



**HAL**  
open science

## Expression de la protéine PGBD5

Linda Beauclair

► **To cite this version:**

Linda Beauclair. Expression de la protéine PGBD5: Où? Quand? Comment?. Sciences du Vivant [q-bio]. 2020. hal-03597097

**HAL Id: hal-03597097**

**<https://ephe.hal.science/hal-03597097>**

Submitted on 4 Mar 2022

**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



École Pratique  
des Hautes Études

PSL 

**INRAE**  
la science pour la vie, l'humain, la terre

**MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE**

**ÉCOLE PRATIQUE DES HAUTES ÉTUDES**

**Science de la Vie et de la Terre**

Mémoire présenté par :

**BEAUCLAIR Linda**

Pour l'obtention du Diplôme de l'École Pratique des Hautes Études (EPHE)

**Expression de la protéine PGBD5**

**Où ? Quand ? Comment ?**

Soutenu le **jeudi 17 décembre 2020** devant le jury suivant :

- **Dr Sophie COUVÉ**, Présidente de jury
- **Dr Yves BIGOT**, Tuteur scientifique
- **Dr Daniel STOCKHOLM**, Tuteur pédagogique
- **Dr Nicolas BUISINE**, Rapporteur
- **Dr Florence ROULEUX-BONNIN**, Examinatrice

Mémoire préparé sous la direction de :

**Dr Yves Bigot**

UMR Physiologie de la Reproduction et des Comportements  
INRAE Centre Val de Loire, site de Nouzilly

**Dr Daniel STOCKHOLM**

Laboratoire Imagerie-Cytométrie  
Généthon, Evry



# Remerciements

---

Voici la partie la plus difficile à écrire à mon sens et je tiens à m'excuser d'avance pour les personnes que j'aurai pu oublier...

Je tiens tout d'abord à remercier les membres de mon jury de diplôme EPHE SVT :

**Sophie COUVÉ**, en tant que présidente du jury, **Nicolas BUISINE**, en tant que rapporteur et **Florence ROULEUX-BONNIN** en tant qu'examinatrice.

Je vous remercie d'avoir accepté d'évaluer mon travail.

Je remercie également **Daniel STOCKHOLM** d'avoir accepté d'être mon tuteur pédagogique et de m'avoir suivie et conseillée durant ces trois années.

Je remercie tout particulièrement **Yves BIGOT**, qui m'a permis de suivre cette formation diplômante qui me tenait à cœur depuis quelques années. Je le remercie pour son encadrement, ses conseils et pour la confiance qu'il m'a témoigné tout le long de ce travail. Un énorme merci pour tout le mal qu'il s'est donné pour initier le projet PGBD5 et qui grâce à sa détermination m'a permis de réaliser ce travail.

Je remercie également les autres membres de mon équipe :

**Florian GUILLOU**, qui m'a encouragée dans cette formation diplômante en participant à son financement en tant de Directeur de l'Unité Physiologie de la Reproduction et des Comportements mais aussi pour ses conseils, son aide et ses corrections lors de la rédaction du mémoire.

**Benoît PIEGU**, pour les bons moments passés ensemble lors des pauses café et des balades postprandiales.

**Laura HELOU**, avec qui j'ai partagé entre autres les joies, les doutes, les données NGS et aussi le bureau ! Merci pour toutes ces discussions aussi bien personnelles que scientifiques, pour l'analyse des données de ChIP alors que je restais émerveillée devant toutes ces lignes de commandes multicolores... Sans oublier notre mission commando pour le sauvetage de plantes !

Un grand merci à **Hugues DARDENTE**, qui a cru dès le départ dans ce projet et qui nous a fourni les différentes isoformes de PGBD5, point de départ de ce travail. Je le remercie aussi pour ses bons conseils dans la résolution de mes problèmes de manips. Je dois avouer qu'il avait raison concernant les lignées cellulaires stables inductibles... Dommage pour moi.

Je remercie également la **Team AP** pour toute sa bonne humeur et son implication dans la prévention. Merci pour tous ces bons moments partagés ensemble !

Je tiens aussi à remercier le personnel animalier de l'unité UEPAO : **Déborah, Aurélie** et **Flora**, merci d'avoir pris soin de mes petites souris.

Un merci spécial à **Peggy JARRIER-GAILLARD**, qui a toujours été là dans les bons comme les mauvais moments et avec qui j'ai pu partager tant de choses...

Je remercie toutes les personnes avec qui j'ai pu échanger au quotidien et lors des différents modules, que ce soit au niveau scientifique ou personnel. Merci pour ces petits moments agréables.

Enfin, je tiens à remercier les personnes qui me sont chères :

mes enfants **Alice** et **Maxime**, pour qui mon absence lors des modules a été difficile à supporter mais qui en sont ressortis grandis, j'espère qu'ils seront fiers de leur maman !

**Vincent** pour avoir géré le quotidien, pour le soutien dans les moments de doutes et pour l'aide apportés tout au long de cette formation et plus encore.

Sans oublier **mes parents** qui ont toujours été là, merci pour votre soutien sans faille.

*« Il arrive souvent de ne rien obtenir parce que l'on ne tente rien »*

*Jacques Deval*

# Table des matières

---

<b>Table des matières</b> .....	<b>3</b>
<b>Liste des illustrations</b> .....	<b>5</b>
<b>Introduction</b> .....	<b>12</b>
I. Organisation d'un génome Eucaryote : Exemple dugénome humain.....	12
I.1. Séquences transcrites (53 %).....	13
I.2. ADN non transcrit (47 %) .....	14
II. Les éléments transposables (ET) .....	16
II.1. Classification des ET .....	16
II.2. Place des ET dans les génomes : diversité et dynamique des ET .....	20
II.3. Impacts des ET sur les génomes et plasticité génomique .....	21
III. Présentation du transposon <i>piggyBac</i> et de ses dérivés.....	25
III.1. L'élément <i>piggyBac</i> .....	25
III.2. Les dérivés de <i>PiggyBac</i> .....	27
IV. PGBD5 : <i>piggyBac</i> derived protein gene element 5.....	29
IV.1. Rôle putatif de <i>pgbd5</i> dans le cerveau et la mémorisation .....	29
IV.2. Rôle présumé de <i>pgbd5</i> dans la fertilité mâle.....	30
IV.3. Rôle présumé de <i>pgbd5</i> en cancérologie .....	30
IV.4. Pourquoi s'intéresser à PGBD5 ? .....	31
<b>Projet d'étude</b> .....	<b>32</b>
<b>Matériels et Méthodes</b> .....	<b>33</b>
I. Matériels biologiques utilisés .....	33
I.1. Les sérums murins anti-PGBD5 d'In-Cell-Art .....	33
I.2. Les lignées cellulaires cancéreuses .....	33
I.3. La lignée de souris <i>Pgbd5<sup>tm1a</sup></i> .....	34
II. Méthodes de biologie cellulaire.....	35
II.1. Transfection transitoire .....	35
II.2. Transfection stable .....	36
II.3. Test de transposition.....	37
II.4. Sélection de clones cellulaires .....	37
II.5. Comptage des clones cellulaires .....	37
II.6. Amplification des clones cellulaires .....	38
II.7. Histologie.....	38
III. Méthodes de biologie moléculaire .....	39
III.1. Clonage.....	39
III.2. RT-PCR semi Quantitative .....	43
III.3. Dosage de la luciférase .....	44
III.4. Mise en évidence des anticorps .....	44
III.5. ChIP-seq : Immunoprécipitation de la chromatine et séquençage haut débit.....	48
III.6. LAM-PCR et séquençage haut débit .....	51
III.7. Fabrication de banques de populations de fragments pour le séquençage Illumina.....	54
IV. Analyse bio-informatique .....	55

IV.1. Analyse qualitative et nettoyage des données .....	55
IV.2. Détection des pics pour le ChIP-séquençage .....	56
IV.3. Détection des jonctions transposons / chromosomes dans les données de LAM-PCR ...	57
V. Analyse statistique .....	58
<b>Résultats.....</b>	<b>59</b>
I. Localisation de l'expression du gène <i>pgbd5</i> dans des tissus sains murins .....	59
I.1. Exploitation des données de transcriptomiques disponibles dans les bases de données....	59
I.2. Expression de l'activité de la $\beta$ -galactosidase comme marqueur de la transcription du gène <i>pgbd5</i> dans un modèle de souris génétiquement modifié.....	60
II. Caractérisation des sérums murins anti-PGBD5 .....	62
II.1. Recherche d'anticorps anti-PGBD5 dans les sérums murins et les anticorps commerciaux .....	62
II.2. Utilisation du sérum anti-PGBD5 en immunofluorescence pour localiser la protéine PGBD5 dans des cellules.....	65
II.3. Tentative de validation du sérum anti-PGBD5 par ChIP-séquençage en utilisant des cellules exprimant une V5-PGBD5.....	66
III. Détermination des sites de liaison de PGBD5 dans la chromatine par ChIP-séquençage .....	70
III.1. Sites de liaisons de PGBD5 à l'ADN dans des tissus sains murins .....	70
III.2 Sites de liaisons à l'ADN de PGBD5 dans des cellules cancéreuses humaines.....	71
IV. Détermination de l'activité de transposition de PGBD5 vis-à-vis des <i>pbles</i> humains .....	73
IV.1 Résultats des tests d'intégration aléatoire .....	73
IV.2 Résultats des tests de transposition .....	74
V. Ré-analyse des données de ChIP-séquençage.....	77
V.1 Dans les tissus sains .....	77
V.2. Dans les cellules cancéreuses .....	77
V.3. Conclusions sur les sites de fixation de la protéine PGBD5 .....	78
<b>Conclusions et discussion .....</b>	<b>79</b>
<b>Perspectives.....</b>	<b>86</b>
<b>Annexes .....</b>	<b>88</b>
<b>Références bibliographiques .....</b>	<b>90</b>

# Liste des illustrations

---

## Figures

---

<b>Figure 1</b> : Schéma de la structuration du génome humain.....	11
<b>Figure 2</b> : Schéma des principaux composants du génome humain.....	11
<b>Figure 3</b> : Schéma simplifié de la structure d'un gène eucaryote .....	12
<b>Figure 4</b> : Schéma simplifié du mode de transposition des éléments transposables de classes I et II .....	14
<b>Figure 5</b> : Classification des différents éléments transposables selon Wicker et al (2007).....	14
<b>Figure 6</b> : Structures des éléments transposables de classe I .....	15
<b>Figure 7</b> : Structures des éléments transposables de classe II .....	17
<b>Figure 8</b> : Modèle théorique et simplifié du mécanisme de transposition des ETs de classe II.....	17
<b>Figure 9</b> : Réarrangement des segments de gènes codant les chaînes H des immunoglobulines .....	21
<b>Figure 10</b> : Représentation des motifs RSS (Signal de séquence de recombinaison).....	21
<b>Figure 11</b> : Structure du transposon <i>PiggyBac</i> .....	24
<b>Figure 12</b> : Les différentes étapes de la transposition de <i>PiggyBac</i> .....	25
<b>Figure 13</b> : Séquence protéique non optimisée de PGBD5 utilisée pour l'immunisation des souris	32
<b>Figure 14</b> : Représentations de la stratégie de création de la lignée de souris <i>pgbd5</i> <sup>tm1a</sup> disponible dans le KOMP (Knockout Mouse Project) Repository.....	33
<b>Figure 15</b> : Séquence nucléotidique murine de <i>pgbd5</i> avec les amorces utilisées pour les RT-PCR nichées .....	39
<b>Figure 16</b> : Représentation des deux isoformes murines de <i>pgbd5</i> .....	39
<b>Figure 17</b> : Carte du plasmide pCS2+MT modifié.....	40
<b>Figure 18</b> : Carte schématique du plasmide pSBtet_ isoforme PGBD5 .....	40
<b>Figure 19</b> : Carte schématique du plasmide pSBtet_V5_ isoforme PGBD5 .....	41
<b>Figure 20</b> : Protocole schématisé de la fabrication de banques Illumina pour le séquençage haut	

débit .....	53
<b>Figure 21</b> : Expression du gène <i>pgbd5</i> chez la souris par analyse transcriptomique.....	58
<b>Figure 22</b> : Expression du gène <i>pgbd5</i> chez l'humain par analyse transcriptomique.....	58
<b>Figure 23</b> : Coupes sagittale de cerveau de souris .....	59
<b>Figure 24</b> : Photographie de cerveaux de souris après coloration de la $\beta$ -galactosidase.....	60
<b>Figure 25</b> : Coupes de cerveau de souris.....	60
<b>Figures 26</b> : Lobules testiculaires de souris <i>Pgbd5</i> <sub>tm1a(KOMP)</sub> - / <i>Pgbd5</i> <sub>tm1a(KOMP)</sub> + après révélation de la $\beta$ -galactosidase avec le substrat X-Gal. ....	60
<b>Figure 27</b> : Coupes histologiques de testicules de souris <i>Pgbd5</i> <sub>tm1a(KOMP)</sub> - / <i>Pgbd5</i> <sub>tm1a(KOMP)</sub> + après révélation de la $\beta$ -galactosidase avec le substrat X-Gal. ....	60
<b>Figure 28</b> : Western Blot sur extraits de cellules HeLa transfectées ou non par les différentes isoformes PGBD5.....	62
<b>Figure 29</b> : Western Blot sur extraits de cellules HeLa transfectées par pCS2_Mm523_Myc hybridé en présence des sérums murins J71. ....	62
<b>Figure 30</b> : Western Blot sur extraits de cellules HeLa transfectées par pCS2_Mm523_Myc hybridé en présence des sérums murins J42. ....	62
<b>Figure 31</b> : Western Blot sur extraits de cellules HeLa transfectées par pCS2_Mm523_Myc hybridé en présence des sérums murins SWISS 3 et SWISS 4. ....	63
<b>Figure 32</b> : Western Blot sur extraits de cellules HeLa transfectées par pCS2_Mm409_Myc. ....	63
<b>Figure 33</b> : Représentation graphique de la présence d'anticorps structuraux des différents sérums murins J71.....	63
<b>Figure 34</b> : Exemples d'observations en microscopie confocale des cellules transfectées par pCS2 seul, pCS2-PGBD5 Mm-409 ou Mm-523. ....	64
<b>Figure 35</b> : Exemple de données obtenues pour la lignée de cellules cancéreuses T98G à partir du site du National Center for Biotechnology Information (NCBI)(Geo Profil) .....	65
<b>Figure 36</b> : Exemples d'observations en microscopie confocale des 6 lignées de cellules cancéreuses après hybridation avec le sérum pré-immun ou le sérum SW. ....	65
<b>Figure 37</b> : Exemple de Western Blot d'extraits protéiques de clones de cellules HeLa_pSBtet_V5_Mm PGBD5 avec et sans induction de doxycycline à 1 $\mu$ g/mL pendant 24h. ....	67
<b>Figure 38</b> : Détection du transgène <i>pgbd5</i> par PCR sur des extraits de clones cellulaires HeLa_pSBtet_Mm409_V5 avec ou sans induction à la doxycycline et dans les différents plasmides pSBtet contenant l'isoforme Mm409.....	67
<b>Figure 39</b> : Analyse de l'expression de <i>pgbd5</i> par RT-PCR semi quantitative dans les cellules	

HeLa_pSBtet_Mm409_V5 avec ou sans induction à la doxycycline.....	67
<b>Figure 40</b> : Western blot sur extraits de cellules cancéreuses transfectées par le plasmide pSBtet_GP ou pSBtet_Mm409_V5 après induction ou non. ....	68
<b>Figure 41</b> : Western blot sur extraits de cellules cancéreuses HEK293 transfectées par le plasmide pSBtet_Mm409_V5 ou pSBtet_Mm523_V5 après induction ou non. ....	68
<b>Figure 42</b> : Comparaison de l'amplitude et de la taille en paire de bases des pics obtenus pour l'hippocampe dorsal et l'hippocampe ventral.....	70
<b>Figure 43</b> : Exemple de données obtenues pour la lignée de cellules cancéreuses HCT116 à partir du site du National Center for Biotechnology Information (NCBI) (Geo Profil) .....	70
<b>Figure 44</b> : Diagramme de Venn représentant les différentes co-localisations des pics obtenus pour la lignée cellulaire cancéreuse T98G. ....	71
<b>Figures 45</b> : Diagrammes de Venn représentant les différentes co-localisations des pics obtenus pour les lignées cellulaires cancéreuses HCT116 et G401. ....	72
<b>Figure 46</b> : Représentation d'un test d'intégration aléatoire en cellules HeLa .....	73
<b>Figure 47</b> : Représentation schématique des extrémités des <i>pbles</i> humains et de <i>ifp2</i> . ....	76

## **Tableaux**

---

<b>Tableau 1</b> : Composition des milieux de culture selon les lignées cellulaires .....	32
<b>Tableaux 2</b> : Conditions de PCR pour le génotypage des souris. ....	34
<b>Tableau 3</b> : Liste des plasmides utilisés dans les tests de transposition. ....	36
<b>Tableau 4</b> : Composition du mélange réactionnel pour la reverse transcription Omniscript.....	39
<b>Tableau 5</b> : Séquences des amorces PCR flanquées de sites de restriction appropriés pour le clonage .....	39
<b>Tableau 6</b> : Liste des enzymes de restriction avec leurs conditions d'utilisation d'après le fournisseur NEB. ....	41
<b>Tableaux 7</b> : Conditions pour la RT-PCR semi-quantitative. ....	43
<b>Tableaux 8</b> : Liste des anticorps commerciaux utilisés en Western Blot, Dot Blot et	

immunocytologie. ....	44
<b>Tableaux 9</b> : Composition des solutions pour le ChIP .....	47
<b>Tableaux 10</b> : Conditions pour les premières étapes de la LAM-PCR. ....	50
<b>Tableaux 11</b> : Conditions pour la suite de la LAM-PCR .....	51
<b>Tableau 12</b> : Séquences des amorces pour les PCR I et PCR II.....	53
<b>Tableau 13</b> : Dosage de la bioluminescence de la luciférase émis par les différents clones HeLa et HeLa_pSBtet_GP après induction ou non pendant 3h de doxycycline 1 µg/mL. ....	66
<b>Tableau 14</b> : Résultats du ChIP-séquençage sur cellules HEK293 transfectées par le plasmide pSBtet_Mm523_V5 après immunoprécipitation par l'anticorps anti_V5 et le sérum SW.....	68
<b>Tableau 15</b> : Résultats du ChIP-séquençage sur tissus sains de souris. ....	70
<b>Tableau 16</b> : Résultats du ChIP-séquençage à partir de différentes lignées cancéreuses.....	71
<b>Tableau 17</b> : Tableau récapitulatif du nombre d'événements de transposition (jonction transposon / chromosome) mis en évidence pour chaque <i>pble</i> en présence de la transposase PGBD5 Mm523.....	75
<b>Tableau 18</b> : Tableau récapitulatif de la conservation des séquences TIR (Terminal Inverted Repeat) et TSD (Target Site Duplication) des transposons <i>pbles</i> après événement de transposition Mm523 PGBD5. ....	75
<b>Tableau 19</b> : Tableau récapitulatif de ré-analyse des données de ChIP-séquençage pour PGBD5 obtenus à partir de tissus sains (HD, HV et cellules Leydig) et à partir de lignées de cellules cancéreuses (T98G, HCT116 et G401).....	77

## **Graphiques**

---

<b>Graphique 1</b> : Résultats du nombre de clones cellulaires obtenus lors du test d'intégration de la cassette de résistance NeoR en présence des transposases (Mm409 et Mm523) ou sans transposase (GFP). ....	73
<b>Graphique 2</b> : Résultats du nombre de clones cellulaires obtenus lors du test de transposition du transposon <i>ifp2</i> -NeoR en présence des transposases (Mm409 et Mm523) ou sans transposase (GFP). ....	74
<b>Graphique 3</b> : Résultats du nombre de clones cellulaires obtenus lors du test de transposition des	

*pbls* en présence des transposases (Mm409 et Mm523) ou sans transposase (GFP)..... 74

## **Annexes**

---

**Annexe 1** : Protocole de transfection par l'ICAFectin441 recommandé par le fournisseur In-Cell-Art pour la transfection de cellules adhérentes.....36

**Annexe 2** : Protocole de transfection avec le JetPEI recommandé par le fournisseur PolyPlus pour la transfection de cellules adhérentes ..... 37

# Liste des abréviations

---

**Aa** : Acide Aminé

**ADN** : Acide DésoxyriboNucléique (ou DNA en anglais)

**ADNc** : ADN complémentaire

**ARN** : Acide RiboNucléique

**ARNm** : Acide RiboNucléique messenger

**BWA** : Burrows-Wheeler Aligner

**CDB** : Chromatine Dilution Buffer (Tampon de dilution pour la chromatine)

**ChIP-seq** : Chromatin ImmunoPrecipitation Sequencing (immunoprécipitation de la chromatine puis séquençage haut débit).

**DMEM** : Dulbecco's Modified Eagle Medium

**EOPS** : Exempt d'Organisme Pathogène Spécifique

**ES (cellules)** : Embryonic Stem (cellules souches embryonnaires)

**ET** : Élément Transposable

**FELASA** : Federation of European Laboratory Animal Science Associations

**Flp** : Flippase recombinase

**GFP** : Green Fluorescent Protein

**HD** : Hippocampe Dorsal

**HMG** : High Mobility Group

**HV** : Hippocampe Ventral

**HPG (axe)** : Hypothalamo-Hypophyso-Gonadique

**IFP2** : Impossible Forming Polyhedra (anciennement TED : Transposon Ellen Dempsey)

**IgG** : immunoglobuline G

**IGV** : Integrative Genomics Viewer

**IR** : Internal Repeat (répétition interne)

**ITR** : Inverted Terminal Repeat (= TIR)

**Kb** : Kiloases

**kDa** : kiloDalton

**LAM-PCR** : Linear - amplification mediated Polymerase Chain Reaction

**LB** : Lysogeny Broth

**LINE** : Long Interspersed Nuclear Elements (Long élément nucléaire intercalé)

**LTR** : Long Terminal Repeats (séquences terminales longues répétées)

**MACS** : Model-based Analysis of ChIP-Seq

**MER** : Medium Reiterated Frequency Repeats  
**MSC** : Multi Site de Clonage  
**NeoR** : Néomycine Résistant  
**ORF** : Open Reading Frame (cadre de lecture ouvert)  
**PAF** : ParaFormaldéhyde  
**Pb** : paire de bases  
**PB** : PiggyBac  
**PBLE** : PiggyBac-Like Element  
**PCR** : Polymerase Chain Reaction (amplification en chaine par enzyme polymérase)  
**PePr** : Peak calling Prioritization pipeline  
**PGBD** : PiggyBac Transposable Element Derived 5  
**PLE** : Penelope-Like Element  
**PRC** : Physiologie de la Reproduction et des Comportements  
**RAG** : Recombinaison Activating Gene  
**RFID** : Radio Frequency IDentification  
**RSS** : Recombinaison Signal Sequence  
**RT** : Reverse Transcriptase  
**SAM** : Sequence Alignment Map  
**SB** : Sleeping Beauty (Transposase)  
**SINE** : Short Intersperced Nuclear Element (petits éléments nucléaires intercalés)  
**SNC** : Système Nerveux Central  
**STR** : Short Tandem Repeats (courte séquence répétée en tandem)  
**SV** : Strutral Variant (Variants structuraux)  
**SVF** : Sérum de Veau Fœtal  
**TAE** (tampon) : Tris Acétate EDTA  
**Tase** : Transposase (enzyme)  
**TIR** : Terminal Inversed Repeat (= ITR) (Répétition Terminale Inversée)  
**TSD** : Target Site Duplication (site cible pour la duplication)  
**UEPAO** : Unité Expérimentale de Physiologie Animale de l'Orfrasière  
**UTR** : Untranslated Transcribed Region (Région non traduite)  
**VNTR** : Variable Number of Tandem Repeats (répétition en tandem à nombre variable en français)  
**WT** : Wild Type (sauvage)  
**ZFD** : Zinc Finger Domain (Domaine à doigts de zinc)

# Introduction

---

L'ADN (Acide Désoxyribo-Nucléique) est la molécule de base de tout être vivant. Il a longtemps été considéré comme une structure inerte peu modifiable en dehors des processus de méiose (recombinaison) des cellules germinales ou lors de translocation de matériel chromosomique entre des chromosomes non homologues, c'est-à-dire n'appartenant pas à la même paire.

Dans les années 1940, Barbara McClintock remet en cause ce concept en observant des variations dans la coloration des grains de maïs. Au lieu d'être uniformément pigmentés, les grains de maïs des lignées qu'elle observait présentaient des zones de pigmentation plus intense. Les tâches variaient en taille et en couleur. En même temps, elle observe que les cellules de ces grains présentent des cassures chromosomiques et d'autres anomalies. Barbara McClintock examine la relation entre la pigmentation des grains et les changements chromosomiques et découvre que l'information génétique n'est pas stable [1]; [2]. Elle démontre que de petits fragments d'ADN sont capables de se déplacer et de se multiplier au sein du génome. Ces petites séquences d'ADN appelées « gènes sauteurs » ou « éléments transposables » sont également connues sous le nom de transposons. Elle constate également que le génome n'est pas simplement une base passive de données d'informations mais un système sensible et dynamique contenant une multitude d'éléments qui interagissent entre eux ainsi qu'avec leur environnement. Ces travaux lui ont valu de recevoir le Prix Nobel en 1983, soit près de 40 ans après la découverte des éléments génétiques mobiles.

## I. Organisation d'un génome Eucaryote : Exemple du génome humain

Le génome humain est l'un des génomes les mieux séquencés. A la fin de l'année 1988, un vaste programme est lancé avec le Projet Génome Humain (Human Genome Project) afin d'établir le séquençage complet de l'ADN du génome humain. Celui-ci prit fin le 14 avril 2003. En septembre 2003, le projet complémentaire ENCODE (Encyclopedia of DNA Elements) est lancé ayant pour but de rechercher les éléments fonctionnels présents dans l'ADN du génome humain. En septembre 2012, les principaux résultats du projet ENCODE sont publiés sous la forme de 30 articles et ont permis de dresser une carte détaillée des fonctions du génome en identifiant plus de 4 millions d'« interrupteurs » génétiques. En effet, contrairement à l'idée initiale d'ADN poubelle, le projet

ENCODE a démontré que 80,4% du génome a une fonction active dans la régulation de la production des protéines [3].

Le génome humain a une taille d'environ 3 400 Mégabases et regroupe l'ensemble de l'information génétique. Il est constitué deux types d'informations génétique selon la localisation : le génome nucléaire et le génome mitochondrial (**Figure 1**). Ce dernier est composé de 37 gènes codant pour 22 ARN de transfert, 2 ARN ribosomiaux et 13 protéines membranaires impliquées dans la production d'énergie par la mitochondrie. Le génome nucléaire est composé à 47 % d'ADN non transcrits regroupant les séquences peu répétées (3 %) et les séquences répétées (44 %). Les 53 % du génome restant correspondent aux séquences transcrites parmi lesquelles l'ADN non codant représente 51,5 % et regroupe les pseudogènes, les fragments de gènes, les introns ainsi que les ARNs de structures et fonctionnels. La faible portion du génome restante (1,5 %) correspond à l'ADN codant les ARN messagers (exons). Cette dernière portion du génome est présente dans la majorité du génome mais sa répartition est cependant hétérogène avec une faible densité d'ADN codant dans les centromères (**Figure 2**).

## *1.1. Séquences transcrites (53 %)*

### *1.1.a. ADN codant (1,5 %)*

Parmi les séquences transcrites, se trouvent les gènes. L'analyse de la séquence du génome humain permet d'estimer leurs nombres à environ 25 000 - 30 000 gènes. Il n'existe pas réellement de définition unique et consensuelle de ce qu'est un gène, il semble néanmoins raisonnable d'admettre comme définition qu'un gène est un enchaînement de nucléotides, c'est-à-dire une portion d'acide désoxyribonucléique (ADN), destiné à être transcrite en acide ribonucléique (ARN). Les gènes eucaryotes sont éminemment divers : par leur fonction, leur taille, leur structure. La plupart du temps, un gène commence par une séquence de nucléotides appelée région promotrice, dont le rôle est de permettre l'initiation mais surtout la régulation (tous les gènes ne sont pas exprimés dans toutes les cellules) de la transcription de l'ADN en ARN. Elle se situe majoritairement en 5' du premier exon et s'étend parfois dans le premier intron. Un gène eucaryote comprend également souvent des séquences introniques absentes de l'ARNm et qui régulent au minimum l'épissage, et se termine par une séquence terminatrice, qui marque la fin de la transcription (**Figure 3**).

### ***1.1.b ADN non codant (51,5 %)***

Outre les gènes codant pour des protéines, certains gènes spécifient des ARN ne codant pas des protéines, tels que les ARN ribosomaux ou les ARN de transfert ainsi que les ARN dits non codants. Ceux-ci portent l'information génétique, sans nécessité que celle-ci soit traduite. Ces derniers participent entre autre aux fonctions de régulation de la transcription, de réplication, de maturation, de modification, de translocation et de stabilité des ARN et/ou de translocation et dégradation des protéines [4]. Les gènes sont en majorité uniques ou potentiellement dupliqués en tandem afin de permettre une production en grandes quantités de matériel. C'est le cas des histones et d'environ 3000 gènes d'ARN non codant comme les ARN ribosomaux par exemple.

Les pseudogènes quant à eux, sont des gènes non fonctionnels qui présentent des similitudes avec un ou plusieurs gènes fonctionnels, provenant le plus souvent de la duplication conforme du gène actif. Cette copie moins soumise aux contraintes de la pression de sélection des gènes évolue vers la non-fonction, en conséquence de mutations dans le promoteur, les régions de régulation ou les exons. Les pseudogènes peuvent être présents en grand nombre au sein d'un génome. Ainsi, chez l'homme, 20 805 gènes actifs et 14 638 pseudogènes ont été recensés en 2017.

L'étendue des fonctions des gènes non codants pour une protéine est donc très vaste et certaines restent encore probablement méconnues.

### ***1.2. ADN non transcrit (47 %)***

Chez l'homme, il représente environ 47 % du génome et se compose de segments de longueurs très diverses (de quelques nucléotides à quelques milliers). Cette fraction d'ADN influence le génome humain en terme de variation, de plasticité et d'évolution.

Dans le génome humain, environ 3 % de ces séquences sont des séquences uniques et les 44 % restantes sont des séquences répétées.

#### ***1.2.a. Les séquences uniques (3 %)***

En dehors des gènes, les séquences uniques non codantes non transcrites sont des séquences très conservées. Leur taux de conservation chez les mammifères est important (96 %). Elles sont généralement situées près des régions régulatrices de gènes des fonctions essentielles de la cellule (gènes du développement, ARNr...). Elles ont un rôle dans la régulation de l'expression des gènes et

dans l'appariement correct des chromatides lors de la division cellulaire.

### ***1.2.b. Les séquences répétées (44%)***

Il existe deux types de séquences répétées : les séquences répétées en tandem et les séquences répétées dispersées.

- Les séquences répétées en tandem : (séquences hautement répétées)

Ce sont des séquences hautement répétées de quelques bases à quelques centaines de bases qui se suivent en tandem (l'extrémité 3' d'une séquence est suivie par l'extrémité 5' d'une même autre séquence).

Elles sont dispersées dans le génome humain (en particulier dans l'hétérochromatine), mais distribuées de manière non aléatoire (toujours positionnées aux mêmes endroits du génome).

Il en existe 3 types :

- les séquences satellites : retrouvées principalement au niveau de l'ADN centromérique,
- les séquences mini-satellites ou VNTR (Variable Number of Tandem Repeats) sont constituées de 10 à 100 paire de bases (pb) répétées 1000 à 2000 fois. Elles sont principalement localisées dans les télomères et sont à la base des premières empreintes génétiques.
- Les séquences micro-satellites ou STR (Short Tandem Repeats) sont constituées de 1 à 8 nucléotides répétés 5 à 50 fois. Elles sont répétées un grand nombre de fois dans tout le génome d'un individu (tous les 25 à 100 kilobases). L'intérêt de ces séquences micro-satellites réside dans le polymorphisme de leur taille hyper variable d'un individu à l'autre. Cette caractéristique en fait un outil majeur en médecine pour les tests de paternité, en biologie des populations pour évaluer le lien de parenté entre individus ou lors de l'analyse de liaison entre une pathologie et un gène, permettant ainsi de localiser la région contenant le gène responsable.

- Les séquences répétées dispersées (Séquences moyennement répétées)

Ce sont des séquences répétées de 80 pb (MADE1) à 180000 pb (Teratorn) dispersées dans tout le génome. Parmi ces séquences se trouvent les éléments transposables (ET) qui sont des séquences d'ADN ayant la capacité de se déplacer dans un chromosome ou même d'un chromosome à l'autre, au sein du génome qui les héberge. Cette particularité est responsable en grande partie de la plasticité du génome, c'est-à-dire la capacité du génome à subir des réarrangements chromosomiques qui peuvent diversifier la combinaison des gènes et leur niveau d'expression.

Les ETs possèdent dans leur séquence le ou les gènes codant les enzymes nécessaires à leur mobilité encore appelée transposition. On distingue les ET autonomes qui possèdent des gènes codant des protéines fonctionnelles de transposition ce qui leur permet d'être mobile et les ET non-autonomes qui contiennent des gènes de transposition mutés ou délétés. Toutefois, ces-derniers peuvent être mobilisés par des éléments autonomes, ce processus étant appelé mobilisation en *trans*.

## II. Les éléments transposables (ET)

### II.1. Classification des ET

Les ET présentent une grande diversité de structures et de mécanismes de transposition, ce qui rend leur classification complexe. En 1989, Finnegan et al. propose une première classification des éléments transposables reposant sur leur mécanisme de transposition. Ceci a donné lieu à la répartition des ET en deux grandes classes :

- Les éléments de classe I, ou rétrotransposons, qui nécessitent un intermédiaire ARN
- Les éléments de classe II, également appelés transposons à ADN, qui transposent via un intermédiaire ADN [5] (**Figure 4**).

Devant la quantité de nouveaux ET découverts chaque année et la diversité de ces éléments, des classifications plus précises et permettant de classer tous les éléments découverts, ont été proposées. Dans les années 2000, deux nouvelles propositions de classifications universelles sont présentées [6]; [7] et proposent de classer les ET en fonction de l'organisation de leurs domaines fonctionnels mais aussi en fonction de leur site d'insertion ou TSD (Target Site Duplication) dans le génome. Ces classifications permettent notamment de classer les ET non-autonomes dont on ne connaît pas systématiquement le mécanisme de transposition. La classification de Wicker et al (**Figure 5**), repose sur l'homologie de séquence avec les ET autonomes permettant leur mobilisation en *trans*. Cependant, il faut noter que ces classifications ne tiennent pas compte de l'histoire évolutive des ET, qui pourraient être catégorisés selon la machinerie enzymatique utilisée pour se déplacer d'un locus à un autre. Dans la classification proposée par Piégu et al en 2015 [8], l'impact des convergences évolutives dans le contexte de la transposition (phénotypes de transposition similaires mais ayant des origines différentes) est mis en avant, de même que le fait, comme les virus, que les ET ne proviennent pas d'un même ancêtre commun mais de nombreuses classes d'origines différentes.

Malgré ces diverses propositions, aucun consensus n'existe à ce jour concernant la classification

des ET. La classification qui sera détaillée ici repose sur celle proposée par Wicker et al., 2007 qui est limitée aux eucaryotes et présente un niveau de complexité suffisant pour la compréhension de ce mémoire.

### ***II.1.a. ET de classe I***

Ces éléments, appelés également rétro-éléments ou rétrotransposons, ont un mécanisme de transposition répllicative qui utilise comme intermédiaire un ARN simple brin. La copie originale reste en place et une nouvelle copie est insérée à un autre emplacement dans le génome selon un mécanisme de « copier-coller ». L'ARN simple brin est rétro-transcrit en ADN double brin par une transcriptase inverse codée par le rétrotransposon puis l'ADNc est ensuite inséré dans un nouveau site dans le génome par une intégrase. De par leur mécanisme de transposition, les rétrotransposons possèdent une capacité d'invasion très importante au sein du génome.

Selon la classification publiée par Wicker et al. en 2007, les éléments de classe I sont divisés en quatre superfamilles d'éléments autonomes auxquelles s'ajoute une superfamille d'éléments non-autonomes (les SINE : Short Intersperced Nuclear Elements). La grande majorité des ET de classe I possède une structure comportant plusieurs ORF encadrées par des régions répétées de même orientation aux extrémités (**Figure 6**).

#### **Les rétrotransposons à LTR**

Ce sont des éléments qui en général ont une taille de 5 à 9 kb. A leurs extrémités sont présentes de longues séquences (de quelques centaines de bases) terminales répétées en orientation directe connues sous le nom de LTR (Long Terminal Repeat). Les rétrotransposons avec LTR ressemblent par leur cycle de réplication et leur structure aux rétrovirus. Comme ces derniers, les rétrotransposons à LTR actifs codent principalement et généralement deux protéines de type GAG et POL. Cependant, les rétrotransposons ont un cycle uniquement intracellulaire, c'est-à-dire qu'ils ne peuvent pas infecter activement la cellule voisine, comme le font les rétrovirus. Pour certains éléments comme *gypsy* chez la drosophile, il n'est pas possible de définir s'il s'agit d'un rétrotransposon à LTR ou d'un rétrovirus sans pathologie associée.

### **Les DIRS-like éléments**

Leur nom DIRS-1 (Dictyostelium Intermediate Repeat Sequence) provient de l'hôte au sein duquel le premier élément a été découvert, l'amibe, *Dictyostelium discoideum*. Ils présentent une grande diversité d'hôte mais sont particulièrement présents dans les génomes de décapodes issus de sources hydrothermales [9];[10]. Les DIRS ne possèdent pas d'intégrase mais une tyrosine recombinase (YR) et ne génèrent pas de duplication au site d'insertion. Ils se différencient des rétrotransposons à LTR par la présence de deux ITR (Inverted Terminal Repeat ou TIR) le premier étant orienté en sens inverse du deuxième.

### **Les LINE**

Les ET de la superfamille des LINE (Long Intersperced Nuclear Elements) comprennent les ET les plus abondants chez les mammifères, les *LI*, en particulier l'homme (16 % du génome) [11]. Ces ET d'une taille de l'ordre de 5 000 pb peuvent contenir des phases ouvertes de lecture sans éléments d'initiation et de contrôle de l'expression.

### **Les PLE**

Les PLE ou Penelope-Like-Element ont été découverts chez la drosophile [12]; [13]). Ces ET constituent un groupe très épars. Leur structure peut contenir une reverse transcriptase (RT) qui ressemble plus à une télomérase qu'aux RT des ET de classe I et les régions terminales peuvent être de même orientation ou de sens inverses.

### **Les SINE**

Ces éléments, non-autonomes et de petite taille, sont généralement inférieurs à 500 pb. Ils sont abondamment présents dans le génome humain où ils sont très diversifiés (~25 % du génome) [14]. Ils ne possèdent pas les gènes codant les enzymes permettant la transposition.

## ***II.1.b. ET de classe II***

Les éléments regroupés dans la classe II (transposons à ADN) étaient à l'origine ceux qui transposent grâce à un mécanisme « couper-coller », sans intermédiaire ARN avec une molécule d'ADN double comme intermédiaire de transposition. La découverte de nouveaux éléments ainsi que de nouveaux mécanismes de transposition a conduit à subdiviser cette classe en deux sous-classes en fonction des caractéristiques de leur mécanisme de transposition. La sous-classe 1 est composée des éléments ayant un mécanisme de transposition classique de type « couper-coller ». La sous-classe 2 comprend les éléments *Hélitrons* au mécanisme de cercle roulant [15] et *Maverick/Polinton* dont le processus de transposition n'est pas connu [16] (**Figure 7**).

### **La sous-classe 1 : les ET à ADN possédant des TIR (Terminal Inversed Repeats)**

Ces éléments se déplacent par un intermédiaire ADN grâce à une enzyme unique codée par l'élément : la transposase (Tase). Celle-ci reconnaît à chacune des extrémités de l'élément de courtes séquences répétées inversées de quelques dizaines à quelques centaines de paires de base (TIR ou ITR : Inverted Tandem Repeat). Lors d'une transposition, la transposase reconnaît des motifs spécifiques dans les ITR, excise l'élément de son locus puis l'insère à un autre locus chromosomique (**Figure 8**).

Les éléments appartenant à l'ordre des TIR possèdent des extrémités terminales inverses de taille variable. Ils codent pour une transposase et génère un TSD lors de leur transposition. Cette sous-classe comprend au moins 10 superfamilles d'éléments : *Tc1-mariner* [17], *hAT* [18], *Mutator* [19], *P* [20], *Merlin* [21], *Transib* [22], *piggyBac* [23], *PIF-Harbinger* [24], *CACTA* [25] et les ET de type *Crypton*. Ces derniers sont classés dans une superfamille à part, du fait de l'absence de TIR et de la présence d'un gène codant une intégrase à tyrosine à la place d'une transposase [26]. Chaque superfamille est caractérisée par sa structure, la taille de la séquence dupliquée au niveau du site d'insertion ou TSD ainsi que des motifs de la transposase DDD ou DDE qui forment le cœur du site catalytique responsable de la transposition [27]; [17].

### **La sous-classe 2 (*Hélitron*, *Maverick*)**

Les éléments *Helitron* décrits dans les génomes d'*Arabidopsis thaliana* et *Caenorhabditis elegans* où ils sont largement répandus [15] ainsi que chez les champignons [9], semblent se répliquer via un mécanisme de cercle roulant identique à celui du transposon bactérien IS91. Ils ne génèrent pas de

duplication du site d'insertion. D'une taille de 5,5 à 17 Kpb, les éléments *Helitron* ne possèdent pas de TIR. Ils codent une tyrosine recombinase similaire à celle trouvée dans des transposons IS91.

Les éléments *Maverick* ont été découverts chez le nématode *Caenorhabditis elegans* ou le poisson zèbre *Danio rerio* [16]. Ces éléments sont de longues séquences de 10 à 20 kilo paires de bases composées de TIR de 150 à 700 pb. Ils codent jusqu'à 11 protéines dont certaines de type viral et dont l'ordre des gènes peut varier suivant les éléments ([28]). Leur excision se fait sous la forme simple brin, l'élément est alors répliqué en ADN double brin puis réintégré au niveau de son site cible. D'après Pritham et al (2007) [16] cet élément pourrait être capable d'auto-réplication.

## ***II.2. Place des ET dans les génomes : diversité et dynamique des ET***

Longtemps présentés comme des parasites génétiques ou de l'ADN inutile (« ADN poubelle »), les éléments transposables sont désormais considérés comme un élément majeur de l'évolution des génomes jouant notamment un rôle dans la création de nouveaux gènes [29]. À de très rares exceptions, les ET ont été trouvés dans tous les génomes chez lesquels ils ont été recherchés. Leur proportion est cependant très variable d'une espèce à l'autre. En effet, les ET peuvent constituer la majeure partie d'un génome donné, comme chez le maïs où ils représentent jusqu'à 90 % du génome. À l'opposé, chez l'abeille, les ET ne constituent qu'une infime partie du génome et ne représentent que 1 % des séquences. Chez les primates (Homme, chimpanzé et macaque), la proportion d'ET est similaire entre espèces et représente environ 45 % du génome. Certaines superfamilles d'ET peuvent être prépondérantes dans certains génomes et il peut y avoir une disproportion pour une classe ou une autre. Par exemple, les génomes végétaux et mammifères semblent dominés par les éléments de classe I [30], [31]. Cependant, la chauve-souris, *Myotis lucifugus* présente une majorité d'ET de classe II tout comme les champignons [32]; [33].

L'évolution d'un ET dans un génome hôte suit une dynamique en trois phases qui sont successivement d'invasion, de maturité et de sénescence. La première étape est la phase d'invasion du génome par un ET actif qui peut être le résultat d'un transfert horizontal inter-population dans une espèce ou même interspécifique. Elle s'effectue d'abord lentement jusqu'à l'augmentation exponentielle du nombre de copies de l'ET actif. Ce point de départ est difficile à mettre en évidence car le nombre d'éléments actifs connus est restreint et les taux de transposition difficile à estimer pour la plupart des ET. La deuxième phase correspond à la « maturité » durant laquelle des systèmes de régulation de l'activité de transposition vont se mettre en place et l'apparition de mutations va limiter

la prolifération des éléments. D'après Kidwell et Lisch [34], au cours de l'évolution, des mécanismes de régulation limitant les effets néfastes de la transposition auraient été sélectionnés. Plusieurs types de mécanismes de régulation sont connus à l'heure actuelle : i) la régulation de la transposition par des mécanismes classiques de répression de l'hôte exerçant un rétrocontrôle, ii) la régulation de la transposition par différents mécanismes d'autorégulation dirigés par le transposon lui-même (agrégation de la Tase (*Sleeping Beauty*, *piggybac*), défaut d'épissage spécifique de certain type cellulaire (*élément P*), présence de silencer (*mariner*)). La régulation épigénétique peut également être impliquée avec la méthylation des histones [35]. L'activité des ET peut être régulée via des ARN anti-sens qui empêche spécifiquement la traduction des transcrits d'ETs [36]. Cette régulation post-transcriptionnelle bloque la production des enzymes permettant la transposition, bien que l'ET soit exprimé [37]; [38]. La phase de sénescence, quant à elle, correspond à une diminution drastique du nombre de copies de l'ET. Elle aboutit à la perte de cet ET à terme, voire de familles entières d'ET au sein d'un génome lorsque le taux de transposition est faible. Deux mécanismes conduisent à la phase de sénescence : l'inactivation verticale correspondant à l'accumulation de mutations (substitutions ou insertions/ délétions), celles-ci conduisant à la perte de fonctionnalité des ET [39]; [40] au cours des générations et la perte stochastique correspondant à des délétions de courts ou longs fragments d'ADN par dérive génétique pouvant entraîner l'élimination d'ET et donc participer à la sénescence.

### ***II.3. Impacts des ET sur les génomes et plasticité génomique***

La transposition des ET induit des modifications génomiques pouvant être dues à l'insertion et / ou à l'excision de l'élément voire simplement à la cicatrice laissée par l'excision. En effet, l'excision des transposons étant très rarement précise, quelques paires de bases peuvent rester au niveau du site d'excision et la fonction initiale du gène en ce site n'est alors pas restaurée [17]. Les modifications résultant des ET ont longtemps été classées en fonction des effets sur le génome hôte, celles-ci étant qualifiées de neutres, délétères ou bénéfiques [34]. Les mutations délétères peuvent entraîner la non viabilité de l'hôte, elles sont généralement contre sélectionnées et sont par conséquent difficilement repérables [41]. Dans les génomes eucaryotes, on observe une tendance à l'accumulation d'ET dans l'hétérochromatine plutôt que dans les séquences codantes (euchromatine). Ce fait n'est peut-être que le reflet de la contre-sélection des ET insérés dans les gènes dit « stables », la sélection génomique cantonnerait en quelque sorte les ET dans des régions où les effets seraient neutres. Toutefois, avec le développement de l'épigénétique, l'observation d'une corrélation entre la présence des ET dans l'hétérochromatine et l'hyperméthylation de cette fraction génomique, suggère que les ET accumulés

n'auraient pas un effet neutre mais un rôle régulateur de l'expression génique [38]. En ce qui concerne les effets bénéfiques, ils sont sélectionnés et par ce fait, détectables. L'ET passe d'un statut de parasite génomique à un statut de gène stable bénéfique à son hôte [42]. Les ETs se retrouvent ainsi "domestiqués" par le génome de l'hôte et sont alors à l'origine de l'apparition de nouveaux gènes fonctionnels tels que la protéine Syncytin impliquée dans la morphogenèse du placenta chez l'homme [43], la protéine SETMAR qui joue un rôle dans la réparation de l'ADN [44] ou encore le système V(D)J du système immunitaire des mammifères [45], ce dernier étant certainement le cas de domestication le plus connu.

### ***II.3.a. Domestication du transposon RAG***

Au cours des vingt dernières années, de plus en plus d'études portent sur les contributions positives des éléments transposables aux échelles de l'évolution des espèces, de la régulation des gènes et dans la création de divers gènes devenus essentiels à l'hôte. C'est ainsi que le gène *Rag* dérive de la transposase d'une ancienne famille de transposons appelée *Transib*.

La co-option des protéines RAG1 et RAG2 (RAG pour *Recombination Activating Gene*) a été un événement crucial dans l'évolution de l'immunité des vertébrés. Ces deux protéines constituent le noyau enzymatique de la machine de recombinaison V(D)J chez les vertébrés à mâchoires [46] ; [47] ; [48] et [49]. Le complexe RAG1-RAG2 catalyse l'assemblage aléatoire de segments des gènes V (pour variable), D (pour diversité), et J (pour jonction) qui sont présents dans le génome en de nombreuses copies et génèrent l'énorme variété des anticorps et des récepteurs d'antigène (**Figure 9**) [50]; [51] et [52]. Le mécanisme qui permet cette recombinaison implique deux protéines RAG1 et RAG 2, ainsi que des RSS (*Recombination Signal Sequences*) qui flanquent les segments géniques V, D et J, à l'image des ITR flanquant les gènes codant les transposases chez les éléments eucaryotes. Chaque RSS est constituée d'une séquence très conservée de 7 paires de bases ( $5' \text{CACAGTG} 3'$ ) et d'une séquence moins conservée de 9 paires de bases ( $5' \text{ACAAAACCC} 3'$ ). Ces deux séquences sont séparées par une séquence espaceur moins conservée de 12 ou 23 paires de bases (12 RSS ou 23 RSS) (**Figure 10**). Pour que la recombinaison soit efficace, il faut qu'un RSS recombine avec un autre RSS. Ces séquences RSS sont reconnues spécifiquement par le complexe enzymatique RAG1/2.

Cette réaction est apparentée à l'étape d'excision de la transposition et met en jeu la protéine RAG1 à cœur catalytique DDE. Les gènes codant les protéines RAG1/2 dérivent des éléments *Transib* ([45]; [53]; [54] et [55]). La recombinaison V(D)J implique une coupure double-brin de l'ADN et la formation d'une structure en épingle à cheveux sur l'extrémité codante.

La recombinaison V(D)J implique également des protéines cellulaires, HMG1 ou HMG2 (High Mobility Group). Le complexe composé des protéines RAG1, RAG2 et HMG1/2 fixe les RSS formant ainsi un complexe stable [56]. Après le clivage de l'ADN, les protéines demeurent associées [57]; [56]. Dans ce complexe, seule la protéine RAG1 peut reconnaître le RSS [58] mais la présence de la protéine RAG2 permet une liaison stable au substrat ainsi que le clivage enzymatique. Elle permet également d'inhiber la liaison non-spécifique de RAG1 sur l'ADN [59].

### ***II.3.b. Domestication des cibles de RAG***

Dans un système où un gène codant une transposase est domestiqué, le niveau de son expression dans la cellule va être contrôlé dans une ou plusieurs fenêtres précises de temps. Il est attendu qu'il en soit de même pour l'accessibilité de ses sites de liaison. C'est ce qu'on observe avec les transposases domestiquées RAG1 et RAG2 qui sont exprimées lors d'une courte fenêtre au cours de la lymphogenèse. C'est aussi ce qui est observé au niveau de leurs gènes cibles qui codent les IgG et les récepteurs des cellules T. En effet, ces gènes sont soumis à des réarrangements chromatinien qui ouvrent et ferment la chromatine grâce à la méthylation de la lysine 4 sur les histones H3 permettant le recrutement de RAG2 [60]. Ce recrutement de RAG2 sur les lysines méthylées va permettre le recrutement de RAG1 sur les RSS, l'activation de RAG1 et la coupure par RAG1 des RSS de façon précise dans le temps. RAG1 et RAG2 pourront ainsi recombiner ces gènes de façon séquentielle (dans le cas des IgG, la recombinaison a d'abord lieu entre les domaines V et D puis entre les domaines D et J, sans que jamais des recombinaisons entre les domaines V et J ne soient autorisées par la structure de la chromatine).

Dans les autres cellules où RAG1 et RAG2 sont exprimées à faible niveau sans fonction apparente (thyroïde et testicule) ou non exprimées, leurs gènes cibles ont été montrés être dans un état de condensation chromatinien empêchant leur accessibilité.

### ***II.3.c. Utilisation des transposons en tant qu'outil naturel de génie génétique***

La capacité naturelle des éléments transposables à se déplacer au sein des génomes et la simplicité de leurs composants en font des candidats de choix pour le développement d'outils pour le transfert de gène. Ces caractéristiques ont attiré l'attention de la communauté scientifique afin de développer des applications de biotechnologies de thérapie génique chez l'Homme et de transgénèse chez les

animaux. Le gène codant la transposase étant alors remplacé par le gène d'intérêt pour devenir un pseudo-transposon pouvant être mobilisé par l'apport en *trans* d'une source de transposase au sein des cellules. La majorité des éléments transposables des vertébrés n'est pas fonctionnelle, du moins dans les cellules somatiques. Ainsi, en l'absence de remobilisation, l'intégration d'un transposon exogène dans les chromosomes peut générer une expression à long terme voire permanente du transgène. Les éléments transposables peuvent donc être exploités en tant que vecteur de transfert de gène pour intégrer de manière stable un gène d'intérêt dans un génome.

A la fin des années 80, l'élément *P* issu de *Drosophila melanogaster*, a été le premier transposon développé comme outil de transgénèse chez ce même organisme [61], révolutionnant ainsi le domaine de la génétique. Le transfert de gène basé sur l'élément *P* a permis de grandes avancées sur la compréhension des voies métaboliques, de réparation de l'ADN et de développement, en premier lieu chez la drosophile, puis par homologie chez d'autres organismes dont les mammifères. Cependant, les généticiens se sont heurtés à l'inefficacité de cet élément transposable à transposer dans les cellules de vertébrés. L'essor des vecteurs transposons chez les mammifères s'est réellement établi en 1997, avec la reconstitution du transposon à ADN *Sleeping Beauty (SB)* de la famille des *IS630-Tc1-mariner*, "endormi" depuis plus de 20 millions d'années [62]. *SB* a été reconstruit par ingénierie moléculaire sur la base d'une séquence consensus obtenue à partir de 12 éléments fossiles *Tc1-like* issus de 8 espèces de poissons de la famille des Salmonidés. Il a été le premier transposon à ADN à être démontré efficace chez la souris et chez l'Homme [62] ; [63]. Par la suite, d'autres transposons naturellement efficaces dans les cellules de mammifères issus de différentes espèces animales ont été découverts comme *Himar1* [64], *Tol2* [65] ou encore le système *piggyBac* [66] qui a été utilisé dans le cadre de ce travail.

# III. Présentation du transposon *piggyBac* et de ses dérivés

## III.1. L'élément *piggyBac*

### III.1.a. Découverte

L'élément *piggyBac* (PB) a été découvert par M.J Fraser par une découverte originale [67]. Cette équipe a observé l'apparition spontanée de baculovirus mutants au cours de la propagation de ces virus dans une lignée cellulaire (TN-368) de la chenille arpeuteuse du chou *Trichoplusia ni*.

L'analyse de ces virus mutants a révélé que de l'ADN de la cellule hôte s'était inséré dans le génome viral entraînant la perte d'une protéine virale de 25 kDa ; la polyhedrine. Ces virus mutants pouvaient revenir au type sauvage par passage en série dans la lignée cellulaire TN-368. Ceci a été associé à la perte de l'ADN inséré dans le génome viral révertant. La taille de ces fragments insérés variait entre 0,5 et 2,8 kb. L'un des fragments les plus fréquemment insérés avait une taille de 2,5 kb et était présent sous forme de copies multiples dans le génome de *Trichoplusia ni*. Ces observations indiquaient que les insertions dans le génome viral étaient des éléments mobiles qui résidaient dans le génome de *Trichoplusia ni*. Les analyses de séquençage de l'ADN inséré ont révélé que les sites d'intégration présentaient des signatures de transposons d'ADN typiques telles que la duplication parfaite du site d'intégration de 4 pb dans l'ADN viral (TTAA) et des répétitions inversées terminales (TIR) dans les fragments insérés. Ils ont ainsi pu caractériser une nouvelle famille d'éléments transposables à ADN : la superfamille *piggyBac* [67], originellement appelée *IFP2* (Impossible forming polyhedra, second transposon (le premier étant TED)). Les éléments *PB* et *PB-like* forment une superfamille à part entière car ils ont la particularité de reconnaître et de s'intégrer dans un site TTAA dans le génome. Ils sont également retrouvés à la fois chez les insectes, les champignons, les crustacés et plusieurs vertébrés [68].

### III.1.b. Organisation de la séquence ADN de l'élément *PB*

D'une longueur de 2475 pb, *PB* est constitué d'un cadre ouvert de lecture de 2,1 kpb codant pour une transposase de 594 acides aminés d'un poids moléculaire de 64 kDa, de deux ITR de 13 pb et de

deux répétitions internes (IR) additionnelles de 19 pb. Ces répétitions sont réparties de manière asymétrique de part et d'autres du gène de la transposase : en 5', l'ITR et l'IR sont séparés de 3 pb tandis que 31 pb séparent les répétitions terminales et internes en 3' (**Figure 11**). Le transposon originel est, quant à lui, pourvu d'un promoteur actif en cellules de mammifères [69].

La protéine possède un signal de localisation nucléaire fonctionnel identifié dans la partie C-terminale [70], qui assure un ciblage efficace de la transposase au noyau, lieu de la transposition. Un domaine de type ZFD est également retrouvé dans cette partie de la protéine [71]. Il pourrait être impliqué dans la fixation de la protéine sur la séquence d'ADN cible ou dans des interactions protéines-protéines conduisant à la dimérisation de la transposase, nécessaire à la transposition. Le site catalytique comporte une triade de résidus aspartate en positions D268, D346, et D447 [72]. Ce type de motif est fréquent chez les enzymes qui clivent l'ADN.

### ***III.1.c. Mécanisme de transposition***

Le transposon *piggyBac* possède une large gamme d'hôtes, ce qui suggère que sa mobilisation pourrait être indépendante de facteurs de l'hôte.

En 2008, Craig et ses collègues ont purifié la transposase *piggyBac* et ont réalisé des réactions d'excision et d'intégration *in vitro* en utilisant la transposase purifiée et un ADN source de transposon [73]. Ils ont ainsi pu mettre en évidence, les différentes étapes de transposition de *piggyBac* (**Figure 12**).

La mobilisation des éléments se fait selon un mécanisme de « couper-coller » en quatre étapes [73]. Lorsque la transposase *piggyBac* se lie à l'extrémité du transposon, la transposition est initiée par une coupure aux extrémités 5' et 3' du transposon, exposant des 3'OH. Ces 3'OH attaquent le brin complémentaire à 4 nucléotides dans l'ADN flanquant générant ainsi une boucle à chaque extrémité du transposon qui est alors excisée. Le site donneur est réparé par complémentarité permettant de reformer précisément la séquence cible TTAA. Les boucles terminales sont ouvertes par la transposase, ce qui re-expose les 3'OH et génère des extrémités sortantes TTAA en 5' du transposon excisé. Les 3'OH vont permettre l'attaque de la séquence cible TTAA/AATT et la production de brèches simple-brin favorisant l'insertion du transposon dans le génome. Après intégration, les brèches sont réparées et conduisent à la duplication du site TTAA.

*PiggyBac* montre ainsi une excision précise et ne laisse généralement pas de cicatrice d'excision après passage des machineries de réparation de l'ADN de l'hôte. Cependant, il y a des échecs occasionnels de réparation du site donneur, qui se traduisent par de petites insertions et/ou délétions,

à une fréquence d'environ 1 % des cas d'excision pendant la transposition chromosomique dans les cellules souches embryonnaires de souris (ES) [74]. Cette propriété unique rend possible l'expression transitoire d'un transgène sans effet mutagène provoqué par l'excision de celui-ci, ce qui fait de *PB* un outil attractif pour des études génétiques aussi bien fonctionnelles que thérapeutiques.

Contrairement aux autres transposons étudiés jusqu'à maintenant, *piggyBac* n'est pas dépendant de la réplication cellulaire et de la synthèse d'ADN pour transposer [73]. De plus, aucun cofacteur de l'hôte n'a été montré être indispensable à sa transposition.

Le système transposon *PB* est également actif dans de nombreuses lignées cellulaires et espèces telles que les insectes, le poisson-zèbre, le poulet, la souris et l'Homme [75]; [76] ; [77] ; [78] ; [79] et [80].

## ***III.2. Les dérivés de PiggyBac***

### ***III.2.a. piggyBac like elements ou PBLE***

Un problème de sécurité pour les patients en thérapie génique reste associé à l'utilisation de *PB* malgré tous ces avantages. Il existe dans le génome humain environ 2 000 éléments *piggyBac-like* (*pble*) qui dérivent d'au moins trois éléments *PB* ancestraux. Ces *pble* se sont accumulés tout au long de l'histoire de l'évolution, depuis les céphalocordés. Ils ne se limitent pas au génome humain et peuvent être partagés avec d'autres génomes de mammifères, amniotes et vertébrés. Ces *pble*, qui contiennent TIR, UTR et éventuellement un ORF ou les traces d'un ORF codant une transposase, sont des éléments pouvant être capables de transposition.

D'après la base de données Repbase [81], qui répertorie les éléments transposables eucaryotes et les répétitions, il existerait 5 *pble* dans le génome humain : MER75, MER75A, MER75B, MER85 (MER : "medium reiterated frequency repeats") et *Looper*. Parmi ces éléments fossiles se distingue l'élément *Looper* (environ 500 copies) qui est un court consensus de multiples séquences divergentes présentes dans le génome humain. Le produit protéique potentiel correspond à la partie N-terminale et ne possède que 25 % de similarité avec la transposase *PB* de *Trichoplusia ni*, mettant ainsi en doute sa fonctionnalité.

### ***III.2.b. piggyBac derived protein gene element ou pgbd***

Le séquençage et l'assemblage de modèle de génome a également permis de révéler l'existence de gènes homologues à celui de la transposase *piggyBac*. Ces éléments *pgbd* (*piggyBac derived protein gene element*) contiennent une séquence dérivée de la transposase *piggyBac* et correspondent à des éléments domestiqués. Ils constituent des gènes de l'hôte à part entière.

Le génome humain contient cinq gènes dérivés d'éléments *piggyBac* domestiqués. Ils sont désignés de *pgbd1* à *pgbd5* [68]. *pgbd1* et *pgbd2* étaient probablement présents chez l'ancêtre commun des mammifères, tandis que *pgbd3* et *pgbd4* l'étaient chez l'ancêtre des primates. *pgbd5* est quant à lui, largement distribué chez les vertébrés, y compris la lamproie et l'amphoxius, suggérant une domestication ancienne concomitante ou antérieure à l'apparition des vertébrés il y a environ 525 millions d'années [68]; [82]. De tels événements de domestication ont été également identifiés chez les ciliés et le xénope (amphibien) où les éléments domestiqués *piggyMac* (*PGM*) et *TPB2* ont été identifiés chez *Paramecium tetraurelia* [83] et *Tetrahymena thermophila* [84], et *pgbd7* chez les anoues (Helou et al 2020, en révision).

Toutes les séquences domestiquées se trouvent en une seule copie dans les génomes. Chez certaines d'entre elles, la séquence codante dérivée de la transposase est fusionnée avec un gène de structure ou un domaine particulier [68].

- ***pgbd1*** est le résultat d'une fusion ancestrale entre une séquence codant une protéine dérivée d'un élément *piggyBac* codant une transposase de 519 acides aminés (aa) et une séquence codant un domaine riche en leucine de 290 aa, dans la région N-terminale.
- ***pgbd2***, correspondant à la fusion de plusieurs exons, code une protéine de taille de 592 aa dont la fonction reste inconnue.
- ***pgbd3***, code une protéine de 593 aa et correspond à une séquence insérée dans le cinquième intron du gène responsable du syndrome du groupe B de Cockayne (CSB). L'épissage alternatif de cette région conduit à la formation d'une protéine de fusion CSB-PGBD3 [85].
- ***pgbd4*** code une protéine de taille de 585 aa, de fonction inconnue.
- ***pgbd5***, contient 7 exons et plusieurs introns. Il code une protéine de taille variant de 409 à 595 aa. Ce gène semble impliqué dans des fonctions neuronales [82].

## IV. PGBD5 : piggyBac derived protein gene element 5

Bien que la majorité des gènes codant pour les enzymes transposases tendent à devenir catalytiquement inactifs et leurs substrats transposons ont tendance à devenir immobiles au cours de l'évolution de l'organisme à cause des mutations accumulées au cours du temps, ceux domestiqués sont sous pression de sélection et peuvent coder une protéine qui a conservé ses activités de transposase [86].

Différents transposons de la superfamille *piggyBac* ont colonisé un large éventail d'organismes [68]. *pgbd5* se distingue des autres gènes humains dérivés de *piggyBac* par le fait qu'il a été domestiqué très tôt dans l'évolution des vertébrés, avant RAG1/RAG2, il y a environ 525 millions d'années chez l'ancêtre commun des céphalocordés et vertébrés [68] ; [82].

PGBD5 est l'une des 500 protéines les plus conservées chez les vertébrés [87]. C'est une recombinase d'ADN codée par un gène dérivé d'un transposon *PiggyBac* dont la séquence est très conservée (> 94 % d'identité entre Homme, rongeurs, bovins, caprins, équins, ovins et porcins). Son gène, *pgbd5*, est soumis à une forte sélection et a donc probablement un rôle physiologique ou métabolique très important chez les vertébrés.

### IV.1. Rôle putatif de *pgbd5* dans le cerveau et la mémorisation

Chez un individu sain, l'expression de PGBD5 chez l'Homme et la souris semble largement limitée aux tissus neuronaux de l'embryon précoce et certains neurones du système nerveux central (SNC) dans le cerveau adulte. Le gène *pgbd5* est exprimé sous forme d'ARN principalement dans le SNC et plus particulièrement dans les neurones pyramidaux de l'hippocampe, de l'amygdale et du bulbe olfactif qui sont impliqués dans les mémoires spatiale ou contextuelle (hippocampe), émotionnelle (amygdale) et olfactive (bulbe olfactif) [68] ; [82].

Une des hypothèses est que PGBD5 serait capable de catalyser, au sein même des neurones pyramidaux de l'hippocampe, de l'amygdale et du bulbe olfactif, des réarrangements génomiques locaux (des excisions et peut-être même des insertions de fragments d'ADN [88] ; [60] d'éléments de régulation transcriptionnelle et épigénétique [89]; [90] présents dans des fragments d'ADN dérivés de transposons *piggyBac*. Les données disponibles suggèrent que la protéine PGBD5 est principalement impliquée dans le réarrangement des gènes codant pour les protéines membranaires

impliquées dans la synaptogénèse. Notre hypothèse est que ces réarrangements permettraient une stabilisation des souvenirs stockés dans ces régions et constitueraient un mécanisme très conservé chez les vertébrés depuis son émergence chez les poissons cartilagineux qui interviendrait dans les processus de mémorisation. La dérégulation du gène *pgbd5* pourrait être à l'origine de certaines maladies neurologiques.

## ***IV.2. Rôle présumé de *pgbd5* dans la fertilité mâle***

La protéine PGBD5 est principalement exprimée dans les neurones pyramidaux du système nerveux central mais également dans plusieurs types de cellules le long de l'axe hypothalamo-hypophyso-gonadique (HPG), notamment les cellules de Leydig et les spermatogonies dans les testicules. Des données obtenues dans l'équipe chez le zebrafish et la souris (non publiées pour le moment) ont montrées que l'inactivation de *pgbd5* par délétion de l'exon 4 aboutie à une stérilité des mâles. Chez le poisson-zèbre, l'allèle invalidé est dominant. Les mâles produisent des gamètes mais leur fécondité est réduite de plus de 1000 fois. Les souris mâles homozygotes sont stériles mais possèdent des testicules de taille normale et produisent des spermatozoïdes. Nous pensons que la protéine PGBD5 présente dans les cellules de Leydig pourrait entraîner des réarrangements de l'ADN, qui auraient un impact sur l'expression des gènes impliqués dans les comportements reproductifs masculins et/ou le fonctionnement des organes sexuels accessoires et/ou de l'épididyme. Si les spermatogonies sont impliquées, la protéine PGBD5 pourrait agir pendant la méiose et/ou l'apoptose pendant la spermatogénèse.

## ***IV.3. Rôle présumé de *pgbd5* en cancérologie***

Une hypothèse parallèle est que PGBD5 est impliquée dans l'apparition et l'évolution des cellules cancéreuses [91] ; [92]. L'activité recombinase de PGBD5 doit être strictement contrôlée car son activité de recombinaison du génome est une source de mutations. En effet, une transposase est aussi une nucléase. En absence de transposon, toute recombinase sur-activée ou sur-concentrée dans un noyau va se comporter comme une ADN nucléase et va potentiellement induire une apoptose ou une dérive des cellules saines en cellules tumorales. Lorsqu'elle est exprimée de manière inappropriée dans les reins des enfants, la protéine PGBD5 déclenche des cancers rhabdoïdes [93]. *pgbd5* est

également sur-exprimé dans la plupart des tumeurs solides chez les enfants. L'analyse de la banque PRIDE (<https://www.ebi.ac.uk/pride/archive/>) de données de protéomique obtenues en spectrométrie de masse dans le cadre du projet « Cancer Genome Atlas » (TCGA) lancé en 2005, soulignent que la protéine PGBD5 est présente dans de multiples cancers et dans plus de 95 % des tumeurs chez les adultes. De plus, les taux de transcription de *pgbd5* dans les tumeurs sont en corrélation avec la survie des patients atteints de gliomes et de tumeurs solides telles que les cancers du côlon. *Pgbd5* pourrait donc être un gène promoteur de cancer dans certaines circonstances.

#### ***IV.4. Pourquoi s'intéresser à PGBD5 ?***

D'après les données PRIDE, le gène *pgbd5* est transcrit et traduit en protéine de manière très localisée. Ce gène codant une protéine transposase PGDB5 semble être sous une très forte pression de sélection au cours de l'évolution ce qui implique que *pgbd5* a un rôle probablement très important chez les vertébrés. A l'heure actuelle, le ou les rôles de PGBD5 ainsi que les mécanismes d'action de cette transposase domestiquée sont encore mal compris.

L'analyse des fichiers dans lequel sont détectés des peptides provenant de PGBD5 révèle que cette protéine est retrouvée dans les liquides corporels : le sérum plasmatique et le liquide cérébro-spinal avec un taux très supérieur en cas de cancers. Comme l'expression de PGBD5 est spécifique des neurones pyramidaux du système nerveux central et dans une moindre mesure des neurones bipolaires de la rétine, et des cellules de Leydig, notre hypothèse est que la présence de PGBD5 dans ces liquides est liée à la mort de cellules des lesquelles elle est exprimée. La protéine PGBD5 pourrait, par ce fait, se montrer être un biomarqueur diagnostique et pronostique d'intérêt pour détecter et suivre l'évolution des maladies neurologiques et / ou de tumeurs cérébrales. Cependant, l'absence d'anticorps commerciaux fiables et efficaces a empêché jusqu'à présent son utilisation comme biomarqueur en analyse médicale.

# Projet d'étude

---

L'ensemble des travaux décrits dans ce mémoire s'inscrit dans un programme de recherche de mon équipe d'appartenance et de ses partenaires visant à développer des sérums ou des fragments d'anticorps reconnaissant la protéine PGBD5 pour la détection et la quantification de celle-ci dans les fluides corporels humains à des fins de dépistages de cancers.

Afin de comprendre un peu mieux le rôle de PGBD5, mon projet d'étude s'est axé autour de deux questions :

- Où s'exprime la protéine PGBD5 dans un organe ou un tissu sain?
- Quels sont les sites de liaisons à l'ADN de la protéine PGBD5 dans le génome de cellules saines et tumorales ? Se fixe-t-elle sur les *pbles* humains ?

Pour répondre à ces deux questions, nous avons dans un premier temps dû nous procurer des outils permettant de détecter la protéine PGBD5 dans des tissus sains et cancéreux afin de mettre en évidence les différentes isoformes de la protéine. Faute d'anticorps commerciaux valides sur le marché, nous avons localisé l'expression de *pgbd5* grâce au gène rapporteur *LacZ* introduit dans un intron par transgénèse dans une lignée de souris mutante. Nous avons également développé des sérums murins par une méthode de vaccination à ADN en utilisant la technologie ICAntibodies™ de la société In-Cell-Art (Nantes) que nous avons ensuite caractérisés pour mettre en évidence la présence d'anticorps reconnaissant des épitopes séquentiels et / ou conformationnels de la protéine PGBD5.

La caractérisation des cibles chromatiniennes de la protéine PGBD5 a été réalisée dans un second temps par une approche de ChIP-séquençage. Les données de séquençage haut débit obtenues après immunoprécipitation de la chromatine à partir de tissus sains murins et de cellules humaines cancéreuses ont été analysées dans le but de déterminer si la protéine PGBD5 est capable de se fixer sur les *pbles* humains. Afin de déterminer si la protéine PGBD5 est une transposase toujours active, des tests de transpositions ont été entrepris afin de savoir si PGBD5 peut transposer les *pbles* humains et une recherche de sites d'insertions d'événements de transposition par la méthode de LAM-PCR a également été réalisée.

L'ensemble de ces travaux permettra d'éclaircir le ou les rôles de la protéine PGBD5 notamment en déterminant les motifs de liaisons de la protéine PGBD5 à l'ADN.

# Matériels et Méthodes

---

## I. Matériels biologiques utilisés

### *I.1. Les sérums murins anti-PGBD5 d'In-Cell-Art*

Pour notre protocole d'immunisation, la séquence de la partie C-terminale de PGBD5 a été utilisée afin de générer un ADN complémentaire (ADNc) optimisé pour l'immunisation génique chez la souris. La séquence de PGBD5 non optimisée de 415 acides aminés correspondant à la position 140 à 554 est utilisée (**Figure 13**). Afin d'initier la synthèse de la protéine l'acide aminé guanine (G) de la position 1 est remplacé par l'acide aminé méthionine (M).

L'ADNc optimisé pour la souris a été synthétisé puis cloné dans un plasmide adapté à l'immunisation génique (pVAX). Dix souris au total ont été vaccinées avec ce plasmide dans le muscle gastrocnémien (mollet) : 5 souris SWISS-CD1 et 5 souris C57BL/6 ont reçues 4 injections au jours (J) 0, 21, 42 et 56. Des échantillons tests d'antisérums de 50 µL par souris ont été prélevés avant chaque immunisation. Au terme du protocole d'immunisation, à J71 après la première immunisation, les souris ont été exsanguinées et le sérum récupéré (environ 500 µL) pour chaque souris.

### *I.2. Les lignées cellulaires cancéreuses*

Six lignées de cellules cancéreuses exprimant une des isoformes de PGBD5 ont été utilisées : trois lignées de cellules de glioblastomes (H4, 8MG et T98G), une lignée de cellules de cancer colorectal (HCT116) ainsi que deux lignées cancéreuses (HeLa et HEK293) n'exprimant pas PGBD5 et utilisées comme témoins négatifs. La composition du milieu de culture pour chaque lignée cellulaire est répertoriée dans le **tableau 1**. Les lignées sont cultivées dans une étuve à 37°C supplémentée de 5 % de CO<sub>2</sub>.

## ***1.3. La lignée de souris $Pgbd5^{tm1a}$***

### ***1.3.a. Origine de la lignée de souris***

La lignée  $Pgbd5^{tm1a}$  provient de la collection du KOMP (Knockout Mouse Project) de l'université de Californie à Davis (Etats-Unis) (**Figure 14A**). Le KOMP est une initiative du NIH visant à générer une ressource publique composée de cellules souches embryonnaires (ES) de souris présentant une mutation nulle pour un gène par recombinaison homologue. Dans le cas de cette lignée, l'insertion du gène rapporteur *LacZ* permet une évaluation rapide des types de cellules qui expriment normalement le gène d'intérêt. La présence d'éléments *cis* tels que les sites loxP et FRT (site de reconnaissance de la Flipase) permettent de rendre la mutation conditionnelle en présence de la recombinaison Cre ou de la flipase recombinase (Flp) (**Figure 14B**).

### ***1.3.b Entretien de la lignée***

Les souris sont hébergées et élevées au sein de l'Unité Expérimentale de Physiologie Animale de l'Orfrasière (UEPAO) dans l'animalerie rongeurs. Cette animalerie respecte les normes standard européennes de référence établies par la FELASA (Federation of European Laboratory Animal Science Associations) qui fonctionne sous forme de recommandations. A chaque statut sanitaire, correspond une liste d'exclusion, c'est-à-dire une liste d'agents infectieux qui doivent être absents de l'animalerie. Le statut « conventionnel », correspondant au niveau d'hébergement de la lignée de souris, est basé sur une liste d'exclusion contenant uniquement les agents zoonotiques majeurs. Lors de l'entrée de la lignée au sein de l'animalerie, les souris ont été placées préalablement en zone de quarantaine pour une période de 2 mois au terme desquels un bilan sanitaire FELASA a été réalisé grâce à des souris sentinelles. Le système de sentinelles consiste à introduire des animaux EOPS (exempt d'organisme pathogène spécifique), mis en contact avec les animaux à contrôler pendant une période suffisamment longue (au minimum 6 semaines) pour être contaminés par les pathogènes éventuellement présents et pour que leur contamination puisse être mise en évidence par le biais des tests mis en œuvre (sérologie, bactériologie...). Le contact peut être effectué de manière directe (animaux allant de cage en cage) ou indirecte (contact avec la litière sale, la nourriture et l'eau de boisson des autres cages).

Trois semaines après une mise bas, les souris sont sevrées et identifiées par puce RFID (radio

frequency identification ou identification par radiofréquence) par le personnel de l'animalerie. Au moment de la pose de la puce, un morceau d'oreille est prélevé afin de déterminer ultérieurement le génotype de l'animal. Les échantillons sont conservés à -20°C en attendant l'extraction d'ADN génomique de l'oreille.

### I.3.c Géotypage des animaux

A partir de ces fragments d'oreilles, l'ADN génomique est extrait en utilisant le kit NucleoSpin® Tissue (Macherey Nagel # 740952.250). Sur chaque échantillon, deux réactions d'amplifications (ou PCR pour Polymérase Chain Reaction) sont réalisées avec le mélange GoTaq (Promega # M3005). Les séquences des amorces utilisées ainsi que les conditions de PCR sont indiquées dans les **tableaux 2**.

Pour amplifier le gène sauvage de *pgbd5*, les amorces CSD-gene-F et CSD-ttR sont utilisées. En présence d'un allèle sauvage du gène *pgbd5*, un amplicon de 377 pb est produit. La seconde PCR amplifie une fraction de la cassette contenant le gène *LacZ*. Les amorces LacZ1 et LacZ2 sont utilisées et conduisent à la production d'un amplicon de 273 pb en présence de l'allèle mutée. Le résultat de ces deux PCR permet de déterminer si la souris a un génotype sauvage (ou WT pour Wild Type) avec deux allèles possédant le gène *pgbd5* sauvage, hétérozygote avec présence d'un allèle sauvage et d'un allèle mutée contenant le gène *LacZ* ou homozygote pour notre construction avec présence de deux allèles mutées.

## II. Méthodes de biologie cellulaire

### II.1. Transfection transitoire

La veille de la transfection, 100 000 cellules vivantes par puits sontensemencées dans une plaque de culture cellulaire de 24 puits. Les cellules sont comptées à l'aide d'une cellule de Malassez en présence de bleu trypan, ce qui permet de différencier les cellules vivantes des cellules mortes. Seules les cellules vivantes apparaissant rondes et brillantes sont comptées, les cellules mortes sont colorées en bleu. Chaque transfection est réalisée en duplicata et deux puits supplémentaires sontensemencés pour le contrôle négatif sans transfection. 45 min avant la transfection, le milieu de culture est

remplacé par le milieu de culture sans sérum. Les cellules sont transfectées avec 500 ng de plasmide avec l'ICAFectin<sup>®</sup>441 (In Cell Art # 441-500) comme agent transfectant en suivant les recommandations préconisées par le fabricant (**Annexe 1**). La transfection est effectuée en incubant les cellules avec le mélange ADN / ICAfectin441 pendant 4h à 37°C avec 5 % de CO<sub>2</sub> puis le milieu est remplacé par 1 mL de milieu adéquat supplémenté de SVF. 24h ou 48h après la transfection, les cellules peuvent être observées ou récupérées selon les besoins après lavages des cellules avec 500 µL de PBS1X. Pour l'analyse des protéines, les cellules sont récupérées en ajoutant 500 µL de tampon Laemmli 2X supplémenté extemporanément de 5 % de 2-mercaptoéthanol. Les échantillons sont congelés à -20 °C en attendant d'être utilisés.

## *II.2. Transfection stable*

Pour la fabrication des lignées cellulaires stables, les cellules HeLa sont préalablement mises en culture dans un milieu DMEM en présence 5 % de SVF (Ozyme # 631107) dépourvu de traces de tétracycline.

Après plusieurs passages en milieu DMEM supplémenté en 5 % de SVF sans tétracycline, 100 000 cellules HeLa sont mises en culture dans des puits de plaque P24.

Le lendemain, les cellules sont co-transfectées par deux plasmides avec un ratio 10 : 1, soit 200 ng de transposon et 20 ng de transposase. Le premier plasmide source de transposon correspond au plasmide pSBtet\_V5\_PGBD5. Le second est un vecteur d'expression plasmique de la transposase Sleeping Beauty hyperactive (plasmide SB100x). Chaque plasmide source de transposon est également transfecté seul afin de déterminer le niveau de bruit de fond correspondant aux intégrations aléatoires dans le génome de la cellule hôte. Le plasmide pSBtet\_GP originel contenant le gène de la luciférase est également transfecté et utilisé comme témoin.

Les cellules sont transfectées avec l'ICAFectin441 comme indiqué précédemment dans la section II.1. Après l'incubation de 4h à 37°C avec 5 % de CO<sub>2</sub>, le milieu est remplacé par 1 mL de milieu DMEM supplémenté de 5 % de SVF sans tétracycline et les cellules sont remises à l'étuve pour 48h avant l'étape de sélection avec la puromycine dihydrochloride (Sigma # P8833). Après 2 à 3 semaines sous sélection en milieu DMEM supplémenté de 5 % de SVF sans tétracycline et 0,5 µg/mL de puromycine, les clones résistants sont prélevés pour être amplifier.

### *II.3. Test de transposition*

100 000 cellules HeLa sont co-transfectées avec un plasmide source de transposase et un plasmide source de transposon. Pour la transfection des cellules, les différents plasmides utilisés sont listés dans le **tableau 3** et un ratio 1:1 est appliqué soit 200 ng de chaque plasmide. Ces plasmides sont transfectés simultanément en présence de l'agent transfectant JetPEI<sup>®</sup> (Polyplus transfection # 101-10) en suivant les recommandations du fournisseur (**Annexe 2**). Après une incubation de 4 heures post transfection, le milieu est retiré puis remplacé par 1 mL de milieu DMEM 5 % SVF classique. Les cellules sont incubées 48 heures à l'étuve à 37°C avec 5 % de CO<sub>2</sub> avant l'étape de sélection en présence de la généticine (G-418) avec le milieu DMEM, 5 % SVF, 0,8 µg/mL G-418. Après 2 à 3 semaines sous sélection, les clones sont colorés pour être comptés ou récupérés selon les besoins.

### *II.4. Sélection de clones cellulaires*

Les cellules à sélectionner sont rincées deux fois avec 1 mL de PBS 1 X. Elles sont ensuite décollées du support en utilisant 100 µL de trypsine par puits. La plaque est incubée 3-4 min maximum à l'étuve à 37 °C avec 5 % de CO<sub>2</sub> afin que les cellules se décollent du fond du puits. Après incubation, 10 volumes de milieu de sélection sont ajoutés et le mélange est transféré intégralement dans une boîte de culture cellulaire de 100 mm de diamètre (ThermoFisher # 172931) contenant 7 mL de milieu de sélection lors d'un test de transposition ou 1, 10 et 100 µL de ce mélange cellulaire est réparti dans 3 boîtes contenant 8 mL de milieu de sélection lors d'une transfection stable. Les boîtes sont ensuite placées dans l'étuve à 37°C avec 5 % de CO<sub>2</sub>. Le milieu de culture est changé toutes les 48 heures pendant une quinzaine de jours jusqu'à obtention de clones cellulaires suffisamment gros pour être comptés ou prélevés.

### *II.5 Comptage des clones cellulaires*

Cette étape a pour objectif de faciliter la visualisation des clones. Le milieu est éliminé, les cellules sont lavées deux fois avec 8 mL de PBS1X puis fixées et colorées avec 8 mL de bleu de méthylène 0,35 %, éthanol 70 % pendant une nuit à température ambiante. Le lendemain, la solution de bleu de

méthylène est récupérée et filtrée pour une utilisation ultérieure et les boîtes sont rincées sous un filet d'eau claire jusqu'à élimination du bleu résiduel sur les boîtes puis elles sont rincées à l'eau déminéralisée. Après séchage des boîtes, les clones cellulaires sont ensuite comptabilisés.

## ***II.6. Amplification des clones cellulaires***

Lorsque les clones cellulaires sont visibles à l'œil nu, une douzaine de clones bien isolés sont ensemencés en puits de plaque P96. Pour cela, le milieu de la boîte est éliminé et le clone est gratté délicatement avec la pointe d'un cône de P200. Celui-ci est ensuite placé dans un puits de plaque P96 contenant 100 µL de milieu de sélection et des aspirations / refoulements du milieu avec la pipette sont réalisées afin de détacher et dissocier mécaniquement les cellules contenues sur le cône.

La plaque P96 est placée dans l'incubateur réglé à 37°C avec 5 % de CO<sub>2</sub> puis la croissance cellulaire est surveillée tous les jours. Lorsque la densité du puits est suffisante, les cellules sont trypsinées puis passées dans un puits plus grand, et ainsi de suite jusqu'au passage dans une flasque T75 (Falcon # 353136). Lorsque les cellules sont à 80-90 % de confluence, elles sont repiquées au 1/10ème et l'excès de cellules est stocké en milieu de culture avec 10 % de diméthyl sulfoxyde (DMSO) (Sigma # D2650) dans des cryotubes (Thermo Fischer # 368632). Ces tubes sont ensuite placés dans une boîte de congélation Mr. Frosty (Thermo Fischer # 5100-0001), puis mis dans un congélateur à -80 °C pour quelques jours. Cette boîte va permettre un refroidissement proche de -1°C par minute qui est la vitesse optimale pour la conservation de cellules. Les cryotubes sont ensuite transférés dans un conteneur à azote liquide pour un stockage sur le long terme.

## ***II.7. Histologie***

### ***II.7.a Fixation des organes***

Les souris adultes sont euthanasiées par élongation cervicale puis les organes sont prélevés, pesés et placés dans une solution de paraformaldéhyde (PAF) 4 %, PBS 1X à température ambiante pendant 2 heures. Après fixation, les organes sont lavés dans deux bains de PBS 1X puis placés pendant une nuit ou plus selon la taille et la nature du tissu dans une solution de sucrose 30 %, PBS 1X afin de les cryopréserver. Lorsque les organes sont parfaitement imbibés de sucrose, ils tombent au fond du

réceptier, ils sont alors inclus dans des moules plastiques contenant du Tissue-Tek® (Sakura # 4583) et placés sur de la carboglace pour une prise rapide. Des coupes d'environ 10 µm d'épaisseur sont ensuite réalisées sur des lames SuperFrost à l'aide d'un cryostat LEICA.

### ***II.7.b. Coloration X-Gal des coupes histologiques***

Les coupes sont réhydratées dans un bain de PBS 1X pendant 5 min puis colorées avec une solution de colorant X-Gal préparée extemporanément. Sur chaque coupe, 200 µL de solution contenant 4 mM de ferrocyanure de potassium, 4 mM de ferricyanure de potassium, 2 mM de MgCl<sub>2</sub> et 0,4 mg/mL de X-Gal dans du PBS1 X sont déposés. Les lames sont ensuite placées à 30°C pendant 4h à l'abri de la lumière. Après incubation, les lames sont rincées 3 fois dans des bains de PBS 1X. Les lames sont ensuite colorées au rouge neutre pendant 5 minutes, lavées à l'eau du robinet, puis déshydratées dans des bains d'alcool 70°, 90° et absolu (trois bains de 5 minutes par type d'alcool). Les lames sont incubées par la suite dans trois bains de toluène et montées entre lame et lamelle avec du Depex (Dibutylphthalate Polystyrene Xylene) Mounting Medium (Sigma # 44581).

### ***II.7.c. Observation au microscope***

Les lames sont observées sous microscope Olympus SZ61 à différents grossissements selon la localisation et l'intensité du signal de la coloration.

## **III. Méthodes de biologie moléculaire**

### ***III.1. Clonage***

#### ***III.1.a. Obtention des isoformes humaines et murines de PGBD5***

Les gènes codant les isoformes de PGBD5 ont été fournies par **Hugues Dardente** (équipe NMR, unité PRC). A partir des données provenant de trois banques de données de RNA-seq : NCBI

(<https://www.ncbi.nlm.nih.gov/gds?term=RNA-seq&cmd=DetailsSearch>), Ensembl (<https://m.ensembl.org/info/genome.html>) et GTEx (<https://gtexportal.org/home/>), la présence d'ARNm codant les différentes isoformes de PGBD5 est vérifiée par RT-PCR nichées en utilisant des amorces permettant de couvrir une ou plusieurs jonctions d'épissage et des ARN totaux de cerveaux entiers humain et souris (**Figure 15**).

Chez l'Humain, 4 isoformes de PGBD5 (Hs554 ; Hs524 ; Hs455 et Hs409) sont mises en évidence et vérifiées par la présence de leurs ARNs. Chez la souris, deux isoformes sont trouvées : Mm409 et Mm523 qui correspondent aux isoformes humaines Hs409 et Hs524 respectivement (**Figure 16**).

Dans ce manuscrit, seules les isoformes murines sont exploitées.

Un cerveau de souris (souche C57Bl6) est utilisé pour extraire les ARN totaux à l'aide de TriReagent® (Sigma-Aldrich # 93289). La synthèse de l'ADNc est réalisée sur 2 µg d'ARN totaux avec le kit RT Omniscript (Qiagen # 205111) et des amorces oligo dT<sub>(15)</sub> (Promega # C1101). Les conditions utilisées lors de la reverse transcription (RT) sont présentées dans le **tableau 4**. Les amorces PCR flanquées de sites de restriction appropriés pour le clonage sont synthétisées par Eurofins Genomics (Ebersberg, Allemagne) (**Tableau 5**). Les PCR sont réalisées avec le kit PCR Master Mix haute-fidélité Phusion (ThermoScientific # F-531S). Après électrophorèse de la totalité des produits PCR sur gel d'agarose à 1,5 % en tampon TAE 1X contenant 1X de GelRed (Biotum # 17G0905-105136), les fragments de PCR d'intérêt sont extraits du gel sous banc UV et purifiés avec le kit QIAquick gel extraction (Qiagen # 28704). La séquence de chaque fragment de PCR est vérifiée par séquençage avant la suite du clonage. Les fragments de PCR purifiés sont ensuite soumis à une restriction enzymatique selon les recommandations du fournisseurs d'enzymes NEB (<https://nebcloner.neb.com/#!/redigest>). L'isoforme Mm523 est digérée par double digestion avec les enzymes *EcoRI/XbaI* et l'isoforme Mm409 par les enzymes *EcoRI/XhoI*. Les fragments PCR purifiés et digérés sont de nouveau purifiés sur colonne avec le kit QIAquick PCR purification kit (Qiagen # 28104) afin d'éliminer toute trace d'enzymes et de petits fragments d'ADN issus de la digestion des extrémités des produits PCR .

### ***III.1.b. Vecteurs plasmidiques d'expression mammalienne pCS2\_x5Myc\_ isoforme PGBD5***

Les fragments PCR digérés et purifiés sont ensuite clonés dans un vecteur d'expression mammalien pCS2+MT (**Figure 17**) préalablement digéré par les enzymes de restriction *EcoRI/XbaI* ou *EcoRI/XhoI*. Le vecteur pCS2 utilisé contient en 5' de son multi-site de clonage (MSC) entre les

sites *ClaI* et *EcoRI*, une séquence de 5 copies de 39 nucléotides (ATGGAGCAAAAGCTCATTCTGAAGAGGACTTGAATGAA) codant un pentamère de l'épitope MYC (N-EQKLISEEDL-C). Les fragments correspondant aux différentes isoformes sont insérés entre les sites de restriction *EcoRI* et *XhoI* pour l'isoforme Mm409 ou *EcoRI* et *XbaI* pour l'isoforme Mm523.

### ***III.1.c. Vecteurs plasmidiques d'expression mammalienne pSBtet\_V5\_ isoforme PGBD5***

Les différentes isoformes de PGBD5 ont également été clonées dans un plasmide pSBtet-GP (Addgene # 60495) après retrait du cadre de fragment d'ADN codant la luciférase (fragment de 1677 pb) par digestion enzymatique avec les enzymes de restriction *ClaI* et *NcoI*. Un MCS (ou polylinker) est ensuite inséré à la place du gène de la luciférase afin d'y insérer les isoformes PGBD5 (**Figure 18A**). La cassette polylinker est composée de deux amorces complémentaires présentant trois sites de restrictions *EcoRI*, *SpeI* et *XhoI* (**Figure 18B**) après hybridation des deux amorces. La cassette d'ADN double brin ainsi créée présente à ses extrémités des « bouts » cohésifs, compatibles avec les extrémités sortantes du plasmide digéré par *NcoI* et *ClaI*.

L'intérêt du plasmide pSBtet-GP est qu'il possède une expression constitutive d'un transactivateur inverse contrôlé par la tétracycline (rtTA). Cette protéine rtTA est une protéine de fusion composée du répresseur TetR (répresseur de la tétracycline) trouvée chez la bactérie *Escherichia coli* et du domaine activateur de la protéine VP16 trouvée dans le virus de l'herpès humain. Dans le cas d'un système Tet-On comme ici, la protéine rtTA est capable de se lier à l'opérateur (tetO) uniquement si elle est liée par une tétracycline ou dérivé de tétracycline. Les séquences TetO couplées au promoteur minimal sont appelées élément de réponse à la tétracycline (TRE) et répondent à la liaison de la protéine rtTA en permettant l'expression du gène (ici *pgbd5*) mis sous contrôle du promoteur. Ainsi, l'introduction de doxycycline (dérivé semi-synthétique de la tétracycline) dans le système déclenche la transcription de l'isoforme *pgbd5* à façon dans des lignées cellulaires stables.

Ce plasmide contient également de part et d'autre du système Tet-On et du gène de résistance à la puromycine, les ITR du transposon *Sleeping Beauty*. En présence de transposase *Sleeping Beauty*, le fragment plasmidique compris entre ces ITR pourra être excisé du vecteur et intégré dans le génome des cellules transfectées. La présence du gène de résistance à la puromycine permet de sélectionner les cellules ayant intégré ce transposon.

Pour la construction des plasmides pSBtet\_V5-PGBD5, un tag correspondant à l'épitope V5

(GKPIPPLLGLDST) est également fusionné en amont du gène *pgbd5* dans le plasmide (**Figure 19A**). L'épitope V5 provient des résidus entre les positions 95 à 108 de l'ARN polymérase  $\alpha$  sous-unité du virus simien de la grippe type 5. Le fragment V5 est récupéré à partir d'un plasmide utilisé dans l'équipe : pCS2\_U5\_V5\_Pokey\_U3. Sa séquence nucléotidique est indiquée en **figure 19B**. Ce plasmide est digéré par l'enzyme de restriction *MfeI* puis par l'enzyme *EcoRI* lors d'une seconde digestion. La présence de ce tag permet d'utiliser un anticorps anti-V5 lors de l'immunoprécipitation de la chromatine.

A chaque étape de clonage, 6 clones bactériens sont vérifiés par digestion enzymatique après miniprep. La séquence des plasmides ayant le profil attendu après digestion enzymatique est vérifiée par séquençage Sanger (Eurofins Genomics, Ebersberg, Allemagne) afin de s'assurer de l'intégrité de l'insert et de l'absence de décalage de cadre de lecture.

### ***III.1.d. Digestion enzymatique et purification des fragments digérés***

Les digestions enzymatiques sont réalisées sur 2  $\mu$ g de plasmide ou 5  $\mu$ L de miniprep lors d'une vérification de clone bactérien dans un volume final de 10  $\mu$ L. Le **tableau 6** regroupe les différentes enzymes de restrictions utilisées lors de ce travail avec les tampons et conditions utilisés. Lors d'une linéarisation de plasmide par digestion enzymatique, une étape de déphosphorylation des extrémités 5' du plasmide est réalisée en incubant le mélange réactionnel de digestion 1 heure à 37°C avec de la phosphatase alcaline (Quick CIP # M0525) afin d'éviter que le plasmide ne se referme sur lui-même lors de l'étape de ligation.

Après le temps d'incubation nécessaire à la digestion enzymatique et la déphosphorylation le cas échéant, le mélange réactionnel est incubé 5 min à 65°C puis déposé sur gel d'agarose 1,5 % et migré en tampon TAE 1X. Pour les besoins d'un clonage, le fragment d'ADN d'intérêt ou le plasmide linéarisé est extrait du gel et purifié à l'aide du kit NucleoSpin® Gel and PCR Clean-up (Macherey Nagel # 740609.250). Les produits purifiés sont ensuite dosés au Nanodrop.

### ***III.1.e. Ligation***

Après quantification de l'insert et du plasmide linéarisé, la ligation est réalisée selon un ratio classique 3 : 1 (3 correspondant à la quantité d'insert et 1 à la quantité de vecteur) à l'aide de la formule suivante : quantité insert requise (ng) = [(ng de vecteur x taille en kb de l'insert) x 3/1] / taille en kb

du vecteur. Une réaction sans insert est également réalisée en parallèle afin de servir de contrôle négatif de ligation (vérification de la qualité de la linéarisation et de la purification du vecteur et éventuellement de la déphosphorylation de ses extrémités). La réaction de ligation est réalisée dans un volume final de 20  $\mu$ L avec la T4 DNA Ligase (NEB # M0202) en suivant les recommandations du fournisseur. Après le temps d'incubation à 4°C (sur le week-end), 16°C (sur la nuit) ou à T° ambiante (1 heure), les tubes sont incubés 10 minutes à 65 °C afin d'inactiver l'enzyme avant la transformation bactérienne.

### ***III.1.f. Transformation bactérienne***

1 à 5  $\mu$ L de produit de ligation sont ensuite introduit par transformation bactérienne dans des bactéries thermo-compétentes *Escherichia coli* One Shot Top10 (Invitrogen # C404010) en suivant les recommandations du fournisseur. Les bactéries ainsi transformées sont ensuite sélectionnées sur boîte de pétri LB agar (pour lysogeny broth) additionné d'antibiotique (ampicilline à 100  $\mu$ g/mL, Kanamycine à 100  $\mu$ g/mL selon le plasmide utilisé). Le lendemain, six colonies obtenues après transformation bactérienne sont mises en culture dans un tube contenant 4 mL de milieu LB additionné d'antibiotique afin de vérifier l'intégration de l'insert dans le plasmide. Les tubes ainsiensemencés sont placés dans un incubateur à 37°C sous agitation rotative (200 rpm) pendant environ 16h.

### ***III.1.g. Extraction d'ADN plasmidique à partir de suspension bactérienne***

Les suspensions bactériennes sont centrifugées en deux fois dans un tube Eppendorf 2mL pendant 5 minutes à 11 000g puis les ADN plasmidiques sont extraits des culots bactériens en utilisant le kit Macherey-Nagel NucleoSpin® Plasmid (# 740588.250).

## ***III.2. RT-PCR semi Quantitative***

Afin de déterminer l'efficacité de l'induction sur la transcription d'ARNm *pgbd5*, des RT-PCR semi quantitatives sont réalisées après induction ou non des clones avec la doxycycline. Les cellules

lavées en PBS 1X sont trypsinées puis récupérées par centrifugation 5 min à 600 rpm dans un tube Eppendorf 1,5 mL RNase free. L'extraction d'ARN est réalisée avec le kit Quick DNA/RNA Microprep Plus (Zymo Research # D7005T) en suivant les recommandations du fournisseur. Les ARN extraits sont ensuite dosés avec le kit Qubit™ RNA HS Assay (Thermo Fischer # Q32855) puis 2 µg d'ARN sont rétro transcrits en ADNc en présence de la transcriptase inverse M-MLV (Thermo Fischer # 28025-013) et d'oligos (dT)<sub>15</sub> (**Tableau 7A**). Les PCR sont réalisées à partir de 2,5 µL d'ADNc synthétisés avec le mix GoTaq® G2 Master (Promega # M7822) (**Tableau 7B**). Le mix PCR est réalisé dans un volume final de 100 µL puis réparti en 4 puits distincts qui subiront un nombre de cycles de PCR de 15, 20, 25 ou 30 cycles après ajout des 2,5 µL d'ADNc. Les amorces utilisées ainsi que la taille théorique de l'amplicon sont répertoriés dans le **tableau 7C**. Après amplification, 10 µL de produits PCR sont déposés sur gel d'agarose 1,5 % supplémenté en GelRed 1X et migrés en tampon TAE 1X.

### *III.3. Dosage de la luciférase*

Les clones de lignée cellulaire HeLa pSBtet\_GP non induits et induits sont testés grâce au kit Bright-Glo Luciferase assay system (Promega # E2610) en ajoutant 200 µL de réactifs dans chaque puits de cellules lavées avec le PBS 1X. Les cellules sont resuspendues puis mises dans un puits de plaque à fond blanc P96. La détection de la luminescence s'effectue entre 2 à 15 min après ajout du réactif sur un lecteur de plaque Mithras<sup>2</sup> 943 (Berthold) avec le logiciel MikroWin 2010.

### *III.4. Mise en évidence des anticorps*

Tous les anticorps primaires commerciaux polyclonaux et monoclonaux utilisés sont présentés **tableau 8A**. Ces derniers ainsi que les sérums murins produits par In-Cell-Art sont testés dans un premier temps en Western Blot. Cette méthode de détection permet de mettre en évidence la présence d'anticorps dirigés contre des épitopes séquentiels, c'est-à-dire reconnaissant une séquence d'acides aminés linéaire. Ces anticorps anti-PGBD5, s'ils sont présents, permettent en milieu dénaturant, de mettre en évidence la protéine d'intérêt dans des extraits protéiques cellulaires mais également de différencier les isoformes de la protéine PGBD5.

Certains anticorps (en gras dans le **tableau 8A**) ainsi que les sérums murins produits par In-Cell-

Art sont testés par la suite en Dot Blot et immunocytochimie afin de mettre en évidence la présence d'anticorps reconnaissant des épitopes conformationnels. Ces anticorps permettent une reconnaissance de la structure secondaire et tertiaire intacte des antigènes contre lesquels ils sont dirigés. Ces anticorps sont nécessaires pour l'immunoprécipitation de la chromatine.

### ***III.4.a. Western Blot***

Les différents anticorps commerciaux et sérums murins sont testés sur des extraits de cellules HeLa transfectées par une des deux isoformes de PGBD5 taguées avec anti-Myc. Ce tag permet de vérifier que les protéines détectées par les anticorps à tester correspondent bien aux protéines attendues.

Les extraits protéiques sont dénaturés à 95°C pendant 5 minutes puis placés dans la glace. 10 µL d'échantillons sont ensuite déposés sur un gel de polyacrylamide dénaturant à gradient de 4 à 15 % (Biorad #4561083). En parallèle, 3 µL de marqueur de poids moléculaire Page Ruler Prestained Protein est déposé. Après migration électrophorétique, les protéines sont transférées en milieu liquide sur une membrane de nitrocellulose à l'aide du kit de transfert Trans-Blot Turbo (Biorad # 1704270). Après transfert, la membrane est bloquée avec le tampon Odyssey Blocking Buffer PBS (LiCor # 927-40000) pendant 1 heure à température ambiante sous agitation. Elle est ensuite incubée avec une dilution de l'anticorps primaire en PBS 1X supplémenté de 0,1 % de Tween pendant une nuit à 4°C sur une plaque rotative. Chaque anticorps est dilué selon les recommandations du fournisseur (**Tableau 8A**). Après 3 lavages de 5 minutes dans du PBS 1X Tween 0,05 %, la membrane est incubée pendant 1 heure en présence d'une dilution d'anticorps secondaire fusionné à un fluorochrome (**Tableau 8B**) en solution PBS-Tween 0,1 % sous agitation à l'obscurité. Après 3 lavages de 5 minutes de PBS-Tween 0,05 %, les membranes sont scannées sur le lecteur Infra-Rouge Imager Odyssey Clx. Cet appareil permet de détecter la fluorescence proche de l'infra-rouge de manière sensible et sans saturation directement à partir de la membrane d'hybridation.

### ***III.4.b. Dot Blot***

Cette méthode permet de détecter les anticorps reconnaissant la structure native de l'antigène quand les extraits bruts sont préparés en condition non dénaturantes.

48H après transfection, les cellules sont grattées dans leur milieu puis récupérées en tube Falcon 15 mL pour être centrifugées 5 min à 270 g à température ambiante. Le culot est lavé une fois avec 5

mL de PBS 1X puis repris dans 1 mL de tampon RIPA [HEPES 20 mM pH 7,5 à 7,9, 150 mM NaCl, 1 % NP40, 10 % Glycérol, 0,25 % sodium cacodylate, 1X PIC (protease inhibitor cocktail)]. Les cellules sont fragmentées par sonication à 4°C pendant 90 secondes (25W) avec une pause de 30 secondes entre chaque traitement de 30 secondes.

Sur une membrane de nitrocellulose posée sur un morceau de papier Wattman 3MM, 3 µL d'extraits cellulaires (équiconcentrés en protéine lorsque plusieurs échantillons différents sont utilisés) sont déposés en duplicat. La membrane est découpée autour des spots de dépôt puis chaque morceau correspondant à un dépôt est mis dans un puits de plaque de culture P6. Les plaques contenant les morceaux de membranes sont ensuite placées à l'étuve à 37°C pour séchage des membranes puis celles-ci sont saturées avec du tampon PBS 1 X, BSA 2 %, Tween20 0,1 % pendant 1 heure à température ambiante sous agitation. Au terme de cette incubation, les membranes sont hybridées avec les sérums à tester dilués au 1/1 000<sup>ème</sup> et un anticorps anti-actine dilué au 1/ 5 000<sup>ème</sup>. Les plaques contenant les membranes sont incubées à 4°C sous agitation pour la nuit.

Le lendemain, le milieu est retiré et chaque membrane est lavée 3 fois 10 min à température ambiante sous agitation avec 1 mL de PBS 1X, Tween20 0,1 % par puits. Les anticorps secondaires anti-souris IRDye 680 (LiCor #926-68072) et anti-lapin IRDye 800 (LiCor # 926-32213) sont dilués au 1/10 000<sup>ème</sup> en PBS 1X, BSA 2 %, Tween20 0,1 % et chaque membrane est hybridée avec 1 mL de cette dilution pendant 1 heure à température ambiante sous agitation. Après cette incubation, les membranes sont de nouveau lavées 3 fois 10 min à température ambiante sous agitation avec 1 mL de PBS 1X, Tween20 0,1 %. Les différentes membranes sont ensuite déposées sur un film transparent type film alimentaire puis scannées sur le lecteur Infra-Rouge Odyssey Clx.

### ***III.4.c. Immunocytologie***

Cette approche est une méthode d'analyse des cellules *in situ* permettant de révéler et cibler l'antigène natif d'intérêt à l'échelle cellulaire permettant ainsi de le localiser.

Sur une lame Lab-Tek 4 puits (Nunc # 177399), 50 000 cellules sontensemencées par puits dans 800 µL de milieu de culture adapté à la lignée de cellules (se reporter au **tableau 1**). Si les cellules doivent être transfectées, 25 000 cellules seulement sontensemencées par puits. Les lames sont incubées dans une étuve à 5 % CO<sub>2</sub>, 37°C pour 24h si la lignée cellulaire a une croissance rapide (HeLa par exemple) ou pour 48h si la lignée cellulaire a une croissance plus lente (HT29 par exemple).

Lorsque les cellules sont à un stade de 80-90 % de confluence, le milieu de culture est retiré et les cellules sont lavées 3 fois pendant 5 min avec 1 mL de PBS 1X. Après le dernier lavage, les cellules sont fixées avec 1 mL de PBS 1X, 4 % de paraformaldéhyde pendant 30 min à température ambiante sous hotte aspirante. La solution de fixation est ensuite éliminée et les cellules sont lavées une fois avec 1 mL de PBS 1X. Les cellules sont ensuite incubées 3 minutes avec 1 mL de PBS 1X, 5 µg/mL de digitonine à température ambiante afin de perméabiliser les membranes cellulaires. Post incubation, le milieu est retiré doucement et les cellules sont lavées 3 fois avec 1 mL de PBS 1X.

Pour limiter le bruit de fond, les lames sont saturées 30 minutes à température ambiante en présence de sérum de l'espèce animale ayant servi à produire l'anticorps secondaire dilué au 1/15<sup>ème</sup> dans du PBS 1X. En absence de sérum spécifique, les lames sont incubées avec 1 mL de PBS 1X, Triton 100X 0,5 %, BSA 2 %. Après incubation, les lames sont lavées deux fois avec 1 mL de PBS 1X, BSA 0,5 % Tween20 0,2 %.

Lorsqu'un anticorps primaire n'a jamais été utilisé, différentes dilutions de celui-ci sont testées (1/20, 1/50, 1/100 et 1/250). Un puits de cellules non hybridées est utilisé comme témoin négatif pour vérifier la fluorescence naturelle des cellules. L'hybridation des cellules avec l'anticorps primaire se fait en milieu PBS 1X, BSA 0,5 %, Tween20 0,2 %. Les lames sont incubées à 4°C pour la nuit en chambre humide. Le lendemain, les lames sont lavées 3 fois 5 minutes avec 1 mL de PBS 1X, BSA 0,5 %, Tween20 0,2 %. Après le dernier lavage, les cellules sont hybridées avec l'anticorps secondaire dilué au 1/500<sup>ème</sup> dans du PBS 1X, BSA 0,5 %, Tween20 0,2 % et incubées 1 h à température ambiante en chambre humide et à l'abri de la lumière. Le milieu d'hybridation est éliminé et les lames sont lavées 3 fois 5 minutes avec 1 mL de PBS 1X, BSA 0,5 %, Tween20 0,2 %. Les lames sont ensuite incubées 2 minutes à température ambiante dans une solution d'eau ultra pure contenant 1 µg/mL d'Hoechst 33258. Cette coloration va permettre de marquer l'ADN des noyaux des cellules. Les lames sont lavées 3 fois pendant 1 minute avec 1 mL d'eau ultra pure puis le montage Lab-Tek est enlevé (puits et joints silicone). Les lames sont séchées à température ambiante puis montées avec une lamelle en ajoutant 200 µL de Fluoromount-G™ (SouthernBiotech # 0100-01), ce qui va permettre le stockage des lames à 4°C plusieurs mois après séchage de 24h à l'obscurité.

### *III.5. ChIP-seq : Immunoprécipitation de la chromatine et séquençage haut débit*

Cette méthode est utilisée pour identifier les régions chromosomiques enrichies en protéine d'intérêt. L'objectif est d'établir une carte du génome indiquant les zones où la protéine s'est liée.

Les différents tampons utilisés pour le ChIP sont répertoriés dans les **tableaux 9**.

#### *III.5.a. Production des échantillons*

Chaque échantillon est réalisé en triplicata à partir de 10 millions de cellules environ.

Les cellules HEK293 devant être transfectées sont mises en culture la veille de la transfection avec 2 millions de cellules. Elles sont ensuite transfectées par 500 ng de plasmide pSBtet\_PGBD5\_Mm524\_V5 avec l'ICAFectin441 puis incubées 24h à l'étuve à 37°C, 5 % de CO<sub>2</sub>. Afin de permettre l'expression de la protéine PGBD5, les cellules HEK transfectées sont induites pendant 24 h avec 1 ug/mL de doxycycline hydrochloride (Sigma # D3447-500MG).

Après le temps d'incubation pour l'induction et lorsque les cellules ont atteint la densité de 80-90 % de confluence, les cellules sont récupérées après traitement à la trypsine puis lavées avec du PBS 1X. Le culot de cellules est ensuite fixé avant congélation à -80°C si besoin.

Les cellules de Leydig de souris sont purifiées à partir de deux testicules de souris âgés de 8 semaines. L'enveloppe blanche constituée de tissu conjonctif correspondant à l'albuginée des testicules est éliminée. Les testicules sont ensuite incubés dans 10 mL de milieu L15 (Eurobio # CM1L15000U) contenant 2 mg/ml de collagénase (Sigma # C0130) et 10 µg/ml de DNase1 (NEB # M0303S) pendant 15 minutes à 35 °C. Les tubes séminifères sont séparés du milieu interstitiel contenant les cellules de Leydig par sédimentation dans une éprouvette de 20 mL. Le surnageant est centrifugé à 80 g pendant 8 minutes et le culot cellulaire contenant les cellules de Leydig est repris par 500 µL de milieu L15 puis déposé sur un gradient de Percoll discontinu de trois concentrations 17%, 42% et 70% (Sigma # P4937). Après centrifugation à 80 g pendant 15 minutes, les cellules de Leydig sont récupérées à l'interface des concentrations 17 % - 42 %. Elles sont ensuite lavées par 20 mL de milieu L15 puis récupérées par centrifugation à 80 g pendant 8 minutes. Le culot de cellules de Leydig est ensuite fixé puis fragmenté ou conservé à - 80 °C après fixation.

### ***III.5.b. Fixation et de fragmentation des cellules***

Avant immunoprécipitation, les temps de fixation et de fragmentation des cellules sont déterminés pour chaque type d'échantillons. Généralement, le temps de fixation est de 5 ou 10 min pour les cellules et petites biopsies. Dans le cadre de ce mémoire, tous les échantillons ont été fixés 10 min. Le temps de fragmentation est quant à lui plus variable selon la nature et la composition de l'échantillon.

Les culots de cellules lavées sont fixés avec une solution de 1 % de formaldéhyde en PBS 1X pendant 10 min à température ambiante sous agitation. L'arrêt de la fixation se fait par ajout de glycine à 1,25M pour une concentration finale de 0,125M. Les cellules sont incubées 5 min à température ambiante sous agitation. Les cellules sont ensuite centrifugées 5 min à 3 000 g puis lavées dans du tampon PBS-C contenant des inhibiteurs de protéase (Roche # 5892791001) avant d'être centrifugées de nouveau. Le culot de cellules est repris dans 300 µL de tampon de lyse SDS puis transféré dans un tube 1,5 mL TPX® (Diagenode # C30010010-300). Les échantillons sont ensuite fragmentés avec un Bioruptor (Diagenode # B01020001) 3 fois 5 min à 6 fois 5 min (paramètres « high », cycles 30s ON / 30 s OFF) selon la lignée de cellules. Après une centrifugation de 5 min à 10600 g, le surnageant contenant la chromatine fragmentée est récupéré pour être directement utilisé ou stocké à -80 °C en vue d'être immunoprécipité.

### ***III.5.c. Analyse de la qualité de la chromatine***

Avant de réaliser l'immunoprécipitation (IP) de la chromatine, la qualité de la fragmentation est évaluée afin de vérifier la taille des fragments d'ADN obtenus après réversion de la fixation par déprotéinisation. Pour cela, 10 µL de chromatine fragmentée sont dilués 4 fois dans une solution finale à 200 mM NaCl. Les échantillons sont incubés sur la nuit dans un Thermomixer Comfort (Eppendorf) sous agitation à 65 °C, 1400 rpm. Les échantillons sont ensuite traités à la RNase A à 10 mg/mL (Thermo Scientific # EN0531) 10 minutes à 45 °C et sous agitation à 1 400 rpm, puis sont traités à la protéinase K à 2 mg/mL (Qiagen # 19131) pendant 1 heure à 45 °C sous agitation à 1 400 rpm. L'ADN est ensuite purifié en utilisant les colonnes d'extraction NucleoSpin® Gel and PCR Clean-up en suivant les recommandations du fournisseur. Une élution de 20 µL est réalisée et 10 µL sont utilisés pour une électrophorèse sur gel d'agarose à 1 % supplémenté en GelRed 1X et migré en tampon TAE 1X. Une chromatine de qualité adéquate est caractérisée sur gel par une trainée ou « smear » de fragments d'ADN plus intense entre 200 pb à 500 pb.

### ***III.5.d. Immunoprécipitation de la chromatine***

L'immunoprécipitation de la chromatine est réalisée à partir d'un protocole adapté de Song et al. [94]. Les différentes étapes de l'immunoprécipitation sont réalisées dans des tubes LoBind (Eppendorf # 0030108051).

Au préalable, des billes magnétiques couplées à la protéine A (Invitrogen Dynabeads-protein A # 10002D) sont lavées deux fois sur un portoir magnétique (Diagenode # B04000003) avec 1 mL de tampon de dilution de la chromatine (CDB). Pour chaque immunoprécipitation, 10 µL de billes sont préparés. Après les lavages, les billes sont resuspendues dans 45 µL de CDB auxquels 5 µL de sérum anti PGBD5 ou d'anticorps commercial anti V5 (Enzo # ABS590-0200) à 0,2 mg/mL sont ajoutés. Les témoins sont immunoprécipités avec le sérum pré immun de l'animal ayant produit les anticorps anti PGBD5 ou avec de l'eau quand cela n'est pas possible. Les billes sont incubées 2 heures à 4 °C sur un portoir rotatif, puis placées sur un portoir magnétique afin d'éliminer le surnageant. Les billes sont lavées trois fois avec 1 mL de CDB. Entre chaque lavage, elles sont incubées dans du milieu CDB 5 minutes à 4 °C sur portoir rotatif. Elles sont ensuite resuspendues dans 50 µL de milieu CDB.

La chromatine fragmentée est diluée au 1/10ème dans le tampon CDB et les 50 µL de billes contenant les anticorps sont ajoutés pour un volume final de 1 mL. Les échantillons sont homogénéisés et incubés sur la nuit sur un portoir rotatif à 4 °C. Ils sont ensuite lavés sur le portoir magnétique deux fois avec 1 mL du tampon de lavage froid à faible concentration en sel, une fois avec 1 mL du tampon de lavage à forte concentration en sel froid, une fois avec 1 mL du tampon de lavage au chlorure de lithium (LiCl) froid, puis une fois avec 1 mL de tampon d'éluion (TE) froid. Ces lavages successifs avec des concentrations en sels croissantes permettent d'éliminer l'ADN non fixé aux billes. Afin de diminuer le bruit de fond, les billes sont transférées dans un nouveau tube et lavées une dernière fois avec 1 mL de tampon TE. Entre chaque lavage, une incubation de 5 minutes à 4 °C sur un portoir rotatif est réalisée. Les billes sont ensuite remises en suspension dans 240 µL de TE pour la réversion de la fixation comme décrit précédemment dans la section III.5.c. Lors de la purification sur colonne, l'éluion finale est faite avec 40 µL de tampon d'éluion. La concentration d'ADN issu de chromatine immuno-précipitée est quantifiée sur 1 µL d'éluat avec le kit Qubit™ 1X dsDNA HS (Invitrogen # Q33231). Les échantillons sont stockés à -80°C en attendant de réaliser les librairies Illumina.

### ***III.6. LAM-PCR et séquençage haut débit***

La méthode de LAM-PCR (*Linear Amplification-Mediated PCR*) est une stratégie de choix pour identifier des populations de sites d'insertions d'un transposon ou d'un virus intégratif, par exemple comme un rétrovirus, ou de localiser dans un génome des populations de séquences répétées. Elle est basée sur l'amplification linéaire des sites d'intégration transposon - ADN génomique à partir d'une amorce unique biotinylée qui s'hybride dans la séquence ADN du transposon. Les fragments amplifiés linéairement par PCR sont ensuite capturés avec des billes magnétiques couplées à la streptavidine, une protéine qui présente une forte affinité pour la biotine. Les fragments PCR d'ADN simple brin ainsi isolés sont d'abord transformés en ADN double brin puis un petit fragment d'ADN dont la séquence est connue (cassette) est liée à leur extrémité localisée dans l'ADN génomique. Cette ancre permettra d'amplifier exponentiellement par PCR la population de fragments d'intérêt. Les fragments ainsi générés sont ensuite séquencés à haut débit après réalisation de banques Illumina à partir de cette population de fragments.

#### ***III.6.a. PCR linéaire***

L'amplification linéaire se fait à partir d'un microgramme d'ADN. Les amorces biotinylées s'hybridant dans le gène de la néomycine présent dans le transposon sont répertoriées en vert dans le **tableau 10A**. Ces amorces vont permettre, après amplification linéaire, de séparer physiquement les fragments d'ADN contenant les sites d'insertion du reste de l'ADN génomique de la cellule hôte. La PCR s'effectue avec un mélange réactionnel contenant 83,5 nM d'amorce biotinylée 3' ou 5', 10 mM de chaque dNTPs, l'ADN à amplifier ainsi que la Taq polymérase Qiagen (Qiagen # 201205) dans un volume final de 50 µL. Le programme PCR utilisé est présenté dans le **tableau 10B**. Au terme des 50 cycles d'amplification, 2,5 unités de Taq polymérase sont ajoutées dans chaque tube puis le programme PCR est répété. Au total 100 cycles d'amplifications linéaires sont réalisés.

#### ***III.6.b. Capture magnétique et synthèse du second brin d'ADN***

Après resuspension des particules magnétiques Dynabeads M-280 Streptavidin à 10 µg/ µL (Invitrogen # 11205D), 20 µL de billes par échantillon sont placées sur un portoir magnétique. Le

milieu de conservation est éliminé en présence du champ magnétique et les billes sont resuspendues dans 40 µL de PBS 1X, BSA 0,1 % à pH 7,5. Cette étape est renouvelée une seconde fois avant de laver les billes avec 20 µL de tampon LiCl BW 3M. La composition des différentes solutions utilisées est indiquée dans le **tableau 10C**. Après élimination du milieu en présence du champ magnétique, les billes sont ensuite resuspendues dans 50 µL de tampon LiCl BW 6 M puis ajoutées aux 50 µL de produits PCR. Le mélange est incubé sur un agitateur à 1 000 rpm pour la nuit à température ambiante. Le lendemain, le surnageant est éliminé à l'aide du portoir magnétique et les billes sont lavées une fois avec 100 µL d'eau distillée. La synthèse du second brin d'ADN se fait en resuspendant les billes avec 10 µL de mélange réactionnel contenant 0,1 µg/µL d'hexanucléotides, 200 µM de chaque dNTP ainsi qu'une unité de polymérase de Klenow. Après une incubation d'une heure à 37°C, 90 µL d'eau distillée est ajoutée au mélange. Les tubes sont placés sur le portoir magnétique et le surnageant est éliminé. Les billes sont lavées une fois avec 100 µL d'eau distillée avant de réaliser les digestions enzymatiques.

### ***III.6.c. Digestion enzymatique de l'ADN génomique de l'hôte et ligation de la cassette linker***

L'ADN double brin est digéré par une ou par un mélange de plusieurs enzymes générant des extrémités cohésives identiques et donc compatibles. Les motifs de restriction contenant des dinucléotides AT étant plus importants que les dinucléotides CG au niveau des sites d'insertion, l'utilisation d'enzymes différentes à motif non CG est recommandée pour leur détection.

Le complexe ADN-billes magnétiques est incubé avec 10 µL de mélange de restriction pendant 1h à 37°C. Deux digestions différentes sont réalisées pour chaque échantillon. Leur composition est présentée dans le **tableau 11A**. Après le temps d'incubation, 90 µL d'eau distillée est ajoutée au mélange de restriction puis le surnageant est éliminé en présence du champ magnétique. Les billes sont ensuite lavées une fois avec 100 µL d'eau distillée avant la ligation avec la cassette linker.

La cassette linker est constituée de deux amorces : oligo I et oligo II (**Tableau 11B**). L'amorce Oligo I est composée d'une séquence longue de 38 nucléotides qui permet de générer deux ancrs pour hybrider des oligonucléotides. Ces ancrs sont utilisées lors des PCR nichées dans l'étape suivante. L'oligo II est quant à lui composé de 29 nucléotides complémentaires à l'amorce oligo I avec en position terminale la séquence complémentaire à l'extrémité cohésive générée par la digestion enzymatique.

Pour la construction de la cassette linker, 40 µL des deux oligos I et II à 100 µM sont incubés 5

minutes à 95°C dans un bloc chauffant en présence de 110 µL de Tris-HCl 250 mM pH7,5 et 100 mM de MgCl<sub>2</sub> dans un volume final de 200 µL. Après l'incubation de 5 min, le bloc chauffant est éteint et les échantillons redescendent en température doucement sur la nuit. Le lendemain, 300 µL d'eau distillée sont ajoutés et l'échantillon est transféré sur une colonne Amicon Ultra 30kDa (Sigma-Aldrich # Z677892) et centrifugé 10 min à 14 000 g à température ambiante. La colonne est ensuite renversée dans un nouveau tube Eppendorf 1,5 mL puis centrifugée à température ambiante 2 min à 1 000 g. Le volume de l'échantillon concentré est complété à 80 µL avec de l'eau distillée puis réparti en aliquot de 10 µL avant d'être conservé à -20°C.

Les complexes ADN-particules magnétiques sont incubés 5 min à température ambiante avec 10 µL de mélange de ligation contenant 2 µL de cassette linker dans le tampon de ligation 2X en présence de 2 unités de DNA ligase (Quick ligation kit NEB # M2200S). Après incubation, l'échantillon est complété à 100 µL avec de l'eau distillée puis le surnageant est éliminé en présence du champ magnétique. Les billes sont ensuite lavées une fois avec 100 µL d'eau distillée avant d'être dénaturées.

### ***III.6.d. Dénaturation et PCRs***

Les complexes ADN-particules magnétique sont incubés à température ambiante sur un agitateur à 1 000 rpm après addition de 5 µL de NaOH 0,1 N fraîchement préparé. Ce traitement dissocie les deux brins d'ADN et libère ainsi dans le surnageant, le brin d'ADN néo synthétisé composé de la cassette linker, de la séquence d'ADN génomique, et de l'extrémité du transposon. Cette récupération du surnageant est effectuée en présence du champ magnétique de façon à retirer l'autre brin biotinylé fixé sur la bille magnétique.

Pour la première PCR, 1 µL de produit de ligation dénaturé est ajouté au mélange réactionnel avec 25 µL de Taq Q5® High-Fidelity 2X Master Mix (NEB # M0491S) contenant une amorce LC1 spécifique de la cassette linker et une amorce biotinylée s'hybridant dans la partie ITR du transposon. Les différentes amorces sont utilisées avec une concentration de de 8,3 µM pour les PCR exponentielles. Les séquences des amorces sont répertoriées dans le **tableau 12**. Après les 35 cycles de PCR, une étape de capture magnétique est réalisée afin d'augmenter la spécificité et la sensibilité de la détection des sites d'insertion. Pour chaque échantillon, 20 µL de particules magnétiques-streptavidine sont lavées comme précédemment avec les tampons LiCl BW puis resuspendues dans 25 µL de tampon LiCl 6M. Cette suspension est mélangée aux 25 µL de produit PCR et incubée au moins 1 heure à température ambiante sur un agitateur à 1 000 rpm. Après ce temps d'incubation, les tubes sont placés sur le portoir magnétique et le surnageant est éliminé. Les billes sont lavées une fois

avec 100  $\mu\text{L}$  d'eau distillée puis les complexes ADN-particules sont dénaturés avec 20  $\mu\text{L}$  de NaOH 0,1 N fraîchement préparée. Le mélange est incubé sous agitation à 1 000 rpm 10 min à température ambiante avant de récupérer le surnageant en présence du champ magnétique. La seconde PCR est ensuite réalisée à partir d'1  $\mu\text{L}$  de produit PCR dénaturé dans un mélange réactionnel de 50  $\mu\text{L}$  avec le mix Q5. Après le programme PCR de 15 cycles, la qualité des produits PCR est contrôlée sur gel d'agarose 1,5 % contenant du GelRed 1X en TAE 1X en faisant migrer 5  $\mu\text{L}$  de produits PCR.

Afin de conserver la diversité des sites d'insertions, seuls les produits PCR présentant des smears de taille allant de 100 à 500 pb environ sur gel sont gardés pour la suite. Si aucun produit n'est visible après la PCR de 15 cycles, la seconde PCR est refaite en ajoutant 5 à 10 cycles supplémentaires.

### *III.7. Fabrication de banques de populations de fragments pour le séquençage Illumina*

#### *III.7.a. Fabrication des banques Illumina*

Les banques de population de fragments pour le séquençage Illumina sont réalisées à partir du kit NEB Next Ultra II DNA Library Prep kit for Illumina (New England Biolabs # E7645S) en suivant les recommandations du fournisseur. Le protocole schématisé est présenté **figure 20**. Pour la majorité des échantillons (provenant de ChIP ou de LAM-PCR), une quantité initiale d'ADN de 10 ng est utilisée pour la réalisation des banques. Lors de l'étape finale d'enrichissement, 7 cycles de PCR sont effectués pour ces échantillons. La quantité d'ADN de départ étant inférieure à 50 ng, aucune sélection de taille n'est réalisée, celle-ci étant effectuée directement par la plateforme de séquençage. Pour certains échantillons, la quantité initiale disponible après immunoprécipitation est de seulement 1 à 8 ng d'ADN. Le nombre de cycle de PCR d'enrichissement est alors adapté selon la quantité d'ADN utilisée au départ du protocole de fabrication des banques et a varié dans nos expériences de 8 à 10 cycles pour les échantillons les plus faibles selon les recommandations du fournisseur. Pour les échantillons présentant une quantité finale faible, 1 à 2 cycles de PCR d'enrichissement supplémentaire ont été effectués en plus des 8 à 10 cycles. Après la PCR d'enrichissement, les produits d'amplification sont nettoyés grâce à des billes magnétiques afin d'éliminer les dimères d'oligonucléotides, puis dosés avec le kit Qubit comme précédemment. A ce niveau, une étape de sélection de taille des fragments peut être introduite.

### ***III.7.b. Séquençage haut débit***

Les échantillons de banque Illumina présentant une concentration supérieure à 5 ng/uL sont envoyés à la plateforme GenomEast de l'Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC) à Illkirch afin d'être séquencés sur un appareil Illumina Hiseq4000.

Pour le séquençage des bibliothèques de CHIP-seq, une seule extrémité des fragments ADN est séquencée (single read), avec une longueur par read de 50 paires de bases. Cette solution est le moyen le plus simple, le plus rapide et le plus économique pour fournir un volume important de données de haute qualité.

Pour le séquençage des bibliothèques de LAM-PCR, les deux extrémités des fragments ADN sont séquencées (paired read), avec une longueur par read de 250 paires de bases. Cette solution est le moyen le plus simple, le plus rapide et le plus économique pour fournir un volume important de données de haute qualité lorsque l'on ne sait pas laquelle des deux extrémités contient l'information d'intérêt (ici l'extrémité du transposon avec le site d'insertion).

L'analyse bio-informatique des données de séquençage haut débit générées ainsi que l'analyse statistique de ces données sont réalisées au sein de l'équipe par **Laura Helou** (étudiante en thèse) et **Yves Bigot** (directeur de recherche).

## **IV. Analyse bio-informatique**

### ***IV.1. Analyse qualitative et nettoyage des données***

Les données fournies par la plateforme ont été transmises sous forme de fichiers compressés au format fastq.gz. Les lectures et les positions nucléotidiques en 5' et 3' de chaque séquence de mauvaise qualité sont nettoyées en utilisant le logiciel TrimmomaticSE. Cet outil permet également de retirer les adaptateurs Illumina utilisés lors de la fabrication des banques. Après analyse qualitative et nettoyage des données, seuls les sites de liaison ayant un seuil de significativité (q-value calculée par PePr (p-value soumise à une correction de Bonferroni, c.a.d. p-value / nombre de pics au-dessous du seuil de sélection ( $p < 0.01$ )) inférieur à 0,05 sont conservés.

## IV.2. Détection des pics pour le ChIP-séquençage

Les lectures de séquence nettoyées sont alignées sur la séquence des chromosomes de la version hg38p10 (de janvier 2017) du génome de référence humain en utilisant le logiciel bowtie2. Cet outil d'alignement demande peu de mémoire disponible pour aligner les séquences des lectures sur de longues séquences de référence. Il est particulièrement efficace pour aligner des lectures d'environ 50 à 100 caractères sur des génomes de taille relativement importante comme le génome humain qui représente environ 3,3 Gigabases. Les alignements générés par bowtie2 sont au format SAM (pour cartographie d'alignement de séquence, en anglais *Sequence Alignment Map*) qui est un format texte permettant de stocker des séquences biologiques alignées sur une séquence de référence. Le format SAM permet également une interopérabilité avec un grand nombre d'autres outils bio-informatiques.

Les alignements obtenus correspondent aux régions contenant le ou les sites de liaison à l'ADN de la protéine PGBD5. Ces régions, représentées par des pics d'enrichissement en lectures, ont été localisées en utilisant en parallèle deux logiciels: PePr1.16 [95] et MACS2 [96]. PePr (pour peak-calling prioritization pipeline) est un outil qui permet d'analyser soit un groupe d'échantillons ChIP-Seq avec des contrôles, soit de comparer deux groupes d'échantillons ChIP-Seq. Il donne la priorité aux pics qui reflètent un plus fort degré d'enrichissement et une plus grande cohérence entre les échantillons. MACS2 (pour Model-based Analysis of ChIP-seq) est un algorithme de calcul qui identifie les sites de transcription, de liaison de facteur protéique ou des modifications d'histones à l'échelle d'un génome à partir des données ChIP-seq. MACS2 se compose de quatre étapes : suppression des lectures redondantes, ajustement de la position de lecture, calcul de l'enrichissement maximal et estimation du taux de fausses découvertes. Ces deux outils diffèrent principalement au niveau du traitement des répliques. PePr prend en compte la variation entre les répliques d'un même groupe alors que MACS2 fusionne les répliques afin d'obtenir une plus grande puissance de détection.

Dans notre cas présent, les deux outils ont fourni des résultats convergents. Seuls les pics obtenus avec une valeur de probabilité  $q$  (taux de fausses découvertes) inférieure à 0,05 avec au moins un des deux logiciels sont conservés et réunis en utilisant la suite d'outils bedtools. L'option merge a permis de fusionner les pics obtenus sur un même intervalle ou se recouvrant avec PePr et MACS2 afin de ne garder qu'un seul pic pour la suite de l'analyse. Les motifs conservés dans les régions chromosomiques couvertes par ces pics sont identifiés en utilisant les logiciels GLAM2, MEME-ChIP-seq et RSAT.

GLAM2 est un programme permettant de trouver des motifs récurrents ou de longueurs variables dans des séquences d'acides aminés ou de nucléotides. La principale innovation de GLAM2 est de permettre la recherche de motifs comportant des insertions ou délétions.

MEME-ChIP est un outil effectuant une analyse complète des motifs (y compris la découverte de motifs) sur de grands ensembles de séquences (généralement des nucléotides).

RSAT regroupe des outils permettant l'analyse des éléments cis-régulateurs, avec par exemple la découverte de motifs, l'analyse des motifs de fixation des facteurs de transcription, la génomique comparative et l'analyse des variations de régulation.

Les éléments *pble* présents dans les régions chromosomiques couvertes par les pics ont été colocalisés en utilisant la suite de logiciels *bedtools*. Les résultats finaux ont été visualisés en utilisant l'interface graphique IGV (Integrative Genomics Viewer)

### ***IV.3. Détection des jonctions transposons / chromosomes dans les données de LAM-PCR***

Les différentes séquences obtenues sont triées afin de récupérer les séquences des jonctions des fragments d'ADN insérés dans les chromosomes des cellules en partant des séquences plasmidiques terminales connues situées à 100 pb en amont et en aval du transposon *pble*-NeoR inséré. Dans un premier temps, les séquences filtrées sont cartographiées par rapport à la séquence du plasmide moins les régions 100 pb flanquant des deux côtés le transposon *pble* NeoR avec le logiciel BWA-MEM. BWA (Burrows-Wheeler Aligner) est un logiciel permettant de cartographier des séquences peu divergentes par rapport à un grand génome de référence, tel que le génome humain. Il se compose de trois algorithmes : BWA-backtrack, BWA-SW et BWA-MEM. BWA-MEM, qui est le plus récent, est généralement recommandé pour les requêtes de haute qualité car il est plus rapide et plus précis. BWA-MEM a également de meilleures performances que BWA-backtrack pour les lectures Illumina de 70-100 pb. Les séquences non cartographiées sont ensuite extraites des séquences qui ne correspondent pas au plasmide avec l'outil SAMtools et *bamToFastq*. Les séquences ainsi récupérées sont alignées en utilisant l'outil BWA contre les séquences des chromosomes hg38 plus celles des transposons flanqués par les régions plasmidiques de 100 pb des deux côtés du transposon NeoR. Le fichier *bam* correspondant à un format de taille réduite sans perte d'informations des données brutes, résultant de l'alignement de chaque ensemble de données est analysé avec le logiciel Lumpy afin d'identifier les séquences fractionnées. Lumpy est un outil permettant de mettre en évidence des différences de structure des chromosomes connues sous le nom de variants structuraux (SV), un terme qui englobe diverses altérations génomiques, notamment la délétion, la duplication, l'insertion, l'inversion, la translocation ou le réarrangement complexe de segments relativement importants (par exemple >100 pb). Les variants structuraux ayant deux positions, l'une génomique et l'autre sur le

transposon ont été extraites en utilisant un script codé en python créé par **Laura Helou**.

## V. Analyse statistique

Les tests statistiques ont été réalisés avec le logiciel PRISM6 (test-t pour comparer deux distributions, test F pour comparer la variabilité de deux échantillons), des macros d'Excel (Chi2 pour comparer une distribution observée à une distribution théorique ; distribution logarithmique pour vérifier l'enrichissement ou la déplétion d'une variable dans une sous population vis à vis de la population dans son entier ; Kruskal-Wallis pour comparer deux distributions) et en shell sous linux (test de permutations pour vérifier si le nombre de co-occurrences entre deux annotations est lié ou non au hasard). Les graphiques ont été réalisés avec le logiciel PRISM6, et à partir de <http://bioinformatics.psb.ugent.be/webtools/Venn/> pour les diagrammes de Venn.

# Résultats

---

Du fait de l'absence d'anticorps anti-PGBD5 valides disponibles, nous avons utilisé deux approches différentes dans notre recherche de localisation de la protéine PGBD5 dans les tissus sains. La première stratégie a consisté à analyser les données transcriptomiques disponibles pour la protéine PGBD5. La seconde stratégie a été de suivre l'expression du gène *pgbd5* grâce au gène rapporteur *LacZ* introduit dans un intron par transgénèse dans une lignée de souris mutante.

## I. Localisation de l'expression du gène *pgbd5* dans des tissus sains murins

### I.1. Exploitation des données de transcriptomiques disponibles dans les bases de données

Selon le site GTEx (<https://www.gtexportal.org/home/gene/ENSG00000177614> ) et le site BioGPS (<http://ds.biogps.org/?dataset=GSE10246&gene=209966> ), les ARNs *pgbd5* sont détectés principalement dans certaines parties du cerveau telles que le cortex préfrontal, le cortex cérébral, l'amygdale, l'hippocampe, le bulbe olfactif, le cervelet, le noyau accumbens et le stratum dorsal. Ils sont également détectés au niveau de la rétine, et dans une moindre mesure dans le testicule et l'ovaire chez la souris (**Figure 21**). C'est également le cas chez l'Homme où les ARNs *pgbd5* sont détectés principalement dans le cerveau. La présence des ARNs *pgbd5* au niveau du testicule est quant à elle, faible. (**Figure 22**). Des recherches dans la banque GEO profiles du NCBI (<https://www.ncbi.nlm.nih.gov/geoprofiles/?term=PGBD5>) ont permis de mettre en évidence la présence d'ARNs *pgbd5* de manière systématique dans le cerveau chez le cochon, le poisson zèbre, le poulet, la vache et le mouton. Pour ces deux dernières espèces, des transcrits de *pgbd5* sont aussi retrouvés dans le testicule.

Afin de déterminer plus précisément où les ARNs *pgbd5* sont transcrits dans le cerveau, nous nous sommes basés sur les données d'hybridation *in situ* sur ARN fournies chez la souris par l'Allen Brain Atlas (<http://mouse.brain-map.org/gene/show/84164>). Les transcrits *pgbd5* sont détectés principalement au niveau de l'hippocampe, des couches corticales du cortex cérébral, du bulbe olfactif, de l'hypophyse et du cervelet (**Figure 23**).

En partant de ces observations, nous avons décidé de concentrer nos travaux sur ces cinq régions du cerveau et sur le testicule.

## ***1.2. Expression de l'activité de la $\beta$ -galactosidase comme marqueur de la transcription du gène *pgbd5* dans un modèle de souris génétiquement modifié***

Les sites de données de protéomiques tel que The Human Protein Atlas indique la présence de la protéine dans les tissus ou cellules en se basant sur la détection des ARNs *pgbd5* (<https://www.proteinatlas.org/ENSG00000177614-PGBD5>). Du fait de l'absence d'anticorps anti-PGBD5 utilisable en Western Blot et immunohistologie nous avons choisi une méthode indirecte pour déterminer les sites d'expression de *pgbd5* chez la souris en utilisant un gène rapporteur inséré dans le gène de PGBD5. Nous possédons depuis avril 2018, une lignée de souris transgénique présentant une insertion de plusieurs sites Lox P, FRT de part et d'autre de l'exon 4 et le gène rapporteur *lacZ* dans l'intron 3 du gène codant pour la protéine PGBD5. Cette lignée a été créée par le Plateau de Biologie Expérimentale de la Souris (ENS de Lyon) à partir de cellules ES dénommées  $Pgbd5^{tm1a(KOMP)wsti}$  produites par le KOMP Repository (UC Davis) (c.f. la section matériel et méthodes **figure 14**). Cette lignée de souris possède le gène rapporteur *lacZ* codant pour une enzyme bactérienne : la  $\beta$ -galactosidase qui est facilement détectable sur des coupes de tissus. L'activité de cette enzyme est détectée par sa capacité à couper un substrat chromogène le 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal). Une coloration bleue apparaît signe de la présence et de l'activité de la  $\beta$ -galactosidase. Comme la transcription du gène *lacZ* est sous le contrôle du promoteur du gène *pgbd5*, la  $\beta$ -galactosidase est théoriquement le reflet de l'expression de la protéine PGBD5. Cette approche a été utilisée pour rechercher l'expression de la  $\beta$ -galactosidase dans le cerveau et le testicule aussi bien chez des souris homozygote ( $Pgbd5^{tm1a(KOMP)}_- / Pgbd5^{tm1a(KOMP)}_-$ ) ou hétérozygote ( $Pgbd5^{tm1a(KOMP)}_- / Pgbd5^{tm1a(KOMP)}_+$ ).

### ***1.2.a. Expression de l'activité de la $\beta$ -galactosidase dans le cerveau***

Un cerveau de souris  $Pgbd5^{tm1a(KOMP)}_- / Pgbd5^{tm1a(KOMP)}_+$  a été prélevé, fixé, puis l'activité de la  $\beta$ -galactosidase est révélée en présence du substrat X-Gal. Un cerveau de souris de génotype sauvage ( $Pgbd5^{tm1a(KOMP)}_+ / Pgbd5^{tm1a(KOMP)}_+$ ) a également été analysé comme contrôle.

L'hydrolyse du substrat X-Gal par la  $\beta$ -galactosidase permet d'observer une coloration bleue au niveau du néocortex pour le cerveau mutant alors que le cerveau contrôle ne présente aucune coloration (**Figure 24**).

Une analyse sur coupes histologiques de cerveau de souris  $Pgbd5^{tm1a(KOMP)}_- / Pgbd5^{tm1a(KOMP)}_+$  montre l'expression de la  $\beta$ -galactosidase dans très peu de structures du cerveau. L'activité  $\beta$ -

galactosidase est localisée dans l'hippocampe (**Figure 25A**), on observe une coloration bleue au niveau de la corne d'Ammon et du gyrus denté.

Dans le cortex cérébral (**Figure 25B**), le signal bleu reflétant l'activité du promoteur *pgbd5* est localisé de manière diffuse et homogène dans les couches II à V. La couche I et VI du cortex cérébral ne présentent pas d'activité de la  $\beta$ -galactosidase.

Au niveau du bulbe olfactif (**Figure 25C**), on observe activité de la  $\beta$ -galactosidase dans une région délimitée au cœur du bulbe olfactif.

Par contre aucune activité de la  $\beta$ -galactosidase n'est détectée dans cervelet (**Figure 25D**), et l'hypophyse (**Figure 25E**).

### ***1.2.b. Expression de l'activité de la $\beta$ -galactosidase dans le testicule***

Au niveau du testicule entier de souris  $Pgbd5^{tm1a(KOMP)} - / Pgbd5^{tm1a(KOMP)} +$ , l'activité de la  $\beta$ -galactosidase est localisée dans au niveau du tissu interstitiel situé entre les tubes séminifères (**Figure 26**).

L'analyse sur coupe histologique montre que l'activité de la  $\beta$ -galactosidase est limitée aux cellules de Leydig ainsi qu'aux spermatogonies ; cellules souches de la spermatogénèse. Elles sont localisées au niveau de la membrane basale du tube séminifère (**Figure 27**).

### ***1.2.c. Conclusions sur les résultats obtenus en expression avec les souris mutées $Pgbd5^{tm1a(KOMP)}$***

Les résultats de localisation de l'expression du gène *pgbd5* dans le cerveau sont concordants avec les données fournies par l'Allen Mouse Brain Atlas où les ARNs *pgbd5* ont été mis en évidence dans ces différentes structures. Pour le cervelet, des ARNs sont présents selon ces sources de données mais nous ne détectons pas l'activité du promoteur du gène *pgbd5* par cette méthode.

Ces résultats nous ont permis de sélectionner le matériel biologique à utiliser pour la réalisation des ChIP-séquençage. Dans un souci de facilité de prélèvement et afin de limiter les coûts du séquençage, deux sources d'échantillon ont été utilisées pour la suite. Notre premier choix se portant sur l'hippocampe en séparant la partie dorsale de la partie ventrale. En effet, bien que ces deux parties partagent des similitudes anatomiques (telle que la corne d'Ammon se trouvant dans ces deux parties) ; la partie dorsale est composée en partie du gyrus denté, qui est absent dans la partie ventrale. Notre deuxième choix s'est porté sur le testicule où seules les cellules de Leydig ont été utilisées pour le ChIP-séquençage.

## II. Caractérisation des sérums murins anti-PGBD5

Il y a trois ans, au début de ce travail, seule une douzaine d'anticorps anti-PGBD5 étaient commercialisés. Ces anticorps étaient décrits comme reconnaissant une protéine de 100 à 130 kDa. Or, aucune des quatre isoformes de PGBD5 présentes chez l'Homme et des deux isoformes présentes chez les autres vertébrés n'ont un poids moléculaire supérieur à 100 kDa. Nous nous sommes donc permis de douter de la validité de ces anticorps. Nous avons cependant vérifié leurs qualités en Western Blot. Au vu des résultats négatifs, nous avons décidé de faire produire des sérums murins par une méthode autre que l'immunisation de souris ou lapin par un peptide de PGBD5. Nous avons utilisé la méthode d'immunisation génique de la société In-Cell-Art qui repose sur l'utilisation de Nanotaxi® qui induisent à la fois une très forte expression de l'antigène mais aussi une stimulation importante du système immunitaire inné grâce à un mécanisme de délivrance unique. Au terme du protocole d'immunisation, dix sérums murins ont été réceptionnés puis caractérisés.

### *II.1. Recherche d'anticorps anti-PGBD5 dans les sérums murins et les anticorps commerciaux*

#### *II.1.a. Recherche d'anticorps reconnaissant des épitopes séquentiels*

Les dix sérums murins ainsi que les douze anticorps commerciaux produits chez le lapin et la souris ont été testés par Western Blot afin de déterminer la présence ou non d'anticorps reconnaissant des épitopes séquentiels (ou linaires), c'est-à-dire reconnaissant une séquence d'acides aminés continue dans la séquence de PGBD5. Des lysats de cellules HeLa transfectées transitoirement avec des vecteurs pCS2 exprimant chacun une des quatre isoformes de PGBD5 fusionnée à un tag MYC ont été utilisés. La présence du tag MYC nous a permis de vérifier l'efficacité de transfection et la production de la protéine de fusion Myc-PGBD5. La protéine MYC a été détectée par Western Blot en utilisant un anticorps anti-MYC.

## **Qualité des anticorps commerciaux**

Aucun des anticorps commerciaux n'a permis de mettre en évidence la protéine PGBD5. La plupart de ces anticorps détectent une protéine d'environ 130 kDa (**Figure 28**), qui ne correspond à aucune des tailles attendues de nos 4 isoformes contrairement à l'anticorps anti-MYC qui révèle les différentes isoformes aux tailles attendues. La bande de 130 kDa détectée par les anticorps commerciaux est également retrouvée dans les cellules HeLa non transfectées. Or le RNAseq réalisé par l'équipe sur des cellules HeLa ne montre aucun transcrite codant une ou plusieurs des isoformes de PGBD5 (résultats en accord avec ceux disponibles sur le site <https://amp.pharm.mssm.edu/Harmonizome/dataset/CCL+Cell+Line+Gene+Expression+Profiles>). La protéine PGBD5 n'est donc pas exprimée dans les cellules HeLa. Nos résultats démontrent que ces anticorps commerciaux anti-PGBD5 ne sont pas spécifiques.

## **Caractérisation des sérums anti-PGBD5 produits en utilisant la méthode de nanotaxi**

Le nombre de sérums produits par la société In-Cell-Art est important (10 souris avec 5 prélèvements au cours du temps) mais la quantité disponible par sérum est limitée (< 500 µL).

Les dix sérums murins arrivés au terme des 4 immunisations (prélevés à 71 jours) ont été analysés par Western Blot sur des extraits de cellules exprimant l'isoforme Mm523. (**Figure 29**). Les témoins du Western blot sont conformes à nos attentes avec une absence de signal dans les cellules HeLa non transfectées et une bande nette dans le témoin positif à environ 90 kDa avec l'anticorps commercial anti-MYC correspondant au poids moléculaire de l'isoforme Mm523. Aucun des dix sérums testés ne montre de signal dans cette condition avec des extraits de cellules Hela transfectées avec le plasmide exprimant la protéine de fusion MYC-Mm523 PGBD5.

Lors d'une immunisation à ADN, la cinétique de production d'anticorps peut être variable en fonction d'au moins trois facteurs: l'agent immunogène, le fond génétique des animaux et la variation entre individus [97]. Les sérums prélevés à J42 lors de la seconde immunisation ont été testés dans les mêmes conditions qu'à J71. Seuls les sérums SWISS 3 et SWISS 4 à J42 présentent une faible bande de même poids moléculaire que la bande détectée par l'anticorps anti-MYC (**Figure 30**). Les sérums SWISS 3 et SWISS 4 prélevés lors du monitoring à J21 et à J56 ont été également analysés. Une faible bande correspondante à la taille attendue de la protéine de fusion MYC-PGBD5 est également détectée pour ces sérums (**Figure 31**). Les sérums provenant des souris SWISS 3 et SWISS 4 contiennent donc possiblement des anticorps capables de reconnaître l'isoforme Mm523 de PGBD5. La cinétique de production d'anticorps semble être constante de J21 à J56 chez ces deux souris, l'absence de détection avec les sérum SWISS 3 et SWISS 4 à J71 paraît donc suspecte.

Les fractions J71 SWISS 3 et SWISS 4 ont été retestées sur des extraits de cellules HeLa transfectées avec un plasmide d'expression de l'isoforme Mm409 fusionnée avec la protéine MYC (**Figure 32**). On observe que la protéine de fusion est détectée avec les sérums J71 des souris SWISS 3 et SWISS 4. Les fractions J71 de ces deux sérums possèdent des anticorps anti-PGBD5 pouvant détecter l'isoforme Mm409 de PGBD5.

Les sérums des souris SWISS3 et SWISS4 permettent une détection spécifique des isoformes Mm409 et Mm523 de la protéine PGBD5, bien que celle-ci soit de faible intensité.

### ***II.1.b. Recherche d'anticorps conformationnels***

Les sérums ont ensuite été testés en dot blot sur membrane en utilisant des extraits protéiques non dénaturés de cellules HeLa transfectées ou non avec le plasmide d'expression de l'isoforme Mm523 de PGBD5 fusionnée à MYC. L'objectif de cette analyse, plus simple et plus rapide à mettre en œuvre vu la quantité de sérums à tester, était de définir si les sérums contenaient des anticorps capables de se lier à des épitopes conformationnels sur la protéine native.

Un premier Dot Blot a été réalisé avec les fractions J71 des sérums murins. Nous avons comparé le signal de fluorescence obtenu lors d'une hybridation avec le sérum pré immun à celui obtenu avec le sérum anti PGBD5 J71 de chaque souris immunisée (**Figure 33**). Quatre sérums ont retenu notre attention, les sérums C57/BL6-4, SWISS 3, SWISS 4 et SWISS 5. Ces sérums J71 présentent un signal en Dot Blot important par rapport à leur contrôle (sérum pré-immun). Ce résultat indique que ces sérums contiennent des anticorps anti-PGBD5 dirigés contre des épitopes conformationnels.

Un autre Dot Blot sur les fractions de sérums J21, J42 et J56 a été réalisé pour déterminer la cinétique de production de ces anticorps anti-PGBD5. Malheureusement aucune de ces autres fractions ne s'est avérée présenter une différence de signal suffisamment significative par rapport à leurs sérums pré-immun comme les fractions J71 des sérums C57/BL6-4, SWISS 3, SWISS 4 et SWISS 5 pour être considérée comme utilisable en ChIP.

### ***II.1.c. Conclusion sur la recherche d'anticorps anti-PGBD5***

Les sérums issus de l'immunisation des souris SWISS 3 et SWISS 4 présentent des anticorps capables de reconnaître des épitopes séquentiels et/ou conformationnels spécifiques des isoformes Mm409 et Mm523 de PGBD5. Pour la suite du projet, les fractions J71 des sérums provenant des souris SWISS 3 et SWISS 4 ont été réunis et sont utilisés comme sérum anti-PGBD5 pour la suite.

## ***II.2. Utilisation du sérum anti-PGBD5 en immunofluorescence pour localiser la protéine PGBD5 dans des cellules***

Le sérum anti-PGBD5 a été testé en immuno-fluorescence sur des cellules HeLa transfectées avec les plasmides d'expression de l'isoforme PGBD5 Mm409 ou PGBD5 Mm523 ainsi que sur des lignées de cellules cancéreuses exprimant PGBD5. Ces expériences ont été effectuées dans l'objectif de vérifier si les anticorps dirigés contre des épitopes conformationnels pouvaient être utilisés en histologie et sur cellules fixées afin de localiser les sites d'expression de la protéine PGBD5 dans les cellules tumorales.

### ***II.2.a. Localisation de la protéine PGBD5 dans les cellules HeLa exprimant l'isoforme Mm409 ou Mm523 de PGBD5***

Lors de l'observation en microscopie confocale, les cellules transfectées avec un plasmide pCS2 "vide" ont été utilisées comme témoin négatif. Les réglages du microscope confocal ont été réalisés sur ces cellules afin de régler le bruit de fond (**Figures 34 A et B**). A partir de culture de cellules HeLa transfectées avec les plasmides exprimant les isoformes Mm409 ou Mm523 fusionnées à la protéine MYC, nous avons observé un marquage cytoplasmique localisé à la périphérie du noyau avec le sérum anti-PGBD5 (flèches vertes **figures 34 C et E**). Avec l'anticorps anti-MYC, on observe également un marquage cytoplasmique (**Figures 34 D et F**), le marquage pour l'isoforme PGBD5 Mm523 est préférentiellement localisé à un pôle du cytoplasme (**Figure 34 F**) alors qu'il est plus diffus dans tout le cytoplasme pour l'isoforme PGBD5 Mm409 (**Figure 34 D**).

### ***II.2.b. Localisation de la protéine PGBD5 dans les lignées cellulaires cancéreuses***

Le sérum anti-PGBD5 ainsi que son sérum pré-immun ont également été utilisés sur cinq lignées de cellules cancéreuses. Pour trois lignées cellulaires (H4, 8MG et T98G) la présence d'ARN codant PGBD5 a été détectée lors d'analyse de données de RNA-seq disponibles dans les banques publiques (**Figure 35**). Les lignées HeLa et HEK293 ont été utilisées comme témoins négatifs. Le sérum pré-immun ne contenant pas d'anticorps anti-PGBD5 a été utilisé comme contrôle négatif et a servi aux réglages du bruit de fond du microscope confocal. Pour les cellules hybridées avec le sérum anti-PGBD5, nous avons pu observer pour les trois lignées de glioblastomes H4, 8MG et T98G, un marquage cytoplasmique plus ou moins diffus. Aucun marquage sur les lignées cellulaires HeLa et HEK293 n'a été observé (**Figure 36**). L'ensemble de ces résultats sont concordants avec nos attentes.

Ces cinq lignées sont utilisées par la suite afin de déterminer les sites de liaisons à l'ADN de la protéine PGBD5 dans un contexte tumoral.

### ***II.2.c. Conclusion au sujet de l'utilisation du sérum anti-PGBD5 en immunocytologie***

Le sérum anti-PGBD5 permet de détecter la présence de la protéine PGBD5 dans les cellules exprimant cette protéine. Cependant, dans cette méthode, le sérum anti-PGBD5 doit être utilisé à une faible dilution pour obtenir un signal. Ne disposant pas d'un volume initial très important pour ce sérum anti-PGBD5, les expérimentations consistant à détecter PGBD5 dans les cellules ont été limitées.

### ***II.3. Tentative de validation du sérum anti-PGBD5 par ChIP-séquençage en utilisant des cellules exprimant une V5-PGBD5***

Un ChIP-séquençage à partir d'un système cellulaire d'expression inductible où la protéine PGBD5 est couplée à un tag correspondant à l'épitope V5 a été réalisé en parallèle d'une immunoprécipitation avec le sérum anti-PGBD5. La présence de ce tag a permis d'immunoprécipiter la chromatine avec un anticorps commercial anti-V5 valide et ainsi de pouvoir comparer les données obtenues avec cet anticorps et ceux obtenus après immunoprécipitation avec le sérum anti-PGBD5.

#### ***II.3.a. Création de lignées cellulaires stables inductibles***

Lors des observations en microscopie, nous avons pu constater que toutes les cellules natives transfectées afin de sur-exprimer une des isoformes PGBD5 n'expriment pas de la même manière et avec la même intensité la protéine PGBD5. Cela peut être le reflet de variations de l'efficacité de la transfection transitoire. Chaque cellule ne localise pas le transgène de façon homogène et en quantité égale dans le noyau. Pour pallier à ce problème et obtenir des cellules exprimant une des isoformes de PGBD5 de façon plus homogène, nous avons entrepris la fabrication de lignées clonales cellulaires stables en utilisant des constructions plasmidiques [produites à partir de pSBtet\_GP (Addgene #60495)] permettant l'expression du gène *pgbd5* avec le système Tet-ON. Cette méthode implique l'intégration stable du plasmide pSBtet\_GP qui est un vecteur d'expression contenant un gène de résistance à l'antibiotique puromycine, ce qui permet de sélectionner les clones cellulaires ayant

intégré par transposition dans leurs chromosomes le plasmide de façon stable. Ce plasmide contient aussi le gène codant pour la luciférase ce qui permet de disposer d'un contrôle positif de l'expression inductible après intégration dans les chromosomes. Dans nos constructions, le gène codant la luciférase est remplacé le gène codant une des isoformes de PGBD5.

### **Vérification de la fonctionnalité du système d'induction par dosage de bioluminescence**

48h après transfection avec le plasmide pSB\_tet\_GP (puromycine +, luciférase+), les cellules HeLa sont sélectionnées dans un milieu DMEM supplémenté de 5 % de SVF sans tétracycline et 0,5 µg/mL de puromycine. Après une quinzaine de jours sous sélection, une dizaine de clones cellulaires sont récupérés puis amplifiés. L'efficacité de l'intégration et de l'expression inductible est vérifiée par mesure de l'activité luciférase. L'activité luciférase est mesurée dans des extraits des clones induits ou non par la doxycycline (analogue semi-synthétique de la tétracycline; **Tableau 13**). Des extraits de cellules HeLa non transfectées ont également été utilisés pour mesurer le bruit de fond. Les résultats présentés dans le tableau 11 démontrent que les clones de cellules HeLa sélectionnées expriment la luciférase 50 à 120 fois plus en présence de doxycycline que les cellules HeLa contrôle. Le système Tet-ON d'induction par la doxycycline est fonctionnel dans nos conditions expérimentales.

### **Vérification de l'induction des lignées HeLa pSBtet V5 isoformes PGBD5 par Western Blot**

Les différents clones de cellules transfectées et sélectionnées avec le plasmide expriment les isoformes Mm409 ou Mm523 ont été induites ou non pendant 24h avec 1 µg/mL de doxycycline. L'expression des isoformes de PGBD5 a été recherchée en Western Blot en utilisant l'anticorps anti-V5 (**Figure 37**). Aucune protéine PGBD5 n'a été détectée dans nos extraits de cellules HeLa non-induites et induites.

### **Vérification de la présence d'ADN *pgbd5* par PCR et de la synthèse d'ARN messager de *pgbd5* par RT-PCR semi-quantitative dans les lignées cellulaires**

Dans cette partie, seule l'isoforme Mm409 est présentée. L'isoforme Mm523 possède dans sa composition des structures quadruplex rendant difficile les amplifications de PCR.

Afin de vérifier si la transcription du gène *pgbd5* est effective dans les clones stables de cellules HeLa produites à partir du plasmide pSBtet\_V5\_Mm409, l'ADN génomique et les ARN totaux des

clones cellulaires induits et non induits par la doxycycline ont été extraits. Des PCR et RT-PCR semi quantitatives ont été réalisées afin de déterminer la présence du transgène *pgbd5* intégré et le niveau d'ARN messager codant l'isoforme Mm409 de PGBD5. Le transgène *pgbd5* Mm409 est détecté à la taille attendue dans l'ADN génomique des cellules HeLa\_pSBtet\_Mm409\_V5 que ce soit en conditions induites ou non induites ainsi que dans les plasmides pSBtet\_Mm409 et pSBtet\_Mm409\_V5 (**Figure 38**).

Au niveau de la transcription, les ARNs de l'isoforme Mm409 sont détectés en quantité plus importante après 30 cycles de PCR lorsqu'il y a une induction qu'en son absence (**Figure 39**). Ce résultat est attendu dans un système d'induction fonctionnel classique. Dans un système Tet-On, comme celui utilisé pour la création de ces clones cellulaires stables, l'expression du transgène n'est normalement pas attendu en absence d'induction. Il semblerait que ce ne soit pas le cas ici, malgré l'usage d'un sérum de veau foetal dépourvu de traces de tétracycline.

### ***II.3.b. Production de la protéine PGBD5 taguée en expression transitoire***

L'obtention de clones de cellules HeLa capables de synthétiser une isoforme de PGBD5 sous promoteur inductible étant impossible dans nos conditions, nous avons donc choisi de revenir à une approche utilisant des transfections transitoires. PGBD5 étant une protéine qui ne s'exprime que dans certains types de cellules (neurones et certaines cellules tumorales) ; quatre lignées cellulaires issues d'origines tumorales différentes ont été testées et transfectées par le plasmide d'expression pSBtet\_Mm409\_V5. En plus des lignées classiques cancéreuses HeLa et HEK293 provenant respectivement de cellules de cancer du col de l'utérus et de cellules cancéreuses embryonnaires de rein, les lignées cellulaires H4 et T98G issues de tumeurs cérébrales sont également testées. Le lendemain de la transfection, les cellules ont été induites pendant 24h par la doxycycline puis récupérées. Des extraits cellulaires transfectés non induites sont utilisés comme contrôle négatif. Les extraits cellulaires ont été analysés par Western Blot et les résultats sont présentés en **figure 40**. Sur les 4 lignées cellulaires testées, seules les cellules HEK293 transfectées par le plasmide pSBtet\_Mm409\_V5 expriment la protéine PGBD5\_Mm409 après induction de façon détectable dans nos conditions. L'isoforme Mm523 a également été testée en Western Blot après transfection transitoire des cellules HEK293 avec et sans induction à la doxycycline et a pu être détectée après induction des cellules par la doxycycline (**Figure 41**)

Au vu de ces résultats, un nouvel essai de fabrication de lignée cellulaire stable inductible a été entrepris à partir de la lignée cellulaire cancéreuse HEK293 mais celui-ci n'a pas permis l'obtention de clones produisant une expression de la protéine PGBD5 de façon stable et inductible.

### ***II.3.c. Résultats du ChIP-séquençage sur les cellules HEK293 transfectées par pSBtet\_Mm523\_V5 avec les anticorps anti-V5 et le sérum anti-PGBD5***

Dans cette partie, deux types de cellules sont utilisées : les cellules HEK293 transfectées de façon transitoire par le plasmide pSBtet\_ isoforme PGBD5\_V5 exprimant PGBD5 après induction et les cellules HEK293 transfectées transitoirement et non induites utilisées comme témoin négatif. L'immuno-précipitation de la chromatine de ces 2 types de cellules a ensuite été réalisée avec un anticorps anti-V5 et en parallèle avec le sérum anti-PGBD5. Après extraction de l'ADN, séquençage et analyse bio-informatique, nous avons pu identifier des sites de liaisons.

Le nombre de sites de liaison pour chaque immunoprécipitation de la chromatine est présenté dans le **tableau 14**. La comparaison des positions de sites de liaisons (pics de ChIP-seq) obtenus avec chacun des deux anticorps révèle que 210 pics de liaisons présentant une q-value inférieure à  $10^{-2}$  sont communs. Ce nombre est réduit à 12 pics lorsque la sélection des pics est effectuée avec une p-value inférieure à  $10^{-5}$ .

Pour déterminer si le nombre de sites de liaison commun est dû au hasard, des tests de permutations à régression linéaire ont été effectués. En utilisant 1 000 permutations, nous avons pu calculer que la probabilité q est de  $4 \cdot 10^{-4}$ . Au seuil de  $q > 0,05$ , notre conclusion est que l'obtention de 210 ou de 12 pics communs entre les deux conditions de ChIP-seq n'est pas liée au hasard. Elle reflète donc la reconnaissance d'une même protéine. Notre interprétation du faible nombre de pics communs entre les deux conditions est que dans chacune d'elle, les épitopes reconnus sont totalement différents et ils pourraient y avoir des présentations d'épitopes totalement différentes à chaque loci en fonction du complexe dans lequel est présent PGBD5.

### ***II.3.d. Conclusions sur la validation du sérum anti-PGBD5***

Les résultats obtenus en Western Blot et en Dot Blot montrent la présence d'anticorps anti-PGBD5 spécifiques de la partie commune (~ 400 acides aminés de la région C-terminale) entre les isoformes Mm409 et Mm523 dans le sérum anti-PGBD5. La comparaison des résultats de ChIP-séquençage obtenus avec l'anticorps anti-V5 et le sérum anti-PGBD5 démontrent que ces anticorps reconnaissent spécifiquement la protéine PGBD5. Le sérum anti-PGBD5 bien que peu concentré, est un sérum qui nous a semblé valide pour la détection de la protéine PGBD5 ; il a été utilisé pour le ChIP-séquençage à partir d'échantillons sains et tumoraux.

### III. Détermination des sites de liaison de PGBD5 dans la chromatine par ChIP-séquençage

Les analyses réalisées ont comme objectif d'identifier le ou les motifs de liaison à l'ADN d'une protéine dont on suppose qu'elle est capable de se lier à l'ADN. L'objectif final de ces analyses est d'établir une carte des sites de liaison de la protéine dans le génome d'une cellule d'un type donné afin de déterminer si les sites de liaison sont situés dans des gènes, des promoteurs ou des séquences non codantes. Une des hypothèses que nous avons formulée est que la protéine PGBD5 pourrait se fixer sur ou à proximité des transposons de type *pbles* humains. En effet, il a été précédemment montré que PGBD5 est capable de se lier à un *pble* phylogénétiquement très éloigné, *Ifp2*, et nous avons donc émis l'hypothèse qu'elle pourrait se lier et mobiliser les *pbles* plus étroitement apparentés sur le plan phylogénétique, tels que les *pbles* humains MER75, MER85 et *Looper*. Il faut noter que dans le génome de la souris, seul le *pble Looper* est présent. L'analyse des résultats a porté dans un premier temps uniquement sur la recherche de co-localisation entre des sites de liaisons de PGBD5 avec les *pbles* MER75, MER85 et *Looper* présents.

Les sites de liaisons de PGBD5 à l'ADN dans les tissus sains ont été ainsi recherchés chez la souris dans l'hippocampe dorsal (HD) et ventral (HV) et dans les cellules de Leydig. Chez l'homme, ils ont été recherchés dans 6 lignées de cellules cancéreuses.

#### III.1. Sites de liaisons de PGBD5 à l'ADN dans des tissus sains murins

Les nombres de sites de liaisons obtenus pour le ChIP-séquençage à partir de biopsies d'hippocampe et des cellules de Leydig sont récapitulés dans le **tableau 15**.

Pour l'hippocampe, les nombres de sites de liaisons obtenus pour chacune des parties ventrale et dorsale sont équivalents. Environ 850 pics sont détectés dans chaque condition mais aucun de ces pics ne colocalise avec le *pble Looper* (seul *pble* présent chez la souris). La comparaison de l'amplitude et de la taille des pics en paire de bases n'a révélé aucune différence majeure entre l'hippocampe dorsal et l'hippocampe ventral. L'amplitude moyenne du signal des pics de liaisons est similaire entre les deux types d'échantillons. La taille moyenne des pics est quant à elle, légèrement plus importante pour les échantillons provenant de l'hippocampe ventral (**Figure 42**). Parmi les 850 séquences obtenues dans les deux échantillons d'hippocampe, un seul pic commun a été trouvé. La

recherche de motifs partagés dans les séquences chromosomiques couvertes par les pics a été effectuée avec les logiciels GLAM2, RSAT et MEME-ChIP. Elle n'a pas permis de mettre en évidence de motifs communs et statistiquement significatifs.

Concernant les pics mis en évidence dans les cellules de Leydig, les 432 séquences chromosomiques couvertes par les pics ne co-localisent pas avec un élément *Looper*. De même, aucun motif commun et statistiquement significatif n'a été mis en évidence avec les outils GLAM2, RSAT et MEME-ChIP.

### ***III.2 Sites de liaisons à l'ADN de PGBD5 dans des cellules cancéreuses humaines***

L'identification des sites chromosomiques de liaison à l'ADN de la protéine PGBD5 chez l'Homme a été réalisée sur des lignées cellulaires humaines dérivées de 3 glioblastomes: H4, 8MG et T98G. Ces lignées ont pour origine des cellules souches neuronales cérébrales qui potentiellement expriment PGBD5. La lignée HCT116 dérivée d'un cancer colorectal a également été analysée. Cette lignée exprime la protéine PGBD5 selon les données fournies par le site du NCBI (**Figure 43**). Les deux lignées cancéreuses HeLa et HEK n'exprimant pas PGBD5 sont utilisées comme contrôle négatifs.

Après analyse qualitative puis nettoyage des données de séquences, une recherche de pics a été effectuée avec l'outil PePr. Comme précédemment, seuls les pics de liaison avec un seuil de significativité inférieur à ( $q < 0,05$ ) ont été conservés (**Tableau 16**). Pour les témoins négatifs, aucun pic n'a été détecté. Ce résultat était attendu puisque PGBD5 n'est pas exprimé dans ces cellules. Pour les deux lignées de glioblastomes H4 et 8MG, aucun pic n'a également été détecté. Par contre pour les lignées cellulaires cancéreuses T98G et HCT116, 283 et 1 306 pics de liaison ont été mis en évidence respectivement avec une q-value  $< 0.05$ .

#### ***III.2.a. Analyse des séquences obtenues pour la lignée cancéreuse T98G***

Dans cette lignée nous avons identifié 283 pics de liaison. L'analyse des séquences chromosomiques couvertes par ces pics avec les logiciels GLAM2, MEME-ChIP-seq et RSAT montre que 66 séquences sont principalement constituées de motifs répétés en tandem de séquence (TGGAA)<sub>n</sub> (**Figure 44**). Ces 66 sites correspondent à des loci composés d'ADN satellite qui sont localisés dans les séquences paracentromériques de plusieurs chromosomes chez l'Homme. Pour les 172 séquences restantes, 15 séquences correspondent à des séquences microsatellites présentes principalement dans les centromères, les télomères et dans certaines régions d'hétérochromatine proche de la membrane

nucléaire. Des tests de permutations ont été effectués afin de déterminer si les séquences détectées étaient spécifiques à la protéine PGBD5 ou s'ils auraient pu être obtenus par chance. La probabilité de 0,074 nous indique que ces séquences ne sont pas statistiquement spécifiques à la protéine PGBD5.

Certaines séquences de liaison de PGBD5 co-localisent avec les 3 *pbles* humains: MER 75, MER85 et *Looper*. Pour MER75, 58 séquences de liaison sont présentes dans celui-ci, 9 séquences sont présentes dans MER85 et 58 séquences sont présentes dans *Looper*. Il est cependant important de noter que le nombre de séquences co-localisant avec les *pbles* recherchés lors de l'analyse de ce ChIP-seq est très faible par rapport aux nombres de *pbles* présents dans le génome humain. En effet, plus de 1 000 annotations sont connues pour chacun de ces 3 *pbles* chez l'Homme. Des tests de permutations ont cependant été réalisés pour vérifier si ces co-localisations étaient liées au hasard. Dans les trois cas de *pbles*, les résultats des tests sont sans ambiguïté ( $p < 1.e^{-100}$ ). Ils indiquent que la co-localisation des pics de liaisons de PGBD5 avec les 3 *pbles* n'est pas liée au hasard. Les 77 séquences restantes n'ont pour l'instant pas de motif connu dans les fichiers d'annotation que nous avons utilisés.

### ***III.2.b. Analyse des séquences obtenues pour la lignée cancéreuse HCT116***

Sur les 1306 séquences de liaisons de PGBD5 détectées pour la lignée de cancer colorectal HCT116, seuls 10 séquences ont pu être identifiées comme des séquences d'ADN satellite ou microsattelites (**Figure 45A**). Les 1296 séquences restantes n'ont pas de motifs connus dans les fichiers d'annotation que nous avons utilisés.

Aucune des séquences de liaisons de PDBD5 ne localise dans les *pbles* MER75, MER85 ou *Looper*.

### ***III.2.c. Ré analyse des séquences disponibles obtenues pour la lignée cancéreuse G401 par Henssen et al (2017)***

Les données de ChIP-séquençage produites par Henssen et al [92] à partir d'une lignée cancéreuse cellulaire G401 ont été ré analysés. La lignée G401 est dérivée d'une tumeur de Wilms dont l'origine est localisée dans les cellules immatures (embryonnaires) du rein. L'immunoprécipitation de la chromatine de ces cellules a été effectuée en 2016 avec un anticorps commercial anti-PGBD5 orb13159 (Biorbyt, Berkeley, CA, USA). Cet anticorps n'a malheureusement pas pu être testé dans le cadre de ce travail car il est non disponible depuis 2017. Les données produites par cette équipe ont été analysées de la même manière que nos données. Sur les 618 pics détectés avec ces cellules, 77 séquences correspondent à des motifs répétés en tandem (TGGAA)<sub>n</sub> et 109 séquences correspondent à des séquences microsattelites ou de l'ADN satellite. Aucune des séquences ne co-localise avec les *pbles* MER75, MER85 et *Looper* (**Figure 45B**).

### ***III.2.d. Conclusions à partir des résultats de ChIP-séquençage sur cellules cancéreuses***

On observe des différences dans le type de séquences de liaison de la protéine PGBD5 dans les lignées cancéreuses T98G, HCT116 et G401. Les lignées de cellules cancéreuses HCT116 et G401 dérivent respectivement de cellules souches coliques et rénales. La lignée de cellules T98G dérive de cellule souche neuronale.

Les résultats provenant de la lignée cellulaire T98G montrent que PGBD5 est capable de se lier au niveau de quelques *pbles* MER75, MER85 et *Looper* présents dans le génome humain. Par contre cela ne semble pas être le cas dans les neurones de l'hippocampe chez la souris avec les éléments *Looper*.

## **IV. Détermination de l'activité de transposition de PGBD5 vis-à-vis des *pbles* humains**

D'après les travaux de Henssen et al. 2015, PGBD5 est une transposase active pouvant médier des réarrangements chromosomiques. Nos résultats montrent que dans les cellules T98G, PGBD5 est capable de se lier sur l'ADN d'un faible nombre de *pbles*. Nous avons donc voulu déterminer si la protéine PGBD5 est également capable de les transposer, c'est-à-dire, de les exciser puis de les intégrer dans une autre région chromosomique de l'ADN.

### ***IV.1 Résultats des tests d'intégration aléatoire***

Les tests d'intégration aléatoire permettent de déterminer si des paramètres tels que la toxicité d'une transposase ou la toxicité des insertions aléatoires impactent sur le résultat obtenu.

Pour déterminer si les isoformes Mm409 et Mm523 de PGBD5 ont un effet toxique, nous avons tout d'abord co-transfecté un plasmide porteur d'un gène de résistance à la néomycine (pCS2-sv40-NeoR) avec i) soit un plasmide servant de source de protéine inerte (la « green fluorescent protein » (GFP) exprimée à partir du plasmide d'expression mammalien pCS2-CMV-GFP) qui ne possède pas de capacité de transposition (ou d'induction de la recombinaison) et qui sera utilisé comme contrôle

négatif, ii) soit un plasmide source de transposase PGBD5 (pCS2-CMV-isoforme PGBD5) (**Figure 46**). Le but étant de comparer le taux d'intégration d'une cassette NeoR qui confère aux cellules une résistance à la néomycine en présence ou non de transposase. Après sélection par la néomycine des cellules ayant acquis la résistance, les clones cellulaires résistants sont quantifiés. Le nombre de clones résistants en présence de transposase est comparé à celui obtenu avec les cellules transfectées par le plasmide témoin codant pour la GFP. Ce dernier étant fixé arbitrairement à 1.

Si le ratio entre le nombre de clones cellulaires obtenus avec chaque protéine transposase et celui obtenu avec la protéine GFP est proche de 1, il sera alors considéré que la transposase n'a pas d'effet. Si le ratio est inférieur à 1, la transposase est supposée avoir un effet toxique sur la survie des cellules. Enfin si le ratio est supérieur à 1, la transposase a un effet actif sur l'intégration aléatoire de la cassette NeoR.

Les résultats obtenus sont présentés dans le **graphique 1**. On observe qu'avec les deux isoformes de transposase PGBD5 Mm409 et Mm523, les ratios sont significativement inférieurs à 1, le nombre de clones obtenus en présence de transposase est environ 1,5 à 3 fois inférieur respectivement au nombre de clones obtenus lors d'une intégration sans transposase (en présence de la protéine inerte GFP). Les isoformes Mm409 et Mm523 ont donc un effet cytotoxique sur les cellules dans lesquelles elles sont exprimées. Un effet similaire avait été observé en cellules HEK293 avec des conditions d'expression identique (promoteur CMV) avec l'isoforme humaine orthologue de Mm523, Hs524 [92].

## **IV.2 Résultats des tests de transposition**

### **IV.2.a. Effet des isoformes Mm409 et Mm523 sur la transposition du transposon *ifp2***

A la différence du test d'intégration aléatoire, le plasmide porteur du gène de résistance à la néomycine est flanqué en 5' et en 3' par les extrémités du transposon *Ifp2* qui contiennent chacune un ITR. Les ITR sont les sites reconnus par la transposase *pble* pour initier la transposition. Dans l'article de Henssen et al [87], les auteurs mettent en évidence la capacité de PGBD5 à transposer *ifp2* dans les cellules HEK293. Nous avons vérifié en utilisant le même transposon que les isoformes de PGBD5 Mm409 et Mm523 sont capables de le transposer dans les cellules HeLa. Les résultats de ces tests sont présentés dans le **graphique 2**. En présence des isoformes de PGBD5 le nombre de clones cellulaires résistants à la néomycine est significativement inférieur au nombre de clones cellulaires obtenus avec la protéine inerte témoin GFP. Les isoformes Mm409 et Mm523 de PGBD5 sont

cytotoxiques et cette caractéristique masque donc les éventuelles intégrations du transposon *Ifp2* par transposition dans ce test.

#### ***IV.2.b. Effet de l'isoforme Mm523 sur la transposition des pbles MER75, MER75B, MER85 et Looper***

Chaque test étant réalisés en triplicata, seule l'isoforme Mm523 est testée afin de limiter la charge de travail. Ici, nous avons vérifié l'effet de l'isoforme Mm523 de PGBD5 sur la transposition des éléments *pbles* humains MER75, MER85 et Looper. Les plasmides pGH\_pble PGBD5\_NeoR contenant le *pble* à tester avec le gène de résistance à la néomycine ont été utilisées comme plasmides sources pour les tests de transposition. Chaque *pble* est testé en parallèle sans transposase avec la protéine inerte GFP. Les résultats sont représentés sous forme de ratio dans le **graphique 3** en utilisant les mêmes modalités que pour les graphiques 1 et 2. Avec l'isoforme Mm523 de PGBD5 le nombre de clones résistants obtenus pour chaque *pble* est significativement inférieur au nombre de clones cellulaires obtenus avec le témoin GFP. Cette diminution varie de 1,4 fois pour le *pble* MER85 à environ 2,5 fois pour *Looper*. La cytotoxicité de l'isoforme Mm523 de PGBD5 masque ici encore les éventuelles intégrations du transposon MER75, MER 85 et *Looper* par transposition dans ce test.

#### ***IV.2.c. Conclusion générale des tests***

On observe dans les tests réalisés un nombre de clones cellulaires résistants à la néomycine plus faible en présence des isoformes Mm409 et Mm523 que dans le témoin avec la protéine GFP. Cela démontre que les deux isoformes Mm409 et Mm523 ont un effet négatif sur la survie des cellules. Cet effet cytotoxique rend les tests ininterprétables dans nos conditions en terme d'évaluation du taux de transposition.

#### ***IV.2.d. Mise en évidence de la transposition des pbles par la méthode de LAM-PCR***

Malgré le problème de cytotoxicité des transposases Mm409 et Mm523, nous avons recherché dans les clones obtenus la présence d'événements de transposition en utilisant la méthode de LAM-PCR (*Linear Amplification-Mediated PCR*) qui est une stratégie de choix pour identifier les sites d'insertions d'un transposon. Dans cette perspective, des tests de transpositions ont été réalisés en cellules HeLa et des populations de clones cellulaires ont été récupérées afin de i) purifier leur ADN chromosomique, ii) réaliser des LAM-PCR pour produire des populations de fragments contenant des sites d'insertion chromosomiques (c.a.d. des jonctions transposon-chromosome) et enfin iii) de séquencer ces populations de fragments à haut débit. En raison du coût de séquençage important,

seule l'isoforme Mm523 est testée avec les transposons *ifp2*, MER75, MER75B, MER85 et *Looper*. Les séquences obtenues après séquençage haut débit sont analysées et seules celles correspondantes à une jonction transposon / chromosome sont conservées. Les résultats pour chaque *pble* sont présentés dans le **tableau 17**.

Les séquences des jonctions en 5' et en 3' des transposons ont ensuite été analysées afin de rechercher si les séquences TIR (Terminal Inverted Repeat) et TSD (Target Site Duplication) des *pbles* sont présents après transposition (**tableau 18**). En effet, pour qu'un transposon soit capable d'être re-mobilisé pour une nouvelle transposition, ces deux séquences doivent être parfaitement conservées de part et d'autre du transposon après insertion. La transposase Mm523 est capable d'insérer le transposon *ifp2* dans les cellules HeLa. Cette transposition permet dans près de 10 % des cas de garder intact le TIR et le TSD d'une extrémité du transposon pour une future mobilisation. On observe également un résultat équivalent avec le *pble* MER75 contrairement à MER75B, MER85 et *Looper* où le pourcentage de conservation de ces deux séquences est de 2 à 4 % environ pour une extrémité du transposon. Plus de 80 % à 90 % des *pbles* ont leur TSD et de leur TIR incomplets après transposition, ce qui préviendra leur mobilisation de nouveau par la transposase.

#### ***IV.2.e. Recherche de motif de reconnaissance***

PGBD5 est a priori capable de mobiliser spécifiquement n'importe quel *pble*. Cela pose la question du motif reconnu spécifiquement dans les extrémités des *pbles*. Un motif commun dans les séquences primaires des extrémités de *pbles* a été recherché. Aucun motif commun n'a pu être mis en évidence au niveau des séquences primaires. La reconnaissance spécifique de la transposase pouvant reposer sur des caractéristiques communes aux *pbles* autre que sur la séquence primaire, nous avons donc recherché dans la séquence des extrémités de chaque *pble* les motifs pouvant adopter des structures secondaires (SII) et tertiaires (SIII), en utilisant des outils publiques de recherche (<https://nonb-abcc.ncifcrf.gov/apps/nBMST/default/>). Nous avons constaté que des répétitions inversées (IR : inverted repeat en anglais) d'au moins 6 nucléotides et séparées par au moins 4 nucléotides étaient présents aux deux extrémités de toutes les séquences de *pbles* (**Figure 47**). La distance entre ces IR et le TIR varie aux deux extrémités de 0 à 360 nucléotides, l'un des deux IR dans chaque *pble* étant toujours à proximité du TIR. Nous avons observé que les *pbles* peuvent soit présenter des IR de séquence identique soit des IR présentant des séquences différentes.

#### ***IV.2.f. Conclusions générales sur l'activité de transposition de Mm523 PGBD5***

Les résultats précédents montrent que l'isoforme Mm523 est capable de transposer le transposon *Ifp2* et les *pbles* humains MER75, MER75B, MER85 et *Looper*. Cependant la transposase Mm523 est peu efficace et ne permet pas une transposition ultérieure des transposons mobilisés lors de ces événements de transposition, les séquences des TIR et du TSD des transposons étant incomplètes après la transposition.

A défaut de reconnaître une séquence primaire dans les transposons *pbles*, l'isoforme Mm523 de PGBD5 semble reconnaître des séquences répétées inversées pouvant former des épingles à cheveux ou des structures cruciformes dans leurs configurations secondaires et tertiaires.

## V. Ré-analyse des données de ChIP-séquençage

Nous avons mis en lumière la présence de répétitions inversées (IR) dans la séquence des *pbles* testés. Au regard de cette nouvelle information, nous avons voulu vérifier si les différents sites de liaisons détectés dans nos expérimentations de ChIP-séquençage pouvaient co-localiser avec ces séquences répétées inversées (IR). Les résultats sont présentés dans le **tableau 19**.

### V.1 Dans les tissus sains

Chez la souris, la grande majorité des séquences de liaisons de PGBD5 obtenus dans l'hippocampe (845 / 851 pour HD et 864 / 873 pour HV) contiennent des IR. Des tests de permutation ont été faits afin de déterminer si cette co-localisation pouvaient être dues au hasard. Les p-value obtenues pour HD et HV montrent que cette co-localisation n'est pas le fruit du hasard (p-value :  $3,29.10^{-8}$  pour HD et  $9,544.10^{-7}$  pour HV). Concernant les cellules de Leydig, la majorité des séquences de liaisons de PGBD5 (420 / 432) co-localise avec des IR au seuil de  $p > 0,05$ , les tests de permutations montrent que cette co-localisation est spécifique ( $p = 0,032688$ ).

### V.2. Dans les cellules cancéreuses

Dans les lignées de cellules cancéreuses T98G, HCT116 et G401, le nombre de séquences co-localisant avec les IR sont majoritaires avec plus de 84 % des pics pour les cellules T98G et plus de 96 % des pics obtenus avec les lignées HCT116 et G401. Les tests de permutations permettent d'affirmer que les séquences de liaison de PGBD5 co-localisant avec des IR pour les cellules HCT116

et G401 ne sont pas dû au hasard. Cependant le nombre de séquence de liaison de PGBD5 qui co-localise avec une IR dans les cellules T98G n'est pas significatif. Le faible nombre de séquences détectées pour les cellules T98G ne permet pas de valider le résultat du test de permutation.

### ***V.3. Conclusions sur les sites de fixation de la protéine PGBD5***

Dans les tissus sains ou les cellules cancéreuses que nous avons analysé, la protéine PGBD5 se lie probablement sur une structure secondaire spécifique constituée de répétitions inversées d'au moins 6 nucléotides et séparées par au moins 4 nucléotides. Ces séquences répétées inversées pouvant former une structure en épingle à cheveu ou cruciforme dans certaines conformations particulières de l'ADN. Leurs rôles et leurs implications dans la transposition reste à déterminer.

# Conclusions et discussion

---

L'objectif premier de mon travail a été de vérifier la spécificité des anticorps anti-PGBD5 commercialisés et de caractériser un anticorps anti-PGBD5 capable de reconnaître spécifiquement la protéine PGBD5. Cet anticorps devait remplir plusieurs critères afin de pouvoir : i) permettre de déterminer par Western Blot les différentes isoformes présentes dans un tissu ou lignée cellulaire, ii) définir le profil d'expression de la protéine en immunohistochimie selon l'organe ou le tissu et iii) immunoprécipiter la protéine afin de localiser les sites de liaison de la protéine PGBD5 sur la chromatine et déterminer si la protéine PGBD5 se lie sur les transposons de types *pbles*.

La protéine PGBD5 possède une séquence très conservée entre espèces (> 94 % d'identité entre l'Homme, les rongeurs, les bovins, les caprins, les équins, les ovins et les porcins). Son gène, *pgbd5*, est soumis à une forte sélection et a donc probablement un rôle physiologique très important chez les vertébrés. Cette caractéristique peut compromettre la production d'anticorps anti-PGBD5. En effet, un organisme peut ne pas produire d'anticorps dirigés contre une protéine reconnue comme faisant partie du soi et PGBD5 est très fortement conservée dans toutes ces espèces.

Les anticorps anti-PGBD5 disponibles dans le commerce sont des anticorps produits par immunisation peptidique. Les Western Blot réalisés avec ces anticorps n'ont pas permis de détecter la protéine PGBD5. Nous avons donc fait produire des anticorps par la société In-Cell-Art. Cette société bio pharmaceutique est spécialisée dans le développement préclinique et pharmaceutique de solutions (vaccins ou thérapeutiques à base d'ADN/ARN) pour le traitement de maladies acquises ou héréditaires à l'aide de systèmes synthétiques de libération qui sont brevetés, appelés Nanotaxi®, et sont composés de différentes familles chimiques (lipides et polymères). Contrairement aux méthodes conventionnelles où les antigènes (peptide ou protéine recombinante) sont injectés avec un adjuvant, la société In-Cell-Art produit ses anticorps par immunisation génétique. L'utilisation d'ADN nu, de biolistique ou l'électroporation en vaccination ADN conduit à une faible induction du titre d'anticorps. L'usage de Nanotaxi® améliore considérablement l'efficacité de la transfection. Il implique un mécanisme physico-chimique conduisant à l'amélioration de la transfection cellulaire de l'ADN plasmidique par un mécanisme de libération directe dans le cytoplasme cellulaire. Cela conduit non seulement à une forte expression de l'antigène mais aussi à une forte stimulation de la réponse immunitaire innée. En conséquence, Nanotaxi® génère des réponses immunitaires puissantes, et par la suite des anticorps polyclonaux d'intérêt même contre des cibles extrêmement difficiles, telles que des protéines membranaires intégrales complexes et des protéines à très haute homologie.

Les anticorps générés par cette société ont été testés par Western Blot et Dot Blot afin de déterminer s'ils possédaient des anticorps capables de reconnaître d'une part des épitopes séquentiels et, d'autre part, des épitopes conformationnels. En Western Blot, sur les dix sérums testés, seuls les sérums SWISS 3 (SW3) et SWISS 4 (SW4) détectent une protéine à la taille attendue pour l'isoforme PGBD5. Ces sérums contiennent donc des anticorps reconnaissant des épitopes séquentiels, c'est-à-dire des acides aminés qui se suivent. Cependant les figures 9 à 11 sont incomplètes, le contrôle négatif avec le sérum pré-immun correspondant à chaque sérum manque. En effet, bien que la bande détectée par les sérums après immunisation corresponde à la taille attendue pour l'isoforme de PGBD5, ce contrôle est nécessaire pour valider que cette bande ne provient pas d'un artefact.

Par Dot Blot, nous avons montré que quatre sérums (C57/BL6-4, SWISS 3, SWISS 4 et SWISS 5) possèdent des anticorps reconnaissant des épitopes conformationnels. Les sérums SWISS 3 et SWISS 4 contiennent donc des anticorps anti-PGBD5 dirigés contre des épitopes conformationnels et/ou séquentiels. Ces résultats obtenus en Western Blot et en Dot Blot sont attendus avec la technologie ICAntibodies. Certaines immunisations peuvent conduire à la production d'une faible quantité d'anticorps dirigés contre des épitopes séquentiels et produire beaucoup plus d'anticorps dirigés contre des épitopes conformationnels. Cela a été le cas notamment avec deux autres transposases domestiquées (ZBED6 et THAP4) pour lesquelles il n'existe pas d'anticorps commerciaux efficaces (ref Arnoaty et al 2013).

L'anticorps anti-PGBD5 devait préférentiellement pouvoir reconnaître des épitopes conformationnels constitués d'acides aminés possiblement non contigus à la surface de la protéine en conformation native (non dénaturée). Bien que le sérum SWISS 5 semble contenir plus d'anticorps conformationnels que les autres sérums, nous avons pris le parti de n'utiliser que les sérums SWISS 3 et SWISS 4. En effet, lors de l'immunoprécipitation, la protéine est en conformation native mais les différentes étapes lors du protocole d'immunoprécipitation peuvent dé-structurer cette conformation secondaire et tertiaire et limiter ainsi la reconnaissance des épitopes conformationnels par l'anticorps. En utilisant les sérums SWISS 3 et SWISS 4, nous avons pensé maximiser les chances de reconnaître la protéine en immunoprécipitant les épitopes séquentiels et/ou conformationnels.

L'anticorps anti-PGBD5 devait également pouvoir être utilisé en immunocytochimie afin de nous renseigner sur la localisation d'expression de PGBD5 dans une cellule ou un organe. La quantité limitée disponible pour chaque sérum ainsi que la faible concentration en anticorps de nos sérums murins ne nous ont pas permis de tester les anticorps sur des coupes de tissus, seules des expérimentations sur cellules fixées ont pu être réalisées. Le sérum anti-PGBD5 nous a permis de mettre en évidence l'expression de PGBD5 dans certaines lignées de cellules cancéreuses. Dans des cellules sur-exprimant les isoformes Mm409 ou Mm523 de PGBD5 couplées à un tag c-MYC, la

protéine PGBD5 a été localisée dans le cytoplasme et plus précisément à la périphérie du noyau. Lors des diverses observations en microscopie confocale, un type de cellules particulières a attiré notre attention. Il semblerait que les cellules en division exprimant fortement PGBD5 concentrent cette protéine autour de la membrane nucléaire. Afin de vérifier si cette concentration de la protéine PGBD5 se limite à la face externe de la membrane nucléaire ou si la protéine PGBD5 se retrouve également au niveau de la face interne de cette membrane, une expérimentation de détection directe de la protéine pourra être envisagée avec un plasmide exprimant Mm409\_GFP ou Mm523\_GFP. Néanmoins cette observation corrobore un résultat obtenu en ChIP-séquençage pour les cellules cancéreuses. En effet, pour la lignée cellulaire T98G, 66 séquences de liaison à l'ADN de PGBD5 sur les 238 séquences détectées sont des motifs répétés en tandem (TGGAA)<sub>n</sub> correspondant à un type de séquence dite ADN satellite centromérique localisé près de la membrane nucléaire. Des résultats similaires ont été obtenus pour la lignée G401 où 77 pics sur les 618 séquences de liaisons correspondent à ce type d'ADN. Lors de la division cellulaire, les chromosomes se séparent et les chromatides sœurs ségréguent de part et d'autre de la cellule, les centromères sont alors dirigés vers la périphérie du noyau la cellule. L'enveloppe nucléaire, quant à elle, s'efface pour permettre la libération des chromosomes, leur capture par les microtubules, puis leur ségrégation entre les deux cellules filles. L'absence de l'enveloppe nucléaire dans les cellules en division permettra ainsi à l'anticorps anti-PGBD5 d'avoir accès à ses cibles plus facilement.

L'anticorps anti-PGBD5 n'étant pas utilisable en histochimie pour permettre de déterminer les organes et les tissus exprimant la protéine PGBD5, nous avons utilisé une stratégie indirecte. Nous avons suivi l'expression du gène *pgbd5* dans les organes de souris grâce au gène rapporteur *LacZ* introduit dans un intron par transgénèse dans une lignée de souris mutante. La transcription du gène *lacZ* étant sous le contrôle du promoteur du gène *pgbd5*, l'enzyme  $\beta$ -galactosidase codée par le gène *LacZ* est théoriquement le reflet de l'expression de la protéine PGBD5. Nous avons ainsi pu vérifier certaines données de RNAseq concernant la présence de PGBD5 dans des parties spécifiques du cerveau telles que l'hippocampe, le cortex cérébral et le bulbe olfactif. Les neurones exprimant PGBD5 dans ces régions sont tous impliqués dans différents types de mémoire.

L'hippocampe joue un rôle central dans la mémoire et la navigation spatiale, il exprimerait fortement la protéine PGBD5 au niveau du gyrus denté et de la corne d'Ammon. Ces deux structures sont composées de nombreux neurones granulaires pour la première et de neurones pyramidaux pour la seconde. Le cortex cérébral, impliqué dans la mémoire sensorielle et motrice, est quant à lui composé de plusieurs couches allant de I à VI. La couche I constituée d'axones et de dendrites ainsi que la couche VI composée de cellules multiformes ne semblent pas exprimer la protéine PGBD5 contrairement aux couches II à V riches en neurones pyramidaux ou granulaires. Le bulbe olfactif

dont la fonction principale est de traiter les informations olfactives en provenance des neurones chimiorécepteurs olfactifs présent au niveau du nez est impliqué dans la mémoire olfactive et la présence de PGBD5 dans cet organe pourrait correspondre à la couche composée de cellules granulaires et/ou de cellules progénitrices.

Outre le fait d'être impliquées dans les processus de mémorisation, ces trois régions du cerveau sont très enrichies en neurones granulaires et / ou neurones pyramidaux. La protéine PGBD5 pourrait donc être exprimée préférentiellement dans ce type de neurones.

Les données de RNAseq montrant une expression des ARNm de PGBD5 dans le cervelet et l'hypophyse n'ont pas pu être vérifiées dans nos conditions expérimentales. Nous n'avons observé aucune expression de la  $\beta$ -galactosidase dans le cervelet, qui est le siège de la régulation de la fonction motrice, tel que le mouvement, la posture et l'équilibre, et dans l'hypophyse, qui sécrète de nombreuses hormones et neuropeptides impliqués par exemple dans le contrôle de la reproduction. Bien que la détection de l'enzyme  $\beta$ -galactosidase est une méthode robuste, cette discordance peut avoir plusieurs origines : premièrement, une expression faible de  $\beta$ -galactosidase. En effet, celle-ci doit être en quantité suffisante afin que la réaction colorimétrique soit détectable. Deuxièmement, bien que les ARNs soient présents, ils peuvent ne pas être traduits en protéines. Dans le cervelet, les séquences analysées en RNAseq pour *pgbd5* sont des séquences anti-sens produits à partir d'un promoteur en 3' du gène ce qui peut expliquer l'absence d'expression de la  $\beta$ -galactosidase. Troisièmement le gène *LacZ* est situé dans un intron. La présence d'un promoteur alternatif actif dans le cervelet et l'hypophyse peut être envisagée, ce qui pourrait conduire à l'absence d'expression du gène *LacZ* et donc de synthèse de l'enzyme  $\beta$ -galactosidase.

Concernant l'expression du gène *pgbd5* dans les gonades, nous avons pu valider les résultats de RNAseq et préciser la localisation de l'expression de la protéine dans les gonades mâles. Le gène *pgbd5* est exprimé dans les cellules de Leydig. Le rôle principal de ces cellules est la production et sécrétion de la testostérone. Cette hormone contrôle le développement et le maintien des caractères sexuels primaires et secondaires. Elle joue un rôle dans le fonctionnement de l'appareil reproducteur masculin et le comportement sexuel. Une expression de la  $\beta$ -galactosidase a été aussi révélée dans les spermatogonies situées au niveau de la membrane basale du tube séminifère. Ces spermatogonies sont à l'origine des spermatozoïdes. Le gène *pgbd5* serait donc transcrit dans les testicules au niveau des cellules de Leydig et des spermatogonies. Cependant il est nécessaire de vérifier la présence de la protéine PGBD5 dans ces types cellulaires. Un indice soutient tout de même que la présence de la protéine PGBD5 est effective, celle-ci étant détectée en spectrométrie de masse dans le testicule [fichier PXD004092 (Testicule humain)] dans la base de données PRIDE.

Concernant les gonades femelles, nous n'avons pas recherché en premier lieu si la protéine

PGBD5 s'exprime dans ces tissus. En effet, l'ovaire est un organe particulier où l'on peut détecter beaucoup de transcrits de gènes qui ne seront pas forcément exprimés par la suite. Dans les banques de données de RNA-seq, les ARNs *pgbd5* ont été détectés seulement dans les ovocytes de souris contrairement aux gonades mâles où les transcrits de *pgbd5* ont été retrouvés également chez l'Homme et d'autres espèces. L'absence de transcrits chez l'Homme et les autres espèces est cependant à considérer avec précaution et pourrait provenir d'un manque de données disponibles.

Grâce à la lignée de souris  $Pgbd5^{tm1a(KOMP)}$  nous avons vérifié les données disponibles de transcriptomique pour la protéine PGBD5 et précisé la localisation de l'expression de la protéine dans les gonades mâles. Il est maintenant important de déterminer le rôle de PGBD5 dans les gonades. Chez la souris, d'après les données fournies par le laboratoire d'Alex Kentsis (Henssen et al), les souris femelles invalidées pour le gène *pgbd5* sont fertiles. Les mâles eux, sont stériles sans qu'il n'y ait de variation de la taille ou d'anomalies en histologie des testicules. Le poids corporel des souris invalidées pour le gène *pgbd5* est cependant plus petit. Les souris présentent également une hypersensibilité aux stress. L'inconvénient des souris de ce génotype est que la délétion du gène *pgbd5* est totale, ce qui peut conduire à de multiples conséquences cognitives, sensorielles, émotionnelles et reproductives. Afin d'étudier l'impact de l'absence de PGBD5 dans certains organes ou tissus cibles, nous avons préféré développer une approche par invalidation conditionnelle. Pour cela, nous avons généré les souris  $Pgbd5^{tm1a(KOMP)}$  qui présentent un site loxP de part et d'autre de l'exon 4. Ces souris seront croisées par la suite avec des souris nestin-cre afin d'invalider le gène dans les neurones et avec des souris AMHR2-Cre afin de l'invalider dans les cellules de Leydig.

Dans un troisième temps, mon objectif a été de déterminer les sites de liaisons de la protéine PGBD5 sur l'ADN afin de définir si la protéine PGBD5 se lie sur ou à proximité des *pbles* humains, de gènes ou de régions spécifiques. Nous avons mis en évidence que la protéine PGBD5 se lie sur des répétitions inversées (IR) particulières. Ces IR auraient la particularité de pouvoir former des structures cruciformes en conformation secondaires et tertiaires. PGBD5 pourrait se fixer sur la boucle formée par ces IRs, qui ne se formerait que dans des conditions bien définies. Si cette hypothèse est exacte la nature et l'origine du matériel biologique utilisé pour nos études seraient la principale source de la grande variabilité des résultats que nous avons obtenus. En effet, la différence de profil obtenue pour les trois lignées cellulaires cancéreuses en est un exemple. Les cellules T98G sont des cellules dérivées d'un contexte neuronal dans lequel la protéine PGBD5 est exprimée naturellement contrairement aux lignées HCT116 et G401 où PGBD5 n'est pas exprimée naturellement. Chez les vertébrés, PGBD5 est la transposase domestiquée la plus ancienne. Sa séquence est la mieux conservée. Dans un système où un gène codant une transposase est domestiqué, son expression dans la cellule va être contrôlée en intensité dans une ou des fenêtres précises de

temps. Il est attendu qu'il en soit de même pour l'accessibilité de ses sites de liaison. Cela est le cas pour les transposases domestiquées RAG1 et RAG2 qui sont exprimées lors d'une courte fenêtre au cours de la lymphogenèse. C'est aussi ce qui est observé au niveau de leurs gènes cibles qui codent les Immunoglobulines G (IgG) et les récepteurs des cellules T. En effet, ces gènes sont soumis à des réarrangements chromatinien qui ouvrent et ferment la chromatine pour permettre à RAG1 et RAG2 de recombinaison ces gènes de façon séquentielle (dans le cas des IgG, la recombinaison a d'abord lieu entre les domaines V et D puis entre les domaines D et J, sans que jamais des recombinaisons entre les domaines V et J ne soient autorisées par la structure de la chromatine). Dans les autres cellules où RAG1 et RAG2 sont exprimées à faible niveau sans fonction apparente (thyroïde et testicule) ou non exprimées, leurs gènes cibles ont été montrés être dans un état de condensation chromatinien empêchant leur accessibilité. Dans un contexte tumoral où l'expression de RAG1 et RAG 2 est dérégulée, il a été montré que ces protéines se fixaient sur des motifs ressemblant aux motifs RSS (Recombinaison Signal Sequence) sur lesquelles les protéines se fixent puis recombinaison localement l'ADN.

Dans notre modèle d'étude, les cellules T98G présenteraient donc des cibles chromosomiques physiologiques accessibles contrairement aux cellules HCT116 et G401 où les cibles chromosomiques physiologiques ne seraient pas accessibles. Dans ces deux derniers types de cellules, la protéine PGBD5 se fixerait de façon aspécifique sur des séquences d'ADN présentant de faible homologie avec les sites spécifiques de cette protéine. Cette hypothèse expliquerait les résultats des ChIP-seq. Cette fixation aspécifique pourrait aussi avoir comme origine la sur-expression de la protéine PGBD5 dans des cellules transfectées. Lors d'une sur-expression, la protéine peut s'accumuler dans le cytoplasme. La fraction de protéine accédant au noyau ne va pas pouvoir se fixer sur des sites de liaisons spécifiques, et elle va alors se lier sur des sites de façon aléatoire dans le génome ou sur des sites à très faible homologie. C'est certainement ce que nous avons observé lors de l'analyse des séquences du ChIP-séquençage des cellules HEK293 sur-exprimant l'isoforme Mm523 de PGBD5. Dans cette expérience nous avons détecté peu de séquences.

D'autres difficultés d'expérimentations ont également été rencontrées au cours de ce travail. Nous nous sommes heurtés à l'absence d'expression de la protéine PGBD5 lors des expérimentations de sur-expression de la protéine ou de fabrication de lignée cellulaire stables. Seules les cellules HEK293 ont permis l'expression de la protéine après transfection transitoire à partir de nos vecteurs d'expression plasmidiques dépendant d'une induction par la DOX.

Dans les autres cellules et tissus où la protéine PGBD5 ne s'exprime pas naturellement, une forte régulation du gène *pgbd5* peut être mis en place. La sur-expression de ce transgène dans un génome de cellule n'exprimant pas la protéine PGBD5 naturellement peut par conséquent déclencher des

systèmes de régulation du transgène et ainsi bloquer la production de la protéine d'intérêt, ce qui peut expliquer la difficulté d'obtention de lignées stables cellulaires exprimant la protéine taguée. Cette forte régulation est nécessaire pour l'organisme ou les cellules car la protéine PGBD5 est une protéine cytotoxique, elle ne doit donc pas s'exprimer n'importe où et n'importe quand. Cette caractéristique a été mise en évidence par Henssen et al [92] et vérifiée dans nos conditions. En effet, dans nos tests d'intégration et de transposition avec la protéine transposase Mm523, nous avons observé un nombre de clones nettement inférieur au nombre de clones obtenus avec la protéine inerte GFP, signe que la transposase est toxique pour les cellules. Pour la mise en œuvre de nos tests, nous nous sommes basés sur les protocoles de transposition de la transposase *PiggyBac*. Or, celle-ci n'est pas ou peu cytotoxique contrairement à la transposase PGBD5.

Néanmoins, nous avons pu mettre en évidence l'existence d'une activité de transposition de PGBD5 grâce à la méthode de LAM-PCR. Nous avons pu déterminer que PGBD5 est capable d'exciser et insérer les *pbles* humains et le transposon *ifp2* de façon plus ou moins proprement avec la conservation des séquences TSD et TIR. D'après nos résultats, la transposase PGBD5 ne pourrait remobiliser que très rarement les transposons une fois insérés contrairement à la transposase *PiggyBac* qui est capable de re-mobiliser plusieurs fois le transposon *ifp2* après insertion. Une différence majeure entre les deux transposases peut expliquer ce résultat. En effet, la transposase PGBD5 est une protéine de l'hôte contrairement à la transposase *PiggyBac*, elle se doit d'être régulée afin de conserver l'intégrité du génome de son hôte. Lors de la domestication de la transposase PGBD5, ce mode de fonctionnement a pu être sélectionné afin de conserver l'activité de la transposition tout en régulant les recombinaisons dans le génome.

En conclusion de mon projet, nous avons pu vérifier les données de RNA-seq disponibles dans les banques pour PGBD5 et localiser avec précision l'expression de la protéine PGBD5 dans le cerveau et les gonades ainsi que son expression dans certaines lignées cellulaires cancéreuses. Ces données montrent que la protéine s'exprime dans des régions du cerveau impliquées dans des processus de mémorisation riches en neurones pyramidaux et cellules granulaires. Selon Henssen et al, la protéine PGBD5 est capable d'événements de recombinaisons. Nous avons pu vérifier également cette caractéristique et mettre en évidence la capacité de la protéine PGBD5 à transposer des *pbles* humains. Pour se faire, la protéine transposase PGBD5 se lierait sur des séquences répétées inversées présentes aux extrémités de ces *pbles*, qui pourrait former des structures cruciformes dans une conformation particulière de l'ADN. La protéine pourrait se fixer également sur d'autres éléments ou gènes présentant ces IRs.

# Perspectives

---

Dans ce travail, nous avons localisé l'expression transcriptionnelle du gène *pgbd5* dans certaines régions du cerveau et dans les gonades grâce à la méthode indirecte du gène rapporteur *lacZ*. Cette méthode étant indirecte, les données obtenues seront donc à vérifier. La présence de la protéine dans les différents types cellulaires de chaque tissu positif devra être vérifiée une fois qu'un anticorps PGBD5 sera obtenu en qualité (permettant de déterminer les isoformes), en localisation (noyau versus cytoplasme) et en quantité suffisante.

Nous avons également pu déterminer grâce aux données obtenues en ChIP-séquençage que la protéine PGBD5 se lie préférentiellement sur des séquences répétées inversées (palindrome). Ces séquences possèdent des caractéristiques leur permettant de former des structures secondaires cruciformes. Il est connu que l'ADN est classiquement formé d'une double hélice que l'on appelle ADN-B. Cependant, il existe d'autres formes *in vivo et in vitro* de formes non-B. Ces structures non B ont été décrites avoir un rôle fonctionnel et peuvent favoriser les cassures ADN et les délétions potentiellement induites par ces cassures, les amplifications, la recombinaison, et les mutations. Il serait important de déterminer si PGBD5 est capable de reconnaître les structures non-B de l'ADN via les IR et les jonctions de Holliday constituées de séquences répétées inversées qui forment des structures cruciformes lors de la recombinaison. En effet, la protéine PGBD5 est une transposase capable d'activité nucléasique dans certaines circonstances. Elle pourrait donc se lier aux jonctions de Holliday et les couper après recombinaison. Il serait également intéressant de déterminer si la protéine PGBD5 agit seule ou si elle requière la présence de co-facteurs telles que les protéines HMG (High Mobility Group) et MLL (myeloid/lymphoid ou mixed-lineage leukemia) également connue sous le nom de KMT2A (lysine N-methyltransferase 2A) pouvant se lier à l'ADN cruciforme.

Les données obtenues en ChIP-séquençage seront également à approfondir pour contextualiser la localisation des séquences détectées afin de mettre en évidence si les sites de liaisons de la protéine se situent sur des gènes ou famille de gènes particuliers ou si la protéine se fixe préférentiellement sur l'hétérochromatine ou séquences non codantes. Cela permettra de déterminer si ces sites de liaisons sont les sites d'action de la protéine ou plutôt des sites de stockage mais surtout de déterminer un motif de coupure privilégié de la protéine.

Enfin, nous avons également démontré que la protéine PGBD5 est capable de transposer les éléments *pbles* humains, c'est-à-dire les exciser et les insérer à un autre locus dans le génome. Cette transposition implique des cassures de l'ADN double brin. Il serait intéressant de regarder si la protéine PGBD5 est liée physiquement à certains partenaires du mécanisme de réparation de l'ADN

tel que la protéine Ku70/80 intervenant dans le processus de réparation NHEJ (Non-Homologous End-Joining) en réalisant une co-immunoprécipitation par exemple avec un anticorps anti-Ku70 et notre anticorps anti-PGBD5. L'importance physiologique de cette interaction pourra être évaluée à partir de notre lignée de souris *Pgbd5*<sup>tm1a(KOMP)</sup>. En effet, afin de mettre en évidence une éventuelle relation entre la protéine PGBD5 et le mécanisme de réparation de l'ADN NHEJ dans le cerveau, la lignée de souris *Pgbd5*<sup>tm1a(KOMP)</sup> pourrait être croisée avec des souris *nestin-cre* qui permettent d'invalider le gène *pgbd5* spécifiquement dans le cerveau des descendants. Cette nouvelle lignée serait ensuite croisée avec des souris *nestin cre / ligase 4* afin d'invalider également la ligase 4 dans le cerveau et ainsi inactiver le mécanisme de réparation de l'ADN NHEJ dans celui-ci. Le phénotype de ces différentes lignées souris serait intéressant à étudier et permettrait très certainement de mettre en lien une éventuelle interaction entre la protéine PGBD5 et le système NHEJ. En effet, la protéine PGBD5 est la seule recombinaison identifiée ce jour présente au niveau des neurones. Les résultats seront cependant à prendre avec précaution. En effet, l'utilisation de la *cre* peut présenter un risque de fuite et donc invalider le gène d'intérêt dans un autre organe autre que l'organe cible et ainsi produire des effets non spécifiques. Les résultats obtenus avec les différentes lignées de souris seront donc à vérifier par une autre méthode d'inactivation du gène *pgbd5* spécifiquement dans le cerveau grâce à la technologie des nanotaxis. Ces particules composées de copolymères et d'ADN plasmidique exprimant un ARN anti-sens contre l'ARN messager de *pgbd5*, permettra d'invalider le gène cible via le phénomène d'interférence par ARN et nous permettra ainsi de valider (ou non) les résultats obtenus avec les lignées de souris.

Les différentes lignées de souris générées permettront par la suite des études sur l'activité et la coordination motrice des souris avec par exemple un test d'agrippage via une grille permettant de déterminer la force de préhension des membres antérieurs ou des études comportementales visant à tester i) la mémoire sociale et l'aptitude aux interactions sociales de souris invalidées ou non dans l'hippocampe via un test de paradigme social à trois chambres par exemple et ii) les mémoires émotionnelles en utilisant des tests d'apprentissages de peur conditionnée. L'ensemble de ces tests permettra de déterminer si PGBD5 a un rôle dans les processus de mémorisations.

## ICAFectin<sup>®</sup> 441: Reagent for DNA transfection in primary and stem cells

### Transfection of adherent cells

---

#### Forward Transfection Procedure

Use the following procedure to transfect adherent cells in a 24-well format. For other formats, see **table 2** Scaling up/down Transfections.

#### Cell Preparation

One day before transfection, plate cells in 1 mL complete growth medium so that cells reach 70-80% confluence at the time of transfection ( $0.5 - 2 \times 10^5$  cells per well).

#### ICAFectin<sup>®</sup> 441 / DNA complex preparation

All amounts and volumes are given to transfect one well in a 24-well format.

1. Thirty minutes before transfection, remove growth medium and add 500  $\mu$ L fresh medium with or without serum (depending on the cells).
2. Dilute 0.5  $\mu$ g of DNA (in a maximal volume of 5  $\mu$ L H<sub>2</sub>O) in 50  $\mu$ L of Opti-MEM I without serum. Mix gently.  
*Note : Do not use the serum-free culture medium in which cells were grown.*
3. Vortex ICAFectin<sup>®</sup> 441 before use. Dilute 0.75  $\mu$ L of ICAFectin<sup>®</sup> 441 in 50  $\mu$ L of Opti-MEM without serum. Mix gently.  
*Note : Do not use the serum-free culture medium in which cells were grown.*
4. Combine the diluted ICAFectin<sup>®</sup> 441 (50.75  $\mu$ L) with the diluted DNA (55  $\mu$ L) by pipetting up and down five times and briefly vortexing.
5. Mix gently and incubate 15 minutes at room temperature.
6. Add the entire volume of complexes (105.75  $\mu$ L) drop-wise to each well containing cells and 500  $\mu$ L of fresh medium with or without serum (depending on the cells). Swirl the plate to ensure the homogeneous distribution of complexes.
7. Incubate cells with transfection complexes under their normal condition growth until analysis. Medium may be supplemented with 50  $\mu$ L serum 100% or changed (depending on the cells) after 2 to 4 hours if cells have been incubated without serum.

**Annexe I :** Protocole de transfection par l'ICAFectin441 recommandé par le fournisseur In-Cell-Art pour la transfection de cellules adhérentes

(extrait de

[https://www.incellart.com/sites/default/files/IMCE/PROTOCOL\\_ICAFectin441\\_Feb16\\_0.pdf](https://www.incellart.com/sites/default/files/IMCE/PROTOCOL_ICAFectin441_Feb16_0.pdf)

# jetPEI<sup>®</sup> *in vitro* DNA Transfection Protocol

## TRANSIENT TRANSFECTION PROTOCOL FOR ADHERENT CELLS (

### 1 CELL CULTURE AND CELL SEEDING

In this protocol, the cells are seeded the day before transfection and the complexes are added subsequently to the cells in serum-containing medium. This standard protocol is referred to as forward protocol and is recommended for routine experiments.

For optimal transfection conditions with jetPEI<sup>®</sup> we recommend using cells 50-70% confluent on the day of transfection. Typically, for transfection in 24-well plates, 50 000 to 100 000 cells are seeded per well 24 hours prior to transfection. Change medium the next morning before performing the experiment and add 1 ml of medium per well. jetPEI<sup>®</sup> is stable in the presence of serum therefore you may use serum containing medium during the entire experiment. For other culture formats, refer to Table 1 for the recommended number of cells to seed the day before transfection.

### 2 PREPARATION OF THE COMPLEXES AND TRANSFECTION

#### Transfection procedure in a 24-well plate:

1. Per well, dilute 1 µg of DNA in 150 mM NaCl to a final volume of 50 µl. Vortex gently and spin down briefly.
2. Per well, dilute 2 µl of jetPEI<sup>®</sup> reagent in 150 mM NaCl to a final volume of 50 µl. Vortex gently and spin down briefly.
3. Add the 50 µl jetPEI<sup>®</sup> solution **to** the 50 µl DNA solution all at once. Please note that mixing the solutions in the reverse order may reduce transfection efficiency.
4. Vortex the solution immediately and spin down briefly.
5. Incubate for 15 to 30 minutes at room temperature.
6. Per well, add the 100 µl jetPEI<sup>®</sup>/DNA mix drop-wise to the cells in 1 ml of serum-containing medium and homogenize by gently swirling the plate.
7. Return the plates to the cell culture incubator.
8. Perform reporter gene assay 24 to 48 h after transfection.

**Annexe 2 :** Protocole de transfection avec le JetPEI recommandé par le fournisseur PolyPlus pour la transfection de cellules adhérentes

(extrait de <https://fnkprddata.blob.core.windows.net/domestic/data/datasheet/PPU/101-10N.pdf>)

# Références bibliographiques

---

- [1] Carnegie Institution of Washington, *Year book - Carnegie Institution of Washington*. Washington, D.C. : Carnegie Institution of Washington, 1948.
- [2] Carnegie Institution of Washington, *Year book - Carnegie Institution of Washington*. Washington, D.C. : Carnegie Institution of Washington, 1962.
- [3] I. Dunham *et al.*, « An integrated encyclopedia of DNA elements in the human genome », *Nature*, vol. 489, n° 7414, Art. n° 7414, sept. 2012, doi: 10.1038/nature11247.
- [4] G. Storz, « An expanding universe of noncoding RNAs », *Science*, vol. 296, n° 5571, p. 1260-1263, mai 2002, doi: 10.1126/science.1072249.
- [5] D. J. Finnegan, « Eukaryotic transposable elements and genome evolution », *Trends Genet.*, vol. 5, n° 4, p. 103-107, avr. 1989, doi: 10.1016/0168-9525(89)90039-5.
- [6] J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, et J. Walichiewicz, « Repbase Update, a database of eukaryotic repetitive elements », *Cytogenet. Genome Res.*, vol. 110, n° 1-4, p. 462-467, 2005, doi: 10.1159/000084979.
- [7] T. Wicker *et al.*, « A unified classification system for eukaryotic transposable elements », *Nat. Rev. Genet.*, vol. 8, n° 12, p. 973-982, déc. 2007, doi: 10.1038/nrg2165.
- [8] B. Piégu, S. Bire, P. Arensburger, et Y. Bigot, « A survey of transposable element classification systems--a call for a fundamental update to meet the challenge of their diversity and complexity », *Mol. Phylogenet. Evol.*, vol. 86, p. 90-109, mai 2015, doi: 10.1016/j.ympev.2015.03.009.
- [9] R. T. M. Poulter et T. J. D. Goodwin, « DIRS-1 and the other tyrosine recombinase retrotransposons », *Cytogenet. Genome Res.*, vol. 110, n° 1-4, p. 575-588, 2005, doi: 10.1159/000084991.
- [10] M. Piednoël et E. Bonnivard, « DIRS1-like retrotransposons are widely distributed among Decapoda and are particularly present in hydrothermal vent organisms », *BMC Evol. Biol.*, vol. 9, p. 86, avr. 2009, doi: 10.1186/1471-2148-9-86.
- [11] T. P. Naas *et al.*, « An actively retrotransposing, novel subfamily of mouse L1 elements », *EMBO J.*, vol. 17, n° 2, p. 590-597, janv. 1998, doi: 10.1093/emboj/17.2.590.
- [12] M. B. Evgen'ev *et al.*, « Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis* », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 94, n° 1, p. 196-201, janv. 1997, doi: 10.1073/pnas.94.1.196.
- [13] M. B. Evgen'ev et I. R. Arkhipova, « Penelope-like elements--a new class of retroelements: distribution, function and possible evolutionary significance », *Cytogenet. Genome Res.*, vol. 110, n° 1-4, p. 510-521, 2005, doi: 10.1159/000084984.
- [14] E. S. Lander *et al.*, « Initial sequencing and analysis of the human genome », *Nature*, vol. 409, n° 6822, p. 860-921, févr. 2001, doi: 10.1038/35057062.
- [15] V. V. Kapitonov et J. Jurka, « Rolling-circle transposons in eukaryotes », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, n° 15, p. 8714-8719, juill. 2001, doi: 10.1073/pnas.151269298.
- [16] C. Feschotte et E. J. Pritham, « Non-mammalian c-integrases are encoded by giant transposable elements », *Trends Genet.*, vol. 21, n° 10, p. 551-552, oct. 2005, doi: 10.1016/j.tig.2005.07.007.
- [17] R. H. Plasterk, Z. Izsvák, et Z. Ivics, « Resident aliens: the Tc1/mariner superfamily of transposable elements », *Trends Genet.*, vol. 15, n° 8, p. 326-332, août 1999, doi: 10.1016/s0168-9525(99)01777-1.
- [18] B. McClintock, « Induction of Instability at Selected Loci in Maize », *Genetics*, vol. 38, n° 6, p. 579-599, nov. 1953.
- [19] D. Lisch, « Mutator transposons », *Trends Plant Sci.*, vol. 7, n° 11, p. 498-504, nov. 2002, doi: 10.1016/s1360-1385(02)02347-6.
- [20] Z. Zachar et P. M. Bingham, « Regulation of white locus expression: the structure of mutant alleles at the white locus of *Drosophila melanogaster* », *Cell*, vol. 30, n° 2, p. 529-541, sept.

- 1982, doi: 10.1016/0092-8674(82)90250-1.
- [21] C. Feschotte, « Merlin, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS1016 insertion sequences », *Mol. Biol. Evol.*, vol. 21, n° 9, p. 1769-1780, sept. 2004, doi: 10.1093/molbev/msh188.
- [22] V. V. Kapitonov et J. Jurka, « Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, n° 11, p. 6569-6574, mai 2003, doi: 10.1073/pnas.0732024100.
- [23] M. J. Fraser, J. S. Brusca, G. E. Smith, et M. D. Summers, « Transposon-mediated mutagenesis of a baculovirus », *Virology*, vol. 145, n° 2, p. 356-361, sept. 1985, doi: 10.1016/0042-6822(85)90172-2.
- [24] J. Jurka et V. V. Kapitonov, « PIFs meet Tourists and Harbingers: a superfamily reunion », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, n° 22, p. 12315-12316, oct. 2001, doi: 10.1073/pnas.231490598.
- [25] R. Kunze et C. F. Weil, « The hAT and CACTA Superfamilies of Plant Transposons », *Mobile DNA II*, p. 565-610, janv. 2002, doi: 10.1128/9781555817954.ch24.
- [26] T. J. D. Goodwin, M. I. Butler, et R. T. M. Poulter, « Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi », *Microbiology (Reading)*, vol. 149, n° Pt 11, p. 3099-3109, nov. 2003, doi: 10.1099/mic.0.26529-0.
- [27] T. G. Doak, F. P. Doerder, C. L. Jahn, et G. Herrick, « A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 91, n° 3, p. 942-946, févr. 1994, doi: 10.1073/pnas.91.3.942.
- [28] V. V. Kapitonov et J. Jurka, « Self-synthesizing DNA transposons in eukaryotes », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, n° 12, p. 4540-4545, mars 2006, doi: 10.1073/pnas.0600833103.
- [29] C. Biémont et C. Vieira, « Junk DNA as an evolutionary force », *Nature*, vol. 443, n° 7111, Art. n° 7111, oct. 2006, doi: 10.1038/443521a.
- [30] E. K. Kentner, M. L. Arnold, et S. R. Wessler, « Characterization of high-copy-number retrotransposons from the large genomes of the Louisiana iris species and their use as molecular markers », *Genetics*, vol. 164, n° 2, p. 685-697, juin 2003.
- [31] C. Vitte, O. Panaud, et H. Quesneville, « LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss », *BMC Genomics*, vol. 8, p. 218, juill. 2007, doi: 10.1186/1471-2164-8-218.
- [32] A. Hua-Van, J. M. Davière, F. Kaper, T. Langin, et M. J. Daboussi, « Genome organization in *Fusarium oxysporum*: clusters of class II transposons », *Curr. Genet.*, vol. 37, n° 5, p. 339-347, mai 2000, doi: 10.1007/s002940050537.
- [33] M.-J. Daboussi et P. Capy, « Transposable elements in filamentous fungi », *Annu. Rev. Microbiol.*, vol. 57, p. 275-299, 2003, doi: 10.1146/annurev.micro.57.030502.091029.
- [34] M. G. Kidwell et D. R. Lisch, « Perspective: transposable elements, parasitic DNA, and genome evolution », *Evolution*, vol. 55, n° 1, p. 1-24, janv. 2001, doi: 10.1111/j.0014-3820.2001.tb01268.x.
- [35] M. Lachner et T. Jenuwein, « The many faces of histone lysine methylation », *Curr. Opin. Cell Biol.*, vol. 14, n° 3, p. 286-298, juin 2002, doi: 10.1016/s0955-0674(02)00335-6.
- [36] D. Zilberman et S. Henikoff, « Silencing of transposons in plant genomes: kick them when they're down », *Genome Biol.*, vol. 5, n° 12, p. 249, 2004, doi: 10.1186/gb-2004-5-12-249.
- [37] G. Abrusán et H.-J. Krambeck, « Competition may determine the diversity of transposable elements », *Theor. Popul. Biol.*, vol. 70, n° 3, p. 364-375, nov. 2006, doi: 10.1016/j.tpb.2006.05.001.
- [38] R. K. Slotkin et R. Martienssen, « Transposable elements and the epigenetic regulation of the genome », *Nat. Rev. Genet.*, vol. 8, n° 4, p. 272-285, avr. 2007, doi: 10.1038/nrg2072.
- [39] A. R. Lohe, E. N. Moriyama, D. A. Lidholm, et D. L. Hartl, « Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements », *Mol. Biol.*

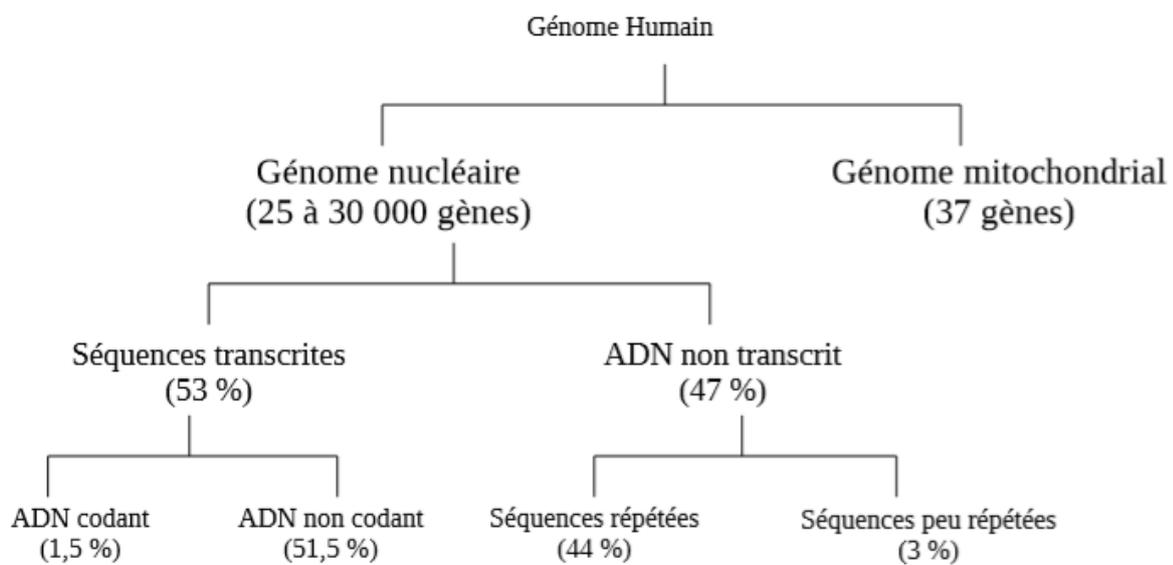
- Evol.*, vol. 12, n° 1, p. 62-72, janv. 1995, doi: 10.1093/oxfordjournals.molbev.a040191.
- [40] A. R. Lohe, D. De Aguiar, et D. L. Hartl, « Mutations in the mariner transposase: the D,D(35)E consensus sequence is nonfunctional », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 94, n° 4, p. 1293-1297, févr. 1997, doi: 10.1073/pnas.94.4.1293.
- [41] S. Venner, C. Feschotte, et C. Biémont, « Dynamics of transposable elements: towards a community ecology of the genome », *Trends Genet.*, vol. 25, n° 7, p. 317-323, juill. 2009, doi: 10.1016/j.tig.2009.05.003.
- [42] W. J. Miller, S. Hagemann, E. Reiter, et W. Pinsker, « P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche* », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 89, n° 9, p. 4018-4022, mai 1992, doi: 10.1073/pnas.89.9.4018.
- [43] S. Mi *et al.*, « Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis », *Nature*, vol. 403, n° 6771, p. 785-789, févr. 2000, doi: 10.1038/35001608.
- [44] D. Liu, J. Bischerour, A. Siddique, N. Buisine, Y. Bigot, et R. Chalmers, « The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase », *Mol. Cell. Biol.*, vol. 27, n° 3, p. 1125-1132, févr. 2007, doi: 10.1128/MCB.01899-06.
- [45] V. V. Kapitonov et J. Jurka, « RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons », *PLoS Biol.*, vol. 3, n° 6, Art. n° 6, mai 2005, doi: 10.1371/journal.pbio.0030181.
- [46] D. G. Schatz, M. A. Oettinger, et D. Baltimore, « Pillars article: the V(D)J recombination activating gene, RAG-1. 1989 », *J. Immunol.*, vol. 180, n° 1, p. 5-18, janv. 2008.
- [47] M. A. Oettinger, D. G. Schatz, C. Gorka, et D. Baltimore, « RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination », *Science*, vol. 248, n° 4962, p. 1517-1523, juin 1990, doi: 10.1126/science.2360047.
- [48] A. Agrawal, Q. M. Eastman, et D. G. Schatz, « Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system », *Nature*, vol. 394, n° 6695, p. 744-751, août 1998, doi: 10.1038/29457.
- [49] M. Melek, M. Gellert, et D. C. van Gent, « Rejoining of DNA by the RAG1 and RAG2 proteins », *Science*, vol. 280, n° 5361, p. 301-303, avr. 1998, doi: 10.1126/science.280.5361.301.
- [50] S. Tonegawa, « Somatic generation of antibody diversity », *Nature*, vol. 302, n° 5909, p. 575-581, avr. 1983, doi: 10.1038/302575a0.
- [51] M. Gellert, « V(D)J recombination: RAG proteins, repair factors, and regulation », *Annu. Rev. Biochem.*, vol. 71, p. 101-132, 2002, doi: 10.1146/annurev.biochem.71.090501.150203.
- [52] D. G. Schatz et P. C. Swanson, « V(D)J recombination: mechanisms of initiation », *Annu. Rev. Genet.*, vol. 45, p. 167-202, 2011, doi: 10.1146/annurev-genet-110410-132552.
- [53] G. Hemmrich, D. J. Miller, et T. C. G. Bosch, « The evolution of immunity: a low-life perspective », *Trends Immunol.*, vol. 28, n° 10, p. 449-454, oct. 2007, doi: 10.1016/j.it.2007.08.003.
- [54] J.-N. Volff, « Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes », *Bioessays*, vol. 28, n° 9, p. 913-922, sept. 2006, doi: 10.1002/bies.20452.
- [55] L. Sinzelle, Z. Izsvák, et Z. Ivics, « Molecular domestication of transposable elements: From detrimental parasites to useful host genes », *Cellular and Molecular Life Sciences*, vol. 66, n° 6, Art. n° 6, mars 2009, doi: 10.1007/s00018-009-8376-3.
- [56] K. Hiom et M. Gellert, « Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination », *Mol. Cell*, vol. 1, n° 7, p. 1011-1019, juin 1998, doi: 10.1016/s1097-2765(00)80101-x.
- [57] A. Agrawal et D. G. Schatz, « RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination », *Cell*, vol. 89, n° 1, p. 43-53, avr. 1997, doi: 10.1016/s0092-8674(00)80181-6.
- [58] M. Ciubotaru, L. M. Ptaszek, G. A. Baker, S. N. Baker, F. V. Bright, et D. G. Schatz, « RAG1-DNA binding in V(D)J recombination. Specificity and DNA-induced conformational

- changes revealed by fluorescence and CD spectroscopy », *J. Biol. Chem.*, vol. 278, n° 8, p. 5584-5596, févr. 2003, doi: 10.1074/jbc.M209758200.
- [59] S. Zhao, L. M. Gwyn, P. De, et K. K. Rodgers, « A non-sequence-specific DNA binding mode of RAG1 is inhibited by RAG2 », *J. Mol. Biol.*, vol. 387, n° 3, p. 744-758, avr. 2009, doi: 10.1016/j.jmb.2009.02.020.
- [60] Y. Liu, R. Subrahmanyam, T. Chakraborty, R. Sen, et S. Desiderio, « A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement », *Immunity*, vol. 27, n° 4, p. 561-571, oct. 2007, doi: 10.1016/j.immuni.2007.09.005.
- [61] A. C. Spradling et G. M. Rubin, « Transposition of cloned P elements into Drosophila germ line chromosomes », *Science*, vol. 218, n° 4570, p. 341-347, oct. 1982, doi: 10.1126/science.6289435.
- [62] Z. Ivics, P. B. Hackett, R. H. Plasterk, et Z. Izsvák, « Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells », *Cell*, vol. 91, n° 4, Art. n° 4, nov. 1997, doi: 10.1016/S0092-8674(00)80436-5.
- [63] G. Luo, Z. Ivics, Z. Izsvák, et A. Bradley, « Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 95, n° 18, p. 10769-10773, sept. 1998, doi: 10.1073/pnas.95.18.10769.
- [64] L. Zhang, U. Sankar, D. J. Lampe, H. M. Robertson, et F. L. Graham, « The Himar1 mariner transposase cloned in a recombinant adenovirus vector is functional in mammalian cells », *Nucleic Acids Res.*, vol. 26, n° 16, p. 3687-3693, août 1998, doi: 10.1093/nar/26.16.3687.
- [65] A. Koga *et al.*, « The medaka fish Tol2 transposable element can undergo excision in human and mouse cells », *J. Hum. Genet.*, vol. 48, n° 5, p. 231-235, 2003, doi: 10.1007/s10038-003-0016-4.
- [66] S. Ding, X. Wu, G. Li, M. Han, Y. Zhuang, et T. Xu, « Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice », *Cell*, vol. 122, n° 3, p. 473-483, août 2005, doi: 10.1016/j.cell.2005.07.013.
- [67] M. J. Fraser, G. E. Smith, et M. D. Summers, « Acquisition of Host Cell DNA Sequences by Baculoviruses: Relationship Between Host DNA Insertions and FP Mutants of *Autographa californica* and *Galleria mellonella* Nuclear Polyhedrosis Viruses », *J. VIROL.*, vol. 47, p. 14, 1983.
- [68] A. Sarkar *et al.*, « Molecular evolutionary analysis of the widespread piggyBac transposon family and related “domesticated” sequences », *Mol. Genet. Genomics*, vol. 270, n° 2, p. 173-180, nov. 2003, doi: 10.1007/s00438-003-0909-0.
- [69] J. Cadiñanos et A. Bradley, « Generation of an inducible and optimized piggyBac transposon system », *Nucleic Acids Res.*, vol. 35, n° 12, p. e87, 2007, doi: 10.1093/nar/gkm446.
- [70] J. H. Keith, T. S. Fraser, et M. J. Fraser, « Analysis of the piggyBac transposase reveals a functional nuclear targeting signal in the 94 c-terminal residues », *BMC Mol. Biol.*, vol. 9, p. 72, août 2008, doi: 10.1186/1471-2199-9-72.
- [71] N. Morellet *et al.*, « Sequence-specific DNA binding activity of the cross-brace zinc finger motif of the piggyBac transposase », *Nucleic Acids Res.*, vol. 46, n° 5, p. 2660-2677, 16 2018, doi: 10.1093/nar/gky044.
- [72] J. H. Keith, C. A. Schaeper, T. S. Fraser, et M. J. Fraser, « Mutational analysis of highly conserved aspartate residues essential to the catalytic core of the piggyBac transposase », *BMC Mol. Biol.*, vol. 9, p. 73, août 2008, doi: 10.1186/1471-2199-9-73.
- [73] R. Mitra, J. Fain-Thornton, et N. L. Craig, « piggyBac can bypass DNA synthesis during cut and paste transposition », *The EMBO Journal*, vol. 27, n° 7, Art. n° 7, avr. 2008, doi: 10.1038/emboj.2008.41.
- [74] K. Yusa, L. Zhou, M. A. Li, A. Bradley, et N. L. Craig, « A hyperactive piggyBac transposase for mammalian applications », *Proceedings of the National Academy of Sciences*, vol. 108, n° 4, Art. n° 4, janv. 2011, doi: 10.1073/pnas.1008322108.
- [75] A. M. Handler, S. D. McCombs, M. J. Fraser, et S. H. Saul, « The lepidopteran transposon

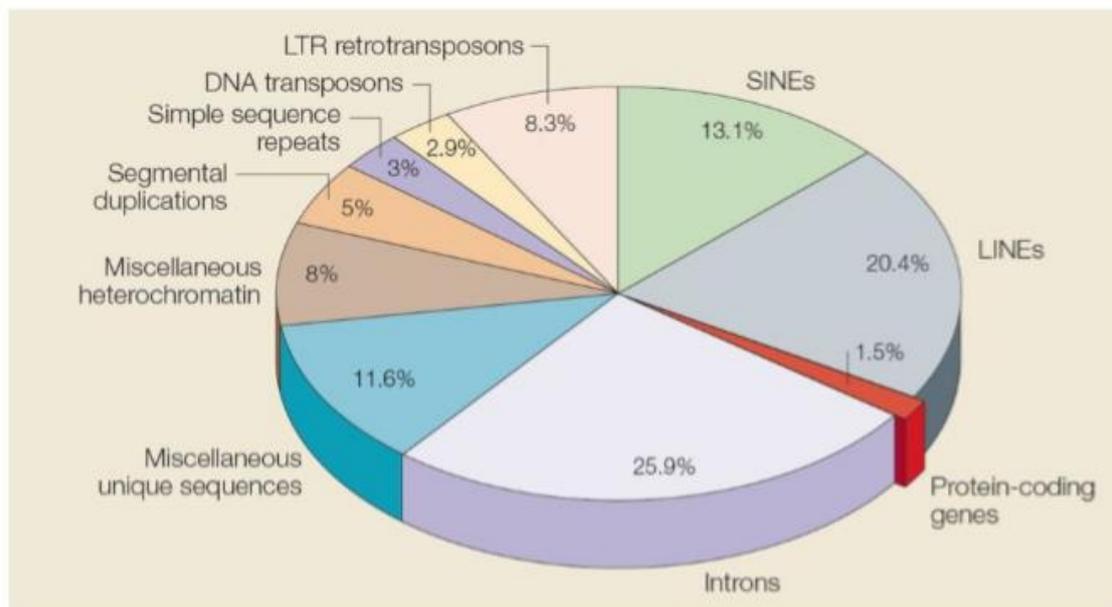
- vector, piggyBac, mediates germ-line transformation in the Mediterranean fruit fly », *Proc. Natl. Acad. Sci. U.S.A.*, vol. 95, n° 13, p. 7520-7525, juin 1998, doi: 10.1073/pnas.95.13.7520.
- [76] N. F. Lobo, T. S. Fraser, J. A. Adams, et M. J. Fraser, « Interplasmid transposition demonstrates piggyBac mobility in vertebrate species », *Genetica*, vol. 128, n° 1-3, p. 347-357, nov. 2006, doi: 10.1007/s10709-006-7165-2.
- [77] Y. Lu, C. Lin, et X. Wang, « PiggyBac transgenic strategies in the developing chicken spinal cord », *Nucleic Acids Res.*, vol. 37, n° 21, p. e141, nov. 2009, doi: 10.1093/nar/gkp686.
- [78] H. Nakanishi, Y. Higuchi, S. Kawakami, F. Yamashita, et M. Hashida, « piggyBac transposon-mediated long-term gene expression in mice », *Mol. Ther.*, vol. 18, n° 4, p. 707-714, avr. 2010, doi: 10.1038/mt.2009.302.
- [79] S. K. Saridey *et al.*, « PiggyBac transposon-based inducible gene expression in vivo after somatic cell gene transfer », *Mol. Ther.*, vol. 17, n° 12, p. 2115-2120, déc. 2009, doi: 10.1038/mt.2009.234.
- [80] M. H. Wilson, C. J. Coates, et A. L. George, « PiggyBac Transposon-mediated Gene Transfer in Human Cells », *Molecular Therapy*, vol. 15, n° 1, Art. n° 1, janv. 2007, doi: 10.1038/sj.mt.6300028.
- [81] K. K. Kojima, « Human transposable elements in Repbase: genomic footprints from fish to humans », *Mob DNA*, vol. 9, p. 2, 2018, doi: 10.1186/s13100-017-0107-y.
- [82] T. Pavelitz, L. T. Gray, S. L. Padilla, A. D. Bailey, et A. M. Weiner, « PGBD5: a neural-specific intron-containing piggyBac transposase domesticated over 500 million years ago and conserved from cephalochordates to humans », *Mobile DNA*, vol. 4, n° 1, Art. n° 1, 2013, doi: 10.1186/1759-8753-4-23.
- [83] C. Baudry *et al.*, « PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia* », *Genes Dev.*, vol. 23, n° 21, p. 2478-2483, nov. 2009, doi: 10.1101/gad.547309.
- [84] C.-Y. Cheng, A. Vogt, K. Mochizuki, et M.-C. Yao, « A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in *Tetrahymena thermophila* », *Mol. Biol. Cell*, vol. 21, n° 10, p. 1753-1762, mai 2010, doi: 10.1091/mbc.e09-12-1079.
- [85] J. C. Newman, A. D. Bailey, H.-Y. Fan, T. Pavelitz, et A. M. Weiner, « An abundant evolutionarily conserved CSB-PiggyBac fusion protein expressed in Cockayne syndrome », *PLoS Genet.*, vol. 4, n° 3, p. e1000031, mars 2008, doi: 10.1371/journal.pgen.1000031.
- [86] M. Muñoz-López et J. L. García-Pérez, « DNA transposons: nature and applications in genomics », *Curr. Genomics*, vol. 11, n° 2, p. 115-128, avr. 2010, doi: 10.2174/138920210790886871.
- [87] A. G. Henssen *et al.*, « Genomic DNA transposition induced by human PGBD5 », *Genes and chromosomes*, p. 20.
- [88] S. Huang *et al.*, « Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination », *Cell*, vol. 166, n° 1, p. 102-114, juin 2016, doi: 10.1016/j.cell.2016.05.032.
- [89] S. Bire *et al.*, « Mariner Transposons Contain a Silencer: Possible Role of the Polycomb Repressive Complex 2 », *PLoS Genet.*, vol. 12, n° 3, p. e1005902, mars 2016, doi: 10.1371/journal.pgen.1005902.
- [90] E. B. Chuong, N. C. Elde, et C. Feschotte, « Regulatory activities of transposable elements: from conflicts to benefits », *Nat. Rev. Genet.*, vol. 18, n° 2, p. 71-86, 2017, doi: 10.1038/nrg.2016.139.
- [91] A. G. Henssen *et al.*, « Forward genetic screen of human transposase genomic rearrangements », *BMC Genomics*, vol. 17, n° 1, Art. n° 1, déc. 2016, doi: 10.1186/s12864-016-2877-x.
- [92] A. G. Henssen *et al.*, « PGBD5 promotes site-specific oncogenic mutations in human tumors », *Nature Genetics*, vol. 49, n° 7, Art. n° 7, mai 2017, doi: 10.1038/ng.3866.
- [93] A. G. Henssen, C. Reed, E. Jiang, et H. D. Garcia, « Therapeutic targeting of PGBD5-induced DNA repair dependency in pediatric solid tumors », *SCIENCE TRANSLATIONAL*

*MEDICINE*, p. 37, 2017.

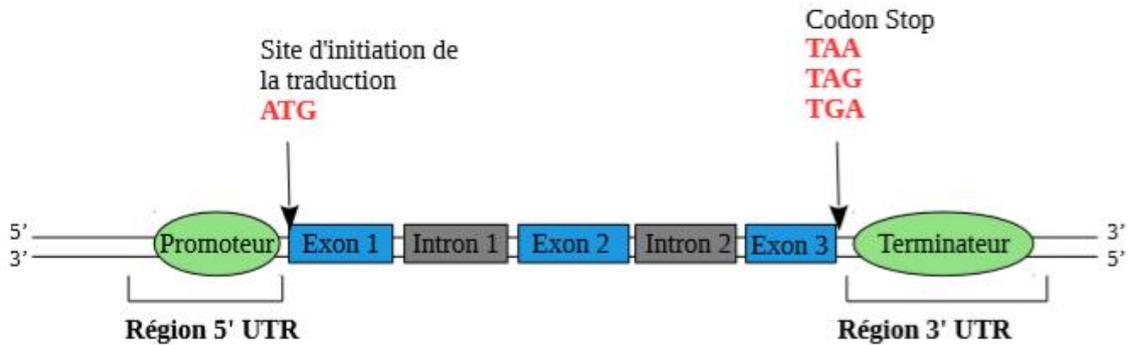
- [94] J. Song, B. Rutjens, et C. Dean, « Detecting histone modifications in plants », *Methods Mol. Biol.*, vol. 1112, p. 165-175, 2014, doi: 10.1007/978-1-62703-773-0\_11.
- [95] Y. Zhang, Y.-H. Lin, T. D. Johnson, L. S. Rozek, et M. A. Sartor, « PePr: a peak-calling prioritization pipeline to identify consistent or differential peaks from replicated ChIP-Seq data », p. 8.
- [96] J. Feng, T. Liu, B. Qin, Y. Zhang, et X. S. Liu, « Identifying ChIP-seq enrichment using MACS », p. 24, 2013.
- [97] A. Arnaoty, V. Gouilleux-Gruart, S. Casteret, B. Pitard, Y. Bigot, et T. Lecomte, « Reliability of the nanospheres-DNA immunization technology to produce polyclonal antibodies directed against human neogenic proteins », *Molecular Genetics and Genomics*, vol. 288, n° 7-8, Art. n° 7-8, août 2013, doi: 10.1007/s00438-013-0754-8.



**Figure 1** : Schéma de la structuration du génome humain



**Figure 2** : Schéma des principaux composants du génome Humain (d'après Nature, 21 octobre 2004)

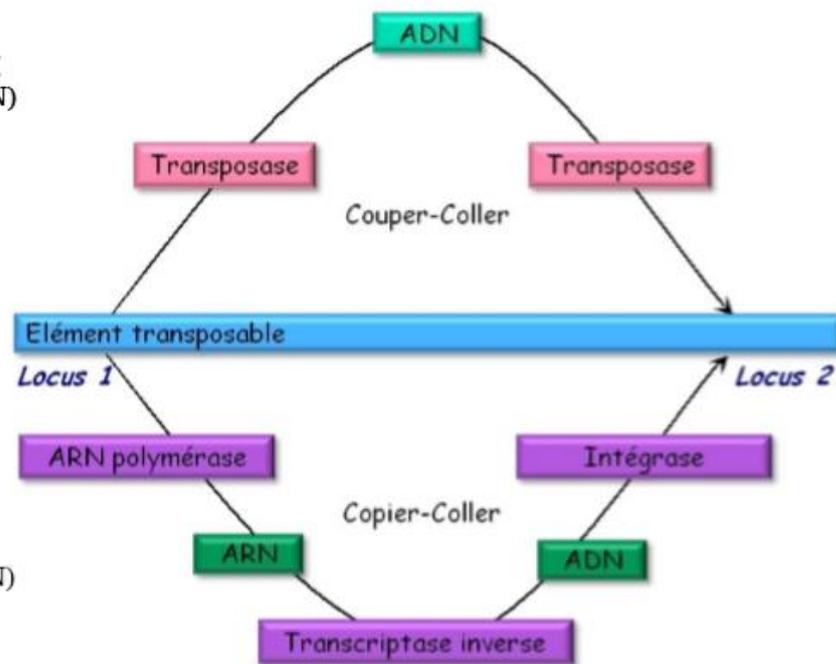


**Figure 3** : Schéma simplifié de la structure d'un gène eucaryote

De part et d'autre de la séquence codante d'un gène eucaryote se trouvent les régions 5' et 3' UTR (*untranslated regions* en anglais). Ces régions ont un rôle important dans la régulation de l'expression du gène. La transcription de la séquence codante démarre au site d'initiation et se termine au niveau du codon Stop (en rouge sur le schéma). Lors de la transcription, les exons (en bleu) et les introns (en gris) sont transcrits puis les introns sont éliminés après l'épissage. Seuls les exons (bleu) sont transcrits et traduits, aboutissant à la synthèse d'une protéine.

Éléments de classe II  
(à intermédiaire ADN)

Éléments de classe I  
(à intermédiaire ARN)



**Figure 4** : Schéma simplifié du mode de transposition des éléments transposables de classe I et II.

Les éléments de classe I ou rétro-éléments, transposent selon un mécanisme de « copier coller ». L'ARN polymérase de l'hôte transcrit le génome du rétro-élément en intermédiaire ARN simple brin. Celui-ci est rétrotranscrit en ADN double brin par une transcriptase inverse et réinséré dans un autre site du génome par l'action de l'intégrase codée par l'élément transposable.

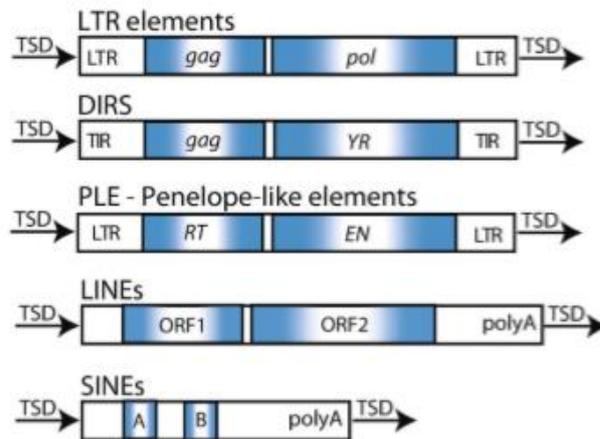
Les éléments de classe II transposent généralement selon un mode de « couper-coller » via un intermédiaire ADN. La transposase, protéine unique codée par l'élément transposable, catalyse l'excision du transposon puis sa réintégration dans un second locus du génome.

Ordre	Super-famille	Structure	TSD
<b>Classe I (rétroéléments : rétrotransposons et rétroposons)</b>			
LTR	Copia	→ GAG AP INT RT RT →	4-6
	Gypsy	→ GAG AP RT RH INT →	4-6
	Bel-Pao	→ GAG AP RT RH INT →	4-6
	Retrovirus	→ GAG AP RT RH INT ENV →	4-6
	ERV	→ GAG AP RT RH INT ENV →	4-6
DIRS	DIRS	→ GAG AP RT RH YR ←	0
	Ngara	→ GAG AP RT RH YR →	0
	VIPER	→ GAG AP RT RH YR →	0
PLE	Penelope	← RT DN →	Variable
LINE	R2	← RT DN →	Variable
	RTE	← APL RT →	Variable
	Jockey	← ORF1 APL RT →	Variable
	L1	← ORF1 APL RT →	Variable
	I	← ORF1 APL RT RH →	Variable
SINE	tRNA	← tRNA →	Variable
	7SL	← 7SL →	Variable
	5S	← 5S →	Variable
<b>Classe II (transposons à ADN) – Sous-classe 1</b>			
TIR	Tc1-Mariner	→ Tase* ←	TA
	hAT	→ Tase* ←	8
	Mutator	→ Tase* ←	9-11
	Merlin	→ Tase* ←	8-9
	Transib	→ Tase* ←	5
	P	→ Tase ←	8
	PiggyBac	→ Tase ←	TTAA
	PIF-Harbinger	→ Tase* ORF2 ←	3
CACTA	→ Tase ORF2 ←	2-3	
Crypton	Crypton	← YR →	0
<b>Classe II (transposons à ADN) – Sous-classe 2</b>			
Helitron	Helitron	← RPA / Y218L →	0
Maverick	Maverick	← C-INT APL CYP POL B ←	6
<b>Motifs</b>		<b>Protéines</b>	
→	LTR (Long Terminal Repeat)	AP : Protéinase Aspartique	RPA : Réplicase A
—	Motif en région non codante	APE / EN : Endonucléase	HEL : Helicase
←	ITR (Inversed Terminal Repeat)	C-INT / INT : Intégrase	CYP : Cystéine protéase
█	Séquence codante	GAG : Protéine capsid	ATP : ATPase
—/—	Région avec un/plusieurs ORFs	RT : Transcriptase inverse	POL B : ADN Polymérase B
—	Séquence non-codante	RH : RNaseH	YR / Y2 : Tyrosine recombinase
		ENV : Protéine d'enveloppe	Tase : Transposase (* avec motif DDE)

**Figure 5** : Classification des différents éléments transposables d'après Wicker et al (2007).

Les éléments transposables sont répartis en deux principales classes selon la présence ou non d'intermédiaire ARN lors du mécanisme de transposition. Ils sont de plus divisés en sous classes, ordres et super-familles. Sur cette figure, la taille du site de duplication (TSD : Target Site Duplication) qui est caractéristique de certaines super-familles est indiquée.

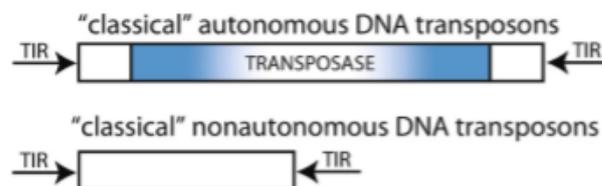
### Éléments transposables Classe I



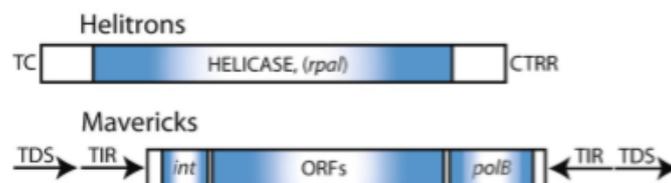
**Figure 6 :** Structures des éléments transposables de classe I (d'après Makalowski et al, 2019)

La grande majorité des ET de classe I possède une structure comportant plusieurs ORF (open reading frame ou cadre de lecture en français) encadrées par des régions terminales répétées (TSD) de même orientation aux extrémités. Les rétrotransposons à LTR codent principalement et généralement deux protéines de type GAG et POL. Les éléments DIRS codent pour une protéine tyrosine recombinase (YR) à la place de la protéine POL. Les éléments PLE ne possèdent qu'une seule ORF et codent pour une protéine contenant une transcriptase réverse (RT) et une endonucléase (EN). Les éléments LINEs possèdent deux ORFs ainsi qu'un promoteur PolII. Les éléments SINEs sont flanqués par des TSD et se terminent par une queue polyA.

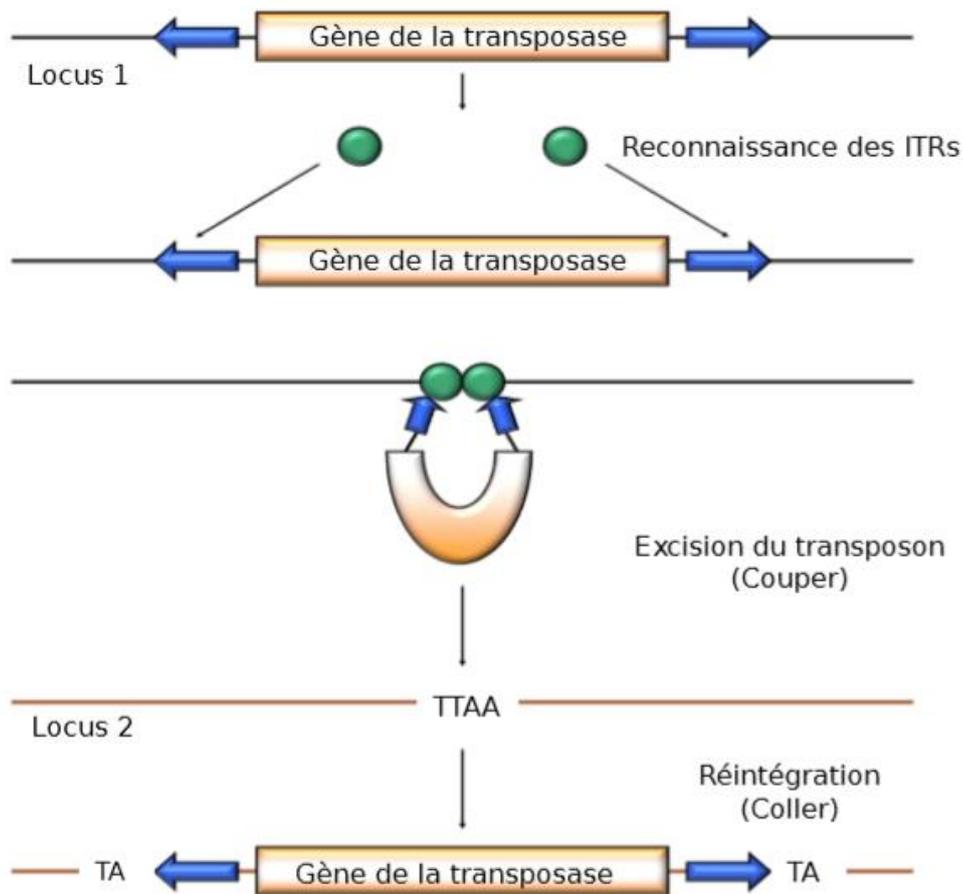
### Éléments transposables Classe II sous classe 1



### Éléments transposables Classe II sous classe 2

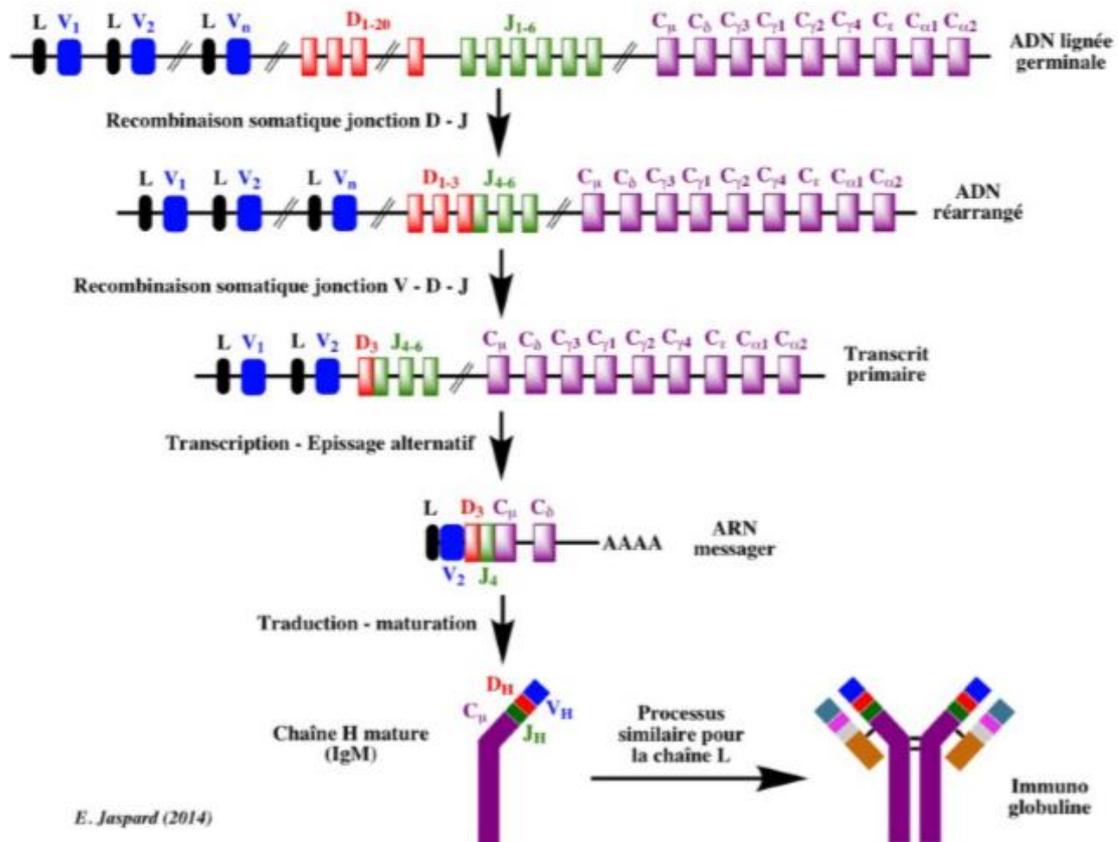


**Figure 7 :** Structures des éléments transposables de classe II (d'après Makalowski et al, 2019)



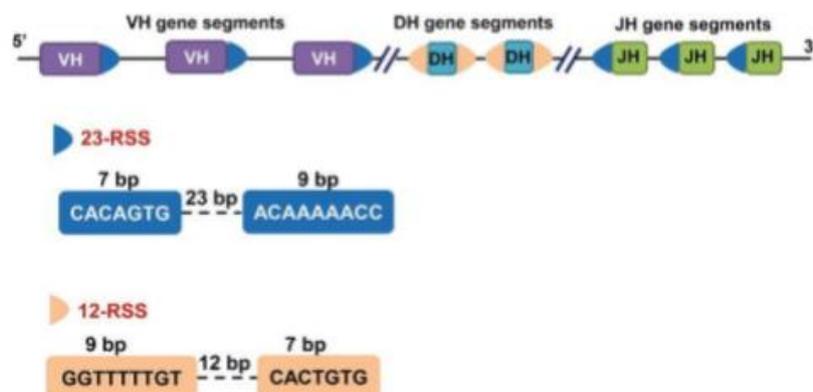
**Figure 8** : Modèle théorique et simplifié du mécanisme de transposition des ETs de classe II.

La transposase (cercle vert) se lie à chaque ITR (flèches bleues) du transposon situé dans l'ADN génomique de l'hôte pour former un complexe synaptique. Ce complexe s'excise de l'ADN puis va reconnaître un site accepteur dans le génome dans lequel il s'intègre. Le site accepteur (ici un tétranucléotide TAA pour l'élément *PiggyBac*) est dupliqué lors de l'insertion du transposon.



**Figure 9** : Réarrangement des segments de gènes codant les chaînes H des immunoglobulines

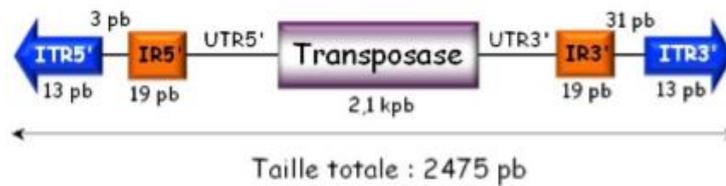
Dans le modèle de réarrangement ordonné, les réarrangements séquentiels conduisent d'abord à la recombinaison d'un des segments D (gène de diversité) avec un des segments J (gène de jonction), puis un segment V (gène variable) est recombinaison avec la recombinaison DJ.



**Figure 10** : Représentation des motifs RSS (Signal de séquence de recombinaison)

(d'après DeFranco AL, Robertson M, Locksley RM. Immunité: la réponse immunitaire dans les maladies infectieuses et inflammatoires. De Boeck; 2009)

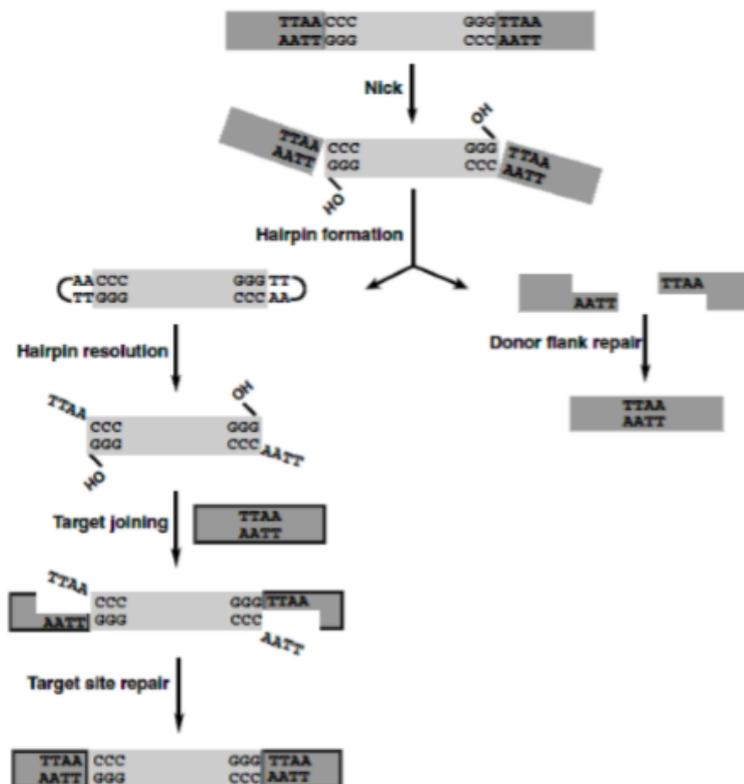
Il existe deux types de RSS, l'un comprend un espaceur de 12 nucléotides [ou paires de bases (bp)] (12-RSS), et l'autre un espaceur de 23 nucléotides (23-RSS). Le 12-RSS et le 23-RSS comprennent tous deux des séquences palindromiques de 7 ou 9 nucléotides très conservées.



**Figure 11** : Structure du transposon *PiggyBac*

Cet élément transposable de 2 475 pb est constitué d'un gène de 2,1 kpb codant une transposase. Le gène est encadré de régions non traduites en 5' et 3' (UTRs : UnTranslated Regions).

À l'extérieur des UTRs se trouvent les IRs (répétitions internes) de 19 pb séparées des ITRs (Inverted Terminal Repeats) par 3 pb en 5' et 31 pb en 3'. Les ITRs délimitent le transposon.



**Figure 12** : Les différentes étapes de la transposition de *PiggyBac* (d'après Mitra et al. 2008)

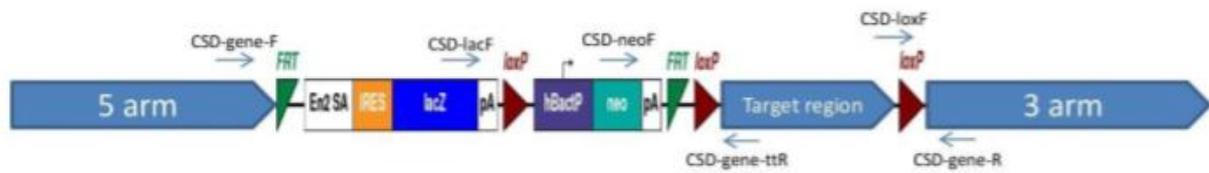
La transposition est initiée par une coupure aux extrémités 5' et 3' du transposon, exposant des 3'OH. Ces 3'OH attaquent le brin complémentaire à 4 nucléotides dans l'ADN flanc, générant ainsi une boucle à chaque extrémité du transposon qui est alors excisée. Le site donneur est réparé par complémentarité permettant de reformer précisément la séquence cible TTAA. Les boucles terminales sont ouvertes par la transposase, ce qui re-expose les 3'OH et génère des extrémités sortantes TTAA en 5' du transposon excisé. Les 3'OH vont permettre l'attaque de la séquence cible TTAA/AATT et la production de brèches simple-brin favorisant l'insertion du transposon dans le génome. Après intégration, les brèches sont réparées et conduisent à la duplication du site TTAA.

**M**GPTRKMPPSASAVDFQFLVDPDNLKNMVVQTNMYAKKFQERFGSDGA  
WVEVTLTEMKAFLGYMISTSISHCESVLSIWSSGGFYNSRSLALVMSQARFEK  
ILKYFHVVAFRSSQTTHGLYKVQPFLDSLQNSFDSAFRPSQTQVLHEPLIDED  
PVFIATCTERELRKRKRKRFSLWVRQCSSTGFIIQIYVHLKEGGGPDGLDAL  
KNKPQLHSMVARSLCRNAAGKNYIIFTGPSITSLTLFEEFEKQGIYCCGLLRA  
RKSDCTGLPLSMLTNPATPPARGQYQIKMKGNMSLICWYNKGHFRFLTNAY  
SPVQQGVIIKRKSGEIPCPLAVEAFAAHL SYICRYDDKYSKYFISHKPNKTW  
QQVFWFAISIAINNAYILYKMSDAYHVKRYRAQFGERLVRELLGLEDASPT  
H

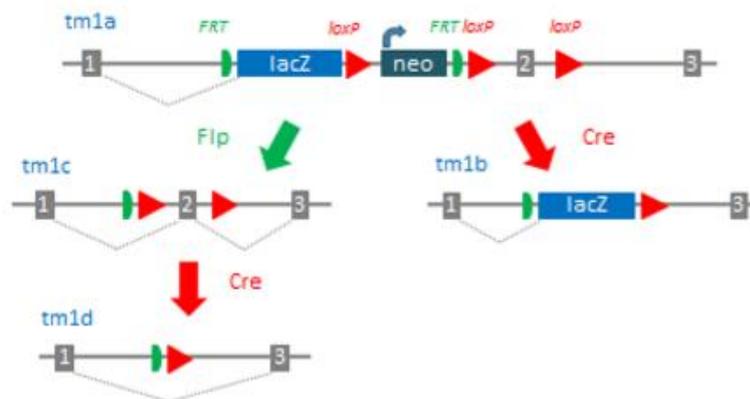
**Figure 13** : Séquence protéique non optimisée de PGBD5 utilisée pour l'immunisation des souris.

Séquence de 415 acides aminés correspondant à la position 140 à 554

A



B



**Figure 14 :** Représentations de la stratégie de création de la lignée *Pgbd5<sup>tm1a</sup>* disponible dans le KOMP (Knockout Mouse Project) Repository

A- Représentation schématique de la cassette présente chez les souris *Pgbd5<sup>tm1a</sup>*. La « target region » correspond à l'exon 4 du gène *pgbd5*.

Les sites FRT correspondent aux sites de reconnaissance de la flipase-recombinase pour l'excision du fragment ADN se situant entre ces sites FRP

Les sites loxP correspondent aux sites de reconnaissance de la Cre-recombinase permettant l'excision de l'élément entre ces sites loxP.

B- Stratégie de délétion de la région souhaitée (« target cible de la figure A), ici indiquée par le chiffre 2.

L'usage de la flipase (Flp) dans un premier temps (partie gauche), permet d'éliminer le gène rapporteur *lacZ* ainsi que le gène de la néomycine (Neo) utilisé lors de la sélection des cellules ES par résistance antibiotique. L'usage de la Cre dans un second temps, permet d'éliminer le gène d'intérêt. Cette construction finale permet d'obtenir des animaux ne possédant plus le gène d'intérêt (Knock-Out : KO)

Dans la partie gauche du schéma B, l'usage de la Cre permet de conserver le gène rapporteur *LacZ* (qui se trouve sous le promoteur du gène d'intérêt) et ainsi suivre l'expression du gène ciblé sans la région (dans notre cas, l'exon 4) d'intérêt du gène.

ATGGCCGAGGGCGGTGGGGCTCACGGAGGAGGGCTCCGGCGCTGCTGGAGGCAGCCCGTGCACGCTACGAGAGCCTGCACATCTCGGACGACGTGTTCCGG  
 CGAGTCGGGCCCGGACAGCGGCGGGAACCCCTTTTACAGCACCTCCGCAGCCTCACGCTCCTCCTCGGCCGCTCCTCGGACGATGAGCGTGAGCGCCCG  
 CGCCTCCGGGGACCGCCCGCCATCCTACGCCCGGATCCCTTGGAACTCGAGGAGGACGAGACCGGTGGGGCTGGAGCGCCGCTGCTGAGGGATCGCCC  
057H →  
 GTCCCCGCGCTTCGAGGACACCGGTGGTCCAACCGAAAGATGCCACCCAGCGCTAGCGCGGTGGACTTCTTCCAGCTCTTCGTCGCCGACAACGTGCTCA  
058H →  
 AGAACATGGTGGTACAGACCAACATGTATGCCAGGAAGTTCCAGGAGCGGTGGTGGCAGTGACGGTGCCTGGGTGGAGGTGACGTTGGCAGAGATGAAGGCC  
059H →  
 TTCCTGGGCTATGTGATCTCCACCAGCGTCTCGACTGTGAGTCGGTACTCAGCATTTGGAGCGGAGGCTTCTATAGCAACCGGAGCCTTGCCTCGTCAT  
 GAGCCAGGCCCGCTTCGAGAAAATCCTCAAGTACTTCCATGTTGTGGCCTTCCGATCCAGCCAGACCACGCATGGGCTCTACAGGTCCAGCCCTTCTCG  
060H →  
 ACTCCCTGCAGAGTGGCTTTGATGCTGCTTTCAGGCCGTCGACTCCCTGCAGAGTGGCTTTGATGCTGCTTTCAGGCCGCTCAGACCCAGGTGCTACATG  
 AACCCCTGATTGACGAGGACCCTGTGTTTATTGCCACGTGCACGGAGCGAGAGCTGCGGAAGAGGAAAAAGCGGAAATTCAGCCTGTGGGTCCGCCAGTGC  
 TTTCAACCGGGTTCATCATCCAGATCTACGTCCACCTAAAGGAAGGTGGTGGCCAGATGGCCTGGACGCTCTGAAGAACAAGCCACAGCTGCACAGCAT  
 GGTGGCCCGGAGCCTGTGCCGGAACGCGGGCCGAAAGAAGTATATCATCTTTACAGGGCCAGCATCACCAGCCTCAATCTGTTGAAAGAATTTGAAAGC  
← 061H  
 AAGGGATCTACTGCTGTGGCCTCCTCAGCTCTAGGAAGAGTGACTGCACAGGTCTCCCTCCATCCATGTTGACCAACCTGCCACCCACTCGCCCGGGGC  
 CAGCACCAGATCAGGACAAAGGGCAACATGTCACTGATCTCTGGTACAACAAGGGGCACCTCCGCTTCTGACCAACGCCTACTCCCTGTGCAGAAAGG  
← 062H  
 CGTCATCATCAAGAGGAGGAGTGGGGAAATCCCTGCCCTCTGGCCGTGGAGGCCCTTGGCCCTCACCTCAGCTACATCTGCAGATATGACGATAAGTACA  
 GCAAGTATTTTCATCTCTACAAGCCAAACAAGACGTGGCAGCAAGTGTCTGGTTCCGATCAGCATCGCCGTCAACAACGCCTACATCCTGTACAAAATG  
 TCAGAGCCTACCACGTGAAGAGTACAGCCGGGCACAGTTCGAGAGAGACTTGTGAGGGAGTTGCTGGGCTTGGAGGACTCATCGCCAGCCACTGA  
← 063H      ← 064H

**Figure 15 :** Séquence nucléotidique murine de *pgbd5* avec les amorces utilisées pour les RT-PCR nichées.

La séquence nucléotidique de l'isoforme Mm523 correspond à la séquence allant du 1<sup>er</sup> ATG (vert) au TGA (rouge).  
 La séquence nucléotidique de l'isoforme Mm409 correspond à la séquence allant du 2<sup>nd</sup> ATG (vert) au TGA (rouge).

Les amorces en orange sont des amorces sens et en violet des amorces anti-sens.

A



NM\_171824.2: Most "consensual" isoform (409 aa).

```
MPPSASAVDFQFLVDPDNLKNNMNVVQTNMYARKFQERFGSDGAWVEVTLAEMKAFLGYVISTSVSHCESVLSIWSSGGFYSNRSLALVMS
QARFEKILKYFHVVAFRSSQITHTGLYKVPFLDSLQSGFDAAFRPSQTQVLHEPLIDEDPVFIATCTERELRKRKRKFLWVRQCSSTGFII
QIYVHLKEGGPDGLDALKNKPKQLHSMVARSLCRNAAGKNYIIFTGPSITSLNLFEEFEKQGIYCCGLLSSRKSCTGLPPSMLTNPATPLAR
GQHQRITKGNMSLICWYNKGHFRFLTNAYSPVQKGVIIKRRSGEIPCPLAVEAFAAHLASYICRYDDKYSKYFISHKPNKTWQQVFWFAISIA
VNNAYILYKMSDAYHVKKYSRAQFGERLVRRELLGLEDS SPAH
```

B



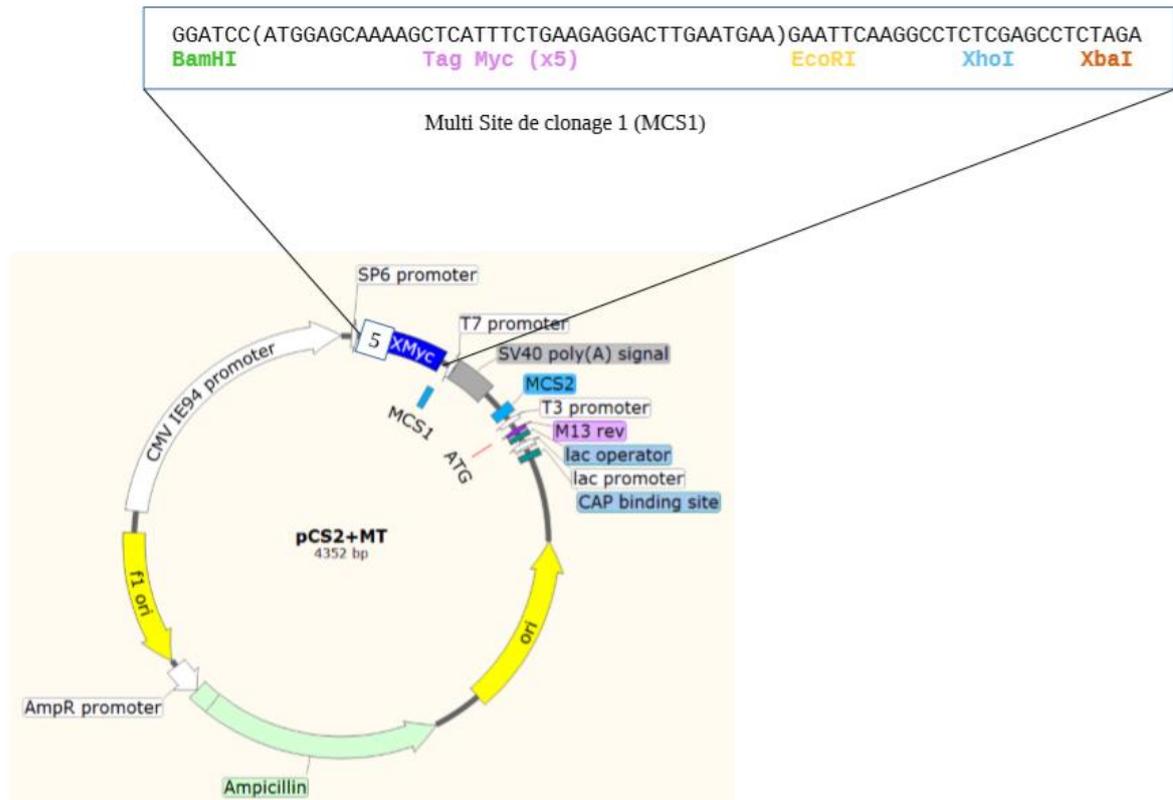
XM\_006530804.1: (523 aa).

```
MAEGGGGRRRAPALLEAARARYESLHISDDVFGESGPDSSGGNPFYSTSAASRSSSAASSDDERERAPPPTAPPSYAADPLELEEDTGGG
WSAVLRDRPSRPFEDTGGPTRKMPPSASAVDFQFLVDPDNLKNNMNVVQTNMYARKFQERFGSDGAWVEVTLAEMKAFLGYVISTSVSHC
ESVLSIWSSGGFYSNRSLALVMSQARFEKILKYFHVVAFRSSQITHTGLYKVPFLDSLQSGFDAAFRPSQTQVLHEPLIDEDPVFIATCTEREL
RKRKRKFLWVRQCSSTGFIIQIYVHLKEGGPDGLDALKNKPKQLHSMVARSLCRNAAGKNYIIFTGPSITSLNLFEEFEKQGIYCCGLLSS
RKSCTGLPPSMLTNPATPLARGQHQRITKGNMSLICWYNKGHFRFLTNAYSPVQKGVIIKRRSGEIPCPLAVEAFAAHLASYICRYDDKYSK
YFISHKPNKTWQQVFWFAISIAVNNAYILYKMSDAYHVKKYSRAQFGERLVRRELLGLEDS SPAH
```

**Figure 16 :** Représentation des deux isoformes murines de *pgbd5*

**A-** Isoforme Mm409 : schéma constitution du gène *pgbd5* (partie haute) et séquence protéique (partie basse)

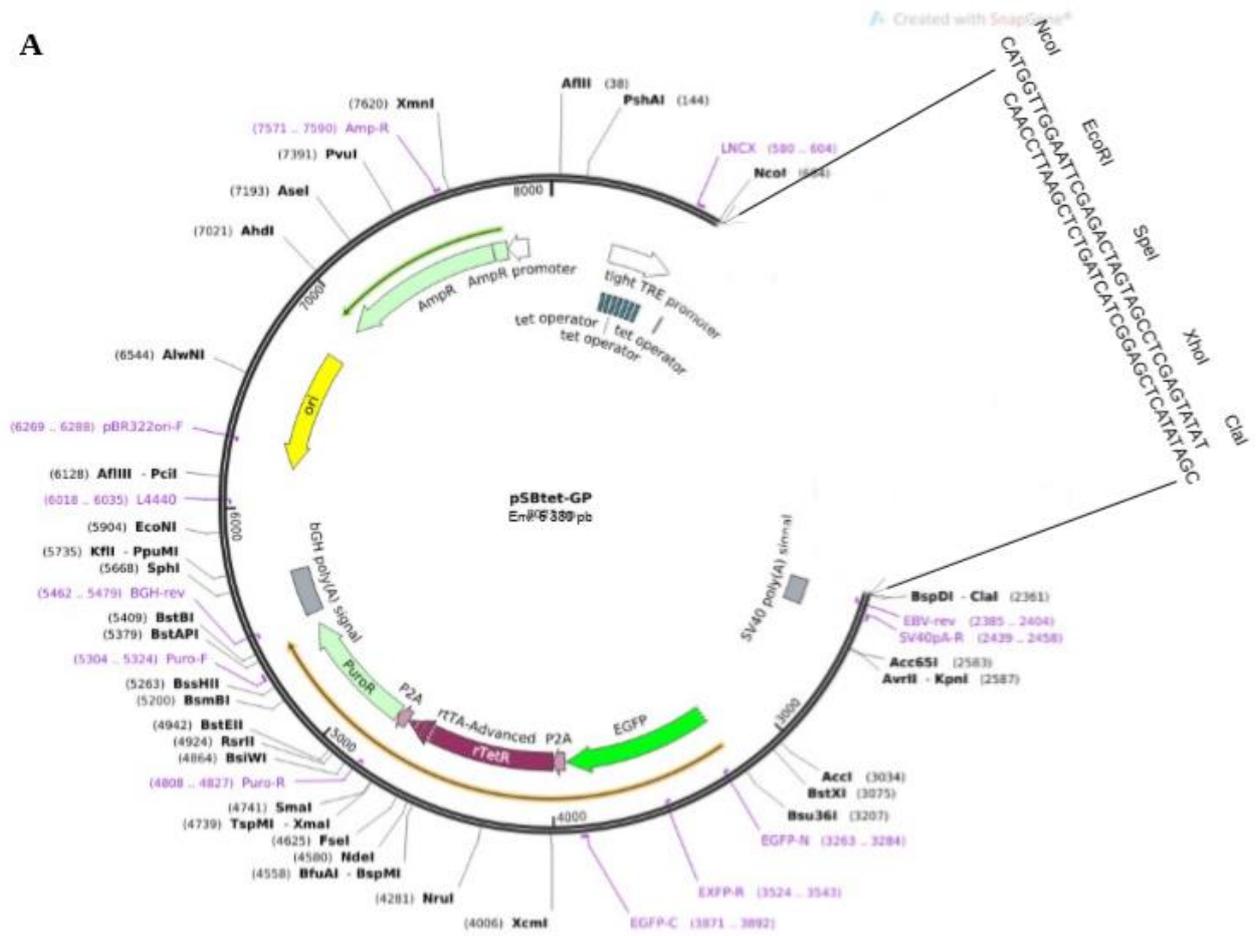
**B-** Isoforme Mm523 : schéma constitution du gène *pgbd5* (partie haute) et séquence protéique (partie basse)



**Figure 17 :** Carte du plasmide pCS2 + MT modifié

Le plasmide originel a été modifié : initialement avec 6X Myc, le dernier tag Myc a été supprimé afin d’insérer un site de restriction pour l’enzyme EcoRI dans le multi site de clonage (MCS 1).

**A**



**B**

Multi site de clonage inséré :

NcoI      EcoRI      SpeI      XhoI      ClaI

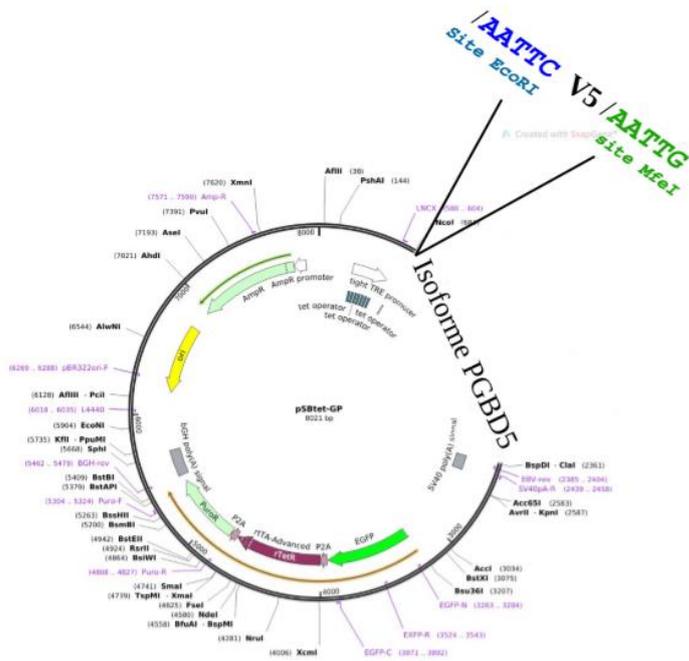
Amorce 1 5' CATGGTTGGAATTCGAGACTAGTAGCCTCGAGTATAT 3'

Amorce 2 3' CAACCTTAAGCTCTGATCATCGGAGCTCATATAGC 5'

**Figure 18 :** Carte schématique du plasmide pSBtet\_iforme PGBD5

Le plasmide originel (Addgene # 60495) a été modifié : après retrait du gène de la luciférase par digestion NcoI et ClaI, le multi-site de clonage (B) est inséré à la place par ligation (A).

A



B

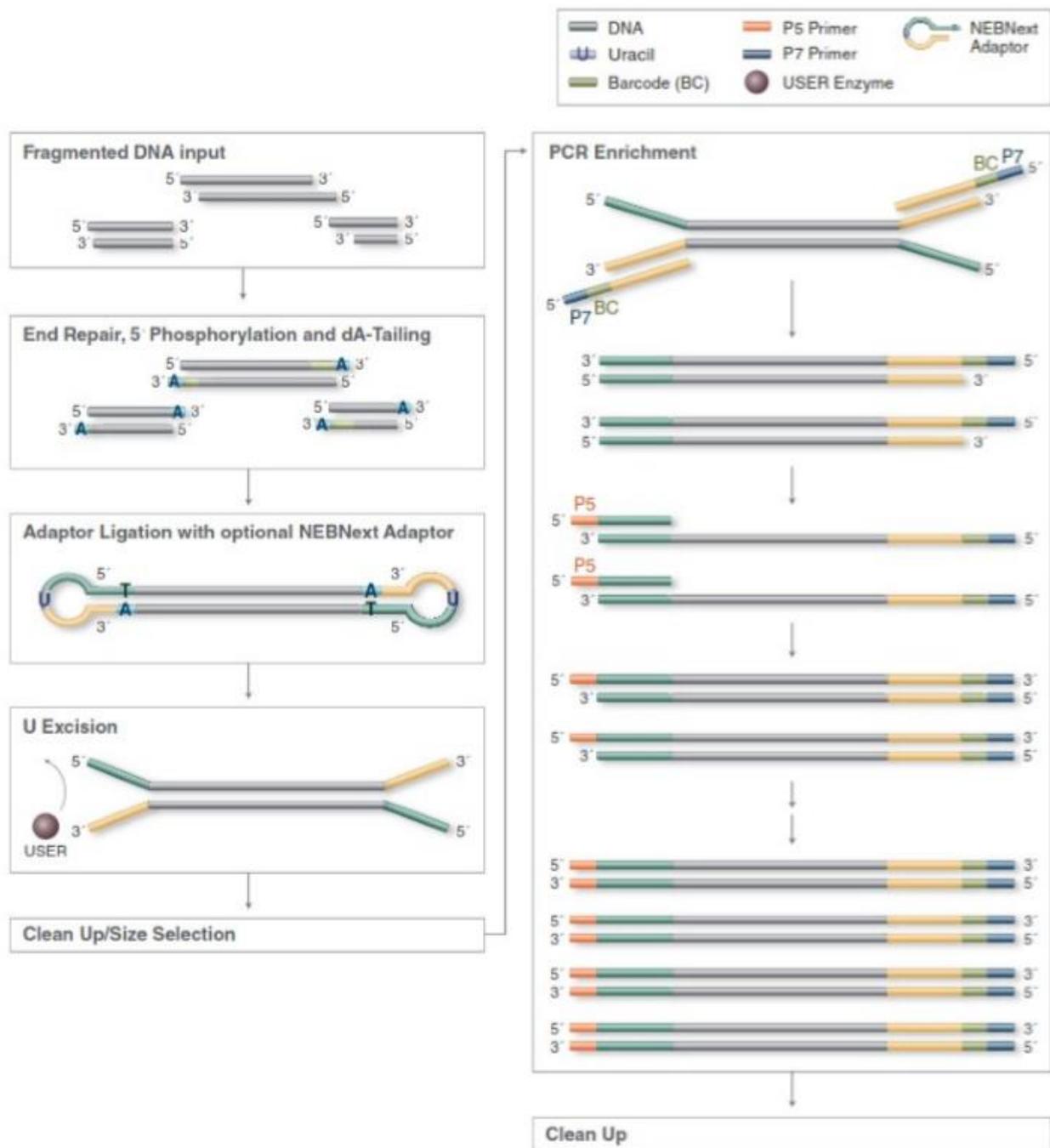
V5 :

GGTAAGCCTATCCCTAACCCCTCTCCTCGGTCTCGA  
TTCTACG

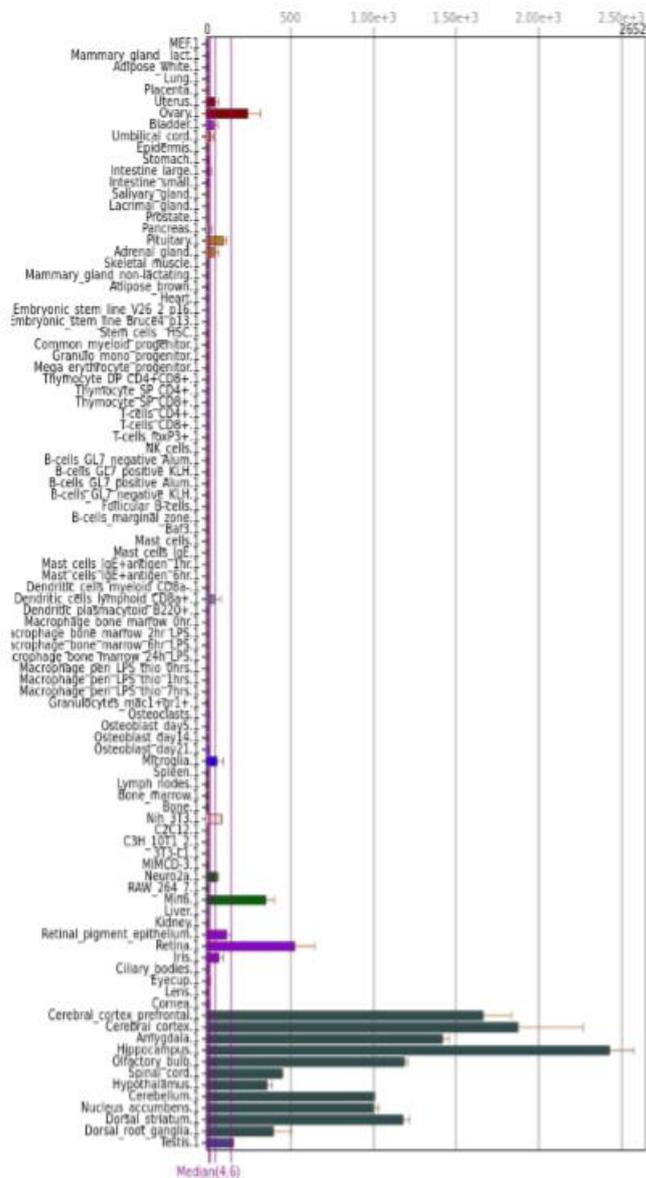
**Figure 19 :** Carte schématique du plasmide pSBtet\_V5\_isoforme PGBD5

Après digestion par EcoRI et MfeI du plasmide pSBtet\_ isoforme PGBD5, le tag V5 est introduit par ligation (A).

La séquence nucléotidique de l'épitope V5 est indiqué en B.

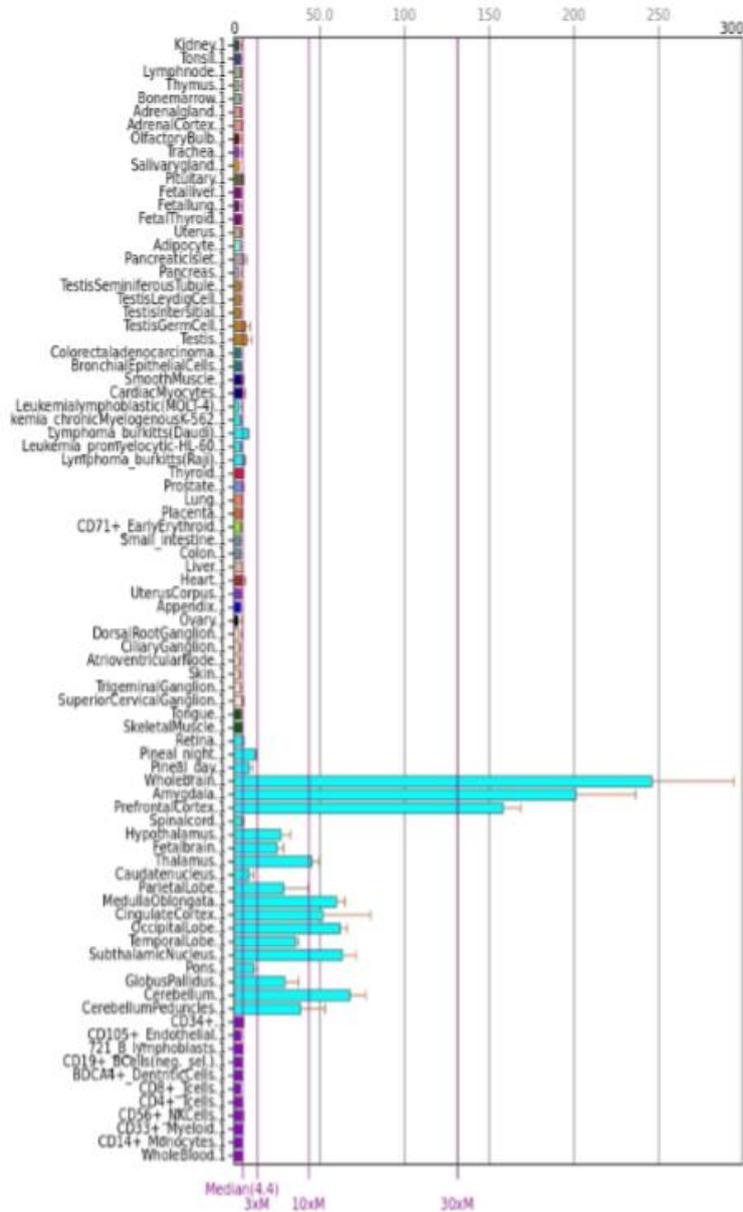


**Figure 20 :** Protocole schématisé de la fabrication de banques Illumina pour le séquençage haut débit.



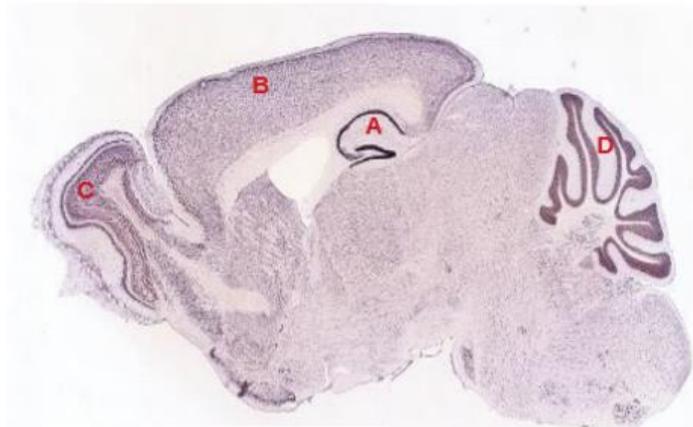
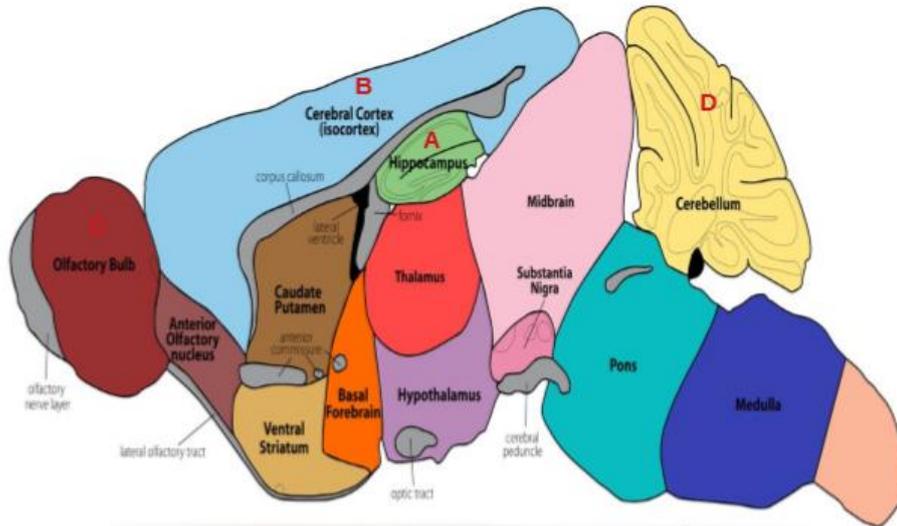
**Figure 21** : Expression du gène *pgbd5* chez la souris par analyse transcriptomique.

(<http://ds.biogps.org/?dataset=GSE10246&gene=209966>)



**Figure 22 : Expression du gène *pgbd5* chez l'humain par analyse transcriptomique.**

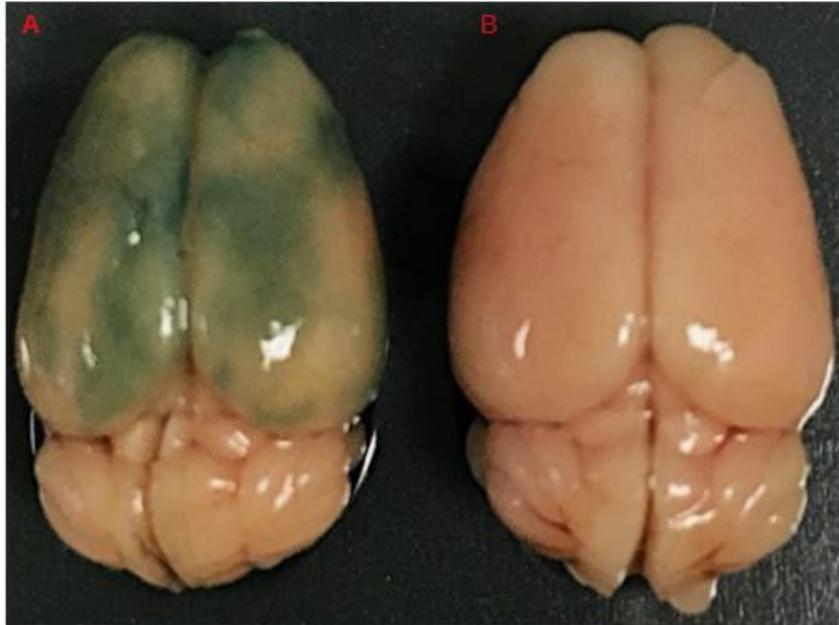
(<http://ds.biogps.org/?dataset=GSE1133&gene=79605>)



**Figure 23 : Coupes sagittale de cerveau de souris.**

En haut, représentation schématique des différentes structures.

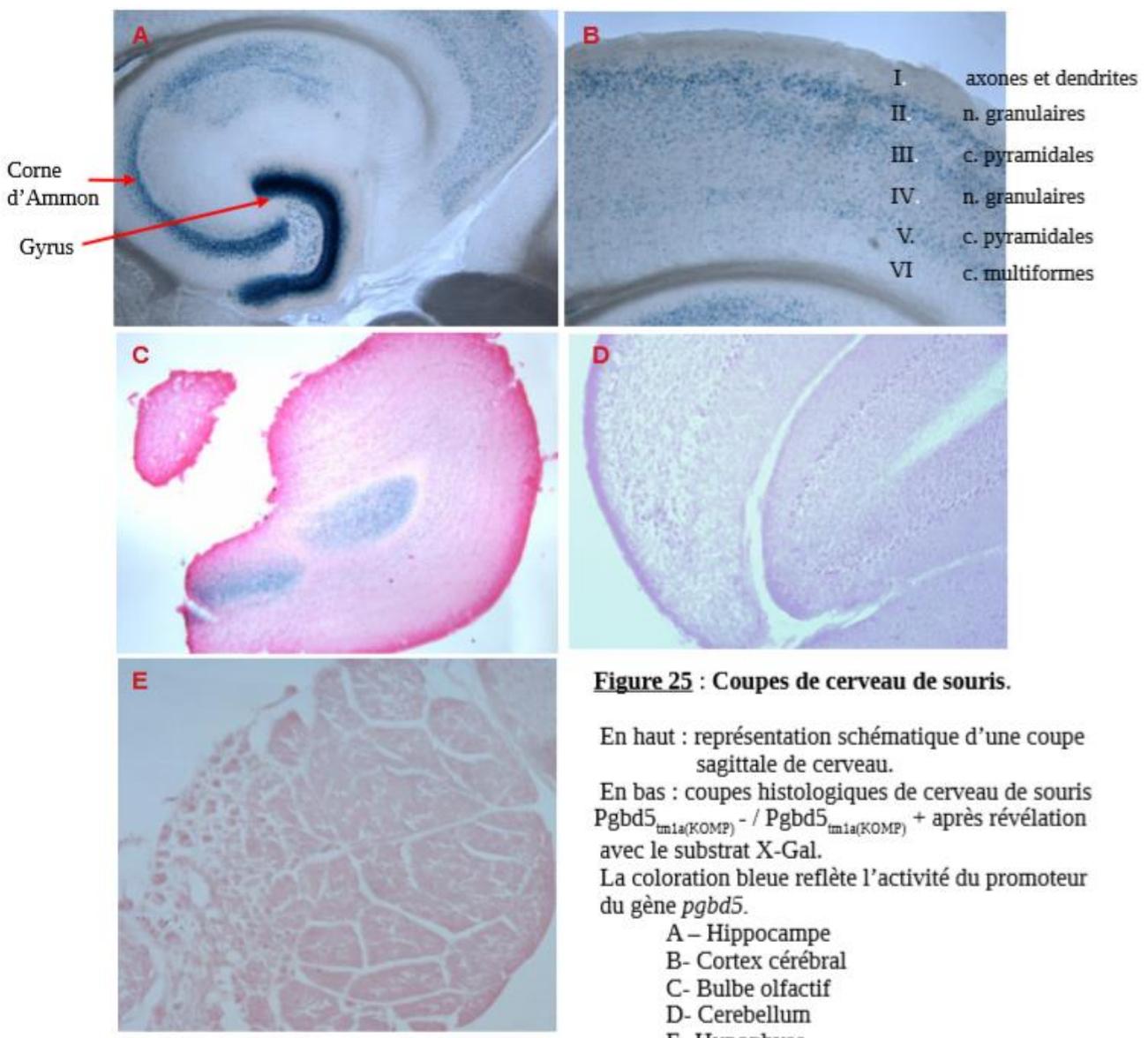
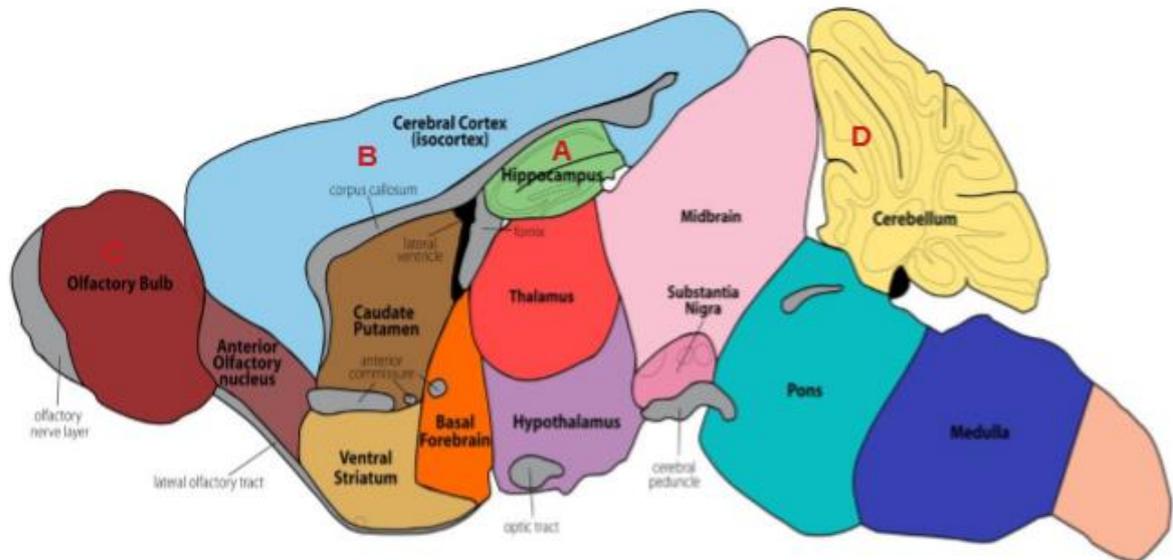
En bas, coupe histologique de cerveau de souris C57BL/6J âgée de 56 jours hybridée avec des ARN *Pgbd5* anti-sens (source [http://mouse.brain-map.org/experiment/thumbnails/70300756?image\\_type=ish&popup=true](http://mouse.brain-map.org/experiment/thumbnails/70300756?image_type=ish&popup=true))



**Figure 24 : Photographie de cerveaux de souris après coloration de la  $\beta$ -galactosidase.**

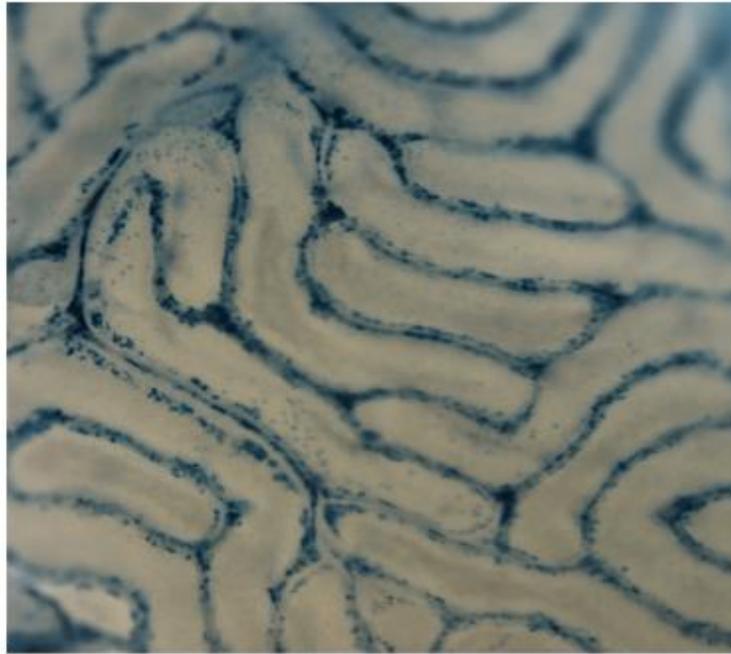
Cerveau de souris adulte muté  $Pgbd5_{tm1a(KOMP)}^- / Pgbd5_{tm1a(KOMP)}^+$  LacZ sous promoteur *pgbd5* (A) et sauvage (B) après coloration en présence de substrat X-Gal.

La coloration bleue reflète l'expression du gène *pgbd5* au niveau du neocortex.



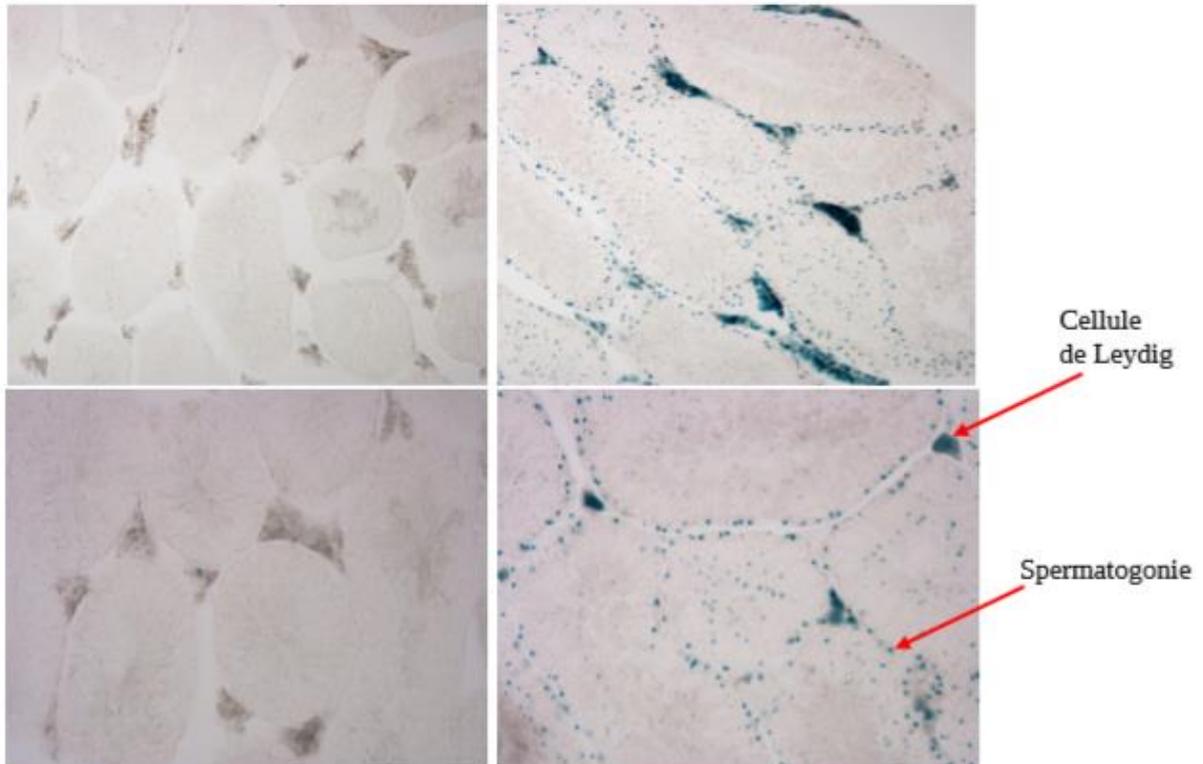
**Figure 25 : Coupes de cerveau de souris.**

En haut : représentation schématique d'une coupe sagittale de cerveau.  
 En bas : coupes histologiques de cerveau de souris  $Pgbd5_{tm1a(KOMP)}^- / Pgbd5_{tm1a(KOMP)}^+$  après révélation avec le substrat X-Gal.  
 La coloration bleue reflète l'activité du promoteur du gène *pgbd5*.  
 A – Hippocampe  
 B- Cortex cérébral  
 C- Bulbe olfactif  
 D- Cerebellum  
 E- Hypophyse



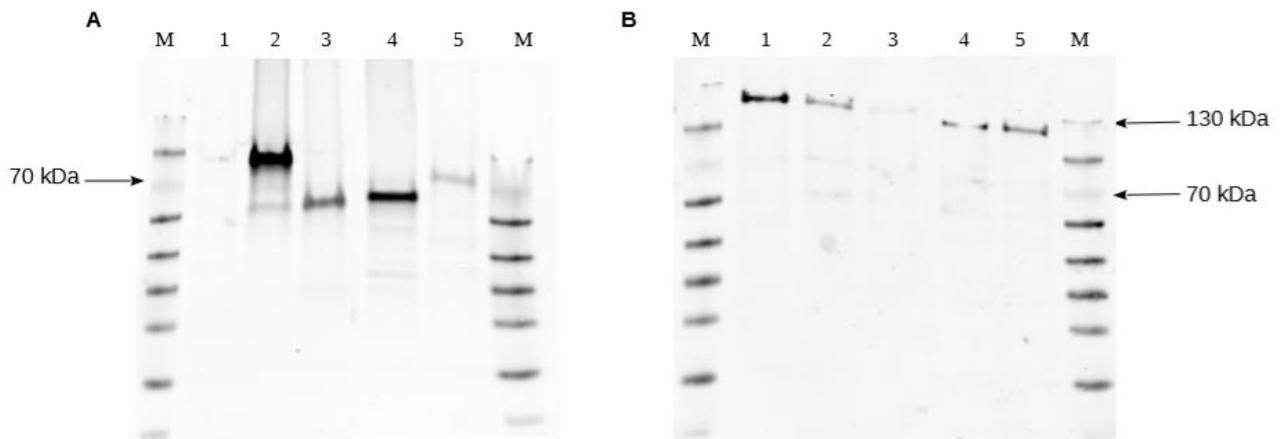
**Figure 26** : Lobules testiculaires de souris  $Pgbd5_{tm1a(KOMP)}^- / Pgbd5_{tm1a(KOMP)}^+$  après révélation de la  $\beta$ -galactosidase avec le substrat X-Gal.

La coloration bleue reflète l'activité du promoteur *pgbd5*.



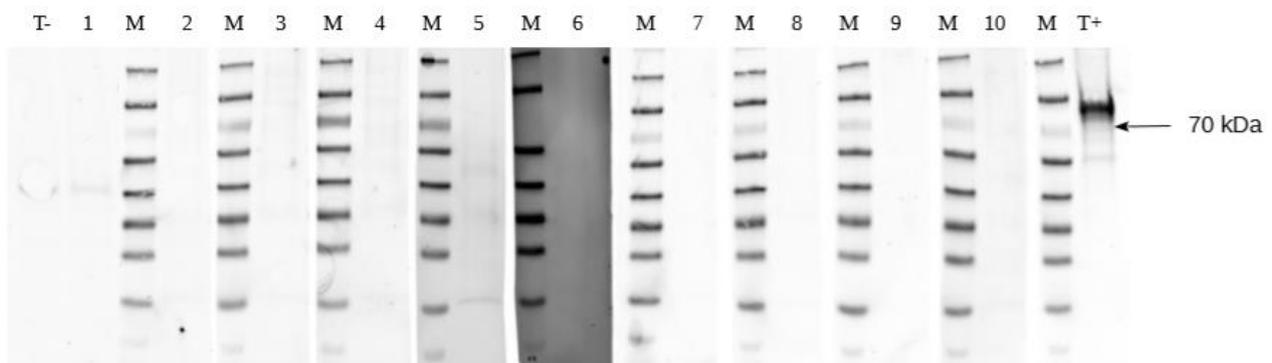
**Figure 27** : Coupes histologiques de testicules de souris  $Pgbd5_{tm1a(KOMP)}^- / Pgbd5_{tm1a(KOMP)}^+$  + après révélation de la  $\beta$ -galactosidase avec le substrat X-Gal.

La coloration bleue reflète l'activité du promoteur *pgbd5*.



**Figure 28 :** Western blot sur extraits de cellules HeLa transfectées ou non par les différentes isoformes PGBD5 en présence d'anticorps anti-myc (A) ou d'anticorps anti-PGBD5 Antibodies-on-line ABIN1112824 (B)

- 1 – extraits de cellules HeLa non transfectées
- 2- extraits de cellules HeLa transfectées par pCS2-PGBD5 Mm-524
- 3- extraits de cellules HeLa transfectées par pCS2-PGBD5 Mm-409
- 4- extraits de cellules HeLa transfectées par pCS2-PGBD5 Hs-554
- 5- extraits de cellules HeLa transfectées par pCS2-PGBD5 Hs- 455



**Figure 29 :** Western blot sur extraits de cellules HeLa transfectées par pCS2\_Mm523\_Myc.

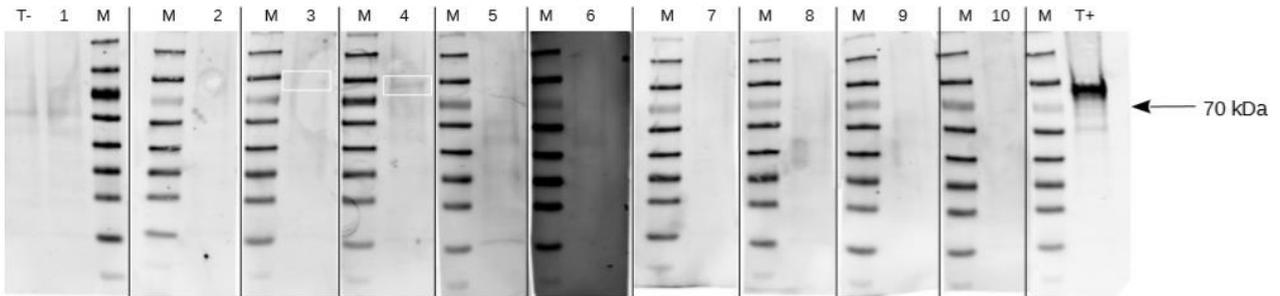
Les membranes sont hybridées avec les dix sérums murins anti-PGBD5 J71 au 1/250ème puis hybridées après lavages avec un anticorps secondaire anti-mouse IRDye 680 au 1/10 000ème.

Après lavages, les membranes sont lues sur le scanner InfraRouge Odyssey.

Témoin négatif (T-) : HeLa non transfectées

Témoin positif (T+) : hybridation avec l'anti-myc au 1/800ème

- |                      |                         |
|----------------------|-------------------------|
| 1- sérum SWISS 1 J71 | 6- sérum C57 BL6 1 J71  |
| 2- sérum SWISS 2 J71 | 7- sérum C57 BL6 2 J71  |
| 3- sérum SWISS 3 J71 | 8- sérum C57 BL6 3 J71  |
| 4- sérum SWISS 4 J71 | 9- sérum C57 BL6 4 J71  |
| 5- sérum SWISS 5 J71 | 10- sérum C57 BL6 5 J71 |



**Figure 30 : Western blot sur extraits de cellules HeLa transfectées par pCS2\_Mm523\_Myc.**

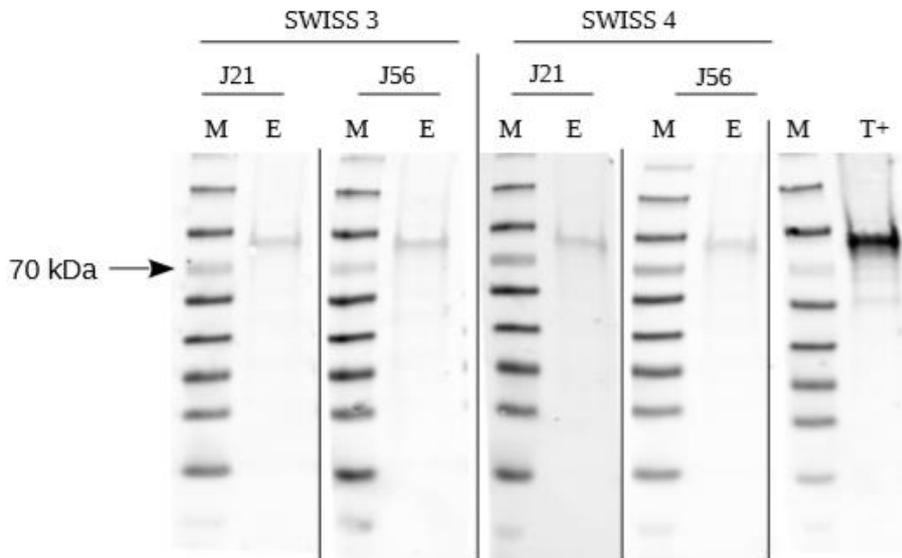
Les membranes sont ré-hybridées avec les dix sérums murins anti-PGBD5 J42 au 1/250ème puis hybridées après lavages avec un anticorps secondaire anti-souris IRDye 680 à 0,1 mg/µL.

Après lavages, les membranes sont lues sur le scanner InfraRouge Odyssey.

Témoin négatif (T-) : HeLa non transfectées

Témoin positif (T+) : hybridation avec l'anti-myc à 5 µg/mL

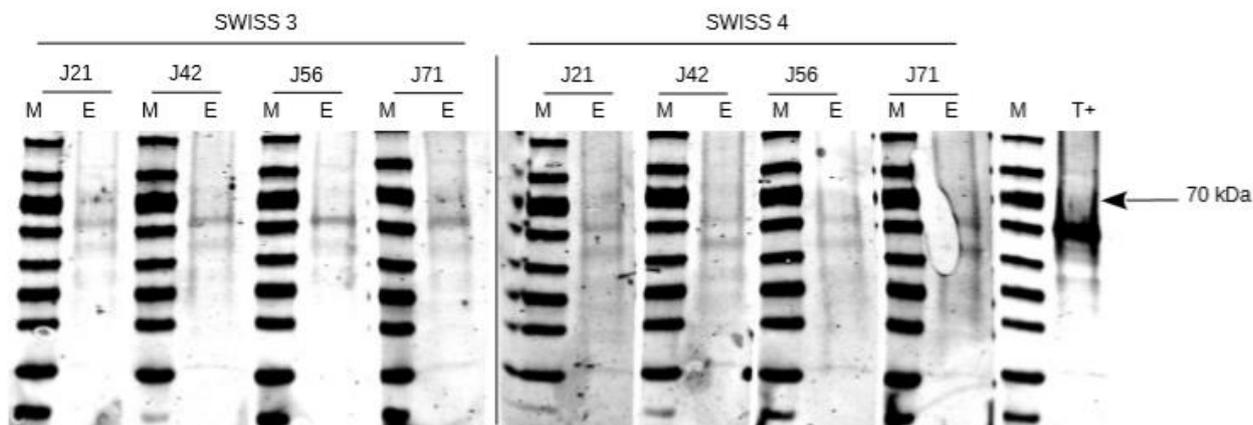
- |                      |                         |
|----------------------|-------------------------|
| 1- sérum SWISS 1 J42 | 6- sérum C57 BL6 1 J42  |
| 2- sérum SWISS 2 J42 | 7- sérum C57 BL6 2 J42  |
| 3- sérum SWISS 3 J42 | 8- sérum C57 BL6 3 J42  |
| 4- sérum SWISS 4 J42 | 9- sérum C57 BL6 4 J42  |
| 5- sérum SWISS 5 J42 | 10- sérum C57 BL6 5 J42 |



**Figure 31 : Western blot sur extraits de cellules HeLa transfectées par pCS2\_Mm523\_Myc.**

Les membranes sont hybridées avec les deux sérums murins anti-PGBD5 SWISS 3 (J21 et J52) et SWISS 4 (J21 et J52) au 1/250ème puis hybridées après lavages avec un anticorps secondaire anti-souris IRDye 680 à 0,1 mg/µL. Après lavages, les membranes sont lues sur le scanner InfraRouge Odyssey.

Témoin positif (T+) : hybridation avec l'anti-myc à 5 µg/mL

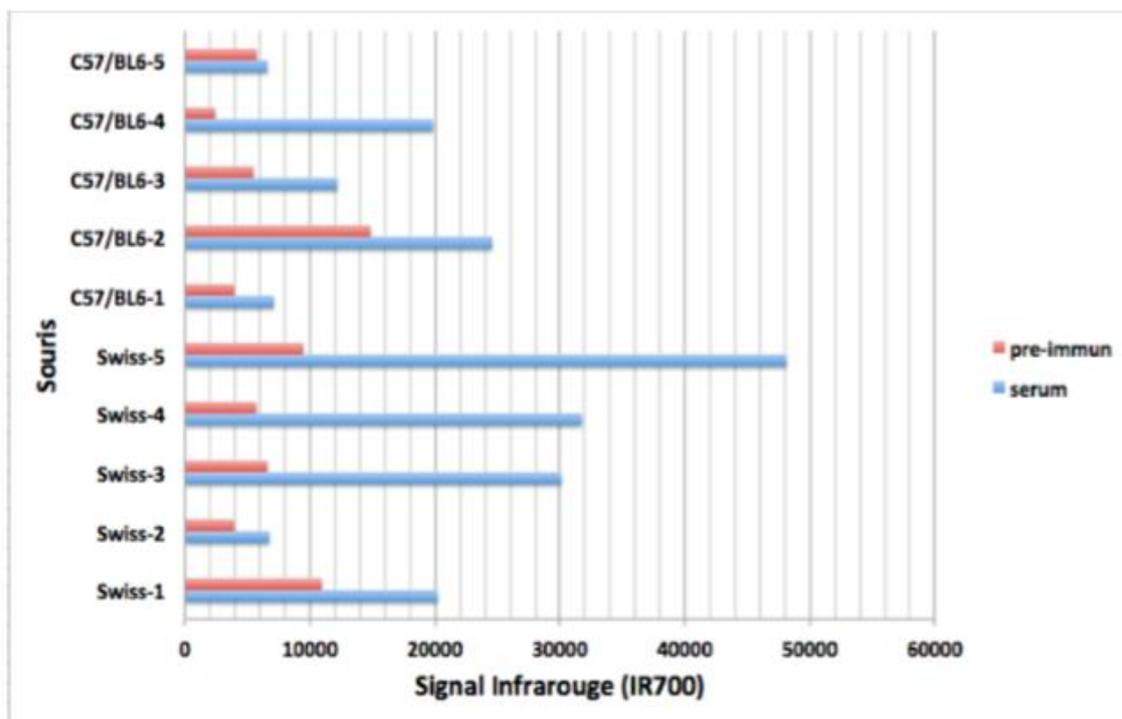


**Figure 32 : Western blot sur extraits de cellules HeLa transfectées par pCS2\_Mm409\_Myc.**

Les membranes sont hybridées avec les sérums murins anti-PGBD5 SWISS 3 (J21 à J71) (A) et SWISS 4 (J21 à J71) (B) au 1/250ème puis hybridées après lavages avec un anticorps secondaire anti-souris IRDye 680 à 0,1 mg/μL.

Après lavages, les membranes sont lues sur le scanner InfraRouge Odyssey.

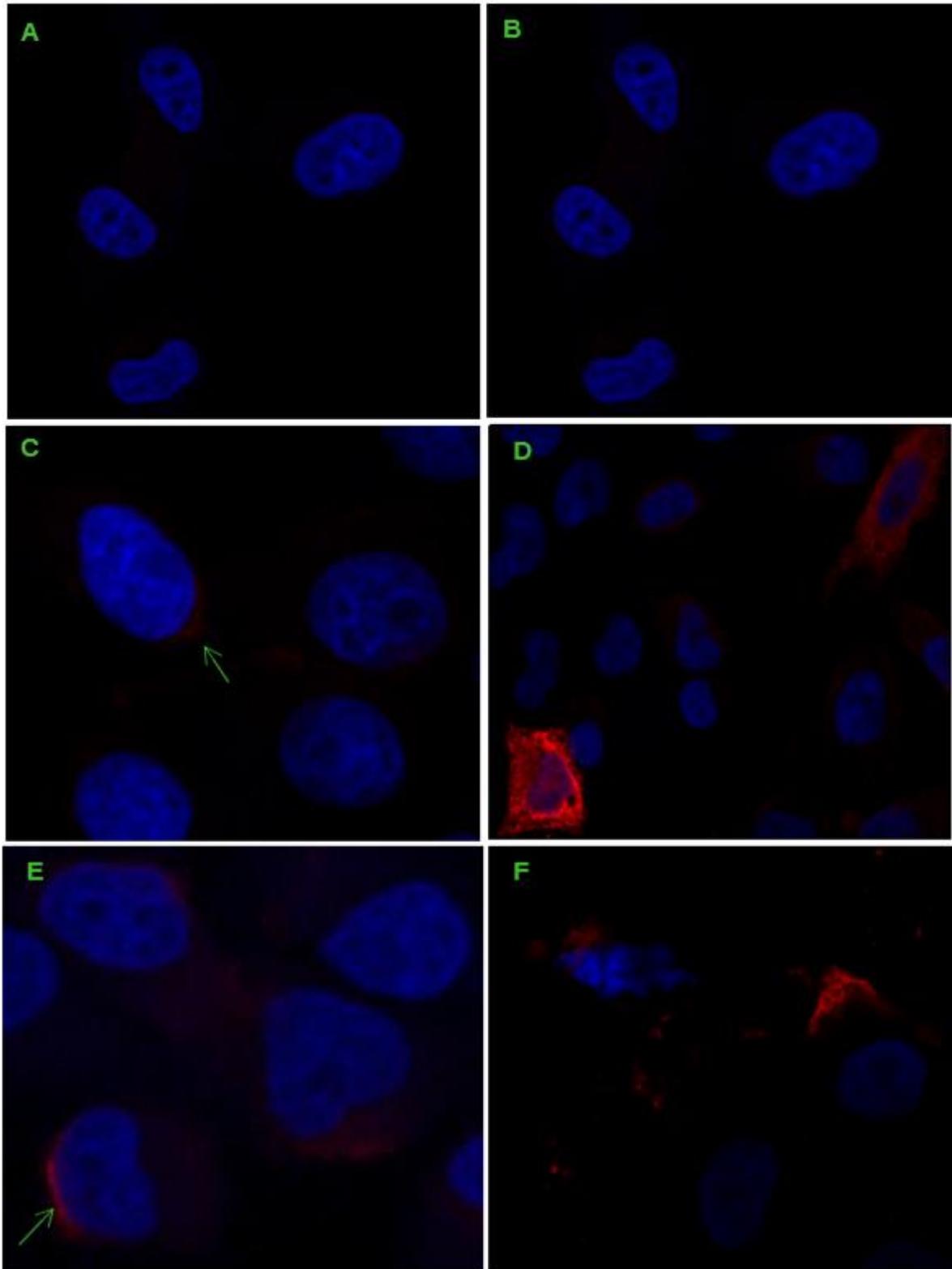
Témoin positif (T+) : hybridation avec l'anti-myc à 5 μg/mL



**Figure 33: Représentation graphique de la présence d'anticorps structuraux des différents sérums murins J71**

Les spots d'extraits protéiques non dénaturés de cellules HeLa transfectées ou non avec pCS2-myc-PGBD5 Mm523 sont hybridés avec le sérum pré-immun ou sérum J71 correspondant à chaque souris au 1/250ème puis hybridés après lavage avec l'anticorps secondaire anti-mouse IRDye 680 à 0,1 mg/μL.

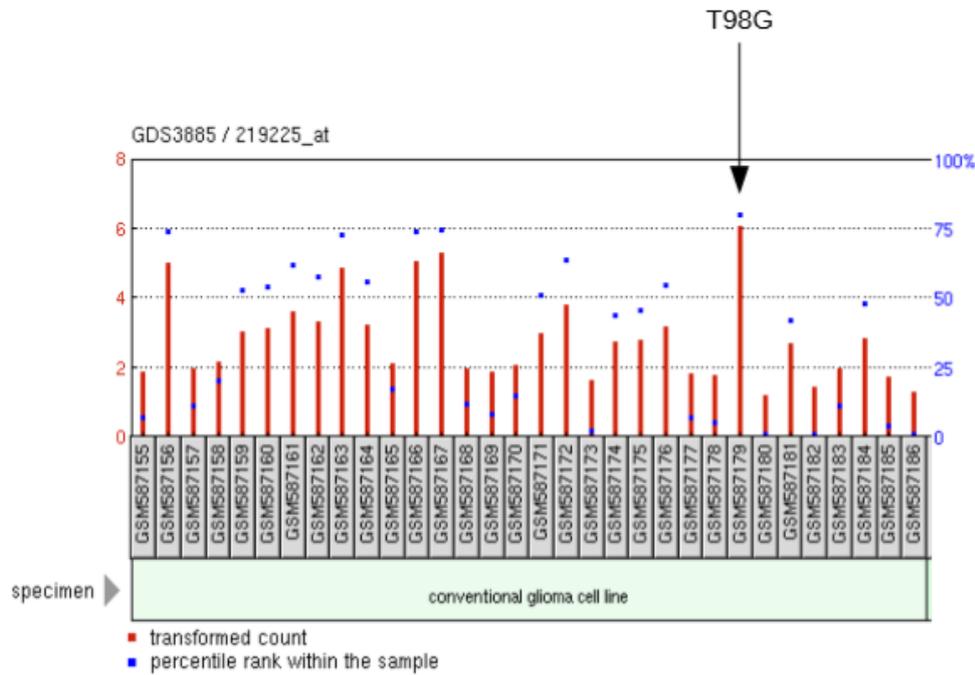
Après lavages, les membranes sont lues sur le scanner InfraRouge Odyssey et la différence de signal entre les deux types de sérums est représentée sur ce graphique.



**Figure 34:** Exemples d'observations en microscopie confocale des cellules transfectées par pCS2 seul (A et B), pCS2-PGBD5 Mm-409 (C et D) ou Mm-523 (E et F) et hybridées avec le sérum anti-PGBD5 (colonne de gauche) ou l'anticorps anti-myc (colonne de droite).

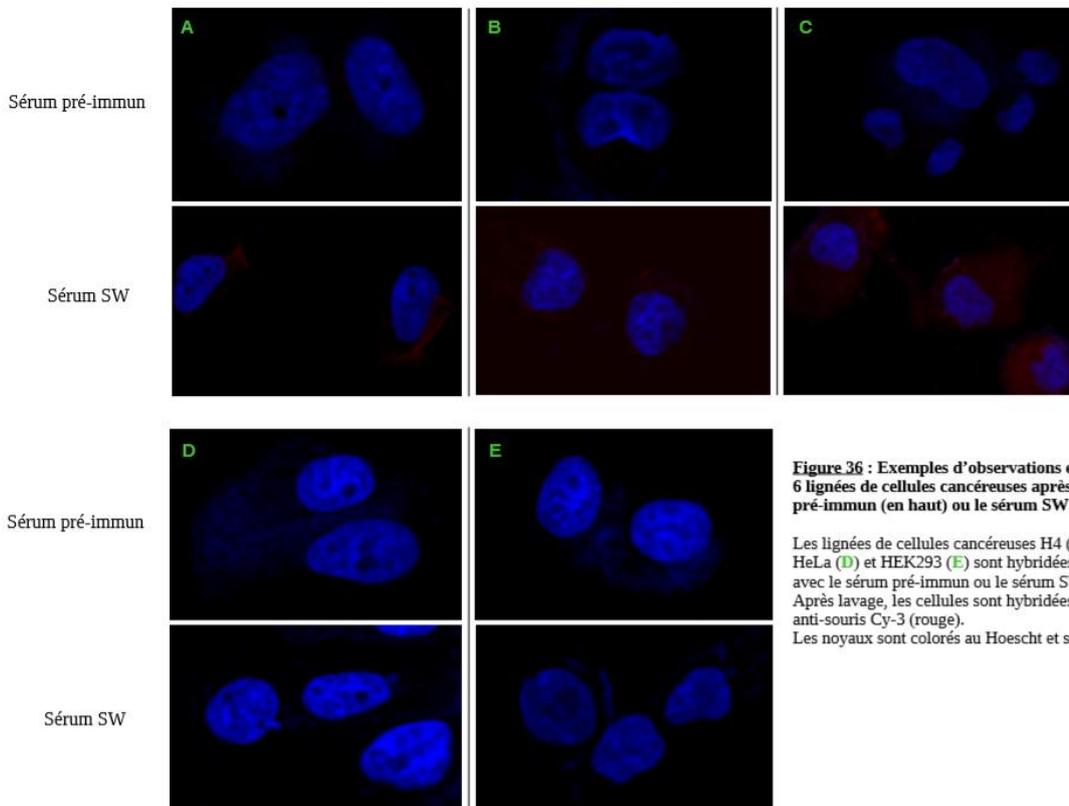
Les cellules sont hybridées 24h après transfection, avec l'anticorps SW au 1/50ème ou l'anticorps anti-myc au 1/50ème. Après lavage, les cellules sont hybridées avec un anticorps secondaire anti-souris Cy-3 (rouge).

Les noyaux sont colorés au Hoescht et sont représentés en bleu.



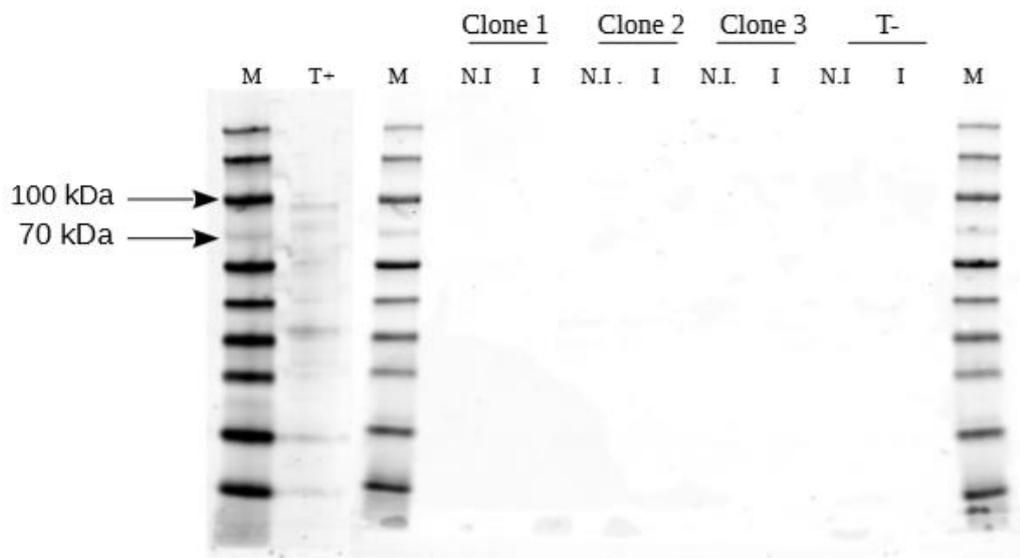
**Figure 35: Exemple de données obtenues pour la lignée de cellules cancéreuses T98G à partir du site du National Center for Biotechnology Information (NCBI)(Geo Profil)**

Adapté d'après : [https://www.ncbi.nlm.nih.gov/geo/tools/profileGraph.cgi?ID=GDS3885:219225\\_at](https://www.ncbi.nlm.nih.gov/geo/tools/profileGraph.cgi?ID=GDS3885:219225_at)



**Figure 36 : Exemples d'observations en microscopie confocale des 6 lignées de cellules cancéreuses après hybridation avec le sérum pré-immun (en haut) ou le sérum SW (en bas).**

Les lignées de cellules cancéreuses H4 (A), 8MG (B), T98G (C), HeLa (D) et HEK293 (E) sont hybridées 24h après transfection, avec le sérum pré-immun ou le sérum SW au 1/50ème. Après lavage, les cellules sont hybridées avec un anticorps secondaire anti-souris Cy-3 (rouge). Les noyaux sont colorés au Hoescht et sont représentés en bleu.



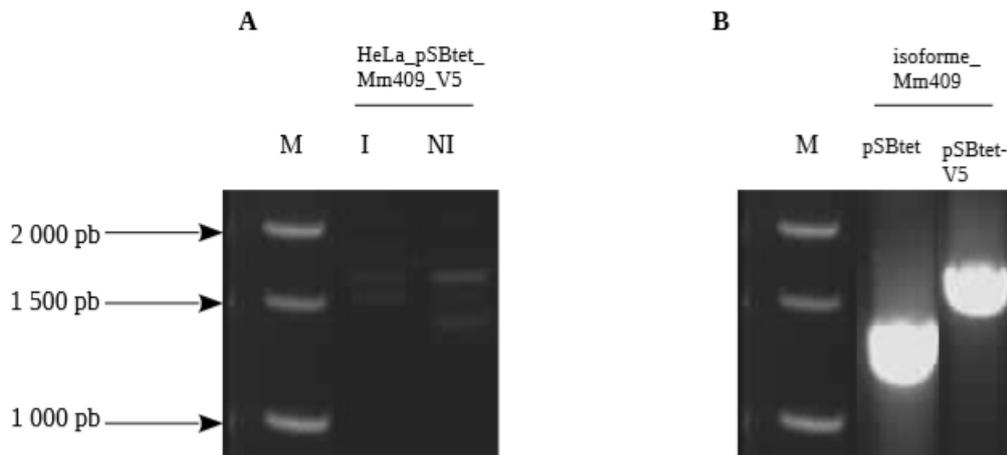
**Figure 37 : Exemple de Western Blot d'extraits protéiques de clones de cellules HeLa pSBtet\_V5 Mm PGBD5 avec et sans induction de doxycycline à 1 µg/mL pendant 24h**

Les extraits protéiques des clones cellulaires non induits (N.I.) et induits (I) sont hybridés en présence d'anticorps anti-V5 R960-25 au 1/5000ème puis hybridés après lavages avec l'anticorps secondaire anti-souris IRDye 680 à 0,1 mg/µL.

T+ : Extraits protéiques de cellules transfectées avec plasmide Metnase\_V5 : taille attendue : env. 90 kDa

T- : Extraits protéiques de cellules HeLa\_pSBtet\_GP

M : Marqueur de poids moléculaire PageRuler™ Prestained Protein Ladder, 10 à 180 kDa

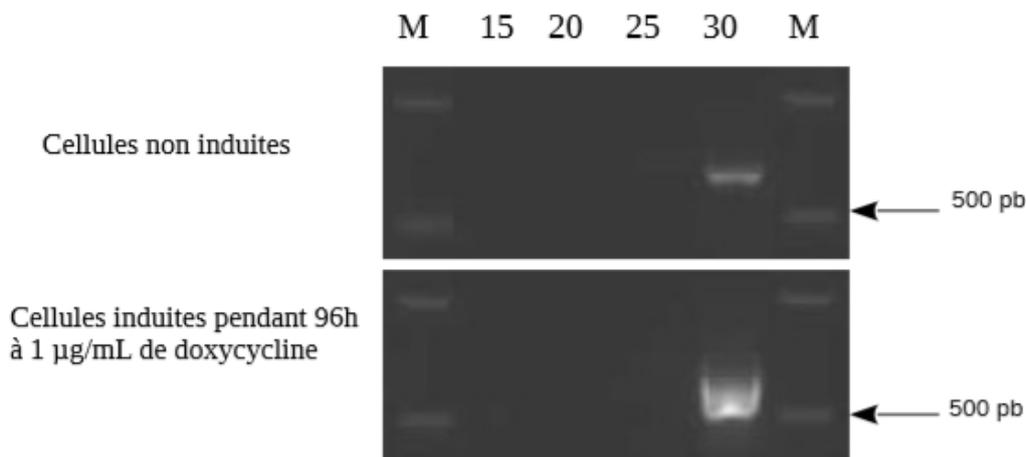


**Figure 38 : Détection du transgène *pgbd5* par PCR sur des extraits de clones cellulaires HeLa\_pSBtet\_Mm409\_V5 avec ou sans induction à la doxycycline (A) et dans les différents plasmides pSBtet contenant l'isoforme Mm409 (B)**

(A) Les cellules sont induites (ou non) avec 1 µg/mL de doxycycline pendant 96h puis récupérées pour extraction de l'ADN génomique. Une amplification du gène *pgbd5* est réalisée sur 35 cycles et les produits sont déposés sur gel d'agarose 1,5 % en présence de GelRed en parallèle d'un marqueur de poids moléculaire.

(B) En contrôle, le transgène *pgbd5* est amplifié dans le plasmide pSBtet\_Mm409 ainsi que dans le plasmide pSBtet\_Mm409\_V5. Ce dernier est utilisé pour la fabrication des clones cellulaires.

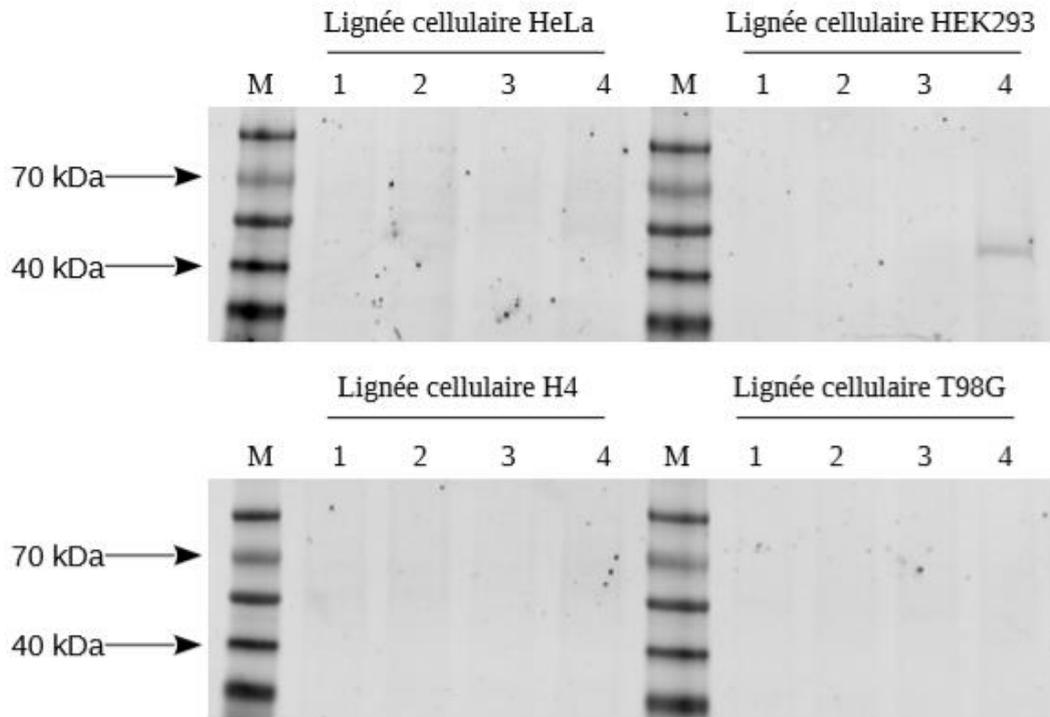
Tailles attendues : pSBtet\_Mm409 : X pb  
pSBtet\_Mm409\_V5 : 1 696 pb



**Figure 39 : Analyse de l'expression de *pgbd5* par RT-PCR semi quantitative dans les cellules HeLa\_pSBtet\_Mm409\_V5 avec ou sans induction à la doxycycline**

Les cellules sont induites (ou non) avec 1 µg/mL de doxycycline pendant 96h puis récupérées pour extraction des ARN. Après une transcription inverse, une amplification du gène *pgbd5* est réalisée et les produits obtenus après 15, 20, 25 et 30 cycles sont déposés sur gel d'agarose 1,5 % en présence de GelRed en parallèle d'un marqueur de poids moléculaire.

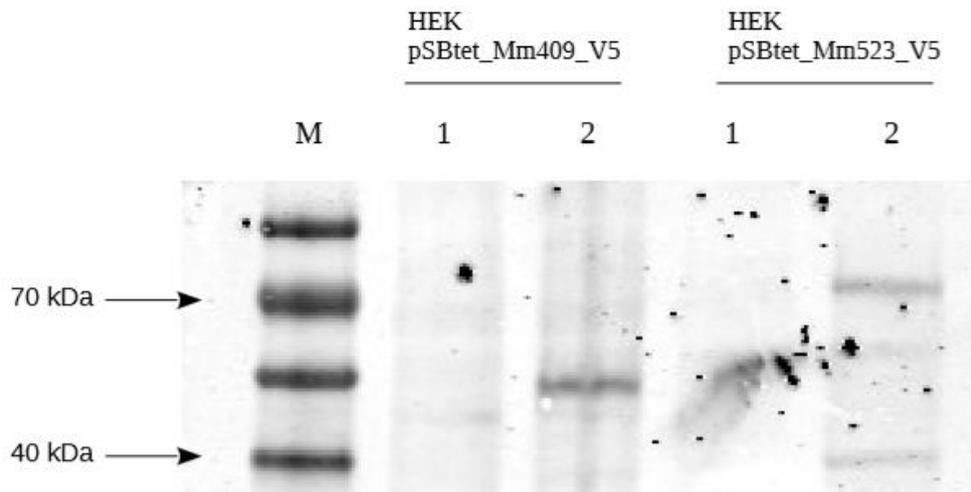
Taille attendue : env. 730 pb



**Figure 40 : Western blot sur extraits de cellules cancéreuses transfectées par le plasmide pSBtet\_GP (puits 1 et 2) ou pSBtet\_Mm409\_V5 (puits 3 et 4) après induction ou non.**

Les extraits protéiques sont hybridés en présence d'anticorps anti-V5 R960-25 au 1/5000ème puis hybridés après lavages avec l'anticorps secondaire anti-souris IRDye 680 à 0,1 mg/μL. L'isoforme Mm409 a un poids moléculaire de 46,5 kDa.

- 1 – extraits de cellules transfectées par pSBtet\_GP non induites
- 2- extraits de cellules transfectées par pSBtet\_GP induites
- 3- extraits de cellules transfectées par pSBtet\_Mm409\_V5 non induites
- 4- extraits de cellules transfectées par pSBtet\_Mm409\_V5 induites
- M- Marqueur de poids moléculaire



**Figure 41** : Western blot sur extraits de cellules cancéreuses HEK293 transfectées par le plasmide pSBtet\_Mm409\_V5 ou pSBtet\_Mm523\_V5 après induction ou non.

Les extraits protéiques sont hybridés en présence d'anticorps anti-V5 R960-25 au 1/5000ème puis hybridés après lavages avec l'anticorps secondaire anti-souris IRDye 680 à 0,1 mg/μL.

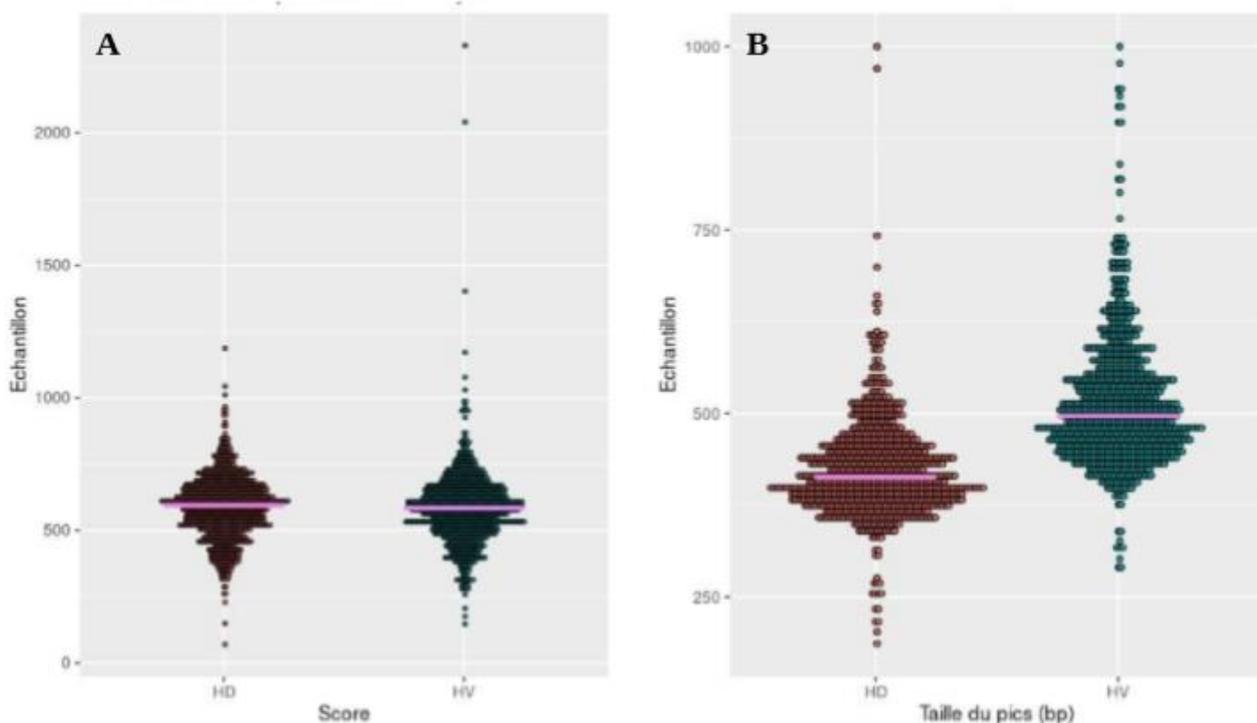
L'isoforme Mm409 a un poids moléculaire de 46,5 kDa.

L'isoforme Mm523 a un poids moléculaire de X kDa.

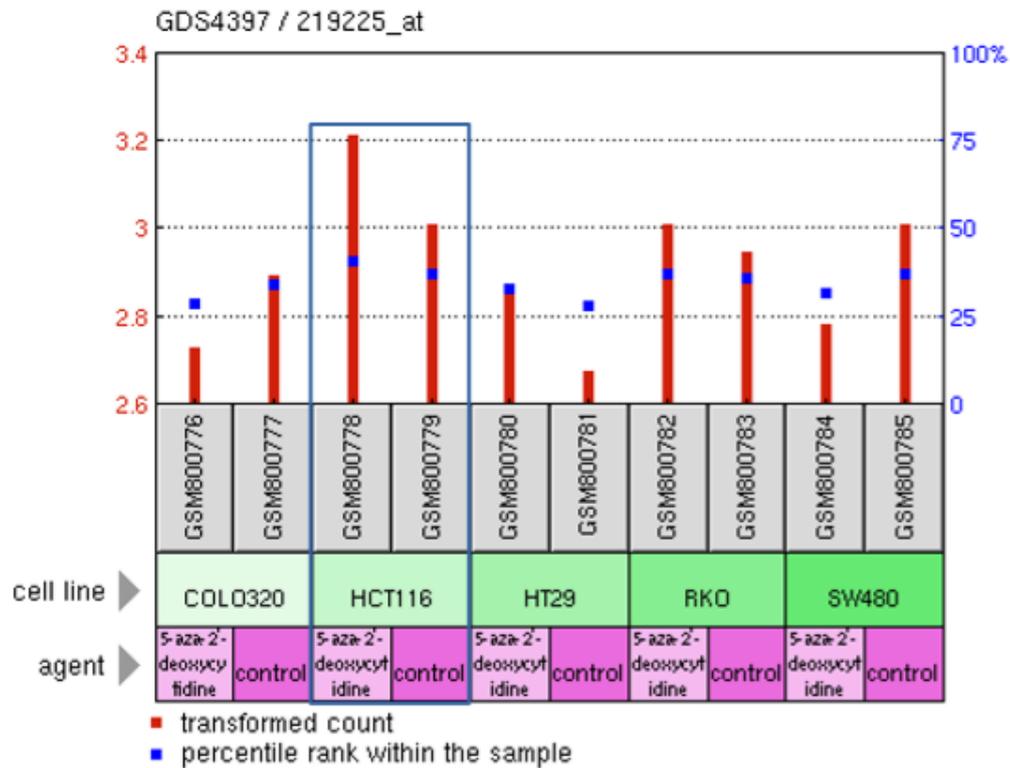
1 – extraits de cellules transfectées non induites

2- extraits de cellules transfectées induites par la doxycycline

M- Marqueur de poids moléculaire

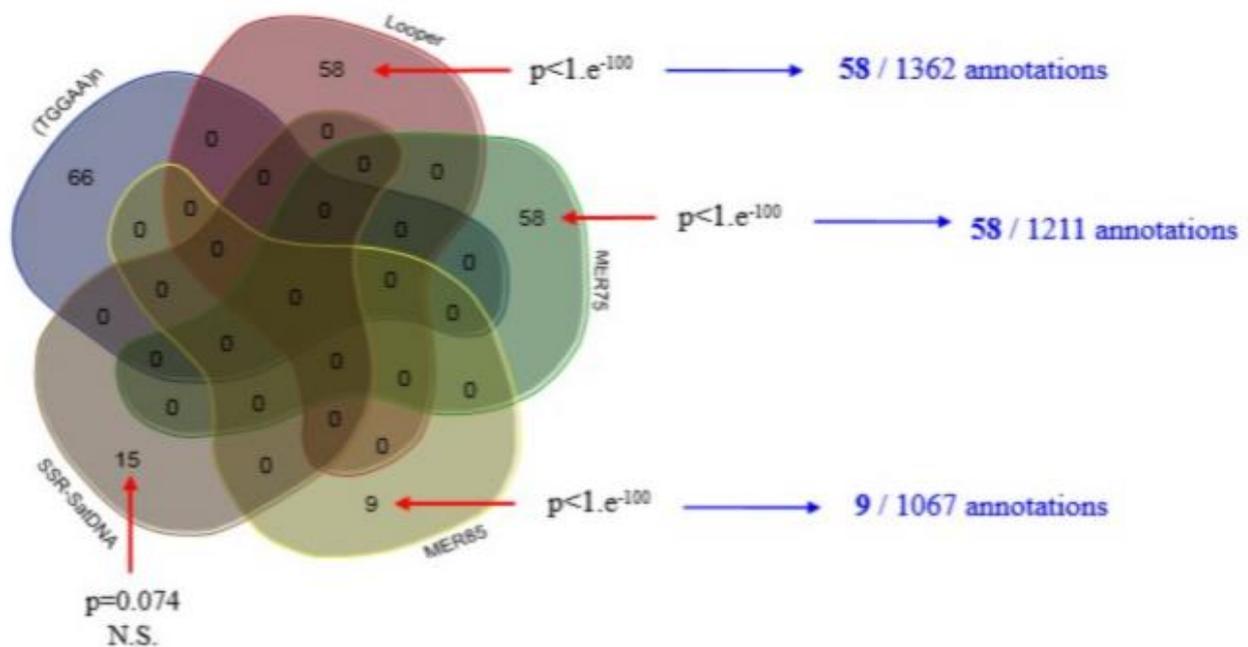


**Figure 42** : Comparaison de l'amplitude (A) et de la taille (B) en paire de bases des pics obtenus pour l'hippocampe dorsal (HD en rose saumon) et l'hippocampe ventral (HV en bleu turquoise)

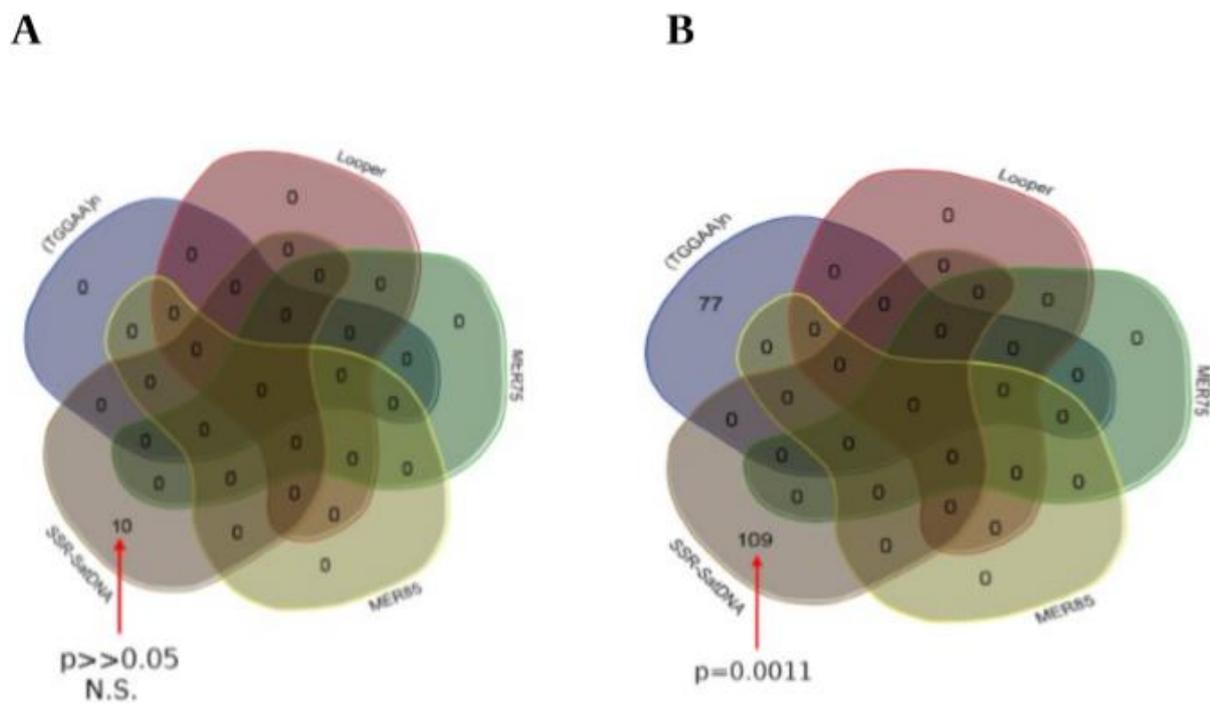


**Figure 43 :** Exemple de données obtenues pour la lignée de cellules cancéreuses HCT116 à partir du site du National Center for Biotechnology Information (NCBI) (Geo Profil)

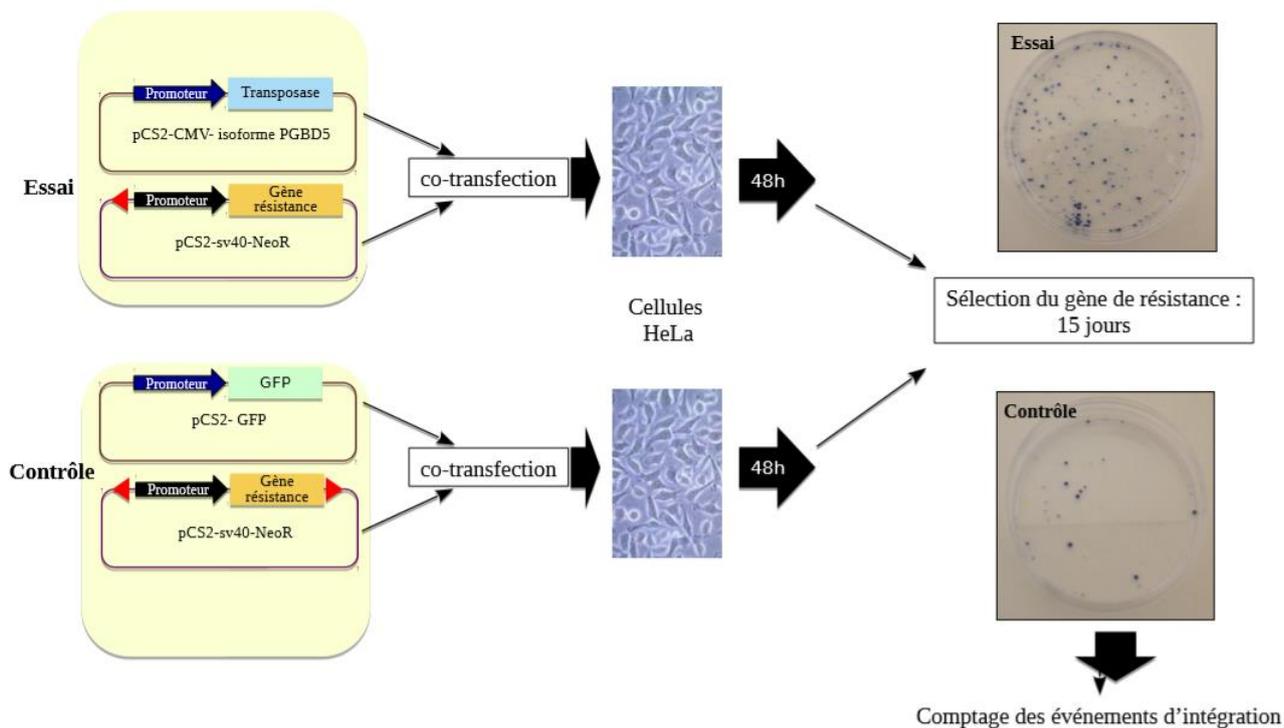
[https://www.ncbi.nlm.nih.gov/geo/tools/profileGraph.cgi?ID=GDS4397:219225\\_at](https://www.ncbi.nlm.nih.gov/geo/tools/profileGraph.cgi?ID=GDS4397:219225_at)



**Figure 44 :** Diagramme de Venn représentant les différentes co-localisations des pics obtenus pour la lignée cellulaire cancéreuse T98G.

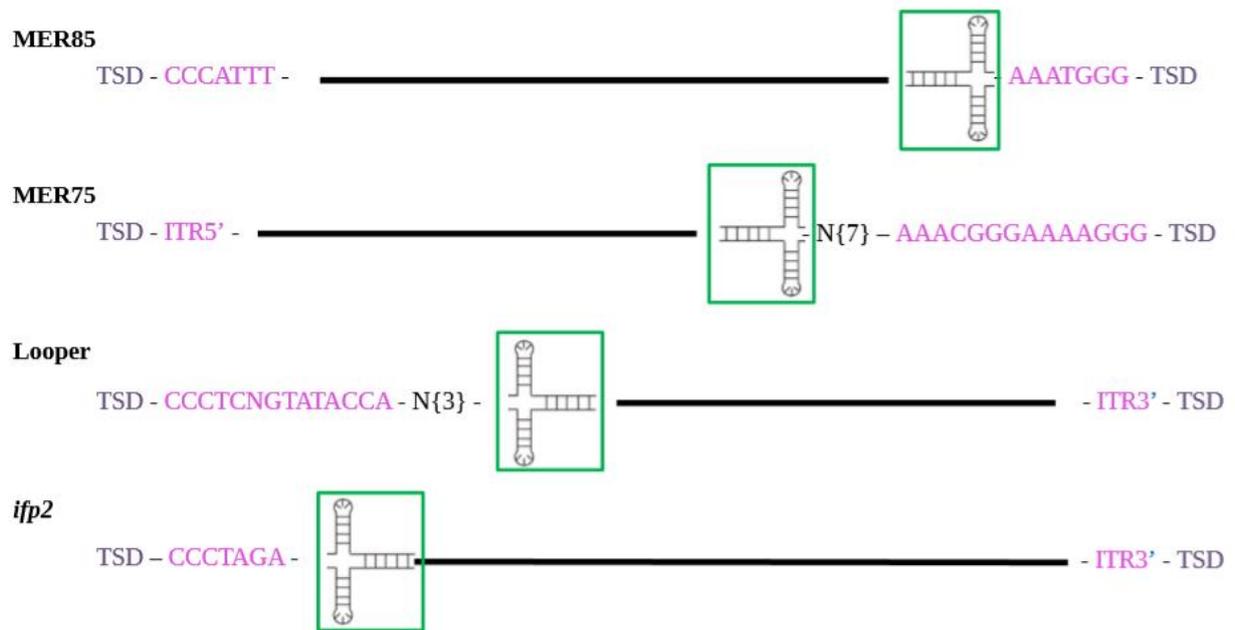


**Figures 45 :** Diagramme de Venn représentant les différentes co-localisations des pics obtenus pour les lignées cellulaires cancéreuses HCT116 (A) et G401 (B) .



**Figure 46 :** Représentation d'un test d'intégration aléatoire en cellules HeLa.

48h après co-transfection, les cellules HeLa sont sélectionnées pendant 15 jours avec du G-418 à 0,8 mg/mL (analogue plus efficace de la néomycine en cellule animale). Les cellules sont ensuite colorées au bleu de méthylène puis comptées.



**Figure 47 :** Représentation schématique des extrémités des *pbls* humains et de *ifp2*.

<b>Lignée cellulaire</b>	<b>Milieu de culture</b>
<b>HeLa HEK HCT116 H4</b>	DMEM
	4 mM Glutamine
	10 % SVF
	100 U / mL pénicilline
	0,1 mg/ mL streptomycine
<b>TG98G 8MG</b>	MEM
	200 mM Glutamine
	10 % SVF
	100 U / mL pénicilline
	0,1 mg/ mL streptomycine

**Tableau 1** : Composition des milieux de culture selon les lignées cellulaires.

DMEM : Dulbecco's Modified Eagle Medium

MEM : Minimal Essential Medium

SVF : Sérum de veau foetal décomplémenté 30 min à 56 °C

**A**

Gène	Nom amorce	Sens	Séquence amorce (5' → 3')
<i>pgbd5</i> sauvage	CSD-gene-F	sens	GCTACCATTACCAGTTGGTCTGGTGTC
	CSD-gene-ttR	Anti-sens	CACAAATAGAATCGAAGGGCACCTC
<i>LacZ</i>	LacZ1	sens	GTCGTTTTACAACGTCGTGACT
	LacZ2	Anti-sens	GCACGGTTACGATGCGCCCATC

**B**

Réactifs	Gène <i>LacZ</i>	Gène <i>pgbd5</i> WT
Tampon GoTaq 5X	1 X	1 X
dNTP (10 mM)	2 mM	0,04 mM
MgCl <sub>2</sub> (25 mM)	2,5 mM	1 mM
Primer sens	0,1 µM	0,2 µM
Primer anti-sens	0,1 µM	0,2 µM
GoTaq	0,25 U	0,25 U
ADN génomique	1 µL	1 µL
eau milliQ	Qsp 50 µL	Qsp 50 µL

**C**

Étapes	Température	Durée	Nombre de cycles
Dénaturation initiale	94 °C	2 min	1 cycle
Dénaturation	94 °C	30 sec	X 35 cycles
Hybridation	60 °C	30 sec	
Extension	72 °C	1 min	
Extension fin	72 °C	5 min	1 cycle
Maintien	4 °C		

**Tableaux 2 :** Conditions de PCR pour le génotypage des souris

**A-** Séquences des amorces utilisées pour l'amplification des gènes *LacZ* et *pgbd5* sauvage (*pgbd5* WT)

**B-** Composition du mélange réactionnel avec le kit GoTaq (Promega) pour la PCR

**C-** Programme PCR utilisé pour l'amplification des gènes *LacZ* et *pgbd5* sauvage

Plasmide	Utilisation
pBS- <i>Ifp2</i> -sv40-NeoR	source de transposon
pGH-MER75-NeoR	source de transposon
pGH-MER75B-NeoR	source de transposon
pGH-MER85-NeoR	source de transposon
pGH-Looper-NeoR	source de transposon
pCS2-myc-Mm523	source de transposase
pCS2-myc-409	source de transposase

**Tableau 3 :** Liste des plasmides utilisés dans les tests de transposition.

En bleu, les plasmides utilisés comme source de transposon (possédant le gène de résistance à la néomycine (NeoR).

En vert, les plasmides utilisés comme source de transposase (contenant l'une des deux transposase murine de *pgbd5*)

Réactifs	Volume / réaction	Concentration finale
Tampon RT 10X	2 µL	1 X
Mix dNTP (5 mM chaque dNTP)	2 µL	0,5 mM chacun
Oligo d(T) (10 µM)	2 µL	1 µM
Inhibiteur RNase (10 U/µL)	1 µL	10 U
Enzyme Omniscript RT	1 µL	4 U
Eau exempte de RNase	variable	
ARN	variable	2 µg
Volume total	Qsp 20 µL	

**Tableau 4 :** Composition du mélange réactionnel pour la reverse transcription Omniscript.

Isoforme	Orientation	Séquence 5' → 3'	Site restriction (vert) pour enzyme
Mm523	Sens	CCGAAGAATTCATGGCCGAGGGCGGTGG	<i>EcoRI</i>
	Anti-sens	CCGTTCTAGATCAGTGGGCTGGCGATGAG	<i>XbaI</i>
Mm409	Sens	CCGAAGAATTCATGCCACCCAGCGCTAGC	<i>EcoRI</i>
	Anti-sens	CCGTTCTCGAGTCAGTGGGCTGGCGATGAG	<i>XhoI</i>

**Tableau 5 :** Séquences des amorces PCR flanquées de sites de restriction appropriés pour le clonage.

Nom Enzyme	Ref. NEB	Type digestion	Tampon	Nombre Unité (U)	Température restriction (°C)	inactivation enzyme
ClaI	R0197	simple	CutSmart	10 U	37°C	65°C 20 min
EcoRI	R0101	simple	NEBbuffer EcoRI	20 U	37°C	65°C 20 min
MfeI	R0589	simple	CutSmart	10 U	37°C	-
NcoI	R0193	simple	NEBbuffer 3.1	10 U	37°C	80°C 20 min
XbaI	R0145	simple	CutSmart	20 U	37°C	65°C 20 min
XhoI	R0146	simple	CutSmart	20 U	37°C	65°C 20 min
ClaI / NcoI	-	double	CutSmart	10 U	37°C	80°C 20 min
EcoRI / XbaI	-	séquentiel	-	-	-	-
EcoRI / XhoI	-	séquentiel	-	-	-	-

**Tableau 6** : Liste des enzymes de restriction avec leurs conditions d'utilisation d'après le fournisseur NEB.  
(<https://nebcloner.neb.com/#!/redigest>)

Les enzymes de restriction sont utilisées en simple digestion (une seule enzyme de restriction), en double digestion (deux enzymes de restriction simultanément) ou en digestion séquentielle (une enzyme de restriction, une purification sur gel ou sur colonne puis une seconde digestion).

**A**

Réactifs	Volume / réaction	Concentration finale
Oligo d(T15) (500 µg/mL)	1 µL	25 µg/mL
ARN totaux	variable	2 µg
Mix dNTP (10 mM)	1 µL	0,5 mM
Eau exempte de RNase	Qsp 12 µL	-
Incuber 5 minutes à 65°C puis maintenir dans la glace		
Tampon 5X	4 µL	1X
DTT 0,1 M	1 µL	5 mM
RNase OUT (40 U / µL)	1 µL	2 U
Incuber 2 minutes à 37°C		
Enzyme M-MLV RT	1 µL	200 U
Incuber 50 minutes à 37°C puis 15 minutes à 70 °C		

**B**

Réactifs	Concentration finale
Tampon GoTaq 5X	1 X
dNTP (10 mM)	2 mM
MgCl <sub>2</sub> (25 mM)	2,5 mM
Primer sens	0,1 µM
Primer anti-sens	0,1 µM
GoTaq	0,25 U
ADN complémentaire	2,5 µL
eau milliQ	Qsp 100 µL

**C**

Isoforme	Nom amorces	Orientation	Séquence 5' → 3'	Taille amplicon (pb)
<b>Mm409</b>	<b>058 H Long</b>	Sens	CGGTGGACTTCTTCCAGCTC	744
	<b>061H Long</b>	Anti-sens	GATTGAGGCTGGTGATGCTG	

**Tableaux 7** : Conditions pour la RT-PCR semi-quantitative

**A-** Conditions pour la transcription inverse (RT) des ARN en ADN complémentaire avec le kit M-MLV.

**B-** Conditions pour la PCR avec le mix GoTaq.

**C-** Amorces utilisés pour la RT-PCR afin d'amplifier la séquence Mm409 de *pgbd5*.

## A

Epitope reconnu	Référence	Fournisseur	Espèce hôte	Type anticorps	concentration ou quantité	dilution WB recommandée
PGBD5	abx019151	Abbexa	Souris	monoclonal	1 mg/mL	1/5 000
	<b>HPA065010</b>	<b>Atlas Antibodies</b>	<b>lapin</b>	<b>polyclonal</b>	-	<b>à tester *</b>
	<b>ABIN1112824</b>	<b>Anticorps-enligne.fr</b>	<b>souris</b>	<b>monoclonal</b>	<b>100 µg</b>	<b>1/500 à 1/ 1 000</b>
	ABIN1854961	Anticorps-enligne.fr	souris	monoclonal	100 µg	à tester *
	ABIN4904741	Anticorps-enligne.fr	souris	monoclonal	100 µg	à tester *
	ABIN1854962	Anticorps-enligne.fr	lapin	polyclonal	100 µg	à tester *
	ABIN1578783	Anticorps-enligne.fr	lapin	polyclonal	100 µL	à tester *
	ABIN1112920	Anticorps-enligne.fr	lapin	polyclonal	100 µg	1 / 10 000 à 1 / 20 000
	ABIN5078778	Anticorps-enligne.fr	lapin	polyclonal	-	à tester *
	nbp2-57708	Novus Biologicals	lapin	polyclonal	-	à tester *
	orb13159	biorbyt	lapin	polyclonal	0,95 mg/ mL	1 / 10 000 à 1 / 20 000
MBS355128	MyBioSource	souris	polyclonal	100 µg	1 / 5 000 à 1 / 10 000	
<b>Myc</b>	<b>MA1-980</b>	<b>Sigma</b>	<b>souris</b>	<b>monoclonal</b>	<b>100 µg</b>	<b>1 / 500 à 1 / 2 000</b>
<b>Actine</b>	<b>ab179467</b>	<b>Abcam</b>	<b>lapin</b>	<b>monoclonal</b>	<b>0.644 - 0.698 mg/ml</b>	<b>1 / 5 000 à 1 / 10 000</b>

## B

Référence	Fournisseur	Espèce hôte	Type anticorps	concentration ou quantité	Utilisation	dilution utilisée
IRD 800 anti Rabbit 926-32211	LiCor	lapin	conjugué Fluorochrome	1 mg/mL	WB ; DB	1/10 000
IRD 680 anti-mouse 926-68072	LiCor	souris	conjugué Fluorochrome	1 mg/mL	WB ; DB	1 / 10 000
Cy-2 711-225-152	Jackson Immuno Research	lapin	Conjugué Cyanine Cy <sup>TM</sup> 2	0,5 mg	Immuno	1 / 500
Cy-2 715-225-151	Jackson Immuno Research	souris	Conjugué Cyanine Cy <sup>TM</sup> 2	0,5 mg	Immuno	1 / 500
Cy-3 715-165-151	Jackson Immuno Research	lapin	Conjugué Cyanine Cy <sup>TM</sup> 3	0,5 mg	Immuno	1 / 500

**Tableaux 8** : Liste des anticorps commerciaux utilisés en Western Blot, Dot Blot et immunocytologie.

**A-** Liste des anticorps primaires polyclonaux et monoclonaux commerciaux utilisés en Western Blot, Dot Blot et immunocytologie

À tester \* : lorsque la concentration ou dilution à utiliser pour le Western Blot n'est pas communiqué par le fournisseur, l'anticorps est utilisé arbitrairement au 1 / 5 000<sup>ème</sup>

Les anticorps **en gras** sont également utilisés en Dot Blot et immunocytologie.

Après recherche, les anticorps de même couleur ont la même origine.

**B-** Liste des anticorps secondaires commerciaux utilisés en Western Blot (WB), Dot Blot (DB) et immunocytologie (Immuno)

Les anticorps conjugués avec Cy-2 fluorescent en vert et ceux avec Cy-3 en rouge.

## A

**Complete Inhibiteur de protéase 100X**  
1 tablette Complete (Roche # 5892791001)  
500 uL eau Ultra Pure

**PBS-C**  
PBS 1 X  
Complete Inhibiteur de protéase 1X

**Tampon de lyse SDS**  
NaCl 0,1M  
EDTA pH8 1 mM  
EGTA pH8 0,5 mM  
HEPES pH8 50 mM  
Formaldéhyde 10 %  
Eau ultra pure

## B

**ChIP Dilution Buffer (CDB)**  
1.1% Triton X-100  
1.2 mM EDTA  
16.7mM Tris-HCl pH  
167mM NaCl  
Eau ultra pure

**Tampon de lavage à faible concentration en sel**  
150mM NaCl  
0.1% SDS  
1% Triton X-100  
2mM EDTA  
20mM Tris-HCl pH 8  
Eau ultra pure

**Tampon de lavage à forte concentration en sel**  
500mM NaCl  
0.1% SDS  
1% Triton X-100  
2mM EDTA  
20mM Tris-HCl pH 8  
Eau ultra pure

**Tampon de lavage LiCl**  
0.25M LiCl  
1% NP-40  
1% sodium deoxycholate  
1mM EDTA  
10mM Tris-HCl pH 8  
H<sub>2</sub>O

**Tampon TE**  
10mM Tris-HCl pH 8  
1mM EDTA  
H<sub>2</sub>O

### **Tableaux 9 :** Composition des solutions pour le ChIP

**A-** Solutions pour la fragmentation de la chromatine

**B-** Solutions pour l'immunoprécipitation de la chromatine

**A**

Amplification	Amorce	Séquence 5' → 3'
Néomycine	(B)-ITR-UTR 5' PB pSV40I	GACTTTCACACCCTAACTGACAC
	(B)-ITR-UTR 3' PB pSV40I	TTCTTCGCCACCCCAACTTGTTTATTGC

**B**

Etapes	Température	Durée	Nombre de cycles
Dénaturation initiale	94 °C	2 min	1 cycle
Dénaturation	94 °C	45 sec	X 50 cycles
Hybridation	60 °C	45 sec	
Extension	72 °C	1 min	
Extension fin	72 °C	5 min	1 cycle
Ajoute 2,5 U de Taq puis x 50 cycles supplémentaires			

**C**

**Tampon LiCl BW 3M**

LiCl 3M  
Tris 10 mM  
EDTA 1 mM  
pH 7,2

**Tampon LiCl BW 6M**

LiCl 6 M  
Tris 10 mM  
EDTA 1 mM  
pH 7,2

**Tableaux 10 :** Conditions pour les premières étapes de la LAM-PCR

**A-** Séquences des amorces biotinylées (B) s'hybridant sur le gène de la Néomycine

**B-** Programme PCR pour la PCR linéaire

**C-** Composition des milieux pour la capture magnétique

**A**

	Digestion 1	Digestion 2
Tampon enzyme 10x	NEB DpnII 1µL	Boeh H
BSA 1µg/µL		1 µl
Enzyme 2U	DpnII (10U/µl) 0,2 µl	PciI (10U/µL) 0,2 µL NcoI (10U/µL) 0,2 µL BspHI (10U/µL) 0,2 µL
QSP eau distillée 10 µL	8,8 µl	8,4 µl

**B**

Amorce	Séquence 5' → 3'
<b>Oligo I</b>	GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG
<b>Oligo II DpnII</b>	GATCCCTAACTGCTGTGCCACTGAATTCAGATC
<b>Oligo II PNB</b>	CATGCCTAACTGCTGTGCCACTGAATTCAGATC

**Tableaux 11** : Conditions pour la suite de la LAM-PCR

**A**- Conditions de digestions enzymatiques des fragments générés à la PCR linéaire

**B**- Séquences des amorces pour la cassette linker

PCR	Amorce	Séquence 5' → 3'
PCR I	LC1	GACCCGGGAGATCTGAATTC
	(B)-ITR-UTR 3'MER75'II	CAITGGCAAAGTTATCTCGGGG
	(B)-ITR-UTR 5'MER75_MER75B II	GATTTGCTTCAGCCTCAAAGAG
	(B)-ITR-UTR 3'MER7B II	CTTGTCTCTAATCCTAATGTAAC
	(B)-Neo 3'MER85 II	CTCATCAATGTATCTTATCATGTC
	(B)-ITR-UTR 5'MER85 II	CACAGAATTCTTGAGCAGTAGG
	(B)-ITR-UTR3'Looper II	CCTTACAAATAGTTAGAACTGCTC
	(B)-ITR-UTR 5'Looper II	GGAATAAAAATTTGGGTTTGCTATG
	(B)-ITR-UTR3'tcr-pble II	GCAGCACTTATGAATCATCTATC
PCR II	LC2	GATCTGAATTCAGTGGCACAG
	ITR-UTR 3' MER75 III	CGCCGGCGAGTATTCTCGG
	ITR-UTR 5' MER75_MER75B III	CGCTGGCAGCGAGCTGCAC
	ITR-UTR 3'MER75B III	GTTACATTAGGATTAGAGACAAGAG
	ITR-UTR 3' MER85 III	CAAGGTCTGATTTTTCACTCATGC
	ITR-UTR 5'MER85 III	CATGCTTAGCGTTCCAATAATGG
	ITR-UTR 3' Looper III	CTAAAATTGGGTCCAATGGACC
	ITR-UTR 5' Looper III	CCAGATGGTAAATGGTGATGAC

**Tableau 12 :** Séquences des amorces pour les PCR I et PCR II

Pour la PCR I, l'amorce LC1 est utilisée en complément d'une autre amorce spécifique du pble à amplifier.

Pour la PCR II, l'amorce LC2 est utilisée en complément d'une autre amorce spécifique du pble à amplifier.

Conditions	lignées cellulaires	clone	Mesure luminescence	ratio	
Non induites	HeLa	Clone 1	3168,1		
		Clone 2	3321,3		
		Clone 3	4223,4		
	HeLa pSBtet-GP	Clone 1	4953,6		
		Clone 2	6528,6		
		Clone 3	5732,7		
		Clone 4	8636,2		
		Clone 5	9864		
		Clone 6	9475,6		
Induites 1 ug/mL doxycycline pendant 3h	HeLa	Clone 1	11669	3,7	
		Clone 2	12461,8	3,8	
		Clone 3	11342	2,7	
	HeLa pSBtet-GP	Clone 1	602076,2	121,5	
		Clone 2	719216,5	110,2	
		Clone 3	431777,9	75,3	
		Clone 4	543233,1	62,9	
		Clone 5	518813,6	52,6	
		Clone 6	633968,4	66,9	

**Tableau 13: Dosage de la bioluminescence de la luciférase émis par les différents clones HeLa et HeLa pSBtet GP après induction ou non pendant 3h de doxycycline 1 µg/mL.**

Le ratio d'induction correspond à la quantité de luminescence émise par le clone induit en présence de doxycycline sur la quantité de luminescence émise par le clone non induit sans doxycycline.

	Nombre de pics détectés		
	p < 10 <sup>-5</sup>		p < 10 <sup>-2</sup>
	MACS14	PePr	PePr
Anticorps anti-V5	16	235	143822
Sérum SW anti-PGBD5	12	650	160567
<b>pics communs</b>	<b>0</b>	<b>12</b>	<b>210</b>

**Tableau 14 : Résultats du chIP-séquençage sur cellules HEK293 transfectées par le plasmide pSBtet\_Mm523\_V5 après immunoprécipitation par l'anticorps anti\_V5 et le sérum SW.**

Deux outils de détection de pics sont utilisés : **MACS14** ([https://usegalaxy.org/root?tool\\_id=toolshed.g2.bx.psu.edu/repos/devteam/mac3/peakcalling\\_mac3/1.0.1/](https://usegalaxy.org/root?tool_id=toolshed.g2.bx.psu.edu/repos/devteam/mac3/peakcalling_mac3/1.0.1/)) et **PePr** (<https://github.com/usnistgov/pepr>)

	Nombre de pics détectés	
	détectés avec PePr	Co-localisant avec Looper
<b>HD</b>	851	0 / 851
<b>HV</b>	873	0 / 873
<b>cellules Leydig</b>	432	0 / 432

**Tableau 15 : Résultats du ChIP-séquençage sur tissus sains de souris**

Les échantillons fixés et fragmentés d'hippocampe dorsal (**HD**), d'hippocampe ventral (**HV**) et **cellules de Leydig** ont été immunoprécipités avec le sérum anti-PGBD5 puis séquencés par haut débit. Les séquences fournies par le séquençage ont été analysées grâce à l'outil de détection de pics **PePr** (<https://github.com/usnistgov/pepr>) avec un seuil de significativité inférieur à 0,05.

Lignées cellulaires	Nombre de pics détectés
HeLa	0
HEK	0
T98G	283
H4	0
8MG	0
HCT116	1306

**Tableau 16 : Résultats du ChIP-séquençage à partir de différentes lignées cancéreuses.**

Les cellules fixées et fragmentées ont été immunoprécipitées avec le sérum anti-PGBD5 puis séquencées par haut débit. Les séquences fournies par le séquençage ont été analysées grâce à l'outil de détection de pics **PePr** (<https://github.com/usnistgov/pepr>) avec un seuil de significativité inférieur à 0,05 et une q-value inférieure à 0,05.

<i>pble</i> (transposon)	nombre jonction transposon / chromosome
<i>ifp2</i>	1479
<i>MER75</i>	389
<i>MER75B</i>	307
<i>MER85</i>	1313
<i>Looper</i>	193

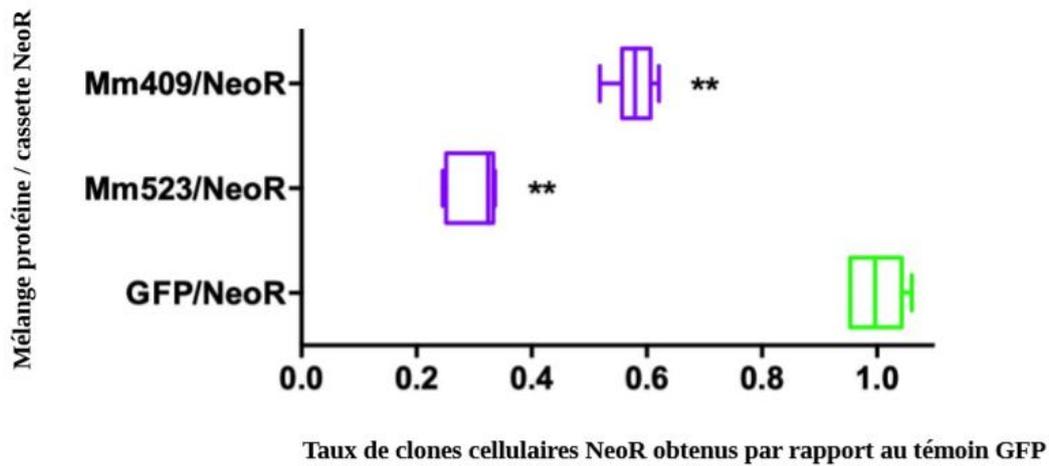
**Tableau 17 : Tableau récapitulatif du nombre d'événements de transposition (jonction transposon / chromosome) mis en évidence pour chaque *pble* en présence de la transposase PGBD5 Mm523.**

Source Transposase	Source transposon	séquences TIR et TSD conservées	Séquences TSD conservées / TIR incomplètes	Séquences TIR conservées / TSD incomplètes	Séquences TSD et TIR incomplètes
Mm523 PGBD5	<i>Ifp2</i> -NeoR	9,9 % (147)	5,4 % (80)	4,3 % (64)	80,4 % (1188)
Mm523 PGBD5	MER75-NeoR	11,0 % (43)	1,5 % (6)	3,9 % (15)	83,6 % (325)
Mm523 PGBD5	MER75B-NeoR	2,3 % (7)	3,9 % (12)	1,9 % (6)	91,9 % (282)
Mm523 PGBD5	MER85-NeoR	2,6 % (34)	2,7 % (36)	0,3 % (4)	94,4 % (1239)
Mm523 PGBD5	Looper-NeoR	3,6 % (7)	3,1 % (6)	1,6 % (3)	91,7 % (177)

**Tableau 18 :** Tableau récapitulatif de la conservation des séquences TIR (Terminal Inverted Repeat) et TSD (Target Site Duplication) des transposons *pbls* après événement de transposition Mm523 PGBD5.

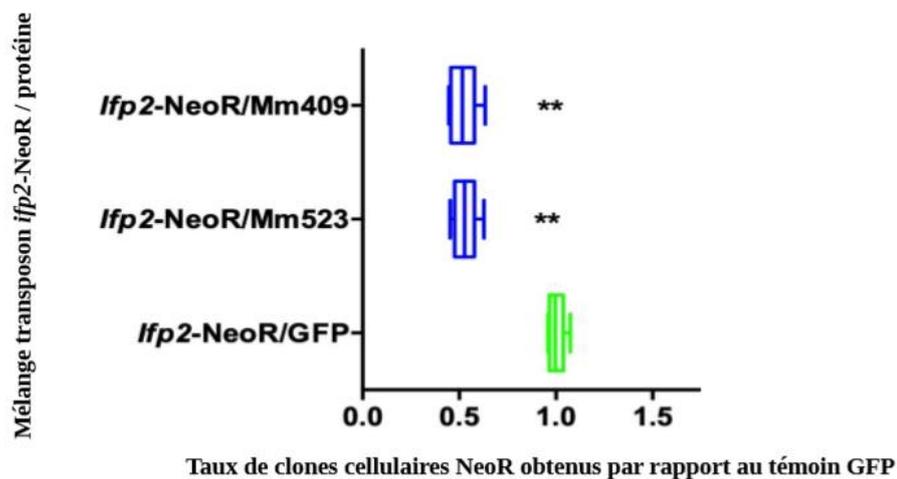
Echantillon	Nombre de pics détectés		p. value (test permutation)
	détectés avec PePr	Co-localisation avec des IR	
<b>HD</b>	851	845 / 851	$3,29.10^{-8}$
<b>HV</b>	873	864 / 873	$9,544.10^{-7}$
<b>cellules Leydig</b>	432	420 / 432	0,032688
<b>T98G</b>	283	239 / 283	0,468
<b>HCT116</b>	1306	1279 / 1306	$< 10^{-100}$
<b>G401</b>	618	594 / 618	$5,008677.e^{-7}$

**Tableau 19 :** Tableau récapitulatif de ré-analyse des données de CHIP-séquençage pour PGBD5 obtenus à partir de tissus sains (HD, HV et cellules Leydig) et à partir de lignées de cellules cancéreuses (T98G, HCT116 et G401).



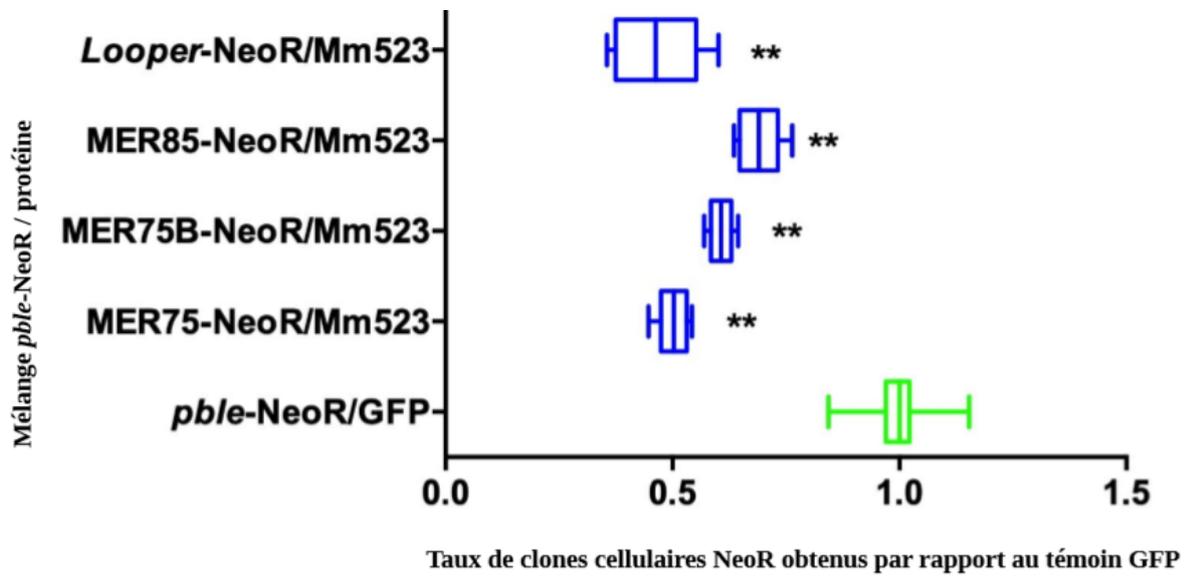
**Graphique 1 :** Résultats du nombre de clones cellulaires obtenus lors du test d'intégration de la cassette de résistance NeoR en présence des transposases (Mm409 et Mm523) ou sans transposase (GFP).

\*\* : différence significative ( $p < 0.05$ )



**Graphique 2 :** Résultats du nombre de clones cellulaires obtenus lors du test de transposition du transposon *ifp2-NeoR* en présence des transposases (Mm409 et Mm523) ou sans transposase (GFP).

\*\* : différence significative ( $p < 0.05$ )



**Graphique 3 :** Résultats du nombre de clones cellulaires obtenus lors du test de transposition des *pbles* en présence des transposases (Mm409 et Mm523) ou sans transposase (GFP).

\*\* : différence significative ( $p < 0.05$ )

# ÉCOLE PRATIQUE DES HAUTES ÉTUDES

Science de la Vie et de la Terre

## Expression de la protéine PGBD5

Où ? Quand ? Comment ?

Mémoire présenté par : **BEAUCLAIR Linda**

Soutenu le jeudi 17 décembre 2020

### RÉSUMÉ

Les éléments transposables (ET) sont une source importante de mutations dans tous les génomes eucaryotes. Au sein de chaque espèce, les ET co-évoluent avec leur environnement et le génome de leurs hôtes pouvant conduire à la récupération par l'hôte d'éléments fonctionnels appartenant à l'ET. Ce processus de domestication permet ainsi parfois la création de nouveaux gènes pour l'hôte. Cinq des 54 néogènes connus chez l'Homme codent des transposases toujours actives : SETMAR, RAG1-RAG2, THAP9 et PGBD5. La protéine PGBD5 est une transposase très conservée chez les vertébrés, exprimée de façon ubiquitaire dans le Système Nerveux Central (SNC) et plus fortement dans les neurones dits pyramidaux et granulaires de l'hippocampe, l'amygdale, du bulbe olfactif et du cervelet chez un individu sain. Elle est capable *in vitro* de catalyser des réarrangements génomiques locaux dans les neurones ce qui pourraient participer à la mémorisation. La protéine PGBD5 est également sur-exprimée dans la plupart des tumeurs solides chez les enfants et est retrouvée dans les liquides corporels tels que le sérum plasmatique et le liquide cérébro-spinal avec un taux très supérieur en cas de cancers. Cependant, l'absence d'anticorps anti-PGBD5 validés empêche pour le moment l'utilisation de la protéine PGBD5 en tant que bio-marqueur de diagnostic pour les maladies neurodégénératives et les cancers.

Dans le cadre de ce travail, nous avons essayé : i) de déterminer les sites d'expression de la protéine en caractérisant un sérum anti-PGBD5 capable de détecter la protéine en Western Blot et immunocytochimie, ii) de localiser les sites de liaison de la protéine PGBD5 vis-à-vis des *PiggyBac*-like elements (*pbles*) sur l'ADN par une approche de ChIP-séquençage et iii) de vérifier la capacité de transposition de la protéine PGBD5 à l'aide d'un test de transposition et de la méthode de LAM-PCR.

Les résultats présentés dans ce manuscrit ont pour but d'éclairer sur le ou les rôles physiologiques potentiels de la protéine PGBD5 et de décrypter la fonction de la protéine PGBD5 dans un noyau des cellules eucaryotes.

**Mots-clés** : Élément transposable (ET), transposase, PGBD5, *pble*, caractérisation anticorps, ChIP-séquençage, test transposition, LAM-PCR.