

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE  
UNIVERSITE SAAD DAHLEB – BLIDA 1



FACULTE DES SCIENCES DE LA NATURE ET DE LA VIE  
DEPARTEMENT DE BIOLOGIE ET PHYSIOLOGIE CELLULAIRE

*MEMOIRE DE FIN D'ETUDE EN VUE DE L'OBTENTION  
DU DIPLOME DE MASTER FILIERE SCIENCES BIOLOGIQUES*

*Spécialité: Bioinformatique et génomique.*

**Thème**

**Etude bioinformatique de deux ARN non  
codant MALAT-1/HOTAIR dans  
l'expression du cancer du poumon**

**Présenté par :**

**DOUIDI Mouna**

**KRITLI Salima**

**Devant le jury composé de**

<b>Mr BEN YAHIA N.</b>	<b>USDB-1</b>	<b>MAA</b>	<b>Président</b>
<b>Mme AISSANI R.</b>	<b>USDB-1</b>	<b>MAA</b>	<b>Examinatrice</b>
<b>Mme ABDUL HUSSAIN A.S.</b>	<b>USDB-1</b>	<b>MCB</b>	<b>Promotrice</b>

**Promotion: 2017**

---

# *Remerciements*

*Nos remerciements vont avant tout, à DIEU tout puissant pour la volonté, la santé et la patience qu'il nous a donnée durant toutes ces longues années d'études afin que nous puissions arriver à ce stade. Et enchallah aller encore plus loin dans un accomplissement personnel et professionnel.*

*Nos remerciements particuliers à notre promotrice, Madame **ABDUL HUSSAIN A.S.** Merci pour votre confiance et votre patience. La confiance que nous a témoignée Mme **ABDUL HUSSAIN**, le soutien et les conseils qu'elle nous a prodigué tout au long de ce modeste travail. Travailler avec elle est une expérience passionnante. C'est une femme que nous estimons pour son honnêteté et son humanisme. Nous espérons pouvoir encore relever de nombreux autres défis avec elle dans les prochaines années.*

*Nous remercions vivement le président du jury Mr **BEN YAHIA** d'avoir accepté de présider notre jury de soutenance et Mme **AISSANI** d'avoir accepté de faire parti du jury et d'évaluer notre mémoire de fin d'études.*

*Nos remerciements vont également à nos familles qui nous ont aidés tout au long de notre chemin avec leur patience, encouragement et confiance.*

---

---

## *Dédicace*

*Nous dédions ce modeste travail ;*

*A nos parents les plus chers a nos cœurs, qui nous ont encouragés à poursuivre nos rêves et soutenus dans chaque étape nos vies, qui ont fait de nous ce que nous somme aujourd'hui, et font tout pour nous rendre heureux. Que Dieu le tout puissant veille sur eux et les protège ;*

*A nos frères et sœurs, aucun mot ne pourra exprimer notre gratitude envers eux. Que Dieu nous unissent a jamais ;*

*Merci Chawki et Moncef pour tous ces moments où vous nous avez fait rire avec vos blagues pourries ;*

*Merci AYA et Dalila pour votre soutien, patiences et tous vos encouragements tout au long de cette année ;*

*A nos grands parents, pour tout l'amour qu'ils nous donnent et toutes leurs invocations ;*

*A nos petits bouts de choux Adem (Power Adem), Saadi Millel, Sara et Mohamed.*

*Et a tous ceux qui nous aiment.*

*Salima*

*Mouna*

---

---

## Abstract

Lung cancer is one of the main causes linked to cancer mortality in the world. Long non-coding RNAs emerge as being one of the key regulatory molecules of the biologic processes. Besides their implication in the normal physiology, their aberrant expression appears to be associated with cancer's development and progression. HOTAIR and MALAT-1 are two LncRNA incriminated in lung cancer and are involved in the P53 pathway, the mutation of which is most frequently found in cancer cells and especially lung cancer with 70% of lung mutations of codons 157 and 158. GenBank, BLAST and COSMIC allowed the study of LncARN and to proceed to the identification of the various mutations (HOTAIR/MALAT-1/P53). Other associated mutations have been revealed and cause other cancers such as bronchi and tongue cancers with a prevalence of 15.04% and that of the kidney with 1.58%. Nonetheless, the functional properties of these two RNAs in lung cancer remain unknown.

**Keywords:** lung cancer, long non coding RNA, P53, HOTAIR, MALAT-1, Bioinformatics.

---

## ملخص

سرطان الرئة هو من الأسباب الأساسية القاتلة في العالم, تتميز جزيئات Long non coding RNA باعتبارها واحدة من الجزيئات التنظيمية الرئيسية للعديد من العمليات البيولوجية، بالإضافة إلى أهميتها في وظائف الأعضاء أثناء نشاطها العادي، فإن أي خلل في التعبير عن هذه الجزيئات يترافق مع تطور مرض السرطان وتقدمه. ويعد كل من HOTAIR وMALAT-1 ثنائي من LncARN والمسببان أيضا للسرطان ويشركان في مسار P53، هذه الطفرة هي الأكثر وجودا وبالأخص في سرطان الرئة بنسبة 70%، أين نجد تغيرات على مستوى الرامزة 157 و 158، Gen Bank, BLAST، وCOSMIC سمح دراسة LncARN والشروع في تحديد مختلف الطفرات (HOTAIR/MALAT-1/P53) الطفرات المترابطة الأخرى أظهرت بأنها تخلق سرطانات أخرى كالقنوات الصدرية واللسان بنسبة 15.04%، بالإضافة إلى سرطان الكلى بنسبة 1.58%. هذه الخصائص الوظيفية لثنائي ARN في سرطان الرئة لا تزال معروفة إلا قليلا.

**الكلمات الرئيسية:** سرطان الرئة، Long non coding RNA, HOTAIR, MALAT-1، إعلام آلي حيوي.

---

## Résumé

Le cancer du poumon est l'une des principales causes de mortalité mondiale liée au cancer. Les longs ARN non codants émergent comme étant des molécules clés dans la régulation de différents processus biologiques. Au-delà de leurs implications dans la physiologie normale, l'expression aberrante des LncARN est souvent associée au développement du cancer et de sa progression. HOTAIR et MALAT-1 sont deux LncARN incriminés dans le cancer du poumon et se trouvent impliqués dans la voie de P53 dont la mutation est le plus fréquemment trouvée dans des cellules cancéreuses et particulièrement le cancer du poumon avec 70 % des cas où l'on retrouve des mutations au niveau des codons 157 et 158. GenBank, BLAST et COSMIC ont permis l'étude des LncARN et procéder à l'identification des différentes mutations (HOTAIR/MALAT-1/P53). d'autres mutations associés ont été révélées et provoquent d'autres cancers tel que celui des bronches et langue avec une prévalence de 15.04% et celui du rein avec 1.58%. Néanmoins, les propriétés fonctionnelles de ces deux ARN dans le cancer du poumon restent peu connues.

**Mots clés :** Cancer du poumon, Long ARN non codant, P53, HOTAIR, MALAT-1, Bioinformatique.

# Liste des Abréviations

**ADN** : Acide DésoxyriboNucléique.

**ARN nc** :ARN Non Codant .

**ARN** : Acide RiboNucléique.

**ARNm** : Acide RiboNucléique Messenger.

**ARNr** : Acide RiboNucléique Ribosomique.

**bcl-2** : B-Cell Lymphoma 2.

**CDS** :Coding sequence .

**CoREST** : REST Corépresseur .

**EZH2** : Enhancer of Zeste Homolog 2.

**GADD45A** : Growth Arrest And DNA Damage Inducible Alpha.

**GLOBOCAN** : estimated cancer incidence mortality and prevalence worldwide in 2012.

**H3K4me3** : lysine 4 de l'histone H3 triméthylée.

**HOTAIR** : HOX transcript antisense RNA..

**HOXC** : homeoboxC.

**HOXD** : homeoboxD.

**LAC** : Lung Adnoma carcinoma .

**lincARN** : Long intergenic RNA.

**LncARN** : Long Non Codant ARN.

**LSD1** : Lysine-Specific Demethylase 1.

**MALAT1** : Metastasis Associated Lung Adenocarcinoma Transcript 1.

**NCBI** : National Center for Biotechnology Information.

**NEAT 2** : Nuclearenriched abundant transcript 2 .

**Nt** : Nucléotide.

**ORF** : Open Reading Frame .

**P53** : Protéine 53.

**PANDA** : P21 Associated ncRNA DNA Damage Activated.

**pb** : paire de bases.

**PRC2** : Polycomb Repressive Complex2.

**SOX2ot** : overlapping transcript.

**TARID** : TCF21 Antisense RNA Inducing Promoter Demethylation.

**TCF21** : Transcription factor 21.

**Tug1** :Taurine Up-regulated 1.

## Liste des figures

<b>01</b>	Structure du poumon.	01
<b>02</b>	Estimations de l'incidence des cancers dans le monde en 2012.	02
<b>03</b>	Estimation de l'incidence des cancers en Algérie en 2012.	03
<b>04</b>	Le dogme central présentant la transcription de l'ADN en ARN messagers traduits par la suite en protéine, et la transcription de l'ADN en ARN non-codants.	06
<b>05</b>	Classification des petits ARN non codants.	07
<b>06</b>	Classification des longs ARNnc.	07 bis
<b>07</b>	Fonction des lncARN.	08 bis
<b>08</b>	Position génomique du gène HOTAIR, Bandes selon Ensemble et localisation selon GeneLoc.	10
<b>09</b>	Schéma du mode d'action de l'ARN non codant HOTAIR.	11
<b>10</b>	Position génomique du gène MALAT1, Bandes selon Ensemble et localisation selon GeneLoc.	12
<b>11</b>	Modifications post-transcriptionnelles de MALAT-1.	13
<b>12</b>	Page d'accueil de GenBank.	16
<b>13</b>	Page d'accueil de BLAST.	17
<b>14</b>	Page d'accueil de COSMIC.	18
<b>15</b>	Page d'accueil de RaptorX.	19
<b>16</b>	Utilisation de GenBank.	21
<b>17</b>	Méthodes d'utilisation de BLAST.	22
<b>18</b>	Dot Matrix View du Transcrit/ variant 1 de HOTAIR.	25
<b>19</b>	Différentes mutations détectées entre le transcrit de HOTAIR et le variant 1.	26
<b>20</b>	Dot Matrix View du transcrit et variant 2 de HOTAIR.	26
<b>21</b>	Mutation au niveau de la base 431...434 entre le transcrit de HOTAIR et le variant 2	27
<b>22</b>	Dot Matrix View du Transcrit et variant 3 de HOTAIR.	28
<b>23</b>	Délétion au niveau du variant 3 de HOTAIR.	28
<b>24</b>	Mutation au niveau de la base 287...290 entre le transcrit de HOTAIR et le variant 3	29

<b>25</b>	Dot Matrix View Transcrit de MALAT-1 avec les 3 variants.	30
<b>26</b>	Surexpression de MALAT-1.	32
<b>27</b>	Mutations au niveau du codon 157 de la P53.	34
<b>28</b>	Mutations au niveau du codon 158 de la P53.	35
<b>29</b>	CDS de la P53 avec les différentes mutations détectées.	36
<b>30</b>	Pourcentage des différentes mutations associées des nucléotides du codon 157.	38
<b>31</b>	Pourcentage des différentes mutations associées des nucléotides du codon 158.	40
<b>32</b>	Pourcentage des différents cancers associés aux mutations des codons 157,158.	41
<b>33</b>	Pourcentage des différentes mutations des codons 157;158 par type de cancer associé.	43

## Liste des tableaux

I	Classification des carcinomes bronchiques.	03
II	Les lncARN associés au cancer du poumon.	04
III	Caryotype humain.	05
IV	Classification des ARN non codants	07
V	Les accessions des différentes séquences étudiées.	16
VI	les résultats de BLAST.	23
VII	Comparaison entre les tailles de MALAT- 1 / HOTAIR.	24
VIII	Les mutations des ARNnc.	24
IX	Les mutations faux sens de la P53.	33
X	Mutations de la p53 provoquant le cancer du poumon.	35
XI	Mutations associées au codon 157 de la p53.	37
XII	Pourcentage des cancers associés a la mutation du codon 157.	39
XIII	Mutations associées au codon 158 de la p53.	39
XIV	Pourcentage des cancers associés à la mutation du codon 158.	40
XV	Pourcentage des différentes mutations nucléotidiques des codons 157,158 de la P53 liées aux cancers associés.	41

---

# Sommaire

Introduction

## CHAPITRE I : RAPPELES BIBLIOGRAPHIQUES

I. Anatomie du poumon.....	01
II. Cancer du poumon.....	02
II.1.Epidémiologie.....	02
II.2.Classification du cancer du poumon.....	03
II.3.LncARN et cancer du poumon.....	04
III. Génomique structurale.....	05
III.1.Caryotype humain.....	05
III.2.L'expression de l'information porte par l'ADN.....	06
III.3.Description des ARN codants et non codants.....	06
IV. Longs ARN non codants.....	08
IV.1.Description.....	08
IV.2.Classification.....	08
IV.3.Fonctions.....	09
V. Deux lncARN : HOTAIR / MALAT1.....	10
V.1.HOTAIR.....	10
V.2.MALAT-1.....	11
VI. Description de P53.....	12

---

---

VI.1.Mutations du gène P53.....	13
---------------------------------	----

VII. Etude bioinformatique.....	14
---------------------------------	----

## **CHAPITRE II : MATERIELS ET METHODES**

II.1.Matériels.....	15
---------------------	----

II.2.Méthodes.....	21
--------------------	----

## **CHAPITRE III : RESULTATS ET DISSCUSION**

III.1.Résultats et Discussion.....	24
------------------------------------	----

Conclusion.....	44
-----------------	----

## **Références bibliographiques**

## **Annexes**

---

---

## Introduction

Le cancer du poumon est l'une des principales causes de mortalité mondiale liée au cancer avec un taux de survie approximatif de 5 ans d'environ 16.6%. (**Wu et al., 2015**).

Les récents progrès concernant les technologies du séquençage permettant des analyses génomiques et transcriptomiques ont révélé que jusqu'à 85% du génome humain est transcrit (**Matthew et al., 2013**). Environ 1% du génome code pour des protéines fonctionnelles. Sur les 99% des portions restantes certaines ont toute fois des fonctions primordiales à l'organisme, notamment les segments du génome humain contenant des ARNnc (**Prensner et al., 2011**).

Les longs ARN non codant sont des transcrit de plus de 200 nucléotides qui ne codent pas pour une protéine (**Mathieu et al., 2014**). Ils peuvent moduler la chromatine ou fonctionner comme régulateurs de l'expression des gènes, et cela dépend de leurs localisations dans le noyau (**Hutchinson et al., 2007**).

Jusqu'à présent, plus de 3000 lncARN ont été identifiés, cependant les fonctions de 1% d'entre eux ont été discutées. L'expression aberrante des lncARN est liée à la tumorigénèse, et leur dysrégulation participe à l'initiation, la progression et aux métastases tumorales dans différents types de cancers humains. (**Zhang H. et al., 2013, Prensner et al., 2011**).

Les lncARN s'impliquent dans la voie de la P53 qui est le gène le plus fréquemment muté dans les cancers humains (70% cancer du poumon). Deux positions de mutation dans la p53 sont spécifiques au cancer bronchique, ces points critiques sont localisés sur les codons 157 et 158.

Par conséquent, durant notre travail, nous nous sommes intéressés à l'étude des deux lncARN (HOTAIR et MALAT-1) ayant une expression aberrante dans le processus tumoral du cancer du poumon et pour procéder à l'identifier des différentes mutations (HOTAIR/MALAT-1/P53), en utilisant plusieurs outils bioinformatiques.

## I. Anatomie du poumon

Les poumons sont situés dans le thorax, de chaque côté du cœur et font partie de l'appareil respiratoire. (Gondé., 2016).

Les poumons sont divisés en plusieurs lobes, eux-mêmes divisés en plusieurs segments. Le poumon gauche comprend 02 lobes et le poumon droit en compte 03. Chaque lobe contient des bronches. Ces dernières sont constituées d'alvéoles reliées entre elles par les bronchioles. Autour des alvéoles se trouvent de très minuscules vaisseaux sanguins (capillaires). Toutes les bronches sont reliées à la trachée : c'est la voie respiratoire en forme de tube qui se trouve dans le cou et le thorax. Elle se divise en 2 tubes chacun entre dans un poumon par une région appelée hile. Les bronches sont tapissées de cellules présentant des prolongements semblables à des cheveux très fins appelés cils. (Gondé., 2016). (Figure 01)

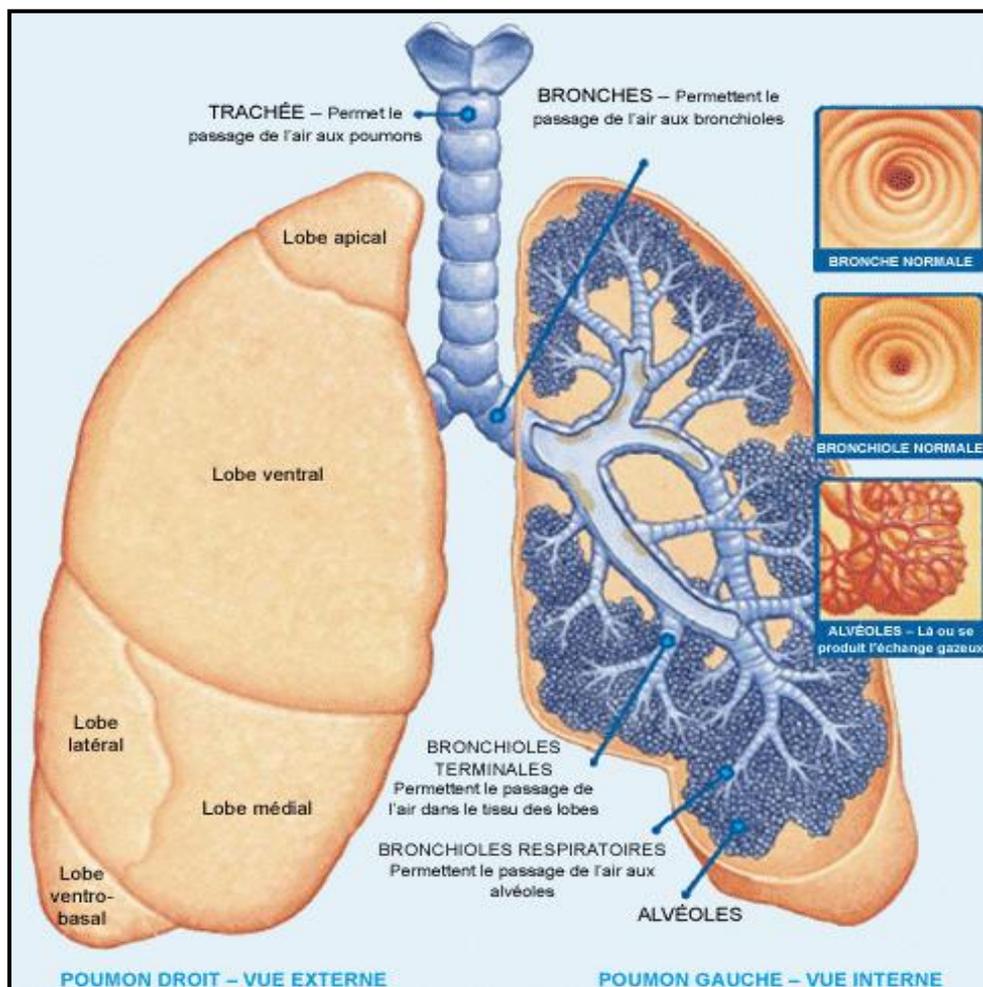
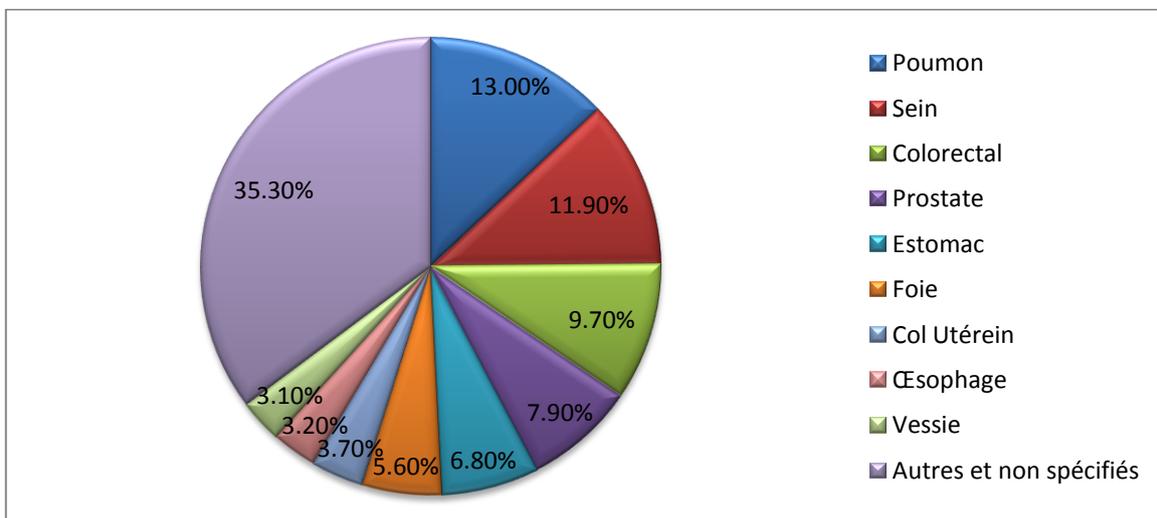


Figure 01 : Structure du poumon. (Gondé., 2016).

## II. Cancer du poumon

### II. 1. Epidémiologie

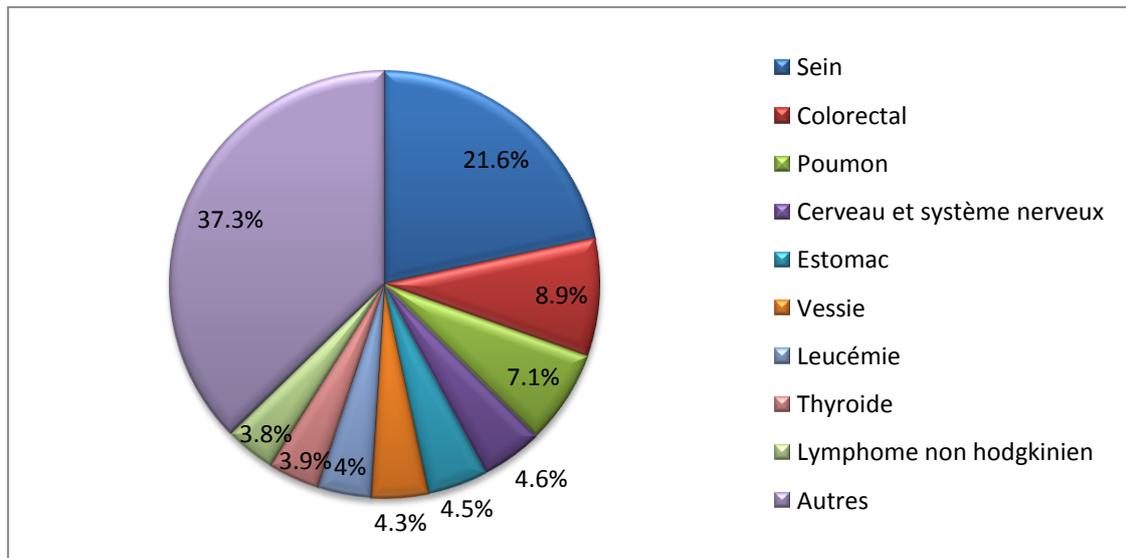
En 2012, le fardeau mondial des cancers tous types confondus s'est élevé à 14.1 millions de nouveaux cas et à 8.2 millions de décès. Le cancer du poumon s'est classé en première position parmi les cancers les plus fréquemment diagnostiqués avec (1.8 millions cas, soit 13%) suivi du cancer du sein avec (1.7 millions cas, soit 11.9%), et du colorectal avec (1.3 millions cas, soit 9.7%). **(Sedkaoui., 2015).** (Figure 02).



**Figure 02 :** Estimations de l'incidence des cancers dans le monde en 2012 **(Globocan., 2012).**

Dans le monde, on a estimé le nombre de nouveaux cas de cancer du poumon à 1.8 millions en 2012 contre 1.35 en 2002 .Ce qui montre une relative augmentation de l'incidence du cancer du poumon **(Chater., 2013).**

En Algérie, l'incidence du cancer du poumon (2707 cas/an soit 7.1%) est classée en troisième position après le cancer du sein (8177 cas/an soit 21.6%) et colorectal avec (3380 cas/an soit 8.9%). **(Sedkaoui., 2015).** (Figure 03).



**Figure 03** : Estimation de l'incidence des cancers en Algérie en 2012. (Globocan., 2012).

## II.2. Classification du cancer du poumon

Les cancers du poumon sont classés en 2 groupes : les cancers à petites cellules (13% des cancers pulmonaires) et les cancers non à petites cellules (87% des cancers pulmonaires). (Brambilla et al., 2001). (Tableau I).

**Tableau I**: Classification des carcinomes bronchiques. (Chater., 2013 ; Rocks et al., 2008).

Cancer bronchique non à petites cellules (plusieurs types)	Cancer bronchique à petites cellules (un seul type)
<ul style="list-style-type: none"> <li>☞ <b>L'adénocarcinome bronchique</b> - 45% des cas. - Observés chez les non fumeurs et les femmes.</li> <li>☞ <b>Le carcinome épidermoïde</b> -30 à 35% des cas. -Prédominance masculine corrélé à la consommation de cigarettes.</li> <li>☞ <b>Le carcinome à grandes cellules</b> -10% des cas. -Sans différenciation malphigienne ou glandulaire nette.</li> </ul>	<ul style="list-style-type: none"> <li>☞ Apparaissent chez les fumeurs.</li> <li>☞ Tumeur d'origine neuroendocrine.</li> <li>☞ Un temps de doublement extrêmement rapide de l'ordre de 30 jours.</li> <li>☞ Un pouvoir métastatique très important, par voie lymphatique et sanguine, rendant inutile une approche thérapeutique chirurgicale.</li> <li>☞ Une très grande sensibilité à la chimiothérapie et à la radiothérapie.</li> </ul>

### II.3.LncARN et cancer du poumon

Reconnu comme une maladie du génome, le cancer peut être provoqué par des mutations, des amplifications, des suppressions ou un dérèglement des produits génomiques, les lncARN règlent des processus cellulaires critiques dans le cancer du poumon, comme la prolifération, l'invasion et la survie (Tableau II). De plus, la dérégulation de l'expression de ces lncARN est corrélée avec les métastases -une surexpression des lncARN engendre une proportion élevée de métastases- (Loewen et al., 2014).

Des altérations de l'expression des lncARN ont été reportées dans plusieurs pathologies et notamment dans les cancers, (Shi et al., 2013)

**Tableau II** : Les lncARN associés au cancer des poumons. (Loewen et al., 2014).

lncARN	Interaction des molécules et des voies.	Processus cellulaire.	Caractéristiques cliniques associées.
HOTAIR	<ul style="list-style-type: none"> <li>➤ Induit par Collagène 1</li> <li>➤ Affecte l'expression des gélatinases.</li> <li>➤ Réprime les gènes de l'adhérence cellulaire.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Prolifération, migration et invasion.</li> <li>➤ Résistance au cisplatine in vitro et in vivo.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Augmente dans le cancer du poumon à non petites cellules.</li> <li>➤ Ganglion lymphatique et métastases cérébrale.</li> <li>➤ Mauvaise survie.</li> <li>➤ Augmente dans le cancer du poumon à petites cellules.</li> <li>➤ Invasion lymphatique.</li> </ul>
TARID	<ul style="list-style-type: none"> <li>➤ Active TCF21 via GADD45A.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Inconnu.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Diminue dans LAC et cancer bronchique à petites cellules</li> </ul>
MALAT1	<ul style="list-style-type: none"> <li>➤ Affecte l'expression de bcl-2 et des gènes liés aux métastases</li> </ul>	<ul style="list-style-type: none"> <li>➤ Croissance tumorale in vivo.</li> <li>➤ Survie.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Augmente dans le cancer du poumon à non petites cellules.</li> <li>➤ Métastases cérébrale.</li> <li>➤ Mauvaise survie.</li> </ul>
SOX2ot	<ul style="list-style-type: none"> <li>➤ Affecte l'expression d'EZH2.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Cycle cellulaire et prolifération.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Augmente le cancer bronchique à petites cellules.</li> <li>➤ Mauvaise survie.</li> </ul>

### III. Génomique structurale

#### III.1. Caryotype humain

Depuis les premières observations des chromosomes en 1880 par Flemming, la génétique est longtemps restée une science marginale, c'est pourquoi ce n'est qu'en 1956 que le nombre de chromosomes de l'espèce humaine a été correctement établi à 46 par Tjio et Levan, 22 paires d'autosomes + 1 paire de gonosomes (XX/XY). (Tableau III).

On estime le nombre de gènes codant une protéine chez l'homme à environ 25000, ces estimations ont été rendues possibles depuis le séquençage du génome humain, ce dernier est de taille approximative de 3 milliards de paires de bases. (Gautheret., 2012).

En 1960, a été établie à Denver la première nomenclature internationale pour la classification des chromosomes, basée sur leur taille et la position de leur centromère (Soukaina.,2015).

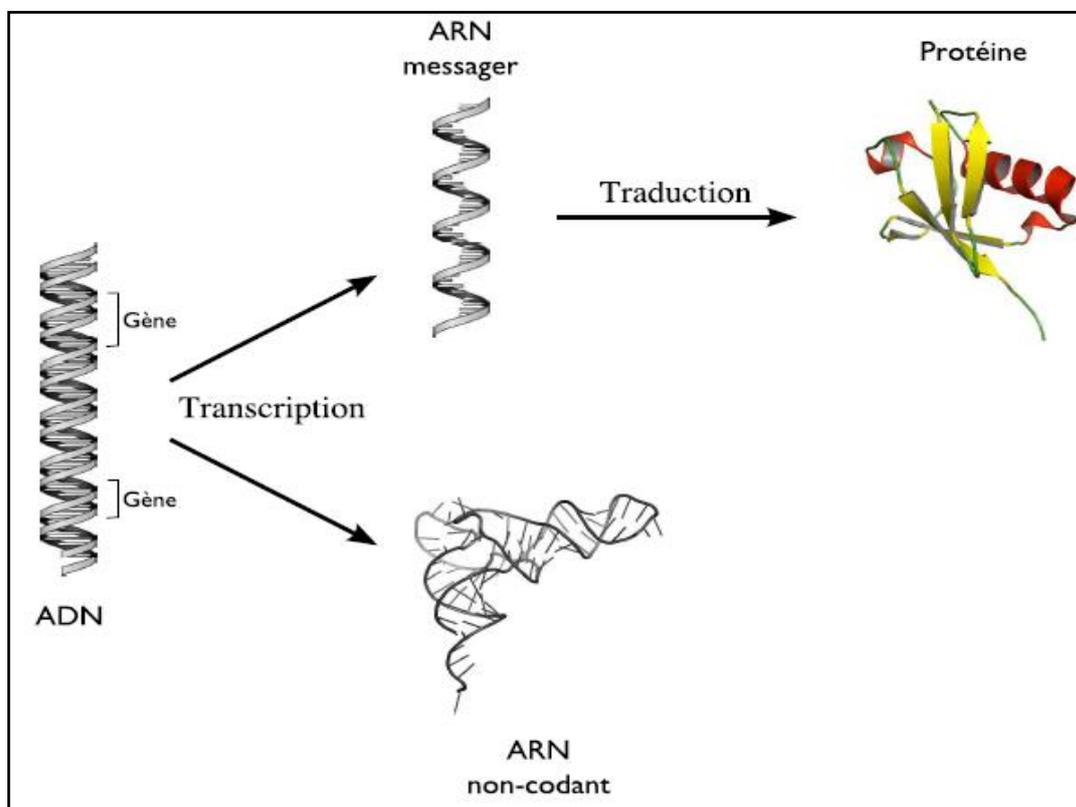
**Tableau. III** : Caryotype humain (NCBI., 2017).

Nr. Chr	Taille (Mb)	Nbr Protéines	ARNr	ARNt	Autres ARN	Gènes	Pseudogènes
1	248.96	11046	17	90	4350	5078	1372
2	242.19	8054	-	8	3638	3862	1166
3	198.3	6790	-	4	2723	2971	887
4	190.22	4374	-	1	2209	2441	799
5	181.54	4599	-	17	2225	2578	766
6	170.81	5338	-	144	2408	3000	876
7	159.35	4843	-	21	2232	2774	896
8	145.14	3962	-	5	1975	2152	661
9	138.4	4572	-	3	2135	2262	702
10	133.8	5237	-	3	2053	2174	631
11	135.09	6187	-	13	2267	2920	835
12	133.28	5648	-	9	2313	2521	680
13	114.36	1946	-	4	1235	1381	477
14	107.04	3252	-	18	1704	2055	583
15	101.99	3436	-	9	1778	1814	555
16	90.34	4453	-	27	1700	1920	451
17	83.26	6004	-	33	2131	2432	541
18	80.37	1812	-	1	1013	988	295
19	56.62	6452	-	6	1750	2481	514
20	64.44	2794	-	-	1276	1349	329
21	46.71	1240	-	1	687	756	202
22	50.82	2493	-	-	965	1172	348
X	156.04	3739	-	4	1196	2158	859
Y	57.23	325	-	-	298	577	395

### III.2.L'expression de l'information porte par l'ADN

L'information génétique contenue dans l'ADN est organisée en segments appelés les gènes. La première étape de l'expression de l'information génétique consiste à transcrire un gène en un ARN. Ce rôle de médiateur de l'information génétique constitue le premier rôle de l'ARN que l'on nomme ARN messager.

Il y a également deux autres types d'ARN : les ARN ribosomiques et les ARN de transfert. Contrairement aux ARN messagers, ces ARN sont des molécules fonctionnelles non traduites en protéine et que l'on regroupe sous le terme d'ARN non-codants. (Fontaine., 2009).



**Figure 04:** Présentation de la transcription de l'ADN en ARN messagers traduits par la suite en protéine, et la transcription de l'ADN en ARN non-codants. (Fontaine., 2009).

### III.3.Description des ARN codants et non codants

Dans le cas des gènes codants, l'ARN messager issu de la transcription, puis de la maturation éventuelle, contient toute l'information nécessaire à la production d'une protéine.

La traduction d'un ARN messager mature en une protéine est un processus complexe qui repose sur une cascade d'assemblages de molécules en interaction avec l'ARN messager à traduire.

A l'inverse des ARN messagers issus de la transcription de gènes codants, les ARN non codants issus de la transcription de gènes à ARN n'ont pas vocation à coder pour des protéines. (Fontaine., 2009).

Les ARN non codants sont classés en deux groupes selon leurs tailles :

- les petits ARN non codant (sncARN) d'une longueur inférieure à 200 nt.
- les longs ARN non codant (lncARN) d'une longueur de plus de 200 nt à plus de 100 kb (Cao., 2014).

**Tableau IV:** Classification des ARN non codants (Read., 2008 ; Fjose., 2001).

#### ARN non codant

Long non codant ARN (lncARN)	Petit ARN non codant (sncARN)
<ul style="list-style-type: none"><li>• LncARN intergénique</li><li>• LncARN intronique</li><li>• LncARN sens, antisens</li><li>• LncARN divergent</li></ul>	<ul style="list-style-type: none"><li>• Les petits ARN nucléaires</li><li>• Les ARN nucléaires</li><li>• Les micro-ARN</li><li>• Les petits ARN interférents</li></ul>

---

## IV. Longs ARN non codants

### IV.1. Description

Les longs ARN non codant sont des transcrit de plus de 200 nt qui ne codent pas pour une protéine (**Mathieu et al., 2014**). Ils ne possèdent généralement pas de cadre de lecture ouvert fonctionnel (ORF).

De nombreux gène des longs ARN non codant possèdent deux ou plusieurs exons, et environ 60% de ces lncRNA ont des queues polyA. Plusieurs d'entre eux possèdent une coiffe à l'extrémité 5' et subissent un épissage (**Fang et al., 2016**). Les lncRNA sont plus enrichis dans le noyau et montrent une plus faible conservation de séquence bien que certains d'entres eux soient fortement conservés (**Ulitsky et al., 2013**).

### IV.2. Classification

La classification des longs ARNnc est en fonction de leurs position par rapport a un gène codant On compte cinq classes de long ARN non codant : intergénique, intronique, sens, anti-sens et divergent. (**Rinn et Chang., 2012**).

La classe intergénique contient des longs ARNnc dont le site d'initiation est généralement éloigné de 5000 pb d'un gène codant pour une protéine. Cette distance empêche tout chevauchement. L'initiation de leur transcription peut se faire dans les deux directions et les longs ARNnc faisant partie de cette classe sont parmi les plus conservés à travers les organismes. (**Guttman et al., 2009**).

Les longs ARNnc étant dans les classes sens et anti-sens sont initiés dans une portion d'un gène codant pour une protéine ou dans sa portion 3'. Ces derniers sont transcrits dans la même direction que le gène (sens) ou dans la direction opposée (anti-sens) et chevauchent généralement au moins un exon codant. (**Rinn et Chang., 2012**).

Finalement, la classe des longs ARNnc divergents, la transcription peut être également effectuée dans les deux directions. Ce qui caractérise cette classe de longs transcrits est que le site d'initiation se fait à proximité d'un promoteur d'un gène codant pour une protéine. Ce site se situe généralement à une centaine de pb du promoteur, mais cette distance n'est toutefois pas encore définitivement établie (figure 06). (**Guttman et al., 2009**).

### IV.3.Fonctions

Les lncARN sont impliqués dans divers fonctions biologiques (Figure 07).

✓ **Régulent la transcription en agissant comme des leurres**

En se liant à une protéine et la titrer sans effectuer aucune fonction. Par exemple, GAS5 (Growth-Arrest Specific 5) qui peut interagir avec le domaine de liaison à l'ADN des récepteurs des glucocorticoïdes (GR), ce qui les empêche de se lier aux éléments de réponse de l'ADN. Ce type de lncARN exerce toujours un rôle régulateur négatif sur un effecteur (Chen et al., 2014).

✓ **Agissent comme molécule d'échafaudage**

En possédant différents domaines fournissant des plates-formes pour assembler différentes molécules effectrices qui fonctionnent ensemble, contrôlant ainsi précisément les interactions intermoléculaires et les événements de signalisation. Par exemple, HOTAIR peut agir comme un échafaudage et des ponts entre PRC2 et le complexe LSD1 / CoREST formant le complexe HOTAIR / PRC2 / LSD1, qui peut supprimer l'expression du gène cible (Tsai et al., 2010).

✓ **Servent aussi de guide**

En se liant à une protéine et dirigeant ensuite la localisation du complexe ribonucléoprotéique vers des cibles spécifiques pour réguler l'expression génique, soit en cis (de gènes voisins), soit en trans (de gènes distants). Par exemple HOTAIR qui dirige les complexes modificateurs de la chromatine, Polycomb Repressive Complex 2 (PRC2) et LSD1, pour cibler les gènes en trans, alors qu'Evf-2 vise sa cible en cis (Chen et al., 2014).

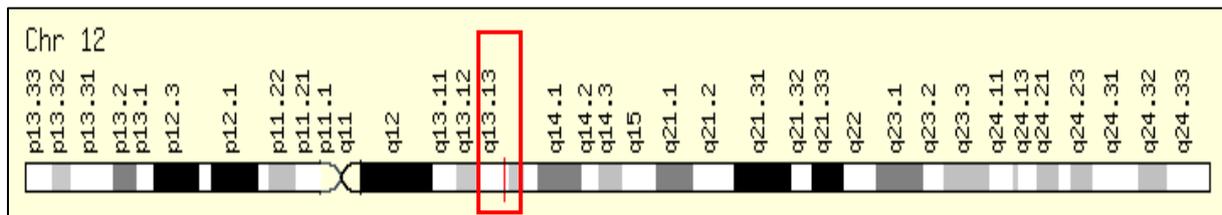
✓ **Servent de signaux moléculaires**

En montrant toujours une expression spécifique des cellules et répondent à divers stimuli, indiquant que leur expression est sous un certain degré de contrôle transcriptionnel. Par exemple, PANDA et Tug1 sont activés par transcription en réponse à l'endommagement de l'ADN par la liaison directe de la p53 à son promoteur (Khalil et al., 2009).

## V. Deux lncARN : HOTAIR / MALAT1

### V.1.HOTAIR (HOX Transcript Antisense Intergenic RNA)

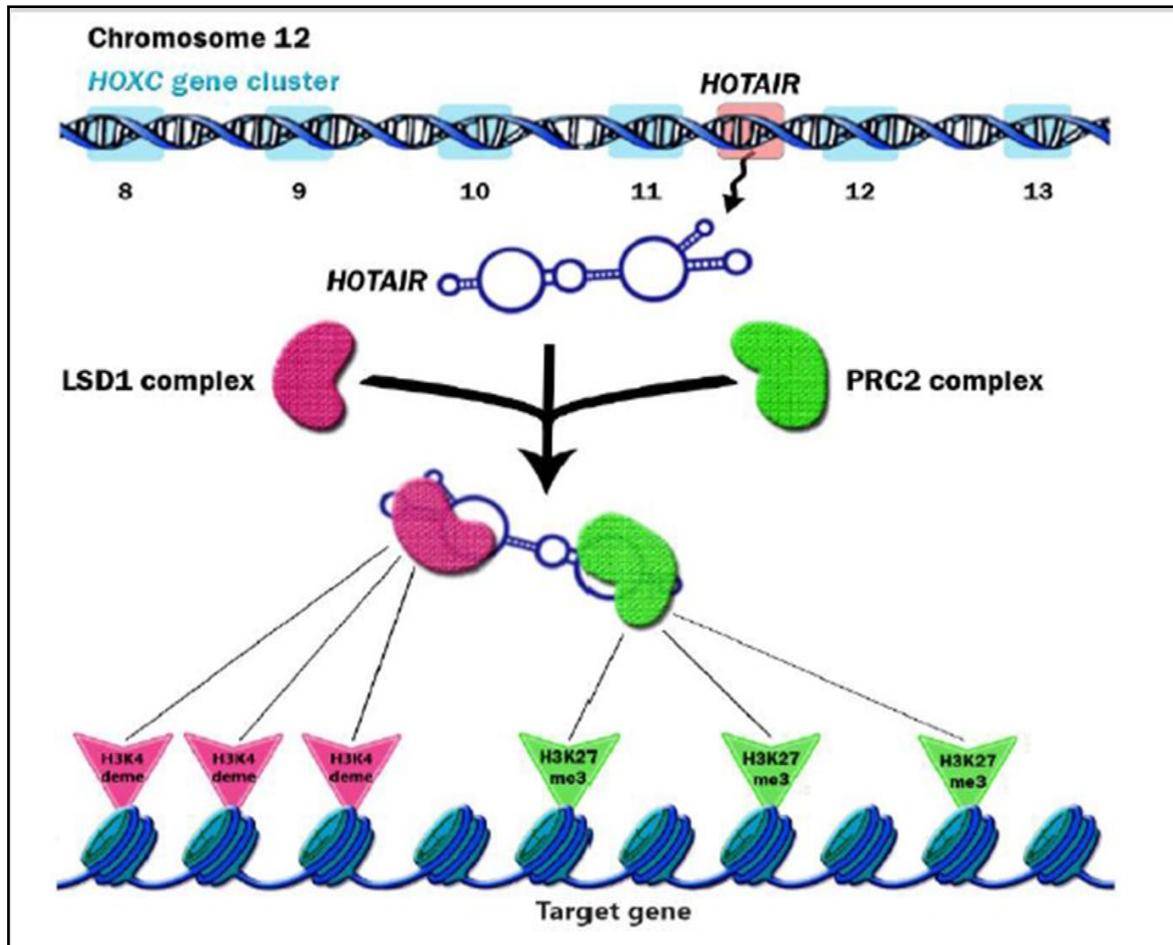
HOTAIR est un ARNnc impliqué dans la régulation des gènes du développement HOX. Ce dernier, décrit pour la première fois en 2007 (**Rinn et al., 2007 ; Woo et Kingston., 2007**), est un ARN de 2,2 kb, épissé (6 exons) et polyadénylé, il est transcrit à partir du chromosome 12 (12q13.13) dans le locus HOXC sur le brin antisens de la région intergénique entre les gènes HOXC 11 et HOXC 12. (Figure 08). HOTAIR est co-exprimé avec l'ensemble du locus HOXC, son profil d'expression est très proche de celui du gène HOXC 11. (**Denoeux., 2015**).



**Figure 08** : Position génomique du gène HOTAIR, Bandes selon Ensemble et localisation selon GeneLoc (2017).

Depuis sa découverte, HOTAIR a été beaucoup étudié du fait que sa surexpression a été observée dans de nombreux cancers et constitue un marqueur du potentiel métastatique. (**Gupta et al., 2010 ; Geng et al., 2011 ; Nakagawa et al., 2013**).

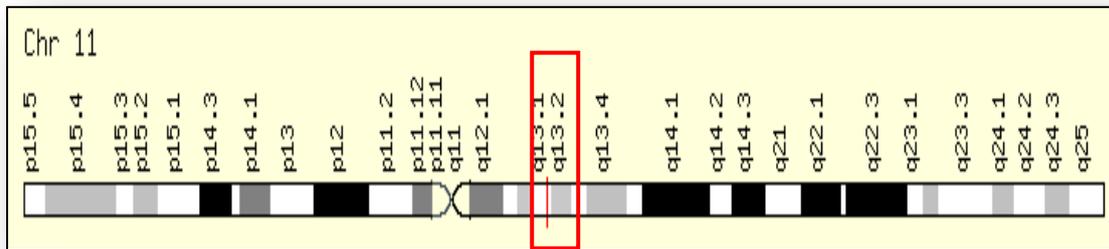
Au niveau du mécanisme d'action, HOTAIR est capable de recruter des complexes modificateurs d'histones. Le recrutement du complexe répresseur Polycomb PRC2 qui apporte la méthylase d'histone EZH2 induit la triméthylation de H3K27 (**Tsai et al., 2010**). Le complexe PRC2 interagit avec HOTAIR grâce à une structure en tige boucle située en 5'. Mais l'ARN HOTAIR possède également une seconde structure cette fois-ci en 3' qui permet le recrutement d'un second complexe protéique, LSD1 (**Tsai et al., 2010**). LSD1 est une histone deméthylase qui va permettre de retirer la marque de l'euchromatine H3K4me3. Ceci permet à HOTAIR d'apporter sur le locus HOXD les complexes permettant de retirer les marques d'activation et d'ajouter les marques du silencing sur les histones (Figure 09).



**Figure 09** : Schéma du mode d'action de l'ARN non codant HOTAIR (Hajjari et Salavaty., 2015).

### V.2.MALAT-1 (Meatastasis Associated Lung Adenocarcinoma Transcript 1)

MALAT-1 est un long ARNnc qui a été initialement découvert du a sa surexpression dans le cancer du poumon de type non petites cellules métastatiques (Ji et al., 2003). Appelé aussi NEAT2 (Nuclear Enriched Abundant Transcript 2), il est hautement conservé et connue pour être localisé dans le noyau et plus précisément dans les granules nucléaires. Ce long ARNnc est d'une longueur d'environ 8.7 kb, transcrit à partir du chromosome 11 (11q13.1) (Chen et al., 2014). (Figure 10).



**Figure 10:** Position génomique du gène MALAT1, Bandes selon Ensemble et localisation selon GeneLoc (2017).

Des études récentes ont révélé que MALAT-1 régule l'épissage alternatif grâce à son interaction avec la famille des phosphoprotéines nucléaires riches en serine / arginine (SR) qui sont impliquées dans le mécanisme d'épissage. Il module la distribution des facteurs d'épissage d'ARN et affecte particulièrement l'état de phosphorylation des protéines SR. (Wapinski et al., 2011).

MALAT-1 a également été lié au contrôle transcriptionnel de l'expression des gènes du cycle cellulaire. (Gutschner et al., 2013).

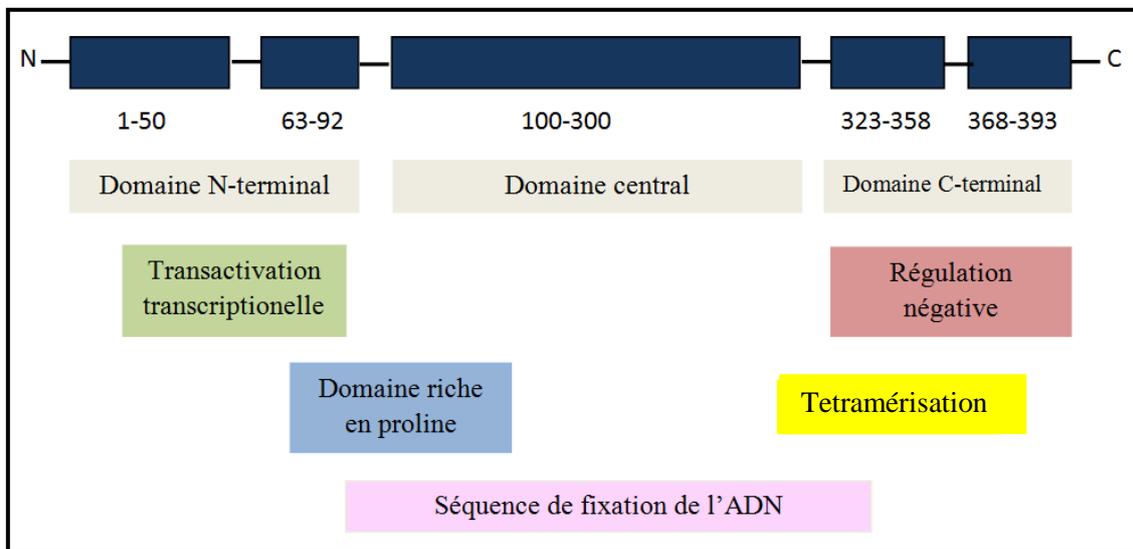
## VI. Description de la P53

Le gène p53 est localisé sur le bras court (p) du chromosome 17, au locus 17p13.1, sa longueur est de 20303 paires de bases réparties en 11 exons.

Il encode pour une protéine TP53 (Tumour Protein of 53 kilodaltons), de 393 acides aminés, cette protéine est codée par les exons 2 à 11 du gène. La majeure partie codante du gène est concentrée au sein d'une zone de 3066 paires de bases. (Chopin et al., 2005).

La protéine p53 comporte plusieurs domaines ayant chacun un rôle différent.

- ☞ La région N-terminale confère à p53 l'activité de facteur de transcription.
- ☞ Le domaine central reconnaît et lie des séquences cibles spécifiques d'ADN grâce à une séquence palindromique de 10 paires de bases.
- ☞ L'extrémité C-terminale reconnaît des séquences monocaténares d'ADN endommagé, avec les quelles elle forme des complexes stables. (Chopin et al., 2005). (Figure 11).



**Figure 11** : Structure schématique de la protéine p53 avec ses 5 domaines fonctionnels actifs. (Jemaa., 2012).

Le gène p53 est le gène le plus fréquemment muté dans les cancers humains. Paradoxalement, dans les cellules tumorales, il existe souvent une surexpression du gène p53, avec, dans de nombreux cas une accumulation de la protéine p53 anormale qui forme des complexes inactifs avec la protéine p53 normale. Le gène p53 est altéré dans plus de 50 % des cancers et dans 70 % des cancers du poumon.

### VI.1. Mutations du gène P53

Pour le gène p53, plus de 300 codons sur les 393 peuvent être modifiés. Ainsi, les mutations de p53 sont pour la plupart des mutations faux-sens et localisées dans la région de liaison à l'ADN, une seule mutation est donc capable de causer une perte de sa fonction. (Soussi et al., 2000).

Quatre positions dans p53 sont fréquemment mutées, ces points critiques sont localisés sur les acides aminés suivants : 175, 248, 249 et 273 et représentent plus de 25% de toutes les mutations faux-sens des cancers humains (Jemaa., 2012).

Deux positions dans la p53 sont spécifiques au cancer bronchique, ces points critiques sont localisés sur les acides aminés suivants 157 et 158 (**Soussi et al., 2000**).

Les mutations sur les positions 248 et 273 empêchent le contact direct de p53 sur l'ADN sans pour autant affecter la conformation même de p53 alors que les mutations sur les sites 175 et 249 induisent des distorsions à l'échelle locale et/ou globale de la structure de p53 et empêchent par conséquent la liaison à l'ADN. Représentent plus de 25% de toutes les mutations faux-sens des cancers humains. Les mutations spécifiques de p53 sont sub-classées en mutations de contact et mutations structurales mais les deux empêchent les fonctions de transactivation pro-apoptotiques de la p53 en empêchant la reconnaissance de son promoteur de ses gènes cibles. (**Jemaa., 2012 ; Donzelli et al., 2008 ; Brosh et Rotter., 2009**).

### **VII. Etude bioinformatique**

La bioinformatique 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. (**Quninkal et Rechenmann., 2004**).

Il s'agit d'une discipline récente, qui prend ses origines au début des années 1980 lorsque le laboratoire européen de biologie moléculaire (EMBL : European Molecular Biology Laboratory) et le département américain de la santé (NIH : National Institute of Health) ont créé les banques de données EMBL et GENBANK pour répertorier les séquences d'ADN découvertes par les biologistes. (**Delgrange., 2009**).

La bioinformatique a plusieurs rôles tel que :

- Organisation et diffusion de l'information génétique.
- Comparaison de séquences.
- Séquençage d'ADN.
- Localisation des gènes.
- Localisation de zones répétitives.
- Prédiction de structures 2D et 3D de l'ARN fonctionnel et des protéines.
- Phylogénie.
- Modélisation des réseaux de régulation.

## ☞ Cadre de l'étude

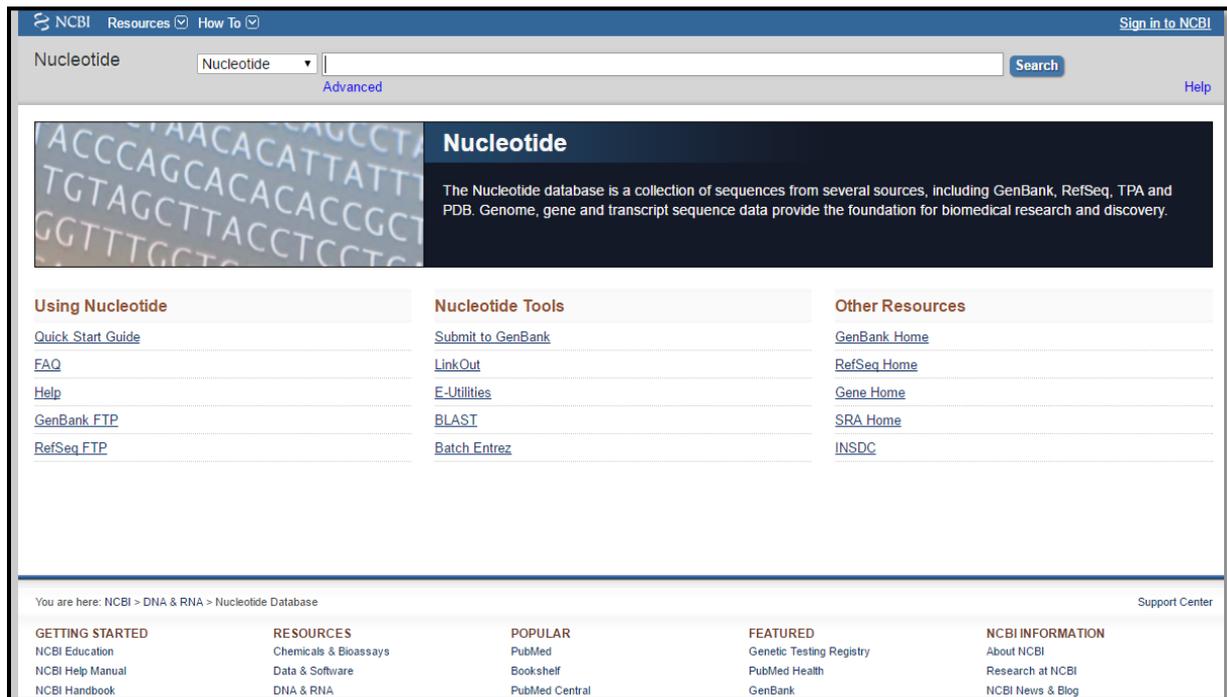
Elle a été réalisée au sein de l'université de SAAD DAHLEB BLIDA-1 dans le cadre de la préparation d'un projet de fin d'étude portant sur « *deux ARN non codons HOTAIR/MALAT-1 dans l'expression du cancer du poumon* » Elle a été menée sur une période de *4mois*, allant du mois de *Mars 2017* au mois *Juillet 2017*.

## I.MATERIEL

Le monde des ARN non codant ne cesse de nous surprendre jour après jour, et surtout après la découverte de la nouvelle classe des Longs ARN non codant, qui crée en elle-même un autre monde très vaste, devenu ces dernières années le sujet de nombreuses recherches les révélant être acteurs dans différents processus physiologiques et cellulaires ainsi que dans les pathologies. Différents outils bioinformatiques étaient a notre porté afin de réaliser notre étude.

### I.1.GenBank

La première partie de notre travail consiste à télécharger les séquences de la base de données nucléotidique GenBank ; base de données généraliste, c'est une collection annotée de toutes les séquences d'ADN accessibles au public (Figure. 12). Cette base de données américaine fait partie de L'INSDC (International Nucleotide Sequence Database). Collaboration qui comprend la DNA Data Bank of Japan (DDBJ), European Bioinformatics Institute (EBI) et The National Center for Biotechnology Information (NCBI).



**Figure 12** : Page d'accueil de GenBank (Capture d'écran .,2017).

## I.2.Séquences

Le tableau suivant résume les séquences étudiées avec leurs numéros d'accèsion sur la banque de donnée NCBI. (Voir séquences entières en annexe I).

**Tableau V** : Les accessions des différentes séquences étudiées.

Nom de la Séquence		Numéro d'accèsion sur NCBI	Origine des séquences
<b>HOTAIR</b>	Transcrit	DQ926657	saine
	Variant 1	NR_047517	mutée
	Variant 2	NR_003716	mutée
	Variant 3	NR_047518	mutée
<b>MALAT- 1</b>	Transcrit	NC_000011	saine
	Variant 1	NR_002819	mutée
	Variant 2	NR_144567	mutée
	Variant 3	NR_144568.	mutée
<b>P53</b>	CDS	X54156	mutée

### I.3. Base de données des mutations du gène p53

Afin d'identifier les mutations du cancer du poumon et des mutations associées nous avons consulté la base de données des mutations du gène p53.

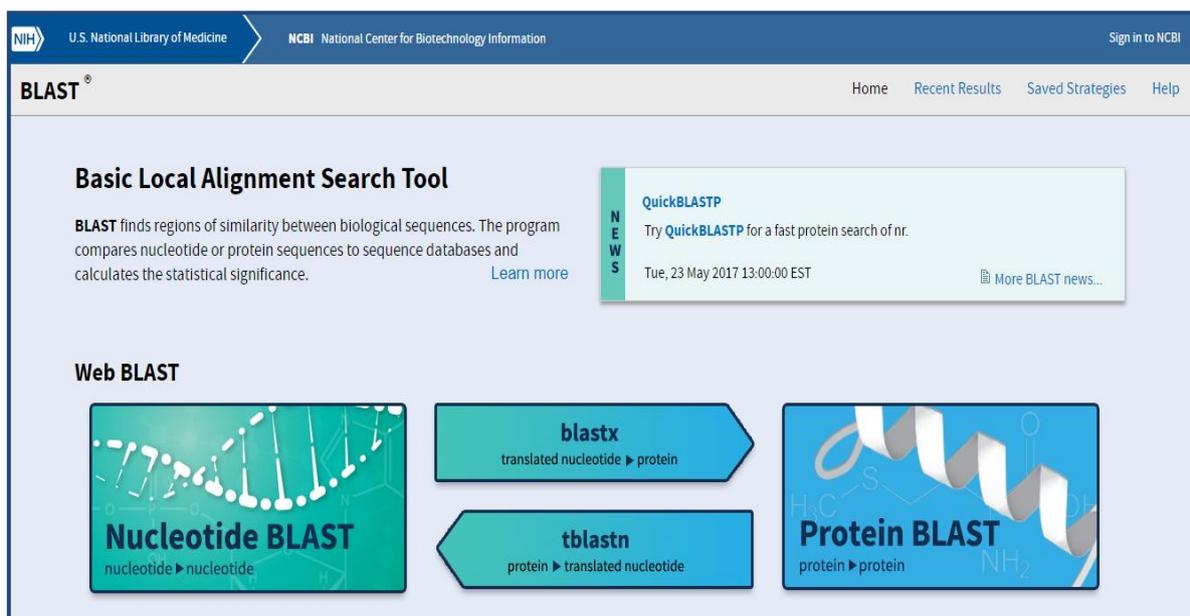
Cette base contient plus de 1505 mutations du gène p53 provenant de patients atteints de plusieurs dizaines de types de cancers différents. La base de données est téléchargeable à l'adresse suivante

(<http://acces.ens-lyon.fr/biotic/gpe/dossiers/p53/html/tabmutp53.htm>).

### I.4. Basic Local Alignment Search Tool (BLAST)

Pour aligner les séquences des transcrits et des variants afin d'identifier les mutations nous avons utilisé l'outil BLAST.

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.



The screenshot shows the BLAST homepage with the following elements:

- Header: NIH U.S. National Library of Medicine and NCBI National Center for Biotechnology Information.
- Navigation: Home, Recent Results, Saved Strategies, Help.
- Main Title: **Basic Local Alignment Search Tool**
- Description: BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.
- QuickBLASTP News: Try QuickBLASTP for a fast protein search of nr. Tue, 23 May 2017 13:00:00 EST.
- Web BLAST Section:
  - Nucleotide BLAST**: nucleotide to nucleotide
  - blastx**: translated nucleotide to protein
  - tblastn**: protein to translated nucleotide
  - Protein BLAST**: protein to protein

**Figure 13:** Page d'accueil de BLAST (Capture d'écran.,2017).

## 1.5. COSMIC (Catalogue Of Somatic Mutations In Cancer)

Le travail sur COSMIC consiste à voir si il y a une surexpression des ARNnc et ou elle se trouve.

COSMIC ; 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.

**COSMIC**  
Catalogue of somatic mutations in cancer

Home ▾ Resources ▾ Curation ▾ Tools ▾ Data ▾ News ▾ Help ▾ About ▾ Search COSMIC... Login ▾

View this page in our beta site

**COSMIC v81, released 09-MAY-17**

COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

Start using COSMIC by searching for a gene, cancer type, mutation, etc. below, or by browsing a region of the human genome using the map to the right.

eg: *Braf*, *COLO-829*, *Carcinoma*, *V600E*, *BRCA-UK*, *Campbell* **SEARCH**

**R Resources**  
Key COSMIC resources

- [COSMIC <sup>Beta</sup>](#)
- [Cell Lines Project](#)
- [COSMIC](#)
- [Cancer Gene Census](#)
- [Drug Sensitivity](#)
- [Mutational Signatures](#)
- [GRCh37 Cancer Archive](#)

**T Tools**  
Additional tools to explore COSMIC

- [Cancer Browser](#)
- [Genome Browser](#)
- [GA4GH Beacon](#)
- [COSMIC-3D](#)
- [COSMIC In BigQuery <sup>New</sup>](#)
- [CONAN](#)

Browse the **genomic landscape** of cancer

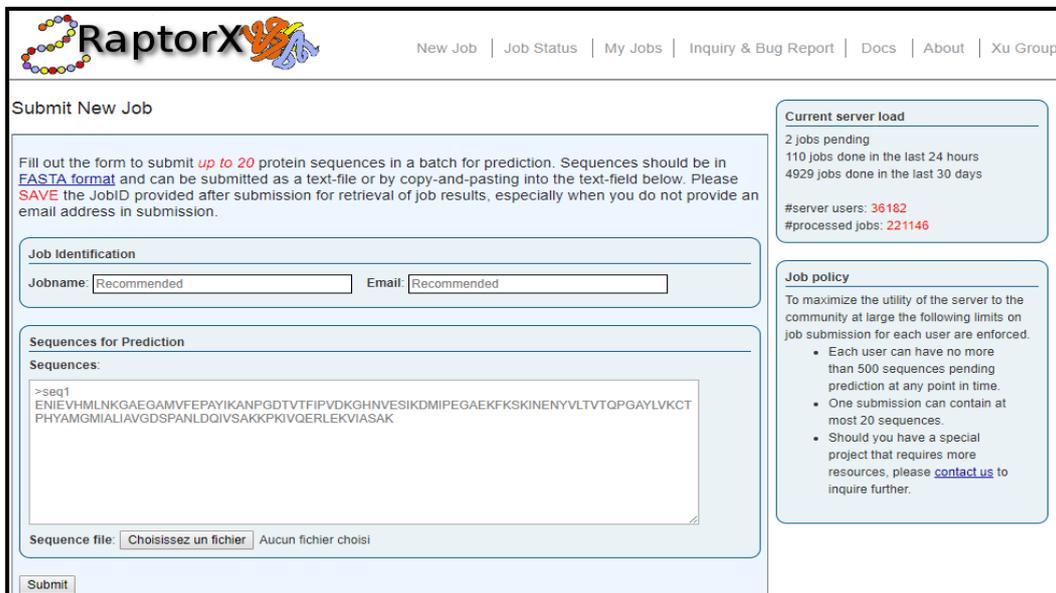
News

**Figure 14** : Page d'accueil de COSMIC. (Capture d'écran.,2017).

## I.6.RaptorX

Pour avoir la structure tridimensionnelle de la protéine P53 nous avons utilisé RaptorX. (Annexe III).

C'est est une plateforme qui a un logiciel intégré (RaptorX Structure Prediction), pour la structure des protéines et la prédiction de leurs fonction. Il est gratuit et non commerciale, c'est l'une des méthodes les plus utilisée pour la prédiction de la structure des protéines.



The screenshot shows the RaptorX website interface. At the top, there is a navigation menu with links: New Job, Job Status, My Jobs, Inquiry & Bug Report, Docs, About, and Xu Group. The main content area is titled 'Submit New Job' and contains a form for submitting protein sequences. The form includes a 'Job Identification' section with 'Jobname' and 'Email' fields, and a 'Sequences for Prediction' section with a text area containing a protein sequence. A 'Submit' button is located at the bottom left of the form. On the right side, there is a 'Current server load' box showing 2 jobs pending, 110 jobs done in the last 24 hours, and 4929 jobs done in the last 30 days. Below this is a 'Job policy' box with a list of rules for users.

**Current server load**  
2 jobs pending  
110 jobs done in the last 24 hours  
4929 jobs done in the last 30 days  
#server users: 36182  
#processed jobs: 221146

**Job policy**  
To maximize the utility of the server to the community at large the following limits on job submission for each user are enforced.

- Each user can have no more than 500 sequences pending prediction at any point in time.
- One submission can contain at most 20 sequences.
- Should you have a special project that requires more resources, please [contact us](#) to inquire further.

Figure 15 : Page d'accueil de RaptorX. (Capture d'écran.,2017).

### Matériel informatique :

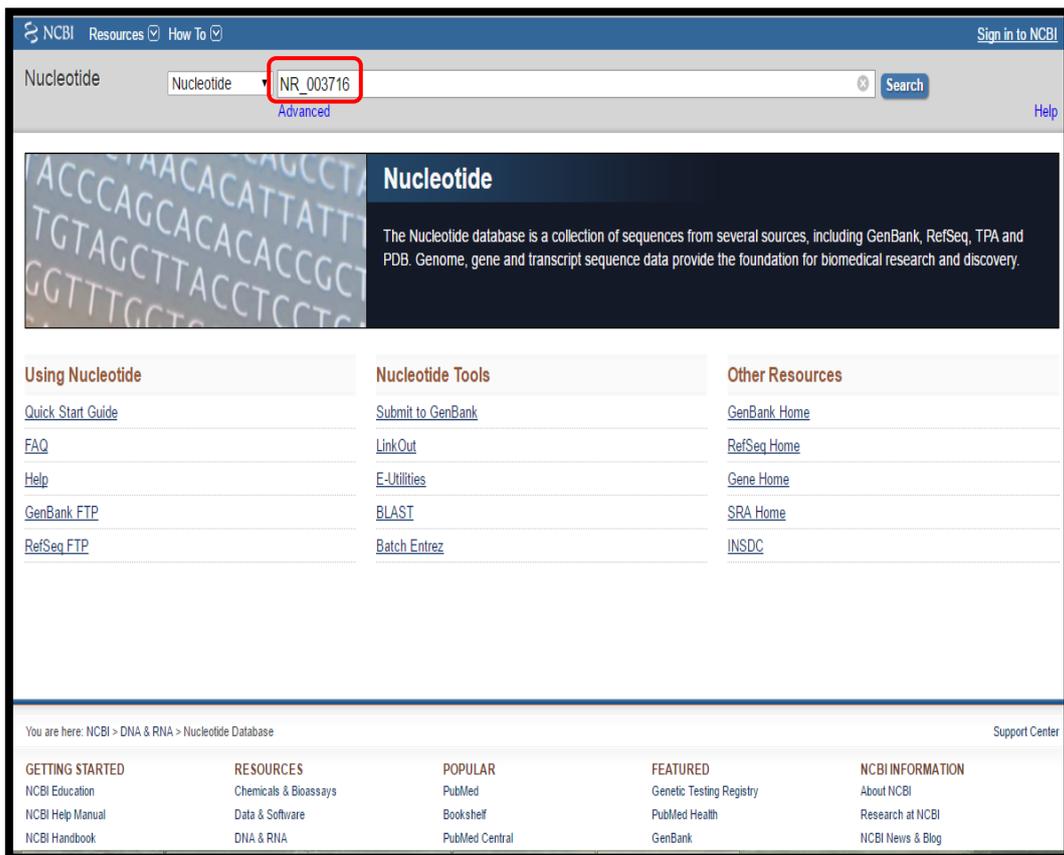
-ordinateur.

-internet.

## II. Méthodes

### ➤ Utilisation de la GenBank

Pour télécharger nos séquences nous avons utilisé Genbank, en passant sur un moteur de recherche, puis insérer un numéro d'accésion comme montrer sur la figure suivante.



**Figure 16** : Utilisation de GenBank (Capture d'écran.,2017).

## ➤ Utilisation de Blastn ; tBlastx

Le logiciel BLAST utilisé pour l'alignement entre les différentes séquences à son tour des sous logiciel dont ceux que nous avons utilisés :

- Blastn : Alignement entre séquences nucléotidiques.
- tBlastx : Alignement d'une séquence nucléotidique "traduite" contre les séquences nucléotidiques "traduites" de la banque.

L'utilisation de BLAST se fait comme suit

- 1- Choisir le type de blast selon le besoin.
- 2- Introduire la séquence a étudié (sous la forme fasta ou directement grâce a un numéro d'accession).
- 3- Choisir la base de comparaison.
- 4- Lancer l'algorithme d'alignement.

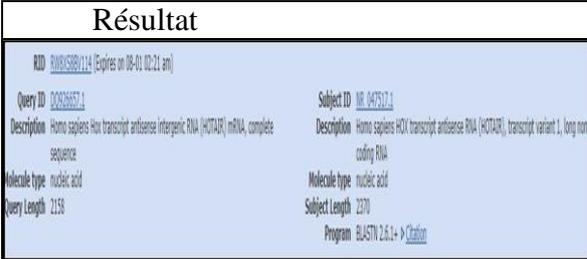
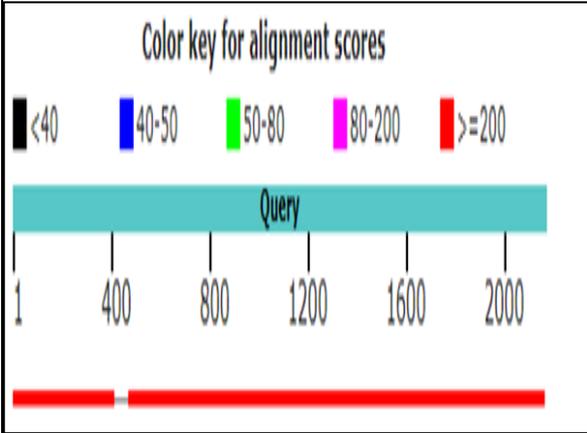
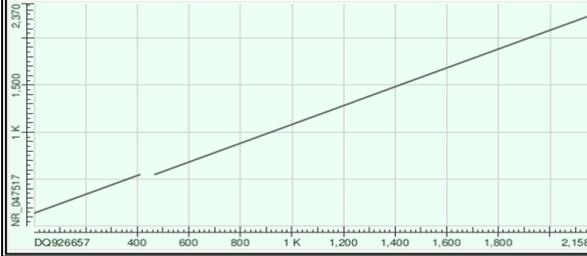
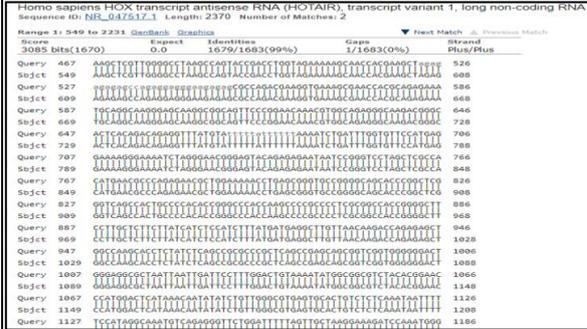
The screenshot shows the 'Standard Nucleotide BLAST' web interface. It is divided into several sections:

- blastn blastp blastx tblastn tblastx**: A navigation bar at the top with a red arrow pointing to 'blastn' and a circled '1' next to it.
- Enter Query Sequence**: A section with a text input field containing 'DQ926657', a 'Clear' button, and a 'Query subrange' section with 'From' and 'To' fields. A red arrow points to the input field with a circled '2'.
- Or, upload file**: A section with a file selection button and a 'Job Title' input field.
- Choose Search Set**: A section with radio buttons for 'Human genomic + transcript', 'Mouse genomic + transcript', and 'Others (nr etc.)'. A dropdown menu is set to 'Nucleotide collection (nr/nt)', with a red arrow pointing to it and a circled '3'.
- Program Selection**: A section with radio buttons for 'Highly similar sequences (megablast)', 'More dissimilar sequences (discontiguous megablast)', and 'Somewhat similar sequences (blastn)'. A 'BLAST' button is at the bottom left, with a red arrow pointing to it and a circled '4'.

**Figure 17** : Méthodes d'utilisation de BLAST. (Capture d'écran.,2017).

Les résultats de BLAST se présentent comme dans le tableau IV

Tableau VI : Les résultats de BLAST.

Résultat	Rôle	Description
	Récapitulatif de la requête.	<ul style="list-style-type: none"> <li>Query : séquence soumise.</li> <li>Database name : banque interrogée.</li> </ul>
	Représentation graphique des résultats.	<ul style="list-style-type: none"> <li>Les traits de couleur correspondent à un alignement entre les deux séquences.</li> <li>La couleur correspond au score.</li> <li>La longueur correspond à la taille de l'alignement.</li> </ul>
	Dot Matrix View.	<ul style="list-style-type: none"> <li>Montre des régions de similarité basées sur les résultats BLAST.</li> </ul>
	Représentation globale des résultats.	<ul style="list-style-type: none"> <li>Chaque ligne correspond à trait coloré de la représentation graphique</li> <li>On trouve dans ce tableau : le score, l'E-value.</li> </ul>
	Description des résultats des alignements.	<ul style="list-style-type: none"> <li>Les grandeurs propres à chaque alignement (score, E-value, % d'identité, % positif, % de gap)</li> </ul>

## A- RESULTATS

### I. Analyse de HOTAIR / MALAT-1

Les différentes études ont démontré que les deux ARNnc MALAT-1 / HOTAIR ont une action directe sur le système de régulation / réparation de l'ADN passant par une action sur la p53 d'où une surexpression de ces deux derniers empêcherait ou inhiberait son action.

Ainsi, la version transcrite de MALAT-1 / HOTAIR extraite d'une cellule saine est d'une taille plus petite que celles des trois variant extrais de cellules tumorales. (Tableau VI).

**Tableau VII** : Comparaison entre les tailles de MALAT-1 / HOTAIR.

	Transcrit (Pb)	Variant 1 (Pb)	Variant 2 (Pb)	Variant 3 (Pb)
<b>HOTAIR</b>	2158	2370	2364	2337
<b>MALAT-1</b>	6816	8779	8545	8302

Aussi notre analyse comparative entre les différents variants a mit en évidence plusieurs mutations de nos deux ARNnc comme le montre le tableau VII.

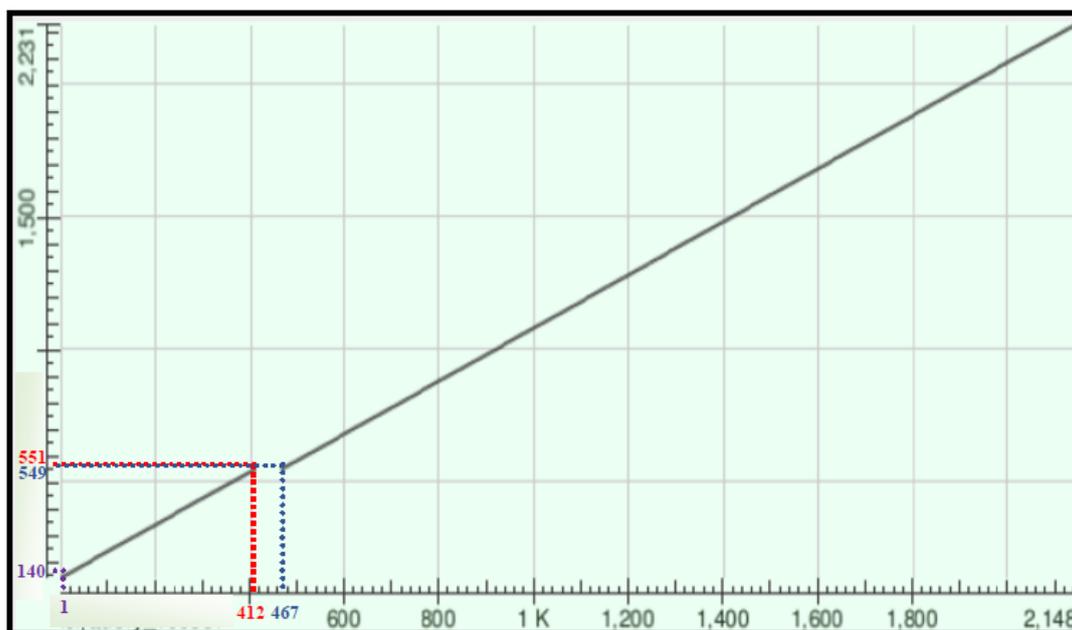
**Tableau VIII** : Les mutations des ARNnc.

	Séquences	N° de la base	La base avant	La base après
<b>HOTAIR</b>	Transcrit/variant1	137	G	A
		1251	C	G
		1759	G	A
		1988	C	T
		1243	-	G
	Transcrit/variant2	137	G	A
		1251	C	G
		1759	G	A
		1988	C	T
		1243	-	G
		431..434	AATT	----
	Transcrit /variant3	1251	C	G
		1759	G	A
		1988	C	T
		1243	-	G
287..290		AGGA	----	
<b>MALAT-1</b>	Transcrit/variant1	Aucun	Aucun	Aucun
	Transcrit/variant2	Aucun	Aucun	Aucun
	Transcrit /variant3	Aucun	Aucun	Aucun

L'analyse comparative des séquences de HOTAIR et MALAT-1 forme transcrite avec leurs trois variants, a mis en évidence des différences structurales qui pourraient expliquer ce phénomène de surexpression ainsi l'appariation de cancers.

## II. Analyse séquentielle de HOTAIR

### II.1. Transcrit HOTAIR /variant1



**Figure 18** : Dot Matrix View du Transcrit/variant 1 de HOTAIR.

Le nombre de lignes présentées dans le plot est identique aux nombres d'alignements trouvés par BLAST. La superposition de HOTAIR au variant 1, nous donne deux alignements :

- ❖ 1<sup>er</sup> alignement : [1-412].
- ❖ 2<sup>ème</sup> alignement : [467-2148].

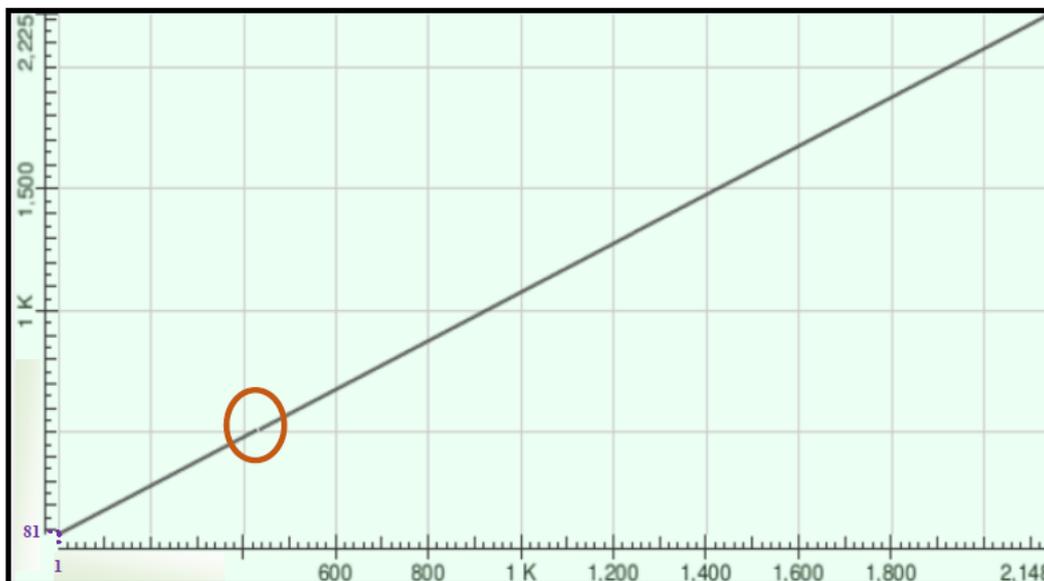
Sachant que le 1<sup>er</sup> nucléotide du transcrit correspond au nucléotide 140 du variant, et 467<sup>ème</sup> correspond au 549<sup>ème</sup> du variant.

Aussi l'alignement des séquences du transcrit de HOTAIR au variant 1 nous a permis l'identification de mutations entre les deux, au niveau des bases (137;1243;1251;1757;1988) comme le montre la figure suivante:



**Figure 19** : Différentes mutations détectées entre le transcrit de HOTAIR et le variant 1.

## II.2. Transcrit HOTAIR /variant 2

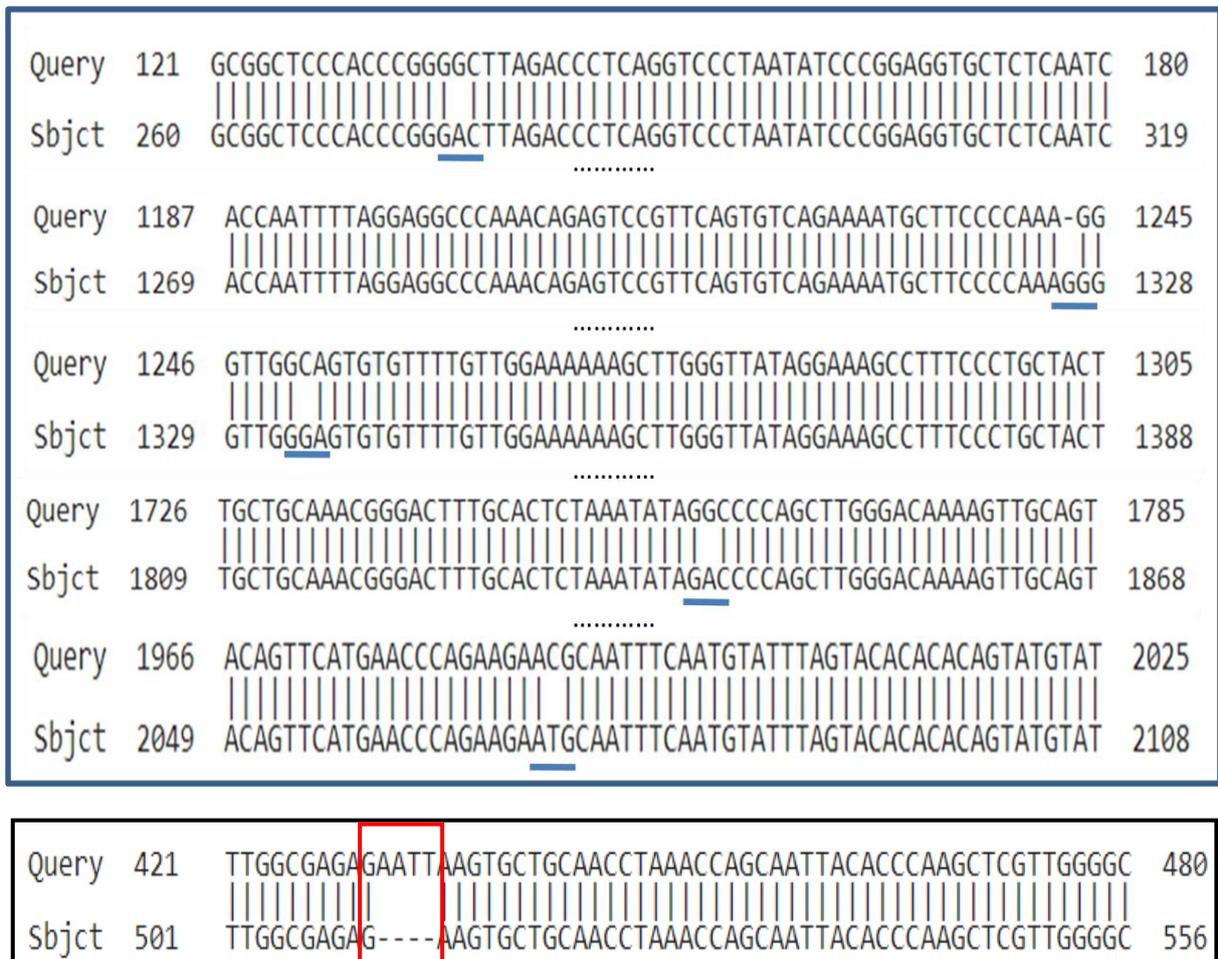


**Figure 20** : Dot Matrix View du transcrit et variant 2 de HOTAIR. (2017).

Nous avons trouvés un seul alignement entre le transcrit de HOTAIR et le variant 2 : [1-2148]. Sachant que le nucléotide 1 du transcrit correspond au nucléotide 81 du variant 2.

La partie encerclée dans la figure 21 montre une délétion du variant 2 mieux identifiée dans la figure 22.

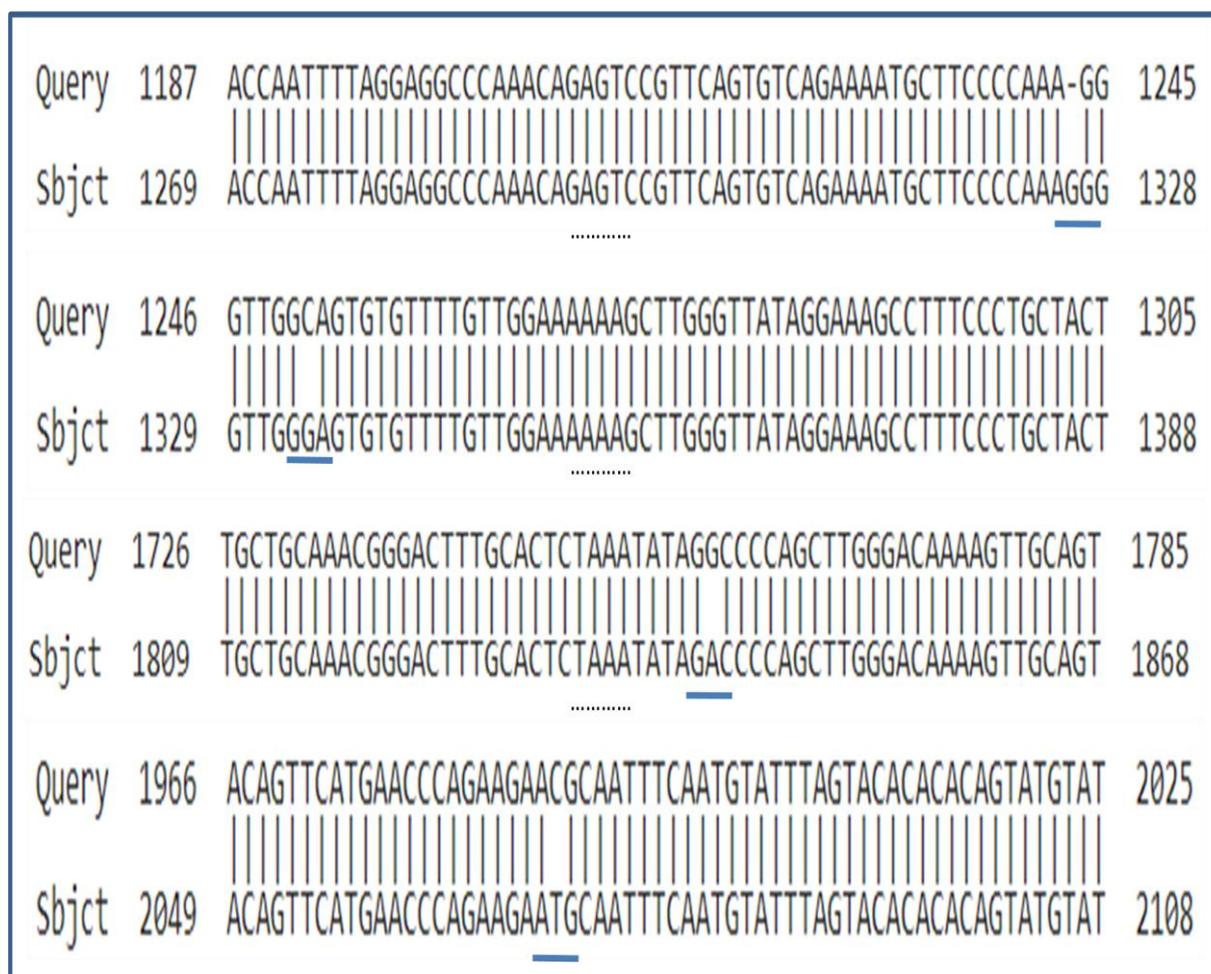
En alignant les séquences du transcrit de HOTAIR au variant 2 Nous avons identifié des mutations similaires aux mutations trouvées entre le transcrit de HOTAIR et le variant1 (137;1243;1251;1757;1988) plus une mutation des bases allant de 431 à 434.



**Figure 21** : Mutation des bases allant de 431 à 434 entre le transcrit de HOTAIR et le variant2.



En alignant les séquences du transcrit de HOTAIR au variant 3 nous avons identifié des mutations similaires aux mutations trouvées entre le transcrit de HOTAIR et variant 1, (1243;1251;1757;1988) plus une mutation des bases allant 287 à 290.



**Figure 24** : Mutation des bases allant 287 à 290 entre le transcrit de HOTAIR et le variant 3.

### III. Analyse de MALAT-1

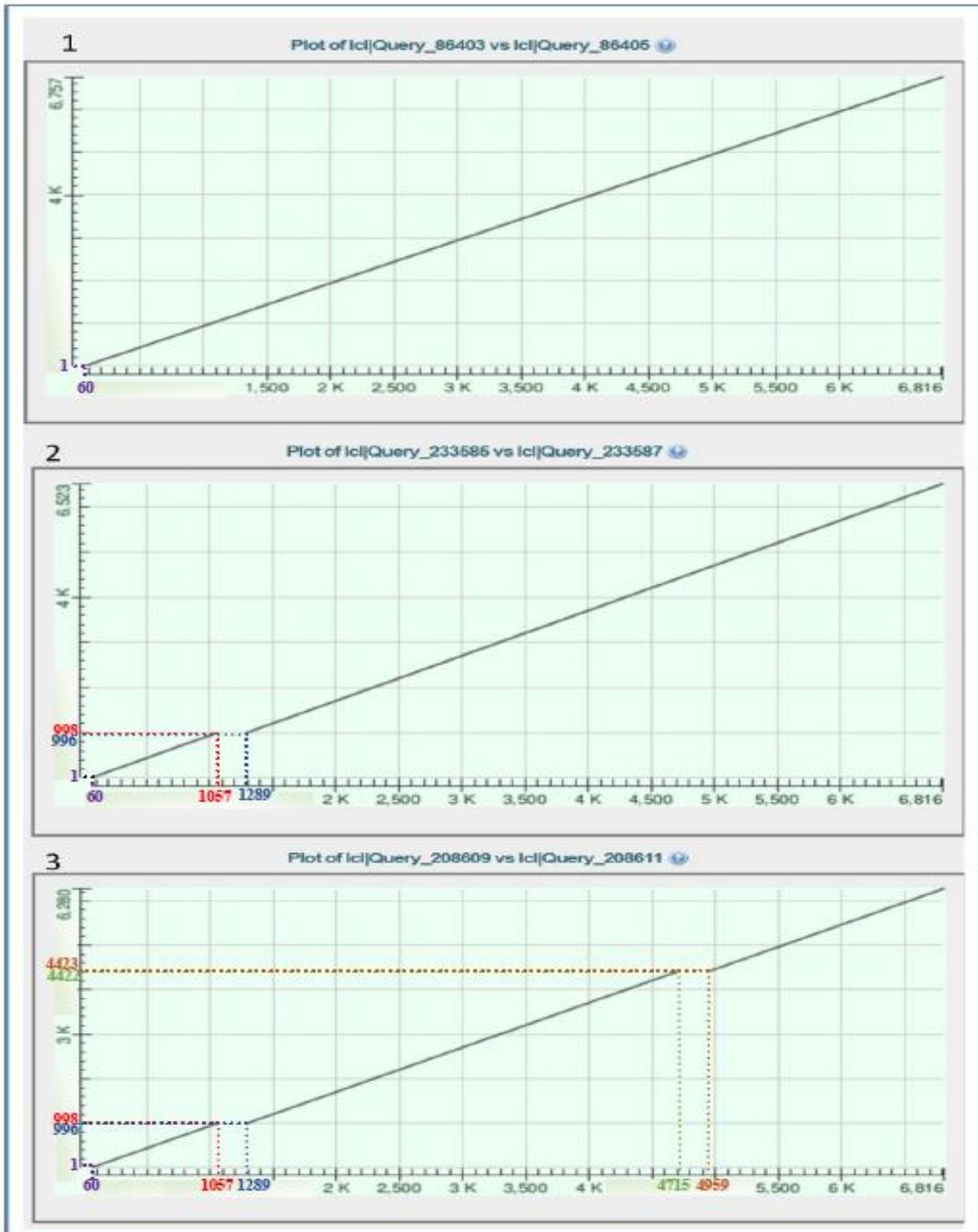


Figure 25 : Dot Matrix View Transcrit de MALAT-1 avec les 3 variants.

Le nombre de lignes présentées dans le plot est identique aux nombres d'alignements trouvés par BLAST. La superposition de MALAT-1 au variant 1 ;

Nous donne un seul alignement :

- ❖ Le 1<sup>er</sup> : [60-6816], sachant que le 60<sup>ème</sup> nucléotide du transcrit correspond au 1<sup>er</sup> nucléotide du variant 1.

Concernant le 2<sup>ème</sup> variant, nous avons trouvé deux alignements :

- ❖ Le 1<sup>er</sup> alignement : [60-1057].
- ❖ Le 2<sup>ème</sup> alignement : [1289-6816].

Le 60<sup>ème</sup> nucléotide du transcrit correspond au 1<sup>er</sup> nucléotide du variant 2 et le 1289<sup>ème</sup> correspond au 996<sup>ème</sup> nucléotide.

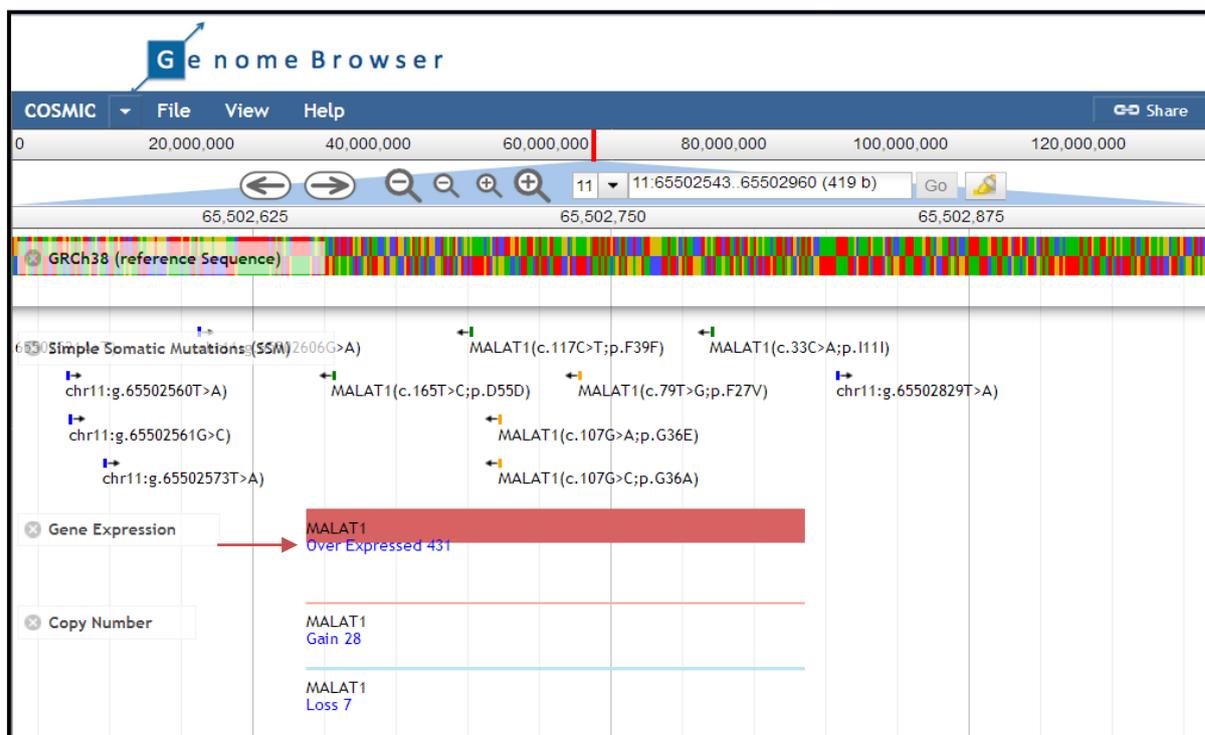
Enfin le 3<sup>ème</sup> variant nous montre trois alignements :

- ❖ Le 1<sup>er</sup> alignement : [60-1057].
- ❖ Le 2<sup>ème</sup> alignement : [1289-4715].
- ❖ Le 3<sup>ème</sup> alignement : [4959-6816].

Le 60<sup>ème</sup> nucléotide du transcrit correspond au 1<sup>er</sup> nucléotide du variant 3, le 1289<sup>ème</sup> correspond au 996<sup>ème</sup> et le 4959<sup>ème</sup> correspond au 4423<sup>ème</sup> nucléotide.

#### IV. Surexpression de MALAT -1 /HOTAIR

Différentes études montre qu'il y a une surexpression de MALAT-1 et HOTAIR dans divers cancers dont le cancer du poumon comme montrer dans la figure suivante



**Figure 26** : Surexpression de MALAT-1.

D'après l'analyse de la base de donnée de COSMIC concernant le phénomène de surexpression de MALAT-1, il a été montré que ce dernier a subit une perte de 7 bases et gains de 28 bases avec une surexpression des bases 431.

Une recherche sur COSMIC a été effectuée pour trouver la surexpression de HOTAIR mais ça n'a pas encore été répertorié.

## V. Protéine P 53

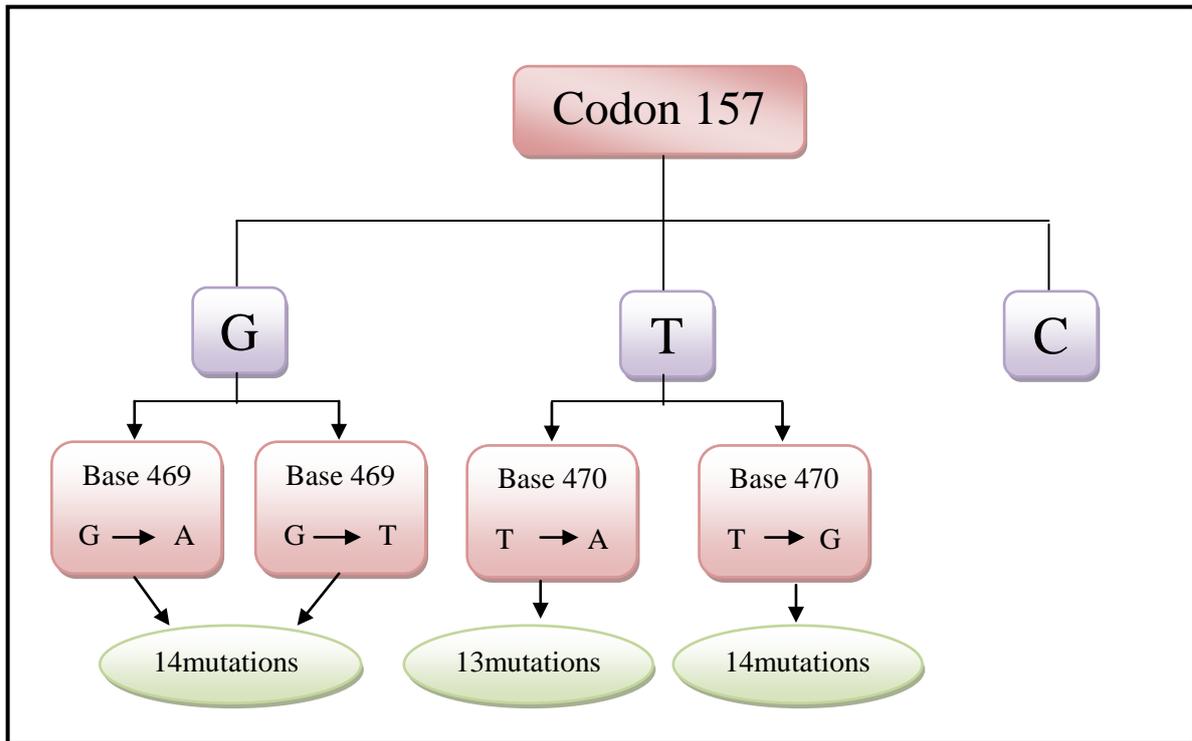
D'après les différentes études et recherches effectuées, plusieurs mutations ont été répertoriées comme faux sens, comme le montre le tableau VIII :

**Tableau IX** : Les mutations faux sens de la P53. (Soussi.,2000).

codon	mutation	séquence 5' → 3'	fréquence mutations
<b>TGG</b>	W53X	GAT ATT GAA CAA <b>TAG</b> TTC ACT GAA	0.6 %
<b>CAG</b>	Q144X	ACC TGC CCT GTG <b>TAG</b> CTG TGG GTT	2.4%
<b>TGG</b>	W146X	CCT GTG GAG CTG <b>TGA</b> GTT GAT TCC	4.3%
<b>CAG</b>	Q192X	CTG GCC CCT CTT <b>TAG</b> CAT CTT ATC	4.1%
<b>CGA</b>	R196X	CAG CAT CTT ATC <b>TGA</b> GTG GAA GGA	11.8%
<b>CGA</b>	R213X	AGA AAC ACT TTT <b>TGA</b> CAT AGT GTG	14.5%
<b>GAG</b>	E298X	GAG CTT CAC CAC <b>TAG</b> CTG CCC CCA	3.3%
<b>CGA</b>	R306X	GGG AGC ACT AAG <b>TGA</b> GCA CTG CCC	7.7%
<b>CAG</b>	Q317X	AGC TCC TCT CCC <b>TAG</b> CCA AAG AAG	1.3%
<b>TAT</b>	Y327X	CTG GAT GGA GAA <b>TAA</b> TTC ACC CTT	0.1%
<b>CGA</b>	R342X	TTC GAG ATG TTC <b>TGA</b> GAG CTG AAT	3.8%
<b>Toutes les mutations étudiées</b>			49.6%

La mutation la plus fréquente est la R213X changeant le codon **CGA** en **TGA** dont la fréquence est de 14.5%, et la mutation la moins fréquente est l'Y327X changeant le codon **TAT** en **TAA** avec une fréquence de 0.1%.

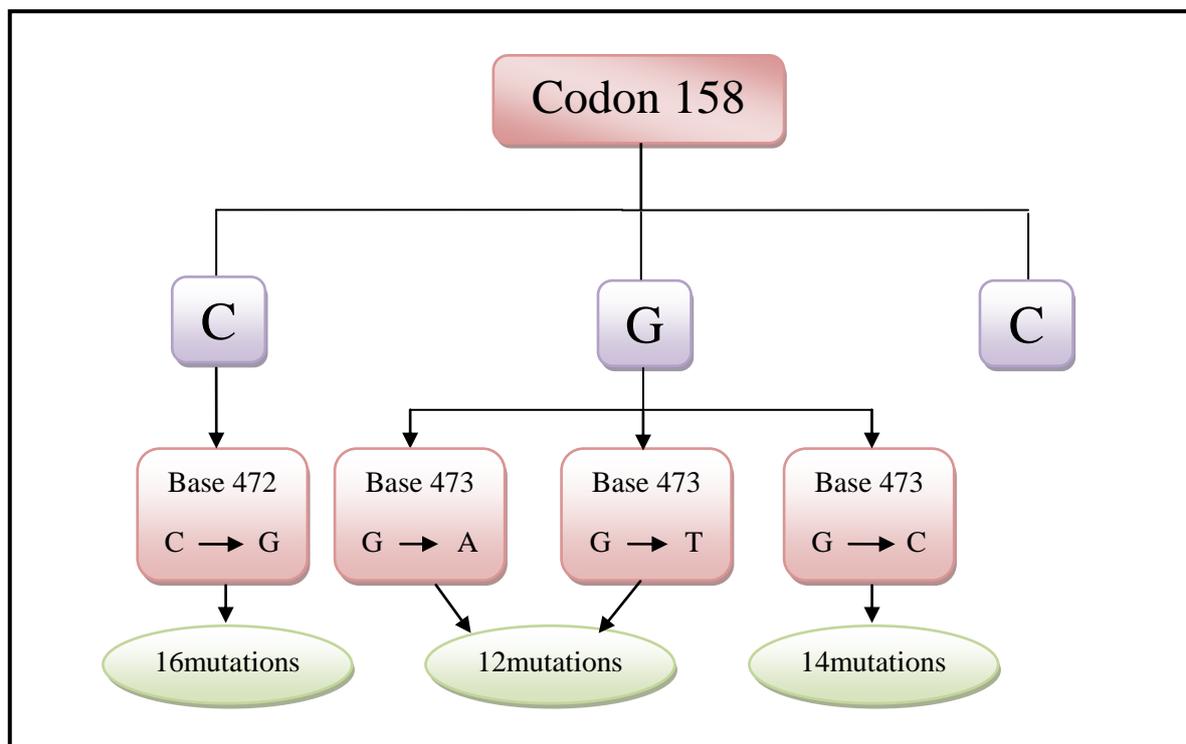
Les mutations spécifiques de la P53 concernant le cancer des poumons sont localisées sur les acides aminés suivants 157 et 158. L'identification de ces mutations nous a conduits à l'identification de plusieurs mutations associées s'exprimant par l'apparition d'autres types de cancers associés secondaires (à savoir des métastases).



**Figure 27** : Mutations au niveau du codon 157 de la P53.

Les mutations au niveau de la base 469 ( $G \rightarrow A$  et  $G \rightarrow T$ ) changeant la Valine en Isoleucine et Phénylalanine (tableau IX) donne 14 mutations associées.

Les mutations de la base 470 ( $T \rightarrow A$  et  $T \rightarrow G$ ) changeant la Valine en Acide aspartique et Glycine (tableau IX) nous ont respectivement révélé 13 et 14 mutations.



**Figure 28** : Mutations au niveau du codon 158 de la P53.

La mutation au niveau de la base 472 (C → G) changeant l’arginine en Glycine (tableau IX) donne 16 mutations associées.

Les mutations de la base 473 (G → A et G → T et G → C) changeant l’arginine en Histidine, Proline et Leucine (tableau IX) nous ont respectivement révélé 12 et 14 mutations.

**Tableau X** : Mutations de la p53 provoquant le cancer du poumon.

N° du codon	N° de la base qui change	Changement	Acide aminé muté
157 Valine	469-470	<u>G</u> T C → A T C	Ile
		<u>G</u> T C → T T C	Phe
		<u>G</u> T C → G A C	Asp
		<u>G</u> T C → G G C	Gly
158 Arginine	472 - 473	<u>C</u> G C → G G C	Gly
		<u>C</u> G C → C A C	His
		<u>C</u> G C → C C C	Pro
		<u>C</u> G C → C T C	Leu

Pour l'identification des codons 157,158, nous avons procéder à la traduction du CDS de la p53, en séquence d'acides aminées (voir séquence ci-dessous) :

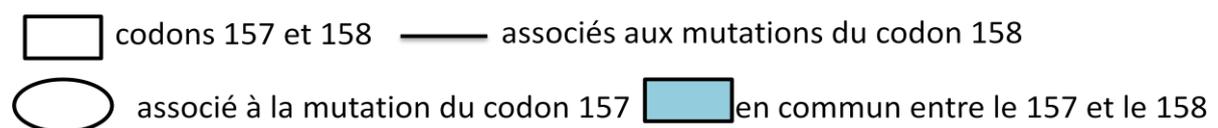
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGPDEA  
 PRMPEAAPRVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTY  
 SPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPPHHERCSDSDG  
 LAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCCTTIHYNMCMSSCMGGMNRR  
 PILTIITLEDSSGNLLGRNSFEVRVCACPGRRDRTEENLRKKGEPHHELPPGSTKRALPNNTSS  
 SPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPGGSRAHSSHLKSKKGQST  
 SRHKKLMFKTEGPDS

Ainsi les lettres colorées correspondent aux codons recherchés.

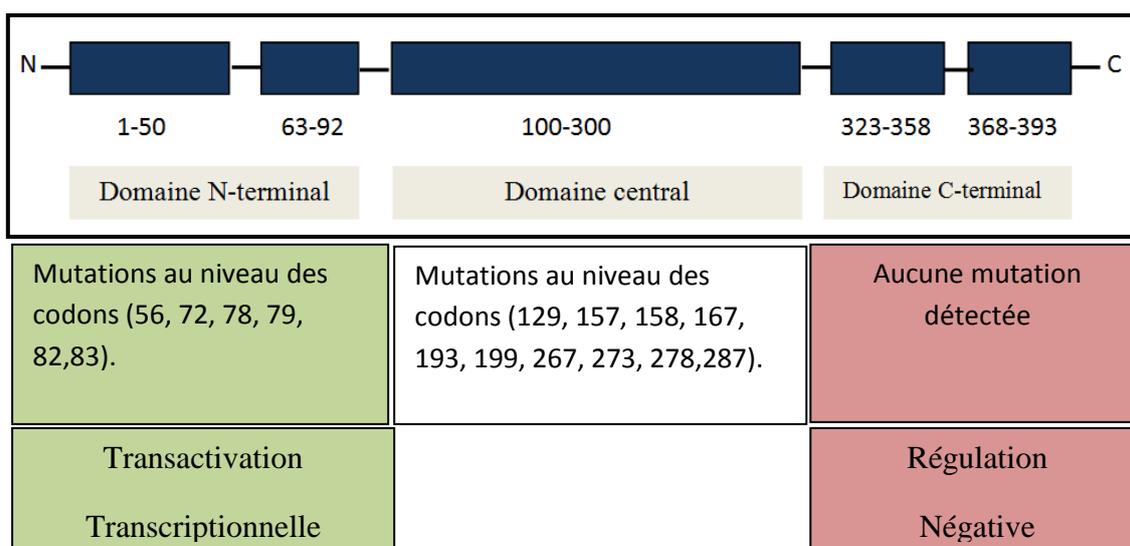
Pour pouvoir identifier la position des mutations au niveau nucléotidique, la séquence a été traduite grâce au logiciel tBlastx (Figure 30).

ATG	GAG	GAG	CCG	CAG	TCA	GAT	CCT	AGC	GTC	GAG	CCC	CCT	CTG	AGT	CAG	GAA	ACA	TTT
TCA	GAC	CTA	TGG	AAA	CTA	CTT	CCT	GAA	AAC	AAC	GTT	CTG	TCC	CCC	TTG	CCG	TCC	CAA
GCA	ATG	GAT	GAT	TTG	ATG	CTG	TCC	CCG	GAC	GAT	ATT	GAA	CAA	TGG	TTC	ACT	GAA	GAC
CCA	GGT	CCA	GAT	GAA	GCT	CCC	AGA	ATG	CCA	GAG	GCT	GCT	CCC	CGC	GTG	GCC	CCT	GCA
CCA	GCA	GCT	CCT	ACA	CCG	GCG	GCC	CCT	GCA	CCA	GCC	CCC	TCC	TGG	CCC	CTG	TCA	TCT
TCT	GTC	CCT	TCC	CAG	AAA	ACC	TAC	CAG	GGC	AGC	TAC	GGT	TTC	CGT	CTG	GGC	TTC	TTG
CAT	TCT	GGG	ACA	GCC	AAG	TCT	GTG	ACT	TGC	ACG	TAC	TCC	CCT	GCC	CTC	AAC	AAG	ATG
TTT	TGC	CAA	CTG	GCC	AAG	ACC	TGC	CCT	GTG	CAG	CTG	TGG	GTT	GAT	TCC	ACA	CCC	CCG
CCC	GGC	ACC	CGC	GTC	CGC	GCC	ATG	GCC	ATC	TAC	AAG	CAG	TCA	CAG	CAC	ATG	ACG	GAG
GTT	GTG	AGG	CGC	TGC	CCC	CAC	CAT	GAG	CGC	TGC	TCA	GAT	AGC	GAT	GGT	CTG	GCC	CCT
CCT	CAG	CAT	CTT	ATC	CGA	GTG	GAA	GGA	AAT	TTG	CGT	GTG	GAG	TAT	TTG	GAT	GAC	AGA
AAC	ACT	TTT	CGA	CAT	AGT	GTG	GTG	GTG	CCC	TAT	GAG	CCG	CCT	GAG	GTT	GGC	TCT	GAC
TGT	ACC	ACC	ATC	CAC	TAC	AAC	TAC	ATG	TGT	AAC	AGT	TCC	TGC	ATG	GGC	GGC	ATG	AAC
CGG	AGG	CCC	ATC	CTC	ACC	ATC	ATC	ACA	CTG	GAA	GAC	TCC	AGT	GGT	AAT	CTA	CTG	GGA
CGG	AAC	AGC	TTT	GAG	GTG	CGT	GTT	TGT	GCC	TGT	CCT	GGG	AGA	GAC	CGG	CGC	ACA	GAG
GAA	GAG	AAT	CTC	CGC	AAG	AAA	GGG	GAG	CCT	CAC	CAC	GAG	CTG	CCC	CCA	GGG	AGC	ACT
AAG	CGA	GCA	CTG	CCC	AAC	AAC	ACC	AGC	TCC	TCT	CCC	CAG	CCA	AAG	AAG	AAA	CCA	CTG
GAT	GGA	GAA	TAT	TTC	ACC	CTT	CAG	ATC	CGT	GGG	CGT	GAG	CGC	TTC	GAG	ATG	TTC	CGA
GAG	CTG	AAT	GAG	GCC	TTG	GAA	CTC	AAG	GAT	GCC	CAG	GCT	GGG	AAG	GAG	CCA	GGG	GGG
AGC	AGG	GCT	CAC	TCC	AGC	CAC	CTG	AAG	TCC	AAA	AAG	GGT	CAG	TCT	ACC	TCC	CGC	CAT
AAA	AAA	CTC	ATG	TTC	AAG	ACA	GAA	GGG	CCT	GAC	TCA	GAC	TGA					

Figure 29 : CDS de la P53 avec les différentes mutations détectées.



L'analyse de la partie codante de la p53 et des mutations au niveau des codons 157, 158 (nucléotides encadrés) à révélé l'existence de plusieurs autres mutations associées (nucléotides soulignés sont associés uniquement aux mutations du codon 158, et celui encerclé est associé qu'à la mutation du codon 157, le reste des nucléotides colorés uniquement sont en commun entre le 157 et le 158).



Remarquons que les mutations associées au niveau des codons (56, 72, 78, 79, 82,83) sont localisées dans le domaine N-terminal de la p53 qui est responsable de la transactivation transcriptionnelle contrairement aux mutations associées au niveau des codons (129, 157, 158, 167, 193, 199, 267, 273, 278,287) qui sont sur le domaine central.

Nous remarquons aussi qu'aucune mutation n'a été détecté dur le domaine c-terminal responsable de la régulation négative.

### V.1.Analyse du codon 157

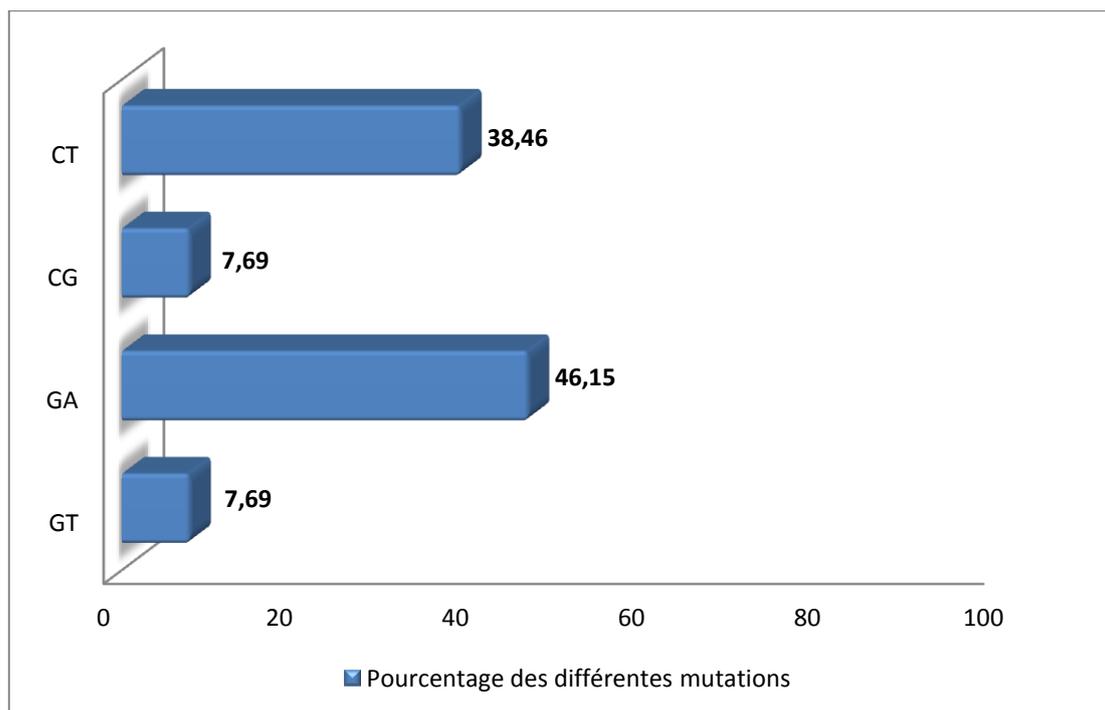
Les mutations du codon 157 ont mené à des mutations associées, ces dernières ont causé des changements d'acides aminés.

**Tableau XI** : Mutations associées au codon 157 de la p53.

N° base	Changement	Séquence avant	Séquence après	A.A. avant	A.A. après	Cancers associés
168	A – G	GAA	GAG	Glu	Glu	---
215	G – C	CGC	CCC	Arg	Pro	---
234	A – G	GCA	GCG	Ala	Ala	---
235	G – A	GCT	ACT	Ala	Ala	---
501	G – C	CAG	CAC	Gln	His	---
579	T – C	CAT	CAC	His	His	---
596	G – T	GGA	GTA	Gly	Val	---
742	C – T	CGG	TGG	Arg	Trp	Tête et cou / Rein / Système hématopoïétique / Vessie / Bronche et langue
743	G – A	CGG	CAG	Arg	Gln	Colon/ tête et cou/ bronches et langue/ œsophage/ vessie/ système hématopoïétique/ foie
800	G - T	CGG	CTG	Arg	Leu	---
818	G – A	CGT	CAT	Arg	His	Œsophage / bronche et langue
832	C – G	CCT	GCT	Pro	Ala	Œsophage
859	G – T	GAG	TAG	Glu	STOP	Bronche et langue
925	C – T	CCC	TCC	Pro	Ser	---

Les mutations au niveau des bases 742, 743, 818, 832 et 859 (Tableau X) qui ont engendré les changements nucléotidiques suivants : C  $\rightarrow$  T / G  $\rightarrow$  A / C  $\rightarrow$  G et G  $\rightarrow$  T ont été identifiées dans divers cancers tels que celui de l'œsophage, colon, rein ... etc.

Les mutations au niveau des autres bases qui ont engendré le changement des nucléotides suivant A  $\rightarrow$  G et T  $\rightarrow$  C n'ont pas été identifiées dans des cancers, fort probablement du non changement de l'acide aminé qui est resté le même après la mutation.



**Figure 30** : Pourcentage des différentes mutations associées des nucléotides du codon 157de la P53.

La mutation la plus fréquente associée aux nucléotides du codon 157 ; est le changement du G  $\rightarrow$  A avec un pourcentage de 46.15%, suivit de la mutation C  $\rightarrow$  T avec un pourcentage 38.46% et enfin les deux mutations C  $\rightarrow$  G et G  $\rightarrow$  T avec 7.69%.

**Tableau XII** : Pourcentage des cancers associés à la mutation du codon 157de la P53.

Cancers associés	Pourcentage
Bronche et langue	15.79%
Œsophage	12.63%
Tête et cou	7.37%
Rein	3.16%
Système hématopoïétique	7.37%
Vessie	7.37%
Colon	4.1%
Foie	4.21%
Autres	38%

Divers cancers sont associés à la mutation du codon 157, en tête est celui des bronches et de la langue avec 15.79% et en dernier celui des reins avec 3.16%. (Tableau XI).

### V.2.Analyse du codon 158

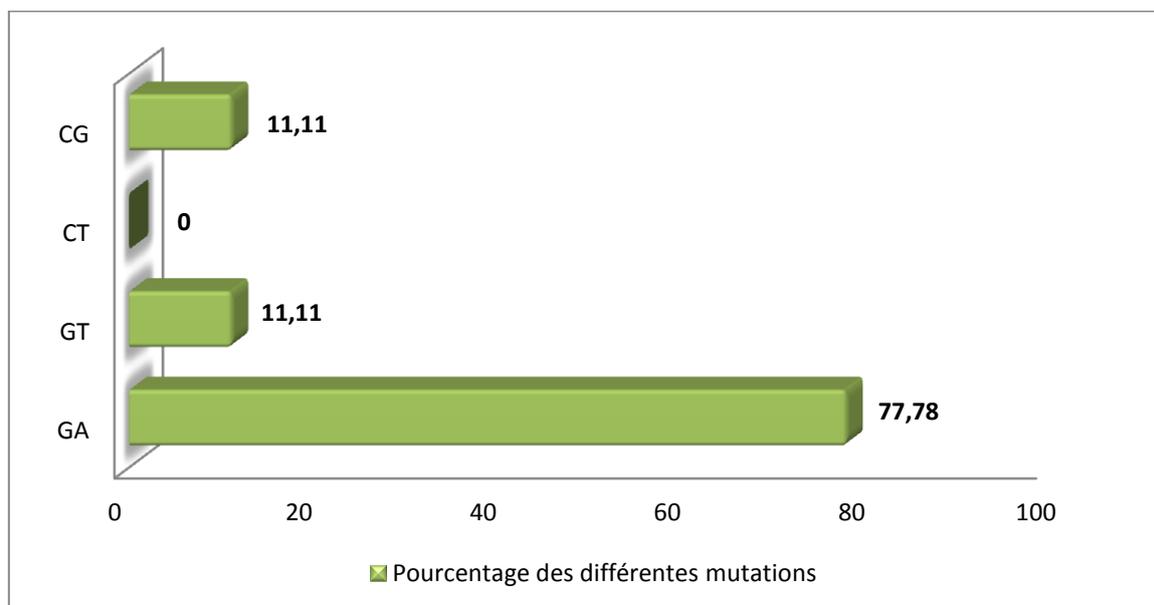
**Tableau XIII** : Mutations associées au codon 158 de la p53.

N° base	Changement	Séquence avant	Séquence après	A.A. avant	A.A. après	Cancers associés
168	A – G	GAA	GAG	Glu	Glu	---
215	G – C	CGC	CCC	Arg	Pro	---
234	A – G	GCA	GCG	Ala	Ala	---
235	G – A	GCT	ACT	Ala	Ala	---
246	G - C	CCG	CCC	Pro	Pro	---
248	C - T	GCG	GTG	Ala	Val	---
385	G - A	GCC	ACC	Ala	Thr	---
501	G – C	CAG	CAC	Gln	His	---
579	T – C	CAT	CAC	His	His	---
596	G – T	GGA	GTA	Gly	Val	---
743	G – A	CGG	CAG	Arg	Gln	Colon/ tête et cou/ bronches et langue/ œsophage/ vessie/ système hématopoïétique/ foie
800	G – T	CGG	CTG	Arg	Leu	---
818	G – A	CGT	CAT	Arg	His	Œsophage / bronche et langue
832	C – G	CCT	GCT	Pro	Ala	Œsophage
859	G – T	GAG	TAG	Glu	STOP	Bronche et langue
925	C – T	CCC	TCC	Pro	Ser	---

Les mutations du codon 158 ont mené à des mutations associées, ces dernières ont causé des changements d'acides aminés.

Les mutations au niveau des bases 743, 818, 832 et 859 (Tableau XII), ont engendré les changements nucléotidiques suivants : G → A / C → G et G → T et ont été identifiées dans divers cancers tels que celui de l'œsophage, colon, foie ...etc.

Les mutations au niveau des autres bases qui ont engendré le changement des nucléotides  $A \longrightarrow G$  et  $T \longrightarrow C$ , n'ont pas été identifiées dans des cancers, fort probablement du non changement de l'acide aminé qui est resté le même après la mutation.



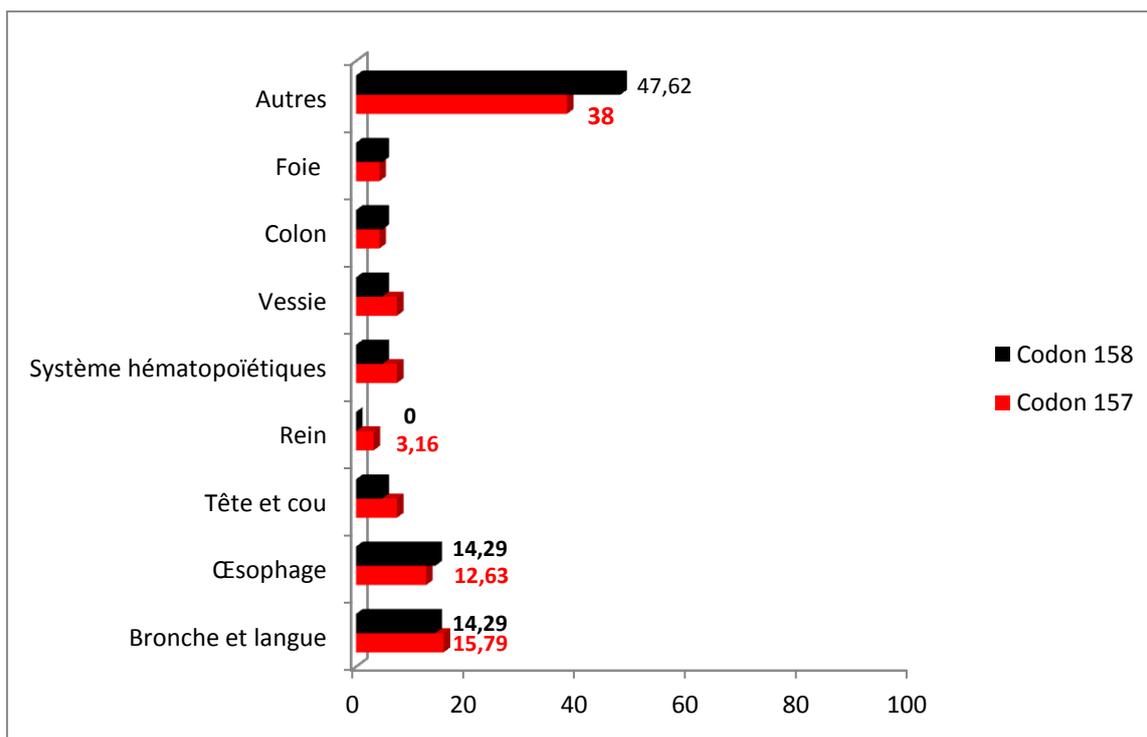
**Figure 31** : Pourcentage des différentes mutations associées des nucléotides du codon 158 de la P53.

La mutation la plus fréquente est associée aux nucléotides du codon 158 avec le changement de  $G \longrightarrow A$  avec un pourcentage de 77.78%, suivit par les mutations  $C \longrightarrow G$  et  $G \longrightarrow T$  avec un pourcentage 11.11%, quand à la mutation  $C \longrightarrow T$ , sans rôle identifié dans l'expression d'éventuels cancers.

**Tableau XIV** : Pourcentage des cancers associés à la mutation du codon 158 de la P53.

Cancers associés	Pourcentage
Bronche et langue	14,29%
Œsophage	14,29%
Tête et cou	4,76%
Rein	0%
Système hématopoïétiques	4,76%
Vessie	4,76%
Colon	4,76%
Foie	4,76%
Autres	47,62%

Divers cancers sont associés à la mutation du codon 158, en tête est celui des bronches, langue et œsophage avec 14.29% suivit de celui de la tête et le cou, en dernier on retrouve, le système hématopoïétique, la vessie, le colon et le foie avec chacun 4.76%. (Tableau XIII).



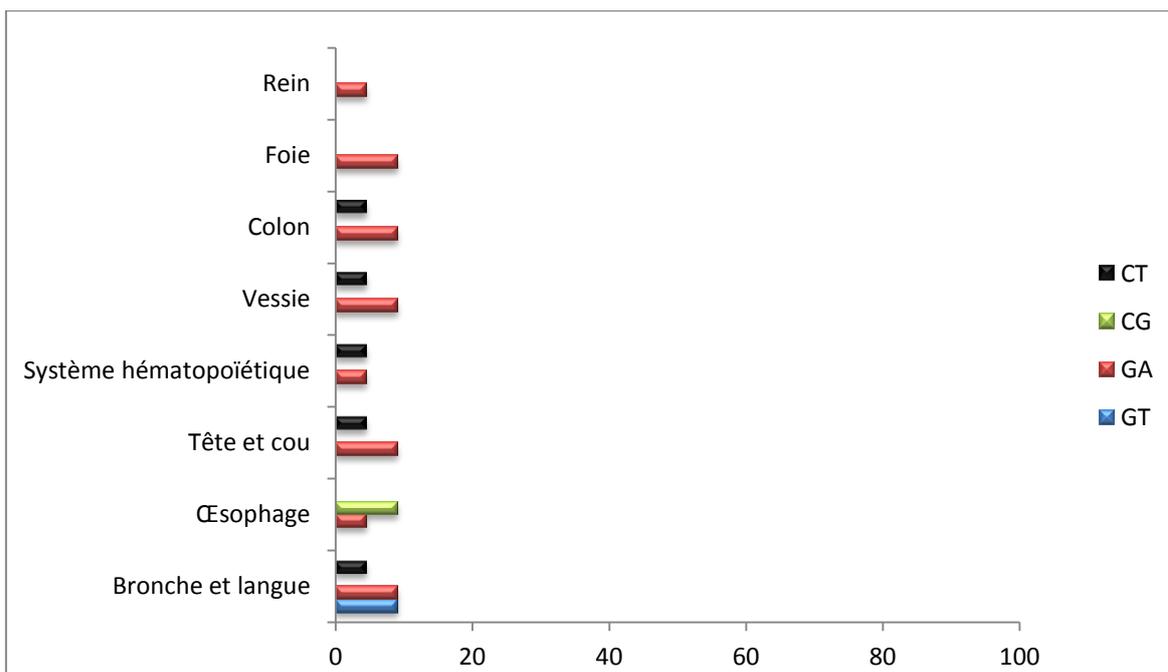
**Figure 32** : Pourcentage des différents cancers associés aux mutations des codons 157,158 de la P53.

En fonction des cancers associés aux mutations des deux codons 157,158, on remarque que les cancers de la bronche/langue et œsophage ont le plus haut pourcentage à savoir 15,79 % et 14,29%.

**Tableau XV** : Pourcentage des différentes mutations nucléotidiques des codons 157,158 de la P53 liées aux cancers associés.

	<b>GT</b>	<b>GA</b>	<b>CG</b>	<b>CT</b>
Bronche/langue <b>15,04%</b>	9,09	9,09	---	4,55
Œsophage <b>13,46%</b>	---	4,55	9,09	---
Tête et cou <b>6,07%</b>	---	9,09	---	4,55
Système hématopoïétiques <b>6,07%</b>	---	4,55	---	4,55
Vessie <b>6,07%</b>	---	9,09	---	4,55
Colon <b>4,43%</b>	---	9,09	---	4,55
Foie <b>4,49%</b>	---	9,09	---	---
Rein <b>1,58%</b>	---	4,55	---	---

D'après les résultats du tableau XIV, il ressort que le cancer des bronches et langue est le plus présent avec (15.04%) et aussi avec la fréquence la plus élevée des mutations à savoir 9.09 pour le nucléotide G qui connaît deux mutations ( $G \rightarrow T$  ;  $G \rightarrow A$ ), alors que le reste des mutations ont la même fréquence qui est de 4.55 pour l'ensemble des autres nucléotides



**Figure 33** : Pourcentage des différentes mutations des codons 157;158 par type de cancer associé.

De la figure 33, différentes mutations ont été retrouvées dans les divers cancers associés à savoir :

- ✓ La mutation G  $\longrightarrow$  A a été identifiée dans tous les cancers associés trouvés avec différents pourcentages, alors que la mutation C  $\longrightarrow$  T a été identifiée que dans 5 cancers (bronche et langue, tête et cou, système hématopoïétique, vessie et colon) avec le même pourcentage.
- ✓ Concernant la mutation G  $\longrightarrow$  T, elle a été identifiée uniquement dans le cancer des bronches et de la langue, contrairement à la C  $\longrightarrow$  G qui n'a été trouvée que dans le cancer de l'œsophage.

## **B-Discussion**

Le génome humain contient un grand nombre d'ARNnc qui s'expriment dynamiquement dans les tissus et les cellules spécifiques et qui ont une influence sur l'expression et la distribution intracellulaire de protéines spécifiques. Ils sont également des facteurs de recrutement des complexes de modification de la chromatine sur des sites spécifiques. Ainsi, une expression aberrante de plusieurs ARNnc est associée au cancer. **(Chang., 2011)**.

Même si multiples études ont démontré l'importance des lncARN dans la pathogénèse du cancer et leur abondance pendant cette dernière, ce qui laisse prédire un rôle potentiel dans son développement. **(Nagano and Fraser, 2011)**, Mais il y a peu d'études qui révèlent l'importance de gènes Hox et HOTAIR dans les tissus pulmonaires à cause de la pénurie d'échantillons tissulaires frais. **(Grier et al.,2005)**.

L'étude d'Hiroshi et al. est le premier rapport sur l'expression de HOTAIR dans des tissus pulmonaires a montré que l'expression HOTAIR est significativement plus haute dans des tumeurs pulmonaires que les tissus normaux.

HOTAIR est impliqué dans la répression transcriptionnelle des gènes HOXD, par un système de Recrutement des complexes répresseurs PRC2 et LSD1, Plusieurs publications ont rapporté une surexpression de HOTAIR dans différents cancers, où il participerait parfois à la formation de métastases **(Mathieu et al., 2014)**.

La P53 est un des gènes suppresseurs de tumeur les plus puissants identifiés dans des cellules humaines. Sa mutation est le plus fréquemment trouvée dans des cellules cancéreuses et particulièrement le cancer du poumon avec 70 % dont ont trouve une mutation P53.

Pour les LncARN surexprimés, nous les trouvons impliqués dans la voie de P53 :

Selon Zhai et al. la p53 est impliqué dans la prolifération cellulaire du cancer du poumon et la migration. Une corrélation négative a été identifiée entre la fonctionnalité p53 et HOTAIR. La nouvelle analyse a découvert que p53 réprime l'expression HOTAIR par l'attache directe à son promoteur. Inversement, lincARN HOTAIR inhibe l'expression de p53 dans des cellules du cancer du poumon.

Des études ont révélé que HOTAIR a modifié le promoteur de p53 et a amélioré histone H3 lysine 27 trimethylation (H3K27me3). Donc, pour la première fois elles révèlent une boucle réglementaire négative entre p53 et HOTAIR dans le cancer du poumon.

La fonction de HOTAIR dans NSCLC n'est toujours pas claire et a besoin de la nouvelle enquête

MALAT-1 est surexprimé dans plusieurs types de cancers, y compris celui du poumon, du sein, du côlon et l'hépatocarcinome. Tous ces cancers sont associés à la mutation de la p53. La surexpression de MALAT-1 dans diverses lignées cellulaires a augmenté la prolifération cellulaire (**Gibb., 2011**). C'est ce qui a été répertorié sur COSMIC, base de données de mutations, où il a été mis en évidence une perte de 7 bases et gains de 28 bases avec une surexpression des bases 431.

MALAT-1, joue un rôle dans la migration cellulaire et la métastase tumorale, il est aussi impliqué dans la régulation de l'épissage alternatif des ARN en interagissant avec des facteurs d'épissage comme ceux de la famille SR contenant des peptides sérine / arginine.

MALAT-1 favorise la prolifération cellulaire et inhibe l'apoptose en inhibant l'activation de la p53 en diminuant son niveau d'acétylation. A l'inverse l'expression forcée du gène p53 supprime le niveau de transcription de MALAT-1, inhibe la prolifération cellulaire et induit une apoptose. (**Chen et al., 2017**).

D'après nos résultats, nous avons retrouvé que les cancers associés aux mutations des codons 157,158 responsable du cancer des poumons étaient de 15.04% (Bronche et

langue) suivit par ceux de l'œsophage 13.48% et la plus faible valeur enregistré pour le cancer du rein avec 1.58%.

Nous remarquons que la mutation G  $\rightarrow$  A, la plus importante avec une implication dans tous les cancers associés. Le reste des mutations des différents nucléotides non répertoriés dans la base de données spécialisées et qui n'ont pas été associé à l'apparition d'autres types de cancers, ainsi il se pourrait que ces dernières puissent être responsables d'autres pathologies.

---

## Conclusion

Durant les dix dernières années, la découverte de la classe des LncARN, a révolutionné le monde des ARN, même si la compréhension biologique de ces molécules avance lentement. Nous sommes à présent sûrs de leur importance dans de nombreux processus cellulaires, parmi eux le processus tumoral **(Rinn and Chang, 2012)**.

En effet, de multiples études ont démontré l'importance des long ARNnc dans la pathogénèse du cancer, par la prédiction de leurs rôles potentiel dans son développement. **(Nagano et Fraser, 2011)**,

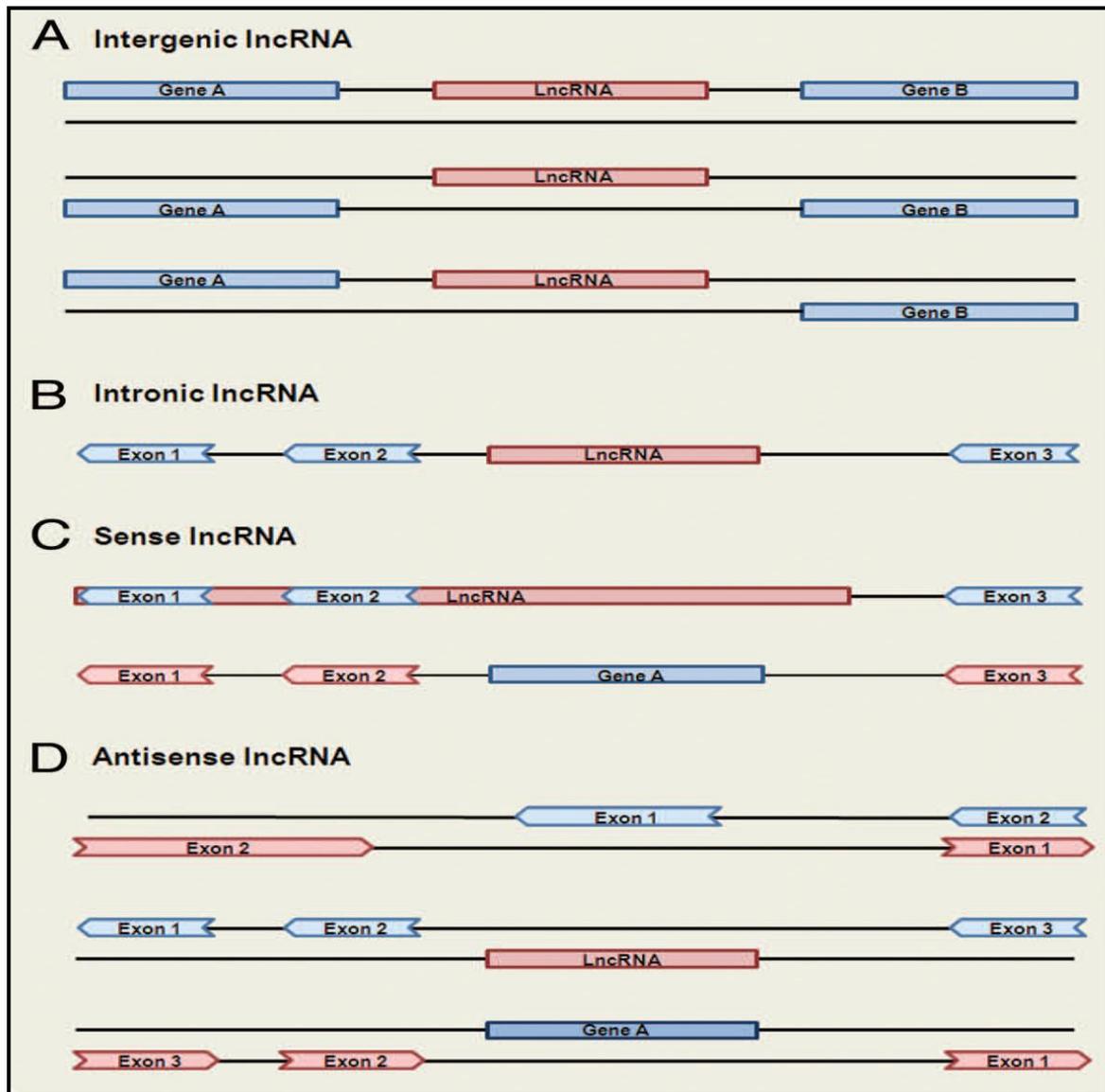
Le premier rapport sur l'expression de HOTAIR dans les tissus pulmonaires, a montré que l'expression HOTAIR est significativement plus haute dans des tumeurs pulmonaires que les tissus normaux. **(Hiroshi et al., 2014)**.

MALAT-1, joue un rôle dans la migration cellulaire et la métastase tumorale, il est aussi impliqué dans la régulation de l'épissage alternatif des ARN en interagissant avec des facteurs d'épissage

MALAT-1 favorise la prolifération cellulaire et inhibe l'apoptose en inhibant l'activation de la p53 en diminuant son niveau d'acétylation. A l'inverse l'expression forcée du gène p53 supprime le niveau de transcription de MALAT-1, inhibe la prolifération cellulaire et induit une apoptose. **(Chen et al., 2017)**.

D'après nos résultats, nous avons retrouvé que les cancers associés aux mutations des codons 157,158 responsable du cancer des poumons étaient de 15.04% (Bronche et langue) suivit par ceux de l'œsophage 13.48% et la plus faible valeur enregistré pour le cancer du rein avec 1.58%.

Ainsi les fonctions détaillées de ces longs ARNnc, HOTAIR/MALAT-1 dans le cancer du poumon n'est toujours pas claire et a besoin de beaucoup plus d'enquêtes aussi bien biologiques que bioinformatiques.



**Figure 06** : Classification des longs ARNnc. (Ma L et al., 2013).

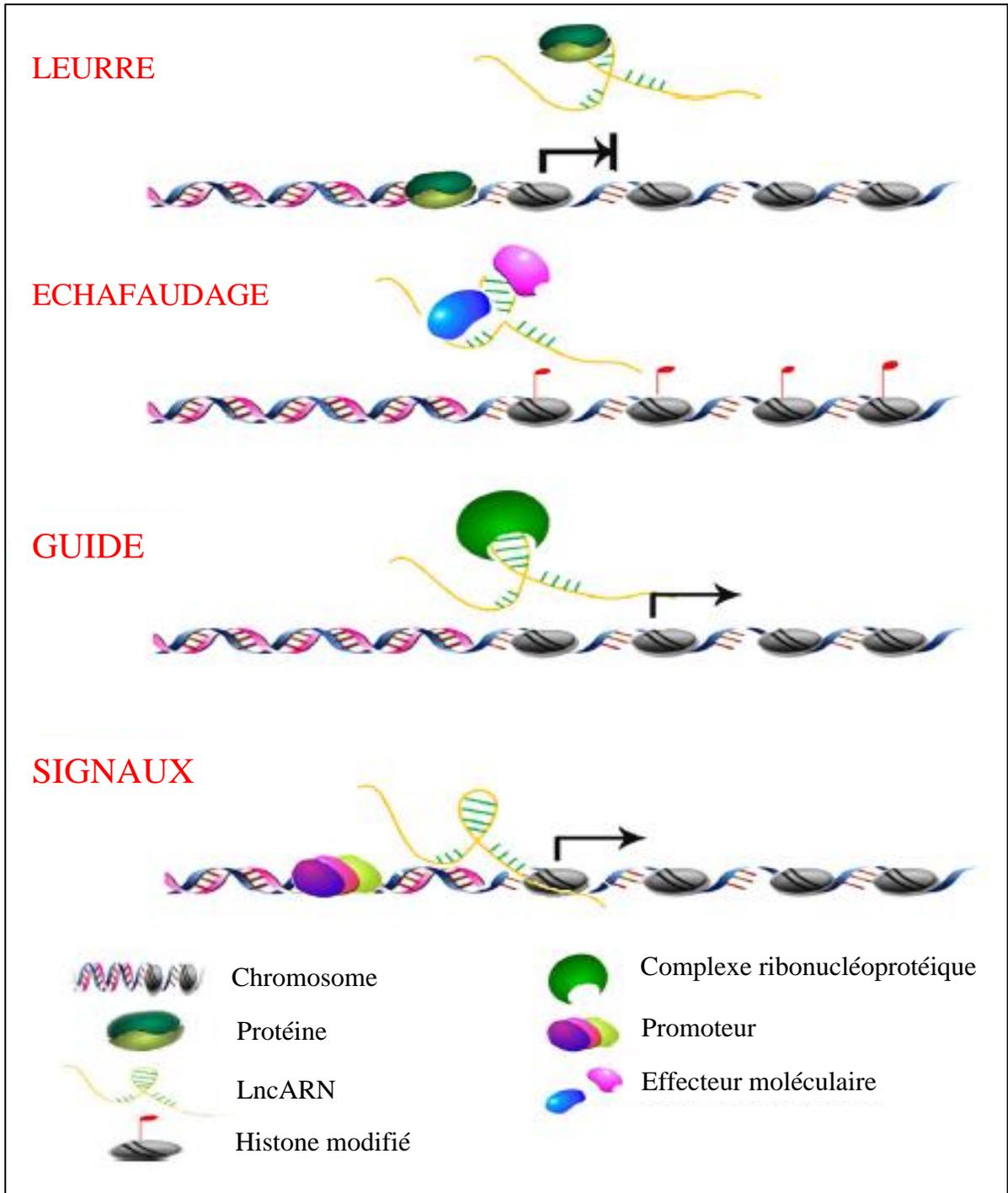
Les gènes codants pour des protéines et leurs exons sont représentés en **bleu**, alors que les long ARNnc et leurs exons sont en **rouge**.

(A) Long ARNnc intergénique.

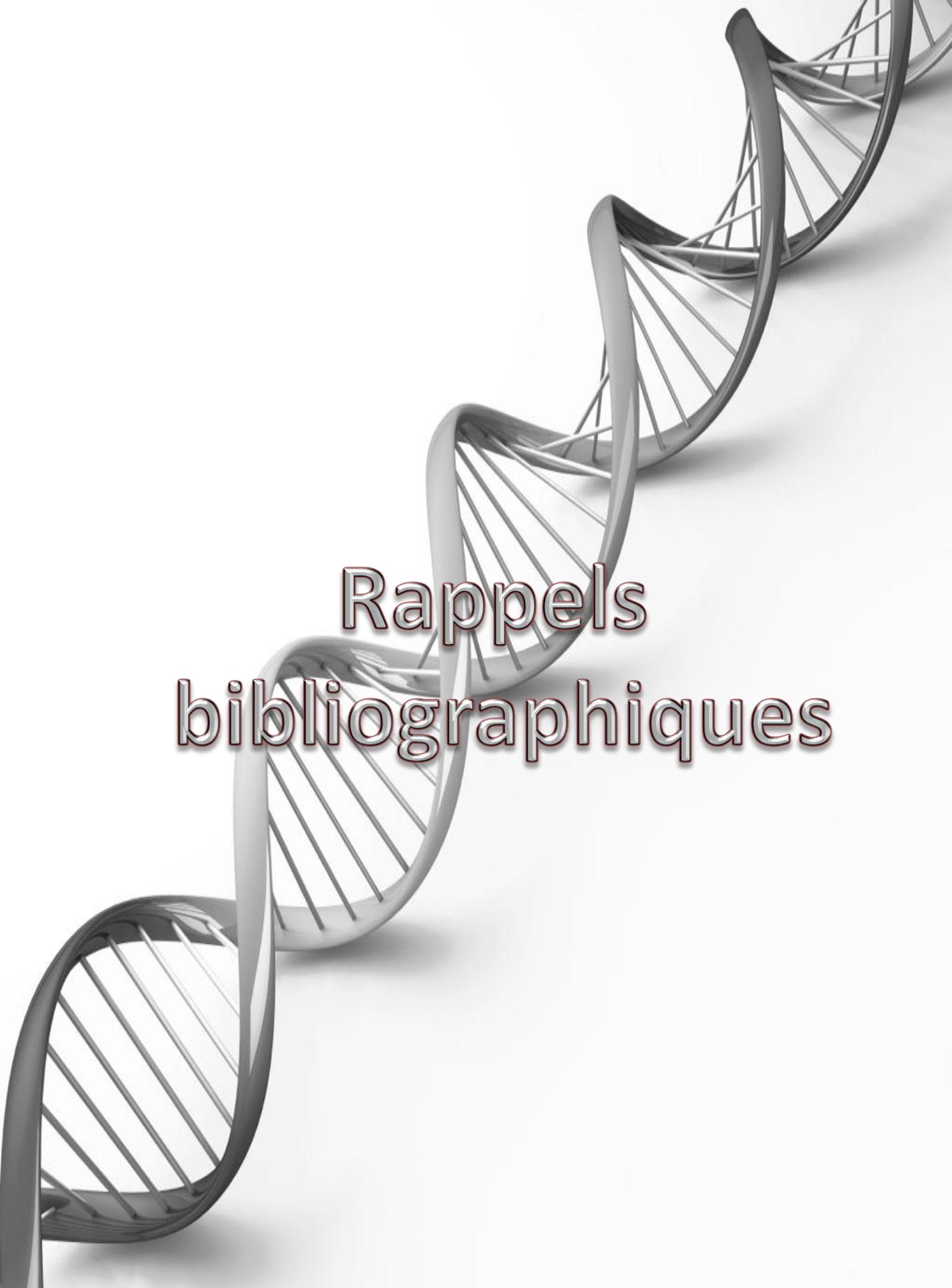
(B) Long ARNnc intronique.

(C) Long ARNnc sens.

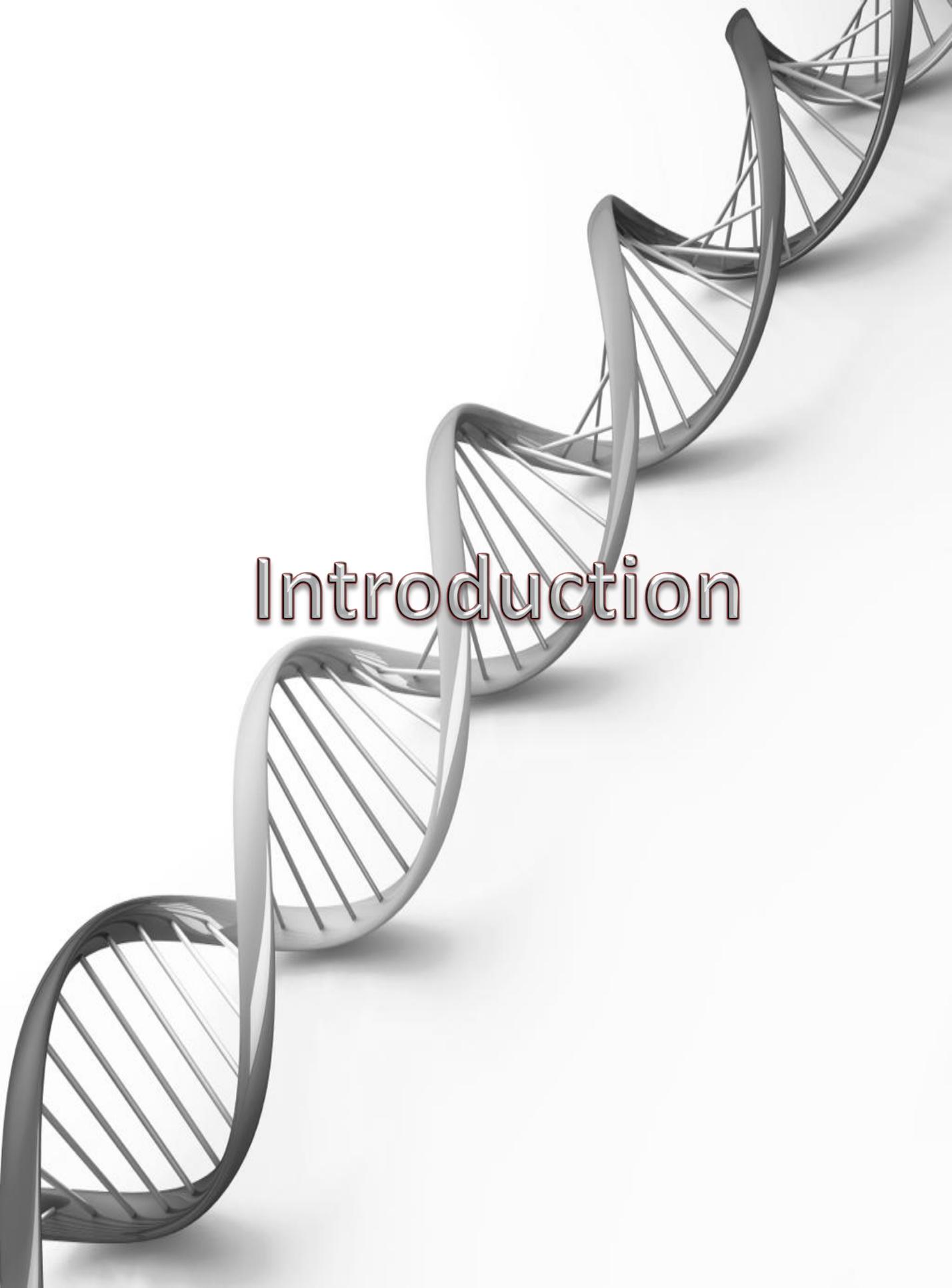
(D) Long ARNnc anti-sens, seule la classe divergent n'est pas représentée dans cette figure.



**Figure 07 : Fonction des lncARN (Chen et al., 2014).**



Rappels  
bibliographiques



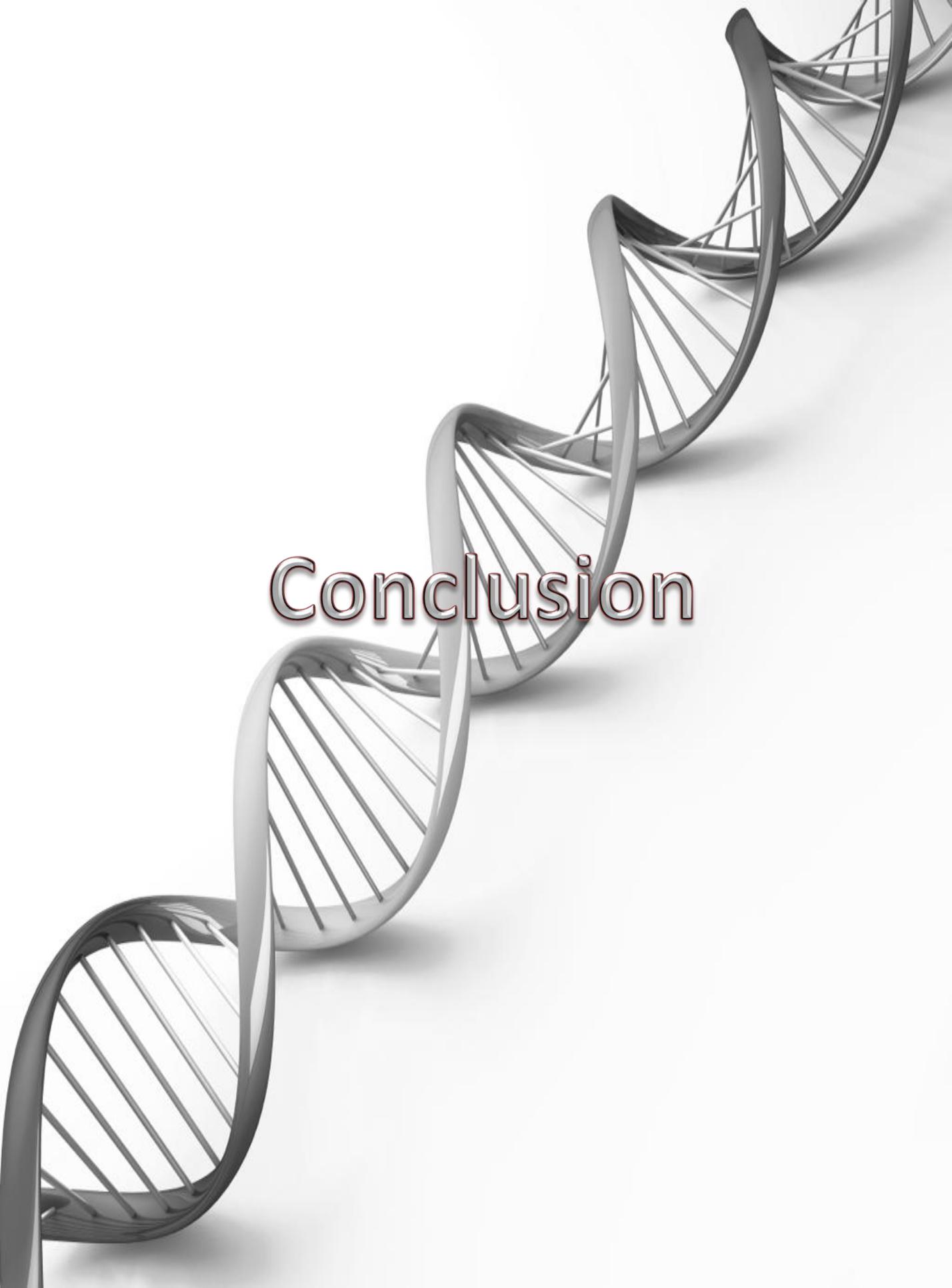
# Introduction



# Matériel et méthodes



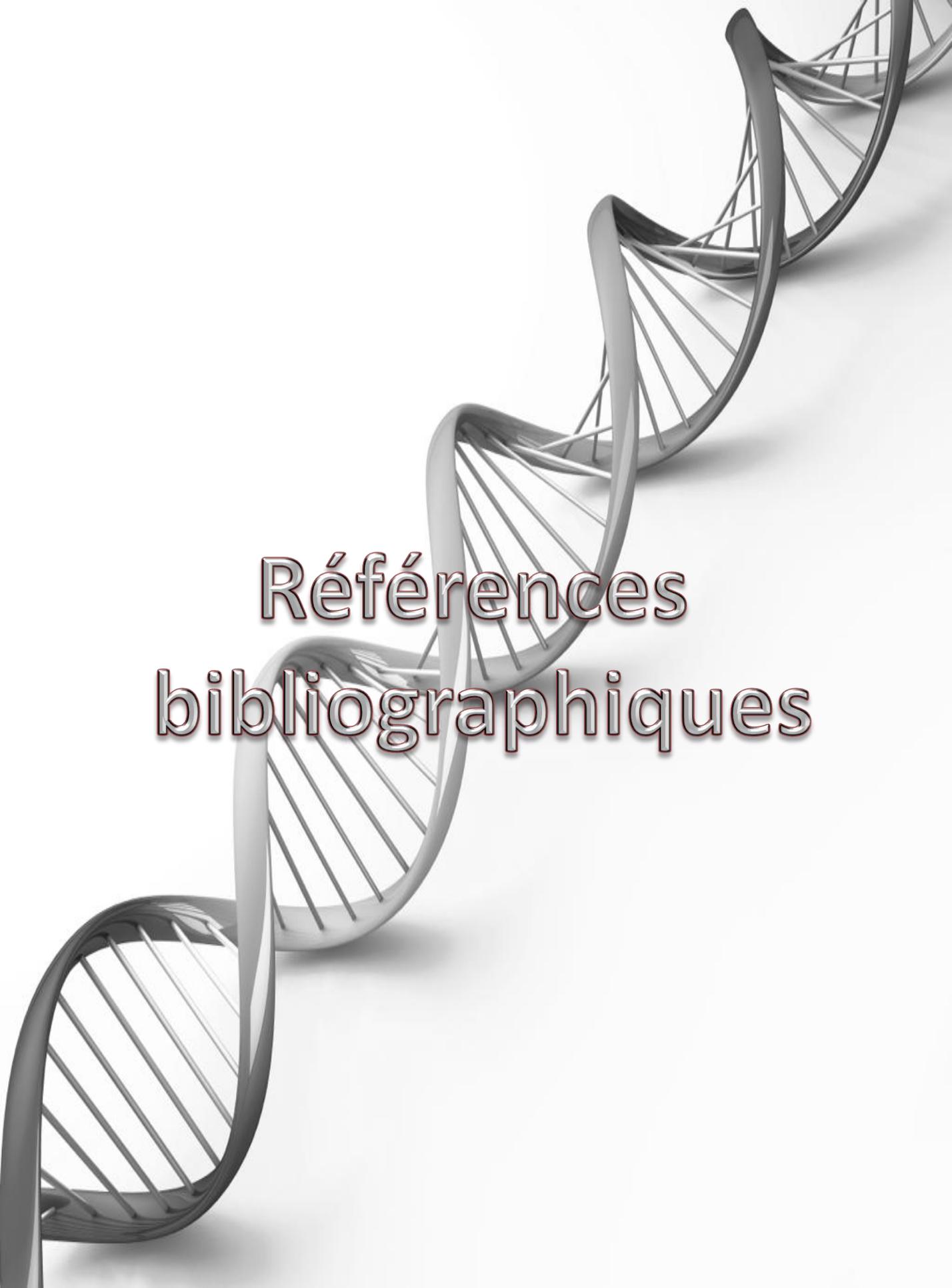
# Résultats et Discussion



Conclusion



# Annexes



Références  
bibliographiques

- ✓ **Brambilla E., Travis W.D., Colby T.V., Corrin B., Shimosato Y.,** 2001: The new World Health Organization classification of lung tumours. *Eur Respir J.* P1059, 1068.
  
- ✓ **Brosh R., Rotter V.,** 2009 : When mutants gain new powers: news from the mutant p53 field. *Nat Rev Cancer.* P 701,713.
  
- ✓ **Cao J.,** 2014: The functional role of long non-coding RNAs and epigenetics. *Biological Procedures Online.*
  
- ✓ **Chater A.,** 2013 : Cancer broncho-pulmonaire et thérapeutique. Thèse de doctorat en médecine. Université Abu bekr belkaid. P118.
  
- ✓ **Chen J., Wang R., Zhang K., Chen L.B.,** 2014: Long non-coding RNAs in non-small cell lung cancer as biomarkers and therapeutic targets. *J Cell Mol Med.* P 2425,2436.
  
- ✓ **Chen R., Yun L., Hao Z., Baicai Y., Kaiwen H., Mingming X., Chunyu H., Huajun G., Xinran Z., Chenxi J., Lingjun L., Yongmei L, Ning Z.,** 2017 : Quantitative proteomics reveals that long non-coding RNA MALAT1 interacts with DBC1 to regulate p53 acetylation. *Nucleic Acids Research.*
  
- ✓ **Chopin D., Cappellen D., Fradvanyi F., Gattegno B.,** 2005: Le gene P53. *Progrès en Urologie.* P 1338-1343.
  
- ✓ **Delgrange O.,** 2009 : Bioinformatique : Un domaine pluridisciplinaire .In *Element : le Magazine de l'Université de Mons.* P 27, 30.
  
- ✓ **Denoeux S.,** 2015 : Etude de la régulation de l'expression des gènes par un ARN antisens. *Biologie cellulaire.* Université Paris-Saclay.
  
- ✓ **Donzelli S., Biagioni F., Fausti F., Strano S., Fontemaggi G., Blandino G.,** 2008 : Oncogenic Approaches in Exploring Gain of Function of Mutant p53. *Curr Genomics.* P200,207.
  
- ✓ **Fang Y., Fullwood M.J.,** 2016: Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics Proteomics Bioinformatics, Volume 14,*P42-54.

- ✓ **Fjose A., Ellingsen S., Wargelius A., Seo H.C.,** 2001: RNA interference: mecanismes and applications *Biotechnology annual. Revievx* ,Volume 7,P31,57.
  
- ✓ **Fontaine A.,** 2009 : Classification d'ARN codants et d'ARN non-codants. *Bio-informatique. Université des Sciences et Technologie de Lille.*
  
- ✓ **Gautheret D.,** 2012 : *Biologie Moléculaire : Structure de la Chromatine et Structure des génomes. Cours 2ème année .Université Paris-Sud. P33.*
  
- ✓ **Geng Y.J., Xie S.L., Li Q., Ma J., Wang G.Y.,** 2011 : Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. *J Int Med Res*, Volume 39, P 2119-2128.
  
- ✓ **Gibb E.A., Brown C.J., Lam W.L.,** 2011: The functional role of long non-coding RNA in human carcinomas. *Mol Cancer.*
  
- ✓ **Gondé H.,** 2016 : Impact économique hospitalier de la robotique dans la chirurgie mini-invasive d'exérèse pulmonaire. *Th.D.Pharm.Rouen . P99.*
  
- ✓ **Gupta R.A., Shah N., Wang K.C., Kim J., Horlings H.M., Wong D.J.,** 2010: Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, Volume 464, P 1071-1076.
  
- ✓ **Gutschner T., Hämmerle M., Eissmann M., Hsu J., Kim Y., Hung G., Revenko A., Arun G., Stentrup M., Gross M., Zörnig M., MacLeod A.R., Spector D.L., Diederichs S.,** 2013: The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res. .P 1180, 1189.*
  
- ✓ **Guttman M., Amit I., Garber M., French C., Lin M.F., Feldser D., Huarte M., Zuk O., Carey B.W., Cassady J.P., Cabili M.N., Jaenisch R., Mikkelsen T.S., Jacks T., Hacohen N., Bernstein B.E., Kellis M., Regev A., Rinn J.L., Lander E.S.,** 2009 : Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*.P223,227.
  
- ✓ **Hajjari M., Salavaty A.,** 2015: HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer Biology & Medicine. P1, 9.*

- ✓ **Hiroshi O., Noriko M., Hiroko N., Eisaku M., Masaru., Masaaki M, Sakae O., Makoto N., Tetsuro H., Naohiko I., Yuichi I.,** 2014 : Long noncoding RNA HOTAIR is relevant to cellular proliferation, invasiveness, and clinical relapse in small-cell lung cancer. *Cancer Medicine*. P632,642.
- ✓ **Hutchinson J.N., Ensminger A.W., Clemson C.M., Lynch C.R., Lawrence J.B., Chess A.,** 2007 : A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* .P 39.
- ✓ **Jemaa M.,** 2012 : Chimiothérapie ciblant les cellules cancéreuses p53 déficientes. *Cancer*. Université Paris Sud - Paris XI.
- ✓ **Ji P., Diederichs S., Wang W., Böing S., Metzger R., Schneider P.M., Tidow N., Brandt B., Buerger H., Bulk E., Thomas M., Berdel W.E., Serve H., Müller-Tidow C.,** 2003: MALAT-1, a novel noncoding RNA, and thymosin 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogène*. P8031, 8041.
- ✓ **Khalil A.M., Guttman M., Huarte M.,** 2009: Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci*. P11667, 11672.
- ✓ **Loewen G., Jayawickramarajah J., Zhuo Y., Shan B.,** 2014: Functions of lncRNA HOTAIR in lung cancer. *Journal of Hematology and Oncology*. P90.
- ✓ **Ma L., Vladimir B. Bajic., Zhang Z.,** 2013: On the classification of long non-coding RNAs, 924–933.
- ✓ **Mathieu E.L., Belhocine M., Dao L.T., Puthier D., Spicuglia S.,** 2014 : Rôle des longs ARN non codants dans le développement normal et pathologique. *Med Sci*, Volume 30, N 8-9, P790–796.
- ✓ **Matthew J. Hangauer., Ian W. Vaughn , Michael T. McManus,** 2013: Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *PLoS Genet*, Volume 9, Issue 6.

- ✓ **Nagano T., Fraser P.,** 2011: No-Nonsense Functions for Long Noncoding RNAs. *Cell*, Volume 145, Issue 2, P178–181.
- ✓ **Nakagawa T., Endo H., Yokoyama M., Abe J., Tamai K., Tanaka N.,** 2013: Large noncoding RNA HOTAIR enhances aggressive biological behavior and is associated with short disease-free survival in human non-small cell lung cancer. *Biochem Biophys Res Commun.* P 319,324.
- ✓ **Prensner J.R., Chinnaiyan A.M.,** 2011: The emergence of lncRNAs in cancer biology. *Cancer Discov.*P 391, 407.
- ✓ **Quinkal I., Rechenmann F.,** 2004 : Entre biologie, informatique et mathématiques : la bioinformatique. *Interstices.info.*
- ✓ **Read A., Donnai D., Sznajer Y., Verloes A.,** 2008: Génétique médicale : De la Biologie à la pratique clinique . Edition De Boeck superieur.
- ✓ **Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo S. L., Xu, X., Bruggmann, S. A.,** 2007: Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* P1311, 1323.
- ✓ **Rocks N., Paulissen G., EL hour M., Quesada F., Crahay C., Gueders M., Foidart J.M., Noel A., Cataldo D.,** 2008 : Emerging roles of ADAM and ADAMTS metalloproteinases in cancer. *Biochimie.* P369, 379.
- ✓ **Sedkaoui Ch.,** 2015 : Chimiothérapie et thérapie ciblée dans le cancer colorectal métastatique. Thèse de doctorat en science médicale.Tizi-Ouzou. Université mouloud mammeri. P201.
- ✓ **Shi X., Sun M., Liu H., Yao Y., Song Y.,** 2013: Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer Lett .* P 159,166.
- ✓ **Soukaina N.,** 2015 : Techniques de cytogénétique conventionnelles et modernes. Thèse de doctorat en médecine. Université Mohammed V-Rabat. P153.
- ✓ **Soussi T., Dehouche K., Bérourd C.,** 2000 : p53 Website and analysis of p53 gene mutations in human cancer : Forging a link between epidemiology and carcinogenesis. *Hum Mutat.* P 105,113.

- ✓ **Tsai M.C., Manor O., Wan Y., Mosammaparast N., Wang J.K., Lan F.,** 2010: Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. P689,693.
- ✓ **Ulitsky I., Bartel D.P.,** 2013 : LincRNAs: Genomics, Evolution, and Mechanisms. *CELL*.P26,46.
- ✓ **Wapinski O.L., Chang H.Y.,** 2011: Long noncoding RNAs and human disease. *Trends Cell Biol*. P 561.
- ✓ **Woo C.J., Kingston R.E.,** 2007: HOTAIR lifts noncoding RNAs to new levels. *Cell*,Volume 129, P 1257-1259.
- ✓ **Wu T., Yin X., Zhou Y., Wang .Z, Shen S., Qiu Y.,** 2015 : Roles of noncoding RNAs in metastasis of nonsmall cell lung cancer: A mini review. *J Can Res* .P 7, 10.
- ✓ **Zhai N., Xia Y., Yin R., Liu R., Gao F.,** 2016: A negative regulation loop of long noncoding RNA HOTAIR and p53 in non-small-cell lung cancer. *OncoTargets and Therapy*. P5713,5720.
- ✓ **Zhang H., Chen Z., Wang X., Huang Z., He Z., Chen Y.,** 2013: Long non-coding RNA: A new player in cancer. *J Hematol Oncol* .P 37.

Sites web consultés :

- ✓ **GLOBOCAN:** <http://globocan.iarc.fr/Default.aspx>
- ✓ **NCBI:** <https://www.ncbi.nlm.nih.gov/>
- ✓ **GeneLoc:** <https://genecards.weizmann.ac.il/geneloc/index.shtml>
- ✓ **COSMIC :** <http://cancer.sanger.ac.uk/cosmic>
- ✓ **RaptorX:** <http://raptorx.uchicago.edu/>
- ✓ **Rfam :** <http://rfam.xfam.org/>

---

## Annexe I

### Les séquences de HOTAIR/MALAT-1/P53

#### I. La séquence du transcrit de HOTAIR

##### **Homo sapiens Hox transcript antisense intergenic RNA (HOTAIR) mRNA, complete sequence**

GenBank: DQ926657.1

[GenBank](#) [Graphics](#)

```
>DQ926657.1 Homo sapiens Hox transcript antisense intergenic RNA (HOTAIR) mRNA, complete sequence
```

```
GACTCGCCTGTGCTCTGGAGCTTGATCCGAAAGCTTCCACAGTGAGGACTGCTCCGTGGGGTAAGAGAG  
CACCAGGCACTGAGGCTGGGAGTTCCACAGACCAACCCCTGCTCCTGGCGGCTCCACCCGGGGCTT  
AGACCCTCAGGTCCCTAATATCCCGGAGGTGCTCTCAATCAGAAAGGTCTGCTCCGCTTCGCAGTGGAA  
TGGAACGGATTTAGAAGCTGCAGTAGGGGAGTGGGGAGTGGAGAGAGGGAGCCAGAGTTACAGACGGC  
GGCGAGAGGAAGGAGGGGCGTCTTTATTTTTTAAAGCCCCAAAGAGTCTGATGTTACAAGACCAGAAA  
TGCCACGGCCGCTCCTGGCAGAGAAAAGGCTGAAATGGAGGACCGGCGCCTTCTTATAAGTATGCACA  
TTGGCGAGAGAATTAAGTGCTGCAACCTAAACAGCAATTACACCCAAGCTCGTTGGGGCTAAGCCAGT  
ACCGACCTGGTAGAAAAAGCAACCACGAAGCTAGAGAGAGAGCCAGAGGAGGGAAGAGAGGCCAGACGA  
AGGTGAAAGCGAACCACGCAGAGAAAATGCAGGCAAGGGAGCAAGGCGGCAGTTCCCGGAACAAACGTGGC  
AGAGGGCAAGACGGGCACTCACAGACAGAGGTTTATGTATTTTTATTTTTAAAATCTGATTTGGTGTT  
CATGAGGAAAAGGGAAAATCTAGGGAACGGGAGTACAGAGAGAATAATCCGGGTCTAGCTGCCACATG  
AACGCCAGAGAACGCTGGA AAAA C T G A G C G G G T G C C G G G C A G C A C C C G G C T C G G G T C A G C C A C T G C C  
CCACACCGGGCCCAAGCCCGCCCCCTCGCGGCCACCGGGCTTCTTGCTCTTCTATCATCTCCAT  
CTTTATGATGAGGCTTGTTAAACAAGACCAGAGAGCTGGCCAAGCACCTCTATCTAGCCGCGCCCGCTCA  
GCCGAGCAGCGGTGGTGGGGGACTGGGAGGCGCTAATTAATTGATTCTTTGGACTGTAAAATATGGC  
GGCGTCTACACGGAACCATGGACTCATAACAATATATCTGTTGGGCGTGAGTGCAGTGTCTCTCAAT  
AATTTTTCCATAGGCAAATGTGAGAGGTTCTGGATTTTTAGTTGCTAAGGAAAGATCAAATGGGACCA  
ATTTTAGGAGGCCAAACAGAGTCCGTTCAAGTGTGAGAAAATGCTTCCCAAAGGTTGGCAGTGTGTTT  
TGTTGGAAAAAGCTTGGGTTATAGGAAAGCCTTCCCTGCTACTTGTGTAGACCCAGCCCAATTTAAGA  
ATTACAAGGAAGCGAAGGGTTGTGTAGGCCGGAAGCCTCTCTGTCCCGGCTGGATGCAGGGGACTTGAG  
CTGCTCCGGAATTTGAGAGGAACATAGAAGCAAAGTCCAGCCTTTGCTTCGTGCTGATTCTAGACTTA  
AGATTCAAAAACAAATTTTTAAAAGTGA AAC C A G C C T A G C C T T T G G A A G C T C T T G A A G T T C A G C A C C C  
ACCCAGGAATCCACCTGCTGTTACACGCCTCTCCAAGACACAGTGGCACCGCTTTTCTAACTGGCAGCA  
CAGAGCAACTCTATAATATGCTTATATTAGGTCTAGAAGAATGCATCTTGAGACACATGGGTAACCTAAT  
TATATAATGCTTGTCCATACAGGAGTGATTATGCAGTGGGACCCTGCTGCAAACGGGACTTTGCACTCT  
AAATATAGGCCCCAGCTTGGGACAAAAGTTGCAGTAGAAAAATAGACATAGGAGAACACTTAAATAAGTG  
ATGCATGTAGACACAGAAGGGGTATTTAAAAGACAGAAATAATAGAAGTACAGAAGAACAGAAAAAAAAT  
CAGCAGATGGAGATTACCATCCCAATGCCTGAACCTCCTCCTGCTATTAAGATTGCTAGAGAATTGTGT  
CTTAAACAGTTCATGAACCCAGAAGAACGCAATTTCAATGTATTTAGTACACACACAGTATGTATATAAA  
CACA ACT C A C A G A A T A T A T T T C C A T A C A T T G G G T A G G T A T G C A C T T T G T G T A T A T A A T A A T G T A T T  
TCCATGCAGTTTTAAAATGTAGATATATTAATATCTGGATGCATTTTCAAAAAAAAAA
```

I.1.La séquence du variant 1 de HOTAIR.

## Homo sapiens HOX transcript antisense RNA (HOTAIR), transcript variant 1, long non-coding RNA

NCBI Reference Sequence: NR\_047517.1

[GenBank](#) [Graphics](#)

```
>NR_047517.1 Homo sapiens HOX transcript antisense RNA (HOTAIR), transcript variant 1,
long non-coding RNA
CCAGTTCTCAGGCGAGAGCCGCGGCTGACAGGGTCTGGGACAGAAGGAAAGCCCTCCAGCCTCCAGGCC
TGCCTTCTGCCTGCACATTCTGCCCTGATTTCCGGAACCTGGAAGCCTAGGCAGGCAGTGGGGAACCTG
ACTCGCTGTGCTCTGGAGCTTGATCCGAAAGCTTCCACAGTGAGGACTGCTCCGTGGGGGTAAGAGAGC
ACCAGGCACTGAGGCTGGGAGTTCCACAGACCAACACCCTGCTCCTGGCGGCTCCACCCGGGACTTA
GACCTCAGGTCCTAATATCCGGAGGTGCTCTCAATCAGAAAGTCTGCTCCGCTTCGAGTGGAAAT
GGAACGGATTTAGAAGCTGCAGTAGGGGAGTGGGAGTGGAGAGAGGGAGCCAGAGTTACAGACGGCG
GCGAGAGGAAGGAGGGGCGTCTTTATTTTTTAAGGCCCAAAGAGTCTGATGTTTACAAGACCAGAAAT
GCCACGGCCGCGTCTGGCAGAGAAAAGGCTGAAATGGAGGACCGGCGCCTTCTTATAAGCTCGTTGGG
GCCTAAGCCAGTACCGACCTGGTAGAAAAAGCAACCACGAAGCTAGAGAGAGAGCCAGAGGAGGGAAGAG
AGCGCCAGACGAAGGTGAAAGCGAACCCAGCAGAGAAATGCAGGCAAGGGAGCAAGGCGGCAGTTCCTGG
AACAAACGTGGCAGAGGGCAAGACGGGCACTCACAGACAGAGGTTTATGTATTTTTATTTTTTAAAATCT
GATTTGGTGTCCATGAGGAAAAGGGAAAATCTAGGGAACGGGAGTACAGAGAGAATAATCCGGGTCTA
GCTCGCCACATGAACGCCAGAGAACGCTGAAAAACCTGAGCGGGTGCCGGGGCAGCACCCGGCTCGGG
TCAGCCACTGCCCCACACCGGGCCACCAAGCCCCGCCCTCGCGGCCACCGGGGCTTCTTGTCTTCT
TATCATCTCCATCTTTATGATGAGGCTTGTAAACAAGACCAGAGAGCTGGCCAAGCACCTCTATCTCAGC
CGGCCCCGCTCAGCCGAGCAGCGGTGGTGGGGGACTGGGAGGCGCTAATTAATTGATTCCTTTGGACT
GTAAAATATGGCGGCGTCTACACGGAACCCATGGACTCATAAACAATATATCTGTTGGGCGTGAGTGCAC
TGTCTCTCAAATAATTTTTCCATAGGCAAATGTCAGAGGGTCTGGATTTTTAGTTGCTAAGGAAAAGATC
CAAATGGGACCAATTTTAGGAGGCCCAAACAGAGTCCGTTCAAGTGTAGAAAATGCTTCCCAAAGGGGT
TGGGAGTGTGTTTTGTTGGAAAAAGCTTGGGTTATAGGAAAGCCTTCCCTGCTACTTGTGTAGACCCA
GCCAATTTAAGAATTACAAGGAAGCGAAGGGTGTGTAGGCCGGAAGCCTCTGTCCCGGCTGGATG
CAGGGGACTTGAGCTGCTCCGGAATTTGAGAGGAACATAGAAGCAAAGGTCCAGCCTTTGCTTCGTGCTG
ATTCCTAGACTTAAGATTCAAAAACAAATTTTTAAAAGTGAACCAGCCCTAGCCTTTGGAAGCTCTTGA
AGGTTCAGCACCCACCCAGGAATCCACCTGCCTGTTACACGCCTCTCCAAGACACAGTGGCACCGCTTTT
CTAACTGGCAGCACAGAGCAACTCTATAATATGCTTATATTAGGTCTAGAAGAATGCATCTTGAGACACA
TGGGTAACCTAATTATATAATGCTTGTTCATACAGGAGTGATTATGCAGTGGGACCCTGCTGCAAACGG
GACTTTGCACTCTAAATATAGACCCAGCTTGGGACAAAAGTTGCAGTAGAAAAATAGACATAGGAGAAC
ACTTAAATAAGTGATGCATGTAGACACAGAAGGGGATTTAAAAGACAGAAATAATAGAAGTACAGAAGA
ACAGAAAAAAAATCAGCAGATGGAGATTACCATCCCAATGCCTGAACTTCTCCTGCTATTAAGATTGC
TAGAGAATTGTCTTAAACAGTTCATGAACCCAGAAGAATGCAATTTCAATGTATTTAGTACACACACA
GTATGTATATAAACACAACCTCACAGAATATATTTCCATACATTGGGTAGGTATGCACTTTGTGTATATA
TAATAATGTATTTCCATGCAGTTTTAAAATGTAGATATATTAATATCTGGATGCATTTTCTGTGCACTG
GTTTTATATGCCTTATGGAGTATATACTCACATGTAGCTAAATAGACTCAGGACTGCACATTCCTTGTGT
AGGTTGTGTGTGTGGTGGTTTTATGCATAAATAAAGTTTTACATGTGGTGAATATAAA
```

I.2.La séquence du variant 2 de HOTAIR.

## Homo sapiens HOX transcript antisense RNA (HOTAIR), transcript variant 2, long non-coding RNA

NCBI Reference Sequence: NR\_003716.3

[GenBank](#) [Graphics](#)

```
>NR_003716.3 Homo sapiens HOX transcript antisense RNA (HOTAIR), transcript variant 2,
long non-coding RNA
CCTCCAGGCCCTGCCTTCTGCCTGCACATTCTGCCCTGATTTCCGGAACCTGGAAGCCTAGGCAGGCAGT
GGGGAACCTCTGACTCGCCTGTGCTCTGGAGCTTGATCCGAAAGCTTCCACAGTGAGGACTGCTCCGTGGG
GGTAAGAGAGACCAGGCCTGAGGCCTGGGAGTTCCACAGACCAACACCCCTGCTCCTGGCGGCTCCCA
CCCGGACTTAGACCTCAGGTCCTAATATCCCGGAGGTGCTCTCAATCAGAAAGGTCTGCTCCGCTT
CGCAGTGGAAATGGAACGGATTTAGAAGCCTGCAGTAGGGGAGTGGGAGTGGAGAGAGGGAGCCAGAGT
TACAGACGGCGGCGAGAGGAAGGAGGGGCGTCTTTATTTTTTAAGGCCCAAAGAGTCTGATGTTTACA
AGACCAGAAATGCCACGGCCGCTCTGGCAGAGAAAAGGCTGAAATGGAGGACCAGCGCCTTCTTTATA
AGTATGCACATTGGCGAGAGAAGTGTGCAACCTAAACCAGCAATTACACCCAAGCTCGTTGGGGCTAA
GCCAGTACCGACCTGGTAGAAAAAGCAACCACGAAGCTAGAGAGAGAGCCAGAGGAGGGGAAGAGAGCGCC
AGACGAAGGTGAAAGCGAACACGCAGAGAAATGCAGGCAAGGGAGCAAGGCGGCAGTTCCCGGAACAAA
CGTGGCAGAGGGCAAGACGGGCACTCACAGACAGAGGTTTATGTATTTTTATTTTTTAAAATCTGATTTG
GTGTTCCATGAGGAAAAGGAAAATCTAGGGAACGGGAGTACAGAGAGAATAATCCGGGCTCTAGCTCGC
CACATGAACGCCAGAGAACGCTGGAAAAACCTGAGCGGGTGCCGGGGCAGCACCCGGCTCGGGTCAAGC
ACTGCCCCACACCGGGCCACCAAGCCCCGCCCTCGCGGCCACCGGGGCTTCTTGTCTTCTTATCAT
CTCCATCTTTATGATGAGGCTTGTTAAACAAGACCAGAGAGCTGGCCAAGCACCTCTATCTCAGCCGCGCC
CGCTCAGCCGAGCAGCGGTGCGTGGGGGGACTGGGAGGCGCTAATTAATTGATTCCTTTGGACTGTAATA
TATGGCGGCGTCTACACGGAACCCATGGACTCATAAACAATATATCTGTTGGGCGTGAGTGCAGTGTCTC
TCAAATAATTTTTCCATAGGCAAATGTCAGAGGGTTCTGGATTTTTAGTTGCTAAGGAAAGATCCAAATG
GGACCAATTTTAGGAGGCCAAACAGAGTCCGTTCAAGTGTGAGAAAATGCTTCCCAAAGGGGTTGGGAG
TGTGTTTTGTTGGAAAAAGCTTGGGTTATAGGAAAGCCTTCCCTGCTACTTGTGTAGACCCAGCCCAA
TTTTAAGAATTACAAGGAAGCGAAGGGGTTGTGAGGCCGGAAGCCTCTCTGTCCCGGCTGGATGCAGGGG
ACTTGAGCTGCTCCGGAATTTGAGAGGAACATAGAAGCAAAGGTCCAGCCTTTGCTTCGTGCTGATTCCT
AGACTTAAGATTCAAAAACAAATTTTTAAAAGTGAAACCAGCCCTAGCCTTTGGAAGCTCTTGAAGGTTT
AGCACCCACCCAGGAATCCACCTGCCTGTTACACGCTCTCCAAGACACAGTGGCACCGCTTTTCTAACT
GGCAGCACAGAGCAACTCTATAATATGCTTATATTAGGTCTAGAAGAATGCATCTTGAGACACATGGGTA
ACCTAATTATATAATGCTTGTTCCATACAGGAGTGATTATGCAGTGGGACCCCTGCTGCAAACGGGACTTT
GCACTCTAAATATAGACCCAGCTTGGGACAAAAGTTGCAGTAGAAAAATAGACATAGGAGAACACTTAA
ATAAGTGATGATGTAGACACAGAAGGGGATTTAAAAGACAGAAATAATAGAAGTACAGAAGAACAGAA
AAAAAATCAGCAGATGGAGATTACCATTCCCAATGCCTGAACTTCTCTGCTATTAAGATTGCTAGAGA
ATTGTGCTTAAACAGTTCATGAACCAGAAGAATGCAATTTCAATGTATTTAGTACACACACAGTATGT
ATATAAACACAACCTCACAGAATATATTTCCATACATTGGGTAGGTATGCACCTTGTGTATATATAATAA
TGTATTTTCCATGCAGTTTTAAAATGTAGATATATTAATATCTGGATGCATTTTCTGTGCACTGGTTTTA
TATGCCTTATGGAGTATACTCACATGTAGCTAAATAGACTCAGGACTGCACATTCCTTGTGTAGGTTG
TGTGTGTGGTGGTTTTATGCATAAATAAAGTTTTACATGTGGTGAATATAAA
```

### I.3. La séquence du variant 3 de HOTAIR.

#### Homo sapiens HOX transcript antisense RNA (HOTAIR), transcript variant 3, long non-coding RNA

NCBI Reference Sequence: NR\_047518.1

[GenBank](#) [Graphics](#)

```
>NR_047518.1 Homo sapiens HOX transcript antisense RNA (HOTAIR), transcript variant 3,
long non-coding RNA
CTAATCAATCCAAAAGGAAAAAGAAAAGAGAGGGGTGGGAAGGCATGGGGTGAAAACTTCAGGTGACAC
AAAGCAAAAAGATTCCAGTCACCAAAAACAAAAACAAAAAAATCTTTGCCCCAGCAAGAATCATTT
GTTTGGGATCTGTTCCAGCCTCCAGGCCCTGCCTTCTGCCTGCACATTCTGCCCTGATTTCCGGAACCTG
GAAGCCTAGGCAGGCAGTGGGGAACCTGACTCGCCTGTGCTCTGGAGCTTGATCCGAAAGCTTCCACAG
TGAGGACTGCTCCGTGGGGTCTGCTCCGCTTCCAGTGGGAATGGAACGGATTTAGAACCTGCAGTAGG
GGAGTGGGGAGTGGAGAGAGGGAGCCAGAGTTACAGACGGCGGCGAGAGGAGGGGCGTCTTTATTTTT
TAAGGCCCAAGAGTCTGATGTTTACAAGACCAGAAATGCCACGGCCGCTCTGGCAGAGAAAAGGCT
GAAATGGAGGACCGGCGCTTCTTATAAGCTCGTTGGGGCCTAAGCCAGTACCGACCTGGTAGAAAAAG
CAACCACGAAGCTAGAGAGAGAGCCAGAGGAGGGAAGAGAGCGCCAGACGAAGGTGAAAGCGAACCCGC
AGAGAAATGCAGGCAAGGGAGCAAGGCGGCAAGTCCCGGAACAAACGTGGCAGAGGGCAAGACGGGCACT
CACAGACAGAGGTTTATGTATTTTTATTTTTAAAATCTGATTTGGTGTTCATGAGGAAAAGGGAAAAT
CTAGGGAACGGGAGTACAGAGAGAATAATCCGGGTCTAGCTCGCCACATGAACGCCAGAGAACGCTGG
AAAAACCTGAGCGGGTCCCGGGGCGAGCACCAGGCTCGGGTCCAGCCACTGCCCCACACCGGGCCCAAG
CCCCGCCCTCGCGGCCACCGGGGCTTCTTGTCTTCTTATCATCTCCATCTTTATGATGAGGCTTGT
AACAAAGACCAGAGAGCTGGCCAAGCACCTCTATCTCAGCCGCGCCGCTCAGCCGAGCAGCGGTGGTGG
GGGACTGGGAGGCGCTAATTAATTGATTCCTTTGGACTGTAAAATATGGCGGCGTCTACACGGAAACCA
TGGACTCATAACAATATATCTGTTGGGCGTGAGTGCCTGTCTCAAATAATTTTTCCATAGGCAAT
GTCAGAGGGTCTGGATTTTTAGTTGCTAAGGAAAGATCCAAATGGGACCAATTTAGGAGGCCCAAACA
GAGTCCGTTCAAGTGTGAGAAAATGCTTCCCAAGGGGTTGGGAGTGTGTTTTGTTGAAAAAAGCTTGG
GTTATAGGAAAGCTTTCCCTGCTACTTGTGTAGACCAGCCCAATTTAAGAATTACAAGGAAGCGAAGG
GGTTGTGTAGGCCGGAAGCCTCTCTGTCGGGCTGGATGCAGGGGACTTGAGCTGCTCCGGAATTTGAGA
GGAACATAGAAGCAAAGTCCAGCCTTTGCTTCGTGCTGATTCCTAGACTTAAGATTCAAAAACAAATTT
TTAAAAGTGAACACAGCCTTAGCCTTTGGAAGCTTTGAAGGTTCCAGCACCCACCCAGGAATCCACCTGC
CTGTTACACGCCTCTCCAAGACACAGTGGCACCCTTTCTAACTGGCAGCACAGAGCAACTCTATAATA
TGCTTATATTAGGTCTAGAAGAATGCATCTTGAGACACATGGGTAACCTAATTATATAATGCTTGTCCCA
TACAGGAGTGATTATGCAGTGGGACCCTGCTGCAAACGGGACTTTGCACTCTAAATATAGACCCAGCTT
GGGACAAAAGTTGCAGTAGAAAAATAGACATAGGAGAACACTTAAATAAGTGTGATGCATGTAGACACAGAA
GGGGTATTTAAAAGACAGAAATAATAGAAGTACAGAAGAACAGAAAAAAATCAGCAGATGGAGATTACC
ATTCCAATGCCTGAACCTCTCCTGCTATTAAGATTGCTAGAGAATTGTGCTTAAACAGTTCATGAAC
CCAGAAGAATGCAATTTCAATGTATTTAGTACACACACAGTATGTATATAAACACAACCTCACAGAATATA
TTTTCCATACATTGGGTAGGTATGCACTTTGTGTATATATAATAATGTATTTCCATGCAAGTTTTAAAT
GTAGATATATTAATATCTGGATGCATTTCTGTGCACTGGTTTTATATGCCTTATGGAGTATATACTCAC
ATGTAGCTAAATAGACTCAGGACTGCACATTCCTGTGTAGGTTGTGTGTGTGGTGGTTTTATGCATA
AATAAAGTTTTACATGTGGTGAATATA
```

I.4.La séquence du transcrit de MALAT-1.

**Homo sapiens chromosome 11, GRCh38.p7 Primary Assembly**

NCBI Reference Sequence: NC\_000011.10

```
>NC_000011.10:65497679-65504494 Homo sapiens chromosome 11, GRCh38.p7 Primary Assembly
TTTGTCCCTGACGCAGCCCCACCGTTGCGCAGTCCCTCCCCGCCCCGCTCTCCCTCCGCAGCCTGCA
GCCCGAGACTTCTGTAAAGGACTGGGGCCCCGCAACTGGCCTCTCTGCCCTCTTAAGCGCAGCGCCATT
TTAGCAACGCAGAAGCCCCGGCGCCGGGAAGCCTCAGCTCGCCTGAAGGCAGGTCCTCTGACGCCTCCG
GGAGCCCAGGTTTCCAGAGTCTTGGGACGCAGCGACGAGTTGTGCTGCTATCTTAGCTGTCCTTATAG
GCTGGCCATTCCAGGTGGTGGTATTTAGATAAAACCACTCAAACCTGTCAGTTTGGTCTTGGGGTTTGGG
GGAAAGCTTTTATTTTCTTCTGCTCCGGTTCAGAAGGTCTGAAGCTCATACTAACCAGGCATAACAC
AGAATCTGCAAAACAAAACCCCTAAAAAAGCAGACCCAGAGCAGTGTAAACACTTCTGGGTGTGTCCCT
GACTGGCTGCCCAAGGTCTCTGTGTCTTGGGAGACAAAGCCATTGCTTAGTTGGTCTACTTTAAAGGC
CACTTGAAGTCTGCTTCCATGGCGATTTGCCTTGTGAGCACTTTCAGGAGAGCCTGGAAGCTGAAAAACG
GTAGAAAAATTTCCGTGCGGGCCGTGGGGGGCTGGCGGCAACTGGGGGGCCGCAGATCAGAGTGGGCCAC
TGGCAGCCAACGCCCCCGGGGCTCAGGCGGGGAGCAGCTCTGTGGTGTGGGATTGAGGCGTTTTCCAAAG
AGTGGGTTTTACGTTTCTAAGATTTCCCAAGCAGACAGCCCGTGTGCTCCGATTTCTGAACAAAAAA
GCAAAACGTGTGGCTGTCTTGGGAGCAAGTCGCAGGACTGCAAGCAGTTGGGGGAGAAAGTCCGCCATTT
TGCCACTTCTCAACCGTCCCTGCAAGGCTGGGGCTCAGTTGCGTAATGGAAAGTAAAGCCCTGAACTATC
ACACTTTAATCTTCTTCAAAGGTGGTAACTATACTACTGTCCCTCAAGAGAACACAAGAAGTGCTT
TAAGAGGTATTTAAAGTTCCGGGGGTTTTGTGAGGTGTTGATGACCCGTTTAAATATGATTTCCAT
GTTTCTTTTGTCTAAAGTTTGCAGCTCAAATCTTCCACACGCTAGTAATTTAAGTATTTCTGCATGTGT
AGTTTGCATTCAAGTTCATAAGCTGTTAAGAAAAATCTAGAAAAGTAAACTAGAACCTATTTTAAACC
GAAGAAGTACTTTTTGCCTCCCTCACAAGGCGGGCGGAAGGTGATCGAATTCGGGTGATGCGAGTTGTTT
TCCGTCTATAAATACGCTCGCCGAGCTGTGCGGTAGGCATTGAGGCAGCCAGCCAGGGGCTTCTGCT
GAGGGGGCAGGCGGAGCTTGAAGGAAACCGCAGATAAGTTTTTTCTCTTGAAGATAGAGATTAATACA
ACTACTTAAAAAATATAGTCAATAGGTTACTAAGATATTGCTTAGCGTTAAGTTTTAACGTAATTTTAA
TAGCTTAAGATTTTAAAGAAAAATATGAAGACTTAGAAGAGTAGCATGAGGAAGGAAAAGATAAAAGGTT
TCTAAACATGACGGAGGTTGAGATGAAGCTTCTCATGGAGTAAAAATGTATTTAAAGAAAAATTGAG
AGAAAGGACTACAGAGCCCCGAATTAATACCAATAGAAGGGCAATGCTTTTAGATTAAAATGAAGGTGAC
TTAAACAGCTTAAAGTTTAGTTTAAAGTTGTAGGTGATTAATAATTTGAAGGCGATCTTTAAAAAG
AGATTAACCGAAGGTGATTAAGAGACCTTGAATCCATGACGCAGGGAGAATTGCGTCATTTAAGCCT
AGTTAACGCATTTACTAAACGCAGACGAAAATGAAAAGATTAATTGGGAGTGGTAGGATGAAACAATTTG
GAGAAGATAGAAGTTTGAAGTGGAAAACGGAAGACAGAAGTACGGGAAGGCGAAGAAAAGAATAGAGAA
GATAGGGAAATTAGAAGATAAAACATACTTTTAGAAGAAAAAGATAAATTTAAACCTGAAAAGTAGGA
AGCAGAAGAAAAAGACAAGCTAGGAAAACAAAAGCTAAGGGCAAAATGTACAACTTAGAAGAAAATTG
GAAGATAGAAAACAAGATAGAAAATGAAAATATTGTCAAGAGTTTCAAGATAGAAAATGAAAAACAAGCTAA
GACAAGTATTGGAGAAGTATAGAAGATAGAAAATATAAAGCCAAAATTTGGATAAAAATAGCACTGAAAA
AATGAGGAAATTATTGGTAACCAATTTATTTTAAAGCCCATCAATTTAATTTCTGGTGGTGCAGAAGTT
AGAAGGTAAGCTTGAAGAAGTAGGGTGTACGTAGACCAGAACCAATTTAGAAGAATACTTGAAGCT
AGAAGGGGAAGTTGGTTAAAAATCACATCAAAAAGCTACTAAAAGGACTGGTGTAAATTTAAAAAACTA
AGGCAGAAGGCTTTTGGAAAGATTAGAAGAATTTGAAGGCCCTAAATATAGTAGCTTAGTTTGAAAAAT
GTGAAGGACTTTCGTAACGGAAGTAATCAAGATCAAGAGTAATTACCAACTTAATGTTTTGCATTGGA
CTTCACTTACATTATTTTTTAAATGCTCAGCACTAGCATTAAATTCAGAGCTGAGGACCTGCTAGCA
```

I.5.La séquence du variant 1 de MALAT-1.

**Homo sapiens metastasis associated lung adenocarcinoma transcript 1 (non-protein coding) (MALAT1), transcript variant 1, long non-coding RNA**

NCBI Reference Sequence: NR\_002819.4

[GenBank](#) [Graphics](#)

```
>NR_002819.4 Homo sapiens metastasis associated lung adenocarcinoma transcript 1 (non-
protein coding) (MALAT1), transcript variant 1, long non-coding RNA
CGCAGCCTGCAGCCCGAGACTTCTGTAAGGACTGGGGCCCCGCAACTGGCCTCTCCTGCCCTCTTAAGC
GCAGCGCCATTTAGCAACGCAGAAGCCCGGCGCCGGGAAGCCTCAGCTCGCCTGAAGGCAGGTCCCCTC
TGACGCCTCCGGGAGCCAGGTTTCCAGAGTCTTGGGACGCAGCGACGAGTTGTGCTGCTATCTTAGC
TGTCTTATAGGCTGGCCATTCAGGTGGTGGTATTTAGATAAAACCACTCAAACCTGCAGTTTGGTCT
TGGGGTTTGGAGGAAAGCTTTATTTTCTTCTGCTCCGGTTCAGAAGTCTGAAGCTCATAACCTAACCC
AGGCATAACACAGAATCTGCAAAACAAAAACCCCTAAAAAAGCAGACCCAGAGCAGTGTAACACTTCTG
GGTGTGTCCCTGACTGGCTGCCCAAGGTCTCTGTGTCTTGGAGACAAAGCCATTGCTTAGTTGGTCTA
CTTTAAAGGCCACTTGAACCTGCTTTCCATGGCGATTTGCCTTGTGAGCACTTTCAGGAGAGCCTGGAA
GCTGAAAAACGGTAGAAAAATTTCCGTGCGGGCCGTGGGGGGCTGGCGGCAACTGGGGGGCCGCAGATCA
GAGTGGGCCACTGGCAGCCAACGGCCCCGGGGCTCAGGCGGGGAGCAGCTCTGTGGTGTGGGATTGAGG
CGTTTTCAAGAGTGGGTTTTACGTTTTCTAAGATTTCCAAGCAGACAGCCGTGCTGCTCCGATTTCT
CGAACAAAAAGCAAAACGTGTGGCTGTCTTGGGAGCAAGTCGCAGGACTGCAAGCAGTTGGGGGAGAAA
GTCCGCCATTTGCCACTTCTCAACCGTCCCTGCAAGGCTGGGGCTCAGTTGCGTAATGGAAAGTAAAGC
CCTGAACTATCACACTTTAATCTTCTTCAAAGGTGGTAACTATACCTACTGTCCCTCAAGAGAACAC
AAGAAGTGCTTTAAGAGGTATTTAAAGTTCGGGGGTTTTGTGAGGTGTTTATGACCCGTTTAAAT
ATGATTTCCATGTTTCTTTGTCTAAAGTTTGCAGCTCAAATCTTCCACACGCTAGTAATTTAAGTATT
TCTGCATGTGTAGTTTGCATTCAAGTTCATAAGCTGTTAAGAAAAATCTAGAAAAGTAAACTAGAACCC
TATTTTAAACCGAAGAACTACTTTTTGCCTCCCTCAAAAGGCGGCGGAAGGTGATCGAATTCGGTGAT
GCGAGTTGTTCTCGTCTATAAATACGCCTCGCCGAGCTGTGCGGTAGGCATTGAGGCAGCCAGCGCAG
GGGCTTCTGCTGAGGGGGCAGGCGGAGCTTGAGGAAACCGCAGATAAGTTTTTTCTTTGAAAGATAG
AGATTAATACAACACTTAAAAAATATAGTCAATAGGTTACTAAGATATTGCTTAGCGTTAAGTTTTTAA
CGTAATTTAATAGCTTAAGATTTTAAAGAAAAATATGAAGACTTAGAAGAGTAGCATGAGGAAGGAAAA
GATAAAAGGTTTCTAAAACATGACGGAGGTTGAGATGAAGCTTCTTATGGAGTAAAAAATGTATTTAAA
AGAAAATTGAGAGAAAGGACTACAGAGCCCCGAATTAATACCAATAGAAGGGCAATGCTTTTAGATTTAA
ATGAAGGTGACTTAAACAGCTTAAAGTTTAGTTTAAAGTTGTAGGTGATTTAAATAATTTGAAGGCGAT
CTTTTAAAGAGATTAAACCGAAGGTGATTTAAAGACCTTGAATCCATGACGCAGGGAGAATTGCGTC
ATTTAAAGCCTAGTTAACGCATTTACTAAACGCAGACGAAAATGGAAAGATTAATTGGGAGTGGTAGGAT
GAAACAATTTGGAGAAGATAGAAGTTTGAAGTGGAAAACCTGGAAGACAGAAGTACGGGAAGGCGAAGAAA
AGAATAGAGAAGATAGGGAATTAGAAGATAAAAACATACTTTTAGAAGAAAAAAGATAAATTTAAACCT
GAAAAGTAGGAAGCAGAAGAAAAAAGACAAGCTAGGAAAACAAAAAGCTAAGGGCAAAATGTACAACTTA
GAAGAAAATTGGAAGATAGAAAACAGATAGAAAATGAAAATATTGTCAAGAGTTTCAGATAGAAAATGAA
AAACAAGCTAAGACAAGTATTGGAGAAGTATAGAAGATAGAAAAATATAAAGCCAAAAATTGGATAAAAT
AGCACTGAAAAAATGAGGAAATTATTGGTAACCAATTTATTTTAAAGCCCATCAATTTAATTTCTGGTG
GTGCAGAAGTTAGAAGGTAAAGCTTGAAGATGAGGGTGTACGTAGACCAGAACCAATTTAGAAGAA
TACTTGAAGCTAGAAGGGGAAGTTGGTTAAAAATCACATCAAAAAGCTACTAAAAGGACTGGTGAATTT
```

I.6.La séquence du variant 2 de MALAT-1.

**Homo sapiens metastasis associated lung adenocarcinoma transcript 1 (non-protein coding) (MALAT1), transcript variant 2, long non-coding RNA**

NCBI Reference Sequence: NR\_144567.1

[GenBank](#) [Graphics](#)

```
>NR_144567.1 Homo sapiens metastasis associated lung adenocarcinoma transcript 1 (non-
protein coding) (MALAT1), transcript variant 2, long non-coding RNA
CGCAGCCTGCAGCCCAGACTTCTGTAAAGGACTGGGGCCCCGCAACTGGCCTCTCCTGCCCTCTTAAGC
GCAGCGCCATTTAGCAACGCAGAAGCCCGGCGCCGGGAAGCCTCAGCTCGCCTGAAGGCAGGTCCCTC
TGACGCCTCCGGGAGCCCAGGTTTCCAGAGTCTTGGGACGCAGCGACGAGTTGTGCTGCTATCTTAGC
TGTCTTATAGGCTGGCCATTCAGGTGGTGGTATTTAGATAAAACCACTCAAACCTCGCAGTTTGGTCT
TGGGGTTTGGAGGAAAGCTTTATTTTCTTCTGCTCCGGTTCAGAAGTCTGAAGCTCATACCTAACCC
AGGCATAACACAGAATCTGCAAAAACAAAACCCCTAAAAAAGCAGACCCAGAGCAGTGTAAACACTTCTG
GGTGTGTCCCTGACTGGCTGCCAAGGTCTCTGTGTCTTCGGAGACAAAGCCATTGCTTAGTTGGTCTA
CTTTAAAAGGCCACTTGAACCTCGCTTCCATGGCGATTTGCCTTGTGAGCACTTTCAGGAGAGCCTGGAA
GCTGAAAAACGGTAGAAAAATTTCCGTGCGGGCCGTGGGGGGCTGGCGGCAACTGGGGGGCCGAGATCA
GAGTGGGCCACTGGCAGCCAACGGCCCCCGGGGCTCAGGCGGGGAGCAGCTCTGTGGTGGGATTGAGG
CGTTTTCAAGAGTGGGTTTTACGTTTTTAAGATTTCCCAAGCAGACAGCCCGTGTGCTCCGATTTCT
CGAACAAAAAAGCAAAACGTGTGGCTGTCTTGGGAGCAAGTCGCAGGACTGCAAGCAGTTGGGGGAGAAA
GTCCGCCATTTGCCACTTCTCAACCGTCCCTGCAAGGCTGGGGCTCAGTTGCGTAATGGAAAGTAAAGC
CCTGAACTATCACACTTTAATCTTCTTCAAAGGTGGTAAACTATACCTACTGTCCCTCAAGAGAACAC
AAGAAGTGCTTTAAGAGGCGGCGGAAGGTGATCGAATTCGGTGATGCGAGTTGTTCTCCGTCTATAAAT
ACGCCTCGCCGAGCTGTGCGGTAGGCATTGAGGCAGCCAGCGCAGGGGCTTCTGCTGAGGGGGCAGGGC
GAGCTTGAGGAAACCGCAGATAAGTTTTTTCTTTGAAAGATAGAGATTAATAACAACACTTAAAAAAA
TATAGTCAATAGGTTACTAAGATATTGCTTAGCGTTAAGTTTTAACGTAATTTAATAGCTTAAGATTT
TAAGAGAAAATATGAAGACTTAGAAGAGTAGCATGAGGAAGGAAAAAGATAAAAAGTTTTCTAAAACATGAC
GGAGGTTGAGATGAAGCTTCTCATGGAGTAAAAATGTATTTAAAAGAAAATTGAGAGAAAGGACTACA
GAGCCCCGAATTAATACCAATAGAAGGGCAATGCTTTAGATTAATAAGGAGTAAACAGCTTAA
AGTTTAGTTTAAAAGTTGTAGGTGATTAATAATTTGAAGGCGATCTTTAAAAAGAGATTAACCGAA
GGTGATTAAGACCTTGAAATCCATGACGCAGGGAGAATTGCGTCATTTAAGCCTAGTTAACGCATTT
ACTAAACGCAGACGAAAATGGAAAGATTAATTGGGAGTGGTAGGATGAAACAATTTGGAGAAGATAGAAG
TTTGAAGTGGAAAACCTGGAAGACAGAAGTACGGGAAGGCGAAGAAAAGAATAGAGAAGATAGGGAAATTA
GAAGATAAAAACATACTTTTGAAGAAAAAGATAAATTTAAACCTGAAAAGTAGGAAGCAGAAGAAAAA
AGACAAGCTAGGAAACAAAAGCTAAGGGCAAAATGTACAACTTAGAAGAAAATTGGAAGATAGAAACA
AGATAGAAAATGAAAATATTGTCAAGAGTTTCAAGATAGAAAATGAAAACAAGCTAAGACAAGTATTGGA
GAAGTATAGAAGATAGAAAATATAAAGCCAAAATTTGGATAAAATAGCACTGAAAAAATGAGGAAATTA
TTGGTAACCAATTTATTTAAAAGCCCATCAATTTAATTTCTGGTGGTGCAGAAAGTTAGAAGGTAAAGCT
TGAGAAGATGAGGGTGTTCAGTAGACCAGAACCAATTTAGAAGAATACTTGAAGCTAGAAGGGGAAGTT
GGTTAAAAATCACATCAAAAAGCTACTAAAAGGACTGGTGTAAATTTAAAAAAAATAAGGCAGAAAGGCTT
TTGGAAGAGTTAGAAGAATTTGGAAGGCCTTAAATATAGTAGCTTAGTTTAAAAAATGTGAAGGACTTTC
GTAACGGAAAGTAATCAAGATCAAGAGTAATTACCAACTTAATGTTTTTGCATTGGACTTTGAGTTAAGA
TTATTTTTTAAATCCTGAGGACTAGCATTAAATTGACAGCTGACCCAGGTGCTACACAGAAGTGGATTCAG
```

I.7.La séquence du variant 3 de MALAT-1.

**Homo sapiens metastasis associated lung adenocarcinoma transcript 1 (non-protein coding) (MALAT1), transcript variant 3, long non-coding RNA**

NCBI Reference Sequence: NR\_144568.1

[GenBank](#) [Graphics](#)

```
>NR_144568.1 Homo sapiens metastasis associated lung adenocarcinoma transcript 1 (non-
protein coding) (MALAT1), transcript variant 3, long non-coding RNA
CGCAGCCTGCAGCCCGAGACTTCTGTAAAGGACTGGGGCCCCGCAACTGGCCTCTCCTGCCCTCTTAAGC
GCAGCGCCATTTTAGCAACGCAGAAGCCCGGCGCCGGGAAGCCTCAGCTCGCCTGAAGGCAGGTCCCTC
TGACGCCCTCCGGGAGCCAGGTTTCCAGAGTCTTGGGACGCAGCGAGGTTGTGTGCTATCTTAGC
TGTCTTATAGGCTGGCCATTCCAGGTGGTGGTATTTAGATAAAACCACTCAAACCTGCAGTTTGGTCT
TGGGGTTTGGAGGAAAGCTTTATTTTTCTTCTGCTCCGGTTCAGAAGTCTGAAGCTCATACCTAACCC
AGGCATAACACAGAATCTGCAAAACAAAACCCCTAAAAAAGCAGACCCAGAGCAGTGTAAACACTTCTG
GGTGTGTCCCTGACTGGCTGCCCAAGTCTCTGTGTCTTGGGAGACAAAGCCATTGCTTAGTTGGTCTA
CTTTAAAAGGCCACTTGAACCTCGCTTCCATGGCGATTTGCCTTGTGAGCACTTTCAGGAGAGCCTGGAA
GCTGAAAAACGGTAGAAAAATTTCCGTGCGGGCCGTGGGGGGCTGGCGGCAACTGGGGGGCCGCAGATCA
GAGTGGGCCACTGGCAGCCAACGGCCCCGGGGCTCAGGCGGGGAGCAGCTCTGTGGTGTGGGATTGAGG
CGTTTTCCAAGAGTGGGTTTTACGTTTCTAAGATTTCCAAGCAGACAGCCCGTGTCTCCGATTTCT
CGAACAAAAAGCAAAACGTGTGGCTGTCTTGGGAGCAAGTGCAGGACTGCAAGCAGTTGGGGGAGAAA
GTCCGCCATTTTGCCACTTCTCAACCGTCCCTGCAAGGCTGGGGCTCAGTTGCGTAATGGAAAGTAAAGC
CCTGAACTATCACACTTTAATCTTCCCTTCAAAGGTGGTAACTATACTACTGTCCCTCAAGAGAACAC
AAGAAGTGCTTTAAGAGGCGGGCGGAAGGTGATCGAATTCGGGTGATGCGAGTTGTTCTCCGTCTATAAAT
ACGCCCTCGCCGAGCTGTGCGGTAGGCATTGAGGCAGCCAGCGCAGGGGCTTCTGCTGAGGGGGCAGGCG
GAGCTTGAGGAAACCGCAGATAAGTTTTTTCTTTGAAAGATAGAGATTAATACAACCTACTTAAAAAA
TATAGTCAATAGGTTACTAAGATATTGCTTAGCGTTAAGTTTTAACGTAATTTAATAGCTTAAGATTT
TAAGAGAAAATATGAAGACTTAGAAGAGTAGCATGAGGAAGGAAAAGATAAAAGGTTTCTAAAACATGAC
GGAGGTTGAGATGAAGCTTCTCATGGAGTAAAAAATGATTTAAAAGAAAATTGAGAGAAAGGACTACA
GAGCCCCGAATTAATACCAATAGAAGGGCAATGCTTTTAGATTTAAAATGAAGGTGACTTAAACAGCTTAA
AGTTTAGTTTAAAAGTTGTAGGTGATTAAAATAATTTGAAGGCGATCTTTAAAAGAGATTAAACCGAA
GGTGATTAAGACCTTGAATCCATGACGCAGGGAGAATTGCGTCATTTAAAGCCTAGTTAACGCATTT
ACTAAACGCAGACGAAAATGGAAAGATTAATTGGGAGTGGTAGGATGAAACAATTTGGAGAAGATAGAAG
TTTGAAGTGGAAAACCTGGAAGACAGAAGTACGGGAAGGCGAAGAAAAGAATAGAGAAGATAGGGAAATTA
GAAGATAAAAACATACTTTTGAAGAAAAAAGATAAATTTAAACCTGAAAAGTAGGAAGCAGAAGAAAAA
AGACAAGCTAGGAAACAAAAGCTAAGGGCAAAATGTACAACTTAGAAGAAAATTGGAAGATAGAAAACA
AGATAGAAAATGAAAATATTGTCAAGAGTTTCCAGATAGAAAATGAAAACAAGCTAAGACAAGTATTGGA
GAAGTATAGAAGATAGAAAATATAAAGCCAAAATTTGGATAAAAATAGCACTGAAAAAATGAGGAAATTA
TTGGTAACCAATTTATTTAAAAGCCCATCAATTTAATTTCTGGTGGTGCAGAAGTTAGAAGGTAAGCT
TGAGAAGATGAGGGTGTACGTAGACCAGAACCAATTTAGAAGAATACTTGAAGCTAGAAGGGGAAGTT
GGTTAAAATCACATCAAAAAGCTACTAAAAGGACTGGTGTAAATTTAAAAAAAATAAGGCAGAAGGCTT
TTGGAAGAGTTAGAAGAATTTGGAAGGCCTTAAATATAGTAGCTTAGTTTGAAAAATGTGAAGGACTTTC
GTAACGGAAGTAATCAAGATCAAGAGTAATTACCAACTAATGTTTTGCATTGGACTTTGAGTTAAGA
TTATTTTTTAAATCCTGAGGACTAGCATTAAATGACAGCTGACCCAGGTGCTACACAGAAGTGGATTGAG
```

---

## II. La séquence codante de la p53 (CDS).

### Homo sapiens p53 gene for transformation related protein p53 (also called transformation-associated protein p53, cellular tumor antigen p53, and non-viral tumour antigen p53)

GenBank: X54156.1

[GenBank](#) [Graphics](#)

```
>(X54156.1:11717-11790, X54156.1:11906-11927, X54156.1:12021-12299, X54156.1:13055-13238,
X54156.1:13320-13432, X54156.1:14000-14109, X54156.1:14452-14588, X54156.1:14681-14754,
X54156.1:17572-17678, X54156.1:18599-18680) Homo sapiens p53 gene for transformation
related protein p53 (also called transformation-associated protein p53, cellular tumor
antigen p53, and non-viral tumour antigen p53)
ATGGAGGAGCCGCGAGTCAGATCCTAGCGTCGAGCCCCCTCTGAGTCAGGAAACATTTTCAGACCTATGGA
AACTACTTCCTGAAAACAACGTTCTGTCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCC
GGACGATATTGAACAATGGTCACTGAAGACCCAGGTCCAGATGAAGCTCCAGAAATGCCAGAGGCTGCT
CCCCGCGTGGCCCCCTGCACCAGCAGCTCCTACACGGCGGCCCTGCACCAGCCCCCTCTGGCCCCCTGT
CATCTTCTGTCCCTTCCCAGAAAACCTACCAGGGCAGCTACGGTTTCCGTCTGGGCTTCTTGCAATTCTGG
GACAGCCAAGTCTGTGACTTGCACGTACTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACC
TGCCCTGTGCAGCTGTGGGTTGATTCCACACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACA
AGCAGTCACAGCACATGACGGAGGTTGTGAGGCGCTGCCCCACCATGAGCGCTGCTCAGATAGCGATGG
TCTGGCCCCCTCCTCAGCATCTTATCCGAGTGAAGGAAATTTGCGTGTGGAGTATTTGGATGACAGAAAC
ACTTTTCGACATAGTGTGGTGGTGCCCTATGAGCCGCTGAGGTTGGCTCTGACTGTACCACCATCCACT
ACAACACATGTGTAACAGTTCCTGCATGGGCGGCATGAACCGGAGGCCATCCTCACCATCATCACACT
GGAAGACTCCAGTGGTAATCTACTGGGACGGAACAGCTTTGAGGTGCGTGTTTGTGCCTGTCCTGGGAGA
GACCGGCGCACAGAGGAAGAGAATCTCCGCAAGAAAGGGGAGCCTCACCACGAGCTGCCCCAGGGAGCA
CTAAGCGAGCACTGCCAACACACAGCTCCTCTCCCCAGCCAAAGAAGAAACCACTGGATGGAGAATA
TTTACCCTTCAGATCCGTGGGCGTGAGCGCTTCGAGATGTTCCGAGAGCTGAATGAGGCCTTGGAACTC
AAGGATGCCAGGCTGGGAAGGAGCCAGGGGGGAGCAGGGCTCACTCCAGCCACCTGAAGTCCAAAAGG
GTCAGTCTACCTCCCGCCATAAAAACTCATGTTCAAGACAGAAGGGCCTGACTCAGACTGA
```

## Annexe II

### Exons de HOTAIR/MALAT-1/P53

#### 1. HOTAIR

<b>Exon1</b> 1-140	cctccaggccctgccttctgcctgcacattctgcctgattccggaacctggaagcctagggcaggcagtgggg aactctgactcgctgtgctctggagcttgatccgaaagcttcacagtgaggactgctccgtggg
<b>Exon2</b> 141-266	ggtaagagagcaccaggcactgaggcctgggagttccacagaccaacaccctgctcctggcggctccac ccgggacttagacctcaggtccctaataccccggaggtgctctcaatcagaaag
<b>Exon3</b> 267-368	gtcctgctccgcttcgcagtggaatggaacggatttagaagcctgcagttaggggagtggggagtgagagag ggagcccagagttacagacggcggcgagag
<b>Exon4</b> 369-492	gaaggaggggcgtctttatTTTTAAGGCCCAAAGAGTCTGATGTTACAAGACCAGAAATGCCACGGCCGC gtcctggcagagaaaaggctgaaatggaggaccggcgccttcctataag
<b>Exon5</b> 493-545	tatgcacattggcgagagaagtgtgcaacctaaaccageaattacaccaag
<b>Exon6</b> 546-2362	ctcggtggggcctaagccagtaccgacctggtagaaaaagcaaccacgaagctagagagagagccagagga gggaagagagcgcagacgaaggtgaaagcgaaccacgcagagaaatgcaggcaaggagcaaggcgg cagttcccggaacaaacgtggcagagggcaagacgggcactcacagacagaggttatgtattttatTTTTAA aatctgatttgggtgtccatgaggaaaaggaaaatctagggaacgggagtagacagagaaataatccgggtcc tagctcgccatgaaccccagagaacgctggaaaaactgagcgggtgccggggcagcaccggctcg ggtcagccactgcccacaccggggccaccaagccccgccctcgcgccaccggggcttccttctcttctt atcatctccatcttatgatgaggcttgttaacaagaccagagagctggccaagcacctctatctcagccgccc cgctcagccgagcagcggcgggtggggggactgggaggcgttaattaattgattccttggactgtaaaatag gcggcgtctacacggaacctatggactcataaacaatatctgttggcgtgagtgcactgtctctcaataatt ttccataggcaaatgtcagaggggttctggatttttagttgctaaggaaagatccaaatgggaccaatttaggag gccccaaacagagtccttcagtgtcagaaaatgcttccccaaagggttgggagtggtgtttgttgaaaaaag cttgggtatagaaagccttccctgctactgtgtagaccagccaatttaagaattacaaggaagcgaagg ggttgttaggccggaagcctctctgcccggctggatgcaggggacttgagctgctccggaattgagagga acatagaagcaaaagtccagccttgcctctgctgattcctagacttaagattcaaaaacaaattttaaaagtga aaccagccctagccttggagctctgaaggtcagcaccaccaggaatccacctgctgttacacgcctct ccaagacacagtgaccgcttttctaactggcagcacagagcaactctataatgcttatattaggtctagaag aatgcatcttgagacacatgggtaacctaatatataatgcttgcatacaggagtgattatgcagtgggacct gctgcaaacgggacttgcactctaaatatagaccccagctgggacaaaagttgcagtagaaaaatagacata ggagaacacttaataagtgtatgatgtagacacagaaggggtatttaaaagacagaaataatagaagtacag aagaacagaaaaaaatcagcagatggagattaccattccaatgcctgaacttctctctgctattaagattgcta gagaattgtgtcttaaacagttcatgaaccagaagaatgcaattcaatgtatttagtacacacacagtatgtatat aacacaactcacagaatataatttccatacattgggtaggtatgcacttgtgtatatataataatgattttccatgc agttttaaagttagatatataatctggatgcaatttctgtgactggtttatgccttatggagtatatactcac atgtagctaaatagactcaggactgcacattcctgtgtaggtgtgtgtgtgtgggttttatgcataataaagtt ttacatgtggtaata

2. MALAT-1

<b>Exon1</b> 1-8779	cgcagcctgcagcccagacttctgtaaaggactggggccccgcaactggcctctcctgcccttfaagcgcagcgcattttag caacgcagaagcccggcggcgggaagcctcagctcgcctgaaggcaggtcccctctgacgcctccgggagcccaggttccc agagtcttgggacgcagcgcaggtgtgctgctatcttagctgtccttataggctggccattccaggtggtggtatttagataaaa ccactcaaacctgcagtttggcttgggggttggaggaaaagctttatcttctctgctccgggtcagaaggtctgaagctcatacc taaccaggcataacacagaatctgcaaaacaaaaaccctaaaaaagcagaccagagcagtgtaaacacttctgggtgtgtcc ctgactggctgcccgaaggtctgtgtcttcggagacaaagccattcgttagtggctactttaaaggccacttgaactcgtttc catggcgatttgccttgtgagcacttccaggagagcctggaagctgaaaaacggtagaaaaatctccgtcggggcctggggggg ctggcggcaactggggggccgcagatcagagtgggcccactggcagccaacggccccgggctcagcggggagcagct ctgtggtgtgggattgagcgtttccaagagtgggtttcacgtttctaaagatttcccaagcagacagcccgtgctcctgattct cgaacaaaaagcaaacgtgtggctgtcttgggagcaagtcgcaggactgcaagcagttgggggagaagtcgccattttg ccacttcaaccgtccctgcaaggtggggctcagttgctaatggaagtaaacctgaactatcacactttaatcttcttcaa aaggtgtaaacctatactactgtccctcaagagaacacaagaagtgctttaaagggtattttaaagggtccgggggtttgtgaggt gtttgatgaccgtttaaataatgattccatgtttctttgtctaaagttgacgtcctaaatcttccacagcctagttaattaaagtattct gcatgtgtgatttgcattcaagttccataagctgttaagaaaaatctagaaaagtaaacctagaactattttaaaccgaagaactact ttttgctccctcacaagcggcgggaaggtgatcgaattccgggtgatgcagttgttctccgtctataatacgcctcggccgag ctgtgcggtagggcattgaggcagccagcgcaggggcttctgctgagggggcagcgggagcttgaggaaaccgcagataagtt ttttctctttaaagatagagattaatacaactactaaaaatatagtcaataggttactaagatattgcttagcgttaaagttttaaactg aatttaataagcttaaagatttaagagaaaaatgaagacttagaagagtagcatgaggaaaggaaaagataaaaggtttctaaaca tgacggaggtgagatgaagcttctcatggagtaaaaaatgtattttaaagaaaattgagagaaaggactacagagccccgaatt aataccaatagaaggcaatgcttttagattaaaatgaaggtgacttaaacagcttaaagtttagtttaaaggtgtaggtgattaaaa taatttgaaggcgtctttttaaagagattaaaccgaaggtgattaaagacctgaaatccatgacgcagggagaattgcgtcat ttaaagcctagttaacgcatttactaaacgcagacgaaaatgaaagattaatgggagtgtaggatgaaacaattggagaaga tagaagttgaaagtgaaaactggaagacagaagtacgggaaaggcgaagaaaagaatagagaagataggaaatagaagat aaaaacatacttttagaagaaaaagataaaftaaacctgaaaagtaggaagcagaagaaaaagacaagctaggaacaaaa agctaagggcaaaatgtacaacttagaagaaaattggaagatagaacaagatagaatgaaaatattgtcaagagtttcagat agaaaatgaaaaacaagctaagacaagtattggagaagtagaagatagaatataaaagccaaaaattggataaaatagca ctgaaaaaatgaggaaatattgtaaccaatftattttaaagccatcaatttaatttctggtggtgcagaagttagaaggtaaagc ttgagaagatgaggggtttacgtagaccagaaccaatttagaagaacttgaagctagaaggggaggttggttaaaaatcacat caaaaagctactaaaaggactggtgtaatttaaaaaaaactaaaggcagaaggcttttggaagagttagaagaatttggaaagcctt aaatatagtagcttagtttgaaaaatgtgaaggacttctgtaacggaaagtaattcaagatcaagaatcaacttaattggtttg attggactttgagttagattttttaaactctgaggactagcattaatgacagctgaccaggtgctacacagaagtggattcag tgaatctaggaagacagcagcagacaggtccaggaaccaggtttgatgaagctaggactgaggagcaagcagcaagca gcagttcgtggtgaagataggaaaagagtcaggagccagtgctgatttggtaaggaagctaggaagaaggaaggagcgtca acgatttgggtgtaagctagaaaaagattccaggaaggagcagtgcaatttgggtgatgaaggtagcagcggccttggctt ggcaaccacacggaggagcgcagcagcgttgtgctgtagaggatcctagaccagcatgccagtggtccaagggcacaggga aagcaggtggttggtaaaaatccgtgaggtcggcaatattgttttctggaacttacttatgtaaccttttatttttctaatataat gggggagttctgactgaggtgtaaaggatftatatggggacgtaggccgatttccgggtgtgtaggttctcttttcagccttat actcatgaatcttctgaagctttttagggcagactgccaagtcctggagaatagtagatgcaagtttgggttttttttttaca cgaattgaggaaaaccaaatgaatttagatgccaatagacaatctcagcaaatctgtaagcagttttagtttagttgggttaa tgaagtattcagttttgtaatagatgacctgttttacttctcaccctgaattcgttttgaatgtagagtttgatgtgtaactgagg cgggggggagttttagtatttttttgggggtgggggcaaaatattttcagttcttttcccttaggtctgctagaatcctaaag gcaaatgactcaaggtgtaacagaaaacaagaaaatccaatatcaggataatcagaccaccaggtttacagttatagaacta gagcagttctcagttgaggtctgtggaagagatgcccattggagaatggctggttagttactttttccccccacccttaatca gactttaaagtgcttaacccttaaacctgttatttttactgaaagcattttgggatggtcttaacagggagaagagaggggtggggg agaaaatgttttttctaaagattttccacagatgctatagctactattgacaaactgggttagagaagaggtgaccgtgtgctgttggc acgaacacctcaggactggagctgctttatccttggaaagagtattccagttgaaagctgaaaagtacagcacagtgagctttg gttcatattcagtcctcaggaactcagaagagcttgagtaggccaatgtgaaagtttaagttttccaataatgtgacttcttaa agttttataaaggggaggggcaaatattggcaattagttggcagtgccctgtacgggtggattgggtgggtggttttaggtaatt gttttagtttagattgcagataaactcatgccagagaacttaagcttagaagttgaaaaagtaagaaatatacaactccaagttgg caagtaactccaatgatttagtttttccccagtttgaattgggaagctgggggaaagttaaatagagccactgggtgtaccag tgcatatttggcaaggaaggtgcataattgatactgtatctgttttcttcaaagtataagccttttgggaaagaaagattga actgggggttggctgacctgagcattactacaattatgggaaatgcaaaagttggttggataggtgtagtgggttct
------------------------	--

ctfttggaaftttttcaggtgatttaataataattaaaactactatagaactgcagagcaaaaggaagtggtctaatgatcctgaagg  
gatttctctgatggtagcttttgattatcaagtaagattctatttcagttgtgtgaagcaagtttttttagtgtaggagaataacttttc  
cattgttfaactgcaaaacaagatgtaaggtatgctcaaaaatttfgtaaattgttttttaacttatctgtttgtaaattgtaactgatt  
aagaattgtgatagttcagcttgaatgtctcttagagggtgggcttttggtagagggaggggaaactttttttctatagacttttt  
cagataacatctctgagtcataaccagcctggcagtagatgagccctagatgcagagaaaacagctccttggtaaftgataagta  
aaggcagaaaagattatgtcatacctcattggggaataagcataaccctgagattcttactactgatgagaacattatctgcata  
tgcaaaaaatttfaagcaaatgaagctaccaatttaagttacggaatctaccatttfaagtttaagttgcttcaagctataaccac  
aaaaataatgaattgatgagaataacaatgaagaggcaatgtccatctcaaaatactgcttttcaaaaagcagaataaaaagcgaaa  
agaaatgaaaatgttactactacattaatcctggaataaaaagaagccgaataaatgagagatgagttgggatcaagtggtgag  
gaggctgtgctgtgccaatgttctgttgcctcagacaggtatctctcgttatcagaagagttgcttattctctgggagcaga  
aacagcaggcagctgttaacagataagtttaactgtcatctgcagtattgcatgttagggataagtgcttatttfaagagctgtgga  
gttcttaaatatcaaccatggcactttctcctgacccttccttaggggatttcaggattgagaaattttccatcagccttttfaaatt  
gtagactgttctgtgggctcagtgatgggatagtacactcactcagagcatttgcatttfaaataatttcttaaaagcctcta  
aagtgatcagtgcttgccttgccttaaggaattttagcattgaatctctgaaggctctatgaaaggaatagcatgatgtgctg  
ttagaatcagatgttactgctaaaatttacctgtgtgatgaaattgtgtaaaaaccattfaaatcattcaaaaataataaactatttttatt  
agagaatgtatacttttagaaagctgtctccttattfaataaaaatagtggttctgtgagttcaggttggggcacttgggggggatt  
cttcttaattcttcagaaacttctcgcgaactcttfaattgaccagatcaggatttgcgggaagaacgaatgtaacttfaagg  
caggaaaagcaaaatttttctcataaagtgatgagcataataaattccaggcacatggcaatagaggccctctaaataaggat  
aaataacctcttagacaggtgggagattatgatcagagtaaaaggttaattacacattttattccagaaagtcaggggtctataaatg  
acagtgattagagtaataacttttccatttcaaaagttgcatgttaacttfaaatgcttacaatcttagagtggtaggcaatgttttaca  
ctattgacctatataaggaagggagggggtgcctgtgggggttfaaagaatttcttgcagaggcatttcatcctcatgaagcc  
atcaggattttgaattgcatatgagtgcttggctctccttctgtctagtgagtgatgagacctgagtgagttatcagcactca  
aaatttttctggaatttggagggatgggaggagggggtgggcttactgtttagcttttttttttacagacttcacagagaat  
gcagttgcttgacttcaggtctgtctgttctgttggcaagtaaatgcagtagtcttctgatcccgtcttattagaatgattgtgaaa  
cgactggagtagatgataaaagtgtgttccccaatgctggagtagtgattgtgaaggaaaaatccagctgagtgataaaggctg  
agtgttaggaaatttctgcagtttaagcagctgatttggtaggagctgagtagcatttctggtgtatttttagttaaagtctttt  
gttcttctgtgtggtggaggggactgaagcctttagtctttccagatgcaaccttaaaatcagtgacaagaacattccaacaa  
gcaacagcttcaagaaataaactggcaagtggaatgttfaaacagttcagtgatctttagtgattgtttagtggtgttctctct  
cccctccttggcttataattctfacatgcaggaaactcagcagacacacgtatgcgaagggccagagaagccagaccagtaa  
gaaaaatagcctatttacttfaataaaacaaacatttcaaatgtggggattgggaaccactagttcttccagatggtattctc  
agactatagaaggagctccagttgaattcaccagtggaacaaatgaggaaaacaggtgaacaagcttttctgtatttaccatacaa  
agtcagatcagttatgggacaatagattgaatagattcagctttatgctggagtaactggcagtgtagcaaaactgtgtggcgtgg  
gggtggaggggtgaggtgggcgctaagccttttttaagattttcaggtaccctcactaaaggcaccgaaggcttaagtagga  
caaccatggagccttctgtggcaggagagacaacaaagcgtattatcctaaggtcaagagaagtgcagcctcactgattttt  
attagtaatgaggacttgcctcaactcctcttctggagtgaagcatccgaaggaatgcttgaagtaccctgggcttcttcaaca  
tttaagcaagctgttttatagcagcttataataaaagcccaaatctcaagcgggtgcttgaaggggaggggaaaggggaaagg  
ggcaaccacttttccctagctttccagaagcctgttaaaagcaaggtctccccacaagcaacttcttccacatcgcacccctg  
gcctttgatctagcacagacccttaccctcactcagatgcagccagtagcttggatccttggggcatgatccataatcggttta  
aggtaacgatggtgctgaggtcttgggtgggtgaactatgttagaaaagccattaatttgcctgcaaatgttaacagaagggtat  
taaaaccacagctaaagtactctattataactattatccagtgactaaaaccaactfaaacagtaagtggagaataacatgttcaa  
gaactgtaatgctgggtgggaacatgtaactgttagactggagaagatagccatttgagtgctgagaggcctttgggtgggaat  
gcaaaaattctctgtaagacttttccaggtgaacataacagacttggccaagctagcatcttagcgggaagctgatctccaatgctct  
tcagtagggctatgaaggttttcttctgagaaaacaacagctattgttctcaggtttgcttttggccttttctagcttaaaaaaa  
aaaaaagcaaaagatgctggtggtggcactcctggttccaggacggggttcaaatccctgcggcgtcttggcttggactactaat  
ctgtctcaggactcttctgtatttctcttctctgcaggtgctagttctggagtttggggaggtgggaggttaacagcacaatc  
ttgaaactatataccttgatgtataattgtcaggagcttactgattgtatattcatattfacacgagaacctaatataactgctt  
tcttttccaggttaatagcctgcagctggtgtttgagaagccctactgctgaaaacttaacaatttggtaataaaaaatggagaagctc  
taaattgtgtggttctttgtgaataaaaaaatcttgattggggaaaaaa



---

## Annexe II

Structure 3D de la protéine P53 par RaptorX et RasMol

