

Popular Democratic Republic of Algeria

Ministry of High Education and Scientific
Research

Abbes Laghrou University- Khenchela-

Natural and life sciences Faculty

Molecular and Cellular Biology Department



N° de série :

MEMOIRE DE FIN D'ETUDES

DE MASTER ACADEMIQUE

Domaine : **Sciences de la nature et de la vie**

Filière : **Sciences Biologiques**

Spécialité : **Génétique**

Présenté par :

RIM Oumaima

BOUZIANE Achouak

Thème

Analyse des signatures génomiques et
transcriptomiques des tumeurs pulmonaires pour des
applications en médecine personnalisée.

Mémoire soutenu publiquement le 18/06/ 2025 Devant le jury composé de :

Dr. MELLAL	H.	Présidente
Dr. HAMADA	Y.	Examineur
Pr. HAMIDECHI	M. A.	Encadrant

Année Universitaire 2024/2025

Remerciements

Au terme de ce travail de recherche, nous tenons à exprimer notre profonde gratitude envers toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce mémoire.

Nos premiers remerciements s'adressent à Allah, le Tout-Puissant, qui nous a accordé la force, la patience et la persévérance nécessaires pour mener à bien cette étude.

Nous tenons à exprimer notre profonde gratitude au Professeur HAMIDECHI, pour sa guidance éclairée et son accompagnement tout au long de ce travail de recherche. Ses conseils avisés, sa disponibilité et son expertise nous ont été d'une aide précieuse. Nous le remercions sincèrement pour sa bienveillance et pour nous avoir guidés avec justesse vers l'aboutissement de ce projet.

Nos remerciements s'étendent également aux doctorants encadrés par Monsieur HAMIDECHI qui ont généreusement partagé leur temps et leurs connaissances. Leur contribution a été inestimable, notamment par la mise à disposition des logiciels nécessaires à cette étude.

Nous adressons également nos sincères remerciements aux membres du jury qui ont accepté d'évaluer ce travail de recherche. Leur expertise et leur temps consacré à l'examen de ce mémoire constituent un honneur et une reconnaissance de nos efforts.

Enfin, que toutes les personnes qui ont contribué, directement ou indirectement, à l'enrichissement de ce travail trouvent ici l'expression de notre gratitude la plus sincère.

DEDICACE

Je tiens à exprimer ma profonde gratitude à Professeur HAMIDECHI , pour sa guidance éclairée et son accompagnement tout au long de ce travail de recherche. Ses conseils m'ont été d'une aide précieuse.

À mon père, Tu as été le pilier silencieux de cette réussite. Tes sacrifices quotidiens, ta confiance indéfectible en mes capacités ont été mes plus grandes forces. Merci pour tes encouragements discrets mais si précieux. Ce mémoire porte en lui ta fierté et ton amour.

À mes chers frères NADJEM EDDINE et ABD ERAFIK, mes compagnons de route et mes soutiens les plus fidèles. Votre présence à mes côtés dans les moments difficiles, vos encouragements et tous les efforts que vous avez déployés pour moi resteront gravés dans mon cœur.

À ma mère bien-aimée, Allah yarhamha qui reste gravée à jamais dans mon cœur. Bien que tu ne sois plus parmi nous, ton amour continue de m'accompagner et de me donner la force d'avancer.

À mon frère Said, parti trop tôt mais jamais oublié. Ton souvenir illumine encore mon cœur et guide mes pas. Je sais que tu aurais été fier de ce parcours. Ce travail porte une partie de l'amour que tu m'as laissée.

À mes tantes et mes cousines Hana, Nahed et Nada, vos encouragements chaleureux et votre fierté palpable à chaque étape franchie ont été des sources inépuisables de motivation. Merci d'avoir cru en mes capacités et d'avoir célébré mes petites victoires comme de grands triomphes.

Chère Achouak, tu as été bien plus qu'une simple binôme. Tu as été ma complice et ma force dans cette aventure. Ensemble, nous avons transformé les défis en victoires partagées. Merci d'avoir fait de ce mémoire notre histoire commune, tissée d'une amitié précieuse.

À Ines, tu as été mon pilier de force dans cette épreuve. Quand mes forces m'abandonnaient et que l'espoir s'amenuisait, tu as été cette voix douce qui me rappelait ma valeur. Merci d'avoir été ma lumière dans l'obscurité et d'avoir transformé mes larmes en sourires par ta tendresse infinie.

À Rayan, tu as été mon phare dans la tempête de ce parcours académique. Ta patience infinie face à mes doutes et ton épaule toujours disponible dans mes moments de faiblesse. Merci d'avoir été cette âme sœur qui comprend mes silences.

À Hazar, merci pour tes encouragements constants et ta bienveillance. Ton soutien, même dans les moments les plus simples, a toujours su raviver ma détermination.

OUMAIMA

DEDICACE

Je dédie ce travail à Monsieur le Professeur HAMIDECHI, pour son encadrement remarquable, sa présence attentive et ses conseils éclairés qui ont grandement contribué à la réalisation de ce mémoire. Merci infiniment.

À mon père , tu es mon héros et la lumière de ma vie. Tu es le meilleur père qu'un enfant puisse avoir. Ton amour inconditionnel et tes sacrifices m'ont façonné chaque jour. Dans mes doutes, tu m'as rassuré , dans mes épreuves, tu m'as porté avec ton cœur immense. Ce travail est le fruit de ton dévouement sans limites. Je te le dédie avec un amour profond qui déborde de mon cœur et une gratitude éternelle. Merci pour tout.

À ma mère , tu es la douceur incarnée et mon premier amour. Tes mains ont séché mes larmes, ton cœur a compris mes silences. Tu m'as entourée de ta bienveillance et inspirée par ta force tranquille. Ton soutien constant et tes conseils précieux ont éclairé mes choix. Par tes prières, je réalise mes rêves. Dans tes yeux, j'ai trouvé ma confiance ; dans tes bras, ma paix. Cette réussite porte ton empreinte tendre. Je te l'offre comme un témoignage de mon amour infini et de ma fierté d'être ta fille.

À mon frère Aissam , merci de m'avoir toujours soutenue. Je suis chanceuse de t'avoir comme frère. Chaque mot de ce travail te remercie pour ton soutien.

À ma chère sœur Ritaj , ma petite sœur et la joie de notre maison, ta présence à mes côtés m'encourage à continuer et à réaliser mes rêves.merci d'être ma force .

À ma nièce Alaa , ma petite princesse , tu es le précieux dépôt que ta maman Ahlam (que Dieu lui fasse miséricorde) m'a laissé, et je te promets de veiller sur toi comme sur le plus précieux des trésors merci d'être dans ma vie .

A ma chère copine Oumaima, ma moitié dans cette belle aventure, Toi qui as marché à mes côtés dans ce long voyage à la recherche du savoir, Chaque étape franchie, chaque défi relevé... nous l'avons fait ensemble, main dans la main. Tu n'as jamais été seulement une camarade d'étude, tu as été un soutien, une force, un sourire dans les moments de doute. Ce mémoire n'est pas juste la fin d'un parcours universitaire C'est le fruit d'une amitié sincère, d'un travail à deux, Et d'une complicité que je chérirai toute ma vie. Comme si la vie avait su exactement qui mettre sur mon chemin au bon moment.

À Ines, Merci d'avoir été ma lumière dans mes moments sombres Merci d'avoir séché mes larmes et partagé mes joies Tu as transformé ma tristesse en sourires Je suis si chanceuse de t'avoir comme amie Ton soutien a été ma plus grande force.

À Hazar, Mon rayon de soleil et ma source d'énergie Toi qui apportes toujours la joie et la bonne humeur Avec ton rire, tu chasses tous mes soucis Tu remplis ma vie de couleurs et d'optimisme Ton énergie positive illumine mes journées.

ACHOUAK

Résumé

Nous avons exploré l'expression de 24 gènes impliqués dans le cancer, l'asthme et la maladie d'Alzheimer chez 1000 patients. Une analyse en composantes principales a révélé une forte corrélation entre les gènes EGFR et MET ($r = 0,899$), suggérant une co-régulation connue dans le cancer du poumon. À l'inverse, GATA3 et GSTP1, deux gènes liés à l'asthme, présentaient une corrélation négative modérée, traduisant possiblement des mécanismes opposés. Le biplot de l'ACP a mis en évidence deux axes majeurs : l'un distinguant des profils immunitaires (GATA3 vs HLA-DRB1), l'autre opposant un profil oncogénique à un profil neuro-inflammatoire. L'analyse différentielle a mis en évidence plusieurs gènes significativement modifiés selon le statut tabagique. Parmi eux, HLA-DRB1 était sur-exprimé, tandis que GATA3 et MAPT étaient sous-exprimés chez les fumeurs. L'enrichissement fonctionnel a révélé une implication significative des gènes dans la voie PI3K, ainsi que dans des fonctions de liaison protéique, de transduction de signal et de liaison ionique, soutenant leur rôle central dans la régulation cellulaire. Ces résultats éclairent les interactions géniques dans des contextes pathologiques complexes et soutiennent le potentiel de certaines cibles moléculaires en médecine personnalisée.

Mots clés : *Maladies génétiques – Expression génique – Analyse différentielle – Médecine personnalisée*

ملخص

قمنا بدراسة التعبير الجيني لـ 24 جيناً مرتبطاً بالسرطان، والربو، ومرض ألزهايمر لدى عينة مكونة من 1000 مريض. وقد كشفت تحليلات المكونات الرئيسية (ACP) عن وجود ارتباط قوي بين الجينين EGFR و MET ($r = 0,899$)، مما يشير إلى تنظيم مشترك معروف في سرطان الرئة. في المقابل، أظهر الجينان GATA3 و GSTP1، المرتبطان بالربو، ارتباطاً سالباً معتدلاً، ما قد يعكس آليات مرضية متعارضة. وقد أبرز تمثيل الـ biplot لتحليل ACP محورين رئيسيين: الأول يميّز بين الأنماط المناعية) مثل GATA3 مقابل- HLA-DRB1، والثاني يفرّق بين نمط ورمي (oncogénique) ونمط التهابي عصبي (neuro-inflammatoire). كما كشفت التحاليل التفاضلية عن تغيرات معنوية في تعبير بعض الجينات حسب حالة التدخين؛ حيث كان الجين HLA-DRB1 مفرط التعبير لدى المدخنين، بينما انخفض تعبير GATA3 و MAPT في نفس الفئة. وأظهرت تحاليل الإثراء الوظيفي انخراط هذه الجينات في مسار PI3K الحيوي، بالإضافة إلى وظائف متعلقة بارتباط البروتينات، ونقل الإشارة، والارتباط بالأيونات، مما يعزّز دورها المحوري في التنظيم الخلوي. تسلّط هذه النتائج الضوء على التفاعلات الجينية في سياقات مرضية معقّدة وتدعم إمكانية استهداف بعض الجزيئات في إطار الطب الشخصي.

الكلمات المفتاحية: الأمراض الوراثية – التعبير الجيني – التحليل التفاضلي – الطب الشخصي

Abstract

We explored the expression of 24 genes involved in cancer, asthma and Alzheimer's disease in 1000 patients. A Principal Component Analysis revealed a strong correlation between EGFR and MET genes ($r = 0.899$), suggesting known co-regulation in lung cancer. Conversely, GATA3 and GSTP1, two asthma-related genes, showed a moderate negative correlation, possibly reflecting opposite mechanisms. The biplot of PCA highlighted two major axes: one distinguishing immune profiles (GATA3 vs HLA-DRB1), the other opposing an oncogenic profile to a neuroinflammatory profile. The differential analysis highlighted several genes significantly modified according to smoking status. Among them, HLA-DRB1 was overexpressed, while GATA3 and MAPT were underexpressed in smokers. Functional enrichment revealed significant involvement of genes in the PI3K pathway, as well as protein binding, signal transduction and ion binding functions, supporting their central role in cell regulation. These results shed light on gene interactions in complex pathological contexts and support the potential of certain molecular targets in personalized medicine.

Keywords: *Genetic diseases – Gene expression – Differential analysis – Personalized medicine*

Table des matières

INTRODUCTION.....	12
PARTIE 1 : Revue bibliographique.....	2
1. LES BASES MOLECULAIRES DU CANCER DU POUMON	2
1.1. Gènes impliqués :	2
1.1.1 Oncogènes	3
1.1.2 Gènes suppresseurs de tumeurs	5
1.1.3 Gènes de réparation de l'ADN :.....	6
1.2. Analyses des variants génétiques	7
1.2.1. Méthodes de détection des variants génétiques	7
1.2.2. Importance clinique de l'analyse des variants génétiques	9
2. ALTERATIONS GENETIQUE ET TRANSCRIPTOMIQUE IMPLIQUÉES	10
2.1. Mutations génétiques associées au cancer du poumon	10
2.1.1 Les altérations du gène EGFR	10
2.1.2 Le gène KRAS.....	10
2.1.3 Le gène ALK	11
2.1.4 Le gène ROS1.....	12
2.1.5 . Le gène BRAF	13
2.1.7 . Le gène MET.....	14
2.1.8 . Le gène P53.....	14
2.1.9 . Le gène RB1	15
2.1.10 .Le gène PTEN	15
2.1.11 Le gène LKB1	16
2.1.12 Les gènes BRCA1/2	16
2.1.13 Le gène ERCC1.....	16
3. APPROCHES BIOINFORMATIQUES ET RESSOURCES POUR L'ÉTUDE DES ALTÉRATIONS MOLÉCULAIRES.....	18
3.1. Ressources en ligne et bases de données génomiques :	18

3.2 Outils bioinformatiques pour l'analyse des altérations :	20
3.2.1 .Variant calling et annotation :	20
3.2.2 .Analyse de l'expression génique :	20
4. LES PRINCIPAUX TYPES DE PUCES ADN	21
4.1 Les types de microarrays	23
4.2 Les types d'études conduites avec une puce à ADN :	23
4.3 Le pré-traitement des données :	24
PARTIE 2 : MATERIEL ET METHODES	25
1. Collecte des données d'expression génique	25
1.1 Sélection des gènes étudiés :	25
1.2 Prétraitement des données :	26
2. Analyse en Composantes Principales (ACP) :	26
2.1 Critères de sélection des gènes :	27
2.2 Visualisation et interprétation des résultats :	27
3. Analyse par carte thermique (Heatmap) :	27
4. Analyse différentielle :	28
5. Analyse bioinformatique d'enrichissement fonctionnel	30
PARTIE 3 : RÉSULTATS & DISCUSSION	31
1. ANALYSE EN COMPOSANTES PRINCIPALES	31
2. ANALYSE DIFFÉRENCIELLE	34
CONCLUSION	40
RÉFÉRENCES BIBLIOGRAPHIQUES	42
LES SITES CONSULTÉS	49

Liste des figures

Figure 1 : Interface principale de la base de données TCGA.....	18
Figure2: Répartition des gènes les plus fréquemment mutés.....	19
Figure 3: Graphique de survie globale pour le cas du gène KRAS.....	20
Figure 4: Représentation schématique des étapes de la réalisation de microarrays.	22
Figure 5: Une portion de l'interface du fichier CSV.	25
Figure 6: Une portion de l'interface du fichier xlsx.	25
Figure 7: Exploration des structures multivariées entre patients et gènes par ACP (24 gènes, 1000 patients).....	32
Figure 8 : Analyse différentielle de l'expression de 24 gènes liés à l'Asthme, au cancer et à l'Alzheimer : représentation Volcano Plot.	34
Figure 9 : Diagramme de Venn illustrant l'analyse comparative des gènes impliqués dans les processus biologiques enrichis.	37
Figure 10 : Visualisation des catégories GO (Biological Process) enrichies pour les gènes différentiellement exprimés.....	38
Figure 11 : Implications des gènes dans les processus biologiques.	38
Figure 12 : Implication des gènes dans les catégories fonctionnelles moléculaires.....	39

Liste des tableaux

Tableau 1: profile d'expression transcriptomique de gènes candidats impliqués dans les voies de signalisation pathologiques étudiées.	26
Tableau 2: Valeurs $\log_2$ du Fold Change (FC) et de $\log_p$ Values pour les 24 gènes analysés.	29
Tableau 3: une portion de la matrice de corrélations	31
Tableau 4 : les analyses différentielles entre les deux classes de patients.	35
Tableau 5 : Analyse de l'expression différentielle des gènes selon le statut de significativité.	35

Liste des abréviations

ALK	Anaplastic Lymphoma Kinase
APOE	Apolipoprotein E
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase
BRCA1	Breast Cancer 1
BRCA2	Breast Cancer 2
CCL5	C-C Motif Chemokine Ligand 5
CD14	Cluster of Differentiation 14
CDH1	Cadherin 1
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
CLU	Clusterin
EGFR	Epidermal Growth Factor Receptor
ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
ERBB4	Erb-B2 Receptor Tyrosine Kinase 4
FGFR2	Fibroblast Growth Factor Receptor 2
FISH	Fluorescence In Situ Hybridization
GATA3	GATA Binding Protein 3
GSTP1	Glutathione S-Transferase Pi 1
HER1	Human Epidermal Growth Factor Receptor 1
HER2	Human Epidermal Growth Factor Receptor 2
HLA-DRB1	Major Histocompatibility Complex, Class II, DR Beta 1

IHC	Immunohistochemie
IL33	Interleukin 33
KIT	KIT Proto-Oncogene, Receptor Tyrosine Kinase
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
LKB1	Liver Kinase B1
LRP2	LDL Receptor Related Protein 2
LRRN1	Leucine Rich Repeat Neuronal 1
MAPT	Microtubule Associated Protein Tau
MET	Mesenchymal-Epithelial Transition factor receptor
NGS	Next-Generation Sequencing
NKX3-1	NK3 Homeobox 1
NTRK2	Neurotrophic Receptor Tyrosine Kinase 2
ORMDL3	ORMDL Sphingolipid Biosynthesis Regulator 3
P53	Tumor Protein P53
PTEN	Phosphatase and Tensin Homolog
RB1	Retinoblastoma 1
RET	Rearranged during Transfection
ROS1	ROS Proto-Oncogene 1, Receptor Tyrosine Kinase
SPINK8	Serine Peptidase Inhibitor Kazal Type 8
STK11	Serine/Threonine Kinase 11
WES	Whole Exome Sequencing

INTRODUCTION

Le développement des approches de médecine personnalisée repose de plus en plus sur une compréhension fine des mécanismes moléculaires impliqués dans les pathologies humaines. Parmi ces mécanismes, les signatures génomiques et transcriptomiques permettent d'identifier des profils spécifiques d'expression génique qui caractérisent les différents états pathologiques, en particulier dans les maladies complexes comme le cancer, les troubles neurodégénératifs et les maladies inflammatoires chroniques. L'essor des technologies à haut débit, telles que les puces à ADN Affymetrix, a rendu possible l'analyse simultanée de l'expression de milliers de gènes, ouvrant ainsi la voie à une meilleure stratification des patients et à des thérapies ciblées (Tung, 2006 ; Schena, 2000 ; Zhang, 2005).

Dans ce travail, nous nous intéressons à l'analyse des signatures génomiques et transcriptomiques associées aux tumeurs pulmonaires, avec pour objectif ultime de contribuer à l'identification de marqueurs moléculaires utiles en médecine personnalisée. Pour cela, nous exploitons un jeu de données transcriptomiques issu de la technologie Affymetrix, couvrant 1000 patients et l'expression de 24 gènes sélectionnés pour leur implication dans trois grandes classes de pathologies :

- Les cancers (incluant le cancer du sein, de la prostate et du poumon).
- L'asthme.
- La maladie d'Alzheimer.

L'analyse de ces gènes à travers différentes pathologies permettra de dégager des signatures spécifiques aux tumeurs pulmonaires, mais aussi de détecter d'éventuels profils partagés ou divergents avec d'autres affections.

Les objectifs principaux de ce travail sont les suivants :

1. Analyser les profils d'expression génique des 24 gènes sélectionnés dans les tumeurs pulmonaires, en les comparant aux autres affections étudiées.
2. Identifier des signatures transcriptomiques caractéristiques des tumeurs pulmonaires à partir des données Affymetrix.

PARTIE 1 : Revue bibliographique

1. LES BASES MOLECULAIRES DU CANCER DU POUMON

Le cancer résulte de la prolifération incontrôlée d'une seule cellule, qui finit par former une tumeur. Au départ, cette tumeur reste localisée dans l'organe d'origine, mais elle peut ensuite s'étendre aux tissus voisins. Elle peut également se propager à d'autres parties du corps sous forme de métastases, ce qui peut entraîner la mort. En théorie, toute cellule d'un organe peut se transformer en cellule cancéreuse sous l'influence de facteurs génétiques ou environnementaux. (Boulet, 2013).

Le cancer du poumon se développe à partir de cellules des bronches. Il existe deux principaux types (catégories) de cancers du poumon en fonction de l'origine cellulaire des bronches dont ils sont issus (Fervers et al., 2023) :

- i. Les cancers bronchiques non à petites cellules (CBNPC), qui représentent près de 85 % des cancers du poumon. Les sous-types majeurs sont : l'adénocarcinome (le plus fréquent chez les non-fumeurs), le carcinome épidermoïde (squameux) et le carcinome à grandes cellules.
- ii. Les cancers bronchiques à petites cellules (CBPC), qui représentent près de 15 % des cancers du poumon. Très agressif, avec une croissance rapide et est souvent diagnostiqué à un stade avancé.

1.1. Gènes impliqués : De nombreux gènes, oncogènes et gènes suppresseurs, sont impliqués dans le cancer bronchique. Les oncogènes et les gènes suppresseurs de tumeurs sont deux grandes catégories de gènes impliqués dans la cancérogénèse, mais agissent de manière opposée dans le contrôle de la prolifération cellulaire. Les oncogènes sont des versions mutées ou sur-exprimées de proto-oncogènes, des gènes normaux qui régulent positivement la croissance et la division cellulaire. Lorsqu'un proto-oncogène subit une mutation activatrice (gain de fonction), il peut devenir un oncogène, conduisant à une prolifération cellulaire incontrôlée. Un seul allèle muté suffit généralement à induire cet effet (mode dominant). La liste de ces gènes inclut, à titre d'exemple, KRAS, MYC et EGFR. Les gènes suppresseurs de tumeurs agissent comme des freins de la prolifération cellulaire. Ils codent pour des protéines impliquées dans la régulation du cycle cellulaire, la réparation de l'ADN ou l'induction de l'apoptose (mort cellulaire programmée). Lorsqu'ils sont inactivés par mutation (souvent une perte de fonction), les

mécanismes de contrôle de la division cellulaire sont levés ; ce qui favorise le développement tumoral. Contrairement aux oncogènes, les deux allèles doivent généralement être inactivés pour entraîner un effet cancérigène, comme dans le cas de TP53, RB1, BRCA1 et PTEN (Tomasini et al., 2025).

1.1.1 Oncogènes : L'EGFR, aussi appelé HER1, est un récepteur membranaire à activité tyrosine kinase impliqué dans la croissance des tumeurs. Codée par un gène localisé sur le chromosome 7, ce gène code pour une protéine qui possède un poids moléculaire de 170 kD, comporte 53 acides aminés et trois ponts disulfures. Cette protéine transmembranaire possède trois domaines à savoir un domaine extracellulaire capable de fixer certains ligands comme l'EGF, un domaine transmembranaire et un domaine intracellulaire. Cette protéine transmembranaire joue un rôle central dans les voies de signalisation responsables de l'apparition et du développement de nombreux cancers. Dans les cellules cancéreuses, les récepteurs de l'EGFR sont souvent surexprimés, le gène peut être amplifié, ou des mutations peuvent survenir. Lorsqu'une mutation activatrice est présente, elle déclenche une cascade de signaux favorisant la prolifération cellulaire, la formation de nouveaux vaisseaux sanguins (angiogenèse) et la capacité des cellules à se disséminer dans l'organisme. Cibler et bloquer ces récepteurs permet ainsi de freiner la multiplication des cellules tumorales (Arcagy-Gineco, 2025).

KRAS (Kirsten Rat Sarcoma Viral Oncogene Homolog), les mutations du gène KRAS figurent parmi les altérations génétiques les plus courantes dans le cancer du poumon non à petites cellules (CPNPC), en particulier dans environ 25 à 30 % des adénocarcinomes pulmonaires. Elles concernent majoritairement les codons 12 et 13, conduisant à une activation permanente de la protéine KRAS. Cette activation déclenche une prolifération cellulaire incontrôlée par le biais des voies de signalisation RAF-MEK-ERK et PI3K-AKT. Contrairement aux mutations de l'EGFR, celles de KRAS sont plus fréquemment observées chez les fumeurs. Elles ont longtemps été associées à un pronostic défavorable et à une résistance aux thérapies ciblées conventionnelles. (Ferrer et al., 2018).

ALK (Anaplastic Lymphoma Kinase) localisé sur le bras court du chromosome 2, plus précisément en position 2p23, code pour une tyrosine kinase réceptrice impliquée dans le développement de certains cancers, notamment le cancer du poumon non à petites cellules (CPNPC). Dans environ 5 % des adénocarcinomes pulmonaires, une translocation chromosomique aboutit à une fusion entre le gène ALK et un autre gène, le plus souvent EML4. Cette fusion génétique entraîne la production d'une protéine de fusion anormale, responsable de

l'activation continue de voies de signalisation favorisant la croissance et la survie des cellules tumorales.

ROS1 situé sur le bras long du chromosome 6 (bande 2q2), code une protéine transmembranaire à activité tyrosine kinase, exprimée dans plusieurs types de cellules tumorales. Ce récepteur est capable d'activer des voies de signalisation clés, intervenant dans des processus cellulaires tels que la différenciation, la prolifération, la croissance et la survie des cellules.

Dans le cancer du poumon non à petites cellules (CPNPC), des réarrangements du gène ROS1 sont identifiés dans 1 à 2 % des cas, principalement chez des non-fumeurs ou des fumeurs légers, en particulier dans les adénocarcinomes pulmonaires.

La protéine ROS1 se compose de trois régions principales : un domaine kinase intracellulaire, une région transmembranaire et une région extracellulaire. Cette dernière permet la liaison à des ligands spécifiques, déclenchant l'activation du domaine intracellulaire par autophosphorylation, ce qui favorise la transmission de signaux à l'intérieur de la cellule. (Zhang & Yang, 2024)

HER2 est un proto-oncogène qui code une protéine transmembranaire présente à la surface des cellules humaines. Il est codé par le gène ERBB2, localisé sur le bras long du chromosome 17. Bien que le ligand spécifique de HER2 n'ait pas encore été identifié, ce récepteur joue un rôle central dans la régulation de la prolifération cellulaire.

Dans certains cancers, la cellule devient anormale et surexprime HER2, c'est-à-dire qu'elle présente un nombre excessif de récepteurs HER2 à sa surface. Cette surexpression entraîne une stimulation excessive de la division cellulaire, favorisant ainsi la croissance des cellules tumorales. On parle alors de cellules HER2-positives (Roche | Cancer Du Sein : Le Statut HER2, n.d. , Eliahiai, Eljjar, Chaib, KHarmoum, & Chraibi, 2023).

BRAF localisé sur le locus chromosomique 7q34, code une protéine impliquée dans la transmission des signaux chimiques de l'extérieur de la cellule vers son noyau. Cette protéine intervient dans la voie de signalisation RAS/MAPK, qui joue un rôle clé dans la régulation de la croissance, de la division et de la différenciation cellulaires, ainsi que dans des mécanismes comme la migration cellulaire et l'apoptose (mort cellulaire programmée).

Ces signaux sont essentiels au développement normal après la naissance. Le gène BRAF appartient à la catégorie des oncogènes, et lorsqu'il subit une mutation, il peut contribuer au

déclenchement de processus cancéreux en perturbant le contrôle de la prolifération cellulaire. (Delage, s.d).

MET localisé sur le bras long du chromosome 7 (7q31), code un récepteur transmembranaire à activité tyrosine kinase dont le ligand principal est le facteur de croissance hépatocytaire (HGF), également connu sous le nom de scatter factor. Ce récepteur joue un rôle fondamental dans la croissance cellulaire, la motilité, l'angiogenèse et le développement embryonnaire. Dans le contexte du cancer bronchique, l'activation anormale de la voie HGF-MET est associée à un pronostic défavorable, en favorisant notamment l'invasion tissulaire, la formation de métastases et la résistance aux thérapies ciblées, notamment chez environ 20 % des patients atteints de cancer du poumon non à petites cellules présentant une mutation EGFR et traités par inhibiteurs de tyrosine kinase. En fonction du type d'anomalie (surexpression, amplification ou mutation), MET peut agir soit comme un oncogène dans les tumeurs où il est altéré, soit comme un facteur de progression tumorale dans les autres cas. (A.-M. Ruppert, M. Beau-Faller, L. Belmont, A. Lavolé, V. Gounant, J. Cadranet, M. Wislez, 2011).

1.1.2 Gènes suppresseurs de tumeurs : RB1 impliqué dans le rétinoblastome, est situé sur le bras long du chromosome 13 (région 13q14.1) et comporte 27 exons. Il code une phosphoprotéine nucléaire, appelée pRb, qui est exprimée de manière constante dans les cellules. Cette protéine joue un rôle clé dans le contrôle du cycle cellulaire, en particulier au niveau de la transition entre les phases G1 et S. La fonction principale de pRb est de réguler les facteurs de transcription de la famille E2F, responsables de l'activation des gènes nécessaires à la réplication de l'ADN. En se liant au domaine de transactivation de ces facteurs, pRb inhibe leur activité transcriptionnelle. Lorsque pRb est phosphorylée, elle devient inactive (Menasria, D., & Radjel, A. 2021)

P53 localisé sur le chromosome 17p13.1, est un gène suppresseur de tumeur essentiel au maintien de l'intégrité du génome. Il comprend 11 exons et contient plusieurs domaines hautement conservés au cours de l'évolution. Ce gène code pour la protéine p53, une phosphoprotéine de 53 kDa, composée de 393 acides aminés, qui agit principalement en tant que facteur de transcription, la protéine p53 joue un rôle clé dans la lutte contre la tumorigenèse. (Chenau, 2009)

PTEN est situé sur le bras long du chromosome 10, plus précisément en position 10q23.3, il code pour une protéine suppresseur de tumeur qui joue un rôle clé dans la régulation de la croissance, de la prolifération et de la survie cellulaires, en inhibant notamment la voie de

signalisation PI3K/AKT -mTOR, impliquée dans la survie cellulaire et la prolifération. L'inactivation de PTEN, par mutation ou délétion, est observée dans environ 10% des CPNPC et conduit à une activation aberrante de cette voie de signalisation. Cette altération est souvent associée à une résistance aux thérapies ciblées, notamment aux inhibiteurs de tyrosine kinase de l'EGFR. (Pérez-Ramírez et al., 2015)

LKB1 également appelé STK11, ce gène est localisé sur le chromosome 19, en position 19p13.3. Sa séquence génomique, d'environ 23 kilobases, est composée de 10 exons, dont 9 sont codants. Chez l'être humain, la protéine LKB1 résultante est constituée de 433 acides aminés, il code pour une kinase impliquée dans la régulation du métabolisme énergétique cellulaire et la polarité cellulaire. Les mutations inactivatrices de STK11 sont retrouvées dans 15-30% des adénocarcinomes pulmonaires et sont fréquemment associées aux mutations de KRAS, ces mutations entraînent une perte de la fonction suppresseur de tumeur de STK11, contribuant ainsi à la progression tumorale et à un phénotype immunosuppresseur. (Skoulidis et al., 2015).

1.1.3 Gènes de réparation de l'ADN : Les défauts dans les systèmes de réparation de l'AND peuvent contribuer à l'accumulation de mutations et à l'instabilité génomique, favorisant ainsi le développement tumoral. Dans le cancer du poumon, plusieurs gènes impliqués dans ces processus sont fréquemment altérés.

BRCA1 et BRCA2, bien connus pour leur implication dans les cancers du sein et de l'ovaire, jouent également un rôle essentiel dans la réparation des cassures double-brin de l'ADN grâce à un mécanisme appelé recombinaison homologue. Même si cela reste relativement rare, des mutations (qu'elles soient héréditaires : germinales ou acquises : somatiques)) ont aussi été observées dans le cancer du poumon. Ces anomalies peuvent rendre les cellules cancéreuses plus sensibles à certains traitements, comme les sels de platine ou les inhibiteurs de PARP, qui ciblent les défauts de réparation de l'ADN. (Paul & Paul, 2014)

ERCC1 est un gène qui code pour une protéine essentielle à la réparation de l'ADN via le mécanisme de réparation par excision de nucléotides (NER). Ce processus permet d'éliminer les lésions importantes de l'ADN causées, entre autres, par les agents alkylants et les rayons UV. Dans le cancer du poumon non à petites cellules (CPNPC), une forte expression d'ERCC1 est liée à une résistance aux traitements à base de sels de platine, car ces médicaments agissent en créant des dommages à l'ADN (adduits) que la voie NER est justement capable de réparer efficacement. (Olaussen et al., 2006)

1.2. Analyses des variants génétiques : En oncologie, avec l'essor des thérapies ciblées, de plus en plus de patients atteints de cancer se voient proposer par leur oncologue ou leur chirurgien des analyses génétiques des cellules tumorales. L'objectif est d'adapter le traitement de manière plus personnalisée. Si un variant pathogène est identifié, le patient peut être orienté vers une consultation d'oncogénétique afin d'évaluer ses caractéristiques génétiques. Ces analyses reposent sur diverses méthodes de détection, chacune présentant des avantages et des limites spécifiques. (Chipoulet et al., 2023).

1.2.1. Méthodes de détection des variants génétiques : Les séquenceurs à lecture longue (long-read) ont grandement amélioré notre compréhension des variants structuraux (VS), qui correspondent à des altérations de l'ADN de plus de 50 paires de bases. Ces VS jouent un rôle clé dans les cellules cancéreuses en modifiant l'activité de gènes essentiels comme les oncogènes ou les gènes suppresseurs de tumeurs, notamment ERBB2 dans le cancer du sein ou ROS1 dans le cancer du poumon. Leur détection reste complexe, car la taille des reads issus du séquençage de nouvelle génération est parfois inférieure à celle des VS eux-mêmes, rendant leur analyse difficile. Toutefois, grâce aux technologies de séquençage à lecture longue, les recherches dans ce domaine ont fait un bond en avant. Depuis 2016, pas moins de 17 VS impliqués dans différents cancers ont été identifiés. Parmi les découvertes marquantes, des chercheurs ont pu, via ce type de séquençage, repérer des VS associés au gène ERBB2, fortement impliqué dans le cancer du sein et lié à des mécanismes de résistance à certains traitements. (Alcimed, 2024).

NGS représente une avancée technologique majeure pour l'analyse moléculaire du cancer du poumon, permettant l'identification simultanée de multiples altérations génétiques avec une sensibilité remarquable. Contrairement aux techniques conventionnelles qui ciblent des mutations spécifiques, le NGS offre une analyse complète du profil moléculaire tumoral, facilitant la détection de variants rares et la quantification de la charge mutationnelle tumorale (TMB), un biomarqueur émergent pour l'immunothérapie. Dans leur étude comparative publiée dans le Journal of Molecular Diagnostics, Sholl et collègues (2016) ont démontré que les panels NGS ciblés permettent d'identifier des altérations cliniquement pertinentes dans plus de 65% des adénocarcinomes pulmonaires, incluant non seulement les mutations classiques dans EGFR et KRAS, mais également des réarrangements complexes et des mutations dans des gènes moins fréquemment testés comme MET, ERBB2 et BRAF. Cette approche pangénomique optimise l'utilisation des échantillons tissulaires limités provenant de biopsies pulmonaires et améliore

considérablement l'accès des patients aux thérapies ciblées appropriées et aux essais cliniques innovants. (Sholl et al., 2016).

WES est une technique de pointe qui cible spécifiquement les régions codantes du génome, appelées exons. Bien qu'elles ne représentent qu'environ 1 à 2 % du génome humain, ces régions concentrent la grande majorité des variants génétiques impliqués dans les maladies. Grâce à cette focalisation, le WES permet une analyse plus précise et économique qu'un séquençage génomique complet, tout en offrant une couverture optimale des mutations les plus pertinentes sur le plan clinique. En pratique, l'analyse génétique des patients repose souvent sur des panels ciblés de gènes, choisis selon les caractéristiques cliniques, mais le WES constitue une alternative puissante pour élargir la recherche de mutations dans un cadre diagnostique ou exploratoire. (Séquençage De L'exome Entier | Macrogen Europe, n.d. , Jacob, 2020).

La FISH en oncologie est une technique permettant d'analyser des régions spécifiques du génome à l'aide de sondes fluorescentes qui s'hybrident précisément sur les zones ciblées. Réalisée sur des coupes tissulaires après analyse histologique en laboratoire d'anatomopathologie, cette méthode vise à détecter des anomalies génomiques de nombre ou de structure ayant une valeur diagnostique, pronostique, et thérapeutique. Dans le cadre du cancer du poumon non à petites cellules, la FISH est utilisée pour rechercher des réarrangements des gènes ALK (environ 5 % des cas), ROS1 (2 %), RET (1 à 2 %) ainsi qu'une amplification du gène MET (3 %), afin d'orienter au mieux la stratégie de traitement personnalisée des patients. (FISH Sur Tissus - Imagenome, 2020).

IHC bien que n'étant pas à proprement parler une méthode d'analyse génétique, l'IHC peut être utilisée comme technique de dépistage pour certaines altérations moléculaires dans le cancer du poumon. L'expression aberrante de protéines, résultant de modifications génétiques sous-jacentes, peut être détectée par des anticorps spécifiques. Par exemple, l'IHC anti-ALK est largement utilisée pour le dépistage des réarrangements ALK, avec une excellente corrélation avec les résultats de la FISH. De même, des anticorps spécifiques dirigés contre les formes mutées de certaines protéines, comme EGFR L858R ou BRAF V600E, ont été développés et peuvent faciliter la détection rapide de ces mutations. (Marchetti et al., 2016).

Biopsie liquide : L'analyse de l'ADN tumoral circulant (ctDNA) présent dans le sang périphérique, également appelée "biopsie liquide", est une approche non invasive qui gagne en importance dans le diagnostic et le suivi des patients atteints de cancer du poumon. Cette méthode permet de détecter les altérations génétiques présentes dans les fragments d'ADN libérés par les cellules tumorales dans la circulation sanguine. La biopsie liquide présente plusieurs avantages : elle évite les risques associés aux biopsies tissulaires invasives, permet de capturer l'hétérogénéité tumorale et facilite le suivi longitudinal de l'évolution moléculaire de la tumeur, notamment pour détecter l'émergence de mécanismes de résistance aux thérapies ciblées. Cependant, sa sensibilité est limitée par la quantité souvent faible de ctDNA, particulièrement dans les stades précoces de la maladie. (Rolfo et al., 2018).

1.2.2. Importance clinique de l'analyse des variants génétiques : L'accès aux tests génétiques en cancérologie devient de plus en plus essentiel à mesure que la recherche progresse, révélant la complexité des cancers auparavant classés selon leur localisation anatomique ou leur aspect histologique. Ces cancers, en réalité, représentent un ensemble de maladies plus rares définies par des altérations moléculaires spécifiques telles que des mutations, des amplifications ou des fusions génétiques. L'analyse de ces altérations nécessite des technologies avancées comme le séquençage de nouvelle génération (NGS). Ce type d'approche ne se limite pas à une meilleure classification des maladies, mais permet aussi d'identifier des cibles thérapeutiques précises, ouvrant ainsi la voie à une médecine personnalisée. Bien que tous les traitements ciblés n'aient pas encore démontré une efficacité optimale, les résultats observés dans des pathologies comme le cancer du poumon valident la pertinence de cette stratégie. Toutefois, l'absence de prise en charge systématique et équitable de ces tests sur le territoire entraîne des disparités d'accès aux soins, soulevant ainsi une question éthique majeure. L'analyse génétique s'impose donc comme un outil central dans le diagnostic, le choix thérapeutique et le suivi des cancers, avec un impact direct sur le pronostic et la qualité de vie des patients. (M. Ducreux, P. Amiel., 2022).

2. ALTERATIONS GENETIQUE ET TRANSCRIPTOMIQUE IMPLIQUÉES

Le cancer du poumon résulte d'altérations génétique et transcriptomique, perturbant les fonctions cellulaires normales. Ces modifications influencent la croissance, la survie et l'adaptabilité des cellules tumorales.

2.1. Mutations génétiques associées au cancer du poumon

2.1.1 Les altérations du gène EGFR : Les mutations somatiques associées au cancer bronchique se trouvent dans les exons 18 à 21 du gène EGFR, qui codent pour l'activité de la tyrosine kinase. Ces altérations se produisent au niveau du site de liaison de l'ATP et incluent des délétions conservant le cadre de lecture, des insertions, ainsi que des mutations ponctuelles de type faux-sens. L'exon 18 est le siège de substitutions affectant le codon 719, où la glycine est remplacée par une cystéine, une alanine ou une sérine (G719X), représentant environ 4 % des altérations identifiées. L'exon 19, quant à lui, est souvent touché par des délétions encadrant les résidus leucine 747 et glutamine 749, représentant près de 44 % des mutations de l'EGFR. Il existe plus d'une vingtaine de variantes de ces délétions, incluant des pertes de 9, 12, 15, 18 ou 24 paires de bases, ainsi que des formes plus complexes comme les délétions-insertions. Des insertions rares ont également été observées dans cet exon, représentant environ 1 % des cas. L'exon 20 est majoritairement impliqué dans des insertions (environ 3 %), la plus fréquente étant la mutation T790M, connue pour conférer une résistance aux inhibiteurs de tyrosine kinase. D'autres mutations ponctuelles rares, telles que V765A et T783A, représentent moins de 1 %. L'exon 21 renferme des mutations activatrices, notamment la substitution L858R, qui représente 41 % de ces altérations, ainsi que L861Q (2 %) et, plus rarement, L861R. À elles seules, la délétion de 15 paires de bases dans l'exon 19 (codons E746 à A750) et la mutation L858R de l'exon 21 comptent pour environ 90 % des mutations retrouvées dans les tumeurs pulmonaires. Par ailleurs, toutes ces mutations n'ont pas la même implication pronostique : les délétions de l'exon 19 sont généralement associées à une meilleure réponse au traitement et à une survie plus prolongée que la mutation L858R de l'exon 21. (Belouadi, Hazam, & Khaldi, 2022)

2.1.2 Le gène KRAS : Les mutations du gène KRAS constituent l'une des altérations génétiques les plus courantes dans le cancer du poumon non à petites cellules (CPNPC), en particulier dans les adénocarcinomes, avec une fréquence estimée entre 20 et 40 % selon les populations. Ces mutations entraînent une activation anormale de la protéine KRAS, une GTPase

impliquée dans les voies de signalisation régulant la prolifération, la différenciation et la survie cellulaire. Lorsque KRAS est muté, il reste dans un état actif permanent, ce qui stimule la croissance incontrôlée des cellules tumorales. Parmi les mutations les plus fréquentes, on retrouve :

a- G12C, qui représente environ 33,6 % des mutations KRAS. Elle résulte d'une substitution de la glycine par une cystéine au niveau du codon 12, rendant la protéine continuellement active. Cette mutation est désormais ciblée par des traitements spécifiques tels que le sotorasib, un inhibiteur de KRAS G12C.

b-G12D, présente dans 23,9 % des cas, implique une substitution de la glycine par un acide aspartique, induisant une activation constante de KRAS.

c-G12V, observée dans 22,1 % des cas, correspond à un remplacement de la glycine par une valine, entraînant une activation similaire (yang et al., 2023).

2.1.3 Le gène ALK : Chez les patients atteints de cancers broncho-pulmonaires non à petites cellules (CBPNPC), il est fréquent d'observer des réarrangements intrachromosomiques impliquant le gène ALK. Bien que ces réarrangements puissent se produire avec différents gènes, la littérature, notamment les travaux de Frankel et al. (2019), met en évidence que le partenaire le plus couramment impliqué est le gène EML4.

Les gènes ALK et EML4 (Echinoderm Microtubule Associated Protein Like 4) sont tous deux localisés sur le bras court du chromosome 2 (2p), mais ils sont transcrits dans des directions opposées. Dans le contexte du cancer du poumon, un réarrangement spécifique, connu sous le nom de variant 1 de la fusion EML4-ALK, est fréquemment observé. Ce réarrangement se produit précisément par une cassure dans le gène EML4, localisée à 3,6 kilobases (kb) en aval de la fin de l'exon 13. Simultanément, une cassure se produit dans le gène ALK, 297 bases en amont du début de l'exon 21. La fusion de ces deux fragments d'ADN crée un nouveau gène chimérique, EML4-ALK variant 1.

La transcription de ce gène de fusion aboutit à la production d'une protéine anormale. Cette protéine chimérique est constituée de la partie N-terminale de la protéine EML4 fusionnée à la partie C-terminale, qui comprend le domaine kinase intracellulaire de la protéine ALK. Fait crucial, cette fusion provoque une activation constitutive (permanente) de l'activité kinase de la partie ALK.

Cette activation anormale de la kinase ALK perturbe plusieurs voies de signalisation intracellulaires essentielles. Ces voies sont normalement impliquées dans la régulation de la croissance et de la division cellulaires, de l'organisation du cytosquelette (la structure interne des cellules), de la migration des cellules et de leur survie. L'activation constitutive de la kinase ALK, due à la fusion EML4-ALK, conduit donc à une prolifération incontrôlée des cellules tumorales, favorise leur capacité à se déplacer et à envahir d'autres tissus, et les rend moins susceptibles de mourir par apoptose (Redouan, 2021).

2.1.4 Le gène ROS1 : Les altérations du gène ROS1 associées au cancer du poumon non à petites cellules (CPNPC) ne sont pas des mutations ponctuelles classiques, comme celles observées pour KRAS ou EGFR, mais plutôt des réarrangements génétiques. Ces réarrangements surviennent lorsqu'un fragment du gène ROS1, situé sur le chromosome 6q22, est fusionné à un autre gène, créant une fusion génique anormale. (Takeuchi et al., 2012 ; Bergethon et al., 2012).

Le point de départ est une cassure de l'ADN au niveau du locus ROS1. Lors de la réparation de cette cassure, le segment terminal (3') de ROS1, qui contient le domaine tyrosine kinase actif, est mal recollé avec la partie initiale (5') d'un autre gène. Ce processus est connu sous le nom de translocation chromosomique. Ce type d'erreur peut se produire spontanément ou à la suite de dommages à l'ADN, provoqués par des facteurs environnementaux ou endogènes (comme le stress oxydatif ou les radiations). (Drilon et al., 2014 ; Yasuda et al., 2012)

Le produit de cette fusion génique est une protéine chimérique qui possède une activité kinasique constitutive — c'est-à-dire que la protéine reste active en permanence, même en l'absence de signal extérieur. Cela provoque une activation incontrôlée de voies de signalisation intracellulaires (comme PI3K/AKT et MAPK), ce qui favorise la prolifération cellulaire, la survie des cellules tumorales et l'échappement à l'apoptose. (Gainor et al., 2013 ; Lin et al., 2017)

Parmi les partenaires de fusion les plus fréquemment observés avec le gène ROS1, on retrouve notamment CD74-ROS1 (la fusion la plus courante), SLC34A2-ROS1, EZR-ROS1 et TPM3-ROS1. Ces réarrangements génétiques entraînent la formation de protéines de fusion dotées d'une activité tyrosine kinase constitutive, favorisant la prolifération tumorale. Il est important de noter que les fusions impliquant ROS1 sont généralement mutuellement exclusives avec d'autres mutations oncogéniques majeures, telles que celles affectant EGFR ou ALK, ce qui signifie qu'un patient porteur d'une fusion ROS1 ne présentera en principe pas de mutation concomitante dans ces autres gènes. (Bergethon et al., 2012 ; Shaw et al., 2014)

2.1.5 Le gène BRAF : Dans le cadre du cancer du poumon non à petites cellules (CPNPC), le gène BRAF peut subir des altérations génétiques, principalement des mutations ponctuelles, qui affectent le fonctionnement de sa protéine codée. Ces mutations prennent généralement naissance à la suite d'erreurs lors de la réplication de l'ADN ou sous l'effet de facteurs endogènes (comme le stress oxydatif) ou exogènes (tels que le tabagisme ou les radiations), qui provoquent des lésions dans l'ADN. L'une des mutations les plus fréquentes est appelée V600E, où une valine est remplacée par un acide glutamique au codon 600. Cette substitution rend la protéine BRAF active en permanence, même en l'absence de signal extracellulaire, ce qui déclenche une activation continue de la voie de signalisation MAPK/ERK. Cela entraîne une croissance cellulaire incontrôlée et favorise la progression tumorale. (Cardarella & Johnson, 2013)

Environ 50 % des mutations observées dans le gène BRAF chez les patients atteints de CPNPC ne sont pas de type V600E. Il existe d'autres variantes telles que G469A, D594G ou K601E, qui se situent généralement dans les exons 11 et 15. Ces mutations présentent des niveaux d'activité enzymatique différents : certaines sont fortement activatrices, tandis que d'autres montrent une activité kinase réduite ou inexistante. Néanmoins, elles ont en commun le potentiel de dérégler les mécanismes de prolifération cellulaire et de réduire la sensibilité des cellules tumorales à l'apoptose. (Paik et al., 2011 ; Planchard et al., 2016)

2.1.6 Le gène HER2 : Les mutations du gène HER2 (également appelé ERBB2) dans le cancer du poumon non à petites cellules (CPNPC) constituent un sous-type moléculaire rare mais cliniquement significatif, représentant environ 1 à 4 % des cas. Ces altérations sont davantage observées chez les femmes, les non-fumeuses, et les patientes présentant un adénocarcinome pulmonaire. (Li et al., 2023)

Le type de mutation le plus fréquent affectant HER2 dans ce contexte est une insertion en cadre de lecture dans l'exon 20 du gène, notamment les variantes A775_G776insYVMA et G778_P780dup. Ces insertions sont le résultat d'erreurs lors de la réplication de l'ADN ou de la réparation de cassures double-brin, pouvant être induites par des stress cellulaires ou des facteurs environnementaux. Ce type de mutation conduit à une activation constitutive de l'activité tyrosine kinase de la protéine HER2, indépendamment de la liaison d'un ligand extracellulaire. En conséquence, les voies de signalisation en aval, notamment PI3K/AKT et MAPK, sont activées de façon permanente, favorisant la prolifération cellulaire incontrôlée et l'échappement à l'apoptose. (Zhu et al., 2020 ; Li et al., 2020)

De plus, les mutations HER2 sont généralement mutuellement exclusives avec d'autres altérations oncogéniques comme les mutations EGFR, les réarrangements ALK ou ROS1, ce qui signifie qu'un patient présentant une mutation HER2 ne présente en principe pas d'autres mutations oncogéniques majeures concomitantes. (Li et al., 2016)

2.1.7 Le gène MET : Les altérations du gène MET, notamment le "skipping" de l'exon 14 (METex14), constituent un mécanisme oncogénique bien caractérisé dans le cancer du poumon non à petites cellules (CPNPC). Localisé sur le chromosome 7q31, le gène MET code un récepteur à activité tyrosine kinase, dont le ligand est le facteur de croissance hépatocytaire (HGF). Lorsqu'une mutation affecte les sites d'épissage entourant l'exon 14, ce dernier est exclu de l'ARNm mature. Cette modification, qui survient généralement par mutation ponctuelle ou délétion, empêche la fixation de la ligase CBL, responsable de la dégradation du récepteur MET. En conséquence, la protéine MET est stabilisée, accumulée à la surface cellulaire, et reste activée de manière prolongée.

Cette activation aberrante déclenche une cascade de signalisation intracellulaire via les voies PI3K/AKT, STAT3 et RAS/ERK, favorisant une prolifération cellulaire incontrôlée, la migration et l'invasion tumorale. Dans le tissu pulmonaire, cela se traduit par une croissance accélérée des cellules épithéliales malignes, une infiltration des structures avoisinantes et une capacité accrue à métastaser (Lu et al., 2019). Cliniquement, les patients atteints de CPNPC avec une mutation METex14 présentent souvent des tumeurs peu différenciées, avec un phénotype agressif, notamment dans les sous-types sarcomatoïdes ou adénoquameux. De plus, ces mutations sont associées à une résistance intrinsèque ou acquise à certaines thérapies ciblées classiques, mais peuvent répondre favorablement à des inhibiteurs spécifiques du MET, comme le capmatinib ou le tepotinib.

Enfin, ces mutations sont généralement mutuellement exclusives avec d'autres altérations majeures (EGFR, ALK, ROS1), ce qui en fait un cible thérapeutique indépendante dans la prise en charge personnalisée du CPNPC. (Awad et al., 2021 ; Guo et al., 2020)

2.1.8 Le gène P53 : Pour le gène p53, il est important de souligner que les mutations qui l'affectent sont parmi les plus fréquentes dans le cancer du poumon (Olivier, Hollstein, & Hainaut, 2010). Ces mutations peuvent se produire à divers niveaux du gène et sont souvent associées à des altérations dans les domaines de la protéine p53. (Vassilev, 2007)

Les mutations de p53 peuvent résulter de différentes erreurs génétiques, notamment des substitutions de bases dans l'ADN, des délétions ou des insertions, qui affectent principalement les exons 5 à 8 du gène (Muller et al., 2006). L'une des mutations les plus courantes dans le cancer du poumon est la mutation dans le codon 175, 245 et 273, où des substitutions de bases entraînent la perte de la fonction de la protéine p53 (Liu et al., 2004). Cela empêche la protéine de jouer son rôle crucial dans l'arrêt du cycle cellulaire et dans la réparation de l'ADN endommagé. (Soussi, 2007)

L'effet de ces mutations est particulièrement délétère, car la perte de fonction de p53 permet à des cellules endommagées de se diviser de manière incontrôlée, favorisant ainsi la formation de tumeurs (Olivier, Hollstein, & Hainaut, 2010). Dans le cas du cancer du poumon, cette altération génétique peut entraîner une progression plus rapide de la maladie, une invasion locale accrue et une résistance aux traitements, notamment à la chimiothérapie. (Muller et al., 2006)

2.1.9 Le gène RB1 : Les mutations du gène RB1 dans le cancer du poumon, sont très fréquentes et représentent un événement majeur dans la tumorigénèse. Ces altérations touchent principalement les exons codant pour le domaine fonctionnel de liaison à E2F, essentiel pour l'inhibition du cycle cellulaire. Les mutations identifiées sont diverses, substitutions de bases, délétions, insertions, mais aussi des mutations non-sens, frameshift et mutations ponctuelles, entraînant souvent une perte de fonction complète de la protéine pRb. La majorité de ces mutations sont inactivantes, elles provoquent soit une troncature de la protéine, soit une altération de son domaine de liaison, ce qui empêche pRb de se fixer efficacement aux facteurs de transcription E2F, entraînant une activation non contrôlée du cycle cellulaire. La localisation des mutations est hétérogène, mais les régions les plus fréquemment touchées sont les exons 17 à 23, qui codent pour le cœur du domaine de liaison à E2F. La conséquence biologique majeure de ces mutations est la perte de contrôle du point de restriction G1/S, ce qui permet aux cellules de proliférer sans régulation. Dans le CPPC, près de 90 % des tumeurs présentent une inactivation biallélique de RB1, associée fréquemment à la perte de TP53, soulignant leur rôle coopératif dans la progression tumorale et la résistance aux thérapies standards. (George et al., 2015 ; Peifer et al., 2012).

2.1.10 Le gène PTEN : Le gène PTEN, est inactivé dans 10-15% des cancers pulmonaires, principalement par perte d'hétérozygotie, mutations ponctuelles ou mécanismes épigénétiques (Zhang et al., 2021). PTEN code pour une phosphatase qui régule négativement la voie de signalisation PI3K/AKT/mTOR, cruciale pour la croissance et la survie cellulaires. La

perte de fonction de PTEN dans les cancers pulmonaires entraîne une activation constitutive de cette voie, contribuant à la prolifération cellulaire, l'angiogenèse et la résistance à l'apoptose. Les altérations de PTEN sont plus fréquentes dans les adénocarcinomes que dans les carcinomes épidermoïdes et sont associées à une résistance aux thérapies ciblées, notamment les inhibiteurs de l'EGFR (Wu et al., 2022). Les patients présentant des altérations de PTEN montrent généralement une survie sans progression réduite et un moins bon pronostic global (Meng et al., 2024).

2.1.11 Le gène LKB1 : Les mutations du gène LKB1/STK11 sont majoritairement inactivatrices et se présentent sous diverses formes, notamment des mutations non-sens, des mutations frameshift (par délétion ou insertion) ainsi que des mutations faux-sens affectant des résidus clés pour l'activité kinase. Ces mutations sont souvent localisées dans les exons codant pour le domaine catalytique de la protéine, en particulier les exons 1 à 9. Elles entraînent la perte de fonction de LKB1, compromettant sa capacité à activer l'AMPK et à réguler le métabolisme cellulaire. L'une des conséquences de cette inactivation est une dérégulation de l'homéostasie énergétique, favorisant un environnement tumoral plus agressif et résistant à l'immunothérapie. Ces altérations sont également associées à une faible infiltration des lymphocytes T et à une expression réduite de PD-L1, ce qui limite l'efficacité des inhibiteurs de points de contrôle immunitaire chez les patients porteurs de mutations LKB1 (Skoulidis et al., 2018 ; Koyama et al., 2016).

2.1.12 Les gènes BRCA1/2 : Les gènes BRCA1 (situé sur le chromosome 17q21) et BRCA2 (localisé sur le chromosome 13q12.3) sont responsables de la production de protéines cruciales pour la réparation des cassures double brin de l'ADN par recombinaison homologue. Les mutations affectant ces gènes, souvent sous forme de délétions, insertions, mutations ponctuelles ou frameshift, entraînent une inactivation des protéines, ce qui perturbe leur fonction réparatrice. Cette défaillance conduit la cellule à recourir à des mécanismes de réparation alternatifs, moins précis, augmentant ainsi le risque d'accumulation d'anomalies génétiques et le développement de cancers. Ces mutations sont héritées selon un mode autosomique dominant, et elles augmentent la probabilité de développer des cancers du sein, des ovaires et, dans certains cas, du poumon. (Docteur Benchimol, s.d. ; Edimark, 2021)

2.1.13 Le gène ERCC1 : Les altérations du gène ERCC1, situé sur le chromosome 19q13.32, se présentent majoritairement sous forme de polymorphismes nucléotidiques simples (SNPs) qui compromettent la fonction de la protéine ERCC1, indispensable au mécanisme de

réparation de l'ADN par excision de nucléotides (NER). Ces variations génétiques affectent fréquemment des régions codantes ou régulatrices impliquées dans l'interaction avec la protéine XPF, perturbant ainsi la formation et l'efficacité du complexe ERCC1-XPF, ce qui diminue la capacité des cellules à corriger les lésions de l'ADN. Cette déficience fonctionnelle conduit à une accumulation de dommages génétiques et à une instabilité du génome, favorisant ainsi l'émergence et la progression du cancer bronchique non à petites cellules (CBNPC). En outre, le niveau d'expression d'ERCC1 constitue un biomarqueur de grande importance : une expression réduite est corrélée à une meilleure réponse au cisplatine, tandis qu'une expression élevée est associée à un meilleur pronostic global, mais à une réponse thérapeutique plus faible (EM-consulte, 2010 ; PMC, 2011).

3. APPROCHES BIOINFORMATIQUES ET RESSOURCES POUR L'ÉTUDE DES ALTÉRATIONS MOLÉCULAIRES

L'essor des technologies à haut débit, telles que le séquençage de nouvelle génération (NGS) et les puces à ADN, a généré une quantité massive de données génomiques et transcriptomiques dans les cancers, notamment celui du poumon. L'analyse et l'interprétation de ces données nécessitent le recours à des outils bioinformatiques spécialisés et à des ressources en ligne permettant de caractériser les altérations moléculaires, d'en explorer les conséquences biologiques, et d'identifier des cibles thérapeutiques potentielles.

3.1. Ressources en ligne et bases de données génomiques : De nombreuses bases de données ont été développées pour centraliser les informations sur les mutations, l'expression génique et les signatures transcriptomiques dans les cancers :

The Cancer Genome Atlas (TCGA) : Hébergé sur le National Institute of Health (USA), il fournit des données multi-omiques sur des milliers d'échantillons tumoraux, dont le cancer du poumon (figure 1). Il a caractérisé moléculairement plus de 20 000 cancers primitifs et a comparé des échantillons normaux couvrant 33 types de cancer .



Figure 1 : Interface principale de la base de données TCGA

(<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>)

La base de données TCGA mises à jour pour tous les types de cancers (Figure 2) en particulier celui du poumon, permet de consulter les fréquences des mutations des gènes associés au cancer du poumon observées dans une cohorte de 1414 cas.

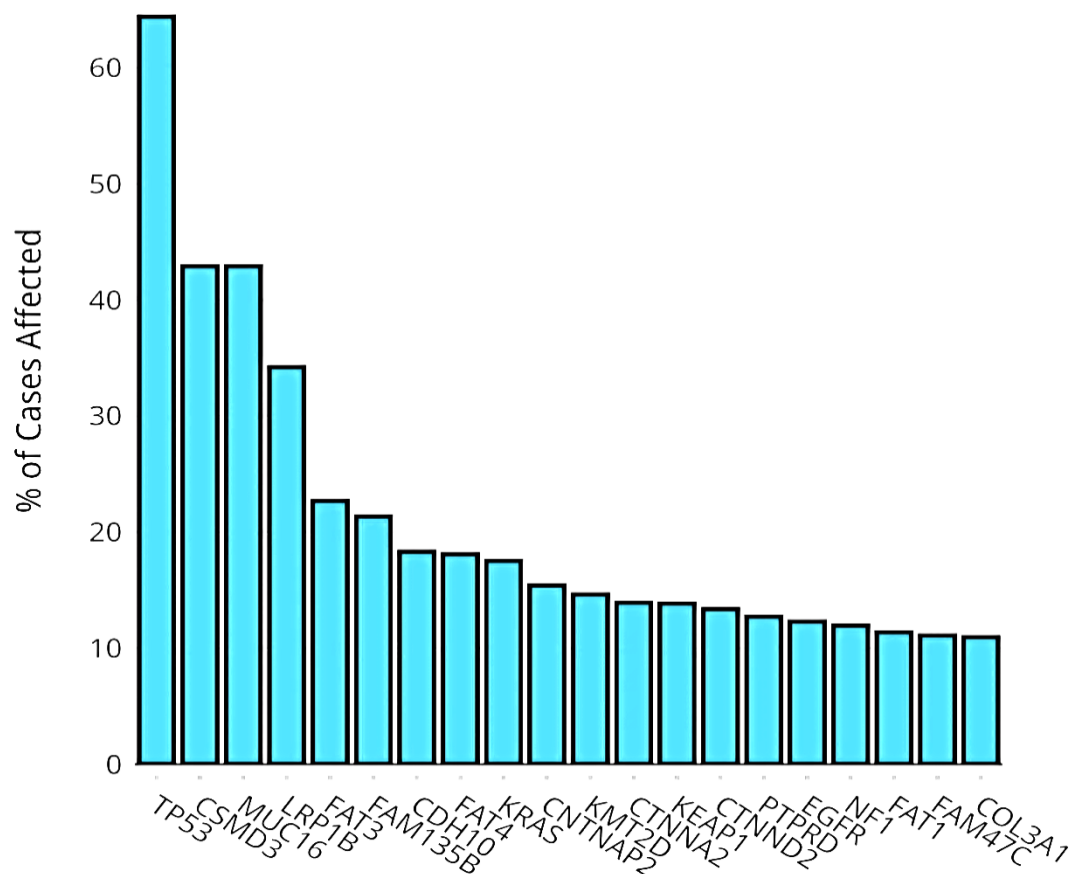


Figure2:Répartition des gènes les plus fréquemment mutés.

(https://portal.gdc.cancer.gov/analysis_page?app=MutationFrequencyApp). Consulté le 28 mai 2025.

Le gène EGFR est muté dans 174 cas sur 1414 sujets, soit une fréquence de mutation de 12,3%. Le KRAS, sur la même cohorte a muté dans 248 cas, soit une fréquence de 17,5%. De plus la survie à ces mutations est différente chez les individus mutés et ceux sauvages. En effet la figure 3 illustre clairement cette distinction pour le cas du gène KRAS. Il est à constater que les patients KRAS-muté ont un taux de survie bien inférieur à celui des non-mutés ; le gène KRAS serait alors associé à un pronostic défavorable ($p=6,45 \times 10^{-2}$).

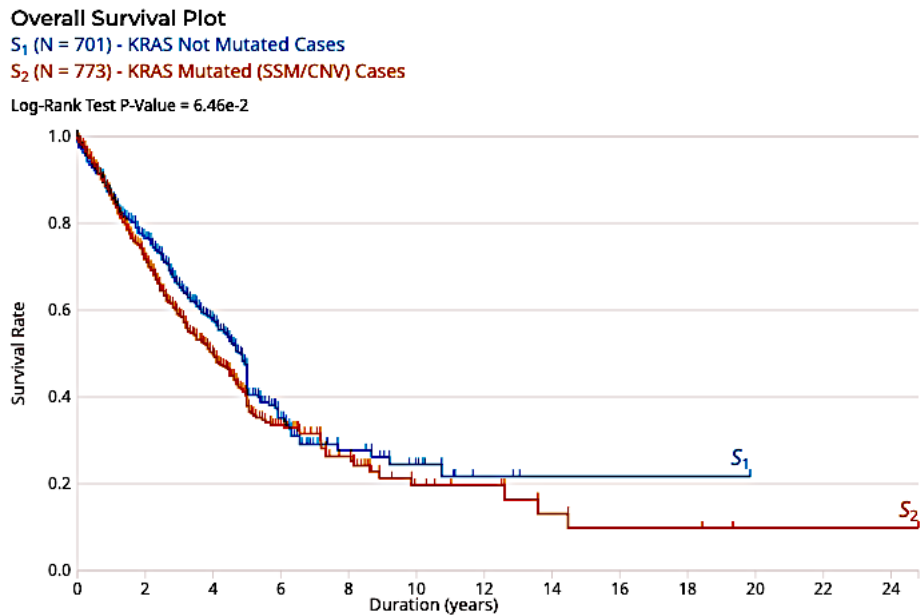


Figure 3: Graphique de survie globale pour le cas du gène KRAS.

(https://portal.gdc.cancer.gov/analysis_page?app=MutationFrequencyApp).

La dbSNP : Répertoire les polymorphismes nucléotidiques et leurs annotations fonctionnelles. Créée en 1998 par le National Center for Biotechnology Information (NCBI), la base de données dbSNP (<https://www.ncbi.nlm.nih.gov/snp>) est une ressource essentielle en génomique pour cataloguer les petites variations génétiques. En 25 ans, dbSNP s'est enrichie de plus de 4,4 milliards de SNP soumis et de 1,1 milliard de SNP de référence unique, fournissant des données essentielles à l'identification des variants génétiques liés aux maladies et à l'étude de la diversité humaine. L'intégration de projets à grande échelle tels que 1000 Genomes, gnomAD, TOPMed et ALFA a permis d'élargir le catalogue de variations génétiques humaines de dbSNP, augmentant ainsi son utilité pour la recherche et les applications cliniques. À l'affût des avancées technologiques telles que le séquençage de nouvelle génération et les infrastructures cloud, dbSNP demeure une pierre angulaire de la recherche génétique, soutenant les découvertes continues en médecine de précision et en génomique des populations.

3.2 Outils bioinformatiques pour l'analyse des altérations : L'exploitation des données issues du séquençage ou des puces génomiques fait appel à une large variété de logiciels et pipelines bioinformatiques.

3.2.1 Variant calling et annotation : les outils tels que GATK et Mutect2 permettent la détection de variants somatiques ; ou ANNOVAR pour l'annotation fonctionnelle des mutations.

3.2.2 Analyse de l'expression génique : les outils DESeq2 et edgeR permettent l'analyse différentielle de l'expression à partir de données RNA-seq, c'est-à-dire des séquences

brutes d'ARN transformées en ADNc et alignées sur un génome référence. Le logiciel Limma est utilisé dans le cas des données de type microarray.

4. LES PRINCIPAUX TYPES DE PUCES ADN

L'étude du profil d'expression génique des cellules et des tissus est devenue un outil majeur de découverte en médecine. Les expériences sur puces à ADN permettent de décrire les variations d'expression à l'échelle du génome, tant en santé qu'en maladie. De plus, une étude impartiale et systématique du profil d'expression génique devrait permettre d'établir une nouvelle taxonomie des syndromes obstétricaux et gynécologiques. Ainsi, une nouvelle ère s'ouvre où les processus et troubles de la reproduction pourraient être caractérisés à l'aide d'outils moléculaires et d'empreintes génétiques.

La conception, l'analyse et l'interprétation des expériences sur puces à ADN requièrent des connaissances spécialisées qui ne font pas partie du cursus standard de notre discipline. Les auteurs ont décrit les types d'études réalisables avec des expériences sur puces à ADN (comparaison de classes, prédiction de classes, découverte de classes) et ont abordé les questions clés relatives à la conception expérimentale, au prétraitement des données et aux méthodes de sélection des gènes. Les types courants de représentation des données sont illustrés. Les pièges potentiels dans l'interprétation des expériences de microarray, ainsi que les atouts et les limites de cette technologie, sont mis en évidence (Tung, 2006).

Les puces à ADN peuvent mesurer simultanément le niveau d'expression de milliers de gènes dans un échantillon d'ARNm particulier. Un tel profilage d'expression à haut débit peut être utilisé pour comparer le niveau de transcription des gènes dans des conditions cliniques afin de :

- a) identifier des biomarqueurs diagnostiques ou pronostiques .
- b) classer les maladies (par exemple, les tumeurs avec un pronostic différent qui sont indiscernables par examen microscopique) .
- c) surveiller la réponse au traitement .
- d) comprendre les mécanismes impliqués dans la genèse des processus pathologiques. Pour ces raisons, les puces à ADN sont considérées comme des outils importants pour la découverte en médecine clinique (Schena, 2000 ; Zhang, 2005).

Une puce à ADN peut être considérée comme une analyse de Southern ou de Northern blot parallèle à grande échelle (au lieu d'un gel, les sondes sont fixées sur une surface inerte, qui

deviendra la puce à ADN). L'ARNm est extrait des tissus ou des cellules, rétrotranscrit et marqué avec un colorant (généralement fluorescent), puis hybridé sur la puce, comme illustré à la figure 4. L'hybridation et les lavages sont réalisés dans des conditions de stringence élevée afin de minimiser le risque d'hybridation croisée entre gènes similaires. L'étape suivante consiste à générer une image par imagerie par fluorescence induite par laser. Le principe de la quantification des niveaux d'expression est que la quantité de fluorescence mesurée à chaque emplacement spécifique de la séquence est directement proportionnelle à la quantité d'ARNm de séquence complémentaire présente dans l'échantillon analysé. Ces expériences ne fournissent pas de données sur le niveau d'expression absolu d'un gène particulier (concentrations réelles d'ARNm), mais sont utiles pour comparer le niveau d'expression entre différentes conditions et gènes sain ou malade (Knudsen, 2004).

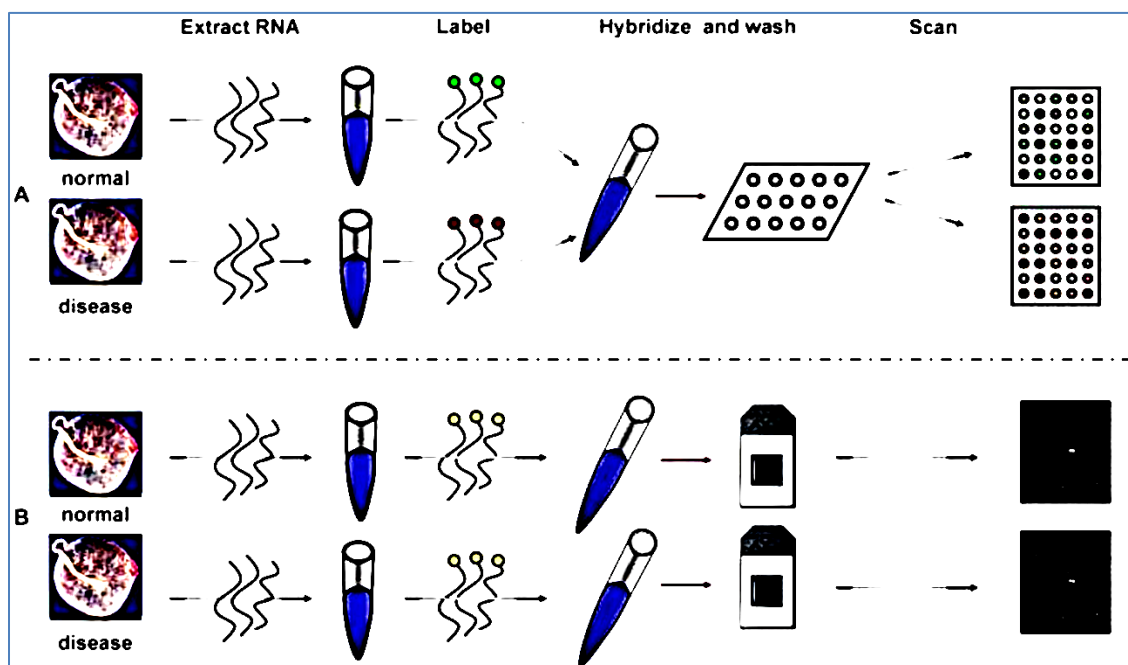


Figure 4: Représentation schématique des étapes de la réalisation de microarrays.

<https://www.sciencedirect.com/science/article/pii/S0002937806008738>

A (le panneau supérieur) illustre la technologie à deux canaux, tandis que B (le panneau inférieur) illustre la technologie à canal unique. Dans le panneau A, les ARNm normaux et pathologiques sont marqués avec deux colorants différents, mélangés puis hybridés sur le même microarray. Après lavage, le microarray est balayé à deux longueurs d'onde différentes pour produire deux images : une pour l'échantillon normal et une pour celui malade. Dans le panneau B (monocanal), chaque échantillon est marqué avec le même colorant fluorescent, mais hybridé indépendamment sur des microarrays différents.

4.1 Les types de microarrays : les micropuces à ADN (ou puces à ADN) peuvent être classées selon trois critères :

i) La longueur des sondes : en fonction de leur longueur, les puces à ADN peuvent être classées en “puces à ADNc”, qui utilisent de longues sondes de plusieurs centaines ou milliers de paires de bases, et en “puces à oligonucléotides”, qui utilisent des sondes courtes (généralement de 50 pb ou moins).

ii) Les méthodes de fabrication comprennent : le « dépôt ou spotting » de séquences précédemment synthétisées et la « synthèse *in situ* ». Généralement, les puces à ADNc sont fabriquées par dépôt, tandis que les puces à oligonucléotides sont fabriquées par des technologies *in situ*. Ces technologies comprennent : la « photolithographie » (par exemple, Affymetrix, Santa Clara, Californie), l’« impression par jet d’encre » (par exemple, Agilent, Palo Alto, Californie) et la « synthèse électrochimique » (par exemple, Combimatrix, Mukilteo, Washington).

iii) Les « puces monocanal » analysent un seul échantillon à la fois, tandis que les « puces multicanaux » peuvent analyser deux échantillons ou plus simultanément. La puce Affymetrix GeneChip est un exemple de puce oligonucléotidique monocanal.

En pratique, le terme « sonde » désigne la séquence nucléotidique fixée à la surface du microarray. Dans les expériences sur microarray, le terme « cible » désigne ce qui est hybridé aux sondes.

4.2 Les types d’études conduites avec une puce à ADN : il existe trois principaux types d'applications des puces à ADN en médecine (Khatri, 2005 ; Tarca, 2006):

1- La première consiste à identifier les différences d'expression entre des groupes prédéfinis d'échantillons. Il s'agit d'une expérience de « comparaison de classes », par exemple, l'identification de gènes différentiellement exprimés dans les placentas de femmes enceintes normales et de femmes atteintes de prééclampsie.

2- Une deuxième application, la « prédiction de classe », consiste à identifier l'appartenance à une classe d'un échantillon en fonction de son profil d'expression génétique (Intelligence Artificielle). Par exemple, il serait possible de prédire si une patiente présente (ou développera) une prééclampsie en fonction de son profil d'expression sanguine. Cela nécessite la construction d'un classificateur (un modèle mathématique) capable d'analyser le profil d'expression génétique d'un échantillon et de prédire son appartenance à une classe. Ce classificateur est construit à

partir d'un ensemble représentatif d'échantillons dont l'appartenance à une classe est connue (par exemple, des femmes ayant une grossesse normale et celles développant ultérieurement une prééclampsie). Ce classificateur sera ensuite utilisé pour évaluer la probabilité de développer une prééclampsie chez les patientes non incluses dans sa construction.

3- Le troisième type d'application consiste à analyser un ensemble donné de profils d'expression génétique afin de découvrir des sous-groupes partageant des caractéristiques communes. Cette application est appelée « découverte de classes ». Par exemple, les profils d'expression d'un grand nombre de femmes atteintes de prééclampsie seront mesurés afin d'identifier des sous-groupes de patientes présentant un profil d'expression génétique similaire. Cet effort vise à générer une taxonomie moléculaire de la maladie. Autrement dit, combien de types moléculaires de prééclampsie (sous-groupes) existent dans un échantillon de femmes atteintes de la maladie ?

4.3 Le pré-traitement des données : Une fois les puces à ADN hybridées, les images obtenues servent à générer un ensemble de données. Cet ensemble de données doit être prétraité avant l'analyse et l'interprétation des résultats.

Le prétraitement est une étape qui extrait ou améliore les caractéristiques significatives des données et les prépare à l'application des méthodes d'analyse. Un exemple typique de prétraitement consiste à calculer le logarithme des valeurs d'intensité brutes. La « normalisation » est un type particulier de prétraitement effectué pour tenir compte des différences systématiques entre les ensembles de données. Un exemple de normalisation consiste à modifier les valeurs d'intensité brutes afin de compenser les différences d'efficacité des colorants lors d'expériences sur puces à ADN à deux canaux utilisant Cy3 (vert) et Cy5 (rouge).

La correction de bruit de fond (Background) est conçue pour ajuster l'hybridation non spécifique, c'est-à-dire l'hybridation de transcrits d'échantillons (cibles) dont les séquences ne correspondent pas parfaitement à celles des sondes sur la puce. Sur les puces à points, l'hybridation non spécifique incluse dans les valeurs d'intensité brutes peut être estimée à partir du niveau de fluorescence à proximité immédiate de la sonde (Yang, 2002).

PARTIE 2 : MATERIEL ET METHODES

1. Collecte des données d'expression génique : L'étude a été réalisée sur un ensemble de données d'expression génique initial de 999 gènes réduit aléatoirement à 24 gènes potentiellement impliqués dans trois maladies distinctes : asthme, cancer (poumon, sein et prostate) et Alzheimer. Les données ont été collectées à partir de 1000 patients enregistrés sur la base de données GEO (Gene Expression Omnibus) du NCBI (National Center for Biotechnology Information) sous format de fichiers CSV (Comma-Separated Values : figure 5) transformés en fichiers tableurs de type excel (figure 6).

Figure 5: Une portion de l'interface du fichier CSV.

Case	LRRN1	3,00 F	TCIM	SPTSSB	INHBB	CLDN1	CXCL3	TNS4	WLS	IER2	JUNB	BTG2	RHOB	EGR3	ATF3	CCN1	EGR1	NR4A1	DUSP1	ZFP63
TCGA-AC-A2QJ	-0.716	1.622	-0.478	-0.282	-0.492	-0.131	-0.045	-0.308	0.315	1.288	1.932	-0.627	-0.560	-0.373	-0.430	1.169	0.440	-0.529	1.483	0.047
TCGA-A2-A3XV	-0.745	-0.447	-0.417	-0.423	-0.701	0.183	-0.353	-0.643	-0.350	1.386	0.358	-0.680	-0.340	-0.574	-0.588	-0.436	-0.502	-0.554	-0.378	-0.577
TCGA-AR-A5QQ	-0.460	8.756	-0.298	-0.395	-0.545	-0.169	0.259	3.432	0.621	0.234	1.800	-0.521	-0.375	-0.355	-0.296	-0.050	-0.404	-0.443	0.624	0.806
TCGA-LL-A6FR	-0.739	-0.407	-0.477	-0.346	3.046	0.793	0.707	1.762	-0.709	2.314	0.065	-0.527	-0.913	-0.530	-0.800	-0.102	-0.488	-0.517	-0.545	-0.428
TCGA-E2-A1L5	-0.140	-0.599	-0.459	0.377	-0.739	0.700	1.150	-0.537	-0.922	0.179	-0.214	0.352	-0.216	0.289	2.073	-0.474	-0.061	0.765	-0.606	0.237
TCGA-AC-A7VC	-0.568	-0.406	-0.025	-0.416	-0.329	-0.296	-0.336	-0.623	1.330	-0.015	0.331	-0.669	-0.571	-0.104	-0.458	-0.184	-0.335	-0.291	0.303	-0.359
TCGA-A2-A4S1	-0.338	0.457	-0.373	-0.430	-0.426	-0.266	0.321	-0.640	2.019	0.134	0.157	-0.614	1.056	-0.283	-0.366	1.118	-0.239	-0.456	3.084	0.369
TCGA-AC-A6XZ	-0.038	-0.599	-0.483	-0.407	-0.232	-0.208	1.634	-0.620	-0.669	-0.039	-0.475	-0.541	-0.652	-0.421	0.242	-0.379	-0.055	-0.517	-0.711	-0.637
TCGA-AC-A2QH	2.981	-0.580	-0.478	-0.422	-0.798	-0.329	-0.523	0.642	2.036	-0.625	-0.824	-0.658	-1.049	-0.432	-0.567	-0.433	-0.067	-0.731	-0.797	-0.738
TCGA-OL-A66P	-0.713	-0.379	-0.388	-0.427	0.477	0.667	-0.130	-0.601	0.009	1.111	1.551	0.299	1.433	-0.273	0.262	1.217	-0.265	0.263	0.515	0.138
TCGA-E9-A5FL	-0.685	-0.049	0.247	-0.403	-0.555	1.221	-0.487	1.243	1.570	3.148	2.303	-0.243	0.021	-0.276	5.626	0.145	0.321	0.395	1.875	3.552
TCGA-EW-A1P8	-0.553	-0.589	0.097	-0.404	-0.617	-0.118	0.322	-0.673	-0.832	0.112	0.037	-0.561	-1.113	-0.599	-0.436	-0.693	-0.610	-0.581	-0.630	-0.540
TCGA-E9-A3X8	-0.301	-0.477	-0.380	0.659	0.335	-0.226	0.152	-0.307	-0.360	4.642	2.945	-0.011	0.574	0.475	0.122	0.488	-0.156	-0.459	-0.147	-0.061
TCGA-OL-A66L	-0.080	1.261	-0.353	-0.109	0.228	-0.102	0.210	-0.215	1.185	18.160	14.803	5.325	5.484	5.773	2.907	7.058	4.060	3.505	7.657	4.772
TCGA-FA-A7DS	-0.748	-0.519	-0.436	-0.425	-0.690	-0.368	0.028	-0.565	-0.846	0.372	-0.223	-0.016	-1.049	0.080	-0.372	-0.628	-0.511	-0.408	-0.556	-0.192
TCGA-LL-A6FP	-0.645	-0.439	-0.225	-0.414	2.004	-0.302	-0.487	-0.428	-0.921	0.351	-0.179	0.356	2.180	-0.612	-0.726	-0.824	-0.609	-0.454	-0.626	-0.691
HCM-CSH-0250-C50	-0.504	-0.713	-0.482	-0.004	3.618	-0.340	-0.544	-0.426	-0.748	0.255	0.288	-0.641	-0.139	-0.619	-0.718	-0.779	-0.344	-0.343	-0.741	-0.440
TCGA-AO-A12B	-0.672	-0.370	-0.451	-0.428	0.337	-0.365	-0.323	-0.602	0.106	-0.374	-0.494	-0.443	-0.156	0.170	-0.519	-0.724	-0.569	-0.494	-0.016	-0.633
TCGA-08-A11T	-0.669	-0.523	-0.401	3.832	-0.408	-0.315	-0.487	2.633	-0.627	-0.631	-0.612	-0.023	0.167	-0.322	-0.295	-0.746	-0.599	-0.529	-0.479	-0.638
TCGA-AO-A0JL	-0.744	-0.404	-0.002	-0.428	0.025	-0.206	-0.099	4.305	0.675	0.641	-0.117	-0.186	-0.666	-0.481	0.186	-0.182	-0.419	-0.433	0.331	-0.405

Figure 6: Une portion de l'interface du fichier xlsx.

1.1 Sélection des gènes étudiés : Au total, 24 gènes parmi 999 ont été analysés et regroupés selon leur association avec les différentes pathologies. Cette sélection est tout à fait aléatoire et n'a fait l'objet d'aucun critère de choix, surtout que l'expression géniques de ces 999 gènes a été réalisée sur une même puce Affymetrix. Ces gènes ont concerné différentes pathologies géniques ; ce qui permet d'atteindre notre objectif de comparaison des expressions des gènes dans les différentes pathologies et de déterminer s'il existe des associations entre ces différentes pathologies.

- a. Cancer du poumon : LRRN1, EGFR, MET, RET, ERBB2, ERBB4, SPINK8, KIT, NTRK2.
- b. Asthme : GSTP1, HLA-DRB1, IL33, CCL5, GATA3, CD14, ORMDL3.
- c. Cancer du sein : CDH1, CDKN2A, FGFR2.
- d. Cancer de la prostate : LRP2, NKX3-1.
- e. Alzheimer : APOE, CLU, MAPT.

1.2 Prétraitement des données : Les niveaux d'expression génique ont été normalisés et représentés sous forme de valeurs numériques comparables entre échantillons.

Les données ont été organisées dans une matrice où chaque ligne correspond à un échantillon patient et chaque colonne à l'expression d'un gène spécifique de l'une des cinq pathologies sus-citées.

Notre base de données finale comporte 1000 cas (patients) avec leurs valeurs d'expression pour les 24 gènes sélectionnés (Tableau 1). Les individus sont partagés en deux groupes : Groupe 1 : comprenant 500 fumeurs et groupe 2 : 500 patients non-fumeurs.

A	B	C	D	E	F	G	H	I	J	K	L
Case	LRRN1	EGFR	MET	RET	ERBB2	ERBB4	SPINK8	KIT	NTRK2	GSTP1	HLA-DRB1
TCGA-AC-A2QJ	-0,716	-0,110	-0,184	-0,583	-0,398	-0,708	-0,297	-0,590	-0,364	3,452	2,746
TCGA-A2-A3XV	-0,745	-0,134	-0,082	-0,582	0,348	-0,691	14,864	-0,620	-0,336	-0,277	-0,241
TCGA-AR-A5QQ	-0,460	0,125	0,189	-0,563	-0,408	-0,684	-0,297	-0,538	-0,300	3,970	1,212
TCGA-LL-A6FR	-0,739	-0,116	0,117	-0,585	-0,272	-0,708	-0,242	-0,226	-0,329	2,205	-0,422
TCGA-E2-A1LS	-0,140	-0,024	0,129	-0,602	-0,423	-0,646	-0,110	4,809	1,466	1,514	-0,859
TCGA-AC-A7VC	-0,568	0,832	-0,178	-0,555	-0,418	-0,707	-0,295	-0,533	-0,101	0,500	-0,142
TCGA-A2-A4S1	-0,338	-0,114	-0,111	-0,593	-0,449	-0,705	-0,297	-0,597	-0,214	0,917	2,659

Tableau 1: profile d'expression transcriptomique de gènes candidats impliqués dans les voies de signalisation pathologiques étudiées.

2. Analyse en Composantes Principales (ACP) : Une Analyse en Composantes Principales a été réalisée à l'aide du logiciel GraphPad Prism pour réduire la dimensionnalité des données et identifier les relations entre les différents profils d'expression génique. L'analyse a été conduite en deux étapes :

a. ACP globale sur l'ensemble des gènes : Application de l'ACP sur la matrice d'expression des 24 gènes pour les 1000 échantillons. Cette analyse a permis d'obtenir une représentation bidimensionnelle des gènes sur un plan défini par les deux premières composantes principales (PC1 et PC2), montrant les corrélations entre les différents gènes et leur contribution à la variabilité des données.

b. ACP ciblée sur gènes sélectionnés : Suite à l'identification des gènes les plus discriminants dans l'analyse initiale, une seconde ACP a été réalisée en focalisant l'analyse sur quatre gènes d'intérêt (APOE, HLA-DRB1, GATA3 et MET), permettant une meilleure discrimination des profils d'expression entre patients.

2.1 Critères de sélection des gènes : Sélection des gènes d'intérêt pour l'analyse ciblée : Suite à l'ACP globale, quatre gènes ont été sélectionnés pour une analyse plus approfondie en fonction de :

- Leur position discriminante sur le graphique d'ACP, indiquant des profils d'expression spécifiques et une forte contribution aux composantes principales
- Leur association connue avec les pathologies étudiées, notamment APOE pour l'Alzheimer, HLA-DRB1 et GATA3 pour l'asthme et MET pour le cancer du poumon.
- Leur importance biologique dans les voies de signalisation impliquées dans les pathologies.

2.2 Visualisation et interprétation des résultats : Les résultats de l'ACP ont été visualisés sous deux formes principales :

a. Graphique de distribution des gènes : Représentation des 24 gènes sur le plan des deux premières composantes principales, permettant d'identifier les gènes qui contribuent le plus à la variabilité des données et leurs relations.

b. Graphique de distribution des patients : Représentation des 1000 patients sur un plan défini par les composantes principales associées aux quatre gènes d'intérêt.

L'analyse a également révélé que la majorité des patients sont regroupés autour de la zone (0,0) (population homogène), indiquant des profils d'expression similaires, avec quelques individus présentant des profils génétiques rares en dehors de ce groupe central.

3. Analyse par carte thermique (Heatmap) : Suite à l'Analyse en Composantes Principales, une analyse complémentaire par carte thermique (heatmap) a été réalisée à l'aide de l'outil en ligne Heatmapper (<http://www.heatmapper.ca/expression/>). Cette approche a permis de :

a. Transformer les données numériques complexes en une représentation visuelle facile à interpréter (figure 8) .

b. Visualiser les profils d'expression des 24 gènes simultanément pour l'ensemble des échantillons.

c. Détecter les sous-groupes de patients partageant des signatures moléculaires communes.

Cette méthode a permis de confirmer et de préciser visuellement les profils d'expression identifiés par l'ACP, notamment en mettant en évidence les relations entre les gènes associés aux différentes pathologies étudiées.

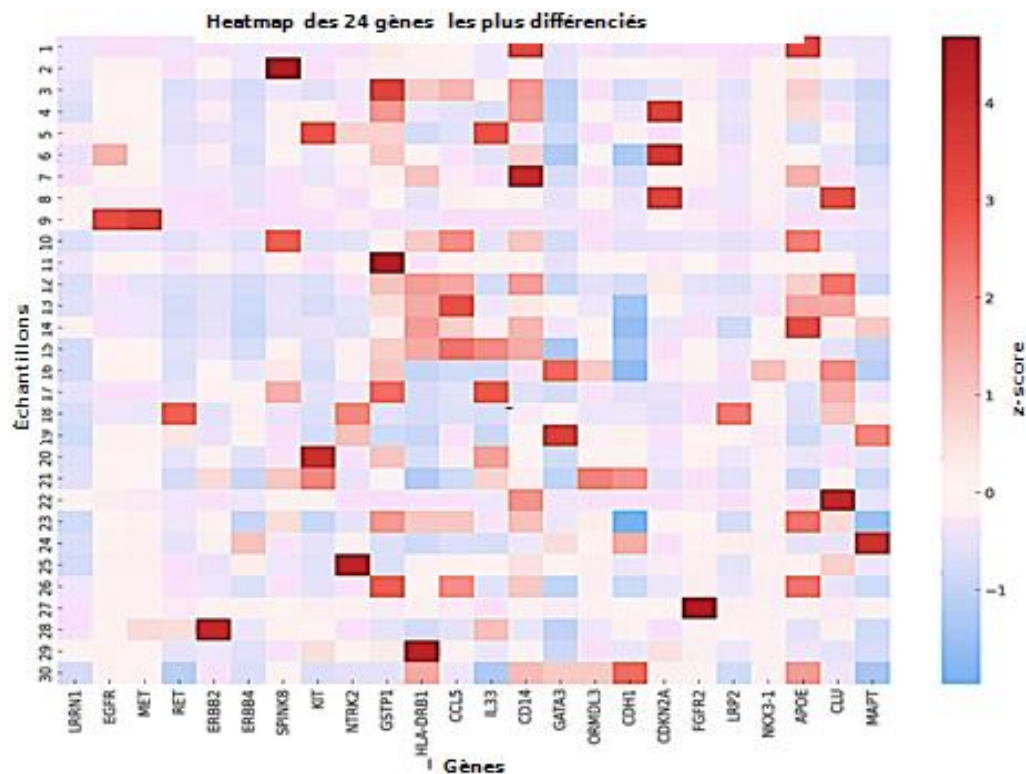


Figure 7:carte thermique représentative des niveaux d'expression génique.

L'intensité de la couleur de la barre de l'échelle représente les valeurs de Z-score normalisées par gène :

- Rouge vif : expression élevée (Z-score positif élevé).
- Bleu vif : expression faible (Z-score négatif).
- Intermédiaire (blanc à rose) : expression moyenne.

HLA-DRB1, CCL5 et CD14 montrent une grande variabilité d'expression selon les individus (fortes traces de couleur).

Ce résultat peut refléter une régulation conditionnelle en fonction du statut tabagique (fumeurs vs non-fumeurs).

4. Analyse différentielle : L'effectif total est partagé en deux sous-groupes de fumeurs et non-fumeurs pour mettre en exergue l'association des gènes impliqués dans le cancer du poumon et par comparaison, les autres types de cancer : Cancer du sein et cancer de la prostate ; d'une part, et une comparaison avec l'expression de gènes liés à deux maladies génétiques : Alzheimer et Asthme.

Cette analyse a été réalisée en transformant les données de notre matrice pour chaque gène :

- i. Calcul du fold change entre les deux groupes, puis faire le $\log_2(\text{fold change})$: $\log_2(\text{moyenne_fumeurs} / \text{moyenne_nonfumeurs})$
- ii. Test statistique (test t de Student), puis transformer la p-value en : $-\log_{10}(\text{p-value})$. Les différentes p-values sont données par l'expression de programmation automatique : $T.TEST(\text{plage_fumeurs}; \text{plage_non_fumeurs}; 2; 3)$; où : 2=test bilatéral et 3=variances non supposées égales. Toutes les p-values ont été transformées par l'équation : $-\log_{10}(\text{p-value})$.

Ces deux nouvelles données (Tableau N°2) ont permis de tracer un graphe volcano pour distinguer les gènes surexprimés, moins exprimés et normaux ; par rapport à une valeur seuil de 1,3 calculée à partir de $-\log_{10}(0,05)$.

Tableau 2: Valeurs $\log_2$ du Fold Change (FC) et de $\log_{10}$ pValues pour les 24 gènes analysés.

Gènes	$\log_2\text{FC}$	$\log_{10}\text{pValues}$	Gènes	$\log_2\text{FC}$	$\log_{10}\text{pValues}$
LRRN1	-0,006	0,757	IL33	-0,145	3,981
EGFR	0,020	0,770	CD14	10,279	0,264
MET	-0,003	1,018	GATA3	-2,017	2,227
RET	-1,488	3,967	ORMDL3	0,426	0,147
ERBB2	0,203	2,219	CDH1	0,140	1,429
ERBB4	-0,340	4,161	CDKN2A	-0,219	0,041
SPINK8	-0,083	2,220	FGFR2	-0,020	0,614
KIT	-0,025	0,201	LRP2	-0,466	5,506
NTRK2	-0,013	2,438	NKX3-1	-0,039	0,339
GSTP1	1,874	0,139	APOE	9,262	0,005
HLA-DRB1	1,464	4,540	CLU	-1,833	1,694
CCL5	0,568	2,846	MAPT	-2,369	2,595

5. Analyse bioinformatique d'enrichissement fonctionnel

Afin d'explorer les fonctions biologiques associées aux gènes d'intérêt, une analyse d'enrichissement fonctionnel a été réalisée à l'aide de l'outil en ligne WebGestalt (WEB-based Gene Set Analysis Toolkit). Cette plateforme permet d'identifier les processus biologiques, les voies de signalisation et les catégories fonctionnelles dans lesquelles les gènes sont sur-représentés par rapport à l'ensemble du génome humain de référence.

La base de données Gene Ontology (GO) a été utilisée, en se concentrant sur les processus biologiques (*Biological Process*). L'enrichissement a été évalué à l'aide du test d'hypergéométrie, et les valeurs de signification ajustées ont été calculées selon la méthode de correction de FDR (False Discovery Rate), avec un seuil de significativité fixé à $FDR \leq 0,05$. Les résultats les plus représentatifs ont été visualisés sous forme d'histogrammes, diagrammes de Venn et tableaux descriptifs.

Le test d'hypergéométrie est une méthode statistique utilisée pour déterminer si le nombre de gènes de notre liste (24 gènes) sont enrichis dans les fonctions biologiques fondamentales telles que la réponse immunitaire. La valeur de probabilité est donnée par l'expression : $p(X = k) =$

$$\sum_{i=9}^{24} \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

Avec : N= 20000 gènes annotés dans le génome humain référence ; K=292 gènes de la base de données GO ; n=24 (notre liste de gènes) et k = 9 gènes dans l'intersection entre notre liste et celle de GO overlap).

PARTIE 3 : RÉSULTATS & DISCUSSION

1. ANALYSE EN COMPOSANTES PRINCIPALES

Dans un premier temps, nous avons évalué la structure de nos données à l'aide d'une ACP.

La matrice de corrélations (Tableau 3) a permis de vérifier l'adéquation des variables (les 24 gènes) et de mettre en évidence les relations linéaires fortes entre certaines d'entre elles. Le graphe biplot (figure 8) illustre graphiquement l'influence de chaque variable sur les deux premiers axes factoriels : les vecteurs longs et proches du cercle unitaire indiquent une contribution importante et une bonne qualité de représentation. De plus, ce biplot positionne chaque échantillon (patients) dans l'espace des deux premières composantes, révélant d'éventuelles structures de regroupement ou de dispersion surtout en fonction du statut tabagique. Ces trois éléments combinés fournissent une vue d'ensemble rigoureuse de la variabilité multivariée et guident l'interprétation ultérieure des axes principaux.

	AP OE	GAT A3	HLA- DRB1	M ET
APOE				
GATA3	- 0,147			
HLA- DRB1	0,46 7	- 0,211		
MET	- 0,049	- 0,127	-0,049	

Tableau 3: une portion de la matrice de corrélations .

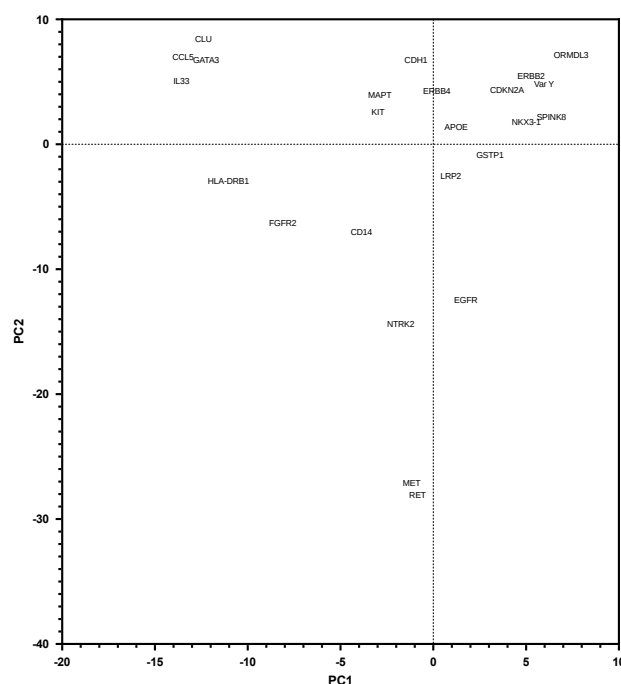


Figure 7: Exploration des structures multivariées entre patients et gènes par ACP (24 gènes, 1000 patients).

Des corrélations positives ($r > 0,5$) ont été observées principalement entre les deux gènes EGFR et MET (gènes associés au cancer du poumon) avec un coefficient $r = +0,899$. Cette forte corrélation basée sur les valeurs d'expression génique Affymetrix suggère une co-régulation marquée ou une interaction fonctionnelle entre ces deux gènes dans les échantillons analysés, tous deux impliqués dans le cancer du poumon.

La forte corrélation ($r = 0,899$) observée entre les gènes *EGFR* et *MET* dans les profils d'expression génique suggère une interaction fonctionnelle ou une régulation conjointe, cohérente avec les mécanismes moléculaires décrits dans les cancers du poumon non à petites cellules. Ces résultats rejoignent les observations de Engelman et al. (2007) et Bean et al. (2007), qui ont mis en évidence la coactivation EGFR/MET comme facteur de progression tumorale et de résistance thérapeutique. En effet, Engelman et al. ont montré que l'activation de MET peut permettre aux cellules tumorales de bypasser l'inhibition d'EGFR dans le NSCLC, confirmant une co-amplification des locus contenant *EGFR* et *MET* dans certains sous-types de NSCLC. De leur côté, Bean et al. (2007) ont décrit la coactivation de EGFR et MET comme mécanisme de progression tumorale et de résistance.

Une corrélation de type négatif a été enregistrée entre les deux gènes associés à l'asthme, GATA3 et GSTP1 où $r = -0,357$, légèrement inférieure à la valeur seuil égale à 0,396 de Pearson pour un ddl de 23 (ddl = $n-1=24-1=23$).

En effet, la corrélation négative modérée ($r = -0,357$) observée entre les gènes GSTP1 et GATA3, tous deux associés à l'asthme, suggère une relation inverse dans leur expression génique. Cette observation peut refléter des mécanismes biologiques distincts ou opposés dans la physiopathologie de l'asthme. La corrélation négative entre ces deux gènes pourrait indiquer que, dans certains contextes, une augmentation de l'activité inflammatoire médiée par GATA3 est associée à une diminution de la capacité antioxydante représentée par GSTP1, ou *vice versa*. Certains auteurs ont confirmé l'augmentation de l'expression de GATA3 dans les voies respiratoires des patients asthmatiques, corrélant avec une inflammation accrue et une hyperréactivité bronchique (Nakamura, 1999).

Le biplot de l'ACP obtenu pour les 24 gènes et 1000 patients offre une vue synthétique de la structure des données, en projetant à la fois les individus (patients) et les variables (gènes) dans un plan défini par les deux premières composantes principales, PC1 et PC2.

La composante principale PC1 semble opposer une expression élevée de GATA3 à une expression élevée de HLA-DRB1. Ce contraste pourrait refléter deux profils immunologiques opposés : PC1 sépare donc potentiellement les patients en fonction de leur type de réponse immunitaire dominante. Le deuxième axe PC2 semble distinguer un profil neurodégénératif (Neuro-inflammation, APOE↑) d'un profil oncogénique (Prolifération tumorale, MET↑).

Les résultats de l'étude de Nakamura et al. (1999) apportent la première preuve d'une augmentation de l'expression du gène GATA-3 en association avec les cellules ARNm IL-5+ dans les voies respiratoires des asthmatiques. Ces résultats étayent une relation causale entre l'augmentation de l'expression du gène GATA-3 et la dysrégulation de l'expression de l'IL-5 dans l'asthme atopique, lequel avec DRB1 peuvent être impliqués dans des mécanismes immunitaires communs, en particulier dans des contextes inflammatoires, auto-immuns, ou allergiques ; les deux gènes participant à la réponse immunitaire adaptative du fait que HLA-DRB1 active les lymphocytes T pour influencer indirectement la production d'IL-5.

2. ANALYSE DIFFÉRENCIELLE

Les résultats obtenus pour les deux classes de patients et pour les types de maladies génétiques ont permis de construire le graphe suivant (figure 9):

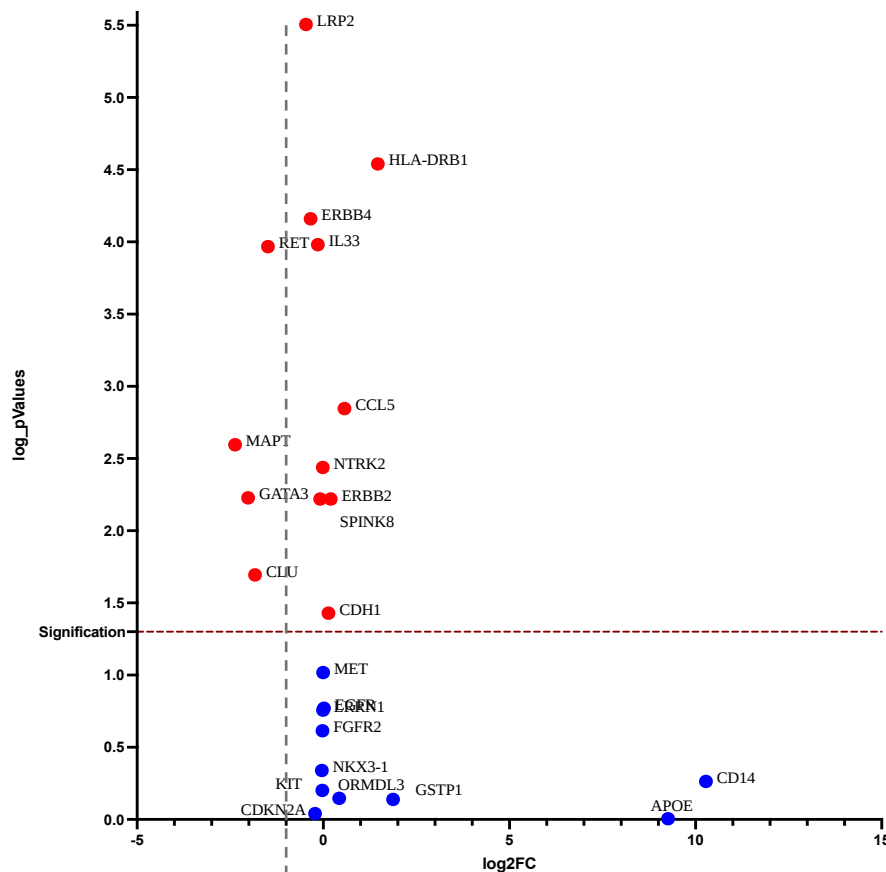


Figure 8 : Analyse différentielle de l'expression de 24 gènes liés à l'Asthme, au cancer et à l'Alzheimer : représentation Volcano Plot.

L'axe des abscisses x ($\log_2(FC)$) représente l'intensité du changement d'expression génique. Les valeurs négatives sont synonyme d'une sous-expression, alors que celles positives expriment une sur-expression.

Quant à l'axe y ($-\log_{10}(p\text{-value})$), il représente la significativité statistique. Plus sa valeur est grande, plus le résultat est significatif. Le seuil de significativité statistique est $y=1,3$ correspondant à $p=0,05$ (risque d'erreur α). Un gène est dit statistiquement significatif si sa p-value est inférieure à un seuil fixé (souvent 0,05).

Le gène CD14 montre une grande sur-expression, mais une statistique non significative ($0,264 < 1,3$). Bien que ce gène semble très fortement sur-exprimé, cette variation n'est pas statistiquement significative, à cause d'une grande variabilité ou à cause de la nature de l'effectif (fumeurs et non-fumeurs).

Le gène LRP2 montre une légère sous-expression avec un $\log_2(\text{FC}) = -0,466$, cependant sa valeur y est 5,506 qui est très significative ($p=10^{5,5}=2 \times 10^{-6}$). Le changement d'expression entre les deux classes (fumeurs et non-fumeurs) est modeste mais fiable.

Le gène MET ne montre aucune variation réelle entre les deux classes ($\log_2\text{FC}=-0,003$) et est non significatif car il présente une significativité inférieure à la valeur seuil ($1,018 < 1,3$; $p=0,096$). Ce gène n'est donc pas différentiellement exprimé entre les deux classes de patients. Le tableau 4 résume les analyses différentielles entre les deux classes de patients.

Tableau 4 :

Tableau 4 : les analyses différentielles entre les deux classes de patients.

Gène	$\log_2 \text{FC}$	$\log_{10} (\text{p-value})$	Significatif ?	Interprétation
CD14	+10,28	0,264	NON	Forte sur-expression, mais non fiable
LRP2	-0,466	5,506	OUI	Sous-expression légère, très significative
MET	-0,003	1,018	NON	Pas de variation, non significatif

Nous pouvons, compte tenu du résultat du volcano plot, classer les gènes analysés sur les deux classes de patients en trois catégories (tableau 5) :

Tableau 5 : Analyse de l'expression différentielle des gènes selon le statut de significativité.

Gènes	$\log_2\text{FC}$	$\log_{-p}\text{Values}$	Gènes	$\log_2\text{FC}$	$\log_{-p}\text{Values}$
LRRN1	-0,006	0,757	IL33	-0,145	3,981
EGFR	0,020	0,770	CD14	10,279	0,264
MET	-0,003	1,018	GATA3	-2,017	2,227
RET	-1,488	3,967	ORMDL3	0,426	0,147
ERBB2	0,203	2,219	CDH1	0,140	1,429
ERBB4	-0,340	4,161	CDKN2A	-0,219	0,041
SPINK8	-0,083	2,220	FGFR2	-0,020	0,614
KIT	-0,025	0,201	LRP2	-0,466	5,506
NTRK2	-0,013	2,438	NKX3-1	-0,039	0,339
GSTP1	1,874	0,139	APOE	9,262	0,005
HLA-DRB1	1,464	4,540	CLU	-1,833	1,694
CCL5	0,568	2,846	MAPT	-2,369	2,595

- i. Gènes significatifs et sur-exprimés (n=4) : HLA-DRB1, CCL5, ERBB2 et NTRK2.
- ii. Gènes significatifs et sous-exprimés (n=7) : ERBB4, RET, GATA3, CLU, MAPT, IL33 et LRP2
- iii. Gènes non significatifs (n=13) : LRRN1, SPINK8, GSTP1, ORMDL3, CDH1, FGFR2, NKX3-1, CD14, APOE, EGFR, MET, KIT, CDKN2A.

Le gène HLA-DRB1 est significativement sur-exprimé dans la condition testée (fumeurs ; non-fumeurs). Cela peut indiquer une activation ou une implication probable de HLA-DRB1 dans le processus pathologique étudié (Asthme). Ce gène, impliqué dans la présentation des gènes aux cellules immunitaires, pourrait jouer un rôle immunologique actif, notamment dans les maladies inflammatoires.

Le gène GATA-3 est significativement moins exprimé chez les fumeurs que chez les non-fumeurs. C'est un facteur de transcription clé pour les cellules T de type Th2, qui participent à la réponse immunitaire de type allergique (asthme, production de cytokines IL-4, IL-5 et IL-13). Dans son travail sur l'expression de GATA-3, Bacha et al. (2024) indiquent que l'expression de ce gène est souvent plus élevée dans les états allergiques ou inflammatoires chroniques à dominante Th2, comme l'asthme non lié au tabac. Le tabac modifie profondément la réponse immunitaire, avec un déséquilibre entre les réponses Th1 et Th2. Une réduction de GATA-3 chez les fumeurs pourrait refléter une suppression de la réponse Th2 ou une réorientation vers des réponses Th1 ou encore une altération de l'immunité muqueuse pulmonaire, fréquente chez les fumeurs.

L'analyse de l'enrichissement des 24 gènes a révélé une probabilité $p = 1,4914 \cdot 10^{-10}$ et un $FDR = 1,2512 \cdot 10^{-7}$. Le test indique qu'il y a neuf (n=9) gènes sur les 24 de notre liste (figure 10) qui appartiennent à une catégorie contenant 292 gènes de la base de données GO. Ce test d'hypergéométrie montre également que notre liste de 24 gènes est très significativement enrichie en gènes appartenant à cette catégorie de 292 gènes recensés dans GO particulièrement la catégorie de la voie biologique 'phosphatidylinositol 3-kinase' (figure 11). Cet enrichissement est statistiquement très peu probable de se produire par hasard car $p = 1,4914 \cdot 10^{-10}$, et il reste significatif même après correction ($FDR < 0,05$).

L'implications des gènes dans les différents processus biologiques tels que la régulation ou la communication cellulaire est rapportée sur la (figure 12) où nous notons un total de 22 gènes sur 24 qui s'expriment dans les fonctions de régulations biologiques et huit (n=8) dans la croissance cellulaire.

Les fonctions biologiques moléculaires auxquelles nos gènes d'intérêt ont été affectés sont particulièrement la fonction de liaisons protéiques (22 gènes sur 24) et ioniques (12 gènes sur 24),

activité de transducteur moléculaire (10 gènes sur 24), c'est-à-dire la capacité de la molécule à convertir un signal d'une forme en une autre à l'intérieur de la cellule ; comme les récepteurs membranaires couplés aux protéines recevant des signaux hormonaux pour activer des voies de signalisation intracellulaire.

La fonction de liaison aux protéines est la fonction la plus fréquente au sein des 24 gènes analysés, impliquant 22 gènes ; ce qui signifie que la majorité de nos gènes codent des produits qui interagissent directement avec d'autres protéines.

Cependant, la liaison aux ions indique un rôle potentiel dans la régulation ionique ou les voies de signalisation dépendantes des ions Ca^{++} et Zn^{++} principalement.

Ces résultats suggèrent que les gènes identifiés sont impliqués dans des fonctions moléculaires centrales telles que la liaison aux protéines, la transduction de signal ou l'activité de liaison aux ions, caractéristiques de protéines jouant un rôle clé dans la signalisation cellulaire et la régulation des processus biologiques complexes tels que cela a bien été noté dans les travaux de Ashburner et al. (2000).

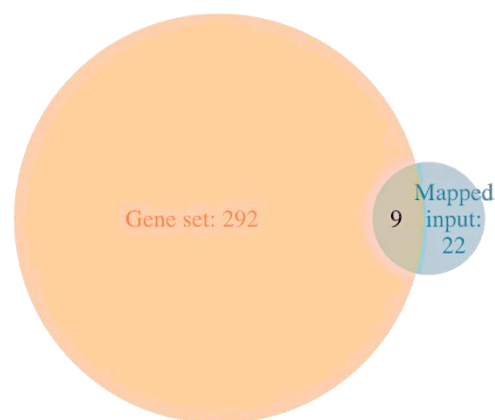


Figure 9 : Diagramme de Venn illustrant l'analyse comparative des gènes impliqués dans les processus biologiques enrichis.

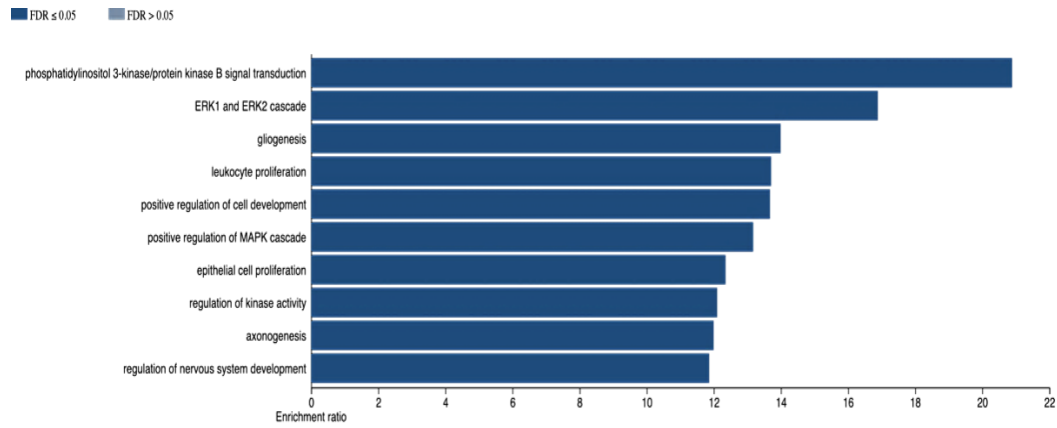


Figure 10 : Visualisation des catégories GO (Biological Process) enrichies pour les gènes différentiellement exprimés

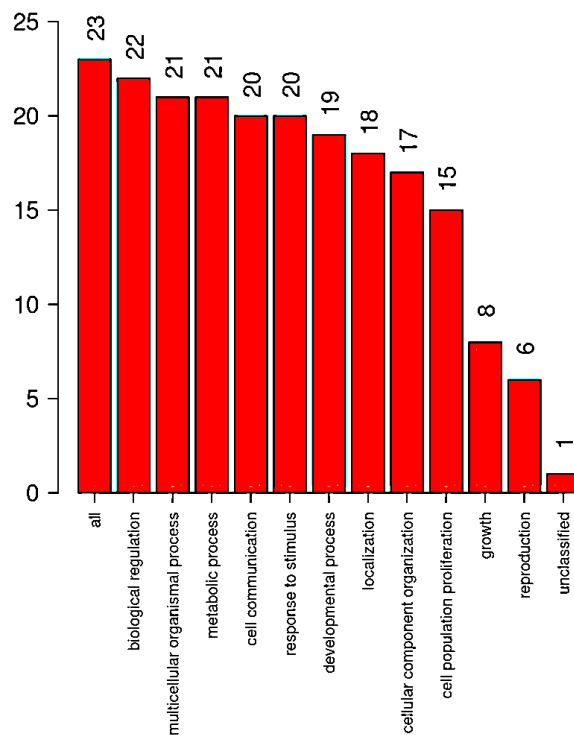


Figure 11 : Implications des gènes dans les processus biologiques.

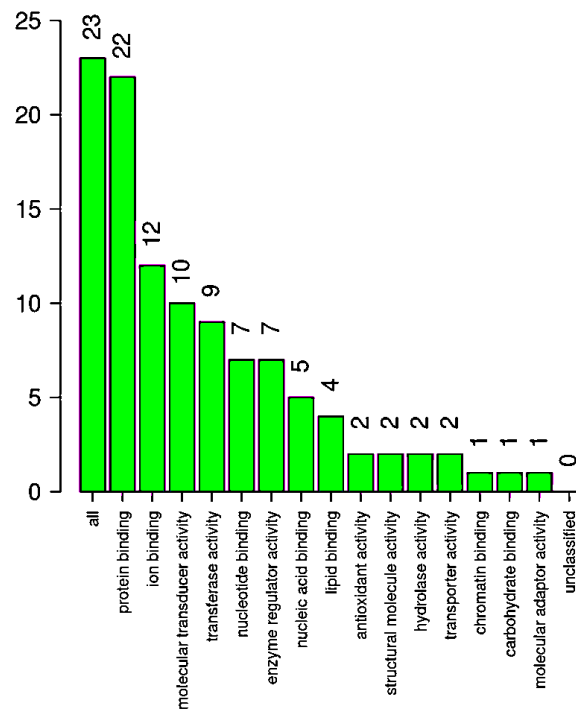


Figure 12 : Implication des gènes dans les catégories fonctionnelles moléculaires.

CONCLUSION

Ce travail de recherche s'est inscrit dans l'objectif de mieux comprendre l'expression différentielle et les relations d'interdépendance entre plusieurs gènes impliqués dans diverses pathologies humaines majeures : les cancers (poumon, prostate, sein), l'asthme, et la maladie d'Alzheimer. À partir d'un jeu de données ciblant 24 gènes connus pour leur implication dans ces maladies, différentes analyses statistiques ont été menées, notamment une analyse en composantes principales (ACP), une analyse de corrélation et une analyse différentielle de l'expression génique (volcano plot) selon le statut tabagique (fumeurs vs non-fumeurs).

L'ACP a révélé que les deux premières composantes principales (CP1 et CP2) capturent l'essentiel de la variabilité des profils d'expression génique. La CP1 a opposé de manière marquée GATA3 (fortement positif), un facteur de transcription impliqué dans les réponses immunitaires Th2 et associé à l'asthme, à HLA-DRB1 (négatif), gène du complexe majeur d'histocompatibilité (CMH) de classe II impliqué dans la présentation des antigènes. Cette opposition pourrait refléter un clivage immunitaire entre réponses allergiques et inflammation chronique de type auto-immune ou tumorale. La CP2, quant à elle, a mis en évidence

une opposition entre APOE, fortement impliqué dans la maladie d'Alzheimer, et MET, proto-oncogène largement étudié dans le cancer du poumon, suggérant une différenciation entre mécanismes neurodégénératifs et oncogéniques.

La matrice de corrélations a mis en évidence des relations géniques notables : une corrélation positive très forte entre EGFR et MET ($r = +0,899$), deux gènes codant pour des récepteurs à activité tyrosine kinase souvent coactivés dans les tumeurs pulmonaires, et une corrélation négative modérée entre GSTP1 et GATA3 ($r = -0,357$), suggérant une possible opposition fonctionnelle ou une régulation croisée dans des contextes inflammatoires ou tumoraux.

L'analyse différentielle des gènes selon le statut tabagique a souligné des profils d'expression contrastés. Le gène HLA-DRB1 est significativement surexprimé chez les fumeurs, ce qui pourrait refléter une stimulation immunitaire chronique induite par l'exposition au tabac. En revanche, GATA3 est significativement sous-exprimé chez les fumeurs, indiquant une altération des réponses Th2 dans ce groupe, cohérente avec une immunité orientée différemment ou un effet immunosuppresseur du tabac. Enfin, le gène MET, bien qu'impliqué dans de nombreux cancers, ne montre ici aucune différence significative entre fumeurs et non-fumeurs, ce qui

suggère que son activation pourrait dépendre d'autres mécanismes que l'expression génique, tels que des mutations ou des modifications post-traductionnelles.

En conclusion, ce travail met en lumière la complexité des interactions géniques en lien avec diverses pathologies humaines. Il souligne l'importance de l'environnement (comme le tabagisme) dans la modulation de l'expression de certains gènes clés, et fournit des pistes pour de futures investigations ciblées sur les voies moléculaires associées à ces maladies. Ces résultats, bien que fondés sur un nombre limité de gènes, constituent une base solide pour des analyses plus larges en transcriptomique ou en médecine personnalisée.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1]- Alcimed. (2024, février 9). *Séquençage long-read, étude des cancers et génomes viraux*. Alcimed .
- [2]- Awad, M. M., Oxnard, G. R., Jackman, D. M., Savic, L. J., Blumenberg, C., Kirschmeier, P., Schrock, A. B., & Jänne, P. A. (2021). MET exon 14 skipping alterations in non-small-cell lung cancer: A review of the literature. *JCO Precision Oncology*, 5, 1–9. <https://doi.org/10.1200/PO.20.00470>
- [3]- Bacha, R., Alwisi, N., Ismail, R., Pedersen, S., & Al-Mansouri, L. (2024). Unveiling GATA3 signaling pathways and disease: Mechanisms, implications and therapeutic potential. *Cells*, 13(24), 2127. <https://doi.org/10.3390/cells13242127>
- [4]- Bean, J., Brennan, C., Shih, J. Y., & Pao, W. (2007). MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. *Proceedings of the National Academy of Sciences*, 104(52), 20932–20937. <https://doi.org/10.1073/pnas.0710370104>
- [5]- Belouadi, M., Hazam, K., & Khaldi, A. (2022). *Étude in silico d'une mutation : Application à la mutation fonctionnelle G719X du gène EGFR retrouvée dans les cancers bronchiques non à petites cellules* [Mémoire de master en biochimie]. Université Aïn Témouchent - Belhadj Bouchaib.
- [6]- Benchimol. (2021, 3 septembre). *Gènes BRCA I et BRCA II. Docteur Benchimol : Gynécologue-obstétricien à Paris, France*. <https://www.docteur-benchimol.com/genes-brca-i-et-brca-ii.html>
- [7]- Bergethon, K., Shaw, A. T., Ou, S. H. I., Katayama, R., Lovly, C. M., McDonald, N. T., ... & Iafrate, A. J. (2012). ROS1 rearrangements define a unique molecular class of lung cancers. *Journal of Clinical Oncology*, 30(8), 863–870. <https://doi.org/10.1200/JCO.2011.35.6345>
- [8]- Boulet, L. (2013). *Physiopathologie respiratoire appliquée*. Presses de l'Université Laval.
- [9]- Brandt-Rauf, P. W., & Santella, R. M. (1993). Mutations of the p53 tumor suppressor gene and their relationship to cancer. *Annals of the New York Academy of Sciences*, 686(1), 13–25. <https://doi.org/10.1111/j.1749-6632.1993.tb39135.x>
- [10]- Cardarella, S., & Johnson, B. E. (2013). The impact of BRAF mutations on treatment of non-small cell lung cancer. *Lung Cancer*, 80(1), 1–3. <https://doi.org/10.1016/j.lungcan.2012.12.003>

- [11]- Chenau, J. (2009, 24 juin). *Le secretome, un modèle adapté à l'étude des protéines sécrétées par les tumeurs : Application à l'analyse du rôle joué par p53 sur la sécrétion de protéines par des cellules du cancer du poumon.*
- [12]- Chipoulet, E., Collet, G., & Couderc, B. (2023). Rôle des médecins dans l'adhésion des patients et des familles à la réalisation d'analyses de caractéristiques génétiques [The role of physicians in patient and family adherence to genetic testing]. *Bulletin du Cancer*, **110**(10), 1002–1014. <https://doi.org/10.1016/j.bulcan.2023.07.008>
- [13]- Drilon, A., Rekhtman, N., Ladanyi, M., & Paik, P. K. (2014). Squamous-cell carcinomas of the lung: Emerging biology, controversies, and the promise of targeted therapy. *The Lancet Oncology*, **15**(10), e418–e426. [https://doi.org/10.1016/S1470-2045\(14\)70301-8](https://doi.org/10.1016/S1470-2045(14)70301-8)
- [14]- Edimark. (2021). *De l'identification des gènes BRCA1 et BRCA2 au développement des inhibiteurs de PARP.* <https://www.edimark.fr/revues/la-lettre-du-cancerologue/n-2-fevrier-2021/de-lidentification-des-genes-brca1etbrca2-au-developpement-desinhibiteurs-de-par>
- [15]- Eliahiai, I., Eljiar, M., Chaib, S., Kharmoum, J., & Chraibi, M. (2023). Récepteur du facteur de croissance épidermique HER2, tests utilisés pour rechercher son amplification dans le cancer du sein : principes et limites. *Bulletin du Cancer*, **110**(12), 1301–1310. <https://doi.org/10.1016/j.bulcan.2023.09.004>
- [16]- Engelman, J. A., Zejnullahu, K., Mitsudomi, T., Song, Y., Hyland, C., Park, J. O., & Lindeman, N. (2007). MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science*, **316**, 1039–1043. <https://doi.org/10.1126/science.1141478>
- [17]- Ferrer, I., Zugazagoitia, J., Herbertz, S., John, W., Paz-Ares, L., & Schmid-Bindert, G. (2018). KRAS-mutant non-small cell lung cancer: From biology to therapy. *Lung Cancer*, **124**, 53–64. <https://doi.org/10.1016/j.lungcan.2018.07.013>
- [18]- Fervers, B., Clin, B., & Pairon, J.-C. (2023). Epidemiology of lung cancer in France. *Revue des Maladies Respiratoires Actualités*, **15**(2), 255–259.
- [19]- Gainor, J. F., Shaw, A. T., Sequist, L. V., Fu, X., Azzoli, C. G., Piotrowska, Z., ... & Engelman, J. A. (2013). EGFR mutations and ALK rearrangements in NSCLC are associated with low response rates to PD-1 pathway blockade. *Journal of Thoracic Oncology*, **11**(2), 171–181.
- [20]- George, J., Lim, J. S., Jang, S. J., Cun, Y., Ozretić, L., Kong, G., ... & Helland, Å. (2015). Comprehensive genomic profiles of small cell lung cancer. *Nature*, **524**(7563), 47–53.

- [21]- Guo, R., Offin, M., Brannon, A. R., Chang, J., Chow, A., Delasos, L., Girshman, J., Wilkins, O., McCarthy, C. G., Makhnin, A., Falcon, C., Scott, K., Tian, Y., Cecchi, F., Hembrough, T., Alex, D., Shen, R., Benayed, R., Li, B. T., Rudin, C. M., Kris, M. G., Arcila, M. E., Rekhtman, N., Paik, P., Zehir, A., & Drilon, A. (2021). MET exon 14-altered lung cancers and MET inhibitor resistance. *Clinical Cancer Research*, 27(3), 799–806.
- [22]- <https://www.sante.fr/cancer-du-poumon-points-cles>
- [23]- Jacob, A. (2020, 22 septembre). *Apport du séquençage d'exome pour le diagnostic et le traitement des maladies complexes*.
- [24]- Khatri P., Draghici S. Ontological analysis of gene expression data: current tools, limitations, and open problems. 2005. *Bioinformatics*. 21:3587–95. .
- [25]- Knudsen S. 2004. Guide to analysis of DNA microarray data. John Wiley & Sons, Inc.; Hoboken, NJ.
- [26]- Koyama, S., Akbay, E. A., Li, Y. Y., Aref, A. R., Skoulidis, F., Herter-Sprie, G. S., Buczkowski, K. A., Liu, Y., Awad, M. M., Denning, W. L., Diao, L., Wang, J., Parra-Cuentas, E. R., Wistuba, I. I., Soucheray, M., Thai, T., Asahina, H., Kitajima, S., Altabef, A., Cavanaugh, J. D., Rhee, K., Gao, P., Zhang, H., Fecci, P. E., Shimamura, T., Hellmann, M. D., Heymach, J. V., Hodi, F. S., Freeman, G. J., Barbie, D. A., Dranoff, G., Hammerman, P. S., & Wong, K. K. (2016). STK11/LKB1 deficiency promotes neutrophil recruitment and proinflammatory cytokine production to suppress T-cell activity in the lung tumor microenvironment. *Cancer Research*, 76(5), 999–1008.
- [27]- Lee E. Y. H. P., Muller W. J. 2010. Oncogenes and Tumor Suppressor Genes. *Cold Spring Harb Perspect Biol*. 2 : a003236.
- [28]- Li, B. T., Michelini, A., Misale, S., Cocco, E., Baldino, L., Cai, Y., ... & Kris, M. G. (2020). HER2-mediated internalization of cytotoxic agents in ERBB2-amplified or mutant lung cancers. *Cancer Discovery*, 10(5), 674–687.
- [29]- Li, B. T., Ross, D. S., Aisner, D. L., Chaft, J. E., Hsu, M., Kako, S. L., ... & Kris, M. G. (2016). HER2 amplification and HER2 mutation are distinct molecular targets in lung cancers. *Journal of Thoracic Oncology*, 11(3), 414–419.
- [30]- Li, K., Bosdet, I., Yip, S., Ho, C., Laskin, J., Melosky, B., Wang, Y., & Sun, S. (2023). Real-World Clinical Outcomes for Patients with EGFR and HER2 Exon 20 Insertion-Mutated Non-Small-Cell Lung Cancer. *Current Oncology*, 30(8), 7099–7111.

- [31]- Lin, J. J., Shaw, A. T., & Ou, S. H. I. (2017). ROS1 fusions in non-small-cell lung cancer. *Journal of Clinical Oncology*, 35(9), 969–977.
- [32]- Liu, T. L., Chen, X., & Zhang, Y. (2004). Mutations of p53 gene in lung cancer and their clinical significance. *Oncology Reports*, 11(2), 465-469.
- [33]- Lu, S., Fang, J., Cao, L., Li, X., Guo, Q., Zhou, J., Wang, Z., Zhang, H., Wang, D., Wang, Y., Zhang, Y., Xu, J., Chen, G., Wang, J., Zhang, L., Wang, M., Yang, J., Zhang, L., Zhou, C., Cheng, Y., Wang, J., Wang, Q., Wu, Y., & Wu, Y.-L. (2019). Preliminary efficacy and safety results of savolitinib treating patients with pulmonary sarcomatoid carcinoma and other types of non-small cell lung cancer harboring MET exon 14 skipping mutations. *Cancer Research*, 79(13 Suppl), CT031.
- [34]- M. Ducreux, P. Amiel, Accès aux tests génétiques en oncologie, Bulletin de l'Académie Nationale de Médecine, Volume 206, Issue 3, 2022,
- [35]- Marchetti, A., Di Lorito, A., Pace, M. V., Iezzi, M., Felicioni, L., D'Antuono, T., Filice, G., Guetti, L., Mucilli, F., & Buttitta, F. (2016). ALK Protein Analysis by IHC Staining after Recent Regulatory Changes: A Comparison of Two Widely Used Approaches, Revision of the Literature, and a New Testing Algorithm. *Journal of Thoracic Oncology*, 11(4), 487–495.
- [36]- Meng, F., Zhang, S., & Song, X. (2024). PTEN loss in lung cancer: Impact on tumor microenvironment and therapeutic strategies. *Nature Reviews Cancer*, 24(1), 45-59.
- [37]- Muller, P. A. J., Vousden, K. H., & Lane, D. P. (2006). P53 mutations in cancer. *Nature Reviews Cancer*, 7(10), 837-846.
- [38]- Nakamura Y., Ghaffar O., Olivenstein R., Taha R. A., Soussi-Gounni A., Zhang D. H. et al. 1999. Gene expression of the GATA-3 transcription factor is increased in atopic asthma. *J. Allergy Clin. Immunol.* 103 (2 Pt1):215-22. DOI:10.1016/s0091-6749(99)70493-8.
- [39]- Ockey, C. H., Vaezi, A., Feldman, R., & Vaezi, A. (2011). ERCC1 and XRCC1 as biomarkers for lung and head and neck cancer. *DNA Repair*, 10(7), 712–719.
- [40]- Olaussen, K. A., Dunant, A., Fouret, P., Brambilla, E., André, F., Haddad, V., Taranchon, E., Filipits, M., Pirker, R., Popper, H. H., Stahel, R., Sabatier, L., Pignon, J. P., Tursz, T., Le Chevalier, T., & Soria, J. C.; IALT Bio Investigators. (2006). DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *The New England Journal of Medicine*, 355(10), 983–991. <https://doi.org/10.1056/NEJMoa060570>

- [41]- Olivier, M., Hollstein, M., & Hainaut, P. (2010). TP53 mutations in human cancers: Origins, consequences, and clinical use. *Cold Spring Harbor Perspectives in Biology*, 2(1), a001008.
- [42]- Paik, P. K., Arcila, M. E., Fara, M., Sima, C. S., Miller, V. A., Kris, M. G., & Ladanyi, M. (2011). Clinical characteristics of patients with lung adenocarcinomas harboring BRAF mutations. *Journal of Clinical Oncology*, 29(15), 2046–2051.
- [43]- Paul, A., & Paul, S. (2014). The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers. *Frontiers in Bioscience (Landmark Edition)*, 19(4), 605–618. <https://doi.org/10.2741/4230>
- [44]- Peifer, M., Fernández-Cuesta, L., Sos, M. L., George, J., Seidel, D., Kasper, L. H., ... & Thomas, R. K. (2012). Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nature Genetics*, 44(10), 1104–1110. <https://doi.org/10.1038/ng.2396>
- [45]- Pérez-Ramírez, C., Cañadas-Garre, M., Molina, M. Á., Faus-Dáder, M. J., & Calleja-Hernández, M. Á. (2015). PTEN and PI3K/AKT in non-small-cell lung cancer. *Pharmacogenomics*, 16(16), 1843–1862. <https://doi.org/10.2217/pgs.15.114>
- [46]- Planchard, D., Besse, B., Groen, H. J. M., Souquet, P. J., Quoix, E., Baik, C. S., ... & Mazieres, J. (2016). Dabrafenib plus trametinib in patients with previously treated BRAF(V600E)-mutant metastatic non-small-cell lung cancer: An open-label, multicentre, phase 2 trial. *The Lancet Oncology*, 17(7), 984–993. [https://doi.org/10.1016/S1470-2045\(16\)30146-2](https://doi.org/10.1016/S1470-2045(16)30146-2)
- [47]- Planchard, D., Olaussen, K.-A., & Soria, J.-C. (2010). ERCC1 et cancer pulmonaire. *Revue des Maladies Respiratoires*, 27(4), 387–394.
- [48]- Redouan, M. (2021). Étude du profil moléculaire des patients marocains atteints de cancer bronchique non à petites cellules : mutation EGFR, réarrangement ALK et l'expression de PD-L1 [Mémoire de master, Université Mohammed V].
- [49]- Rolfo, C., Mack, P. C., Scagliotti, G. V., Baas, P., Barlesi, F., Bivona, T. G., Herbst, R. S., Mok, T. S., Peled, N., Pirker, R., Raez, L. E., Reck, M., Riess, J. W., Sequist, L. V., Shepherd, F. A., Sholl, L. M., Tan, D. S., Wakelee, H. A., Wistuba, I. I., ... Gandara, D. R. (2018). Liquid biopsy for advanced non-small cell lung cancer (NSCLC): A statement paper from the IASLC. *Journal of Thoracic Oncology*, 13(9), 1248–1268.
- [50]- Ruppert A-M, Beau-Faller M, Belmont L, Lavolé A, Gounant V, Cadranel J, Wislez M Un regard simple sur la biologie du cancer bronchique : MET. *Revue des Maladies Respiratoires*. 28(10), 2011, e139–e145. doi:10.1016/j.rmr.2011.07.006

- [51]- Schena, M. (2000). *Microarray biochip technology*. Eaton Publishing.
- [52]- Shaw, A. T., Ou, S. H. I., Bang, Y. J., Camidge, D. R., Solomon, B. J., Salgia, R., ... & Jänne, P. A. (2014). Crizotinib in ROS1-rearranged non-small-cell lung cancer. *New England Journal of Medicine*, 371(21), 1963–1971.
- [53]- Sholl, L. M., Do, K., Shivdasani, P., Cerami, E., Dubuc, A. M., Kuo, F. C., Garcia, E. P., Jia, Y., Davineni, P., Abo, R. P., Pugh, T. J., van Hummelen, P., Thorner, A. R., Ducar, M., Berger, A. H., Nishino, M., Janeway, K. A., Church, A., Harris, M., ... Lindeman, N. I. (2016). Institutional implementation of clinical tumor profiling on an unselected cancer population. *JMD: The Journal of Molecular Diagnostics*, 18(5), 683-696
- [54]- Skoulidis, F., Byers, L. A., Diao, L., Papadimitrakopoulou, V. A., Tong, P., Izzo, J., Behrens, C., Kadara, H., Parra, E. R., Canales, J. R., Zhang, J., Giri, U., Gudikote, J., Cortez, M. A., Yang, C., Fan, Y., Peyton, M., Girard, L., Coombes, K. R., ... Heymach, J. V. (2015). Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. *Cancer Discovery*, 5(8), 860-877.
- [55]- Skoulidis, F., Goldberg, M. E., Greenawalt, D. M., Hellmann, M. D., Awad, M. M., Gainor, J. F., Schrock, A. B., Hartmaier, R. J., Trabucco, S. E., Gay, L., Ali, S. M., Elvin, J. A., Singal, G., Ross, J. S., Fabrizio, D., Szabo, P. M., Chang, H., Sasson, A., Srinivasan, S., ... Heymach, J. V. (2018). STK11/LKB1 mutations and PD-1 inhibitor resistance in KRAS-mutant lung adenocarcinoma. *Cancer Discovery*, 8(7), 822–835.
- [56]- Soussi, T. (2007). p53 mutations in human cancer: Database and predictive factors. *Human Mutation*, 28(6), 619–627.
- [57]- Takeuchi, K., Soda, M., Togashi, Y., Suzuki, R., Sakata, S., Hatano, S., ... & Mano, H. (2012). RET, ROS1 and ALK fusions in lung cancer. *Nature Medicine*, 18(3), 378–381.
- [58]- Tarca, A. L., Romero, B., & Draghici, S. (2006). Analysis of microarray experiments of gene expression profiling. *American Journal of Obstetrics and Gynecology*, 195(2), 373–388.
- [59]- Tomasini, P., Zhou, C., Novello, S., Garrido, P., Doms, C., Alatorre-Alexander, J., *et al.* (2025). Qualité de vie liée à la santé (HRQoL) et symptômes chez les patients de l'essai LIBRETTO-431 atteints de cancer bronchique non à petites cellules (CBNPC) avancé avec fusion du gène RET. *Revue des maladies respiratoires Actualités*, 17(1), 242.

- [60]- Vassilev, L. T. (2007). Small-molecule inhibitors of the p53-MDM2 interaction: Reactivating the p53 pathway for cancer therapy. *Clinical Cancer Research*, **13**(2), 567-572.
- [61]- Wang, C., Xu, Z., Zhang, C., Wang, Y., Zhang, Y., Zhang, Y., & Christiani, D. C. (2015). Polymorphisms in ERCC1 and ERCC2/XPD genes and carcinogen susceptibility. *DNA Repair*, **36**, 1–8.
- [62]- Yang, Y. H., Buckley, M. J., Dudoit, S., & Speed, T. P. (2002). Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics*, **11**, 108–136
- [63]- Yang, Y., Zhang, H., Huang, S., & Chu, Q. (2023). KRAS mutations in solid tumors: Characteristics, current therapeutic strategy, and potential treatment exploration. *Journal of Clinical Medicine*, **12**(2), 709. <https://doi.org/10.3390/jcm12020709>
- [64]- Yasuda, H., Park, E., Yun, C. H., Sng, N. J., Lucena-Araujo, A. R., Yeo, W. L., ... & Jänne, P. A. (2012). Structural, biochemical, and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer. *Science Translational Medicine*, **5**(216), 216ra177. <https://doi.org/10.1126/scitranslmed.3005758>
- [65]- Zhang X, Jafari N, Barnes RB, Confino E, Milad M, Kazer RR. 2005. Studies of gene expression in human cumulus cells indicate pentraxin 3 as a possible marker for oocyte quality. *Fertil Steril*. **83**:1169–79. .
- [66]- Zhang, Y., & Yang, N. (2024). *New anti-cancer drug development and evaluation*. Springer Nature.
- [67]- Zhu, C., Chen, Z., Zhao, Y., Gao, Z., & Li, W. (2020). HER2 exon 20 insertion mutations in non-small-cell lung cancer are sensitive to the irreversible pan-HER receptor tyrosine kinase inhibitor pyrotinib. *Annals of Oncology*, **31**(7), 903–911.

Les sites consultés

- [68]- Altérations des gènes ALK et ROS1 dans les cancers du poumon | Fiche santé HCL. (n.d.). <https://www.chu-lyon.fr/alterations-des-genes-alk-et-ros1-dans-les-cancers-du-poumon?>
- [69]- Delage, V. (n.d.). Dictionnaire médical de l'Académie de Médecine. <https://www.academie-medecine.fr/le-dictionnaire/index.php?q=BRaf%20gene>
- [70]- FISH sur tissus - Imagenome. (2020, January 30). Imagenome. <https://imagenome.fr/definitions/fish-sur-tissus/>
- [71]- Institut National Du Cancer. (n.d.). Cancer du poumon : points-clés - Cancer du poumon. <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-poumon/Les-points-cles>.
- [72]- Roche | Cancer du sein : le statut HER2. (n.d.). <https://www.roche.fr/articles/cancer-sein-statut-her2>
- [73]- Séquençage de l'exome entier | MacroGen Europe. (n.d.). <https://www.macrogen-europe.com/fr-fr/service/whole-exome-sequencing>