant de participer à mon jury. Je vous exprime ma profonde gratitude.

A Monsieur le Pr. Mohamed ADNAOUI, Doyen de la Faculté de Médecine et de Pharmacie de Rabat, Université Mohammed V pour m'avoir inscrit au sein de cet établissement.

A Monsieur le Pr. Taoufiq DAKKA, Directeur du Centre d'Etudes Doctorales des Sciences de la Vie et de la Santé à la Faculté de Médecine et de Pharmacie de Rabat, Université Mohammed V, pour votre aide administrative.

Je remercie le Centre National pour la Recherche Scientifique et Technique (CNRST) de m'avoir accordé une bourse d'excellence pour la recherche scientifique pendant ces trois années de thèse.

Je remercie également l'Union Européenne de m'avoir accordé une bourse Erasmus+ pour effectuer un stage de recherche de 8 mois au sein de "Interuniversity Institute of Bioinformatics in Brussels" sous la direction du Pr. Wim Vranken.

Je tiens à remercier mes parents et mon frère pour leur soutien permanent, constant et surtout sans aucune faille avec la patience et la confiance que vous avez toujours placée en moi, vous m'avez permis d'avancer en toute quiétude. J'espère qu'ils trouveront en ce modeste travail une récompense de ce qu'ils ont fait pour moi.

Finalement, au terme de ce travail je tiens à remercier toutes les personnes que j'ai rencontrés lors de mes études doctorales et qui ont contribué de près ou de loin au bon déroulement de cette thèse.

*'Promesse de la science : la science moderne a pour but aussi peu de douleur que possible'*

*Friedrich Nietzsche*

## Résumé

### **Analyses bio-informatiques des données génomiques, transcriptomiques et épigénétiques des gliomes grade II et III**

**Auteur :** Bendahou Mohammed Amine

**Mots clés :** Génomique, Transcriptomique, Épigenétique, Bio-informatique, Gliome de bas grade.

L'utilisation des approches omiques a révolutionné la biologie et elle a conduit à l'émergence d'approches systémiques y compris la génomique, la transcriptomique et l'épigénétique.

Dans le premier chapitre, nous avons ciblé des mutations impliquées dans les gliomes, suivi d'une analyse bio-informatique pour étudier l'impact de ces mutations sur la structure et la fonction des 3 protéines IDH1, IDH2 et p53. Nous avons obtenu une mutation qui a un effet sur le site catalytique de la protéine IDH1 comme R132H. D'autres mutations de la p53 ont été identifiées comme K305N, qui est une mutation pathogène ; R175H, qui est une mutation bénigne ; et R158G, qui perturbe la conformation structurale de la protéine suppresseur de tumeur.

Dans le 2<sup>ème</sup> chapitre, on a identifié des biomarqueurs spécifiques pour les gliomes en analysant des données génomiques et épigénomiques. La prévalence des mutations IDH dans le gliome de bas grade nous a guidé à cibler leurs conséquences épigénomiques en aval, telles que G-CIMP. On a trouvé que les astrocytomes sont caractérisés par une hyperméthylation au niveau du gène MYT1L, et les oligodendrogliomes au niveau du gène PLCG1.

Dans le 3<sup>ème</sup> chapitre, on a identifié les gènes et les miRNAs différentiellement exprimés au niveau des gliomes. Ces biomarqueurs sont sélectionnés sur la base d'une combinaison de seuil de changement d'expression et de seuil de score, et nous avons proposé un nouveau réseau de miARNs qui combinent à la fois des gènes et des lncARNs spécifique pour les gliomes de bas grade.

Dans le chapitre 4 on a démontré que les gliomes présentent également des altérations épigénétiques, parmi eux on trouve la méthylation de l'ADN et leur interaction avec les modifications d'histones.

Enfin, on a réussi à caractériser les gliomes de bas grade en analysant les différentes données omiques et on a mis en place un workflow bio-informatique précis et efficace pour analyser ces données et avoir une stratégie prometteuse pour mieux comprendre le développement du gliome.

## Abstract

### **Bioinformatics analyzes of genomic, transcriptomic and epigenetic data of grade II and III gliomas**

**Author:** Bendahou Mohammed Amine

**Keywords:** Genomics, Transcriptomics, Epigenetics, Bioinformatics, Low grade glioma.

The use of omics approaches has revolutionized biology and has led to the emergence of systemic approaches including genomics, transcriptomics and epigenetics.

In the first chapter, we targeted mutations involved in gliomas, starting with the sequencing of genomic regions encompassing isocitrate dehydrogenase 1 (IDH1) exon 4 and isocitrate dehydrogenase 2 (IDH2) and the four exons (5-6 and 7-8) of TP53 from 32 samples, followed by a bioinformatic analysis to study the impact of these mutations on the structure and function of the 3 proteins IDH1, IDH2 and p53. We have obtained a mutation which has an effect on the catalytic site of IDH1 protein such as. Other p53 mutations have been identified as K305N, which is a pathogenic mutation; R175H, which is a benign mutation; and R158G, which disrupts the structural conformation of the tumor suppressor protein.

In the 2nd chapter, we identified specific biomarkers for gliomas by analyzing genomic and epigenomic data. The prevalence of IDH mutations in low-grade glioma has guided us to target their downstream epigenomic consequences, such as G-CIMP. It has been found that astrocytomas are characterized by hypermethylation at the level of the MYT1L gene, and oligodendrogliomas at the level of the PLCG1 gene.

In the 3rd chapter, we identified the genes and miRNAs differentially expressed in gliomas. These biomarkers are selected on the basis of a combination of expression change threshold and score threshold, and we have proposed a new network of miRNAs that combine to both genes and lncRNAs specific for low grade gliomas.

In Chapter 4 it was shown that low grade gliomas also exhibit epigenetic alterations, among them we find DNA methylation and their interaction with histone modifications.

In the end, we succeeded in characterizing low-grade gliomas by analyzing the different data omics and we set up a precise and efficient bioinformatic workflow to analyze these data and have a promising strategy to better understand the development of the glioma.

## ملخص

### الثانية الدرجة من الدبقية للأورام والجينية والنسخية الجينومية للبيانات الحيوية المعلوماتية تحليلات والثالثة

أمين بندوح محمد :المؤلف

الدرجة منخفض الدبقي الورم ، الحيوية المعلوماتية ، التخلق علم ، النسخ علم ، الجينوم علم :المفتاحية الكلمات .

أحدث استخدام نهج omics ثورة في علم الأحياء وأدى إلى ظهور مناهج منهجية بما في ذلك علم الجينوم وعلم النسخ وعلم التخلق لتعزيز فهمنا للآليات والمسارات الجزيئية للأورام الدبقية منخفضة الدرجة بما في ذلك الأورام النجمية والأورام الدبقية قليلة التغصن وتحديد المؤشرات الحيوية المحتملة المشاركة في أورام الدماغ هذه .

في الفصل الأول ، استهدفنا الطفرات المتضمنة في الأورام الدبقية منخفضة الدرجة ، بدءًا من تسلسل المناطق الجينومية التي تشمل الأكسون 4 من IDH1 , IDH2 والأكسونات الأربعة (6-5 و 8-7 ) من TP53 من 32 عينة ، متبوعًا بتحليل المعلومات الحيوية لدراسة تأثير هذه الطفرات على بنية ووظيفة البروتينات الثلاثة IDH1 و IDH2 و p53. لقد حصلنا على طفرة لها تأثير على الموقع التحفيزي لبروتين IDH1 مثل R132H. تم التعرف على طفرات p53 الأخرى على أنها K305N ، وهي طفرة ممرضة ؛ R175H ، وهي طفرة حميدة ؛ و R158G ، مما يعطل التكوين الهيكلي للبروتين الكابت للورم.

في الفصل الثاني ، حددنا مؤشرات حيوية محددة للأورام النجمية والأورام الدبقية قليلة التغصن من خلال تحليل البيانات الجينومية والتخلقية. قادنا انتشار طفرات IDH في الورم الدبقي منخفض الدرجة إلى استهداف عواقبها اللاجينومية اللاحقة ، مثل G-CIMP. لقد وجد أن الأورام النجمية تتميز بفرط الميثيل على مستوى جين MYT1L ، وأورام الدبقية قليلة التغصن على مستوى جين PLCG1.

في الفصل الثالث ، حددنا الجينات و miRNAs المعبر عنها تفاضليًا في الأورام النجمية والأورام الدبقية قليلة التغصن. يتم اختيار هذه المؤشرات الحيوية على أساس مزيج من عتبة تغيير التعبير وعتبة النتيجة ، والتي تستند عمومًا إلى قيم P التي تم إنشاؤها بواسطة النمذجة الإحصائية ، وقد اقترحنا شبكة جديدة من miRNAs التي تتحد مع كل من الجينات و IncARNs الخاصة بـ منخفضة الصف الورم الدبقي.

في الفصل الرابع ، تبين أن الأورام الدبقية منخفضة الدرجة تظهر أيضًا تغيرات جينية ، من بينها نجد مثيلة الحمض النووي وتفاعلها مع تعديلات هيستون.

في النهاية ، نجحنا في توصيف الأورام الدبقية منخفضة الدرجة من خلال تحليل بيانات omics المختلفة وقمنا بإعداد سير عمل معلوماتي بيولوجي دقيق وفعال لتحليل هذه البيانات ولدينا استراتيجية واعدة لفهم تطور الورم الدبقي بشكل أفضل.

# SOMMAIRE

Avant-propos.....	2
Dédicaces .....	3
Remerciement .....	6
Résumé.....	10
Abstract .....	11
ملخص .....	12
Sommaire.....	13
Liste des figures.....	17
Liste des tableaux.....	19
Liste des abréviations .....	20
Production scientifique.....	22
Chapitre 1.....	24
I – Introduction générale.....	25
1- Définition des gliomes de bas grade.....	25
2- Caractéristiques principales et histoire naturelle.....	31
3 - Tests de diagnostic moléculaire.....	33
II - Recommandations actuelles en matière de traitement .....	34
1 - La chirurgie.....	34
2 – La radiothérapie .....	36
3 - Chimiothérapie et chimio radiothérapie .....	38
Chapitre 2.....	41
I – La bio-informatique .....	42
1 – Définition .....	42
2 – Objectifs de la bio-informatique.....	42
3 – L’impact de la bio-informatique .....	44
4 – Les limites de la bio-informatique.....	45
5 – Perspectives de la bio-informatique .....	46
II – Les bases de données .....	46
1 – Définition .....	47
2 – Types de bases de données .....	47
2 – 1 – Bases de données relationnelles .....	48
2 – 2 – Bases de données orientées objet.....	48
2 – 3 – Bases de données biologiques.....	49
2 – 3 – 1 – Bases de données primaires .....	50
2 – 3 – 2 – Bases de données secondaires .....	50

2 – 3 – 3 – Bases de données spécialisées.....	51
3 – Interconnexion entre les bases de données biologiques.....	52
<b>III – Biomarqueurs et médecine personnalisées .....</b>	<b>53</b>
1 – Biomarqueurs du cancer .....	53
2 – L'intérêt du traitement ciblé : la médecine de précision.....	54
3 – The cancer genome atlas (TCGA) .....	55
3 – 1 – Impact de la tcga .....	55
3 – 2 – Ethique et politiques de la tcga.....	56
<b>IV – Les approches omiques .....</b>	<b>57</b>
1 – Définition .....	57
2 – Du génome au protéome.....	57
3 – Les applications des approches omiques .....	58
3 – 1 – La génomique .....	58
3 – 2 – La transcriptomique .....	61
3 – 3 – L'épigénomique.....	62
3 – 4 – La protéomique.....	63
3 – 5 – La métabolomique .....	64
4 – Approches basées sur le séquençage de la prochaine génération.....	65
5 – L'intégration des données omiques .....	67
6 – Avenir et défis des études multi-omiques .....	68
<b>Chapitre 3.....</b>	<b>70</b>
<b>Analyse computationnelle des mutations IDH1, IDH2 et TP53 dans les gliomes de bas grade, y compris les oligodendrogliomes et les astrocytomes.....</b>	<b>71</b>
<b>I - Introduction.....</b>	<b>71</b>
<b>II- Objectifs .....</b>	<b>72</b>
<b>III – Matériel et méthodes.....</b>	<b>72</b>
<b>1- Déclaration éthique .....</b>	<b>72</b>
<b>2- Conception de l'étude et collecte d'échantillons .....</b>	<b>73</b>
<b>3 - Préparation d'adn, amplification des gènes idh1, idh2 et tp53 et séquençage.....</b>	<b>73</b>
<b>4 - Analyse computationnelle .....</b>	<b>74</b>
<b>IV - Résultats .....</b>	<b>78</b>
<b>V - Discussion .....</b>	<b>85</b>
<b>VI – Conclusion.....</b>	<b>88</b>
<b>Identification des biomarqueurs potentiels pour les gliomes de bas grade en analysant les profils génomiques et le phénotype CpG island methylator des patients dans la base de données TCGA .....</b>	<b>90</b>
<b>I - Introduction.....</b>	<b>90</b>
<b>II – Objectifs.....</b>	<b>90</b>

III – Matériel et méthodes.....	90
1– Collecte de données .....	90
2- Analyse bio-informatique .....	91
IV – Résultats .....	94
1 – Distribution des gènes impliqués dans les gliomes de bas grade .....	94
2 – Paysage mutationnel des altérations somatiques des gliomes de bas grade .....	95
3 – Classification des variantes .....	97
4 – Interaction des gènes .....	100
5 – Analyse de survie des patients .....	100
6 – Visualisation du glioma cpg island methylator phenotype .....	101
V- Discussion .....	103
VI- Conclusion.....	105
<b>Analyse bio-informatique des gènes et des miARNs différentiellement exprimés dans les gliomes de bas grade.....</b>	<b>106</b>
I - Introduction.....	106
II – Objectifs.....	107
III – Matériel et méthodes.....	107
1 – Traitement des données transcriptomiques .....	107
2 – Analyse de séquence d'ARN et des miARNs .....	108
2 – 1 – Analyse différemment exprimée.....	108
2 – 2 – Analyse d'enrichissement fonctionnel.....	109
2 – 3 – Analyse de survie globale des patients .....	109
3 – 4 – Construction du réseau des miarns .....	110
3 – 4 – 1– Calcule de test hypergéométrique .....	110
3 – 4 – 2– Régulation similaire.....	111
3 – 4 – 3– Corrélation de sensibilité.....	111
IV – Résultats .....	112
1– Identification des gènes et des miARNs différentiellement exprimés .....	112
2 – Analyse des voies signalisation des gènes différentiellement exprimés.....	112
3 – Analyse de la survie globale des patients .....	116
4 – Construction des réseaux des gènes et de miarns .....	119
V - Discussion .....	122
VI – Conclusion.....	124
<b>Profil global des modifications épigénétiques impliquées dans les gliomes de bas grade.....</b>	<b>125</b>
I – Introduction .....	125
II – L’hyperméthylation d'adn induite par la mutation idh.....	125
III – Modifications des histones dans les gliomes de bas grade .....	126
IV – La méthylation de l'adn affectant les contacts de la chromatine .....	129

<b>V- Conclusion .....</b>	<b>129</b>
<b>Conclusion générale.....</b>	<b>130</b>
<b>Références.....</b>	<b>137</b>

## Liste des figures

<i>Figure 1 : La classification histologique et moléculaire intégrée des gliomes selon l'OMS 2016</i>	27
<i>Figure 2 : Les différents sous-domaines de la bioinformatique. Les outils bioinformatique se répartissent en trois domaines : l'analyse des séquences, l'analyse des structures et l'analyse des fonctions</i>	44
<i>Figure 3 : L'algorithme de prédiction de polyphen 2</i>	76
<i>Figure 4 : Le protocole I-TASSER pour l'identification de la structure et de la fonction des protéines</i>	77
<i>Figure 5 : Prédiction des effets possibles des mutations IDH1, IDH2 et TP53</i>	80
<i>Figure 6 : Le profil rWCN (ligne bleue) et le profil ConSurf (ligne noire) de la protéine IDH. Les gènes mutés (A) IDH1, (B) IDH2 et (C) TP53 sont marqués dans un cercle rouge. Le rWCN et les scores de conservation sont normalisés à leurs scores respectifs. Profils normalisés du facteur B d'I-TASSER pour (D) IDH1, (E) IDH2 et (F) TP53</i>	82
<i>Figure 7 : Diagramme de Ramachandran des modèles (A) IDH1, (B) IDH2 et (C) p53. La région centrale, la région autorisée et la région générale sont colorées respectivement en rouge, jaune et beige</i>	84
<i>Figure 8 : Chromatogrammes des séquences des gènes IDH1, IDH2 et TP53 mutés et analyse structurale des protéines portant les mutations (A) R132H, (B) R172M (B) et (C) R158G</i>	84
<i>Figure 9 : Distribution of genes involved in low-grade gliomas</i>	93
<i>Figure 10 : Altération mutationnelle de 501 échantillons atteints de gliome de bas grade</i>	94
<i>Figure 11 : Les gènes impliqués dans les astrocytomes (A) barplot montrent les 10 premiers mutés qui sont classés en fonction de leur fréquence de mutation. (B) Classification des variantes. (C) Type de variante. (D) Classe SNV. (E) Tracé de transition et de transversion affichant la distribution des SNV en six événements de transition et de transversion. Le diagramme à barres empilées (en bas) montre la distribution des spectres de mutation pour chaque échantillon. (F) paires de gènes mutuellement exclusifs et cooccurrents affichés sous forme de matrice triangulaire</i>	97
<i>Figure 12 : Les gènes impliqués dans les oligodendrogliomes (A) barplot montrent les 10 premiers mutés qui sont classés en fonction de leur fréquence de mutation. (B) Classification des variantes. (C) Type de variante. (D) Classe SNV. (E) Tracé de transition et de transversion affichant la distribution des SNV en six événements de transition et de transversion. Le diagramme à barres empilées (en bas) montre la distribution des spectres de mutation pour chaque échantillon. (F) paires de gènes mutuellement exclusifs et cooccurrents affichés sous forme de matrice triangulaire</i>	98
<i>Figure 13 : Estimations de Kaplan–Meier de la survie globale des patients atteints d'astrocytomes et d'oligodendrogliomes classés en fonction de la mutation IDH1 et du statut de type sauvage IDH1</i>	99

<i>Figure 14 : (A) La valeur moyenne de la methylation d'ADN chez les patients atteints d'astrocytome et d'oligodendrogliome. (B) Carte thermique presentant toute les poitions hypermethylyé et hypomethylyé chez les patients atteints d'astrocytome et d'oligodendrogliome</i>	102
<i>Figure 15 : (A) Diagramme de volcan des DEG. (B) Diagramme à colonnes des DEG. (C) Diagramme à points de l'analyse d'enrichissement des OG</i>	113
<i>Figure 16 : (A) Les gènes hautement exprimés. (B) Courbes de Kaplan-Meier pour la survie globale des patients en fonction de gènes hautement exprimés</i>	112
<i>Figure 17 : (A) Les gènes faiblement exprimés. (B) Courbes de Kaplan-Meier pour la survie globale des patients en fonction de gènes hautement exprimés</i>	113
<i>Figure 18 : Heatmap des miARNs différentiellement exprimés impliqués dans les gliomes de bas grade, incluant les astrocytomes et les oligodendroglomes</i>	115
<i>Figure 19: Courbes de Kaplan-Meier pour la survie globale des patients en fonction des miARNs fortement exprimés (A) et des miARNs faiblement exprimés (B)</i>	117
<i>Figure 20 : (A) Le réseau des gènes faiblement exprimés (B) Le réseau des gènes hautement exprimés</i>	119
<i>Figure 21 : Le réseau des miARNs. (A) Les carrés rouges indiquent des miARNs hautement exprimés, les cercles bleu foncé représentent les gènes et les cercles bleu clair représentent les lncARNs. (B) Les carrés verts indiquent les miARNs exprimés bas, les cercles bleu foncé représentent les gènes et les cercles bleu clair représentent les lncARNs</i>	120
<i>Figure 22: Profils moyens des marqueurs d'histones H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K9ac et H3K9me3</i>	127
<i>Figure 23: Carte thermique des marques d'histones H3K4me1, H3K4me3, H3K27ac, H3K9ac, H3K9me3, H3K27me3 et H3K36me3 pour les tissus cérébraux</i>	128

## Liste des tableaux

<i>Tableau 1 : Mutations des gènes IDH1, IDH2 et TP53</i>	79
<i>Tableau 2 : Les 8 miARNs les plus exprimés et les moins exprimés dans les gliomes de bas grade</i>	117
<i>Tableau 3 : Les 8 miARNs les plus exprimés et les moins exprimés dans les gliomes de bas grade</i>	117

## Liste des abréviations

*ADN* Acide désoxyribonucléique  
*ARN* Acide ribonucléique  
*ARNm* ARN messenger  
*ATRX* Alpha thalassemia/mental retardation syndrome X-linked  
*BioLiP* Database of biologically relevant ligand–protein interactions  
*BP* Processus biologiques  
*CC* Composant cellulaire  
*CGH* Comparative genomic hybridization  
*Chip-seq* Combines chromatin immunoprecipitation sequencing  
*COBRA* Common Object Request Broker Architecture  
*DDBJ* DNA Data Bank of Japan  
*DGE* Differential gene expression  
*EDTA* Éthylènediaminetétraacétique  
*EMBL* Laboratoire européen de biologie moléculaire  
*EORTC* Organisation européenne pour la recherche et le traitement du cancer  
*ERK* Signal extracellulaire de la kinase  
*FFPE* Blocs de tissus inclus en paraffine  
*G-CIMP* Glioma CpG island methylator phenotype  
*GEO* Gene Expression Omnibus  
*OG* Ontology gene  
*HuR* Human antigen R  
*I-TASSER* Iterative Threading ASSEmby Refinement  
*IDH* Isocitrate déshydrogénase  
*IMRT* Radiothérapie à modulation d'intensité  
*Institut Nationale du Cancer*  
*IRM* Imagerie par résonance magnétique  
*LGG* Low grade glioma  
*lncRNA* Long non-coding RNAs  
*MF* Fonction moléculaire  
*MGMT* O6-methylguanine-DNA methyltransferase  
*miARN* MicroRNA  
*MIGS* Minimum information about a genome sequence

*NCBI National Center for Biotechnology Information*  
*NHGRI Institut Nationale de Recherche sur le Génome Humain*  
*NIH National Institutes of Health*  
*NOS Non spécifié autrement*  
*OMS Organisation mondiale de la santé*  
*PCR Polymerase chain reaction*  
*PCV Procarbazine, lomustine et vincristine*  
*PDB Protein Data Bank*  
*PIR Protein Information Resource*  
*PTM Modifications post-traductionnelles*  
*QV Qualité de vie*  
*RMN Résonance magnétique nucléaire*  
*RTOG Radiation Therapy Oncology Group*  
*rWCN Réciproque du profil WCN*  
*SG Survie globale*  
*SNP Single-nucleotide polymorphism*  
*SNV Single-nucleotide variant*  
*SSP Survie sans progression*  
*TAD Domaines d'association topologique*  
*TAIR The Arabidopsis Information Resource*  
*TCGA The Cancer Genome Atlas*  
*TERT Telomerase Reverse Transcriptase*  
*TET2 Ten-Eleven Translocation-2*  
*TM-score Score de modélisation*  
*TP53 Tumor protein p53*  
*VIH Virus de l'immunodéficience humaine*  
*WCN Numéro de contact pondéré*  
*WGBS Whole-genome bisulfite sequencing*  
*XML eXtensible Markup Language*  
*2-HG 2-hydroxyglutarate*  
*5-mC 5-méthylcytosine*  
*5caC 5-carboxylcytosine*  
*5fC 5-formylecytosine*

## Production scientifique

### Liste des publications

**Bendahou MA**, Ibrahimi A, Boutarbouch M, et al. Computational Analysis of IDH1, IDH2, and TP53 Mutations in Low-Grade Gliomas Including Oligodendrogliomas and Astrocytomas. *Cancer Informatics*. (2020)19:1176935120915839.

**Bendahou MA**, Ibrahimi A, Boutarbouch M. Identification of Potential Biomarkers for Low-Grade Gliomas by Analyzing the Genomics Profiles and CpG Island Methylator Phenotype of Patients in the TCGA Database. *Int J Appl Biol Pharm*. (2020):188-199.

**Bendahou MA**, Ibrahimi A, Boutarbouch M. Bioinformatics analysis of differentially expressed genes and miRNAs in low-grade gliomas. *Cancer Informatics*. (2020)10.1177/1176935120969692.

**Bendahou MA**, Ibrahimi A, Boutarbouch M. Global profile of epigenetic modifications involved in low-grade gliomas. En cours de publication.

Lahlou L, El Mrimar N, Laamarti M, Alouane T, **Bendahou MA**, Bssaibis F, Ben Lahlou Y, Zegmout A, El Hafidi N, ElJaoudi R, Frikh M, Lemnouar A, Elouennass M, Ibrahimi A. Whole-Genome Shotgun Sequences of Three Multidrug-Resistant Mycobacterium tuberculosis Strains Isolated from Morocco. *Genome Announc*. 2017 Nov 16;5(46):e01275-17. doi: 10.1128/genomeA.

### Communications orales

**Bendahou MA**, Ibrahimi A, Boutarbouch M. Molecular characterization of gliomas : identification of biomarkers in astrocytomas and oligodendrogliomas. 1st edition of the international congress "Computational Systems Biology of Cancer". Pierre and Marie Curie University Paris. September 24th - 29th.

**Bendahou MA**, Ibrahimi A, Boutarbouch M. Molecular characterization of gliomas : identification of biomarkers in astrocytomas and oligodendrogliomas. 8th research days Faculty of Dental Medicine of Rabat. April 19, 2018.

**Bendahou MA**, Ibrahimi A, Boutarbouch M . Molecular characterization of gliomas in the Moroccan population. 1st edition of the international conference Technology, Innovation and Information Systems. National School of Applied Sciences of Tetouan. March 17-18, 2018.

**Bendahou MA**, Ibrahimi A, Boutarbouch M . Molecular characterization of gliomas in the Moroccan population. European Researchers' Night. Maison des sciences de l'Homme, University of Burgundy Dijon. September 28, 2018.

### **Communications affichées**

**Bendahou MA**, Ibrahimi A, Boutarbouch M. Molecular characterization of diffuse gliomas : identification of biomarkers in astrocytomas and oligodendrogliomas. 8th Scientific Days of CEDoc-SVS, 5th Scientific Days of AMADOC SVS. Faculty of Medicine and Pharmacy Rabat. March 27-31, 2018.

### **Concours francophone international ma thèse en 180 secondes**

- 1er Prix du Jury de la finale de l'Université Mohammed V de Rabat. 11 Juin 2018
- Deux Prix lors de la Finale Nationale Marocaine. 28 Juin 2018.
- Représentant du Maroc à la Nuit Européenne des Chercheurs à Dijon 2018.

## **CHAPITRE 1**

*L'objectif de ce chapitre est de donner une vue globale sur le gliome de bas grade, ses caractéristiques principales et son histoire naturelle, les tests de diagnostic moléculaire ainsi que les recommandations actuelles en matière de traitement notamment la chirurgie, la radiothérapie, la chimiothérapie et la chimio radiothérapie.*

## I – Introduction générale

### 1- Définition des gliomes de bas grade

Au cours des trois dernières années, de nouvelles observations concernant les précurseurs moléculaires et les diagnostics moléculaires chez les adultes atteints de gliomes de bas grade ont entraîné un changement dans la classification histopathologique de tous les gliomes.

Afin de garantir l'uniformité des critères de diagnostic dans le monde entier, la classification et le classement de ces tumeurs cérébrales ont été basés sur les conclusions d'un groupe de travail international publiés par l'Organisation Mondiale de la Santé (OMS). La première édition de la classification OMS des tumeurs du système nerveux a été publiée en 1979 (Zulch KJ et al. 1979), suivie d'une deuxième édition en 1993 qui a introduit l'analyse immunohistochimique dans le diagnostic (Kleihues P et al. 1993). En 2000, la troisième édition a introduit des dossiers génétiques comme aide supplémentaire pour définir les tumeurs du cerveau (International Agency for Research on Cancer. 2000), tandis que la quatrième édition en 2007 a ajouté plusieurs nouvelles entités histopathologiques (Louis DN et al. 2007).

En 2016, l'OMS a publié une mise à jour de la quatrième édition de la classification des tumeurs du SNC qui représente le consensus de 117 contributeurs et qui pour la première fois, utilise des paramètres moléculaires en plus de l'histologie traditionnelle pour diagnostiquer les tumeurs du SNC (Louis DN et al. 2016 ; International Agency for Research on Cancer. 2016). L'évaluation diagnostique standard des gliomes de bas grade doit maintenant inclure une évaluation moléculaire des mutations de l'isocitrate déshydrogénase (IDH) et la co-délétion des bras chromosomiques 1p et 19q pour être considérée comme complète. (International Agency for Research on Cancer. 2016 ; Reuss DE et al. 2015 ; Cancer Genome Atlas Research Network et al. 2015).

En fait, ces marqueurs de diagnostic moléculaire remettent en question les hypothèses antérieures concernant la définition des gliomes de bas grade, et les données disponibles soutiennent maintenant l'idée que la classification primaire devrait être basée principalement sur l'état mutationnel plutôt que sur le grade. En outre, les observations en laboratoire et les résultats des essais cliniques ont remis en question les paradigmes de traitement traditionnels. Deux essais de phase III ont indiqué que, bien qu'un traitement initial par chimiothérapie ou radiothérapie puisse produire des résultats globalement similaires car les résultats varient en fonction du diagnostic moléculaire (Baumert BG et al. 2016 ; Eckel-Passow JE et al. 2015). En outre, la chimiothérapie offre un avantage de survie substantiel par rapport à la radiothérapie seule (Buckner JC et al. 2016). En raison de ces changements dans l'évaluation diagnostique et

les approches thérapeutiques, la gestion des patients atteints de gliome de bas grade devrait être prise en compte, et les nouveaux résultats devraient être intégrés dans le paradigme global de la gestion des patients.

Les gliomes de bas grade (LGG) sont un groupe de tumeurs neuroépithéliales hétérogènes provenant des cellules gliales de soutien du système nerveux central. L'Organisation Mondiale de la Santé (OMS) a traditionnellement classé les gliomes en quatre grades, en fonction de caractéristiques histopathologiques telles que l'atypie, l'anaplasie, l'activité mitotique, la prolifération microvasculaire et la présence de nécrose (Kleihues P et al. 2002).

Classiquement, les gliomes se composaient de tumeurs de grade I de l'OMS, dépourvues de toutes ces caractéristiques, et de tumeurs de grade II de l'OMS, présentant seulement une atypie. (Louis DN et al. 2007). Cependant, ces deux sous-catégories se sont avérées très différentes sur le plan clinique et moléculaire (Rueda E et al. 2011). Les gliomes de grade I de l'OMS sont des tumeurs réellement bénignes qui peuvent être soignées par une ablation chirurgicale. En revanche, les gliomes de grade II de l'OMS sont des lésions intracérébrales diffuses et infiltrantes rarement guérissables (Duffau H et al. 2015). Les gliomes de grade II de l'OMS comprennent l'astrocytome et l'oligodendrogliome, ce sont des affections qui se développent lentement et qui présentent un potentiel de transformation maligne et qui évoluent presque invariablement vers un gliome de haut grade (Duffau H et al. 2015 ; Wessels PH et al. 2003). En 2014, la Société internationale de neuropathologie s'est réunie aux Pays-Bas et elle a établi des directives pour l'intégration de paramètres moléculaires dans la classification des entités tumorales du cerveau (Louis DN et al. 2014). Cette classification combinée de diagnostique histopathologique et génotypique est le plus grand changement introduit entre la mise à jour de la classification de l'OMS de 2007 et celle de 2016. Dans la nouvelle classification, tous les gliomes infiltrant différemment sont regroupés indépendamment de la cellule d'origine, astrocytes ou oligodendrocytes, car ils partagent tous des mutations génétiques et des facteurs de pronostique spécifiques (Figure 1) (Louis DN et al. 2016).

Nous nous sommes concentré sur les gliomes de bas grade, qui représentent 15 % de tous les gliomes. Ces tumeurs sont également connues sous le nom de gliomes diffus en raison de leur nature infiltrante et invasive. Le gliome diffus a été défini par Duffau comme "...une maladie progressive, invasive et chronique du système nerveux central" (Duffau H et al. 2015), ces tumeurs semblent croître lentement mais de façon continue dans le temps le long des fibres de la substance blanche des hémisphères. Une fois que la différenciation maligne se produit, elles se comportent comme des tumeurs de grade III ou IV et une détérioration neurologique rapide est attendue avant le décès (Wessels PH et al. 2003).

Le terme "non spécifié autrement" (NOS) a été introduit dans la classification de l'OMS 2016 et il est réservé aux gliomes de bas grade pour lesquels la caractérisation moléculaire, telle que le statut mutationnel de l'IDH et/ou le statut de co-délétion 1p/19q, fait défaut parce que soit l'analyse n'a pas pu être effectuée, soit les résultats n'étaient pas concluants (International Agency for Research on Cancer, 2016), en raison du risque de mauvaise classification d'entités cliniques et biologiques distinctes et du risque de mauvaise reproductibilité et de la forte variabilité inter observateurs pour certains diagnostics, en particulier parmi les gliomes diffus tels que l'oligoastrocytome. Toutefois, les tests moléculaires peuvent ne pas être disponibles dans de grandes parties du monde et peuvent parfois échouer, pour diverses raisons (par exemple, en raison d'un oligodendrogliome calcifiés qui nécessitent une décalcification à long terme), et une approche pragmatique est nécessaire pour ces situations. Les tumeurs astrocytaires et oligodendrogiales diffuses de grade II de l'OMS partagent non seulement un modèle de croissance à infiltration diffuse caractéristique et une tendance intrinsèque à la progression maligne, mais aussi une mutation génétique sous-jacente qui affecte les gènes IDH1 et IDH2.

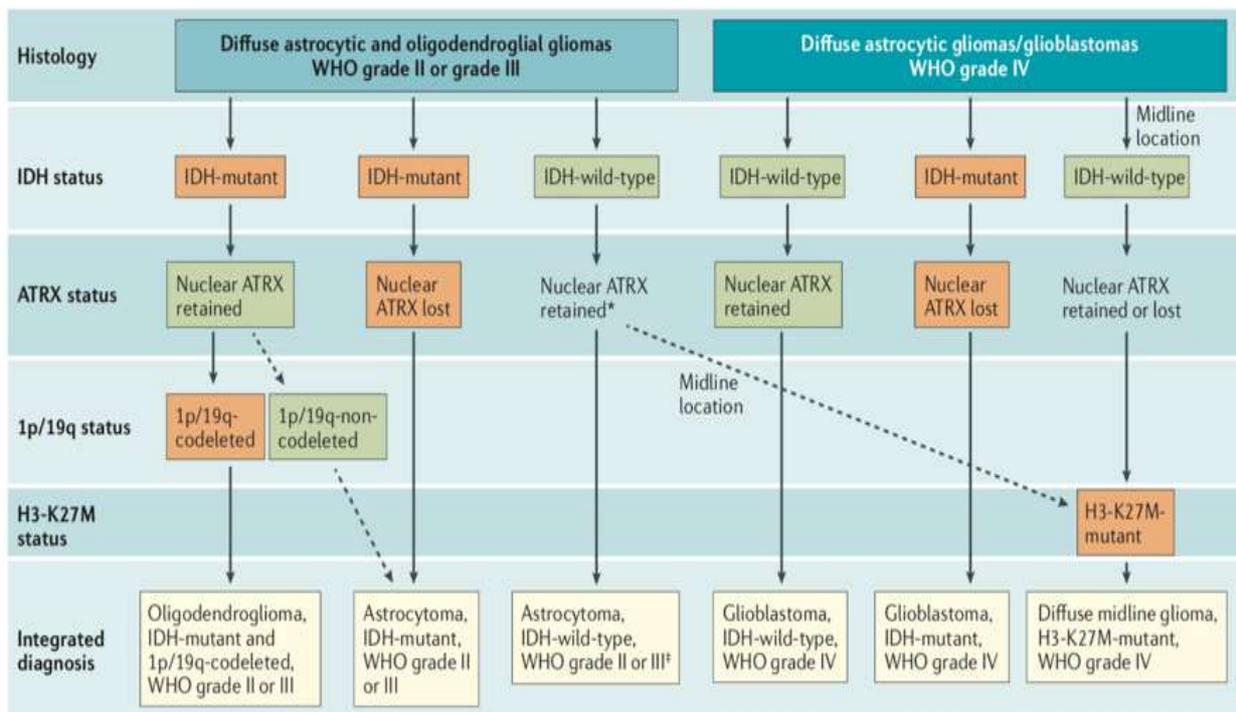


Figure 1: La classification histologique et moléculaire intégrée des gliomes selon l'OMS 2016

Par définition, la classification "Astrocytome diffus, mutant IDH" (grade II de l'OMS) fait référence à un astrocytome à infiltration diffuse avec une mutation dans le gène IDH1 ou IDH2. Ce type de tumeur est typiquement composé de cellules présentant un pléomorphisme modéré, et présente une différenciation astrocytaire avancée et une croissance lente. Le diagnostic

d'astrocytome diffus est confirmé par la présence des mutations TP53 et ATRX. Histologiquement, une tumeur qui ressemble morphologiquement à un oligodendrogliome est compatible avec un diagnostic d'astrocytome diffus mutant IDH, à condition que la codélétion 1p/19q soit absente.

L'astrocytome diffus, de type IDH sauvage, est un astrocytome à infiltration diffuse sans mutations dans les gènes IDH. La classification 2016 de l'OMS des tumeurs du SNC note que ce diagnostic est rare et il est considéré comme une entité provisionnelle. Ce type de tumeur comprend probablement une variété d'entités ; en effet chez les adultes, de nombreuses tumeurs ayant cet aspect histo-pathologique peuvent être reclassées en d'autres entités tumorales après des analyses génétiques supplémentaires (Reuss DE. et al. 2015 ; Cancer Genome Atlas Research Network et al. Comprehensive. 2015).

L'astrocytome diffus, "NOS" est une tumeur présentant les caractéristiques histo-pathologiques d'un astrocytome diffus dans lequel le statut de mutation IDH n'a pas été entièrement évalué. Les gliomes diffus de bas grade doivent être distingués des gliomes de ligne médiane à double fusion avec une mutation Lys27Met dans l'histone H3 (qui présentent des caractéristiques histologiques conformes au grade II de l'OMS et comprennent environ 10 % de gliomes intrinsèques diffus) car le grade histologique ne permet pas de prédire l'issue des tumeurs très agressives avec cette mutation. Cette entité peut être identifiée par la combinaison d'une tumeur située à la ligne médiane et de la détection immunohistochimique du mutant de l'histone H3 Lys27Met avec un anticorps spécifique de la mutation.

L'astrocytome gémistocytaire, mutant IDH" (grade II de l'OMS) reste le seul sous-type histopathologique d'astrocytome et représente environ 10 % des astrocytomes diffus de grade II de l'OMS. Au moins 20 % des cellules néoplasiques de ce type de tumeur contiennent un cytoplasme éosinophile intracellulaire abondant et des noyaux excentriques. Des rapports précédents ont indiqué que les astrocytomes gémistocytaires subissent une progression maligne plus fréquemment que les autres astrocytomes diffus, mais il n'existe pas de données dans lesquelles le pronostic a été évalué dans le contexte d'une mutation IDH.

L'entité classée par l'OMS comme "oligodendrogliome, mutant IDH et co-délétion 1p/19q" est un gliome à infiltration diffuse et à croissance lente présentant une mutation IDH1 ou IDH2 et une co-délétion des bras chromosomiques 1p et 19q par une translocation t(1;19)(p10;q10) (Jenkins RB et al. 2006).

Histopathologiquement, la tumeur est typiquement composée de cellules à noyau rond et uniforme et d'un cytoplasme clair gonflé (également appelé halo périnucléaire), une

morphologie qui ressemble à celle des oligodendrocytes ; cependant, une morphologie astrocytaire est compatible avec le même diagnostic, à condition qu'une mutation IDH1 ou IDH2 et une co-délétion des bras chromosomiques 1p et 19q soient présentes. Des microcalcifications et un réseau vasculaire ramifié caractéristique sont typiques de cette classe de tumeurs.

Le terme "oligodendrogliome, NOS" est réservé aux tumeurs qui présentent une histologie oligodendroliale classique, mais pour lesquelles les tests moléculaires de détection d'une mutation IDH et/ou d'une co-délétion 1p/19q n'ont pas pu être effectués ou n'ont pas été concluants. L'oligodendrogliome NOS doit être distingué de l'oligodendrogliome dépourvu d'IDH ou une co-délétion 1p/19q, à savoir un oligodendrogliome de type pédiatrique, qui représente un petit sous-ensemble d'oligodendrogliomes histologiquement classiques qui surviennent chez les enfants et les adolescents et ne présentent pas les altérations moléculaires typiques après des tests moléculaires appropriés.

Dans la classification 2016 de l'OMS, le terme "oligo-astrocytome, NOS" est exclusivement réservé aux tumeurs pour lesquelles les analyses moléculaires n'ont pas pu être réalisées ou n'ont pas été concluantes, et pour lesquelles la présence d'un mélange conique de deux types de cellules présentant des caractéristiques oligodendrocytaires ou astrocytaires exclut toute classification histologique en tant qu'astrocytome diffus NOS ou oligodendrogliome NOS. La grande majorité des tumeurs précédemment diagnostiquées comme oligoastrocytomes ou gliomes mixtes du point de vue moléculaire ont un profil génétique correspondant soit à un "astrocytome diffus, mutant IDH" (en combinaison avec des mutations dans TP53 et ATRX), soit à un "oligodendrogliome, mutant IDH et co-déleté 1p/19q" (en combinaison avec une mutation dans le promoteur TERT)(Cancer Genome Atlas Research Network et al. 2015 ; Eckel-Passow JE. et al. 2015).

Il existe de rares exceptions dans lesquelles un gliome diffus présente un mélange de populations de cellules tumorales astrocytaires et oligodendrocytaires avec des preuves moléculaires d'un double génotype. Dans ces cas, un sous-ensemble de cellules tumorales présente un génotype astrocytaire, une mutation IDH combinée à une mutation TP53 et/ou une accumulation nucléaire de p53, une perte d'expression nucléaire ATRX et une absence de co-délétion 1p/19q, alors que d'autres cellules tumorales ont un génotype oligodendrogliol dans lequel une mutation IDH est combinée à une co-délétion 1p/19q et les cellules conservent l'expression nucléaire ATRX, ne présentent pas de mutation TP53 et ne présentent pas d'accumulation nucléaire p53.

Bien que le terme "gliome de bas grade" (lorsqu'il désigne un gliome diffus et non des lésions de grade I ou un épépendymome de l'OMS) ait traditionnellement fait référence à des lésions de grade II, le terme "gliome de bas grade" englobe les lésions de grade II et de grade III. Les lésions de grade III présentent les caractéristiques morphologiques de l'anaplasie (qui comprennent principalement une activité mitotique accrue, mais peuvent inclure des modifications microvasculaires et/ou une nécrose dans l'oligodendrogliome anaplasique), et ils présentent souvent des altérations du nombre de copies chromosomiques plus fréquentes que les lésions de grade II. Le pathologiste peut discerner des lésions typiques de grade II et de grade III, mais de nombreux patients ont des tumeurs dont les caractéristiques morphologiques couvrent le spectre des gliomes de grade II et de grade III, même lorsque ces tumeurs sont échantillonnées en plusieurs endroits. Sur le plan du pronostic, l'importance de la différenciation entre le grade II et le grade III est incertaine dans les gliomes mutants IDH ; en effet, les combinaisons de mutations dans les gènes IDH, ATRX, le promoteur TERT, et la co-délétion 1p/19q prédisent plus efficacement le comportement et l'issue que les tumeurs de grade (Cancer Genome Atlas Research Network et al. 2015 ; Jenkins RB et al. 2006; Eckel-Passow JE. et al. 2015). De manière surprenante, les mutations du promoteur TERT se produisent dans les gliomes présentant le meilleur pronostic (comme l'oligodendrogliome avec mutation IDH et co-délétion 1p/19q) et ceux présentant le pire pronostic (comme le glioblastome, type IDH-wild-type). La présence d'une mutation du promoteur TERT indique un mauvais pronostic dans les astrocytomes de type IDH-wild, les astrocytomes anaplastiques et les glioblastomes (Eckel-Passow JE et al. 2015). La pertinence du grade peut varier considérablement pour chacun des groupes moléculaires, et constitue un domaine d'étude important pour l'avenir.

Actuellement, la survie moyenne des gliomes de bas grade est généralement inférieure à 10 ans à partir du diagnostic (Rueda E et al. 2011 ; Duffau H et al. 2015 ; Wessels PH et al. 2003).

L'impact sur la survie sans progression (SSP), la survie globale (SG) et la qualité de vie (QV) des différentes options de traitement des gliomes est controversé et fait l'objet de recherches. Cette affection a été traditionnellement considérée comme une maladie chronique et d'une certaine manière bénigne, avec peu ou pas d'impact sur la qualité de vie des patients, si ce n'est la nécessité de contrôler les crises. Une politique d'attente a été préconisée de manière classique, en grande partie parce que les gliomes affectent généralement de jeunes adultes sans défauts neurologiques majeurs, présentant des tumeurs qui ont tendance à se développer dans des zones éloquentes du cerveau où même la biopsie était considérée comme dangereuse (Duffau H et al. 2015).

## 2- Caractéristiques principales et histoire naturelle

Les patients atteints de gliome de bas grade présentent généralement trois stades cliniques : pré-symptomatique, une période de durée inconnue au cours de laquelle la tumeur s'infiltré lentement dans le cerveau mais le patient reste largement asymptomatique ; symptomatique, une période de 7 ans en moyenne pendant laquelle le patient présente généralement une épilepsie et peut-être des troubles cognitifs subtils, généralement compatibles avec une vie sociale et professionnelle presque normale ; et enfin une transformation maligne, une période d'environ 2 à 3 ans pendant laquelle la tumeur accélère sa croissance et se dédouble en un gliome de grade III ou IV selon l'OMS (Rueda E et al. 2011 ; Duffau H et al. 2015 ; Wessels PH et al. 2003). Au cours de cette phase, les patients voient leurs fonctions neurologiques se détériorer (l'épilepsie s'aggrave généralement et de nouveaux défauts neurologiques s'ensuivent) jusqu'à ce que la mort survienne, malgré le traitement. Bien que la survie soit influencée par certains facteurs de pronostique, la durée moyenne de survie à partir du diagnostic est de 5-6 ans, allant de 3 à 10 ans (Duffau H et al. 2015). Selon ces chiffres, cette affection ne peut pas être considérée comme une lésion bénigne. Historiquement, les patients porteurs de gliome se voyaient souvent proposer une observation de la lésion, ou simplement une biopsie pour exclure une tumeur maligne, ainsi que des médicaments pour l'épilepsie, qui est présente dans au moins 90% des cas (Duffau H et al. 2015 ; Chang EF et al. 2008).

Les gliomes s'infiltrent de façon diffuse dans la structure normale du cerveau. Comme le rythme de progression des gliomes de bas grade est le plus souvent de l'ordre de plusieurs années, la structure physique brute et la structure du réseau neuronal du cerveau sont souvent bien préservées jusqu'à ce que la progression entraîne des dommages structurels suffisants pour modifier le fonctionnement du cerveau. Même lorsque ce processus se produit, la lenteur de la progression permet une réorganisation compensatoire du réseau neuronal, ce qui retarde la manifestation clinique de la maladie (Duffau H et al. 2006 ; Pascual-Leone A et al. 2005).

Ces manifestations dépendent de la localisation de la maladie dans le cerveau et sont soit insidieuses, soit paroxysmique. L'apparition insidieuse peut se produire parce que la progression pathologique est lente et que les réseaux compensatoires évoluent, quelle que soit la localisation de la tumeur, de sorte que les premiers déficits cliniques peuvent être imperceptibles et difficiles à distinguer des variations normales du fonctionnement du cerveau, comme le fait d'égarer un objet dans la maison ou de mal prononcer un mot. Cette imperceptibilité est particulièrement vraie lorsque les changements fonctionnels sont de nature cognitive ou sensorielle. Cependant, la perturbation progressive des réseaux neuronaux affecte

parfois la stabilité électrique d'un groupe ou le fonctionnement en réseau des neurones, ce qui peut entraîner une forme de crise d'épilepsie.

La manifestation d'une crise dépend de la localisation de son origine dans le cerveau, ainsi que de l'ampleur et de la rapidité de la propagation de décharges électriques anormales à travers les systèmes interconnectés du cerveau. Les facteurs qui régissent ou limitent la distribution des décharges de la crise dans le cerveau ne sont pas bien compris (Martinet LE et al. 2015).

Des décharges épileptiques limitées peuvent être associées à des symptômes très soudains mais brefs, le plus souvent de nature cognitive ou sensorielle, qui peuvent être difficiles à différencier de la variation normale des fonctions cérébrales. Par définition, de tels événements sont de nature subjective et ne peuvent donc pas être observés par d'autres personnes, et la description que fait l'individu de ses perceptions est souvent mal comprise ou mal interprétée. On peut citer comme exemple la perte temporaire d'attention ou la perception soudaine d'une odeur ou d'une émotion en l'absence d'un stimulus externe. Des crises plus fréquentes peuvent être associées à des expériences subjectives plus intenses, mais elles sont souvent difficiles à expliquer et à interpréter. Certaines crises sont associées à une perte de conscience transitoire. Souvent, la personne qui vit un tel événement n'est pas consciente de cette perte et ne sait que vaguement qu'il se passe quelque chose. Si ces événements impliquent des régions du cerveau qui sont liées aux fonctions du langage, une certaine altération de la compréhension ou de l'expression de la parole se produit. Même dans cette situation, les témoins peuvent rester inconscients de ce qui se passe réellement jusqu'à ce que ces événements aient atteint une intensité ou une durée suffisante, ou continuer à impliquer des systèmes moteurs ou des réseaux cérébraux globaux, auquel cas une crise avec manifestations motrices ou une crise généralisée se produit.

De nouvelles preuves suggèrent que les patients atteints d'un gliome de grade II muté par l>IDH sont plus susceptibles d'avoir des crises d'épilepsie comme élément du diagnostic que les patients atteints de tumeurs IDH de type sauvage. La mutation du gène IDH1 entraîne une augmentation de la production de D-2-hydroxyglutarate, un analogue du glutamate. Le glutamate est un neurotransmetteur excitateur, de sorte qu'une augmentation de la signalisation des récepteurs du glutamate déclenche des décharges neuronales aberrantes, qui entraînent des crises. L'utilisation individuelle ou combinée de la résection chirurgicale, de la radiothérapie et de la chimiothérapie réduit la charge des cellules tumorales chez les patients, ce qui entraîne une diminution de la fréquence des crises (Yang Y et al. 2016 ; Huberfeld G & Vecht CJ. 2016 ; Koekkoek JA et al. 2015).

En outre, l'ampleur de la résection chirurgicale est associée à la maîtrise des crises : les patients qui subissent une résection tumorale totale brute, avec ou sans résection du cortex ou de

l'hippocampe adjacent qui semble normale à l'IRM, ont une fréquence de crises plus faible que ceux qui subissent une résection tumorale subtotala (Englot D J et al. 2012).

Lorsque les symptômes se présentent après que l'infiltration dans le cerveau a atteint une grande ampleur, la tumeur peut avoir un effet de masse sur les structures cérébrales voisines, ce qui entraîne des modifications fonctionnelles plus manifestes des fonctions cognitives, motrices et sensorielles qui sont parfois associées à des céphalées focales ou généralisées progressives. Dans les gliomes de bas grade, ces changements évoluent le plus souvent lentement sur des semaines ou des mois. La céphalée peut généralement être différenciée de toute céphalée syndromique antérieure (comme les céphalées de tension, les migraines ou les céphalées en grappes) par la nature de la progression et les schémas de variation des symptômes dans le temps. Dans la maladie avancée à effet de masse, les facteurs qui augmentent la pression intracrânienne (comme la toux ou l'éternuement) peuvent exacerber les céphalées.

Dans de rares cas, les tumeurs qui présentent un effet de masse régional peuvent se manifester par des symptômes neurologiques paroxystiques qui ne sont pas de nature épileptique, mais qui se traduisent par une perte transitoire de fonction neurologique semblable à un accident vasculaire cérébral. Ces événements ne sont pas des séquelles d'une altération du flux sanguin cérébral, comme dans les phénomènes ischémiques typiques, et le mécanisme physiopathologique est largement inconnu. Cependant, ces événements représentent souvent une source de confusion clinique dans l'évaluation initiale des patients atteints de tumeurs cérébrales.

La lenteur de la présentation de la maladie et le développement d'une réorganisation compensatoire du réseau neuronal ont une conséquence importante pour la prise en charge initiale des gliomes de bas grade : le volume potentiel d'une résection chirurgicale maximale sûre dépasse souvent de loin ce que l'on pourrait estimer d'après l'anatomie et la physiologie normales du cerveau (Duffau H. et al. 2003).

Des techniques telles que l'IRM fonctionnelle, F-DOPA PET, la craniotomie en état de veille, complétées par l'IRM peropératoire, permettent d'optimiser l'équilibre entre une résection maximale sûre et le risque de perte de la fonction neurologique (Pascual-Leone A et al. 2005).

### 3 - Tests de diagnostic moléculaire

Les gliomes qui présentent une immunohistochimie négative pour la mutation commune IDH1 Arg132H nécessitent une évaluation complète du statut mutationnel de l'IDH, avec une analyse de la séquence du codon 132 de l'IDH1 et du codon 172 de l'IDH2. L'évaluation du statut de co-délétion 1p/19q peut alors être effectuée selon les besoins. Diverses approches peuvent être

suivies pour évaluer l>IDH et le statut 1p/19q. L'algorithme de diagnostic pratique commence par l'évaluation du statut IDH1 Arg132His et ATRX par immunohistochimie. Les gliomes diffus qui présentent une expression mutante de l>IDH1 et une absence d'expression de l'ATRX sont classés comme astrocytomes diffus et aucun test supplémentaire n'est effectué, car < 3 % de ces tumeurs présentent une co-délétion 1p/19q (Eckel-Passow, J. E. et al. 2015 ; Leeper, H. E. et al. 2015).

À l'inverse, dans les gliomes diffus qui présentent une expression mutante d>IDH1 et conservent l'expression d'ATRX, le statut 1p/19q est évalué. Il est important de noter que les astrocytomes présentant des mutations dans les gènes IDH, TP53 et ATRX présentent fréquemment de petites délétions des bras 1p et 19q. Beaucoup de ces petites délétions sont détectées par les sondes généralement utilisées dans l'hybridation in situ par fluorescence clinique pour tester les co-délétions 1p/19q du bras entier, et des résultats faussement positifs peuvent se produire (Eckel-Passow JE et al. 2015).

Actuellement, l'analyse moléculaire donne aux médecins l'information dont ils ont besoin pour repérer les gènes qui se sont transformés (gènes mutés). En connaissant ces mutations, ils peuvent choisir des options thérapeutiques adaptées au profil individuel des patients. Il est maintenant possible de reconnaître des combinaisons particulières de biomarqueurs caractéristiques de la tumeur et, à partir de cette information, de diagnostiquer certains types de tumeurs, d'établir le pronostic (le destin probable de la tumeur) et de traiter ces cancers. Le type et le nombre de mutations peuvent nous éclairer sur la façon dont le patient répondra à des médicaments donnés. En fin, l'objectif de l'analyse moléculaire est de concevoir des traitements à la carte, hautement ciblés et efficaces qui peuvent améliorer le devenir des patients, par l'identification des biomarqueurs qui sont des molécules qui révèlent des signes et des processus normaux ou anormaux dans l'organisme. Ils peuvent être aussi responsable à une réponse, à l'absence de réponse ou à la résistance à certaines options thérapeutiques. Avec cette information, on peut élaborer des traitements ciblés, qui seront plus efficaces pour un profil tumoral particulier.

## **II - Recommandations actuelles en matière de traitement**

### **1 - La chirurgie**

L'utilisation de la chirurgie pour établir un diagnostic histologique du gliome de bas grade avant de commencer le traitement est une partie acceptée de la gestion et elle est essentielle pour établir le génotype moléculaire (Ragel BT et al. 2015).

Toutefois, le rôle de la chirurgie pour le contrôle des symptômes et la cytoréduction reste controversé. La plupart des cliniciens seraient d'accord pour dire qu'il faut procéder à une résection tumorale maximale sûre, en particulier dans les cas où le volume de la tumeur est associé à un effet de masse important. Cependant, cette affirmation peut être interprétée de manière variable en ce qui concerne le rôle de la cytoréduction pour la survie. Un guide de pratique de 2015 a révélé que 26 des 32 études publiées soutenaient une association entre l'augmentation de la résection et la survie globale, mais la plupart des études se limitaient aux preuves de classe III (Aghi MK et al. 2015).

Les preuves à l'appui d'une association entre l'augmentation de l'étendue de la résection et la survie sans progression étaient un peu plus fortes (preuves de classe II). En particulier, seule une minorité d'études ont utilisé une véritable évaluation volumétrique assistée par ordinateur de l'étendue de la résection et de la tumeur résiduelle (Castellano A et al. 2012 ; Sanai N & Berger MS. 2010 ; Smith JS et al. 2008).

Ces études ont uniformément démontré une association entre une plus grande étendue de la résection et une survie globale plus longue chez les patients atteints de gliome.

Ces études non randomisées n'abordent pas le biais de sélection : les patients pour lesquels une résection chirurgicale agressive est tentée et/ou réalisée pourraient représenter une population différente de celle des patients qui ne subissent qu'une biopsie. Les résultats ont suggéré que les tumeurs qui se prêtent à une large résection ont tendance à impliquer les lobes frontaux et temporaux, dans lesquels les tumeurs mutantes à l'IDH sont plus fréquentes que les autres types de tumeurs. En outre, l'association entre une plus grande étendue de la résection et une plus grande survie globale des patients pourrait refléter une fréquence exceptionnellement élevée de mutations IDH favorables dans les premières études, qui ne comprenaient pas de tests pour la mutation IDH (Kizilbash SH et al. 2014). Toutefois, à l'opposé de cet argument, une association volumétrique a été démontrée entre l'étendue de la résection et la survie globale chez les patients atteints de gliomes de bas grade de la région insulaire, une population de tumeurs qui sont presque universellement mutantes à l'IDH et qui étaient auparavant considérées comme largement inopérables (Koekkoek JA et al. 2015).

De plus, il a été démontré que l'étendue de la résection d'une tumeur augmentant le contraste est en corrélation avec la survie globale, quel que soit l'état de mutation IDH. Il est intéressant de noter que l'étendue de la résection des tumeurs ne renforçant pas le contraste était également fortement corrélée à la survie globale, mais uniquement pour les patients atteints de tumeurs mutantes IDH (Beiko J et al. 2014).

D'autres études ont permis de mieux caractériser les gliomes diffus sur la base d'une analyse de l'expression des gènes (par exemple, la TP53 et ATRX) ou de la présence ou de l'absence de mutation IDH. On ne sait pas si la valeur pronostic de l'étendue de la résection varie avec ces différents sous-ensembles moléculaires qui font actuellement l'objet de recherches approfondies (Shigeo Ohba et al. 2020).

Dans l'ensemble, la majorité des données actuelles soutiennent une association entre une plus grande étendue de résection et une survie globale accrue pour les patients atteints de gliome de bas grade. Les guides de pratique publiés au cours des dernières années indiquent que maximiser l'étendue de la résection est une option recommandée (recommandation de niveau III). Une étude plus approfondie est nécessaire sur l'étendue volumétrique de la résection et la contribution des groupes moléculaires et des facteurs de traitement prédictifs à cette procédure.

## 2 – La radiothérapie

Les progrès de la radiothérapie moderne ont considérablement amélioré le rapport risque-bénéfice de ce traitement pour les patients atteints de gliome de bas grade. Ces patients connaissent souvent une survie prolongée, de sorte que le maintien de la qualité de vie et des fonctions cognitives aussi longtemps que possible est un objectif important de la thérapie. La radiothérapie est associée à un déclin des fonctions cognitives et à une réduction de la qualité de vie chez certains patients atteints d'une tumeur cérébrale, principalement les patients qui reçoivent une radiothérapie du cerveau entier pour une maladie métastatique. Cependant, les patients atteints d'un gliome de bas grade qui reçoivent une radiothérapie focale (même avec des techniques quelque peu historiques, comme la radiothérapie conventionnelle conforme) présentent des changements cognitifs très légers qui ne sont pas détectables par une évaluation neuropsychiatrique formelle avant plus de 12 ans après le traitement (Laack NN et al. 2005 ; Douw L et al. 2009).

Grâce à des techniques encore plus précises, comme la radiothérapie à modulation d'intensité (IMRT) et la protonthérapie, ainsi qu'à des améliorations de l'imagerie et de l'administration du traitement, les effets à long terme de la radiothérapie devraient encore s'atténuer.

Les améliorations apportées aux logiciels d'imagerie et d'enregistrement des images, ainsi qu'aux systèmes de positionnement des patients et au guidage quotidien des images, permettent de réduire encore les volumes cibles de la radiothérapie par rapport à ceux que permettait la technologie disponible auparavant. Les technologies plus récentes réduisent le volume du cerveau normal qui est irradié et diminuent le risque de complications à long terme. Une IRM postopératoire est fusionnée au scanner de planification du traitement et utilisée pour la

localisation de la tumeur ; la tumeur est mieux visualisée sur une récupération par inversion d'atténuation des fluides ou une IRM pondérée en T2. Le volume cible doit également inclure la cavité de résection chirurgicale et toute anomalie détectée lors de la récupération par inversion de l'atténuation des fluides ou par IRM en pondération T2 au-delà des bords de la cavité. Une zone d'expansion de 1 cm, limitée par des barrières anatomiques à la propagation, telles que le tentorium ou la dure-mère, doit être ajoutée pour tenir compte de l'infiltration subclinique de la tumeur. Avec l'utilisation d'un guidage d'image quotidien et de tables robotisées pour la position du patient, la marge d'incertitude pour la mise en place peut être limitée à 3 mm. Les essais d'escalade de dose de radiothérapie n'ont montré aucune différence dans les résultats pour des doses de 45,0-64,8 Gy (REFS 35,36). La dose utilisée dans l'essai RTOG (Radiation Therapy Oncology Group) 9802 était de 54 Gy en fractions de 1,8 Gy par jour (Buckner JC et al. 2016). Les études qui ont évalué la radiothérapie avec la chimiothérapie concomitante, comme le témozolomide, ont utilisé une dose de radiothérapie réduite de 50,4 Gy pour tenir compte d'une possible augmentation de la radiosensibilité résultant de la chimiothérapie. Une taille de fraction de <2 Gy par jour a été utilisée pour minimiser le risque d'effets cognitifs à long terme (Fisher BJ et al. 2013).

La radiochirurgie est généralement un traitement à forte dose, de 1 à 5 fractions, qui est prescrit pour une maladie grave et qui a une chute brutale de la dose de rayonnement pour minimiser le risque des fortes doses administrées, qui peuvent entraîner des lésions au cerveau adjacent. Cette technique a un rôle très limité dans la fixation des gliomes de bas grade, en raison de la nature infiltrante des gliomes diffus - dont l'invasion des cellules tumorales est indétectable à l'IRM - et des petits champs de traitement étroitement circonscrits associés à la radiochirurgie. La radiochirurgie a été utilisée dans un contexte de sauvetage et pour des séries limitées de patients atteints de maladies de faible volume. Cependant, aucune donnée ne soutient actuellement l'utilisation de la radiochirurgie avec une chimiothérapie concomitante ou adjuvante, de sorte que la radiochirurgie joue un rôle minime dans le traitement des tumeurs nouvellement diagnostiquées (Ryken TC et al. 2015).

Bien que la toxicité cognitive après une radiothérapie moderne et conforme soit légère, les doses élevées de radiothérapie pour les hippocampes sont de plus en plus appréciées pour être associées à une altération de la mémoire à court terme après radiothérapie (Gondi V et al. 2013). La radiothérapie à photons et à protons est une technique courante qui permet de réduire considérablement la dose de radiothérapie à l'hippocampe (Gondi V et al. 2013 ; Shih HA et al. 2015). En utilisant plusieurs faisceaux et un contrôle informatisé des feuilles de blocage à l'intérieur du faisceau, le volume de radiothérapie à forte dose se conforme étroitement aux

frontières irrégulières de la tumeur et minimise la dose de rayonnement aux structures critiques proches comme l'hippocampe (Kazda T et al. 2014). L'utilisation de photons dans cette technique signifie que l'exposition du cerveau à une radiothérapie à très faible dose augmente souvent par rapport aux techniques conventionnelles. La radiothérapie par protons, en particulier les nouvelles formes qui utilisent un faisceau de crayon ou une technologie de balayage, présente l'avantage de la conformité de l'IMRT sans l'augmentation des rayonnements à faible dose pour le cerveau, car les protons ont une dose de sortie limitée au-delà de la cible combinée aux photons. Un suivi à long terme est nécessaire pour confirmer si ces résultats se traduiront ou non par des améliorations de la fonction cognitive à long terme (Kazda T et al. 2014).

### 3 - Chimiothérapie et chimio radiothérapie

Les gliomes de bas grade sont connus pour être chimiosensibles. Le taux de réponse à la chimiothérapie avec la procarbazine, la lomustine et la vincristine (PVC) comme traitement initial varie de 52% à 100% (Buckner JC et al. 2003 ; Stege EM et al. 2005). Les taux de réponse au témozolomide chez les patients atteints d'oligodendrogliomes de bas grade sans ou avec radiothérapie préalable sont respectivement de 31-58% (Hoang-Xuan K et al. 2004 ; Brada M et al. 2003) et 47-61% (Pace A et al. 2003).

Sur la base de la chimiosensibilité avérée des gliomes de bas grade au PCV, les chercheurs ont mené un essai de phase III chez des adultes présentant un risque élevé de progression précoce de la maladie, définis comme des patients de plus de 40 ans ou ceux dont la résection totale de la tumeur est inférieure à l'estimation du neurochirurgien traitant, quel que soit leur âge. Les patients ont été répartis au hasard dans les 8 semaines suivant le diagnostic histologique initial pour recevoir soit une radiothérapie seule, soit une radiothérapie plus une chimiothérapie contre le PVC. La survie médiane sans progression était de 4 ans contre 10,4 ans pour la radiothérapie seule contre la chimiothérapie, avec une survie sans progression de 21% et 51%, respectivement. La survie globale était également supérieure dans le groupe chimio-radiation. La survie médiane était de 7,8 ans contre 13,3 ans pour la radiothérapie seule contre la chimio-radiation, et la survie à 10 ans était de 40 % contre 60 %, respectivement. La taille de l'effet du traitement était la plus importante chez les patients atteints d'oligodendrogliome et de tumeurs présentant la mutation IDH1 Arg132His, les rapports de risque en faveur de la chimioradiation étaient respectivement de 0,43 et 0,42 pour les deux types de tumeurs (Quinn JA et al. 2003 ; Buckner JC et al. 2016).

Dans cet essai, la survie sans progression et la survie globale ont été les plus longues rapportées à ce jour dans un essai de traitement du gliome de grade II, et les seules à avoir connu une survie prolongée; par conséquent, les résultats de cet essai sont devenus la norme pour la comparaison avec d'autres approches de traitement. De nombreux fournisseurs de traitements préconisent le témozolomide comme traitement initial après une intervention chirurgicale, afin de reporter le traitement par radiations et potentiellement de retarder le déclin cognitif associé induit par les radiations. Dans l'essai 22033 de l'Organisation Européenne pour la Recherche et le Traitement du Cancer (EORTC), les patients ont été répartis au hasard pour recevoir soit du témozolomide seul, soit une radiothérapie seule comme thérapie initiale (Baumert BG et al. 2016). Les patients ne recevaient pas de traitement à moins d'être âgés de  $\geq 40$  ans et d'avoir une progression tumorale radiologique, une aggravation des symptômes neurologiques ou des crises réfractaires. Bien qu'aucune différence n'ait été détectée dans la médiane globale de la survie sans progression dans l'ensemble de la cohorte, les patients atteints de tumeurs mutantes IDH et non co-déleté 1p/19q qui ont reçu du témozolomide ont connu un délai de progression plus court que ceux qui ont reçu une radiothérapie (36 mois contre 55 mois, respectivement). Un suivi supplémentaire continu fournira des données de survie pour cet essai.

Dans un petit échantillon de patients ( $n = 36$ ) traités dans le cadre de la CODEL, un essai randomisé de radiothérapie plus le PCV, radiation plus témozolomide ou témozolomide seul pour un oligodendrogliome anaplasique co-déleté 1p/19q, la randomisation a été interrompue par le comité de contrôle de la sécurité des données en raison de la progression de la tumeur chez une plus grande proportion de patients dans le bras témozolomide seul ( $n = 6/12$ ; 50%) que dans les bras de radiothérapie ( $n = 2/24$ ; 8%) (Jaeckle K et al. 2015).

Dans une série de cas de patients atteints d'oligodendrogliome de bas grade, des échantillons de tissus tumoraux appariés étaient disponibles à partir du diagnostic initial et de la croissance tumorale récurrente; les tumeurs étaient hypermutées lors de la récurrence dans une plus grande proportion de patients traités au témozolomide seul que de ceux qui n'ont reçu aucun traitement d'intervalle. Les altérations génétiques qui en résultent ressemblent à celles observées dans le glioblastome, notamment la perturbation de la voie de suppression des tumeurs de la protéine associée à la rétinoblastome et l'activation de la voie Akt-mTOR. À ce jour, aucun rapport clinique n'indique que l'état de tumeur hypermutée donne lieu à des tumeurs chimiorésistantes ou radorésistantes, ou que la survie après une maladie récurrente diffère selon qu'il s'agit de tumeurs hypermutées ou non. Cependant, les données font craindre que le temozolomide seul puisse accélérer la transformation de la tumeur en un phénotype plus agressif. En outre, aucune preuve n'a démontré que l'administration d'une chimiothérapie seule après une intervention

chirurgicale entraîne une survie sans progression ou supérieure à celle observée avec la chimio-radiation pour les patients de tout groupe de diagnostic histologique ou moléculaire. Bien que la radiothérapie puisse logiquement être posé pour préserver la fonction cognitive et maintenir la prolongation connue de la survie avec la chimioradiothérapie, aucune donnée d'essai clinique prospective ne confirme actuellement cette hypothèse (Johnson BE et al. 2014).

## **CHAPITRE 2**

*L'objectif de ce chapitre est de donner une vue globale sur l'application de la bio-informatique dans le domaine de la neuro-oncologie en citant les différentes approches omiques qui existent notamment la génomique, la transcriptomique, l'épigénomique, la protéomique et la métabolomique.*

## **I – La Bio-informatique**

### **1 – Définition**

La bio-informatique est un domaine de recherche interdisciplinaire qui relie l'informatique et les sciences biologiques. Il existe une variété de définitions dans la littérature; certains sont plus inclusifs que d'autres. Ici, nous adoptons la définition proposée par (Luscombe et al. 2001) pour définir la bio-informatique comme une union de la biologie et de l'informatique: la bio-informatique implique la technologie qui utilise des ordinateurs pour le stockage, la récupération, la manipulation et la distribution d'informations liées aux macromolécules biologiques telles que l'ADN, l'ARN et les protéines. L'accent est mis ici sur l'utilisation des ordinateurs, car la plupart des tâches de l'analyse des données génomiques sont hautement répétitives ou mathématiquement complexes. L'utilisation d'ordinateurs est absolument indispensable dans l'extraction de génomes pour la collecte d'informations et le renforcement des connaissances.

La bio-informatique diffère d'un domaine connexe connu sous le nom de biologie computationnelle. La bio-informatique se limite à l'analyse séquentielle, structurelle et fonctionnelle des gènes et des génomes et de leurs produits correspondants et elle est souvent considérée comme la biologie moléculaire computationnelle. Cependant, la biologie computationnelle englobe tous les domaines biologiques qui impliquent le calcul. Par exemple, la modélisation mathématique des écosystèmes, la dynamique de la population, l'application de la théorie des jeux dans les études comportementales et la construction phylogénétique, mais n'impliquent pas nécessairement des macromolécules biologiques. Outre cette distinction, il convient de noter qu'il existe d'autres points de vue sur la relation entre les deux termes. Par exemple, une version définit la bio-informatique comme le développement et l'application d'outils de calcul dans la gestion de toutes sortes de données biologiques, tandis que la biologie computationnelle se limite davantage au développement théorique d'algorithmes utilisés pour la bio-informatique. La confusion actuelle sur la définition peut refléter en partie la nature de ce nouveau domaine dynamique qui est en évolution rapide (Luo J. 2014).

### **2 – Objectifs de la bio-informatique**

Le but ultime de la bio-informatique est de mieux comprendre une cellule vivante et son fonctionnement au niveau moléculaire. En analysant la séquence moléculaire brute et les données structurelles, la recherche en bio-informatique peut générer de nouvelles perspectives et fournir une perspective «globale» de la cellule. La raison pour laquelle les fonctions d'une

cellule peuvent être mieux comprises en analysant les données de séquence est finalement parce que le flux d'informations génétiques est dicté par le «dogme central» de la biologie dans lequel l'ADN est transcrit en ARN, qui est traduit en protéines. Les fonctions cellulaires sont principalement réalisées par des protéines dont les capacités sont finalement déterminées par leurs séquences. Par conséquent, la résolution de problèmes fonctionnels à l'aide d'approches séquentielles et parfois structurelles s'est avérée être une entreprise fructueuse (Can T. 2014 ; Searls DB. 2010). La bio-informatique comprend deux sous-domaines: le développement d'outils et de bases de données informatiques et l'application de ces outils et bases de données pour générer des connaissances biologiques afin de mieux comprendre les systèmes vivants. Ces deux sous-groupes sont complémentaires l'un de l'autre. Le développement d'outils comprend l'utilisation de logiciels pour l'analyse séquentielle, structurelle et fonctionnelle, ainsi que la construction et la conservation de bases de données biologiques. Ces outils sont utilisés en génomique et en biologie moléculaire: l'analyse de séquence moléculaire, l'analyse structurale moléculaire et l'analyse fonctionnelle moléculaire. Les analyses de données biologiques génèrent souvent de nouveaux problèmes et défis qui, à leur tour, stimulent le développement de nouveaux et meilleurs outils informatiques. Les domaines de l'analyse de séquence comprennent l'alignement de séquences, la recherche dans la base de données de séquences, la recherche de gènes et de promoteurs, la reconstruction de relations évolutives et l'assemblage et la comparaison de génomes. Les analyses structurales comprennent l'analyse, la comparaison, la classification et la prédiction de la structure des protéines et des acides nucléiques. Les analyses fonctionnelles comprennent le profil d'expression génique, la prédiction de l'interaction protéine-protéine, la prédiction de la localisation subcellulaire des protéines, la reconstruction de la voie métabolique et la simulation.

Les aspects de l'analyse bio-informatique ne sont pas isolés mais interagissent souvent pour produire des résultats intégrés (Figure. 2). Par exemple, la prédiction de la structure des protéines dépend des données d'alignement de séquence; le regroupement des profils d'expression génique nécessite l'utilisation de méthodes de construction d'arbres phylogénétiques dérivées de l'analyse de séquence. La prédiction du promoteur basée sur la séquence est liée à l'analyse fonctionnelle des gènes co-exprimés. L'annotation des gènes implique un certain nombre d'activités, qui comprennent la distinction entre les séquences codantes et non codantes, l'identification des séquences protéiques traduites et la détermination

de la relation évolutive du gène avec d'autres gènes connus; la prédiction de ses fonctions cellulaires utilise des outils de ces groupes d'analyses (Chen R et al. 2015)

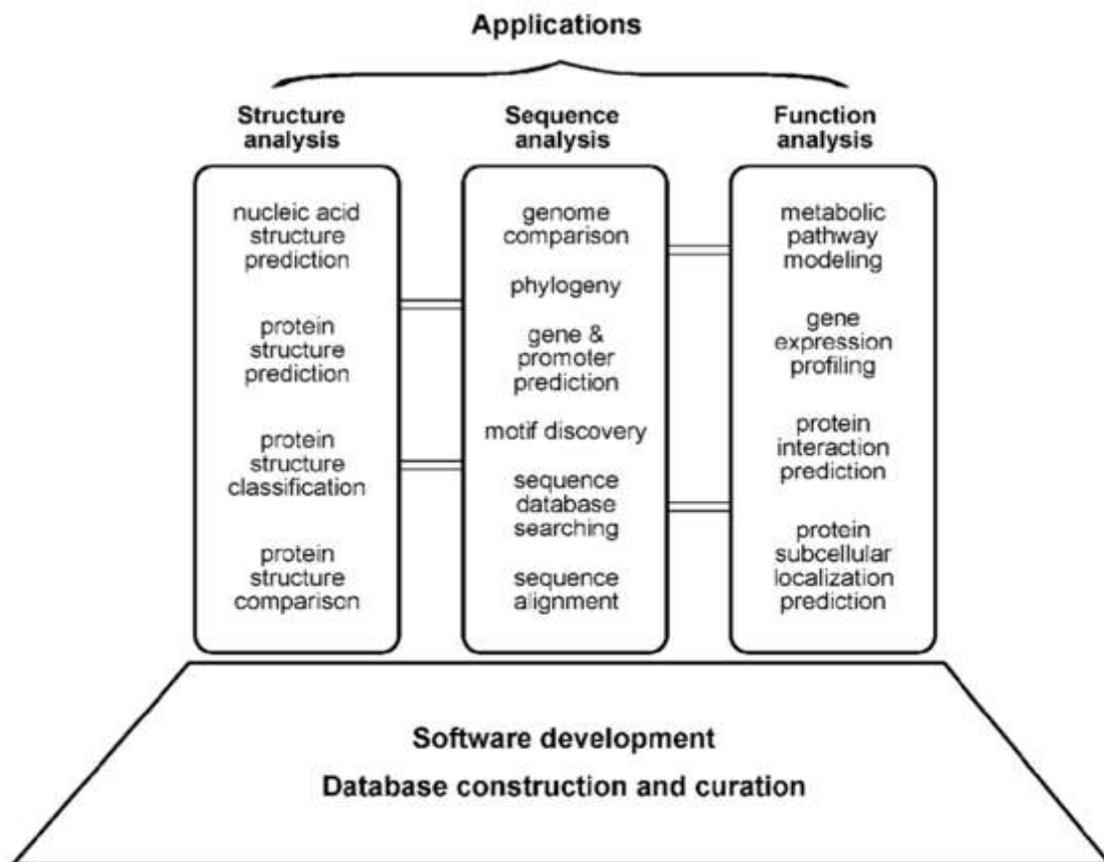


Figure 2 : Les différents sous-domaines de la bioinformatique. Les outils bioinformatique se répartissent en trois domaines : l'analyse des séquences, l'analyse des structures et l'analyse des fonctions

### 3 – L'impact de la bio-informatique

La bio-informatique est non seulement devenue essentielle pour la recherche fondamentale en biologie génomique et moléculaire, mais elle a un impact majeur sur de nombreux domaines de la biotechnologie, les sciences biomédicales et l'oncologie. Elle a des applications, par exemple, dans la conception de médicaments, l'analyse médico-légale de l'ADN et la biotechnologie agricole. Les études informatiques des interactions protéine-ligand fournissent une base rationnelle pour l'identification rapide de nouvelles pistes pour les drogues de synthèse. La connaissance des structures tridimensionnelles des protéines permet de concevoir des molécules capables de se lier au site récepteur d'une protéine cible avec une grande affinité et spécificité. Cette approche basée sur l'informatique, réduit considérablement le temps et le coût nécessaires pour développer des médicaments avec une puissance plus élevée, moins d'effets secondaires et moins de toxicité que l'utilisation de l'approche traditionnelle par essais et erreurs. En médecine légale, les résultats de l'analyse phylogénétique moléculaire ont été

acceptés comme preuves dans les tribunaux pénaux. Certaines statistiques bayésiennes sophistiquées et méthodes fondées sur la vraisemblance pour l'analyse de l'ADN ont été appliquées à l'analyse de l'identité judiciaire. Il convient de mentionner que la génomique et la bio-informatique sont maintenant sur le point de révolutionner notre système de santé en développant une médecine personnalisée. Le séquençage génomique à haut débit couplé à une technologie informatique permettra de séquencer rapidement le génome d'un patient et de détecter facilement les mutations nocives potentielles et de s'engager dans un diagnostic précoce et un traitement efficace des maladies. Des outils de bio-informatique sont également utilisés en agriculture. Les bases de données sur le génome des plantes et les analyses des profils d'expression des gènes ont joué un rôle important dans le développement de nouvelles variétés de cultures qui ont une productivité plus élevée et une plus grande résistance aux maladies (Diniz WJ et al. 2017 ; Keerthikumar S. 2017).

#### 4 – Les limites de la bio-informatique

Ayant reconnu la puissance de la bio-informatique, il est également important de prendre conscience de ses limites et d'éviter de trop compter sur les résultats générés.

En effet, la bio-informatique a un certain nombre de limites inhérentes. À bien des égards, le rôle de la bio-informatique dans la recherche en génomique et en biologie moléculaire peut être comparé au rôle de la collecte de renseignements sur les champs de bataille. Le renseignement est clairement très important pour conduire à la victoire sur un champ de bataille. Mener une bataille sans intelligence est inefficace et dangereux. Le fait de disposer d'informations supérieures et d'une intelligence correcte permet d'identifier les faiblesses de l'ennemi et de révéler sa stratégie et ses intentions. Les informations recueillies peuvent ensuite être utilisées pour diriger les forces pour engager l'ennemi et gagner la bataille. Cependant, s'appuyer entièrement sur l'intelligence peut également être dangereux si l'intelligence est d'une précision limitée. Une confiance excessive dans des renseignements de mauvaise qualité peut entraîner des erreurs coûteuses, voire des échecs complets.

Par analogie, combattre des maladies ou d'autres problèmes biologiques en utilisant la bio-informatique revient à mener des batailles avec intelligence. La bio-informatique et la biologie expérimentale sont des activités indépendantes mais complémentaires. La bio-informatique dépend de la science expérimentale pour produire des données brutes à analyser. À son tour, Elle fournit une interprétation utile des données expérimentales et des pistes importantes pour de nouvelles recherches expérimentales. Les prédictions bio-informatiques ne sont la preuve formelle d'aucun concept. Ils ne remplacent pas les méthodes traditionnelles de recherche

expérimentale consistant à tester des hypothèses. De plus, la qualité des prédictions bio-informatiques dépend de la qualité des données et de la sophistication des algorithmes utilisés. Les données de séquence provenant d'analyses à haut débit contiennent souvent des erreurs. Si les séquences sont incorrectes, les résultats de l'analyse en aval sont également trompeurs. C'est pourquoi il est si important de maintenir une perspective réaliste du rôle de la bio-informatique (Wenzhi Mao et al. 2015).

## 5 – Perspectives de la bio-informatique

La bio-informatique est un domaine qui recèle un grand potentiel pour révolutionner la recherche biologique dans les prochaines décennies. Actuellement, le domaine est en pleine expansion. En plus de fournir des outils de calcul plus fiables et plus rigoureux pour l'analyse séquentielle, structurelle et fonctionnelle, le principal défi pour le développement futur de la bio-informatique est de développer des outils pour élucider les fonctions et les interactions de tous les biomarqueurs dans l'organisme. Cela représente un défi énorme car il nécessite l'intégration de domaines disparates de connaissances biologiques et une variété d'outils mathématiques et statistiques complexes. Pour acquérir une compréhension plus approfondie des fonctions cellulaires, des modèles mathématiques sont nécessaires pour simuler une grande variété de réactions et d'interactions intracellulaires au niveau de la cellule entière. Cette simulation moléculaire de tous les processus cellulaires est appelée biologie des systèmes. L'atteinte de cet objectif représentera un bond en avant vers la compréhension complète d'un système vivant. C'est pourquoi la simulation et l'intégration au niveau du système sont considérées comme l'avenir de la bio-informatique. Modéliser de tels réseaux complexes et faire des prédictions sur leur comportement présentent d'énormes défis et opportunités pour les bio-informaticiens (Ambrosino L et al. 2019).

## II – Les bases de données

L'une des caractéristiques des approches omiques moderne est la génération d'énormes quantités de données brutes. À mesure que le volume de ces données augmente, des méthodologies de calcul sophistiquées sont nécessaires pour gérer le déluge de données. Le tout premier défi est de stocker et de gérer l'énorme volume d'informations et de traiter la grande quantité de données biologiques moléculaires.

## 1 – Définition

Une base de données est une archive informatisée utilisée pour stocker et organiser les données de telle manière que les informations peuvent être facilement récupérées via une variété de critères de recherche. Les bases de données sont composées de matériel informatique et de logiciels pour la gestion des données. Chaque enregistrement, également appelé entrée, doit contenir un certain nombre de champs contenant les éléments de données réels. Pour récupérer un enregistrement particulier de la base de données, un utilisateur peut spécifier une information particulière, appelée valeur, à trouver dans un champ particulier et s'attendre à ce que l'ordinateur récupère l'intégralité de l'enregistrement de données. Ce processus s'appelle faire une requête. Bien que la récupération des données soit le principal objectif de toutes les bases de données, les bases de données biologiques ont souvent un niveau d'exigence plus élevé, connu sous le nom de découverte de connaissances, qui se réfère à l'identification de connexions entre des éléments d'information qui n'étaient pas connus lors de la première saisie des informations. Par exemple, les bases de données contenant des informations de séquence brutes peuvent effectuer des tâches de calcul supplémentaires pour identifier l'homologie de séquence ou les motifs conservés. Ces caractéristiques facilitent la découverte de nouvelles connaissances biologiques à partir de données brutes (Christensson P. 2009 ; Shobha S et al. 2018).

## 2 – Types de bases de données

À l'origine, les bases de données utilisaient toutes un format de fichier plat, qui est un long fichier texte contenant de nombreuses entrées séparées par un délimiteur, un caractère spécial tel qu'une barre verticale. Dans chaque entrée se trouvent un certain nombre de champs séparés par des tabulations ou des virgules. À l'exception des valeurs brutes de chaque champ, le fichier texte complet ne contient aucune instruction cachée permettant aux ordinateurs de rechercher des informations spécifiques ou de créer des rapports basés sur certains champs de chaque enregistrement. Le fichier texte peut être considéré comme un seul tableau. Ainsi, pour rechercher dans un fichier plat une information particulière, un ordinateur doit lire l'intégralité du fichier, un processus manifestement inefficace. Cela est gérable pour une petite base de données, mais à mesure que la taille de la base de données augmente ou que les types de données deviennent plus complexes, ce style de base de données peut devenir très difficile pour la récupération d'informations. En effet, les recherches dans ces fichiers provoquent souvent des pannes de l'ensemble du système informatique en raison de la nature gourmande en mémoire de l'opération.

Pour faciliter l'accès et la récupération des données, des logiciels pour l'organisation, la recherche et l'accès aux données ont été développés. Ils sont appelés systèmes de gestion de bases de données. Ces systèmes contiennent non seulement des enregistrements de données brutes mais également des instructions opérationnelles pour aider à identifier les connexions cachées entre les enregistrements de données. L'établissement d'une structure de données a pour but de faciliter l'exécution des recherches et de combiner différents enregistrements pour former des rapports de recherche finaux. Selon les types de structures de données, ces systèmes de gestion de bases de données peuvent être classés en deux types: les systèmes de gestion de bases de données relationnelles et les systèmes de gestion de bases de données orientées objet (NCBI Resource Coordinators. 2015 ; Hagen MS et al. 2010).

### 2 – 1 – Bases de données relationnelles

Au lieu d'utiliser une seule table comme dans une base de données de fichiers plats, les bases de données relationnelles utilisent un ensemble de tables pour organiser les données. Chaque table, également appelée relation, est composée de colonnes et de lignes. Les colonnes représentent des champs individuels. Les lignes représentent des valeurs dans les champs des enregistrements. Les colonnes d'une table sont indexées selon une caractéristique commune appelée attribut, afin qu'elles puissent être croisées dans d'autres tables. Pour exécuter une requête dans une base de données relationnelle, le système sélectionne les éléments de données liés dans différentes tables et combine les informations dans un seul rapport. Par conséquent, des informations spécifiques peuvent être trouvées plus rapidement à partir d'une base de données relationnelle qu'à partir d'une base de données de fichiers plats (De Silva NH. 2017 ; Rice M et al. 2004).

### 2 – 2 – Bases de données orientées objet

L'un des problèmes des bases de données relationnelles est que les tables utilisées ne décrivent pas les relations hiérarchiques complexes entre les éléments de données. Pour surmonter le problème, des bases de données orientées objet ont été développées qui stockent les données sous forme d'objets. Dans un langage de programmation orienté objet, un objet peut être considéré comme une unité qui combine des données et des routines mathématiques qui agissent sur les données. La base de données est structurée de telle sorte que les objets sont liés par un ensemble de pointeurs définissant des relations prédéterminées entre les objets. La recherche dans la base de données implique la navigation dans les objets à l'aide des pointeurs reliant différents objets. Les langages de programmation comme C ++ sont utilisés pour créer

des bases de données orientées objet. Le système de base de données orienté objet est plus flexible; les données peuvent être structurées en fonction de relations hiérarchiques. Ce faisant, les tâches de programmation peuvent être simplifiées pour les données qui sont connues pour avoir des relations complexes. Cependant, ce type de système de base de données n'a pas la base mathématique rigoureuse des bases de données relationnelles. Il existe également un risque que certaines des relations entre les objets soient déformées. Certaines bases de données actuelles ont donc incorporé des caractéristiques des deux types de programmation de base de données, créant le système de gestion de base de données relationnelle objet (Hebart MN et al. 2019 ; Pornputtapong N et al. 2014)

### 2 – 3 – Bases de données biologiques

Les bases de données biologiques actuelles utilisent tous les types de structures de base de données: fichiers plats, relationnels et orientés objet. Malgré les inconvénients évidents de l'utilisation de fichiers plats dans la gestion des bases de données, de nombreuses bases de données biologiques utilisent toujours ce format. La justification de cette mesure est que ce système implique un minimum de conception de base de données et que les résultats de la recherche peuvent être facilement compris par les biologistes. En fonction de leur contenu, les bases de données biologiques peuvent être divisées en trois catégories: bases de données primaires, bases de données secondaires et bases de données spécialisées. Les bases de données primaires contiennent des données biologiques originales. Ce sont des archives de séquences brutes ou de données structurelles soumises par la communauté scientifique. GenBank et Protein Data Bank (PDB) sont des exemples de bases de données primaires. Les bases de données secondaires contiennent des informations traitées par ordinateur ou organisées manuellement, sur la base des informations originales des bases de données primaires. Les bases de données de séquences de protéines traduites contenant des annotations fonctionnelles appartiennent à cette catégorie, les exemples sont SWISS-Prot et Protein Information Resource (PIR). Les bases de données spécialisées sont celles qui répondent à un intérêt de recherche particulier. Par exemple, Flybase, la base de données de séquences du VIH et le projet de base de données ribosomale sont des bases de données spécialisées dans un organisme particulier ou un type particulier de données. L'Atlas du génome du cancer (TCGA), une base de donnée de génomique du cancer, qui regroupe plus de 20 000 cancers primitifs et correspond à des échantillons normaux couvrant 33 types de cancers (Cannataro M et al. 2014 ; Helmy M et al. 2016).

### 2 – 3 – 1 – Bases de données primaires

Il existe plusieurs bases de données de séquences publiques qui stockent des données brutes sur les séquences d'acides nucléiques produites et soumises par des chercheurs du monde entier : GenBank, la base de données du laboratoire européen de biologie moléculaire (EMBL) et la DNA Data Bank of Japan (DDBJ). La plupart des données dans les bases de données sont fournies directement par les auteurs avec un niveau minimal d'annotation. Un petit nombre de séquences, en particulier celles publiées en 1980, ont été saisies manuellement à partir de la littérature publiée par le personnel de gestion de la base de données (Benson DA et al. 2010 ; Stoesser G et al. 2002). Actuellement, la soumission de séquences à GenBank, EMBL ou DDBJ est une condition préalable à la publication dans la plupart des revues scientifiques pour garantir que les données moléculaires fondamentales soient mises à disposition gratuitement. Ces bases de données publiques collaborent étroitement et échangent quotidiennement de nouvelles données. Elles constituent la collaboration internationale de base de données de séquences nucléotidiques. Cela signifie qu'en se connectant à l'une de ces bases de données, il faut avoir accès aux mêmes données de séquence nucléotidique. Bien que ces bases de données contiennent toutes les mêmes ensembles de données brutes, chacune des bases de données individuelles a un type de format légèrement différent pour représenter les données.

Heureusement, pour les structures tridimensionnelles des macromolécules biologiques, il n'y a qu'une seule base de données centralisée, la Protein Data Bank (PDB) (Berman HM et al. 2000). Cette base de données archive les coordonnées atomiques des macromolécules (protéines et acides nucléiques) déterminées par cristallographie aux rayons X et la résonance magnétique nucléaire (RMN). Elle utilise un format de fichier plat pour représenter le nom de la protéine, les auteurs, les détails expérimentaux, la structure secondaire, les cofacteurs et les coordonnées atomiques. L'interface Web de PDB fournit également des outils de visualisation pour une manipulation d'image simple (Zou D et al. 2015).

### 2 – 3 – 2 – Bases de données secondaires

Les informations d'annotation de séquence dans la base de données primaire sont souvent minimales. Pour transformer les informations de séquence brutes en connaissances biologiques plus sophistiquées, un post-traitement important des informations de séquence est nécessaire. Cela nécessite le besoin de bases de données secondaires, qui contiennent des informations de séquence traitées par calcul dérivées des bases de données primaires. La quantité de travail de traitement informatique varie considérablement entre les bases de données secondaires; certains

sont de simples archives de données de séquence traduites à partir de cadres de lecture ouverts identifiés dans l'ADN, tandis que d'autres fournissent des annotations et des informations supplémentaires liées à des niveaux plus élevés d'informations concernant la structure et les fonctions.

Un exemple frappant de bases de données secondaires est SWISS-PROT (Bairoch A et al. 2000), qui fournit une annotation de séquence détaillée qui inclut la structure, la fonction et l'affectation de la famille des protéines. Les données de séquence sont principalement dérivées de TrEMBL, une base de données de séquences d'acide nucléique traduites stockées dans la base de données EMBL. L'annotation de chaque entrée est soigneusement organisée par des experts et est donc de bonne qualité. L'annotation des protéines comprend la fonction, la structure du domaine, les sites catalytiques, la liaison au cofacteur, la modification post-traductionnelle, les informations sur la voie métabolique, l'association avec la maladie et la similitude avec d'autres séquences. La plupart de ces informations sont obtenues à partir de la littérature scientifique et entrées par les conservateurs de la base de données. L'annotation apporte une valeur ajoutée significative à chaque enregistrement de séquence d'origine. L'enregistrement de données fournit également des liens de renvoi à d'autres ressources en ligne d'intérêt. D'autres caractéristiques telles qu'une très faible redondance et un haut niveau d'intégration avec d'autres bases de données primaires et secondaires rendent SWISS-Prot très populaire auprès des biologistes.

Un effort récent pour combiner SWISS-PROT, TrEMBL et PIR a conduit à la création de la base de données UniProt (UniProt Consortium. 2008), qui a une couverture plus large que n'importe laquelle des trois bases de données tout en conservant la fonction SWISS-PROT originale de faible redondance, et une annotation de haute qualité (Mewes HW et al. 2011 ; Zou D et al. 2015).

### 2 – 3 – 3 – Bases de données spécialisées

Les bases de données spécialisées servent pour la recherche spécifique ou se concentrent sur un organisme particulier. Le contenu de ces bases de données peut être des séquences ou d'autres types d'informations. Les séquences de ces bases de données peuvent chevaucher une base de données principale, mais peuvent également avoir de nouvelles données soumises directement par les auteurs. Parce qu'ils sont souvent organisés par des experts dans le domaine, ils peuvent avoir des organisations uniques et des annotations supplémentaires associées aux séquences. De nombreuses bases de données génomiques spécifiques à la taxonomie entrent dans cette

catégorie. Les exemples incluent Flybase, WormBase, TAIR et TCGA (McQuilton P et al. 2012 ; Todd W Harris et al. 2020 ; Rhee SY et al. 2003 ; Krasnov GS et al. 2019).

### 3 – Interconnexion entre les bases de données biologiques

Les bases de données primaires sont des référentiels centraux et des distributeurs d'informations brutes sur la séquence et la structure. Par conséquent, dans la communauté biologique, il est souvent nécessaire que les bases de données secondaires et spécialisées se connectent aux bases de données primaires et continuent de télécharger les informations de séquence. De plus, un utilisateur a souvent besoin d'obtenir des informations des bases de données primaires et secondaires pour effectuer une tâche car les informations d'une seule base de données sont souvent insuffisantes. Au lieu de permettre aux utilisateurs de visiter plusieurs bases de données, il est pratique que les entrées d'une base de données soient croisées et liées à des entrées connexes dans d'autres bases de données qui contiennent des informations supplémentaires. Le principal obstacle à la liaison de différentes bases de données biologiques est l'incompatibilité des formats, les structures de base de données hétérogènes limitent la communication entre les bases de données. Une solution pour mettre en réseau les bases de données consiste à utiliser un langage de spécification appelé Common Object Request Broker Architecture (COBRA) (Zhang ZJ et al. 2009), qui permet aux programmes de bases de données à différents emplacements de communiquer dans un réseau via un «interface broker» sans avoir à comprendre la structure de la base de données de chacun. Il fonctionne d'une manière similaire au langage HTML (HyperText Markup Language) pour les pages Web, étiquetant les entrées de base de données à l'aide d'un ensemble de balises courantes. Un protocole similaire appelé eXtensible Markup Language (XML) aide également à faire le pont entre les bases de données. Dans ce format, chaque enregistrement biologique est décomposé en petits composants de base qui sont étiquetés avec une imbrication hiérarchique des étiquettes. Cette structure de base de données améliore considérablement la distribution et l'échange d'annotations de séquences complexes entre les bases de données. Récemment, un protocole spécialisé pour l'échange de données bioinformatiques a été développé. Il s'agit du système d'annotation distribué, qui permet à un ordinateur de contacter plusieurs serveurs et de récupérer des informations d'annotation de séquence dispersées liées à une séquence particulière et d'intégrer les résultats dans un seul rapport combiné (Li X et al. 2003).

### III – Biomarqueurs et médecine personnalisées

#### 1 – Biomarqueurs du cancer

Les biomarqueurs sont des molécules qui indiquent un processus normal ou anormal dans l'organisme. Des signes, des substances ou des processus anormaux peuvent indiquer la présence d'une maladie ou d'une affection sous-jacente. Plusieurs types de molécule, notamment l'ADN, les protéines et l'ARN, peuvent servir de biomarqueurs. Les biomarqueurs sont produits par le tissu cancéreux ou par d'autres cellules du corps en réponse au cancer. Les biomarqueurs peuvent être présents dans le sang, l'urine et le tissu cancéreux, ainsi que dans d'autres tissus et liquides organiques. Les biomarqueurs ne se limitent pas au cancer. Il existe des biomarqueurs pour bien d'autres maladies, comme les maladies du cœur ou la sclérose en plaques. Les biomarqueurs sont aussi appelés « marqueurs moléculaires », il existe plus d'un type de biomarqueurs dans le cancer. Selon les caractéristiques particulières de la molécule, les biomarqueurs remplissent différentes fonctions et peuvent réagir de manière spécifique à certains traitements. La caractérisation des biomarqueurs peut se faire selon quatre dimensions : (i) la nature du paramètre mesuré, peut être biologique, anatomique, physiologique, biochimique, moléculaire) (ii) l'aire thérapeutique ciblée comme par exemple l'oncologie, la neurologie, l'immunologie, l'infectiologie, la cardiologie, (iii) la technologie de support qui repose essentiellement sur des technologies révélant les altérations d'un gène ou de son produit (ARN, protéine) mais également sur l'imagerie ou l'informatique, (iv) le niveau de complexité entrera aussi en compte selon qu'il s'agit d'un marqueur monoparamétrique ou multiparamétrique par exemple une détermination d'une signature biologique). La notion de biomarqueur est un concept polyvalent et va être associé à une ou plusieurs étapes du cycle de vie du médicament. Dans le cadre de la Recherche&Développement (R&D), la découverte de nouvelles molécules thérapeutiques sera essentiellement associée à l'utilisation de biomarqueurs de recherche pour lesquels deux types sont définis : (i) les biomarqueurs mécanistiques, ils permettent d'étudier et/ou découvrir les mécanismes physiopathologiques d'intérêts (ii) les biomarqueurs de stade qui permettent de faire la distinction entre les différents stades de la maladie (Goossens N et al. 2015 ; George AP et al. 2019). Au cours de l'étape d'Évaluation Thérapeutique des nouvelles molécules identifiées, comprenant les études précliniques et cliniques qui seront spécifiquement utilisés des biomarqueurs d'efficacité et de toxicité ainsi que des biomarqueurs pharmacogénomiques. Les biomarqueurs d'efficacité/toxicité sont des paramètres biologiques se substituant à un critère clinique et permettent une évaluation provisoire du rapport bénéfice/risque du traitement. Les

biomarqueurs pharmacogénomiques, quant à eux, vont permettre l'étude des mécanismes génétiques et des variations individuelles de la réponse aux médicaments. Enfin, certains biomarqueurs seront utilisés de manière transversale au cours de la chaîne de valeur du médicament. On retrouve les biomarqueurs de diagnostique et/ou prédictif associés à chaque étape du cycle de vie (R&D, évaluation thérapeutique et commercialisation). Ces derniers permettent d'identifier la pathologie et des populations cibles, et donnent une indication précoce sur la réponse au traitement. Enfin, les biomarqueurs de pronostique et de suivi, associés à l'évaluation thérapeutique et à la commercialisation, permettront de prédire l'évolution de la maladie et de suivre sa progression (Selleck MJ et al. 2017 ; Sethi S et al. 2013 ; Mohammed A et al. 2017).

## 2 – L'intérêt du traitement ciblé : la médecine de précision

L'Institut national du cancer définit la médecine de précision ou « médecine personnalisée » comme une « forme de médecine qui utilise l'information sur les gènes, les protéines et l'environnement d'une personne pour prévenir, diagnostiquer et traiter la maladie ». La médecine de précision a fait son apparition au cours des 20 dernières années avec l'avènement et l'amélioration des techniques d'analyse moléculaire. Dans les traitements classiques du cancer, ils utilisent des médicaments qui sont toxiques pour les cellules (agents cytotoxiques) pour empêcher rapidement la division cellulaire en altérant l'ADN et les mécanismes de division cellulaire. Dans les traitements ciblés du cancer fondés sur l'analyse moléculaire, aussi appelés thérapies moléculaires ciblées, ils utilisent des médicaments ou d'autres substances pour réguler la fonction de cibles moléculaires précises agissant sur la signalisation, la prolifération, le métabolisme et la nécrose des cellules. La plupart des tumeurs sont associées à de multiples mutations plutôt qu'à une ou à deux mutations soupçonnées au départ. Cette découverte importante des dernières années expliquent pourquoi des traitements ciblant une seule mutation ne donnent pas toujours les résultats escomptés. Aujourd'hui, les chercheurs et les médecins ont la difficile mission d'utiliser l'information fournie par le profilage moléculaire et d'en déterminer l'incidence sur les traitements ciblés. Les traitements ciblés pourraient ainsi être plus efficaces, causer moins d'effets secondaires et être plus susceptibles de guérir la maladie, ou à tout le moins de mieux la prendre en charge (Ginsburg GS et al. 2018 ; Gameiro GR et al. 2018 ; Bilkey GA et al. 2019).

Les stratégies de recherche actuelles visent à : trouver des traitements adaptés au profil génétique individuel des cellules tumorales, inclure plusieurs agents qui ciblent des molécules

spécifiques dans le même protocole de traitement, les patients sont traités par des agents précis en fonction des anomalies moléculaires détectées dans leurs tumeurs.

Dans certaines formes de cancer, l'analyse moléculaire a largement contribué à améliorer les taux de survie. Cette amélioration s'explique par : la compréhension des caractéristiques moléculaires des tumeurs, la mise au point de tests diagnostiques qui détectent des biomarqueurs et une approche modernisée de mise au point des médicaments où l'on cible des biomarqueurs ou des mécanismes cellulaires précis (Roden DM et al. 2013 ; Love-Koh J et al. 2018 ; Maier M. 2019 ; Goetz LH et al. 2018).

### 3 – The Cancer Genome Atlas (TCGA)

Ce programme est une initiative des instituts nationaux de la santé qui vise à cartographier les altérations génomiques fondamentales en cause dans les principaux types et sous-types de cancer. L'Atlas se veut un outil de référence pour la recherche de pointe et elle a été conçue pour favoriser la mise en place de stratégies plus efficaces pour prévenir, diagnostiquer et traiter le cancer. Il sert aussi de modèle pour d'autres projets de cartographie des génomes. La TCGA a généré plus de 2,5 pétaoctets de données génomiques, épigénomiques, transcriptomiques et protéomiques. Les données qui ont déjà permis d'améliorer la capacité à diagnostiquer, traiter et prévenir le cancer, resteront accessibles au public pour toute personne de la communauté des chercheurs. La TCGA a aidé à établir l'importance de la génomique du cancer, a amélioré la compréhension du cancer et a même commencé à changer la façon dont la maladie est traitée en clinique. L'impact va encore plus loin, atteignant les technologies de la santé et des sciences, la biologie computationnelle et d'autres domaines de recherche. Après 12 ans, les contributions de plus de 11 000 patients et les efforts incroyables de milliers de chercheurs, TCGA a produit un riche ensemble de données d'une valeur incommensurable. Ces données restent accessibles au public en tant que référence de confiance qui sera exploitée pendant de nombreuses années (Tomczak K et. 2015 ; Cancer Genome Atlas Research Network. 2014).

#### 3 – 1 – Impact de la TCGA

Approfondir la compréhension du cancer grâce à plusieurs caractéristiques moléculaires

- Les chercheurs ont détecté des aberrations dans les séquences d'ADN, l'expression génique, l'épigénétique (par exemple, miARN, lncARN et la méthylation), l'expression et la structure des protéines impliquant les différentes conséquences fonctionnelles.

- Bien que des milliers d'altérations aient été identifiées, elles peuvent être mieux comprises dans le contexte des voies fonctionnelles ou regroupées en tant que signatures de mutation distinctes.
- Les différents tissus cancéreux peuvent partager les mêmes altérations et être biologiquement plus similaires les uns aux autres qu'aux autres tumeurs du même tissu d'origine (Wang Z et al. 2016).

Renforcé le domaine de la biologie computationnelle

- La grande quantité de données et de nombreux types de données produits par TCGA ont stimulé une croissance énorme dans le domaine de la biologie computationnelle.
- Les chercheurs développent des outils pour un large éventail d'objectifs, comme identifier des mutations somatiques et germinales, prédire des gènes d'importance pronostique, construire des réseaux de réglementation et établir des workflow pour l'analyse bioinformatique de données TCGA (Zhu Q et al. 2018).

### 3 – 2 – Éthique et politiques de la TCGA

Les principes de partage et de publication des données génomiques ont été élaborés lors d'une réunion à Fort Lauderdale parrainée par le Wellcome Trust. Ces «principes de Fort Lauderdale» ont informé les directives de publication originales de la TCGA, qui équilibrent la mise à disposition immédiate des données pour la recherche et la publication.

Une explication des facteurs pris en compte dans la décision de l'Institut Nationale du Cancer (INC) et l'Institut Nationale de Recherche sur le Génome Humain (NHGRI) de publier les données et les informations de 'Cancer Cell Line Encyclopedia' en tant que données en accès libre.

La TCGA a collecté de nombreux types de données pour chacun des 20 000 échantillons de tumeurs et tissus normaux. Chaque étape du pipeline de caractérisation du génome a généré de nombreux points de données, tels que: les enseignements cliniques et les données de caractérisation moléculaire. Cette base de données a catalysé une croissance et des progrès considérables dans le domaine de la biologie computationnelle en soutenant le développement de technologies de caractérisation génomique à haut débit, en générant une quantité massive de données et en mettant en place des équipes de chercheurs pour analyser les données, ces technologies basées sur le séquençage à haut débit d'ADN et d'ARN dominaient le domaine de la biologie moléculaire pour caractériser et identifier des biomarqueurs cancéreux (Han J et al. 2018 ; Kaur P et al. 2019 ; Sariyar M et al. 2015).

## IV – Les approches Omiques

### 1 – Définition

L'acquisition de marques de cancer nécessite des altérations moléculaires à plusieurs niveaux, notamment le génome, l'épigénome, le transcriptome, le protéome et le métabolome. Au cours de la dernière décennie, de nombreuses tentatives ont été faites pour démêler les mécanismes moléculaires de la cancérogenèse en utilisant des approches omiques uniques, comme le balayage du génome à la recherche de mutations spécifiques au cancer et l'identification des paysages épigénétiques modifiés dans les cellules cancéreuses ou en explorant l'expression différentielle de l'ARNm et de la protéine par des techniques de génomique, d'épigénomique et de transcriptomique, respectivement. Si ces approches omiques à un seul niveau ont contribué à l'identification de mutations spécifiques au cancer, d'altérations épigénétiques et de sous-typage moléculaire des tumeurs sur la base de l'expression des gènes, elles peuvent pas établir la relation entre les signatures moléculaires et la manifestation phénotypique des caractéristiques du cancer (Subramanian I et al. 2020). En revanche, les approches multi-omiques impliquant l'intégration des cellules/tissus cancéreux dans de multiples dimensions ont le potentiel de découvrir le mécanisme moléculaire complexe qui regroupe les différentes manifestations phénotypiques des caractéristiques du cancer telles que les métastases et l'angiogenèse. En outre, les approches multi-omiques peuvent être utilisées pour disséquer la réponse cellulaire à la chimiothérapie ou à l'immunothérapie, ainsi que pour découvrir des candidats moléculaires ayant une valeur diagnostique ou pronostique. L'accumulation d'énormes données multi-omiques spécifiques au cancer dans des bases de données comme le TCGA offre une occasion pour l'approche de la biologie systémique de s'attaquer à la complexité des cellules cancéreuses par l'unification des données expérimentales et des modèles informatiques/mathématiques (Pinu FR et al. 2019 ; Conesa A et al. 2019).

### 2 – Du génome au protéome

Les différents niveaux des approches omiques, génomique, transcriptomique, épigénomique et protéomique varient considérablement dans leur complexité, qui est largement déterminée par la dynamique spatiale et/ou temporelle et les modifications chimiques. Le flux d'informations de l'ADN à l'ARN et finalement aux protéines s'accompagne d'une augmentation exponentielle de la complexité. L'information héréditaire stockée dans le génome sous la forme de 4 nucléotides reste en grande partie statique mais une dynamique temporelle est introduite dans le processus de transcription par lequel les gènes sont transcrits en ARN. La régulation

temporelle de l'expression des gènes en fonction de signaux développementaux, environnementaux et extracellulaires via des réseaux de régulation des gènes fait du transcriptome une entité hautement dynamique (Hasin H et al. 2017). Les ARNm sont engagés dans des systèmes de codage de d'information biologique complexe : processus de traduction où les ARNm codent pour des protéines comprenant 20 acides aminés. Après la synthèse, les protéines sont généralement repliées en de nombreuses conformations possibles en fonction des séquences d'acides aminés primaires et de la modification chimique des résidus d'acides aminés connue sous le nom de modifications post-traductionnelles (PTM). Les protéines subissent un grand nombre de PTM (par exemple, phosphorylation, acétylation et glycosylation) qui peuvent affecter directement leur structure et leur fonctionnalité. De plus, contrairement aux ARNm qui sont synthétisés dans le noyau et traduits dans le cytoplasme, les protéines ont différentes localisations subcellulaires telles que la membrane cellulaire, le cytoplasme, et différents organites subcellulaires liés à la membrane du noyau, aux mitochondries, aux réticulum endoplasmique, etc. L'ensemble de ces événements confère une énorme complexité au protéome (Manzoni C et al. 2018 ; Hager GL et al. 2009).

### 3 – Les applications des approches omiques

#### 3 – 1 – La génomique

Les technologies et les données génomiques ont été les premières données omiques répandues disponibles pour une application en médecine de précision. Par conséquent, la plupart des premiers efforts en matière de médecine de précision se sont principalement concentrés sur l'utilisation d'informations sur les séquences génomiques pour diagnostiquer les patients, prédire le risque de développement de maladies chez certains patients et évaluer si des traitements spécifiques sont appropriés et susceptibles de réussir chez certains patients. Au cours de la dernière décennie, les études d'association à l'échelle du génome et, plus récemment, les données de séquençage de l'exome entier et du génome entier de la prochaine génération ont fourni un vaste ensemble de variantes de séquences d'ADN qui sont associées à une pléthore de maladies et de traits chez l'homme. Malgré l'expansion rapide de ces ensembles de données, les progrès réalisés au niveau de l'identification des variantes de séquences spécifiques à l'origine des maladies et des mécanismes physiopathologiques par lesquels elles affectent le développement ou la progression des maladies ont été lents. L'application des approches de score de risque polygénique à une série de maladies communes a été une stratégie pour fournir des informations potentiellement utiles et exploitables cliniquement à partir de ces données génétiques sans qu'il soit nécessaire d'identifier les changements de séquence causale

individuels et leurs modes d'action. Toutefois, l'application de ces scores de risque a été limitée à quelques maladies seulement. Les nombreuses informations sur les variantes génétiques ont également été utilisées en pharmacogénomique pour prédire les réponses aux médicaments chez des patients individuels. Tout comme les scores de risque, cette stratégie a eu un succès limité. Il est clair que des progrès supplémentaires sont nécessaires pour utiliser efficacement les informations génomiques disponibles et évaluer les avantages et les limites de la prédiction des risques, du diagnostic des maladies et de la prise de décision en matière de traitement (Gallo Cantafio ME et al. 2018 ; Alexandrov LB et al. 2013).

En oncologie, les approches génomiques se sont concentrées sur le séquençage de l'ADN pour identifier les mutations spécifiques au cancer, et pour analyser les réarrangements chromosomiques afin de caractériser les classes ou sous-types de cancer. L'ADN isolé des cellules cancéreuses est porteur d'un large éventail de modifications somatiques et potentiellement héréditaires, allant des polymorphismes, insertions et délétions d'un seul nucléotide aux variations du nombre de copies et aux réarrangements structuraux plus importants. Le nombre total de mutations et/ou de réarrangements chromosomiques varie considérablement entre les types de tumeurs, et entre les mêmes types de tumeurs chez différents individus, ce qui rend difficile l'identification des changements fonctionnels clés qui sont essentiels à la formation et à la croissance des tumeurs. Le schéma global des mutations dans le cancer humain a été largement étudié au cours de la dernière décennie (Burrell RA et al. 2013 ; Tomczak K et al. 2015).

Les premières études d'oncologie génétique se sont concentrées sur l'identification des mutations héréditaires qui augmentent considérablement le risque de cancer. Les premières études sur les familles présentant un regroupement de types de cancer particuliers ont permis d'identifier un certain nombre de gènes et de mutations qui, lorsqu'elles sont héritées, augmentent le risque de certains types de cancer. Les premiers exemples comprennent le syndrome de Lynch, un cancer colorectal souvent causé par des mutations dans les gènes impliqués dans la réparation des anomalies de l'ADN (MLH1, MSH2, MSH6, PMS1, PMS2, (Silva FC et al. 2009), le syndrome de Li-Fraumeni, un syndrome héréditaire le plus souvent causé par des mutations dans le gène suppresseur de tumeur TP53 qui peut conduire au développement d'un large éventail de cancers, y compris les sarcomes, la leucémie et les cancers du cerveau (Guha T et al. 2017) et le syndrome héréditaire du cancer du sein et des ovaires, souvent causé par des mutations dans les gènes BRCA1 et BRCA2 (Kurian AW. 2010). Les personnes présentant des mutations dans l'un de ces gènes ont un risque nettement plus élevé de développer un cancer que les personnes non porteuses. Ces découvertes ont facilité

l'utilisation des informations génétiques pour les décisions de traitement chez les personnes concernées et ont conduit à la création des prestataires qui proposent des tests génétiques ciblés pour les patients atteints de cancer. Dans le prolongement de ces premières études, l'association pan-génomique et, plus récemment, le développement de scores de risque polygénique ont permis d'augmenter le nombre de variantes de séquences et de mutations associées à un risque accru de cancer. Un nombre important de ces variantes sont aujourd'hui systématiquement testées comme diagnostiques complémentaires pour la caractérisation des tumeurs, et la prise de décision en matière de traitement (Sud A et al. 2017).

D'autres études ont porté sur l'analyse des réarrangements chromosomiques et de la variation du nombre de copies dans les cellules cancéreuses. De nombreuses technologies, dont l'hybridation in situ par fluorescence, l'hybridation génomique comparative (CGH, pour les petits réarrangements chromosomiques et la variation du nombre de copies) et, depuis peu, le séquençage du génome entier de la prochaine génération, permettent de caractériser en détail les anomalies chromosomiques prédisposant au développement du cancer et d'identifier les altérations spécifiques des cellules cancéreuses qui contribuent à expliquer la croissance cellulaire anormale ou la résistance au traitement, comme dans le cas de la formation de gènes de fusion tels que BCR-ABL (chromosome de Philadelphie) ou bien la mutation R132H au niveau du gène IDH1 impliqué dans la croissance des gliomes (Vinhas R et al. 2018). Dernièrement, le séquençage du génome entier de la prochaine génération a également été utilisé pour caractériser la conversion des gènes par recombinaison inter-chromosomique en comparant les tissus tumoraux de plusieurs cancers avec les tissus tumoraux adjacents. Malgré ces progrès considérables dans les essais cliniques pour les formes de cancer héréditaires, les progrès vers une meilleure compréhension des mécanismes génétiques sous-jacents contribuant à la majorité des cancers ont été plus difficiles. Les progrès récents dans les technologies de séquençage permettent d'analyser efficacement les tumeurs individuelles, y compris l'analyse des séquences au niveau d'une seule cellule. L'ensemble croissant de données disponibles à partir de ces efforts a révélé que le profil sous-jacent des mutations somatiques s'accumulent dans les cellules tumorales, et il a été difficile de distinguer les mutations potentielles (qui sont responsables de la formation et de la croissance de tumeurs) à partir de mutations «passagers» qui sont phénotypiquement neutres. La compréhension de la charge mutationnelle des tumeurs a toutefois permis de mieux définir les mécanismes physiopathologiques contribuant au développement et à la progression des tumeurs (Hu T et al. 2018 ; Gibson G. 2019).

### 3 – 2 – La transcriptomique

Approche qui examine les niveaux d'ARN, à la fois qualitativement (transcriptions présentes, identification des sites d'épissage) et quantitativement (quantité exprimée de chaque transcription). Le dogme central de la biologie considère l'ARN comme un intermédiaire moléculaire entre l'ADN et les protéines, qui sont considérées comme la principale lecture fonctionnelle de l'ADN. D'autres exemples de fonction de l'ARN, tels que les complexes structurels (par exemple, les complexes ribosomiques), ou régulateurs (par exemple, l'ARN non codant Xist dans l'inactivation de ChrX). L'avènement des grandes études de transcriptomique au cours de la dernière décennie a montré que si seulement ~3% du génome encode les protéines (Consortium EP. 2012). Les études de séquençage de l'ARN ont identifié des milliers de nouvelles isoformes et ils ont montré une complexité du transcriptome codant pour les protéines plus grande que ce qui était apprécié auparavant (Trapnell C et al. 2010). Cependant, une contribution encore plus importante de ces études a été le développement du domaine de l'ARN non codant. Il est désormais clair que des milliers de longs ARN non codants transcrits dans des cellules de mammifères jouent un rôle essentiel dans de nombreux processus physiologiques, par exemple la différenciation de l'adiposité brune (Alvarez-Dominguez JR et al. 2015), la régulation endocrinienne (Knoll M et al. 2015) et le développement des neurones (Yao B et al. 2016). La dérégulation des ARNs longs non codants a été impliquée dans diverses maladies, telles que l'infarctus du myocarde (Ishii N et al. 2006), le diabète (Arnes L et al. 2016 ; Moran I et al. 2012), le cancer (Gupta RA et al. 2010) et d'autres (Schmitz SU et al. 2016). En plus des ARNs longs non codants, le séquençage à haut débit permet l'interrogation des ARNs courts comme les miRNAs (Feinberg MW et al. 2016). Tout comme pour les ARN longs non codants, un nombre croissant de preuves indiquent un dérèglement des ARNs courts dans la maladie (Chen Y et al. 2016 ; Li LJ et al. 2016) et leur utilisation potentielle comme biomarqueurs ou comme cibles thérapeutiques.

À ce jour, près de 800 000 ensembles de données sur l'expression des gènes liés au cancer ont été déposés dans la base de données Gene Expression Omnibus (GEO) du centre nationale de biotechnologie (NCBI). L'initiative TCGA a également utilisé la transcriptomique pour fournir une caractérisation détaillée de l'expression génique de tumeurs individuelles et de tissus de plus de 11 000 patients (Cancer Genome Atlas Research Network. 2013). De nombreuses autres études ont également révélé des signatures d'expression génique caractéristiques qui peuvent aider à prédire les résultats pour les patients et la réponse au traitement pour divers types de cancer, et plusieurs tests cliniques sont maintenant sur le marché pour prédire le pronostic ou

le risque de récurrence pour les patients atteints de cancer du sein, de cancer colorectal, de glioblastome multiforme et de cancer du poumon (Prat A et al. 2011 ; Duarte CW et al. 2012 ; Li W et al. 2012 ; Botling J et al. 2013). Des études récentes ont étendu l'analyse de l'expression génétique à des cellules isolées (Song Q et al. 2019) et les données qui en résultent sont susceptibles d'élargir notre compréhension de l'hétérogénéité cellulaire dans le cancer (Suva ML et al. 2019) et peuvent influencer notre prise de décision clinique.

### 3 – 3 – L'épigénomique

Approche qui se concentre sur la caractérisation à l'échelle du génome des modifications réversibles de l'ADN ou des protéines associées à l'ADN, telles que la méthylation de l'ADN ou l'acétylation des histones. Les modifications covalentes de l'ADN et des histones sont des régulateurs majeurs de la transcription des gènes et, par conséquent, du destin cellulaire (Piunti A et al. 2016). Ces modifications peuvent être influencées à la fois par des facteurs génétiques et environnementaux, ils peuvent être durables et parfois héréditaires (Gut P et al. 2013 ; Liu L et al. 2008 ; Taudt A et al. 2016). Bien que le rôle des modifications épigénétiques en tant que médiateurs des effets environnementaux transgénérationnels reste controversé (Barres R et al. 2016 ; Nestler EJ. 2016) leur importance dans les processus biologiques et le développement des maladies est évidente d'après les nombreuses études d'association à l'échelle de l'épigénome qui ont été rapportées. Par exemple, les régions d'ADN différemment méthylées peuvent être utilisées comme indicateurs du statut de la maladie pour le syndrome métabolique (Horvath S. 2013 ; Multhaup ML et al. 2015) les maladies cardiovasculaires (Kim M et al. 2010) le cancer (Baylin SB et al. 2001) et de nombreux autres états physiopathologiques (Raghuraman S et al. 2016).

La méthylation de l'ADN a été utilisée pour établir des priorités dans les schémas de traitement de certaines affections. Le statut de méthylation de l'ADN au niveau du promoteur MGMT est actuellement utilisé pour déterminer si le témozolomide, un agent d'alkylation de l'ADN, sera efficace dans le traitement du glioblastome (Hegi ME et al. 2005). Le MGMT est une enzyme de réparation de l'ADN, et lorsque le promoteur du gène est méthylé, il est exprimé à de faibles niveaux. Le traitement au témozolamide est efficace dans ce cas car les dommages aux cellules cancéreuses induits par le traitement ne peuvent être réparés, ce qui entraîne la mort des cellules. À l'inverse, lorsque le promoteur MGMT n'est pas méthylé, le gène est exprimé et l'effet néfaste du témozolamide est limité par la capacité de la cellule à réparer les dommages (El-Awady RA et al. 2015), les ADN méthyltransférases ou les inhibiteurs d'histone désacétylases qui peuvent interférer avec ces mécanismes ont été testés comme médicaments anticancéreux présumés et

ont montré des résultats précliniques prometteurs dans le cancer du pancréas (Neureiter, D et al. 2014). De même, les profils d'expression de molécules spécifiques de miARN et de lncARN peuvent aider à identifier les cellules tumorales résistantes aux médicaments chez les patients soumis à une chimiothérapie standard, les identifiant comme des patients nécessitant un traitement supplémentaire. Enfin, les miARNs et les lncARNs peuvent offrir des possibilités de nouvelles approches thérapeutiques pour les cancers qui ne sont pas efficacement ciblés par les options thérapeutiques actuellement disponibles, comme le glioblastome (Teplyuk, N.M et al. 2016). Les signatures épigénétiques sont souvent spécifiques aux tissus et plusieurs grands consortiums s'emploient à établir des cartes épigénomiques complètes dans de nombreux tissus humains. En plus des connaissances acquises grâce à l'identification des modifications épigénétiques en corrélation avec les maladies, les données générées par ces études ont un grand potentiel pour améliorer notre interprétation fonctionnelle des variantes génétiques résidant dans ces régions.

### 3 – 4 – La protéomique

Au fur et à mesure que l'analyse de l'expression des gènes a progressé, les recherches se sont tournées vers l'analyse des protéines cellulaires, les "produits" de traduction des transcriptions d'ARN et les principaux médiateurs de la fonction cellulaire. Alors que la transcriptomique permet de quantifier le produit immédiat du génome d'une cellule à un moment donné (et donc de "mesurer" l'activité directe du génome, ou toute modification de celui-ci dans différentes conditions ou au cours du développement de la maladie), les premières études à grande échelle des protéomes cellulaires ont montré une corrélation relativement faible entre les niveaux d'expression des protéines et les niveaux d'expression correspondants de l'ARNm (Washburn MP et al. 2003). Il est clair que la fonction des protéines est médiée et altérée par une myriade de mécanismes. Les modifications post-traductionnelles des protéines (par exemple, la phosphorylation) sont souvent nécessaires à l'activité ou à la signalisation. Le pliage et le traitement post-traductionnel des pré-protéines, ainsi que la formation de complexes multi-protéiques sont nécessaires pour créer la machinerie qui préforme les fonctions cellulaires requises. Enfin, la localisation cellulaire d'une protéine détermine souvent si une protéine exprimée est active ou inactive. Par conséquent, aucune méthodologie n'est en mesure d'évaluer correctement le "protéome" sous toutes ses facettes, et la plupart des méthodologies de protéomique se concentrent sur l'identification et la quantification de protéines individuelles, généralement par spectrométrie de masse ou par des réseaux de protéines basés sur l'affinité. D'autres méthodes analytiques, telles que la RMN et la cristallographie par rayons X, peuvent

fournir des informations sur la structure des protéines et des complexes protéiques dans les cellules et les tissus.

En protéomique du cancer, le consortium TCGA a représenté le premier effort à grande échelle pour établir le profil du protéome de la tumeur. Cependant, l'analyse a été réalisée à l'aide de réseaux de protéines en phase inverse, et s'est donc limitée à l'analyse ciblée de quelques centaines de protéines. Cependant, plusieurs études récentes ont utilisé des approches de spectrométrie de masse de pointe pour identifier des biomarqueurs spécifiques au cancer de l'ovaire (Swiatly A et al. 2018) et ont utilisé des données protéomiques pour aider à la classification du cancer du sein (Yanovich G et al. 2018). D'autres analyses ont également utilisé des données protéomiques pour prédire la sensibilité aux médicaments et identifier des protéines supposées être des médiateurs de la résistance aux médicaments (Ali M et al. 2018 ; Cruz IN et al. 2017). Ces études protéomiques complètent la classification immunohistochimique traditionnelle des types de tumeurs, telle que la caractérisation de l'expression des récepteurs d'œstrogènes dans les tumeurs cancéreuses du sein.

### 3 – 5 – La métabolomique

La métabolomique est l'analyse des petites molécules présentes dans une cellule, un tissu ou un fluide. Les métabolites sont souvent considérés comme les produits de processus cellulaires, médiés par des protéines ; par conséquent, les changements dans les métabolites sont présumés refléter les changements de fonction des enzymes et des protéines médiatrices. Les progrès des techniques de chromatographie en phase liquide avec spectrométrie de masse en tandem (LC-MS/MS) au niveau de la quantification précise et de l'analyse des données en ont fait une plateforme solide pour l'identification et la quantification simultanées des métabolites.

La grande majorité des analyses métabolomiques se sont concentrées sur l'analyse d'échantillons de plasma ou de sérum de patients, souvent pour tenter d'identifier des biomarqueurs spécifiques au cancer ou à la tumeur qui peuvent être utilisés dans le diagnostic sans nécessiter un échantillon de biopsie tumorale invasive. Cependant, comme dans la plupart des études liées à la maladie, il reste difficile de déterminer la spécificité d'un changement caractéristique des métabolites du plasma pour une maladie particulière ou, dans le cas du cancer, pour un type de tumeur, surtout que les métabolites identifiés comme biomarqueurs dans une maladie sont également altérés dans d'autres maladies.

Les études limitées en biologie du cancer ont identifié certains biomarqueurs métaboliques présumés, tels que les glucides modifiés dans la leucémie myéloïde aiguë (Chaturvedi A et al. 2013) et les acides gras libres insaturés dans le cancer colorectal (Zhang Y et al. 2016). En

revanche, les études sur le cancer de la prostate n'ont révélé que des modifications du citrate (Giskeodegard GF et al. 2013) et des acides aminés à chaîne ramifiée (Mayers JR et al. 2014) qui ont été signalées pour d'autres maladies ce qui suggère que les biomarqueurs observés ne sont probablement pas spécifiques au cancer de la prostate et reflètent plus probablement des mécanismes pathologiques sous-jacents communs. En outre, il reste à voir si les changements identifiés dans les métabolites plasmatiques sont représentatifs des changements dans les cellules tumorales (Ferguson JF et al. 2016 ; Giesbertz P et al. 2016).

#### 4 – Approches basées sur le séquençage de la prochaine génération

Ces dernières années, les techniques de génomique ont surtout été consacrées au séquençage du génome d'un individu afin de comprendre les variations interindividuelles au niveau de la lignée germinale et au niveau somatique. Le passage éventuel des technologies de séquençage de Sanger au séquençage massivement parallèle basé sur les NGS a permis aux chercheurs de séquencer le génome/exome d'intérêt pour caractériser les paysages mutationnels d'un tissu donné. Par exemple, dans le projet atlas du génome du cancer (TCGA), les scientifiques ont utilisé le NGS couplé à une analyse bioinformatique en aval pour découvrir le paysage de mutation somatique dans des milliers d'échantillons de tumeurs représentant les principaux types de cancer, en partant de l'hypothèse que ces études de mutation à l'échelle du génome seraient essentielles pour comprendre la complexité des différents cancers (Kandoth C et al. 2013 ; Lawrence MS et al. 2013).

L'épigénomique est définie par l'identification à l'échelle du génome de modifications chimiques telles que la méthylation et l'acétylation de l'ADN et/ou des protéines histones liant l'ADN. Les modifications épigénétiques des protéines de l'ADN et des histones servent de mécanisme de régulation majeur contrôlant l'expression des gènes et les phénotypes cellulaires (Piunti A and Shilatifard A. 2016). Les études épigénomiques ont joué un rôle essentiel dans la découverte des marqueurs épigénétiques associés à la maladie. Les techniques d'épigénomique qui sont largement utilisées comprennent les essais d'immunoprécipitation de la chromatine (ChIP) couplés à des séquençage à haut débit communément appelés ChIP-séquençage ou ChIP-seq et l'analyse de méthylation par le séquençage du génome entier à base de bisulfites/particules. ChIP-Seq a été développé comme un outil puissant pour l'identification des sites de liaison à l'ADN des facteurs de transcription et des protéines d'histone à l'échelle du génome afin de construire des cartes à haute résolution des marques de modification des histones à l'échelle du génome. ChIP-seq suit un protocole simple dans lequel les protéines liées à l'ADN sont généralement immunoprécipitées par un anticorps spécifique, puis extraites,

purifiées et séquencées. Ces dernières années, l'application de ChIP-seq nous a permis de mieux comprendre les événements de régulation des gènes responsables de diverses maladies et voies biologiques, telles que la progression et le développement du cancer. En comparant ces profils pan-génomiques des marques de modification des histones entre le cancer et les tissus normaux, il a été possible de comprendre comment la dérégulation épigénétique s'est manifestée dans divers cancers tels que le sein et le gliome. En outre, les modifications des histones, les modifications chimiques de certaines bases de l'ADN peuvent avoir des effets épigénétiques. Par exemple, la modification chimique du résidu de cytosine dans la séquence d'ADN promoteur des gènes peut modifier leur expression. En exploitant la puissance des NGS, il est maintenant possible d'analyser le méthylome à l'échelle du génome à une résolution nucléotidique unique (Raj U et al. 2017 ; Li T et al. 2015).

Le séquençage du bisulfite du génome entier (WGBS) est une technologie puissante qui permet de détecter les bases méthylées de la cytosine dans l'ADN génomique. En bref, la méthode implique le traitement de l'ADN génomique avec du bisulfite de sodium suivi d'un séquençage pour construire une carte de la cytosine méthylée à l'échelle du génome avec une résolution à base unique. Une autre technique connue sous le nom "the minimum information about a genome sequence" (MIGS) a été utilisée pour analyser le schéma de méthylation du génome entier (Serre D et al. 2010), cette technique repose sur la précipitation de l'ADN méthylé par le domaine de liaison méthyl-CpG, suivie d'un séquençage. Les techniques de transcriptomique sont engagées dans la détection de la présence et la quantification des transcriptions d'ARN, notamment les ARNm, mais peuvent également être étendues à d'autres types de transcriptions d'ARN non codants tels que les transcriptions des lncARNs et les miARNs. Le transcriptome dans une cellule particulière comprend toutes les molécules d'ARN qui sont transcrites à partir du génome à un moment donné. Contrairement au génome, qui est statique par nature, le transcriptome est soumis à des changements en fonction de stimuli cellulaires, environnementaux, extracellulaires et développementaux de manière temporelle. Avant l'avènement des NGS, les puces à ADN étaient utilisées comme technique de laboratoire classique pour détecter à haut débit les modifications des niveaux d'ARNm dans les cellules à différents stades. Les biopuces peuvent généralement être utilisées pour quantifier l'abondance relative des ARNm de milliers de gènes simultanément. Grâce à cette technique, il est possible de construire des profils d'expression génique cellulaire ou tissulaire entre des états normaux et cancéreux, ce qui peut faciliter l'identification des gènes qui présentent une expression différentielle entre les états normaux et cancéreux. Grâce au développement de techniques NGS efficaces, le séquençage de l'ARN (ARN-seq) permet d'identifier la présence et

l'abondance des transcriptions d'ARN. Contrairement aux puces à ADN, la technologie ARN-seq ne repose pas sur des sondes spécifiques aux transcriptions et peut donc effectuer avec succès la détection non biaisée de nouvelles transcriptions. L'analyse de l'ARN-Seq a montré que le paysage transcriptionnel des mammifères est beaucoup plus complexe qu'on ne l'imaginait auparavant. En plus d'une gamme variée d'ARN codant pour les protéines et d'ARN régulateurs bien établis tels que les miARNs, différents types d'ARN non codants sont transcrits à partir de la grande majorité des régions non codantes du génome, y compris les séquences intergéniques et introniques (Djebali S et al. 2012). L'afflux récent d'énormes quantités de données sur les ARN-séquences a révélé un modèle d'expression génique différentiel entre divers types de tissus cancéreux et leurs homologues normaux, ce qui permet de découvrir les mécanismes moléculaires complexes permettant de comprendre la progression du cancer (Li M et al. 2017). Les énormes dépôts de données tels que la TCGA offrent la possibilité de ré-analyser les données omiques par une approche pan-cancéreuse où différents types de cancers peuvent être comparés et contrastés en termes de paysages génomiques, épigénomiques et transcriptomiques (Cancer Genome Atlas Research. 2013)

## 5 – L'intégration des données omiques

Le criblage à haut débit est l'une des principales technologies d'exploration de la dynamique intracellulaire complexe en biologie moderne (Ashraf MA and Sarfraz M. 2016). Les données génomiques sont apparues dans le cadre du projet du génome humain pour obtenir le schéma d'une information génétique humaine complète, ensuite le transcriptome et le protéome deviennent également disponibles pour mesurer l'abondance de l'expression de l'ARNm et de la protéine, respectivement (Vo TV et al. 2016). Dernièrement, l'épigénomique a été développée pour étudier la "matière noire" du génome (par exemple, les éléments régulateurs potentiels situés dans les séquences non codantes) (Pickersgill M et al. 2013 ; Romanoski CE et al. 2015). Outre la compréhension approfondie de l'association génotype-phénotype, les métabolites ont été largement utilisés pour établir un pont entre le génome et le phénotype en raison de leur rôle de régulation (Xu B et al. 2018). Parallèlement, les interactions entre les différentes molécules sont également confirmées et rassemblées dans des bases de données, qui fournissent les métadonnées sur les réseaux de molécules, appelés interactomes (Rolland T et al. 2014 ; Li Y et al. 2016). L'intégration est un concept efficace pour résoudre les problèmes complexes ou comprendre le système biologique compliqué (Rapport DJ et al. 2013). D'un point de vue informatique, l'intégration des données peut utiliser pleinement les informations complémentaires (Wen Y et al. 2016) fournir un résumé du facteur caché (Weischenfeldt J et

al. 2013) réaliser une correction de biais dans l'analyse (Shen R et al. 2012) et introduire un modèle de données commun et diversifié (Zeng T et al. 2014). Dans le domaine de la biologie, l'intégration des données est une étude multidimensionnelle sur l'exhaustivité et la complexité du système biologique. En particulier dans les études génomiques à haut débit sur le cancer, les résultats de l'analyse d'ensembles de données uniques souffrent souvent d'un manque de reproductibilité en raison de la petite taille des échantillons, et les études de référence ont révélé l'hétérogénéité et les compromis existant dans l'analyse des données omiques (Shi X et al. 2014 ; Shi X et al. 2015). Pour résoudre ces problèmes, l'analyse intégrative peut efficacement combiner et examiner de nombreux ensembles de données de manière rentable afin d'améliorer la reproductibilité.

En bref, les méthodes actuelles d'analyse intégrative sur les données biologiques ont deux modes : l'un est l'intégration ascendante (la combinaison des données avec une intégration manuelle de suivi), et l'autre est l'intégration descendante (la fusion des données avec une intégration *in silico* de suivi). Dans l'intégration ascendante la combinaison de grandes quantités de données publiques peut nous permettre d'examiner les relations dynamiques générales au cours de la régulation des gènes (Bebek G et al. 2012 ; Zhang S et al. 2012) par exemple, la combinaison de différents types de données fournit un modèle plus complet de la cellule cancéreuse que celui offert par un seul type. Ces analyses combinatoires devraient intégrer les diverses données pour reconstruire des réseaux biologiquement significatifs et fournir potentiellement un aperçu plus fiable des mécanismes biologiques sous-jacents (Hieke S et al. 2016). En revanche, dans l'intégration descendante, l'idée générale d'intégration est basée sur la fusion d'informations, où différents types de données peuvent offrir des perspectives complètes sur le même phénomène biologique. Les approches intégratives seraient plus puissantes lorsqu'elles peuvent incorporer tous les types de données simultanément et générer une seule affectation intégrée de groupes d'échantillons, comme les méthodes basées sur les statistiques (Kirk P et al. 2012 ; Mo Q et al. 2013 ; Huang YT et al. 2015 ; Wang W et al. 2013) les méthodes basées sur l'apprentissage automatique (Liu Y et al. 2013 ; Speicher NK and Pfeifer N. 2015 ; Narayanan M et al. 2010) et les méthodes basées sur les matrices (Zhang S et al. 2012 ; Le Van T et al. 2016).

## 6 – Avenir et défis des études multi-omiques

Les récents progrès de l'analyse bio-informatique des données omiques ont permis de découvrir l'ensemble du système cellulaire et démêler les mécanismes moléculaire de la cancérologie. Grâce à de puissants outils de calcul, il est possible d'identifier le lien entre les aberrations

génomiques et les ARNm, les protéines et les métabolites exprimés différemment qui sont associés à une perturbation cellulaire due au cancer. L'intégration des données multi-omiques fournit une plate-forme permettant de relier les altérations génomiques/épigénomiques aux réseaux interconnectés de transcriptomes, de protéomes et de métabolomes, qui sous-tend la réponse cellulaire à une perturbation. Cette vision offre une opportunité pour mieux comprendre la réponse cellulaire au niveau des systèmes, mais pose en même temps un défi pour la modélisation axée sur la biologie des systèmes. La prochaine phase de la recherche en biologie systémique se concentrera sur des modèles qui peuvent traiter de manière dynamique des milliers de changements d'ARNm, de protéines et de métabolites. L'approche de la biologie systémique a le potentiel de développer des stratégies efficaces pour administrer une thérapie personnalisée contre le cancer (Werner HMJ et al. 2014). L'objectif de l'approche de la biologie des systèmes est de développer des modèles prédictifs qui sont affinés et limités par des validations expérimentales. Ces modèles prédictifs seront particulièrement utiles pour sélectionner et stratifier les patients sur la base de données multi-omiques personnalisées afin de déterminer qui sont les plus susceptibles de bénéficier de thérapies ciblées (D. Guhathakurta et al. 2013). En bref, les modèles de biologie systémique basés sur les données multi-omiques peuvent contribuer à accroître l'efficacité des onco-médicaments et à surmonter le phénotype de résistance à la chimio/immunothérapie des cellules cancéreuses, ce qui les rend vulnérables aux thérapies ciblées et en fin à améliorer la qualité de vie des patients (Hillmer RA. 2015 ; Eicher T et al. 2020).

## CHAPITRE 3

L'objectif de ma thèse est avoir une caractérisation moléculaire complète pour les gliomes de bas grade y compris les astrocytomes et les oligodendrogliomes en identifiant des biomarqueurs moléculaire spécifiques pour chaque tumeur cérébrale sur le plan génomique, transcriptomique et épigénétique.

Ce chapitre est composé de quatre études, chaque étude traite un type de biomarqueur particulier.

Etude 1: Analyse computationnelle des mutations IDH1, IDH2 et TP53 dans les gliomes de bas grade, y compris les oligodendrogliomes et les astrocytomes.

Etude 2: Identification des biomarqueurs potentiels pour les gliomes de bas grade en analysant les profils génomiques et le phénotype CpG Island Methylator des patients dans la base de données TCGA.

Etude 3: Analyse bioinformatique des gènes différentiellement exprimés et des miARN dans les gliomes de bas grade.

Etude 4: Profil global des modifications épigénétiques impliquées dans les gliomes de bas grade.

## **Analyse computationnelle des mutations IDH1, IDH2 et TP53 dans les gliomes de bas grade, y compris les oligodendrogliomes et les astrocytomes.**

### **I - Introduction**

Les gliomes de bas grade sont un groupe de tumeurs cérébrales hétérogènes aux propriétés biologiques et cliniques distinctes; dans le passé, les sous-groupes et la classification des gliomes étaient définis par des caractéristiques histologiques. Cependant, cette méthode de classification ne reflète pas l'hétérogénéité tumorale, elle ne permet pas un diagnostic précis. Au cours des trois dernières années, de nouvelles découvertes de biomarqueurs moléculaires et de diagnostic chez les adultes atteints de gliomes de bas grade ont entraîné un changement dans la classification pathologique de tous les gliomes tels que les astrocytomes et les oligodendrogliomes (Laug D et al. 2018 ; The Cancer Genome Atlas Research Network. 2015). La classification actualisée de l'Organisation Mondiale de la Santé des tumeurs du système nerveux central en 2016 insiste sur le fait que l'évaluation diagnostique standard des gliomes de bas grade devrait désormais inclure une évaluation moléculaire des gènes impliqués, comme les mutations des isocitrates déshydrogénases IDH1 et IDH2 et le gène suppresseur de tumeur TP53 (Hoshida R et al. 2016 ; Gussyatiner O et al. 2018). L'IDH catalyse la conversion de l'isocitrate en alpha-cétoglutarate (KG), il est présent sous forme de trois isoformes: IDH1 situé dans le cytoplasme, IDH2 situé dans les mitochondries et IDH3 qui fait partie du cycle de l'acide tricarboxylique (Gussyatiner O et al. 2017). Les patients atteints de gliomes de bas grade, notamment les astrocytomes et les oligodendrogliomes, sont caractérisés par des mutations dans l'IDH1 en position R132 et IDH2 en position R172. Dans le cas muté, IDH1 et IDH2 catalysent la conversion de l'alpha-KG en bêta-hydroxyglutarate (2-HG); des niveaux supranormaux de 2-HG intracellulaire entraînent une hyperméthylation de gènes cibles tels que TET2 impliqués dans la différenciation cellulaire (Gussyatiner O et al. 2018 ; Huang J et al. 2019). Le gène TP53 de type sauvage a un rôle indispensable dans plusieurs processus cellulaires, notamment la suppression des tumeurs, l'apoptose, la réparation de l'ADN, l'autophagie, etc, mais il est souvent muté ou inactif au stade précoce des gliomes de bas grade. Dans environ 30% à 40% des cas d'astrocytomes, une perte de la fonction p53 peut être observée à un stade précoce. Des mutations de p53 dans certains résidus tels que le R158G peuvent perturber les propriétés structurelles et fonctionnelles de la protéine à des degrés divers et affecter le pronostic des patients (Sarma PP et al. 2017).

## **II- Objectifs**

- Identifier les mutations des gènes impliqués dans les gliomes de bas grade IDH1, IDH2 et TP53.
- Étudier l'effet de ces mutations sur la fonction des protéines.
- Étudier l'effet de ces mutations sur la structure des protéines.

## **III – Matériel et méthodes**

### **1- Déclaration éthique**

Cette étude a été approuvée par le Comité d'éthique de la recherche biomédicale de l'Université Mohammed V aux motifs suivants : conditions satisfaisantes de validité, pertinence scientifique, intérêt de recherche, pertinence éthique, conditions satisfaisantes de protection des personnes, intelligibilité de la note d'information et conformité des modalités de recueil du consentement. Le comité d'éthiques pour la recherche biomédicale se base ses délibérations sur :

- La loi N° 28-13 relative à la protection des personnes participants aux recherches biomédicales (B.O.N° 9396 du 17 septembre 2015).
- La déclaration d'Helsinki (version 2008).
- Les lignes directrices internationales d'éthiques pour la recherche biomédicale sciences médicales (CIOMS - version 2002).
- La directive 2001/20/CE du parlement européen et du conseil du 4 avril 2001 concernant le rapprochement des dispositions législatives, réglementaires et administratives des états membres relatives à l'application de bonnes pratiques cliniques dans la conduite d'essais cliniques de médicaments à usage humain.
- La loi française fixant le règlement intérieur type devant être adopté par les comités de protection des personnes (arrêté du 13 janvier 2010).
- Le règlement intérieur du CERB.
- La réglementation marocaine concernant l'exercice de la médecine.
- Les spécificités socioculturelles et religieuses du Maroc.

Ce comité d'Éthique pour la Recherche Biomédicale de rabat est enregistré auprès de l'office de la recherche et la protection humaine, département de la santé et des services sociaux des États-Unis sous le n°IORG0006594.

## 2- Conception de l'étude et collecte d'échantillons

Cette étude a été réalisée sur 32 échantillons, dont 24 tumeurs et 8 contrôles. Pour les 24 tumeurs, nous avons 20 blocs de tissus inclus en paraffine (14 astrocytomes et 6 oligodendrogliomes) et 4 échantillons de sang (2 astrocytomes et 2 oligodendrogliomes). Pour les 8 échantillons de contrôle, nous avons 6 blocs de tissus inclus en paraffine appartenant aux mêmes patients qui ont une tumeur (On a coupé la partie saine de chaque patient, 3 astrocytomes et 3 oligodendrogliomes), et 2 échantillons de sang indépendants des autres patients. Entre 2013 et 2017, 20 blocs tissulaires de patients inclus en paraffine ont été collectés rétrospectivement au département d'anatomie pathologique, Hôpital spécialisé, CHU Ibn Sina, Rabat. 5 ml de sang périphérique ont été collectés dans des tubes contenant de l'acide Éthylènediaminetétraacétique (EDTA) de 4 patients d'une manière prospective au département de neurochirurgie, Hôpital spécialisé, CHU Ibn Sina, Rabat, le consentement éclairé a été remplis par les patients suivant le protocole approuvés par le comité d'Éthique pour la Recherche Biomédicale de Rabat.

## 3 - Préparation d'ADN, amplification des gènes IDH1, IDH2 et TP53 et séquençage

26 (20 patients, 6 témoins) blocs de tissus inclus en paraffine (FFPE) fixés au formol ont été préparés par un microtome semi-automatique (HM 340E; Thermo Fisher Scientific), des coupes de 8  $\mu$ M ont été coupées séquentiellement à partir de chaque bloc, et chaque section a été collectée sur une diapositive séparée. Pour chaque lame, 25 mg de tissu ont été raclés puis placés dans un tube de 1,5 ml pour extraire l'ADN en utilisant la méthode phénol-chloroforme. Le déparaffinage a été effectué avec du xylène 1 ml; après cela, plusieurs lavages à l'éthanol (1 ml) ont été effectués. La lyse cellulaire a été réalisée par l'addition de 300  $\mu$ L de tampon d'extraction. La digestion et l'élimination des protéines ont été effectuées, en ajoutant 30  $\mu$ L de protéinase K (10 mg / ml) et 300  $\mu$ L de phénol-chloroforme respectivement; 250  $\mu$ L d'acétate d'ammonium ont été utilisés pour précipiter l'ADN. Concernant les 6 échantillons de sang, le kit MyTaq Blood-PCR (BIO-25054) a été utilisé directement pour effectuer une réaction en chaîne par polymérase (PCR) sans l'étape d'extraction d'ADN. La quantification de l'ADN a été évaluée avec le système QuantiFluor dsDNA (Promega E4871); la qualité d'ADN des échantillons a également été évaluée par électrophorèse. Pour identifier les mutations IDH1, IDH2 et TP53, les régions génomiques englobant l'exon 4 d'IDH1 et IDH2 et les 4 exons (5-6 et 7-8) de TP53 ont été conçus par Primer3 (PCR primer design tool version 2.0). Les séquences d'amorces suivantes ont été utilisées: IDH1 forward, 5'-TGATGAGAAGAGGGTTGAGGA-3' et reverse, 3'-ATCCCCATAAGCATGACGAC-5'; IDH2 forward, 5'-CAAGCTGAAGAAGA

TGTGGAA-3' et reverse, 3'-CAGAGACAAGAGGATGGCTA-5'; 5ème et 6ème exons de TP53 forward, 5'-AATGGTTCCTGAAGACCCA-3' et reverse, 3'-AAAAGTTGACACGTTATTCAATT-5'; et 7ème et 8ème exons de TP53 forward, 5'-TAGGTTGGCTCTGACTGTA-3' et reverse, 3'-AAGTTGAATGTTATAAAAGTT-5'. Environ 5 à 50 ng d'ADN ont été utilisés pour effectuer la PCR (SimpliAmp Thermal Cycler A24811) avec 1 µL de chaque amorce et 12,5 µL de MyTaq Mix (Bioline, BIO-25041). Les conditions de PCR pour toutes les séquences sont les suivantes: 30 cycles avec dénaturation à 95 ° C pendant 30 secondes; hybridation pendant 40 secondes à 51 ° C, 50 ° C, 52 ° C et 54 ° C pour IDH1, IDH2, 5ème et 6ème exons de TP53, et 7ème et 8ème exons de TP53, respectivement; et l'extension finale à 72 ° C pendant 40 secondes.

Les produits de PCR ont été purifiés en utilisant le produit de PCR ExoSAP-IT Express (75001.1.ML; Thermo Fisher Scientific). La réaction de séquençage a été soumise à 25 cycles d'amplification consistant en une dénaturation à 96 ° C pendant 10 secondes, une hybridation à 50 ° C pendant 5 secondes et une élongation à 60 ° C pendant 4 minutes dans un volume total de 10 µL. Le produit purifié est séquencé à l'aide du kit de séquençage de cycle BigDye X-Terminator v3.1 (Numéro de catalogue 4376486; Applied Biosystems) conformément au protocole du fournisseur.

#### 4 - Analyse computationnelle

SnapGene viewer software version 4.1.9 a été utilisé pour visualiser les électrophérogrammes de séquence générés par le séquenceur 'applied biosystems 3500xl genetic analyzer' (GSL Biotech; available at [snapgene.com](http://snapgene.com)). La plupart des variations génétiques humaines sont représentées par des polymorphismes mononucléotidiques (SNP), et on pense que de nombreux SNP provoquent des différences phénotypiques entre les individus humains. Nous nous concentrons spécifiquement sur les SNP non synonymes (nsSNPs), c'est-à-dire les SNP situés dans des régions codantes et entraînant des changements d'acides aminés dans les produits protéiques des gènes. Pour comprendre les effets des mutations trouvées sur la structure et la fonction de la protéine, nous avons utilisé PolyPhen-2 version 2.0.2 (Adzhubei I et al. 2013). On a utilisé 8 fonctions prédictives basées sur la séquence et 3 fonctions prédictives basées sur la structure, qui ont été sélectionnées automatiquement par un algorithme glouton itératif. La majorité de ces caractéristiques impliquent la comparaison d'une propriété de l'allèle de type sauvage et de la propriété correspondante de l'allèle mutant (dérivé, causant la maladie), qui définissent ensemble un remplacement d'acide aminé. Les caractéristiques les plus informatives

caractérisent dans quelle mesure les deux allèles humains s'intègrent bien dans le modèle de remplacements d'acides aminés dans l'alignement à séquences multiples de protéines homologues, à quelle distance la protéine abritant la première déviation de l'allèle de type sauvage humain est de la protéine humaine, et si l'allèle mutant provenait d'un site hyper mutable. On a sélectionné l'ensemble de séquences homologues pour l'analyse à l'aide d'un algorithme de regroupement, pour effectuer un alignement multiple. La signification fonctionnelle d'un remplacement d'allèle est prédite à partir de ses caractéristiques individuelles par le classificateur Naïve Bayes. Le logiciel utilisé deux paires d'ensembles de données pour entraîner et tester les fichiers d'entrée. Nous avons compilé la première paire, HumDiv, à partir des 3155 allèles dommageables ayant des effets connus sur la fonction moléculaire à l'origine des maladies mendéliennes humaines, présentes dans la base de données UniProt, ainsi que 6321 différences entre les protéines humaines et leurs homologues mammifères étroitement apparentés, supposés être non-dommageable. La deuxième paire, HumVar, comprend toutes les 13 032 mutations humaines causant des maladies d'UniProt, ainsi que 8 946 nsSNP humains sans implication annotée dans la maladie, qui ont été traitées comme non dommageables (Yang J et al. 2014).

Pour une mutation, PolyPhen-2 calcule la probabilité postérieure Naïve Bayes que cette mutation est dommageable et rapporte des estimations de faux positif (la probabilité que la mutation soit classée comme dommageable alors qu'elle est en fait non dommageable) et de vrai positif (la probabilité que la mutation est classée comme dommageable alors qu'elle est effectivement dommageable) (Figure 3) (Kumar P et al. 2018) .

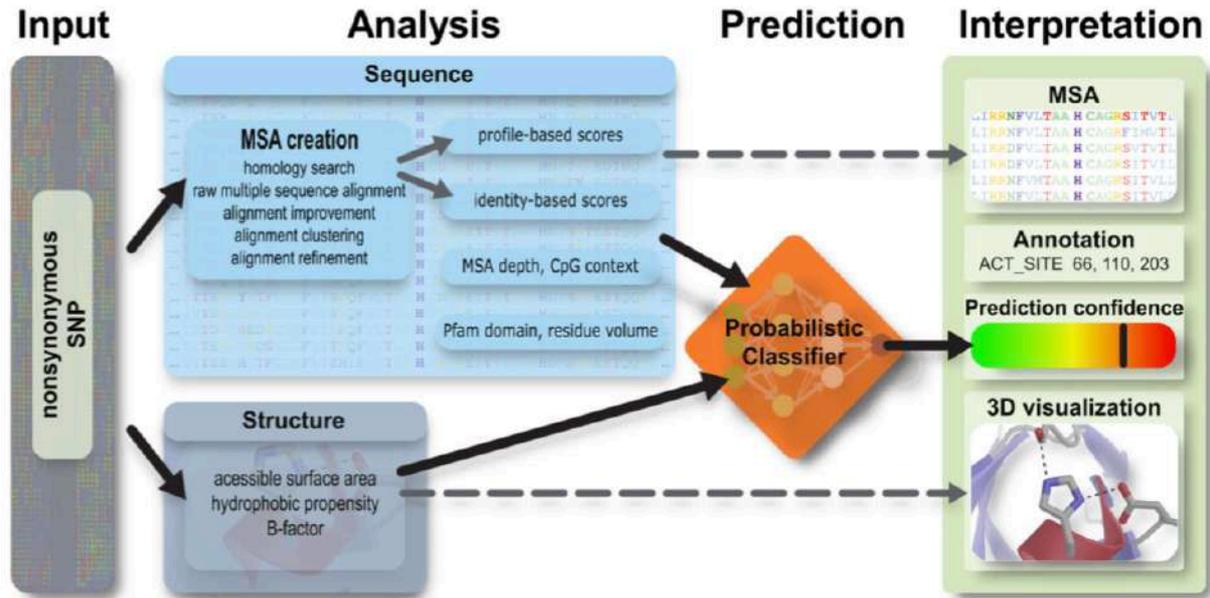


Figure 3 : L'algorithme de prédiction de polyphen 2

Iterative Threading ASSEmbly Refinement server (I-TASSER) version 5.1 a été utilisé pour modéliser les structures protéiques qui contiennent toutes les mutations: R132H dans IDH1, R172M dans IDH2 et R158G, R175H et K305N dans TP53. Le protocole I-TASSER constitue trois étapes: l'identification du modèle structurel, l'assemblage de la structure itérative et l'annotation des fonctions basées sur la structure. En partant de la séquence d'acides aminés, on identifie d'abord les modèles de structure homologues à partir de la bibliothèque PDB à l'aide de LOMETS, un algorithme de méta-threading qui se compose de plusieurs programmes de threading individuels. La topologie des modèles pleine longueur est ensuite construite en réassemblant les structures de fragments alignées en continu excisées des modèles LOMETS et des segments de structure super-secondaires. Les conformations d'énergie libre la plus basse sont identifiées par SPICKER à travers le regroupement des trajectoires de simulation de Monte Carlo. À partir des clusters SPICKER, un deuxième cycle de réassemblage de la structure est effectué pour affiner les modèles structurels, les conformations à faible énergie libre étant affinées par des simulations atomiques complètes en utilisant FG-MD et ModRefiner qui sont des algorithmes pour le raffinement de la structure des protéines à haute résolution (Zhang W et al. 2016 ; Yang J and Zhang Y. 2015).

Pour dériver la fonction biologique des protéines cibles, les modèles I-TASSER sont mis en correspondance avec les protéines de la bibliothèque BioLiP (Yang J et al 2015), qui est une base de données de fonction protéique. Les informations fonctionnelles, y compris la liaison de ligand, la commission enzymatique et l'ontologie des gènes, sont déduites par les modèles BioLiP qui sont classés en fonction d'un paramètre de notation composite combinant la similitude structurelle globale et locale, la conservation des caractéristiques chimiques et les alignements de profils de séquence (Yang J et al. 2016 ; Wang Y et. 2017 ; Zhang Y. 2007).

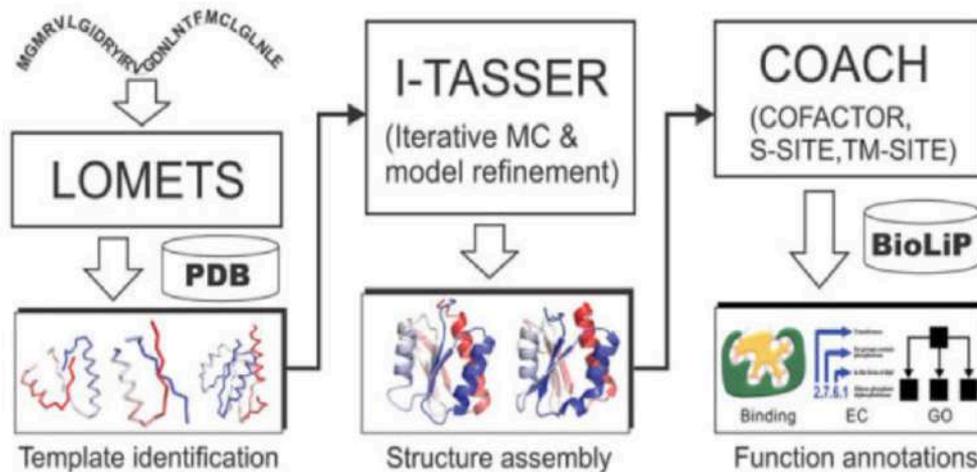


Figure 4 : Le protocole I-TASSER pour l'identification de la structure et de la fonction des protéines

Le score de confiance de chaque modèle (score C), est mesurée quantitativement en fonction de l'importance des alignements de modèles de filetage et des paramètres de convergence des simulations d'assemblage de structure. Le score de modélisation (TM-score), la déviation quadratique moyenne des positions atomiques et le facteur B normalisé sont estimés sur la base du score C et de la longueur de la protéine suivant la corrélation observée entre ces qualités. La validation structurelle du modèle protéique a été réalisée par le programme PROCHECK 3.5.4 dont le but est d'évaluer dans quelle mesure la géométrie des résidus dans une structure protéique donnée est normale, ou au contraire inhabituelle, par rapport aux paramètres stéréochimiques dérivés de structures à haute résolution bien raffinées (Figure 4).

Le modèle protéique a été validé à l'aide du tracé de Ramachandran qui offre une vue simple de la conformation d'une protéine et affiche également une visualisation des régions énergétiques, pour lesquelles nous trouvons le pourcentage de chaque région: la région favorisée, la région autorisée et la région aberrante. Ces régions doivent être incluent dans un

intervalle spécifique pour la validation de la structure protéique (Chinthakunta N et al. 2018 ; Sunil L et al. 2017).

La compaction protéique a été décrite en utilisant le numéro de contact pondéré (WCN), et les scores de conservation spécifiques à la séquence qui ont été calculés en suivant le protocole de ConSurf, qui est un outil bio-informatique pour estimer la conservation évolutive du résidu dans la protéine basée sur les relations phylogénétiques entre les séquences homologues. Le degré auquel une position d'acide aminé (ou nucléique) est conservée dans l'évolution (c'est-à-dire sa vitesse d'évolution) dépend fortement de son importance structurelle et fonctionnelle. Ainsi, l'analyse de conservation des positions parmi les membres d'une même famille peut souvent révéler l'importance de chaque position pour la structure ou la fonction de la protéine. La réciproque du profil WCN (rWCN) est utilisée pour la comparaison avec le profil de conservation de séquence (Pereira GRC et al. 2019 ; Huang TT et al. 2015). A la fin, la visualisation de la structure tridimensionnelle (3D) a été effectuée à l'aide de PyMOL version 2.3.1 (Rigsby RE et al. 2016).

#### **IV - Résultats**

Tous les échantillons inclus dans l'étude ont été anonymisés et aucune information sur l'identité d'un individu n'était disponible pendant l'analyse. Une mutation R132H du gène IDH1 a été détectée dans 11 des 32 échantillons, dont 9 astrocytomes et 2 oligodendrogliomes (Tableau 1). La mutation R172M dans IDH2 a été détectée dans 1 oligodendrogliome et les mutations R158G, R175H et K305N du gène TP53 ont été identifiées dans 5 astrocytomes (Tableau 1).

<b>Samples</b>	<b>Tumors/Controls</b>	<b>Type of tumor</b>	<b>IDH1</b>	<b>IDH2</b>	<b>TP53</b>
1	Tumor	FFPE Astrocytoma	R132H	WT	R175H
2	Tumor	FFPE Astrocytoma	R132H	WT	K305N R158G
3	Tumor	FFPE Astrocytoma	R132H	WT	R175H
4	Tumor	FFPE Astrocytoma	R132H	WT	K305N
5	Tumor	FFPE Astrocytoma	R132H	WT	WT
6	Tumor	FFPE Astrocytoma	R132H	WT	K305N
7	Tumor	FFPE Astrocytoma	R132H	WT	K305N
8	Tumor	FFPE Astrocytoma	R132H	WT	R175H
9	Tumor	FFPE Astrocytoma	R132H	WT	R158G
10	Tumor	FFPE Astrocytoma	WT	WT	R158G
11	Tumor	FFPE Astrocytoma	WT	WT	R158G
12	Tumor	FFPE Astrocytoma	WT	WT	R158G
13	Tumor	FFPE Astrocytoma	WT	WT	K305N
14	Tumor	FFPE Astrocytoma	WT	WT	WT
15	Tumor	FFPE Astrocytoma	WT	WT	WT
16	Tumor	Blood Astrocytoma	WT	WT	R175H
17	Tumor	Blood Astrocytoma	WT	WT	WT
18	Tumor	FFPE Oligodendroglioma	R132H	WT	WT
19	Tumor	FFPE Oligodendroglioma	R132H	WT	WT
20	Tumor	FFPE Oligodendroglioma	WT	R172M	WT
21	Tumor	FFPE Oligodendroglioma	WT	WT	WT
22	Tumor	FFPE Oligodendroglioma	WT	WT	WT
23	Tumor	Blood Oligodendroglioma	WT	WT	WT
24	Tumor	Blood Oligodendroglioma	WT	WT	WT
25	Control	FFPE Astrocytoma	WT	WT	R175H
26	Control	FFPE Astrocytoma	WT	WT	WT
27	Control	FFPE Astrocytoma	WT	WT	WT
28	Control	FFPE Oligodendroglioma	WT	WT	WT
29	Control	FFPE Oligodendroglioma	WT	WT	WT
30	Control	FFPE Oligodendroglioma	WT	WT	WT
31	Control	Blood	WT	WT	WT
32	Control	Blood	WT	WT	WT

*Tableau 1 : Mutations des gènes IDH1, IDH2 et TP53*

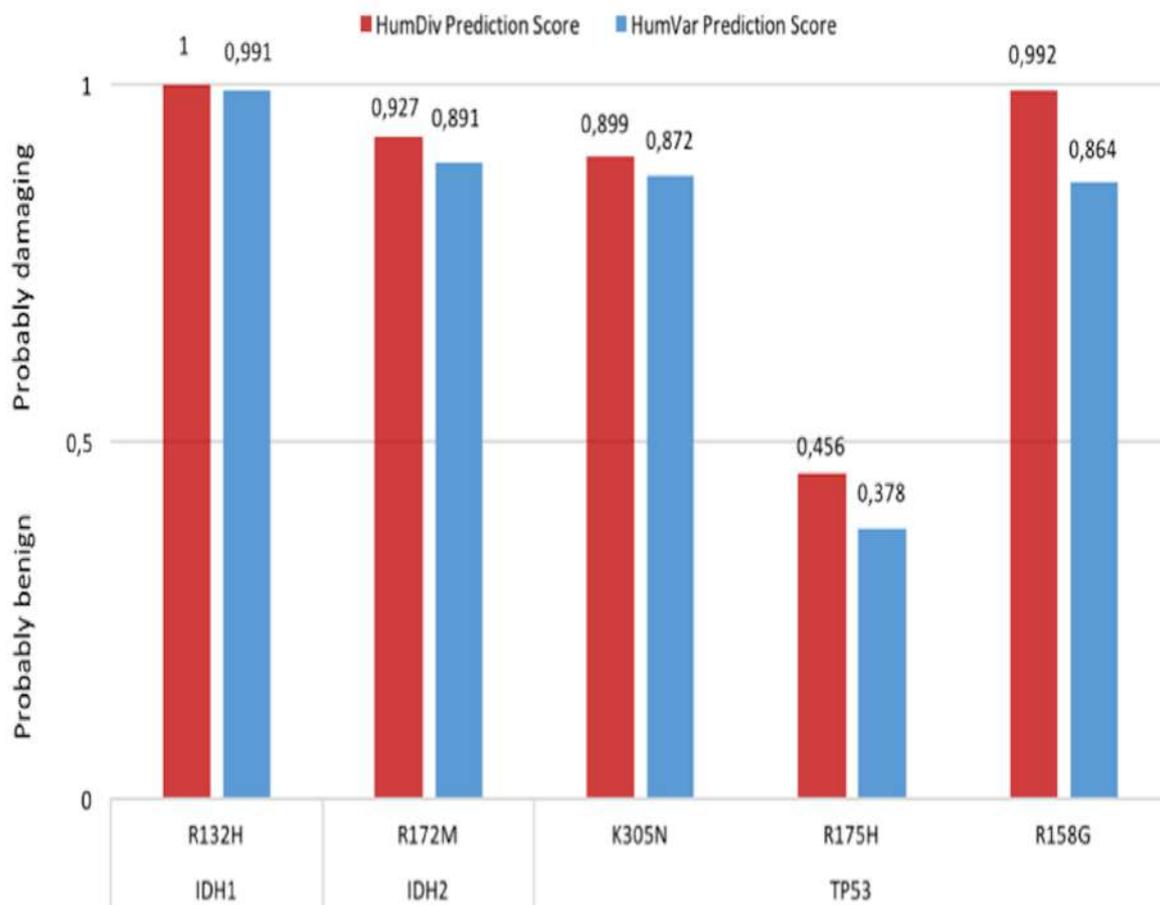
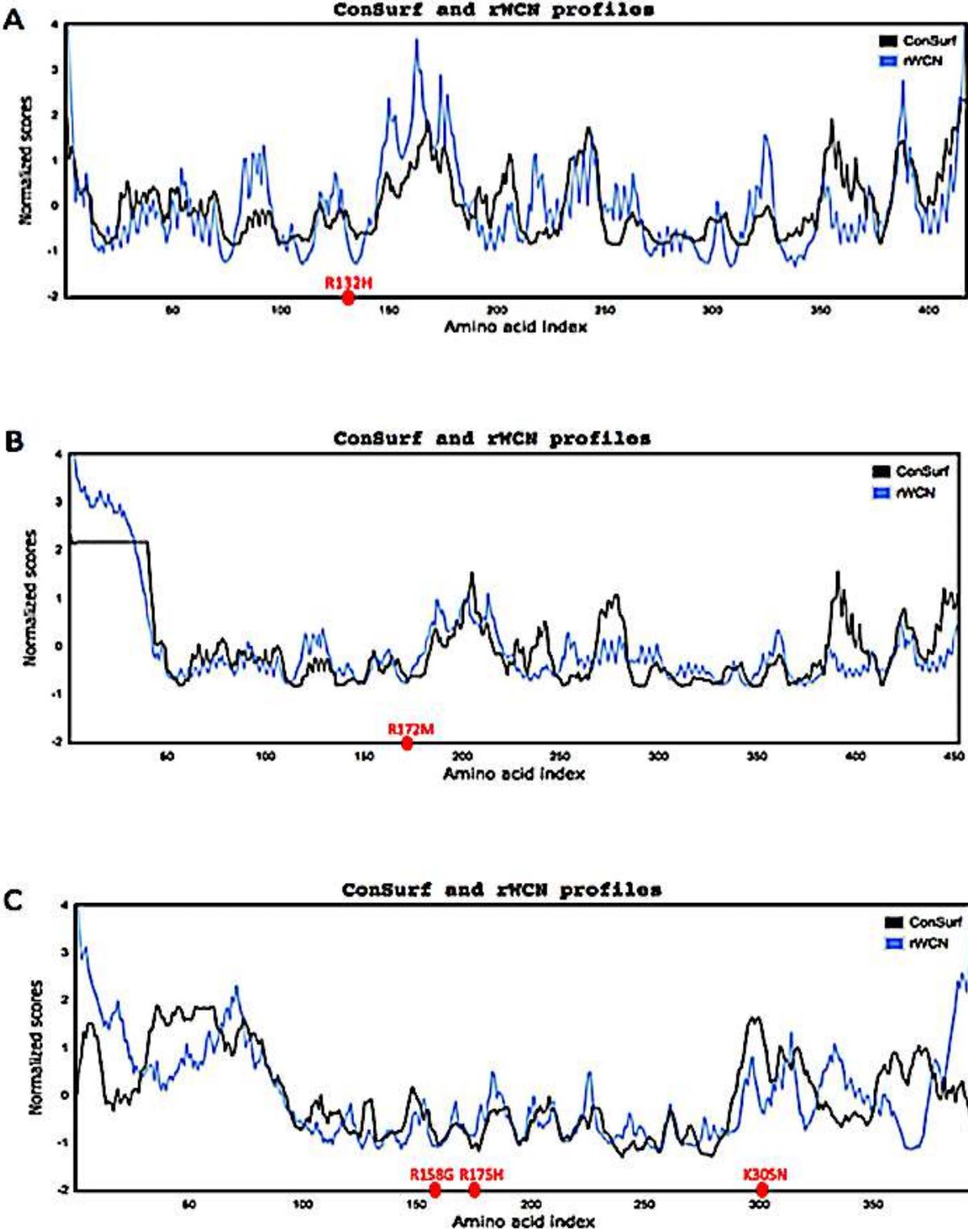


Figure 5 : Prédiction des effets possibles des mutations IDH1, IDH2 et TP53

La figure 5 montre la pathogénicité des substitutions d'acides aminés au niveau des gènes IDH1, IDH2 et TP53 prédites par PolyPhen-2. Le score PolyPhen-2 représente la probabilité qu'une substitution soit dommageable, avec un score allant de 0,0 (toléré) à 1,0 (délétère). PolyPhen-2 utilise 2 bases de données pour former le modèle complet. La première est HumDiv, qui est utilisée pour évaluer les allèles rares de loci potentiellement impliqués dans des phénotypes complexes et les mutations nuisibles ayant des effets connus sur les maladies dans les bases de données UniProtKB. Le second est HumVar, qui contiennent tous les polymorphismes mononucléotidiques humains communs et les protéines non dommageables. Toutes les mutations IDH1 R132H, IDH2 R172M et TP53 K305N, R158G et R175H présentent un score supérieur à 0,5, de sorte que ces variantes devraient être dommageables ; cependant, R175H dans p53 est une mutation bénigne.

Protein Structure Prediction Server (PS) 2 version 3.0 a été utilisé pour calculer les scores de conservation. La série de scores de conservation d'une séquence est appelée profil de

conservation. En particulier, dans le profil de conservation d'une protéine, le résidu qui a un score de conservation inférieur, est plus conservé que celui d'un score de conservation élevé.



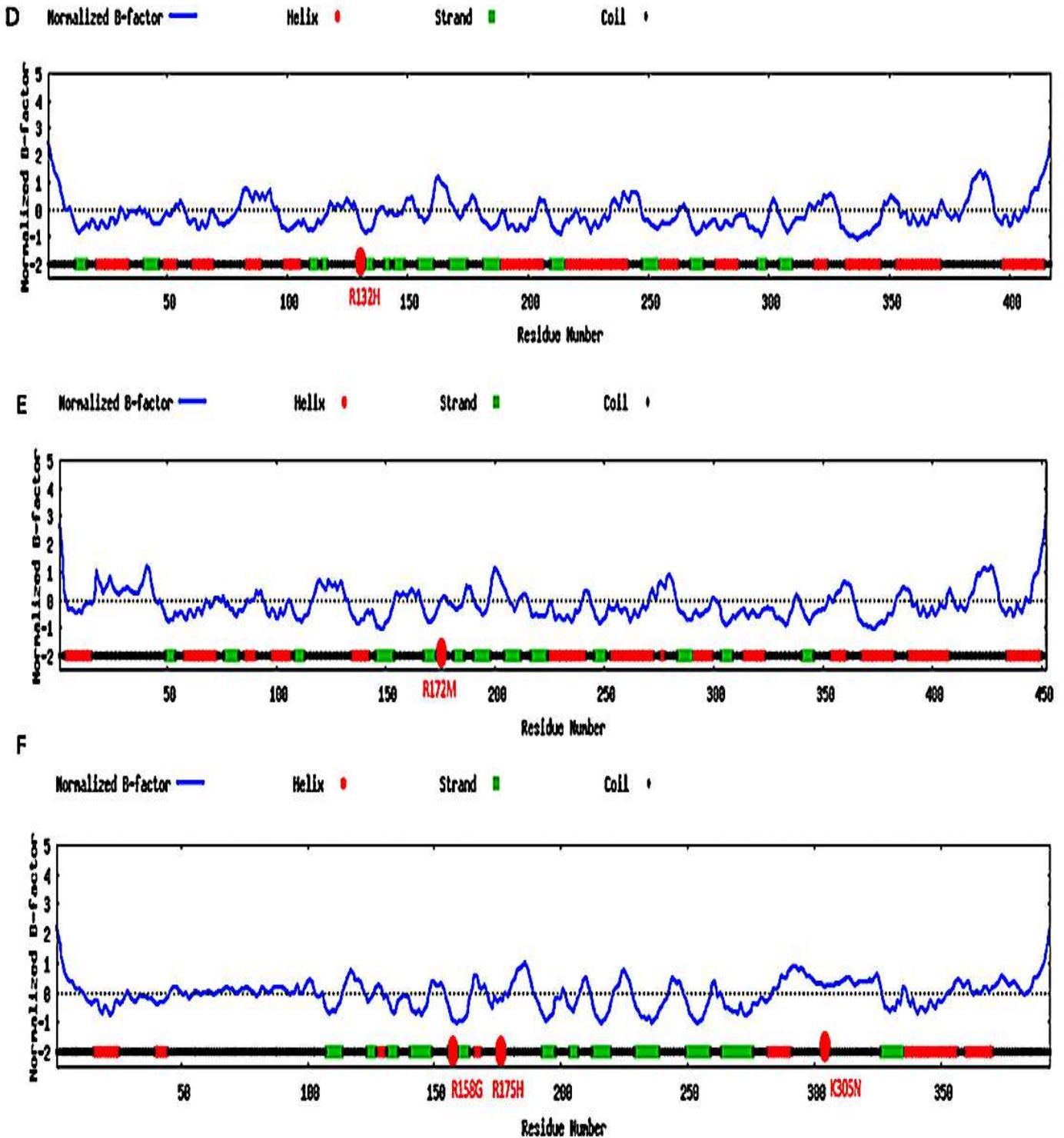


Figure 6 : Le profil rWCN (ligne bleue) et le profil ConSurf (ligne noire) de la protéine IDH. Les gènes mutés (A) IDH1, (B) IDH2 et (C) TP53 sont marqués dans un cercle rouge. Le rWCN et les scores de conservation sont normalisés à leurs scores respectifs. Profils normalisés du facteur B d'I-TASSER pour (D) IDH1, (E) IDH2 et (F) TP53

En effet, la figure 6 montre que le profil rWCN et le profil de conservation se chevauchent extrêmement bien dans les protéines IDH1, IDH2 et p53. Ici, nous montrons que parmi les

mutations pathogènes, ces résidus (IDH1 R132H, IDH2 R172M et p53 R175H) sont conservés et situés dans des environnements plus compacts, il est donc raisonnable de supposer que ces mutations sont fortement associées à la fonctionnalité et la structure de la protéine. En revanche, la mutation K305N au niveau de la protéine p53 (figure 6C) est situé dans une zone moins préservée car il a un score positif. L'autre partie (figure 6D à F) présente une valeur pour indiquer l'étendue de la mobilité thermique inhérente des résidus / atomes dans les protéines, ainsi que la précision locale des résidus dans le brin, l'hélice et la bobine. Les mutations R132H IDH1 et R158G p53 se trouvent dans la partie brin, et les mutations R172M IDH2 , R175H et K305N P53 se trouvent dans la position de la bobine. Pour estimer la qualité des modèles prédits par I-TASSER, le score C est calculé en fonction de l'importance des alignements de modèles de filetage et des paramètres de convergence de la structure. Il est typiquement de l'ordre de  $[-5, 2]$ : le score C d'IDH1 est de 1,31, d'IDH2 est de -0,84 et de p53 c'est -2,58. Pour le TM-score, qui est une échelle récemment proposée pour mesurer la similitude structurelle entre les structures, ce seuil ne dépend pas de la longueur des protéines. Un TM-score > 0,5 indique un modèle de topologie correcte comme IDH1 TM =  $0,90 \pm 0,06$ , IDH2 TM =  $0,61 \pm 0,14$  et p53 TM =  $0,42 \pm 0,14$ ; cependant, dans le cas où le score TM est < 0,17, il existe une similitude aléatoire.

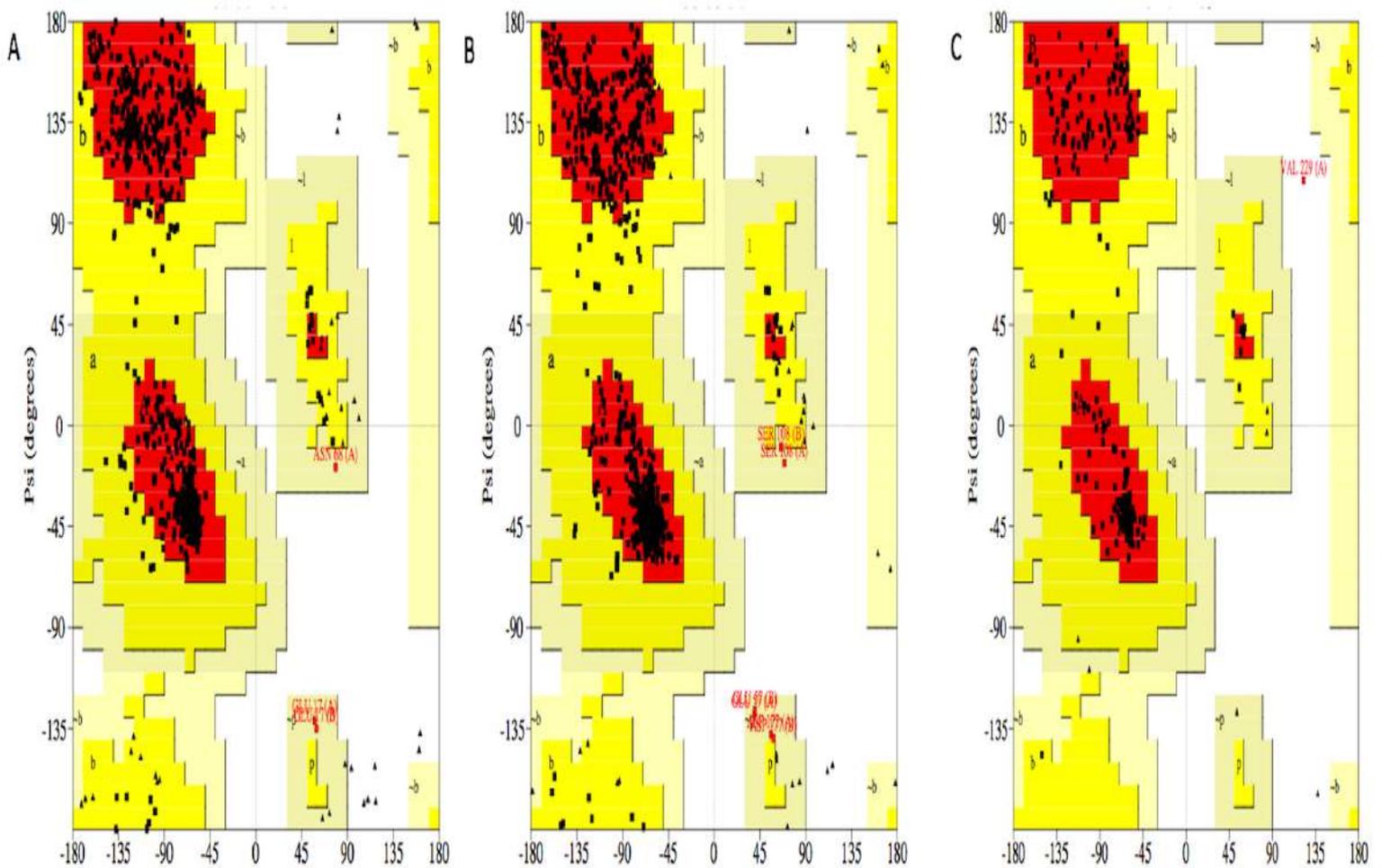


Figure 7 : Diagramme de Ramachandran des modèles (A) IDH1, (B) IDH2 et (C) p53. La région centrale, la région autorisée et la région générale sont colorées respectivement en rouge, jaune et beige

Les protéines modélisées sont validées par le tracé de Ramachandran généré par Procheck. La valeur du tracé de Ramachandran pour IDH1 (figure 7A) s'est avérée être de 91,8% avec 628 résidus se trouvant dans la région favorisée, 7,7% des résidus se trouvant dans la région autorisée et 0,4% se situant dans la région aberrante. Aucun résidu n'est localisé dans la région disqualifiée. Le nombre de résidus glycine est de 58 et le nombre de résidus proline est de 24. La figure 7B montre que l'IDH2 avec 647 résidus (91,9%) se trouvant dans la région favorisée, 7,2% des résidus se trouvant dans la région autorisée et 0,9% se trouvant dans la région aberrante. Aucun résidu n'est localisé dans la région disqualifiée. Le nombre de résidus glycine est de 66 et le nombre de résidus proline est de 30. Pour p53, la figure 7C montre que 202 résidus (93,1%) se trouvant dans la région favorisée, 6,5% des résidus se trouvant dans la région autorisée et 0% se trouvant dans la région aberrante. Environ 0,5% seulement des résidus se

trouvaient dans la région disqualifiée. Le nombre de résidus glycine est de 15 et le nombre de résidus proline est de 9.

En fin, les modèles générés par I-TASSER au format PDB ont été visualisés en utilisant PyMOL (Figure 8).

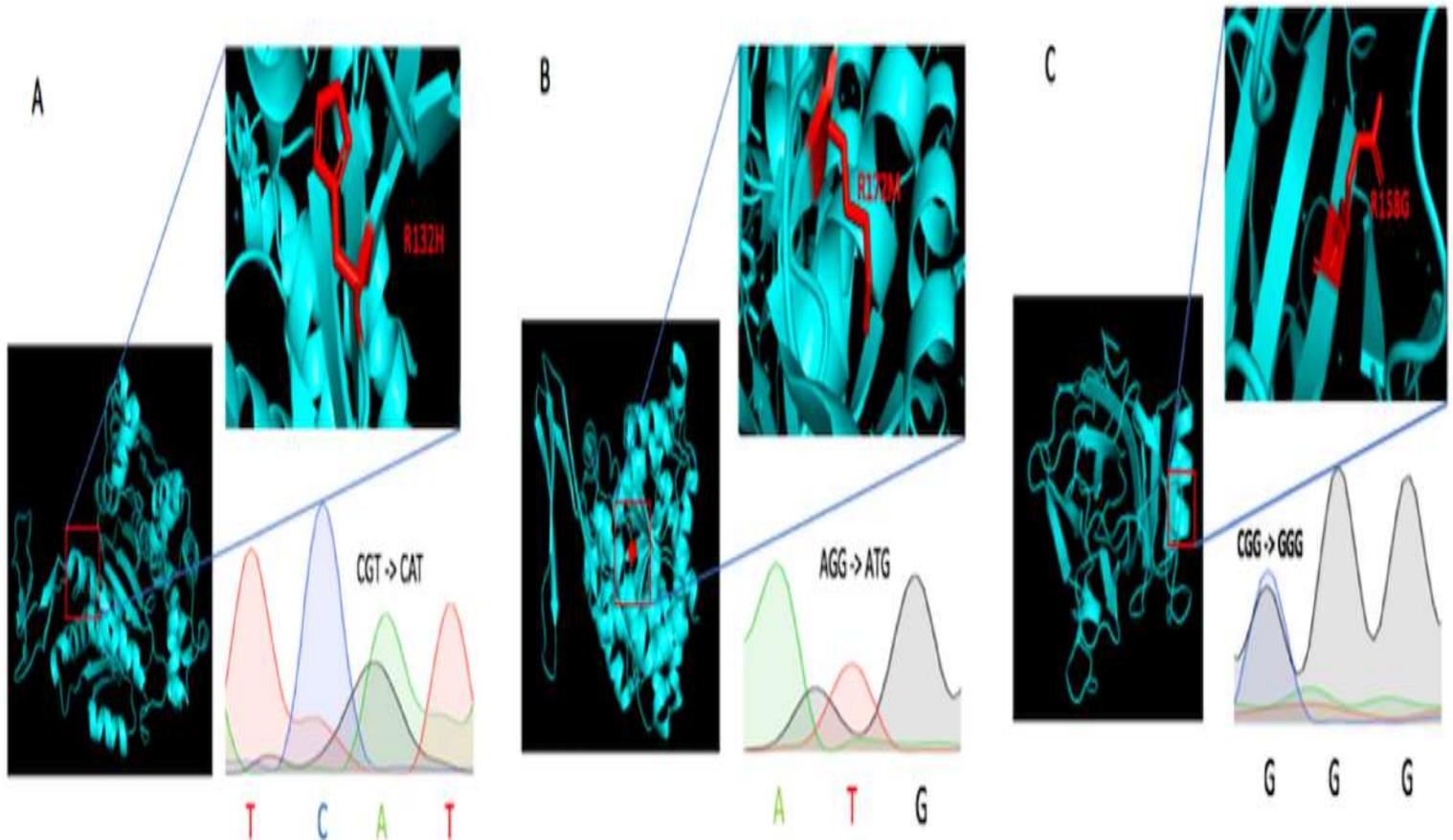


Figure 8 : Chromatogrammes des séquences des gènes IDH1, IDH2 et TP53 mutés et analyse structurale des protéines portant les mutations (A) R132H, (B) R172M (B) et (C) R158G

## V - Discussion

Récemment, de nouveaux biomarqueurs impliqués dans les gliomes de bas grade ont conduit à un changement de classification pathologique.

De plus, plusieurs études ont étudié par l'approche computationnelle l'impact des mutations dans IDH1 comme I147S, V444A et D375Y et dans IDH2 comme N439D et R140G (Thirumal Kumar D et al. 2018 ; Borger DR et al. 2012) dans notre étude, nous avons pu identifier quelques substitutions d'acides aminés différentes dans les protéines IDH1, IDH2 et p53, telles

que IDH1 R132, IDH2 R173M et p53 R175H, R158G et K305N (Adzhubei I et al. 2013 ; Wang S et al. 2017).

Les mutations IDH (figures 7A et B) survenant dans le site catalytique sont susceptibles de provoquer une perte d'activité catalytique; ces gènes mutés codent pour D2HG qui est proposé comme un onco-métabolite dans le cytoplasme pour IDH1 et dans les mitochondries pour IDH2; il inhibe de manière compétitive les enzymes dépendantes du KG, y compris la famille TET de 5-méthylcytosine hydroxylases et la famille JmjC d'histones lysine déméthylases, entraînant une différenciation cellulaire (Avellaneda Matteo D et al. 2017 ; Horbinski C et al. 2009). (Thirumal et al. 2018) révèlent que la thérapeutique médicamenteuse pour La D-2-hydroxyglutarate déshydrogénase est très limitée; par conséquent, la compréhension de la nature de la structure moléculaire causée par ces mutations servira pour le développement de nouvelles cibles pour une nouvelle thérapie médicamenteuse pour la D-2-hydroxyglutarate. Concernant la mutation de R158G p53 (figure 7C), la substitution de l'arginine par la glycine en position 158 de p53 affecte la conformation structurelle de la protéine et peut empêcher la liaison à l'ADN (Sarma PP et al. 2017 ; Modrek AS et al. 2017). Ce changement de résidu est très fréquent dans les cancers humains tels que les gliomes. (Anasuya Pal et al. 2015) ont étudié l'implication des mutations TP53 dans le cancer du sein et ont constaté que la plupart des mutations ponctuelles TP53 se produisent dans le domaine de liaison à l'ADN et peuvent être classées comme des mutations structurelles, telles que R248W, Y234C et H179R.

Dans notre étude, la plupart de ces mutations (IDH1 R132, IDH2 R173M et p53 R175H, R158G et K305N) ont été identifiées dans des échantillons de gliome de grade II et III, suggérant un rôle synergique possible dans la gliomagenèse.

Cependant, (Borger et al. 2012) ont identifié pour la première fois une fréquence élevée de mutations dans les gènes IDH1 et IDH2 dans les cholangiocarcinomes spécifiquement d'origine intrahépatique, ce qui indique que l'IDH1 et l'IDH2 sont impliqués dans d'autres types de cancer. L'analyse informatique a montré que les mutations de l'IDH1, l'IDH2 et R158G, K305N de p53 et étaient probablement pathogènes; au contraire, la substitution R175H est une mutation bénigne (Figure 5). Ces résultats ont été calculés par PolyPhen-2, qui est un outil de prédiction de l'impact possible de la substitution des acides aminés sur la fonction d'une protéine humaine (Figure 5). Cependant, dans d'autres études tel que (Thirumal et al. 2018), ils ont utilisé l'outil SIFT pour prédire la pathogénicité de la mutation trouvée.

Il a été démontré dans plusieurs études que l'impact des variants alléliques d'acides aminés sur la structure et la fonction des protéines peut être prédit par l'analyse d'alignements de séquences multiples et de structures 3-D de protéines. Comme nous l'avons démontré précédemment, ces

prédictions sont en corrélation avec un excès d'allèles rares parmi les allèles prédits fonctionnels (Robert F, Pelletier J. 2018). Par conséquent, les prédictions au niveau moléculaire révèlent des SNP affectant les phénotypes réels. PolyPhen-2 est un outil automatique de prédiction de l'impact possible d'une substitution d'acide aminé sur la structure et la fonction d'une protéine humaine, les prédictions automatisées de ce type sont essentielles pour interpréter de grands ensembles de données de variantes génétiques rares, qui ont de nombreuses applications dans la recherche moderne en génétique humaine. Les utilisations dans la recherche récente incluent l'identification des allèles rares qui causent la maladie mendélienne (Bamshad MJ et al 2011), la recherche d'allèles potentiellement médicalement exploitables dans le génome d'un individu et le profilage du spectre des variations rares découvertes par des séquençage de grandes populations (Tennessen JA et al 2012 ; Adzhubei I et al 2010). La prédiction est basée sur un certain nombre de caractéristiques comprenant la séquence et les informations phylogénétiques et structurales caractérisant la substitution. Pour une substitution d'acides aminés donnée dans une protéine, nous avons extrait diverses caractéristiques basées sur la séquence et la structure du site de substitution et on les envoyé vers un classificateur probabiliste. Des effets précoces d'un mauvais pronostic des mutations du hot spot p53 ont été mis en évidence dans les astrocytomes et les oligodendrogliomes de bas grade (Dong C et al. 2018).

Les informations sur la structure et la fonction des protéines fournissent ainsi des conseils importants pour comprendre les voies de signalisation et développer de nouvelles thérapies pour réguler les processus de la vie des patients. Bien que de nombreuses études de biologie structurale aient été consacrées à la révélation de la structure et de la fonction des protéines, les procédures expérimentales sont généralement lentes et coûteuses. Alors que les méthodes de calcul ont le potentiel de créer des modèles de structure et de fonction rapides et à grande échelle, la précision et la fiabilité sont souvent un problème. Des progrès significatifs ont été observés au cours des deux dernières décennies dans les prévisions de structure informatisées (Kryshtafovych et al. 2014), une avancée notable, par exemple, est que les serveurs informatiques automatisés peuvent désormais générer des modèles avec une précision comparable à la meilleure modélisation qui combine une variété d'inspections manuelles et d'analyses structurelles et fonctionnelles (Huang Met al. 2014). Le protocole I-TASSER, construit sur la base de simulations d'assemblage de fragments itératifs (Yang J et al. 2014), représente l'une des méthodes les plus réussies démontrées pour les prédictions automatisées de la structure et de la fonction des protéines.

Les mutations de la p53 se produisent dans une zone très compacte de la protéine, comme cela a été démontré en utilisant ConSurf, également utilisé par (Thirumal et al. 2018) pour calculer le score de conservation des protéines. Dans ConSurf, le taux d'évolution est estimé en se basant sur la relation évolutive entre la protéine (ADN / ARN) et ses homologues et en considérant la similitude entre les acides aminés (nucléiques) tels que reflétés dans la matrice de substitutions. L'un des avantages de ConSurf par rapport aux autres méthodes est le calcul précis du taux d'évolution en utilisant soit une méthode bayésienne empirique, soit une méthode du maximum de vraisemblance (Ashkenazy H et al. 2016 ; Ben Chorin A et al. 2020).

Ce type de mutation qui stabilise la structure de liaison à l'ADN est appelé «mutation structurelle»; elle affecte l'architecture globale de la surface de la liaison à l'ADN et modifie la conformation de la protéine, contrairement à la protéine p53 de type sauvage qui agit comme un suppresseur de tumeur en raison de son activité de liaison à l'ADN. Enfin, bien que nous identifions l'effet des mutations au niveau des gènes IDH1, IDH2 et TP53, les limites de cette étude se concrétisent dans la manipulation difficile qui est due à la durée de conservation des échantillons tissulaires inclus en paraffine. De plus, le spectre des biomarqueurs étudiés doit être élargi dans les futures études pour inclure les mutations du promoteur TERT et la co-délétion 1p / 19q afin de fournir une valeur supplémentaire à l'étude et à sa tentative de comprendre la gliomagenèse (Pereira GRC et al. 2019 ; Killela PJ et al. 2013).

La combinaison de nouvelles méthodes moléculaires avec l'analyse de la pathologie anatomique doit faire partie des soins cliniques quotidiens, cette combinaison est un outil essentiel pour le diagnostic des gliomes de bas grade et la médecine personnalisés. Elle fournit des informations plus précises et peut également réduire le besoin de chirurgies supplémentaires et fournir des informations importantes pour d'autres traitements de radiothérapie ou de chimiothérapie tant que le traitement varie en fonction du diagnostic. Malgré ces avancées, il est important de noter qu'il reste beaucoup à faire dans ce domaine, notamment pour développer des modèles informatiques innovants et améliorés pour la recherche dite translationnelle (Li X et al. 2017 ; Buckner J et al. 2017).

## **VI – Conclusion**

Les découvertes de la dernière décennie ont complètement changé notre vision de la génomique des gliomes humains. Grâce à l'analyse computationnelle, nous avons pu étudier la pathogénicité des mutations IDH1, IDH2 et TP53 et modéliser les protéines générées en 3D, ce qui nous a permis de mieux comprendre la gliomagenèse. Ces techniques informatiques

compléteront l'analyse de la pathologie anatomique, ce qui améliorera le diagnostic des astrocytomes et des oligodendrogliomes.

# **Identification des biomarqueurs potentiels pour les gliomes de bas grade en analysant les profils génomiques et le phénotype CpG Island Methylator des patients dans la base de données TCGA**

## **I - Introduction**

Les gliomes de bas grade sont les tumeurs cérébrales primaires les plus courantes qui affectent le système nerveux central (SNC), elles provoquent une morbidité considérable et représentent un défi thérapeutique en raison de l'hétérogénéité de leur comportement clinique. Ils sont plus fréquents chez l'adulte et se caractérisent par un mauvais diagnostic. La différenciation entre les astrocytomes et les oligodendrogliomes reste un défi majeur en neuro-oncologie car ils ont une apparence morphologique similaire et des réponses biologiques et thérapeutiques distinctes (Gittleman H et al. 2019 ; Su J et al. 2019).

Les caractéristiques histologiques sont insuffisantes pour discriminer systématiquement entre les deux sous-types, c'est pourquoi la mise à jour 2016 de la classification des tumeurs du SNC de l'Organisation mondiale de la santé (OMS) représente un changement dans le diagnostic des tumeurs cérébrales en intégrant des caractéristiques moléculaires et phénotypiques dans la classification des tumeurs en réduisant ainsi les sous-groupes définis. Parmi les altérations génomiques et épigénétiques associées aux gliomes de bas grade, la mutation de l'isocitrate déshydrogénase (IDH) et le phénotype de méthylation de l'îlot CpG du gliome (G-CIMP) respectivement (Paul Y et al. 2017 ; Malta TM et al. 2018).

Dans cette étude, les données ont été obtenues pour les patients atteints de gliomes de bas grade de la base de donnée The Cancer Genome Atlas (TCGA), un total de 516 échantillons ont été analysés par TCGA biolinks R Bioconductor.

## **II – Objectifs**

- Identifier les profils génomiques des patients atteints des gliomes de bas grade.
- Identifier les profils épi-génomiques des patients atteints des gliomes de bas grade.
- Reclasser les patients atteints des gliomes de bas grade selon la nouvelles classification des tumeurs cérébrales de l'OMS 2016.

## **III – Matériel et méthodes**

### **1– Collecte de données**

Les données ont été téléchargées à partir de la base de données TCGA (<https://gdc-portal.nci.nih.gov>), y compris 516 échantillons de LGG (gliomes de bas grade). Les données

génomiques (Illumina HiSeq 2000) et les données de méthylation (Illumina Infinium Human Methylation 450 BeadChip) ont été obtenues par le package TCGA biolinks R (version 2.12.6) (Colaprico et al. 2016).

```
query <- TCGAquery(tumor = c("LGG"), level = 1,  
  
  platform = c("Illumina HiSeq 2000", "Illumina Infinium Human  
Methylation 450 BeadChip"), samples = c("TCGA-19-4065", "TCGA-E1-  
5322-01A-01D-1467-05"), version = list(c("Illumina HiSeq  
2000", "LGG", 1), c("HumanMethylation450", "LGG", 5)))
```

Après normalisation, 501 échantillons ont été retenus pour les données génomiques et 496 échantillons pour les données de méthylation.

```
TCGAanalyze_Normalization (tabDF, geneInfo, method = "geneLength")
```

Toutes les informations cliniques sur les patients ont été utilisées pour chaque type de tumeur.

## 2- Analyse bio-informatique

Le package ggplot2 (version 3.2.1) est utilisé pour décrire la distribution des gènes mutés chez les patients atteints des gliomes de bas grade (H. Wickham. 2009 ; Maag JLV. 2018)

```
library(ggplot2)  
ggplot(housing, aes (x = Home.Value)) + geomhistogram ()
```

En se basant sur la nouvelle classification de l'OMS 2016, nous avons reclassé tous les échantillons en deux type de tumeurs (astrocytomes et oligodendrogliomes) par le package maftools (version 3.9) puis nous avons générer un oncolplot qui présente le paysage mutationnel de tous les patients avec leurs données cliniques (Mayakonda A et al. 2018).

```
if (!require("BiocManager"))  
  install.packages("BiocManager")  
BiocManager::install("maftools")  
library(maftools)  
laml.maf = system.file('extdata', 'tcga_laml.maf.gz', package = 'maftools')  
laml.clin = system.file('extdata', 'tcga_laml_annot.tsv', package = 'maftools')  
  
laml = read.maf(maf = laml.maf, clinicalData = laml.clin)  
oncoplot(maf = laml, top = 10)
```

Ensuite, nous avons utilisé la fonction "maf summary" pour afficher le nombre de variantes de chaque échantillon sous forme de graphique à barres empilées et les types de variantes sous forme de boîte à moustaches.

```
plotmafSummary(maf = lam1, rmOutlier = TRUE, addStat = 'median', dashboard = TRUE, titvRaw = FALSE)
```

La fonction `titv` classe les SNP en transitions et transversions, les données récapitulatives de la boîte à moustaches indiquant la distribution générale de six conversions différentes.

```
lam1.titv = titv(maf = lam1, plot = FALSE, useSyn = TRUE)  
plotTiTv(res = lam1.titv)
```

Puis, nous avons utilisés le fisher's exact test et le package `cometExactTest` pour identifier les gènes qui s'excluent mutuellement ou coexistent, et les gènes potentiellement altérés respectivement.

```
somaticInteractions(maf = lam1, top = 25, pvalue = c(0.05, 0.1))
```

Enfin, pour estimer la survie des patients à l'aide des données cliniques en fonction du biomarqueur IDH1 (muté/sauvage), nous avons utilisé le package `Survival KM` (Colaprico A et al. 2016).

```
mafSurvival(maf = lam1, genes = 'DNMT3A', time = 'days_to_last_followup', Status = 'Overall_Survival_Status', isTCGA = TRUE)
```

Concernant les données de méthylation, nous avons utilisé le TCGA Workflow Bioconductor (version 1.8.1), nous avons commencé par vérifier la méthylation moyenne de l'ADN des différents groupes d'échantillons reclassé (astrocytomes et oligodendrogliomes) via le TCGA `Visualize_MeanMethylation` fonction (Colaprico A et al. 2016 ; Gu Z et al. 2016).

```

astro.samples <- matchedMetExp("astrocytomas", n = 10)
oligo.samples <- matchedMetExp("oligodendrogliomas", n = 10)
samples <- c(astro.samples,oligo.samples)
query <- GDCquery(project = c("astrocytomas",
"oligodendrogliomas"),
  data.category = "DNA methylation",
  platform = "Illumina Human Methylation 450",
  legacy = TRUE,
barcode = samples)
GDCdownload(query)
met <- GDCprepare(query, save = FALSE)

met <- subset(met,subset = as.character(seqnames(met)) %in%
c("chr9"))
data(elmerExample)
met <- met[rowSums(is.na(assay(met))) == 0,]
df <- data.frame(
  "Sample.mean" = colMeans(assay(met), na.rm = TRUE),
  "groups" = met$project_id)

library(ggpubr)
ggpubr::ggboxplot(df,
  y = "Sample.mean",
  x = "groups",
  color = "groups",
  add = "jitter",
  ylab = expression(paste("Mean DNA methylation (", beta, "-
values)")),
  xlab = "") +
stat_compare_means()

```

les régions différentiellement méthylées ont été visualisées sous forme de cartes thermiques par le logiciel ComplexHeatmap (version 2.0.0) (Colaprico A et al. 2016 ; Gu Z et al. 2016).

```

library(ComplexHeatmap)
clinical <- plyr::rbind.fill(astro_clin,oligo_clin)
status.col <- "status"
probes<- rownames(dmc)[grep("hypo|hyper",dmc$status,ignore.case
= TRUE)]
sig.met <- met[probes,]
clinical.order<clinical[match(substr(colnames(sig.met),1,12),c
clinical$bcr_patient_barcode),]
ta = HeatmapAnnotation(df = clinical.order[, c("disease",
"gender", "vital_status", "race")],
col = list(disease = c("astro" = "red", "oligo" = "bleu"),
gender = c("male" = "blue", "female" = "pink")))
ra = rowAnnotation(df = dmc[probes, status.col],
col = list("status.TCGA.astro.TCGA.oligo" =
c("Hypomethylated" = "bleu", "Hypermethylated" = "red")),width
= unit(1, "cm"))

heatmap <- Heatmap(assay(sig.met), name = "DNA methylation", col
= matlab::jet.colors(200), show_row_names = FALSE, cluster_rows
= TRUE, cluster_columns = FALSE, show_column_names = FALSE,
bottom_annotation = ta, column_title = "DNA Methylation")
png("heatmap.png",width = 600, height = 400)
draw(heatmap, annotation_legend_side = "bottom")
dev.off()

```

## IV – Résultats

### 1 – Distribution des gènes impliqués dans les gliomes de bas grade

Afin de montrer la distribution des gènes, nous avons utilisé le package ggplot2. La forme asymétrique de kurtosis indique que les gènes varient d'un patient à l'autre et nous constatons que certains gènes sont plus présents que d'autres (figure 9).

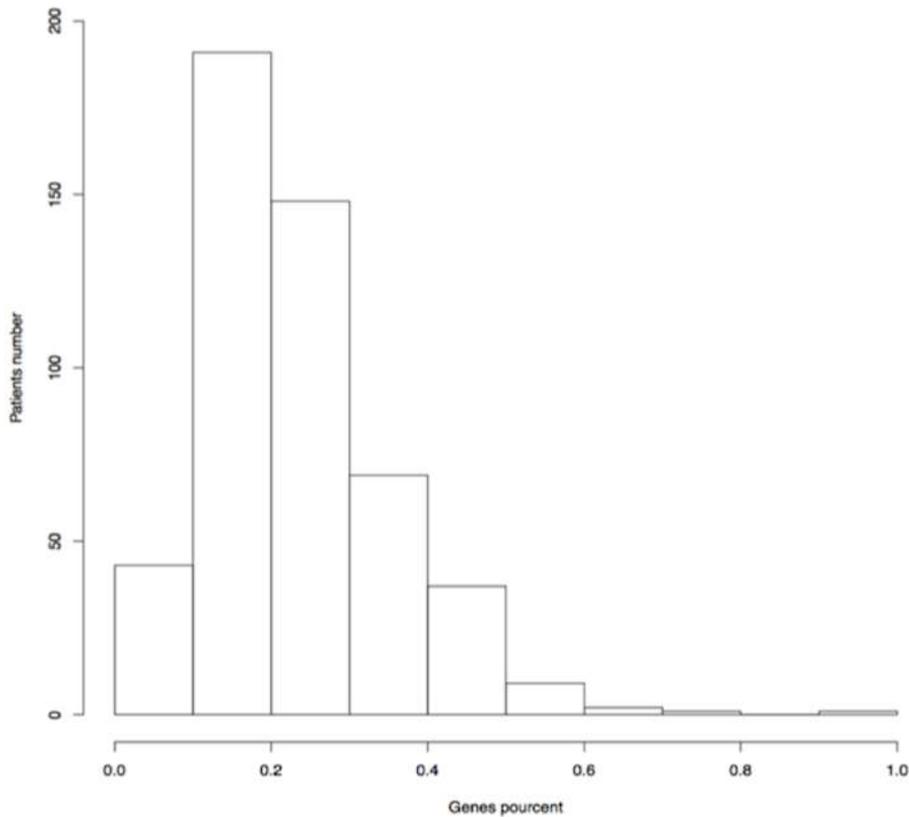


Figure 9 : Distribution des gènes impliqués dans les gliomes de bas grade

## 2 – Paysage mutationnel des altérations somatiques des gliomes de bas grade

Pour identifier des différences significatives des mutations génétiques entre les deux sous-types astrocytomes et oligodendrogliomes, nous avons utilisé le package Maftools (Figure 10). Pour les patients atteints d'astrocytomes, les gènes les plus mutés trouvés sont IDH1 78%, TP53 73%, ATRX 63%, TTN 12%, MUC16 6%, NF1 6% et PTEN 6% (Figure 11 A), de l'autre côté, les patients avec les oligodendrogliomes, la plupart des gènes mutés sont IDH1 81%, CIC 56%, FUBP1 23%, NOTCH1 12%, PIK3CA 12%, IDH2 9% et PIK3R1 7% (Figure 12 A). Dans une cohorte de 501 échantillons répartis en trois catégories selon la classification de l'OMS 2007, 186 astrocytomes, 182 oligodendrogliomes et 96 oligoastrocytomes. Ces patients ont été reclassés en deux catégories selon la nouvelle classification OMS de 2016, 290 astrocytomes et 174 oligodendrogliomes, et il reste 37 patients NOS (not otherwise specified) dont le séquençage doit être répété pour avoir un diagnostique précis. La majorité des patients ont plus de 41 ans et le nombre d'hommes est de 157 avec des astrocytomes et 98 avec des oligodendrogliomes, pour les femmes, nous avons 132 astrocytomes et 76 oligodendrogliomes (Figure 10).

Altered in 462 (92.2%) of 501 samples.

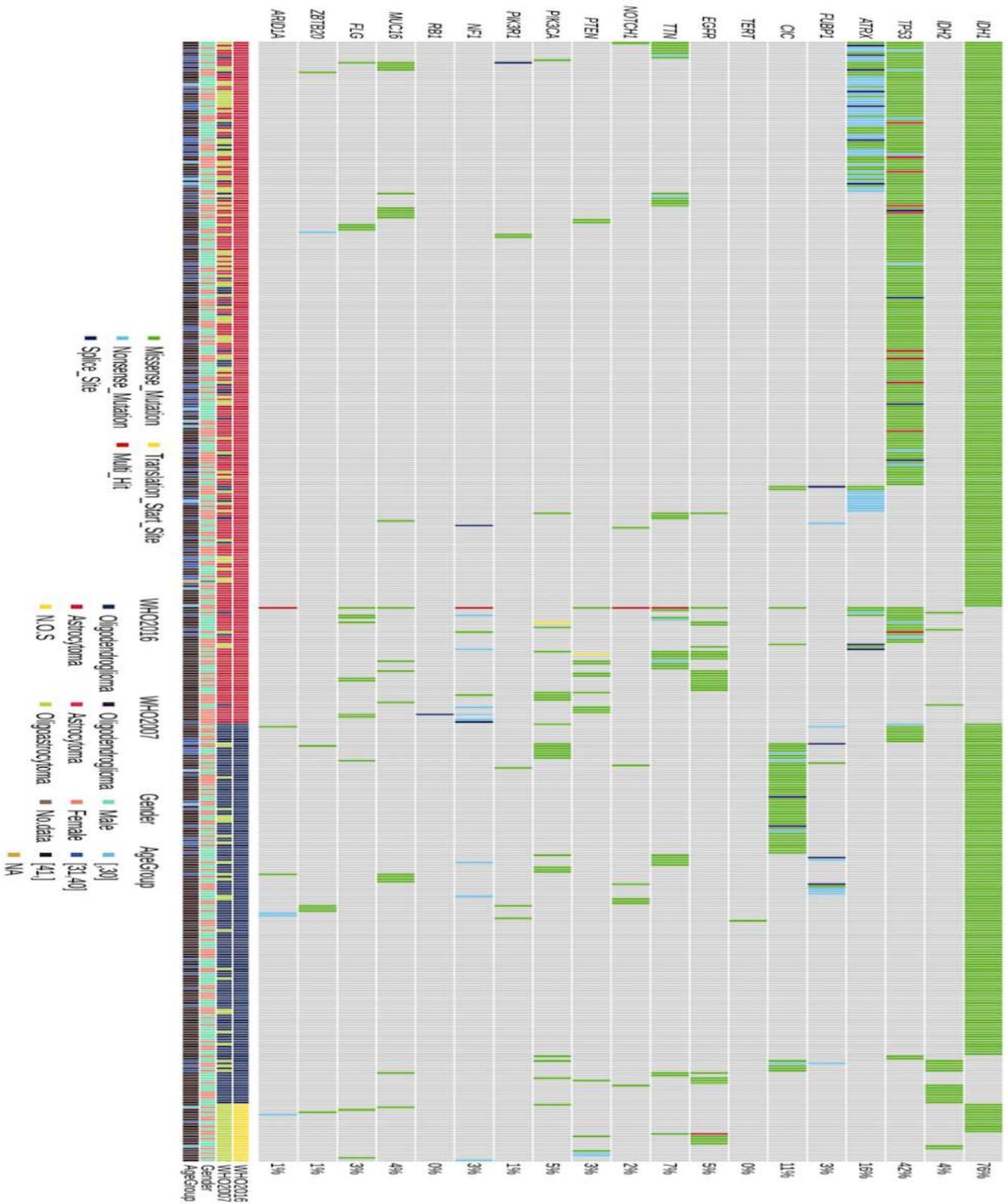


Figure 10 : Altération mutationnelle de 501 échantillons atteints de gliome de bas grade

### 3 – Classification des variantes

Afin d'établir la classification des variantes, nous avons utilisé la fonction "maf summary" du package Maftools, les deux sous-types sont caractérisés par un taux très élevé de mutations faux-sens<sup>1</sup> représentant 13834 mutations pour les astrocytomes et 2906 mutations pour les oligodendrogliomes, puis nous trouvons la non-sens mutation<sup>2</sup> qui représente 949 mutations pour les astrocytomes et 154 mutations pour les oligodendrogliomes suivies du frame shift del<sup>3</sup> qui représente 484 mutations, 301 mutations et le frame shift ins<sup>3</sup> représentent 258 mutations, 57 mutations pour les astrocytomes et les oligodendrogliomes respectivement. Le site d'épissage représente 360 mutations pour les astrocytomes et 95 mutations pour les oligodendrogliomes (Figure 11 B et Figure 12 B). Concernant les type des variantes, les plus présents sont les SNP suivis des délétions et enfin les insertions présentant respectivement 22631, 792, 321, 4635, 555 et 80 mutations pour les astrocytomes et les oligodendrogliomes (Figure 3 C, Figure 4 C). La classe SNV (Figure 11 D, Figure 12 D) et le diagramme de transition transversion (Figure 11 E, Figure 12 E) montre que le type de mutation le plus courant dans les deux sous-types est la mutation de transition 74% pour les astrocytomes (C> T, T> C ) et 76% pour les oligodendrogliomes (C> T, T> C) par rapport aux mutations transversales qui présentent 26% et 24% pour les astrocytomes (C> A, C> G, T> A et T> G) et les oligodendrogliomes (C> A, C> G, T> A et T> G) respectivement.

---

1. Mutation faux-sens se produit lorsque le changement d'une seule paire de bases provoque la substitution d'un acide aminé différent dans la protéine résultante.

2. Mutation non-sens est une substitution d'une seule paire de bases qui conduit à l'apparition d'un codon stop là où il y avait auparavant un codon spécifiant un acide aminé.

3. Frameshift est un type de mutation impliquant l'insertion ou la suppression d'un nucléotide dans lequel le nombre de paires de bases supprimées n'est pas divisible par trois. "Divisible par trois" est important car la cellule lit un gène par groupes de trois bases. Chaque groupe de trois bases correspond à l'un des 20 acides aminés différents utilisés pour construire une protéine. Si une mutation perturbe ce cadre de lecture, la traduction sera incorrecte.

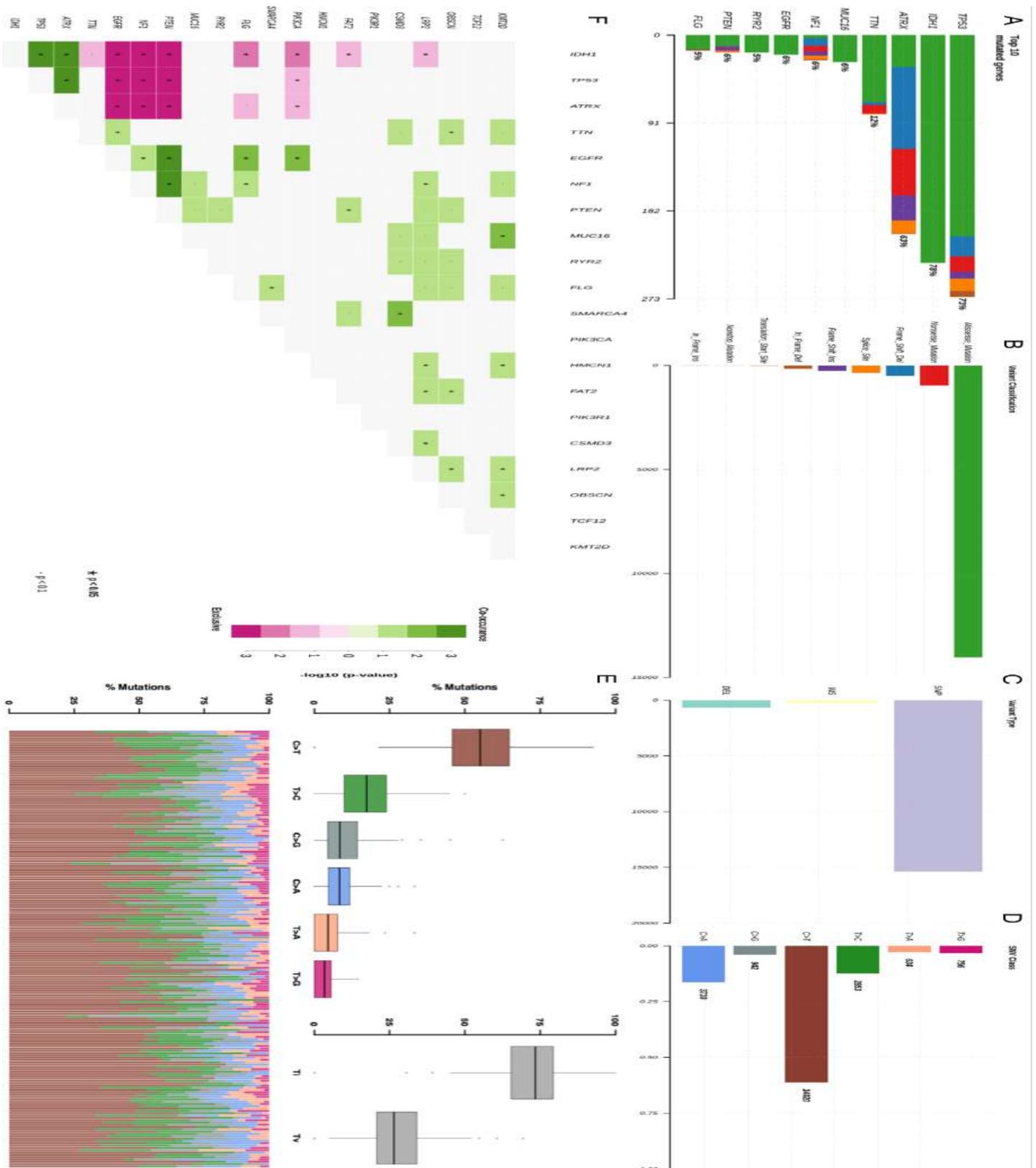


Figure 11 : Les gènes impliqués dans les astrocytomes (A) barplot montrent les 10 premiers mutés qui sont classés en fonction de leur fréquence de mutation. (B) Classification des variantes. (C) Type de variante. (D) Classe SNV. (E) Tracé de transition et de transversion affichant la distribution des SNV en six événements de transition et de transversion. Le diagramme à barres empilées (en bas) montre la distribution des spectres de mutation pour chaque échantillon. (F) Paires de gènes mutuellement exclusifs et cooccurrents affichés sous forme de matrice triangulaire.



#### 4 – Interaction des gènes

La génération de la matrice présentée dans les (Figure 11 F et Figure 12 F), montre l'interaction entre les gènes en fonction de la distribution binomiale cumulative. Elle prend en compte à la fois la longueur des gènes et le nombre total de mutations dans les gènes identifiés comme potentiellement significatifs. Un score est calculé pour chaque grappe, le vert indique une tendance à la co-occurrence, le rose indique une tendance à l'exclusivité (Tamborero D et al. 2013). Dans le cas des patients atteints d'astrocytomes, les gènes qui ont une co-occurrence significative de  $P < 0,05$  sont: IDH1-TP53; IDH1-ATRX; ATRX-TP53; PTEN-EGFR; PTEN-NF1 puis les gènes exclusifs qui présentent une signification de  $P < 0,05$  sont: IDH1-PTEN; IDH1-NF1; TP53-EGFR; ATRX-PTEN; TP53-NF1.

D'autre part, les oligodendrogliomes sont caractérisés par d'autres gènes, nous constatons que les gènes qui ont une co-occurrence significative de  $P < 0,05$  sont: IDH1-CIC; IDH1-FUBP1; CIC-FUBP; NOTCH1-ZNF292, dans l'autre sens, les gènes exclusifs qui présentent une signification de  $P < 0,05$  sont: IDH1-IDH2; IDH1-EGFR; CIC-EGFR; CIC-TP53.

#### 5 – Analyse de survie des patients

Pour estimer la survie globale des patients, nous avons utilisé le package Survival KM, notre analyse des données cliniques a montré que les personnes atteintes d'astrocytomes avaient une survie globale plus courte que celles des patients atteints d'oligodendrogliomes (Figure 5 A). Plus précisément, les patients qui ont un mutant IDH1 ont une survie plus longue que ceux qui ont un type sauvage IDH1 et cela est vrai pour les astrocytomes et les oligodendrogliomes (Figure 13 B C).

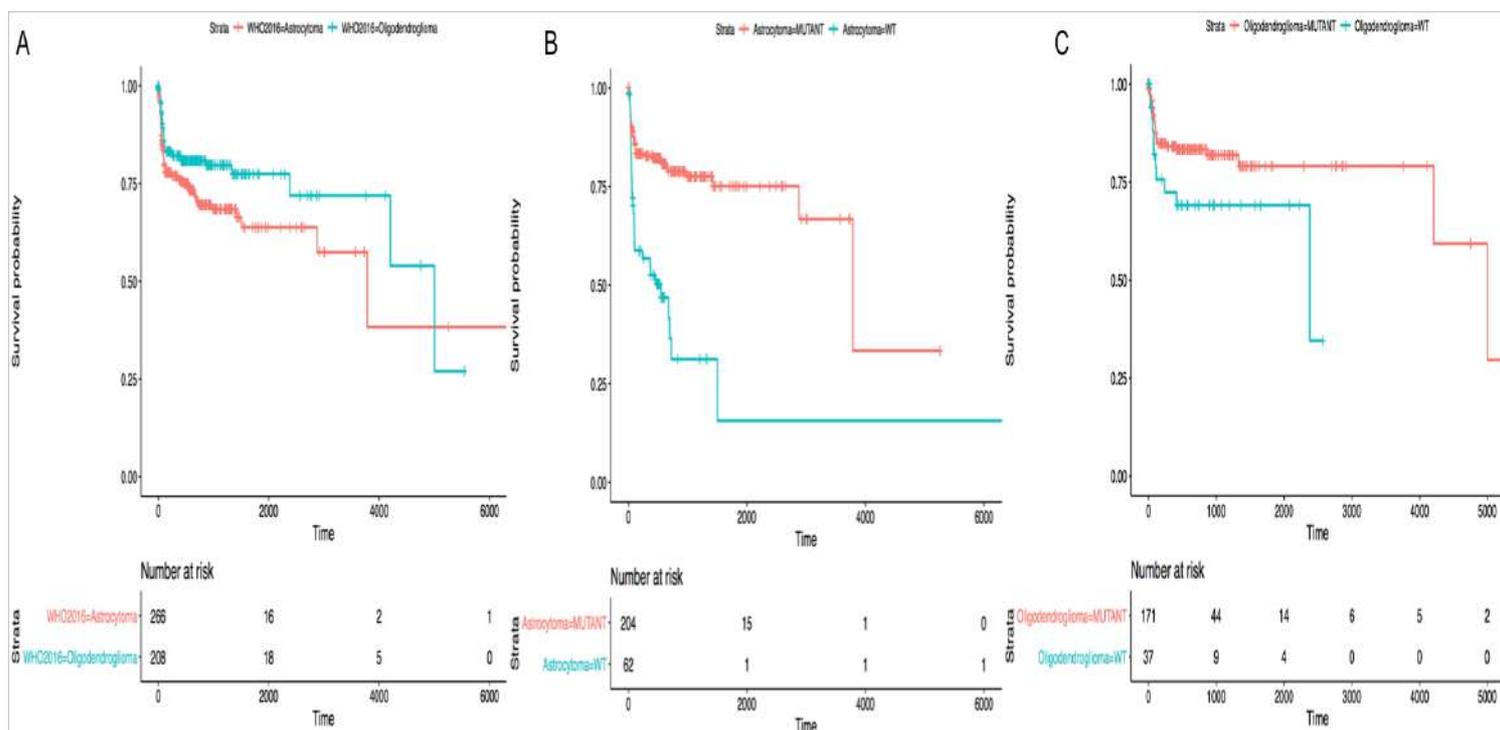


Figure 13 : Estimations de Kaplan–Meier de la survie globale des patients atteints d'astrocytomes et d'oligodendrogliomes classés en fonction de la mutation IDH1 et du statut de type sauvage IDH1

## 6 – Visualisation du Glioma CpG Island Methylator Phenotype

Pour détecter la valeur moyenne de la méthylation des patients, nous avons utilisé le TCGA Workflow Bioconductor, la (Figure 14 A) montre un graphe qui présente cette valeur pour chaque échantillon du groupe des astrocytomes (314 échantillons) et une valeur moyenne pour chaque échantillon du groupe des oligodendrogliomes (182 échantillons). L'analyse à grande échelle met en évidence une différence entre les deux sous-types, on constate que les oligodendrogliomes présentent une hyperméthylation par rapport aux astrocytomes.

Pour visualiser les régions différentiellement méthylées, nous avons utilisé le logiciel ComplexHeatmap, la (Figure 14 B) montre une carte thermique qui visualise le niveau de méthylation de l'ADN de tous les échantillons. Au total, nous trouvons 361340 locus CpG, pour les astrocytomes, les gènes hypométhylés sont REST en position cg26694713, FSCN1 en positions cg23080179, cg06123544 et cg04618002, IMPDH1 en position cg15740366, CYHR1 en positions cg14831162 et ATL3 en positions cg27042081 et cg15814736. En revanche, les patients présentant des oligodendrogliomes ne présentent pas de gènes hypométhylés identiques à ceux des patients atteints d'astrocytomes, on trouve trois gènes codant pour les protéines FGGY en position cg17393296, RTN4 en position cg14640066 et PDE7B en positions

cg08378567 et cg27306443, il y a aussi des gènes qui codent pour les lincARN comme AC073283.4 en position cg25653638, LINCO1551 en position cg03505995 et cg06779110, et le gène FOXG1-AS1 en position cg00946992 de type antisens qui interagit avec l'ARNm pour inhiber la synthèse de la protéine correspondante. En cas d'hyperméthylation, les patients atteints d'un astrocytome ont trois gènes codant pour des protéines, MYT1L en position cg20435238, ELAVL4 en position cg26888153 et cg26888153 et PSD3 en position cg22276811 et RP11-1102P16.1 qui est un gène qui code pour un LincRNA en position cg 08758568. Les patients atteints d'oligodendrogliomes n'ont pas de gènes hyperméthylés identiques aux patients atteints d'astrocytomes, on trouve le gène PLCG1 en position cg03222834 et STAB1 en position cg04654429 et un gène qui code pour U91319.1 LincRNA en position cg04654429.

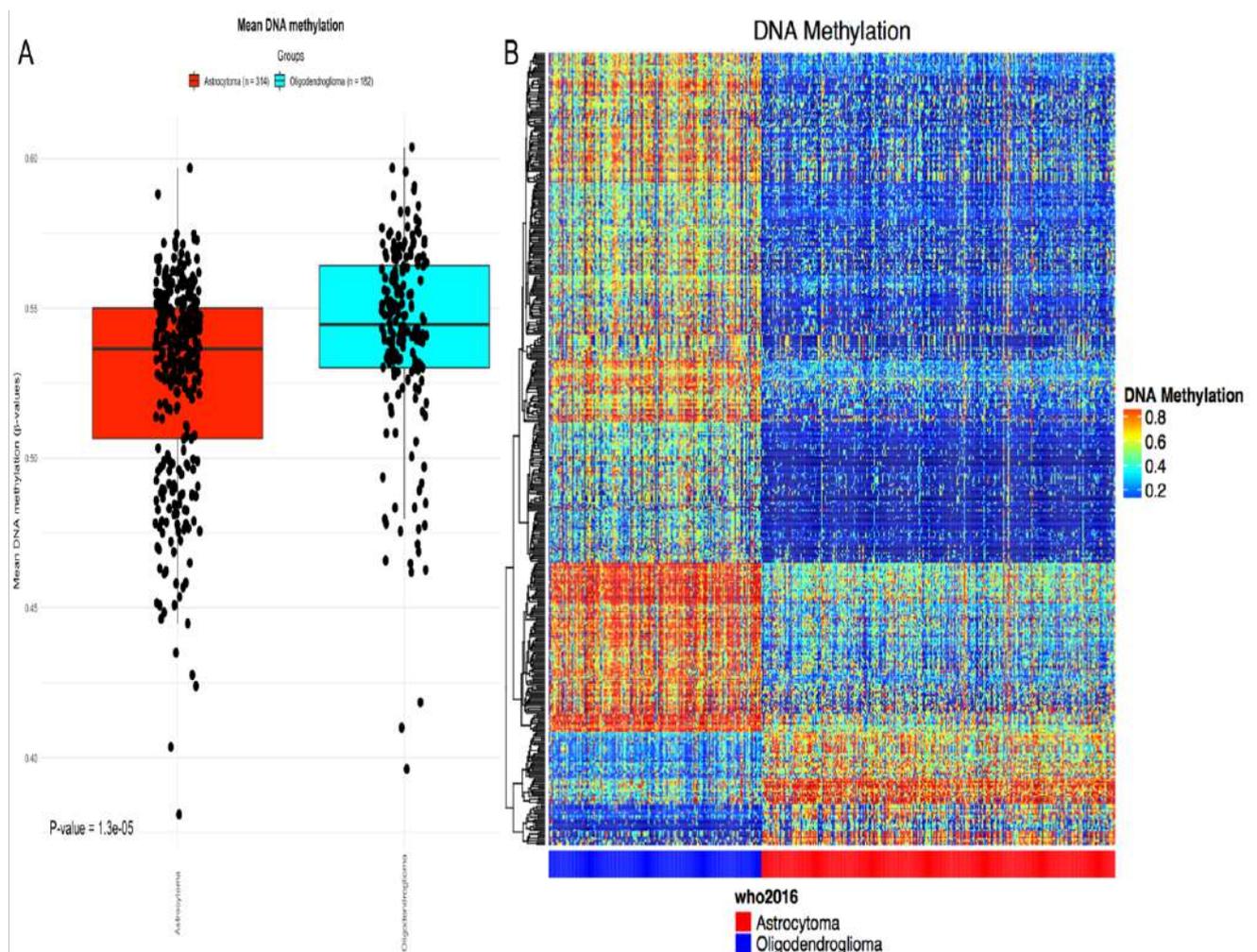


Figure 14 : (A) La valeur moyenne de la méthylation d'ADN chez les patients atteints d'astrocytome et d'oligodendrogliome. (B) Carte thermique présentant toute les positions hyperméthylé et hypométhylé chez les patients atteints d'astrocytome et d'oligodendrogliome

## V- Discussion

Dans cette étude, notre analyse des données génomiques des gliomes de bas grade de la base de données TCGA était basée sur la nouvelle classification de l'OMS de 2016. Tous les patients qui avaient une classification de l'OMS de 2007 ont été reclassés en deux types de tumeurs (astrocytomes et oligodendrogliomes), par contre le profil des oligoastrocytomes (gliome mixte) a été éliminé. Les patients qui ne présentent pas de biomarqueurs de classification tels que TP53, ATRX, CIC et FUBP1 ont été nommés NOS (not otherwise specified) dont le diagnostic n'est pas complet ce qui incite à un nouveau séquençage afin de mieux caractériser leurs profils génomiques. Pour le biomarqueur IDH1, un gène commun entre les deux types de tumeurs. Dans le cas d'une mutation, il catalyse la conversion de l'alpha-cétoglutarate en bêta-hydroxyglutarate (2-HG) (Gusyatiner et al. 2018). Étant donné que la structure du 2-HG est similaire à celle de l' $\alpha$ -KG, le 2-HG inhibe une variété de dioxygénases dépendantes de l' $\alpha$ -KG (Xu et al. 2011; Nonoguchi et al. 2013), parmi eux, on trouve le Ten-Eleven Translocation-2 (TET2) qui induit une déméthylation globale de l'ADN en catalysant la conversion de la 5-méthylcytosine (5-mC) en 5-hydroxyméthylcytosine (5-hmC). Le mutant IDH1 provoque une augmentation des concentrations de 5mC, au lieu d'une diminution de 5hmC (Bardella et al. 2016). La mutation IDH favorise également la méthylation de l'ADN par l'inhibition de TET2, résultant en un phénotype de méthylateur d'îlot de gliome CpG (G-CIMP).

(Malta et al. 2018; Turcan et al. 2012), ont démontré que le phénotype G-CIMP dans le glioblastome est formé après l'introduction d'une mutation IDH1 dans les astrocytes humains normaux, indiquant que le mutant IDH a induit la suppression de TET2, suivi par G-CIMP, dans les cellules cancéreuses. Comme les patients atteints de gliomes IDH mutants, les patients atteints de gliomes G-CIMP survivent plus longtemps que ceux sans G-CIMP.

Pour les astrocytomes, nous trouvons la présence des gènes mutés TP53 et ATRX. Le gène TP53 de type sauvage joue un rôle essentiel dans plusieurs processus cellulaires, notamment la suppression des tumeurs, l'apoptose, la réparation de l'ADN et l'autophagie, ce qui suggère que la mutation au niveau TP53 est définie par une perte de la fonction protéique P53 (Aubrey et al. 2016).

L'inactivation du gène ATRX dans les gliomes peut être due à des mutations, des suppressions, des fusions de gènes ou un amalgame de ces causes (Bunda et al. 2019). Pour les oligodendrogliomes, il existe la présence des gènes CIC et FUBP1; CIC est un répresseur transcriptionnel qui neutralise l'activation des gènes en aval du récepteur de la tyrosine kinase (RTK) (Carpentier et al. 2012) et le gène FUBP1 qui a un rôle putatif de suppresseur de tumeur

dans l'oligodendrogliomagénèse (Wang et al. 2019). La prévalence des mutations IDH dans les gliomes de bas grade incite à cibler les enzymes mutantes elles-mêmes ou leurs conséquences métaboliques et épigénomiques en aval, comme le G-CIMP, cependant, les patients atteints d'astrocytomes présentent une hyperméthylation au niveau du gène MYT1L qui est un facteur de transcription qui joue un rôle clé dans la différenciation neuronale en réprimant spécifiquement l'expression des gènes non neuronaux, (Wong et al. 2019) ont montré que ce gène supprime directement l'expression de YAP1, une protéine qui favorise la prolifération et la croissance des glioblastomes, lorsqu'elle est hyperméthylée elle perd son fonctionnement (Wong et al. 2019). Le gène ELAVL4 hyperméthylé, également présent dans les astrocytomes, (Huang et al. 2016) a indiqué que lors de la différenciation neuronale, ELAVL2, ELAVL3 et ELAVL4 ont induit une polyadénylation alternative de HuR (antigène humain R) et par conséquent supprimé sa traduction, conduisant à un état non prolifératif. Dans le scénario des glioblastomes, il existe une hypothèse que la régulation des gènes ELAVL2, ELAVL3 et ELAVL4 améliore encore la traduction de HuR, conduisant à un état moins différencié et hautement prolifératif des cellules cancéreuses. En revanche, les patients atteints d'oligodendrogliomes sont caractérisés par la présence du gène PLCG1 qui favorise la production de diacylglycérol et d'inositol 1,4,5-trisphosphate (IP3). (Su et al. 2013) ont révélé que cette réaction utilise le calcium comme cofacteur et joue un rôle important dans la transduction intracellulaire des activateurs de la tyrosine kinase, ainsi que le gène STAB1 codant pour une protéine qui lie la matrice nucléaire et les ADN d'échafaudage via un seul noyau nucléaire architecture. (Cancer Genome Atlas Research Network, 2015), ont montré que la protéine recrute des facteurs de remodelage de la chromatine pour réguler la structure de la chromatine et l'expression des gènes, l'hyperméthylation de ces gènes conduit à la réduction de cette activité.

Notre analyse génomique intégrative avec le statut G-CIMP a montré que les deux sous-types de gliomes de bas grade ont un point commun tel que la mutation IDH1 ou l'exclusivité EGFR-IDH1, mais aussi des points de divergence comme les gènes mutés ATRX pour les astrocytomes ou CIC pour les oligodendrogliomes, également au niveau des gènes hyperméthylés qui codent pour des protéines telles que le gène MYT1L pour les astrocytomes et le gène PLCG1 pour les oligodendrogliomes. L'intégration des données cliniques dans l'analyse pour l'étude de la survie des patients a montré que les patients qui ont muté IDH1 ont une survie plus longue que les patients avec le statut sauvage de l'IDH1.

Cette étude a certaines limites, l'analyse des données était basée sur un algorithme de bio-informatique utilisant des données TCGA publiques et aucune autre validation expérimentale

n'a été menée. Contrairement à de nombreuses études ont montré que la classification histopathologique des gliomes de bas grade est sujette à de fortes variations inter-observateurs, cela fournit un aperçu d'identification de biomarqueurs supplémentaires qui compléteront les diagnostics histopathologiques et de neuro-imagerie afin d'élaborer une stratégie de traitement pour chaque patient.

## **VI- Conclusion**

La prévalence du gène IDH1 muté dans les gliomes de bas grade incite à cibler soit les enzymes mutantes elles-mêmes, soit leurs conséquences métaboliques et épigénomiques en aval, telles que le G-CIMP. Les mutations des gènes ATRX, CIC, FUBP1, TP53, NOTCH1 et EGFR sont également impliquées dans la gliomagenèse, avec spécificité et une prévalence importante dans les gliomes de bas grade incluant les astrocytomes et les oligodendrogliomes. Ces biomarqueurs se prêtent à un système de classification basé sur la biologie moléculaire et la clinique, qui doit améliorer la concordance entre les observateurs. Ce type de système semble également susceptible de réduire le diagnostic d'oligoastrocytome et la confusion qui pourrait en résulter et développer de nouvelles stratégies thérapeutiques à un stade précoce avant la transformation maligne.

## **Analyse bio-informatique des gènes et des miARNs différentiellement exprimés dans les gliomes de bas grade.**

### **I - Introduction**

Le gliome de bas grade est une tumeur cérébrale qui provient des cellules gliales qui soutiennent et protègent le système nerveux du cerveau, ils causent des difficultés considérables et représentent un défi thérapeutique en raison de l'hétérogénéité de leur comportement clinique (Wen PY and Huse JT. 2016).

La récente classification de l'OMS de 2016 a mis à jour la définition de ces tumeurs pour inclure la mutation isocitrate déshydrogénase (IDH) et la co-délétion 1p / 19q en tant que biomarqueur essentiel qui caractérise les gliomes. En effet, il est devenu crucial d'identifier et de caractériser d'autres biomarqueurs de gliomes candidats à l'aide d'une analyse bio-informatique, comme la numérisation du génome pour détecter des mutations spécifiques du cancer et l'identification de paysages épigénétiques modifiés dans les cellules cancéreuses ou en explorant l'expression différentielle des gènes, des miARNs et des protéines par des techniques de transcriptomique et de protéomique (Wang H et al. 2019 ; Zhao Z et al. 2019).

Parmi ces techniques à haut débit, le séquençage d'ARN (RNA-seq) est une méthode efficace pour la caractérisation robuste du microenvironnement tumoral. L'utilisation croissante de l'ARN-seq dans des contextes cliniques et scientifiques offre une puissante opportunité d'accéder à de nouveaux biomarqueurs thérapeutiques tels que les miARNs qui sont de petits ARN non codants d'environ 19-25 nucléotides de longueur, ils peuvent réguler divers gènes cibles et sont donc impliqués dans la régulation d'une variété de processus biologiques et pathologiques, il a été démontré que les miARNs dérégulés affectent les caractéristiques du cancer, y compris le maintien de la signalisation proliférative, en passant des suppresseurs de croissance, la résistance à la mort cellulaire, l'activation de l'invasion et des métastases et l'induction de l'angiogenèse. Les miARNs sont reconnus comme des biomarqueurs potentiels pour le diagnostic du cancer humain, le pronostic et le développement d'outils thérapeutiques précis (Luca FRG and Ian JM. 2019 ; Peng Y and Croce C. 2016). Certaines recherches ont montré, par exemple, que le hsa-mir-21 est régulé à la hausse dans les gliomes et favorise l'invasion ciblant des inhibiteurs spécifiques des métalloprotéases matricielles et hsa-mir-7 est régulée à la baisse dans les gliomes et inhibe l'angiogenèse (Babae N et al. 2014). Nous avons également de longs ARN non codants (lncRNA) qui constituent une grande classe de transcrits  $\geq 200$  nucléotides de longueur, les lncRNAs peuvent agir comme des amplificateurs, des échafaudages ou des leurres en interagissant physiquement avec d'autres ARN ou protéines, ce

qui a un impact direct sur la signalisation cellulaire (Sanchez Calle A et al. 2019). Ce sont des régulateurs de la transcription cis, des modulateurs du traitement de l'ARNm, du contrôle post-transcriptionnel et de l'activité protéique et de l'organisation des domaines nucléaires. Ils ont également une implication dans le paysage du cancer en raison de leur influence sur le gain ou la perte de signatures oncogènes (Zhao Z et al. 2019 ; Sanchez Calle A et al. 2018). Par exemple, certains chercheurs ont montré que HOTTIP est un facteur clé dans les métastases d'hypoxie et de gliome, ou MALAT1 qui est associé pour gliomes dont la croissance est régulée par le signal extracellulaire de la kinase (ERK) et de la protéine kinase activée (Li Z et al. 2017). Les miARNs ne s'hybrident pas parfaitement avec leurs cibles, ils peuvent donc cibler des milliers de gènes et de lncARNs dont les produits protéiques fonctionnent dans une variété de voies de signalisation et peuvent donc affectent de multiples voies et réseaux dans la cancérogenèse. Les énormes référentiels de données de la base de données TCGA offrent la possibilité d'analyser ces données épitranscriptomiques. Cependant, d'autres études ont travaillé sur le mécanisme mono-axe des miARNs, des lncARNs et des gènes. Par conséquent, il est important d'étudier l'implication du réseau dans la régulation d'autres biomarqueurs tels que les gènes et les lncRNAs. Dans cette étude, nous avons analysé l'expression des gènes et des miARNs et nous avons théoriquement proposé un nouveau réseau, qui décrit l'association des miARNs différemment exprimés avec des gènes et des lncARNs impliqués dans les gliomes de bas grade.

## II – Objectifs

- Identifier les gènes différemment exprimés dans les gliomes de bas grade.
- Annoter les voies biologiques des gènes différemment exprimés.
- Identifier les miRNAs différemment exprimés dans les gliomes de bas grade.
- Étudier la survie globale des patients atteints des gliomes de bas grade en fonctions des gènes et des miRNAs identifié .
- Créer un network de biomarqueurs spécifiques pour les gliomes de bas grade qui contient à la fois les miRNAs, les gènes et les lncRNAs.

## III – Matériel et méthodes

### 1 – Traitement des données transcriptomiques

L'ensemble des données des miRNAs et l'expression d'ARN de 516 patients atteints des gliomes de bas grade incluant les astrocytomes et les oligodendrogliomes, ont été téléchargés à partir de

TCGA (<https://gdc-portal.nci.nih.gov>) par le package TCGAbiolinks R version 2.12.6 (Colaprico A et al. 2016).

```
project <- 'TCGA-LGG'
rnadir <- paste(project, 'RNAseq', sep='/')
mirdir <- paste(project, 'miRNAs', sep='/')
gdcRNADownload(project.id      = 'TCGA-CHOL',
               data.type      = 'RNAseq',
               write.manifest = FALSE,
               method         = 'gdc-client',
               directory       = rnadir)
gdcRNADownload(project.id      = 'TCGA-CHOL',
               data.type      = 'miRNAs',
               write.manifest = FALSE,
               method         = 'gdc-client',
               directory       = mirdir)
clinicaldir <- paste(project, 'Clinical', sep='/')
gdcClinicalDownload(project.id = 'TCGA-CHOL',
                    write.manifest = FALSE,
                    method         = 'gdc-client',
                    directory       = clinicaldir)
```

Ensuite, Les données ont été normalisé par le logiciel R.

```
library(DT)
data(rnaCounts)
data(mirCounts)
rnaExpr <- gdcVoomNormalization(counts = rnaCounts, filter = FALSE)
mirExpr <- gdcVoomNormalization(counts = mirCounts, filter = FALSE)
##### Parse and filter RNAseq metadata #####
metaMatrix.RNA <- gdcParseMetadata(project.id = 'TCGA-LGG',
                                  data.type  = 'RNAseq',
                                  write.meta = FALSE)
metaMatrix.RNA <- gdcFilterDuplicate(metaMatrix.RNA)
metaMatrix.RNA <- gdcFilterSampleType(metaMatrix.RNA)
datatable(as.data.frame(metaMatrix.RNA[1:5,]), extensions = 'Scroller',
          options = list(scrollX = TRUE, deferRender = TRUE, scroller = TRUE))
```

## 2 – Analyse de séquence d'ARN et des miRNAs

### 2 – 1 – Analyse différemment exprimée

Comparé au groupe normal avec les patients atteints des gliomes de bas grade, le package Limma version 3.42.0 dans le logiciel R (Ritchie ME et al. 2015), a été utilisé pour identifier les gènes différemment exprimés avec des seuils d'une valeur  $P < 0,05$  et un changement de pli | FC |  $> 1$  et les miARNs différemment exprimés (Fig. 2) avec une valeur  $P < 0,01$  et un changement de pli | FC |  $> 2$ .

```

DEGAll <- gdcDEAnalysis(counts      = rnaCounts,
                       group       = metaMatrix.RNA$sample_type,
                       comparison  = 'PrimaryTumor-SolidTissueNormal',
                       method      = 'limma')
datatable(as.data.frame(DEGAll),
          options = list(scrollX = TRUE, pageLength = 5))
deALL <- gdcDEReport(deg = DEGAll, gene.type = 'all')
deLNC <- gdcDEReport(deg = DEGAll, gene.type = 'long_non_coding')
dePC <- gdcDEReport(deg = DEGAll, gene.type = 'protein_coding')
ceOutput <- gdcCEAnalysis(lnc      = rownames(deLNC),
                         pc       = rownames(dePC),
                         lnc.targets = 'starBase',
                         pc.targets  = 'starBase',
                         rna.expr   = rnaExpr,
                         mir.expr   = mirExpr)

gdcVolcanoPlot(DEGAll)
gdcBarPlot(deg = deALL, angle = 45, data.type = 'RNAseq')
degName = rownames(deALL)
gdcHeatmap(deg.id = degName, metadata = metaMatrix.RNA, mirna.expr =
mirnaExpr)

```

## 2 – 2 – Analyse d'enrichissement fonctionnel

Afin d'analyser les processus biologiques et l'enrichissement fonctionnel, le logiciel ClusterProfiler (Yu G et al. 2013), a été utilisé avec un niveau significatif (valeur P <0,01).

```

gdcEnrichPlot(enrichOutput, type='bubble', category='GO', num.terms = 10)

```

## 2 – 3 – Analyse de survie globale des patients

Pour identifier les valeurs pronostiques des gènes et des miARN trouvés en combinant les données de patients cliniques, le logiciel Kaplan-Meier de survie (Lumley TSR. 2019), a été utilisé avec une valeur P <0,05.

```

survOutput <- gdcSurvivalAnalysis(gene      = rownames(deALL),
                                 method     = 'coxph',
                                 rna.expr   = rnaExpr,
                                 metadata   = metaMatrix.RNA)
survOutput <- gdcSurvivalAnalysis(gene      = rownames(deALL),
                                 method     = 'KM',
                                 rna.expr   = rnaExpr,
                                 metadata   = metaMatrix.RNA,
                                 sep        = 'median')

gdcKMPlot(gene      = 'LOX',
          rna.expr   = rnaExpr,
          metadata   = metaMatrix.RNA,
          sep        = 'median')
gdcKMPlot(mirna     = ' hsa-mir-18a-5p ',
          mirna.expr = mirnaExpr,
          metadata   = metaMatrix.RNA,
          sep        = 'median')

```

### 3 – 4 – Construction du réseau des miARNs

Nous avons construit le réseau d'expression génique impliqué dans les gliomes de bas grade par la plateforme GeneMANIA Cytoscape, en implémentant l'algorithme GeneMANIA qui utilise une approche d'association pour dériver des prédictions à partir d'une combinaison de sources de données potentiellement hétérogènes. Il étend la plate-forme de visualisation et d'analyse du réseau Cytoscape et les fonctionnalités du site Web de prédiction de la fonction génique GeneMANIA. Le réseau prédit résultant de relations fonctionnelles entre les gènes interrogés et prédits est ensuite disponible sous forme de réseau Cytoscape annoté pour une analyse plus approfondie (Warde-Farley D et al. 2013).

Les réseaux sont regroupés en six catégories: co-expression, co-localisation, interaction génétique, interaction physique, domaine protéique prédit et partagé et voie de signalisation commune. Les données proviennent d'un large éventail de sources, y compris des études individuelles et de grandes bases de données telles que BIOGRID, GEO, I2D (Brown et Jurisica, 2005). Le réseau des miARNs, a été construit sur le raisonnement selon lequel ces ARN non codants peuvent interagir sur la régulation de l'activité génique (Sana J et al. 2018). Afin de déterminer les interactions entre les miARNs-gènes et miARNs-lncARNs, il faut établir trois critères par le package GDCRNATools version 1.7.0 dans le logiciel R (Li R et al. 2018), l'expression des miRNAs doit être positivement corrélée, les miARNs devraient jouer des rôles similaires dans la régulation de l'expression de l'ARNm et de lncARN et en fin les lncARNs et l'ARNm doivent partager un nombre important de miARN.

Le test hypergéométrique des miARNs partagés, l'analyse de corrélation d'expression entre lncARN-miARN et gène-miARN et l'analyse du modèle de régulation des miARNs partagés sont tous implémentés dans la fonction `gdcCEAnalysis ()`.

```
ceOutput <- gdcCEAnalysis(lnc      = rownames(deLNC) ,
                          pc       = rownames(dePC) ,
                          lnc.targets = lncTarget ,
                          pc.targets  = pcTarget ,
                          rna.expr   = rnaExpr ,
                          mir.expr   = mirExpr)
```

#### 3 – 4 – 1– Calcul de test hypergéométrique

Nous avons effectué le test pour tester si un lncARN et un gène partagent de manière significative de nombreux miARNs.

$$p = 1 - \sum_{k=0}^m \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

M est le nombre de miARN partagés, N est le nombre total de miARN dans la base de données, n est le nombre de miARN ciblant lncARN, K est le nombre de miARN ciblant le gène codant la protéine.

### 3 – 4 – 2– Régulation similaire

Nous avons défini un score de similarité de régulation pour vérifier la similitude entre la corrélation d'expression miARNs-lncARN et la corrélation d'expression miARN-gène.

$$\text{Regulation similarity score} = 1 - \frac{1}{M} \sum_{k=1}^M \left[ \frac{|\text{corr}(m_k, l) - \text{corr}(m_k, g)|}{|\text{corr}(m_k, l)| + |\text{corr}(m_k, g)|} \right]^M$$

M est le nombre total de miARN partagés, le kème miARN partagé, corr (mk, l) et corr (mk, g) représente la corrélation de Pearson entre miARN-lncARN et miARN-gene, respectivement

### 3 – 4 – 3– Corrélation de sensibilité

La corrélation de sensibilité est définie par (Paci et al. 2014) pour mesurer si la corrélation entre un lncARN et un gène est médiée par un miARN dans le triplet lncARN-miARN-gène. Nous prenons la moyenne de tous les triplets d'une paire lncRNA-ARNm et leurs miARNs partagés comme corrélation de sensibilité entre un lncRNA et le gène sélectionnés.

$$\text{Sensitivity correlation} = \text{corr}(l, g) - \frac{1}{M} \sum_{k=1}^M \frac{\text{corr}(l, g) - \text{corr}(m_k, l)\text{corr}(m_k, g)}{\sqrt{1 - \text{corr}(m_k, l)^2} \sqrt{1 - \text{corr}(m_k, g)^2}}$$

M est le nombre total de miARN partagés, k est le kème miARN partagé, corr (l, g), corr (mk, l) et corr (mk, g) représente la corrélation de Pearson entre lncARN-gene, miARN-lncARN, miRNA-gene respectivement.

Des bases de données telles que starBase v2.0 (Li et al. 2014), miRcode (Jeggari Marks and Larsson. 2012) et mirTarBase release 7.0 (Chou et al. 2017) sont utilisées pour collecter les interactions miARN-gène et miARN-lncARN. Les identifiants géniques de ces bases de données sont mis à jour avec la dernière annotation Ensembl 90 du génome humain et les noms des miARNs sont mis à jour avec la nouvelle version des identifiants miRBase (Zhou RS et al. 2019).

Les interactions lncARN-miARN et miARN-gène ont été signalées par la fonction `gdcExportNetwork ()` et visualisées par Cytoscape version 3.6.0 (Otasek D et al. 2019). Les edges ont été importées en tant que réseau et les nœuds en tant que table d'entités.

```
ceOutput2 <- ceOutput[ceOutput$hyperPValue<0.01 &
  ceOutput$corPValue<0.01 & ceOutput$regSim != 0,]

edges <- gdcExportNetwork(ceNetwork = ceOutput2, net = 'edges')
nodes <- gdcExportNetwork(ceNetwork = ceOutput2, net = 'nodes')

write.table(edges, file='edges.txt', sep='\t', quote=F)
write.table(nodes, file='nodes.txt', sep='\t', quote=F)
```

## IV – Résultats

### 1– Identification des gènes et des miARNs différentiellement exprimés

Nous avons exploré 1082 gènes (593 gènes hautement exprimés et 489 gènes faiblement exprimés) (Figure 15 A. B), les 8 principaux gènes hyper exprimés sont DSCAML1, PEA15, ALDOC, PTPRT, LUZP2, NDRG2, MASP1 et BCAN (Figure 16 A) , d'autre part, les 8 principaux gènes faiblement exprimés sont VEGFA, NAP1L1, COL1A1, NKD2, NNMT, ITGA2, ANXA1 et LOX. (Figure 17 A) (Tableau 2). Parmi les gènes identifiés pour les gliomes de bas grade, on note une expression différente des gènes entre astrocytomes et oligodendrogliomes. Les gènes MASP1, LOX, ITGA2, LUZP2, ANXA1, COL1A1 et NNMT sont plus exprimés chez les patients atteints d'astrocytomes que chez les patients atteints d'oligodendrogliomes. Cependant, les patients atteints d'oligodendrogliomes sont caractérisés par une forte expression des gènes DSCAML1, NDRG2, ALDOC, BCAN, PEA15, PTPRT, NAP1L1 et NKD2 par rapport aux patients atteints d'astrocytomes. Néanmoins, le gène VEGFA est exprimé de la même manière pour les deux sous-types de tumeurs (figure 16 A et figure 17 A). En effet, nous avons identifié 1954 miARNs (1343 miARNs hautement exprimés et 611 miARNs faiblement exprimés) qui s'expriment de manière similaire entre les deux sous-types de gliomes de bas grade dont les astrocytomes et les oligodendrogliomes (Figure 18).

### 2 – Analyse des voies signalisation des gènes différentiellement exprimés

Nous avons annoté les gènes différentiellement exprimés pour fournir des informations systématiques et complètes sur les voies et les fonctions biologiques, cette analyse a montré que les changements dans les processus biologiques (BP) qui est un complexe de changements dans la granularité de la cellule, étaient significativement enrichis en déméthylation de l'histone lysine, déméthylation des histones, déméthylation des protéines, désalkylation des protéines, déméthylation et initiation de la traduction. Les changements dans le composant cellulaire (CC) des DEG, qui sont les structures complexes à partir desquelles les cellules sont formées, ont été principalement enrichis en complexe histone méthyltransférase et en méthyltransférase. Les changements dans la fonction moléculaire (MF) qui module l'activité d'un complexe de gènes

tels que les régulateurs enzymatiques et les régulateurs de canal, ont été principalement enrichis en activité histone déméthylase, activité déméthylase et activité dioxygénase (figure 1C).

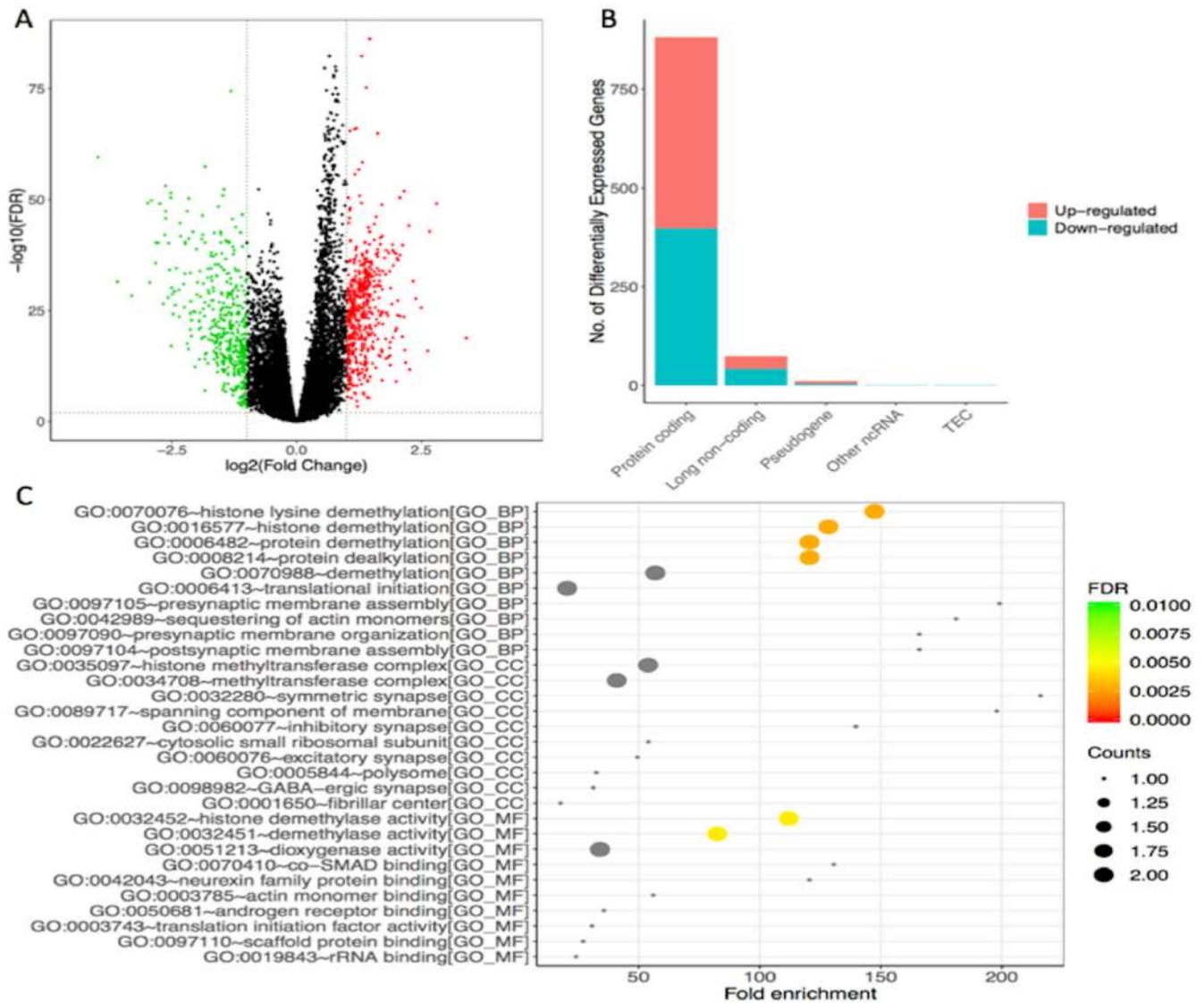


Figure 15 : (A) Diagramme de volcan des DEG. (B) Diagramme à colonnes des DEG. (C) Diagramme à points de l'analyse d'enrichissement des OG

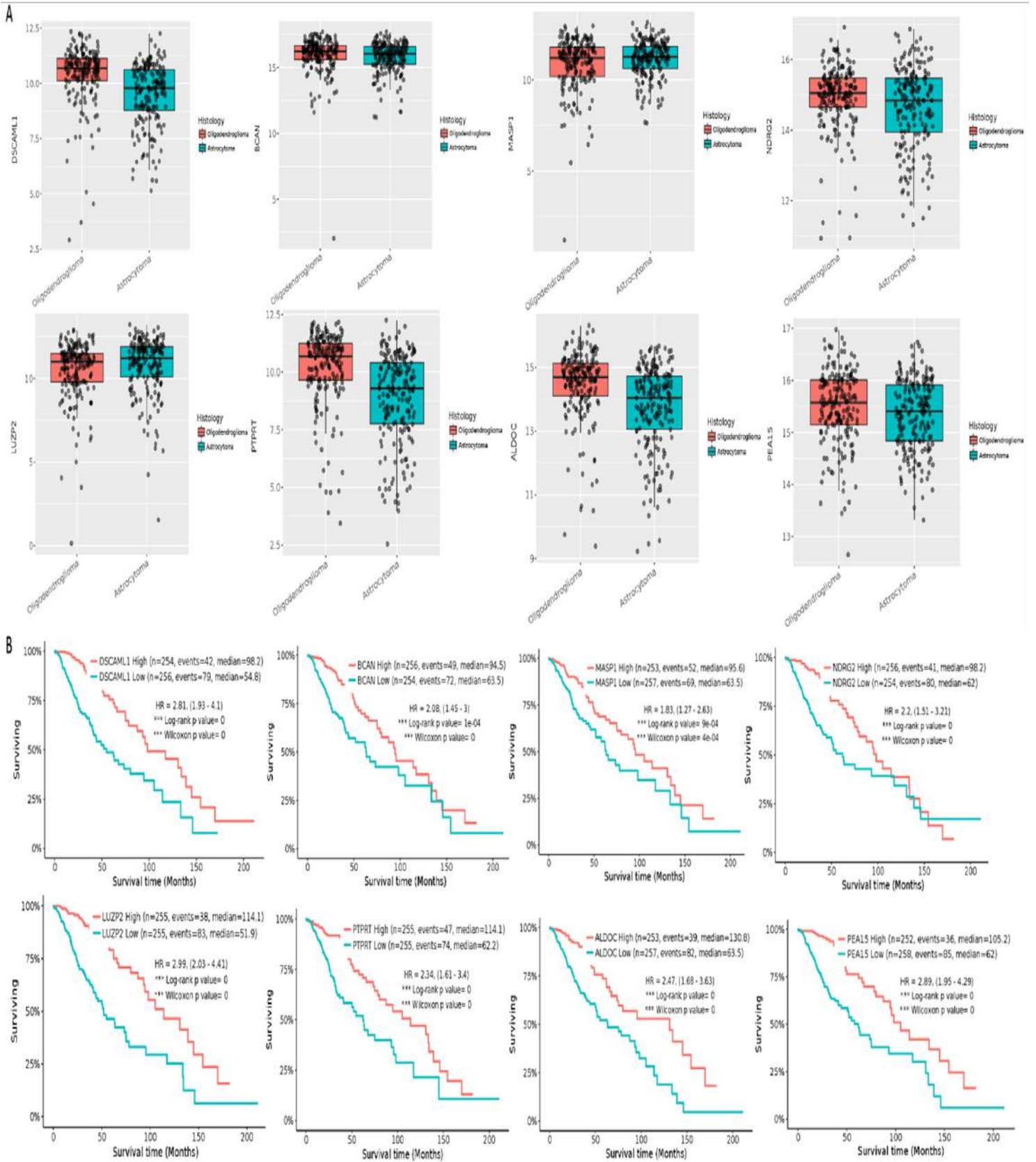


Figure 16 : (A) Les gènes hautement exprimés. (B) Courbes de Kaplan-Meier pour la survie globale des patients en fonction de gènes hautement exprimés

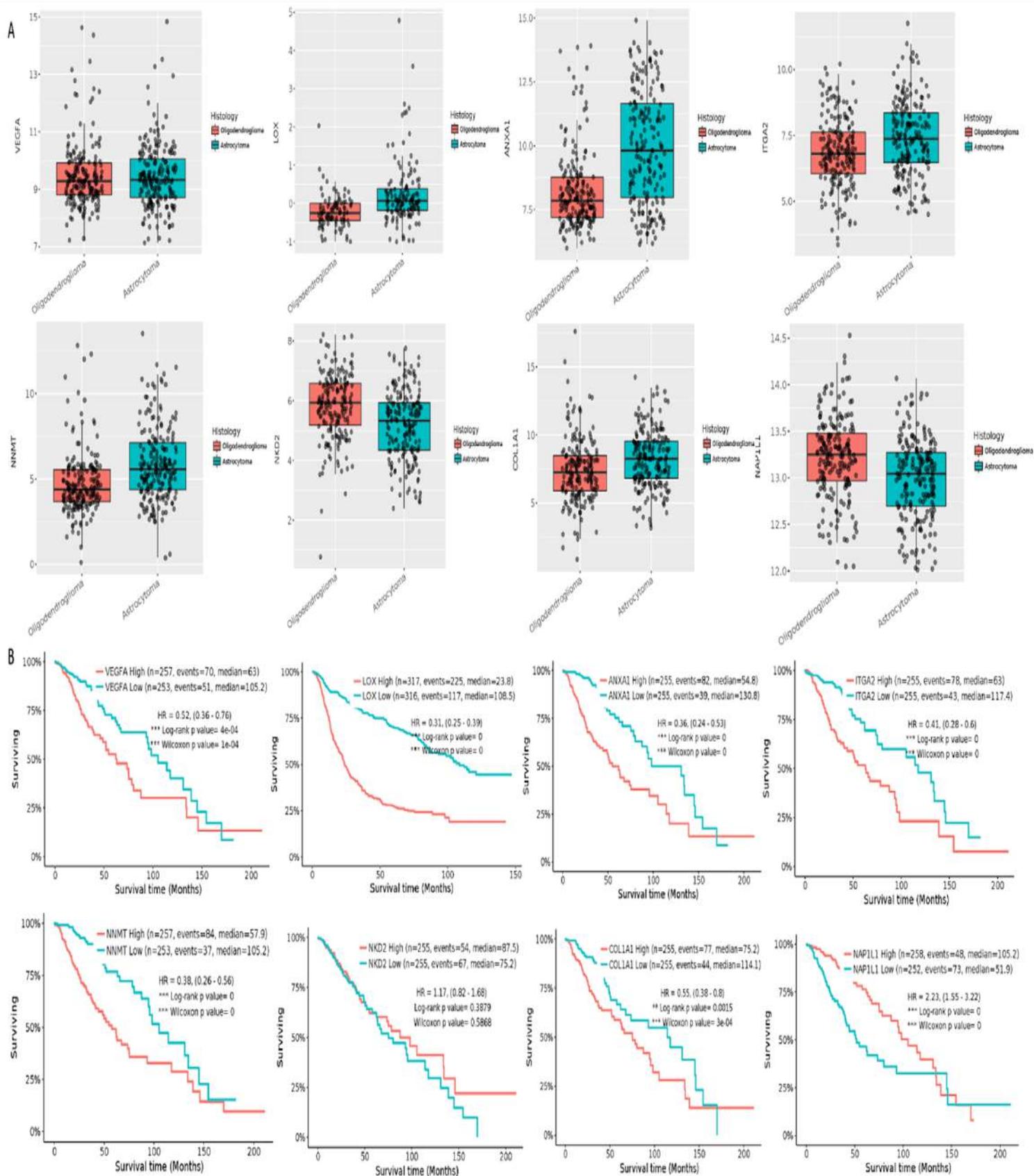


Figure 17 : (A) les gènes faiblement exprimés. (B) Courbes de Kaplan-Meier pour la survie globale des patients selon les gènes à faible expression

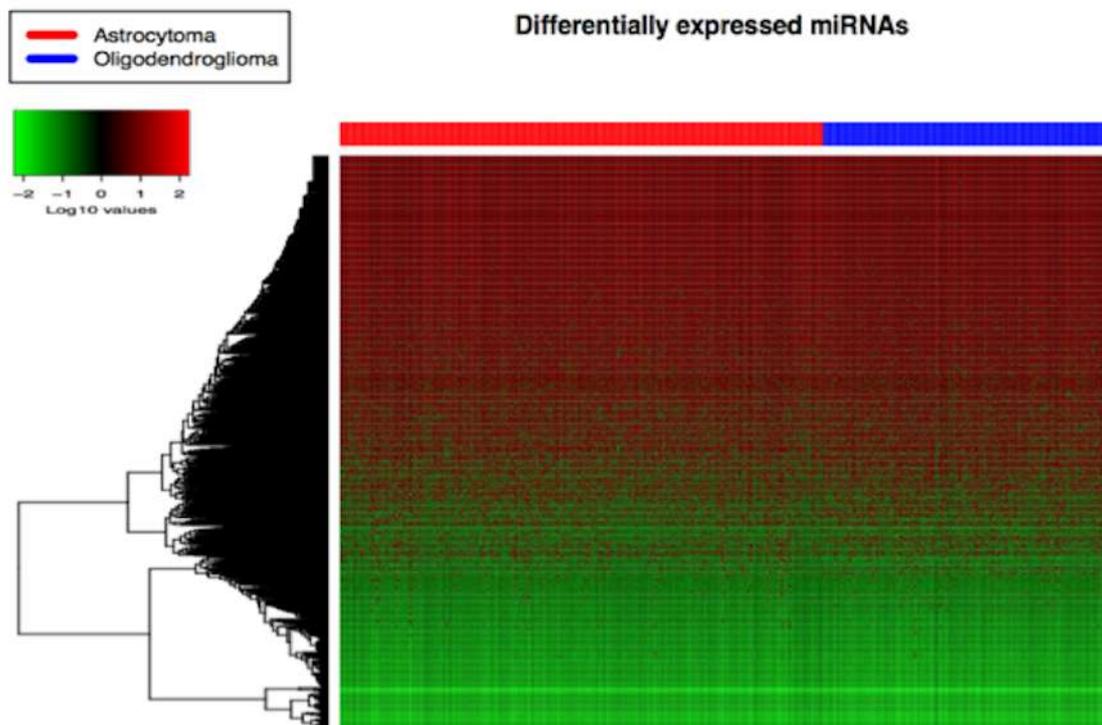


Figure 18 : Heatmap des miARNs différentiellement exprimés impliqués dans les gliomes de bas grade, incluant les astrocytomes et les oligodendroglomes

### 3 – Analyse de la survie globale des patients

L'analyse de la survie globale des patients en fonction des gènes différentiellement exprimés a montré que les patients qui ont une faible expression des gènes LOX et NAP1L1 présentent une longue survie, en revanche, les patients qui sont caractérisés par une faible expression des gènes NKD2 et COL1A1 et une forte expression du gène NAP1L1 montrent une faible survie.

Pour les miARNs, la survie globale des patients a été analysée selon les 8 premiers miARNs hautement exprimés (hsa-mir-18a-5p, hsa-mir-155-5p, hsa-mir-128-3p, hsa-mir-21-5p, hsa-mir-124-3p, hsa-mir-137, hsa-mir-451a, hsa-mir-221-3p) et 8 premiers miARNs faiblement exprimés (hsa-mir-31-5p, hsa-mir-222-3p, hsa-mir-17-5p, hsa-mir-330-5p, hsa-mir-330-3p, hsa-mir-346, hsa-mir-125b-5p, hsa-mir-16-5p) qui ont été exprimés de manière significativement différentielle dans les tests de survie (Figure 19 ; Tableau 3). La valeur pronostique des miARNs identifiés, a été examinée par la méthode Kaplan–Meier et le test du log-rank selon les caractéristiques cliniques, il a été constaté que hsa-mir-18a-5p, hsa-mir-330-3p, hsa-mir-15-5p et hsa-mir-137 ont un meilleur pronostique que les autres miARNs identifiés.

	Genes	Ensembl ID	Log2(Mean expression)	P-value	FDR
highly expressed genes	DSCAML1	ENSG00000177103	3.91	5.28e-05	1.69e-02
	PEA15	ENSG00000162734	4.88	4.24e-04	2.57e-03
	ALDOC	ENSG00000109107	2.97	1.09e-04	1.89e-07
	PTPRT	ENSG00000196090	8.42	1.47e-12	2.32e-05
	LUZP2	ENSG00000187398	5.71	1.36e-09	4.41e-11
	NDRG2	ENSG00000165795	9.13	3.17e-07	3.28e-02
	MASP1	ENSG00000127241	10.11	1.82e-02	2.16e-05
	BCAN	ENSG00000132692	6.01	1.22e-04	1.90e-08
low expressed genes	VEGFA	ENSG00000112715	3.31	1.37e-06	2.07e-12
	NAP1L1	ENSG00000187109	2.52	2.71e-06	2.15e-02
	COL1A1	ENSG00000108821	5.39	3.29e-04	6.19e-04
	NKD2	ENSG00000145506	8.07	2.31e-12	2.40e-06
	NNMT	ENSG00000166741	6.85	1.85e-02	5.16e-09
	ITGA2	ENSG00000169429	4.61	1.11e-17	2.76e-09
	ANXA1	ENSG00000135046	8.29	2.72e-09	6.13e-05
	LOX	ENSG00000113083	7.37	3.15e-01	1.66e-03

Tableau 2 : Les 8 gènes les plus exprimés et les moins exprimés dans les gliomes de bas grade

	miRNAs	Log2(Mean expression)	P-value	FDR
highly expressed miRNA	hsa-mir-18a-5p	2.16	3.87e-05	2.76e-04
	hsa-mir-155-5p	5.09	1.25e-14	1.11e-04
	hsa-mir-128-3p	10.84	1.32e-02	1.43e-04
	hsa-mir-21-5p	14.65	1.56e-09	4.68e-05
	hsa-mir-124-3p	8.12	1.09e-02	4.90e-02
	hsa-mir-137	5.72	1.38e-04	7.31e-01
	hsa-mir-451a	9.97	2.03e-02	2.03e-02
	hsa-mir-221-3p	6.13	2.29e-07	3.30e-06
low expressed miRNAs	hsa-mir-31-5p	0.74	1.29e-03	5.68e-03
	hsa-mir-222-3p	3.93	1.66e-07	1.23e-05
	hsa-mir-17-5p	8.04	1.13e-01	4.54e-01
	hsa-mir-330-5p	5.55	1.31e-01	2.07e-01
	hsa-mir-330-3p	4.19	1.68e-01	3.21e-01
	hsa-mir-346	2.85	1.75e-06	2.11e-08
	hsa-mir-125b-5p	11.79	3.34e-02	9.79e-02
	hsa-mir-16-5p	8.39	1.38e-05	1.53e-04

Tableau 3 : Les 8 miARNs les plus exprimés et les moins exprimés dans les gliomes de bas grade

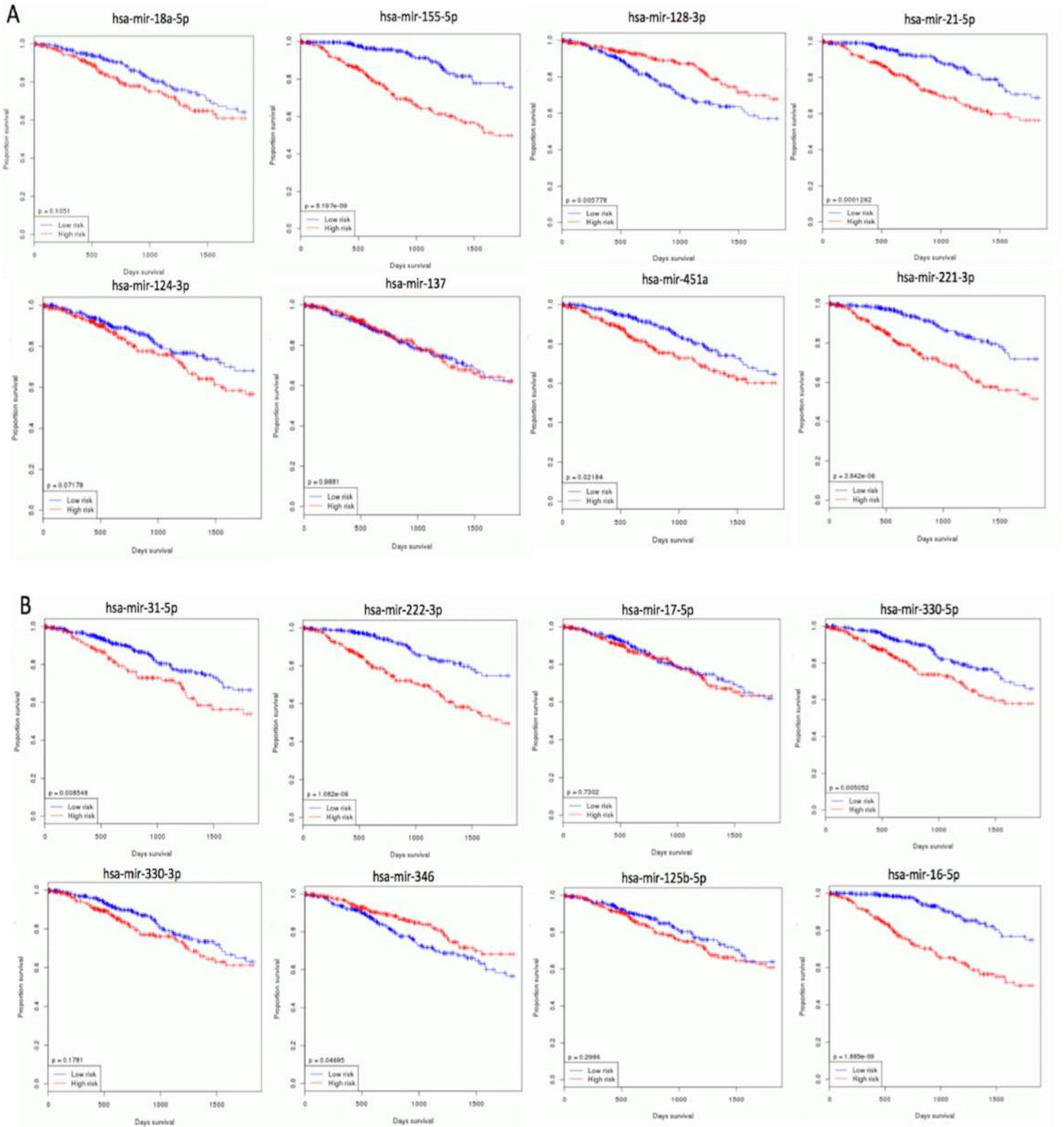


Figure 19: Courbes de Kaplan-Meier pour la survie globale des patients en fonction des miARNs fortement exprimés (A) et des miARNs faiblement exprimés (B)

#### 4 – Construction des réseaux des gènes et de miARNs

Nous avons construit deux réseaux, le premier (figure 20A) contient des gènes faiblement exprimés et le second (figure 20B) contient des gènes hautement exprimés. Ces réseaux sont caractérisés par plusieurs types d'interactions entre les gènes identifiés, on retrouve dans le réseau (Fig 20A) des interactions physiques (couleur rose), des domaines protéiques partagés (couleur gris foncé), la co-localisation des gènes (couleur bleue), les gènes impliqués dans la même voie (couleur gris clair) et les relations fonctionnelles prédites entre les gènes (couleur orange).

Le réseau (figure 20B) est caractérisé par un autre type d'interaction comme la co-expression des gènes (couleur violette) et l'interaction génétique entre les gènes (couleur verte).

Nous avons construit le réseau basé sur les modèles d'expression des miARNs chez les patients atteints de LGG (Figure 21). Pour les miARNs hautement exprimés, un total de 8 nœuds miARNs, 22 nœuds de gènes et 90 bords ont été identifiés comme des profils exprimés différemment, de l'autre côté, nous avons 8 nœuds miARNs, 22 nœuds lncARNs et 60 bords ont été identifiés comme des profils exprimés différemment. Pour les miARNs faiblement exprimés, un total de 8 nœuds miARNs, 22 nœuds de gènes et 80 bords ont été identifiés comme des profils exprimés différemment, mais l'autre partie du réseau à 8 miARNs, 22 nœuds lncARNs et 55 bords ont été identifiés comme des profils exprimés différemment.





## V - Discussion

Le développement d'approches omiques telles que la transcriptomique et l'épigénétique ouvre de nouveaux champs dans la compréhension de la carcinogenèse. Le contenu global de l'expression de l'ARN et des miARNs impliqués dans les gliomes de bas grade téléchargés à partir de la base de données TCGA, a été analysé à l'aide de différents types de méthodes à grande échelle, de différents nombres d'échantillons de tumeurs analysés, ainsi que de divers échantillons de référence (tissu cérébral non tumoral). Dans cette étude, nous avons identifié les gènes différentiellement exprimés impliqués dans les gliomes de bas grade dont les astrocytomes et les oligodendrogliomes, nous trouvons des gènes plus exprimés dans les astrocytomes par rapport aux oligodendrogliomes tels que LUZP2, LOX, NNMT et MASP1, et des gènes qui sont fortement exprimés dans les oligodendrogliomes comme NKD2, PTPRT, ALDOC et BCAN. Nous avons identifié les voies et les fonctions dans lesquelles l'expression différentielle des gènes est impliquée. Des études antérieures ont montré que certaines des voies annotées étaient associées à d'autres types de cancer tels que le cancer du poumon, le médulloblastome, le cancer du sein et le cancer de la vessie (Otasek D et al. 2019 ; Li J et al. 2017 ; Janardhan A et al. 2018). L'analyse de survie globale a montré que les patients qui ont une expression élevée du gène NAP1L1 et une faible expression des gènes NKD2 et COL1A1 ont une faible survie, contrairement aux patients qui sont caractérisés par une faible expression des gènes LOX et NAP1L1 avec une longue survie. En effet, nous avons identifié 16 miARNs, significativement liés à la survie globale des patients atteints de LGG. Cependant, il devient de plus en plus évident que le cancer peut rarement être attribué à une seule mutation génomique. Par exemple, les gènes sont associés à d'autres types de biomarqueurs dans les voies de signalisation biologique sous la forme de «réseaux».

Le réseau qui relie les gènes différentiellement exprimés identifiés montre plusieurs types d'interactions, le lien entre les gènes NKD2 et FUS est une interaction physique dans laquelle deux produits géniques interagissent dans une interaction protéine-protéine, le lien de co-localisation entre le VEGFA et Gènes KDR, signifie que les deux gènes sont liés s'ils sont tous deux exprimés dans le même tissu ou si leurs produits géniques sont identifiés dans le même emplacement cellulaire. L'interaction des deux gènes COL1A1 et GH1, qui participent à la même réaction, se retrouve également dans une voie de métabolisme des glucides. Nous avons un autre type d'interaction que les relations fonctionnelles prédites entre les gènes comme dans le cas entre le gène NAP1L1 avec les gènes HP1BP3 et H2AFJ. La liaison de la co-expression des gènes révèle que les deux gènes ont des niveaux d'expression similaires, elle se présente à

la fois entre le gène DSCAML1 et le gène PTPRT et entre le gène LUZP2 et le PEA15. On note la présence d'une interaction génétique entre le gène BCAN et les deux gènes MLC1 et GPR68, ce qui indique que les gènes sont fonctionnellement associés si les effets de la perturbation d'un gène se sont avérés modifiés par des perturbations d'un second gène. Ce nouveau raisonnement suggère que tous les transcrits devraient être combinés via des ARM partagés (éléments de réponse miARN), ce qui pourrait clarifier les interactions entre les biomarqueurs et leurs implications pour la maladie. Une analyse approfondie des miARNs peut clarifier les progrès du LGG, nous avons donc construit un réseau d'ARN basé sur des données TCGA à grande échelle. De nombreuses études ont rapporté que les miARNs forment des réseaux de régulation complexes dans le développement cellulaire, la différenciation et l'homéostasie, ils interviennent en se liant à d'autres ARN pour réguler leurs structures secondaires, les ciblant pour l'épissage, le transport, la traduction et la dégradation. La dérégulation de la fonction des miARNs est associée à un nombre croissant de maladies humaines, en particulier le gliome (Zhong ME et al. 2019 ; Zhou RS et al. 2019 ; Lei L et al. 2019). Chaque réseau contient des MRE pour une combinaison de différents miARNs et peuvent avoir un impact sur plusieurs gènes et cibler les lncRNAs ( Zhong ME et al. 2019 ; Papageorgiou I et al. 2017).

Dans cette étude, 16 miARNs interagissent avec 44 gènes et 44 lncRNAs. Certains miARN sont cités dans d'autres études comme hsa-mir-18a-5p qui a été identifié par (Chen X et al. 2018) comme régulateur de l'expression génique dans les gliomes du tronc cérébral. (Roberta M, et al. 2019) ont montré que hsa-mir-155-5p a des effets anti-apoptotiques et régule la sensibilisation du gliome aux médicaments anti-tumoraux. Ensuite (Huo L et al.2019) ont identifié que miR-128-3p comme un inhibiteur de la croissance du gliome en attaquant la protéine pentraxine-1 neuronale. Également (Wu Q et al.2018) ont mentionné que le has-mir-124-3p réprime la croissance cellulaire et la motilité cellulaire en ciblant EphA2 dans le gliome. (Mahmoudi E et Cairns MJ. 2017) ont révélé que le mir-137 contribue à plusieurs types de cancer, tels que le neuroblastome et le glioblastome multiforme. (Roberta M et al.2019) ont cité que les gliomes expriment deux miARNs, mir-155-5p et mir-221-3p, et que ceux-ci sont liés à des effets anti-apoptotiques. Les miARNs affichés dans le réseau interagissent avec des gènes et des lncRNAs, le gène PTEN est lié à has-mir-221- 3p, has-mir-18a-5p, has-mir-21-5p, has-mir-155-5p et has-mir128-3p, dans d'autres études (Wong ST et al.2012) ont montré que l'inhibition du mir-21 augmentait la chimiosensibilité du glioblastome humain au taxol, activait les caspases -3 et -9, et supprimait la croissance des cellules de gliome indépendamment de régulation du PTEN. Nous trouvons également (Zhang Z et al.2017) qui ont indiqué que miR-137 peut cibler le 3'-UTR du récepteur du facteur de croissance épidermique (EGFR) pour

réduire sa traduction protéique, ce qui entraîne la suppression de la signalisation EGFR dans les cellules de glioblastome. Concernant le gène SIRT1, il est lié dans notre réseau avec hsa-mir-221-3p, hsa-mir-18a-5p, hsa-mir-124-3p et hsa-mir-155-5p, d'autre part (Romeo SG, et coll. 2016) ont suggéré un rôle pour miRNA-34a, miRNA-132 et miRNA-217 dans le contrôle épigénétique de Sirt-1 au cours de la gliomagenèse et de la progression. Car le gène CCND1 est lié avec hsa-mir-330- 5p, hsa-mir-16-5p et hsa-mir-17-5p et dans l'étude de (Munoz JL et al. 2019), ils ont identifié mir-19, 23a / b, 93b, 193 et 373 comme cible du CCND1. Il existe également des interactions entre les miARNs et les lncARNs tels que KCNQ1OT1 qui agit comme un régulateur de mir-504.33-35, (Zhou RS et al. 2019) a découvert que KCNQ1OT1 pourrait provoquer le développement d'un carcinome épidermoïde de la langue et inhibent la sensibilité de la tumeur au cisplatine (Zhang JX et al. 2013 ; Zhang K et al. 2015). Un autre biomarqueur important, MALAT1, qui joue un rôle dans la promotion de la prolifération cellulaire et inactive la signalisation ERK / MAPK (Li J et al. 2018 ; Han Y et al. 2016) et HOTAIR qui se lie à EZH2 et favorisent la croissance cellulaire (Zhang JX et al. 2013 ; Zhang K et al. 2015). XIST favorise une migration cellulaire et interagit avec mir-152 dans notre réseau, il interagit avec hsa-mir-137, hsa-mir-21-5p, hsa-mir-18a-5p et Gas5, il exerce des fonctions suppressives de tumeur dans les cellules de gliome humain en ciblant les interactions des miARNs. Certains de ces transcriptions clés ont également été trouvés dans d'autres réseaux impliqués dans d'autres types de cancer (Zhao X et al. 2015). Cependant, l'étude était basée sur une analyse bio-informatique utilisant des données publiques du TCGA et aucune autre validation expérimentale n'a été effectuée. La recherche sur les transcriptions est toujours en cours de développement et de nombreux aspects doivent être améliorés. Plusieurs miARNs doivent être cliniquement validés pour améliorer le diagnostique et le pronostique des gliomes de bas grade.

## VI – Conclusion

En conclusion, cette étude a appliqué avec succès l'analyse bio-informatique des données épitranscriptomiques pour établir un réseau de miARN spécifique pour les gliomes de bas grade, ce qui devrait nous aider à mieux comprendre la régulation des gènes et des lncARNs par les miARNs et à avoir une idée claire des interactions entre ces biomarqueurs participant à la gliomagenèse de ces tumeurs cérébrales.

## **Profil global des modifications épigénétiques impliquées dans les gliomes de bas grade.**

### **I – Introduction**

Le gliome de bas grade est un groupe de tumeurs neuroépithéliales hétérogènes résultant des cellules gliales du système nerveux central commun chez les adultes. La prise en charge de ces tumeurs cérébrales reste un défi à cause de la nature infiltrante de la tumeur qui empêche le traitement curatif après une résection totale ou supratotale (Wen PY et al. 2016).

Au cours de la dernière décennie, les mutations des gènes régulateurs épigénétiques ont été identifiées comme les principaux biomarqueurs des sous-types des gliomes de bas grade ayant des caractéristiques cliniques distinctes. En plus des altérations génétiques connues, l'initiation et la progression des gliomes de bas grade est étroitement associées aux aberrations épigénétiques, y compris la méthylation de l'ADN, les modifications des histones, l'ARN non codant (lncARN), la microARN (miARN) et le remodelage de la chromatine (Baylin SB et al. 2016). À cet égard, initier une altération génétique telle que des mutations IDH peut coopérer avec des événements épigénétiques ultérieurs au cours de la progression tumorale (Mazor T et al. 2015 ; Lai A et al. 2011 ; Watanabe T et al. 2009). Ces altérations épigénétiques sont utilisées comme biomarqueurs et jouent un rôle central dans la classification des gliomes (WHO 2016) et les décisions de traitement. L'objectif de cette étude est de présenter un aperçu globale sur les biomarqueurs de méthylation de l'ADN impliqué dans les gliomes de bas grade et leurs impact sur les facteurs de transcription, l'expression génique globale et la conformation de la chromatine. En plus de résumer l'impact de la méthylation globale de l'ADN sur la pathobiologie des gliomes, nous avons analysé les données Chip-Seq de 516 patients téléchargés à partir des bases de données TCGA pour l'étude des modifications des histones.

### **II – L'hyperméthylation d'ADN induite par la mutation IDH**

Il existe trois ADN méthyltransférases (DNMT), DNMT1, 3a et 3b. DNMT1 est essentiel pour maintenir la méthylation de l'ADN (Hermann A et al. 2004), tandis que DNMT3a est impliquée dans la méthylation des promoteurs de l'expression des gènes et DNMT3b qui est impliquée dans la méthylation des séquences entourant les centromères (Ehrlich M et al. 2013). Les sites d'ADN méthylés peuvent être liés et reconnus par les protéines de liaison à la méthyl cytosine (Baylin SB et al. 2011). Cependant, les protéines de translocation à onze (TET) participent activement à la déméthylation de l'ADN par l'oxydation de 5mC en 5-hydroxyméthylcytosine (5hmC), 5-formylcytosine (5fC) et 5-carboxylcytosine (5caC) (Wu X et al. 2017). Les gliomes

avec des mutations dans les gènes métaboliques IDH1 ou IDH2 présentent une signature frappante d'hyperméthylation de l'ADN qui est complètement différente des gliomes IDHwt et elle a été appelé phénotype de méthylateur de l'îlot Glioma CpG (G-CIMP) (Noushmehr H et al. 2010). Les mutations IDH sont des lésions précoces dans le développement des gliomes et se regroupent dans le site de liaison au substrat de ces enzymes, au codon 132 d'IDH1 ou au codon 172 d'IDH2, respectivement (Hartmann C et al. 2009 ; Pietrak B et al. 2011) . Ces mutations favorisent une réaction néomorphique catalysant la conversion de l' $\alpha$ -cétoglutarate en D-2-hydroxyglutarate (2HG) (Figuroa ME et al. 2010). Le 2HG agit comme un oncométabolite en s'accumulant à des concentrations élevées qui inhibent les enzymes dépendantes de l' $\alpha$ -cétoglutarate. Les enzymes dépendantes de l' $\alpha$ -cétoglutarate comprennent des modificateurs épigénétiques tels que l'enzyme TET2 impliquée dans la déméthylation de l'ADN ou l'histone déméthylase spécifique à la lysine KDM2A (Zhang YW et al. 2017 ; Scourzic L et al. 2015 ; Xu W et al. 2011).

### III – Modifications des histones dans les gliomes de bas grade

Une histone est un assemblage de huit petites protéines (octamère). L'enrôlement d'ADN autour de chaque octamère conduit à la condensation du brin d'ADN en chromatine (collier de perles). Les différentes modifications de condensation des chromatines permettent de moduler l'accessibilité aux gènes en fonction des besoins cellulaires. Les histones fonctionnent comme des commutateurs provoquant l'accessibilité des gènes. Il existe plusieurs formes d'histones avec des marques de méthylation différentes qui permettent le recrutement ou non de facteurs de transcription et donc de lecture génomique (Prensner R et al. 2011). Le nucléosome est composé de huit protéines histones, deux copies chacune de H2A, H2B, H3 et H4 et 147 paires de bases d'ADN. Le mutant IDH1 augmente les marques répressives telles que l'histone H3K9me2 / 3 déméthylée / triméthylée H3 (H3K9me2 / 3) et H3K27me3, ainsi que les marques actives telles que H3K4me2 / 3 dans les astrocytes humains immortalisés (Turcan S et al. 2012 ; Venneti S et al. 2013).

Dans notre étude, les ensembles de données Chip-Seq de 516 patients atteints de gliomes de bas grade ont été téléchargés à partir du TCGA (<https://gdc-portal.nci.nih.gov>) par le package TCGAAbiolinks R version 2.12.6 (Colaprico A et al. 2016).

```
downloadGEObedFiles(genome="hg19", destDir="hg19")
gsm <- hg19$gsm[sample(nrow(hg19), 10)]
downloadGSMbedFiles(gsm, destDir="hg19")
```

Le package ChIPseeker version 1.24.0 (Yu G et al. 2015), a été utilisé pour visualiser les changements d'histones qui sont enrichis sur les régions hypométhylées et hyperméthylées des amplificateurs, des promoteurs et une activation accrue des éléments génomiques (Figure 22), et concernant (Figure 23), elle indique que la plupart des pics qui chevauchent les sondes ne sont pas spécifiques au cerveau. Les réactions de méthylation en K9 et K27 sur l'histone 3 (H3) sont catalysées par les histones méthyltransférases et par le transfert de groupes méthyle qui répriment la transcription.

```
data("tagMatrixList")
plotAvgProf(tagMatrixList, xlim=c(-3000, 3000))
data("tagMatrixList")
plotAvgProf(tagMatrixList, xlim=c(-3000, 3000))
tagHeatmap(tagMatrixList, xlim=c(-3000, 3000), color=NULL)
peakAnnoList <- lapply(files, annotatePeak, TxDb=txdb,
```

Par exemple, la méthylation de H3K9 est un changement épigénétique qui marque l'hétérochromatine. Si la triméthylation (H3K9me3) est impliquée avec Suv39h, la mono- et la déméthylation (H3K9me1 et H3K9me2) sont principalement liées à l'histone lysine méthyltransférase G9a et à la protéine de type G9a (GLP) (Collins BE et al. 2019 ; Shinkai Y et al. 2011). BIX-01294 supprime l'activité enzymatique G9a et diminue les niveaux de H3K9me2 dans la chromatine (Underwood CJ et al. 2018). Cependant, l'activité homologue de l'histone méthyltransférase activatrice (EZH2) est modifiée par la mutation K27M et entraîne également une réduction des taux de H3K27me3 et une activation transcriptionnelle, entraînant ainsi des changements globaux de l'état épigénétique des cellules (Sturm D et al. 2012).

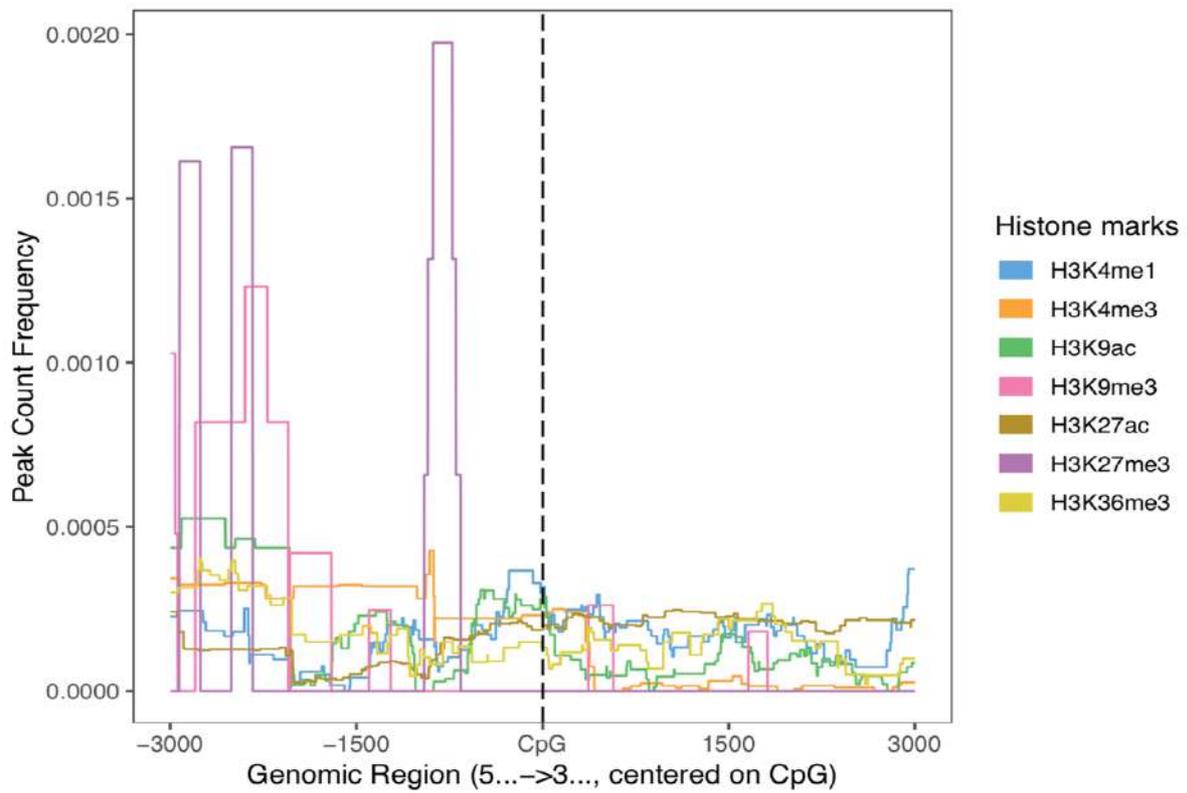


Figure 22: Profils moyens des marqueurs d'histones H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K9ac et H3K9me3

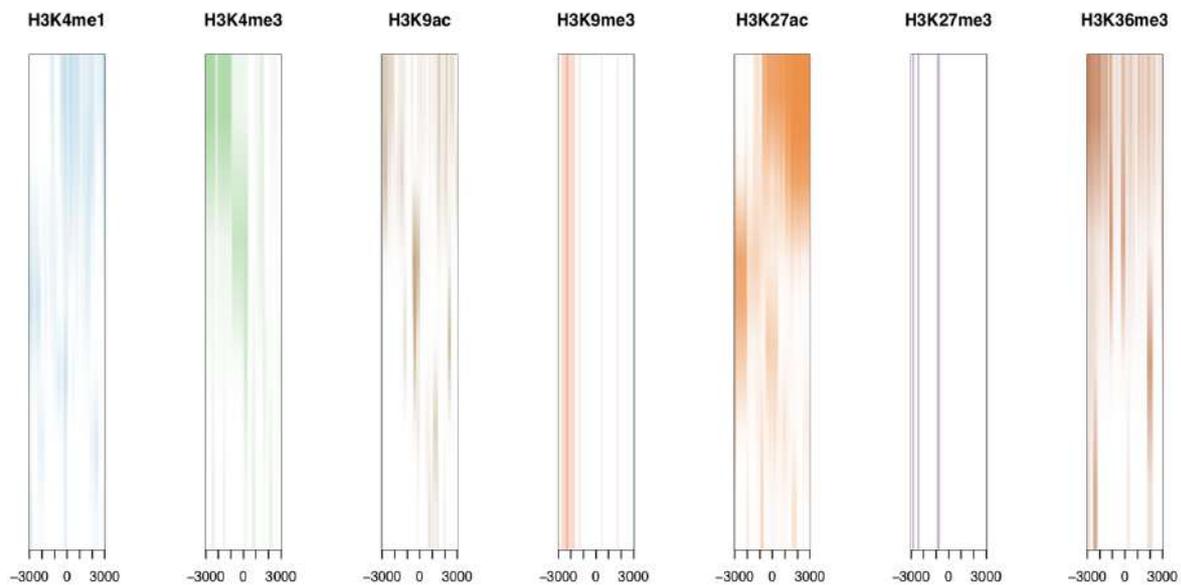


Figure 23: Carte thermique des marques d'histones H3K4me1, H3K4me3, H3K27ac, H3K9ac, H3K9me3, H3K27me3 et H3K36me3 pour les tissus cérébraux

#### **IV – La méthylation de l'ADN affectant les contacts de la chromatine**

La structure tridimensionnelle de la chromatine comprend certaines zones appelées domaines d'association topologique (TAD). Ces TAD contiennent des séquences d'ADN (plusieurs milliers à plusieurs millions de bases d'ADN) qui interagissent physiquement les unes avec les autres, suggérant que les gènes situés dans ces zones fonctionnent ensemble. Les extrémités des TAD sont maintenues par le facteur de liaison CCCTC (CTCF). (Flavahan WA et al. 2016) ont rapporté que l'hyperméthylation de l'ADN perturbe la liaison au CTCF, provoquant une expression aberrante d'oncogènes tels que PDGFRA. À l'inverse, un traitement in vitro avec un agent déméthylant de l'ADN réduit l'hyperméthylation de l'ADN au niveau du motif CTCF, qui est associée à l'occupation du CTCF et à la restauration de l'isolement PDGFRA (Verhaak RG et al. 2010). D'autres aberrations fréquentes dans les gliomes sont la perte de mutations fonctionnelles d'ATRX, qui code pour le régulateur de chromatine du type SWI / SNF (Cancer Genome Atlas Research. 2015 ; Jiao Y et al. 2012 ; Schwartzentruber J et al. 2012) . Les sites de liaison CTCF autour du gène SOX2 sont affectés par les changements de méthylation dans les cellules neurales mutantes IDH1 / TP53 / ATRX. il bloque soudainement la différenciation normale de ces cellules souches. On peut en déduire que les cellules souches neurales passent à une voie oncogénique par l'activation d'un facteur de transcription au mauvais moment. Cependant, des études suggèrent que la restauration de l'expression de SOX2 peut être suffisante pour sauver la différenciation des cellules souches neurales (Modrek AS et al. 2017).

#### **V- Conclusion**

L'analyse épigénétique de la méthylation de l'ADN doit être mise en œuvre comme un outil puissant pour caractériser l'hétérogénéité et éclairer les décisions cliniques. Il est important de noter que cette approche a également le potentiel de révéler des entités tumorales cérébrales précédemment non reconnues et de trouver des approches thérapeutiques efficaces contre les aberrations épigénétiques qui caractérisent les gliomes de bas grade.

## Conclusion générale

Comprendre la base moléculaire des gliomes de bas grade et identifier les biomarqueurs impliqués dans ce type de tumeur cérébrale est susceptible de tirer de nouvelles cibles ou voies d'intervention thérapeutique. Bien que les progrès dans ce domaine aient été limités jusqu'à présent, les initiatives fournissant des données moléculaires et clinico-pathologiques intégrées sur le gliome de bas grade vont accélérer les progrès de recherche neuro-oncologique. Les résultats collectifs de cette thèse soulignent d'abord l'importance et la nécessité de la caractérisation des astrocytomes et les oligodendrogliomes qui ont un aspect morphologique similaire mais deux voies moléculaires différentes en utilisant l'approche bio-informatique pour une meilleure compréhension des processus moléculaires et cellulaires impliqués dans la tumorigenèse, et pour plus d'efficacité de diagnostic, de traitement approprié et de pronostic amélioré.

Les gliomes de bas grade sont des néoplasmes infiltrants qui surviennent le plus souvent dans les hémisphères cérébraux des adultes et comprennent les astrocytomes et les oligodendrogliomes. En raison de leur nature hautement invasive, la résection neurochirurgicale complète est difficile, et la présence de la tumeur résiduelle entraîne une progression maligne. Un sous-ensemble de ces gliomes évoluera en glioblastome (gliomes de grade IV de l'OMS) en quelques mois, tandis que d'autres resteront stables pendant des années. De même, la survie varie largement, de 1 à 15 ans, certains gliomes de bas grade ont une sensibilité thérapeutique impressionnante.

Dans le passé, la classification des gliomes était définie par des caractéristiques histologiques, cependant, cette méthode de classification souffre d'une forte variabilité intra-observateur et inter-observateur, ne reflète pas l'hétérogénéité de la tumeur et elle ne permet pas de prédire adéquatement les résultats cliniques. Au cours des 3 dernières années, de nouvelles découvertes de biomarqueurs moléculaires ont entraîné une modification de la classification pathologique de tous les gliomes tels que les astrocytomes et les oligodendrogliomes. Les chercheurs et les cliniciens comptent de plus en plus sur la classification génétique pour guider la décision clinique. Des mutations dans IDH1 et IDH2 (deux gènes très similaires, ci-après dénommés collectivement IDH) caractérisent la majorité des gliomes de bas grade chez l'adulte et définissent un sous-type associé à un pronostic favorable. Dans le premier chapitre, on s'est intéressé aux algorithmes génomiques pour identifier les mutations tumorales et les outils de modélisation moléculaire pour prédire la structure tridimensionnelle des protéines, afin de faciliter la compréhension des mécanismes dynamiques impliqués dans la pathogenèse des

gliomes de bas grade y compris les oligodendrogliomes et les astrocytomes. Nous avons ciblé des mutations impliquées dans les gliomes de bas grade, en commençant par le séquençage des régions génomiques englobant l'exon 4 de l'isocitrate déshydrogénase 1 (IDH1) et l'isocitrate déshydrogénase 2 (IDH2) et les quatre exons (5-6 et 7-8) de TP53 à partir de 32 échantillons, suivi d'une analyse informatique pour étudier l'impact de ces mutations sur la structure et la fonction de 3 protéines IDH1, IDH2 et p53. Nous avons obtenues une mutation qui a un effet sur le site catalytique de la protéine IDH1 comme R132H et sur le site catalytique de la protéine IDH2 comme R172M. D'autres mutations de la p53 ont été identifiées comme K305N, qui est une mutation pathogène; R175H, qui est une mutation bénigne; et R158G, qui perturbe la conformation structurelle de la protéine suppresseur de tumeur. Les mutations dans IDH1, IDH2 et TP53 qu'on a identifié peuvent être la clé de la progression tumorale car elles ont un effet sur la fonction de la protéine comme les mutations R132H dans IDH1 et R172M dans IDH2, qui changent la fonction de l'enzyme alpha-cétoglutarate, ou R158G dans TP53, qui affecte la structure de la protéine générée, d'où leur importance dans la compréhension de la gliomagenèse et pour un diagnostic plus précis complémentaire aux tests de pathologie anatomique.

Au cours de ma thèse, on a identifié d'autres biomarqueurs potentiels pour les gliomes de bas grade incluant les astrocytomes et les oligodendrogliomes en analysant des données génomiques et épigénomiques. Nous avons utilisé une approche bio-informatique complète pour délimiter les fondements biologiques des gliomes de bas grade en se basant sur la nouvelle classification des tumeurs cérébrales de l'OMS 2016 qui incorpore des données moléculaires dans le diagnostic pathologique. Les signatures moléculaires du gliome de bas grade se prêtent à un système de classification biologiquement basé sur la pratique qui devrait améliorer la concordance entre les observateurs. Nous avons reclassé tous les patients inclus dans la base de données TCGA qui ont un diagnostic 2007 de l'OMS (astrocytome, oligodendrogliome et oligoastrocytome) en deux type de tumeur (astrocytome et oligodendrogliome) en réduisant le diagnostic d'«oligoastrocytome» et la confusion liée à sa prise en charge. De plus Notre analyse de la survie globale des patients a montré que les personnes qui avaient des gliomes de bas grade avec une mutation IDH avaient une survie globale plus longue que celles qui avaient des gliomes de bas grade avec IDH1 sauvage.

IDH1 et IDH2 sont des enzymes homodimériques qui catalysent la conversion de l'isocitrate en  $\alpha$ -cétoglutarate et produisent le nicotinamide adénine dinucléotide phosphate NADPH. Des mutations dans IDH1 se produisent à l'arginine 132, avec des mutations dans IDH2 se produisant à l'arginine 172 comme on a révélé dans le chapitre 1.

Les mutations IDH-1 et IDH-2 associées au gliome entraînent un gain de fonction des enzymes métaboliques isocitrate déshydrogénase 1 et 2. La protéine mutante perd son activité enzymatique normale et commence la production d'oncométabolite R-2-hydroxyglutarate, qui consomme du NADPH. Le R-2-HG inhibe également de manière compétitive les dioxygénases dépendantes de l' $\alpha$ -cétoglutarate, avec une inhibition épigénétique ultérieure des histones lysine déméthylases et de la famille translocalisation (TET) de la 5-méthyl cytosine hydroxylase. Ainsi, l'IDH mutant modifie l'environnement redox inhérent en modifiant le rapport NADPH / NADP, la déplétion du NAD étant une caractéristique de sensibilité métabolique des tumeurs mutantes IDH1.

La présence et l'identification de l'oncométabolite R-2-HG, nous conduit à l'hypothèse actuelle de mutations IDH étant oncogènes. La différenciation cellulaire restreinte et l'hyperméthylation de l'ADN sont les caractéristiques des mutants associés à l'IDH, et il a un effet sur l'inhibition de la déméthylation des histones et de l'hyperméthylation de l'ADN comme on a signalé dans le chapitre 4.

Les gliomes ont un phénotype de méthylateur d'îles CpG (G-CIMP) avec des aberrations épigénomiques distinctives et étendues. Celles-ci ont été directement corrélées aux mutations IDH1 avec une hyperméthylation généralisée dans les régions promotrices du gène faisant inhiber l'expression des facteurs de différenciation cellulaire. Les altérations épigénomiques résultant du mutant IDH1 activent des programmes d'expression génique clé dans les gliomes au G-CIMP.

La prévalence des mutations IDH dans le gliome de bas grade nous a guider à cibler leurs conséquences métaboliques et épigénomiques en aval, telles que G-CIMP. On a trouvé que les astrocytomes sont caractérisé par une méthylation au niveau du gène MYT1L, et au niveau du gène PLCG1 chez les oligodendrogliomes.

Les mutations dans les gènes TP53 et ATRX pour les astrocytomes et les gènes CIC et FUBP1 pour les oligodendrogliomes n'ont été impliquées que récemment dans le gliome de bas grade. Les astrocytomes diffus ont la signature moléculaire d'une mutation IDH 1/2, avec des mutations TP53 et ATRX. Le type le plus fréquent d'astrocytome est la combinaison moléculaire de la perte de la fonction du gène ATRX et de la mutation IDH.

Le facteur ATRX de remodelage de la chromatine dépendant du TP53 ( $\alpha$  thalassémie / retard mental lié à l'X) a été lié aux processus nucléaires d'expression génique, de réparation, de réplication et de recombinaison de l'ADN. ATRX code une protéine essentielle de remodelage de la chromatine (ATRX), dont la carence a été associée à une dérégulation épigénomique et à un dysfonctionnement des télomères. Un déficit en ATRX est associé à une instabilité

génomique généralisée et peut induire la mort cellulaire dépendant de p53 chez les astrocytomes.

Les pertes alléliques dans les régions de 19q et 1p qui abritent la mutations des gènes CIC et FUBP1 sont une caractéristique commune des oligodendrogliomes. La surexpression de FUBP1 peut stimuler l'expression de MYC, le complexe protéique FUBP1-PUF60 régule négativement l'expression de MYC. Les mutations qui inactivent les gènes CIC et FUBP1 dans une proportion significative d'oligodendrogliomes peuvent fournir un aperçu de la pathogénie moléculaire de ces tumeurs. La spécificité de ces biomarqueurs et leurs prévalences avec la mutation IDH soutiennent les rôles centraux dans l'oncogénèse et plaident en faveur d'une caractérisation approfondie des réseaux de signalisation associés pour faciliter le développement thérapeutique comme on a démontré dans le chapitre 2. Cependant, l'utilisation de la classification moléculaire peut être intégrée à d'autres données cliniques, neuro-imagerie et pathologiques pour concevoir une stratégie de traitement pour les patients individuels.

L'analyse du profilage de l'expression génique est une méthode utile pour identifier les gènes liés à la tumeur notamment les astrocytomes et les oligodendrogliomes, du diagnostique moléculaire à la classification pathologique, de l'évaluation thérapeutique à la prédiction du pronostic. Ces dernières années, de grandes échelles de profilage génétique ont été faites pour identifier le nombre écrasant de gènes, des analyses statistiques compliquées et systémiques devraient être effectuées pour permettre la validation significative des biomarqueurs identifiés.

La technologie RNA-seq permet l'identification détaillée des isoformes géniques, des événements de translocation, des variations nucléotidiques et des modifications post-transcriptionnelles des bases. Dans le chapitre 3, on a atteint notre objectif par l'identification des gènes différentiellement exprimés au niveau des astrocytomes et des oligodendrogliomes. Ces gènes sont sélectionnés sur la base d'une combinaison de seuil de changement d'expression et de seuil de score, qui sont généralement basés sur les valeurs P générées par la modélisation statistique.

Le niveau d'expression de chaque unité d'ARN est mesuré par le nombre de fragments séquencés qui correspondent au transcrit, ce qui devrait être en corrélation directe avec son niveau d'abondance.

Nous avons produit diagramme de volcan qui présente la différence d'expression génique et nous avons sélectionné 16 gènes avec un degré de connectivité plus élevé parmi les gènes sélectionnés au début. Ensuite, nous avons analysé les voies du composant cellulaire (CC), du processus biologique (BP), de la fonction moléculaire (MF), ainsi que l'analyse de la survie

globale des patients (OS), puis, nous avons établi un réseau d'interaction des DEG. Ces analyses ont révélé qu'il y a des gènes qui sont plus exprimés dans les astocytomes que les oligodendrogliomes tels que LUZP2, LOX, NNMT et MASP1, et des gènes qui sont fortement exprimés dans les oligodendrogliomes comme NKD2, PTPRT, ALDOC et BCAN. Nous avons aussi montré que les patients qui ont une expression élevée du gène NAP1L1 et une faible expression des gènes NKD2 et COL1A1 ont une faible survie, contrairement aux patients qui sont caractérisés par une faible expression des gènes LOX et NAP1L1 ont une longue survie. Cette étude démontre que pour les ensembles de données avec un grand nombre de gènes, le package Limma est bien adapté pour détecter les gènes différentiellement exprimés et que la modélisation des données de comptage de gènes comme une distribution log normale avec les pseudo-dénombrements appropriés ce qui nous a permis la création et la mise en place d'un workflow bio-informatique précis et efficace pour l'analyse des données transcriptomique.

Un nouveau profil d'expression génique pour aborder la caractérisation moléculaire des gliomes a mis en évidence des altérations épigénétiques. Il a été rapporté que le profilage d'expression des microARNs (miARN) affecte la fonction des gènes impliqués dans la gliomagenèse, la croissance tumorale, la prolifération, l'apoptose et la régulation post-transcriptionnelle des anti-oncogènes. En conséquence, il est suggéré que le profilage des miARNs peut être un moyen disponible pour identifier l'origine du gliome et son pronostic. Cependant, il devient de plus en plus évident que le cancer peut rarement être attribué à une seule mutation génomique ou l'expression de certains gènes. Par exemple, les gènes sont associés à d'autres types de biomarqueurs dans les voies de signalisation biologique sous la forme de «réseaux».

L'analyse des voies et des réseaux présente un certain nombre d'avantages par rapport à l'analyse des données génomiques au niveau des gènes individuels. Premièrement, ces techniques agrègent les événements moléculaires sur plusieurs gènes dans la même voie ou voisinage de réseau, augmentant ainsi la probabilité que les événements passent un seuil de détection statistique et réduisant le nombre d'hypothèses testées. Deuxièmement, les résultats sont souvent plus faciles à interpréter car les altérations génomiques sont liées à des concepts familiers tels que le cycle cellulaire ou l'apoptose. Troisièmement, des mécanismes causaux potentiels peuvent être identifiés, par exemple en prédisant un microARN particulier ou un facteur de transcription qui explique les différences d'expression entre les échantillons de tumeurs et les témoins. Quatrièmement, les résultats obtenus à partir d'ensembles de données connexes peuvent devenir plus comparables puisque les informations de chemin permettent une interprétation dans un espace de caractéristiques commun. Enfin, les techniques informatique facilitent l'intégration de divers intrants tels que les données génomiques, transcriptomiques et

épigénomique dans une vue unifiée de la biologie tumorale, améliorant ainsi le pouvoir statistique et interprétatif. Ceci nous a guidé à identifier 16 miARNs, significativement liés aux gliomes de bas grade et à proposer un nouveau réseau de miARNs qui combinent à la fois des gènes et des lncARNs afin de mieux caractériser ces tumeurs cérébrales, mieux comprendre la gliomagenèse, et établir un workflow bio-informatique adapté à l'analyse et l'identification de ce genre de biomarqueurs. La plupart des miARNs surexprimés dans les tumeurs étaient impliqués dans les processus de propagation du gliome. Ils semblaient agir comme un oncogène impliqué dans les processus de transformation maligne du gliome. Nos analyses computationnelles ont révélé une expression élevée de hsa-mir-21 dans la plupart des gliomes par rapport au cerveau normal, et le hsa-mir-211 avait le potentiel de réguler environ 70 gènes cibles communs et d'exercer un effet coopératif sur la fonction via la voie AKT.

Dans le chapitre 4, on a démontré que les gliomes de bas grade ne partagent pas que certaines aberrations de séquence d'ADN spécifiques, mais ils présentent également des altérations communes de la chromatine. Parmi les caractéristiques épigénétiques spécifiques, on trouve la méthylation de l'ADN et leur interaction avec les modifications d'histones qui peuvent affecter la liaison des facteurs de transcription, l'expression globale des gènes et la conformation de la chromatine. La mutation IDH modifie non seulement la méthylation globale de l'ADN, mais modifie également les modifications des histones. Les gliomes G-CIMP associés au statut de mutation IDH présentent un phénotype hyperméthylé, provoquant un état épigénétique répressif et un blocage de la différenciation. Les changements dans la méthylation de l'ADN des gliomes de bas grade modifient l'affinité de liaison de plusieurs facteurs de transcription, y compris SOX. On trouve aussi le CTCF qui est un isolant important pour le maintien de la structure de la chromatine. Les changements de méthylation de l'ADN dans les sites de liaison du CTCF conduisent à des altérations globales de la structure de la chromatine.

La méthylation globale de l'ADN dans les gliomes pourrait bientôt devenir un marqueur diagnostique et pronostique. Comme les changements dans la méthylation de l'ADN sont si bien décrits, ce sera la prochaine étape, pour lier la méthylation de l'ADN au niveau d'une seule molécule avec d'autres altérations de l'ADN.

Ces études ont conduit à la mise en place d'une approche d'analyse bio-informatique intégrative pour les données génomiques, transcriptomiques et épigénétiques des gliomes de bas grade.

Ces tumeurs cérébrales représentent une classe très hétérogène de tumeurs humaines. Les mécanismes moléculaires à l'origine du développement des gliomes sont mal compris. L'application de ces méthodes innovantes de biologie computationnelle et analyse bio-

informatique représente une stratégie prometteuse pour mieux comprendre le développement du gliome.

Les simulations bio-informatiques nous donne des indices sur les biomarqueurs impliqués dans les mécanismes moléculaire et les voies de signalisation des gliomes de bas grade. Nos perspectives sont le développement des méthodes et des outils pour l'analyse intégrative de l'ensembles des données moléculaires des gliomes, la validation des biomarqueurs identifiés et le test de divers scénarios des origines de ces tumeurs.

## Références

Adzhubei I, Jordan DM, Shamil R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;7:7–20.

Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248-9.

Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer [published correction appears in *Nature*. 2013 Oct 10;502(7470):258. Imielinsk, Marcin. *Nature*. 2013;500(7463):415-421.

Ali M, Khan SA, Wennerberg K, Aittokallio T. Global proteomics profiling improves drug sensitivity prediction: results from a multi-omics, pan-cancer modeling approach. *Bioinformatics*. 2018;34(8):1353-1362.

Alvarez-Dominguez JR, Bai Z, Xu D, Yuan B, Lo KA, Yoon MJ, et al. De novo reconstruction of adipose tissue transcriptomes reveals long non-coding RNA regulators of brown adipocyte development. *Cell Metab*. 2015;21:764–76.

Ambrosino L, Tangherlini M, Colantuono C, et al. Bioinformatics for Marine Products: An Overview of Resources, Bottlenecks, and Perspectives. *Mar Drugs*. 2019;17(10):576.

Amir M, V Kumar, M Taj, et al. Investigation of deleterious effects of nsSNPs in the POT1 gene: a structural genomics-based approach to understand the mechanism of cancer development. *J Cell Biochem*. 2018;120:10281-10294.

Arnes L, Akerman I, Balderes DA, Ferrer J, Sussel L. *Betalinc1* encodes a long noncoding RNA that regulates islet beta-cell formation and function. *Genes Dev*. 2016;30:502–7.

Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016;44(W1):W344-W350.

Ashraf MA, Sarfraz M. Biology and evolution of life science. *Saudi J Biol Sci.* 2016;23(1):S1-S5.

Aubrey BJ, Strasser A, Kelly GL. Tumor-Suppressor Functions of the TP53 Pathway. *Cold Spring Harb Perspect Med.* 2016;6(5):a026062.

Avellaneda MD, Gonzalez J, Eric R, et al. Molecular mechanisms of isocitrate dehydrogenase 1 (IDH1) mutations identified in tumors: The role of size and hydrophobicity at residue 132 on catalytic efficiency. *J Biol Chem.* 2017;292:7971-7983.

Awad AJ, Burns TC, Zhang Y, et al. Targeting MET for glioma therapy. *Neurosurgical Focus.* 2014;37:10.

Babae N, Bourajjaj M, Liu Y, et al. Systemic miRNA-7 delivery inhibits tumor angiogenesis and growth in murine xenograft glioblastoma. *Oncotarget.* 2014;5:6687-700.

Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 2000;28(1):45-48.

Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet.* 2011;12(11):745-755.

Bardella C, Al-Dalahmah O, Krell D, et al. Expression of IDH1 R132H in the Murine Subventricular Zone Stem Cell Niche Recapitulates Features of Early Gliomagenesis. *Cancer Cell.* 2016;30(4):578-594.

Barres R, Zierath JR. The role of diet and exercise in the transgenerational epigenetic landscape of T2DM. *Nat Rev Endocrinol.* 2016;12:441-51.

Baumert BG, Hegi ME, van den Bent MJ, et al. Temozolomide chemotherapy versus radiotherapy in high-risk low-grade glioma (EORTC 22033-26033): a randomised, open-label, phase 3 intergroup study. *Lancet Oncol.* 2016;17(11):1521-1532.

Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum Mol Genet.* 2001;10:687–92.

Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer.* 2011;11(10):726-734.

Baylin SB, Jones PA. Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol.* 2016;8(9):a019505.

Bebek G, Koyutürk M, Price ND, Chance MR. Network biology methods integrating biological data for translational science. *Brief Bioinform.* 2012;13(4):446-459.

Beiko J, Suki D, Hess KR, et al. IDH1 mutant malignant astrocytomas are more amenable to surgical resection and have a survival benefit associated with maximal surgical resection. *Neuro Oncol.* 2014;16(1):81-91.

Ben Chorin A, Masrati G, Kessel A, et al. ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Sci.* 2020;29(1):258-267.

Bendahou MA, Arrouchi H, Lakhili W, et al. Computational Analysis of IDH1, IDH2, and TP53 Mutations in Low-Grade Gliomas Including Oligodendrogliomas and Astrocytomas. *Cancer Inform.* 2020;19:1176935120915839.

Bendahou MA, Ibrahim A, Boutarbouch M. Identification of Potential Biomarkers for Low-Grade Gliomas by Analyzing the Genomics Profiles and CpG Island Methylator Phenotype of Patients in the TCGA Database. *Int J Appl Biol Pharm.* (2020):188-199.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2010;38(Database issue):D46-D51.

Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;28(1):235-242.

Bilkey GA, Burns BL, Coles EP, Mahede T, Baynam G, Nowak KJ. Optimizing Precision Medicine for Public Health. *Front Public Health*. 2019;7:42.

Borger DR, Tanabe KK, Fan KC, et al. Frequent Mutation of Isocitrate Dehydrogenase (IDH)1 and IDH2 in Cholangiocarcinoma Identified Through Broad-Based Tumor Genotyping. *The Oncologist*. 2012;17:72-9.

Botling J, Edlund K, Lohr M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clin Cancer Res*. 2013;19(1):194-204.

Brada M, Viviers L, Abson C, et al. Phase II study of primary temozolomide chemotherapy in patients with WHO grade II gliomas. *Ann Oncol*. 2003;14(12):1715-1721.

Buckner J, Giannini C, Eckel-Passow J, et al. Management of diffuse low-grade gliomas in adults - Use of molecular diagnostics. *Nat Rev Neurol*. 2017;13:340-351.

Buckner JC, Gesme D Jr, O'Fallon JR, et al. Phase II trial of procarbazine, lomustine, and vincristine as initial therapy for patients with low-grade oligodendroglioma or oligoastrocytoma: efficacy and associations with chromosomal abnormalities. *J Clin Oncol*. 2003;21(2):251-255.

Buckner JC, Shaw EG, Pugh SL, et al. Radiation plus Procarbazine, CCNU, and Vincristine in Low-Grade Glioma. *N Engl J Med*. 2016;374(14):1344-1355.

Bunda S, Heir P, Metcalf J, et al. CIC protein instability contributes to tumorigenesis in glioblastoma. *Nat Commun*. 2019;10(1):661.

Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*. 2013;501(7467):338-345.

Can T. Introduction to bioinformatics. *Methods Mol Biol*. 2014;1107:51-71.

Cancer Genome Atlas Research Network, Brat DJ, Verhaak RG, et al. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med.* 2015;372(26):2481-2498.

Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013;45(10):1113-1120.

Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 2013, 45, 1113–1120.

Cannataro M, Guzzi PH, Tradigo G, Veltri P. Biological Databases. In: Kasabov N. (eds) *Springer Handbook of Bio-/Neuroinformatics.* Springer Handbooks. Springer, Berlin, Heidelberg. 2014

Capper D, Weissert S, Balsl J, et al. Characterization of R132H mutation-specific IDH1 antibody binding in brain tumors. *Brain Pathol.* 2010;20(1):245-254.

Castellano A, Bello L, Michelozzi C, et al. Role of diffusion tensor magnetic resonance tractography in predicting the extent of resection in glioma surgery. *Neuro Oncol.* 2012;14(2):192-202.

Chang EF, Potts MB, Keles GE, et al. Seizure characteristics and control following resection in 332 patients with low-grade gliomas. *J Neurosurg.* 2008;108(2):227-235.

Chaturvedi A, Araujo Cruz MM, Jyotsana N, et al. Mutant IDH1 promotes leukemogenesis in vivo and can be specifically targeted in human AML. *Blood.* 2013;122(16):2877-2887.

Chen R. On bioinformatic resources. *Genomics Proteomics Bioinformatics.* 2015;13(1):1-3.

Chen X, Dong D, Pan C, et al. Identification of Grade-associated MicroRNAs in Brainstem Gliomas Based on Microarray Data. *J Cancer.* 2018;9:4463-4476.

Chen Y, Li C, Tan C, Liu X. Circular RNAs: a new frontier in the study of human diseases. *J Med Genet.* 2016;53:359–65.

Chinthakunta N, Cheemanapalli S, Chinthakunta S, et al. A new insight into identification of in silico analysis of natural compounds targeting GPR120. *Netw Model Anal Heal Informatics Bioinforma.* 2018;7:8.

Chou CH, Shrestha S, Yang CD, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 2018;46(D1):D296-D302.

Christensson, P. (2009, October 27). Database Definition. Retrieved 2020, Aug 31, from <https://techterms.com>

Colaprico A, Silva TC, Olsen C, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 2016;44(8):e71.

Collins BE, Greer CB, Coleman BC, Sweatt JD. Histone H3 lysine K4 methylation and its role in learning and memory. *Epigenetics Chromatin.* 2019;12(1):7.

Conesa A, Beck S. Making multi-omics data accessible to researchers. *Sci Data.* 2019;6(1):251.  
Cruz IN, Coley HM, Kramer HB, et al. Proteomics Analysis of Ovarian Cancer Cell Lines and Tissues Reveals Drug Resistance-associated Proteins. *Cancer Genomics Proteomics.* 2017;14(1):35-51.

De Silva NHND. Relational Databases and Biomedical Big Data. *Methods Mol Biol.* 2017;1617:69-81.

De Souza CF, Sabedot TS, Malta TM, et al. A Distinct DNA Methylation Shift in a Subset of Glioma CpG Island Methylator Phenotypes during Tumor Recurrence. *Cell Rep.* 2018;23(2):637-651.

Desai M, Chauhan JB. Predicting the functional and structural consequences of nsSNPs in human methionine synthase gene using computational tools. *Syst Biol Reprod Med.* 2019;65:288-300.

- Diniz WJ, Canduri F. REVIEW-ARTICLE Bioinformatics: an overview and its applications. *Genet Mol Res.* 2017;16(1):10.4238/gmr16019645. Published 2017 Mar 15
- Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature.* 2012;489(7414):101-108.
- Dong C, Yuan Z, Li Q, et al. The clinicopathological and prognostic significance of TP53 alteration in K27M mutated gliomas: an individual-participant data meta-analysis. *Neurol Sci.* 2018;39:1191-1201.
- Douw L, Klein M, Fagel SS, et al. Cognitive and radiological effects of radiotherapy in patients with low-grade glioma: long-term follow-up. *Lancet Neurol.* 2009;8(9):810-818.
- Duarte CW, Willey CD, Zhi D, et al. Expression signature of IFN/STAT1 signaling genes predicts poor survival outcome in glioblastoma multiforme in a subtype-specific manner. *PLoS One.* 2012;7(1):e29653.
- Duffau H, Capelle L, Denvil D, et al. Functional recovery after surgical resection of low grade gliomas in eloquent brain: hypothesis of brain compensation. *J Neurol Neurosurg Psychiatry.* 2003;74(7):901-907.
- Duffau H, Taillandier L. New concepts in the management of diffuse low-grade glioma: Proposal of a multistage and individualized therapeutic approach. *Neuro Oncol.* 2015;17(3):332-342.
- Duffau H. Brain plasticity: from pathophysiological mechanisms to therapeutic applications. *J Clin Neurosci.* 2006;13(9):885-897.
- Eckel-Passow JE, Lachance DH, Molinaro AM, et al. Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *N Engl J Med.* 2015;372(26):2499-2508.
- Ehrlich M, Lacey M. DNA hypomethylation and hemimethylation in cancer. *Adv Exp Med Biol.* 2013;754:31-56.

Eicher T, Kinnebrew G, Patt A, et al. Metabolomics and Multi-Omics Integration: A Survey of Computational Methods and Resources. *Metabolites*. 2020;10(5):202.

El-Awady RA, Hersi F, Al-Tunaiji H, et al. Epigenetics and miRNA as predictive markers and targets for lung cancer chemotherapy. *Cancer Biol Ther*. 2015;16(7):1056-1070.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.

Englot DJ, Han SJ, Berger MS, Barbaro NM, Chang EF. Extent of surgical resection predicts seizure freedom in low-grade temporal lobe brain tumors. *Neurosurgery*. 2012;70(4):921-928.

Er TK, Chen CC, Chien YH, et al. Development of a feasible assay for the detection of GAA mutations in patients with Pompe disease. *Clin Chim Acta*. 2014;429:18–25.

Feinberg MW, Moore KJ. MicroRNA regulation of atherosclerosis. *Circ Res*. 2016;118:703–20.

Ferguson JF, Wang TJ. Branched-Chain Amino Acids and Cardiovascular Disease: Does Diet Matter?. *Clin Chem*. 2016;62(4):545-547.

Figueroa ME, Abdel-Wahab O, Lu C, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*. 2010;18(6):553-567.

Figueroa ME, Abdel-Wahab O, Lu C, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*. 2010;18(6):553-567.

Fisher BJ, Hu C, Macdonald DR, et al. Phase 2 study of temozolomide-based chemoradiation therapy for high-risk low-grade gliomas: preliminary results of Radiation Therapy Oncology Group 0424. *Int J Radiat Oncol Biol Phys*. 2015;91(3):497-504.

Flavahan WA, Drier Y, Liau BB, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 2016;529(7584):110-114.

Gallo Cantafio ME, Grillone K, Caracciolo D, et al. From Single Level Analysis to Multi-Omics Integrative Approaches: A Powerful Strategy towards the Precision Oncology. *High Throughput*. 2018;7(4):33.

Gameiro GR, Sinkunas V, Liguori GR, Auler-Júnior JOC. Precision Medicine: Changing the way we think about healthcare. *Clinics (Sao Paulo)*. 2018;73:e723.

George AP, Kuzel TM, Zhang Y, Zhang B. The Discovery of Biomarkers in Cancer Immunotherapy. *Comput Struct Biotechnol J*. 2019;17:484-497.

Gibson G. On the utilization of polygenic risk scores for therapeutic targeting. *PLoS Genet*. 2019;15(4):e1008060.

Giesbertz P, Daniel H. Branched-chain amino acids as biomarkers in diabetes. *Curr Opin Clin Nutr Metab Care*. 2016;19(1):48-54.

Ginsburg GS, Phillips KA. Precision Medicine: From Science To Value. *Health Aff (Millwood)*. 2018;37(5):694-701.

Giskeødegård GF, Bertilsson H, Selnæs KM, et al. Spermine and citrate as metabolic biomarkers for assessing prostate cancer aggressiveness. *PLoS One*. 2013;8(4):e62375.

Gittleman H, Sloan AE, Barnholtz-Sloan JS. An independently validated survival nomogram for lower-grade glioma. *Neuro Oncol*. 2020;22(5):665-674.

Goetz LH, Schork NJ. Personalized medicine: motivation, challenges, and progress. *Fertil Steril*. 2018;109(6):952-963.

Gondi V, Hermann BP, Mehta MP, Tomé WA. Hippocampal dosimetry predicts neurocognitive function impairment after fractionated stereotactic radiotherapy for benign or low-grade adult brain tumors. *Int J Radiat Oncol Biol Phys*. 2013;85(2):348-354.

Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. *Transl Cancer Res.* 2015;4(3):256-269.

Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32(18):2847-2849.

Guha T, Malkin D. Inherited TP53 Mutations and the Li-Fraumeni Syndrome. *Cold Spring Harb Perspect Med.* 2017;7(4):a026187.

Guhathakurta D, Sheikh NA, Meagher TC, Letarte S, Trager JB. Applications of systems biology in cancer immunotherapy: from target discovery to biomarkers of clinical outcome. *Expert Rev Clin Pharmacol.* 2013;6(4):387-401.

Guo B, Zhang Q, Wang H, et al. KCNQ1OT1 promotes melanoma growth and metastasis. *Aging (Albany NY).* 2018

Guo J, Lin F, Zhang X, et al. NetLand: quantitative modeling and visualization of Waddington's epigenetic landscape using probabilistic potential. *Bioinformatics.* 2015;33:1583-1585.

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature.* 2010;464:1071–6.

Gusyatiner O, Hegi ME. Glioma epigenetics: From subclassification to novel treatment options. *Semin Cancer Biol.* 2018;51:50–58.

Gut P, Verdin E. The nexus of chromatin regulation and intermediary metabolism. *Nature.* 2013;502:489–98.

H. Wickham. *ggplot2: elegant graphics for data analysis.* Springer New York, 2009.

Hagen MS, Lee EK. BIOSPIDA: A Relational Database Translator for NCBI. *AMIA Symposium.* 2010 Nov;2010:422-426.

Hager GL, McNally JG and T. Misteli. Transcription Dynamics. *Molecular Cell*, vol. 35, no. 6, pp. 2009:741–753.

Han J, Puri RK. Analysis of the cancer genome atlas (TCGA) database identifies an inverse relationship between interleukin-13 receptor  $\alpha 1$  and  $\alpha 2$  gene expression and poor prognosis and drug resistance in subjects with glioblastoma multiforme. *J Neurooncol.* 2018;136(3):463-474.

Han Y, Wu Z, Wu T, et al. Tumor-suppressive function of long noncoding RNA MALAT1 in glioma cells by downregulation of MMP2 and inactivation of ERK/MAPK signaling. *Cell Death Dis.* 2016;7:2123.

Hasin Y, Seldin M, Lusic A. Multi-omics approaches to disease. *Genome Biol.* 2017;18(1):83.

Hebart MN, Dickter AH, Kidder A, et al. THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLoS One.* 2019;14(10):e0223792.

Hegi ME, Diserens AC, Gorlia T, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med.* 2005;352(10):997-1003.

Helmy M, Crits-Christoph A, Bader GD. Ten Simple Rules for Developing Public Biological Databases. *PLoS Comput Biol.* 2016;12(11):e1005128.

Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem.* 2004;279(46):48350-48359.

Hieke S, Benner A, Schlenl RF, Schumacher M, Bullinger L, Binder H. Integrating multiple molecular sources into a clinical risk prediction signature by extracting complementary information. *BMC Bioinformatics.* 2016;17(1):327.

Hillmer RA. Systems biology for biologists. *PLoS Pathog.* 2015;11(5):e1004786.

Hoang-Xuan K, Capelle L, Kujas M, et al. Temozolomide as initial treatment for adults with low-grade oligodendrogliomas or oligoastrocytomas and correlation with chromosome 1p deletions. *J Clin Oncol.* 2004;22(15):3133-3138.

Horbinski C, Kofler J, Kelly LM, et al. Diagnostic use of IDH1/2 mutation analysis in routine clinical testing of formalin-fixed, paraffin-embedded glioma tissues. *J Neuropathol Exp Neurol*. 2009;68:1319-25.

Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14:R115.

Hoshida R, Jandial R. World Health Organization Classification of Central Nervous System Tumors: An Era of Molecular Biology. *World Neurosurg*. 2016;94:561–562.

Hu T, Kumar Y, Shazia I, et al. Forward and reverse mutations in stages of cancer development. *Hum Genomics*. 2018;12(1):40.

Huang J, Yu J, Tu L, Huang N, Li H, Luo Y. Isocitrate Dehydrogenase Mutations in Glioma: From Basic Discovery to Therapeutics Development. *Front Oncol*. 2019;9:506.

Huang M, Parker MJ, Stubbe J. Choosing the right metal: case studies of class I ribonucleotide reductases. *J Biol Chem*. 2014;289(41):28104-28111.

Huang SC, Zhang L, Sung YS, et al. Recurrent CIC Gene Abnormalities in Angiosarcomas: A Molecular Study of 120 Cases With Concurrent Investigation of PLCG1, KDR, MYC, and FLT4 Gene Alterations. *Am J Surg Pathol*. 2016;40(5):645-655.

Huang TT, Hwang JK, Chen CH, et al. (PS)2: Protein structure prediction server version 3.0. *Nucleic Acids Res*. 2015;43:338-42.

Huang YT. Integrative modeling of multi-platform genomic data under the framework of mediation analysis. *Stat Med*. 2015;34(1):162-178.

Huberfeld G, Vecht CJ. Seizures and gliomas--towards a single therapeutic approach. *Nat Rev Neurol*. 2016;12(4):204-216.

Huo L, Wang B, Zheng M, et al. miR-128-3p inhibits glioma cell proliferation and differentiation by targeting NPTX1 through IRS-1/PI3K/AKT signaling pathway. *Exp Ther Med*. 2019;17:2921-2930.

Idbaih A, Ducray F, Dehais C, et al. SNP array analysis reveals novel genomic abnormalities including copy neutral loss of heterozygosity in anaplastic oligodendrogliomas. *PLoS One*. 2012;7(10):e45950.

International Agency for Research on Cancer; Kleihues P, Cavenee WK, eds. World Health Organization classification of tumours: pathology and genetics of tumours of the nervous system. Lyon, France: IARC Press, 2000.

International Agency for Research on Cancer; Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, editors. WHO classification of tumours of the central nervous system. 4th ed. Lyon, France: World Health Organization, 2016.

Ishii N, Ozaki K, Sato H, Mizuno H, Saito S, Takahashi A, et al. Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J Hum Genet*. 2006;51:1087–99

Jadhav A, Dash RC, Hirwani R, et al. Sequence and structure insights of kazal type thrombin inhibitor protein: Studied with phylogeny, homology modeling and dynamic MM/GBSA studies. *Int J Biol Macromol*. 2018;108:1045-1052.

Jaeckle K, Vogelbaum M, Ballman K, et al. ATCT-16: CODEL (ALLIANCE-N0577; EORTC-26081/2208; NRG-1071; NCIC-CEC-2): PHASE III RANDOMIZED STUDY OF RT VS. RT + TMZ VS. TMZ FOR NEWLY DIAGNOSED 1p/19q-CODELETED ANAPLASTIC GLIOMA. ANALYSIS OF PATIENTS TREATED ON THE ORIGINAL PROTOCOL DESIGN. *Neuro Oncol*. 2015;17(Suppl 5):v4-v5.

Janardhan A, Kathera C, Darsi A, et al. Prominent role of histone lysine demethylases in cancer epigenetics and therapy. *Oncotarget*. 2018;9:34429-34448.

Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*. 2012;28(15):2062-2063.

Jenkins RB, Blair H, Ballman KV, et al. A t(1;19)(q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma. *Cancer Res*. 2006;66(20):9852-9861.

Jiao Y, Killela PJ, Reitman ZJ, et al. Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas. *Oncotarget*. 2012;3(7):709-722.

Johnson BE, Mazar T, Hong C, et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science*. 2014;343(6167):189-193.

Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013;502(7471):333-339.

Kargar F, Mortazavi M, Savardashtaki A, et al. Genomic and protein structure analysis of the luciferase from the Iranian bioluminescent beetle, *Luciola* sp. *Int J Biol Macromol*. 2019;124:689-698.

Kauffman EC, Robinson BD, Downes MJ, et al. Role of androgen receptor and associated lysine-demethylase coregulators, LSD1 and JMJD2A, in localized and advanced human bladder cancer. *Mol Carcinog*. 2011;50:931-44.

Kaur P, Porras TB, Ring A, Carpten JD, Lang JE. Comparison of TCGA and GENIE genomic datasets for the detection of clinically actionable alterations in breast cancer. *Sci Rep*. 2019;9(1):1482.

Kazda T, Jancalek R, Pospisil P, et al. Why and how to spare the hippocampus during brain radiotherapy: the developing role of hippocampal avoidance in cranial radiotherapy. *Radiat Oncol*. 2014;9:139. Published 2014 Jun 16.

Keerthikumar S. An Introduction to Proteome Bioinformatics. *Methods Mol Biol*. 2017;1549:1-3.

Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW. DNA methylation as a biomarker for cardiovascular disease risk. *PLoS One*. 2010;5:e9692.

Kirk P, Griffin JE, Savage RS, Ghahramani Z, Wild DL. Bayesian correlated clustering to integrate multiple datasets. *Bioinformatics*. 2012;28(24):3290-3297.

Kizilbash SH, Giannini C, Voss JS, et al. The impact of concurrent temozolomide with adjuvant radiation and IDH mutation status among patients with anaplastic astrocytoma. *J Neurooncol.* 2014;120(1):85-93.

Kleihues P, Burger PC, Scheithauer BW. The new WHO classification of brain tumours. *Brain Pathol* 1993;3(3):255Y268.

Kleihues P, Louis DN, Scheithauer BW, et al. The WHO classification of tumors of the nervous system. *J Neuropathol Exp Neurol.* 2002;61(3):215-229.

Knoll M, Lodish HF, Sun L. Long non-coding RNAs as regulators of the endocrine system. *Nat Rev Endocrinol.* 2015;11:151–60.

Koekkoek JA, Kerkhof M, Dirven L, Heimans JJ, Reijneveld JC, Taphoorn MJ. Seizure outcome after radiotherapy and chemotherapy in low-grade glioma patients: a systematic review. *Neuro Oncol.* 2015;17(7):924-934.

Koekkoek JA, Kerkhof M, Dirven L, Heimans JJ, Reijneveld JC, Taphoorn MJ. Seizure outcome after radiotherapy and chemotherapy in low-grade glioma patients: a systematic review. *Neuro Oncol.* 2015;17(7):924-934.

Krasnov GS, Kudryavtseva AV, Snezhkina AV, et al. Pan-Cancer Analysis of TCGA Data Revealed Promising Reference Genes for qPCR Normalization. *Front Genet.* 2019;10:97.

Kryshtafovych A, Monastyrskyy B, Fidelis K, Moulton J, Schwede T, Tramontano A. Evaluation of the template-based modeling in CASP12. *Proteins.* 2018;86 Suppl 1(Suppl 1):321-334.

Kumar P, Mahalingam K. In silico approach to identify non-synonymous SNPs with highest predicted deleterious effect on protein function in human obesity related gene, neuronal growth regulator 1 (NEGR1). *3 Biotech.* 2018;8:466.

Kurian AW. BRCA1 and BRCA2 mutations across race and ethnicity: distribution and clinical implications. *Curr Opin Obstet Gynecol.* 2010;22(1):72-78.

Laack NN, Brown PD, Ivnik RJ, et al. Cognitive function after radiotherapy for supratentorial low-grade glioma: a North Central Cancer Treatment Group prospective study. *Int J Radiat Oncol Biol Phys.* 2005;63(4):1175-1183.

Lai A, Kharbanda S, Pope WB, et al. Evidence for sequenced molecular evolution of IDH1 mutant glioblastoma from a distinct cell of origin. *J Clin Oncol.* 2011;29(34):4482-4490.

Laug D, Glasgow M, Deneen B. A glial blueprint for gliomagenesis. *Nat Rev Neurosci.* 2018;19:393–403.

Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499(7457):214-218.

Le Van T, van Leeuwen M, Carolina Fierro A, et al. Simultaneous discovery of cancer subtypes and subtype features by molecular data integration. *Bioinformatics.* 2016;32(17):i445-i454.

Leeper HE, Caron AA, Decker PA, Jenkins RB, Lachance DH, Giannini C. IDH mutation, 1p19q codeletion and ATRX loss in WHO grade II gliomas. *Oncotarget.* 2015;6(30):30295-30305.

Lei L, Xiaohui Liu, Xiaoye Ma, et al. Identification of key candidate genes and pathways in glioblastoma by integrated bioinformatical analysis. *Exp Ther Med.* 2019;18:3439–3449.

Li C, Miao R, Zhang J, et al. Long non-coding RNA KCNQ1OT1 mediates the growth of hepatocellular carcinoma by functioning as a competing endogenous RNA of miR- 504. *Int J Oncol.* 2018;52:1603–12.

Li J, Ching T, Huang S, Garmire LX. Using epigenomics data to predict gene expression in lung cancer. *BMC Bioinformatics.* 2015;16 Suppl 5(Suppl 5):S10.

Li J, Tao X, Shen J, et al. The molecular landscape of histone lysine methyltransferases and demethylases in non-small cell lung cancer. *Int J Med Sci.* 2019;16:922-930.

Li J, Zhu Y, Wang H, et al. Targeting Long Noncoding RNA in Glioma: A Pathway Perspective. *Mol Ther Nucleic Acids.* 2018;13:431-441.

Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 2014;42(Database issue):D92-D97.

Li LJ, Huang Q, Pan HF, Ye DQ. Circular RNAs and systemic lupus erythematosus. *Exp Cell Res.* 2016;346:248–54.

Li M, Sun Q, Wang X. Transcriptional landscape of human cancers. *Oncotarget.* 2017;8(21):34534-34551.

Li R, Qu H, Wang S, et al. GDCRNATools: an R/Bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC. *Bioinformatics.* 2018;34:2515- 2517.

Li W, Wang R, Yan Z, Bai L, Sun Z. High accordance in prognosis prediction of colorectal cancer across independent datasets by multi-gene module expression profiles. *PLoS One.* 2012;7(3):e33653.

Li X, We J, Yixiong L, et al. Primary Astrocytic Tumours and Paired Recurrences have Similar Biological Features in IDH1, TP53 and TERTp Mutation and MGMT, ATRX Loss. *Sci Rep.* 2017;7:13038.

Li X, Zhang Y. Bioinformatics data distribution and integration via Web Services and XML. *Genomics Proteomics Bioinformatics.* 2003;1(4):299-303.

Li Y, Sahni N, Yi S. Comparative analysis of protein interactome networks prioritizes candidate genes with cancer signatures. *Oncotarget.* 2016;7(48):78841-78849.

Li Z, Xu C, Ding B, et al. Long non-coding RNA MALAT1 promotes proliferation and suppresses apoptosis of glioma cells through derepressing Rap1B by sponging miR-101. *J Neurooncol.* 2017;134:19-28.

Liu L, Li Y, Tollefsbol TO. Gene–environment interactions and epigenetic basis of human diseases. *Curr Issues Mol Biol.* 2008;10:25–36

Liu Y, Devescovi V, Chen S, Nardini C. Multilevel omic data integration in cancer cell lines: advanced annotation and emergent properties. *BMC Syst Biol.* 2013;7:14.

Louis DN, Ohgaki H, Wiestler OD, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* 2007;114(2):97-109.

Louis DN, Perry A, Burger P, et al. International Society Of Neuropathology--Haarlem consensus guidelines for nervous system tumor classification and grading. *Brain Pathol.* 2014;24(5):429-435.

Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* 2016;131(6):803-820.

Love-Koh J, Peel A, Rejon-Parrilla JC, et al. The Future of Precision Medicine: Potential Impacts for Health Technology Assessment. *Pharmacoeconomics.* 2018;36(12):1439-1451.

Luca FRG and Ian JM. Regulation of microRNA function in animals . *Nat Rev Mol Cell Biol.* 2019;20: 21–37.

Lumley TSR. Package survival. 2019.

Luo J. Teaching the ABCs of bioinformatics: a brief introduction to the Applied Bioinformatics Course. *Brief Bioinform.* 2014;15(6):1004-1013.

Maag JLV. gganatogram: An R package for modular visualisation of anatograms and tissues based on ggplot2. *F1000Res.* 2018;7:1576.

Mahmoudi E, Cairns MJ. MiR-137: an important player in neural development and neoplastic transformation. *Mol Psychiatry.* 2017;22:44-55.

Maier M. Personalized medicine-a tradition in general practice!. *Eur J Gen Pract.* 2019;25(2):63-64.

Malta TM, de Souza CF, Sabedot TS, et al. Glioma CpG island methylator phenotype (G-CIMP): biological and clinical implications. *Neuro Oncol.* 2018;20(5):608-620.

Manzoni C, Kia DA, Vandrovцова J, et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Brief Bioinform.* 2018;19(2):286-302.

Martinet LE, Ahmed OJ, Lepage KQ, Cash SS, Kramer MA. Slow Spatial Recruitment of Neocortex during Secondarily Generalized Seizures and Its Relation to Surgical Outcome. *J Neurosci.* 2015;35(25):9477-9490.

Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018;28(11):1747-1756.

Mayers JR, Wu C, Clish CB, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat Med.* 2014;20(10):1193-1198.

Mazor T, Pankov A, Johnson BE, et al. DNA Methylation and Somatic Mutations Converge on the Cell Cycle and Define Similar Evolutionary Histories in Brain Tumors. *Cancer Cell.* 2015;28(3):307-317.

McQuilton P, St Pierre SE, Thurmond J; FlyBase Consortium. FlyBase 101--the basics of navigating FlyBase. *Nucleic Acids Res.* 2012;40(Database issue):D706-D714.

Mewes HW, Ruepp A, Theis F, et al. MIPS: curated databases and comprehensive secondary data resources in 2010. *Nucleic Acids Res.* 2011;39(Database issue):D220-D224.

Mo Q, Wang S, Seshan VE, et al. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc Natl Acad Sci U S A.* 2013;110(11):4245-4250.

Modrek AS, Golub D, Khan T, et al. Low-Grade Astrocytoma Mutations in IDH1, P53, and ATRX Cooperate to Block Differentiation of Human Neural Stem Cells via Repression of SOX2. *Cell Rep.* 2017;21:1267-1280.

Mohammed A, Biegert G, Adamec J, Helikar T. Identification of potential tissue-specific cancer biomarkers and development of cancer versus normal genomic classifiers. *Oncotarget*. 2017;8(49):85692-85715.

Moran I, Akerman I, van de Bunt M, Xie R, Benazra M, Nammo T, et al. Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab*. 2012;16:435–48.

Multhaup ML, Seldin MM, Jaffe AE, Lei X, Kirchner H, Mondal P, et al. Mouse–human experimental epigenetic analysis unmasks dietary targets and genetic liability for diabetic phenotypes. *Cell Metab*. 2015;21:138–49.

Munoz JL, Walker ND, Mareedu S, et al. Cycling Quiescence in Temozolomide Resistant Glioblastoma Cells Is Partly Explained by microRNA-93 and -193-Mediated Decrease of Cyclin D. *Front Pharmacol*. 2019;10:134.

N. M. Luscombe, D. Greenbaum, M. Gerstein . What is Bioinformatics? A Proposed Definition and Overview of the Field . 2001. *Methods of Information in Medicine* 40(4):346-58.

Nahed BV, Redjal N, Brat DJ, et al. Management of patients with recurrence of diffuse low grade glioma: A systematic review and evidence-based clinical practice guideline. *J Neurooncol*. 2015;125(3):609-630.

Narayanan M, Vetta A, Schadt EE, Zhu J. Simultaneous clustering of multiple gene expression and physical interaction datasets. *PLoS Comput Biol*. 2010;6(4):e1000742.

NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2015;43(Database issue):D6-D17.

Nestler EJ. Transgenerational Epigenetic Contributions to Stress Responses: Fact or Fiction?. *PLoS Biol*. 2016;14(3):e1002426.

Neureiter D, Jäger T, Ocker M, Kiesslich T. Epigenetics and pancreatic cancer: pathophysiology and novel treatment aspects. *World J Gastroenterol*. 2014;20(24):7830-7848.

Nonoguchi N, Ohta T, Oh JE, Kim YH, Kleihues P, Ohgaki H. TERT promoter mutations in primary and secondary glioblastomas. *Acta Neuropathol.* 2013;126(6):931-937.

Noushmehr H, Weisenberger DJ, Diefes K, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell.* 2010;17(5):510-522.

Ohba S, Kuwahara K, Yamada S, Abe M, Hirose Y. Correlation between IDH, ATRX, and TERT promoter mutations in glioma. *Brain Tumor Pathol.* 2020;37(2):33-40.

Otasek D, Morris JH, Bouças J, et al. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol.* 2019;20:185.

Pace A, Vidiri A, Galiè E, et al. Temozolomide chemotherapy for progressive low-grade glioma: clinical benefits and radiological response. *Ann Oncol.* 2003;14(12):1722-1726.

Pal A, Gonzalez-Malerva L, Eaton S, et al. Abstract P4-05-07: Functional genomics of TP53 mutations and its impact in breast cancer progression. *Cancer Research.* 2015;75.

Papageorgiou I, Court MH. Identification and validation of the microRNA response elements in the 3'-untranslated region of the UDP glucuronosyltransferase (UGT) 2B7 and 2B15 genes by a functional genomics approach. *Biochem Pharmacol.* 2017;146:199-213.

Park JW, Turcan Ş. Epigenetic Reprogramming for Targeting IDH-Mutant Malignant Gliomas. *Cancers (Basel).* 2019;11(10):1616.

Pascual-Leone A, Amedi A, Fregni F, Merabet LB. The plastic human brain cortex. *Annu Rev Neurosci.* 2005;28:377-401.

Paul Y, Mondal B, Patil V, Somasundaram K. DNA methylation signatures for 2016 WHO classification subtypes of diffuse gliomas. *Clin Epigenetics.* 2017;9:32.

Peng Y and Croce C. The role of MicroRNAs in human cancer. *Sig Transduct Target Ther.* 2016;1:15004

Pereira GRC, Da Silva ANR, Do Nascimento SS, et al. In silico analysis and molecular dynamics simulation of human superoxide dismutase 3 (SOD3) genetic variants. *J Cell Biochem.* 2019;120:3583–3598.

Pereira GRC, Da Silva ANR, Do Nascimento SS, et al. In silico analysis and molecular dynamics simulation of human superoxide dismutase 3 (SOD3) genetic variants. *J Cell Biochem.* 2019;120:3583-3598.

Pickersgill M, Niewöhner J, Müller R, Martin P, Cunningham-Burley S. Mapping the new molecular landscape: social dimensions of epigenetics. *New Genet Soc.* 2013;32(4):429-447.

Pietrak B, Zhao H, Qi H, et al. A tale of two subunits: how the neomorphic R132H IDH1 mutation enhances production of  $\alpha$ HG. *Biochemistry.* 2011;50(21):4804-4812.

Pinu FR, Beale DJ, Paten AM, et al. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites.* 2019;9(4):76.

Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science.* 2016;352(6290):aad9780.

Pornputtpong N, Wanichthanarak K, Nilsson A, Nookaew I, Nielsen J. A dedicated database system for handling multi-level data in systems biology. *Source Code Biol Med.* 2014;9:17.

Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol.* 2011;9(1):48-57.

Prensner JR, Chinnaiyan AM. Metabolism unhinged: IDH mutations in cancer. *Nat Med.* 2011;17(3):291-293.

Quinn JA, Reardon DA, Friedman AH, et al. Phase II trial of temozolomide in patients with progressive low-grade glioma. *J Clin Oncol.* 2003;21(4):646-651.

Ragel BT, Ryken TC, Kalkanis SN, Ziu M, Cahill D, Olson JJ. The role of biopsy in the management of patients with presumed diffuse low grade glioma: A systematic review and evidence-based clinical practice guideline. *J Neurooncol.* 2015;125(3):481-501.

Raghuraman S, Donkin I, Versteyhe S, Barres R, Simar D. The emerging role of epigenetics in inflammation and immunometabolism. *Trends Endocrinol Metab.* 2016;27:782–95.

Raj U, Aier I, Semwal R, Varadwaj PK. Identification of novel dysregulated key genes in Breast cancer through high throughput ChIP-Seq data analysis. *Sci Rep.* 2017;7(1):3229.

Rapport DJ and Maffi L. A call for integrative thinking. *Science* 2013;339:6123-1032.

Reuss DE, Kratz A, Sahm F, et al. Adult IDH wild type astrocytomas biologically and clinically resolve into other tumor entities. *Acta Neuropathol.* 2015;130(3):407-417.

Rhee SY, Beavis W, Berardini TZ, et al. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* 2003;31(1):224-228.

Rice M, Gladstone W, Weir M. Relational databases: a transparent framework for encouraging biology students to think informatically. *Cell Biol Educ.* 2004;3(4):241-252.

Rigsby RE, Parker AB. Using the PyMOL application to reinforce visual understanding of protein structure. *Biochem Mol Biol Educ.* 2016;44:433–437.

Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:47.

Robert F, Pelletier J. Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. *Front Genet.* 2018;9:507. Published 2018 Oct 30.

Roberta M, Eleonora B, Enrica F, et al. Targeting miR-155-5p and miR-221-3p by peptide nucleic acids induces caspase-3 activation and apoptosis in temozolomide-resistant T98G glioma cells. *Int J Oncol.* 2019;55:59–68.

Roden DM, Tyndale RF. Genomic medicine, precision medicine, personalized medicine: what's in a name?. *Clin Pharmacol Ther.* 2013;94(2):169-172.

Rolland T, Taşan M, Charlotteaux B, et al. A proteome-scale map of the human interactome network. *Cell.* 2014;159(5):1212-1226.

Romanoski CE, Glass CK, Stunnenberg HG, Wilson L, Almouzni G. Epigenomics: Roadmap for regulation. *Nature.* 2015;518(7539):314-316.

Romeo SG, Conti A, Polito F, et al. miRNA regulation of Sirtuin-1 expression in human astrocytoma. *Oncol Lett.* 2016;12:2992-2998.

Rueda E, Sierra M, Infante J, et al. Aspectos controvertidos en el manejo de gliomas hemisféricos OMS grado II: revisión de la bibliografía reciente [Controversial aspects in WHO grade II gliomas management: review of recent literature]. *Rev Neurol.* 2011;53(12):747-757.

Ryken TC, Parney I, Buatti J, Kalkanis SN, Olson JJ. The role of radiotherapy in the management of patients with diffuse low grade glioma: A systematic review and evidence-based clinical practice guideline. *J Neurooncol.* 2015;125(3):551-583.

Sana J, Busek P, Fadrus P, et al. Identification of microRNAs differentially expressed in glioblastoma stem-like cells and their association with patient survival. *Sci Rep.* 2018;8:2836.

Sanai N, Polley MY, Berger MS. Insular glioma resection: assessment of patient morbidity, survival, and tumor progression. *J Neurosurg.* 2010;112(1):1-9.

Sanchez Calle A, Kawamura Y, Yamamoto Y, et al. Emerging roles of long non-coding RNA in cancer. *Cancer Sci.* 2018;109:2093-2100.

Sankar S, Saravanan N, Rajendiran P, et al. Identification of B- and T-cell epitopes on HtrA protein of *Orientia tsutsugamushi*. *J Cell Biochem.* 2019;120:5869–5879.

Sariyar M, Schluender I, Smee C, Suhr S. Sharing and Reuse of Sensitive Data and Samples: Supporting Researchers in Identifying Ethical and Legal Requirements. *Biopreserv Biobank.* 2015;13(4):263-270.

Sarma PP, Dutta D, Mirza Z, et al. Point mutations in the DNA binding domain of p53 contribute to glioma progression and poor prognosis. *Mol Biol.* 2017;51:293–299.

Schmitz SU, Grote P, Herrmann BG. Mechanisms of long noncoding RNA function in development and disease. *Cell Mol Life Sci.* 2016;73:2491–509.

Schwartzentruber J, Korshunov A, Liu XY, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature.* 2012;482(7384):226-231.

Scourzic L, Mouly E, Bernard OA. TET proteins and the control of cytosine demethylation in cancer. *Genome Med.* 2015;7(1):9.

Searls DB. The roots of bioinformatics. *PLoS Comput Biol.* 2010;6(6):e1000809. Published 2010 Jun 24. doi:10.1371/journal.pcbi.1000809.

Selleck MJ, Senthil M, Wall NR. Making Meaningful Clinical Use of Biomarkers. *Biomark Insights.* 2017;12:1177271917715236.

Serre D, Lee BH, Ting AH. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res.* 2010;38(2):391-399.

Sethi S, Ali S, Philip PA, Sarkar FH. Clinical advances in molecular biomarkers for cancer diagnosis and therapy. *Int J Mol Sci.* 2013;14(7):14771-14784.

Shen R, Mo Q, Schultz N, et al. Integrative subtype discovery in glioblastoma using iCluster. *PLoS One.* 2012;7(4):e35236.

Shi X, Shen S, Liu J, Huang J, Zhou Y, Ma S. Similarity of markers identified from cancer gene expression studies: observations from GEO. *Brief Bioinform.* 2014;15(5):671-684.

Shi X, Yi H, Ma S. Measures for the degree of overlap of gene signatures and applications to TCGA. *Brief Bioinform.* 2015;16(5):735-744.

Shih HA, Sherman JC, Nachtigall LB, et al. Proton therapy for low-grade gliomas: Results from a prospective trial. *Cancer.* 2015;121(10):1712-1719.

Shinkai Y, Tachibana M. H3K9 methyltransferase G9a and the related molecule GLP. *Genes Dev.* 2011;25(8):781-788.

Shobha Sharma, Stacy Ciufu, Elena Starchenko, Dakshesh Darji, Larry Chlumsky, Ilene Karsch-Mizrachi, Conrad L Schoch, The NCBI BioCollections Database, Database, Volume 2018, 2018, bay006.

Silva FC, Valentin MD, Ferreira Fde O, Carraro DM, Rossi BM. Mismatch repair genes in Lynch syndrome: a review. *Sao Paulo Med J.* 2009;127(1):46-51.

Smith JS, Chang EF, Lamborn KR, et al. Role of extent of resection in the long-term outcome of low-grade hemispheric gliomas. *J Clin Oncol.* 2008;26(8):1338-1345.

SnapGene software (from GSL Biotech; available at [snapgene.com](http://snapgene.com)).

Song Q, Hawkins GA, Wudel L, et al. Dissecting intratumoral myeloid cell plasticity by single cell RNA-seq. *Cancer Med.* 2019;8(6):3072-3085.

Speicher NK, Pfeifer N. Integrating different data types by regularized unsupervised multiple kernel learning with application to cancer subtype discovery. *Bioinformatics.* 2015;31(12):i268-i275.

Stege EM, Kros JM, de Bruin HG, et al. Successful treatment of low-grade oligodendroglial tumors with a chemotherapy regimen of procarbazine, lomustine, and vincristine. *Cancer.* 2005;103(4):802-809.

Stoesser G, Baker W, van den Broek A, et al. The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.* 2002;30(1):21-26.

Sturm D, Witt H, Hovestadt V, et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell.* 2012;22(4):425-437.

Su J, Long W, Ma Q, et al. Identification of a Tumor Microenvironment-Related Eight-Gene Signature for Predicting Prognosis in Lower-Grade Gliomas. *Front Genet.* 2019;10:1143.

Su Y, Xiong J, Bing Z, et al. Identification of novel human glioblastoma-specific transcripts by serial analysis of gene expression data mining. *Cancer Biomark*. 2013;13(5):367-375.

Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform Biol Insights*. 2020;14:1177932219899051.

Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer*. 2017;17(11):692-704.

Sunil L, Prasanna V. In silico designing of therapeutic protein enriched with branched-chain amino acids for the dietary treatment of chronic liver disease. *J Mol Graph Model*. 2017;76:192-204.

Suvà ML, Tirosh I. Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging Challenges. *Mol Cell*. 2019;75(1):7-12.

Swiatly A, Horala A, Matysiak J, Hajduk J, Nowak-Markwitz E, Kokot ZJ. Understanding Ovarian Cancer: iTRAQ-Based Proteomics for Biomarker Discovery. *Int J Mol Sci*. 2018;19(8):2240.

Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics*. 2013;29(18):2238-2244.

Taudt A, Colome-Tatche M, Johannes F. Genetic sources of population epigenomic variation. *Nat Rev Genet*. 2016;17:319–32.

Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012;337(6090):64-69. doi:10.1126/science.1219240

Teplyuk NM, Uhlmann EJ, Gabriely G, et al. Therapeutic potential of targeting microRNA-10b in established intracranial glioblastoma: first steps toward the clinic. *EMBO Mol Med*. 2016;8(3):268-287.

The Cancer Genome Atlas Research Network. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med*. 2015;372:2481–2498.

Thirumal Kumar D, Jerushah Emerald L, George Priya Doss C, et al. Computational approach to unravel the impact of missense mutations of proteins (D2HGDH and IDH2) causing D-2-hydroxyglutaric aciduria 2. *Metabolic Brain Disease*. 2018;33:1699-1710.

Thirumal Kumar D, Sneha P, Uppin J, et al. Investigating the Influence of Hotspot Mutations in Protein–Protein Interaction of IDH1 Homodimer Protein: A Computational Approach. *Advances in Protein Chemistry and Structural Biology*. *Adv Protein Chem Struct Biol*. 2018;111:243-261.

Todd W Harris, Valerio Arnaboldi, Scott Cain, et al. WormBase: a modern Model Organism Information Resource, *Nucleic Acids Research*, Volume 48, Issue D1, 08 January 2020, Pages D762–D767.

Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015;19(1A):A68-A77.

Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511-515.

Turcan S, Rohle D, Goenka A, et al. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*. 2012;483(7390):479-483.

Underwood CJ, Choi K, Lambing C, et al. Epigenetic activation of meiotic recombination near *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. *Genome Res*. 2018;28(4):519-531.

UniProt Consortium. The universal protein resource (UniProt). *Nucleic Acids Res.* 2008;36(Database issue):D190-D195.

Venneti S, Felicella MM, Coyne T, et al. Histone 3 lysine 9 trimethylation is differentially associated with isocitrate dehydrogenase mutations in oligodendrogliomas and high-grade astrocytomas. *J Neuropathol Exp Neurol.* 2013;72(4):298-306.

Verhaak RG, Hoadley KA, Purdom E, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell.* 2010;17(1):98-110.

Vinhas R, Lourenço A, Santos S, et al. A novel BCR-ABL1 mutation in a patient with Philadelphia chromosome-positive B-cell acute lymphoblastic leukemia. *Onco Targets Ther.* 2018;11:8589-8598.

Vo TV, Das J, Meyer MJ, et al. A Proteome-wide Fission Yeast Interactome Reveals Network Evolution Principles from Yeasts to Human. *Cell.* 2016;164(1-2):310-323.

Wan J, Huang M, Zhao H, et al. A novel tetranucleotide repeat polymorphism within KCNQ1OT1 confers risk for hepatocellular carcinoma. *DNA Cell Biol.* 2013;32:628- 34.

Wang B, Li D, Yao Y, et al. The crucial role of DNA-dependent protein kinase and myelin transcription factor 1-like protein in the miR-141 tumor suppressor network. *Cell Cycle.* 2019;18(21):2876-2892.

Wang H, Diaz AK, Shaw TI, et al. Deep multiomics profiling of brain tumors identifies signaling networks downstream of cancer driver genes. *Nat Commun.* 2019;10:3718.

Wang S, Jin F, Fan W, et al. Gene expression meta-analysis in diffuse low-grade glioma and the corresponding histological subtypes. *Sci Rep.* 2017;7:1–13.

Wang W, Baladandayuthapani V, Morris JS, Broom BM, Manyam G, Do KA. iBAG: integrative Bayesian analysis of high-dimensional multiplatform genomics data. *Bioinformatics.* 2013;29(2):149-159.

Wang Y, Virtanen J, Xue Z, et al. I-TASSER-MR: automated molecular replacement for distant-homology proteins using iterative fragment assembly and progressive sequence truncation. *Nucleic Acids Res.* 2017;45:429-434.

Wang Z, Jensen MA, Zenklusen JC. A Practical Guide to The Cancer Genome Atlas (TCGA). *Methods Mol Biol.* 2016;1418:111-141.

Warde-Farley D, Sylva L, Ovi C, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010;38:W214–W220.

Washburn MP, Koller A, Oshiro G, et al. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A.* 2003;100(6):3107-3112.

Watanabe T, Nobusawa S, Kleihues P, Ohgaki H. IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas. *Am J Pathol.* 2009;174(4):1149-1153.  
Weischenfeldt J, Simon R, Feuerbach L, et al. Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell.* 2013;23(2):159-170.

Wen PY, Huse JT. 2016 World Health Organization Classification of Central Nervous System Tumors. *Continuum (Minneap Minn).* 2017;23:1531-1547.

Wen PY, Huse JT. 2016 World Health Organization Classification of Central Nervous System Tumors. *Continuum (Minneap Minn).* 2017;23(6, Neuro-oncology):1531-1547.

Wen Y, Wei Y, Zhang S, et al. Cell subpopulation deconvolution reveals breast cancer heterogeneity based on DNA methylation signature. *Brief Bioinform.* 2017;18(3):426-440.

Wenzhi Mao, Cihan Kaya, Anindita Dutta, Amnon Horovitz, Ivet Bahar, Comparative study of the effectiveness and limitations of current methods for detecting sequence coevolution, *Bioinformatics*, Volume 31, Issue 12, 15 June 2015, Pages 1929–1937.

Werner HM, Mills GB, Ram PT. Cancer Systems Biology: a peek into the future of patient care?. *Nat Rev Clin Oncol.* 2014;11(3):167-176.

Wessels PH, Weber WE, Raven G, Ramaekers FC, Hopman AH, Twijnstra A. Supratentorial grade II astrocytoma: biological features and clinical course. *Lancet Neurol*. 2003;2(7):395-403.

Wong KK, Rostomily R, Wong STC. Prognostic Gene Discovery in Glioblastoma Patients using Deep Learning. *Cancers (Basel)*. 2019;11(1):53.

Wong ST, Zhang XQ, Zhuang JT, et al. MicroRNA-21 inhibition enhances in vitro chemosensitivity of temozolomide-resistant glioblastoma cells. *Anticancer Res*. 2012;32:2835-41.

Wu Q, Xu L, Wang C, et al. MicroRNA-124-3p represses cell growth and cell motility by targeting EphA2 in glioma. *Biochem Biophys Res Commun*. 2018;503:2436-2442.

Wu X, Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet*. 2017;18(9):517-534.

Xu B, Liu Y, Lin C, Dong J, Liu X, He Z. Reconstruction of the Protein-Protein Interaction Network for Protein Complexes Identification by Walking on the Protein Pair Fingerprints Similarity Network. *Front Genet*. 2018;9:272.

Xu L, Chen L, Li F, et al. Over-expression of the non-coding RNA HOTTIP inhibits glioma cell growth by BRE. *J Exp Clin Cancer Res*. 2016;35:162.

Xu W, Yang H, Liu Y, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of  $\alpha$ -ketoglutarate-dependent dioxygenases. *Cancer Cell*. 2011;19(1):17-30.

Yang J, Yan R, Roy A, et al. The I-TASSER suite: Protein structure and function prediction. *Nat Methods*. 2014;12:7–8.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. 2015;12(1):7-8.

Yang J, Zhang W, Baoji H, et al. Template-based protein structure prediction in CASP11 and retrospect of I-TASSER in the last decade. *Proteins*. 2016;84:233–246.

Yang J, Zhang Y. Protein Structure and Function Prediction Using I-TASSER. *Curr Protoc Bioinformatics*. 2015;52:5.8.1-5.8.15.

Yang Y, Mao Q, Wang X, et al. An analysis of 170 glioma patients and systematic review to investigate the association between IDH-1 mutations and preoperative glioma-related epilepsy. *J Clin Neurosci*. 2016;31:56-62.

Yanovich G, Agmon H, Harel M, Sonnenblick A, Peretz T, Geiger T. Clinical Proteomics of Breast Cancer Reveals a Novel Layer of Breast Cancer Classification. *Cancer Res*. 2018;78(20):6001-6010.

Yao B, Christian KM, He C, Jin P, Ming GL, Song H. Epigenetic mechanisms in neurogenesis. *Nat Rev Neurosci*. 2016;17:537–49.

Yao Y, Ma J, Xue Y, et al. Knockdown of long non-coding RNA XIST exerts tumor-suppressive functions in human glioblastoma stem cells by up-regulating miR- 152. *Cancer Lett*. 2015;359:75-86.

Yu G, Wang LG, Han Y, et al. clusterProfiler : an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284-7.

Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31(14):2382-2383.

Zeng T, Wang DC, Wang X, Xu F, Chen L. Prediction of dynamical drug sensitivity and resistance by module network rewiring-analysis based on transcriptional profiling. *Drug Resist Updat*. 2014;17(3):64-76.

Zhang JX, Han L, Bao ZS, et al. HOTAIR, a cell cycle-associated long noncoding RNA and a strong predictor of survival, is preferentially expressed in classical and mesenchymal glioma. *Neuro Oncol*. 2013;15:1595-603.

Zhang K, Sun X, Zhou X, et al. Long non-coding RNA HOTAIR promotes glioblastoma cell cycle progression in an EZH2 dependent manner. *Oncotarget*. 2015;6:537-46.

Zhang S, Liu CC, Li W, Shen H, Laird PW, Zhou XJ. Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Res*. 2012;40(19):9379-9391.

Zhang W, Yang J, Baoji H, et al. Integration of QUARK and I-TASSER for Ab Initio Protein Structure Prediction in CASP11. *Proteins*. 2016;84:76–86.

Zhang Y, He C, Qiu L, et al. Serum Unsaturated Free Fatty Acids: A Potential Biomarker Panel for Early-Stage Detection of Colorectal Cancer. *J Cancer*. 2016;7(4):477-483.

Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008;9:1–8.

Zhang YW, Wang Z, Xie W, et al. Acetylation Enhances TET2 Function in Protecting against Abnormal DNA Methylation during Oxidative Stress. *Mol Cell*. 2017;65(2):323-335.

Zhang Z, Song X, Tian H, et al. MicroRNA-137 inhibits growth of glioblastoma through EGFR suppression. *Am J Transl Res*. 2017;9(3):1492-1499.

Zhang Z, Weaver DL, Olsen D, et al. Long non-coding RNA chromogenic in situ hybridisation signal pattern correlation with breast tumour pathology. *J Clin Pathol*. 2016;69:76-81.

Zhang ZJ, Sun AY. *Nan Fang Yi Ke Da Xue Xue Bao*. 2009;29(10):2021-2026.

Zhao ML, Wang W, Nie H, et al. In silico structure prediction and inhibition mechanism studies of AtHDA14 as revealed by homology modeling, docking, molecular dynamics simulation. *Comput Biol Chem*. 2018;75:120–130.

Zhao X, Wang P, Liu J, et al. Gas5 Exerts Tumor-suppressive Functions in Human Glioma Cells by Targeting miR-222. *Mol Ther*. 2015;23:1899-911.

Zhao Z, Zhang K, Wang Z, et al. A comprehensive review of available omics data resources and molecular profiling for precision glioma studies. *Biomed Rep*. 2019;10:3- 9.

Zhong ME, Chen Y, Zhang G, et al. LncRNA H19 regulates PI3K–Akt signal pathway by functioning as a ceRNA and predicts poor prognosis in colorectal cancer: integrative analysis of dysregulated ncRNA-associated ceRNA network. *Cancer Cell Int.* 2019;19:148.

Zhou RS, Zhang EX, Sun QF, et al. Integrated analysis of lncRNA-miRNA-mRNA ceRNA network in squamous cell carcinoma of tongue. *BMC Cancer.* 2019;19:779.

Zhou RS, Zhang EX, Sun QF, et al. Integrated analysis of lncRNA-miRNA-mRNA ceRNA network in squamous cell carcinoma of tongue. *BMC Cancer.* 2019;19:779.

Zhu Q, Sun Y, Zhou Q, He Q, Qian H. Identification of key genes and pathways by bioinformatics analysis with TCGA RNA sequencing data in hepatocellular carcinoma. *Mol Clin Oncol.* 2018;9(6):597-606.

Zou D, Ma L, Yu J, Zhang Z. Biological databases for human research. *Genomics Proteomics Bioinformatics.* 2015;13(1):55-63.

Zulch, K. J. *Histological Typing of Tumours of the Central Nervous System.* Geneva, Switzerland: World Health Organization. *International Histological Classification of Tumours,* No. 21.1979.

## Résumé

L'utilisation des approches omiques a révolutionné la biologie et elle a conduit à l'émergence d'approches systémiques y compris la génomique, la transcriptomique et l'épigénétique. Dans le premier chapitre, nous avons ciblé des mutations impliquées dans les gliomes, suivi d'une analyse bio-informatique pour étudier l'impact de ces mutations sur la structure et la fonction des 3 protéines IDH1, IDH2 et p53. Nous avons obtenu une mutation qui a un effet sur le site catalytique de la protéine IDH1 comme R132H. D'autres mutations de la p53 ont été identifiées comme K305N, qui est une mutation pathogène ; R175H, qui est une mutation bénigne ; et R158G, qui perturbe la conformation structurelle de la protéine suppresseur de tumeur. Dans le 2<sup>ème</sup> chapitre, on a identifié des biomarqueurs spécifiques pour les gliomes en analysant des données génomiques et épigénomiques. La prévalence des mutations IDH dans le gliome de bas grade nous a guidé à cibler leurs conséquences épigénomiques en aval, telles que G-CIMP. On a trouvé que les astrocytomes sont caractérisés par une hyperméthylation au niveau du gène MYT1L, et les oligodendrogliomes au niveau du gène PLCG1. Dans le 3<sup>ème</sup> chapitre, on a identifié les gènes et les miRNAs différenciellement exprimés au niveau des gliomes. Ces biomarqueurs sont sélectionnés sur la base d'une combinaison de seuil de changement d'expression et de seuil de score, et nous avons proposé un nouveau réseau de miARNs qui combinent à la fois des gènes et des lncARNs spécifique pour les gliomes de bas grade. Dans le chapitre 4 on a démontré que les gliomes présentent également des altérations épigénétiques, parmi eux on trouve la méthylation de l'ADN et leur interaction avec les modifications d'histones. Enfin, on a réussi à caractériser les gliomes de bas grade en analysant les différents données omiques et on a mis en place un workflow bio-informatique précis et efficace pour analyser ces données et avoir une stratégie prometteuse pour mieux comprendre le développement du gliome.

**Mots clés :** Génomique, Transcriptomique, Épigenétique, Bio-informatique, Gliome de bas grade.

## Abstract

The use of omics approaches has revolutionized biology and has led to the emergence of systemic approaches including genomics, transcriptomics and epigenetics. In the first chapter, we targeted mutations involved in gliomas, starting with the sequencing of genomic regions encompassing isocitrate dehydrogenase 1 (IDH1) exon 4 and isocitrate dehydrogenase 2 (IDH2) and the four exons (5-6 and 7-8) of TP53 from 32 samples, followed by a bioinformatic analysis to study the impact of these mutations on the structure and function of the 3 proteins IDH1, IDH2 and p53. We have obtained a mutation which has an effect on the catalytic site of IDH1 protein such as. Other p53 mutations have been identified as K305N, which is a pathogenic mutation; R175H, which is a benign mutation; and R158G, which disrupts the structural conformation of the tumor suppressor protein. In the 2nd chapter, we identified specific biomarkers for gliomas by analyzing genomic and epigenomic data. The prevalence of IDH mutations in low-grade glioma has guided us to target their downstream epigenomic consequences, such as G-CIMP. It has been found that astrocytomas are characterized by hypermethylation at the level of the MYT1L gene, and oligodendrogliomas at the level of the PLCG1 gene. In the 3rd chapter, we identified the genes and miRNAs differentially expressed in gliomas. These biomarkers are selected on the basis of a combination of expression change threshold and score threshold, and we have proposed a new network of miRNAs that combine to both genes and lncRNAs specific for low grade gliomas. In Chapter 4 it was shown that low grade gliomas also exhibit epigenetic alterations, among them we find DNA methylation and their interaction with histone modifications. In the end, we succeeded in characterizing low-grade gliomas by analyzing the different data omics and we set up a precise and efficient bioinformatic workflow to analyze these data and have a promising strategy to better understand the development of the glioma.

**Keywords:** Genomics, Transcriptomics, Epigenetics, Bioinformatics, Low grade glioma.

**Structure de recherche :** Laboratoire de biotechnologie médical