

الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université–Ain Temouchent- Belhadj Bouchaib
Faculté des Sciences et de Technologie
Département de Biologie



Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en : Biochimie
Domaine : Sciences de la Nature et de la Vie
Filière : Sciences Biologiques
Spécialité : Biochimie
Thème

Etude *in silico* d'une mutation: application à la mutation
fonctionnelle G719X du gène EGFR retrouvée dans les cancers
bronchiques non à petites cellules

Présenté Par :

- 1) Melle Belouadi Meriem
- 2) Melle Hazam Khaoula
- 3) Melle Khaldi aicha

Devant le jury composé de :

Dr.CHERIF Nadjib	MCB UAT.B.B (Ain Temouchent)	Pésident
Dr.GHEMBAZA Nassira	MCB UAT.B.B (Ain Temouchent)	Examineur
Dr.MOGHTIT Fatima Zohra	MCB UAT.B.B (Ain Temouchent)	Encadrant

Année Universitaire 2021/2022

Remerciements

Nous tenons tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce modeste travail.

Avant de commencer la présentation de ce travail, nous profitons de l'occasion pour remercier toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce projet de fin d'études.

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à ce travail en acceptant d'examiner ce mémoire et de l'enrichir par leurs suggestions.

*Nos remerciements les plus respectueux vont à **M. Cherif Nadjib** qui a bien voulu nous faire l'honneur d'examiner ce travail. Veuillez trouver ici Monsieur le témoignage de notre reconnaissance et notre respect le plus profond.*

*Nous remercions également **Mme. Ghembaza Nassira** pour l'honneur qu'elle nous a fait en acceptant d'examiner ce travail et de siéger parmi cet honorable jury. Veuillez accepter ce travail, en preuve de notre grand respect et de toutes nos profondes reconnaissances.*

*Nous remercions tous particulièrement notre encadrant **Mme. Moghtit Fatima Zohra**. Vous avez bien voulu nous confier ce travail riche d'intérêt et nous guider à chaque étape de sa réalisation. Vous nous avez toujours réservé le meilleur accueil, malgré vos obligations professionnelles. Vos encouragements inlassables, votre confiance, votre gentillesse, votre patience méritent toute admiration. Aucune expression de gratitude ne sera suffisante pour vous exprimer notre respect et notre reconnaissance.*

Nous tenons à exprimer nos sincères remerciements à tous les professeurs qui nous ont enseigné et qui par leurs compétences nous ont soutenus dans la poursuite de nos études.

N'oublions pas de remercier nos très chers parents, frères et sœurs, nos collègues et amis pour leur contribution, leur soutien, leurs encouragements et leur patience

A toutes et tous, un grand merci !

Dédicace

Je dédie ce mémoire :

Avec l'aide de dieu tout puissant, j'ai pu achevé ce modeste travail que je dédie à toutes les personnes qui me sont chère :

A ma très chère mère, source inépuisable de tendresse, de patience et de sacrifice. Ta prière et ta bénédiction m'ont été d'un grand secours tout au long de ma vie. Quoique je puisse dire et écrire, je ne pourrais exprimer ma grande affection et ma profonde reconnaissance. J'espère ne jamais te décevoir, ni trahir ta confiance et tes sacrifices. Puisse Dieu tout puissant, te préserver et t'accorder santé, longue vie et bonheur.

A mon très cher père *ربي برحمه*

J'espère réaliser l'un de tes rêves. Aucune dédicace ne saurait exprimer mes respects, ma reconnaissance et mon profond amour

A mes frères Hamza et sa femme , Djamel et Omar et sa femme qui sont toujours à mes côtés je vous remercie d'être toujours là pour moi. Rien ne saurait exprimer tous les sentiments que je vous porte. Que dieu vous garde pour moi.

A mes petits princes houari, Mohamed et saido puisse Allah vous protéger

A ma sœur Meriem

J'implore Allah de vous réserver un avenir meilleur

Je te remercie de tes encouragement tout le long de ce travail, puisse Allah te protégé.

khaoula

Dédicace

Je dédie ce mémoire à :

A ma chère maman pour tous ses sacrifices, son amour, sa tendresse et son soutien tout au long de mes études .Tu as toujours été pour moi un exemple de mère respectueuse et honnête.

Grâce à toi, maman, j'ai appris le sens du travail je t'aime maman.

A mon cher père (رحمه الله), qui m'a soutenu et encouragé durant ces années d'études.

A mon fiancé qui m'a toujours soutenu et encouragé durant ces années universitaires.

A ma petite sœur Loudjainet mon Binôme et ma sœur Khaoula pour ses encouragements et son soutien moral.

Merci d'être toujours là pour moi.

Meriem

Dédicace

Je dédie ce mémoire :

J'ai le grand plaisir de dédier ce modeste travail

A mes chères parents ma mère et mon père

Pour leur patience, leur amour, leur soutien, et leurs encouragements

A mes frères

A mes amies et mes camarades

Et tout qui m'aide et compluse se modeste travail

Sans oublier tout les professeurs que ce soit du primaire ,de moyen ,

Du secondaire ou l'enseignement supérieur.

ASGHA

Résumé

Le gène *EGFR* est un gène suppresseur de tumeur et son altération moléculaire peut augmenter le risque de cancer bronchique. Afin de contribuer à l'étude moléculaire, nous nous sommes intéressés à l'étude des effets délétères de trois mutations du gène *EGFR*, à savoir la mutation c.2155G>A, la mutation c.2156G>C et c.2155G>T.

Ainsi, les effets délétères des trios mutations ont été prédites en utilisant un protocole d'étude *in silico* constituée de différents logiciels (I-Mutant 2.0 ; SIFT ; Polyphen-2 ; Project HOPE).

Les mutations c.2155G>A,c.2156G>C,c.2155G>T sont prédites comme étant délétères altérant la protéine EGFR. En fait, les changements physico-chimiques entre les acides aminés sauvages et mutés peuvent perturber la stabilité des protéines et entraîner une perte d'interactions avec d'autres molécules. Nos résultats montrent que ces différents logiciels sont cohérents et que leur combinaison peut améliorer les performances de prédiction des effets de ces mutations.

Cette étude nous permet de mieux comprendre l'impact des mutations et met en lumière l'importance de la recherche prédictive dans les logiciels de bioinformatique.

Mots clés : Protéine EGFR, Mutation, effet, *in silico*, Bioinformatique

Abstract

The EGFR gene is a tumor suppressor gene and its molecular alteration may increase the risk of breast cancer. In order to contribute to the molecular study, we were interested in the study of the deleterious effects of three mutations of the EGFR gene, namely the c.2155G>A mutation, the c.2156G>C mutation and c. 2155G>T.

Thus, the deleterious effects of the three mutations were predicted using an *in silico* protocol consisting of different software (I-Mutant 2.0; SIFT; Polyphen-2; Project HOPE).

The mutations c.2155G>A, c.2156G>C, c.2155G>T are predicted to be deleterious altering the EGFR protein. In fact, physicochemical changes between wild-type and mutated amino acids can disrupt protein stability and lead to loss of interactions with other molecules. Our results show that these different softwares are consistent and that their combination can improve the performance of predicting the effects of these mutations.

This study allows us to better understand the impact of mutations and highlights the importance of predictive research in bioinformatics software.

Keywords. EGFR protein, Mutation, effect, *in silico*, Bioinformatic

الملخص

إن الجين EGFR هو جين كابح للورم وقد يؤدي تغييره الجزيئي إلى زيادة خطر الإصابة بسرطان الشعب الهوائية. من أجل المساهمة في الدراسة الجزيئية، قد اهتمنا في هذا البحث بدراسة (in silico) لوظائف ثلاث طفرات في جين EGFR، وهي c.2155G>A، و c.2156G>C، و c.2155G>T. قد تم التنبؤ بالآثار الضارة للطفرات باستخدام بروتوكول دراسة *in silico* الذي يتكون من برامج مختلفة I- Mutant 2.0 (SIFT Polyphen-2 Project HOPE). من المتوقع أن تكون الطفرات c.2155G>A، و c.2156G>C، و c.2155G>T ضارة في تغيير البروتين. في الواقع، تؤدي التغييرات الفيزيائية والكيميائية بين الأحماض الأمينية الأصلية والمتحورة إلى تعطيل استقرار البروتين ويمكن أن تؤدي إلى فقدان التفاعل مع الجزيئات الأخرى. كما أظهرت نتائجنا أن هذه البرامج المختلفة متناسقة وأندمجها يمكن أن يزيد من أداء التنبؤ بتأثير هذه الطفرات. سمحت لنا هذه الدراسة بفهم تأثير الطفرات بشكل أفضل وكشفت عن أهمية دراسات التنبؤ بواسطة برامج المعلوماتية الحيوية. الكلمات المفتاحية: EGFR, in silico, المعلوماتية الحيوية.

Table des matières

Liste des figures	I
Liste des tableaux	III
Liste des abréviations.....	IV
Introduction	2
Revue Bibliographique	3
I. Description de gène EGFR	4
II. La protéine EGFR	4
II-1 La structure de la protéine	4
III. Rôle et fonction	6
IV. Les mutations du gène de l'EGFR.....	8
V. Généralités sur le cancer bronchique.....	9
V-1 Définition et épidémiologie.....	9
V-2 les types des cancers bronchiques	10
VI. Les thérapies.....	12
VI-1 La Chimiothérapie.....	12
VI-2 Les thérapies ciblées	12
VI-3 La radio thérapie	12
VI-4 La chirurgie	12
VI-5 Immunothérapie	12
VII. Généralités sur la bioinformatique	13
VII-1 Définition	13
VII-2 les outils bioinformatiques.....	14
Matériels et méthodes	
I. Matériels.....	17
II. Méthodes.....	17

II-1 Prédiction de l'effet de la mutation sur la stabilité de la protéine par le logiciel	
I-Mutant 2.0.....	18
II-2 Prédiction des effets des mutations par le logiciel Polymorphisme Phenotyping-2	
(Poly Phen-2).....	18
II-3 Prédiction de l'effet délétère des mutations exoniques sur la protéine par le logiciel	
SortingIntolerantFromTolerant (SIFT).....	20
II-4 Project HOPE (Have (y)Our ProteinExplained).....	21
Résultat et discussion	
I .Résultat.....	23
I.1 Résultat obtenu par le logiciel I-Mutant 2.0.....	23
I.2 Résultats obtenus par le logiciel POLYPHEN-2.....	24
I.3 Comparaison des propriétés physicochimiques entre les résidus normaux et mutés	
(Project HOPE) pour La mutation c.2156G>C (p.Gly719Ala)	25
I.4 Comparaison des propriétés physicochimiques entre les résidus normaux et mutés	
(Project HOPE) pour La mutation c.2155G>T (p.Gly719Cys)	26
I.5 Comparaison des propriétés physicochimiques entre les résidus normaux et mutés	
(Project HOPE) pour La mutation c.2155G>A (p.Gly719Ser).....	27
I.6 Résultats obtenus par SIFT.....	28
II Discussion	29
Conclusion et perspective	
Conclusion	32
Références bibliographiques	34

Liste des figures

Figure 1: Localisation du gène EGFR sur le chromosome 7.....	4
Figure 2: Famille des récepteurs de l'EGF	5
Figure 3 :La structure du récepteur EGF.....	6
Figure 4: Les voies de signalisation.....	8
Figure 5 : Types histologiques les plus fréquents de carcinome bronchique primitif.....	11
Figure 6: Thérapie ciblée.....	13
Figure 7:Interface du logiciel I-Mutant2.0	18
Figure 8 : Interface du logiciel Poly Phen-2	19
Figure 9: Interface du logiciel sift	20
Figure 10 : Interface du logiciel Project Hope	21
Figure 11: Résultat obtenu par le logiciel I-Mutant 2.0 pour la mutation c.2156G>C (p.Gly719Ala)	23
Figure 12 : Résultat obtenu par le logiciel I-Mutant 2.0 pour la mutationc.2155G>T (p.Gly719Cys).....	23
Figure 13 : Résultat obtenu par le logiciel I-Mutant 2.0 pour la mutation.....	24
Figure 14 : Résultat obtenu par le logiciel PolyPhen-2 pour les mutations c.2155G>A (p.Gly719Ser) c.2156G>C (p.Gly719Ala) c.2155G>T (p.Gly719Cys).....	24
Figure 15: Les structures schématiques de la Glycine et l'alanine	25
Figure 16 : Les structures schématiques de la cystéine et glycine.	26
Figure 17 : Les structures schématiques de glycine et Sérine.....	27
Figure 18: Résultat obtenu par le logiciel SIFT pour la mutation.....	28

Liste des tableaux

Tableau I : Liste des mutations analysées	17
Tableau II: Les propriétés physicochimiques de deux aa la glycine et l'alanine	26
Tableau III: les propriétés physique chimiques des deux AA cystéine et glycine.	26
Tableau IV: les propriétés physique chimiques des deux AA Sérine et glycine.....	27
Tableau V: L'hydrophobicité des acides aminés	28
Tableau VI : Les différentes propriétés physico-chimiques correspondant aux couleurs des aa obtenus par SIFT pour la mutation c.2155G>A (p.Gly719Ser) c.2156G>C (p.Gly719Ala) c.2155G>T (p.Gly719Cys).....	28

Liste des abréviations

AA:acide aminé

AKT:proteinekinase B .

Align GVGD: web based program that combines the biophysical characters of amino acids.

ARNm:acideribonucléique messenger.

ATP: adenosine triphosphate.

BLAST: basic local alignment search tool.

BRCA1:breast cancer 1.

C- myc,C -fos,ELK-1,C-jun : des facteurs de transcriptions codée avec des protéines.

CBNPC:cancer bronchique non a petite cellule.

CBPC:cancer bronchique a petit cellule.

CPCR:collectif des parties civiles pour le Rwanda.

CR1: cestirnrich .

DDG: différence d'énergie libre entre la protéine sauvage et mutée.

E746-A750: deletion mutation

EGF: epidermal growth factor.

EGFR:epidermal growth factor receptor.

ERbB:thefamille is comprised of four receptor tyrosine kinase.

G719A:mutation Glycine Alanine.

G719C:mutation Glycine Cystéine.

G719S: mutationGlycine Sérine.

GPCR: récepteurs couplés aux protéines G.

GRB2/sos : growth factor receptor bound protein 2/son of sevenless.

HB-EGF:heparin-binding EGF-like growth factor.

HER:human epidermal growth factor receptor .

IGF1-R: insulin like growth factor 1 receptor.

I-Mutant: predicting stability changes upon mutation from the proteinsquence or structure.

Kb: kilo base.

KD4v mutation tasting: comprehensible knowledge discovery system for missense variant.

KDa: kilo dalton.

MAPK: mitogenactivatedproteien kinases.

NIH: national institutes of health.

OMS: organisation mondiale de la santé.

p: bras court du chromosome.

pb: paire de base.

PDB: protein DATA BanK.

PHD SNP: predictor of human deleterious single nucleotide polymorphisms.

PI3 kinase: phosphoinositide 3-kinase.

PI3K: phosphatidyl-inositol 3 kinase.

PLC gamma: phospholipase C gamma.

Polyphen-2: polymorphism phenotyping-2.

Projecthope: have our protein explained.

PROTEINE pRb: protéine du rétinoblastome.

PyK2: proline-rich tyrosine kinase 2.

REGfvIII: forme tronquée de REGF.

SH2: src homology 2.

SIFT: sorting intolerant from tolerant.

SMV: support vector machine.

Src: service renseignement de la confédération.

STAT: signal transducer and activator of transcription.

STAT3: signal transducer and activator of transcription 3.

TAIR: the Arabidopsis information resource.

TGF-alpha: transforming growth factor alpha.

UniProtKB: base de données universelle de protéine.

Introduction

Le cancer est un problème sans cesse croissant prenant des vies dans le monde entier à un rythme alarmant. C'est une maladie caractérisée par la présence d'une tumeur maligne formée à partir d'une prolifération cellulaire anormalement importante au sein d'un tissu normal de l'organisme. Cette prolifération peut toucher d'autres organes formant ainsi des métastases (Lacave,, 2005).

Les progrès récents de la biologie moléculaire ont démontré que le cancer est une maladie de l'ADN qui résulte de l'accumulation d'évènements génétiques (mutations ponctuelles et réarrangements chromosomiques) qui altèrent le fonctionnement normal de certains gènes. Il est bien établi que la cancérogénèse résultent de l'inactivation des gènes supprimeurs de tumeurs et activation des proto-oncogène(Blay, 2010). Ainsi, EGFR est un des proto-oncogènes les plus fréquemment activés dans les cancers du poumon.

Plusieurs mutations somatiques touchant le gène EGFR ont été mises en évidence et elles sont fréquentes dans le cancer du poumon. Ces mutations sont localisées dans les 4 exons du gène de l'EGFR (exons 18 à 21) .Elles correspondent à des délétions respectant le cadre de lecture, des insertions ou des mutations ponctuelles faux-sens(Raymond, 2009).

Au cours de ce travail, nous nous sommes intéressées à l'étude *in silico* des fonctionnalités de quelques mutations du gène EGFR au niveau protéique par l'utilisation de différents logiciels de prédiction. Les effets délétères des mutations ont été prédits en utilisant plusieurs logiciels de prédiction : I-Mutant 2.0, SIFT, Polyphen-2, Align GVGD, KD4v Mutation Tasting et Project HOPE.

Revue
Bibliographique

I. Description du gène EGFR :

Le gène du récepteur de facteur de croissance épidermique est appelé gène EGFR (Epidermal Growth Factor Receptor) est un des tout premiers proto-oncogène cellulaires humains identifié et le premier protooncogène membranaire .Il est situé au niveau du bras court du chromosome 7 dans la région 7p11.2 (Figure 1), et s'étend sur environ 200kb. Ce gène est constitué de 32 exons et 31 introns .L'intron 1 représente à lui seul 123kb. Son principal transcrit est un ARN messager (ARNm) de 9905 paire de base (pb)(Reiter et al. 2001).

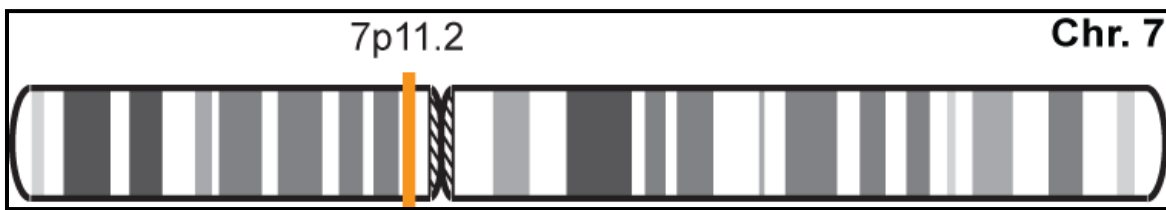


Figure 1:Localisation du gène sur le chromosome 7

II La protéine EGFR

II-1 Structure

Le gène EGFR code pour une glycoprotéine transmembranaire de 170KDa composée de 1210 acides aminés. Il est le premier récepteur membranaire à avoir identifié et cloné. Ce dernier appartient à la famille des récepteurs de tyrosine kinase(de La Motte Rouge, Valent,et al, 2007).

Cette protéine est présente dans les cellules d'origine méso- et ectodermiques à un taux de 2×10^4 à 2×10^5 récepteurs par cellule. Au niveau de l'appareil respiratoire, l'expression de l'EGFR est forte dans les cellules épithéliales bronchiques et modérée dans les cellules épithéliales alvéolaires. Elle est également surexprimée dans de nombreuses tumeurs solides comme les tumeurs gliales, les Carcinomes mammaires, les carcinomes coliques et les carcinomes bronchiques non à petites cellules (Carlson, Garrison, et al, 2009).

L'EGFR s'intègre à la famille des facteurs de croissance épidermiques erbB/HER comprenant erbB1=Her1 nommé communément EGFR, erbB2/Her2/neu, erbB3/Her3 et erbB4/Her4 (Figure 02) .Leur structure est commune et comporte quatre domaines :

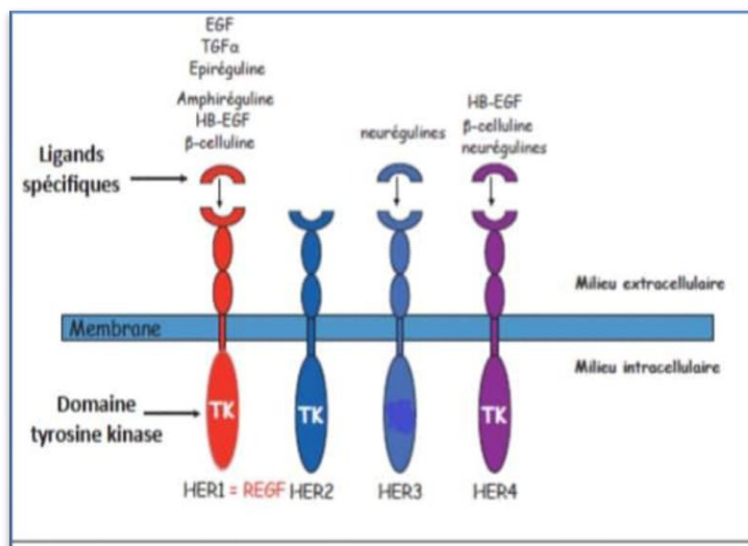


Figure 1: Famille des récepteurs de l'EGF (Lièvre et Laurent-Puig 2010).

-Le domaine extracellulaire (621 aa) composé du site de fixation des ligands entourés de deux séquences riches en cystéine qui lui confèrent sa structure tertiaire. Ce domaine est composé de 4 sous-domaines (I à IV) répétés 2 à 2, se succédant en alternance. Les domaines I et III sont analogues à ceux de l'Insuline-like Growth Factor type-1 Receptor (IGF1-R), un autre récepteur à activité tyrosine kinase. Ces deux Domaines ont été identifiés comme étant le siège de la fixation du ligand, d'où leur Dénomination L1 et L2. les domaines II et IV sont caractérisés par leur abondance en résidus cystéine d'où leur appellation CR1 et 2 (CR=Cystéine riche).

-Le domaine transmembranaire (23 aa) lipophile permettant l'ancrage du récepteur dans la membrane plasmique

-Le domaine intracellulaire (542aa) est constitué du domaine tyrosine kinase (environ 250aa) dont l'activité nécessite la fixation d'une molécule d'ATP (adénosine triphosphate) au niveau de la lysine 721.

-Un domaine de régulation C-terminal qui comporte plusieurs résidus tyrosine se phosphorylant de manière spécifique après liaison du ligand. (Figure 3)

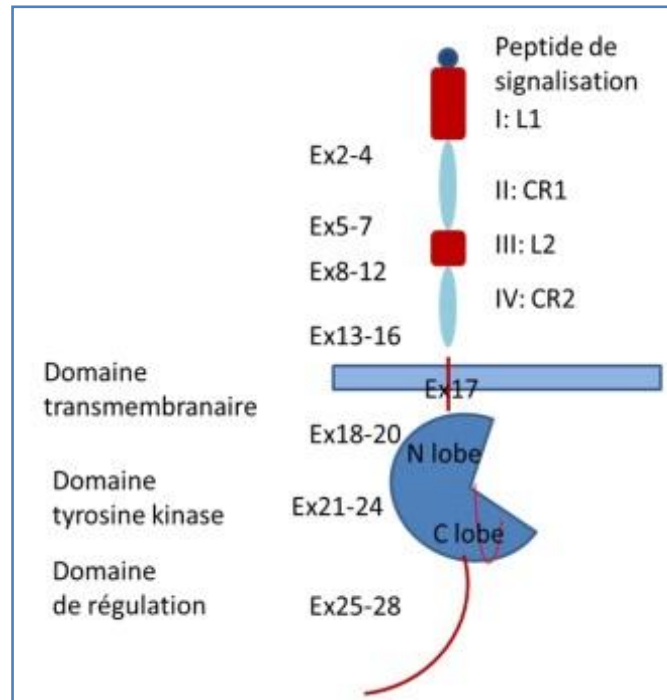


Figure 2 :La structure du récepteur EGF (

III- Rôle et fonction

Le récepteur EGFR est impliqué dans les voies de signalisation de l'apoptose, de la prolifération et de la différenciation cellulaires.

L'activation du récepteur se fait par la fixation d'un ligand au niveau de son site de liaison extracellulaire. Il existe six ligands pour l'EGFR : l'épidermal Growth factor (EGF), le transforming growth factor alpha (TGF α), l'héparin-binding (HB)-EGF, l'amphiréguline, la beta cellulaire et l'épiréguline dont la fixation contrôle pour une part la nature du dimère HER formé. Ces ligands endogènes sont synthétisés sous forme de proligands à faible affinité pour le récepteur. Sous l'action de métallo-protéases membranaires, ils subissent un clivage protéolytique qui les transforme en ligand fonctionnel de haute affinité pour le récepteur (Azria & Spano, 2006). Dans sa forme inactive, l'EGFR est un monomère. La fixation d'un des ligands précédemment cités provoque sa dimérisation. L'appariement de deux récepteurs identiques constitue une homodimérisation par exemple, HER1/HER1) alors qu'une hétérodimérisation met en jeu deux récepteurs HER différents. Cette dimérisation provoque un changement conformationnel du domaine intracellulaire qui déclenche l'activité tyrosine Kinase. La fixation de l'ATP sur la lysine 721 du site catalytique fournit le substrat phosphate permettant la phosphorylation. Ensuite, l'autophosphorylation de résidus tyrosine situés dans la partie C-terminale du récepteur permet le recrutement de molécules possédant des domaines SH2 (Src homology 2^o) ou des domaines de liaison à des tyrosines telles que

GRB2/SOS (Growth Factor receptorboundprotein 2/Son of sevenless), PLC gamma(phospholipase C gamma) ,STAT (Signal transducer and Activator of transcription) et la P13K (phosphatidyle-inosito 3 kinase).L'interaction de ces protéines adaptatrices avec le récepteur déclenche la transduction du signal des différentes voies de signalisation régulées par L'REGF. Parmi ces voies, trois ont une importance biologique fondamentale dans le tissu normal. (Meert et al., 2002)

- La voie des MitogenActivatedProtein Kinases (MAPK) qui induit l'activation de facteur de transcription tels que c-mys, c-fos, c-jun, ELK-1, ect, est plus particulièrement impliquée dans la prolifération et la différenciation cellulaire. En effet, en fonction des types cellulaire concernés et/ou du type de dimère formé, leseffets biologiques observés sont différents orientant la cellule soit vers une voie de prolifération soit de différenciation.
- La voie de P13 kinass est une voie impliquée dans la survie cellulaire. En stimulant la protéine kinase AKT, elle est responsable de la régulation de voies antiapoptotiques.
- La voie STAT3 et la voie de la PLC gamma via l'activation de la protéine kinase C, régulent la transcription de gène impliquée dans la prolifération, l'apoptose et l'angiogenèse.

Dans l'organisme vivant pluricellulaire, le signal de prolifération régulé par l'EGFR est connecté à d'autres voies de signalisation cellulaire. La coexistence des récepteurs de la famille HER avec d'autres récepteurs (aux cytokines, neuromédiateurs...) implique forcément la présence d'un réseau complexe de régulations positives ou négatives dont les mécanismes sous-jacents font l'objet de nombreuses recherches(Pierre & Sandrine, 2007). Deux types d'interaction ont déjà été rapportés et sont illustrés dans la (Figure 4) :

- Une protéine G, activée par un récepteur couplé à une protéine G(GPCR), déclenche une mobilisation du calcium intracellulaire qui provoquant l'activation de Pyk2. Cette protéine kinase cytosolique va phosphoryler Src, tyrosine kinase située au carrefour de différentes voies de signalisation et va indirectement, activer la voie de L'REGF (Haley & Gullick, 2009).
- Ce même récepteur GPCR est aussi responsable de l'activation de métallo protéases de la matrice extracellulaire, qui vont pouvoir cliver HB-EGF, réservoir d'EGF qui sera ainsi libéré pour activer le récepteur.
- En cancérologie, l'REGF est impliqué de façon formelle dans l'oncogenèse de différents-types de tumeurs. Ce récepteur est un proto-oncogène dont le rôle dans

l'oncogenèse précoce résulte toujours d'un gain de fonction. Ce gain de fonction peut découler de plusieurs mécanismes :

- D'une surexpression du récepteur au niveau de la membrane cellulaire et /ou d'une augmentation de l'expression du ligand
- D'une activation du récepteur indépendante du ligand. C'est le cas du récepteur mutant tranquéREGFvIII, qui présente une délétion des exons 2 à 7
- D'une augmentation de l'activité intrinsèque de phosphorylation en présence de quantité normale de ligand, en raison de l'existence de mutations au niveau du domaine catalytique (Acton, 2013).

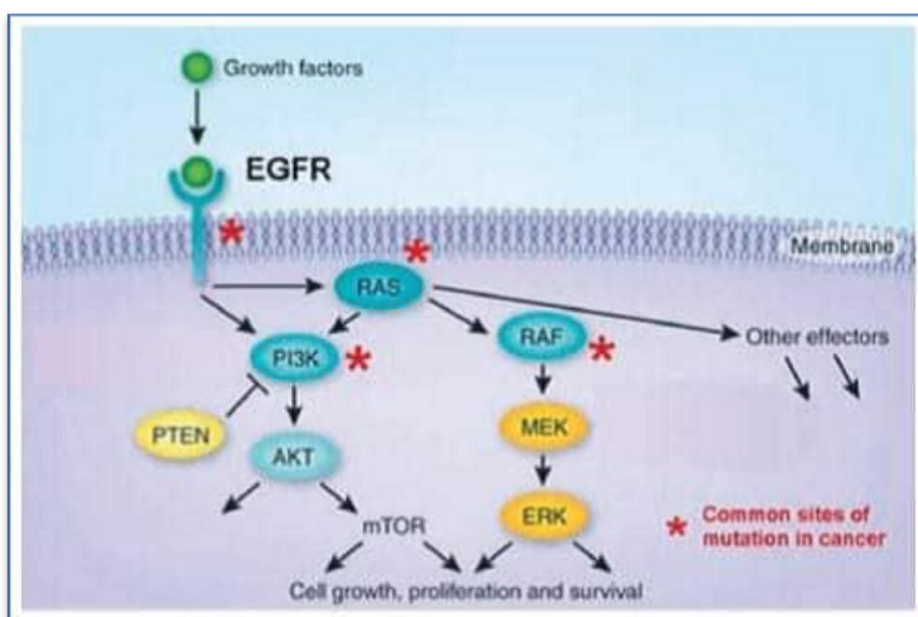


Figure 3: Les voies de signalisation

IV. Les mutations du gène de l'EGFR

Les mutations somatiques impliquées dans le cancer bronchiques sont localisées dans les 4 exons du gène de l'EGFR codant pour l'activité tyrosine kinase (codons 18 à 21). Elles se situent au niveau du site de fixation de l'ATP. Elles correspondent à des délétions respectant le cadre de lecture, des insertions ou des mutations ponctuelles faux-sens.

- Exon 18 :

Au sein de l'exon 18, il s'agit de mutations touchant le codon 719 et à l'origine une substitution d'une glycine par une cystéine, alanine ou sérine (G719X) représentant 4 % des mutations (Mitsudomi& Yatabe, 2010) .

- Exon 19 :

Au sein de l'exon 19, il s'agit de délétions, intéressant toujours la leucine en 747 et la glutamine en 749 et représentant 44 % des mutations de l'EGFR. Il existe plus de 20 variantes de délétion de l'exon 19 (9, 12, 15, 18, 24 pb ou délétions-insertions, ...) (de La Motte Rouge et al., 2007). Il existe également des insertions représentant 1 % des mutations.

-Exon 20 :

Au sein de l'exon 20, il s'agit d'insertions représentant 3% des mutations, dont la plus fréquente est la substitution T790M (Bronte et al., 2010). Moins de 1% des mutations concernent les mutations ponctuelles V765A et T783A (Bronte et al., 2010).

- Exon 21 :

La mutation ponctuelle L858R de l'exon 21 représente 41 % des mutations activatrices de l'EGFR, la mutation ponctuelle L861Q 2 % (Gazdar, 2009). La mutation L861R a aussi été recensée dans l'exon 21.

La délétion de 15 pb retrouvée dans l'exon 19 du codon 746 au codon 750 (E746- A750) et la mutation ponctuelle L858R de l'exon 21 représentent 90 % des mutations des tumeurs bronchiques (Gazdar, 2009). Toutes les mutations ne seraient pas équivalentes du point de vue du pronostic. Les délétions de l'exon 19 seraient associées à une meilleure survie que la mutation ponctuelle L858R de l'exon 21.

V. Généralités sur le cancer bronchique

V.1 Définition et épidémiologie

Le cancer bronchique est une tumeur maligne originaire de la muqueuse tapissant les bronches. On parle de cancer primitif du poumon ou plus communément du cancer du poumon.. Il représente aujourd'hui la première cause de mortalité par cancer. Il est à l'origine de plus de 90% des tumeurs décelées dans les poumons. Même si durant la dernière décennie la mortalité chez l'homme s'est stabilisée, on observe actuellement une augmentation de la mortalité chez la femme et on estime qu'il supplante actuellement le cancer du sein aux Etats Unis. Ceci s'explique par la forte consommation de tabac chez la femme depuis 20 ans. En effet, le tabac est le facteur majeur du cancer du poumon. Le risque de cancer bronchique lié au tabac est en rapport avec le nombre de cigarettes fumées par jour et avec la durée d'exposition. Il diminue pendant plusieurs années après l'arrêt du tabagisme mais ne s'annule pas. Ce risque concerne non seulement les carcinomes épidermoïdes (99%) et les cancers à petites cellules (90%), mais également les adénocarcinomes et les carcinomes à grandes

cellules (70%). D'autres facteurs peuvent être impliqués dans la survenue de cette affection à savoir : les facteurs d'exposition professionnelle, les facteurs environnementaux ainsi que les facteurs génétiques (Robert, 2006).

V.2 Les types des cancers bronchiques

Le cancer bronchiques représentent environ 75 à 85 % des cancers du poumon et se soignent plus facilement que les CBPC car ils progressent plus lentement (Herbst et al. 2008). Ils regroupent trois sous- types histologiques très différents selon le type de cellules concernées. Les carcinomes épidermoïdes, les adénocarcinomes et les carcinomes à grandes cellules (Brambilla et al ., 2001).

❖ Adénocarcinomes

Les adénocarcinomes sont des tumeurs épithéliales avec différenciation glandulaire. Cette différenciation peut être reconnue morphologiquement par la présence de glandes et d'une mucosécrétion. (Figure4)(Rothschilda et al., 2010). Ils sont caractérisés par un aspect macroscopique blanchâtre, brillant et se déclarent en périphérie du poumon sous forme de nodules solitaires émergeant des glandes bronchiques et des cellules alvéolaires. Ils représentent environs 60% des CBNPC. Etant donné l'hétérogénéité histologique rencontrée, les adénocarcinomes sont divisés en quatre sous-groupes en fonction des éléments prédominants : acineux, papillaires, bronchiole-alvéolaires et solides. Le taux de survie à cinq ans des individus atteints d'adénocarcinome est d'environ 12%. Actuellement, ils sont les plus fréquemment diagnostiqués chez les non-fumeurs, dans la population féminine et la population asiatique (Brambilla et al ., 2001).

❖ Carcinomes épidermoïdes (squameux)

Les carcinomes épidermoïdes sont caractérisés par des cellules tumorales avec noyaux fortement atypiques, polymorphes et cytoplasme kératinisé à certains endroits(Figure4)(Rothschilda et al., 2010) . Ils siègent majoritairement sur des bronches centrales sous forme de masses térébrantes et hémorragiques. La différenciation épidermoïde se manifeste par l'existence de nombreux desmosomes et, éventuellement, par des foyers de kératinisation (globes cornés). Le taux de survie à cinq ans des individus atteints est d'environ 25%. Ce type de cancer constitue environ 30 % des CBNPC. Ils ont une prédominance masculine et sont fortement corrélés à la consommation de tabac (Brambilla et al ., 2001) .

❖ **Carcinomes à grandes cellules :**

Les carcinomes à grandes cellules sont caractérisés généralement par des noyaux plus gros (Figure 04) (Rothschilda et al., 2010) qualifiés de « non-différenciés ». En effet, ce type de tumeur est défini par l'OMS (Organisation Mondiale de la Santé) comme n'ayant pas les caractéristiques des carcinomes épidermoïdes, des adénocarcinomes ou des tumeurs à petites cellules. Ils représentent 10% des CBNPC et le taux de survie à cinq ans est de 13%(Rothschilda et al., 2010) .

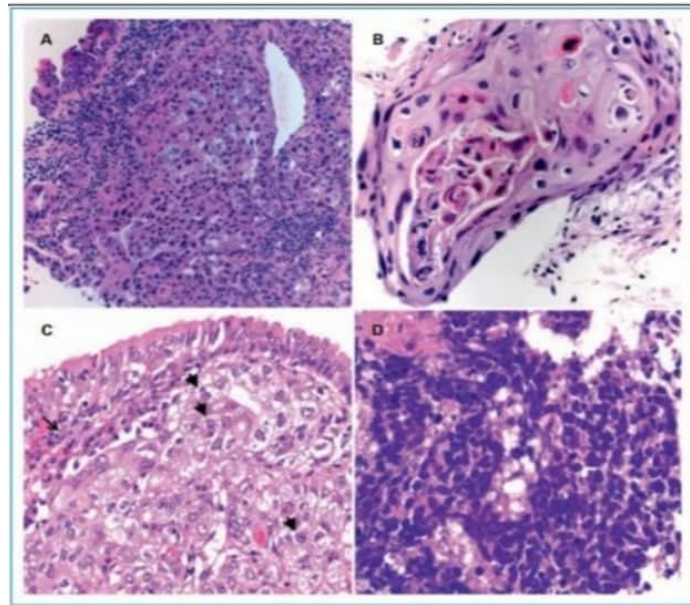


Figure 4 : Types histologiques les plus fréquents de carcinome bronchique primitif.(Rothschilda et al. 2010)

A:Adénocarcinomes dans une biopsie bronchique. Glandes bien reconnaissables avec mucus intraluminal. Agrandissement 200 fois.

B :carcinome épidermoïde dans une biopsie bronchique : cellules tumorales avec noyaux fortement atypiques, polymorphes et cytoplasme kératinisé à certain endroits (en rouge) .Agrandissement 400 fois.

C :Carcinome pulmonaire à grandes cellules (large celllungcarcinoma) dans la muqueuse bronchique. Les noyaux des cellules tumorales (pointe de la flèche sont généralement plus gros que trois petits noyaux lymphocytaires (flèche)). Aucune autre caractéristique de différenciation dans le tissu cancéreux. Agrandissement 400fois.

D :carcinome à petites cellules : cellules tumorales avec noyaux relativement petits mais fortement hyper chromatiques et polymorphes. Peu de cytoplasme, artefacts typiques d'écrasement mécanique des noyaux. Agrandissements 400 fois.

VI. Les thérapies

VI.1. La chimiothérapie

Un traitement par chimiothérapie consiste à injecter des molécules qui seront toxiques pour les cellules cancéreuses. Cependant, Cette action thérapeutique n'agit pas de façon spécifique sur les cellules cancéreuses et touche aussi des cellules saines comme les cellules du sang (Rotow, Cheema and al ; 2022).

VI.2. La radiothérapie

La radiothérapie thoracique est une arme thérapeutique indispensable à la prise en charge des formes localisées inopérables des carcinomes bronchiques non à petites cellules (CBNPC) qui s'articule avec la chirurgie et les traitements généraux(Durdux et al,2018) .

VI.3.La chirurgie

La chirurgie, ablation de la tumeur, reste le seul traitement curatif du cancer bronchique non à petites cellules (CBNPC). La chirurgie est le traitement de choix dans les cancers de stade I et II, et a une place importante pour certains cancers de stade III, en association à la chimiothérapie et la radiothérapie.

VI.4.Immunothérapie

L'immunothérapie consiste à stimuler le système immunitaire pour qu'il détruise les cellules tumorales. Ces traitements ne se contentent plus de bloquer le développement de la tumeur mais ont vocation à la détruire. Les cellules immunitaires (lymphocytes) ainsi stimulées par le traitement d'immunothérapie ne sont plus inactivées par les cellules cancéreuses et peuvent ainsi les détruire(Mazieres et al, 2020).

VI.5.Les thérapies ciblées

De nouvelles stratégies thérapeutiques, appelées thérapies ciblées sont aujourd'hui possibles grâce aux progrès récents en oncologie. Ces traitements atteignent plus spécifiquement les cellules cancéreuses en perturbant les mécanismes biologiques qui conduisent à leur multiplication : ils agissent essentiellement au niveau des récepteurs du facteur de croissance. Les inhibiteurs de tyrosine kinase sont de petites molécules qui agissent

spécifiquement à l'intérieur des cellules tumorales en bloquant l'activation des tyrosines kinases. Elles font partie de la nouvelle génération de médicaments anticancéreux appelés thérapies ciblées.(Figure 05) (Rotow, Cheema and al ;2022).

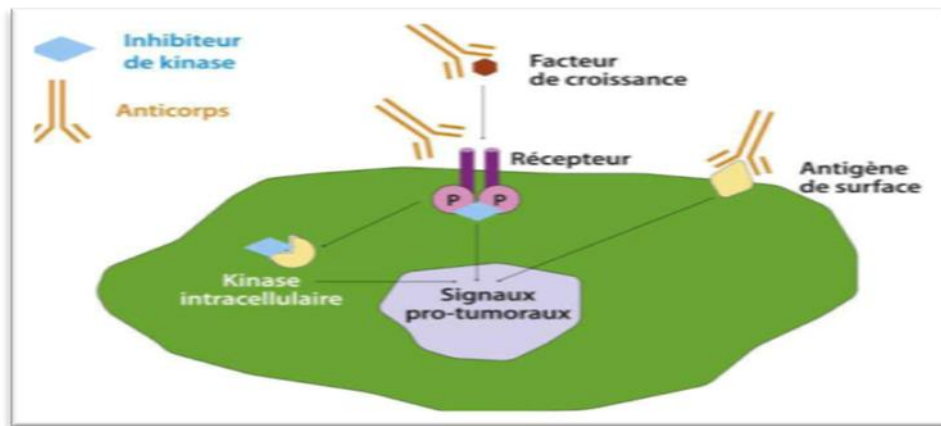


Figure 6: la thérapie ciblée (Faure,2015)

VII Généralités sur la bioinformatique

VII.1 Définition

Les données biologiques s'accumulent extrêmement rapidement. Pour exploiter toutes ces données, une nouvelle science est née : la bioinformatique.

La bioinformatique est la science de l'utilisation de l'ordinateur dans l'acquisition, le traitement et l'analyse de l'information biologique. Les données traitées par la bioinformatique sont toutes celles qui intéressent le biologiste : séquences d'acides nucléiques, séquences protéiques, références bibliographiques, résultats expérimentaux.....etc. Le terme bioinformatique, aussi appelé en anglais « computational ou *in silico* biology ».

L'origine de la bioinformatique est relativement ancienne. En effet, le terme "bioinformatique" a été utilisé dès 1977 par la biologiste allemande Paulien Hogeweg qui a établi un groupe de recherche de bioinformatique à l'Université d'Utrecht (Hogeweg et Hesper., 1978). Cependant, le mot bioinformatique n'a eu véritablement lieu dans la littérature scientifique que vers 1990 (Claverie., 2007).

Initialement, la bioinformatique consistait à créer et à maintenir des bases de données pour le stockage de l'information biologique comme les séquences nucléotidiques et d'acides aminés. Le passage de la recherche en biologie de l'étude d'un seul gène à l'étude de génomes et de protéomes d'organismes complets, a nécessité des moyens de calculs de plus en plus importants, le rassemblement des banques de données, le développement de programmes performants et le développement de nouvelles techniques. L'analyse de cette grande quantité de données n'est rendue possible que par l'utilisation de la bioinformatique. En effet, la bioinformatique est un élément essentiel pour le passage de la biologie à l'ère post génomique et aux études à haut débit (Aube et al., 2004).

V.2 les outils bioinformatiques

La bioinformatique permet d'analyser les données de génomique, de transcriptomique et de métabolomique et d'exploiter la quantité de données issues des analyses omiques. De nombreuses banques de données ont été créées pour regrouper et mettre en commun les données collectées sous un format unique, harmonisé et annoté : La protéine Data Bank (pdb) pour les structures protéiques, Genbank du NIH pour les séquences d'ADN, Uniprot pour les séquences protéiques ; Mais aussi des banques de données spécialisées en fonction des organismes modèles : The Arabidopsis Information Resource (TAIR). Pour les séquences d'ADN, ARN et protéines d'Arabidopsis...

Les outils d'informatique disponibles permettent aujourd'hui d'aligner un très grand nombre de séquences entre elles par une technique appelée Basic local Alignment Search Tool (BLAST) mais aussi aligner deux séquences, c'est chercher à savoir comment passer de l'une à l'autre : on comptabilise les modifications nécessaires pour écrire la seconde en partant de la première. On peut ainsi rechercher toutes les séquences similaires à une séquence d'intérêt.

BLAST nous indiquera pour chaque séquence comparée à la nôtre (et il peut le faire sur toutes les séquences de la base de données interrogée, soit plusieurs dizaines de millions!) Par ordre décroissant de ressemblance :

- Le pourcentage de couverture (% de la longueur de la séquence d'intérêt qui s'aligne avec la séquence interrogée).
- Le pourcentage d'identité (nombre de nucléotides identiques entre les deux séquences / nombre total de nucléotides alignés).

- La E- value (probabilité d'observer cet alignement par le seul fait de hasard : plus cette valeur est faible, plus l'alignement.

La bioinformatique consiste aussi à rechercher des domaines de recherche des domaines ou des motifs protéiques particuliers en utilisant des chaînes de Markov ou même de modéliser une structure protéique *de novo* à partir de sa séquence par homologie.

D'autres outils bioinformatiques permettent de prédire les régions hydrophobes (transmembranaires par exemple), des régions accessibles (antigènes potentiels pour la synthèse d'anticorps de détection), des régions d'interaction avec d'autres molécules comme les motifs d'interaction avec l'ADN. Ainsi la bioinformatique offre un panel d'outils dans le domaine de l'analyse de séquence des protéines, qui judicieusement combinés sont des moyens efficaces et rapides pour l'étude d'une protéine dont seule la séquence est connue. Néanmoins les déductions obtenues *in silico* doivent être vérifiées par des approches expérimentales *in vitro* et *in vivo* (Dubrin et al. 2002).

Les conséquences des mutations faux-sens ou des délétions/insertions en phase (y compris les anomalies de l'épissage) sont moins certaines. Ces mutations sont responsables d'un changement de la séquence protéique et peuvent notamment affecter la stabilité, l'adressage intracellulaire, la maturation de la protéine, les sites importants pour l'activité enzymatique ou encore les interactions fonctionnelles avec des ligands et d'autres protéines. Mais le changement de la séquence de la protéine peut également être sans conséquence sur sa fonction. Ainsi, lorsqu'un changement nucléotidique de type faux-sens ou une délétion/insertion en phase surviennent au sein de la séquence codante d'un gène, un faisceau d'arguments est nécessaire pour déterminer les conséquences de cette anomalie dans la pathologie. De ce fait, de nombreux logiciels ont été développés afin de prédire les différents effets et impacts de ces mutations délétères. La modélisation tridimensionnelle (3D) vient s'ajouter à ces nombreux logiciels de prédiction et permet une meilleure appréciation des défauts moléculaires.

*Matériel et
méthodes*

I- Matériel

Les mutations du récepteur du facteur de croissance épidermique EGFR sont fréquemment retrouvées chez les patients atteints du CBNPC. Elles favorisent l'analyse anarchique des cellules tumorales. Dans la présente étude nous nous sommes intéressés à la à trois (03) mutations du gène EGFR. Ces mutations siègent sur l'exon 18. Elles sont résumées dans le tableau I. Ces mutations conduisent à la substitution de l'acide aminé Glycine en position 719 de la protéine

Tableau I : Liste des mutations analysées

Exon/Intron	Type de mutation	DNAc	Protéine
Exon 18	Faux –sens	c.2155G>A	(p.Gly719Ser)
	Faux –sens	c.2156G>C	(p.Gly719Ala)
	Faux –sens	c.2155G>T	(p.Gly719Cys)

II- Méthode

Les variations au niveau des gènes conduisent le plus souvent à un dysfonctionnement des protéines synthétisées, conduisant à l'apparition de plusieurs pathologies. Ces variations peuvent se produire aussi bien au niveau des régions géniques codantes (exons), qu'au niveau des introns et de ce fait avoir un effet sur la structure et/ou la fonction de la protéine. Cela explique, la diversité des logiciels destinés à prédire les différents effets délétères des mutations étudiées. Les effets délétères des mutations G719A, G719C, G719S ont été explorés à l'aide de quatre logiciels de bioinformatique : SIFT, PolyPhen-2, Mutation Taster et Project HOPE.

II.1 Prédiction de l'effet de la mutation sur la stabilité de la protéine par le logiciel I-Mutant 2.0 :

En tenant compte de la structure tertiaire de la protéine, l'effet stabilisant ou déstabilisant des mutations faux sens a été évalué en utilisant le logiciel I-Mutant 2.0 (<http://folding.biofold.org/i-mutant/i-mutant2.0.html>). I-Mutant 2.0 est basé sur une approche de type Support Vector Machine (SVM) qui estime la différence d'énergie libre (DDG) entre la protéine sauvage et mutée (Capriotti et al, 2005). Une valeur positive (DDG>0) implique une augmentation de la stabilité des protéines, tandis qu'une valeur négative (DDG<0) suggère une mutation déstabilisante. Figure (06)

Figure 5: Interface du logiciel I-Mutant2.0

II.2. Prédiction des effets des mutations par le logiciel Polymorphism Phenotyping-2 (PolyPhen-2) :

Il correspond à un logiciel gratuit qui a été réalisé par Shamil et VasilyRamensky (figure 7). Il prédit l'impact possible d'une substitution d'acide aminé sur la fonction protéique (Shamil et al., 2001) en se basant sur des règles empiriques faisant appel à 3 types d'informations issues des banques de données et de calcul :

- Les informations pour les substitutions connues à une position donnée, contenant les annotations de séquences.
- L'alignement multiple.

- Les informations structurales provenant de banque de données de structure (PDB).

L'algorithme de Polyphen a quatre principaux arguments d'entrées et un optionnel :

- L'identification SwissProt permettant de retrouver la séquence et les annotations.
- La position de la substitution sur la séquence protéique.
- L'acide aminé original.
- L'acide aminé variant.
- La séquence protéique. Cette donnée est optionnelle car PolyPhen-2 à un module pour rechercher la séquence cible dans une banque de séquences protéiques.

Les prédictions que donne PolyPhen-2 sont : « error » (absence de score), « benign », « possiblydamaging » et « probablydamaging ». La prédiction dite « benign » signifie que la fonction de cette protéine ne semble pas être affectée par cette variation. A l'inverse, la prédiction dite « probablydamaging » signifie que la protéine a une forte probabilité d'être altérée par cette variation(Sediki.F.Z. 2015).



Figure 6 : Interface du logiciel Poly Phen-2

II.3. Prédiction de l'effet délétère des mutations exoniques sur la protéine par le logiciel Sorting Intolerant From Tolerant (SIFT) :

Ce logiciel est gratuitement accessible en ligne et s'appuie sur l'homologie de séquences. Il trie la substitution d'un acide aminé dit intolérant par rapport au tolérant et prédit l'effet phénotypique de cette substitution sur l'activité de la protéine (Pauline., 2003)

L'Algorithme de SIFT peut être décrit en trois phases :

- Recherche de séquences homologues, par rapport à la séquence cible. Cette recherche permet de construire un set de 5 séquences minimum.
- Le set de séquences obtenu est aligné par BLAST.
- Le calcul de probabilité : SIFT élimine les séquences 100% identiques à la séquence cible. Ainsi, il élimine la possibilité que la séquence protéique contenant la substitution étudiée soit déjà dans la base de données et soit donc prévu comme tolérée. La probabilité, pondérée en fonction du nombre de séquences alignées est calculée. La valeur seuil permettant la prédiction est 0,05 (Poch.Olivier.M.2012). En effet, les probabilités sont normalisées sur l'acide aminé consensus pour chaque position, qui est l'acide aminé avec le plus grand score.

En réalité, trois différents scores basés sur la similarité entre deux acides aminés sont obtenus :

- Intolérant (Intolérant): [0.00-0.05].
- Potentiallytolerant (potentiellement tolérant): [0.051-0.20].
- Tolérant (tolérant): [0.201-1.00].

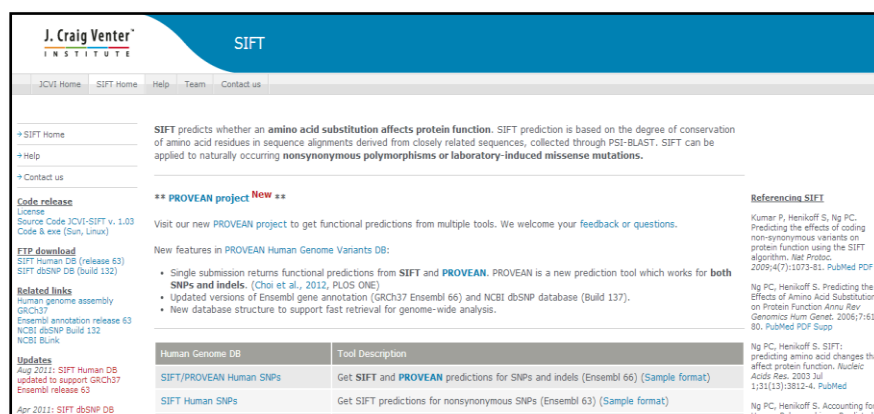


Figure 7: Interface du logiciel sift

II.4 Project HOPE (Have (y) Our ProteinExplained):

Project HOPE est un programme gratuit accessible via le lien suivant (<http://www.cmbi.ru.nl/hope/input?3>). Il analyse l'impact structurel et fonctionnel des mutations ponctuelles. Le système HOPE utilise toutes les informations collectées, combinant les propriétés connues des aa de type sauvage et mutant, telles que la taille, la charge et l'hydrophobicité, pour prédire l'effet des mutations sur la structure et la fonction des protéines. (Elshaikh, Ismaiel, Osman et Shockley, 2016).

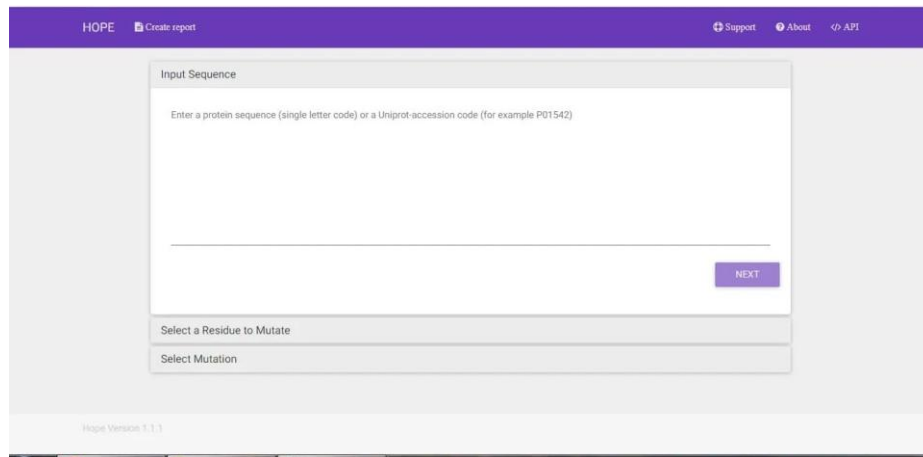


Figure 8 : Interface du logiciel Project Hope

Résultats
Et discussion

I- Résultats :

I.1. Résultat obtenu par le logiciel I-Mutant 2.0 :

D’après les analyses effectuées par logiciel I-Mutant 2.0 afin de tester les effets des mutations c.2156G>C, c.2155G>T et c.2155G>A sur la stabilité de protéine EGFR. Le résultat obtenu est représenté sur les figures (10), (11), (12)

```
*****
**                               **
**               I-Mutant v2.0    **
**   Predictor of Protein Stability Changes upon Mutations   **
**                               **
**                               **
SEQ File: fileseq.txt
*****
Position WT NEW Stability RI  pH  T
719      G  A  Decrease 7  7.0  25

WT: Aminoacid in Wild-Type Protein
NEW: New Aminoacid after Mutation
RI: Reliability Index
T: Temperature in Celsius degrees
pH: -log[H+]

*****
* Capriotti E, Fariselli P and Casadio R (2005). I-Mutant2.0: predicting *
* stability changes upon mutation from the protein sequence or structure. *
* Nucl. Acids Res. 33: W306-W310. *
* http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi *
*****
```

Figure 9: Résultat obtenu par le logiciel I-Mutant 2.0 pour la mutation c.2156G>C (p.Gly719Ala)

```
*****
**                               **
**               I-Mutant v2.0    **
**   Predictor of Protein Stability Changes upon Mutations   **
**                               **
**                               **
SEQ File: fileseq.txt
*****
Position WT NEW Stability RI  pH  T
719      G  C  Decrease 6  7.0  25

WT: Aminoacid in Wild-Type Protein
NEW: New Aminoacid after Mutation
RI: Reliability Index
T: Temperature in Celsius degrees
pH: -log[H+]

*****
* Capriotti E, Fariselli P and Casadio R (2005). I-Mutant2.0: predicting *
* stability changes upon mutation from the protein sequence or structure. *
* Nucl. Acids Res. 33: W306-W310. *
* http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi *
*****
```

Figure 10 : Résultat obtenu par le logiciel I-Mutant 2.0 pour la mutation c.2155G>T (p.Gly719Cys)

```

*****
** I-Mutant v2.0                               **
** Predictor of Protein Stability Changes upon Mutations **
*****
SEQ File: fileseq.txt
Position WT NEW Stability RI pH T
       719 G S Decrease 7 7.0 25
WT: Aminoacid in Wild-Type Protein
NEW: New Aminoacid after Mutation
RI: Reliability Index
T: Temperature in Celsius degrees
pH: -log[H+]
*****
* Capriotti E, Fariselli P and Casadio R (2005). I-Mutant2.0: predicting
* stability changes upon mutation from the protein sequence or structure.
* Nucl. Acids Res. 33: 4306-4310.
* http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi
*****
    
```

Figure 11 : Résultat obtenu par le logiciel I-Mutant 2.0 pour la mutation

Le principe de logiciel I-Mutant 2.0 est de calculer la différence d'énergie libre notée DDG entre la protéine sauvage et mutante. Les mutations testées sont déstabilisantes pour la structure de la protéine EGFR. Cela indique que la valeur DDG calculée est négative.

I. 2 Résultats obtenus par le logiciel POLYPHEN-2 :

Le logiciel POLYPHEN-2 nous a permis d'avoir une prédiction de l'impact de la mutation c.2156G>C, c.2155G>T et c.2155G>A sur la protéine EGFR. Le résultat obtenu est présenté sur les Figures (13)

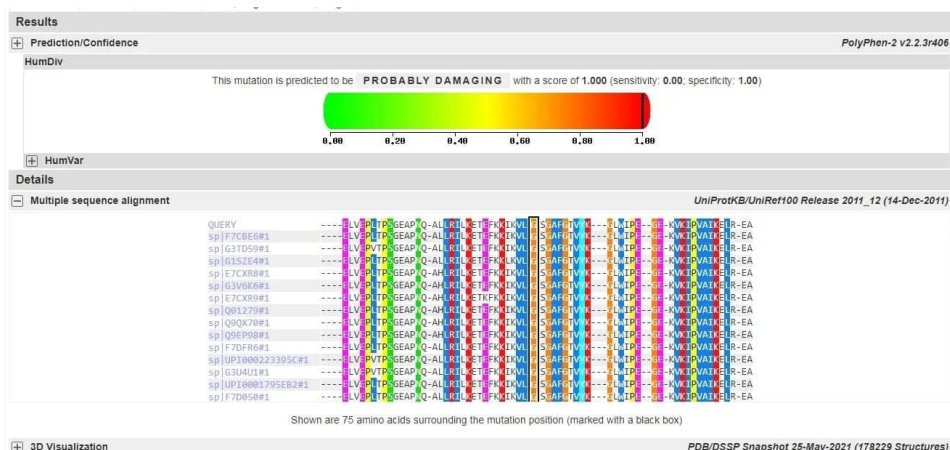


Figure 12 : Résultat obtenu par le logiciel PolyPhen-2 pour les mutations c.2155G>A (p.Gly719Ser) c.2156G>C (p.Gly719Ala) c.2155G>T (p.Gly719Cys).

Selon le résultat de logiciel PolyPhen -2 la mutation de G (Glycine à Alanine), G (glycine à cystéine) et G (glycine à sérine) sont probablement délétères (Probablydamaging) de

la structure protéique, cela indique que ces mutations ont un effet sur la structure et la fonction protéique.

D'autre part, ce même logiciel nous a permis d'obtenir le résultat d'alignement multiple par comparaison de la protéine EGFR avec toutes les espèces enregistrées sur UniProtKB. À partir de l'alignement multiple, nous avons trouvés que ces mutations siègent dans une région conservée entre plusieurs espèces. Les régions conservées sont considérées comme les régions les plus fonctionnelles des protéines.

I.3 Comparaison des propriétés physicochimiques entre les résidus normaux et mutés (Project HOPE) :

- La mutation c.2156G>C

La mutation c.2156G>C résulte de remplacement de la glycine en position 719 par aa l'alanine (G719A). Les structures de ces deux aa sont illustrées à la figure (14). Alors que les propriétés physicochimiques de chaque résidu sont résumées dans le tableau (II).

La figure ci-dessous montre les structures schématiques de l'acide aminé d'origine (la glycine) à gauche et du mutant (l'alanine) à droite. Le squelette, qui est le même pour chaque acide aminé, est coloré en rouge. La chaîne latérale, unique pour chaque acide aminé, est colorée en noir.



Figure 13: Les structures schématiques de la Glycine et l'alanine

Tableau II: Les propriétés physicochimiques de deux aa la glycine et l'alanine

Résidus	La taille	La charge
La Glycine	75Da	Neutre
L'alanine	89 Da	Twisterions(peuvent contenir des charges positives et négatives)

Chaque acide aminé a sa propre taille, sa charge et sa valeur d'hydrophobicité. Le résidu original de type sauvage et le résidu mutant nouvellement introduit diffèrent souvent par ces propriétés. Le résidu mutant (l'alanine) est plus grand que le résidu de type sauvage (Glycine). Le résidu mutant est plus hydrophobe que le résidu de type sauvage. La différence de propriétés entre le type sauvage et la mutation peut facilement entraîner une perte d'interactions avec le ligand parce que la liaison au ligand est souvent importante pour la fonction de la protéine, cette fonction pourrait être perturbée par cette mutation.

-La mutation c.2155G>T

Les analyse des effets structurels de la mutation c.2155G>T (p.Gly719Cys) dans la séquence protéique EGFR. Qui responsable de la substitution de glycine par cystéine, le résultat présenter dans la figure (15) par logiciel.



Figure 14 : Les structures schématiques de la cystéine et glycine.

Tableau III: les propriétés physique chimique des deux AA cystéine et glycine.

Résidu	Taille	Charge
cystéine	121Da	Neutre polaire
Glycine	75Da	Neutre non polaire

Les structures de ces deux acides aminés sont représentées dans la figure. Par logiciel Project Hope, tandis que les propriétés physicochimiques de chaque résidu ce soit leur taille ainsi que leur charge et leur capacité de former des liaisons avec d'autres acides aminés talques des jonctions di sulfure pour la cystéine.Ces différences sont la raison du changement de conformations mais aussi de fonction lorsqu'ils touchent des sites importants.

-La mutation c.2155G>A

La mutation c.2155G>A (p.Gly719Ser) est responsable de la substitution de l'aaGlycinepar l'aa sérine. La figure ci-dessous montre les structures schématiques de l'acide aminé d'origine (à gauche) et du mutant (à droite). Le squelette, qui est le même pour chaque acide aminé, est coloré en rouge. La chaîne latérale, unique pour chaque acide aminé, est colorée en noir.



Figure 15 : Les structures schématiques de glycine et Sérine

Tandis que les propriétés physicochimiques de chaque résidu sont résumées dans le tableau :

Tableau IV: les propriétés physique chimique des deux AA Sérine et glycine

Résidus	La taille	La charge
La Glycine	75Da	Neutre
Sérine	105,09 Da	Polaires et non chargés

Tableau V: L'hydrophobicité des acides aminés

Les acides aminés	Alanine	Cystéine	Serine	Glycine
L'hydrophobicité	hydrophobe	hydrophobe	Hydrophile	hydrophobe

I.6 Résultats obtenus par SIFT

Le résultat obtenu par le logiciel SIFT est représenté sur la figure(17) nous avons obtenu le même résultat pour les trois mutations, Ce résultat est présenté sous forme de deux colonnes. La colonne de gauche regroupe les aa qui sont prédits intolérants pour une position donnée. A l'inverse, celle de droite regroupe les aa qui sont prédits tolérants. Les aa sont colorés en fonction de leurs propriétés physico-chimiques (tableau V).

Predict Not Tolerated	Position	Seq Rep	Predict Tolerated
wcfmyivhl	701Q	0.88	gdntaePrSKQ
wyhfncmqdkp	702A	0.88	gsltiEVA
	703L	0.88	cwpdmekQngrtsivaHLFY
dghwnyresckqpt	704L	0.88	fivAML
wfmcyihvdlenq	705R	0.88	tgakPSR
hwqnprcgeksyfnta	706I	0.88	ldVI
cwngdhrska	707L	0.88	tyPaEmvFiL
cmfiyvlhp	708K	0.88	taqsrGWneDK
ywvtsrqpnmlkihgfedca	709E	0.88	E
hcrqmeqdykfpn	710T	0.88	liWasVT
cwmfyihlvrrntgskaq	711E	0.88	dPE
dgcwnehspkqyta	712F	0.88	mRiVFL
cwfmvvhgdln	713K	0.88	taqSPERK
cwdfmiyvgpshnlataeq	714K	0.88	RK
hwqngprcskyE	715I	0.88	fmDtaLIV
cwdfmiyvgpshnlataeq	716K	0.88	TRK
hwqdnpercgskytatfm	717V	0.88	LIV
ywvtsrqpnmlkihgfedca	718L	0.88	L
ywvtsrqpnmlkihgfedca	719G	0.88	G
whyfmrqndekcilpt	720S	0.88	atQS
ywvtsrqpnmlkihgfedca	721G	0.88	G
whyfmrqndekcilpt	722A	0.88	gVSA
ywvtsrqpnmlkihgfedca	723F	0.88	F
ywvtsrqpnmlkihgfedca	724G	0.88	G
whycferqdgkpnivla	725T	0.88	sMT
hwdngarvksqpcfntal	726V	0.88	IV

Figure 16: Résultat obtenu par le logiciel SIFT pour la mutation

Tableau VI : Les différentes propriétés physico-chimiques correspondant aux couleurs des aa obtenus par SIFT

Couleurs des aa	Propriété physico-chimique
Rouge	Basique
Vert	Polaire non chargé
Noire	Non polaire
Bleu	Acide

II. Discussion

Les progrès de la bioinformatique ont permis d'accélérer le taux de découverte et de réduire le besoin d'expériences de laboratoire coûteuses ou d'essais cliniques. L'évolution de cette discipline est actuellement indispensable au développement de la biologie moléculaire (Luscombe et al, 2001).

Dans cette étude, nous nous sommes intéressées à l'étude *in silico* des fonctionnalités des mutations du gène EGFR au niveau protéique par l'utilisation de différents logiciels de prédiction.

Les effets délétères des mutations sur la structure/fonction de la protéine EGFR ont été prédits en utilisant plusieurs logiciels de prédiction : I-Mutant 2.0, SIFT, Polyphen-2, Project HOPE.

Les résultats obtenus par le logiciel I-Mutant 2.0 sont en faveur d'un effet déstabilisant des trois mutations sur la protéine. De même, les logiciels (SIFT, Polyphen2) ont prédit des effets délétères de ces mutations sur la fonction de la protéine. En effet, ces mutations siègent dans une région conservée entre plusieurs espèces. Elle est considérée comme délétère et responsable de l'apparition des maladies. Elles ont un impact sur la fonction protéique car elles se trouvent dans une région importante de la protéine plus exactement au niveau de : domaine tyrosine kinase (Cartot et al. 2007) qui est codé par les exons 18 à 24 jouant un rôle important dans la transmission des signaux qui contrôlent le cycle cellulaire, la migration, l'adhésion, le métabolisme, la différenciation, la prolifération. De ce fait, ces mutations inhibent d'une manière réversible l'activité kinase de l'EGFR, empêchant ainsi l'autophosphorylation des résidus de tyrosine associés au récepteur et accentuant ainsi l'inhibition de la signalisation en aval qui favorise la survie et la prolifération des cellules.

La présence d'une mutation EGFR est un facteur de bon pronostic et un facteur prédictif de la réponse aux inhibiteurs de tyrosine kinase (ITK). Cependant, ce paramètre ne permet pas de prédire le bénéfice en termes de survie lié autrement par ITK.

Les informations concernant les propriétés physico-chimiques des aa suite à une mutation faux-sens ont été prédites par Project HOPE. Concernant la taille du résidu muté des trois variations est plus grande que le résidu sauvage. Cette différence pourrait engendrer un

changement de conformation de la protéine ce qui pourrait entraîner un changement au niveau de la structure et la fonction de la protéine.

De même, la charge et hydrophobicité sont différentes, Ces différences pourraient perturber la protéine entraînant par exemple la création des ponts disulfures (deux groupements thiols sont susceptibles de s'oxyder pour former une liaison covalente Cys-S-S-Cys)dans le cas de mutation G719C(Sharma et al.2007).

Il est essentiel de combiner plusieurs outils de prédiction *in silico* afin d'avoir une meilleure appréciation des effets délétères des mutations et de les confronter avec ceux d'études fonctionnelles. Les résultats combinées d'études *in silico* et fonctionnelles auraient une retombée directe sur l'interprétation des analyses de diagnostic moléculaires et permettraient également d'améliorer la connaissance de la physiopathologie des cancers ce qui est essentiel au développement de stratégies thérapeutiques adaptées au patient.

Conclusion et perspectives

Dans ce travail, nous avons contribué à la compréhension des mécanismes moléculaires à l'origine du défaut de la protéine EGFR. Les effets délétères de trois mutations (.2155G>A ; c.2156G>C ; c.2155G>T.) ont été étudiés en utilisant d'une combinaison de logiciels d'étude *in silico*.

Les mutations faux sens ont été prédites comme délétères au niveau structural et fonctionnel de la protéine EGFR, et les changements physicochimiques entre les acides aminés natifs et mutés perturbent fortement la stabilité de la protéine. Ces mutations siègent au niveau du domaine qui est nécessaire à la protéine EGFR pour la transmission des signaux de prolifération et de survie cellulaires.

Ces mutations peuvent entraîner un changement de conformation de la protéine pRb qui aura un impact sur sa fonction

Cette étude a permis de démontrer l'efficacité de l'utilisation associée de plusieurs outils prédictifs afin d'évaluer les effets délétères des mutations au niveau protéique. Ce protocole pourra être utilisé pour la prédiction des effets délétères des mutations responsables de plusieurs maladies génétiques.

Dans les cancers, la compréhension des mécanismes moléculaires associés aux différents dysfonctionnements de la protéine EGFR est essentielle pour le diagnostic, le pronostic et au développement de nouvelles stratégies thérapeutiques. Les études *in silico* permettent d'évaluer les effets des mutations sur la structure et la fonction de la protéine. Il est toutefois intéressant de réaliser des études fonctionnelles afin de confirmer ces résultats.

*Références
bibliographiques*

-Abdi.M.(2014).contribution à l'étude moléculaire de l'hémophilie

-Acton, Q. A. (2013). EGFR Inhibitors—Advances in Research and Application: 2013

Edition:ScholarlyBrief: ScholarlyEditions.

-Aube, A., Cabarrocas, J., Neunlist, M., Doulay, F., Liblau, R., & Galmiche, J. (2004).Alteredintestinal motility and permeability in a transgenic mice model of glia disruption.Paperpresented at the Gastroenterology .

-Azria, D., & Spano, J. P. (2006). Thérapies ciblées en cancérologie: John Libbey Eurotext.

-Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A., & Wheeler, D. L.(2002). GenBank. Nucleic Acids Research, 30(1), 17-20.

-Blay, J. Y. (2010). Les cibles membranaires de la cellule tumorale - 1ere partie, John Libbey Eurotext.

-Blanc, G., Ogata, H., Robert, C., Audic, S., Suhre, K., Vestris, G., . . . Raoult, D. (2007). Reductive genome evolution from the mother of Rickettsia.

-Brambilla, E., W. D. Travis, T. V. Colby, B. Corrin, et Y. Shimosato.(2001). « The New World Health Organization Classification of Lung Tumours ». The European Respiratory Journal 18 (6): 1059-1068.

-Bronte, G., Rizzo, S., La Paglia, L., Adamo, V., Siragusa, S., Ficorella, C., . . Russo, A. (2010). Drive mutations and differential sensitivity to targeted therapies: a new approach to the treatment of lung adenocarcinoma. Cancer Treat Rev, 36 Suppl 3, S21-29. doi:10.1016/s0305-7372(10)70016-5

-CartotA.Role du Récepteur à l'EGF dans la survenue et le traitement des cancers bronchiques non à petites cellules.La lettre du pneumologue 2007.Vol 06

-Carlson, J. J., Garrison, L. P., Ramsey, S. D., & Veenstra, D. L. (2009). Epidermal growth factor receptor genomic variation in NSCLC patients receiving tyrosine kinase inhibitor therapy: a systematic review and meta-analysis.

- De La Motte Rouge, T., Valent, A., Ambrosetti, D., Vielh, P., & Lacroix, L. (2007). [Clinical and molecular predictors of response to EGFR tyrosine kinase inhibitors in non-small cell lung cancer]. *Ann Pathol*, 27(5), 353-363. doi:10.1016/s0242-6498(07)78274-3
- Durdux C., Bibault J.-E, Jaffré F., Principes et techniques de la radiothérapie thoracique des carcinomes bronchiques non à petites cellules (CBNPC), Volume 6277, Issue 3, 10/2018, Pages 181-465, ISSN 1877-1203, [http://dx.doi.org/10.1016/S1877-1203\(18\)30025-9](http://dx.doi.org/10.1016/S1877-1203(18)30025-9)
- Elshaikh, A., Ismaiel, M. M., Osman, M. M., & Shokri, S. A. I. (2016). Computational analysis of single nucleotide polymorphism (SNPs) in human GRM4 gene. *American Journal of Biomedical Research*, 4(3), 61-73.
- Etienne, B., Denis, M.-S., Thierry Le, C., Gilbert, M., Jean-Yves, D., Serge, T., . . . Jean-Léon, L. (2001). Standards, Options et Recommandations pour la prise en charge thérapeutique des patients atteints d'un cancer bronchopulmonaire primitif non à petites cellules localement avancé. *Bulletin du Cancer*, 88(4), 369-387.
- Faure S., Thérapies ciblées anticancéreuses. *Actualité pharmaceutiques* 2015.57-61
- Haley, J. D., & Gullick, W. J. (2009). *EGFR Signaling Networks in Cancer Therapy*: Humana Press.
- Herbst, Roy S., John V. Heymach, et Scott M. Lippman. « Lung Cancer ». (2008). *The New England Journal of Medicine* 359 (13): 1367-1380. doi:10.1056/NEJMra0802714.
- Glass, H. C., Glidden, D., Jeremy, R. J., Barkovich, A. J., Ferriero, D. M., & Miller, S. P. (2009). Clinical neonatal seizures are independently associated with outcome in infants at risk for hypoxic-ischemic brain injury. *The Journal of pediatrics*, 155(3), 318-323.
- Gazdar, A. F. (2009). Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene*, 28 Suppl 1(Suppl 1), S24- S31. doi:10.1038/onc.2009.198
- Gilles C. Les acides aminés. 2016. Planet et vie.

- Lacave R, Larsen JC, Robert J. *Cancérogénologie fondamentale*. John Libbey Eurotext 2005.
- Mazieres Julien, Ritmier A, Gadgeel S, Toyooki H, et al. Atezolizumab Versus Docetaxel in pretreated patients with NSCLC : Final results from the randomized Phase 2 POPLAR and phase 3 OAK clinical trials. *Journal of Thoracic oncology* 2020, Volume 16
- Meert, A.-P., Martin, B., Delmotte, P., Berghmans, T., Lafitte, J.-J., Mascaux, C., . . . Sculier, J.-P. (2002). The role of EGF-R expression on patient survival in lung cancer: a systematic review with meta-analysis. *European Respiratory Journal*, 20(4), 975-981. doi:10.1183/09031936.02.00296502
- Mitsudomi, T., & Yatabe, Y. (2010). Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. *FEBS J*, 277(2), 301-308. doi:10.1111/j.1742-4658.2009.07448.x
- Pierre, H., & Sandrine, F. (2007). *Le ciblage des récepteurs de la famille ERB-B/HER*: John Libbey Eurotext.
- Poch, Olivier. M. (2012). Développement d'une infrastructure d'analyse multi-niveaux pour la découverte des relations entre génotype et phénotype dans les maladies génétiques humaines. Thèse de doctorat. ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ. Université de Strasbourg ; Allemagne.
- Roger Lacave, C.-J.L. (2005). *Cancérogénologie fondamentale* (Vol. 1). John Libbey Eurotext.
- Raymond, É. (2009). *Le concept de cible en cancérologie*, John Libbey Eurotext.
- Reiter, J. L., Threadgill, D. W., Eley, G. D., Strunk, K. E., Danielsen, A. J., Sinclair, C. S., Maihle, N. J. (2001). Comparative genomic sequence analysis and isolation of human and mouse alternative EGFR transcripts encoding truncated receptor isoforms. *Genomics*, 71(1), 1-20. doi:10.1006/geno.2000.6341
- Rothschild, Sacha, Daniel, B., C. Betticher, Adrian Ochsenbein, Rolf Stahel, Lukas Bubendorfer, Mathias Gugger, Martin Brutsch, Miklos Plesch, et Oliver Gautschi. (2010). « Importance de l'histologie pour le traitement du carcinome bronchique non à petites cellules à un stade avancé », 384–388.

-Robert Jacques. Cent ans d'innovation diagnostiques et thérapeutiques en Cancérologie : Centenaire de la société française du cancer.2006.John Libbey Eurotext.138p

-Rotow J. Cheema.P, Pisapia P, TronconeG.FastFcats : EGFR Exon 20 insertionmutations in NSCLC.2022.Karger Medical and scientific Publishers.60p

-Sediki.F.Z.(2015).étude génétique de la mucoviscidose dans un échantillon de la population Algérienne. Université d'Oran

-Stoesser, G., Sterk, P., Tuli, M. A., Stoehr, P. J., & Cameron, G. N. (1997). The EMBL nucleotide sequence database. Nucleic Acids Research, 25(1), 7-13.

-Shaema SV, Bell DW, Settleman J, HavberDA.Epidermal Growth factor receptor mutations in Lung cancer .Nat Rev Cancer.2007 ; 7/ 169-181.

Résumé

Le gène *EGFR* est un gène suppresseur de tumeur et son altération moléculaire peut augmenter le risque de cancer bronchique. Afin de contribuer à l'étude moléculaire, nous nous sommes intéressés à l'étude des effets délétères de trois mutations du gène *EGFR*, à savoir la mutation c.2155G>A, la mutation c.2156G>C et c.2155G>T.

Ainsi, les effets délétères des trios mutations ont été prédites en utilisant un protocole d'étude *in silico* constituée de différents logiciels (I-Mutant 2.0 ; SIFT ; Polyphen-2 ; Project HOPE).

Les mutations c.2155G>A,c.2156G>C,c.2155G>T sont prédites comme étant délétères altérant la protéine *EGFR*. En fait, les changements physico-chimiques entre les acides aminés sauvages et mutés peuvent perturber la stabilité des protéines et entraîner une perte d'interactions avec d'autres molécules. Nos résultats montrent que ces différents logiciels sont cohérents et que leur combinaison peut améliorer les performances de prédiction des effets de ces mutations.

Cette étude nous permet de mieux comprendre l'impact des mutations et met en lumière l'importance de la recherche prédictive dans les logiciels de bioinformatique.

Mots clés .Protéine *EGFR*, Mutation, effet, *in silico*, Bioinformatique