

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE SAAD DAHLAB, BLIDA -1-
FACULTE DES SCIENCES DE LA NATURE ET DE LA VIE
DEPARTEMENT DE BIOLOGIE ET PHYSIOLOGIE CELLULAIRE



Mémoire de fin d'études

En vue de l'obtention du diplôme de Master

En Sciences de la Nature et de la Vie

Filière : SCIENCES BIOLOGIQUES

Option : Génétique

THÈME

Etude bio-informatique et génétique du diagnostic de quelques types de cancer des muqueuses chez les adultes

Présenté par:

Soualah.Mohamed Amine

Soutenu le : 06/07/2020

Devant le jury

Mme HAMZI W.

MAA

USDB 1

Présidente

Mr R.MOHMED SAID.

MCA

USDB 1

Examineur

Mme ABDUL HUSSAINAS

MCA

USDB 1

Promotrice

Promotion : 2019 – 2020.

Remerciements

Mes vifs remerciements sont d'abord adressés à Mme ABDUL HUSSAIN. A. S qui m'a fait l'honneur d'encadrer ce travail. Je tiens à lui exprimer ma gratitude et mon profond respect.

Je tiens également à remercier les membres de Jury :

Mme HAMZI.W de m'avoir honoré de présider le jury de la soutenance.

Mr R. MOHAMED SAID d'avoir aimablement accepté d'examiner ce travail.

Dédicace

Je dédie ce travail :

*A mes parents qui m'ont soutenu et encouragé durant ces années
d'études*

Qu'ils trouvent ici le témoignage de ma profonde reconnaissance

A mes frères et sœurs, source de joie et de bonheur.

A tous mes amis

Liste des figures

<i>Figure 1 : Types des mutations génétiques</i>	3
<i>Figure2 : La localisation de gène p53.....</i>	13
<i>Figure3 : Répartition des mutations le long de la molécule p53.....</i>	13
<i>Figure4 : Structure génomique de TP53 et structure de la protéine p53 correspondante.....</i>	14
<i>Figure5 : Localisation de gène BRCA</i>	15
<i>Figure 6 : Utilisation de GenBank.....</i>	22
<i>Figure 7 : Tableau de bord de la requête sur NCBI</i>	23
<i>Figure 8: Utilisation de BLAST sur la plate-forme NCBI.....</i>	24
<i>Figure 9 : Utilisation d'Ensemble.....</i>	25
<i>Figure 10 : Utilisation d'ORF Finder</i>	26
<i>Figure11 : Utilisation de Swiss-model</i>	27
<i>Figure ;12 Structure cristalline d'un tétramère P53 humain multidomaine lié à l'élément de réponse P53 naturel.....</i>	29
<i>Figure13: la mutation de la base C/G 523 a détecté chez une des séquences analysées</i>	30
<i>Figure14: Traduction de la séquence mutée.....</i>	31
<i>Figure15 : Comparaison de séquence nucléotidique avec angine.....</i>	35
<i>Figure16 ;Carte d' information d'alignement portique de p53 sur blaste.....</i>	36
<i>Figure17: Séquence protéique du gène P53.....</i>	37
<i>Figure18: Structure 3D de la protéine p53.....</i>	38
<i>Figure 19: Structure de domaine de liaison à l'ADN de p53.....</i>	39
<i>Figure 20 : Représentation de 4 p53 liée à à deux motifs consécutifs de demi-sitesd'ADN.....</i>	39
<i>Figure 21: Représentation graphique circulaire de la relation phénotype génotype de la p53..</i>	40

Liste des tableaux

<i>Tableau I : les types des mutations et leur action.....</i>	<i>13</i>
<i>Tableau II: les Principaux types de cancers.....</i>	<i>6</i>
<i>TableauIII: les différents types de cancers selon le tissu dont ils sont originaires.....</i>	<i>9</i>
<i>TableauIV: Séquences étudiées.....</i>	<i>21</i>
<i>TableauV: Résultats de BLAST de gène p53.....</i>	<i>30</i>
<i>TableauVI : Les meilleurs résultats pBLAST pour la séquence p53 dans le génome. d'Homo sapiens</i>	<i>31</i>
<i>Tableau VII: Echantillon de mutation détectée après l'utilisation de BLAST p53(personnelle2020).....</i>	<i>34</i>
<i>TableauVIII :Les résultats de l'Ensemble se présentant comme dans le tableau.....</i>	<i>41</i>

Sommaire

<i>Introduction</i>	1
<i>Partie théorique</i>	2
<i>Chapitre I :Généralités</i>	2
<i>I.1 Définition</i>	2
<i>I.2 Gènes et cancer :</i>	2
<i>I.3 Mutation et Cancer</i>	2
<i>I.3.1Types des mutations génétiques</i>	3
<i>I.4. La carcinogenèse</i>	3
<i>.4.1Carcinogènes chimique</i>	4
<i>I.4.2. Les virus à cancer</i>	4
<i>I.4.3Les carcinogènes physiques</i>	4
<i>I.5 Types de gène cancer</i>	5
<i>I.6. Principaux types de cancer</i>	6
<i>I.7 Cancer broncho-pulmonaire</i>	7
<i>I.7.1Facteurs de risque</i>	7
<i>I.7.2Principes de traitement</i>	7
<i>I.8 Cancer du sein</i>	8
<i>I.8.1 Facteurs de risque du cancer du sein</i>	8
<i>I.8.2 Principe de traitement</i>	8
<i>9. Le cancer colorectal</i>	9
<i>I.9.1. Facteurs de risque du cancer colorectal</i>	9
<i>I.9.2. Principe de traitement</i>	9
<i>I.10. Classification histologique</i>	10
II.Épidémiologie	11
<i>II.1. Épidémiologie dans le monde</i>	11
<i>II.2. Épidémiologie dans l'Algérie</i>	12
III. génétique moléculaire	14
<i>III.1. Le gène p53</i>	14
<i>III.2. Structure localisation de gène p53</i>	14

III.3. <i>Le gène BRCA</i>	16
III .4. <i>La structure et localisation de gène BRCA</i>	16
IV. <i>Etude bio-informatique</i>	18
IV.1. <i>Historique</i>	18
IV 2. <i>Description de quelques outils bio-informatiques utilisés dans les recherches biomédicales</i>	15
Chapitre II : I Matériels et méthodes	22
I.1. <i>Matériels</i>	22
I.2. <i>Séquence du gène p53</i>	23
I.4. <i>methode d'utilisation Différents outils et logiciels bio-informatique</i>	23
Chapitre III : Résultats et discussion	30
Chapitre III : Conclusion	46
Références bibliographiques	43

Liste des abréviations

ADN :acide désoxyribonucléique

ARN :AcideRiboNucléique

BLAST :Basic Local AlignmentSearchTool

BPDE :benzo[a]pyrène-(trans)-7,8-dihydrodiol-9,10-époxyde

BRCA :breast cancer

DDBJ :DNA Data Bank of Japan

EBI :European Bioinformatics Institute

EBV :Epstein–Barr virus

EMBL :European Molecular Biologie Laboratory

ID: Identifiant

kDa :kilodalton

NCBI :National Center for Biotechnology Information

NIH :NationalInstitutue of Health

NMD :Nonsense-Mediated mRNA Decay

NMD :Nonsense-Mediated mRNA Decay

OMIM :OnlineMendelian Inheritance in ManCOSMIC

ORF :OpenRedingFram

P53 : protein53

PC :poste de commande

SV-40 :simianvacuolating virus 40

TP53 : tumorprotein p53 (or p53).

Résumé

Les études conduites au cours de ces 30 dernières années ont constitués un apport majeur pour la compréhension de la transformation tumorale.

Plus de 70 gènes ont été identifiés et une grande partie d'entre eux font l'objet de tests génétiques. L'indication des tests, la prise en charge en aval des sujets à risque doivent être réfléchies de façon collégiale. Les efforts sont à poursuivre pour identifier les facteurs génétiques et non génétiques modificateurs des risques tumoraux

Le cancer Broncho-pulmonaires, le cancer de sein et le cancer colorectal représentent les trois premiers cancers dans le monde. Ces types de cancer touchent plus d'un tiers de la population ; De nombreuses mutations génétiques sont la principale cause de ces types de cancer et les plus connues sont : la P53 et la BRCA

Notre recherche est basée sur la localisation des mutations de nos deux gènes au niveau de leur séquence à l'aide des outils bio-informatiques.

Ainsi les mutations les plus connues en Algérie concernant le gène p53 avec le changement : C523G et CGC175GGC le nucléotide C qui était à la 523ème position sur le gène a été remplacé par le nucléotide A qui provoque un changement de la substitution des acides aminés.

Mots clé :

Cancer, mutation, gène, bio-informatique, séquence

Abstract

The studies conducted over the past 30 years have been a major contribution in understanding tumor transformation.

Over 70 genes have been identified and many of them are undergoing genetic testing. The indication for the tests and the downstream management of subjects at risk must be considered in a collegial manner. Efforts should continue to identify the genetic and non-genetic factors that modify tumor risks

The Broncho pulmonary cancer and breast cancer and Colorectal cancer are the top three cancers in the world. These types of cancer affect more than a third population; Many genetic mutations are the main cause of these cancers and are the more known worldwide: the P53 and BRCA gene.

Our research is based on the location of mutations at the nucleotide sequence level using bioinformatics tools, we found that the mutations are located on several sites distributed on the nucleotide chain.

the most known mutation in Algeria are p53: C523G CGC175GGC the nucleotide C which was at the 523rd position on the gene has been replaced by the nucleotide A which causes a change in the substitution of amino acids.

Keywords:

Cancer mutation bio-informatique gènes

الملخص

ساهمت الدراسات التي أجريت على مدى السنوات الثلاثين الماضية بشكل كبير في فهم تحول الورم. تم تحديد أكثر من 70 جيناً والعديد منها يخضع للاختبارات الجينية التي يجب أن تؤخذ على محمل الجد. كما يجب أن تستمر الجهود لتحديد العوامل الوراثية وغير الوراثية التي تعدل مخاطر الورم. يمثل سرطان القصبات الرئوية وسرطان الثدي وسرطان القولون والمستقيم أكبر ثلاث سرطانات في العالم. تؤثر هذه الأنواع من السرطان على أكثر من ثلث السكان ؛ العديد من الطفرات الجينية هي السبب الرئيسي لهذه الأنواع من السرطان وأشهرها هي P53 و BRCA. يعتمد بحثنا على توطين الطفرات في جيناتنا على مستوى تسلسلها باستخدام أدوات المعلوماتية الحيوية. وبالتالي فإن أكثر الطفرات شهرة في الجائز تتعلق بجين p53 مع التغيير: C523G و CGC175GGC تم استبدال النوكليوتيد C الذي كان في المركز 523 على الجين بالنيوكليوتيد A الذي يسبب تغييراً في استبدال الأحماض الأمينية.

الكلمات الدالة:

السرطان ، الطفرات ، الجينات ، المعلوماتية الحيوية ، التسلسل

Introduction

Le cancer est une maladie génétique de la cellule. C'est un dogme dont les trentes dernières années de recherche n'ont cessé de vérifier (1)(2). En effet, le caractère tumoral d'une cellule est héritable et transmissible aux cellules filles par des modifications génétiques ou épigénétiques du génome. La transformation tumorale est un processus le plus souvent d'origine clonal, dans lequel la cellule acquière au fur et à mesure des divisions successives par rapport à ses congénères (survie, prolifération, indépendance aux facteurs de croissance, etc). Ces étapes sont acquises par des altérations très variées du génome (mutations ponctuelles, délétions, translocations, gains ou pertes de chromosomes, modifications épigénétiques). Ces altérations génétiques ont deux conséquences bien différentes. Elles peuvent conduire à l'augmentation de l'activité de certains gènes favorisant au sens large la croissance tumorale ; ces gènes sont appelés oncogènes. A l'inverse, elles peuvent inactiver d'autres gènes dont l'activité physiologique s'oppose à la transformation tumorale, d'où leur nom de gènes suppresseurs de tumeurs(3). Au niveau de la cellule tumorale, les oncogènes ont une action dominante : l'activation d'un seul allèle est généralement suffisante à sa contribution au phénotype tumoral ; à l'inverse, les gènes suppresseurs de tumeurs sont récessifs : l'inactivation des deux allèles est nécessaire. Il faut cependant nuancer ces notions, car d'une part la réduction d'activité de certains gènes suppresseurs, par exemple par haplo insuffisance, suffit parfois à conférer un avantage sélectif vers la malignité. D'autre part, certaines mutations de gènes suppresseurs de tumeurs ont une action dominante, tout en inhibant la fonction du gène (action dominante négative). L'exemple le plus classique est le gène TP53 dont des mutations inactivatrices conduisent à stabiliser la protéine mutante et inhiber de façon dominante le complexe transcriptionnel tétramérique P53 (4).

Chapitre I :Généralités**I .1 Définition du cancer**

Le cancer est une maladie caractérisée par une prolifération cellulaire anormale au sein d'un tissu normal de l'organisme. Ces cellules dérivent toutes d'un même clone, la cellule initiatrice du cancer qui a acquis certaines caractéristiques lui permettant de se diviser indéfiniment et de pouvoir former des métastases.

Le cancer est un terme général pour n'importe quelle maladie pour lesquelles certaines cellules du corps humain se divisent d'une manière incontrôlée. Les nouvelles cellules résultantes peuvent former une tumeur maligne (un néoplasme) ou se propager à travers le corps et former des métastases.(5)

I.2Gènes et cancer

Tous les cancers sont causés par un changement qui survient dans des gènes ou par une lésion des gènes. Les gènes sont présents dans chaque cellule du corps. Les gènes de chaque cellule indiquent à celle-ci comment se développer, fonctionner, se diviser et mourir. Les gènes sont des particules d'ADN présentes à l'intérieur des chromosomes dans toutes nos cellules. Lorsqu'un gène change ou subit une mutation (ce qu'on appelle une mutation génétique), les instructions qu'il donne à la cellule peuvent empêcher celle-ci d'agir correctement. Cette mutation peut entraîner un développement anormal du corps ou causer un trouble.(6)

I .3 Mutation et Cancer :

Le cancer est le résultat de la panne des contrôles qui régulent les cellules. Les causes de la panne incluent toujours des changements dans des gènes importants. Ces changements

sont souvent le résultat de mutations, de changements dans la séquence d'ADN des chromosomes.

Les mutations peuvent être de très petits changements, n'affectant que quelques nucléotides ou elles peuvent être très importantes, entraînant des changements majeurs dans la structure des chromosomes .Les petites et les grandes mutations peuvent affecter le comportement des cellules. Des combinaisons de mutations dans des gènes importants peuvent conduire au développement d'un cancer.(7)

I .3.1. Types des mutations génétiques

le tableau suivant représente les types des mutations génétiques:

Tableau I : les types de mutation et leur action.

Mutation	Action
L'insertion	se produit lorsqu'un fragment supplémentaire d'ADN est ajouté à un gène.(8)
La duplication	se produit lorsqu'une partie de l'ADN d'un gèneest copiée une ou plusieurs fois.(9)
La délétion	se produit quand un fragment d'ADN est enlevé d'un gène, lorsqu'il manque un gène complet.(10.)
La translocation	signifie qu'un fragment d'un chromosome se casse et se fixe à un autre chromosome.(11)
L'inversion	se produit quand un chromosome se casse à 2 endroits et qu'un fragment d'ADN est inversé et réinséré dans le chromosome(12)

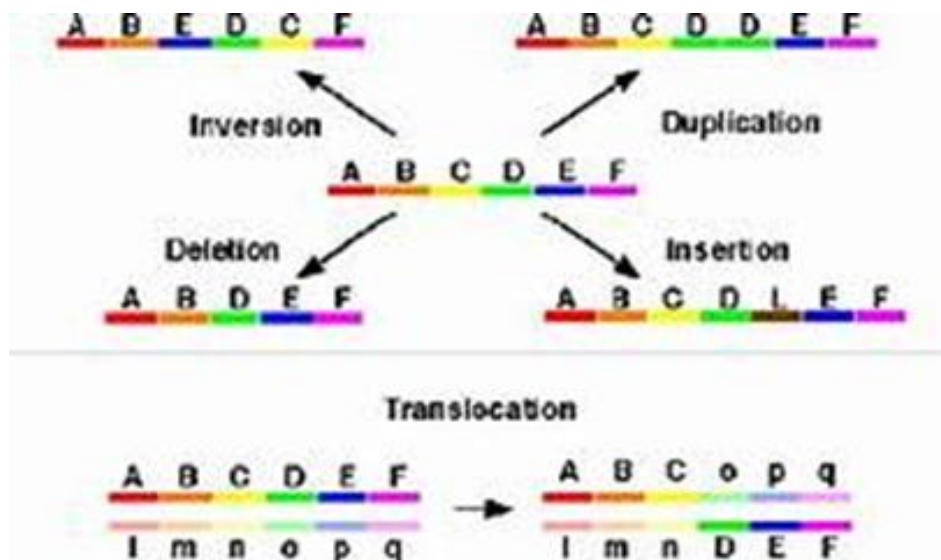


Figure 1 : Types des mutations génétiques.(8)

I .4. La carcinogenèse :

I .4.1Carcinogèneschimiques :

Les carcinogènes chimiques humains connus sont variés dans leur structure chimique et leur origine. Ils peuvent être issus des produits industriels ou être des produits naturels, et être présent dans la nourriture, l'air ou l'eau, mais ils peuvent aussi être endogènes quand il s'agit de certains produits du métabolisme (13).

I .4.2.Les virus à cancer :

Les rétrovirus ou virus à ARN sont à développement exclusivement cellulaire. Leur enzymologie particulière (transcriptase inverse) et leur lien avec les oncogènes font un sujet d'étude intéressant en cancérologie. De nombreux rétrovirus tumeur gènes possédant des gènes transformant v-onc, homologues de gènes cellulaires normaux c-onc, ont été identifiés. L'étude systématique des rétrovirus capables de transformer les cellules in vitro a permis d'identifier de nombreux oncogènes viraux correspondant chacun à la version un peu modifiée d'un (quelquefois plusieurs) oncogène(s) cellulaire(s). Leur origine animale ainsi que l'emplacement chromosomique de leur homologue humain sont connus. Les oncogènes viraux proviennent de la capture de gènes de l'hôte vertébré dans lequel le virus s'est répliqué. Les adénovirus ou virus à ADN, dont certains sont impliqués dans la genèse de certains cancers, comme le SV-40 et le virus du polyome, les papillomavirus, le virus EBV, le virus de l'hépatite B.(13)(14).

I .4.3Les carcinogènes physiques :

Deux types d'irradiation semblent favoriser le développement des cancers : les radiations ionisantes et les ultraviolets. Deux grands effets induits par les irradiations peuvent être observés : la mort cellulaire et la transformation maligne de la cellule non atteinte de façon létale, chacun de ces effets s'observant pour des doses et des circonstances variées.(15).

I .5 Types de gènes du cancer

Il y a 3 principaux types de gènes du cancer qui contrôlent la croissance cellulaire et qui peuvent causer l'apparition du cancer.

- **Les oncogènes**

Sont des gènes ayant mutés, qui incitent les cellules à croître de façon désordonnée et qui peuvent ainsi engendrer un cancer. Les proto-oncogènes sont des gènes normaux qui contrôlent la croissance cellulaire, mais qui peuvent se transformer en oncogènes s'ils subissent une mutation. Les proto-oncogènes et les oncogènes agissent comme des interrupteurs marche-arrêt. Un proto-oncogène est habituellement inactif (arrêt)(16).

- **Les gènes suppresseurs de tumeur**

Sont des gènes normaux qui ralentissent la croissance et la division des cellules, réparent les erreurs dans l'ADN et disent aux cellules quand mourir (un processus normal appelé apoptose ou mort cellulaire programmée). Ils contribuent à nous protéger contre le cancer.

Les gènes suppresseurs de tumeur fonctionnent correctement lorsqu'ils sont actifs. Ils empêchent les cellules de se diviser trop rapidement. Mais lorsque ces gènes subissent une mutation, ils deviennent inactifs. Cela incite les cellules à croître de façon désordonnée, ce qui peut engendrer un cancer. (17)(18)

Ex :p53

- **Les gènes de réparation de l'ADN :**

Réparent les erreurs qui peuvent se produire dans d'autres gènes lorsque l'ADN est copié. Quand les gènes de réparation de l'ADN subissent une mutation, ils ne peuvent plus réparer les erreurs dans les oncogènes et les gènes suppresseurs de tumeur, et cela peut mener au cancer.(19)

I .6. Principaux types de cancer :

Les cancers broncho-pulmonaires, du sein chez la femme et du côlon-rectum sont les trois principaux types de cancer en termes d'incidence et se classent parmi les cinq premiers en termes de mortalité (premier, cinquième et deuxième, respectivement).

Tableau II : les Principaux types de cancers(20)

Le type de cancer	Le nombre sur le monde	Pourcentage
* Les cancers broncho-pulmonaire. *cancer de sein.	On estime à environ 2,1 millions le nombre de diagnostics de chacun de ces cancers	qui représentent environ 11,6 % du fardeau total de l'incidence du cancer.
Le cancer colorectal	1,8 million de cas	10,2 % du total
le cancer de la prostate	1,3 million de cas	7,1 % du total
le cancer de l'estomac	1,0 million de cas	soit 5,7 % du total
Le cancer du poumon	1,8 million de décès	18,4 % du total

I.7Cancerbroncho-pulmonaire:

Le cancer broncho-pulmonaire 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 cellules de la muqueuse des voies respiratoires ou des alvéoles pulmonaires se multiplient de manière incontrôlée et forment des tumeurs malignes. De tous les types de cancer, le cancer du poumon est celui qui conduit le plus souvent à la mort.(21)

I .7.1 Les facteurs de risque :

Parmi les facteurs de risque des cancers broncho-pulmonaires, le tabagisme a un rôle prédominant, étant retrouvé dans 85% des cas chez l'homme et 60% des cas chez la femme (22).

L'évolution de l'incidence et de la mortalité du cancer broncho-pulmonaire reflète l'évolution de la consommation de tabac depuis l'après-guerre, avec une latence comprise entre 10 et 20 ans(23).

Néanmoins, seuls 10% à 15% des fumeurs développent un cancer du poumon, suggérant l'existence d'un polymorphisme génétique qui, à exposition égale, favorise la survenue de ce cancer (22).

I.7.2Principes de traitement :

Le traitement actuel des cancers broncho-pulmonaires non à petites cellules repose sur la chirurgie lorsque la tumeur est ré sécable (moins de 20% des cas), sur la radiothérapie en cas de tumeur localement avancée non ré sécable, et sur la chimiothérapie en cas de tumeur métastatique.

Malgré l'association séquentielle et/ou concomitante de ces modalités thérapeutiques et les optimisations techniques de ces procédures, le taux de survie global des cancers bronchiques à petites cellules est inférieur à 20% à 5 ans, soit du fait d'une résistance de la tumeur au traitement, soit du fait de récives locales et/ou systémiques précoces.

En cas de tumeur métastatique au diagnostic, la survie ne dépasse pas 15 mois avec les protocoles les plus récents de poly-chimiothérapie (24).

I.8 cancer du sein :

La notion de « cancer du sein » relève d'une nomenclature générique qui fait référence à tout un ensemble de proliférations néoplasiques de la glande mammaire qui diffèrent tant dutpoint de vue histologique qu'en ce qui concerne leur comportement évolutif.

Le terme de cancer du sein ne désigne que les tumeurs maligne potentiellement agressives ; du sein tandis que le terme de « tumeur du sein » design à la fois les tumeurs maligne et les tumeurs bénignes(25)

I .8.1 les facteurs de risque du cancer du sein :

Il existe suffisamment de preuves permettant d'affirmer que la susceptibilité génétique, l'exposition à des facteurs environnementaux et à des facteurs liés au style de vie jouent un rôle important dans l'étiologie de cette maladie Les facteurs de risque génétiques représentent 5 à 10 % des cas de cancer du sein (26).

Il existe d'autres facteurs associés à un risque élevé de cancer du sein : le tabagisme, certains aspects de la nutrition (consommation de viande et de graisse), absence d'activité physique et le stress psychologique (26).

I .8.2Principe de traitement :

Le traitement locorégional dans les premiers temps est l'acte chirurgical. Cet acte peut être radical (mastectomie) ou conservateur (tumorectomie) et est associé à un geste auxillaire homolatéral (procédure du ganglion sentinelle et/ou curage auxillaire).

Le traitement systémique (chimiothérapie et/ou hormonothérapie), dit « adjuvant » lorsqu'il est administré après le geste chirurgical, ou « néoadjuvant » lorsqu'il est administré en premier lieu, avant l'intervention chirurgicale.(27).

I .9. Le cancer colorectal :

Le cancer colorectal ou colo-rectal est une tumeur maligne de la muqueuse du côlon ou du rectum. Le cancer colorectal peut toucher tous les segments anatomiques du gros intestin comme le côlon ascendant, le côlon transverse, le côlon descendant, le côlon sigmoïde et le rectum mais ne concerne pas le cancer du canal anal qui est une entité distincte(28)

I .9.1. les facteurs de risque du cancer colorectal

L'âge :Le cancer colorectal touche surtout les personnes de plus de 50 ans. C'est pourquoi le dépistage organisé recommande un test tous les deux ans, de 50 à 74 ans

Les antécédents personnels ou familiaux de maladie du côlon :

Une maladie du côlon personnelle ou un antécédent familial augmente le risque de cancer du côlon.

Les maladies inflammatoires de l'intestin :

les polypes nouvellement formés au niveau du rectum ou du côlon ne sont pas cancéreux, en revanche, certaines maladies de l'intestin, comme notamment une polypose rectologique qui se manifeste par de très nombreux polypes, exposent à un risque très élevé de cancer colorectal.

la bactérie Fusobacteriumnucleatum: inhibe l'action des cellules immunitaires, ce qui pourrait faciliter la cancérogenèse(29)

I .9.2. Principe de traitement :

Le premier traitement est la chirurgie, qui permet d'enlever la tumeur et les ganglions environnants (curage ganglionnaire). Les cancers très superficiels sont parfois totalement résectionnés par voie endoscopique, sans chirurgie complémentaire nécessaire s'il n'y a pas franchissement de la muqueuse musculaire.

On y associe des traitements adjuvants : chimiothérapie, si le curage contient des métastases ganglionnaires ou s'il existe des métastases hépatiques ou pulmonaires ; radiothérapie si l'exérèse chirurgicale n'a pu être totale, avec reliquat tumoral individualisé sur un organe vital.(29).

I .10. Classification histologique

Il existe plusieurs types de cancers, qui sont déterminés en fonction de l'histologie, autrement dit la nature du tissu dans lequel ils se développent.

Voici un tableau qui illustre bien les différents types de cancers selon le tissu dont ils sont originaires :

Tableau III: Les différents types de cancers selon le tissu dont ils sont originaires (30)

Principaux types de cancers	Tissu d'origine de la tumeur	Fréquence	Localisations
Adénocarcinome	Épithélium (tissu de recouvrement des glandes).	85 % de tous les cancers.	Sein, foie, rein, prostate, ovaire, thyroïde, colon, estomac, glandes salivaires, poumon.
Carcinome épidermoïde	Épithélium malpighien (peau, muqueuses, épiderme).	85 % de tous les cancers.	Peau, voies digestives, poumon, sphère ORL (larynx, pharynx, cavité buccale), col utérin.
Sarcome	Tissu de soutien ou musculo-squelettique (os, muscles, tissu conjonctif ou graisseux...).	2 % de tous les cancers.	Os, cartilage, tissu graisseux, vaisseaux
Lymphome de Hodgkin	Lymphocytes B ou T, cancer caractérisé par la présence de grosses cellules atypiques	5 à 7 % de tous les cancers	Ganglions, rate
Lymphome non-Hodgkinien	Lymphocytes B ou T	5 à 7 % de tous les cancers.	Ganglions, voies digestives, peau, cerveau, os, organes génitaux, poumon.
Leucémie.	Cellules de la moelle osseuse (blastés).	4 % de tous les cancer	Sang
Myélome	Cellules de la moelle osseuse (plasmocytes)	4 % de tous les cancer.	Moelle osseuse

II. Épidémiologie

II.1. Épidémiologie dans le monde

II.1.1. Le cancer broncho-pulmonaire

-Le cancer broncho-pulmonaire représente la première cause de mortalité par le cancer dans le monde, avec plus de 1,2 million de décès annuels., et avec plus de 30 000 nouveaux cas par an, le cancer broncho-pulmonaire est le second cancer chez l'homme et le quatrième chez la femme(31).

Si l'incidence et la mortalité de ce cancer est actuellement 6 fois supérieures chez les hommes, il existe une augmentation continue de ces indicateurs chez les femmes : respectivement 4,5% et 2,9% par an depuis 1980 (contre respectivement 0,6% et 0,7% chez les hommes) (32).

Une accélération de cette tendance est actuellement observée chez les femmes jeunes, et les estimations indiquent que la mortalité par cancer broncho-pulmonaire sera triplée d'ici 2020, dépassant chez les femmes la mortalité des cancers du sein..

II.1.2. Le cancer du sein

Les projections de GLOBOCAN 2012 estiment une augmentation substantielle de 19,3 millions de nouveaux cas de cancer par an d'ici à 2025 en raison de la croissance démographique et du vieillissement de la population mondiale. Environ 1,7 million de femmes ont un diagnostic de cancer du sein chaque année et en 2012, 6,3 millions de femmes vivaient avec un cancer du sein diagnostiqué au cours des cinq années précédentes.(33).

La répartition du cancer du sein est très inégale d'un pays à l'autre, d'un continent à un autre comme le montre les données suivantes :

- * 27% Amérique du Nord.
- * 26% en Europe de l'Ouest.
- * 20% en Europe de l'Est.
- * 17% en Afrique du Sud – Est.

- * 15% en Afrique de l'Ouest.

- * 12% au Japon.

II.1.3. Le cancer colorectal

Le cancer colorectal est la troisième cause de mortalité dans le monde, après le cancer du poumon chez l'homme et du sein chez la femme, avec plus de 600000 cas de décès soit 8% de l'ensemble des décès. Géographiquement, il a été constaté une grande variation de son incidence à travers le monde dont près de 55 des cas surviennent dans les pays les plus développés.

Ainsi, les plus hauts taux sont observés en Australie / nouvelle- Zélande (44.8 et 32.2 par 100.000 habitants chez les hommes et les femmes, respectivement), et les plus faibles en Afrique de l'Ouest (4.5 et 3.8 pour 100.000 habitants)(34).

II.2. Épidémiologie dans l'Algérie

II.2.1 Cancer broncho-pulmonaires :

le cancer bronchique occupe la première place chez l'homme il représente des cancer masculins suivi du cancer colorectal et cancer de la vessie. les données statistique fournies par le registre de cancer concernant des cancer la région de Sétif ,estiment que la cancer broncho-pulmonaire représente 18% avec un taux d'incidence standardisé 15.5/100 000 le registre de la wilaya d'Oran incidence standardisé est de 105.4/100 000chez les personnes de sexe masculin et 111.5/100 000 de sexe féminin et dans la population de l'ouest algérien avec une pourcentage d'atteinte de 10.9 %avec un pourcentage égale 23. %de sexe masculin(35).

II.2.2. Cancer du sein

la modification démographique avec une augmentation de l'espérance de vie, la transformation de l'environnement et le changement de mode de vie sont a la faveur de l'apparition de nouvelles pathologies comme le cancer du sein est devenu un problème de santé publique majeur avec une réelle urgence d'intervention et de prise en charge.

Chaque année, 7500 cas de cancer du sein sont enregistrés avec environ 3500 décès enregistrés chaque année.

- 40,4% ont moins de 50 ans.

-
- 02% de femmes présentant un cancer du sein n'ont jamais eu d'enfant.
 - 10% de l'ensemble des cancers sont familiaux.
 - 70% ont subi l'amputation
 - 30% ont bénéficié de traitements conservateurs.(36)

II.2.3.Le cancer colorectal :

En Algérie, le cancer colorectal est au deuxième rang, pour les deux sexes. Chez l'homme, la prévalence est de 1180 cas soit 7,1% après le cancer du poumon. Chez la femme, la prévalence a été de 1082 soit 7,1% après le cancer du sein et du colutérin (37).

III. génétique moléculaire

III.1.Le gène Tp53

Le P53 est une protéine de gène suppresseur de tumeur(tp53) qui contrôle la croissance et la division des cellules. Le gène TP53 envoie aussi des signaux à d'autres gènes pour aider à réparer l'ADN endommagé.

Le gène Tp53 est situé sur le chromosome 17p13 et le domaine de transactivation se trouve dans la région N terminale à proximité du domaine de fixation séquence spécifique de l'ADN et du domaine d'oligomérisation C terminal. Les mutations du gène Tp53 correspondent aux anomalies génétiques les plus souvent observées dans les cancers et particulièrement dans les cancers pulmonaires.

Ces mutations concernent le plus souvent les exons 5 à 8 et concernent les codons 157, 245, 248 et 273, les codons 157, 248 et 273 étant le siège d'adduits BPDE au niveau de leurs sites CpG dans les cellules bronchiques en culture soumises aux carcinogènes du tabac comme le benzo(a)pyrène (38).

III.2. Structure localisation de gène p53 :

La protéine p53 est encodée par le gène TP53 qui est situé sur le petit bras du chromosome 17 chez l'humain. Ce gène est constitué de 20 000 paires de bases et comporte 11 exons. Une fois traduite, p53 est constituée de 393 acides aminés (a. a.) et a un poids moléculaire de 53kDa, dont elle tire d'ailleurs son nom (39).

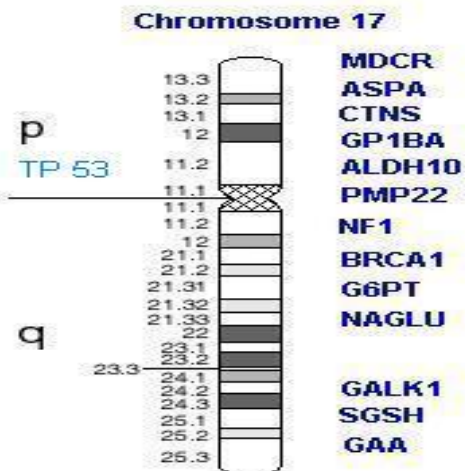


Figure 2: Localisation de gène p53 (Weston A. 1997 J)

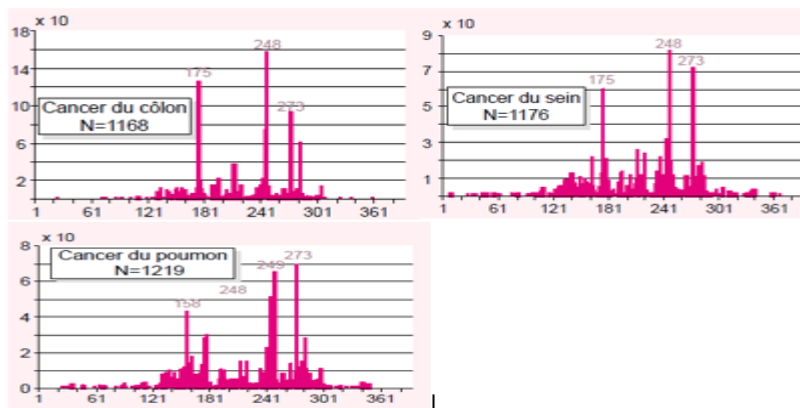


Figure 3 : Répartition des mutations le long de la molécule p53(40).

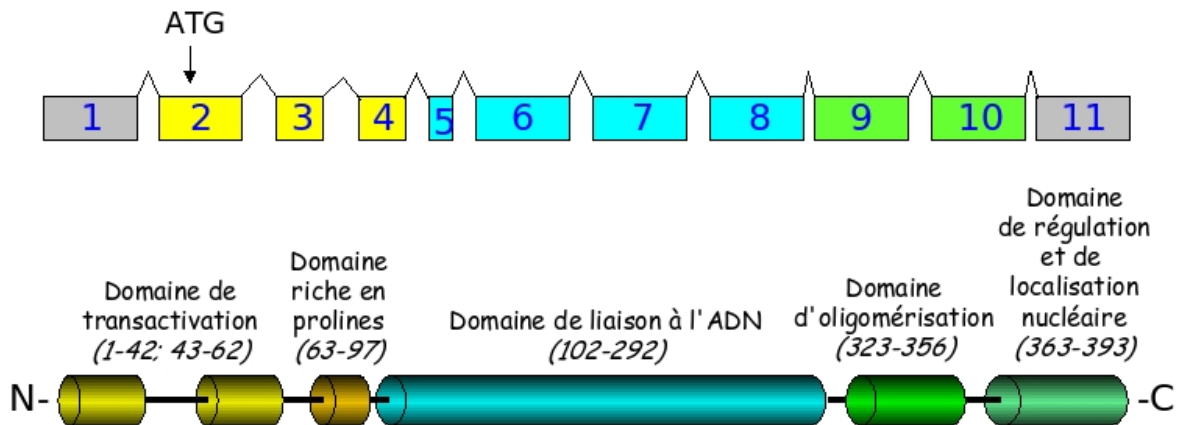


Figure 4 : Structure génomique de TP53 et structure de la protéine p53 correspondante(41)

III.3.Le gène BRCA :

Les gènes BRCA sont des gènes suppresseurs de tumeurs qui aident habituellement à prévenir l'apparition du cancer. Ils contrôlent la croissance et la division des cellules et aident à la réparation des dommages causés à l'ADN. Toutefois, des gènes BRCA qui ont subi une mutation peuvent accroître le risque d'apparition de certains types de cancer.

Il existe 2 mutations des gènes BRCA qui sont connues pour provoquer le cancer – la mutation du **gène BRCA1** et la **mutation du gène BRCA2**.

Ces mutations génétiques font augmenter le risque d'une femme d'être un jour atteinte d'un cancer du sein et d'un cancer de l'ovaire.

Les mutations du gène BRCA2 sont aussi liées à une hausse du risque de cancer du sein et du cancer de la prostate chez l'homme. Les mutations du gène BRCA2 rendent aussi l'homme et la femme un peu plus susceptibles d'être un jour atteints d'un cancer du pancréas(42).

III.4. La structure et localisation de gène BRCA 1

Le gène BRCA1 est situé sur le bras long (q) du chromosome 17 au niveau de la bande 21, entre les paires de base 38 449 843 et 38 530 933.

Ce gène composé de 22 exons code, par l'intermédiaire d'un ARN messager, une protéine formée de 1 863 acides aminés(42).

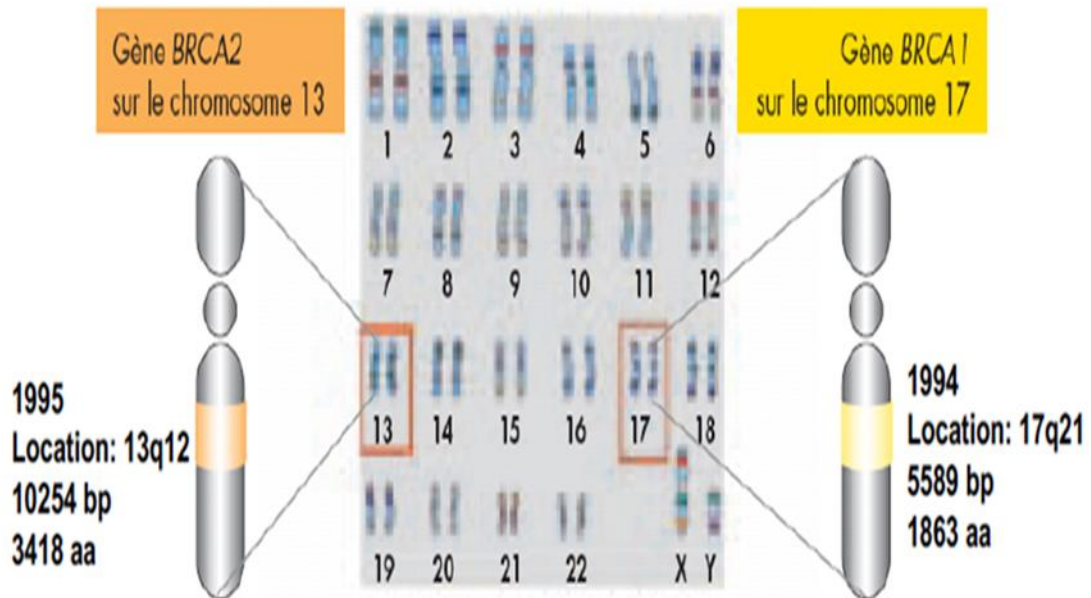


Figure 5: Représente la localisation de gène BRCA (42).

IV .Étude bio-informatique

IV.1.Historique :

La bio-informatique se définit actuellement comme un domaine de recherche qui analyse et interprète des données biologiques, au moyen de méthodes informatiques, afin de créer de nouvelles connaissances en Biologie.

En effet, la bio-informatique peut être utilisée pour créer et partager un large éventail de bases de données et de logiciels, ce qui peut constituer un outil scientifique

puissant, ce qui permet la comparaison, l'analyse et la visualisation des données biologiques provenant de l'étude de plusieurs organismes (43).

Pour ce nouveau domaine scientifique, tout a vraiment commencé dans les laboratoires publics au milieu des années 1980, lorsque le laboratoire européen de biologie moléculaire (EMBL : European Molecular Biologie Laboratory) et le département américain de la santé (NIH : National Institut of Health) ont créé les banques de données EMBL et GENBANK pour répertorier les séquences d'ADN découvert par les biologistes, quatre ans plus tard ,une nouvelle banque généralisée est apparue PIR (Protéine Information Ressource) en USA et SwissProt(constituée à l'université de Genève à partir de 1986) (44). En 1985, est eu la création d'un Programme d'alignement local de séquences "FASTA" En 1990, Programme d'alignement local de séquences "BLAST" (45) .

IV.2.Description de quelques outils bio-informatiques utilisés dans les recherches biomédicales :

Les variations de séquences au niveau des gènes sont le plus souvent responsables de troubles fonctionnels des protéines concernées menant à l'apparition de nombreuses maladies. 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 le processus d'épissage alternatif, ce qui se traduit par diverses conséquences sur la structure et/ou la fonction de la protéine.

L'étude in silico permet une meilleure compréhension de l'effet et de l'impact de différentes mutations sur le dysfonctionnement de la protéine.

De nombreux logiciels et bases de données sont été développés afin de prédire les différents effets et impacts de ces mutations délétères, par exemple.

- **GenBank :**

Base de données généraliste américaine, qui contient des séquences nucléotidiques disponibles au public pour 370 000 espèces formellement décrits, Elle est construite et distribué par le Centre national pour l'information biotechnologique (NCBI), cette banque de donnée américain travaille en collaboration avec la DNA.

-
- **Data Bank of Japon (DDBJ) :**EuropéenBio-informatiques Institute (EBI). (44)

- **BLAST : (Basic Local Aligement Search Tools) :**

Il Permet de trouver les régions similaires entre deux ou plusieurs séquences de nucléotides ou d'acides aminés et de réaliser un alignement de ces régions homologues. Ce programme permet de retrouver rapidement dans des bases de données, les séquences ayant des zones de similitude avec une séquence donnée (44)

- **OMIM : (Online Mendélien Inheritance in Man) :**

Est une base de données de référence pour les maladies génétiques, née dans les années 1960 grâce au travail de Victor McKusick qui est souvent surnommé "the Feather of médical génétiques" (45), est le principal dépôt d'informations complètes, dont le commissaire sur les gènes et les phénotypes génétiques et les relations entre eux, un aperçu des caractéristiques cliniques de phénotypes(46).

- **Swiss model :**

Est un serveur Web de bio-informatique structurale dédié à la modélisation d'homologie de structures protéiques 3D. La modélisation par homologie est actuellement la méthode la plus précise pour générer des modèles de structure protéique tridimensionnels fiables. Elle est couramment utilisée dans de nombreuses applications pratiques.

Les méthodes de modélisation d'homologie (ou comparative) utilisent des structures de protéines expérimentales ("modèles") pour construire des modèles de protéines liées à l'évolution ("cibles").

- **COSMIC (the Catalogue Of Somatic Mutations In Cancer) :**

Une base de données en ligne. Elle met en disposition les données des articles de la littérature scientifique. Elle est la ressource la plus vaste et la plus complète au monde pour explorer l'impact des mutations somatiques dans le cancer humain. Cette base de données est librement disponible sans restriction via son site web.

- **Ensemble :**

Ensemble est un système bio-informatique d'annotation automatique de génomes. C'est un projet conjoint de l'European Bioinformatics Institute (EBI) et du Wellcome Trust Sanger Institute dont l'idée centrale est d'organiser de vastes champs d'information biologique autour de séquences génomiques.

Pour chaque génome analysé, d'identifier par un processus automatique l'ensemble des gènes qu'il contient. Il s'appuie pour cela sur des données de séquences existantes (ARN, protéines), qu'il « raccroche » sur le génome, pour en déduire la structure des gènes. (48).

Matériel et méthodes

Chapitre II : matériels et méthode

II.1.Matériels :

Cette étude a été réalisée grâce à :

- PC
- Connexion internet
- Différents outils bio-informatiques

Bases de données et logiciels bio-informatiques :

- **GenBank** : La première étape de cette étude consiste à télécharger les séquences de la base de données nucléotidique GenBank.

Basic Local Alignment SearchTool (BLAST) : pour l'alignement des séquences.

- **ORF Finder** : pour avoir la séquence protéique.

Online MendelianInheritance in Man(OMIM) : pour la détection des mutations et pour voir la relation entre le génotype et le phénotype.

- **Swiss model** : est utilisé pour la visualisation des structures tridimensionnelles des protéines.
- **Ensembl** : Base de données spécialisée complémentaire à NCBI

Tableau IV: Séquences étudiés

Séquence	Position	Numéro d'accèsion sur NCBI	Variant	ID sur NCBI
Gène P53	117479963..1 17668665	AB082923.1	-	-
ARNm P53	1...6132		-	-

II .2. Méthodes d'utilisation Différents outils et logiciels bio-informatique

II .2.1.Utilisation de GenBank

pour récupérer la séquence de gène p53 et les séquences des exons et télécharger la on doit passer sur un moteur de recherche et faire comme suit :

1-Insérer le numéro d'accèsion. (Voir figure 6)

2-Lancer la recherche (voir figure 6)

The screenshot shows the NCBI homepage with a search bar at the top. The search bar contains the text 'Nucleotide' and 'AB082923'. A red box labeled '1' is placed over the search input field, and another red box labeled '2' is placed over the 'Search' button. Below the search bar, there is a red banner with a COVID-19 warning. The main content area is divided into three columns. The left column is a navigation menu with items like 'NCBI Home', 'Resource List (A-Z)', 'All Resources', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', 'Homology', 'Literature', 'Proteins', 'Sequence Analysis', 'Taxonomy', and 'Training & Tutorials'. The middle column is titled 'Welcome to NCBI' and contains a description of the center, a link to 'About the NCBI | Mission | Organization | NCBI News & Blog', and three main sections: 'Submit' (Deposit data or manuscripts into NCBI databases), 'Download' (Transfer NCBI data to your computer), and 'Learn' (Find help documents, attend a class or watch a tutorial). The right column is titled 'Popular Resources' and lists 'PubMed', 'Bookshelf', 'PubMed Central', 'BLAST', 'Nucleotide', 'Genome', 'SNP', 'Gene', 'Protein', and 'PubChem'. Below this is a 'NCBI News & Blog' section with a recent article about RefSeq release 200.

Figure 6: Utilisation de GenBank (personnelle, 2020).

II .2.2.A.Téléchargement des séquences

Une fois que la recherche de la séquence est lancée sur la plateforme NCBI, un tableau de bord concernant la requête demandée est affiché sur la page de navigation, d'où on a la possibilité de télécharger les séquences sous format FASTA.

The screenshot shows the NCBI GenBank interface. At the top, there is a search bar with the text 'Nucleotide' and a search button. Below the search bar, there is a red banner with a warning icon and text: 'COVID-19 is an emerging, rapidly evolving situation. Get the latest public health information from CDC: <https://www.coronavirus.gov>. Get the latest research from NIH: <https://www.nih.gov/coronavirus>'. Below the banner, the main content area displays the search results for 'Homo sapiens mRNA for P53, complete cds'. The results include the GenBank accession number AB082923.1, the FASTA format, and a table of sequence details. The table includes fields such as LOCUS, DEFINITION, ACCESSION, VERSION, KEYWORDS, SOURCE, ORGANISM, REFERENCE, and JOURNAL. The ORGANISM field is 'Homo sapiens (human)'. The REFERENCE field includes the title 'Mutated p53 gene encodes a nonmutated epitope recognized by HLA-B*4601-restricted and tumor cell-reactive CTLs at tumor site' and the journal 'Cancer Res. 63 (4), 854-858 (2003)'. The JOURNAL field includes the PubMed ID '12591737'. On the right side of the interface, there are several analysis tools and options, including 'Change region shown', 'Customize view', 'Analyze this sequence', 'Run BLAST', 'Pick Primers', 'Highlight Sequence Features', 'Find in this Sequence', and 'Articles about the TP53 gene'.

Figure 7: Tableau de bord de la requête sur NCBI (personnelle, 2020).

II .3.Utilisation de Blast

Le Blaste est utilisé pour l'alignement entre les différentes séquences. Nous avons utilisé pour notre étude le BLASTn et le BLAST comme le montre la figure 8.

Nous avons procédés comme suit :

- 1-Choisir le type de blast selon le besoin.
- 2-Introduire le numéro d'accèsion ou bien la séquence sous forme FASTA.
- 3-Choisir la base de comparaison.
- 4-Lancer l'algorithme d'alignement.

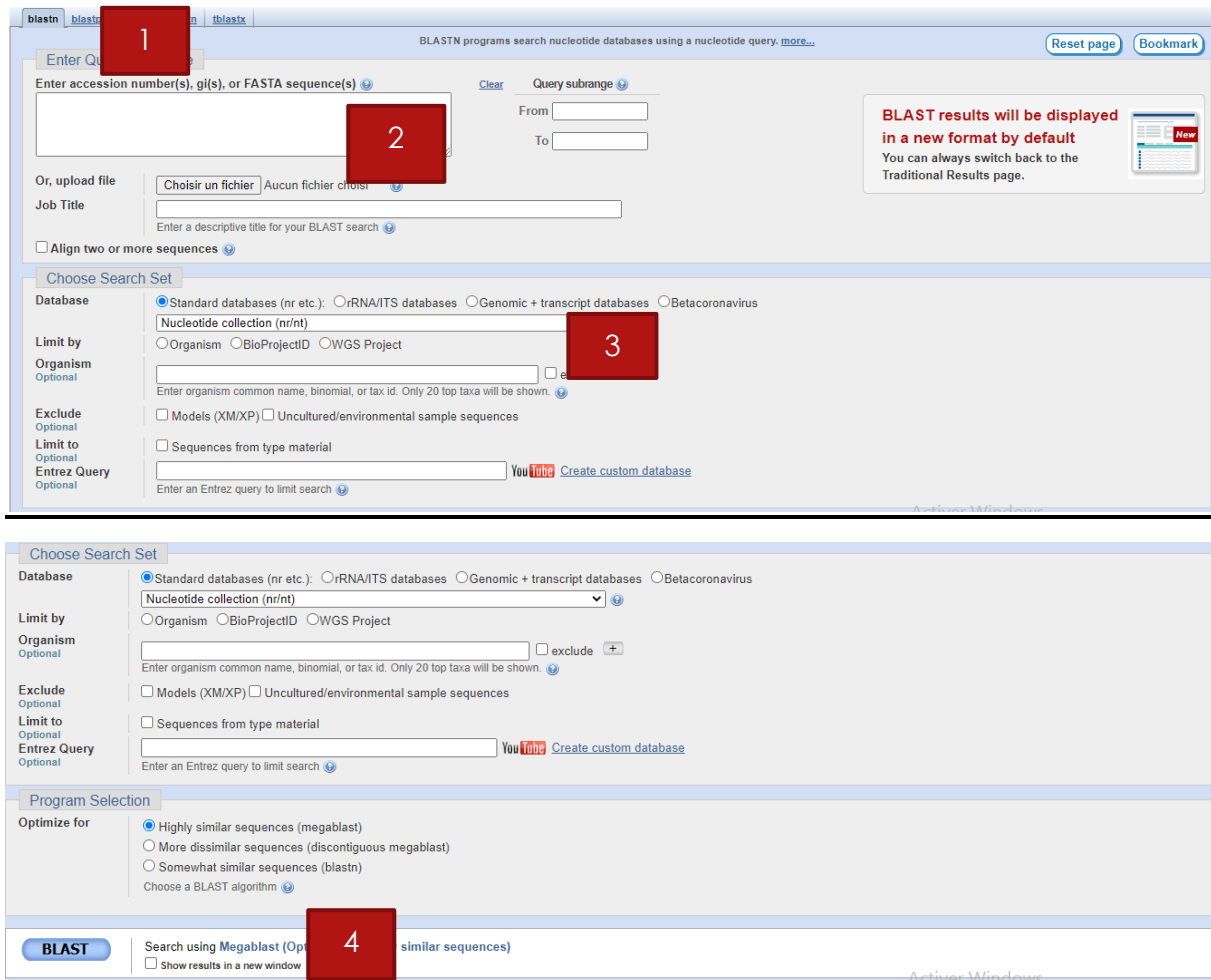


Figure 8: Utilisation de BLAST sur la plate-forme NCBI (Personnelle, 2020).

II .4.Utilisation d'Ensembl

On utilise cette base de données pour chercher une éventuelle relation entre la structure et la fonction de notre gène, pour cela, On passe sur la page d'accueil de la base de données Ensembl.

- 1-Choisir l'espèce Humaine.
- 2-Insérer le nom de gène.
- 3-Lancer la recherche

The screenshot shows the Ensembl genome browser homepage. At the top, there is a navigation bar with links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located in the top right corner. Below the navigation bar, there are four main sections: Tools, BioMart, BLAST/BLAT, and Variant Effect Predictor. The search bar is highlighted with a red box labeled '1'. The search input field contains 'Human' and 'p53', with a red box labeled '2' around the input field. The search button is labeled 'Go' and is highlighted with a red box labeled '3'. Below the search bar, there are sections for 'All genomes' and 'Favourite genomes'. The 'All genomes' section has a dropdown menu for species selection. The 'Favourite genomes' section lists 'Human GRCh38.p13', 'Mouse GRCm38.p6', and 'Zebrafish'. On the right side, there is a section for 'Ensembl Release 100 (April 2020)' with a list of updates and news. At the bottom right, there is a watermark for 'Activier Windows'.

Figure 9: Utilisation d'Ensembl (Personnelle, 2020).

II .5.Utilisation d'ORF Finder

Algorithme utilisé pour avoir les cadres de lectures ‘Open Frame Read’ de notre séquence. en premier passe sur la page d'accueil d'ORF Finder puis

- 1-Insérer la séquence.
- 2-Les ORFs peuvent commencer avec tous les codons.
- 3-Chercher les ORFs dans les cadre de lecture 1,2 et 3 du sens direct/indirect.
- 4-Ne renvoie que des ORF d'au moins 60 codons de long.
- 5-Utiliser le code génétique standard.
- 6-Lancer la recherche.

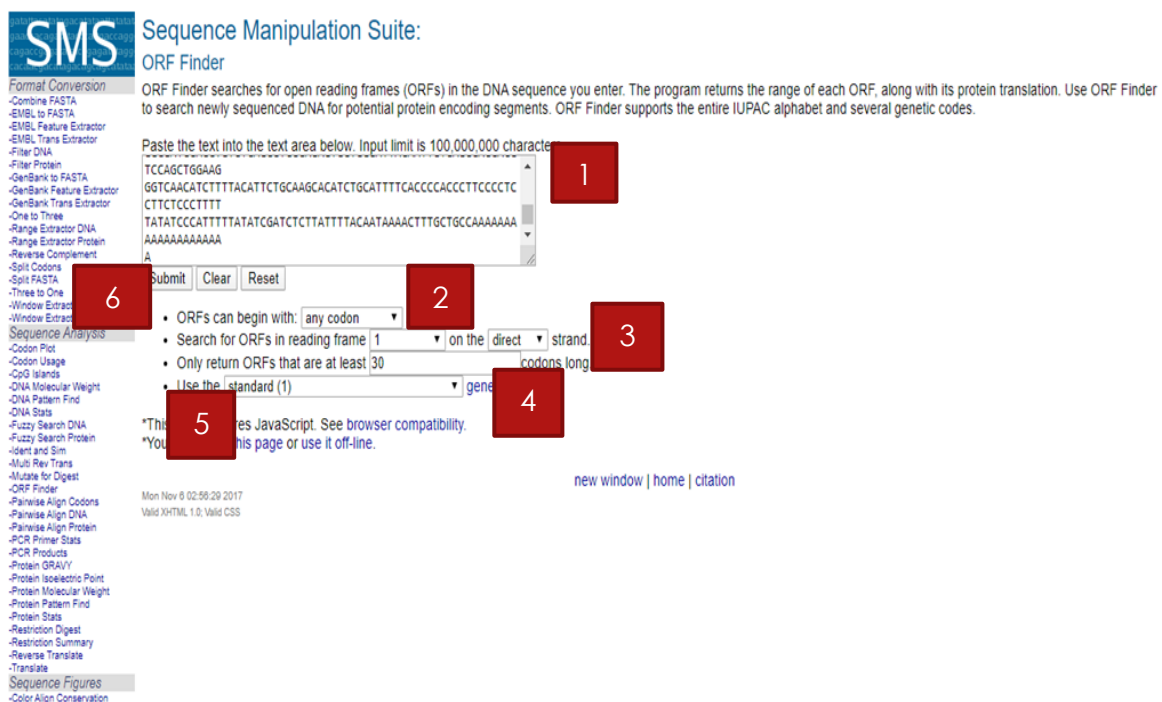


Figure 10 : Utilisation d'ORF Finder (Personnelle, 2020).

II .5.Utilisation de swiss-model

C'est une base de données spécialisée dans la quelle on retrouve les différentes informations concernant notre requête (Structure 3D, séquence protéique, alignement multiple...etc) On passe sur la page d'accueil de swiss-prot pour commencer la modélisation. (Figure 20)

1-insérer la séquence protéique ou bien le numéro d'accession.

2-Lancer la recherche de modèle.

The screenshot shows the SWISS-MODEL web interface for starting a new modelling project. The page has a header with the SIB and BIOZENTRUM logos, the text 'SWISS-MODEL', and navigation links for 'Modelling', 'Repository', 'Tools', 'Documentation', 'Log in', and 'Create Account'. The main content area is titled 'Start a New Modelling Project'. It features a 'Target Sequence(s):' section with a text input field containing a red box with the number '1'. Below the input field are buttons for '+ Upload Target Sequence File...' and 'Validate'. To the right, a 'Supported Inputs' dropdown menu is open, showing options: 'Sequence(s)', 'Target-Template Alignment', 'User Template', and 'DeepView Project'. Below the 'Target Sequence(s)' section are input fields for 'Project Title:' (containing 'Untitled Project') and 'Email:' (containing 'Optional'). At the bottom are two large blue buttons: 'Search For Templates' and 'Build Model'. A red box with the number '2' is positioned to the left of the 'Build Model' button.

Figure 11: Utilisation de Swiss-model (Personnelle ,2020).

RESULTATS ET DISSCUSSION

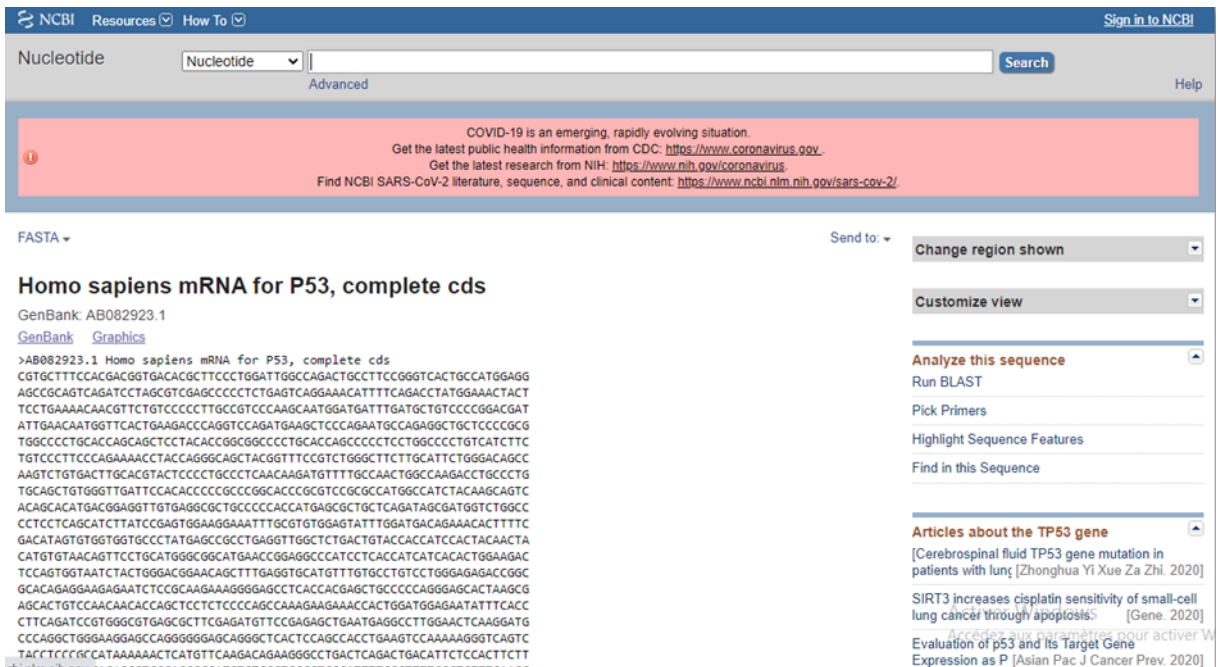
Chapitre III : Résultats et Discussion

I. Les résultats de GenBank

Les résultats obtenus sous forme FASTA débutent par une ligne commençant par :

Le signe « > » il est recommandé que toutes les lignes de texte soient plus courtes que 80 caractères.

- L'identifiant du gène.
- Quelques commentaires.
- Puis sur la deuxième ligne début la chaîne de séquence.



The screenshot shows the NCBI GenBank interface. At the top, there's a search bar with 'Nucleotide' selected. Below it, a red banner contains COVID-19 related information. The main content area displays the FASTA sequence for 'Homo sapiens mRNA for P53, complete cds' (GenBank: AB082923.1). The sequence is shown in a monospaced font, with the first line starting with '>AB082923.1 Homo sapiens mRNA for P53, complete cds'. To the right of the sequence, there are several interactive options: 'Send to', 'Change region shown', 'Customize view', 'Analyze this sequence' (with sub-options like 'Run BLAST', 'Pick Primers', 'Highlight Sequence Features', 'Find in this Sequence'), and 'Articles about the TP53 gene' (with sub-options like 'Cerebrospinal fluid TP53 gene mutation in patients with lung cancer', 'SIRT3 increases cisplatin sensitivity of small-cell lung cancer through apoptosis', 'Evaluation of p53 and Its Target Gene', 'Expression as P').

Figure 12 : Séquence d'ARNm du gène p53 sous format FASTA

Chapitre : III Résultats et discussion

II. Les résultats de Blast n de l'ARNm :

Les résultats de Blast n (alignement global nucléotidique) sont montrés dans le tableau

Tableau V: Résultats de BLAST de gène p53

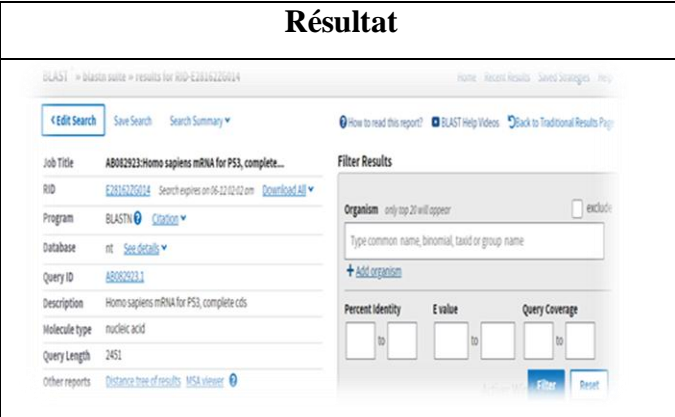
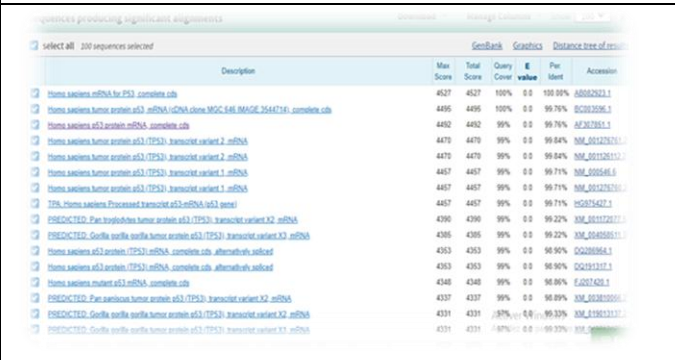
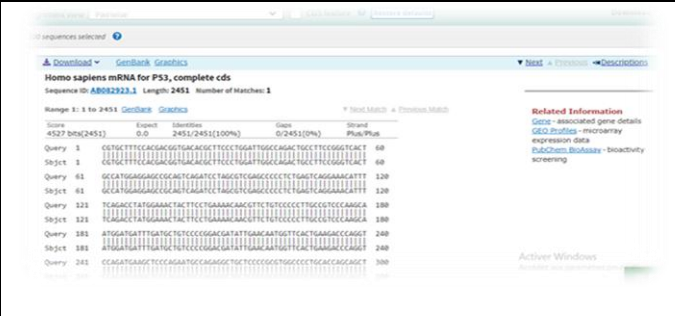
Résultat	Rôles	Rôles
	Récapitulatif de la requête.	Séquence soumise. Identifiant. Longueur . Type. Banque de donnée interrogée. programme utilisé.
	Représentation graphique des résultats.	les traits de couleur représentent à un alignement entre la séquence soumise et une séquence de la banque. La couleur correspond au score. La longueur correspond à la taille de l'alignement.
	Représentation global des résultats.	Chaque ligne correspond à un trait colorée la représentation graphique. Une liste des séquences avec leur numéro d'accèsion, le score d'alignement et l E-valu
	Description des résultats, des alignements	Query : la séquence soumise. Subject : la séquence trouvée dans la base de données. Les grandeurs propre à chaque alignement (score, E-valu, identité, positif, de gap).

Tableau VI : des meilleurs résultats pBLAST pour la séquence p53 dans le génome d'Homo sapiens : Score maximum, score total, valeur E, couverture de requête et pourcentage d'identité correspondants sont donnés pour faire correspondre les numéros d'accession des correspondances et leurs descriptions.

numéros d'accession	Description	Score max	Score total	valeur E, couverture de requête	pourcentage d'identité
[NP_000537.3]	<u>cellular tumor antigen p53 isoform a</u> [Homo sapiens] 813	<u>613</u>	<u>100</u>	<u>0</u>	<u>100.00</u>
[NP_001119590.1]	<u>cellular tumor antigen p53 isoform g</u> [Homo sapiens] 737	<u>737</u>	<u>90</u>	<u>0</u>	<u>100.00</u>
[NP_001119585.1]	<u>cellular tumor antigen p53 isoform c</u> [Homo sapiens] 689	<u>689</u>	<u>86</u>	<u>0</u>	<u>97.65</u>
[NP_001119586.1]	<u>cellular tumor antigen p53 isoform b</u> [Homo sapiens] 688	<u>688</u>	<u>84</u>	<u>0</u>	<u>100.00</u>
[NP_001263624.1]	<u>cellular tumor antigen p53 isoform h</u> [Homo sapiens] 612	<u>612</u>	<u>76</u>	<u>0</u>	<u>97.35</u>
[NP_001263625.1]	<u>cellular tumor antigen p53 isoform i</u> [Homo sapiens] 612	<u>612</u>	<u>74</u>	<u>0</u>	<u>100.00</u>
[NP_001119587.1]	<u>cellular tumor antigen p53 isoform d</u> [Homo sapiens] 549	<u>549</u>	<u>66</u>	<u>0</u>	<u>100.00</u>
[NP_001263626.1]	<u>cellular tumor antigen p53 isoform j</u> [Homo sapiens] 491	<u>491</u>	<u>59</u>	<u>3E-149</u>	<u>100.00</u>
[NP_001119588.1]	<u>cellular tumor antigen p53 isoform e</u> [Homo sapiens] 422	<u>422</u>	<u>50</u>	<u>2 E-149</u>	<u>100.00</u>
[NP_001119589.1]	<u>cellular tumor antigen p53 isoform f</u> [Homo sapiens] 422	<u>422</u>	<u>53</u>	<u>3 E-149</u>	<u>96.17</u>

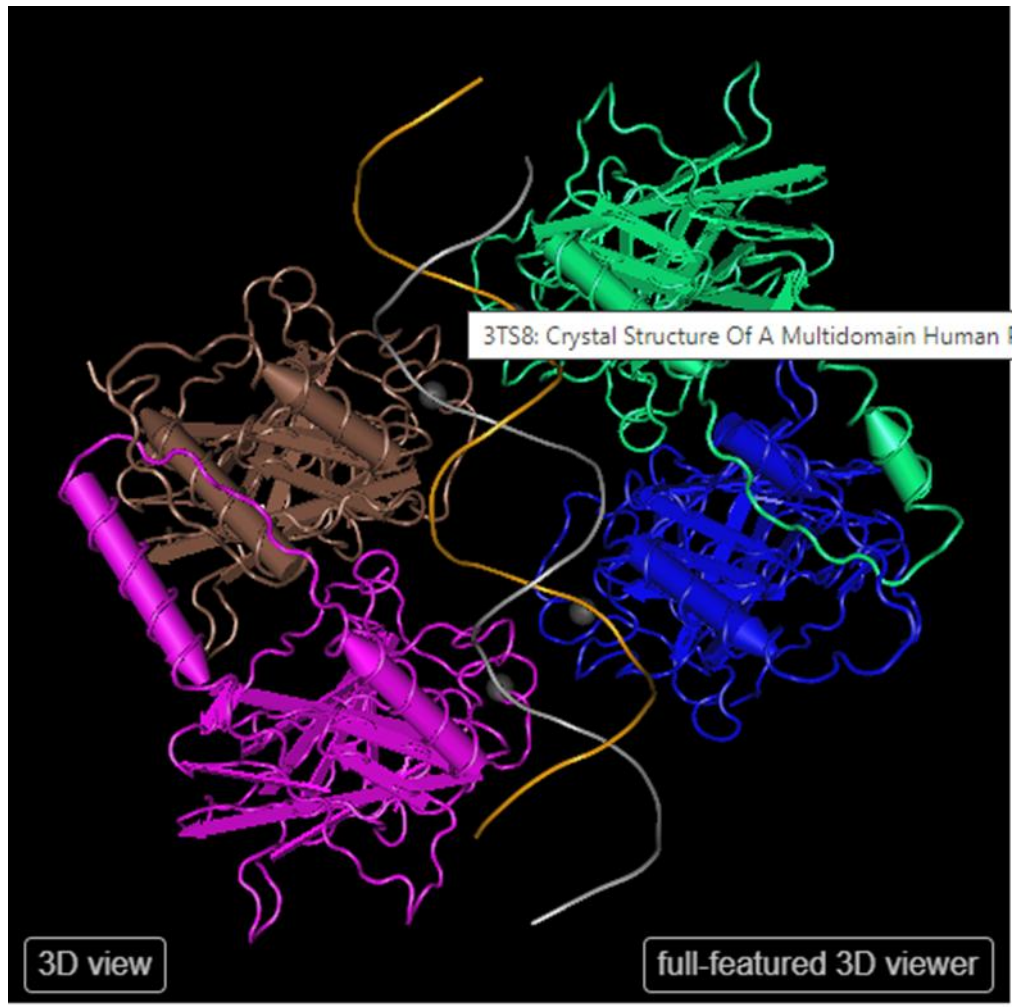


Figure 12 Structure cristalline d'un tétramère P53 humain multidomaine lié à l'élément de réponse P53 naturel (perssonelle2020)

III Résultats d'alignement nucléotidique :

Après l'utilisation de l'outil bio-informatique BLAST l'alignement nous a permis de faire ressortir les régions non similaires. L'objectif de l'alignement est de disposer les composants (acides nucléotidique) pour identifier les zones de concordance.

Chapitre : III Résultats et discussion

Ces alignements sont réalisés par un outil bio-informatiques dont l'objectif est de maximiser le nombre de coïncidences entre les nucléotides ou les acides aminés dans les différentes séquences, effectivement nous avons détectée toutes les mutations de chaque séquence

Un exemple d'une mutation détecté dans la figure L'analyse moléculaire de la séquence nous à permis d'identifier une mutation de la base C/A523 a détecté chez une des séquences analysé de gène p53.

```
Query 421 AAGTCTGTGACTTGACGTA CTCCCCTGCCCTCAACAAGATGTTTGGCCAACTGGCCAAG 480
          |||
Sbjct 421 AAGTCTGTGACTTGACGTA CTCCCCTGCCCTCAACAAGATGTTTGGCCAACTGGCCAAG 480

Query 481 ACCTGCCCTGTGCAGCTGTGGGTTGATTCCACACCCCGCCCGGCACCCGCGTCCGCGCC 540
          |||
Sbjct 481 ACCTGCCCTGTGCAGCTGTGGGTTGATTCCACACCCCGCCCGGCACCCGCGTCCGCGCC 540
```

Figure13: la mutation de la base C/G 523 est détectée chez une des séquences analysées (personnelle 2020)

Le nucléotide C qui était à la 523ème position sur le gène a été remplacée par le nucléotide A qui provoque un changement de la substitution des acides aminés

III .1La localisation des mutations

Notre recherche est basée sur la localisation des mutations au niveau de la séquence nucléotidique à l'aide des outils bio-informatique (BLAST.Anagène), nous avons constaté que les mutations sont localisées sur plusieurs sites distribués sur la chaîne nucléotidique. Les différentes études des séquences nucléotidiques de gène p53

Chapitre : III Résultats et discussion

Allèles	Séquence	Nucléique	Polypeptide	Type de mutation	Phénotype clinique
p53mut248d	G743-	CGG248del	Arg248del	délétion cadre de lecture modifié	cancer des bronches
p53mut175c	C523A	CGC175AGC	Arg175Ser	faux-sens	cancer des bronches
p53mut175b	C523G	CGC175GGC	Arg175Gly	faux-sens	cancer des bronches
p53mut273d	G818C	CGT273CCT	Arg273Pro	faux-sens	cancer du sein
p53mut273e	817T	CGT273Ins	Arg273Ins	Insertion cadre de lecture modifié	cancer du sein
p53mut175a	G524A	CGC175CAC	Arg175His	faux-sens	cancer du colorectal

TableauVII: quelque mutation détectée après l'utilisation de BLAST p53 (personnelle2020).

III .1.Les résultats dans ORF Finder de la mutation :

Le nucléotide C qui était à la 523ème position sur le gène a été remplacé par le nucléotide A qui provoque un changement de la substitution des acides aminés.

```
>Translation of ORF number 1 in reading frame 1 on the direct strand.  
RAFHDGDTLPWIGQTAFRVTAMEEPOSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQA  
MDDLMLSPDDIEQWFTEDPGPDEAPRMPEAAPRVAPAPAAPTPAAPAPAPSWPLSSSVPS  
QKTYQGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCQLAKTQPVQLWVDSTPPPQTRVRA  
MAIYKQSQHMTEVVRRCPPHHESSDSDGLAPPQHILIRVEGNLRVEYLDDRNTFRHSVVVF  
YEPPEVGSDCCTTIHYNMCMSSCMGGMNRRPILTIITLEDSSGNLLGRNSFEVHVCACPG  
RDRRTEENLRKKGEPHHELPPGSTKRALSNNNTSSSPQPKKKPLDGEYFTLQIRGRERFE  
MFRELNEALELKDAQAGKEPGGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD*
```

Figure14: Traduction de la séquence mutée (personnelle 2020)

Le remplacement :

Arg  Ser

Type de mutation dans ce cas est : faux-sens

Chapitre : III Résultats et discussion

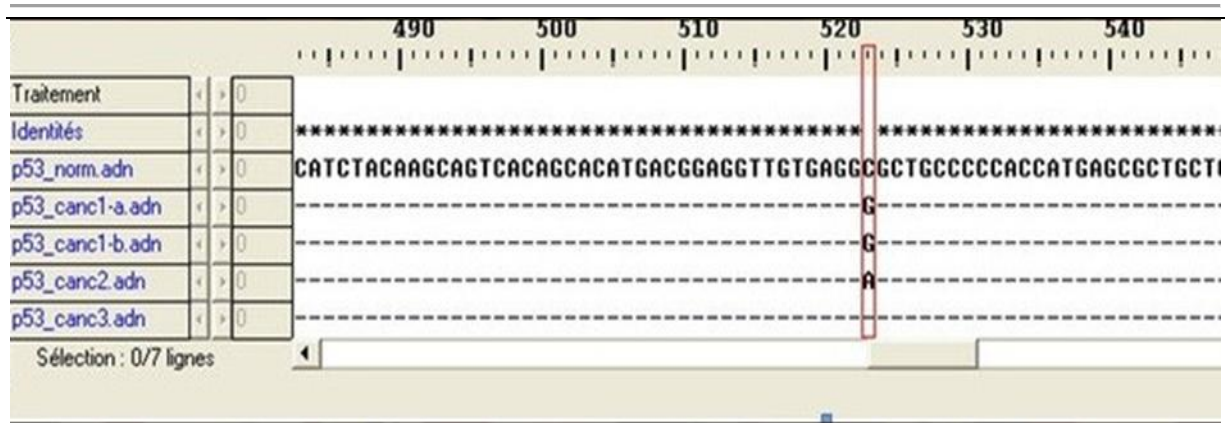


Figure15 :comparaison de séquence nucléotidique avec angine

Dans ce cas nous remarquons que le gène p53 est composé de 4 chaînes peptidiques différentes qui s'enroulent sur la partie lésée de l'ADN et la réparer. On comprend donc que lorsque ce gène est défectueux, cancer puis métastase ne peuvent que proliféré

III .Résultats d'alignement protéiques :

La séquence protéique soumise et la séquence de la base de données :

- Sont des molécules de type acide aminé.
- Leur taille est 393acides aminés.
- Le programme utilisé est le blast P.
- Le score d'alignement de cette séquence entre 80-200

Chapitre : III Résultats et discussion

Job Title	AB082923:Homo sapiens mRNA for P53, complete...
RID	FNAB9KUC014 Search expires on 07-01 10:54 am Download All ▼
Program	BLASTP ? Citation ▼
Database	nr See details ▼
Query ID	lcl Query_37436
Description	None
Molecule type	amino acid
Query Length	393

Figure16 ;carte d' information d'alignement portique de p53 sur blaste(perssonelle2020)

III.1.Les résultats deswiss-model :

Lesprotéines restent difficiles à étudier sur la structure moléculaire. Le problème majeur posé par l'enseignement des structures enzymatiques est la perception des objets dans l'espace à 3 dimensions. Pour les molécules de faibles tailles rencontrées surtout en chimie, on utilise un outil pour modéliser des structures et des mouvements des macromolécules comme les enzymes et les protéines.

III.2. Séquence protéique de p53

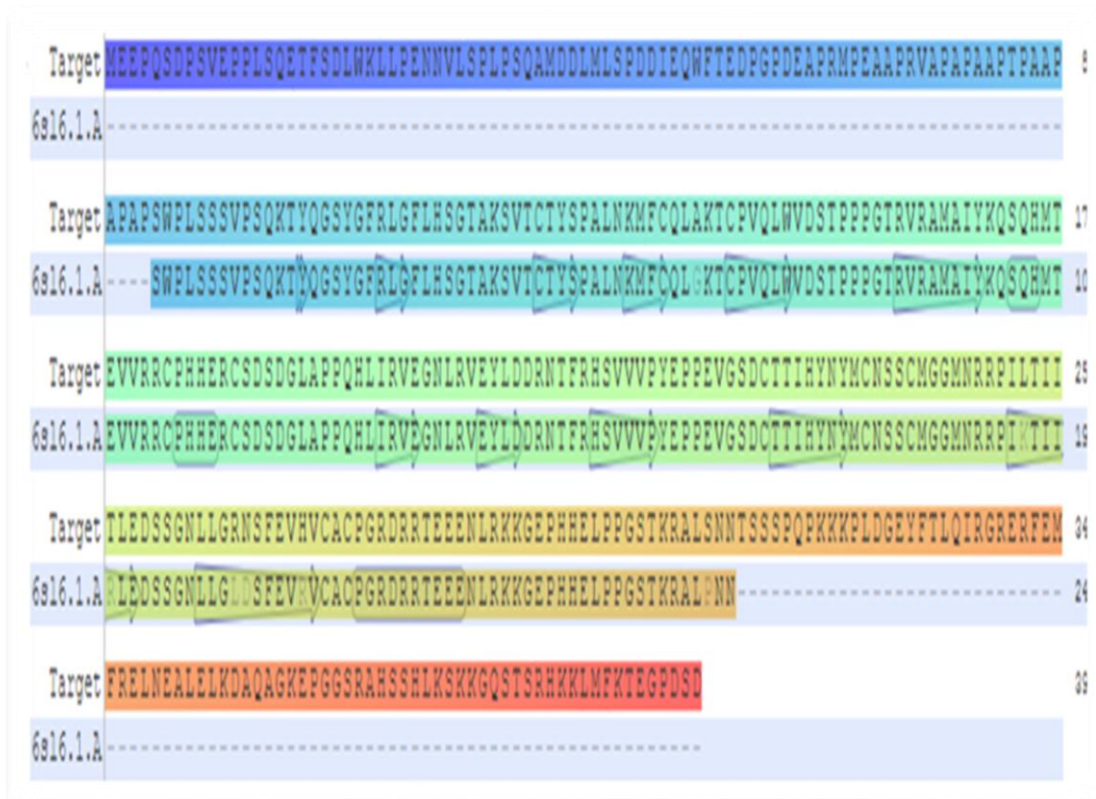


Figure17: Séquence protéique du gène P53(personnelle, 2020).

III.3. Structure tridimensionnelle de la protéine p53 :

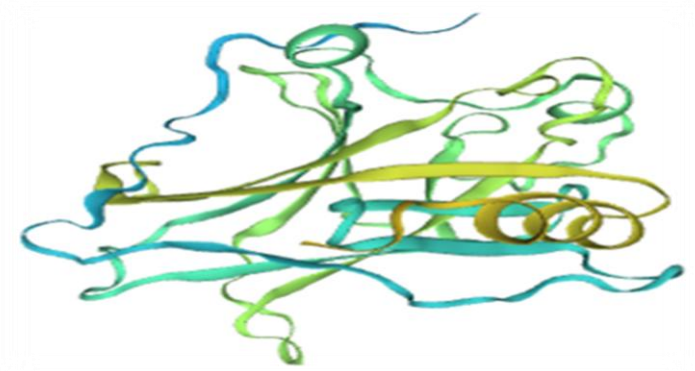


Figure18: Structure 3D de la protéine p53. (Personnelle, 2020).

- Chaque couleur présent dans la séquence protéique représente la même couleur dans la structure 3D.
- Pour savoir la position d'un n'importe quel acide amine sur la structure Tridimensionnel on clique sur l'acide aminé, il va apparaître sur la structure.

III .4. Structure de domaine de liaison à

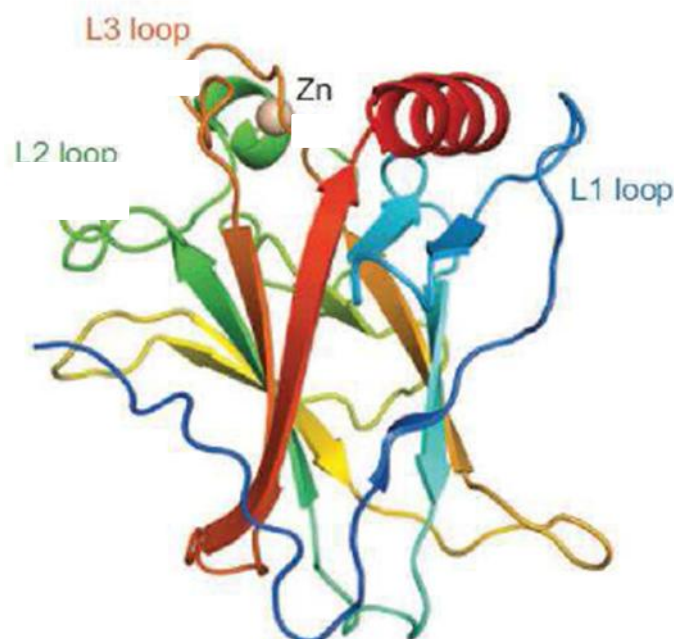


Figure 19 : Structure de domaine de liaison à l'ADN de p53(personnelle 2020)

Chapitre : III Résultats et discussion

- B : Structure de p53 sans ADN représentée avec un dégradé de couleur arc-en-ciel, du bleu au N-terminal au rouge au C-terminal ; PDB :2OCJ.

Le domaine p53C est très instable. Il a été montré que dans ce domaine de nombreux groupes polaires enfouis ne sont pas stabilisés par des liaisons hydrogènes, ce qui pourrait expliquer cette instabilité. En effet la substitution d'un tel couple enfoui (Tyr236 et Thr253) avec les résidus hydrophobes correspondants, trouvés dans les paralogues p63 et p73

Des études phylogénétiques suggèrent que les gènes des paralogues p63 et p73 sont plus anciens que p53, cela impliquerait que l'instabilité de p53 soit une évolution permettant une flexibilité et une rapide transition entre l'état replié et non replié. Cette flexibilité de p53C fournit la plasticité structurelle nécessaire pour une fonctionnalité optimale (49).

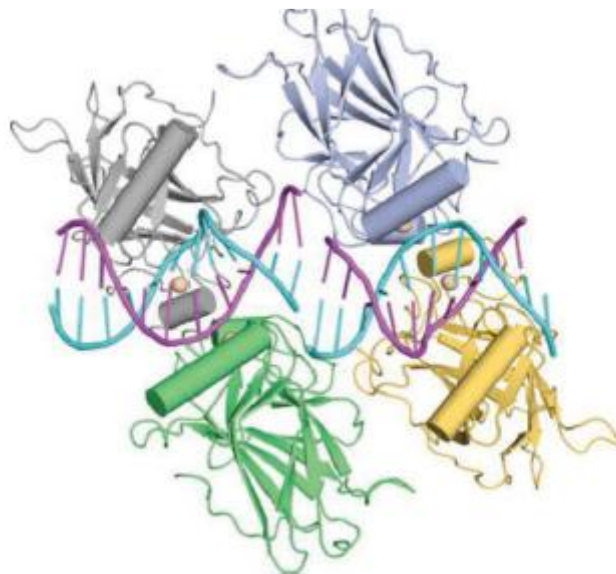


Figure 20 : Représentation de 4 p53 lié à deux motifs consécutifs de demi-sites d'ADN (personnelle 2020)

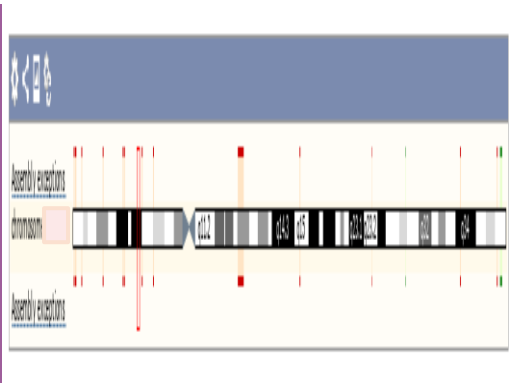

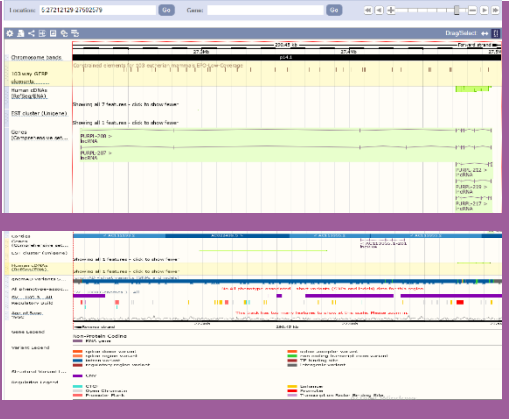
- Il a été démontré que 4 domaines p53 se lient à l'ADN de manière très coopérative donnant un complexe avec une stoechiométrie 1 ADN pour 4 p53.
- Une séquence spécifique d'ADN double brin sur laquelle se lie p53 afin de réguler la

Chapitre : III Résultats et discussion

Transcription des différents gènes.

IV. Les résultats d'Ensemble

Tableau VIII: Les résultats de l'Ensembl se présentant comme dans le tableau :

Résultat	Rôle	Descriptifs
	Image chromosomique	<ul style="list-style-type: none"> *Les régions de chromosome. *Le numéro de chromosome. *La région d'intérêt (rectangle en rouge).
	Image vue d'ensemble	<ul style="list-style-type: none"> *Les bandes de chromosome. *La séquence génomique. *La région d'intérêt. *La position de gène.
	Représentation de région	<p>La transcription de la région d'intérêt :</p> <ul style="list-style-type: none"> *les traits correspondent aux exons. *les lignes correspondent aux introns. *La séquence de la région d'intérêt

V .Les résultats dans OMIM

Représentation graphique des relations phénotype / gène associées à cette entrée. Les séries phénotypiques (lorsqu'elles sont disponibles) sont affichées avec les gènes pertinents et les phénotypes suivants à une profondeur de 12 nœuds

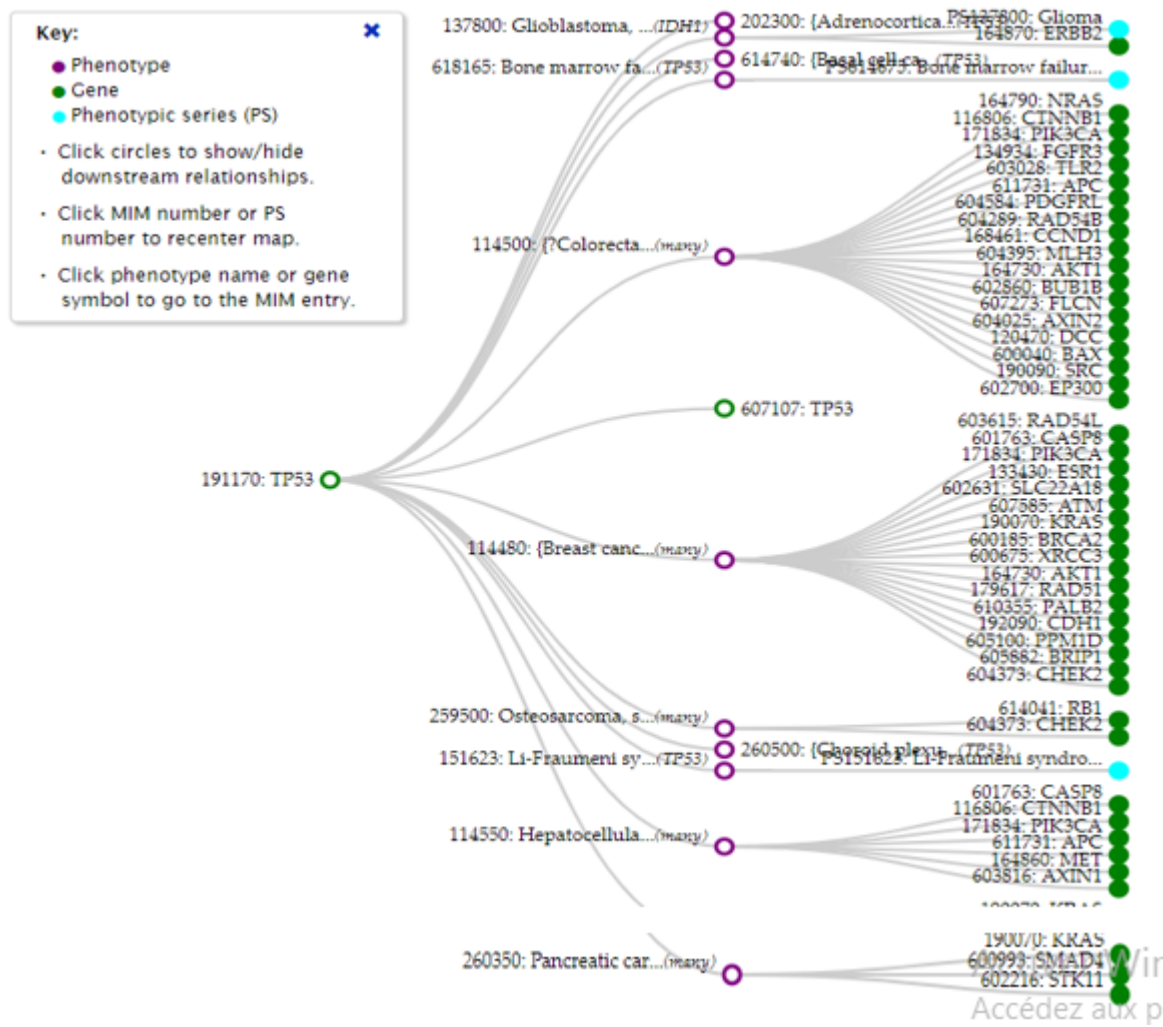


Figure 21: Représentation graphique circulaire de la relation phénotype génotype de la p53. (personnelle,2020)

Conclusion

Dans ce travail nous avons proposé de faire des études bio-informatique sur des séquences nucléotidiques prises d'une base de données spécialisé aléatoirement anonymes des gens atteints de cancer avec des états phénotypiques déférentes.

L'analyse des mutations du gène p53 est un exemple frappant de cette archéologie moléculaire qui nous permet de définir l'origine des mutations pour certains cancers. Dans un futur proche, d'autres gènes pourront probablement être utilisés comme sonde pour ce type d'analyse. Par ailleurs, une meilleure connaissance des signatures moléculaires associées à l'exposition à des carcinogènes devrait nous permettre de mieux affiner ces études. Il est important maintenant de définir quelles conséquences ces études peuvent avoir en terme de santé publique.

Il est remarquable de voir la vitesse à laquelle l'épidémiologie moléculaire a évolué au cours de la dernière décennie. Il ne fait plus de doute maintenant qu'un grand nombre de cancers humains sont provoqués par des mutations induites par des carcinogènes exogènes.

Reference bibliographiques

- 1-Comings DE (1973) : A general theory of carcinogenesis. *ProcNatlAcadSci U S A* 70: 3324-3328.
- 2-Nowell PC (1976) : The clonal evolution of tumor cell populations. *Science* 194:23-28.
- 3-Vogelstein B, Kinzler KW (2004) : Cancer genes and the pathways they control. *Nat Med* 10: 789-799.
- 4-Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, et al. (2006) : The consensus coding sequences of human breast and colorectal cancers. *Science* 314: 268-274.
- 5-Ko LJ, Prives C (1996) : p53: puzzle and paradigm. *Genes Dev* 10: 1054-1072.
- 6-Douglas Hanahan et Robert A. Weinberg, « The hallmarks of cancer », *Cell*, vol. 100, 7 janvier 2000, p. 57-70 (PMID 10647931).
- 7-Genetic testing for cancer susceptibility, *Journal of Clinical Oncology*, vol.14, pp.1730-1736, 1996.
- 8-Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Reece, J. B. (2017). *Campbell Biology* (11th ed.). Pearson.
- 9-Banavali, Nilesh K. (2013). "Partial Base Flipping is Sufficient for Strand Slippage near DNA Duplex Termini". *Journal of the American Chemical Society*. 135 (22): 8274–8282
- 10-Zhang J (2003)."Evolution by gene duplication: an update" (PDF).*Trends in Ecology& Evolution*. 18 (6)
- 11-Lewis, R. (2004). *Human Genetics: Concepts and Applications* (6th ed.). McGraw Hil
- 12-www.eurogentest.org.Retrieved March 29, 2019.
- 13-Wogan, G.N., Hecht, S.S., Felton, J.S., Conney, A.H. & Loeb, L.A. (2004).Environmental and chemical carcinogenesis. *Seminars in Cancer Biology*, 14, 473-486.
- 14-Deregulation of the Notch pathway as a common road in viral carcinogenesis.Vázquez-UlloaE,Lizano M, Sjöqvist M, Olmedo-Nieva L, Contreras-Paredes A. *Rev Med Virol*. 2018 Sep;28(5):e1988. Epub 2018 Jun 28.
- 15-Hall, J. & Angele, S. (1999). Radiation, DNA damage and cancer.*Molecular Medecin Today*, 5, 157-164
- 16-Wilbur B, ed. (2009). *The World of the Cell* (7th ed.). San Francisco, C.
Croce CM (January 2008). "Oncogenes and cancer".*The New England Journal of Medicine*.358 (5): 502–11. doi:10.1056/NEJMra072367. PMID 18234754.
- 17-Wen Xue W. et coll. Senescence and tumour clearance is triggered by p53 restoration un murine liver carcinomas (textecomplet). *Nature* (2007) 445:656-660.
- 18-Bulletin du Cancer Volume 104, Issue 11, November 2017, Pages 962-970 Altération de la réparation de l'ADN et cancer.
- 19-Centre Inter national de recherch: 2001

Chapitre : III Résultats et discussion

- 20-Centre inter nationale de recherche de cancer communique de pressen° 263 12 septembre 201.
- 21-Chia-Lin Wei, Qiang Wu, Vinsensius B. Vega, Kuo Ping Chiu, Patrick Ng, Tao Zhang, Atif Shahab, How Choong Yong, YuTao Fu, ZhipingWeng, JianJun Liu, Xiao Dong Zhao, Joon-Lin Chew, Yen Ling Lee, Vladimir A. Kuznetsov, Wing-Kin Sung, Lance D. Miller, Bing Lim, Edison T. Liu, Qiang Yu, Huck-Hui Ng, and YijunRuan ; A GlobalMap of p53 Transcription-Factor Binding Sites in the HumanGenome ; 2006 ; Cell 124, 207–219.
- 22-Christiani DC.Smoking and the molecular epidemiology of lung cancer.Clin Chest Med 2000;21:87-93.
- 23-Eilstein D, Uhry Z, Cherie-Challine L, Isnard H. Mortalité par cancer du poumon chez les femmes en France.Analyse et tendance de projection de 1975 à 2014. Rev EpidemiolSantePublique 2005;53:167-81
- 24-Sandler A, Gray R, Perry MC, et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung.
- 25-Dr. Marc Espié , Guide Affection longue durée – Cancer du sein, (janvier2010).
- 26-Bray F, Ren JS, Masuyer E, Ferlay J (2013).Global estimates of cancer prevalence for 27-sites in the adult population in 2008.Int J Cancer, 132(5):1133–1145.
- 28-Revue Médicale Suisse N° -507 publiée le 25/08/2004
- 29-Dimitrios Raptis, Ignaz Schneider, Klaus E. Matzel et Oliver Ott, « The Differential Diagnosis and Interdisciplinary Treatment of Anal Carcinoma », DeutschesAerzteblatt Online, 3 avril 2015.
- 30-calassification-cancer 2018
- 31-Ecrit par: La rédaction de Doctissimo Révision médicale : Dr Jesus Cardenas, Directeur médical de Doctissimo, 16 septembre 2016 Sources : Dossier de presse de la semaine nationale contre le cancer colorectal.
32. Sasco A. Epidémiologie descriptive des cancers du poumon. Rev Mal Respir 1999;16:S44-5 . Molinie F, Velten M, Remontet L, et al. Progression de l'incidence du cancer du poumon en France (1978-2000). Rev Mal Respir 2006;23:127-34
33. Eilstein D, Uhry Z, Cherie-Challine L, Isnard H. Mortalité par cancer du poumon chez les femmes en France.Analyse et tendance de projection de 1975 à 2014. Rev EpidemiolSantePublique 2005;53:167-81.
- 34-Ferlay J, Soerjomataram I, Ervik M. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base Lyon, France; 2015. 11(1)
- 35-Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 SedkaouiC.Chimiothérapie et thérapie ciblée dans le cancer colorectal métastatique, thèse de doctorat, Université Mouloud Mammeride Tizi-Ouzou. Faculté de Médecine 2015 ; 203

Chapitre : III Résultats et discussion

36-Epidémiologie du cancer du sein en algérie Mahnane.A-HamdiCherifM- laboratoire santé environnement des hauts plateaux sétifiens 18 FEV 2012

37-Boumedjar A, Smaili F. GROS PLAN sur le cancer colorectal. Santé – Mag 2012 ; 3 :28

38-Ramet M, Castren K, Jarvinen K, Pekkala K, Turpeenniemi-Hujanen T, Soini Y, Paakko P, Vahakangas K. p53 protein expression is correlated with benzo[a]pyrene-DNA adducts in carcinoma cell lines. Carcinogenesis.1995, 16: 2117-24.

39-Weston A. and Godbold J. H., Polymorphisms of H-ras-1 and p53 in Breast Cancer and Lung Cancer: A Metaanalysis, 1997, Environmental Health Perspectives Vol 105.

Institut français de l'éducation plateforme acces.

40-Thierry Soussi Karim Dehouche Christophe BéroutL'analyse des mutations du gène p53 dans les cancers humains : le lien entre l'épidémiologie et la carcinogenèse médecine/sciences 2000 ; 16 : 1387-96.

41-S.J. Boulton, « Cellular functions of the BRCA tumour-suppressor proteins », Biochemical Society Transactions, vol. 34, no 5, 2006, p. 633–645.

42-Panagiotis K ;Stasinakis ;DespoinaNicolaou ; Modeling of DNA and Protein Organization Levels with Cn3D Software ; Biochemistry and Molecular Biology Education ;vol00 ;page 1-4 ;2016.

43-Boudjellal Sara, Fettelhadjer, étude bioinformatique de trois forme alléliques du gène APOintervenant dans l'expression de la maladie d'Alzheimer forme sporadique, mémoire de master, Université de Blida 1 ;département biologie et physiologie cellulaire, spécialité bioinformatique, soutenue 2017.

44-Sylvain Legrand, Blast Basic Local AlignmentSearchTool, cour, université de Lille 1 science et technologie, page 1-52, Septembre 2016.

45-Dennis A. Benson, Mark Cavanaugh, Karen Clark, Ilene Karsch-Mizrachi, David J. Lipman,JamesOstell and Eric W. Sayers ; GenBank ; Nucleic Acids Research, Vol. 45,p37-42,2017.

46-Simona Rossi , AristotelisTsirigos , Antonio Amoroso , NicolettaMascellani , IsidoreRigoutsos ,George A. Calin , Stefano Volinia, OMiR: Identification of associations between OMIM diseases and microRNAs ; Genomics vol 97 ,page 71–76.2011.

47-,George A. Calin , Stefano Volinia, OMiR: Identification of associations between OMIM diseases and microRNAs ; Genomics vol 97 ,page 71–76.2011.

48-Joanna S. Amberger and Ada Hamosh ; Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes ; CurrProtoc Bioinformatics ;vol58 ;p1- 20.2018

Chapitre : III Résultats et discussion

49-Cañadillas, J.M.P., Tidow, H., Freund, S.M.V., Rutherford, T.J., Ang, H.C., and Fersht, A.R. (2006). Solution structure of p53 core domain: Structural basis for its instability. Proc Natl Acad Sci U S A 103,2109–2114.

www.liguepulmonaire.ch Liguespulmonairescantionales(suiss)

Ecrit par: La rédaction de Doctissimo Révision médicale: Dr Jesus Cardenas, Directeur médical de Doctissimo, 16 septembre 2016 Sources : Dossier de presse de la semaine nationale contre le cancer colorectal.

<https://fqc.qc.ca/fr/informatio/le-cancer/calassification-cancer-2018>