



HAL
open science

Les séquences alu : synthèse bibliographique et étude de la réactivité d'une sonde commerciale par hybridation in situ sur les tissus de plusieurs espèces animales

Océane Juillard

► To cite this version:

Océane Juillard. Les séquences alu : synthèse bibliographique et étude de la réactivité d'une sonde commerciale par hybridation in situ sur les tissus de plusieurs espèces animales. Médecine vétérinaire et santé animale. 2024. dumas-04810907

HAL Id: dumas-04810907

<https://dumas.ccsd.cnrs.fr/dumas-04810907v1>

Submitted on 29 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Année 2024

**LES SÉQUENCES *ALU* : SYNTHÈSE
BIBLIOGRAPHIQUE ET ÉTUDE DE LA RÉACTIVITÉ D'UNE
SONDE COMMERCIALE PAR HYBRIDATION *IN SITU* SUR
LES TISSUS DE PLUSIEURS ESPÈCES ANIMALES**

THÈSE

pour obtenir le diplôme d'État de

DOCTEUR VÉTÉRINAIRE

présentée et soutenue publiquement devant

la Faculté de Médecine de Créteil (UPEC)

le 25 octobre 2024

par

Océane, Cécile, Claudette JUILLARD

sous la direction de

Édouard REYES-GOMEZ

JURY

Présidente du jury :	Mme Fanny PILOT-STORCK	Professeur à Vet Agro Sup
Directeur de thèse :	M. Édouard REYES-GOMEZ	Maître de conférences à l'EnvA
Examinatrice :	Mme Lucie CHEVALLIER	Maître de conférences à l'EnvA

Liste des enseignants intervenant dans l'encadrement des thèses de Doctorat vétérinaire



version Octobre 2024

Liste des Professeurs et Maîtres de conférences titulaires de l'HDR

M	Adjou	Karim	Professeur	DPASP
M	Audigié	Fabrice	Professeur	DEPEC
M	Bellier	Sylvain	Professeur	DSBP
Mme	Benchekroun	Ghita	Maître de conférences HDR	DEPEC
M	Blaga	Radu	Professeur	DSBP
M	Blot	Stéphane	Professeur	DEPEC
M	Boulouis	Henri-Jean	Professeur émérite	DSBP
Mme	Chahory	Sabine	Professeur	DEPEC
Mme	Chastant	Sylvie	Professeur	DPASP
M	Chateau	Henry	Professeur	DSBP
Mme	Chetboul	Valerie	Professeur	DEPEC
Mme	Crepeaux	Guillemette	Maître de conférences HDR	DSBP
Mme	Crevier-Denoix	Nathalie	Professeur	DSBP
M	Degueurce	Christophe	Professeur	DSBP
M	Denoix	Jean-Marie	Professeur émérite	DEPEC
M	Desquilbet	Loïc	Professeur	DSBP
Mme	Dufour	Barbara	Professeur émérite	DPASP
M	Fayolle	Pascal	Professeur émérite	DEPEC
M	Federighi	Michel	Professeur	DPASP
M	Fontbonne	Alain	Professeur	DEPEC
Mme	Gilbert	Caroline	Professeur	DSBP
Mme	Grimard-Ballif	Bénédicte	Professeur émérite	DPASP
Mme	Haddad-Hoang Xuan	Nadia	Professeur	DPASP
M	Jouvion	Gregory	Professeur	DSBP
M	Kohlhauer	Matthias	Professeur	DSBP
Mme	Le Poder	Sophie	Professeur	DSBP
Mme	Le Roux	Delphine	Professeur	DSBP
M	Manassero	Mathieu	Professeur	DEPEC
Mme	Maurey-Guénec	Christelle	Professeur	DEPEC
M	Millemann	Yves	Professeur	DPASP
M	Ponter	Andrew	Professeur	DPASP
Mme	Risco-Castillo	Véronica	Professeur	DSBP
Mme	Rivière	Julie	Professeur	DPASP
Mme	Robert	Céline	Professeur	DSBP
M	Tiret	Laurent	Professeur	DSBP
M	Tissier	Renaud	Professeur	DSBP
M	Verwaerde	Patrick	Professeur	DEPEC
Mme	Viateau	Véronique	Professeur	DEPEC

Liste des Maîtres de conférences et ingénieurs de recherche DMV

M	Arné	Pascal	Maître de conférences	DPASP
Mme	Barassin	Isabelle	Maître de conférences	DPASP
Mme	Barbarino	Alix	Ingénieur de recherche (DMV)	DEPEC
M	Béguin	Jérémy	Maître de conférences	DEPEC
Mme	Bertal	Mileva	Maître de conférences associé	DEPEC
Mme	Bertoni	Lelia	Maître de conférences	DEPEC
Mme	Canonne-Guibert	Morgane	Maître de conférences	DEPEC
Mme	Chevallier	Lucie	Maître de conférences	DSBP
Mme	Cochet-Faivre	Noelle	Maître de conférences	DEPEC
Mme	Cordonnier-Lefort	Nathalie	Maître de conférences	DSBP
Mme	Coudry	Virginie	Ingénieur de recherche (DMV)	DEPEC
M	Crozet	Guillaume	Maître de conférences	DPASP
Mme	De Paula Reis	Alline	Maître de conférences	DPASP
M	Delsart	Maxime	Maître de conférences	DPASP
Mme	Denis	Marine	Ingénieur de recherche (DMV)	DPASP
M	Deshuillers	Pierre	Maître de conférences	DSBP
Mme	Goutchat	Rebecca	Maître de conférences	DSBP
Mme	Guérin	Virginie	Maître de conférences	DSBP
Mme	Guélin-Poirier	Valentine	Maître de conférences	DPASP
Mme	Jacquet	Sandrine	Ingénieur de recherche (DMV)	DEPEC
M	Kurtz	Maxime	Ingénieur de recherche (DMV)	DEPEC
Mme	Lagrée	Anne-Claire	Maître de conférences	DSBP
Mme	Le Dudal	Marine	Ingénieur de recherche (DMV)	DSBP
Mme	Legrand	Chantal	Maître de conférences associé	DSBP
M	Mammeri	Mohamed	Maître de conférences	DSBP
Mme	Manguin	Estelle	Maître de conférences	DEPEC
Mme	Marignac	Genevieve	Maître de conférences	DSBP
Mme	Marotto	Stéphanie	Ingénieur de recherche (DMV)	DEPEC
Mme	Maurice	Emeline	Ingénieur de recherche (DMV)	DEPEC
Mme	Mespoules-Rivière	Céline	Ingénieur de recherche (DMV)	DEPEC
M	Mortier	Jérémy	Maître de conférences associé	DEPEC
Mme	Mtmet	Narjès	Maître de conférences	DPASP
M	Munir	Muhammad Tanveer	Maître de conférences	DPASP
M	Nudelmann	Nicolas	Maître de conférences	DEPEC
M	Pignon	Charly	Ingénieur de recherche (DMV)	DEPEC
M	Polack	Bruno	Maître de conférences	DSBP
Mme	Quéré	Émilie	Ingénieur de recherche (DMV)	DEPEC
M	Radière	Guéric	Maître de conférences associé	DSBP
Mme	Ravary-Plumioën	Bérangère	Ingénieur de recherche (DMV)	DPASP
M	Reyes-Gomez	Edouard	Maître de conférences	DSBP
Mme	Rose	Hélène	Maître de conférences associé	DSBP
M	Tanquerel	Ludovic	Maître de conférences	DEPEC
Mme	Touzot-Jourde	Gwenola	Maître de Conférences	DEPEC

Remerciements

Au Président du Jury de cette thèse, Madame Fanny PILOT-STORCK, Professeur à Vet Agro Sup

Pour nous avoir fait l'honneur d'accepter la présidence de ce jury de thèse.

Hommages respectueux.

A Monsieur Édouard REYES-GOMEZ, Maître de conférences à l'EnvA,

Pour m'avoir proposé et confié ce sujet.

Un grand merci pour votre encadrement et votre disponibilité tout au long de ce travail.

Très sincères remerciements.

A Madame Lucie CHEVALLIER, Maître de conférences à l'EnvA,

Pour avoir accepté de faire partie de ce jury de thèse.

Sincères remerciements.

A Madame Anne Girard-Luc ainsi que toute l'équipe de VetDiagnostics,

Pour la précieuse aide que vous nous avez apportée.

Sincères remerciements.

A Madame Sophie CHATEAU-JOUBERT,

Pour le temps que vous avez consacré à nos expériences.

Pour nos discussions sur le sujet.

Sincères remerciements.

A ma famille, à mes amis,

Pour votre présence et votre soutien.

Je vous remercie infiniment.

Table des matières

Liste des figures.....	3
Liste des tableaux.....	5
Liste des abréviations.....	7
Introduction.....	9
Première partie : bibliographie.....	11
1. Origine des séquences <i>Alu</i>	11
A. <i>Origine du nom</i>	11
B. <i>Un élément de la famille des SINE</i>	11
C. <i>Théories actuelles sur l'émergence de la séquence <i>Alu</i></i>	11
D. <i>Structures primaire et secondaire des séquences <i>Alu</i></i>	12
a. <i>Structure primaire des séquences <i>Alu</i></i>	12
b. <i>Structure secondaire des séquences <i>Alu</i></i>	13
c. <i>Conservation des structures primaire et secondaire</i>	14
2. Classification des séquences <i>Alu</i>	15
A. <i>Les différentes sous-familles</i>	15
B. <i>Composition des différentes sous-familles</i>	16
C. <i>Activité des différentes sous-familles au sein du génome humain</i>	16
3. Localisation et abondance des séquences <i>Alu</i> dans le génome humain.....	17
A. <i>Proportion et distribution des séquences <i>Alu</i></i>	17
a. <i>Proportion au sein du génome humain</i>	17
b. <i>Distribution dans le génome humain</i>	17
B. <i>Mécanisme d'amplification des séquences <i>Alu</i></i>	17
a. <i>Modèles hypothétiques de l'expansion des éléments <i>Alu</i></i>	17
b. <i>Modèle de rétrotransposition</i>	18
c. <i>Facteurs régulateurs de la transcription</i>	20
4. Conséquences de la présence de séquences <i>Alu</i>	21
A. <i>Conséquences fonctionnelles</i>	21
B. <i>Conséquences délétères des insertions <i>Alu</i> aléatoires</i>	22
C. <i>Marqueurs génétiques en hybridation <i>in situ</i></i>	24
a. <i>Principe de l'hybridation <i>in situ</i></i>	24
b. <i>Limites de l'hybridation <i>in situ</i> chromogénique</i>	25
c. <i>Séquences <i>Alu</i> comme marqueurs de tissus humains en hybridation <i>in situ</i></i>	25
5. Étude de la séquence <i>Alu</i> chez d'autres espèces.....	26
A. <i>Phylogénie des primates</i>	26
B. <i>Les séquence <i>Alu</i> au sein du génome simien</i>	28
a. <i>Chez le chimpanzé</i>	28
b. <i>Chez l'orang-outan</i>	28
c. <i>Chez le macaque rhésus</i>	28
d. <i>Chez les ouistitis</i>	29
e. <i>Chez les lémuriers</i>	29
f. <i>Chez les galagos</i>	29
C. <i>Dans le génome d'autres espèces</i>	30
a. <i>Au sein du génome murin</i>	30
b. <i>Au sein du génome des hamsters</i>	30
c. <i>Au sein du génome de la musaraigne arboricole</i>	31
d. <i>Au sein du génome de la myxine de bord de mer</i>	31
e. <i>Evolution des séquences dérivées de l'ARN 7SL</i>	31

Deuxième partie : étude expérimentale	33
1. Introduction et contexte de l'étude	33
2. Matériels et méthodes	34
A. Sélection des échantillons	34
a. Pour les espèces primates	34
b. Chez les espèces non primates	34
B. Technique du "Tissue Micro-Array"	34
C. Hybridation <i>in situ</i>	36
D. Évaluation du marquage de chaque échantillon.....	36
a. Lecture des lames.....	36
b. Comptage des cellules positives	37
3. Résultats	38
A. Chez les espèces primates.....	38
a. Description de l'échantillon.....	38
b. Résultats généraux	39
c. Étude de différents facteurs pouvant influencer la proportion de cellules marquées dans les échantillons.....	40
B. Chez les espèces non primates.....	45
a. Description de l'échantillon.....	45
b. Evaluation du marquage de la sonde <i>Alu</i>	45
4. Discussion	46
A. Limites de l'étude.....	46
a. Âge des différents prélèvements	46
b. Choix des prélèvements.....	46
c. Méthode de dénombrement	46
B. Discussion des résultats	47
a. Interprétation du marquage partiel	47
b. Interprétation chez les espèces primates	49
c. Interprétation chez les espèces non primates	49
d. Les séquences <i>Alu</i> : un marqueur utilisable en médecine légale ?	50
Conclusion.....	51
Liste des références bibliographiques	53
Annexe 1 : Protocole expérimental	59
Annexe 2 : Données concernant le marquage des cellules dans les tissus	63

Liste des figures

Figure 1 : Excision de la séquence <i>Fossil Alu Monomer</i> à partir de l'ARN 7SL.....	11
Figure 2 : Apparition de la séquence <i>Alu</i> originale à partir de la séquence <i>Fossil Alu Monomer</i>	12
Figure 3 : Représentation de la structure secondaire d'un ARN <i>Alu</i>	13
Figure 4 : Classification des éléments <i>Alu</i> en différentes sous-familles	15
Figure 5 : Rétrotransposition de l' <i>Alu</i> médiée par LINE1	19
Figure 6 : Technique de l'hybridation <i>in situ</i>	24
Figure 7 : Arbre phylogénétique des primates	26
Figure 8 : Arbre phylogénétique hypothétique de l'apparition des séquences <i>Alu</i> et apparentées .	32
Figure 9 : Schéma de la technique <i>Tissue Micro-Array</i>	35
Figure 10 : Photographie d'un bloc <i>Tissue Micro-Array</i>	35
Figure 11 : Détail de deux noyaux dont un marqué par la sonde <i>Alu</i>	37
Figure 12 : Répartition des organes testés au cours des deux expériences	39
Figure 13 : Exemple d'un tissu avec marquage positif	39
Figure 14 : Exemple d'un tissu avec marquage négatif.....	40
Figure 15 : Pourcentages de cellules marquées au sein de chaque échantillon	40
Figure 16 : Pourcentages de cellules marquées en fonction de la date de prélèvement du tissu...	41
Figure 17 : Répartition des moyennes de cellules marquées en fonction de l'organe	42
Figure 18 : Exemple de marquage important des cholangiocytes.....	43
Figure 19 : Pourcentages de cellules marquées en fonction des conditions analytiques	43
Figure 20 : Exemple d'un cas avec interprétation difficile du marquage.....	47

Liste des tableaux

Tableau 1 : Exemples de maladies génétiques liées à des séquences <i>Alu</i>	23
Tableau 2 : Photographies de plusieurs espèces de grands primates	27
Tableau 3 : Individus et organes testés au cours des deux expériences	38
Tableau 4 : Résultats du test de Wilcoxon concernant les conditions analytiques	44
Tableau 5 : Répartition des organes testés par espèce non primate	45

Liste des abréviations

ADN : Acide DésoxyriboNucléique
ARN : Acide RiboNucléique
DAB : 3,3-diaminobenzidine
EnvA : Ecole Nationale Vétérinaire d'Alfort
FAM : *Fossil Alu Monomer*
FLAM : *Free Left Arm Monomer*
FRAM : *Free Right Arm Monomer*
HES : Hémalun-Eosine-Safran
ISH : *In Situ Hybridization* ou hybridation *in situ*
LINE : *Long Interspersed Nuclear Elements*
LTR : *Long Terminal Repeat*
RTE : *Retrotransposon-like Element*
SINE : *Short Interspersed Nuclear Elements*
SRP : *Signal Recognition Particle*
TPRT : *Target-Primed Reverse Transcriptase*

Introduction

En génétique, les éléments intercalés courts (SINE pour *short interspersed repetitive elements*) sont des séquences répétées d'ADN rassemblant plusieurs familles dont la plus abondante dans le génome humain est la famille *Alu*, avec environ 1,2 millions de copies (Zhang *et al.*, 2021). Pour mieux comprendre les séquence *Alu*, il est important de faire quelques rappels sur les éléments transposables. Les éléments transposables sont des répétitions intercalées représentant plus de la moitié du génome chez les organismes supérieurs (Tang *et al.*, 2018). Parmi eux se trouvent les rétrotransposons qui représentent la grande majorité des éléments transposables dans le génome humain. Ces éléments peuvent être classés en deux catégories : les rétrotransposons comportant une longue répétition terminale (LTR) ou les rétrotransposons ne la possédant pas (non-LTR). Dans la deuxième catégorie se trouvent principalement les éléments intercalés longs (LINE pour *long interspersed repetitive elements*) et les éléments SINE, qui représentent respectivement 20,4 % et 13,1 % du génome humain. Ce sont des éléments qui ont conservé leur capacité de mobilité au sein du génome (Lander *et al.*, 2001).

La grande proportion de séquences *Alu* au sein du génome humain, alors que ces mêmes séquences sont absentes dans les espèces animales domestiques, en fait des marqueurs fiables des tissus humains. Des sondes spécifiques de ces séquences, appelées sondes *Alu*, sont ainsi utilisées comme témoins positifs de réaction en hybridation *in situ* sur tissus humains. Ces séquences sont également utiles dans de nombreux domaines, tels que l'étude des populations humaines (Hammer, 1994) ou le marquage de certains cancers (Shaban *et al.*, 2022).

Les séquences *Alu* ne sont cependant pas propres à l'espèce humaine. Si elles ont été très étudiées dans l'espèce humaine afin de les recenser et de comprendre leurs fonctions et impacts potentiels, plusieurs études ont en effet mis en évidence ces séquences dans le génome d'autres primates. Des séquences communes avec l'espèce humaine ont été découvertes, mais aussi d'autres séquences, spécifiques de l'espèce humaine ou d'une espèce de primate en particulier.

Dans cette thèse expérimentale, le travail consistait à rechercher la présence de séquences *Alu*, préalablement identifiées dans le génome humain, dans les tissus de diverses espèces animales primates et non primates à l'aide d'une sonde *Alu* commerciale. Le but était ainsi de déterminer si cette sonde commerciale pouvait être utilisée comme un outil diagnostique rapide permettant de distinguer les tissus humains des tissus de primates non humains.

Dans une première partie, nous nous intéresserons aux données bibliographiques concernant les séquences *Alu*. Dans une seconde partie, nous présenterons le travail de thèse ayant porté sur la recherche de cette séquence *Alu* dans les tissus de dix espèces de primates et huit espèces animales non primates.

Première partie : bibliographie

1. Origine des séquences *Alu*

A. Origine du nom

Le nom de séquence *Alu* provient d'*Arthrobacter luteus*, une bactérie Gram positif appartenant au genre *Arthrobacter*. Il s'agit du premier organisme à partir duquel ces séquences ont été isolées. L'endonucléase *AluI* tient également son nom du même organisme puisqu'il s'agit de l'enzyme de restriction ayant permis l'isolement de cette séquence (Jay et Wu, 1976).

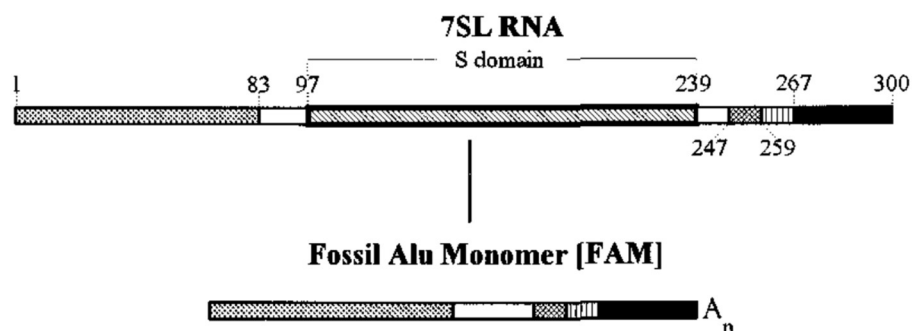
B. Un élément de la famille des SINE

Les séquences *Alu* sont des rétrotransposons appartenant à la famille des SINE (*Short Interspersed Nuclear Elements*), c'est-à-dire de petits éléments nucléaires courts et dispersés. Ce sont des séquences ADN non codantes de quelques centaines de paires de bases. Les éléments de cette famille ne sont pas autonomes pour leur réplication, à la différence des LINE (*Long Interspersed Nuclear Elements*) qui sont également des éléments répétés et dispersés de quelques milliers de paires de bases. SINE et LINE appartiennent à la famille des rétroéléments dépourvus de longues répétitions terminales, ou LTR (*Long Terminal Repeat*) (Lander *et al.*, 2001).

C. Théories actuelles sur l'émergence de la séquence *Alu*

D'après Ullu et Tschudi (1984), le monomère fossile *Alu* appelé *FAM* (*Fossil Alu Monomer*), serait le monomère *Alu* originel dérivé d'un gène de l'ARN 7SL puisque leurs séquences primaires sont similaires. L'ARN 7SL est un ARN cytoplasmique abondant faisant partie du complexe SRP appelé particule de reconnaissance du signal (*Signal Recognition Particle*). Ce complexe intervient lors de la synthèse de protéines dans le réticulum endoplasmique. Ces auteurs émettent l'hypothèse que la séquence serait apparue par délétion du « domaine S », une partie centrale de 141 dinucléotides, d'un gène de l'ARN 7SL, ainsi que par la fixation d'une terminaison poly-adénylate à l'extrémité 3' (Figure 1). Ils suggèrent que l'excision de *FAM* aurait été réalisée par un rétrovirus, et aurait pu se produire plusieurs fois au cours de l'évolution. Aujourd'hui, les séquences *FAM* sont peu abondantes dans le génome humain.

Figure 1 : Excision de la séquence *Fossil Alu Monomer* à partir de l'ARN 7SL



Modèle proposé pour l'apparition de la séquence *FAM* à partir de l'ARN 7SL, par excision du domaine S (« S domain ») (Mighell *et al.*, 1997).

Quentin (1994) a également supposé que la séquence *FAM* provienne de l'ARN 7SL, mais il émet une hypothèse différente quant à l'apparition de cette séquence. Il mit en évidence des changements conformationnels de l'ARN 7SL au cours de certaines phases du cycle. Ensuite, il remarqua que les zones de l'ARN qui restent protégées coïncident avec la frontière entre les deux domaines *Alu* et S. Il émet donc l'hypothèse que l'excision de *FAM* au sein de l'ARN 7SL serait le résultat d'une rupture accidentelle de l'ARN 7SL lors de changements conformationnels. Selon ce même auteur, cet évènement serait probablement survenu avant la radiation des mammifères, donc il y a au moins 66 millions d'années environ.

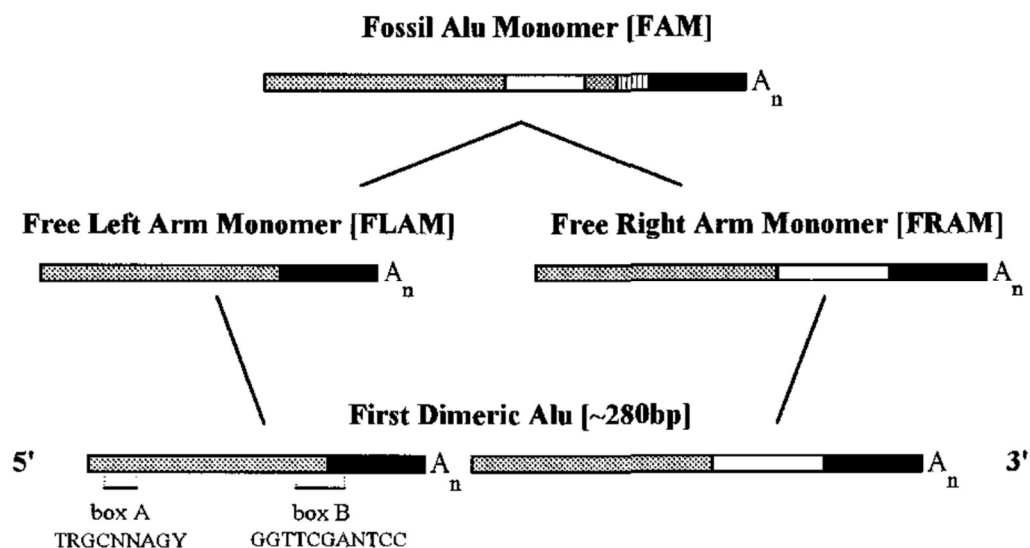
D. Structures primaire et secondaire des séquences *Alu*

a. Structure primaire des séquences *Alu*

Les séquences *Alu* sont longues d'environ 280 à 300 paires de bases. Ces séquences sont riches en dinucléotides C et G. En effet, elles comportent environ 7,1 millions de dinucléotides CpG, ce qui représente plus de 23 % de tous les CpG du génome humain (Luo *et al.*, 2014). Elles sont composées de deux monomères semblables mais non identiques appelés *FLAM* (*free left arm monomer*) d'environ 140 paires de bases, et *FRAM* (*free right arm monomer*) d'environ 170 paires de bases (Ziętkiewicz *et al.*, 1998).

Selon ces auteurs, les monomères *FLAM* et *FRAM* proviendraient du monomère *FAM* évoqué précédemment, par délétion de 42 et 11 paires de bases respectivement. La séquence *Alu* originale serait apparue à la suite d'une dimérisation de *FLAM* et *FRAM*, reliés par une séquence riche en dinucléotides A d'une trentaine de paires de bases (Figure 2). La plupart des séquences *Alu* possèdent également une terminaison polyadénylate de longueur variable en région 3', qui pourrait faciliter la transcription inverse des séquences transcrites par l'ARN polymérase III, et ainsi faciliter son expansion dans le génome (Quentin, 1992a).

Figure 2 : Apparition de la séquence *Alu* originale à partir de la séquence *Fossil Alu Monomer*



*Apparition de la première séquence *Alu* à partir des séquences *FLAM* et *FRAM*, provenant elles-mêmes de *FAM*. Les boîtes A et B sont visibles sur la séquence (box A et box B) (Mighell *et al.*, 1997).*

Le mécanisme de dimérisation de *FLAM* et *FRAM* est encore inconnu aujourd'hui. Cet évènement aurait eu lieu il y a environ 60 millions d'années, avant la divergence prosimienne

(Ziętkiewicz *et al.*, 1998). Ces mêmes auteurs suggèrent que l'apparition de la séquence *Alu* originale ne serait apparue qu'une seule fois au cours de l'évolution. A la suite de l'apparition de la première séquence, il y aurait eu une amplification massive dans le génome. Les séquences *FLAM* isolées sont retrouvées dans le génome humain (Quentin, 1988), ainsi que quelques séquences *FRAM* isolées en plus petite quantité (Quentin, 1992b).

Le monomère *FLAM* comporterait deux éléments promoteurs de l'ARN polymérase III appelés boîte A et boîte B, d'une longueur d'environ dix dinucléotides chacune. Ces deux éléments se trouveraient en position 10-25 et 70-90 respectivement (Fuhrman *et al.*, 1981). Les auteurs rapportent que la boîte B serait impliquée dans le mécanisme de rétrotransposition de ces séquences.

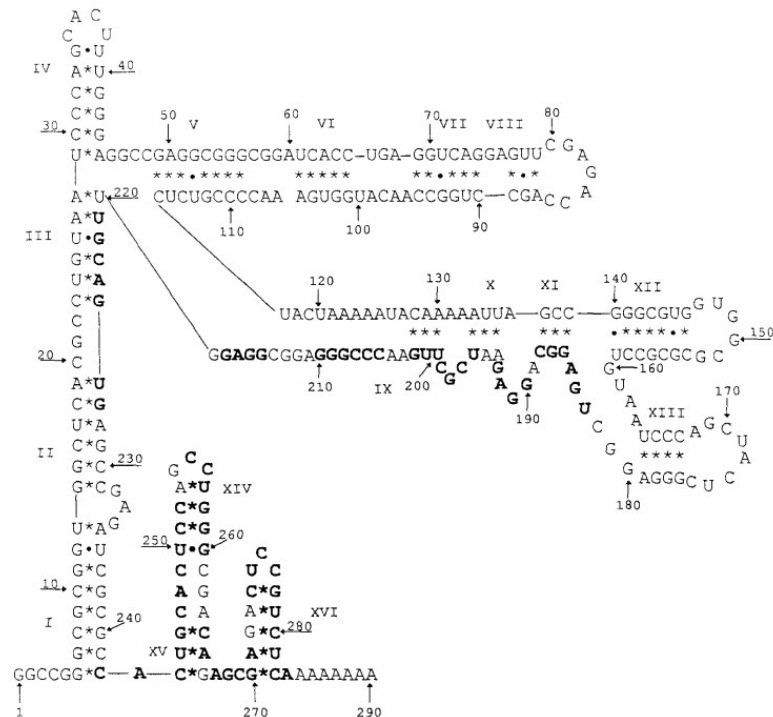
Des études postérieures ont prouvé l'existence de deux versions de *FLAM* : *FLAM-A* et *FLAM-C*. *FLAM-A*, aussi appelé *FLA*, serait apparue à partir de la séquence *FAM*, alors que *FLAM-C* serait un descendant de *FLA*. La séquence *Alu* serait le produit de la fusion entre *FRAM* et *FLAM-C* (Quentin, 1992b).

Au final, ce sont les 98 premiers et 60 derniers dinucléotides de l'ARN 7SL, regroupés sous le nom de domaine *Alu*, qui sont à l'origine des séquences *Alu*. Ce domaine, associé à deux protéines, est associé à la fonction d'arrêt de l'élongation de la SRP (Ullu et Tschudi, 1984).

b. Structure secondaire des séquences *Alu*

Jurka et Milosavljevic (1991) ont étudié la structure secondaire des ARN *Alu* par analyse phylogénétique. Selon eux, cet ARN présente une structure cruciforme composée de seize hélices (Figure 3). Les hélices IV à XVI peuvent être renforcées ou affaiblies en fonction des différents éléments *Alu* et des variations nucléotidiques locales spécifiques. Malgré ces quelques variations, les auteurs ont rapporté une conservation de structure secondaire entre les différents éléments, permise grâce à la présence de mutations compensatoires. En particulier, lorsqu'une mutation se produit, une autre survient fréquemment, permettant ainsi le maintien des liaisons hydrogènes essentielles à la stabilisation de la structure.

Figure 3 : Représentation de la structure secondaire d'un ARN *Alu*



Les différentes hélices sont numérotées à l'aide de chiffres romains (Jurka et Milosavljevic, 1991).

c. Conservation des structures primaire et secondaire

Les séquences *Alu* sont classées en différentes sous-familles, dont la classification sera explicitée ultérieurement. Malgré des divergences de séquence entre ces différentes sous-familles, toutes les séquences *Alu* conservent une homologie de séquence substantielle. En effet, les comparaisons entre les membres de la famille *Alu* démontrent une identité de séquence d'environ 70 % en moyenne (Shen *et al.*, 1991). D'après les articles de Ziętkiewicz *et al.* (1998) et de Rowold et Herrera (2000), cette homologie de séquence pourrait être le témoin de l'existence d'une pression de sélection s'exerçant sur la structure primaire des séquences *Alu*.

Comme explicité précédemment, la structure secondaire de l'ARN *Alu* est également très conservée parmi les différentes sous-familles (Jurka et Milosavljevic, 1991). En effet, la présence de mutations compensatoires permet la conservation et la stabilisation de la structure secondaire de ces ARN. De plus, d'après (Labuda et Striker, 1989), les deux sous-unités des ARN *Alu* conservent les motifs de structure secondaire de l'ARN 7SL.

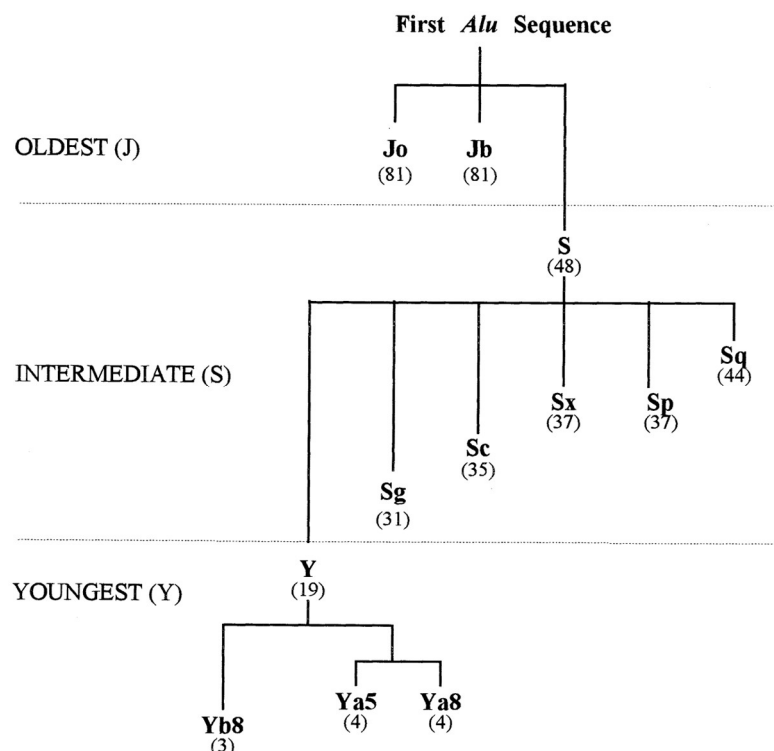
2. Classification des séquences *Alu*

A. Les différentes sous-familles

Willard *et al.* (1987) ainsi que Britten *et al.* (1988) ont démontré que les séquences *Alu* pouvaient être classées en différentes sous-familles en fonction de leur degré de divergence. Jurka et Milosavljevic (1991) proposent une classification en fonction de leurs résultats ainsi que des études menées antérieurement. Ils proposent un classement avec deux sous-familles principales : la sous-famille *AluJ* qui regroupe les éléments *Alu* les plus anciens, et la famille *AluS*, qui regroupe les éléments *Alu* d'âge intermédiaire. Ensuite, ils divisent la sous-famille *AluJ* en deux branches principales appelées *AluJo* et *AluJb*, et la sous-famille *AluS* en cinq sous-branches appelées *AluSx*, *AluSq*, *AluSp*, *AluSc* et *AluSb*. Une nouvelle branche de cette sous-famille *AluS*, appelée *AluSg*, a été découverte puis ajoutée à cette classification (Jurka, 1995).

En 1996, Batzer *et al.* proposent une nouvelle nomenclature à partir des nomenclatures existantes. Selon eux, la sous-famille *AluSb* représente en réalité une troisième sous-famille des séquences *Alu*, et ils la nomment *AluY* (Figure 4). Ensuite, ils divisent cette nouvelle sous-famille en plusieurs lignées, et les nomment selon une nomenclature bien définie. Ainsi, pour le génome humain, une lignée commence par la lettre Y en majuscule en référence à la sous-famille. Ensuite une lettre minuscule est ajoutée en fonction de l'ordre de découverte de la lignée en prenant l'ordre alphabétique. Enfin, un chiffre arabe est ajouté, et indique le nombre de modifications diagnostiques par rapport à la séquence consensus de la sous-famille *AluY*. Pour les primates non humains, la nomenclature est la même mais les lettres du genre et de l'espèce du primate en question sont ajoutées en italique après le chiffre arabe. Dans le génome humain, trois sous-branches principales de la sous-famille *AluY* ont été mises en évidence : Ya5, Ya8 et Yb8. La plupart des séquences *Alu* spécifiques à l'Homme appartiennent aux sous-familles *AluYa5* et *AluYb8*.

Figure 4 : Classification des éléments *Alu* en différentes sous-familles



Synthèse de Mighell *et al.* (1997) à partir des travaux précédents, et notamment de l'étude de Kapitonov et Jurkal (1996) pour les âges moyens des différentes sous-familles (indiqués en millions d'années entre parenthèses).

B. Composition des différentes sous-familles

Les éléments *Alu* les plus récemment intégrés dans le génome sont plutôt homogènes entre eux et sont presque des copies exactes les uns des autres (Batzer *et al.*, 1990), tandis que les séquences les plus anciennes divergent d'au moins 20 % environ (Shen *et al.*, 1991). Les sous-familles les plus anciennes ont une composition réduite en CpG. Il est possible que les gènes *Alu* soient soumis à une pression de sélection qui mènent les séquences à contenir de plus en plus de doublets CpG par séquence. De plus, les séquences présentent davantage de dégénérescence au niveau des séquences adjacentes et des queues poly-adénylates. Ces caractéristiques peuvent s'expliquer par des mutations accumulées au cours du temps depuis leur rétroposition (Britten *et al.*, 1988 ; Jurka et Milosavljevic, 1991).

Pour chaque sous-famille, une séquence consensus a été déterminée. Il s'agit de la séquence la plus représentative, soit celle qui partage le plus de similarités entre tous les membres d'une même sous-famille (Kapitonov et Jurkal, 1996).

Les séquences spécifiques de l'Homme appartenant aux deux sous-familles *AluYa5* et *AluYb8* auraient évolué depuis la divergence homme-chimpanzé, et elles diffèrent grandement aujourd'hui du gène source ancestral (Waterson *et al.*, 2005).

C. Activité des différentes sous-familles au sein du génome humain

Depuis leur apparition, les séquences *Alu* ont réussi à s'amplifier et à se diversifier dans le génome des primates. Leur taux d'amplification n'a pas été constant depuis leur apparition mais il a fluctué dans le temps. Cette amplification aurait commencé il y a environ 65 millions d'années avec l'amplification des familles *AluJ* et *AluS* (Shen *et al.*, 1991). La famille *AluJ* aurait été active il y a 65 à 40 millions d'années, la famille *AluS* aurait été intensément active il y a 35 à 40 millions d'années, et la famille *AluY* aurait été active de 30 millions d'années avant notre ère jusqu'à aujourd'hui (Shen *et al.*, 1991 ; Kapitonov et Jurkal, 1996 ; Bailey *et al.*, 2003). Plus précisément, Liu *et al.* (2009) ont estimé que la période d'amplification la plus intense pour le génome humain est située entre 35 et 55 millions d'années avant notre ère, soit entre la divergence des prosimiens et l'émergence des anthropoïdes. La famille *AluS* serait responsable de la grande majorité de l'enrichissement du génome humain en séquences *Alu*, puisque plus d'un tiers de tous les événements de rétroposition se serait produit pendant cette période (Bailey *et al.*, 2003). Lors de ce pic d'amplification, les auteurs ont estimé qu'une nouvelle insertion *Alu* survenait à chaque naissance (Cordaux et Batzer, 2009).

De nos jours, l'activité des séquences *Alu* a considérablement diminué, et leur rétroposition est devenu un événement plus rare, estimé à une nouvelle insertion pour 21 naissances, et limité à quelques éléments actifs appartenant à la jeune famille *AluY* (Xing *et al.*, 2009). Rowold et Herrera (2000) suggèrent que cette baisse d'activité était attendue car une charge génétique trop importante pourrait amener une espèce à ne plus être génétiquement viable. Le génome humain aurait pu atteindre cette limite, se trouvant actuellement dans un équilibre dynamique entre les nouvelles insertions et les séquences perdues, par suppression ou adoption d'une fonction régulatrice.

3. Localisation et abondance des séquences *Alu* dans le génome humain

A. Proportion et distribution des séquences *Alu*

a. Proportion au sein du génome humain

Les séquences *Alu* sont les éléments répétitifs les plus abondants du génome humain. D'après Athanasiadis *et al.* (2007), il existerait plus de 1 100 000 copies de la séquence, ce qui représenterait plus de 10 % du génome. En moyenne, on retrouve une séquence toutes les 3 000 à 6 000 paires de bases, même si la distribution n'est pas uniforme en réalité.

b. Distribution dans le génome humain

Les anciennes séquences *Alu* semblent se situer préférentiellement dans les régions riches en dinucléotides G et C, ainsi que dans les gènes (Korenberg et Rykowski, 1988). En effet, des répétitions *Alu* ont été trouvées dans 5 % des 1616 ADN complémentaires (ADNc) (Yulug *et al.*, 1995). Dans ces 5 %, 82 % ont été retrouvés dans l'UTR 3', 14% dans l'UTR 5', et très rarement dans la région codante. En effet, comme pour les autres SINE, les séquences *Alu* situées dans les gènes se localisent principalement dans les régions non codantes de gènes à copie unique. Il est rare de retrouver des séquences *Alu* dans les exons codants pour des protéines, car il existe de nombreux codons « stop » dans le sens de lecture et l'antisens, ce qui aboutirait à des protéines tronquées. Étant donné que les anciennes séquences *Alu* sont également retrouvées dans les gènes, les éléments *Alu* sont extrêmement répandus dans les molécules d'ARN. Environ 5 à 10 % de tous les ARNm possèderaient un ou plusieurs éléments *Alu* dans leur extrémité 3' (Deininger, 2011). En moyenne, une molécule de pré-ARNm possèderait environ 16 séquences *Alu* (DeCerbo et Carmichael, 2005).

Contrairement aux anciennes séquences, les nouvelles séquences *Alu* sont principalement localisées dans les régions pauvres en gènes, et riches en dinucléotides A et T, proches d'éléments L1 appartenant à la famille des LINE-1 (Waterson *et al.*, 2005).

B. Mécanisme d'amplification des séquences *Alu*

a. Modèles hypothétiques de l'expansion des éléments *Alu*

Plusieurs modèles d'expansion des séquences *Alu* ont été proposés. Nous allons détailler quatre mécanismes d'expansion, dont le plus communément accepté est celui des séquences maîtresses.

- *Modèle de séquences maîtresses*

L'existence des différentes sous-familles suggère que les séquences *Alu* ont été amplifiées à partir d'un petit sous-ensemble de séquences que les auteurs ont appelé gènes sources (Britten *et al.*, 1988), séquences maîtresses (Labuda et Striker, 1989 ; Quentin, 1994) ou gènes maîtres (Batzner *et al.*, 1990). Ainsi, les familles *Alu* seraient composées d'un grand nombre de copies incapables de se transposer, mais aussi de quelques éléments dotés de la capacité de transposition. Il est supposé que chaque sous-famille soit le produit de l'amplification d'un seul gène maître, et que la séquence consensus de chaque famille serait la meilleure estimation de ce gène maître (Quentin, 1994). Les différentes séquences maîtresses seraient codées par des *loci* distincts (Quentin, 1988). Elles se seraient succédé au cours du temps, à l'origine des différentes sous-familles d'*Alu* (Quentin, 1994). L'analyse de ces séquences a montré que les séquences maîtresses possédaient toutes un promoteur de l'ARN Pol III et une structure secondaire similaire.

Quentin (1994) a défini deux périodes dans l'évolution des séquences maîtresses. La première période correspond à l'émergence de FAM à partir de l'ARN 7SL. Cette séquence aurait subi de multiples modifications à l'origine d'un remodelage de la structure secondaire, notamment de l'hélice

V, et des mutations ayant permis l'amélioration de la boîte B et la stabilisation de la séquence *Alu*. Cette première phase aurait eu lieu avant la divergence entre les primates et les rongeurs, et a pris fin indépendamment dans chaque lignée avec l'apparition des premières séquences maîtresses. La deuxième période correspond à la succession de périodes d'amplification et de stabilisation des séquences maîtresses. Celles-ci auraient évolué par des mutations par insertion, délétion et substitution ne concernant pas le promoteur de l'ARN pol III et ne modifiant pas la structure secondaire de l'ARN. Comme les séquences *Alu* sont riches en éléments CpG, les mutations ponctuelles se produisent fréquemment. Le remplacement des anciennes séquences maîtresses par de nouvelles pourrait être le résultat d'une compétition entre les différentes séquences vis-à-vis de la même machinerie de rétroposition (Schmid et Maraia, 1992). Les premières séquences maîtresses pourraient s'être éteintes dans toutes les lignées exceptées celles des primates et des rongeurs, où elles se sont largement amplifiées en deux familles différentes mais apparentées, *Alu* et *B1* (Quentin, 1994).

Bien que ce modèle soit le plus communément accepté, il est remis en cause par l'observation de l'activité transpositionnelle des éléments de la sous-famille *AluYb*. La lignée *AluYb* fait partie des lignées *Alu* les plus actives du génome humain (Cordaux *et al.*, 2004). Cette sous-famille serait apparue postérieurement à la divergence entre les singes de l'Ancien et du Nouveau Monde, il y a environ 18 à 25 millions d'années. D'autres résultats ont montré que cette lignée avait été peu voire pas du tout active durant les 10 millions d'années suivant son apparition, mais qu'elle a pourtant conservé son potentiel de rétrotransposition au cours de cette période prolongée. Pour expliquer cette capacité non expliquée par le modèle de séquences maîtresses, d'autres modèles de rétrotransposition ont été proposés. D'autres modèles ont également été suggérés.

- *Autres modèles d'expansion des éléments Alu*

Le modèle du transposon suggère que tous les éléments *Alu* ont une capacité similaire à générer de nouvelles copies, y compris les éléments mutés (Cordaux *et al.*, 2004).

Le modèle intermédiaire est un modèle qui, comme son nom l'indique, est intermédiaire entre le modèle de séquences maîtresses et le modèle du transposon. Il suggère que de nombreux éléments *Alu* seraient très actifs pour l'expansion (Deininger *et al.*, 1992 ; Cordaux *et al.*, 2004).

Le modèle furtif suit le modèle des séquences maîtresses, mais permet d'expliquer l'amplification aberrante des éléments *AluYb*. Certains éléments hyperactifs muteraient et perdraient leur capacité de rétrotransposition relativement rapidement par sélection. Les autres éléments moins actifs conserveraient leur activité de rétrotransposition pendant de longues périodes puis produiraient des éléments hyperactifs à courte durée de vie. Ce modèle expliquerait l'expansion différente des éléments de la sous-famille *AluYb* (Han *et al.*, 2005).

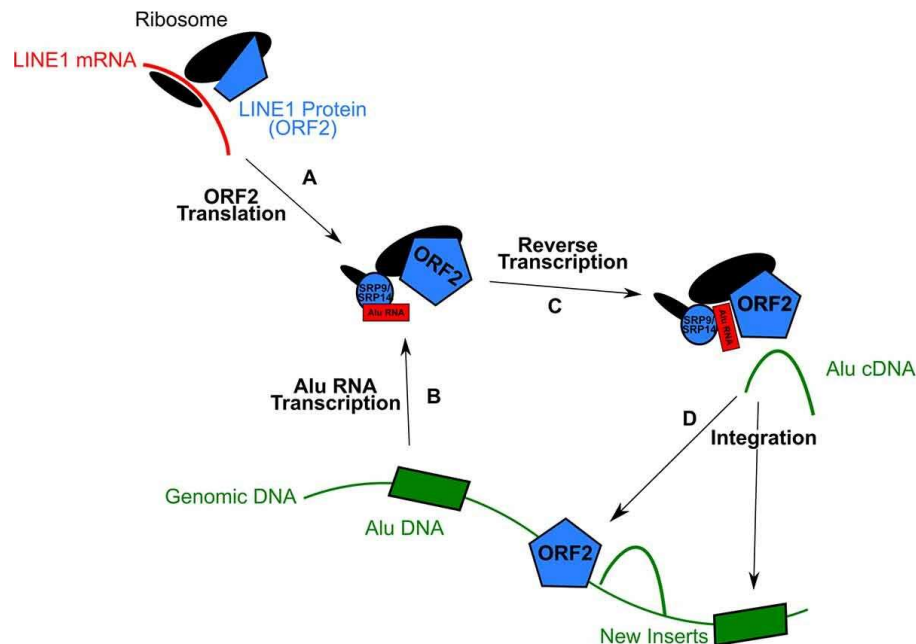
b. Modèle de rétrotransposition

Les éléments de la famille des SINE sont des éléments qui se propagent par le processus de rétroposition. Ces éléments ne sont pas autonomes pour leur réplication, et, par conséquent, ils utilisent la machinerie de transposition des éléments LINE-1, une famille appartenant aux LINE. Dans ce modèle, les séquences sont transcrites puis réinsérées sous forme d'ADN à divers endroits du génome (Smit *et al.*, 1995). D'après une étude, les éléments de la famille *Alu* ne font pas défaut à ce principe (Dewannieux *et al.*, 2003).

La traduction des ARNm LINE1 produirait deux protéines nécessaires à la rétrotransposition des séquences *Alu*, appelées ORF1p et ORF2p. La protéine ORF2p confère son activité de transcriptase inverse et d'endonucléase, et le rôle de la protéine ORF1p n'est pas encore clairement élucidé (Dewannieux *et al.*, 2003). La transcription des séquences *Alu* serait réalisable grâce à un promoteur pour l'ARN polymérase III dans leur moitié gauche, permettant la formation d'un intermédiaire ARN *Alu* (Fuhrman *et al.*, 1981). Ensuite, l'hétérodimère SRP9/14 se lierait à cet ARN *Alu* afin de l'associer à la petite unité ribosomique 40S. Le positionnement de l'ARN *Alu* par rapport

à ORF2p permettrait sa transcription inverse en commençant par la queue poly-A, générant ainsi un ADNc *Alu*. Enfin, l'activité endonucléase d'ORF2p permettrait à ce nouvel ADNc de s'insérer sur un nouveau site génomique en utilisant le processus de transcription inverse amorcée par la cible, ou TPRT (*Target-Primed Reverse Transcriptase*). Pour le mécanisme de l'insertion, ORF2p utilise son activité endonucléase sur un premier brin dans la séquence consensus 5'-TTTT/AA-3'. Ensuite, un mécanisme inconnu produit une deuxième coupure du brin complémentaire, puis la synthèse du deuxième brin est amorcée en utilisant le premier brin de l'élément *Alu* comme matrice (Figure 5). L'élément *Alu* sera entouré par de courtes répétitions directes de 7 à 20 paires de bases de part et d'autre de l'élément, appelées duplications de sites cibles (Deininger, 2011).

Figure 5 : Rétrotransposition de l'*Alu* médiée par LINE1



Sur ce schéma, les ribosomes sont représentés en noir, l'ADN en vert, l'ARN en rouge, et les protéines en bleu. L'étape (a) représente la traduction de l'ARNm LINE1, à l'origine de la protéine ORF2. L'étape (b) représente la transcription de l'élément *Alu*, à l'origine d'un ARN *Alu* lié à l'hétérodimère SRP9/14 pour le localiser dans la petite sous-unité ribosomique 40S. L'étape (c) représente la transcription inverse de l'ARN *Alu* par ORF2, à l'origine d'un ADNc *Alu*. L'étape (d) représente l'insertion de l'ADNc *Alu* par ORF2 dans l'ADN génomique (Gussakovsky et McKenna (2021), d'après la description de Dewannieux et al. (2003)).

Les séquences *Alu* ne semblent pas contenir de signal pour la terminaison de la transcription, ce qui suggère que les transcrits sont susceptibles de s'étendre sur une distance variable dans la séquence adjacente (Fuhrman et al., 1981). L'ARN polymérase III arrête son élongation lorsqu'elle rencontre quatre dinucléotides T successifs dans l'ADN adjacent à la séquence (Dewannieux et Heidmann, 2005).

Au cours de la transcription inverse, la queue poly-A est lue comme faisant partie intégrante de la séquence, elle n'est pas ajoutée post-transcriptionnellement. Cela a pour conséquence l'apparition de modifications dans la longueur de la queue poly-A terminale des éléments *Alu* (Dewannieux et al., 2003). Par conséquent, chaque ARN *Alu* généré par l'ARN polymérase III est unique en termes de mutations accumulées, et de longueur et hétérogénéité de la séquence accumulée dans la queue poly-A (Deininger, 2011).

L'analyse des transcrits *Alu* dans des cellules cultivées montre que la majorité des transcrits, environ 66 %, provenaient de sous-familles anciennes (Sinnott et al., 1992) et que moins de 1 % des transcrits provenaient de la sous-famille *AluYa5* (Shaikh et al., 1997). Etant donné que la

majorité des séquences amplifiées appartiennent à la grande famille *AluY*, cela signifie que de nombreux ARN sont transcrits mais non rétrotransposés. Le manque d'activité rétrotranspositionnelle des éléments appartenant aux familles les plus anciennes ne provient pas d'un manque de transcription.

c. Facteurs régulateurs de la transcription

De nombreux facteurs de régulation semblent influencer l'activité de rétrotransposition des éléments *Alu*, à l'origine d'une différence d'activité marquée entre les différents éléments. Plusieurs facteurs ont été étudiés et identifiés.

Premièrement, l'accumulation de mutations aléatoires au sein de l'élément *Alu* peuvent altérer la rétrotransposition. Par exemple, des mutations aléatoires apparaissant dans les boîtes A ou B, au sein du monomère droit (Aleman, 2000) ou un écart de séquence de plus de 10 % de la séquence consensus (Bennett *et al.*, 2008) peuvent diminuer l'activité de rétrotransposition. D'autres mutations apparaissant dans la queue poly-A peuvent entraîner des répercussions majeures dans la rétrotransposition, comme un raccourcissement important de cette partie, ou l'apparition de mutations C et G en particulier (Comeaux *et al.*, 2009). En effet, les études actuelles suggèrent que la queue poly-A est nécessaire au mécanisme de rétrotransposition. Les protéines LINE reconnaîtraient leur substrat à travers cette séquence, et la transcription inverse débiterait par cette séquence poly-A. La reconnaissance de cette séquence pourrait être directe, ou bien se faire par l'intermédiaire de protéines stabilisatrices liées à cette région (Dewannieux *et al.*, 2003). De plus, d'autres mutations aléatoires apparaissant à divers endroits de l'élément pourraient impacter la stabilité des ARN *Alu* et diminuer la rétrotransposition.

Deuxièmement, une capacité de fixation diminuée à la sous-unité SRP9/14 peut diminuer l'activité de rétrotransposition (Bennett *et al.*, 2008). En effet, la sous-unité SRP9/14 assurerait la stabilisation du transcrit, et augmenterait l'efficacité de la rétrotransposition en adressant l'ARN *Alu* aux ribosomes, à proximité des protéines L1 naissantes (Dewannieux *et al.*, 2003).

Troisièmement, le site d'insertion de l'élément *Alu* exerce également une grande influence sur sa rétrotransposition. Tout d'abord la distance entre la queue poly-A et le terminateur pol III situé en aval a une influence sur le niveau d'activité rétrotranscriptionnelle. Il semblerait que plus cette distance est réduite, plus l'activité est grande (Comeaux *et al.*, 2009). De plus, étant donné que la transcription des ARN Pol II et III partagent de nombreux facteurs régulateurs, l'insertion fortuite d'un élément *Alu* dans une région codante active peut augmenter la production de transcrits pol III *Alu* (Chesnokov et Schmid, 1996). Ensuite, le degré de méthylation de l'ADN impacte l'activité de rétrotransposition (Schmid et Maraia, 1992). Enfin, la présence de *loci* éloignés peut influencer l'amplification des éléments *Alu*. Par exemple, certains gènes peuvent coder pour des facteurs pouvant impacter la stabilité des ARN *Alu* ou des protéines associées à la machinerie transcriptionnelle, comme la protéine La (Rowold et Herrera, 2000).

4. Conséquences de la présence de séquences *Alu*

A. Conséquences fonctionnelles

Longtemps considérées comme de l'ADN inutile, plusieurs études ont démontré que les séquences *Alu* pouvaient avoir un rôle. Toutes les séquences *Alu* ne possèdent cependant pas les mêmes fonctions, et certaines ne semblent d'ailleurs pas en avoir, à moins que ces rôles n'aient pas encore été découverts. Les exemples ci-dessous ne sont pas exhaustifs mais ils permettent de saisir l'importance fonctionnelle de ces séquences.

Tout d'abord, les séquences *Alu* peuvent jouer un rôle activateur ou inhibiteur dans la transcription des gènes. L'étude de Norris *et al.* (1995) a démontré que lorsqu'elles étaient insérées dans un gène, certaines séquences *Alu* jouaient un rôle amplificateur de transcription sous dépendance des récepteurs œstrogènes. Ces auteurs ont suggéré qu'il est fortement possible que d'autres sous-classes de séquences *Alu* puissent jouer ce rôle, avec éventuellement des dépendances à d'autres molécules. L'étude de Brini *et al.* (1993) a démontré que certaines séquences *Alu* comportaient les éléments activateurs et inhibiteurs du promoteur du gène de la chaîne gamma du récepteur de haute affinité pour l'IgE.

Les séquences *Alu* peuvent également agir comme source de variabilité protéique. L'étude de Makałowski *et al.* (1994) a mis en évidence des insertions *Alu* dans des introns, à l'origine d'une perturbation de l'épissage alternatif. Cela a abouti à la création de nouveaux ARNm. L'étude de Sorek *et al.* (2002) a montré que tous les exons dérivés de séquences *Alu* étaient alternativement épissés. D'un point de vue évolutif, ces séquences augmentent la variabilité du transcriptome.

Les séquences *Alu* peuvent également jouer le rôle de composant structural des gènes. Par exemple, une étude a prouvé que le domaine cytoplasmique de l'un des quatre récepteurs aux interférons présente une grande similarité de séquence avec les séquences *Alu*. Les auteurs ont suggéré que l'apparition d'une cassette *Alu* au sein de cette protéine aurait pu avoir lieu grâce à l'épissage alternatif (Mullersman et Pfeffer, 1995).

Enfin, comme expliqué précédemment, les séquences *Alu* présentent une forte homologie de séquences primaire et secondaire avec la SRP, en particulier les polypeptides SRP9 et SRP14 qui correspondent aux régions 3' et 5' de l'ARN 7SL. Par cette forte conservation de séquences, plusieurs auteurs ont suggéré que la structure secondaire de l'ARN *Alu* pouvait révéler un rôle fonctionnel de cet ARN. Des études ont démontré que, chez les rongeurs et les humains, des ARN *Alu* cytoplasmiques se liaient à une protéine homologue à la SRP14 et à l'hétérodimère SRP9/14 respectivement. Cela suggère que ces ARN *Alu* pourraient jouer un rôle dans la production des protéines, de manière homologue à la SRP (Chang *et al.*, 1994 ; Bovia *et al.*, 1995). Par ailleurs, cette structure secondaire pourrait être un facteur important dans l'amplification par rétroposition de ces éléments (Labuda et Striker, 1989).

B. Conséquences délétères des insertions *Alu* aléatoires

La rétrotransposition des séquences *Alu* est à l'origine d'insertion de nouvelles séquences à des endroits aléatoires du génome. Certaines insertions peuvent perturber l'activité des gènes environnants et entraîner des conséquences délétères pour l'organisme. L'étude de Szmulewicz *et al.* (1998) décrit de nombreuses possibilités de perturbations en fonction du site d'insertion.

L'insertion de séquences *Alu* dans des séquences non codantes peut altérer les gènes en perturbant les séquences régulatrices. Par exemple, l'insertion de séquences au sein d'un promoteur peut engendrer une incapacité de reconnaissance du promoteur par l'ARN polymérase ou par d'autres facteurs. L'insertion d'une séquence entre deux promoteurs peut altérer leurs fonctions en modifiant la distance entre ces deux régions. L'insertion au sein de séquences amplificatrices d'un promoteur peut altérer leur fonction. Les séquences *Alu* peuvent également modifier directement les séquences codantes des gènes en s'insérant dans les introns ou dans les exons. L'insertion au sein d'un intron peut n'avoir aucune conséquence si l'élément est excisé, ou bien cela peut engendrer une modification des sites d'excision avec maintien de l'élément *Alu*. L'insertion de séquences au sein d'un exon peut entraîner la production de nouvelles protéines ayant des fonctions similaires ou différentes, engendrer des réarrangements génétiques, ou bien un site d'épissage non conventionnel est activé, aboutissant à l'excision de l'élément *Alu* de l'ARN. Enfin, l'insertion de séquences au sein d'un domaine sous forme d'hétérochromatine peut entraîner une accumulation de séquences *Alu* dans la région.

Tous ces mécanismes peuvent produire des altérations génétiques, des perturbations fonctionnelles, et être à l'origine de maladies génétiques (Tableau 1). Environ 0,1 % des troubles génétiques humains seraient liés à des séquences *Alu* (Kim *et al.*, 2016).

Tableau 1 : Exemples de maladies génétiques liées à des séquences *Alu*

Gène	Maladie humaine	Famille <i>Alu</i>	Type de séquence
<i>COL4A3, COLA5</i>	Syndrome d'Alport	Indéterminée, <i>AluY</i>	Intron
<i>ATM</i>	Ataxie-télangiectasie	<i>AluSg</i>	Intron
<i>CNTN1, MED13</i>	Cancer du cerveau	Plusieurs	ARNm mature et intron
<i>BC200, BRCA1, MDM2, MDM1</i>	Cancer du sein	Plusieurs	Intron et ARNm mature
<i>PMM2</i>	Troubles congénitaux du syndrome de glycosylation – la	<i>AluJb</i>	Intron
<i>DMD</i>	Dystrophie musculaire de type Duchenne	Indéterminée	Intron
<i>PDZK1</i>	Cancer gastrique	<i>AluSg</i>	ARNm mature
<i>HIPK3</i>	Gliomes	<i>AluSz, AluSq2, AluSg</i>	Intron, ARNm mature
<i>AVOINE</i>	Atrophie giratoire	<i>AluSx4</i>	Intron
<i>F8</i>	Hémophilie A	<i>AluY</i>	Intron
<i>PSMB2</i>	Carcinome hépatocellulaire	<i>AluSc, AluJb</i>	ARNm mature
<i>NOSIP</i>	Maladie de Hirschsprung	<i>AluSp</i>	ARNm mature
<i>PTS</i>	Hyperphénylalaninémie	<i>AluSq2</i>	Intron
<i>LRRC56</i>	Altération de la clairance mucociliaire	<i>AluY</i>	ARNm mature
<i>AKR1A1</i>	Cancer du larynx	<i>AluJo</i>	ARNm mature
<i>XIAP</i>	Leucémies	Indéterminée	ARNm mature
<i>GLI1</i>	Myélome multiple	<i>AluY, AluSx</i>	ARNm mature
<i>TOMM40</i>	Maladies neurodégénératives	<i>AluY</i>	Intron
<i>SMN</i>	Amyotrophie spinale	<i>AluY, AluSx1, AluSq</i>	Intron
<i>ATP7B</i>	Maladie de Wilson	<i>AluSq</i>	Intron

Le tableau ci-dessus est une liste non exhaustive de maladies génétiques humaines liées à l'insertion d'une séquence *Alu*. La famille *Alu* impliquée ainsi que son site d'insertion sont mentionnés (Gussakovsky et McKenna, 2021).

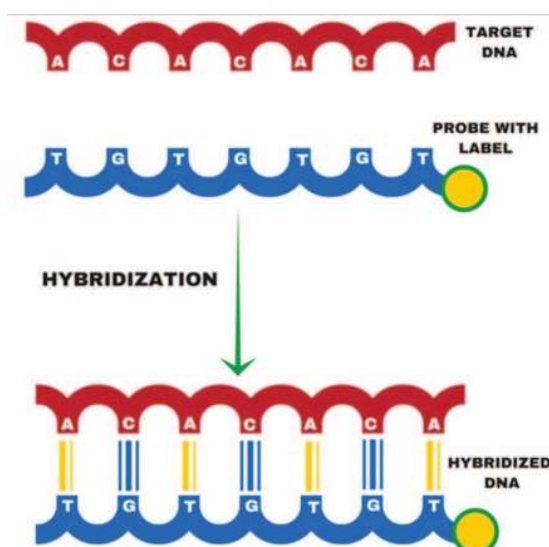
C. Marqueurs génétiques en hybridation *in situ*

a. Principe de l'hybridation *in situ*

Très utilisée dans le domaine de l'anatomopathologie, l'hybridation *in situ* (*In Situ Hybridization*, ou ISH) est une technique de détection de séquences nucléotidiques au sein de tissus. Son principe repose sur la complémentarité de bases entre une sonde nucléotidique et une séquence ARN ou ADN spécifique (Figure 6). Il existe plusieurs variantes de cette technique engendrant des différences de marquage, comme l'hybridation *in situ* par fluorescence (*Fluorescence In Situ Hybridization*, ou FISH) ou par l'hybridation *in situ* chromogénique (*Chromogenic In Situ Hybridization*, ou CISH). Nous allons détailler ici la technique CISH.

Les sondes utilisées peuvent être des sondes ADN simple- ou double-brin, ARN, ou oligonucléotidiques. Les sondes sont marquées avec des molécules comme la biotine ou la digoxigénine. Ensuite a lieu l'étape d'hybridation de la sonde, incluant l'étape de dénaturation de l'ADN lorsque l'ADN est la cible, ou que la sonde utilisée est une sonde ADN double-brin. Cette dénaturation peut se faire de différentes manières : une température élevée, un changement de concentration des sels en solution, des valeurs de pH extrêmes associées à la présence de certains composés organiques comme le diméthylsulfoxyde et le formamide (Ehtisham et al., 2016). Un rétablissement des conditions initiales permet la renaturation de l'ADN (Wang et al., 2014). Après l'hybridation, un anticorps conjugué à une enzyme, comme la phosphatase alcaline ou la peroxydase de raifort (*horseradish peroxydase*), reconnaît et marque la sonde hybridée. Enfin, un substrat chromogénique est ajouté, comme la diaminobenzidine, ce qui va engendrer une réaction enzymatique et l'apparition d'un produit de réaction coloré, qui traduit ainsi la présence de la sonde hybridée et donc de la séquence recherchée. La lecture des résultats est facilement réalisable par microscopie optique.

Figure 6 : Technique de l'hybridation *in situ*



Dans ce schéma, l'ADN recherché est représenté en rouge (*target DNA*), et la sonde avec marqueur est représentée en bleu (« *probe with label* »). La première étape du schéma représente l'ADN déjà dénaturé avec la sonde non hybridée. La deuxième étape du schéma représente l'hybridation entre l'ADN et la sonde (« *hybridized DNA* ») (Ehtisham et al., 2016).

Les principaux avantages de cette technique sont la visualisation de la distribution du marquage au sein des cellules (Coulton et De Belleruche, 1992), l'absence de dégradation des lames après marquage permettant une conservation à long terme (Halling et Wendel, 2009), la simplicité de traitement, la spécificité de la technique, et l'interprétation simple des résultats (Veselinová et al., 2021). Afin d'optimiser le marquage, les cellules doivent être conservées de

manière optimale afin que les séquences nucléotidiques soient intactes. Les tissus congelés ou fixés au formol pendant plusieurs années peuvent être utilisés pour l'ISH (Jensen, 2014).

b. Limites de l'hybridation in situ chromogénique

Il existe certaines limites à la technique CISH. Tout d'abord, cette technique n'est pas totalement sensible, et les résultats peuvent parfois être difficiles à lire à cause du faible signal, différemment des résultats obtenus avec la technique FISH. En particulier, des expériences de détection de faibles copies de l'ADN recherché au sein d'une cellule peuvent manquer de sensibilité. Ensuite, différemment de la technique FISH, les sondes ne peuvent cibler qu'une seule séquence par expérience. Cette limite est relative et plus ou moins contraignante en fonction de l'expérience menée (Halling et Wendel, 2009).

c. Séquences Alu comme marqueurs de tissus humains en hybridation in situ

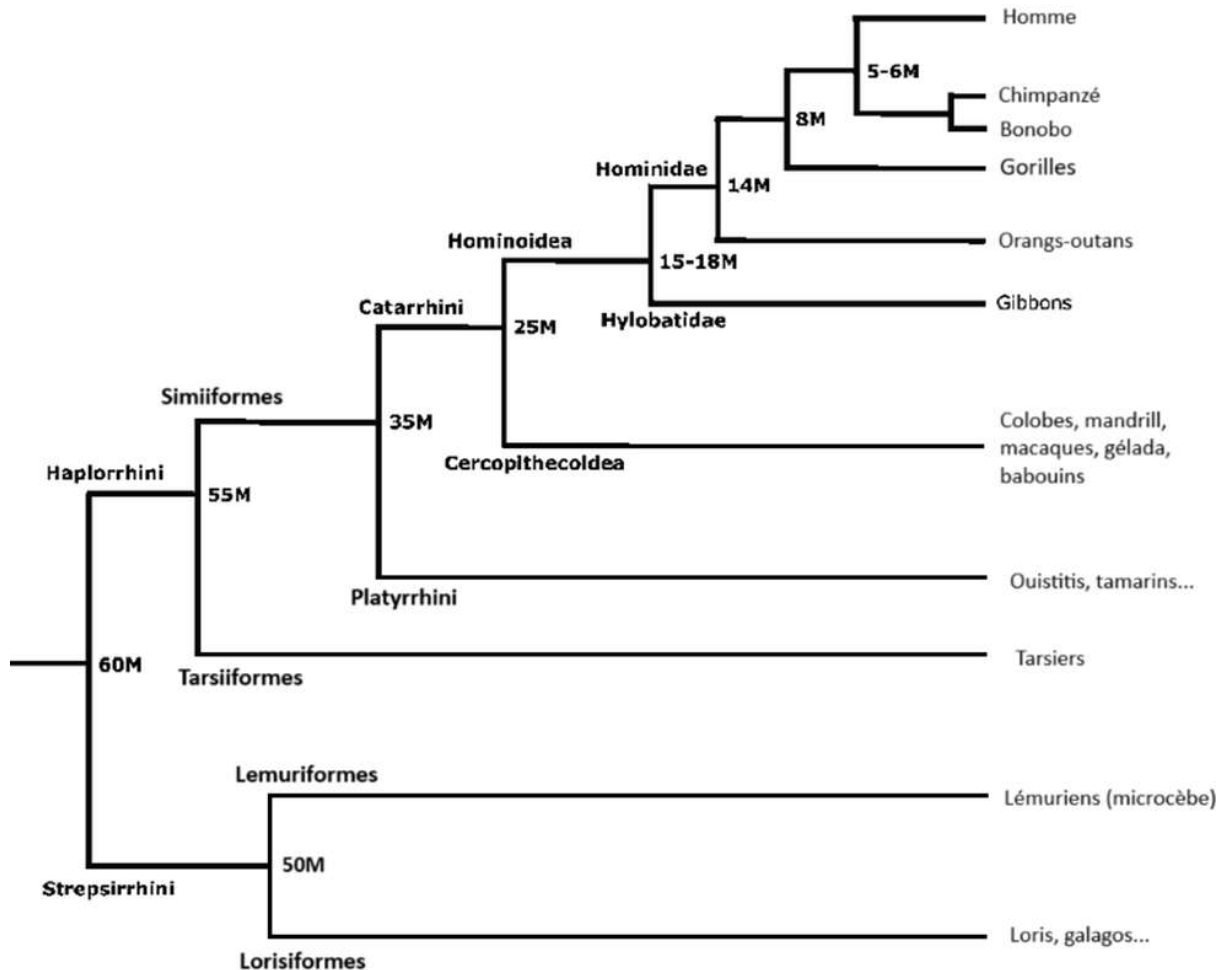
L'étude d'Allard *et al.* (2014) a recherché la présence de séquences *Alu* et de deux autres marqueurs à travers différents tissus humains par la technique d'hybridation *in situ*. Les résultats ont démontré un fort signal *Alu* dans tous les tissus testés. Parmi les trois marqueurs, le marqueur *Alu* était le plus intense et le plus présent dans tous les tissus, ce qui démontre son utilité en tant que marqueur des tissus humains.

5. Étude de la séquence *Alu* chez d'autres espèces

A. Phylogénie des primates










L'ordre des primates est le cinquième ordre le plus diversifié avec ses 233 espèces, dont l'espèce humaine *Homo sapiens*. Il se divise en deux sous-ordres monophylétiques : les *Strepsirrhini* et *Haplorrhini*. Le sous-ordre des *Strepsirrhini* se divise en deux infra-ordres appelés Lémuriformes (incluant les lémurs, comme le microcèbe) et Lorisiformes (incluant les loris et galagos). Les *Haplorrhini* se divisent en deux infra-ordres appelés *Tarsiiformes* (incluant les tarsiers) et *Simiiformes* (incluant les singes). Les *Simiiformes* se divisent en deux micro-ordres appelés *Platyrrhini* (incluant les singes du Nouveau Monde, comme les ouistitis et tamarins) et *Catarrhini* (incluant les singes de l'Ancien Monde). Les *Catarrhini* se divisent en deux super-familles appelées *Cercopithecoidea* (incluant les colobes, mandrill, macaques, gélada et babouins) et *Hominoidea*. Les *Hominoidea* se divisent en deux familles appelées *Hylobatidae* (incluant le gibbon) et *Hominidae* (incluant les orangs-outans, gorilles, chimpanzés et l'Homme) (Goodman *et al.*, 1998 ; Rocchi *et al.*, 2009). La Figure 7 récapitule cette phylogénie des primates, et le Tableau 2 répertorie des photographies de plusieurs de ces grands primates.

Figure 7 : Arbre phylogénétique des primates



Arbre phylogénétique des primates, incluant les classifications, les espèces représentatives, et la date approximative de divergence entre deux branches (inspiré de Rocchi *et al.*, 2009).

Tableau 2 : Photographies de plusieurs espèces de grands primates

<p><i>Gorilla sp.</i></p>		<p><i>Pan troglodytes</i></p>	
<p><i>Pan paniscus</i></p>		<p><i>Macaca mulatta</i></p>	
<p><i>Papio anubis</i></p>		<p><i>Hylobates lar</i></p>	
<p><i>Mandrillus sphinx</i></p>		<p><i>Pongo sp.</i></p>	
<p><i>Theropithecus gelada</i></p>		<p><i>Colobus guereza</i></p>	

Photographies issues du site internet « Wisconsin National Primate Research Center »

B. Les séquences *Alu* au sein du génome simien

Hwu *et al.* (1986) ont estimé la proportion de séquences *Alu* présentes dans le génome de trois espèces de grands primates : le génome des chimpanzés posséderait environ 330 000 copies, celui des gorilles en posséderait environ 410 000 copies, et celui des orangs-outans environ 580 000 copies.

Certaines séquences *Alu* sont partagées entre l'humain et les primates, par exemple entre l'humain et le chimpanzé (Sawada *et al.*, 1985), ou entre l'humain et l'orang-outan (Koop *et al.*, 1986). D'autres séquences sont spécifiques à l'humain, tandis que d'autres séquences sont spécifiques à une ou plusieurs espèces de primates (Minghetti et Dugaiczky, 1993). Vis-à-vis de ces observations, Hwu *et al.* (1986) concluent que de multiples copies de la séquence *Alu* ont été ajoutées ou supprimées des génomes des primates au cours des derniers millions d'années. Selon eux, de nombreux événements d'insertion se seraient produits, mais la régularité de ces événements reste incertaine. L'expérience de Liu *et al.* (2009) démontre que les taux d'amplification des séquences *Alu* varient de façon spécifique à l'espèce ou à la lignée à laquelle elle appartient.

En plus des séquences *Alu*, certains primates dont les humains possèdent également un ARN, appelé ARN BC200, qui présente de fortes similarités avec le monomère *Alu* gauche (FLAM) (Watson et Sutcliffe, 1987). Cet ARN est long d'environ 200 paires de bases. Il possède une région centrale riche en adénine, et sa séquence n'a aucune similitude avec d'autres séquences ADN connues. Il est principalement exprimé dans le système nerveux, bien que sa fonction ne soit pas encore bien connue (Tiedge *et al.*, 1993).

a. Chez le chimpanzé

La plupart des séquences *Alu* spécifiques au chimpanzé (genre *Pan*) appartiennent à la sous-famille appelée *AluYc1*. Ces séquences sont très similaires au gène source du dernier ancêtre commun homme-chimpanzé (Liu *et al.*, 2009), ce qui est probablement secondaire au faible niveau de divergence entre les séquences présentes dans cette espèce, en comparaison avec la grande hétérogénéité des séquences *Alu* au sein du génome humain (Sawada *et al.*, 1985).

Dans cette espèce, il semblerait que la réplication des séquences *Alu* soit trois fois moins active que chez l'Homme, alors qu'elle est restée aussi active chez d'autres espèces de primates, comme chez le babouin par exemple. Cela suggère un déclin général de l'activité de transposition des séquences *Alu* chez cette espèce (Waterson *et al.*, 2005).

La comparaison des génomes humains et du chimpanzé montre que, depuis la divergence hommes-chimpanzés il y a environ 6 millions d'années, il y a eu environ 2400 insertions dans le génome des chimpanzés, contre environ 5000 dans le génome humain (Hedges *et al.*, 2004 ; Mills *et al.*, 2006).

b. Chez l'orang-outan

L'orang-outan, genre *Pongo*, montre une diminution marquée du taux d'amplification des séquences *Alu* dans cette espèce car ce taux est bien inférieur à celui des chimpanzés ou des humains. En effet, son génome n'aurait acquis que 250 insertions spécifiques à l'espèce au cours des 12 derniers millions d'années (Locke *et al.*, 2011). Comme pour le chimpanzé, les séquences de l'espèce sont très proches des séquences humaines, en lien avec le faible niveau de divergence des séquences entre elles (Koop *et al.*, 1986).

c. Chez le macaque rhésus

Le macaque rhésus (*Macaca mulatta*) posséderait environ 110 000 insertions génomiques spécifiques à cette espèce. Une nouvelle sous-famille appartenant à la grande sous-famille *AluY* et partagée entre les macaques et les babouins a été identifiée par Liu *et al.* (2009). Cette nouvelle

sous-famille a évolué selon quatre lignées actuellement actives : *AluYRa1-4*, *AluYRb1-4*, *AluYRc1-2* et *AluYRd1-4* (Han *et al.*, 2007).

d. Chez les ouistitis

Les ouistitis, genre *Callithrix*, possèdent une sous-famille spécifique à leur espèce, appartenant à la grande sous-famille des *AluS*. De plus, leur génome possède une grande abondance de répétitions *Alu*, environ égale à 188 répétitions par mégabase (Liu *et al.*, 2009).

e. Chez les lémuriens

Des séquences *Alu* appartenant à la sous-famille *AluJo* ont été identifiées chez les lémuriens. Liu *et al.* (2009) ont identifié sept nouvelles sous-familles de séquences *Alu* spécifiques aux lémuriens, appartenant à la grande sous-famille *AluJ*. Ils ont également démontré que cette espèce avait une faible densité de répétitions *Alu*, environ égale à 55 répétitions par mégabase.

f. Chez les galagos

Deux familles de SINE, similaires à la famille *Alu*, ont été découvertes chez le galago. Elles ont été nommées familles *Alu* type I et type II.

La famille *Alu* de type I est étroitement liée à la famille *Alu* humaine, et montre une grande similarité de séquence avec le dimère humain, même s'il existe des divergences spécifiques de séquences entre les deux espèces. Cette famille serait apparue bien avant la divergence entre les lignées humaines et du galago. Le galago en posséderait environ 259 000 copies au sein de son génome. La divergence moyenne entre les éléments de cette famille est d'environ 17 % (Daniels et Deininger, 1983). Une étude a montré que cette famille était présente en grand nombre chez d'autres primates, tels que le lémurien, le singe hibou, le singe rhésus, et l'Homme (Daniels et Deininger, 1991). Les auteurs ne peuvent pas conclure quant à la nature réelle de ces séquences (Daniels et Deininger, 1985), tandis que d'autres pensent qu'elles n'appartiennent à la famille *Alu* (Daniels *et al.*, 1983).

La famille *Alu* de type II est spécifique au galago. Elle serait apparue plus récemment dans l'évolution, probablement ultérieurement à la divergence entre les lignées humaine et celle du galago car aucune séquence similaire n'a été trouvée dans les génomes de l'humain ou du lémurien. Sa structure est hybride. En effet, dans sa partie 5', ses éléments sont composés d'une séquence, d'environ 100 paires de bases, étroitement similaire à une autre famille de SINE, appelée famille *Monomère*, spécifique au galago et non apparentée à la famille *Alu*. Dans sa partie 3', il y a une séquence d'environ 160 paires de bases qui est très apparentée aux éléments de la séquence *Alu* de type I. Une hypothèse concernant l'apparition de ces éléments serait l'intégration d'un élément de la famille *Monomère* au centre d'un élément *Alu* de type I, ce qui aurait engendré la famille des éléments *Alu* de type II dans le génome du galago. Cette famille semble beaucoup plus active pour la rétroposition que la famille *Alu* de type I du galago. Ceci pourrait s'expliquer par l'intégration du promoteur de l'ARN polymérase III de l'élément *Monomère* dans sa moitié gauche, qui serait alors plus efficace que le promoteur présent dans les éléments *Alu* type I. De plus, il semblerait que cette famille soit divisée en plusieurs sous-familles, déterminée par le partage de mutations ponctuelles communes, d'insertions ou de délétions par rapport à la séquence consensus. A partir de cette observation, il a été supposé que les éléments *Alu* type II ont été amplifiés à partir de séquences maîtresses, sur le même modèle que les séquences *Alu* humaines (Daniels et Deininger, 1991). La divergence moyenne entre les éléments de cette famille est d'environ 14 %. La plus grande homogénéité des éléments au sein de cette famille, par rapport à la famille de type I, s'explique par son apparition plus récente (Daniels et Deininger, 1983).

C. Dans le génome d'autres espèces

Les éléments SINE dérivés de l'ARN 7SL forment un groupe appelé SINE1. Ces éléments sont similaires aux séquences *Alu*, et ont été retrouvés dans le génome de diverses espèces appartenant au Super-Ordre des *Euarthontoglires*, excepté pour une espèce que nous détaillerons dans ce paragraphe.

a. Au sein du génome murin

Le génome des rongeurs (ordre des *Rodentia*) contient un grand nombre de SINE appartenant à une famille appelée *B1*, une famille d'éléments apparentés à la famille *Alu*. La divergence entre les primates et les rongeurs aurait eu lieu il y a environ 100 millions d'années (Novacek, 1992). Le nombre d'éléments *B1* dans la sous-famille des *Murinae* a été estimé à presque un million de copies. Les éléments *B1* ont également été retrouvés chez 22 autres familles appartenant à l'ordre des *Rodentia* (Veniaminova *et al.*, 2007). Par ailleurs, aucun dimère humain n'a été trouvé chez les rongeurs, et aucun monomère murin n'a été trouvé chez les humains (Schmid et Shen, 1985).

□ Structure des séquences *B1*

Différemment des éléments *Alu*, les éléments *B1* sont des monomères composés d'environ 140 paires de bases, suivi d'une région riche en A (Krayev *et al.*, 1980). Ces éléments sont semblables à l'ARN 7SL malgré quelques divergences. En plus de la délétion centrale du domaine S, ils comportent une répétition en tandem interne de 29 paires de bases, et une délétion de neuf paires de bases entre les positions 65 et 73. De plus, il existe une séquence d'ARN 4,5S, spécifique au génome des rongeurs, qui est similaire à un élément *B1* tronqué, avec une portion nucléotidique très divergente (Haynes *et al.*, 1981). Cet ARN est composé de 90 à 94 nucléotides, avec une courte séquence oligo(dU) de longueur variable à son extrémité 3' (Jelinek et Leinwand, 1978 ; Harada et Kato, 1980). Cet ARN pourrait provenir d'une séquence *B1* (Rogers, 1985).

□ Origine des séquences *B1*

Quentin (1994) a identifié quelques séquences ne possédant pas la duplication de 29 paires de bases, très similaires à la séquence *FLAM-A* des primates. Cette famille a été nommée *Proto-B1*, appelée *PB1*. Plusieurs sous-familles ont été étudiées au sein de cette famille *PB1*, comme les familles *PB1D7* et *PB1D10*. Les éléments de ces familles sont délétés dans la même région que les éléments *B1* par rapport à l'ARN 7SL, mais uniquement de sept et dix paires de bases respectivement. De plus, cet auteur a mis en évidence une similarité de séquence entre les éléments de la famille *PB1D10* et l'ARN 4,5S, suggérant alors la présence d'une séquence ancestrale commune. *FLAM-A/PB1* serait probablement né chez l'ancêtre commun des *Euarthontoglires* (Nishihara *et al.*, 2002).

b. Au sein du génome des hamsters

Les hamsters représentent la famille des *Cricetinae*, appartenant à l'ordre des *Rodentia* (Haynes *et al.*, 1981). Dans leur génome, il existe deux sous-familles de séquences similaires aux familles *Alu*, appelées Type I et Type II. Il s'agit de séquences monomériques mesurant environ 140 et 96 paires de bases respectivement. Les séquences de type I sont analogues aux séquences *B1*. Les éléments de type I et de type II sont analogues aux moitiés droite et gauche du dimère *Alu* humain, respectivement. La famille de type II semble produire des ARN polymérase III, différemment de la famille de type I (Daniels et Deininger, 1983). Le rat possède une famille très similaire à la famille de type II (Leinwand *et al.*, 1982).

c. Au sein du génome de la musaraigne arboricole (Nishihara et al., 2002)

La musaraigne arboricole est une espèce faisant partie de l'ordre *Scandentia*, un ordre dérivé du Super-Ordre des *Euarchontoglires*, de même que l'ordre des *Glires*. Deux familles de SINE, similaires à la famille *Alu*, ont été trouvées dans le génome de cette espèce. Elles ont été nommées Tu type I et Tu type II. Les régions 3' de ces éléments présentent plus de 85 % de similarité de séquence à l'ARN 7SL humain, suggérant que les éléments de ces familles seraient dérivés de cet ARN. Les éléments de type Tu I sont composés d'une région 5' de 96 paires de bases, et d'une région 3' dérivée de l'ARN 7SL. Cette structure hybride se rapproche de la structure des éléments *Alu* de type II du galago. Les éléments de type Tu II sont composés d'une région 5' de 86 paires de bases semblable à la région 5' des éléments Tu I, et d'une région 3' composé d'un dimère provenant d'un monomère, lui-même dérivé de l'ARN 7SL, ce qui en fait une région similaire à la séquence *Alu* des primates. La musaraigne arboricole posséderait environ une centaine de copies des types I et II au sein de son génome. D'après les auteurs, ces éléments pourraient avoir perdu leur activité rétropositionnelle.

d. Au sein du génome de la myxine de bord de mer (Kojima, 2020)

La myxine de bord de mer (*Eptatretus burgeri*) est un vertébré appartenant à la Super-Classe des agnates. Une nouvelle famille de SINE1, appelée *SINE1-1_EBu*, a été découverte au sein de son génome. Il s'agit de la première découverte d'un élément SINE1 issu d'un organisme n'appartenant pas au Clade des *Euarchontoglires*.

Les éléments *SINE1-1_EBu* sont composés de cinq parties différentes, dont les deux situées en 5' dérivent de l'ARN 7SL. Les autres parties seraient dérivées d'autres rétrotransposons non-LTR de type RTE (*retrotransposon-like element*) appartenant à la famille des LINE. Concernant l'origine de ces éléments, ces auteurs supposent que cet élément chimérique serait formé par la réunion des régions 5' et 3' de l'ARN 7SL, constituant le domaine *Alu*, ainsi que par deux éléments non autonomes dérivés de type RTE, RTE-4_EBu et RTE-2_EBu, après avoir subi une délétion interne. De plus, au sein de la structure secondaire de ces éléments se trouve une structure similaire au repliement des ARN *Alu*. Cette homologie de structure secondaire pourrait refléter l'importance fonctionnelle du domaine *Alu* dans la prolifération des SINE.

Vis-à-vis de la grande homogénéité entre les différents membres SINE1 (identité de séquence supérieure à 95 % entre les différents membres), les auteurs concluent qu'il s'agit d'une jeune famille de SINE. Une différence majeure des *SINE1-1_EBu* par rapport aux autres *SINE1* se trouve dans la machinerie d'amplification de ces séquences. En effet, contrairement aux éléments *Alu* et *B1*, les éléments *SINE1-1_EBu* seraient transposés par une famille de rétrotransposons non-LTR appelée *RTE-2_EBu*, à la place de la machinerie *L1*. A partir de ces constatations, les auteurs supposent que le *SINE1-1_EBu* serait apparu et aurait évolué indépendamment de *Alu* et *B1*.

e. Evolution des séquences dérivées de l'ARN 7SL

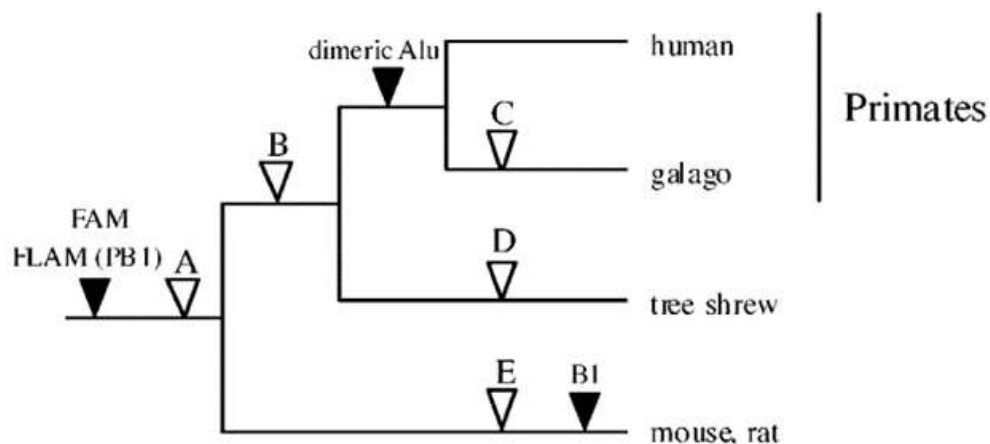
Selon l'étude de (Kojima, 2020), toutes les séquences *SINE1* retrouvées chez les *Euarchontoglires* proviendraient d'une séquence ancestrale apparue chez un ancêtre commun de ce Super-Ordre. Cette famille aurait pu ressembler à *FLAM-A/PB1*.

Plusieurs auteurs ont proposé un ordre d'apparition pour ces différentes séquences issues de l'ARN 7SL. L'apparition de la séquence *FAM* serait le premier évènement, suivi de l'apparition des séquences *FLAM* et *FRAM* qui en découlent. Ensuite, par son étroite similarité de séquence avec *FLAM*, la famille *PB1* serait apparue proche de cette première bifurcation. De cette dernière famille serait ensuite apparu *PB1D* par délétion interne, qui aurait elle-même engendré la famille *B1* par duplication interne. Il en découle l'hypothèse selon laquelle l'apparition des séquences *FAM* et *FLAM* aurait eu lieu chez un ancêtre commun aux primates et aux rongeurs. Par la suite, les séquences *Alu* et *B1* auraient évolué indépendamment dans chaque lignée (Quentin, 1994).

Nishihara et al. (2002) ont étudié plusieurs arbres phylogénétiques incluant l'apparition des séquences *Alu* et apparentées, en fonction de la position phylogénétique de la musaraigne

arboricole. D'après les études postérieures qui ont clarifié sa position (Fan *et al.*, 2019), nous allons exposer les hypothèses de l'apparition des séquences *Alu* et apparentées. En accord avec l'étude de Veniaminova *et al.* (2007), les auteurs suggèrent que l'apparition de *FLAM-A/PB1* à partir de l'ARN 7SL se soit produite une seule fois dans le génome d'un ancêtre commun aux rongeurs, primates, et aux musaraignes arboricoles. Ensuite, ils proposent deux hypothèses concernant l'apparition de *PB1D* : soit la séquence serait apparue chez un ancêtre commun aux rongeurs et à la musaraigne arboricole, soit elle serait apparue indépendamment dans chaque lignée. Les deux hypothèses sont discutables : la première car aucune séquence de ce type n'a été retrouvée dans le génome humain, la deuxième car cela requerrait une délétion identique du même nombre de dinucléotides à partir de l'ARN 7SL au sein des deux lignées différentes, ce qui paraît improbable. L'apparition des séquences *PB1D* n'est donc pas encore clairement résolue. Ensuite, les auteurs ont émis deux hypothèses concernant les séquences *Tu* de type I de la musaraigne et *Alu* de type II du galago. La première hypothèse serait que ces séquences seraient apparues chez un ancêtre commun à la musaraigne arboricole et aux primates. Cette hypothèse impliquerait un événement de délétion de la séquence *Alu* galago type II engendrant l'apparition de *Tu* type I dans la lignée de la musaraigne arboricole. La deuxième hypothèse serait que ces séquences soient apparues indépendamment dans chaque lignée. Selon ces auteurs, la première hypothèse est moins probable car aucune séquence de ce type n'a été retrouvée dans le génome humain. Ils suggèrent alors que concernant les deux situations, l'hypothèse d'apparition indépendante dans chaque lignée soit préconisée. Les différentes hypothèses sont récapitulées dans la Figure 8.

Figure 8 : Arbre phylogénétique hypothétique de l'apparition des séquences *Alu* et apparentées



*Cet arbre phylogénétique propose des événements d'apparition des séquences *Alu* et apparentées dans différentes lignées. Les triangles représentés en noirs représentent des hypothèses d'apparition communément partagées, et les triangles représentés en blanc correspondent à des hypothèses d'apparition de ces auteurs. « Dimeric *Alu* » fait référence à la séquence *Alu* humaine, et « B1 » aux séquences B1. Le point A représente l'hypothèse de l'apparition de *PB1D* chez un ancêtre commun à la musaraigne arboricole (tree shrew) et aux rongeurs (mouse, rat), et les points D et E représentent l'hypothèse complémentaire selon laquelle *PB1D* serait apparue indépendamment dans les deux lignées. Le point B représente des séquences *Tu* type I et *Alu* type II du galago chez un ancêtre commun à la musaraigne arboricole et aux primates, et les points C et D représentent l'hypothèse complémentaire selon laquelle ces séquences seraient apparues indépendamment dans ces deux lignées (Nishihara *et al.*, 2002).*

L'ARN BC200 serait apparu du côté des primates, juste avant l'apparition de *FLAM-C*. L'ARN BC200 serait apparu par incorporation de séquences en aval du site d'intégration d'un élément *FLAM* (Tiedge *et al.*, 1993).

Deuxième partie : étude expérimentale

1. Introduction et contexte de l'étude

L'idée à l'origine de cette thèse provient d'une demande d'analyse inhabituelle adressée il y a quelques années au laboratoire d'anatomocytopathologie de l'Ecole Nationale Vétérinaire d'Alfort (EnvA).

Près d'une zone d'accident de la route, un cœur de taille compatible avec un cœur humain a été retrouvé à côté d'un cadavre humain par ailleurs décapité par la violence de l'accident. L'hypothèse selon laquelle le cœur aurait pu appartenir à ce cadavre, et aurait été expulsé du thorax par le cou a été initialement considérée par les médecins légistes comme impossible. Aussi, afin de déterminer s'il s'agissait d'un cœur humain, la gendarmerie a réquisitionné le service d'anatomocytopathologie qui dispose d'un automate d'hybridation *in situ* avec une sonde *Alu*. Cette sonde, qui correspond en fait au témoin positif de réaction fourni par le fabricant de cet automate destiné initialement au diagnostic humain, a été détournée de son utilisation initiale et largement utilisée au sein du laboratoire de l'EnvA pour le marquage des tumeurs humaines greffées sur des souris (xénogreffes) dans un contexte de recherche biomédicale. Dans le cas du cœur d'origine inconnue, un marquage positif a été obtenu avec la sonde *Alu* et suggérait donc qu'il s'agissait bien d'un cœur humain. La séquence de cette sonde n'étant cependant pas publique et sa réelle spécificité vis-à-vis de l'espèce humaine n'étant pas documentée, il n'était pas possible d'exclure formellement qu'il puisse s'agir d'un cœur de primate non humain. En effet, nos recherches bibliographiques exposées en première partie ont montré que les primates possédaient également des séquences *Alu* au sein de leur génome mais que celles-ci pouvaient être différentes des séquences humaines. Si la réponse finale a pu être apportée par séquençage de l'ADN prélevé sur le cœur et comparaison avec celui prélevé sur le cadavre (ce qui a montré une identité parfaite), la question de la spécificité de la sonde restait en suspens.

L'objectif de cette étude était donc de déterminer si cette sonde *Alu* commerciale était spécifique des tissus humains ou si elle reconnaissait également des tissus de primates non humains et, en particulier, de grands singes dont la taille et la forme des organes se rapprochent de celles de l'espèce humaine. Il était ainsi espéré que les résultats de cette étude permettent de valider ou non l'utilisation de cette sonde à des fins médico-légales.

Par ailleurs, afin de confronter nos résultats à nos recherches bibliographiques, nous avons également testé cette sonde sur des tissus d'autres espèces animales non primates.

2. Matériels et méthodes

La sonde commerciale a été testée dans deux groupes d'espèces animales : primates et non primates.

A. Sélection des échantillons

a. Pour les espèces primates

Dans un premier temps, nous avons voulu tester la sonde sur différents tissus issus de plusieurs espèces de primates, en ciblant préférentiellement de grands singes et en cherchant à obtenir pour chaque espèce des tissus de deux individus. La disponibilité de tissus de primates étant très limitée, nous nous sommes rapprochés d'un laboratoire d'anatomie pathologique vétérinaire recevant fréquemment des prélèvements issus de zoos et avons fait une recherche rétrospective dans la base de données du laboratoire, croisée avec la liste de prélèvements de primates envoyés par un zoo français, ce qui nous a permis de sélectionner des échantillons provenant de plusieurs espèces de primates. Les prélèvements récupérés sous forme de blocs de paraffine avaient tous été fixés en solution de formol à 4 %. La durée de fixation n'était pas connue.

Le premier critère de sélection des espèces s'est porté sur la taille des différentes espèces. Etant donné que nous souhaitions étudier préférentiellement la réactivité de la sonde *Alu* chez des primates dont les organes sont de taille similaire aux organes humains, nous avons sélectionné les espèces de primates appartenant au micro-ordre des *Catarrhini*. Nous avons alors réalisé une liste des espèces avec ce critère d'inclusion. Le deuxième critère de sélection s'est porté sur la disponibilité des échantillons présents au laboratoire avec lequel nous avons travaillé.

En tenant compte de la difficulté d'accès à ce type de prélèvement, nous avons choisi de sélectionner deux blocs de paraffine provenant de deux individus distincts par espèce, lorsque cela était possible. Parmi les prélèvements présents au laboratoire, nous avons privilégié des blocs contenant des organes très cellulaires, comme la rate ou le foie. Si la nature du tissu ou de l'organe n'était pas un critère de sélection en soi, considérant que le matériel génétique était identique dans tous les tissus d'un même individu, nous voulions privilégier des organes riches en cellules afin de pouvoir effectuer des comptages cellulaires. Nous avons également veillé à lire les comptes-rendus histologiques de chaque bloc afin de sélectionner des organes bien conservés (exclusion des organes autolysés et/ou putréfiés). Si nous avions le choix entre plusieurs prélèvements, les blocs les plus récents étaient choisis.

b. Chez les espèces non primates

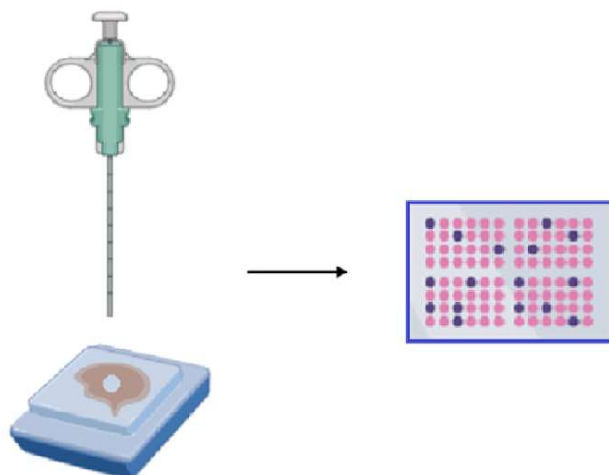
Nous voulions tester la réactivité de la sonde *Alu* chez diverses espèces animales non primates, incluant des espèces domestiques et non domestiques. Le choix des espèces s'est fait en fonction de la disponibilité des échantillons présents dans le service d'anatomopathologie de l'EnvA. Afin de limiter les coûts et en tenant compte de nos recherches bibliographiques qui indiquaient l'absence de séquences *Alu* dans ces espèces, nous avons choisi de ne tester qu'un seul individu par espèce. Lorsque cela était possible, nous avons choisi des organes très cellulaires et les plus récents possible.

B. Technique du "Tissue Micro-Array"

Pour les tissus de primates, des blocs ont été préparés à partir de la technique du « Tissue Micro-Array », aussi appelée TMA, qui permet de rassembler plusieurs prélèvements issus de différents blocs au sein d'un même bloc paraffine afin de tester plusieurs échantillons en même temps et dans les mêmes conditions. Pour cela, nous avons utilisé un kit TMA (Tissue-Tek, Quick-

Ray) pour réaliser des carottes de 5 mm de diamètre dans chaque organe choisi que nous avons ensuite déposées dans les puits dédiés d'un bloc paraffine fourni avec le kit. Pour chaque puits, l'espèce et le tissu correspondant étaient notés.

Figure 9 : Schéma de la technique *Tissue Micro-Array*



Représentation schématique de la fabrication du bloc Tissue Micro-Array. La partie gauche de la figure montre l'étape de carottage d'un organe présent au sein d'un bloc de paraffine à l'aide de l'instrument dédié. La partie droite de la figure montre le dépôt de chaque carotte d'organe au sein d'un puits du bloc Tissue Micro-Array.

Nous avons choisi d'utiliser cette technique afin de gagner en temps, en coût de réactif, et en reproductibilité, étant donné que tous les échantillons sont traités uniformément au cours de cette technique. Un aperçu d'un bloc TMA finalisé est visible sur la Figure 10.

Figure 10 : Photographie d'un bloc *Tissue Micro-Array*



Le témoin positif de l'expérience se trouve dans le puits inférieur droit, et les autres puits comportent les tissus provenant des différentes espèces incluses dans l'expérience.

Nous avons fabriqué deux blocs TMA, comportant chacun des tissus provenant des mêmes individus, mais en prélevant des organes différents entre les deux blocs lorsque cela était possible. De cette façon, l'expérience a pu être réalisée sur deux organes différents pour chaque individu. Les deux organes en question ont donc été testés au cours de deux expériences distinctes.

C. Hybridation *in situ*

Pour les primates, les blocs testés étaient les blocs de TMA produits. Pour les autres espèces, les blocs initiaux ont été utilisés sans avoir recours au TMA.

Les blocs paraffines ont été coupés à 4 µm et les sections recueillies sur lames de verre pour coloration hémalun-éosine-safran (HES) standard et pour hybridation *in situ*. Pour les espèces non primates, certaines lames pouvaient contenir plusieurs sections provenant de blocs et donc d'espèces différentes.

L'hybridation *in situ* a été effectuée sur automate Ventana Discovery XT (Roche Diagnostics). La sonde *Alu* utilisée était de référence 05278694001, produite par le laboratoire Roche Diagnostics, et dont la séquence n'était pas publique. Le détail des étapes du protocole expérimental est répertorié dans l'annexe 1. Le système de révélation utilisait la 3,3-diaminobenzidine (DAB), produisant un précipité brun dans le noyau des cellules marquées. Une contre-coloration à l'hématoxyline était utilisée.

Pour notre protocole, nous avons choisi d'inclure des témoins positif et négatif. La sonde de notre étude étant une sonde habituellement utilisée comme témoin positif des réactions d'hybridation *in situ* sur tissu humain, nous avons utilisé comme témoin positif une xénogreffe de tumeur mammaire humaine sur une souris. Le témoin négatif correspondait au tissu murin adjacent à cette tumeur.

Après obtention des résultats initiaux sur les tissus de primates, nous avons réalisé dans un second temps une nouvelle série d'expériences selon quatre protocoles différents. Les conditions initiales ont été répétées, afin d'avoir un point de comparaison fiable, puis trois nouvelles conditions expérimentales ont été testées. Les nouveaux protocoles proposés étaient les suivants :

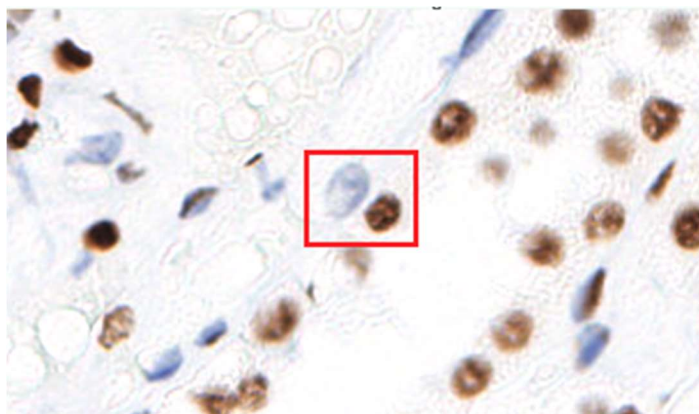
- un mode CC2UC, c'est-à-dire le paramètre CC2 (condition initiale) en mode ultra court
- un mode CC2C, c'est-à-dire la paramètre CC2 (condition initiale) en mode court associé au paramètre temps de protéase réglé à huit minutes
- un mode CC2C, c'est-à-dire la paramètre CC2 (condition initiale) en mode court associé au paramètre temps de protéase à seize minutes.

D. Évaluation du marquage de chaque échantillon

a. Lecture des lames

L'évaluation du marquage des tissus a été réalisée à l'aide d'un microscope photonique. Cette étape s'est déroulée à l'aveugle, ce qui signifie que le lecteur des lames ne connaissait ni l'espèce ni l'individu dont provenait l'échantillon. Les cellules marquées par la sonde étaient repérées par leur marquage nucléaire brun foncé (Figure 11). Lorsqu'une cellule présentait un marquage nucléaire brun, nous l'avons considérée positive, quelle que soit l'intensité de marquage. Un marquage autre que nucléaire était considéré comme aberrant et non pris en compte.

Figure 11 : Détail de deux noyaux dont un marqué par la sonde *Alu*



Photographie d'une lame de rein de babouin (Gx40) montrant le détail de deux noyaux. Le noyau de droite a été coloré par la 3,3-diaminobenzidine. Le noyau de gauche n'a pas été marqué et apparaît de couleur bleu clair avec la contre-coloration à l'hématoxyline.

b. Comptage des cellules positives

Nous avons réalisé une estimation de la proportion de cellules marquées sur chaque échantillon de chaque expérience. Pour cela, chaque lame a été scannée par un scanner NanoZoomer (Hamamatsu). Nous avons compté le nombre de cellules marquées sur dix champs répartis de manière uniforme sur chaque échantillon, puis nous en avons fait une moyenne. Le comptage a été réalisé informatiquement à l'aide du logiciel NDP.view 2 (Hamamatsu). Les champs de lecture ont été tracés informatiquement à l'aide du logiciel au grossissement le plus avancé (x40), et les cellules ont été comptées dans ce champ, sur une surface environ égale à 0,04 mm².

Afin d'estimer le nombre de cellules marquées dans l'échantillon entier, nous avons réalisé des intervalles de confiance à 95 % à partir de nos moyennes. Nous avons noté n le nombre de cellules dénombrées dans l'échantillon, et p l'estimation de cellules marquées. Nous avons alors calculé l'écart-type σ à l'aide de la formule suivante :

$$\sigma = \sqrt{(p(1 - p) \div n)}$$

Nous avons ensuite vérifié que les produits $n * p$ ou $n * (1 - p)$ étaient supérieurs à 5, puis nous avons calculé l'intervalle de confiance à 95 %, IC_{95%} à l'aide de la formule suivante :

$$IC_{95\%} = p \pm 1,96 * \sigma$$

Cet intervalle de confiance nous permet de faire de l'inférence, et nous indique que, sous l'hypothèse que l'estimation ne soit pas biaisée par du biais d'estimation, il y a 95 % de chances pour que la valeur réelle du pourcentage de cellules marquées dans l'organe de l'espèce se trouve dans cet intervalle de valeurs.

3. Résultats

A. Chez les espèces primates

a. Description de l'échantillon

Nous avons collecté des prélèvements provenant de dix espèces différentes : une espèce de gorille (*Gorilla sp.*), le chimpanzé (*Pan troglodytes*), le bonobo (*Pan paniscus*), une espèce d'orang-outan (*Pongo sp.*), une espèce de babouin (*Papio sp.*), le mandrill (*Mandrillus sphinx*), une espèce de colobe (*Colobus sp.*), une espèce de macaque (*Macaca sp.*), une espèce de gibbon (*Hylobates sp.*) et le gélada (*Theropithecus gelada*). Pour tous les individus, hormis les chimpanzés, le bonobo, les mandrills et les colobes, nous n'avions pas connaissance de l'espèce précise, mais uniquement du genre. Pour chaque espèce, nous avons pu collecter des blocs de deux individus différents, sauf pour le bonobo où un seul bloc était disponible. Le Tableau 3 récapitule les différentes espèces testées, avec la date de fixation de chaque prélèvement, et les organes correspondants.

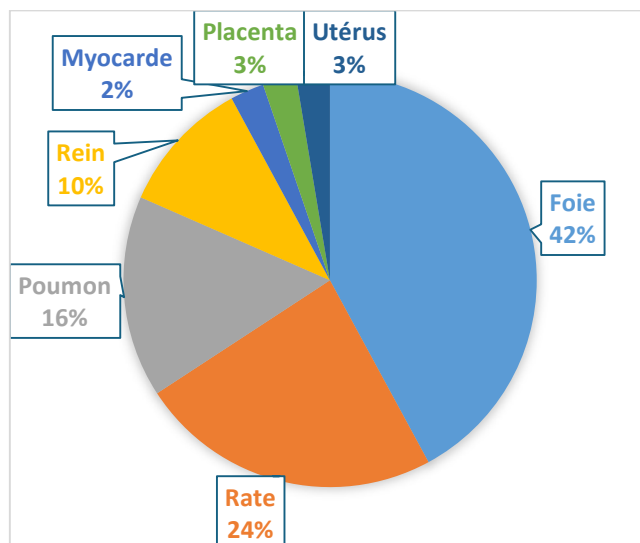
Tableau 3 : Individus et organes testés au cours des deux expériences

Espèce	Année de prélèvement	Numéro de bloc	Organe	Espèce	Année de prélèvement	Numéro de bloc	Organe
<i>Gorilla sp. 1</i>	2016	1	Rate	<i>Mandrillus sphinx 1</i>	2023	2	Foie
<i>Gorilla sp. 1</i>		2	Poumon	<i>Mandrillus sphinx 2</i>	2022	1	Poumon
<i>Gorilla sp. 2</i>	2017	1	Foie	<i>Mandrillus sphinx 2</i>		2	Utérus
<i>Gorilla sp. 2</i>		2	Rein	<i>Colobus sp. 1</i>	2016	1	Rate
<i>Pan troglodytes 1</i>	2021	1	Foie	<i>Colobus sp. 1</i>		2	Foie
<i>Pan troglodytes 1</i>		2	Rein	<i>Colobus sp. 2</i>	2017	1	Foie
<i>Pan troglodytes 2</i>	2021	1	Placenta	<i>Colobus sp. 2</i>		2	Poumon
<i>Pan troglodytes 2</i>		2	Foie	<i>Macaca sp. 1</i>	2023	1	Rein
<i>Pongo sp. 1</i>	2019	1	Poumon	<i>Macaca sp. 1</i>		2	Foie
<i>Pongo sp. 1</i>		2	Rate	<i>Macaca sp. 2</i>	2023	1	Rate
<i>Pongo sp. 2</i>	2022	1	Rate	<i>Macaca sp. 2</i>		2	Foie
<i>Pongo sp. 2</i>		2	Poumon	<i>Hylobates sp. 1</i>	2023	1	Rate
<i>Pan paniscus</i>	2015	1	Rate	<i>Hylobates sp. 1</i>		2	Foie
<i>Pan paniscus</i>		2	Foie	<i>Hylobates sp. 2</i>	2023	1	Foie
<i>Papio sp. 1</i>	2023	1	Rate	<i>Hylobates sp. 2</i>		2	Poumon
<i>Papio sp. 1</i>		2	Foie	<i>Theropithecus gelada 1</i>	2017	1	Foie
<i>Papio sp. 2</i>	2024	1	Rein	<i>Theropithecus gelada 1</i>		2	Foie
<i>Papio sp. 2</i>		2	Foie	<i>Theropithecus gelada 2</i>	2020	1	Foie
<i>Mandrillus sphinx 1</i>	2023	1	Myocarde	<i>Theropithecus gelada 2</i>		2	Rate

Le numéro suivant le nom d'espèce permet de différencier les deux individus appartenant à la même espèce. Les numéros de bloc font référence au numéro d'expérience. Par exemple, les prélèvements du bloc 1 étaient tous présents sur le même bloc TMA, et ont donc été marqués au cours de la même expérience. Les âges des différents prélèvements sont également répertoriés dans le tableau.

Pour certaines espèces, aucun prélèvement récent n'était disponible, nous avons donc choisi les organes les mieux conservés. Les deux expériences ont été menées sur deux organes différents pour chaque individu, excepté pour le gélada 1 où le foie a été marqué dans les deux expériences. La Figure 12 récapitule la répartition des organes testés.

Figure 12 : Répartition des organes testés au cours des deux expériences

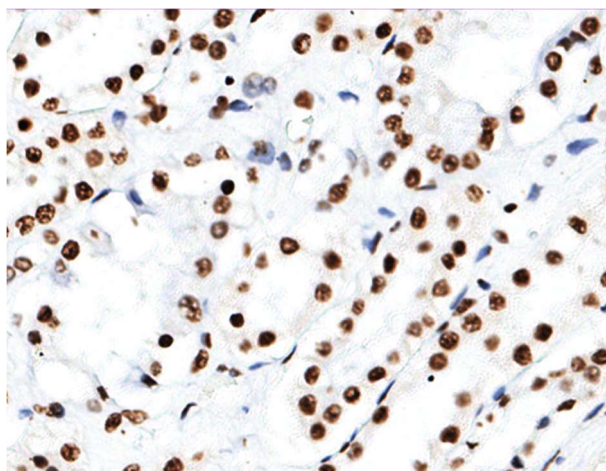


La figure ci-dessus montre la répartition des organes testés au cours des deux expériences, avec une prédominance de foie, puis de rate, qui étaient des organes recherchés pour notre étude.

b. Résultats généraux

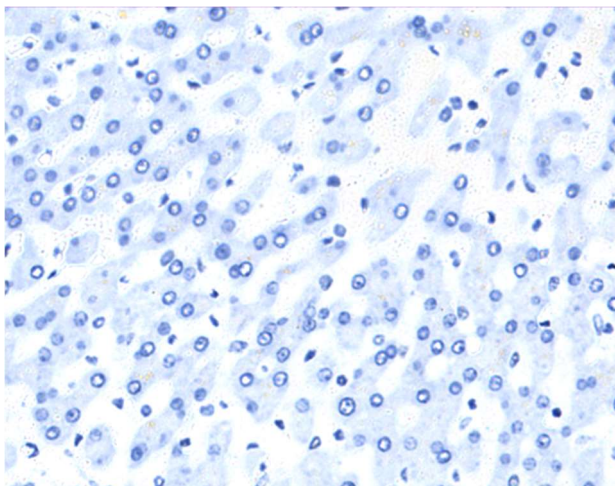
Etant donné que toutes les cellules d'un même organisme possèdent le même matériel génétique, nous nous attendions à obtenir soit un marquage total des cellules de l'échantillon, soit une absence de marquage sur l'entièreté du tissu. Nous avons obtenu des résultats inattendus, puisque la presque totalité des échantillons ont été partiellement marqués, y compris les témoins positifs. Certains tissus ont montré un marquage fort (Figure 13) alors que pour d'autres celui-ci était faible. Seuls les tissus du bonobo et le témoin négatif n'ont montré aucun marquage pour les deux expériences (Figure 14).

Figure 13 : Exemple d'un tissu avec marquage positif



Sur cette photographie de rein de babouin (Gx40), nous notons une prédominance de noyaux marqués, ainsi que la présence de quelques noyaux non marqués par la DAB

Figure 14 : Exemple d'un tissu avec marquage négatif



Cette photographie de foie de bonobo (Gx40) correspond au seul individu de l'expérience qui a montré une absence complète de marquage nucléaire par la DAB (hors le témoin négatif).

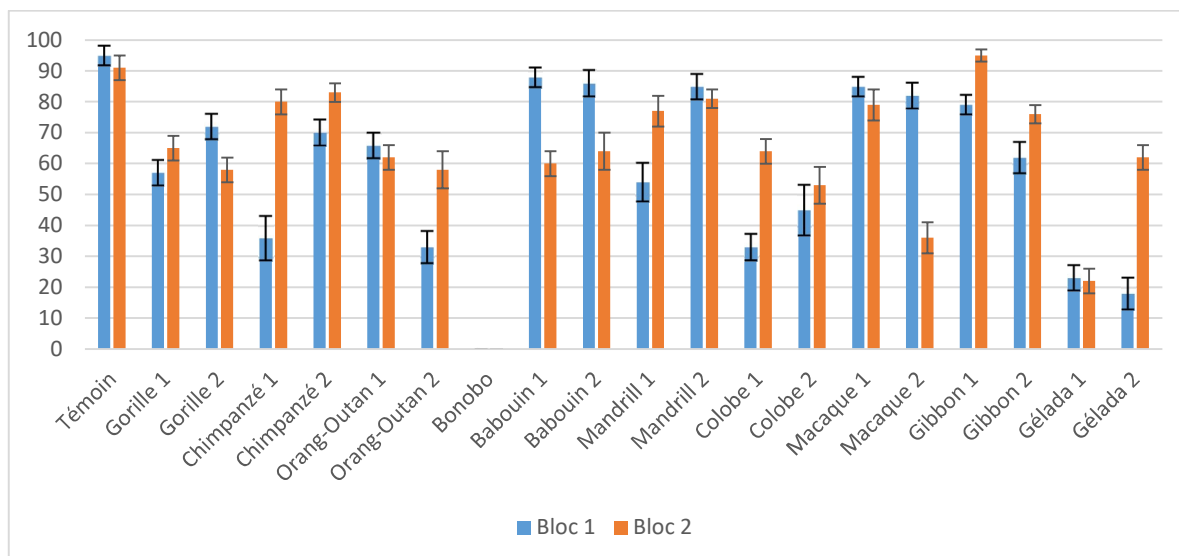
Afin de trouver une explication à ce marquage partiel prédominant, nous avons étudié l'influence de plusieurs facteurs sur la proportion de cellules marquées au sein des échantillons.

c. Étude de différents facteurs pouvant influencer la proportion de cellules marquées dans les échantillons

- *Proportion de cellules marquées par échantillon*

Les résultats de dénombrement des cellules marquées dans chaque échantillon, associés à leur intervalle de confiance à 95 %, sont répertoriés dans la Figure 15.

Figure 15 : Pourcentages de cellules marquées au sein de chaque échantillon



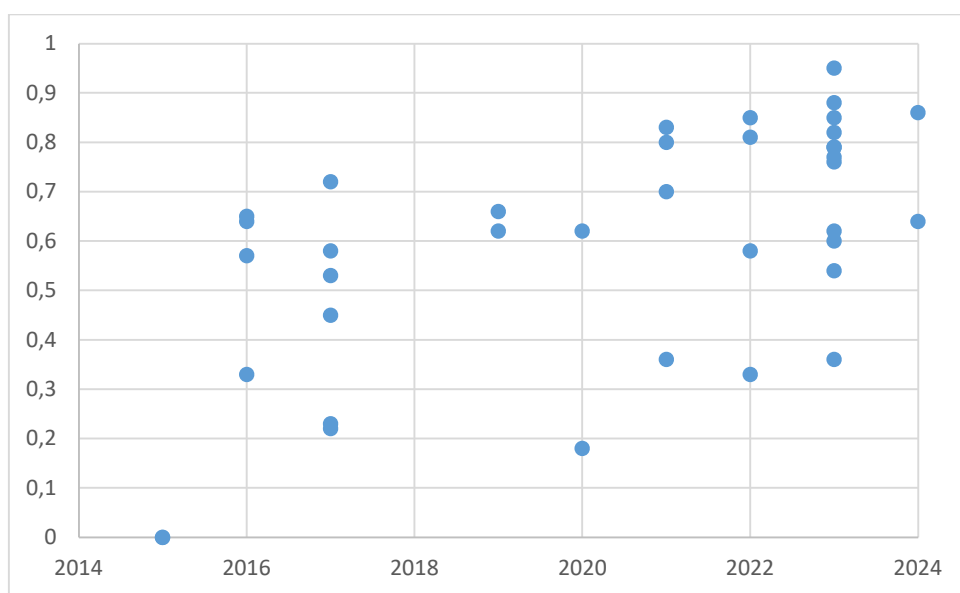
Le graphique ci-dessus représente le pourcentage de cellules marquées pour chaque échantillon. Les colonnes bleue et jaune appariées à un animal représentent la proportion de cellules marquées pour chacune des deux expériences. Les intervalles de confiance à 95 % sont associés à chaque valeur et représentés par les traits noirs. Le tableau complet avec les données à l'origine de ce graphique se trouve en annexe 2.

Les pourcentages de cellules marquées entre deux organes d'un même individu étaient similaires pour le témoin, le gorille 1, l'orang-outan 1, le bonobo, le mandrill 2, le colobe 2, le macaque 1, et le gélada 1. Les autres individus, représentés par le gorille 2, les deux chimpanzés, l'orang-outan 2, les deux babouins, le mandrill 1, le colobe 1, le macaque 2, les deux gibbons et le gélada 2 n'ont pas montré de pourcentages similaires de cellules marquées entre les deux organes testés. Pour la majorité des individus, la proportion de noyaux marqués entre les deux organes testés était donc différente.

- *Étude de l'influence de la date de fixation sur la proportion de cellules marquées*

Nous avons étudié une potentielle influence de l'âge du prélèvement sur la proportion de cellules marquées. Pour cela, nous avons représenté la distribution du marquage des différents échantillons en fonction de leur date de fixation, toutes espèces et organes confondus (Figure 16).

Figure 16 : Pourcentages de cellules marquées en fonction de la date de prélèvement du tissu



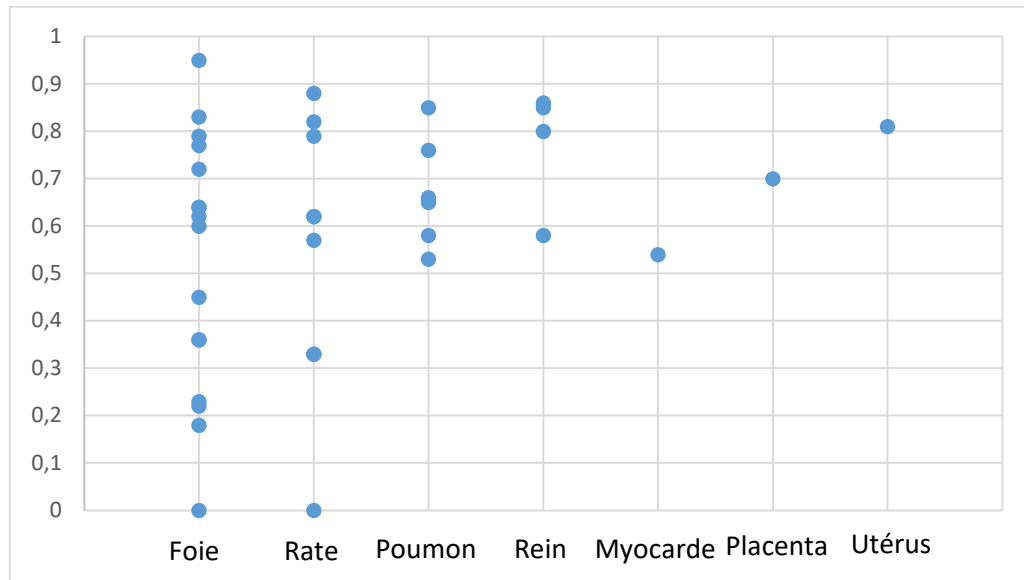
Sur le graphique ci-dessus représentant le pourcentage de cellules marquées en fonction de la date de prélèvement de l'échantillon, un point représente la moyenne de cellules marquées pour un échantillon.

Tous les points appartenant à ce graphique ne sont pas clairement indépendants car certains appartiennent au même animal ou à la même espèce par exemple. Par ailleurs, comme annoncé précédemment, tous les organes sont confondus. Etant donné les potentiels multiples facteurs pouvant influencer le marquage, et l'impossibilité d'étudier un seul facteur variable à la fois, nous ne pouvons pas réaliser de test statistique interprétable afin de vérifier la présence d'une association statistique entre le pourcentage de cellules marquées et la date de fixation de l'échantillon.

- *Étude de l'influence de l'organe sur la proportion de cellules marquées*

Nous avons étudié une potentielle influence de l'organe sur la proportion de cellules marquées dans un échantillon. Ainsi, nous avons analysé la répartition des cellules marquées en fonction du type d'organe, toutes espèces et années confondues (Figure 17).

Figure 17 : Répartition des moyennes de cellules marquées en fonction de l'organe



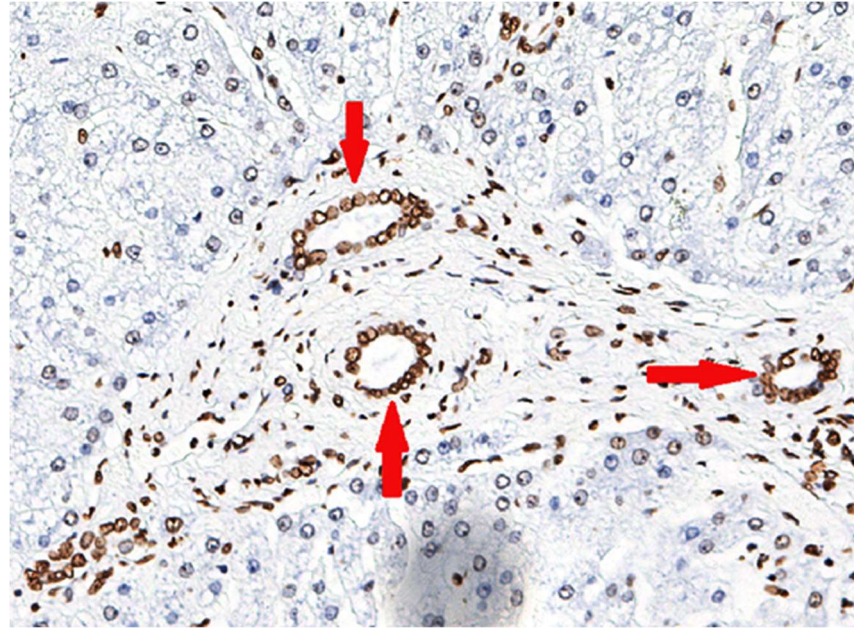
Sur le graphique ci-dessus représentant le pourcentage de cellules marquées en fonction de l'organe utilisé, chaque point représente la moyenne des cellules marquées pour un échantillon.

De même que pour la figure 16, étant donné les potentiels multiples facteurs pouvant influencer le marquage, et l'impossibilité d'étudier un seul facteur variable à la fois, nous ne pouvons pas réaliser de test statistique interprétable afin de vérifier la présence d'une association statistique entre le pourcentage de cellules marquées et l'organe testé. Cependant, subjectivement et de façon non formelle, nous ne relevons pas de différence de marquage frappante entre la proportion de cellules marquées et l'organe utilisé.

- *Étude de l'influence du type de cellules sur le marquage des cellules*

Pour plusieurs prélèvements de foie, nous avons remarqué que les canaux biliaires ont été particulièrement marqués, comme illustré par la Figure 18.

Figure 18 : Exemple de marquage important des cholangiocytes

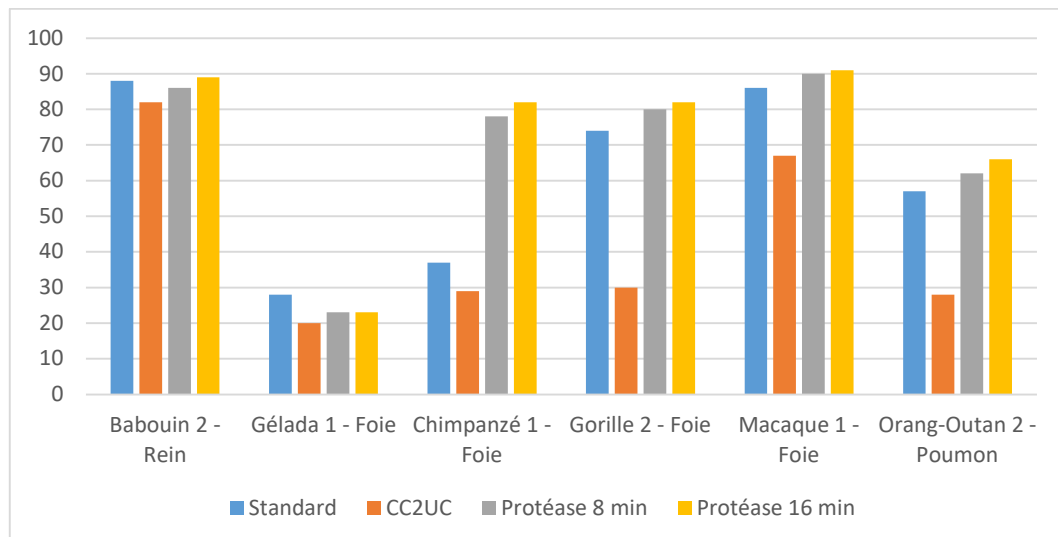


Sur cette photographie de foie de macaque (Gx40), les cholangiocytes sont fortement marqués (mis en évidence par les flèches rouges), tandis que les cellules présentes sur le reste de l'échantillon le sont très peu voire pas du tout.

- *Étude de l'influence des conditions analytiques sur la proportion de cellules marquées*

Nous avons étudié une potentielle influence des conditions analytiques sur la proportion de cellules marquées dans les échantillons. Pour cela, nous avons comparé le marquage de six prélèvements selon différentes conditions analytiques détaillées dans le paragraphe D de la rubrique « Matériels et méthodes ». Les résultats du dénombrement du nombre de cellules marquées en fonction des conditions analytiques sont récapitulés dans la Figure 19.

Figure 19 : Pourcentages de cellules marquées en fonction des conditions analytiques



L'histogramme ci-dessus représente les quatre moyennes de cellules marquées en fonction des quatre différentes conditions analytiques, par animal.

Afin d'étudier la présence d'une association statistique entre la proportion de cellules marquées dans un échantillon et les conditions analytiques de l'expérience, nous avons tout d'abord déterminé si nos données suivaient une distribution normale. Pour cela, nous avons réalisé le test statistique de Shapiro-Wilk. Notre population cible est représentée par l'ensemble des tissus de grands primates marqués par la sonde *Alu* en hybridation *in situ*. Nous avons défini l'hypothèse H_0 telle que : « dans notre population cible, la proportion de cellules marquées suit une distribution normale ». Nous avons réalisé le test de Shapiro-Wilk sur le site internet « StatsKingdom », et nous avons obtenu une p-value égale à $6,64 \times 10^{-10}$. Cette p-value étant inférieure à 0,05, nous avons rejeté l'hypothèse H_0 . Par conséquent, nous faisons l'hypothèse que, sous réserve d'absence de biais d'association, nos données ne suivaient pas une répartition normale dans notre échantillon.

Ensuite, nous avons vérifié la présence d'une association statistique entre le pourcentage de cellules marquées d'un échantillon et les différentes conditions analytiques à l'aide d'un test statistique. Pour cela, nous avons réalisé plusieurs moyennes du nombre de cellules marquées sur un même échantillon selon différentes conditions analytiques. Nous avons ensuite cherché un test statistique permettant la comparaison de deux médianes d'un même échantillon traité sous deux conditions analytiques différentes, sachant que la distribution des moyennes des cellules marquées ne suit pas de distribution normale, et nous avons réalisé le test de Wilcoxon. Notre population cible est représentée par l'ensemble des tissus de grands primates marqués par la sonde *Alu* par immunohistochimie. Nous avons défini l'hypothèse nulle H_0 telle que : « dans la population cible, la médiane des pourcentages de cellules marquées dans le traitement standard est égale à la médiane des pourcentages de cellules marquées sous différentes conditions analytiques ». Nous avons réalisé le test pour comparer le test standard à chaque nouvelle condition analytique. Les résultats de chaque p-value se trouvent dans le Tableau 4.

Tableau 4 : Résultats du test de Wilcoxon concernant les conditions analytiques

Condition analytique	p-value
CC2UC	$p = 0,036$
CC2C + protéase 8 min	$p = 0,248$
CC2C + protéase 16 min	$p = 0,115$

Résultats du test de Wilcoxon ayant permis de vérifier la présence ou absence de différence de marquage significative entre le test standard et chacune des autres conditions analytiques.

Nous avons constaté que la p-value du test statistique comparant le marquage en conditions standard avec le marquage en condition CC2UC est inférieure à 0,05. Par conséquent, nous rejetons l'hypothèse H_0 au risque d'erreur α à 5 %. Nous en concluons que, dans notre étude, la médiane de cellules marquées avec les conditions standards était significativement différente de la médiane de cellules marquées avec la condition CC2UC. Sous l'hypothèse d'absence de biais d'association, il y a de grandes chances pour que, dans la population cible, il existe une association réelle entre la médiane des cellules marquées et les conditions analytiques.

En ce qui concerne les autres p-values, elles sont toutes les deux supérieures à 0,05. Par conséquent, nous acceptons l'hypothèse H_0 au risque d'erreur β inconnue pour chacune des conditions analytiques. Nous en concluons que, dans notre échantillon, il n'y avait pas de différence de marquage significative entre ces différentes conditions analytiques. Sous l'hypothèse d'absence de biais d'association, dans notre population cible, il y a des chances pour qu'il n'existe pas de différence de marquage réelle entre les différentes conditions analytiques, mais nous ne pouvons absolument pas en être certains.

B. Chez les espèces non primates

a. Description de l'échantillon

Des prélèvements de huit espèces ont été testés : le rat (*Rattus norvegicus*), le lapin (*Oryctolagus cuniculus*), le chien (*Canis lupus familiaris*), le porc (*Sus scrofa domesticus*), le cheval (*Equus caballus*), le bovin (*Bos taurus*), le wallaby (*Macropus rufogriseus*) et une espèce d'ours (*Ursidae*). La plupart des échantillons des espèces domestiques, soit le rat, le chien, le porc, le cheval et le bovin, sont des prélèvements plutôt récents, fixés après 2020, tandis que les autres échantillons sont plus anciens, prélevés entre 2011 et 2015 pour le wallaby et l'ours. Le Tableau 5 récapitule les organes testés pour chaque espèce.

Tableau 5 : Répartition des organes testés par espèce non primate

Espèces	Organes
<i>Rattus norvegicus</i>	Amygdale
<i>Oryctolagus cuniculus</i>	Rate et nœud lymphatique
<i>Canis lupus familiaris</i>	Thymus
<i>Sus scrofa domesticus</i>	Rate
<i>Equus caballus</i>	Rate
<i>Bos taurus</i>	Peau
<i>Macropus rufogriseus</i>	Rate, foie, cœur et poumon
Ursidé	Rate et foie

Certaines expériences ont été réalisées sur plusieurs tissus d'un même individu lorsque le bloc comprenait plusieurs organes.

b. Evaluation du marquage de la sonde Alu

A l'examen microscopique, la totalité des tissus des différentes espèces n'a montré aucun marquage.

4. Discussion

A. Limites de l'étude

Les primates sont des espèces très protégées, ce qui rend l'accès à leurs prélèvements difficile. Dans le temps qui nous était imparti pour la réalisation de cette étude, nous avons été contraints de mener une étude rétrospective et non prospective, ce qui a augmenté le risque de biais.

a. Âge des différents prélèvements

Le choix de mener une étude rétrospective a engendré une disparité notable concernant les dates de prélèvements de nos différents échantillons, et ce pour les deux études (primates et non primates). Nous supposons que l'hétérogénéité des âges des différents prélèvements est une limite importante quant à l'interprétation de la proportion de cellules marquées. Nous étudierons ce point un peu plus loin dans la discussion. Par ailleurs, pour nos espèces non primates, l'âge précis n'était pas connu pour de nombreux prélèvements, ce qui limite les analyses et la discussion.

b. Choix des prélèvements

Le fait de mener une étude rétrospective a limité les prélèvements disponibles utilisables pour notre étude. Nous n'avons pas d'uniformité des organes testés, ce qui a pu engendrer des biais. De plus, nous n'avons pas non plus l'espèce précise pour de nombreux prélèvements de primates car les informations n'étaient pas disponibles et très difficilement retrouvables rétrospectivement.

c. Méthode de dénombrement

Le dénombrement des cellules marquées a pu être biaisé par du biais de mesure, expliqué notamment par les différents points ci-dessous.

- *Subjectivité de lecture*

Il était parfois difficile de déterminer si un noyau était coloré par la DAB ou par l'hématoxyline. De ce fait, la lecture des lames pouvait admettre une certaine subjectivité notamment en cas de marquage faible. Afin de limiter ce biais, nous avons choisi de lire les lames dans les champs où la coloration des noyaux apparaissait la plus univoque.

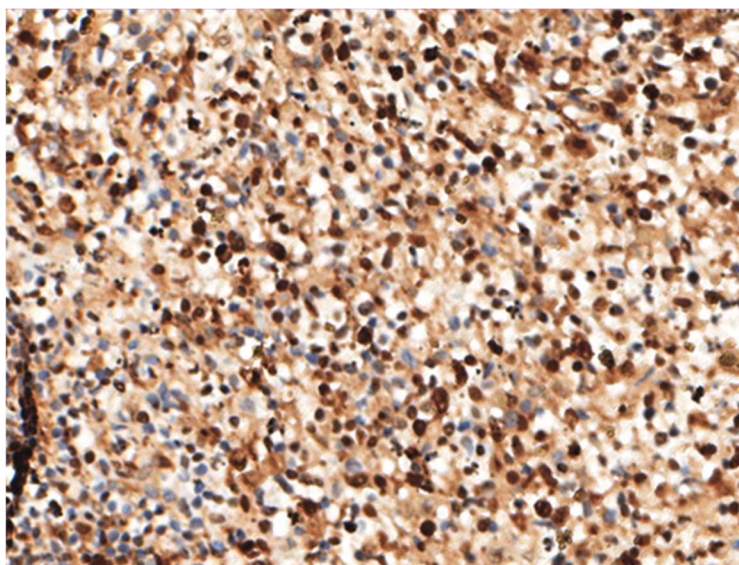
- *Marquage aberrant*

Nous avons eu plusieurs cas de marquage aberrant sur nos lames, c'est-à-dire un marquage cytoplasmique à la place d'un marquage nucléaire. L'origine de cette coloration cytoplasmique n'est pas connue. Ne sachant pas si ce marquage était significatif, nous avons décidé de ne pas les comptabiliser dans le dénombrement afin d'écarter toute erreur de classification.

- *Pigments*

Certains foies et rates possédaient des pigments et bruits de fond pouvant gêner la lecture des lames, ce qui a pu engendrer des incertitudes de dénombrement (Figure 20). Pour éviter ce biais, nous avons tenté de lire certaines lames dans des champs dépourvus de pigments, mais cela n'était pas possible pour tous les échantillons.

Figure 20 : Exemple d'un cas avec interprétation difficile du marquage



Photographie de rate de macaque (Gx40). Un fort bruit de fond rend difficile la lecture du marquage.

Par ailleurs, la non-inclusion des cellules non interprétables a pu engendrer du biais d'échantillonnage, mais nous considérons que ce biais est négligeable par rapport aux moyennes estimées dans les différents échantillons.

B. Discussion des résultats

a. Interprétation du marquage partiel

Compte-tenu de l'uniformité du matériel génétique dans les cellules d'un même individu, la présence d'un marquage partiel des cellules (présence de noyaux marqués et de noyaux non marqués) sur la presque totalité des échantillons de primates a été un résultat inattendu dont l'interprétation a été difficile. Pour expliquer ces observations, notre principale hypothèse est que ce marquage partiel représente un artefact reflétant une perte d'ADN dans les tissus. Nous émettons plusieurs théories quant à l'origine de cette perte : soit l'ADN contenu dans les noyaux a été altéré à la suite d'une mauvaise fixation des tissus, soit les échantillons ont subi une perte d'ADN avec le temps, soit les conditions analytiques n'étaient pas optimales sur ces échantillons ce qui aurait été à l'origine d'une perte d'ADN par éclatement des noyaux.

- ***Hypothèses en faveur d'un marquage partiel d'origine artificielle***

Premièrement, n'ayant pas eu accès aux protocoles de fixation des différents tissus de notre étude, nous émettons l'hypothèse que le marquage partiel des tissus pourrait être dû à une mauvaise fixation des prélèvements. Cette mauvaise fixation aboutirait à une perte d'ADN par altération des noyaux, et donc à un marquage faussement négatif. Cette altération pourrait toucher les cellules les plus fragiles en priorité, comme le suggère la différence claire de marquage entre les hépatocytes et les cholangiocytes présents sur les échantillons de foie, comme souligné par la figure 18. Les hépatocytes sont des cellules très actives qui conservent leur ADN sous forme de chromatine plutôt décondensée, à l'inverse des cholangiocytes qui sont des cellules moins actives avec un noyau plus dense et riche en hétérochromatine. Il est possible que l'ADN des hépatocytes soit ainsi davantage exposé aux dégradations du matériel génétique, ce qui pourrait expliquer le marquage partiel de ce type cellulaire. Une mauvaise fixation et/ou conservation des tissus pourrait alors expliquer en partie la présence de marquage partiel dans nos expériences. Pour valider cette hypothèse, nous pourrions comparer le marquage d'un foie fixé de différentes manières, et étudier

la différence de marquage entre les hépatocytes et cholangiocytes. Par ailleurs, nous pouvons également ajouter qu'à l'inverse, une surfixation des tissus aurait également pu aboutir à un pontage entre les protéines, et par conséquent à un marquage partiel.

Deuxièmement, nous suggérons que l'âge avancé des prélèvements pourrait conduire à une perte d'ADN et donc à un marquage partiel. Une des explications possibles serait que, malgré la fixation au formol à 4 %, l'ADN contenu dans les noyaux pourrait s'altérer avec le temps, et cette altération pourrait engendrer des marquages faussement négatifs de certains noyaux. A cause du faible nombre d'échantillons disponibles, nous n'avons pas pu étudier statistiquement cette hypothèse. Cependant, nous avons tenté de dégager une tendance globale concernant nos résultats de la figure 16. Nous pouvons constater que, d'une part, les échantillons marqués à plus de 80 % ont tous été fixés en 2021 ou plus récemment et que, d'autre part, le seul échantillon totalement négatif était le tissu le plus ancien parmi. Cependant, nous pouvons aussi remarquer que l'échantillon ayant montré le marquage partiel le plus faible datait de 2020, et que parmi les tissus fixés entre 2021 et 2024, trois d'entre eux ont montré un faible marquage à moins de 40 %. De façon subjective, nous avons une tendance pour une corrélation positive entre le pourcentage de cellules marquées au sein d'un échantillon et l'année de fixation du tissu. Par conséquent, nous supposons que plus l'échantillon serait âgé, plus le risque de marquage faussement négatif serait élevé. Cette corrélation n'a pas pu être démontrée statistiquement, mais nous conservons cette hypothèse dans les explications probables de la présence de marquage partiel. Idéalement, cette hypothèse pourrait être confirmée si nous pouvions comparer le marquage de plusieurs échantillons provenant d'une même espèce mais d'âges différents.

Troisièmement, après discussion avec le fabricant de l'automate d'hybridation *in situ*, nous avons émis l'hypothèse que les conditions analytiques de l'expérience auraient pu être à l'origine d'un éclatement des cellules, et donc d'un marquage partiel par perte d'ADN. Afin d'étudier cette hypothèse, nous avons contacté le laboratoire pour obtenir de nouveaux protocoles moins agressifs pour les cellules, et nous avons comparé les résultats obtenus selon chaque protocole. Le test de Wilcoxon n'a montré aucune différence significative de marquage entre les différentes conditions analytiques, sauf pour la condition CC2UC qui a engendré une dégradation du marquage nucléaire. Aucune condition analytique n'a permis d'amélioration significative du marquage des cellules dans notre échantillon. Nous faisons alors l'hypothèse que les conditions analytiques standards ne sont pas à l'origine du marquage partiel des tissus dans notre expérience. Cependant, il serait intéressant de réaliser de nouveau cette expérience avec des temps de protéase très augmentés afin d'étudier si cette nouvelle condition pourrait avoir un impact sur le marquage partiel (notamment pour l'hypothèse de la surfixation des tissus).

La dernière hypothèse quant à la présence de ce marquage partiel serait qu'il ne représente pas un artefact mais un patron particulier aux tissus de primates. Même si cette hypothèse doit être évoquée, elle semble peu vraisemblable compte-tenu de l'uniformité du matériel génétique au sein d'un même individu.

- *Marquage partiel concernant les témoins positifs de l'expérience*

Sur les lames du témoin positif (xénogreffes), nous avons également observé un marquage partiel que nous pouvons interpréter différemment. En effet, d'après la disposition et la morphologie des cellules non marquées, il apparaît que les cellules non marquées par la DAB étaient des cellules stromales murines au sein de la xénogreffe humaine.

- *Conclusion sur l'interprétation du marquage partiel*

Au final, nous considérons comme hypothèse la plus probable que le marquage partiel correspond à un artefact traduisant une perte d'ADN dans certaines cellules. Dans ce contexte, l'observation de cellules présentant un marquage nucléaire dans un tissu, quelle qu'en soit la proportion, indiquerait la présence de la séquence ADN complémentaire de la sonde *Alu* dans le génome de l'individu testé.

Afin de confirmer cette hypothèse, nous avons tenté d'obtenir la séquence de la sonde commerciale afin de réaliser des alignements entre sa séquence et le génome publié pour chaque espèce. Malheureusement, pour des raisons de confidentialité, nous n'avons pu avoir accès à cette séquence.

b. Interprétation chez les espèces primates

- *Pour tous les primates hormis le bonobo*

Toutes les espèces testées ont montré un marquage partiel compris entre 18 et 95 %. Avec nos hypothèses précédentes, nous considérons que le marquage est positif pour la séquence de la sonde commerciale utilisée. Par conséquent, d'après notre étude, le gorille, le chimpanzé, l'orang-outan, le babouin, le mandrill, le colobe, le macaque, le gibbon et le gélada possèderaient des séquences *Alu* complémentaires de la sonde commerciale utilisée.

- *Pour le bonobo*

Les deux tissus de l'unique bonobo testé n'ont montré aucun marquage spécifique par la sonde. Nous avons émis plusieurs hypothèses face à ce résultat.

La première interprétation serait l'absence de la séquence *Alu* complémentaire de la sonde au sein du génome du bonobo. Cette hypothèse impliquerait une perte de cette séquence chez cette espèce, qui aurait eu lieu postérieurement à la divergence avec le chimpanzé.

La deuxième hypothèse serait l'absence de cette séquence chez cet individu particulier. Le bonobo est la seule espèce pour laquelle nous n'avons obtenu qu'un seul prélèvement, donc nous avons testé cette sonde sur un seul individu de l'espèce. Même si cela reste très improbable, nous émettons l'hypothèse que cet individu particulier ait subi une perte ou une modification de la séquence, et que la détection soit alors revenue négative. Cela suggérerait que, chez le bonobo, la séquence *Alu* complémentaire de la sonde, serait présente en très peu voire en un seul exemplaire, car la perte de nombreuses séquences de façon concomitante est improbable.

La dernière hypothèse serait une perte d'ADN par altération de l'échantillon, possiblement due à une mauvaise conservation. En effet, l'échantillon provenant du bonobo était le plus vieil échantillon de notre étude (2015). Si l'ADN contenu dans ses noyaux s'est altéré au cours du temps à cause d'une mauvaise fixation, les résultats de la détection de la séquence *Alu* ont pu être faussement négatifs.

Il serait intéressant de renouveler l'expérience avec des tissus récents de bonobo de plusieurs individus afin de statuer sur la présence ou l'absence de la séquence *Alu* testée.

c. Interprétation chez les espèces non primates

Pour tous les échantillons testés, aucun marquage n'a été observé. Nous considérons que l'absence de marquage est fiable et interprétable pour les échantillons les plus récents, soit le rat, le chien, le porc, le cheval et le bovin. Pour les autres prélèvements, la question de la qualité du matériel testé en lien avec la fixation et la conservation peut se poser. Ces échantillons ont néanmoins tous été fixés dans les mêmes conditions par le service d'anatomocytologie de l'EnvA. Nous considérons donc que l'absence de marquage est interprétable pour l'ours et le wallaby. Au final, nous concluons que le rat, le lapin, le chien, le porc, le cheval, le bovin, le wallaby et l'ours ne possèdent pas la séquence *Alu* complémentaire de la sonde utilisée. Pour ces espèces, ces résultats sont concordants avec les résultats de la première partie, qui indiquent que les séquences *Alu* sont spécifiques des primates.

d. Les séquences Alu : un marqueur utilisable en médecine légale ?

Des résultats de notre étude, il apparaît que la sonde *Alu* commerciale utilisée reconnaisse, en plus des tissus humains, les tissus de tous les grands primates étudiés à l'exception du bonobo pour lequel nous ne pouvons pas conclure. Dans ce contexte, elle n'apparaît pas comme un outil fiable pour distinguer des tissus humains de tissus d'autres primates, notamment dans un contexte médico-légal.

Pour conforter ces conclusions, ce travail préliminaire devrait être complété par des études prospectives sur des tissus de grands primates, dont des bonobos, avec des conditions de fixation homogènes. Ces expériences pourraient par ailleurs répondre à nos questions sur le marquage partiel observé sur les tissus de primates : artéfact ou patron particulier aux primates ?

Conclusion

Cette étude rétrospective a montré un marquage positif, bien que partiel, par une sonde *Alu* commerciale chez neuf espèces de grands primates : le gorille, le chimpanzé, l'orang-outan, le babouin, le mandrill, le colobe, le macaque, le gibbon et le gélada. Pour ces espèces, ce marquage est en faveur de la présence de séquences *Alu* complémentaires de la sonde au sein de leur génome. Aucun marquage n'a été mis en évidence chez l'unique bonobo étudié, mais plusieurs facteurs limitants ont pu engendrer un faux négatif : ce prélèvement était le plus anciens de nos échantillons et les conditions de fixation et de conservation des tissus étaient inconnues. Par ailleurs, un seul individu a été testé contre deux individus pour les autres espèces. Il est donc à ce stade difficile de conclure sur la reconnaissance des tissus de bonobo par la sonde étudiée et sur la présence de séquences complémentaires dans le génome du bonobo.

En ce qui concerne notre questionnement initial sur l'utilité de la sonde commerciale *Alu* en anatomopathologie, et plus particulièrement dans un contexte médico-légal pour l'identification de tissus humains, il semble que cette sonde ne soit pas spécifique et qu'elle ne permette pas par conséquent de distinguer un tissu humain d'un tissu provenant d'une espèce de grands primates.

En ce qui concerne les espèces non primates, et conformément aux données de la littérature indiquant que les séquences *Alu* sont spécifiques des primates, aucun marquage n'a été détecté dans les espèces testées, soit le rat, le lapin, le porc, le chien, le cheval, le bovin, le wallaby, et l'ours.

Par ailleurs, notre étude a montré qu'il était préférable d'utiliser des prélèvements les plus récents possible, et qu'il n'y avait pas de différence de marquage significative entre les différents organes testés. De plus, les différents protocoles testés n'ont montré aucune différence significative de marquage dans notre échantillon, hormis la condition CC2UC qui a engendré une dégradation du marquage nucléaire.

En ce qui concerne les perspectives de cette étude, il serait intéressant de statuer sur la présence de la séquence *Alu* complémentaire de la sonde testée chez le bonobo. Pour cela, nous pourrions réaliser de nouveaux marquages sur des tissus récents de plusieurs individus de cette espèce. La plus grande difficulté sera l'obtention de tels prélèvements. Enfin, si nous parvenions à obtenir la séquence de la sonde commerciale, nous pourrions comparer nos résultats aux génomes des espèces testées afin d'étudier la correspondance entre la présence de la séquence dans un génome et le marquage des cellules par hybridation *in situ*.

Liste des références bibliographiques

- ALEMAN, C. (2000) Cis-acting influences on Alu RNA levels. *Nucleic Acids Research* vol. 28, n° 23, p. 4755-4761. [<https://doi.org/10.1093/nar/28.23.4755>]
- ALLARD, J., LI, K., LOPEZ, X.M., *et al.* (2014) Immunohistochemical Toolkit for Tracking and Quantifying Xenotransplanted Human Stem Cells. *Regenerative Medicine* vol. 9, n° 4, p. 437-452. [<https://doi.org/10.2217/rme.14.26>]
- ATHANASIADIS, G., ESTEBAN, E., VIA, M., *et al.* (2007) The X chromosome Alu insertions as a tool for human population genetics: data from European and African human groups. *European Journal of Human Genetics* vol. 15, n° 5, p. 578-583. [<https://doi.org/10.1038/sj.ejhg.5201797>]
- BAILEY, J.A., LIU, G., EICHLER, E.E. (2003) An Alu Transposition Model for the Origin and Expansion of Human Segmental Duplications. *The American Journal of Human Genetics* vol. 73, n° 4, p. 823-834. [<https://doi.org/10.1086/378594>]
- BATZER, M.A., DEININGER, P.L., HELLMANN-BLUMBERG, U., *et al.* (1996) Standardized nomenclature for Alu repeats. *Journal of Molecular Evolution* vol. 42, n° 1, p. 3-6. [<https://doi.org/10.1007/BF00163204>]
- BATZER, M.A., KILROY, G.E., RICHARD, P.E., *et al.* (1990) Structure and variability of recently inserted Alu family members. *Nucleic Acids Research* vol. 18, n° 23, p. 6793-6798. [<https://doi.org/10.1093/nar/18.23.6793>]
- BENNETT, E.A., KELLER, H., MILLS, R.E., *et al.* (2008) Active *Alu* retrotransposons in the human genome. *Genome Research* vol. 18, n° 12, p. 1875-1883. [<https://doi.org/10.1101/gr.081737.108>]
- BOVIA, F., FORNALLAZ, M., LEFFERS, H., *et al.* (1995) The SRP9/14 subunit of the signal recognition particle (SRP) is present in more than 20-fold excess over SRP in primate cells and exists primarily free but also in complex with small cytoplasmic Alu RNAs. *Molecular Biology of the Cell* vol. 6, n° 4, p. 471-484. [<https://doi.org/10.1091/mbc.6.4.471>]
- BRINI, A.T., LEE, G.M., KINET, J.P. (1993) Involvement of Alu sequences in the cell-specific regulation of transcription of the gamma chain of Fc and T cell receptors. *Journal of Biological Chemistry* vol. 268, n° 2, p. 1355-1361. [[https://doi.org/10.1016/S0021-9258\(18\)54083-7](https://doi.org/10.1016/S0021-9258(18)54083-7)]
- BRITTEN, R.J., BARON, W.F., STOUT, D.B., *et al.* (1988) Sources and evolution of human Alu repeated sequences. *Proceedings of the National Academy of Sciences* vol. 85, n° 13, p. 4770-4774. [<https://doi.org/10.1073/pnas.85.13.4770>]
- CHANG, D.-Y., NELSON, B., BILYEU, T., *et al.* (1994) A Human *Alu* RNA-Binding Protein Whose Expression Is Associated with Accumulation of Small Cytoplasmic *Alu* RNA. *Molecular and Cellular Biology* vol. 14, n° 6, p. 3949-3959. [<https://doi.org/10.1128/mcb.14.6.3949-3959.1994>]
- CHESNOKOV, I., SCHMID, C.W. (1996) Flanking sequences of an Alu source stimulate transcription in vitro by interacting with sequence-specific transcription factors. *Journal of Molecular Evolution* vol. 42, n° 1, p. 30-36. [<https://doi.org/10.1007/BF00163208>]
- COMEAX, M.S., ROY-ENGEL, A.M., HEDGES, D.J., *et al.* (2009) Diverse *cis* factors controlling *Alu* retrotransposition: What causes *Alu* elements to die? *Genome Research* vol. 19, n° 4, p. 545-555. [<https://doi.org/10.1101/gr.089789.108>]
- CORDAUX, R., BATZER, M.A. (2009) The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics* vol. 10, n° 10, p. 691-703. [<https://doi.org/10.1038/nrg2640>]
- CORDAUX, R., HEDGES, D.J., BATZER, M.A. (2004) Retrotransposition of Alu elements: how many sources? *Trends in Genetics* vol. 20, n° 10, p. 464-467. [<https://doi.org/10.1016/j.tig.2004.07.012>]
- COULTON, G.R., DE BELLEROCHE, J. (Éd.) (1992) *In Situ Hybridization: Medical Applications*. Dordrecht, Springer Netherlands. [<https://doi.org/10.1007/978-94-011-2984-8>]

- DANIELS, G.R., DEININGER, P.L. (1991) Characterization of a third major SINE family of repetitive sequences in the galago genome. *Nucleic Acids Research* vol. 19, n° 7, p. 1649-1656. [https://doi.org/10.1093/nar/19.7.1649]
- DANIELS, G.R., DEININGER, P.L. (1985) Integration site preferences of the Alu family and similar repetitive DNA sequences. *Nucleic Acids Research* vol. 13, n° 24, p. 8939-8954. [https://doi.org/10.1093/nar/13.24.8939]
- DANIELS, G.R., DEININGER, P.L. (1983) A second major class of Alu family repeated DNA sequences in a primate genome. *Nucleic Acids Research* vol. 11, n° 21, p. 7595-7610. [https://doi.org/10.1093/nar/11.21.7595]
- DECERBO, J., CARMICHAEL, G.G. (2005) SINEs point to abundant editing in the human genome. *Genome Biology* vol. 6, n° 4, p. 216. [https://doi.org/10.1186/gb-2005-6-4-216]
- DEININGER, P. (2011) Alu elements: know the SINEs. *Genome Biology* vol. 12, n° 12, p. 236. [https://doi.org/10.1186/gb-2011-12-12-236]
- DEININGER, P.L., BATZER, M.A., HUTCHISON, C.A., et al. (1992) Master genes in mammalian repetitive DNA amplification. *Trends in Genetics* vol. 8, n° 9, p. 307-311. [https://doi.org/10.1016/0168-9525(92)90262-3]
- DEWANNIEUX, M., ESNAULT, C., HEIDMANN, T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics* vol. 35, n° 1, p. 41-48. [https://doi.org/10.1038/ng1223]
- DEWANNIEUX, M., HEIDMANN, T. (2005) Role of poly(A) tail length in Alu retrotransposition. *Genomics* vol. 86, n° 3, p. 378-381. [https://doi.org/10.1016/j.ygeno.2005.05.009]
- EHTISHAM, M., WANI, F., WANI, I., et al. (2016) Fundamentals of in situ hybridization: A review. *International Research Journal of Clinical Medicine*, 1, 23-29.
- FAN, Y., YE, M.-S., ZHANG, J.-Y., et al. (2019) Chromosomal level assembly and population sequencing of the Chinese tree shrew genome. *Zoological Research*. vol. 40, n° 6, p. 506-521. [https://doi.org/10.24272/j.issn.2095-8137.2019.063]
- FUHRMAN, S.A., DEININGER, P.L., LAPORTE, P., et al. (1981) Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. *Nucleic Acids Research* vol. 9, n° 23, p. 6439-6456. [https://doi.org/10.1093/nar/9.23.6439]
- GOODMAN, M., PORTER, C.A., CZELUSNIAK, J., et al. (1998) Toward a Phylogenetic Classification of Primates Based on DNA Evidence Complemented by Fossil Evidence. *Molecular Phylogenetics and Evolution* vol. 9, n° 3, p. 585-598. [https://doi.org/10.1006/mpev.1998.0495]
- GUSSAKOVSKY, D., MCKENNA, S.A. (2021) Alu RNA and their roles in human disease states. *RNA Biology* vol. 18, n° sup2, p. 574-585. [https://doi.org/10.1080/15476286.2021.1989201]
- HALLING, K.C., WENDEL, A.J. (2009) In Situ Hybridization: Principles and Applications. In Allen, T.C., Cagle, P.T. (Éd.), *Basic Concepts of Molecular Pathology, Molecular Pathology Library*. Boston, MA, Springer US, p. 109-118. [https://doi.org/10.1007/978-0-387-89626-7_12]
- HAMMER (1994) A recent insertion of an alu element on the Y chromosome is a useful marker for human population studies. *Molecular Biology and Evolution*. [https://doi.org/10.1093/oxfordjournals.molbev.a040155]
- HAN, K., KONKEL, M.K., XING, J., et al. (2007) Mobile DNA in Old World Monkeys: A Glimpse Through the Rhesus Macaque Genome. *Science* vol. 316, n° 5822, p. 238-240. [https://doi.org/10.1126/science.1139462]
- HAN, K., XING, J., WANG, H., et al. (2005) Under the genomic radar: The Stealth model of Alu amplification. *Genome Research* vol. 15, n° 5, p. 655-664. [https://doi.org/10.1101/gr.3492605]
- HARADA, F., KATO, N. (1980) Nucleotide sequences of 4.5S RNAs associated with poly (A) - containing RNAs of mouse and hamster cells. *Nucleic Acids Research* vol. 8, n° 6, p. 1273-1286. [https://doi.org/10.1093/nar/8.6.1273]
- HAYNES, S.R., TOOMEY, T.P., LEINWAND, L., et al. (1981) The Chinese Hamster Alu -Equivalent Sequence: a Conserved Highly Repetitious, Interspersed Deoxyribonucleic Acid Sequence in Mammals Has a Structure Suggestive of a Transposable Element. *Molecular and Cellular Biology* vol. 1, n° 7, p. 573-583. [https://doi.org/10.1128/mcb.1.7.573-583.1981]

- HEDGES, D.J., CALLINAN, P.A., CORDAUX, R., *et al.* (2004) Differential *Alu* Mobilization and Polymorphism Among the Human and Chimpanzee Lineages. *Genome Research* vol. 14, n° 6, p. 1068-1075. [<https://doi.org/10.1101/gr.2530404>]
- HWU, H.R., ROBERTS, J.W., DAVIDSON, E.H., *et al.* (1986) Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. *Proceedings of the National Academy of Sciences* vol. 83, n° 11, p. 3875-3879. [<https://doi.org/10.1073/pnas.83.11.3875>]
- JAY, E., WU, R. (1976) Nucleotide sequence analysis of DNA. 28. *Arthrobacter luteus* restriction endonuclease recognition sequence and its cleavage map of SV40 DNA. *Biochemistry* vol. 15, n° 16, p. 3612-3620. [<https://doi.org/10.1021/bi00661a032>]
- JELINEK, W., LEINWAND, L. (1978) Low molecular weight RNAs hydrogen-bonded to nuclear and cytoplasmic poly(a)-terminated RNA from cultured chinese hamster ovary cells. *Cell* vol. 15, n° 1, p. 205-214. [[https://doi.org/10.1016/0092-8674\(78\)90095-8](https://doi.org/10.1016/0092-8674(78)90095-8)]
- JENSEN, E. (2014) Technical Review: *In Situ* Hybridization. *The Anatomical Record* vol. 297, n° 8, p. 1349-1353. [<https://doi.org/10.1002/ar.22944>]
- JURKA, J., MILOSAVLJEVIC, A. (1991) Reconstruction and analysis of human *Alu* genes. *Journal of Molecular Evolution* vol. 32, n° 2, p. 105-121. [<https://doi.org/10.1007/BF02515383>]
- KAPITONOV, V., JURKAL, J. (1996) The age of *Alu* subfamilies. *Journal of Molecular Evolution* vol. 42, n° 1, p. 59-65. [<https://doi.org/10.1007/BF00163212>]
- KIM, S., CHO, C.-S., HAN, K., *et al.* (2016) Structural Variation of *Alu* Element and Human Disease. *Genomics & Informatics* vol. 14, n° 3, p. 70. [<https://doi.org/10.5808/GI.2016.14.3.70>]
- KOJIMA, K.K. (2020) Hagfish genome reveals parallel evolution of 7SL RNA-derived SINES. *Mobile DNA* vol. 11, n° 1, p. 18. [<https://doi.org/10.1186/s13100-020-00210-2>]
- KOOP, B.F., MIYAMOTO, M.M., EMBURY, J.E., *et al.* (1986) Nucleotide sequence and evolution of the orangutan ϵ globin gene region and surrounding *Alu* repeats. *Journal of Molecular Evolution* vol. 24, n° 1-2, p. 94-102. [<https://doi.org/10.1007/BF02099956>]
- KORENBERG, J.R., RYKOWSKI, M.C. (1988) Human genome organization: *Alu*, LINES, and the molecular structure of metaphase chromosome bands. *Cell* vol. 53, n° 3, p. 391-400. [[https://doi.org/10.1016/0092-8674\(88\)90159-6](https://doi.org/10.1016/0092-8674(88)90159-6)]
- KRAYEV, A.S., KRAMEROV, D.A., SKRYABIN, K.G., *et al.* (1980) The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Research* vol. 8, n° 6, p. 1201-1216. [<https://doi.org/10.1093/nar/8.6.1201>]
- LABUDA, D., STRIKER, G. (1989) Sequence conservation in *Alu* evolution. *Nucleic Acids Research* vol. 17, n° 7, p. 2477-2491. [<https://doi.org/10.1093/nar/17.7.2477>]
- LANDER, E.S., LINTON, L.M., BIRREN, B., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* vol. 409, n° 6822, p. 860-921. [<https://doi.org/10.1038/35057062>]
- LEINWAND, L.A., WYDRO, R.M., NADAL-GINARD, B. (1982) Small RNA Molecules Related to the *Alu* Family of Repetitive DNA Sequences. *Molecular and Cellular Biology* vol. 2, n° 11, p. 1320-1330. [<https://doi.org/10.1128/mcb.2.11.1320-1330.1982>]
- LIU, G.E., ALKAN, C., JIANG, L., *et al.* (2009) Comparative analysis of *Alu* repeats in primate genomes. *Genome Research* vol. 19, n° 5, p. 876-885. [<https://doi.org/10.1101/gr.083972.108>]
- LOCKE, D.P., HILLIER, L.W., WARREN, W.C., *et al.* (2011) Comparative and demographic analysis of orang-utan genomes. *Nature* vol. 469, n° 7331, p. 529-533. [<https://doi.org/10.1038/nature09687>]
- LUO, Y., LU, X., XIE, H. (2014) Dynamic *Alu* Methylation during Normal Development, Aging, and Tumorigenesis. *BioMed Research International* vol. 2014, p. 1-12. [<https://doi.org/10.1155/2014/784706>]
- MAKAŁOWSKI, W., MITCHELL, G.A., LABUDA, D. (1994) *Alu* sequences in the coding regions of mRNA: a source of protein variability. *Trends in Genetics* vol. 10, n° 6, p. 188-193. [[https://doi.org/10.1016/0168-9525\(94\)90254-2](https://doi.org/10.1016/0168-9525(94)90254-2)]
- MIGHELL, A.J., MARKHAM, A.F., ROBINSON, P.A. (1997) *Alu* sequences. *FEBS Letters* vol. 417, n° 1, p. 1-5. [[https://doi.org/10.1016/S0014-5793\(97\)01259-3](https://doi.org/10.1016/S0014-5793(97)01259-3)]

- MILLS, R.E., BENNETT, E.A., ISKOW, R.C., *et al.* (2006) Recently Mobilized Transposons in the Human and Chimpanzee Genomes. *The American Journal of Human Genetics* vol. 78, n° 4, p. 671-679. [<https://doi.org/10.1086/501028>]
- MINGHETTI, P.P., DUGAICZYK, A. (1993) The emergence of new DNA repeats and the divergence of primates. *Proceedings of the National Academy of Sciences* vol. 90, n° 5, p. 1872-1876. [<https://doi.org/10.1073/pnas.90.5.1872>]
- MULLERSMAN, J.E., PFEFFER, L.M. (1995) An Alu Cassette in the Cytoplasmic Domain of an Interferon Receptor Subunit. *Journal of Interferon & Cytokine Research* vol. 15, n° 9, p. 815-817. [<https://doi.org/10.1089/jir.1995.15.815>]
- NISHIHARA, H., TERAJ, Y., OKADA, N. (2002) Characterization of Novel Alu- and tRNA-Related SINEs from the Tree Shrew and Evolutionary Implications of Their Origins. *Molecular Biology and Evolution* vol. 19, n° 11, p. 1964-1972. [<https://doi.org/10.1093/oxfordjournals.molbev.a004020>]
- NORRIS, J., FAN, D., ALEMAN, C., *et al.* (1995) Identification of a New Subclass of Alu DNA Repeats Which Can Function as Estrogen Receptor-dependent Transcriptional Enhancers. *Journal of Biological Chemistry* vol. 270, n° 39, p. 22777-22782. [<https://doi.org/10.1074/jbc.270.39.22777>]
- NOVACEK, M.J. (1992) Mammalian phylogeny: shaking the tree. *Nature* vol. 356, n° 6365, p. 121-125. [<https://doi.org/10.1038/356121a0>]
- QUENTIN, Y. (1994) Emergence of master sequences in families of retroposons derived from 7sl RNA. *Genetica* vol. 93, n° 1-3, p. 203-215. [<https://doi.org/10.1007/BF01435252>]
- QUENTIN, Y. (1992a) Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. *Nucleic Acids Research* vol. 20, n° 13, p. 3397-3401. [<https://doi.org/10.1093/nar/20.13.3397>]
- QUENTIN, Y. (1992b) Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Research* vol. 20, n° 3, p. 487-493. [<https://doi.org/10.1093/nar/20.3.487>]
- QUENTIN, Y. (1988) The Alu family developed through successive waves of fixation closely connected with primate lineage history. *Journal of Molecular Evolution* vol. 27, n° 3, p. 194-202. [<https://doi.org/10.1007/BF02100074>]
- RICHARD SHEN, M., BATZER, M.A., DEININGER, P.L. (1991) Evolution of the master Alu gene(s). *Journal of Molecular Evolution* vol. 33, n° 4, p. 311-320. [<https://doi.org/10.1007/BF02102862>]
- ROBERT, W. (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* vol. 437, n° 7055, p. 69-87. [<https://doi.org/10.1038/nature04072>]
- ROCCHI, M., STANYON, R., ARCHIDIACONO, N. (2009) Evolutionary New Centromeres in Primates. In Ugarkovic, D. (Éd.), *Centromere, Progress in Molecular and Subcellular Biology*. Berlin, Heidelberg, Springer Berlin Heidelberg, p. 103-152. [https://doi.org/10.1007/978-3-642-00182-6_5]
- ROGERS, J.H. (1985) The Origin and Evolution of Retroposons. In *International Review of Cytology*. Elsevier, p. 187-279. [[https://doi.org/10.1016/S0074-7696\(08\)61375-3](https://doi.org/10.1016/S0074-7696(08)61375-3)]
- ROWOLD, D.J., HERRERA, R.J. (2000) [No title found]. *Genetica* vol. 108, n° 1, p. 57-72. [<https://doi.org/10.1023/A:1004099605261>]
- SAWADA, I., WILLARD, C., SHEN, C.-K.J., *et al.* (1985) Evolution of alu family repeats since the divergence of human and chimpanzee. *Journal of Molecular Evolution* vol. 22, n° 4, p. 316-322. [<https://doi.org/10.1007/BF02115687>]
- SCHMID, C., MARAIA, R. (1992) Transcriptional regulation and transpositional selection of active SINE sequences. *Current Opinion in Genetics & Development* vol. 2, n° 6, p. 874-882. [[https://doi.org/10.1016/S0959-437X\(05\)80110-8](https://doi.org/10.1016/S0959-437X(05)80110-8)]
- SHABAN, S., AL-RAHIM, A., SULEIMAN, A. (2022) ALU repeat as potential molecular marker in the detection and prognosis of different cancer types: A systematic review. *Molecular and Clinical Oncology* vol. 16, n° 4, p. 86. [<https://doi.org/10.3892/mco.2022.2519>]

- SHAIKH, T.H., ROY, A.M., KIM, J., *et al.* (1997) cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *Journal of Molecular Biology* vol. 271, n° 2, p. 222-234. [https://doi.org/10.1006/jmbi.1997.1161]
- SINNETT, D., RICHER, C., DERAGON, J.-M., *et al.* (1992) Alu RNA transcripts in human embryonal carcinoma cells. *Journal of Molecular Biology* vol. 226, n° 3, p. 689-706. [https://doi.org/10.1016/0022-2836(92)90626-U]
- SMIT, A.F.A., TÓTH, G., RIGGS, A.D., *et al.* (1995) Ancestral, Mammalian-wide Subfamilies of LINE-1 Repetitive Sequences. *Journal of Molecular Biology* vol. 246, n° 3, p. 401-417. [https://doi.org/10.1006/jmbi.1994.0095]
- SOREK, R., AST, G., GRAUR, D. (2002) *Alu* -Containing Exons are Alternatively Spliced. *Genome Research* vol. 12, n° 7, p. 1060-1067. [https://doi.org/10.1101/gr.229302]
- SZMULEWICZ, M.N., NOVICK, G.E., HERRERA, R.J. (1998) Effects of *Alu* insertions on gene function. *ELECTROPHORESIS* vol. 19, n° 8-9, p. 1260-1264. [https://doi.org/10.1002/elps.1150190806]
- TANG, W., MUN, S., JOSHI, A., *et al.* (2018) Mobile elements contribute to the uniqueness of human genome with 15,000 human-specific insertions and 14 Mbp sequence increase. *DNA Research* vol. 25, n° 5, p. 521-533. [https://doi.org/10.1093/dnares/dsy022]
- TIEDGE, H., CHEN, W., BROSIUS, J. (1993) Primary structure, neural-specific expression, and dendritic location of human BC200 RNA. *The Journal of Neuroscience* vol. 13, n° 6, p. 2382-2390. [https://doi.org/10.1523/JNEUROSCI.13-06-02382.1993]
- ULLU, E., TSCHUDI, C. (1984) Alu sequences are processed 7SL RNA genes. *Nature* vol. 312, n° 5990, p. 171-172. [https://doi.org/10.1038/312171a0]
- VENIAMINOVA, N.A., VASSETZKY, N.S., KRAMEROV, D.A. (2007) B1 SINEs in different rodent families. *Genomics* vol. 89, n° 6, p. 678-686. [https://doi.org/10.1016/j.ygeno.2007.02.007]
- VESELINYOVÁ, D., MAŠLANKOVÁ, J., KALINOVÁ, K., *et al.* (2021) Selected In Situ Hybridization Methods: Principles and Application. *Molecules* vol. 26, n° 13, p. 3874. [https://doi.org/10.3390/molecules26133874]
- WANG, X., LIM, H.J., SON, A. (2014) Characterization of denaturation and renaturation of DNA for DNA hybridization. *Environmental Health and Toxicology* vol. 29, p. e2014007. [https://doi.org/10.5620/eht.2014.29.e2014007]
- WATERSON, R.H., LANDER, E.S., WILSON, R.K. (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* vol. 437, n° 7055, p. 69-87. [https://doi.org/10.1038/nature04072]
- WATSON, J.B., SUTCLIFFE, J.G. (1987) Primate Brain-Specific Cytoplasmic Transcript of the Alu Repeat Family. *Molecular and Cellular Biology* vol. 7, n° 9, p. 3324-3327. [https://doi.org/10.1128/mcb.7.9.3324-3327.1987]
- WILLARD, C., NGUYEN, H.T., SCHMID, C.W. (1987) Existence of at least three distinct Alu subfamilies. *Journal of Molecular Evolution* vol. 26, n° 3, p. 180-186. [https://doi.org/10.1007/BF02099850]
- XING, J., ZHANG, Y., HAN, K., *et al.* (2009) Mobile elements create structural variation: Analysis of a complete human genome. *Genome Research* vol. 19, n° 9, p. 1516-1526. [https://doi.org/10.1101/gr.091827.109]
- YULUG, I.G., YULUG, A., FISHER, E.M.C. (1995) The Frequency and Position of Alu Repeats in cDNAs, as Determined by Database Searching. *Genomics* vol. 27, n° 3, p. 544-548. [https://doi.org/10.1006/geno.1995.1090]
- ZHANG, X.-O., PRATT, H., WENG, Z. (2021) Investigating the Potential Roles of SINEs in the Human Genome. *Annual Review of Genomics and Human Genetics* vol. 22, n° 1, p. 199-218. [https://doi.org/10.1146/annurev-genom-111620-100736]
- ZIĘTKIEWICZ, E., RICHER, C., SINNETT, D., *et al.* (1998) Monophyletic Origin of Alu Elements in Primates. *Journal of Molecular Evolution* vol. 47, n° 2, p. 172-182. [https://doi.org/10.1007/PL00006374]

Annexe 1 : Protocole expérimental

Protocole N° 225 : ALU DAB (07/02/2023)

Procédure: Recherche ISH DAB Map XT

Module Discovery XT

Etape N°	Désignation des étapes
1	***** Début des étapes synchronisées*****
2	***** Sélectionner EZ Prep *****
3	Chauffer les Lames à 75 Deg C, et Incuber 8 Minutes
4	Dépôt EZPrep Volume Adjust (DXT)
5	Application Depar Volume Adjust (DXT)
6	Dépôt Coverslip (DXT)
7	Incuber 8 Minutes
8	Rincer la lame
9	Application Depar Volume Adjust (DXT)
10	Dépôt Coverslip (DXT)
11	Chauffer les Lames à 37 Deg C, et Incuber 4 Minutes
12	***** Sélectionner Rinçage SSC *****
13	***** Sélectionner EZ Prep *****
14	Rincer la lame
15	Jet Drain (DXT)
16	Dépôt CC2 Long (DXT)
17	Appliquer CC Coverslip (DXT)
18	Dépôt CC2 Long (DXT)
19	Appliquer CC Coverslip (DXT)
20	Chauffer les Lames à 95 Deg C, et Incuber 8 Minutes
21	Dépôt Coverslip (DXT)
22	Dépôt CC2 Moyen (DXT)
23	Chauffage des lames désactivé
24	Incuber 8 Minutes
25	Rincer la lame avec Reaction Buffer
26	Ajuster le volume sur la lame avec Reaction Buffer
27	Dépôt Coverslip (DXT)
28	Chauffer les Lames à 37 Deg C, et Incuber 4 Minutes
29	Chauffer la lame à [37 Deg C], et Incuber 4 Minutes (Prétraitement 2 Temp RB)
30	Rincer la lame avec Reaction Buffer
31	Ajuster le volume sur la lame avec Reaction Buffer
32	Déposer une goutte de [PROTEASE 3] (Enzyme), Appliquer Coverslip, et Incuber [20 Minutes]
33	Rincer la lame avec Reaction Buffer
34	Ajuster le volume sur la lame avec Reaction Buffer
35	Dépôt Coverslip (DXT)
36	Chauffer les Lames à 37 Deg C, et Incuber 4 Minutes
37	***** Sélectionner EZ Prep *****
38	Chauffage des lames désactivé
39	Rincer la lame
40	Ajuster le volume sur la lame
41	***** Sélectionner Rinçage SSC *****
42	Rincer la lame
43	Ajuster le volume sur la lame
44	Déposer une goutte de [Alu POS CTL] (Sonde), Appliquer Coverslip, et Incuber 4 Minutes
45	Chauffer les Lames à 37 Deg C, et Incuber 8 Minutes

* Une goutte correspond à une dose de réactif
Ecole National Veterinaire Alfort, 7, Avenue du General de Gaulle Maisons-Alfort
NexES v10.6

Protocole N° 225 : ALU DAB (07/02/2023)
Procédure: Recherche ISH DAB Map XT
Module Discovery XT

Etape N°	Désignation des étapes
46	Dépôt Coverslip (DXT)
47	Chauffer la lame à [85 Deg C], et Incuber [8 Minutes] (Dénaturation)
48	Chauffer la lame à [50 Deg C], et Incuber 4 Minutes (Hybridation)
49	Incuber [1 Heure] (Hybridation)
50	Rincer la lame
51	Ajuster le volume sur la lame
52	Dépôt Coverslip (DXT)
53	Rincer la lame
54	Jet Drain (DXT)
55	Rincer avec User Wash #1 [2,0X SSC]
56	Dépôt Coverslip (DXT)
57	Chauffer la lame à [45 Deg C], et Incuber [4 Minutes] (Lavage de Stringence 1)
58	Rincer la lame
59	Jet Drain (DXT)
60	Rincer avec User Wash #2 [2,0X SSC]
61	Dépôt Coverslip (DXT)
62	Chauffer la lame à [45 Deg C], et Incuber [4 Minutes] (Lavage de Stringence 2)
63	Rincer la lame
64	Ajuster le volume sur la lame
65	Dépôt Coverslip (DXT)
66	Chauffer les Lames à 37 Deg C, et Incuber 4 Minutes
67	Rincer la lame
68	Ajuster le volume sur la lame
69	Dépôt Coverslip (DXT)
70	Déposer une goutte de Inhibitor D, et incuber 4 Minutes
71	Rincer la lame avec Reaction Buffer
72	Ajuster le volume sur la lame avec Reaction Buffer
73	Dépôt Coverslip (DXT)
74	Déposer une goutte de [DETECTION 4] (Détection 2), et incuber [32 Minutes]
75	Rincer la lame avec Reaction Buffer
76	Ajuster le volume sur la lame avec Reaction Buffer
77	Dépôt Coverslip (DXT)
78	Déposer une goutte de Blocker D, et incuber 4 Minutes
79	Déposer une goutte de SA-HRP D, et incuber 16 Minutes
80	Rincer la lame avec Reaction Buffer
81	Ajuster le volume sur la lame avec Reaction Buffer
82	Dépôt Coverslip (DXT)
83	Rincer la lame avec Reaction Buffer
84	Ajuster le volume sur la lame avec Reaction Buffer
85	Dépôt Coverslip (DXT)
86	Rincer la lame avec Reaction Buffer
87	Ajuster le volume sur la lame avec Reaction Buffer
88	Dépôt Coverslip (DXT)
89	Rincer la lame avec Reaction Buffer
90	Ajuster le volume sur la lame avec Reaction Buffer

* Une goutte correspond à une dose de réactif
 Ecole National Veterinaire Alfort, 7, Avenue du General de Gaulle Maisons-Alfort
 NexES v10.6

Imprimé le 16/08/2024 10:10:50
 Page 2 sur 3

Protocole N° 225 : ALU DAB (07/02/2023)
Procédure: Recherche ISH DAB Map XT
Module Discovery XT

Etape N°	Désignation des étapes
91	Déposer une goutte de DAB D et une goutte de DAB H2O2 D, Déposer Huile de Paraffine, et incubé 8 Minutes
92	Rincer la lame avec Reaction Buffer
93	Ajuster le volume sur la lame avec Reaction Buffer
94	Déposer une goutte de Copper D, Déposer Huile de Paraffine, et incubé 4 Minutes
95	Rincer la lame avec Reaction Buffer
96	Ajuster le volume sur la lame avec Reaction Buffer
97	Dépôt Coverslip (DXT)
98	***** Sélectionner EZ Prep *****
99	Rincer la lame avec Reaction Buffer
100	Ajuster le volume sur la lame avec Reaction Buffer
101	Dépôt Coverslip (DXT)
102	Incuber 4 Minutes
103	Chauffage des lames désactivé
104	***** Stations de mixage désactivées *****
105	***** Attendre la puce d'enregistrement (Mise en attente pour la contre-coloration/nettoyage)*****
106	***** Stations de mixage activées *****
107	***** Sélectionner EZ Prep *****
108	Chauffer les Lames à 37 Deg C, et Incuber 4 Minutes
109	Rincer la lame avec Reaction Buffer
110	Ajuster le volume sur la lame avec Reaction Buffer
111	Déposer une goutte de [HEMATOXYLIN II] (Contre-coloration), Appliquer Coverslip, et incubé [4 Minutes]
112	Rincer la lame avec Reaction Buffer
113	Ajuster le volume sur la lame avec Reaction Buffer
114	Dépôt Coverslip (DXT)
115	Rincer la lame avec Reaction Buffer
116	Ajuster le volume sur la lame avec Reaction Buffer
117	Déposer une goutte de [BLUING REAGENT] (Post Contre-Coloration), Appliquer Coverslip, et incubé [4 Minutes]
118	Rincer la lame avec Reaction Buffer
119	Ajuster le volume sur la lame avec Reaction Buffer
120	Dépôt Coverslip (DXT)
121	Rincer la lame avec Reaction Buffer
122	Rincer la lame avec Reaction Buffer

* Une goutte correspond à une dose de réactif
 Ecole National Veterinaire Alfort, 7, Avenue du General de Gaulle Maisons-Alfort
 NexES v10.6

Annexe 2 : Données concernant le marquage des cellules dans les tissus

Cases	Prélèvement	Bloc	Organe	% de cellules marquées	Nombre de cellules comptées	Ecart-type	Valeur inférieure	Valeur supérieure
A1	Humain	1	Tumeur mammaire	0,94	216	0,01616	0,91	0,97
A1	Humain	2	Tumeur mammaire	0,88	233	0,02129	0,84	0,92
B1	Gorille 1	1	Rate	0,74	554	0,02167	0,70	0,78
B1	Gorille 1	2	Poumon	0,78	533	0,02287	0,74	0,82
C1	Gorille 2	1	Foie	0,34	311	0,01920	0,29	0,39
C1	Gorille 2	2	Rein	0,65	214	0,02294	0,59	0,71
D1	Chimpanzé 1	1	Foie	0,36	172	0,03660	0,29	0,43
D1	Chimpanzé 1	2	Rein	0,47	281	0,02033	0,41	0,53
E1	Chimpanzé 2	1	Placenta	0,7	578	0,01906	0,66	0,74
E1	Chimpanzé 2	2	Foie	0,83	582	0,01557	0,80	0,86
A2	Orang-Outan 1	1	Poumon	0,66	640	0,01872	0,62	0,70
A2	Orang-Outan 1	2	Rate	0,62	718	0,01811	0,58	0,66
B2	Orang-Outan 2	1	Rate	0,43	158	0,02431	0,35	0,51
B2	Orang-Outan 2	2	Poumon	0,46	211	0,02883	0,39	0,53
C2	Bonobo	1	Rate	0	234	0,00000	0,00	0,00
C2	Bonobo	2	Foie	0	297	0,00000	0,00	0,00
D2	Babouin 1	1	Rate	0,7	548	0,01958	0,66	0,74
D2	Babouin 1	2	Foie	0,6	587	0,02022	0,56	0,64
E2	Babouin 2	1	Rein	0,86	299	0,02007	0,82	0,90
E2	Babouin 2	2	Foie	0,64	250	0,03036	0,58	0,70
A3	Mandrill 1	1	Myocarde	0,59	70	0,03056	0,47	0,71
A3	Mandrill 1	2	Foie	0,52	148	0,02533	0,44	0,60
B3	Mandrill 2	1	Poumon	0,85	359	0,01885	0,81	0,89
B3	Mandrill 2	2	Utérus	0,81	489	0,01774	0,78	0,84
C3	Colobe 1	1	Rate	0,67	235	0,02195	0,61	0,73
C3	Colobe 1	2	Foie	0,64	465	0,02226	0,60	0,68
D3	Colobe 2	1	Foie	0,45	151	0,04049	0,37	0,53
D3	Colobe 2	2	Poumon	0,53	227	0,03313	0,47	0,59
E3	Macaque 1	1	Rein	0,85	454	0,01676	0,82	0,88
E3	Macaque 1	2	Foie	0,79	291	0,02388	0,74	0,84
A4	Macaque 2	1	Rate	0,82	458	0,01795	0,78	0,86
A4	Macaque 2	2	Foie	0,36	402	0,02394	0,31	0,41
B4	Gibbon 1	1	Rate	0,85	405	0,01825	0,82	0,88
B4	Gibbon 1	2	Foie	0,78	122	0,01105	0,71	0,85
C4	Gibbon 2	1	Foie	0,62	345	0,02613	0,57	0,67
C4	Gibbon 2	2	Poumon	0,76	665	0,01656	0,73	0,79
D4	Gélada 1	1	Foie	0,23	451	0,01982	0,19	0,27
D4	Gélada 1	2	Foie	0,22	456	0,01940	0,18	0,26
E4	Gélada 2	1	Foie	0,18	216	0,02614	0,13	0,23
E4	Gélada 2	2	Rate	0,62	728	0,01799	0,58	0,66

LES SÉQUENCES *ALU* : SYNTHÈSE BIBLIOGRAPHIQUE ET ÉTUDE DE LA RÉACTIVITÉ D'UNE SONDE COMMERCIALE PAR HYBRIDATION *IN SITU* SUR LES TISSUS DE PLUSIEURS ESPÈCES ANIMALES

AUTEUR : Océane JUILLARD

RÉSUMÉ :

Les séquences *Alu* sont un type de séquences répétées présentes au sein du génome des primates. Il en existe plusieurs types dont certains sont spécifiques d'une espèce. Dans un contexte médico-légal, il est parfois demandé de déterminer la nature humaine ou animale d'un organe trouvé, et l'utilisation de sondes pour hybridation *in situ* reconnaissant les séquences *Alu* peut permettre de répondre à ces questions. Idéalement, la sonde ne devrait reconnaître que les tissus humains et pas ceux d'autres espèces primates, notamment les espèces dont la taille est proche de l'espèce humaine (grands primates).

L'objectif de ce travail était 1) d'effectuer une synthèse bibliographique sur les séquences *Alu* et 2) d'étudier la réactivité d'une sonde *Alu* commerciale sur des tissus de diverses espèces de grands primates et d'autres espèces animales non primates par hybridation *in situ* afin de déterminer, notamment, si cette sonde *Alu* permettait de différencier les organes humains de ceux des grands primates.

Des prélèvements issus de dix espèces de grands singes et huit espèces non primates ont été inclus dans cette étude rétrospective. Chaque espèce était représentée par deux individus, excepté pour le bonobo et les espèces non primates où un seul individu était disponible. L'hybridation *in situ* par la sonde *Alu* a été réalisée par un automate sur des organes fixés au formol et conservés en bloc de paraffine.

Tous les tissus de primates, hormis ceux du bonobo, ont été partiellement marqués par la sonde *Alu*, avec entre 18 et 95 % de cellules positives. Les tissus des espèces non primates n'ont pas montré de marquage par la sonde *Alu*.

Cette étude a montré une réactivité partielle de la sonde *Alu* commerciale étudiée chez neuf espèces de grands primates : le gorille, le chimpanzé, l'orang-outan, le babouin, le mandrill, le colobe, le macaque, le gibbon et le gélada. L'absence de marquage sur les tissus du seul bonobo étudié est d'interprétation difficile étant donné qu'il s'agissait des prélèvements les plus anciens et que les tissus d'un seul individu étaient disponibles. L'absence de marquage chez les espèces non primates est en accord avec la littérature sur la spécificité des séquences *Alu* pour les primates. L'explication privilégiée pour la présence d'un marquage partiel chez les primates étudiés est une mauvaise fixation des tissus ou une dégradation du matériel génétique due à l'ancienneté des prélèvements testés. Dans notre étude, le marquage ne semblait pas dépendant de l'organe testé.

Il semble donc que la sonde *Alu* étudiée ne puisse être utilisée comme un marqueur spécifique des tissus humains, en particulier dans un contexte médico-légal.

MOTS CLÉS : ADN / PRIMATES / SÉQUENCES *ALU* / HYBRIDATION *IN SITU* / SONDE / GÉNÉTIQUE

JURY :

Présidente : Pr Fanny PILOT-STORCK
Directeur de thèse : Dr Édouard REYES-GOMEZ
Examinatrice : Dr Lucie CHEVALLIER

ALU SEQUENCES: LITERATURE REVIEW AND STUDY OF THE REACTIVITY OF A COMMERCIAL PROBE BY *IN SITU* HYBRIDIZATION ON TISSUES FROM SEVERAL ANIMAL

AUTHOR: Océane JUILLARD

SUMMARY:

Alu sequences are a type of repeated sequences found within the genome of primates. There are several types, some of which are species-specific. In a forensic context, it is sometimes necessary to determine whether a found organ is of human or animal origin, and the use of probes for in situ hybridization that recognize *Alu* sequences can help answer these questions. Ideally, the probe should only recognize human tissues and not those of other primate species, especially species whose size is close to that of humans (great apes).

The objectives of this work were 1) to conduct a literature review on *Alu* sequences and 2) to study the reactivity of a commercial *Alu* probe on tissues from various species of great apes and other non-primate animal species through in situ hybridization, in order to determine whether this *Alu* probe could differentiate human organs from those of great apes.

Samples from ten species of great apes and eight non-primate species were included in this retrospective study. Each species was represented by two individuals, except for the bonobo and the non-primate species, where only one individual was available. In situ hybridization using the *Alu* probe was performed automatically on formalin-fixed and paraffin-embedded organs.

All primate tissues, except those from the bonobo, were partially marked by the *Alu* probe, with between 18 % and 95 % of cells testing positive. Tissues from non primate species showed no staining by the *Alu* probe.

This study demonstrated partial reactivity of the commercial *Alu* probe in nine species of great apes: gorilla, chimpanzee, orangutan, baboon, mandrill, colobus, macaque, gibbon, and gelada. The absence of positive reaction in the tissues of the single bonobo studied is difficult to interpret, given that these were the oldest samples, and only one individual's tissues were available. The absence of staining in non primate species is consistent with the literature on the specificity of *Alu* sequences to primates. The most likely explanation for the partial staining in the primates studied is poor tissue fixation or degradation of genetic material due to the age of the samples tested. In our study, the staining did not appear to be dependent on the organ tested.

It therefore seems that the *Alu* probe studied cannot be used as a specific marker for human tissues, particularly in a forensic context.

KEYWORDS: DNA / PRIMATES / ALU SEQUENCES / IN SITU HYBRIDIZATION / PROBE / GENETICS

JURY:

Chairperson: Pr Fanny PILOT-STORCK
Thesis Director: Dr Édouard REYES-GOMEZ
Reviewer: Dr Lucie CHEVALLIER