

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la
Propriété Intellectuelle
Bureau international



(43) Date de la publication internationale
4 février 2016 (04.02.2016)

WIPO | PCT

(10) Numéro de publication internationale
WO 2016/016358 A1

(51) Classification internationale des brevets :
C12N 9/22 (2006.01) C12N 15/90 (2006.01)

(21) Numéro de la demande internationale :
PCT/EP2015/067472

(22) Date de dépôt international :
30 juillet 2015 (30.07.2015)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
1457396 30 juillet 2014 (30.07.2014) FR

(71) Déposants : CENTRE NATIONAL DE LA RE-
CHERCHE SCIENTIFIQUE (CNRS) [FR/FR]; 3, rue
Michel Ange, F-75016 Paris (FR). INSTITUT NATION-
AL DE LA RECHERCHE AGRONOMIQUE
[FR/FR]; 147, rue de l'Université, F-75007 Paris (FR).

(72) Inventeurs : BIGOT, Yves; 9, Allée des Cytises, F-37550
Saint Avertin (FR). CARPENTIER, Guillaume; 9 im-
passe Marconi, F-49300 Cholet (FR). CASTERET, So-
phie; 11, rue de la Buffeterie, F-37190 Vallères (FR). PIE-
GU, Benoît; La Touche, F-37360 Rouziers de Touraine
(FR).

(74) Mandataire : REGIMBEAU; 20, rue de Chazelles, F-
75847 Paris Cedex 17 (FR).

(81) États désignés (sauf indication contraire, pour tout titre
de protection nationale disponible) : AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) États désignés (sauf indication contraire, pour tout titre
de protection régionale disponible) : ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, RU,
TJ, TM), européen (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

Publiée :

- avec rapport de recherche internationale (Art. 21(3))
- avec la partie de la description réservée au listage des sé-
quences (règle 5.2.a))

(54) Title : TARGETING OF NON-VIRAL INTEGRATIVE VECTORS IN THE NUCLEOLAR DNA SEQUENCES OF EUKA-
RYOTES

(54) Titre : CIBLAGE DE VECTEURS INTEGRATIFS NON-VIRAUX DANS LES SEQUENCES D'ADN NUCLEOLAIRES
CHEZ LES EUCARYOTES

(57) Abstract : The invention relates to a system for integrating a nucleic acid molecule of interest in a chromosome region associa-
ted with the nucleolus. Said system includes at least: - one non-viral integrative vector located in the nucleolus of the cell nucleus
and lacking a functional recombinase source; and - one functional recombinase source that is located in the nucleolus of the cell nu-
cleus and provided via transfer from said non-viral integrative vector.

(57) Abrégé : L'invention vise un système pour l'intégration d'une molécule d'acide nucléique d'intérêt dans une région chromoso-
mique associée au nucléole, comprenant au moins: - Un vecteur intégratif non viral se localisant dans le nucléole du noyau cellu-
laire, et dépourvu de source de recombinaison fonctionnelle; et - Une source de recombinaison fonctionnelle se localisant dans le nu-
cléole du noyau cellulaire, fournie en trans dudit vecteur intégratif non viral.



WO 2016/016358 A1

CIBLAGE DE VECTEURS INTEGRATIFS NON-VIRAUX DANS LES SEQUENCES D'ADN NUCLEOLAIRES CHEZ LES EUCARYOTES

De nombreuses techniques de transfert de séquences nucléotidiques exogènes ont
5 été proposées dans le domaine de la thérapie génique, ou pour fabriquer des cellules
et organismes génétiquement modifiés à des fins de production de protéines
d'intérêt économique ou thérapeutique, ou encore pour la recherche fondamentale.

Ces techniques font notamment appel à systèmes comprenant des vecteurs
10 intégratifs non viraux couplés à des enzymes spécifiques. Les vecteurs intégratifs non
viraux sont des vecteurs comprenant des séquences susceptibles d'être intégrées
dans le génome. Cette intégration requiert généralement l'intervention d'enzymes
capables de catalyser ou de stimuler l'intégration de la séquence d'intérêt dans
l'ADN. Il existe plusieurs types de systèmes vecteurs non viraux/enzymes
15 d'intégration, qui fonctionnent selon des mécanismes variés. A titre d'exemple, on
peut ainsi citer les éléments transposables (TE), dont les transposons et rétro-
transposons, qui permettent l'intégration de séquences d'intérêt lorsqu'ils sont
couplés à des recombinases telles que des transposases ou des intégrases. Parmi les
systèmes vecteurs non viraux/enzymes d'intégration, on trouve aussi les systèmes
20 plasmide/nucléase dans lesquels l'enzyme stimule l'intégration de la séquence
d'intérêt par recombinaison homologue.

De manière générale, les éléments transposables (TE) ou éléments génétiques
mobiles (EGM) sont des fragments d'ADN de petite taille, capables de se déplacer
25 d'un site chromosomique à un autre. Il existe diverses classes de TE, selon les
mécanismes d'intégration dans l'ADN mis en jeu par ces derniers. Dans le cas de
transposons, qui sont des TE dits de classe II, ces fragments d'ADN sont caractérisés
par des séquences répétées inversées (ITR) situées en positions 5' et 3' terminales.
Une enzyme codée par les TE eux-mêmes, la transposase, catalyse le processus de
30 transposition de ces derniers.

Des transposons ont été identifiés tant chez les procaryotes que chez les eucaryotes.
Chez les procaryotes, un grand nombre de transposons ont été répertoriés à ce jour.
On peut citer, par exemple, des séquences d'insertion telles que IS1, et des
35 transposons, tel Tn5. Chez les eucaryotes, les éléments de classe II comprennent dix

familles : P, PiggyBac (incluant le transposon Pokey), hAT, hélitron, Harbinger, En/Spm, Mutator, Transib, Pogo et IS630-Tc1-mariner (incluant Sleeping Beauty).

Les systèmes plasmides/nucléases sont aussi régulièrement utilisés pour transférer
5 des gènes d'intérêt dans le génome de cellules eucaryotes de manière ciblée. Ces
systèmes sont, comme les systèmes transposon/transposase, des systèmes bipartites
qui comprennent en général un plasmide dont la séquence est définie de façon à
favoriser la transposition d'un transgène par recombinaison homologue, et une
nucléase de type Méganuclease, Zinc finger Nucléase, TALEN, ou CRISPR/Cas9 pour
10 cibler spécifiquement l'intégration à un site chromosomique à une fréquence élevée.

Le principe du ciblage de ces derniers vecteurs repose donc sur la capacité de clivage
de l'ADN à des sites spécifiques par les nucléases en question. Ce clivage stimule la
recombinaison homologue à proximité de ces sites, permettant ainsi l'intégration de
15 séquences exogènes. En outre, les bases moléculaires régissant le lien entre la
structure de ces nucléases et les sites qu'elles reconnaissent ont été bien étudiées. Il
est donc désormais possible avec ces systèmes de produire des nucléases à façon
ciblant spécifiquement les séquences d'ADN dans lesquelles on souhaite effectuer
l'intégration.

20

Toutefois, le transfert d'un ADN exogène à l'aide de ces moyens connus, qu'il
s'agisse des systèmes dérivés des transposons ou des systèmes reposant sur des
nucléases, n'est pas sans poser des difficultés, notamment du fait d'une efficacité,
en particulier vis à vis de la taille du fragment à intégrer, et d'une spécificité
25 d'intégration du transgène qui restent malgré tout très limitées.

En effet, ces systèmes n'excluent pas les insertions aléatoires. Or, celles-ci sont
susceptibles d'avoir des effets délétères en mutant des gènes pour lesquels il
n'existe qu'une copie dans le génome. En outre, il a été montré que les insertions
30 aléatoires sont éteintes à 95% dans les cellules HeLa (Garisson et al., 2007). Ainsi, il
semblerait que certains transgènes, selon le site dans lequel ils ont été insérés, sont
finalement reconnus par la cellule comme appartenant au non-soi et éteints. Ces
résultats soulignent l'importance du site d'insertion en tant que tel pour la stabilité
ultérieure de l'expression du transgène.

35

Finalement, les propriétés du couple ADN intégré/site d'intégration sont essentielles pour obtenir une expression durable et contrôlable d'un transgène.

Or, les recherches ont montré que les séquences codant pour les ARN ribosomiaux
5 étaient des sites privilégiés permettant une expression optimale des transgènes chez
les eucaryotes. De plus, ces séquences présentent l'avantage d'être regroupées
spatialement au niveau du nucléole, dans des régions génomiques spécifiques
comprenant, entre autre, des clusters de séquences répétées (de gènes codant pour
les ARNr 5S, 5.8S, 18S et 28S). Ainsi, l'insertion de molécules d'acide nucléique dans
10 ces régions est moins susceptible de donner lieu à des mutations délétères. Un tel
site d'intégration a d'ailleurs déjà été évalué dans le cadre de techniques sans aucun
lien avec la mise en œuvre d'un outil transposable. Il a ainsi été démontré que
l'inactivation, par intégration d'ADN, de 50% des gènes codant les ARNr 18S-5.8S-28S
n'affecte pas la viabilité de la cellule ou de l'organisme. En effet, bien que leur
15 nombre puisse varier d'un facteur 1 à 2 dans le génome d'une même espèce, ces
séquences sont fortement répétées car elles sont essentielles à la synthèse protéique
et, donc, à la viabilité de l'organisme. Utiliser les répétitions en tandem contenant
les gènes des ARNr 18S-5.8S-28S comme cibles pour l'intégration permet en outre de
limiter, voire supprimer, la diffusion indésirable des transgènes à des cultures non
20 visées et des populations naturelles végétales et animales.

Malgré cet intérêt pour l'intégration de transgènes au niveau du nucléole, et en
particulier dans les régions comprenant des gènes codant les ARNr 18S-5.8S-28S,
c'est-à-dire pour les régions chromosomiques associées au nucléole, il n'existe pas à
25 l'heure actuelle de système performant garantissant une bonne efficacité
d'intégration à ces sites, en particulier dans les génomes animaux et de plantes.

L'expression de transgènes intégrés à ces sites a été confirmée en cellules de
mammifère dans un contexte de thérapie génique en utilisant la recombinaison
30 homologue (Wen et al., 2008). D'après les travaux rapportés, par recombinaison
homologue, mécanisme d'intégration passif et aléatoire, l'efficacité d'intégration
atteint 10^{-4} - 10^{-5} , c'est-à-dire qu' 1 cellule sur 10 000 ou 1 cellule sur 100 000
possède le transgène correctement intégré dans le site chromosomique cible (Wen et
al., 2008 ; Liang et al., 2007).

35

Il est donc clair que les niveaux d'efficacité d'intégration susceptibles d'être obtenus par les différentes méthodes décrites jusque-là restent encore trop faibles pour les applications visées, et qu'il existe bien un besoin pour des moyens et méthodes permettant l'intégration de molécules d'acides nucléiques spécifiquement dans ces sites d'intérêt, à savoir dans les régions chromosomiques associées au nucléole.

Les moyens et méthodes de l'invention permettent de résoudre ce problème technique d'une manière simple et élégante, facile à mettre en œuvre et avec des taux d'intégration dans les sites visés particulièrement intéressants.

LEGENDE DES FIGURES

Figure 1 : Analyse de la qualité des complexes plasmide PNA par retard sur gel d'agarose. La totalité des digestions ont été déposées sur gel d'agarose 1%, 1XTAE, 1X GelRed. Puits 1 : marqueur de taille λ DNA digéré par HindIII; Puits 2 : pBS-PB-PNABS-pPol1h-NeoR-pASV40 digéré par BamHI/SpeI; Puits 3 : complexe [pBS-PB-PNABS-pPol1h-NeoR-pASV40/PNA-NoLS] hybridé en excès de X60 en oligonucléotide puis digéré par BamHI/SpeI.

Figure 2 : Analyse en microscopie à épifluorescence de cellules HeLa transfectées depuis 24 heures avec des complexes [plasmide/PNA-NoLS] colorés au YOYO-1. A : observation en épifluorescence des complexes ADN-PEI; B: localisation en lumière blanche des nucléoles dans les noyaux; C: la fusion des deux images (épifluorescence et lumière blanche) permet de visualiser la concentration de fluorescence dans les nucléoles.

Figure 3 : Analyse en microscopie à épifluorescence de cellules HeLa transfectées depuis 24 heures avec PB-GFP. A : les nucléoles apparaissent comme des zones sombres dans le noyau. B : témoins de localisation nucléolaire fusionné au DsRED. C : La fusion des deux images révèle que la transposase PB est activement internalisée dans les noyaux mais est exclue des nucléoles.

Figure 4 : Analyse en microscopie à épifluorescence de cellules HeLa transfectées depuis 24 heures avec NoLS PB-GFP. A : les nucléoles apparaissent comme des zones

fluorescente dans le noyau à cause de la présence de NoLS PB-GFP. B: témoins de localisation nucléolaire fuisonné au DsREd. C : La fusion des deux images révèle que la transposase NoLS-PB est activement internalisée dans les nucléoles.

- 5 Figure 5 : Résultats de tests de transposition réalisés sans transposase (témoin) ou avec (PB et NoLS-PB) et deux vecteurs contenant respectivement deux cassettes de résistance à la néomycine dépendantes de deux enzymes de transcription différentes : pSV40 = ARN polymérase II et pPolh = ARN polymérase I. En ordonnée est indiqué le nombre de colonies NéoR stables obtenues après 2 semaines de sélection. Ce nombre
10 est proportionnel à l'efficacité de la transposition. Les témoins ont été obtenus avec un plasmide pCS2 vide comme source négative de transposase.

- Figure 6 : Pourcentage d'évènements d'insertion différents du vecteur localisés dans l'ADN chromosomique du compartiment nucléaire nucléoplasmique. Deux qualités
15 nucléiques sont utilisées pour chaque vecteur comme source de transposase et de transposon : ADNp/ADNp et ARNm/ADNp respectivement. Les combinaisons de plasmides et d'ARN sont indiquées en ordonnées. Les barres grises indiquent le pourcentage de loci affectés par une insertion du vecteur dans ce compartiment. Les barres noires indiquent le pourcentage de séquences décrivant une insertion du
20 vecteur dans ce compartiment. Les deux graphes du haut (A) décrivent des résultats obtenus par IPCR alors que ceux du bas (B) ont été obtenus par LAM-PCR. Dans chaque graphe, les deux barres grise et noire indiquent les pourcentages théoriques attendus dans le compartiment nucléaire considéré.

- 25 Figure 7 : Pourcentage d'évènements d'insertion différents du vecteur localisés dans l'ADN chromosomique du compartiment nucléaire LAD. Deux qualités nucléiques sont utilisées pour chaque vecteur comme source de transposase et de transposon : ADNp/ADNp et ARNm/ADNp respectivement. Les combinaisons de plasmides et d'ARN
30 sont indiquées en ordonnées. Les barres grises indiquent le pourcentage de loci affectés par une insertion du vecteur dans ce compartiment. Les barres noires indiquent le pourcentage de séquences décrivant une insertion du vecteur dans ce compartiment. Les deux graphes du haut (A) décrivent des résultats obtenus par IPCR alors que ceux du bas (B) ont été obtenus par LAM-PCR. Dans chaque graphe, les deux barres grise et noire indiquent les pourcentages théoriques attendus dans le
35 compartiment nucléaire considéré.

Figure 8 : Pourcentage d'évènements d'insertion différents du vecteur localisés dans l'ADN chromosomique du compartiment nucléaire nucléoplasmique. Deux qualités nucléiques sont utilisées pour chaque vecteur comme source de transposase et de transposon : ADNp/ADNp et ARNm/ADNp respectivement. Les combinaisons de plasmides et d'ARN sont indiquées en ordonnées. Les barres grises indiquent le pourcentage de loci affectés par une insertion du vecteur dans ce compartiment. Les barres noires indiquent le pourcentage de séquences décrivant une insertion du vecteur dans ce compartiment. Les deux graphes du haut (A) décrivent des résultats obtenus par IPCR alors que ceux du bas (B) ont été obtenus par LAM-PCR. Dans chaque graphe, les deux barres grise et noire indiquent les pourcentages théoriques attendus dans le compartiment nucléaire considéré.

Figure 9 : Pourcentage d'évènements d'insertion différents du vecteur localisés dans l'ADN chromosomique du compartiment nucléaire LAD. Deux qualités nucléiques sont utilisées pour chaque vecteur comme source de transposase et de transposon : ADNp/ADNp et ARNm/ADNp respectivement. Les combinaisons de plasmides et d'ARN sont indiquées en ordonnées. Les barres grises indiquent le pourcentage de loci affectés par une insertion du vecteur dans ce compartiment. Les barres noires indiquent le pourcentage de séquences décrivant une insertion du vecteur dans ce compartiment. Les deux graphes du haut (A) décrivent des résultats obtenus par IPCR alors que ceux du bas (B) ont été obtenus par LAM-PCR. Dans chaque graphe, les deux barres grise et noire indiquent les pourcentages théoriques attendus dans le compartiment nucléaire considéré.

Figure 10 : Pourcentage d'évènements d'insertion différents du vecteur localisés dans l'ADN chromosomique du compartiment nucléolaire. Deux qualités nucléiques sont utilisées pour chaque vecteur comme source de transposase et de transposon : ADNp/ADNp et ARNm/ADNp respectivement. Les combinaisons de plasmides et d'ARN sont indiquées en ordonnées. Les barres grises indiquent le pourcentage de loci affectés par une insertion du vecteur dans ce compartiment. Les barres noires indiquent le pourcentage de séquences décrivant une insertion du vecteur dans ce compartiment. Les deux graphes du haut (A) décrivent des résultats obtenus par IPCR alors que ceux du bas (B) ont été obtenus par LAM-PCR. Dans chaque graphe, les deux barres grise et noire indiquent les pourcentages théoriques attendus dans le compartiment nucléaire considéré.

Figure 11 : Pourcentage d'évènements d'insertion différents du vecteur localisés dans les régions NAD contenus dans le compartiment nucléolaire. Deux qualités nucléiques sont utilisées pour chaque vecteur comme source de transposase et de transposon : ADNp/ADNp et ARNm/ADNp respectivement. Les combinaisons de plasmides et d'ARN sont indiquées en ordonnées. Les barres grises indiquent le pourcentage de loci affectés par une insertion du vecteur dans ce compartiment. Les barres noires indiquent le pourcentage de séquences décrivant une insertion du vecteur dans ce compartiment. Les deux graphes du haut (A) décrivent des résultats obtenus par IPCR alors que ceux du bas (B) ont été obtenus par LAM-PCR. Dans chaque graphe, les deux barres grise et noire indiquent les pourcentages théoriques attendus dans le compartiment nucléaire considéré.

Figure 12 : Pourcentage d'évènements d'insertion différents du vecteur localisés dans les régions flanquantes des rDNA limitrophes dans le compartiment nucléolaire. Deux qualités nucléiques sont utilisées pour chaque vecteur comme source de transposase et de transposon : ADNp/ADNp et ARNm/ADNp respectivement. Les combinaisons de plasmides et d'ARN sont indiquées en ordonnées. Les barres grises indiquent le pourcentage de loci affectés par une insertion du vecteur dans ce compartiment. Les barres noires indiquent le pourcentage de séquences décrivant une insertion du vecteur dans ce compartiment. Les deux graphes du haut (A) décrivent des résultats obtenus par IPCR alors que ceux du bas (B) ont été obtenus par LAM-PCR. Dans chaque graphe, les deux barres grise et noire indiquent les pourcentages théoriques attendus dans le compartiment nucléaire considéré.

Figure 13 : Pourcentage d'évènements d'insertion différents du vecteur localisés dans les rDNA contenus dans le compartiment nucléolaire. Deux qualités nucléiques sont utilisées pour chaque vecteur comme source de transposase et de transposon : ADNp/ADNp et ARNm/ADNp respectivement. Les combinaisons de plasmides et d'ARN sont indiquées en ordonnées. Les barres grises indiquent le pourcentage de loci affectés par une insertion du vecteur dans ce compartiment. Les barres noires indiquent le pourcentage de séquences décrivant une insertion du vecteur dans ce compartiment. Les deux graphes du haut (A) décrivent des résultats obtenus par IPCR alors que ceux du bas (B) ont été obtenus par LAM-PCR. Dans chaque graphe, les deux barres grise et noire indiquent les pourcentages théoriques attendus dans le compartiment nucléaire considéré.

Figure 14 : Efficacité de la l'IPCR à détecter des vecteurs insérés dans les 3 principaux compartiments nucléaires, nucléoplasme, LAD, et nucléoles. En ordonnée sont représentés les résultats obtenus avec 8 combinaisons moléculaires de vecteur.

- 5 Figure 15 : Efficacité de LAM-PCR à détecter des vecteurs insérés dans les 3 principaux compartiments nucléaires, nucléoplasme, LAD, et nucléoles. En ordonnée sont représentés les résultats obtenus avec 8 combinaisons moléculaires de vecteur.

10

DESCRIPTION

Les inventeurs ont mis au point un système moléculaire bi-partite permettant l'intégration de molécules d'acides nucléiques d'intérêt dans les régions du génome formant le nucléole. Le système moléculaire des inventeurs permet une intégration
15 spécifique de site dans les cellules hôtes eucaryotes, plus particulièrement dans les régions chromosomiques associées au nucléole. Le nucléole étant le site de transcription le plus actif du noyau, les séquences intégrées grâce au système moléculaire de l'invention sont exprimées de manière stable dans le temps.

- 20 Le système moléculaire de l'invention comprend d'une part un vecteur intégratif non viral, dépourvu d'activité enzymatique permettant son intégration dans le génome, et, d'autre part, une enzyme permettant cette intégration.

Cette enzyme peut ainsi être par exemple, et selon le type de vecteur intégratif
25 choisi, une recombinase, une transposase, une intégrase, une nucléase.

En outre, le système moléculaire de l'invention est particulièrement original en ce que ces deux éléments, à la fois le vecteur intégratif non viral et la source de recombinase fonctionnelle, possèdent tous deux la capacité de se localiser à
30 proximité des sites ciblés. En effet, les inventeurs ont élaboré des outils moléculaires, en particulier des vecteurs intégratifs non viraux et des sources de recombinase fonctionnelle (c'est à dire des recombinases ainsi que leurs équivalents fonctionnels), se localisant dans le noyau cellulaire, et même au nucléole. Cette localisation spécifique des moyens du système moléculaire de l'invention permet

d'augmenter de façon très importante le taux d'intégration dans les régions chromosomiques associées au nucléole.

Pour obtenir ce résultat, les inventeurs ont conçu en particulier des vecteurs
5 intégratifs non viraux spécifiques, dont la capacité de localisation au noyau et plus particulièrement au nucléole repose sur l'utilisation de signaux de localisation peptidique NoLS. Cette solution technique est particulièrement surprenante. En effet, les signaux de localisation NoLS sont des peptides classiquement utilisés pour permettre la localisation de peptides et de protéines. Il est donc étonnant de
10 recourir à ce type d'outils pour localiser des molécules telles que des vecteurs, qui sont des polynucléotides. Le recours à une telle solution technique peut sembler contre-intuitif. En effet, sa mise en œuvre nécessite l'élaboration d'éléments de liaisons additionnels, permettant de greffer le vecteur intégratif non viral au signal de localisation NoLS, ce qui peut la rendre *a priori* complexe. Cependant, contre
15 toute attente, les vecteurs intégratifs non viraux de l'invention se révèlent être particulièrement efficaces pour ce qui est d'intégrer des séquences nucléiques d'intérêt dans les régions chromosomiques constituant le nucléole.

Ainsi, l'invention vise en premier lieu un système moléculaire pour l'intégration
20 d'une molécule d'acide nucléique d'intérêt dans une région chromosomique associée au nucléole, comprenant au moins :

- un vecteur intégratif non viral se localisant dans le nucléole du noyau cellulaire, et dépourvu de source de recombinaison fonctionnelle; et
- une source de recombinaison fonctionnelle se localisant dans le nucléole du noyau
25 cellulaire, fournie en *trans* dudit vecteur intégratif non viral.

Par « système moléculaire », on entend au sens de l'invention un ensemble de moyens de biologie moléculaire. Au sens de l'invention, ces moyens sont en particulier adaptés à l'intégration de molécules d'acides nucléiques d'intérêt au sein
30 d'une région chromosomique. Plus particulièrement, au sens de l'invention, les moyens du système moléculaire selon l'invention comprennent au moins (a) un vecteur intégratif non viral et (b) une source d'enzyme recombinante fonctionnelle ou une source d'un équivalent fonctionnel de celle-ci.

De manière générale, on désignera ci-après dans la demande les diverses enzymes visées par l'invention par leur caractère fonctionnel, en les regroupant sous les termes « source de recombinaison fonctionnelle », étant entendu que cette source peut être constituée d'une recombinaison en tant que telle, ou d'un équivalent
5 fonctionnel, tel que par exemple une intégrase, une transposase ou une nucléase.

Par « source de recombinaison fonctionnelle », on entend au sens de l'invention une source d'enzyme capable de remplir la même fonction qu'une recombinaison.

10 Les termes et expressions « activité », « fonction », « activité biologique » et « fonction biologique » sont équivalents et répondent à l'acception usuelle dans le domaine technique de l'invention. En particulier, la recombinaison comprise dans le système moléculaire selon l'invention est considérée « fonctionnelle » si elle est apte à catalyser le processus d'intégration de la ou des séquences nucléotidiques
15 exogènes d'intérêt dans la région chromosomique cible.

L'homme du métier pourra aisément vérifier que la recombinaison d'intérêt est « fonctionnelle » au sens de l'invention, par des tests *in vitro* bien connus dans le domaine. Ce type de tests consiste à réaliser une analyse biochimique des activités
20 de liaison et de coupure aux extrémités de l'ADN par la recombinaison ainsi que l'analyse de la capacité de la recombinaison à transférer des brins d'ADN. Dans le cas d'une nucléase, l'homme du métier pourra aisément vérifier que l'enzyme d'intérêt est fonctionnelle par des tests *in vitro* de spécificité de clivage d'ADN.

25 La source de recombinaison fonctionnelle est de préférence choisie parmi les transposases, les nucléases, les intégrases et les recombinaisons. Notamment, l'homme du métier saura choisir, parmi les différentes enzymes disponibles, celle qui sera le mieux adaptée à l'intégration de la séquence d'intérêt, par exemple en fonction du vecteur intégratif non viral utilisé. Par exemple, lorsque le vecteur intégratif non
30 viral est un transposon ou pseudo-transposon, l'homme du métier pourra choisir de préférence une transposase pour permettre la transposition de la séquence d'intérêt. De préférence, la source de recombinaison fonctionnelle est une transposase.

Avantageusement, la transposase selon l'invention est choisie parmi la famille de
35 transposases DD[E/D], laquelle comprend notamment les transposases issues des

transposons Sleeping Beauty, Frog Prince, MuA, *Mos-1*, Pokey, piggyBac, Tol2, Tn3, Tn5, Tn7, Tn10, et Hermes. Cette famille de transposases a notamment été décrite en détail dans Nesmelova et al. (Adv Drug Deliv Rev.; 30;62(12):1187-95; 2010). Avantageusement, l'intégrase selon l'invention est choisie parmi les intégrases
5 de lentivirus ou rétrovirus. Avantageusement, la recombinaise selon l'invention est choisie parmi les recombinaises de phages (telle que par exemple PhiC31). Avantageusement, la nucléase selon l'invention est choisie parmi les nucléases TALEN (pour « *Transcription Activator-Like (TAL) Effector Nucleases* »), ou les nucléases à doigts de zinc (ZFNs, pour « *Zinc-finger nucleases* »), la Méganuclease, ou les
10 nucléases de type Cas9 (utilisée dans le système CRISPR/Cas9).

Par «source de recombinaise fonctionnelle se localisant dans le nucléole du noyau cellulaire», on entend au sens de l'invention que la source de recombinaise selon l'invention est associée à un système moléculaire permettant sa localisation dans le
15 nucléole du noyau cellulaire. Ainsi, par «se localisant dans le nucléole du noyau cellulaire», on entend que les molécules de « source de recombinaise fonctionnelle », qu'il s'agisse de transposases, nucléases, recombinaises ou intégrases, sont activement localisées dans la cellule eucaryote de telle sorte qu'une très grande majorité d'entre elles, voire la totalité, se situent dans le nucléole du noyau
20 cellulaire. Cet adressage des molécules est facilement suivi, notamment par des techniques de microscopie de fluorescence usuelles.

Le noyau cellulaire est facilement détectable dans les cellules eucaryotes, et peut notamment être observé par microscopie via un grand nombre de techniques
25 usuelles. Le nucléole est un organe dépourvu de membrane. Malgré cela, le nucléole est individualisé dans le noyau cellulaire et les échanges moléculaires sont actifs et strictement régulés, et réfractaires à la diffusion des macromolécules biologiques. Le nucléole est le plus souvent basophile et donc visible en microscopie optique via des colorants basiques (Pyronine en rouge et bleu de Giemsa). On peut
30 l'observer plus précisément en ayant recours à la microscopie électronique. La localisation de la recombinaise dans le nucléole du noyau cellulaire est ainsi facilement évaluable.

Les systèmes moléculaires permettant la localisation ciblée de protéines dans le
35 nucléole du noyau cellulaire sont bien connus de l'homme de l'art, et ne nécessitent

pas d'être détaillés. Ces systèmes moléculaires comprennent notamment les peptides signaux de localisation nucléolaire NoLS. Ces peptides signaux sont bien connus dans le domaine de la biologie moléculaire et de nombreuses séquences ont été décrites dans la littérature (voir notamment Emmott and Hiscox, 2009; Christophe et al, 2000).

Préférentiellement, la source de recombinaison fonctionnelle selon l'invention comprend un motif ou signal de localisation nucléolaire, ou NoLS, qui permet sa localisation au nucléole.

10

Par « fournie en *trans* dudit vecteur intégratif non viral », on entend au sens de l'invention que la source de recombinaison fonctionnelle est fournie en complément du vecteur intégratif non viral, c'est-à-dire qu'elle est fournie sous une forme distincte du vecteur intégratif non viral. Par exemple, la source de recombinaison fonctionnelle est fournie en *trans* via un plasmide d'expression qui exprimera la protéine dans la cellule, ou encore via un ARN messager. Alternativement, la source de recombinaison fonctionnelle peut être apportée en *trans* sous la forme d'une protéine, par exemple d'une protéine native isolée, ou, préférentiellement, d'une protéine recombinante purifiée. Les plasmides d'expression et/ou les protéines recombinantes peuvent être produits selon des méthodes d'ingénierie moléculaire et de production de protéines recombinantes usuelles, bien connues de l'homme de l'art.

Par « vecteur intégratif non viral » on entend au sens de la présente invention un vecteur dont tout ou partie de la séquence nucléotidique est destinée à s'intégrer dans le génome. Au sens de l'invention, les vecteurs intégratifs non viraux comprennent par exemple les dérivés d'éléments génétiques mobiles tels que les transposons ou pseudo-transposons, mais aussi les plasmides dont la séquence est définie de façon à favoriser la transposition d'un transgène par recombinaison homologue stimulée par une nucléase dont la coupure est spécifique. Dans le cas des transposons, l'intégration de la séquence nucléotidique est généralement stimulée par la transposase pour laquelle ils codent. Dans le cas des plasmides visant une recombinaison homologue, l'intégration de la séquence nucléotidique est généralement stimulée par une nucléase de type Méganuclease, Zinc finger Nucléase, TALEN, ou CRISPR/Cas9. Dans le cas des vecteurs intégratifs non viraux au sens de

l'invention, l'intégration de la séquence nucléotidique est stimulée par la recombinaison apportée en *trans*.

Par « pseudo-transposon », on entend au sens de l'invention un transposon dépourvu de l'activité transposase. Le plus souvent, le pseudo-transposon est un transposon dont le gène codant pour la transposase d'origine a été supprimé ou muté par des techniques de biologie moléculaire. Un pseudo-transposon possède donc les extrémités ITR et UTR mais est dépourvu de l'activité transposase. Il a par conséquent perdu la capacité à transposer, sauf à être associé à une activité transposase extérieure (fournie en *trans*).

De préférence, le pseudo-transposon de l'invention est un transposon dépourvu de l'activité transposase, et choisi parmi les transposons Sleeping Beauty, Frog Prince, piggyBac, MuA, *Mos-1*, *Himar1*, *Hsmar1*, Pokey, Tol2, Tn3, Tn5, Tn7, Tn10, et Hermes et leur dérivés. De préférence, le pseudo-transposon de l'invention est un transposon dépourvu de l'activité transposase choisi parmi piggyBac et ses dérivés.

Par « vecteur intégratif non viral se localisant dans le nucléole du noyau cellulaire », on entend au sens de l'invention que le vecteur intégratif non viral selon l'invention dispose de, ou est associé à, un système moléculaire permettant sa localisation active dans le nucléole du noyau cellulaire. Ainsi, par « localisé dans le noyau cellulaire », on entend que les molécules de vecteur intégratif non viral sont localisées dans la cellule eucaryote de telle sorte qu'une majorité d'entre elles, voire la totalité, se situent dans le nucléole du noyau cellulaire. Cette localisation majoritaire des molécules est facilement identifiable, notamment par des techniques de microscopie de fluorescence usuelles.

Les inventeurs ont déterminé que le signal moléculaire le plus efficace permettant de cibler le vecteur intégratif non viral dans le nucléole du noyau cellulaire, et, ainsi, de favoriser l'intégration de la molécule d'acide nucléique d'intérêt dans une région chromosomique associée au nucléole est, de façon particulièrement surprenante, le signal de localisation nucléaire NoLS. Préférentiellement, le vecteur intégratif non viral selon l'invention comprend un motif de localisation nucléolaire, ou NoLS, qui permet sa localisation active dans le nucléole du noyau cellulaire.

Ainsi, selon un mode de réalisation de l'invention, ledit vecteur intégratif non viral se localise au nucléole et/ou ladite source de recombinaison se localise au nucléole via un signal de localisation nucléolaire NoLS. Selon un mode de réalisation de l'invention, le signal de localisation nucléolaire NoLS est un peptide de séquence SEQ ID No.1 : RQARRNRRRRWRERQRQI.

Par « *via* un signal de localisation NoLS », on entend au sens de l'invention que le vecteur intégratif non viral selon l'invention dispose de, ou est associé à, un peptide NoLS.

10

Le signal de localisation NoLS est un peptide ; il est donc constitué d'acides aminés. Le vecteur intégratif non viral quant à lui est classiquement une molécule constituée d'acides nucléiques. Ainsi, en pratique, l'association du vecteur intégratif non viral selon l'invention avec le signal de localisation NoLS s'effectue par le biais d'un élément de liaison permettant un ancrage chimique ou biochimique. Au sens de l'invention, le système moléculaire de l'invention comprend en outre un élément de liaison dudit signal de localisation NoLS audit vecteur intégratif non viral.

15

De préférence, au sens de l'invention, cet élément de liaison est couplé au signal de localisation NoLS. En outre, l'élément de liaison selon l'invention est capable de s'hybrider au vecteur intégratif non viral, en particulier pour former une triple hélice. Au sens de l'invention, toute molécule capable de s'hybrider à une molécule d'ADN et capable d'être couplée à un peptide peut être utilisée comme élément de liaison.

25

Par hybridation, on entend le processus au cours duquel, dans des conditions appropriées, deux polynucléotides, par exemple un polynucléotide et un ollgonucléotide, se lient avec des liaisons hydrogènes stables et spécifiques pour former un complexe double brin. Dans le cadre de l'invention, une molécule est « capable de s'hybrider avec une molécule d'ADN » si ladite molécule forme un complexe avec la molécule d'ADN dans des conditions stringentes, selon la définition généralement acceptées dans le domaine. L'hybridation de deux polynucléotides peut être totale (le complexe double brin obtenu lors de cette hybridation comprend uniquement des liaisons A-T et des liaisons C-G), ou partielle (le complexe double brin obtenu comprend des bases non liées à une base complémentaire). L'hybridation

30

35

entre deux polynucléotides dépend des conditions opératoires. Toutes ces données sont bien connues et les conditions appropriées peuvent être déterminées par l'homme du métier. En général, selon la longueur des polynucléotides que l'on souhaite hybrider, la température d'hybridation est comprise entre environ 20 et 5 70°C, en particulier entre 35 et 65°C dans une solution saline à une concentration d'environ 0,5 à 1 M.

De nombreux domaines protéiques de liaison à l'ADN ont déjà été décrits, tels que par exemple les domaines de liaison à l'ADN de LexA (Schnarr et al., 1988 ; Oertel- 10 Buchheit et al., 1993), N57SB10 (Izsvak et al., 2002), et pZFD^{bcr-abl} (Mc Namara et al., 2000).

Toutefois, les inventeurs ont découvert, de façon tout à fait surprenante, que les molécules les plus efficaces en tant qu'élément de liaison du signal de localisation 15 NoLS audit vecteur intégratif non viral sont les molécules choisies parmi les oligonucléotides formant des triplex (OFT), en particulier les acides nucléiques peptidiques et leurs dérivés. Préférentiellement, le vecteur intégratif non viral selon l'invention comprend une séquence pour l'hybridation de l'élément de liaison.

20 Avantagement, l'élément de liaison selon l'invention est un oligonucléotide formant des triplex (OFT), de préférence choisi dans le groupe des PNA (acides nucléiques peptidiques), des LNA (acides nucléiques verrouillés, « locked nucleic acids »), des hybrides PNA:PNA (« bisPNA ») et des hybrides LNA:LNA (« Zorro LNA »). En pratique les inventeurs ont déterminé que l'utilisation d'éléments de liaison de 25 type bisPNA oligonucléotide permet la formation de complexes vecteur intégratif non viral /signal de localisation NoLS très stables. De manière avantageuse, l'élément de liaison est un bisPNA couplé au signal de localisation NoLS.

Les acides nucléiques peptidiques (PNA pour « peptide nucleic acids ») et leurs 30 dérivés ont notamment pour avantage de posséder une grande affinité pour l'ADN ou l'ARN. Cette grande affinité permet d'obtenir une hybridation importante avec l'ADN cible, en l'occurrence l'ADN du pseudo-transposon, y compris avec les acides nucléiques peptidiques de petite taille. En outre, les acides nucléiques peptidiques sont des molécules synthétiques qui n'existent pas à l'état naturel, et ne sont pas

reconnues par les nucléases ou les protéases. Ces molécules sont par conséquent particulièrement résistantes aux dégradations enzymatiques.

5 Selon un mode de réalisation de l'invention, le système moléculaire permettant la localisation du vecteur intégratif non viral au nucléole comprend un élément de liaison bisPNA couplé à un signal de localisation NoLS, ledit élément de liaison bisPNA couplé audit signal de localisation NoLS ayant pour séquence la séquence : KKKLLTTCTTCTTTTLLTCTTCTTLLKKKQARRNRRRRWRERQRQI, dans laquelle K, L, I, R, Q, A, N, W et E sont des acides aminés, C et t sont des nucléotides.

10

Le couplage entre l'élément de liaison et le signal de localisation pourra facilement être réalisé via des techniques de biologie moléculaire connues de l'homme de l'art, telle que par exemple l'hybridation par TACH (pour «*Temperature-Assisted Cyclic Hybridization* », détaillée dans Oprea et al., *Mol Biotechnol.*;45(2):171-9; 2010).

15

Par « région chromosomique », on entend au sens de l'invention une portion de chromosome de taille variable. Les régions chromosomiques au sens de l'invention peuvent notamment comprendre en tout ou partie, les régions chromosomiques p et q (autrement appelées « bras p » et « bras q » des chromosomes) telles que classiquement définies en biologie cellulaire, bien qu'elles ne se limitent pas à celles-ci.

20

Il est bien connu dans le domaine technique de l'invention que le nucléole comprend différents types de régions chromosomiques, et notamment les NOR (pour *Nucleolus Organizer Region*) qui sont des régions chromosomiques qui participent à la formation du nucléole. Les NOR comportent plusieurs copies en tandem des gènes codant pour les ARN ribosomiques, ainsi que les domaines de chromatine associés au nucléole (ou « *Nucleolus-Associated Chromatin Domains* », NADs). Ces domaines et régions ont été décrits dans Németh et al. (*PLoS Genet.* 6:e1000889. 2010) et dans Stults et al. (Genome Res. 18:13-18, 2008).

25

30

Préférentiellement, par « région chromosomique associée au nucléole », on entend au sens de l'invention une région comprenant une région organisatrice du nucléole NOR.

Les NOR ont été identifiées chez un grand nombre de mammifères. Chez l'homme, les NOR contiennent des gènes pour les ARNr 5.8S, 18S et 28S et sont localisées sur les bras courts des chromosomes 13, 14, 15, 21 et 22 (autrement dit sur les bras p des chromosomes acrocentriques). Les séquences codant pour l'ARN ribosomique 18S
5 de plus de 270 000 espèces vivantes sont recensées dans la base de donnée « *Ribosomal database project* », et consultables à l'adresse rdp.cme.msu.edu.

Les régions associées au nucléole peuvent par conséquent facilement être visualisées par microscopie, selon des techniques de cytogénétique classiques bien établies. Par
10 exemple, les NOR peuvent être localisées indirectement, en identifiant ces régions par la technique d'hybridation in situ (FISH) sur chromosomes mitotiques. Par ailleurs, les NOR peuvent être identifiées par caryotype, suite à une coloration à l'argent. En outre, il est relativement aisé de localiser par cette même technique les séquences d'intérêt que l'on a tenté d'intégrer dans le génome de la cellule.

15 Ainsi, l'homme du métier peut aisément identifier les régions chromosomiques associées au nucléole et vérifier l'efficacité du système moléculaire de l'invention, quelle que soit l'espèce d'origine de la cellule eucaryote dans laquelle on souhaite effectuer l'insertion de la séquence d'intérêt.

20 Par « intégration », on entend au sens de l'invention le processus de recombinaison conduisant à l'insertion une molécule d'acide nucléique dans une autre. Ainsi, au sens de l'invention le système moléculaire est capable de conduire à l'intégration de ladite molécule d'acide nucléique d'intérêt dans la région chromosomique cible de
25 ladite cellule hôte eucaryote, à savoir dans une région chromosomique associée au nucléole.

Par « cellule hôte eucaryote », on entend au sens de l'invention une cellule hôte comprenant un noyau individualisé. Dans le cadre de la présente invention, la cellule
30 hôte eucaryote est choisie parmi les cellules de fungi, de plantes et d'animaux. Au sens de la présente invention, le terme « animaux » comprend les membres du règne animal, et notamment les mammifères, en particulier l'homme.

Par « molécule d'acide nucléique d'intérêt », on entend au sens de l'invention toute
35 molécule d'acide nucléique dont l'introduction dans le génome de la cellule

eucaryote selon l'invention est souhaitable, telle que par exemple une molécule d'acide nucléique codant pour un ARN ou un peptide dont on souhaite obtenir une expression, de préférence stable, ou encore une molécule d'acide nucléique induisant une mutation d'intérêt.

5

Selon un mode de réalisation préféré, la molécule d'acide nucléique est un gène, en particulier un gène thérapeutique, fonctionnel (on pourra parler de « transgène »). Par « gène thérapeutique », on entend que le gène lui-même ou son produit d'expression présente un intérêt thérapeutique. Au sens de l'invention, un gène est
10 dit « fonctionnel » si la séquence nucléotidique correspondante code pour un ARN, tel que par exemple un ARN structural, une enzyme nucléique (RNaseP, ribozyme), un ARN impliqué dans les mécanismes de l'ARN interférant (miRNA, shRNA), ou code pour un peptide. Selon un mode de réalisation, le gène thérapeutique code pour un peptide s'il comprend au moins un cadre ouvert de lecture (ORF), c'est-à-dire une
15 séquence codante, apte à donner lieu à une séquence en acides aminés présentant une activité thérapeutique. Le gène thérapeutique fonctionnel peut être un gène d'intérêt sauvage ou muté, c'est-à-dire un gène sauvage comprenant une ou plusieurs mutations, dès lors que le produit du gène muté conserve une activité thérapeutique. Ainsi, dans cette définition de « gène fonctionnel », on englobe
20 également les séquences codantes dépourvues de leur promoteur. A titre d'exemple, la molécule d'acide nucléique d'intérêt peut être un gène de résistance à un antibiotique, dépourvu ou non de son promoteur (par exemple, le gène de résistance à la néomycine), ou n'importe quel autre marqueur de sélection approprié.

25 Le nucléole est le site privilégié de la synthèse des molécules d'ARN ribosomal, et comprend essentiellement des molécules d'ARN polymérase de type I. Ainsi, pour faciliter l'expression de la séquence d'intérêt, lorsque les séquences codantes des gènes fonctionnels d'intérêt sont accompagnées d'un promoteur, il s'agit préférentiellement d'un promoteur dépendant de l'ARN polymérase de type I.

30

Au sens de l'invention, une « mutation » est conforme à l'acception usuelle en biotechnologie. Ainsi, une mutation peut être une substitution, une addition ou une délétion d'une ou plusieurs bases dans une séquence nucléotidique, ou d'un ou plusieurs acides aminés dans une séquence protéique. Une « mutation » peut
35 notamment désigner une substitution d'au moins une base d'un codon d'une

séquence nucléotidique, ladite substitution entraînant par exemple, lors de la traduction de la séquence nucléotidique en cause, l'incorporation d'un acide aminé différent aux lieu et place de l'acide aminé natif, dans la séquence protéique résultante. En règle générale, dans le cas d'un gène fonctionnel, on préférera que la
5 ou les mutations n'entraîne(nt) pas de perte de la fonction biologique du produit muté. En revanche, une diminution d'activité pourra éventuellement être tolérée. De manière générale, une « modification génétique » équivaut à une ou plusieurs mutations. Si une séquence codante est génétiquement modifiée, alors, typiquement, elle contient une ou plusieurs mutations.

10

Il a été montré qu'en l'absence de contre-sélection, l'intégration de vecteurs dans les répétitions en tandem codant les ARNr est labile dans les cellules germinales. Ce phénomène est dû à un mécanisme de contrôle appelé restauration, qui élimine, dans la lignée germinale (les cellules de la reproduction), en 2 à 5 générations, les
15 copies des gènes contenant des intégrations ou des délétions.

20

Cette labilité peut être mise en profit pour prévenir une diffusion intempestive des transgènes. Toutefois, il peut être souhaitable dans certains cas de stabiliser l'intégration du transgène dans le temps.

Pour cela, il peut être utile d'effectuer des sélections parmi les organismes clonaux obtenus grâce au système moléculaire de l'invention, selon les techniques et méthodes classiquement utilisées dans le domaine.

25 Le nucléole est l'unique siège de transcription par l'ARN polymérase I et l'ARN polymérase II y est absente, sa localisation étant réduite au nucléoplasme (Malyavantham et al 2008; Németh et al 2010). Cette caractéristique présente un intérêt très particulier dans le contexte de l'utilisation de vecteurs ciblés dans les régions nucléolaires.

30

Les inventeurs ont utilisé certains aspects de la localisation des séquences chromosomiques cibles pour déterminer des modes de réalisation des systèmes moléculaires de l'invention qui sont particulièrement avantageux.

Dans un mode de réalisation, le vecteur intégratif non-viral selon l'invention comprend un gène rapporteur dont l'expression dépend d'un promoteur de l'ARN polymérase de type I.

- 5 L'expression de ce premier gène rapporteur, qui dépend de l'ARN polymérase de type I, pourra ultérieurement être utilisée pour sélectionner les cellules dans lesquelles le système moléculaire de l'invention s'est effectivement intégré au site visé, à savoir le nucléole. Ainsi, l'expression de ce premier gène rapporteur, qui dépend de l'ARN polymérase de type I, pourra ultérieurement être utilisée pour
10 effectuer une « sélection positive ».

Dans un autre mode de réalisation, le vecteur intégratif non-viral selon l'invention comprend un gène rapporteur dont l'expression dépend d'un promoteur de l'ARN polymérase de type II.

- 15 L'expression de ce deuxième gène rapporteur, qui dépend de l'ARN polymérase de type II, pourra ultérieurement être utilisée pour éliminer les cellules dans lesquelles le transgène s'est intégré dans une région autre que le nucléole. Dans cette perspective, l'expression de ce deuxième gène rapporteur, qui dépend de l'ARN polymérase de type II, pourra ultérieurement être utilisée pour effectuer une
20 « sélection négative ».

- A titre d'exemple, on utilisera de préférence comme gène permettant une sélection positive des gènes de résistance à des antibiotiques spécifiques, ou encore des gènes
25 codant pour des protéines pouvant être facilement observées (fluorescence, activité enzymatique détectable). On pourra par ailleurs utiliser comme gène permettant une sélection négative le gène codant la thymidine kinase de l'herpesvirus 1 humain, laquelle rendra le clone sensible à une sélection par ganciclovir, ou bien encore un gène codant pour une protéine fluorescente, qui permettra le tri des clones par
30 cytométrie de flux.

- Selon un mode de réalisation particulièrement avantageux, le vecteur intégratif non-viral selon l'invention comprend un gène rapporteur dont l'expression dépend d'un promoteur de l'ARN polymérase de type I et un gène rapporteur dont l'expression
35 dépend d'un promoteur de l'ARN polymérase de type II.

Les inventeurs ont développé un procédé, mettant en œuvre les systèmes de transposition de l'invention, et permettant d'augmenter de manière très sensible le taux de cellules eucaryotes effectivement génétiquement modifiées par l'intégration de séquences d'intérêt au niveau du nucléole. Le procédé de l'invention repose sur
5 la sélection de caractéristiques d'intérêt par des méthodes usuelles simples et rapides à mettre en œuvre, et qui contribuent à stabiliser l'intégration du transgène dans le temps. La mise en œuvre du procédé de l'invention permet ainsi d'obtenir une population dans laquelle 99,99% des cellules possèdent une modification génétique spécifique. Par « modification génétique spécifique », on entend au sens
10 de la demande que la modification génétique est spécifiquement localisée dans une région chromosomique associée au nucléole de la cellule hôte.

L'invention vise aussi un procédé de modification génétique d'au moins une cellule hôte eucaryote, comprenant :

- 15 a) L'introduction dans ladite cellule d'au moins un système selon l'invention, comprenant ledit vecteur intégratif non viral et ladite source de recombinaise fonctionnelle, tel que décrit plus haut ;
- b) La sélection de cellules de l'étape a), dans lesquelles la molécule d'acide nucléique d'intérêt est exprimée ;
- 20 c) La sélection de cellules de l'étape b), dans lesquelles le vecteur intégratif non viral est intégré spécifiquement dans une région chromosomique associée au nucléole.

Le procédé de l'invention se réfère à la cellule hôte telle que précédemment définie
25 dans la présente demande.

L'introduction du système de transposition dans la cellule hôte eucaryote peut être faite selon toute technique usuelle connue de l'homme de l'art. A titre d'exemple, les techniques usuelles comprennent notamment la transfection par phosphate de calcium, la transfection par liposomes, la transfection par agents polycationiques,
30 DEAE-Dextran, l'électroporation, et les propriétés particulières de réactifs comme le GeneCellin, le Jet PEI ou les blocks copolymères.

En pratique, on pourra transfecter la cellule hôte eucaryote simultanément ou séparément avec le vecteur intégratif non viral et la recombinaise, et de préférence avec un élément de liaison couplé au signal de localisation NoLS.

- 5 Par exemple, le vecteur intégratif non viral (accompagné ou non de l'élément de liaison couplé au signal de localisation NoLS) et la recombinaise peuvent être transfectés conjointement ou séparément.

Par ailleurs il est possible d'assembler le vecteur intégratif non viral et l'élément de
10 liaison couplé au signal de localisation NoLS préalablement à la transfection.

La sélection de l'étape a) pourra être réalisée selon des méthodes classiques, par exemple si le vecteur intégratif non viral comprend un gène rapporteur, en sélectionnant les cellules exprimant ce gène rapporteur. A titre d'exemple, on
15 utilisera de préférence comme gène rapporteur des gènes de résistance à des antibiotiques spécifiques, ou encore des gènes codant pour des protéines pouvant être facilement observées (fluorescence, activité enzymatique détectable). Ainsi, de préférence, pour cette étape, on sélectionnera les cellules selon l'expression d'un gène permettant une sélection positive. Dans ce cas, les cellules sélectionnées à
20 l'issue de l'étape b) sont celles qui expriment le gène permettant la sélection positive.

Dans le cas où le gène rapporteur est un gène de résistance à un antibiotique, l'utilisation de cet antibiotique permettra la sélection des clones dans lesquels la
25 séquence nucléotidique d'intérêt est exprimée.

La sélection de l'étape b) pourra être effectuée là encore selon des méthodes classiques, par exemple si le vecteur intégratif non viral comprend un gène rapporteur dont l'expression dépend d'un promoteur de l'ARN polymérase de type II
30 et permettant une sélection négative. Dans ce cas, les cellules sélectionnées à l'issue de l'étape c) sont celles n'exprimant pas le gène rapporteur. Les cellules exprimant le gène rapporteur dont l'expression dépend d'un promoteur de l'ARN polymérase de type II et permettant une sélection négative sont éliminées.

En fonction des modes de réalisation, l'intégration de la molécule d'acide nucléique d'intérêt est réalisée *in vitro* ou *ex vivo* ou encore *in vivo*. Dans les applications *in vivo*, la cellule hôte est une cellule eucaryote de préférence choisie parmi les cellules de fungi, de plantes, de modèles animaux.

5

Ainsi, l'invention a aussi pour objet l'utilisation d'au moins un système de transposition selon l'invention pour la modification génétique d'au moins une cellule hôte eucaryote, en particulier pour l'intégration contrôlée d'au moins une molécule d'acide nucléique d'intérêt dans le génome d'une cellule hôte eucaryote.

10

Par « intégration contrôlée », on entend au sens de l'invention que l'intégration de la molécule nucléique d'intérêt est contrôlée spatialement. En particulier, par « intégration contrôlée », on entend que l'intégration est contrôlée pour intervenir dans une région chromosomique associée au nucléole.

15

De plus, l'invention vise aussi une cellule hôte eucaryote génétiquement modifiée susceptible d'être obtenue par le procédé de l'invention.

En outre, l'invention vise aussi une cellule hôte eucaryote génétiquement modifiée:

- caractérisée en ce qu'elle contient au moins un système moléculaire selon l'invention; et/ou
- dans laquelle au moins une molécule nucléotidique d'intérêt est intégrée dans une région chromosomique associée au nucléole, conformément au procédé décrit plus haut.

20

25 En particulier, une telle cellule hôte eucaryote exprimera la molécule d'acide nucléique d'intérêt.

L'invention vise aussi un organisme eucaryote transgénique non humain, en particulier un animal transgénique, dont au moins une cellule est une cellule génétiquement modifiée selon l'invention. Par « organisme eucaryote », on désigne ici un fungus, une plante ou un animal. Cette définition ne prétend pas viser spécifiquement les variétés végétales et les races animales en tant que telles, mais peut inclure des individus appartenant à des variétés végétales ou des races animales. Un tel organisme sera, en particulier, un modèle animal notamment utile
35 comme bio-marqueur animal (applicable, par exemple, dans des analyses chimiques

30

environnementales ou de matériaux). Des exemples de modèles animaux sont, de manière non limitative : des mammifères (e.g., bovins, porcins, caprins, ovins, équins, rongeurs tels que souris, rats, hamsters, animaux domestiques tels que chats, chiens), des volatiles (oiseaux, volaille), des poissons.

5

L'invention vise aussi un kit pour l'intégration d'au moins une molécule d'acide nucléique d'intérêt dans le génome d'une cellule hôte eucaryote en vue de son expression, caractérisé en ce qu'il comprend au moins :

- un système moléculaire, et/ou
- 10 - une cellule hôte génétiquement modifiée, et/ou
- un organisme eucaryote transgénique non humain, selon l'invention.

En outre, un tel kit pourra comprendre un ou plusieurs éléments supplémentaires, 15 tels que notamment, une solution tampon compatible avec la recombinaison ou son équivalent fonctionnel, , un ou plusieurs ADN contrôles (témoins de réaction), des oligonucléotides utiles pour le séquençage permettant de contrôler l'efficacité de la réaction, des bactéries compétentes, une notice d'utilisation.

20 Le kit selon la présente invention peut par exemple être utilisé pour modifier le génome de cellules eucaryotes. Ces modifications peuvent par exemple permettre de modifier le métabolisme des cellules cibles, afin d'augmenter leur rendement en bioproduction. Alternativement, les modifications effectuées peuvent avoir des effets thérapeutiques.

25

Le kit selon la présente invention peut être notamment utilisé pour produire une protéine d'intérêt, particulier une protéine d'intérêt cosmétique, une protéine d'intérêt thérapeutique ou une cellule exprimant une telle protéine. La « protéine d'intérêt » est ici le produit d'expression de la molécule d'acide nucléique d'intérêt 30 intégrée, grâce aux moyens de l'invention, dans le génome de la cellule hôte. Une « cellule exprimant une protéine d'intérêt thérapeutique » peut être qualifiée d'implant, utile notamment en thérapie régénérative.

L'invention vise par ailleurs une composition pharmaceutique comprenant au moins :

- 35 - un système moléculaire selon l'invention, et/ou

- une cellule hôte génétiquement modifiée selon l'invention
en association avec un support acceptable d'un point de vue pharmaceutique.

En tous les cas, les nombreuses applications des moyens de l'invention, pour
5 lesquelles ceux-ci se révèlent d'un intérêt considérable et dont la liste fournie ici est
loin d'être exhaustive, font appel à des techniques conventionnelles de biologie
moléculaire bien connues de l'homme du métier.

Les exemples ci-après sont fournis à titre purement illustratif et ne limitent en
10 aucune façon l'objet de la présente invention.

EXEMPLES

Les travaux pour obtenir les résultats suivants, concernant les systèmes de
15 l'invention ont été réalisés sur la lignée HeLa de cellule humaine.

1. Principe du ciblage dans le nucléole des deux composants du vecteur piggybac

La source physique de transposase peut être un plasmide ou un ARN messenger
20 synthétisé à partir de ce plasmide. Ces deux types de source peuvent être
transfectés dans les cellules avec un agent de transfection. L'usage de JetPEI permet
d'avoir un seul protocole de transfection pour ces deux types de source. Pour être
ciblé dans le nucléole, le gène codant la transposase piggybac a été fusionnée en N-
terminal avec un oligonucléotide particulier codant un NoLS (RQARRNRRRRWRERQRQI
25 ; Cochrane et al 1990).

La source physique de vecteur transposon est un plasmide pBS contenant les
extrémités d'un vecteur piggybac avec en son centre une cassette codant des
marqueurs et un site d'hybridation pour un bis-PNA oligonucléotide couplé à un NoLS
30 (PNA-NoLS ; KKK-LLTTCTTCTTTLLTTTCTTCTT-LLL-KKK-RQARRNRRRRWRERQRQI
dans lequel K, L, I, R, Q, A, N, W et E sont des acides aminés ; C et T sont des
nucléotides). L'hybridation du PNA-NoLS se fait en utilisant un protocole similaire à
une PCR. Il est appelé TACH (Temperature-assisted cyclic hybridization ; Oprea et al
2010). La purification des complexes [plasmide/PNA-NoLS] s'effectue en utilisant des
35 kits de mini-purification de plasmides ou de produits de PCR de type Qiagen,

Masherey-Nagel, ou Promega. Le NoLS hybridé sur le plasmide permet l'adressage du plasmide au noyau et au nucléole. Ces complexes [plasmide/PNA-NoLS] sont transfectés dans les cellules avec un agent de transfection (jetPEI).

- 5 L'hybridation d'un bis-PNA oligonucléotide sur l'ADN forme une structure résultant d'une double hybridation sur un brin d'ADN qui est beaucoup plus stable que celle résultant de l'hybridation simple d'un PNA oligonucléotide. Cette double hybridation permet de produire des complexes [Plasmide/bis-PNA] qui sont beaucoup plus stables et spécifiques au moment de l'hybridation que les complexes [Plasmide/PNA]

10

2. Vérification du fonctionnement du ciblage dans le nucléole des deux composants des vecteurs piggybac

- 15 La séquence des plasmides pBS présentant un site de liaison pour l'oligonucléotide PNA-NoLS et contenant un vecteur piggybac dans lequel est inséré un transgène correspondent aux séquences SEQ ID No.2 et SEQ ID No.3.

2.1. Protocole d'hybridation du bis-PNA sur le plasmide source de transposon

20 a - Hybridation

Le bis-PNA-NoLS est ajouté 20, 40 ou 60 fois en excès par rapport au plasmide pour l'hybridation. L'hybridation est effectuée dans des tubes PCR 0.5mL :

	T	x20	x40	x60
25 PBSK Néo (0.5µg/µL final)	5.26	5.26	5.26	5.26
Bis-PNA 400µM	0	0.5	1	1.5
Tampon Phosphate 0.1M pH 5.8	4	4	4	4
EtOHabs	4	4	4	4
H2O	6,74	6.24	5.74	5.24
30 Vf	20 µL	20µL	20µL	20µL

- Une goutte d'huile minérale est ajoutée à la surface de chaque mix pour éviter l'évaporation pendant la réaction d'hybridation. La réaction d'hybridation est réalisée dans un thermocycler. Le programme utilisé est constitué de 20 cycles de deux phases.
- 35

- 30s 80°C (dés-hybridation)
 - 2min 60°C (hybridation)
- 5 À la fin du programme, l'ensemble du mix contenu dans le tube de 0,5 ml est déposé sur un morceau de parafilm afin de séparer l'huile minérale de la phase aqueuse contenant l'ADN. La séparation des deux phases s'effectue en faisant couler l'échantillon sur le parafilm. L'huile se sépare alors de la phase aqueuse que l'on peut récupérer avec un cône puis dans un micro tube de 1,5 mL.
- 10
- b - Purification des complexes bis-PNA/plasmide
- La purification est effectuée avec les réactifs Utilisation du kit Wizard SV Gel ou le PCR Clean-Up System de Promega :
- 15
- 200µL Membrane Binding Solution (pour 2,5 µg d'ADN)
 - Agiter doucement
 - Déposer sur la colonne munie d'un tube collecteur
 - Incuber sur la colonne pendant 1 min à température ambiante
 - Centrifuger 1 min à 14 000 rpm à température ambiante
- 20
- Éliminer l'éluat contenu dans le tube collecteur puis remettre la colonne sur le tube collecteur
 - Déposer sur la colonne 700 µL Membrane Wash Solution
 - Centrifuger 1min à 14 000 rpm
 - Éliminer l'éluat contenu dans le tube collecteur puis remettre la colonne dans
- 25 le tube collecteur
- Déposer sur la colonne 500 µL Membrane Wash Solution
 - Centrifuger 5 min à 14 000 rpm
 - Éliminer l'éluat contenu dans le tube collecteur puis remettre la colonne dans le tube collecteur
- 30
- Centrifuger 1 min à 14 000 rpm pour sécher la colonne (l'élimination des traces d'EtOH est cruciale)
 - Transférer la colonne sur un nouveau microtube 1 de 0.5 mL
 - Éluer les complexes bis-PNA/ADN en incubant la colonne 1 min à température ambiante avec 50 µL « Nucléase-free » water
- 35
- Centrifuger 1 min à 14 000 rpm

- Re-éluer avec le premier éluat en réitérant les deux dernières étapes
- Dosage de la concentration des ARNs au BioSpec Nano avec 2 μ L

c - Contrôle de la qualité des complexes bis-PNA/plasmide

- 5 Les complexes bis-PNA/ADN sont stables pendant au moins 2-3 semaines à 20°C. A tout moment, la qualité d'un stock de complexe peut être vérifiée en utilisant la procédure décrite au ci-dessous.

- Les sites PNABS-560 (fragments de 250 pb) étant clonés entre deux sites uniques dans nos constructions BamHI-SpeI, des gels retards par électrophorèse sur gel d'agarose
10 peuvent être effectués pour vérifier la qualité de l'hybridation spécifique du bis-PNA sur le plasmide.

Etape I : Digestion 5 μ L d'éluat par BamHI-SpeI

Dans un microtube de 1,5 ml, mettre :

- 15 - 0,5 μ g de complexe
- 1 μ L Tampon E 10x Promega
- 1 μ L BSA 1 mg/mL
- 0.5 μ L BamHI
- 0.5 μ L SpeI
20 - Qsp H₂O à 10 μ L

Incuber 1h 37°C puis laisser à 4°C

La totalité des digestions sont déposées sur gel d'agarose 1%, 1XTAE, 1X GelRed

Puits 1 : marqueur de taille λ DNA digéré par HindIII

- 25 Puits 2 : pBS-PB-PNABS-pPol1h-NeoR-pASV40 digéré par BamHI/SpeI

Puits 3 : complexe [pBS-PB-PNABS-pPol1h-NeoR-pASV40/PNA-NoLS] hybridé en excès de X60 en oligonucléotide puis digéré par BamHI/SpeI

Les résultats sont présentés à la figure 1.

- 30 Conclusions : La comparaison du profile central et du profil de droite (B et C) montre l'apparence qu'a le retard lorsque le PNA est hybridé sur l'ADN du plasmide (Disparition de la bande de 250 pb et apparition d'une traînée s'étalant du vecteur depuis le lieu de dépôt de l'échantillon). La méthode d'assemblage et de purification des complexes est donc fiable.

2.2. Vérification de la fonctionnalité du ciblage du plasmide source de transposon

Des complexes [plasmide/PNA-NoLS] colorés avec le fluorochrome intercalant YOYO-1 ont été utilisés pour transfecter des cellules HeLa dans l'objectif de vérifier s'ils ont la propriété d'être activement localisés dans les nucléoles.

5

Les résultats sont présentés à la figure 2.

Conclusions : Nos résultats ont permis de vérifier que les complexes [plasmide/PNA-NoLS] sont activement localisés dans les nucléoles. Cependant, cette vérification n'a pu être effectuée que dans des conditions où de faibles quantités de complexes sont transfectées (50 ng). Lorsque de fortes quantités sont utilisées, le système est saturé et des complexes [plasmide/PNA-NoLS] sont visualisés dans tous les compartiments du noyau.

10

15 2.3. Préparation des sources de transposases

Deux types de transposases ont été utilisées (PB et HIV-NoLS PB) à partir de deux supports : un plasmide vecteur d'expression (Annexe 3) ou un ARN messager. Les ADN plasmiques ont été préparés en utilisant un kit Masherey-Nagel de purification (endotoxine-free).

20

a - Préparation des plasmides

Les d'ARN messager à préparer sont clonés dans le vecteur d'expression plasmidique pCS2+ (c.f. carte ci-dessous).

25 Pour être utilisé comme matrice de transcription, le plasmide doit être linéarisé par clivage enzymatique avec une enzyme de restriction coupant en aval de l'ORF. Pour que la transcription soit efficace, en particulier sa terminaison, il faut que le clivage produise des fragments avec des extrémités franches ou décalées 5' sortantes. Ici, les enzymes utilisées sont localisées dans le multi-site de clonage (MCS) le plus proximal du promoteur CMV : NotI et Acc65I (isoschizomère de KpnI à extrémités 5'

30 sortantes). L'enzyme retenue pour la digestion ne devra pas couper dans l'ORF.

Protocole de linéarisation du plasmide.

- 10 µg d'ADN du plasmide
- 35 - 10 µL de 10X de tampon de digestion enzymatique (ici, NEB 3)

- 1 µL De BSA 10mg/ml (final : 100 µg/ml)
- 5 µL d'enzyme de restriction (NotI ou Acc65I selon le plasmide), c.a.d. 50 unité (excès de 5 fois)
- qsp 100 µL

5

Incuber 1h30 heures à 37°C, puis prélever une aliquote de 5 µL pour contrôler la digestion sur minigel (1X TAE, 0,8% agarose, 1X GelRed ; Témoins : 500 ng de plasmide non digéré et 1 marqueur de taille).

- 10 - Si la digestion est à saturation, incuber la digestion 5 min à 65°C, puis 1 min dans la glace.

Déprotéinisation du plasmide linéarisé.

- 5 µL SDS 20%
- 15 - x µL protéinase K (final 100 µg/ml)
- qsp 200 µL avec de l'eau stérile

Incuber 1 h heures à 37°C puis 5 minutes dans la glace

- Extraire les protéines en condition volume/volume avec du phénol saturé (pH 20 6 à 8)/chloroforme (100 µL phénol + 100 µL chloroforme/alcool iso-amylque 24/1) : vortexer 1 minute, puis centrifuger 5 minutes à 13500 rpm (15000 G), et enfin récupérer la phase aqueuse.
- Extraire les traces de phénol en condition volume/volume avec du chloroforme (+ 200 µL chloroforme/alcool isoamylque 24/1) : vortexer 1 minute, 25 puis centrifuger 5 minutes à 13500 rpm (15000 G), et enfin récupérer la phase aqueuse.
- Précipiter les acides nucléiques en ajustant à 0,3 M acétate de sodium (solution mère à 3 M, donc plus de 20 µL) et 3 volumes d'éthanol absolu. Incuber 1 heure à -80°C ou 1 nuit à -20°C, puis centrifuger 15 min à 13500 rpm (15000 G), 4°C.
- 30 Décanter le culot, laver avec 750 µL d'EtOH 70% (Le culot est lavé avec 250 µL d'EtOH 70 préparé avec de l'EtOH absolu + H2O Nucléase free (kit)) puis séché à la lampe sur la paillasse ou 10 min dans la cloche à vide.
- Re-suspendre le culot de plasmide dans 20 µL d'eau RNase-Free
- Quantifier le plasmide sur le BioSpec Nano avec 2 µL

35

b - Synthèse des ARN messagers (ARNm)

Etape 1 : Transcription et coiffage des ARN

- 1 µg de plasmide linéarisé
- 2 µL mix ATP/CTP/GTP1/5/UTP
- 5 - 2 µL ARCA (tube indépendant du kit de synthèse)
- 2 µL réaction Buffer 10X
- 2 µL enzyme mix
- qsp 20 µL avec H₂O RNase-Free

10 Incubation 5 heures à 37°C

- 1 µL Turbo DNase

Incubation 15 min à 37°C

- Prélever 1 µL pour le contrôle ultérieur sur minigel (contrôle synthèse d'ARN)

Etape 2 : Polyadénylation des ARNs

- 15 - 37 µL H₂O
- 20 µL 5X-E-PAP Buffer
- 10 µL MnCl₂ 25 mM
- 10 µL ATP 10 mM
- 4 µL E-PAP (E. coli Poly(A) Polymerase I)

20 Vf = 100 µL

Incubation 2h à 37°C

- ajouter 60 µL LiCl 7,5M (2,8125 M final)

Précipitation sur la nuit à -20°C

- Centrifuger 15 min à 13500 rpm (15000 G), 4°C.
- 25 - Décanter le culot
- Le culot est lavé avec 250 µL d'EtOH 70% préparé avec de l'EtOH absolu + H₂O nucléase « free » (kit)
- Sécher à la lampe sur la paillasse ou 10 min dans la cloche à vide.
- Le culot est repris dans 20 µL d'H₂O nucléase free (kit de transcription)

30 Etape 3 : Quantification des ARNm

- Dosage de la concentration des ARNs au BioSpec Nano avec 2 µL

Etape 4 : Contrôle de la qualité des ARNm

- Préparer un minigel en 1X TAE (spécial ARN), 0,8% Argarose, 1X GelRed

- Préparer les échantillons d'ARN pour l'électrophorèse : 1 µl d'extrait + 5 µl d'H₂O nucléase free + 6 µl Gel Loading Buffer II (2X ; Ambion). Incuber 5 min à 65°C puis 1 min sur glace
 - Déposer les échantillons dans le gel
 - 5 - Faire migrer 1 h à 90 V, un marqueur de PM ADN et ARN si disponible en parallèle des 2 échantillons : 1 µl Contrôle Synthèse d'ARN et 1 µl d'ARN finaux.
- Etape 5 : Conservation des ARNm à -80°C
- Aliquoter les ARNm en 4 échantillons de 4 µL
 - Conserver à -80°C

10

2.2. Vérification de la fonctionnalité du ciblage du plasmide source de transposon

a - Fonctionnalité du ciblage dans les nucléoles

- Des plasmides pCS2 exprimant des fusions PB-GFP, HIV-NoLS-PB-GFP et DSRed-NoLS (marqueur de localisation nucléolaire (Becherel et al. 2006) ont été utilisés pour transfecter des cellules HeLa afin de vérifier si la protéine HIV-NoLS-PB a la propriété d'être activement localisée dans les nucléoles.
- 15 Les résultats sont présentés aux figures 3 et 4.

- 20 Conclusions : Nos résultats ont permis de vérifier que la transposase HIV-NoLS-PB est activement localisée dans les nucléoles. Cependant, cette vérification n'a pu être effectuée que dans des conditions de faible expression de NoLS-PB. Lorsque de fortes expressions sont utilisées, le système est saturé et NoLS-PB est visualisée dans tous les compartiments nucléaires.

25

b - Fonctionnalité de la transposase HIV-NoLS-PB

- De nombreuses transposases (Sleeping Beauty, Tol2, Mos1, ...) et intégrases ont une activité de recombinaison qui est diminuée ou annulée lorsqu'elles sont fusionnées avec un peptide ou un domaine protéique en N-terminale et/ou en C-terminale.
- 30 L'activité de la fusion N-terminale HIV-NoLS-PB a été comparée à celle de la transposase native PB dans des tests de transposition en cellule HeLa (voir protocole dans la section suivante ; 2 plasmides sont utilisés dans chaque condition de test : une source de transposase et une source de transposon). Deux types de sources de vecteurs transposon ont été utilisés, dont les séquences correspondent aux séquences
- 35 des SEQ ID No.3 et No.4.

Les résultats sont présentés figure 5.

Conclusions : Nos résultats indiquent que l'activité de transposition des deux transposases n'est pas différente quelle que soit la cassette de sélection utilisée dans les vecteurs transposons.

5

3.3. Conditions de transposition utilisées.

Deux types de tests de transposition ont été utilisés. Le premier test implique deux plasmides (une source de transposase et une source de transposon). Dans ce test, l'expression de la transposase a lieu pendant 7 à 8 jours, le temps que le plasmide source de transposase soit naturellement éliminé par la cellule. Le second type de test implique 1 ARNm (comme source de transposase) et un plasmide (comme source de transposon). Dans ce test, l'expression de la transposase a lieu pendant 16 heures, une durée inférieure à celle du temps de division de cellule HeLa fraîchement transfectées.

Le détail des caractéristiques de ces deux sources de transposases PB et les conditions des deux types de tests de transposition sont décrits dans Bire et al (2013). Brièvement, pour chaque test, une source de transposase (~ 200 ng) et une source de vecteur transposon (~ 200 ng) sont transfectées pendant 4 heures dans 50.000 cellule HeLa en utilisant du jetPEI comme agent de transfection. Dans la perspective où le ciblage fonctionne très précisément, nous avons retenu des conditions non-optimales de transposition (en terme de quantité de source de vecteur transposon) de façon à pouvoir observer à côté du ciblage du bruit de fond en contrôle. Les cellules sont ensuite cultivées pendant 48 h avant d'être soumise à une sélection en G418 pendant deux semaines. À la fin de la sélection, les cellules clones de cellules résistants au G418 (c.a.d. exprimant stablement le transgène intégré via le vecteur transposon dans les chromosomes) sont fixées et colorées avant d'être dénombrées ou trypsinisées, lavées et culottées afin de purifier leur ADN génomique. Chaque test est tripliqué. Les conditions utilisées permettent l'obtention de cellules génétiquement modifiées par une intégration moyenne de 2 à 3 vecteurs transposons / génome haploïde de clone cellulaire.

4. Vérification de la spécificité d'insertion des vecteurs piggybac dans des unités répétées en tandem contenant les gènes codant les ARNr 18S-5.8S-28S

4.1. Production de populations de cellules HeLa génétiquement modifiées

Quatorze tests de transposition (indiqués dans le tableau 1) ont été effectués avec les sources dont les séquences correspondent aux séquences SEQ ID No.2 à 4.

N° du test	Source de transposase	Source de transposon	Finalité du test
1	pCS2-vide	pBS-PB-PNABS-pSV40-NeoR-pASV40	Témoins négatif
2	pCS2-vide	pBS-PB-PNABS-pPol1h-NeoR-pASV40	Témoins négatif
3	pCS2-vide	Complexe [pBS-PB-PNABS-pSV40-NeoR-pASV40/PNA-NoLS]	Témoins négatif
4	pCS2-vide	Complexe [pBS-PB-PNABS-pPol1h-NeoR-pASV40/PNA-NoLS]	Témoins négatif
5	ARNm-GFP	pBS-PB-PNABS-pSV40-NeoR-pASV40	Témoins négatif
6	ARNm-GFP	pBS-PB-PNABS-pPol1h-NeoR-pASV40	Témoins négatif
7	ARNm-GFP	Complexe [pBS-PB-PNABS-pSV40-NeoR-pASV40/PNA-NoLS]	Témoins négatif
8	ARNm-GFP	Complexe [pBS-PB-PNABS-pPol1h-NeoR-pASV40/PNA-NoLS]	Témoins négatif
9	pCS2-PB	Complexe [pBS-PB-PNABS-pPol1h-NeoR-pASV40/PNA-NoLS]	Transposition non ciblée sur 8 jours
10	PCS2-HIV-NoLS-PB	Complexe [pBS-PB-PNABS-pPol1h-NeoR-pASV40/PNA-NoLS]	Transposition ciblée sur 8 jours
11	ARNm HIV-NoLS-PB	pBS-PB-PNABS-pSV40-NeoR-pASV40	Transposition non ciblée sur 16 heures
12	ARNm HIV-NoLS-PB	pBS-PB-PNABS-pPol1h-NeoR-pASV40	Transposition non ciblée sur 16 heures
13	ARNm HIV-NoLS-PB	Complexe [pBS-PB-PNABS-pSV40-NeoR-pASV40/PNA-NoLS]	Transposition ciblée sur 16 heures
14	ARNm HIV-NoLS-PB	Complexe [pBS-PB-PNABS-pPol1h-NeoR-pASV40/PNA-NoLS]	Transposition ciblée sur 16 heures

Les tests 1 à 4 sont des témoins de transposition. Une moyenne de ~ 30 clones a été observé pour les tests N° 1, 3, 5 et 7. Une moyenne de ~ 0 à 2 clones a été observé pour les tests N° 2, 4, 6 et 8. Ces témoins correspondent à des tests d'intégration des plasmides sources de transposon par recombinaison aléatoire. Dans nos conditions
5 expérimentales, ils indiquent que les promoteurs cryptiques de l'ARN polymérase 2 dans le promoteur à l'ARN polymérase 1 (pPol1h) sont inefficaces pour exprimer un transgène dans le nucléoplasme.

Des moyennes ~1000 et 2000 clones ont été respectivement obtenues avec les tests
10 N° 9 et 10. Des moyennes ~50, 300, 200 et 300 clones ont été respectivement obtenues avec les tests N° 11, 12, 13 et 14. Ces résultats indiquent que l'intégration de vecteurs dans les chromosomes par transposition dure probablement plus longtemps lorsque la transposition est effectuée avec deux plasmides plutôt qu'avec un ARNm et un plasmide. Cette différence entre les deux tests de transposition avait
15 déjà été observée dans les tests de transposition non ciblés décrits dans Bire et al (2014).

4.2. Vérification de la spécificité d'insertion des vecteurs ciblés.

20 a - Purification des ADN génomique des populations de clones

Des lots de 1000 clones de cellules ont été récoltés à partir des tests de transposition N°9, 10, 11, 12, 13 et 14. L'ADN génomique de ces populations de clones a ensuite été purifié en évitant les kits commerciaux et en utilisant la procédure décrite dans Ahmad et al. 1995.

25

b - Détermination des populations de sites d'insertions dans chaque population de clones

IPCR. Le protocole d'IPCR utilisé est celui décrit dans Wang et al. 2012a, le protocole détaillé ayant été à notre demande transmis par les auteurs (Wang et al. 2012b). De
30 nouveaux oligonucléotides adaptés aux vecteurs piggybac ont été définis utilisés pour l'amplification des populations de sites d'insertion sur leur extrémité 5', puis leur marquage Illumina. Pour chaque échantillons, 3 types de digestions par enzymes de restriction a été réalisé : [Tail], [BsaW1 + BsrF1 + XmaI] (BBX) et [Sau3AI]. Après purification des produits de digestion, les fragments d'ADN à la concentration de ~1
35 ng/μL ont été circularisés pendant la nuit à 16°C par la T4 DNA ligase. Des

oligonucléotides ont été conçus et synthétisés pour amplifier les produits d'IPCR puis les marquer pour le séquençage en Illumina (tableau 2).

Nom de l'oligonucléotide	Séquence	SEQ ID
ITR-UTR PB 5'sens	ATAAACCTCGATATACAGACC	SEQ ID No.5
ITR-UTR PB 5'rev	CGATAAACACATGCGTCAAT	SEQ ID No.6

TABLEAU 2

- 5 Le tableau 3 récapitule les oligonucléotides pour le marquage Illumina des populations de site d'insertion en 5' des vecteurs piggybac.

Nom de l'oligonucléotide	Séquence	SEQ ID
Mega-LinkerIllumina	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCC TGCTGAACCGCTCTTCCGATCTAGTGGCACAGCAGTTAGG	SEQ ID No.7
OligoITR5oute rBCN°1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTACGCGATAAAACACATGCGTCA AT	SEQ ID No.8
OligoITR5oute rBCN°2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTCTACGATAAAACACATGCGTCA AT	SEQ ID No.9
OligoITR5oute rBCN°3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTGATCGATAAAACACATGCGTCA AT	SEQ ID No.10
OligoITR5oute rBCN°4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTTGCGGATAAAACACATGCGTCA AT	SEQ ID No.11

TABLEAU 3

- 10 LAM-PCR. Le protocole de LAM-PCR utilisé est celui décrit dans Bartholomae et al. 2012. Ce protocole a été défini pour amplifier des sites d'insertion de lentivirus. De nouveaux oligonucleotides adaptés aux vecteurs piggybac ont été définis utilisés pour l'amplification des populations de sites d'insertion sur leur extrémité 5', puis leur marquage Illumina. Pour chaque échantillons, 3 types de digestions par enzymes de restriction a été réalisé : [DpnII], [PciI + NcoI + BspHI] (PNB) et [SpeI + AvrII + NheI +
- 15

XbaI] (SANX). Des oligonucléotides pour chacun de ces types de coupure ont été désignés et synthétisés.

Le tableau 4 récapitule les oligonucléotides biotynylés (B) ou non pour les 3
5 amplifications nichées de l'extrémité 5' des vecteurs piggybac.

Nom de l'oligonucléotide	Séquence	SEQ ID
(B)-ITR-UTR 5' PB pSV40 I	(B)- GACTTTCACACCCTAACTGACAC	SEQ ID No.12
(B)-ITR-UTR 5' PB pPol1h I	(B)- GATCCATGAATTCGTCGACATCG ATACCA	SEQ ID No.13
(B)-ITR-UTR PB 5'II	(B)- ATAAACCTCGATATACAGACC	SEQ ID No.14

TABLEAU 4

10 Le tableau 5 récapitule les oligonucléotides utilisés pour chaque type de coupure par enzyme de restriction

Nom de l'oligonucléotide	Séquence	SEQ ID
Oligo I	GACCCGGGAGATCTGAATTCAGTGGCACAGCAGT TAGG	SEQ ID No.15
Oligo II DpnII	GATCCCTAACTGCTGTGCCACTGAATTCAGATC	SEQ ID No.16
Oligo II PNB	CATGCCTAACTGCTGTGCCACTGAATTCAGATC	SEQ ID No.17

TABLEAU 5

Le tableau 6 récapitule les oligonucléotides pour le marquage Illumina des
15 populations de site d'insertion en 5' des vecteurs piggybac.

Nom de l'oligonucléotide	Séquence	SEQ ID
Mega-LinkerIllumina	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGC ATTCCTGCTGAACCGCTCTTCCGATCTAGTGGCA CAGCAGTTAGG	SEQ ID No.18
OligoITR5outerB CN°1	AATGATACGGCGACCACCGAGATCTACACTCTTT CCCTACACGACGCTCTTCCGATCTACGCGATAAA ACACATGCGTCAAT	SEQ ID No.19
OligoITR5outerB CN°2	AATGATACGGCGACCACCGAGATCTACACTCTTT CCCTACACGACGCTCTTCCGATCTCTACGATAAA ACACATGCGTCAAT	SEQ ID No.20
OligoITR5outerB CN°3	AATGATACGGCGACCACCGAGATCTACACTCTTT CCCTACACGACGCTCTTCCGATCTGATCGATAAA ACACATGCGTCAAT	SEQ ID No.21
OligoITR5outerB CN°4	AATGATACGGCGACCACCGAGATCTACACTCTTT CCCTACACGACGCTCTTCCGATCTTGCCGATAAA ACACATGCGTCAAT	SEQ ID No.22
Oligo II SANX	CTAGCCTAACTGCTGTGCCACTGAATTCAGATC	SEQ ID No.23

TABLEAU 6

Le séquençage des populations de fragments correspondant à des sites d'insertion de vecteur piggybac a été réalisé en Illumina MiSeq, 300-pb par lecture. Environ 2 millions de lectures ont été obtenues par échantillon.

c - Procédure d'analyse des séquences Mi-Seq

Préambule sur les limites de la LAM-PCR.

La LAM-PCR est une approche dont l'objectif est de déterminer le profil d'insertion d'un vecteur dans les chromosomes. C'est une approche qualitative et peu quantitative. En effet, les amplifications faites par PCR conduisent systématiquement à l'amplification préférentielle de fragments qui représentent pour les 20 plus nombreux 90% des lectures. Ce biais d'amplification est constant. En conséquence, si une quantification stricte de l'efficacité du ciblage d'un vecteur ne peut pas être effectuée, la comparaison de résultats de LAM-PCR permet de déterminer s'il existe des différences entre échantillons provenant de test de transposition utilisant des vecteurs intégratif à insertion ciblée ou non.

Préambule sur les cibles chromosomiques du ciblage.

Le ciblage des vecteurs de l'invention est dirigé vers les régions chromosomiques associées aux nucléoles.

- 5 Dans un génome haploïde contenant au maximum 500 unités de répétition des gènes codant les ARNr 18S-5.8S-28S, celles-ci représentent ~0,69 % du génome d'une cellule HeLa. Leurs régions flanquantes et les NADs représentent respectivement ~0,55% et ~2,92% du génome. Les séquences liées à la membrane nucléaire, appelées lamina-associated domains (LADs), représentent ~36,83 % du génome (Guelen et al
10 2008). Leur caractéristique d'être compactée sur la membrane nucléaire fait que ces séquences sont, en théorie, moins accessibles à la transposition. Elles ont donc été retenues comme témoins interne à nos expériences de ciblage.

- Dans un test de transposition dans lequel l'insertion est aléatoire (non ciblée), il est
15 donc attendu à ce que 4.4 à 4.5 % des insertions aient lieu dans ces séquences chromosomiques nucléolaires, si et seulement si ces séquences sont aussi accessibles que n'importe quelle autre séquence contenue dans le nucléoplasme. Cependant, comme l'entrée et la sortie de macromolécules dans le nucléole sont fortement contrôlées et nécessitent des mécanismes actifs de transport, il est attendu que le
20 vecteur n'ait pas ou très peu la capacité de diffuser dans cet organite. En conséquence, il est attendu que le niveau d'insertion des vecteurs non ciblés dans les régions chromosomiques nucléolaires soit faible, voir nul, du moins dans les tests dont la durée de transposition est inférieure à celle d'un cycle de division des cellules.

25

Préambule sur les propriétés des populations de sites d'insertion obtenues par PCR.

- Impact de l'approche moléculaire. L'approche LAM-PCR (idem pour l'IPCR) produit des résultats qualitatifs sur la capacité ou non d'un vecteur à intégrer une cible. Cette approche est limitée par des défauts inhérents (sensibilité au taux de GC des
30 séquences à amplifier, les rDNA étant très riche en GC% (60 à 80%), vis à vis des autres séquences dont le GC% moyen est de 45% ; Aird et al 2011, Oyola et al 2012 ; van Heesch et al 2013) dès qu'il s'agit d'obtenir des résultats qualitatifs. Dans le cas des vecteurs de l'invention, où les séquences ciblées sont nucléolaires (au total ~4,16 % du génome), les résultats ne vont strictement pouvoir répondre qu'aux questions
35 suivantes :

- 1 - Les différentes combinaisons de sources de vecteurs, ciblées ou non, sont-elles ciblées dans les séquences nucléolaires ?
- 2 - Les combinaisons ciblées sont-elles plus efficaces pour cibler dans les séquences nucléolaires ?
- 5 Les réponses à ces deux questions sont cependant nécessaires et suffisantes pour démontrer l'efficacité du ciblage.

Impact de l'analyse informatique. Une chaîne complète d'analyse de banques de séquences d'ADN décrivant les populations de sites d'insertions obtenues par LAM-PCR ou par IPCR est disponible (HASTI ; Arens et al 2012). Cependant, HASTI ne permet pas l'automatisation de la séparation des séquences en fonction des tags (spécifiques de chaque conditions) et n'a pas une qualité de cartographie des locus suffisante pour nos travaux. Une chaîne de traitement maison a donc été créée pour traiter nos données.

15 Tout d'abord, au cours des étapes 1 à 3, sont retirés des séquences les fragments suivants :

- tag : 8 possibilités (acg cta gat tgc atc cct gta tgc) – obligatoire (les séquence ne le possédant pas sont exclues du traitement ultérieur)
- 20 • amorce : 2 possibilité (5' ou 3') – obligatoire
- fragment piggybac (PB) : 1 possibilité mais différente suivant le type d'amorce (5' ou 3') – obligatoire
- TSD : le TSD canonique de PB est TTAA
- le linker illumina 3' : peut être présent

25 Une fois ces fragments retirés, les séquences restantes (la séquence flanquant le locus d'insertion) sont cartographiées sur le génome humain en utilisant bwa (Li and Durbin, 2010). Les locus cartographiés sont ensuite filtrés : ne sont conservés que les locus pour lesquels l'alignement de la séquence couvre au moins 95 % de celle-ci.

30 d - Résultats de l'analyse qualitative du ciblage des prototypes de vecteur
Les analyses ont été réalisées avec les deux types de population de séquences produites par IPCR et LAM-PCR à partir d'amplification faites sur l'extrémité 5' du vecteur. Les résultats sont présentés sous forme graphique. Pour chaque prototype de vecteur, les résultats sont exprimés en pourcentages de locus (barres noires) et en
35 pourcentages de séquences cartographiées (barres grises) dans chacune de des 6

catégories nucléaires suivantes : Nucléoplasme, « Lamina-associated domains », « Nucleolus-associated domains », « rDNA repeats », « rDNA flanking regions » et Total Nucléoles. Il a été considéré que les deux pourcentages devaient avoir la même tendance dans les 4 essais pour être pris en compte. En références ont été utilisés les
5 pourcentages attendus dans chacun des compartiments si la transposition avait eu lieu de façon non ciblée et de façon équivalente dans tous les compartiments nucléaires. Le prototype de vecteur totalement ciblé pour ses deux composants et dont le gène marqueur dépend d'un promoteur nucléolaire (Poll) est indiqué en italique.

10

d1 - Contrôle de l'aptitude de nos approches expérimentales à détecter tropismes d'insertion des vecteurs piggyBac en fonction de l'accessibilité à l'ADN chromosomique.

Les LADs sont des compartiments nucléaires qui contiennent des domaines
15 chromosomiques fortement hétérochromatinisés. Il est donc attendu que les régions d'ADN génomique contenues dans les LADS soient moins accessibles à l'intégration de vecteurs intégratifs, en particulier lors des phases G0 et G1 du cycle cellulaire. Les profils d'insertion de 6 prototypes de vecteurs ont donc d'abord été analysés dans cette perspective (voir figure 6 et 7).

20

La première observation est que les six prototypes de vecteurs tendent à significativement plus s'insérer dans l'ADN contenu dans le nucléoplasme. Les taux d'intégration dans le nucléoplasme sont systématiquement plus élevés que le taux attendu par le simple effet du hasard.

25

La seconde observation est que les six prototypes de vecteurs tendent à significativement moins s'insérer dans l'ADN contenu dans les LADs.

Ces résultats ont été confirmés en utilisant non pas la quantité d'ADN présent dans
30 chaque compartiment nucléaires, mais la quantité de sites d'insertion potentiels pour nos vecteurs piggyBac (tétranucléotide « TTAA »).

Proportions des catégories nucléaires / de la quantité d'ADNg ou du nombre de motifs TTAA dans la version hg19 du génome humain.

35

Compartiments nucléaires	% dans l'ADNg total	% des motifs TTAA totaux	# de motifs TTAA	Densité en TTAA / Mpb
Nucléoplasme	59,0057	53,0563	38092044	12304,85
LADs	36,8351	43,5537	16590508	14540,64
NADs	2,9292	3,0033	1144020	12608,90
Flanking rDNA	0,5359	0,1898	72290	4354,82
rDNA	0,6941	0,1969	75000	3488,45
Totale Nucléole	4,1591	3,3899	1291310	12028,86

TABLEAU 7

Les résultats obtenus (voir les figures 8 et 9) en prenant en compte le nombre de sites TTAA par catégorie de domaines nucléaires confirment et même amplifient les conclusions tirées de la première analyse sur la différence d'accessibilité aux vecteurs intégratifs entre LADs et nucléoplasme. Ces contrôles ont permis de vérifier que les prédictions faites sur l'accessibilité de l'ADN chromosomique aux vecteurs intégratifs en fonction sa localisation dans le nucléoplasme ou les LADs sont systématiquement vérifiées dans les résultats de notre approche expérimentale. Les taux d'intégration sont systématiquement moins élevés que le taux aléatoire attendu dans le nucléoplasme.

d2 - Présence d'insertion dans l'ADN génomique contenu dans l'ensemble des régions nucléolaires.

Les régions nucléolaires ont représentées par les « Nucleolus-associated domains », les « rDNA repeats », et les « rDNA flanking regions ».

L'analyse des résultats (figure 10) fournis quatre informations.

- 1 - L'ADN génomique contenu dans le nucléole est difficile d'accès pour un vecteur intégratif, même ciblé.
- 2 - Le prototype de vecteur intégratif dont les deux composants sont ciblés dans le nucléole (pCS2-HIV-PB x pPol1h-Neo-PNA) et dont l'expression du marqueur est dépendante d'un promoteur nucléolaire (ARN polymérase I) est le plus efficace pour s'intégrer dans les séquences nucléolaire (ratio relatif de 5 à 300 fois, selon le test).
- 3 - Un vecteur ciblé ou non dans le nucléole et contenant un marqueur de sélection dont l'expression est placée sous la dépendance d'un promoteur nucléaire non nucléolaire (pSV40 - ARN polymérase II) est incapable d'être retenu par le processus

de sélection cellulaire. Cette incapacité provient de l'absence de l'ARN polymérase II dans le nucléole.

4 - Les transfections faites avec des vecteurs dont les deux composants sont plasmidiques seraient a priori un peu moins efficace pour le ciblage dans les
5 séquences nucléolaires.

d3 - Régions nucléolaires accessibles aux vecteurs intégratifs.

Pour ces analyses, les régions nucléolaires ont été subdivisées en trois catégories : les « Nucleolus-associated domains », les « rDNA repeats », et les « rDNA flanking
10 regions ». L'analyse des résultats (figure 11) permet d'accéder à deux informations concernant l'accessibilité des trois catégories de régions nucléolaires aux vecteurs intégratifs utilisés, l'impact de la méthode de transfection.

C'est dans les régions NAD (catégorie de région nucléolaire) que les vecteurs ciblés,
15 dont le marqueur de sélection dépend de l'ARN polymérase I, s'intègrent le mieux et à un niveau supérieur de ceux des autres prototypes de vecteurs. Dans ces séquences, les résultats obtenus en IPCR suggèrent que la différence d'efficacité du ciblage nucléolaire est plus à l'avantage du vecteur ciblé lorsque le prototype est constitué d'un ARNm et d'un plasmique que de deux composants plasmidiques.

20

Les conclusions tirées des résultats sur les régions chromosomiques flanquantes (figure 12) des gènes codant les ARNs ribosomiaux 18S-5.8S-28S sont identiques à celles des NADs.

25 Dans les limites d'efficacité de la préparation des banques de sites d'insertion des vecteurs par IPCR ou LAM-PCR, et de la sensibilité de nos analyses in silico, les résultats (figure 13) indiquent que nos prototypes de vecteurs sont incapables d'accéder aux gènes répétés en tandem codant les ARN ribosomiaux 18S-5.8S-28S et à leurs espaceurs non transcrits. Trois hypothèses non exclusives permettent
30 d'expliquer ce phénomène :

1 - Les gènes codant les ARN ribosomiaux 18S-5.8S-28S sont beaucoup moins accessibles que les autres séquences nucléolaires, du fait de leur rôle essentiel dans la transcription par l'ARN polymérase I.

2 - Les ADNs polymérase utilisées en PCR sont beaucoup moins efficaces pour l'amplification sur des matrices d'ADN ayant un taux de GC de 60 à 80% que sur des matrices avec un taux de GC de ~45%. De fait, les amplifications de site d'insertion des rDNA sont dramatiquement sous représentées dans les produits finaux d'IPCR et de LAM-PCR.

Dans cette perspective, une analyse des taux d'efficacité d'amplification des séquences par PCR a été effectuée. Les résultats (figures 14 et 15) indiquent que l'IPCR dans nos conditions expérimentales amplifie de façon homogène les séquences, quelle que soit leur localisation nucléolaire et l'échantillon. Cependant, l'intérêt de cette qualité d'amplification est minoré de par le fait que seulement 0,1 à 1% des séquences produites ne proviennent pas de produits d'amplification artéfactuelle. À contrario, les résultats indiquent que la LAM-PCR dans nos conditions expérimentales amplifie de façon biaisée les séquences en fonction de leur localisation nucléolaire et de l'échantillon. Cette propriété est à mettre en parallèle du fait que les séquences récupérées à partir de produits de LAM-PCR sont de 70 à 99% cartographiables in silico.

3 - La composition en séquences répétées (en particulier en séquence *Alul* (300-bp) des espaceurs non transcrits (eux-mêmes à 60 à 80% de taux de GC) présents entre les régions transcrites en ARN ribosomiaux 18S-5.8S-28S dans les rDNA ne permet pas de les cartographier correctement in silico. D'autre part la séquence exacte des régions contenant les unités répétées codant pour les ARNr n'est pas connue.

4. Conclusions sur l'efficacité des vecteurs ciblés

Les résultats expérimentaux permettent de tirer différentes conclusions.

- 1 - Les vecteurs intégratifs de l'invention peuvent être efficacement ciblés dans l'ADN nucléolaire.
- 2 - L'utilisation des vecteurs ciblés peut se faire avec deux composants plasmidique ou 1 ARNm et un plasmide.

5. Analyse complémentaires sur l'aptitude des vecteurs ciblés à s'intégrer dans les gènes codant les ARN ribosomiaux 18S-5.8S-28S

L'impossibilité de détecter des insertions de vecteurs dans les gènes codant les ARNs ribosomiaux 18S-5.8S-28S peut avoir deux origines. La première serait que les vecteurs ne peuvent accéder à ces séquences chromosomiques car elles seraient protégées dans les nucléoles et nécessiteraient un mécanisme spécifique d'accès à ces séquences. La seconde serait liée à certaines propriétés d'amplification des polymérases utilisées en PCR et qui sont d'amplifier avec une plus grande efficacité les séquences les plus riches en AT aux dépens de celles riches en GC. La seconde hypothèse a été testée en utilisant une stratégie de LAM-PCR ciblée sur la région transcrite des gènes codant les ARNs ribosomiaux 18S-5.8S-28S des oligonucléotides spécifiquement ancrés d'une part dans cette région et, d'autre part, dans les extrémités 5' et 3' des vecteurs utilisés.

Caractéristiques des oligonucléotides

Primers ancrés dans les ADNr : Chaque paire de primers a été définie de façon à permettre une amplification nichée dans les ADNr. Ainsi dans le 18S, la première amplification est effectuée avec le 18F ou le 18R, la seconde étant respectivement faite avec les 18S F2 et 18S R2.

- 20 18S R - SEQ ID No.24
5' GCTGAACGCCACTTGTCCTCTAAGAAGT 3'
18S F - SEQ ID No.25
5' GGACACGGACAGGATTGACAGATTGATAGC 3'
18S R2 - SEQ ID No.26
25 5' GCTATCAATCTGTCAATCCTGTCCGTGTCC 3'
18S F2 - SEQ ID No.27
5' ACTTCTTAGAGGGACAAGTGGCGTTCAGC 3'
- 28S Rmid - SEQ ID No.28
30 5' GGGCTAGTTGATTCCGGCAGGTGAGTTG 3'
28S Fmid - SEQ ID No.29
5' CTAGTAGCTGGTCCCTCCGAAGTTTCCCT 3'
- 28S Rmid2 - SEQ ID No.30
35 5' GGGAACTTCGGAGGGAACCAGCTACTA 3'

28S Fmid2 - SEQ ID No.31

5' GTAACAACTCACCTGCCGAATCAACTAGCC 3'

28S Rend - SEQ ID No.32

5 5' CGACCCAGAAGCAGGTCGTCTACGAAT 3'

28S Fend - SEQ ID No.33

5' GGCGAAGCTACCATCTGTGGGATTATGACTG 3'

28S Rend2 - SEQ ID No.34

10 5' CAGTCATAATCCCACAGATGGTAGCTTCGC 3'

28S Fend2 - SEQ ID No.35

5' CGCTAAACCATTCTGTAGACGACCTGCTTCTG 3'

Primers ancrés aux extrémités du vecteur piggybac

15 ITR 5'

ITR5a - SEQ ID No.36 - 5' GAATTCGTCGACATCGATACCAAAAGTTTTGTTAC 3'

ITR5b - SEQ ID No.37 - 5' CCTCGATATACAGACCGATAAAACACATGCGTCA 3'

ITR 3'

ITR3a - SEQ ID No.38 - 5' GCGGCGACTGAGATGTCCTAAATGCAC 3'

20 ITR3b - SEQ ID No.39 - 5' GCGACGGATTGCGCTATTTAGAAAGAGAG 3'

Echantillons utilisés

Deux paires d'échantillons d'ADN génomiques provenant de vecteurs non-ciblés et ciblés ont été utilisés : N°9 et N°10, et N°12 et N°14.

25

Procédure d'amplification des vecteurs intégrés dans les ADNr

a. PCR linéaire (1 échantillon = 2 PCR linéaire en 5' et en 3')

Diluer les oligos (100 µM) au 1/100 = 1 µM

30 ADNg 50 à 100 ng

Tampon 5X GC One Taq 10 µl

dNTPs (10 mM each) 1 µl

Primer (B) transposon I (1 µM) 4,18 µl (83,5 nM)

One Taq polymérase (NEB 5U/µl) 0,5 µl

35 H2O QSP 50 µl

Notes :

-Primer (B) transposon I

(B)-ITR-UTR 5' PB pSV40 I	(B)-GACTTTCACACCCTAACTGACAC	SEQ ID No.12
(B)-ITR-UTR PB 3' pSV40 I	(B)- TTCTTCGCCCACCCCAACTTGTTTATTGC	SEQ ID No.40
(B)-ITR-UTR 5' PB pPol1h	(B)- GATCCATGAATTCGTCGACATCGATACCA	SEQ ID No.13
(B)-ITR-UTR PB 3' pPol1h	(B)- GGACTCGGAAACGAATTCGATATCAAGCT	SEQ ID No.41

- 5 -Faire un contrôle avec ADN_g issus de cellules non transfectées
- 1- 95°C 2 min
 - 2- 95°C 45 sec l
 - 3- 60°C 45 sec l x50
 - 4- 68°C 6 min l
 - 10 5- 68°C 10 min

Ajouter 2,5 U (0,5 µl) de Taq polymérase par tube et réitérer tout le cycle de PCR

b. Capture magnétique

- 15 - Exposer 20 µl de particules magnétiques (10 µg/µl) (Dynabeads M-280 Streptavidin (Invitrogen, Carlsbad, USA) 60 secondes au champ magnétique du Single Place Magnetic Stand (Invitrogen, Carlsbad, USA) à TA.
- Enlever le surnageant en présence du champ magnétique.
 - Resuspendre les particules dans 40 µl PBS-0,1% BSA (pH 7,5). Eliminer le surnageant
 - 20 en présence du champ magnétique. Recommencer cette étape une deuxième fois. (= 2 lavages des billes magnétiques en PBS-0.1%BSA)
 - Laver les particules avec 20 µl de LiCl 3M BW buffer (même procédure magnétique qu'avec les lavages en PBS-0.1%BSA).
 - 25 - Les resuspendre les billes dans 50 µl de LiCl 6M BW buffer.
- Note : A partir des étapes suivantes, procéder avec le Side Skirted Magnetic device.

- Transférer 50 µl de solution de particules magnétiques dans chaque produit de PCR linéaire (le ratio de produit PCR et la solution de LiCl doit toujours être de 1:1). Mélanger doucement.
- Incuber sur un agitateur à 300 rpm à TA (entre 20° C et 43° C) OVN (8h-48h).
- 5 - Exposer l'échantillon 60 sec à un champ magnétique, éliminer le surnageant en présence du champ magnétique et laver les billes une fois avec 100 µl d'eau distillée.

c - 1ère Amplification

Chaque PCR linéaire est amplifiée par 6 jeux d'oligonucléotides (1 extrémité du transposon puis 6 amorces rDNA)).

Par échantillons, on s'attend donc à récolter 12 produits de PCR (6 en 5' et 6 en 3'). Avec 4 échantillons on aura donc à ce stade 48 tubes de PCR. La 1ère PCR est réalisée sur 5 µL d'ADN linéarisé et purifié. Remettre en solution les billes dans 70 µl d'eau distillée

15

	Mix
ADNss purifié	5µL
Tampon 5X GC One Taq	5µL
dNTP mix 10mM	0.5µL
Primer 1 10µM	0.5µL
Primer 2 10µM	0.5µL
OneTaq polymerase	0.25µL
H ₂ O	xµL
Vf	25µL

Primer 1 = ITR5a ou ITR3a

Primer 2 = 18SF, 18SR, 28SFmid, 28SRMid, 28S Fend, 28SRend

20 Programme du thermocycleur

Hold à 94° C

30s à 94° C

30s à 94° C

30s à 62° C

} X 35 cycles

25 6 min à 68° C

5min à 68°C

Hold à 4°C

- Contrôle sur gel de la qualité des produits de PCR (5 µl) avec un gel 1.5 % Agarose, 1X TBE, 1X Gelred.

c. 2ème Amplification

Chaque tubes de PCR a son mix de primer niché

	Mix
PCR	5µL
Tampon 5X GC One Taq	5µL
dNTP mix 10mM	0.5µL
Primer 1 10µM	0.5µL
Primer 2 10µM	0.5µL
OneTaq polymerase	0.25µL
H ₂ O	xµL
Vf	25µL

Primer 1 = ITR5b ou ITR3b

- 10 Primer 2 = 18SF2, 18SR2, 28SFmid2, 28SRMid2, 28S Fend2, 28SRend2

Programme du thermocycleur

Hold à 94°C

30s à 94°C

30s à 94°C

- 15 30s à 62°C } X 35 cycles

6 min à 68°C

5min à 68°C

Hold à 4°C

- 20 Contrôle sur gel de la qualité des produits de PCR (5 µl) avec un gel 1.5 % Agarose, 1X TBE, 1X Gelred.

d. Purification des produits de PCR et séquençage

Pour chaque paire de primers, une comparaison des produits d'amplification a été

- 25 effectuée entre d'une part les échantillons 9 et 10 et, d'autre part, les échantillons

- 12 et 14. Les fragments d'amplification présents dans les produits provenant des échantillons 10 ou 14, et respectivement absents dans les produits provenant des échantillons 9 ou 12 ont été purifiés puis séquencés par un prestataire par la méthode de Sanger.
- 5
- e. Résultats et conclusion
- Une trentaine de produits d'amplification ont été séquencés. Leur séquence a permis de démontrer que le vecteur est capable de s'intégrer dans différents motifs TTAA contenus dans les gènes codant les ARNs ribosomiaux 18S-5.8S-28S.
- 10 Ces résultats indiquent que les vecteurs ciblés dans le nucléole avec l'invention sont capables de s'intégrer dans toutes les séquences nucléolaires.

REFERENCES

- 15 Ahmad NN, Cu-Unjieng AB, Donoso LA. 1995. Modification of standard proteinase K/phenol method for DNA isolation to improve yield and purity from frozen blood. *J Med Genet.* 32:129-130.
- Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12:R18.
- 20 Arens A, Appelt JU, Bartholomae CC, Gabriel R, Paruzynski A, Gustafson D, Cartier N, Aubourg P, Deichmann A, Glimm H, von Kalle C, Schmidt M. 2012. Bioinformatic clonality analysis of next-generation sequencing-derived viral vector integration sites. *Hum Gene Ther Methods.* 23:111-118.
- 25 Bartholomae CC, Glimm H, von Kalle C, Schmidt M. 2012. Insertion site pattern: global approach by linear amplification-mediated PCR and mass sequencing. *Methods Mol Biol.* 859:255-265.
- Becherel OJ, Gueven N, Birrell GW, Schreiber V, Suraweera A, Jakob B, Taucher-Scholz G, Lavin MF. 2006. Nucleolar localization of aprataxin is dependent on interaction with nucleolin and on active ribosomal DNA transcription. *Hum Mol Genet.* 15:2239-2249.
- 30 Bire S, Ley D, Mermoud N, Casteret S, Bigot Y, Rouleux-Bonnin F (2014) Monitoring the biosafety of the piggyBac transposon for gene delivery using mRNA and insulators. *PloSOne.* 8:e82559.

- Cadiñanos J, Bradley A. 2007. Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res.* 35:e87.
- Catez F, Erard M, Schaerer-Uthurralt N, Kindbeiter K, Madjar JJ, Diaz JJ. 2002. Unique motif for nucleolar retention and nuclear export regulated by phosphorylation. *Mol Cell Biol.* 22:1126-1139.
- Cochrane AW, Perkins A, Rosen CA. 1990. Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function. *J Virol.* 64:881-885.
- Di Matteo M, Mátrai J, Belay E, Firdissa T, Vandendriessche T, Chuah MK. (2012) PiggyBac toolbox. *Methods Mol Biol.* 859:241-254.
- Emmott E, Hiscox JA. (2009) Nucleolar targeting: the hub of the matter. *EMBO Rep.* 10:231-238.
- Floutsakou I, Agrawal S, Nguyen TT, Seoighe C, Ganley AR, McStay B. 2013. The shared genomic architecture of human nucleolar organizer regions. *Genome Res.* Oct 31. [Epub ahead of print ; doi : 10.1101/gr.157941.113]
- Gonzalez IL, Sylvester JE. 1995. Complete sequence of the 43-kb human ribosomal DNA repeat: analysis of the intergenic spacer. *Genomics.* 27:320-328.
- Ide S, Miyazaki T, Maki H, Kobayashi T. 2010. Abundance of ribosomal RNA gene copies maintains genome integrity. *Science* 327:693-696.
- Li H, Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 26 :589-595.
- Li MA, Turner DJ, Ning Z, Yusa K, Liang Q, Eckert S, Rad L, Fitzgerald TW, Craig NL, Bradley A. 2011. Mobilization of giant piggyBac transposons in the mouse genome. *Nucleic Acids Res.* 39:e148.
- Malyavantham KS, Bhattacharya S, Barbeitos M, Mukherjee L, Xu J, Fackelmayer FO, Berezney R. 2008. Identifying functional neighborhoods within the cell nucleus: proximity analysis of early S-phase replicating chromatin domains to sites of transcription, RNA polymerase II, HP1gamma, matrin 3 and SAF-A. *J Cell Biochem.* 105:391-403.
- Németh A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Péterfia B, Solovei I, Cremer T, Dopazo J, Längst G. 2010. Initial genomics of the human nucleolus. *PLoS Genet.* 6:e1000889.
- Oprea II, Simonson OE, Moreno PM, Viola JR, Lundin KE, Smith CI. (2010) Temperature-assisted cyclic hybridization (TACH): an improved method for supercoiled DNA hybridization. *Mol Biotechnol.* 45:171-179.

- Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, Campino S, Turner DJ, Macinnis B, Kwiatkowski DP, Swerdlow HP, Quail MA. 2012. Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics*. 13:1.
- 5 Palazzoli F, Carnus E, Wells DJ, Bigot Y. 2008. Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration. *Curr Gene Ther*. 8:367-390.
- Palazzoli F, Testu FX, Merly F, Bigot Y. 2010. Transposon tools: worldwide landscape of intellectual property and technological developments. *Genetica*. 138:285-299.
- 10 Palmer TD, Miller AD, Reeder RH, McStay B. 1993. Efficient expression of a protein coding gene under the control of an RNA polymerase I promoter. *Nucleic Acids Res*. 21:3451-3457.
- Stults DM, Killen MW, Pierce HH, Pierce AJ. 2008. Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Res*. 18:13-18.
- 15 Sylvester JE, Whiteman DA, Podolsky R, Pozsgay JM, Respass J, Schmickel RD. 1986. The human ribosomal RNA genes: structure and organization of the complete repeating unit. *Hum Genet*. 73:193-198.
- van Heesch S, Mokry M, Boskova V, Junker W, Mehon R, Toonen P, de Bruijn E, Shull JD, Aitman TJ, Cuppen E, Guryev V. 2013. Systematic biases in DNA copy number originate from isolation procedures. *Genome Biol*. 14:R33.
- 20 Wang H, Mayhew D, Chen X, Johnston M, Mitra RD. 2012a. "Calling cards" for DNA-binding proteins in mammalian cells. *Genetics*. 190:941-949.
- Wang H, Mayhew D, Chen X, Johnston M, Mitra RD. 2012b. "Calling cards" for DNAbinding proteins in mammalian cells : protocol.

REVENDICATIONS

1. Système moléculaire pour l'intégration d'une molécule d'acide nucléique d'intérêt dans une région chromosomique associée au nucléole, comprenant au moins :
- un vecteur intégratif non viral comprenant un motif de localisation nucléolaire, et dépourvu de source de recombinaison fonctionnelle; et
 - une source de recombinaison fonctionnelle comprenant un motif de localisation nucléolaire, fournie en *trans* dudit vecteur intégratif non viral.
2. Système selon la revendication 1, caractérisé en ce que le vecteur intégratif non viral est un pseudo transposon.
3. Système selon la revendication 1 ou 2, caractérisé en ce que ledit vecteur intégratif non viral et ladite source de recombinaison fonctionnelle se localise au nucléole comprenant un motif de localisation nucléolaire.
4. Système selon l'une quelconque des revendications 1 à 3, caractérisé en ce que ledit motif de localisation nucléolaire est un motif de localisation nucléolaire NoLS.
5. Système selon l'une quelconque des revendications 1 à 4, caractérisé en ce qu'il comprend en outre un élément de liaison dudit signal de localisation audit vecteur intégratif non viral.
6. Système selon la revendication 5, caractérisé en ce que ledit élément de liaison est un oligonucléotide formant des triplex (OFT), de préférence choisi dans le groupe des PNA (acides nucléiques de peptides, « peptide nucleic acids »), des LNA (acides nucléiques verrouillés, « locked nucleic acids »), des hybrides PNA:PNA (« bisPNA ») et des hybrides LNA:LNA (« Zorro LNA »).
7. Système selon la revendication 6, caractérisé en ce que ledit élément de liaison est un bisPNA.

8. Système selon l'une quelconque des revendications 1 à 7, caractérisé en ce que ledit vecteur intégratif non viral est un pseudo-transposon choisi parmi les transposons Sleeping Beauty, Frog Prince, piggyBac, MuA, *Mos-1*, Pokey, , Tol2, Tn3, Tn5, Tn7, Tn10, Hermes et leurs dérivés.
- 5
9. Système selon l'une quelconque des revendications 4 à 8, caractérisé en ce que ledit signal de localisation nucléolaire NoLS est un peptide de séquence SEQ ID No.1 : RQARRNRRRRWRERQRQI.
- 10
10. Système selon la revendication 9, caractérisé en ce que le ledit élément de liaison bisPNA couplé à audit signal de localisation NoLS a pour séquence la séquence: KKKLLTTCTTCTTTTLLTTTCTTCTTLLKKKRQARRNRRRRWRERQRQI, dans laquelle K, L, I, R, Q, A, N, W et E sont des acides aminés, C et T sont des nucléotides.
- 15
11. Procédé *in vitro* de modification génétique d'au moins une cellule hôte eucaryote, comprenant :
- a) l'introduction dans ladite cellule d'au moins un système selon l'une des revendications 1 à 10;
- 20
- b) la sélection de cellules de l'étape a), dans lesquelles la molécule d'acide nucléique d'intérêt est exprimée ;
- c) la sélection de cellules de l'étape b), dans lesquelles le vecteur intégratif non viral est intégré dans une région chromosomique associée au nucléole.
- 25
12. Cellule hôte eucaryote génétiquement modifiée caractérisée en ce qu'elle contient au moins un système selon l'une quelconque des revendications 1 à 10.
- 30
13. Cellule hôte eucaryote génétiquement modifiée susceptible d'être obtenue par le procédé selon la revendication 11, caractérisée en ce qu'au moins une molécule d'acide nucléique d'intérêt est intégrée dans une région chromosomique associée au nucléole.

14. Organisme eucaryote transgénique non humain, en particulier animal transgénique, dont au moins une cellule est une cellule selon la revendication 12 ou 13.
- 5 15. Kit pour l'intégration d'au moins une molécule d'acide nucléique d'intérêt dans le génome d'une cellule hôte eucaryote en vue de son expression, caractérisé en ce qu'il comprend au moins :
- un système moléculaire selon l'une quelconque des revendications 1 à 10 ; et/ou
 - une cellule hôte selon la revendication 12 ou 13 ; et/ou
- 10 - un organisme eucaryote transgénique non humain selon la revendication 14.

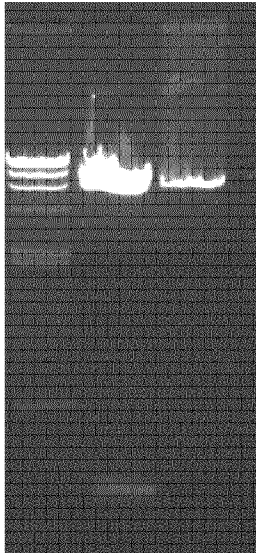


Fig. 1

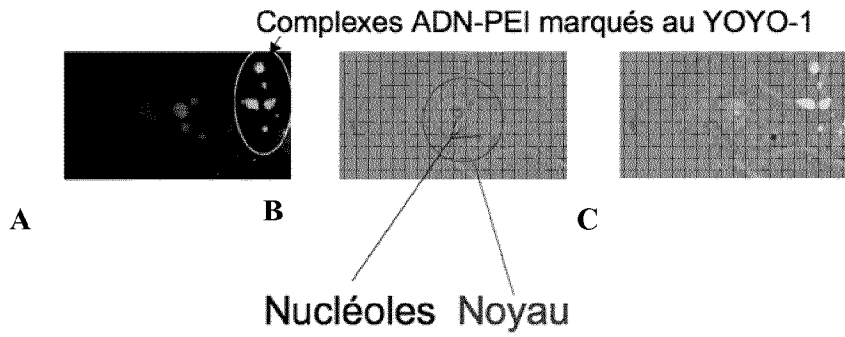


Fig. 2

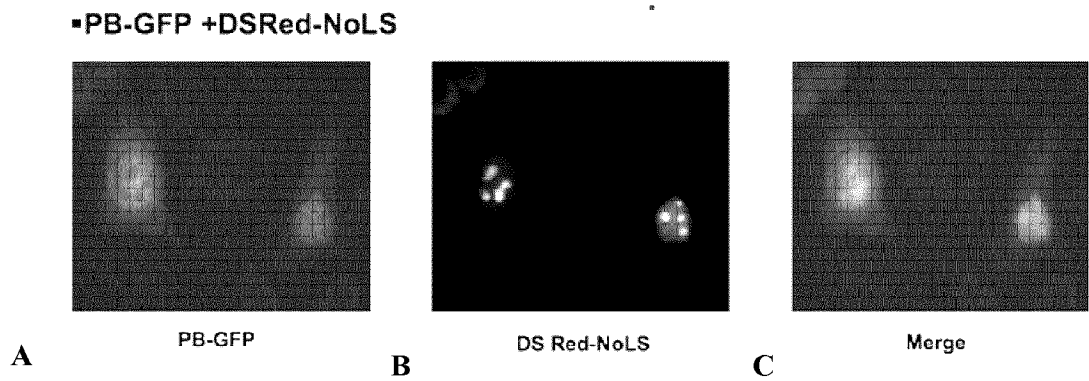


Fig. 3

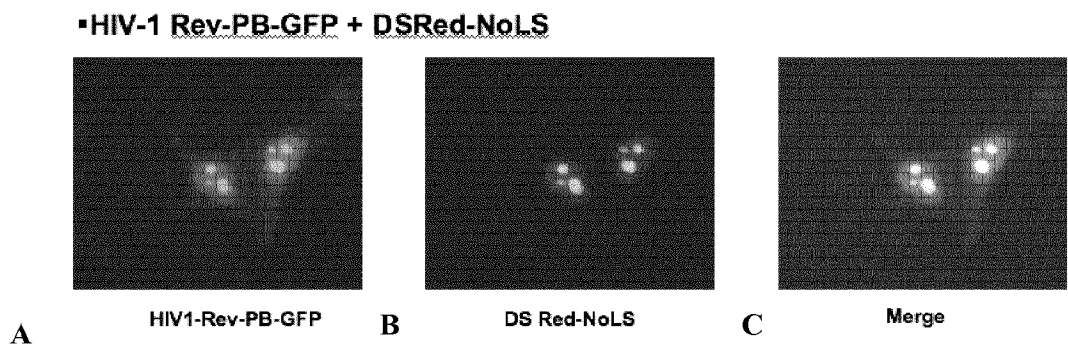


Fig. 4

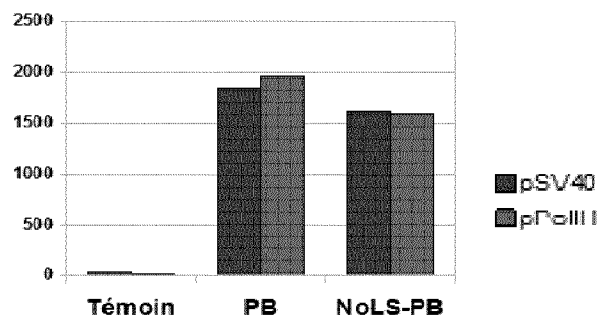


Fig. 5

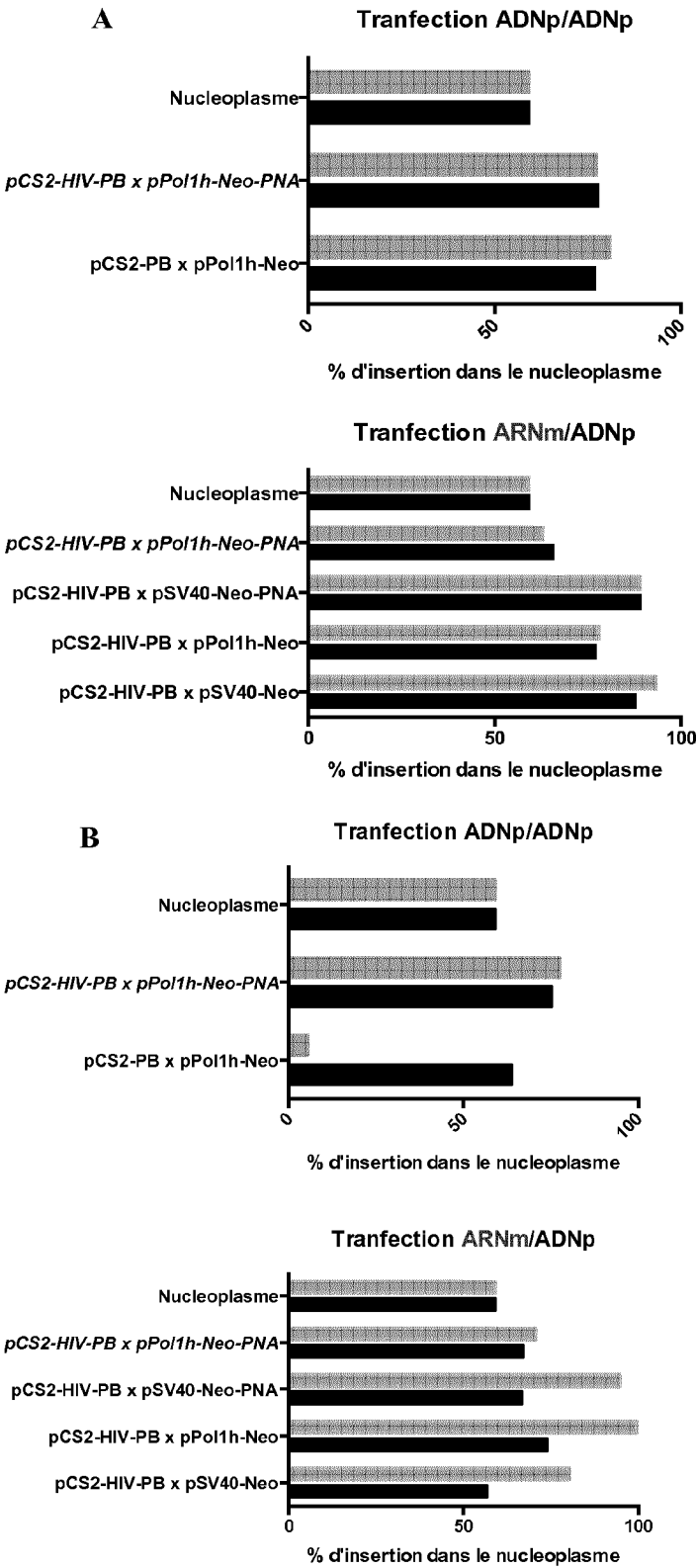


Fig. 6

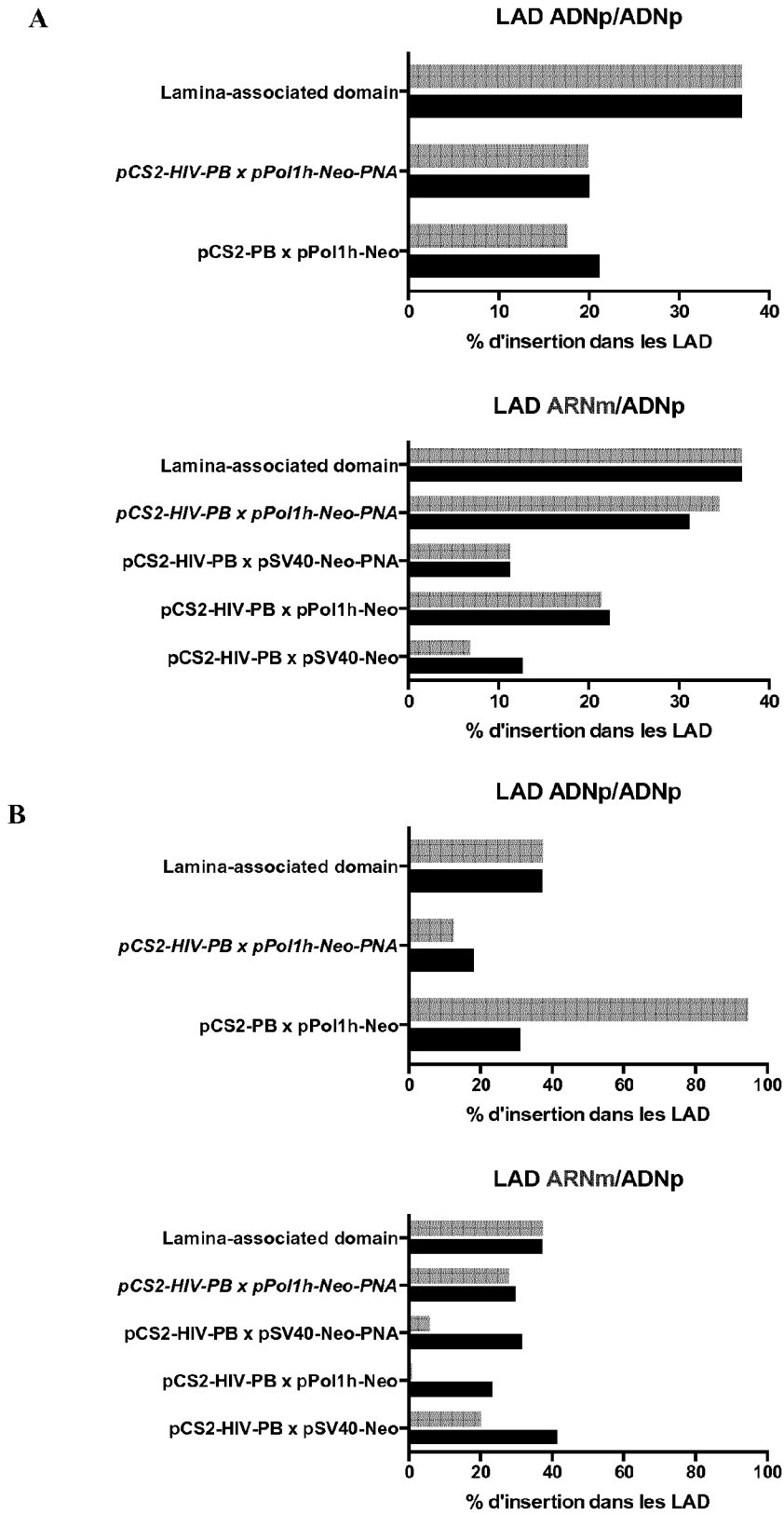


Fig. 7

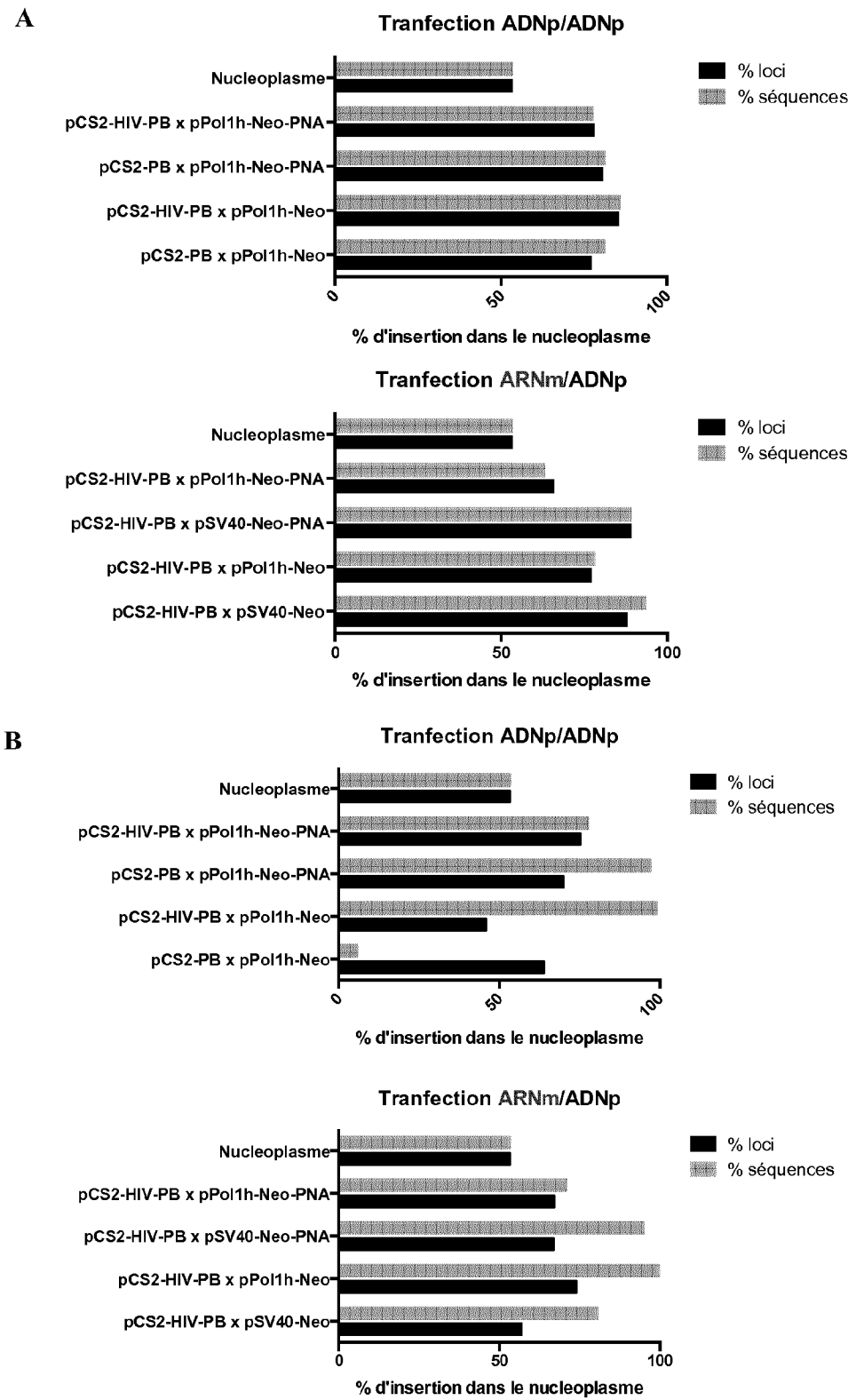


Fig. 8

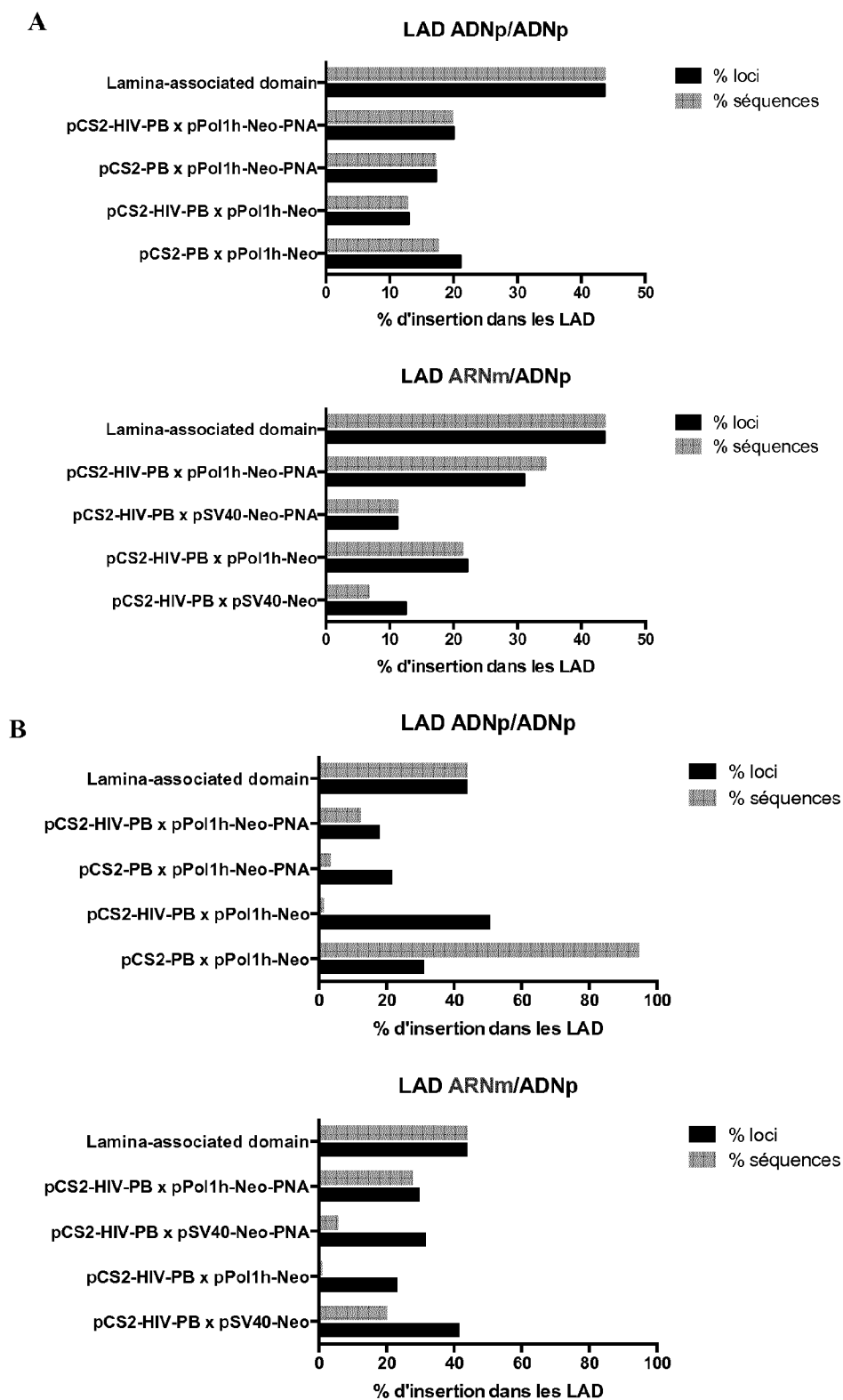
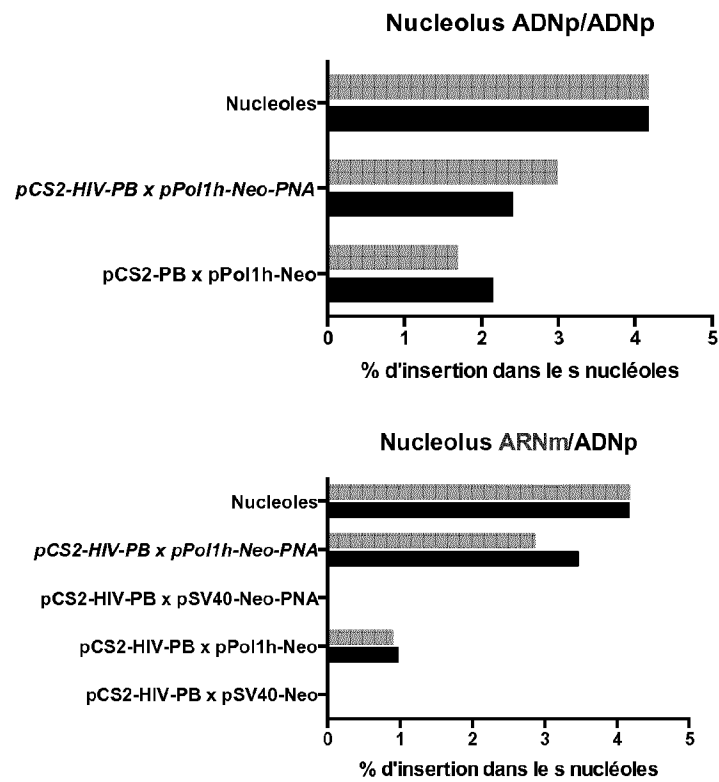


Fig. 9

A



B

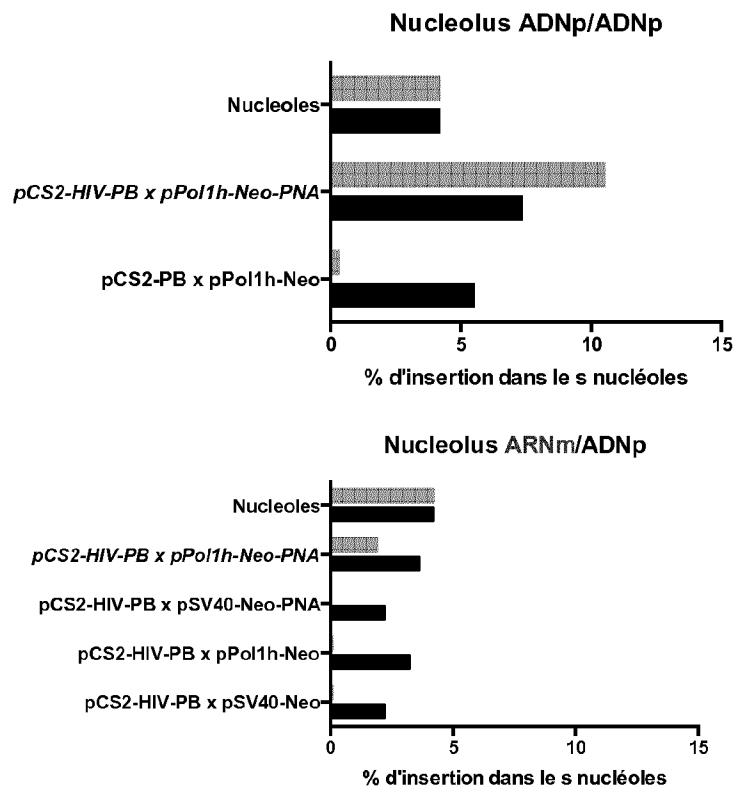
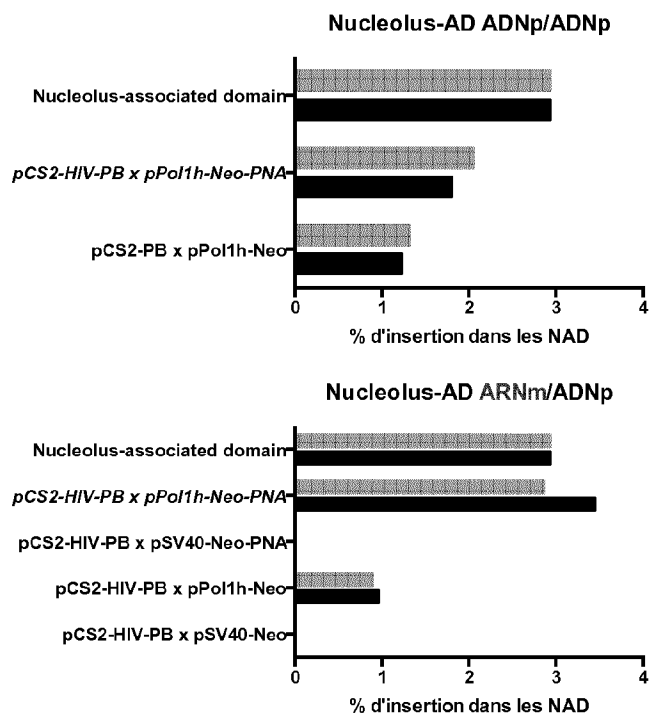


Fig. 10

A



B

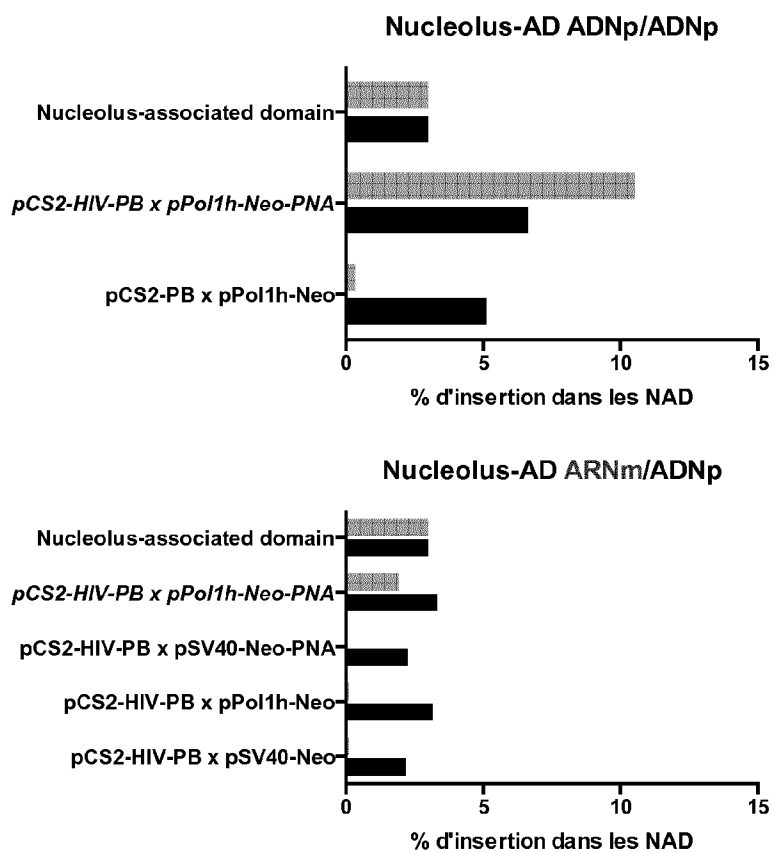


Fig. 11

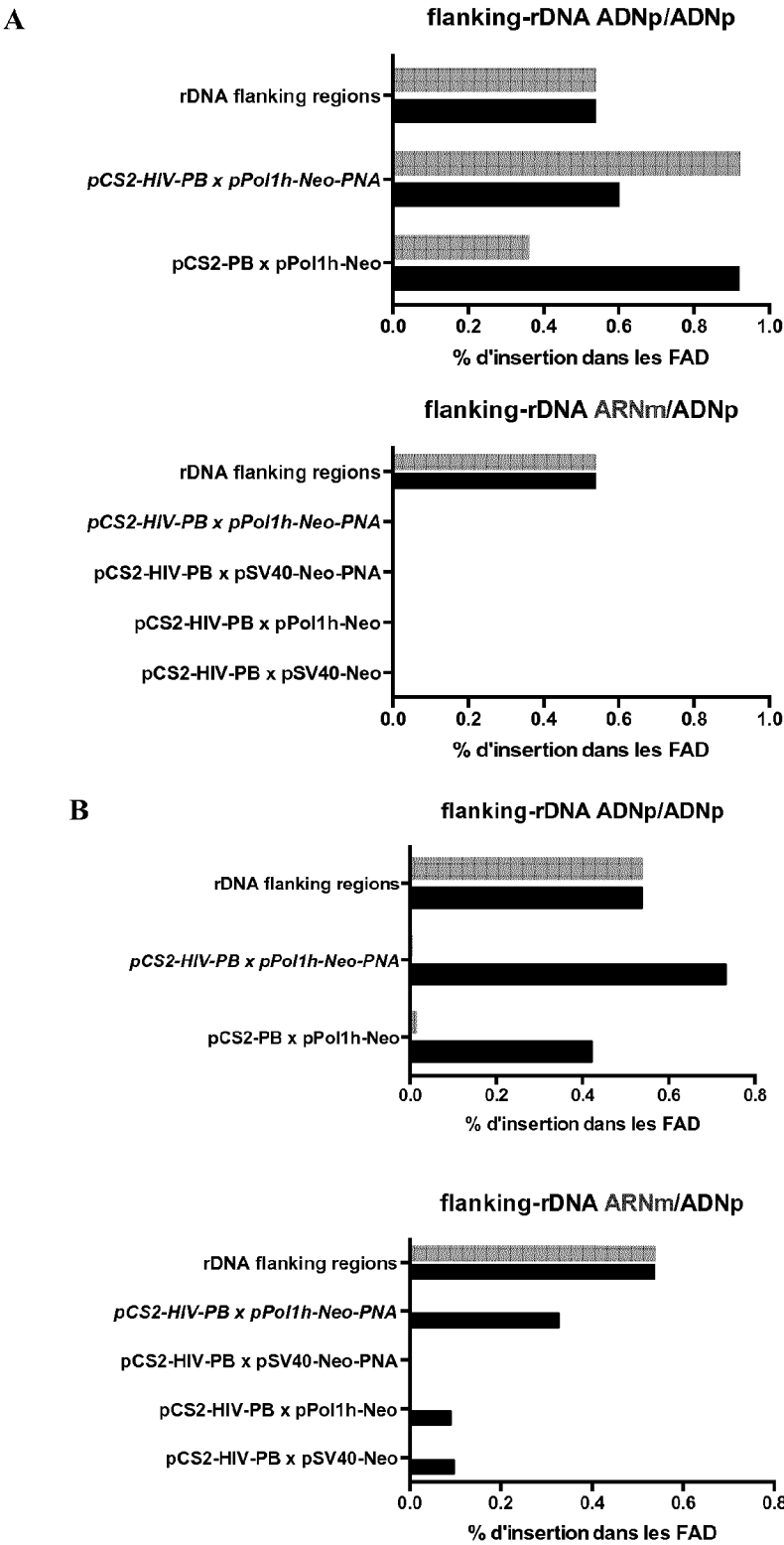
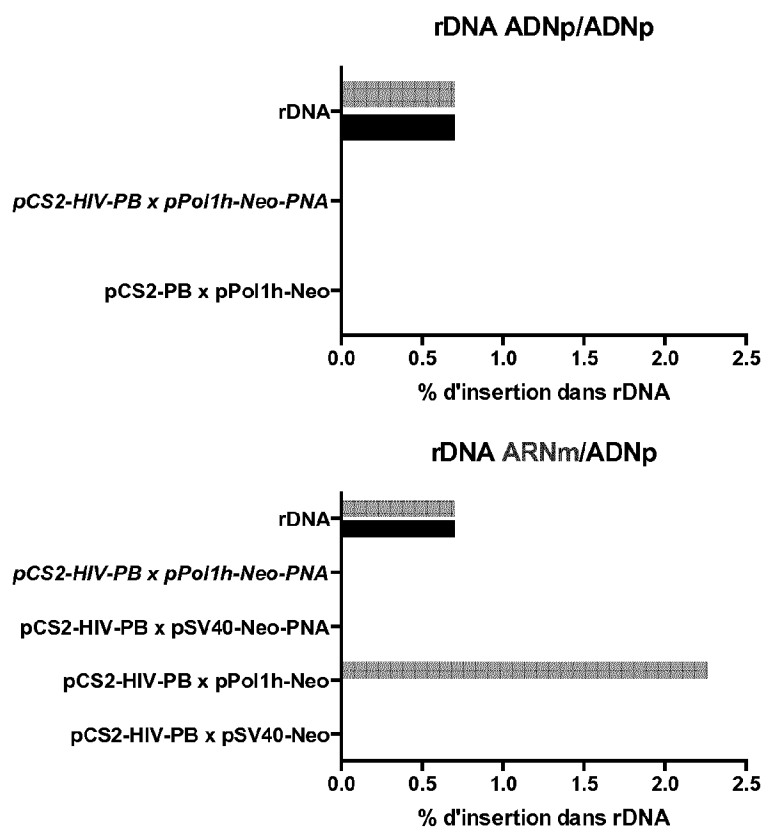


Fig. 12

A



B

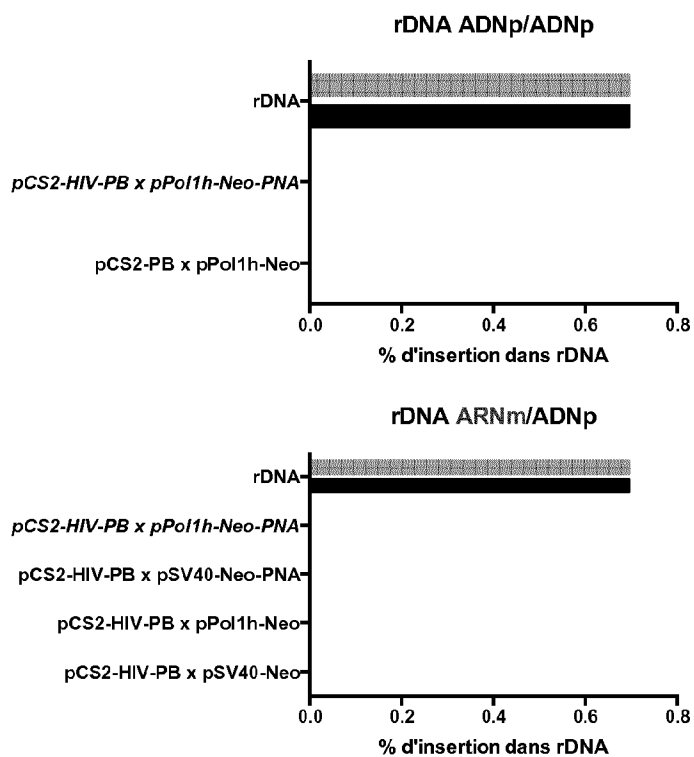


Fig. 13

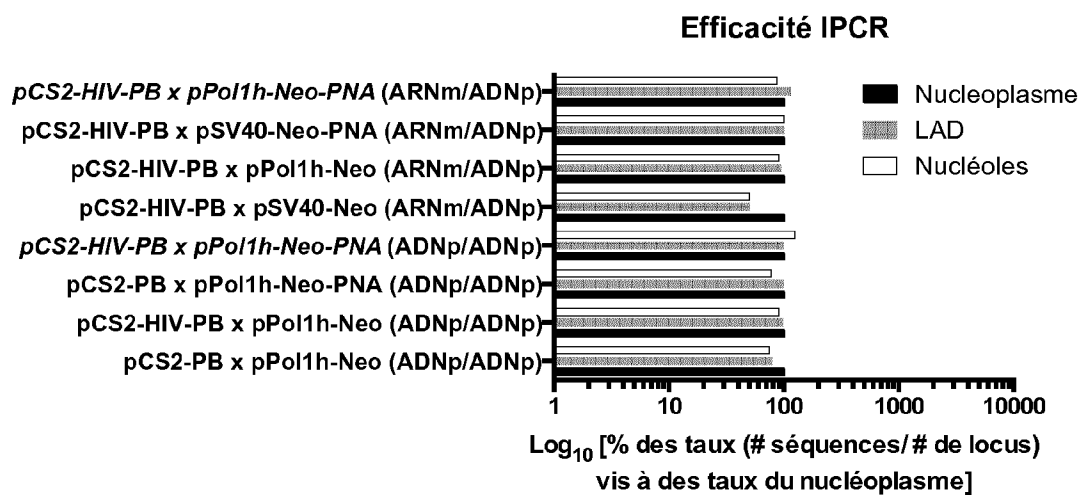


Fig. 14

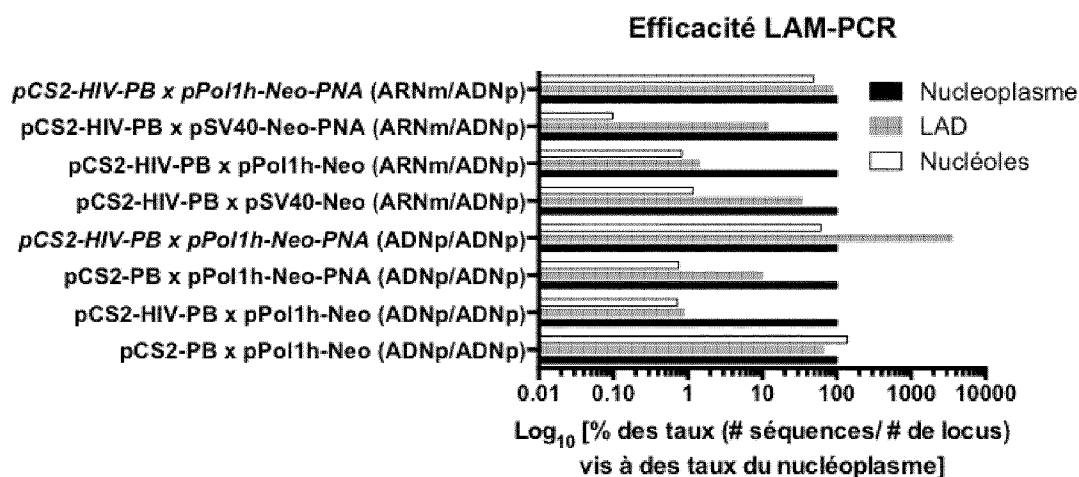


Fig. 15

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/067472

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N9/22 C12N15/90
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N A01K A61K C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EP0-Internal, Sequence Search, WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2004/069995 A2 (MAX DELBRUECK CENTRUM [DE]; IVICS ZOLTAN [DE]; IZSVAK ZSUZSANNA [DE]) 19 August 2004 (2004-08-19)	1-5,8, 11-15
Y	page 14, paragraph 2; figure 1c ----- -/--	9



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

28 September 2015

Date of mailing of the international search report

09/10/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Lonnoy, Olivier

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/067472

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WEN ET AL: "Efficient protein expression from the endogenous RNA polymerase I promoter using a human ribosomal DNA targeting vector", BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 367, no. 4, 14 January 2008 (2008-01-14), pages 846-851, XP022449883, ISSN: 0006-291X, DOI: 10.1016/J.BBRC.2008.01.018 page 850, column 2, paragraph 2; figures 3,4 -----	12-15
X	WO 94/24300 A1 (TRANSGENE SA [FR]; JACOBS ERIC [FR]) 27 October 1994 (1994-10-27) claims 20-22 -----	12-15
Y	WO 2008/060483 A2 (CELLUMEN INC [US]; TAYLOR LANSING D [US]; GIULIANO KENNETH A [US]; PRE) 22 May 2008 (2008-05-22) figure 15 -----	9
Y	-& DATABASE Geneseq [Online] 24 July 2008 (2008-07-24), "Human immunodeficiency virus Rev protein localization domain, SEQ ID 4.", XP002740862, retrieved from EBI accession no. GSP:ARW31639 Database accession no. ARW31639 the whole document -----	9
A	JIN-BON HONG ET AL: "A Nucleolus-Predominant piggyBac Transposase, NP-mPB, Mediates Elevated Transposition Efficiency in Mammalian Cells", PLOS ONE, vol. 9, no. 2, 24 February 2014 (2014-02-24), page e89396, XP055195631, DOI: 10.1371/journal.pone.0089396 abstract -----	1-15
A	BIRE SOLENNE ET AL: "Transposase concentration controls transposition activity: Myth or reality?", GENE, vol. 530, no. 2, 10 November 2013 (2013-11-10), pages 165-171, XP028728402, ISSN: 0378-1119, DOI: 10.1016/J.GENE.2013.08.039 abstract -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2015/067472

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2004069995 A2	19-08-2004	AT 497540 T DK 1594971 T3 EP 1594971 A2 JP 2006517104 A US 2006236413 A1 WO 2004069995 A2	15-02-2011 23-05-2011 16-11-2005 20-07-2006 19-10-2006 19-08-2004
WO 9424300 A1	27-10-1994	AU 686156 B2 AU 6571994 A CA 2160697 A1 EP 0694072 A1 FR 2703996 A1 JP H08508878 A US 6346414 B1 WO 9424300 A1	05-02-1998 08-11-1994 27-10-1994 31-01-1996 21-10-1994 24-09-1996 12-02-2002 27-10-1994
WO 2008060483 A2	22-05-2008	EP 2095119 A2 US 2010112602 A1 WO 2008060483 A2	02-09-2009 06-05-2010 22-05-2008

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°
PCT/EP2015/067472

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
INV. C12N9/22 C12N15/90
ADD.

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
C12N A01K A61K C07K

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)
EPO-Internal, Sequence Search, WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	WO 2004/069995 A2 (MAX DELBRUECK CENTRUM [DE]; IVICS ZOLTAN [DE]; IZSVAK ZSUZSANNA [DE]) 19 août 2004 (2004-08-19)	1-5,8, 11-15
Y	page 14, alinéa 2; figure 1c -----	9
X	WEN ET AL: "Efficient protein expression from the endogenous RNA polymerase I promoter using a human ribosomal DNA targeting vector", BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 367, no. 4, 14 janvier 2008 (2008-01-14), pages 846-851, XP022449883, ISSN: 0006-291X, DOI: 10.1016/J.BBRC.2008.01.018 page 850, colonne 2, alinéa 2; figures 3,4 ----- -/--	12-15

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date de dépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

28 septembre 2015

Date d'expédition du présent rapport de recherche internationale

09/10/2015

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Lonnoy, Olivier

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°

PCT/EP2015/067472

C(suite). DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	WO 94/24300 A1 (TRANSGENE SA [FR]; JACOBS ERIC [FR]) 27 octobre 1994 (1994-10-27) revendications 20-22 -----	12-15
Y	WO 2008/060483 A2 (CELLUMEN INC [US]; TAYLOR LANSING D [US]; GIULIANO KENNETH A [US]; PRE) 22 mai 2008 (2008-05-22) figure 15	9
Y	-& DATABASE Geneseq [Online] 24 juillet 2008 (2008-07-24), "Human immunodeficiency virus Rev protein localization domain, SEQ ID 4.", XP002740862, extrait de EBI accession no. GSP:ARW31639 Database accession no. ARW31639 le document en entier -----	9
A	JIN-BON HONG ET AL: "A Nucleolus-Predominant piggyBac Transposase, NP-mPB, Mediates Elevated Transposition Efficiency in Mammalian Cells", PLOS ONE, vol. 9, no. 2, 24 février 2014 (2014-02-24), page e89396, XP055195631, DOI: 10.1371/journal.pone.0089396 abrégé -----	1-15
A	BIRE SOLENNE ET AL: "Transposase concentration controls transposition activity: Myth or reality?", GENE, vol. 530, no. 2, 10 novembre 2013 (2013-11-10), pages 165-171, XP028728402, ISSN: 0378-1119, DOI: 10.1016/J.GENE.2013.08.039 abrégé -----	1-15

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale n°

PCT/EP2015/067472

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 2004069995 A2	19-08-2004	AT 497540 T DK 1594971 T3 EP 1594971 A2 JP 2006517104 A US 2006236413 A1 WO 2004069995 A2	15-02-2011 23-05-2011 16-11-2005 20-07-2006 19-10-2006 19-08-2004
WO 9424300 A1	27-10-1994	AU 686156 B2 AU 6571994 A CA 2160697 A1 EP 0694072 A1 FR 2703996 A1 JP H08508878 A US 6346414 B1 WO 9424300 A1	05-02-1998 08-11-1994 27-10-1994 31-01-1996 21-10-1994 24-09-1996 12-02-2002 27-10-1994
WO 2008060483 A2	22-05-2008	EP 2095119 A2 US 2010112602 A1 WO 2008060483 A2	02-09-2009 06-05-2010 22-05-2008