



Développement de nouvelles approches d'édition du génome à l'aide de nucléases artificielles (TALENs et CRISPR/Cas9)

Marine Charpentier

► To cite this version:

Marine Charpentier. Développement de nouvelles approches d'édition du génome à l'aide de nucléases artificielles (TALENs et CRISPR/Cas9). Génétique. École pratique des hautes études - EPHE PARIS, 2016. Français. NNT : 2016EPHE3106 . tel-02105233

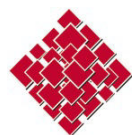
HAL Id: tel-02105233

<https://theses.hal.science/tel-02105233>

Submitted on 20 Apr 2019

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

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



École Pratique
des Hautes Études

Mention « Systèmes intégrés, environnement et biodiversité »

École doctorale de l'École Pratique des Hautes Études
[CNRS UMR 7196 – INSERM U1154 / Équipe GE2R]

Développement de nouvelles approches d'édition du génome à l'aide de nucléases artificielles (TALENs et CRISPR/Cas9)

Par Marine CHARPENTIER

Thèse de doctorat de Systèmes intégrés, environnement et biodiversité

Sous la direction de M. Christophe TERZIAN et M. Andràs PALDI
Sous l'encadrement de M. Jean-Paul CONCORDET

Soutenue le 19 décembre 2016

Devant un jury composé de :

Mme. Anne PLESSIS, Professeur d'université à Paris Diderot, Paris
Mme. Pascale BERTRAND, Chargé de recherche au CEA, Fontenay-aux-Roses
M. Fabien NOGUE, Chargé de recherche à l'INRA, Versailles
M. Andràs PALDI, Directeur d'études EPHE, Paris
M. Jean-Paul CONCORDET, Chargé de recherche au MNHN, Paris
M. Hervé TOSTIVINT, Professeur au MNHN, Paris

Remerciements

Je remercie tout d'abord Mme Anne Plessis et Mme Pascale Bertrand d'avoir accepté d'être rapporteurs de cette thèse. Merci également à M Fabien Nogué et M Hervé Tostivint d'avoir accepté d'être membres de ce jury et d'évaluer mon travail.

Je remercie le Docteur Carine Giovannangeli et le Professeur Jean-François Riou de m'avoir donné l'opportunité de travailler dans leur unité et leur équipe, de m'avoir fait confiance et de m'avoir aidé et donné des conseils pendant ces années.

Je remercie tout particulièrement le Docteur Jean-Paul Concordet de m'avoir permis de travailler à ses côtés sur ce beau sujet qu'est l'édition du génome. La confiance qu'il m'a accordée ainsi que les nombreuses discussions m'ont permis d'avancer et de m'épanouir.

Je tiens aussi à remercier le Docteur Christophe Terzian sans qui cette thèse n'aurait pu être possible, merci à lui pour ses encouragements, sa gentillesse et son enthousiasme. Vous serez manqué.

Je remercie aussi le Docteur Andràs Paldi pour s'être occupé de moi ces derniers mois et de m'avoir permis de soutenir cette thèse.

Je ne remerciais jamais assez toutes les personnes que j'ai eu la chance de côtoyer pendant ces quelques années : Charlotte, Armêl, Karine, Hind, Loic, Gildas, Anthony, Anne, Erika, Jean-Baptiste, François... Merci à tous pour vos sourires, vos conseils, votre présence et vos encouragements !

Merci à l'équipe dont j'ai déjà cité quelques noms. Merci pour les réunions du jeudi, pour vos conseils et votre bonne humeur.

Des remerciements tout particuliers à Armêl, Charlotte et Karine. Trois personnes au grand cœur qui sont devenues de véritables amies ! Cette thèse n'aurait certainement pas été la même sans vous ! Merci pour vos encouragements, votre connerie, les pauses goûter, les danses dans le L2, la déco du bureau, les weekend boulot ou sorties... J'en passe, j'en oublie et il me faudrait rédiger quelques dizaines de pages pour vous dire à quel point je vous suis reconnaissante pour tout ce que vous avez fait pour moi. BIA***

Je tiens aussi à remercier ma famille, pour m'avoir soutenue pendant cette thèse. Merci à Lilou et Tissou pour leurs câlins, leurs vidéos de bébés chiens, les photos de Brisca et Nala mes bébés et les messages d'amour, c'est un bonheur d'avoir un frère et une sœur aussi géniaux. Je suis tellement fière de vous ! Merci à mes parents de toujours avoir été là pour moi, de m'avoir supportée les weekend quand je venais reprendre mon souffle en Bretagne, merci de m'avoir accueillie pour le début de ma rédaction et merci pour tout, tout, tout et tout, vous êtes des parents formidables et c'est une chance indescriptible de vous avoir. Sans vous rien de tout ce que j'ai fait n'aurait été possible.

Merci aussi à ma famille au sens plus large, oncles, tantes, cousins... Une famille aussi soudée et présente c'est de l'amour constamment et ça permet d'avancer. Mention spéciale à mes grands-parents, Paulette et Jeannot vous être en or, ainsi qu'à Valérie, ma reine.

Je voudrais terminer par remercier en vrac quelques petites choses qui m'ont permis de tenir en cas de coup dur : le chocolat, les Pringles, les pizza, l'ibuprofène, les chiens, Daryl, Tony, Josh, et Jon.

Je dédie cette thèse à ma famille

Abréviations

53BP1	<i>p53-binding protein 1</i>
APLF	<i>Aprataxin and PNKP Like Factor</i>
ARNg	<i>ARN guide</i>
ATM	<i>Ataxia Telangectasia Mutated</i>
ATR	<i>ATM- and Rad3-related</i>
ATRIP	<i>ATR-Interacting Protein</i>
BER	<i>Base Excision Repair</i>
BLM	<i>Bloom</i>
BRCA1	<i>Breast Cancer 1</i>
BRCA2	<i>Breast Cancer 2</i>
Cas9	<i>CRISPR associated protein 9</i>
CDB	<i>Cassure Double Brin</i>
CDK	<i>Cyclin-dependent kinase</i>
Chk2	<i>Checkpoint 2</i>
Cpf1	<i>Clustered Regularly Interspaced Short Palindromic Repeats from Prevotella and Francisella 1</i>
CRISPR	<i>Clustered Regularly Interspaced Palindromic Repeats</i>
CtIP	<i>C-terminal-binding protein interacting protein</i>
DMD	<i>Dystrophie Musculaire de Duchenne</i>
DNA-PKcs	<i>DNA-dependent Protein Kinase catalytic subunit</i>
DRGFP	<i>Direct Repeat Green Fluorescent Protein</i>
ERCC1	<i>Excision Repair Cross Complementation group 1</i>
ESC	<i>Embryonic Stem Cell</i>
EXO1	<i>Exonuclease1</i>
GFP	<i>Green Fluorescent Protein</i>

HE	<i>Homing Endonuclease</i>
hSSB1	<i>human Single Strand Binding protein1</i>
hPSC	<i>human Pluripotent Stem Cells</i>
iPSC	<i>induced Pluripotent Stem Cells</i>
KI	<i>Knock In</i>
KO	<i>Knock Out</i>
KRAB	<i>Krüppel-associated box</i>
MDC1	<i>Mediator of DNA damage Checkpoint protein 1</i>
MMEJ	<i>Micro-homology mediated end joining</i>
MMR	<i>mismatch repair</i>
MRN	<i>Mre11-Rad50-Nbs1</i>
Nbs1	<i>Nibrin</i>
NER	<i>Nucleotide Excision Repair</i>
NHEJ	<i>Non Homologous End Joining</i>
NLS	<i>Nuclear Localization Signal</i>
PAM	<i>Protospacer Adjacent Motif</i>
PARP	<i>Poly(ADP-ribose) Polymerase</i>
PAXX	<i>PAralog of XRCC4 and XLF</i>
PIKK	<i>Phosphatidylinositol 3-Kinase-like protein Kinase</i>
RDA	<i>Réponse aux Dommages de l'ADN</i>
RH	<i>Recombinaison Homologue</i>
RNF	<i>Ring Finger protein</i>
RPA	<i>Replication Protein A</i>
TALE	<i>Transcription Activator Like Effector</i>
TALEN	<i>TALE Nuclease</i>
UCHL5	<i>Ubiquitine thioesterase L5</i>
USP4	<i>Ubiquitin Specific Protease 4</i>
VIH	<i>Virus de l'Immunodéficience Humaine</i>
XLF	<i>XRCC4-Like Factor</i>

XPF	<i>Xeroderma Pigmentosum complementation group F</i>
XRCC4	<i>X-ray Repair Cross-Complementing Protein 4</i>
ZFN	<i>Zinc Finger Nuclease</i>

Table des matières

1	INTRODUCTION	12
1.1	Principe et applications de l'édition du génome	13
1.1.1	Applications de l'édition du génome	15
1.2	Les nucléases artificielles	21
1.2.1	Historique	21
1.2.2	Les nucléases à doigts de zinc	22
1.2.3	Les TALE nucléases	23
1.2.4	Le système CRISPR/Cas9	26
1.2.5	Nouveaux types de nucléases artificielles utilisés	30
1.2.6	Difficultés liées à l'utilisation des nucléases	31
1.3	Les systèmes de réparation des CDB	35
1.3.1	Les cassures double-brin de l'ADN	35
1.3.2	Les systèmes de réparation des cassures double brin	36
1.3.3	Régulation de la réparation des cassures double brin par NHEJ et RH	47
1.3.4	53BP1 et BRCA1, deux protéines clés dans la réparation des cassures double brin	50
1.4	Optimisation des approches d'édition du génome	54
1.4.1	Approches développées pour augmenter le taux de mutations	54
1.4.2	Approches développées pour augmenter le taux de recombinaison homologue	55
1.4.3	Bilan	62
2	PROJET DE THESE ET RESULTATS	64
2.1	Article : Stimulation du taux d'insertion d'un ADN exogène en fusionnant la partie Nter de la protéine CtIP à la nucléase cas9.	65
2.2	Stimulation du taux de mutations en fusionnant l'exonucléase Trex2 à la nucléase Cas9	67
2.2.1	Surexpression et fusion de l'exonucléase Trex2 à la nucléase Cas9	67
		8

2.2.2	Recrutement de l'exonucléase Trex2 à la cassure double brin par fusion à la nickase Cas9 (D10A) 69	
2.2.3	Conclusion et discussion	74
2.3	Assemblage de TALEN pour l'édition du génome	78
2.3.1	Méthode d'assemblage des répétitions TALE	78
2.3.2	Collaborations	83
2.3.3	Conclusion	97
3	DISCUSSION ET PERSPECTIVES	99
3.1.1	Mécanismes d'action possibles du domaine L2	101
3.1.2	Conclusion	108
3.1.3	Effet d'autres protéines sur le taux de KI	111
4	ANNEXE	116
4.1.1	Séquence du plasmide pSKduo contenant la répétition HD seulement	117
4.1.2	Banque des tétramères de répétitions TALE	119
4.1.3	Séquence du plasmide pCS2 contenant la demi-répétition NG	120
4.1.4	Article : Edition du génome de rat par Knock-In	123
4.1.5	Article : Edition du génome de poisson zèbre par Knock-Out	137
5	BIBLIOGRAPHIE	146

Tables des figures

Figure 1. Principe de l'édition du génome.....	14
Figure 2. Nucléases à doigts de zinc ou Zinc Finger Nucleases (ZFN).	23
Figure 3. Transcription Activator-Like Effector (TALE) (figure adaptée de Boch et al., 2009).....	24
Figure 4. TALE nucléases (TALEN).....	25
Figure 5. Fonctionnement du système CRISPR/Cas naturel (Adapté de Mali et al., 2013).....	27
Figure 6. Système CRISPR-Cas9 (Figure adaptée de Jinek et al., 2012).....	29
Figure 7. Cpf1 (Clustered Regularly Interspaced Short Palindromic Repeats from Prevotella and Francisella 1).	30
Figure 8. Réparation des cassures double brin par NHEJ.....	38
Figure 9. Réparation des cassures double brin par RH.	41
Figure 10. Les 5 phases du cycle cellulaire.....	47
Figure 11. Régulation des systèmes de réparation selon le cycle cellulaire.	50
Figure 12. Effet de l'exonucléase Trex2 sur le taux de mutations après coupure par le système CRISPR/Cas9.	68
Figure 13. Effet de la fusion de l'exonucléase Trex2 à la nucléase Cas9(D10A) sur le taux de mutations dans les cellules U2OS.	70
Figure 14. Exemple de gel agarose (2,5%) obtenu après transfection d'U2OS avec Cas9 nickase fusionnée à Trex2 ou Trex2 muté et test T7E1.	71
Figure 15. Effet de l'exonucléase Trex2 sur le taux de mutations après coupure par le système CRISPR/Cas9(D10A) dans les cellules RG37DR.	73
Figure 16. Exemple de gel de test T7E1 obtenu après transfection de RG37DR avec Cas9 nickase fusionnée à Trex2 ou Trex2 muté.	74
Figure 17. Assemblage des extrémités des enzymes compatibles NheI et SpeI. Après digestion, les extrémités produites par les enzymes sont compatibles et permettent une ligation faisant disparaître les sites de restriction de ces enzymes.....	79
Figure 18. Etapes d'assemblage des répétitions avec le Unit Assembly.	80
Figure 19. Digestions nécessaires au protocole optimisé d'assemblage des répétitions TALE.....	80
Figure 20. Principe d'assemblage des répétitions TALE.	81
Figure 21. Principe d'assemblage d'une TALEN.....	82

<i>Figure 22. Comparaison de l'efficacité des différentes nucléase au site DUX4 dans les cellules U2OS pour inactiver le site de polyadénylation (en rouge).</i>	<i>96</i>
<i>Figure 23. Composition du domaine L2 de la protéine CtIP.....</i>	<i>100</i>
<i>Figure 24. Taux relatif d'insertion par KI d'un donneur GFP exogène au site AAVS1 obtenu dans des HEK293 à l'aide du système CRISPR/Cas9 et de l'ARNg T2.....</i>	<i>102</i>
<i>Figure 25. Effet de l'inhibition du protéasome sur l'effet de la Cas9-L2 sur le taux de mutations et le taux d'insertion par homologie d'un donneur exogène exprimant la GFP dans des HEK293.....</i>	<i>106</i>
<i>Figure 26. Western blot après transfection dans des HEK293 de la Cas9 wt et de la Cas9-L2 avec inhibition du protéasome par différentes doses de MG132.</i>	<i>107</i>
<i>Figure 27. Modèle de recrutement de CtIP par le domaine L2 de CtIP fusionné à la Cas9 wt permettant de favoriser la réparation des CDB par la RH.....</i>	<i>110</i>
<i>Figure 28. Effet de la surexpression ou du recrutement à la coupure de la protéine Rad18 sur le taux de mutations et le taux de KI.....</i>	<i>112</i>
<i>Figure 29. Effet de la surexpression ou du recrutement à la coupure de la protéine Nbs1.....</i>	<i>114</i>

1 Introduction

1.1 Principe et applications de l'édition du génome

Ma thèse a porté sur l'optimisation des techniques d'édition du génome. L'édition du génome repose sur l'utilisation de nucléases artificielles pour introduire une cassure double brin (CDB) à une position précise du génome et permettre, au cours de sa réparation, d'introduire des modifications de la séquence à cet endroit. Les modifications peuvent être des insertions ou délétions aléatoires de petite taille ou encore l'insertion précise d'une séquence d'intérêt, comme un transgène. Ces différentes modifications sont dépendantes du système de réparation choisi par la cellule pour réparer les CDB induites par les nucléases artificielles. Deux de ces systèmes sont principalement utilisés et permettent d'établir les grands principes de l'édition du génome (Figure 1) : le *Non Homologous End Joining* (NHEJ), qui permet une ligation directe des extrémités de la CDB induisant de petites mutations aléatoires si les extrémités ont été modifiées avant réparation et la Recombinaison Homologue (RH), qui elle permet de réparer la CDB en copiant les informations à partir d'un ADN homologue à la séquence à réparer. Dans ce dernier cas la cellule peut utiliser la chromatide sœur ou un ADN exogène permettant aussi d'insérer toute mutation de manière précise.

L'objectif a été ici de développer différentes approches pour augmenter le taux de mutations au site de coupure des nucléases ou bien pour augmenter le taux d'insertion précise d'un transgène. Je vais tout d'abord présenter dans cette introduction, le principe général de l'édition du génome ainsi que ses différentes applications. Ensuite je présenterai les différentes nucléases artificielles qui peuvent être utilisées, puis les différents systèmes de réparation qui permettent de corriger les cassures infligées par ces nucléases. Enfin je présenterai les approches qui ont été développées jusqu'ici pour optimiser les techniques d'édition du génome.

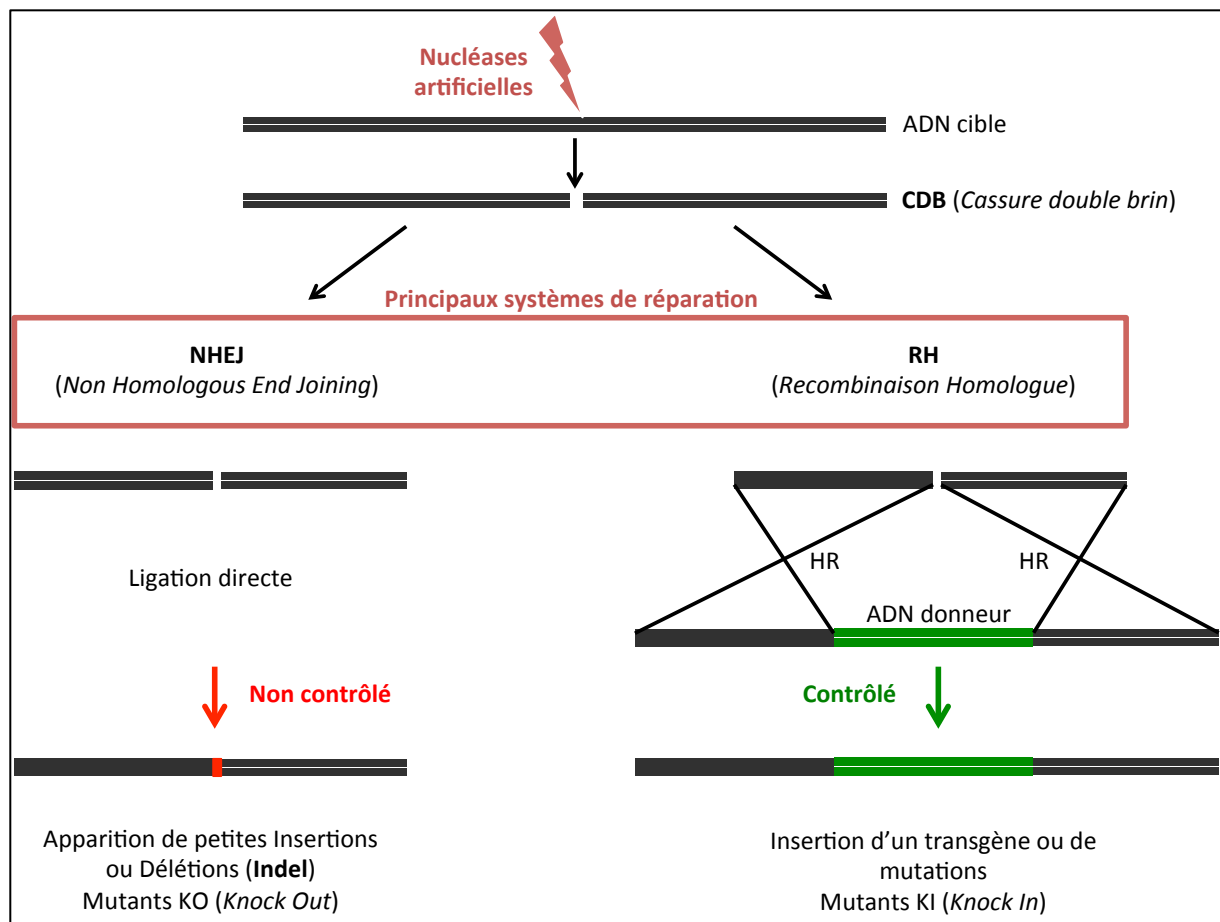


Figure 1. Principe de l'édition du génome.

Deux principaux systèmes de réparation, NHEJ ou RH, peuvent réparer les cassures double brin induites par les nucléases artificielles et être utilisés à notre avantage. Dans le cas de la réparation par NHEJ il est possible d'obtenir de petites mutations en recoupant plusieurs fois et d'induire un changement du cadre de lecture, en ciblant un gène, aboutissant à l'inactivation de celui-ci (KO). Dans le cas de la réparation par RH, la cellule utilise la chromatide sœur pour réparer l'ADN endommagé, mais si on insère un ADN donneur avec des bras d'homologie avec la séquence à réparer la cellule peut le prendre comme matrice de réparation. Il est ainsi possible d'insérer n'importe quelle modification ou transgène de manière précise (KI).

Dans la partie Résultats, je présenterai en premier lieu un manuscrit concernant l'augmentation du taux d'insertion précise à l'aide de la protéine CtIP. Nous verrons ensuite qu'il a été possible d'augmenter le taux de mutations engendré par les nucléases artificielles en utilisant une exonucléase, Trex2. Et pour finir cette partie, je décrirai une nouvelle technique d'assemblage des nucléases artificielles de type TALEN, qui ont été

utilisées pour différentes collaborations dont je vous présenterai rapidement les résultats.

1.1.1 Applications de l'édition du génome

Les techniques d'édition du génome ont évolué rapidement ces dernières années, permettant l'apparition de nombreuses applications et avancées dans différents domaines. Je vais ici présenter des applications possibles et des avancées qui ont été faites jusqu'à présent dans le domaine de la recherche fondamentale, de la recherche appliquée à la médecine ou dans l'industrie.

1.1.1.1 Applications en recherche fondamentale

En recherche fondamentale, les nouvelles techniques d'édition du génome représentent des outils absolument extraordinaires qui ont été très rapidement adoptés dans de très nombreux domaines de la biologie. Plusieurs applications sont plus particulièrement intéressantes, comme l'établissement de lignées cellulaires modifiées pour étudier la fonction des gènes ou l'utilisation des nucléases pour visualiser des séquences d'ADN ou moduler leur expression.

1.1.1.1.1 L'édition du génome permet l'étude de gènes

Grâce aux nucléases artificielles et à l'induction de mutations il est devenu simple d'inactiver spécifiquement un gène, dans une lignée cellulaire par exemple, et d'en étudier les effets. On peut aussi créer de nouvelles lignées cellulaires exprimant un transgène choisi. Un exemple de l'insertion d'une séquence peut être l'insertion d'une étiquette, ou « tag », à une protéine. Il sera alors assez simple de cibler ce tag à l'aide d'un anticorps spécifique et de voir évoluer la protéine d'intérêt (Li et al., 2015). Cet ajout de tag à la fin d'une protéine peut aussi servir à purifier ses partenaires.

Il est aussi possible d'utiliser ces approches sur un organisme entier (lapin, drosophile, souris, rat, cochon...), en inactivant le gène souhaité, ou en insérant un transgène, au stade œuf fécondé et en laissant l'organisme se développer. Dans ce cas, l'organisme entier contiendra la modification et il sera alors possible d'en étudier les effets sur l'animal (Guo et al., 2016; Kato et al., 2013; Lv et al., 2016; Nakagawa et al., 2016).

1.1.1.1.2 Les nucléases artificielles et leurs applications dans la visualisation de locus génomiques

Dans le cas où il est question de visualiser un locus génomique, ce n'est pas la capacité de clivage qui est intéressante chez les nucléases artificielles mais leur spécificité pour une séquence unique. On utilise alors leur domaine de reconnaissance de l'ADN, dépourvu de son domaine nucléase, qui sera fusionné à une protéine fluorescente par exemple, comme la GFP (*Green Fluorescent Protein*). Il est alors possible de visualiser des séquences répétées (100 à plusieurs milliers de répétitions) par microscopie, puisqu'il y aura accumulation de la protéine fluorescente à ce locus (Chen et al., 2013). La visualisation d'un locus unique reste, quant à elle, difficile à accomplir.

1.1.1.1.3 Les nucléases artificielles et leurs applications dans la régulation de l'expression des gènes

Il est possible de recruter à une séquence donnée, à l'aide des capacités de ciblage des nucléases artificielles, d'autres protéines que la GFP. Il est intéressant en effet, de recruter des protéines modifiant l'expression de gènes, comme un domaine activateur de la transcription tel que VP64 (VP64 : 4 répétitions du domaine VP16 du virus Herpes simplex) ou un domaine répresseur comme KRAB (*Krüppel-associated box*). Ces protéines, une fois recrutées au niveau du promoteur d'un gène, vont pouvoir moduler son expression (Agne et al., 2014; Chavez et al., 2015; Gao et al., 2014; Hu et al., 2014). Des ADN méthyltransférases peuvent aussi être recrutées de la même manière pour faire varier le taux de méthylation des histones à un locus précis (Agne et al., 2014).

1.1.1.2 Applications en recherche appliquée

Les nouvelles techniques d'édition du génome peuvent aussi trouver de nombreuses applications en recherche appliquée pouvant aider à comprendre ou trouver des solutions aux maladies humaines.

1.1.1.2.1 L'édition du génome permet la création de nouveaux modèles animaux

Une application très intéressante des techniques d'édition du génome est de pouvoir créer des modèles animaux plus simplement qu'avant, notamment pour étudier des maladies génétiques humaines. Le gène FH (Fumarate Hydratase) a notamment été muté chez le poisson zèbre pour l'étudier et ainsi mieux comprendre son implication dans la maladie humaine qu'est l'acidurie fumorique (Hwang et al., 2013). Cette maladie cause chez l'enfant le plus souvent une mort précoce ou, s'il survie, de nombreuses atteintes neurologiques.

En 2014, un nouveau modèle pour la myopathie de Duchenne a été créé, chez le rat (Larcher et al., 2014). En effet, différents modèles existaient jusqu'ici, chez le chien ou le cochon, des animaux modèles chers et difficiles à étudier, ou la souris, dont le phénotype est très atténué par rapport aux patients humains. Le rat reste un petit animal, 10 fois plus gros que la souris et dont le phénotype, après mutation du gène DMD, est proche de celui observé chez l'homme, ce qui en fait un modèle très pertinent et prometteur.

1.1.1.2.2 L'édition du génome permet la correction de mutations dans des cellules de patients

Il est aussi possible de modifier un génome pour corriger une mutation afin de combattre ou guérir une maladie génétique. Ce type d'approche a pour le moment été principalement mené dans des cellules en culture, dont certaines issues de prélèvements de patients. Dans ce cas la réparation se doit d'être précise. On introduit donc un ADN donneur portant la correction souhaitée pour obtenir une modification précise et programmée par RH.

La mutation $\Delta F408$ du gène CFTR (*Cystic fibrosis transmembrane conductance regulator*) cause une fibrose kystique diminuant drastiquement l'espérance de vie des personnes atteintes. Cette mutation, retrouvée chez plus de 80% des patients a pu être corrigée dans des cellules souches de patients (Schwank et al., 2013). Le gène de la dystrophine, inactivé dans la Dystrophie Musculaire de Duchenne (DMD) a quant à lui été corrigé dans des iPSC (*induced Pluripotent Stem Cells*) dérivées de fibroblastes de patients (Li et al., 2015; Ousterout et al., 2015; Wojtal et al., 2016).

Ce type d'exemples laisse espérer qu'un jour il sera possible de réinjecter ces cellules corrigées et obtenir une amélioration clinique chez les patients.

1.1.1.2.3 L'édition du génome fera partie des futures thérapies

S'il est possible de modifier des cellules extraites de patients pour corriger les mutations responsables de la maladie, avant de pouvoir réinjecter ces cellules aux patients, plusieurs étapes restent à franchir. Il faut en effet vérifier la faisabilité chez l'animal et entrer en essai clinique chez l'homme. La thérapie génique ou cellulaire par édition du génome serait un progrès important pour la médecine.

L'hémophilie A, par exemple, pourrait trouver son traitement dans l'édition du génome puisqu'en corrigeant le gène F8 muté chez les patients, il serait possible de rétablir l'expression du facteur VIII et de les guérir. Cela a été réalisé chez la souris. Des cellules ont été extraites de patients humains, transformées en iPSCs et à l'aide de nucléases artificielles, entre 4 et 7% de ces cellules ont été corrigées. Les cellules ont ensuite été différenciées puis injectées dans des souris atteinte d'hémophilie A pour observer la sécrétion du facteur VIII et l'amélioration de l'état des animaux (Park et al., 2015), démontrant ainsi que, dans ce modèle, même un taux relativement faible de correction peut suffire à améliorer la tableau clinique.

Un autre exemple marquant est le début des premiers essais cliniques reposant sur l'édition du génome comme approche thérapeutique dans le traitement du VIH (Virus d'immunodéficience Humaine). Les nucléases peuvent en effet être utilisées pour modifier des lymphocytes T CD4+ de patients atteints du VIH dans le but d'inactiver le

récepteur CCR5 (*C-C chemokine receptor type 5*), nécessaire à l'entrée du virus dans les cellules, permettant ainsi de rendre les cellules résistantes. La réinjection de ces cellules chez le patient pourrait contribuer au rétablissement de ses défenses immunitaires (Tebas et al., 2014).

Une autre application médicale de l'édition du génome pourrait jouer un rôle majeur, elle concerne la xénotransplantation. En effet, il est connu depuis longtemps que la morphologie et la physiologie de plusieurs organes du porc sont très semblables aux nôtres. L'inactivation de plusieurs antigènes responsables du rejet des xénogreffes pourrait relancer les études sur les xénogreffes. Dans une première étape, le gène GGTA1 (*glycoprotein, alpha-galactosyltransferase 1*) a été inactivé dans des fibroblastes de porcs pour créer des porcs GGTA1 KO par transfert de noyaux de cellules somatiques (Xin et al., 2013). Des porcelets GGTA1 KO ont ainsi été obtenus, dont les fibroblastes ne déclenchent pas d'attaque immunitaire quand ils sont mis en contact avec du sérum humain.

1.1.1.3 Place de l'édition du génome dans l'industrie

De nombreuses applications de l'édition du génome sont en cours dans le monde de l'industrie agroalimentaire, notamment pour augmenter les capacités nutritionnelles ou la résistance aux stress des principales céréales à la base de l'alimentation dans le monde. Les revues de Kamthan et de Sovová donnent une vue assez globale des avancées des dernières années dans ce domaine (Kamthan et al., 2016; Sovová et al., 2016). Les approches utilisant les nucléases artificielles chez les plantes ont débuté assez rapidement après leur découverte. Différents gènes de la famille DCL (*Dicer like*) ont par exemple été inactivés (Curtin et al., 2011). Ces gènes sont impliqués dans le système d'ARN interférence. Après les preuves de concept, notamment chez *Arabidopsis thaliana* (Qi et al., 2013), de nombreuses approches ont été développées visant par exemple à simplifier les conditions de cultures de différentes plantes. Des gènes de

résistance aux moisissures ont notamment été insérés dans le génome de blé (Wang et al., 2014). D'autres études ont montré qu'il était possible par édition du génome d'améliorer le produit fini issu de plantes, comme l'huile de soja (Haun et al., 2014). Différents gènes impliqués dans la production d'acides gras poly-insaturés, menace pour la santé publique, ont alors été inactivés par KO.

D'autres applications de l'édition du génome sont étudiées en agronomie. La modification d'animaux dans différents buts, comme l'amélioration du produit fini ou la résistance à des maladies ou pathogènes (Sovová et al., 2016). Le gène produisant la myostatine a par exemple été inactivé chez la vache ou le cochon permettant d'augmenter leur masse musculaire (Guo et al., 2016; Lv et al., 2016). La myostatine est en effet un puissant régulateur négatif de la production de muscle. Le gène *blg* a aussi été inactivé chez la vache, permettant de réduire la production de la beta-lactoglobuline, l'allergène dominant du lait.

Comme nous avons pu le voir ici, beaucoup d'applications voient le jour depuis l'avènement des techniques d'édition du génome. Il reste cependant des inquiétudes et des difficultés à dépasser avant de réellement voir apparaître les organismes ou cellules génétiquement modifiés par les nucléases artificielles dans notre quotidien. Il y a notamment des inquiétudes d'ordre éthique à utiliser des animaux ou des plantes génétiquement modifiés ou encore des inquiétudes concernant les conséquences de ces modifications sur l'homme.

1.2 Les nucléases artificielles

1.2.1 Historique

Au début des années 80, et ce depuis le début de ses travaux sur la transformation de *Saccharomyces cerevisiae*, J. Szostak a montré l'importance des coupures double brin de l'ADN dans l'initiation des événements de recombinaison (Orr-Weaver and Szostak, 1983, p.).

La découverte de l'*homing endonuclease* (HE) I-SceI, qui est codé par un élément mobile pouvant promouvoir sa mobilité. Il a joué un grand rôle dans le développement des nouvelles méthodes d'édition du génome. En effet, les HE sont aussi appelées méganucléases parce que ce sont des endonucléases possédant un site long de reconnaissance de l'ADN, de 14 à 44pb (Chevalier and Stoddard, 2001), qui est en général absent du génome de mammifères, TAGGGATAACAGGGTAAT pour I-SceI. Dès lors, il a été possible de tester l'importance de l'induction d'une cassure double brin unique pour initier la modification ciblée du génome, notamment chez la levure dans un premier temps (Dujon, 1989; Plessis et al., 1992). En 1994 Rouet et al., ont testé l'utilisation de I-SceI dans des cellules de mammifère, COS1 et 3T3, et observé une stimulation de la recombinaison extrachromosomale en faisant recombiner 2 plasmides entre eux (Rouet et al., 1994). Après avoir inséré le site de reconnaissance de l'enzyme I-SceI dans des cellules murines (NIH-3T3 et PCC7-S), Chouluka et al ont quant à eux stimulé la réparation par recombinaison homologue au site de coupure dans le génome murin avec l'aide d'un ADN donneur (Chouluka et al., 1995). Cependant, les méganucléases disponibles ont une séquence déterminée et ne permettent pas de cibler n'importe quelle séquence souhaitée. Des tentatives de mutagenèse ont vu le jour pour obtenir des méganucléases ciblant d'autres séquences que leur cible naturelle (Arnould et al., 2006; Epinat et al., 2003; Grizot et al., 2010; Smith et al., 2006), mais les résultats

n'ont pas été très nombreux. Alors que les HE semblaient de plus en plus difficiles à utiliser, une nouvelle génération de nucléase a vu le jour, les nucléases à doigts de zinc.

1.2.2 Les nucléases à doigts de zinc

En effet, l'étude des doigts de zinc a permis de concevoir une nouvelle génération de nucléases, plus simple à modifier. Les protéines à doigts de zinc représentent la plus grande famille de facteurs de transcription. L'étude cristallographique de leur liaison à l'ADN a révélé qu'un doigt de zinc fixait 3 paires de bases de l'ADN de manière spécifique (Pavletich and Pabo, 1991). Il a fallu ensuite comprendre quel était le code des doigts de zinc, identifier lesquels reconnaissaient quelle séquence de 3pb (Desjarlais and Berg, 1992) et les assembler entre eux pour cibler une séquence d'ADN prédéterminée. Une fois ce nouveau domaine de reconnaissance de l'ADN assemblé, il a été utilisé pour recruter le domaine endonucléase de l'enzyme de restriction FokI (Kim et al., 1996), les ZFN (*Zinc Finger Nucleases*) étaient nées. Le domaine endonucléase de FokI n'est pas spécifique de séquences (Li et al., 1992; Sugisaki and Kanazawa, 1981). Il peut donc être fusionné à d'autres domaines de reconnaissance de l'ADN pour rediriger la coupure vers de nouveaux sites (Bitinaite et al., 1998).

Les ZFN sont obtenues par assemblage des doigts de zinc et du domaine de clivage de FokI (Figure 2). Elles fonctionnent en dimère et permettent de reconnaître une séquence de deux fois 9, 12 ou 15pb, selon le nombre de doigts de zinc assemblés, qui sera clivée en son centre. Il était alors possible, en principe, de cibler n'importe quelle séquence d'ADN précise et unique. Cependant, les mécanismes de reconnaissance de l'ADN par les doigts de zinc sont plus complexes que décrit au départ (Fairall et al., 1993). En effet chaque doigt de zinc reconnaît bien 3 pb mais lorsqu'ils sont assemblés, ils influencent la reconnaissance de leurs voisins et malheureusement il n'existe donc pas un doigt de zinc pour chaque triplet de nucléotides. Ce système, est compliqué à utiliser puisqu'en pratique il est nécessaire de tester plusieurs dizaines de ZFN potentielles. Cependant de nombreuses découvertes ont été faites à l'aide des ZFN et l'ère de l'édition du génome a

réellement débuté avec ces nucléases. Elles ont par exemple été utilisées pour corriger le gène IL2R γ , impliqué dans le SCID (*Severe Combined Immune Deficiency*) dans des lignées cellulaires humaines ou encore pour inactiver des gènes (gène *golden* et gène *no tail/Brachyury*) de manière dirigée chez le poisson zèbre pour la première fois (Doyon et al., 2008; Urnov et al., 2005).

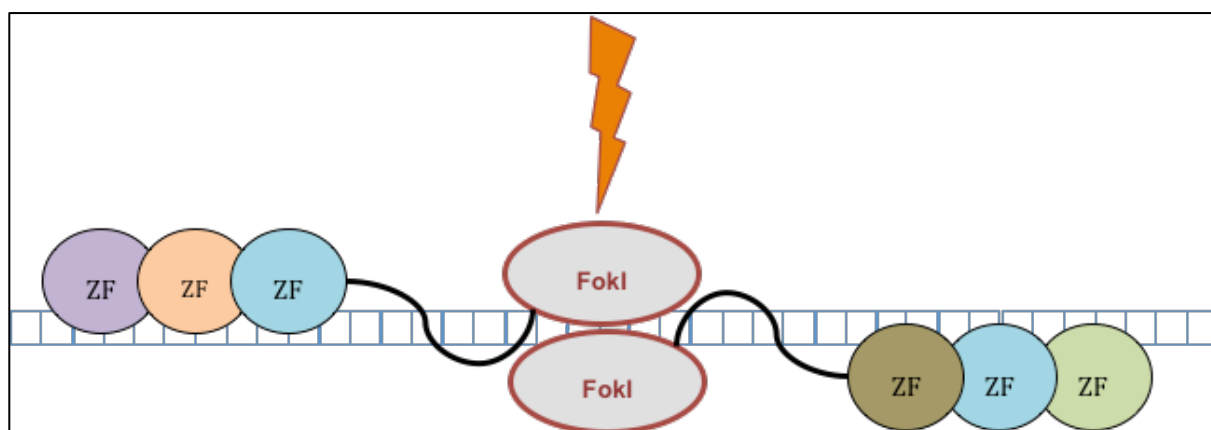


Figure 2. Nucléases à doigts de zinc ou *Zinc Finger Nucleases* (ZFN).

Une ZFN est composée de 3 doigts de zinc, reconnaissant chacun 3pb de l'ADN, et d'un domaine de clivage de l'ADN non spécifique, celui de l'enzyme FokI fonctionnant en dimère. Deux ZFN sont donc nécessaires pour effectuer une cassure double brin de la séquence d'ADN cible.

1.2.3 Les TALE nucléases

Les TALEs (*Transcription-activator-like effector*) représentent la plus grande famille des effecteurs de type III provenant des bactéries pathogènes de plantes de genre *Xanthomonas* (Bogdanove et al., 2010; Kay and Bonas, 2009; Römer et al., 2010). Ces protéines sont produites par la bactérie et injectées directement dans des cellules infectées pour modifier l'expression de ses gènes de sorte à favoriser l'infection. Les TALEs contiennent plusieurs domaines fonctionnels : en C-terminal, un domaine activateur de la transcription, des NLS (*Nuclear Localization Domain*) et en son centre un domaine de reconnaissance de l'ADN constitué d'un nombre variable de répétitions dites TALE (Figure 3). Ces répétitions TALE sont chacune composées de 34 acides

aminés conservés dont seuls les aa 12 et 13 diffèrent pour conférer à chaque répétition sa spécificité de reconnaissance d'une paire de base de l'ADN.

Le code des domaines TALE a été « déchiffré » (Boch et al., 2009) et on sait les acides aminés hypervariables nécessaires pour reconnaître chaque base de l'ADN (Figure 3).

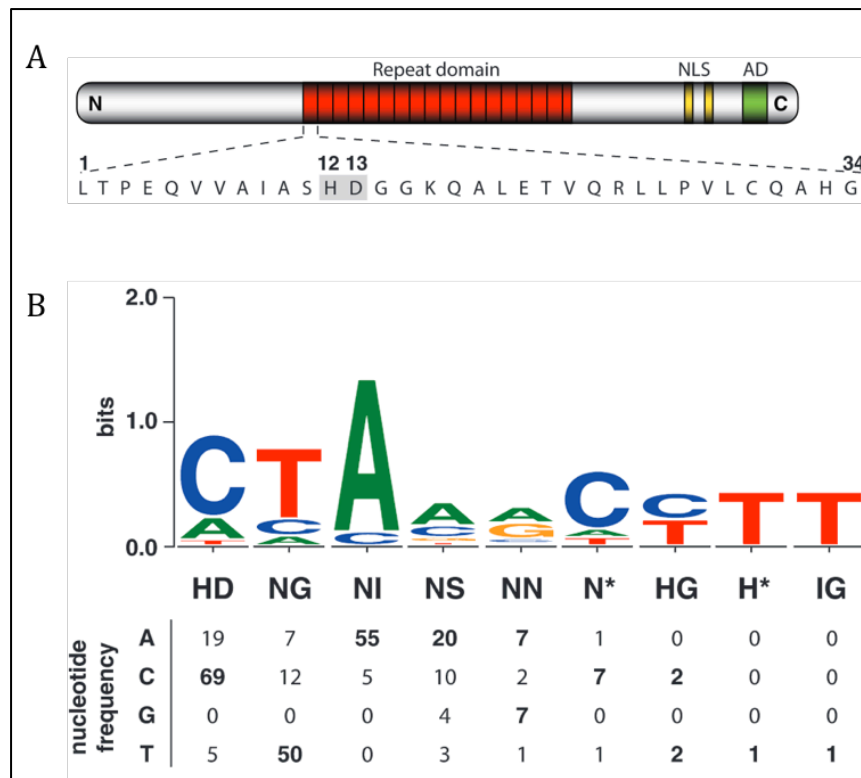


Figure 3. *Transcription Activator-Like Effector* (TALE) (figure adaptée de Boch et al., 2009).

(A) La protéine TALE naturelle est composée d'un domaine central de reconnaissance de l'ADN, de signaux de localisation nucléaire (NLS) et d'un domaine d'activation (AD). Le domaine de fixation à l'ADN est lui-même composé de plusieurs répétitions de 34 aa dont seulement les aa 12 et 13 varient (aa dits hypervariables) pour conférer la spécificité de reconnaissance d'une base d'ADN (B) Bases de l'ADN reconnues par chaque combinaison d'aa hypervariables.

En plus des répétitions, le domaine de liaison à l'ADN comporte une demi-répétition reconnaissant elle aussi une base selon le même code et un domaine supplémentaire N-terminal reconnaissant une thymidine. La séquence reconnue par un domaine TALE peut donc être simplifiée comme suit : $TN_{(n)}N$ (n indiquant le nombre de répétitions

TALE et le dernier N correspondant à la base reconnue par la demi-répétition). Les répétitions TALE sont totalement indépendantes les unes des autres au cours de la fixation à l'ADN : le code TALE permet donc de prédire la séquence à laquelle se fixe un domaine TALE avec précision. Inversement, il est aussi possible d'assembler ces répétitions TALE pour cibler une séquence déterminée. Ce domaine TALE modulable peut, tout comme les doigts de zinc, être couplé au domaine de clivage de l'enzyme FokI et ainsi former une TALE Nucléases (TALEN) (Figure 4).

Les TALEN sont plus simples à assembler que les ZFN et plusieurs techniques ont vu le jour pour assembler ces domaines de liaison à l'ADN à façon. Il y a la technique du Golden Gate (Cermak et al., 2015), celle du FLASH (*fast ligation-based automatable solid-phase high-throughput*) (Reyon et al., 2012), celle du LIC (*ligation-independent cloning*) (Schmid-Burgk et al., 2015) et enfin celle du Unit assembly (Huang et al., 2014).

La plupart des protocoles permettent d'assembler 16 répétitions et ainsi d'avoir 18pb reconnues par chaque sous-unité TALEN (16 répétitions + la demi-répétition + le T en 5') et 36pb au final de par l'utilisation de deux sous-unités. Pour que les domaines FokI puissent dimériser, il est important d'avoir un spacer de plus de 12pb entre les sites de fixation des deux sous-unités TALEN.

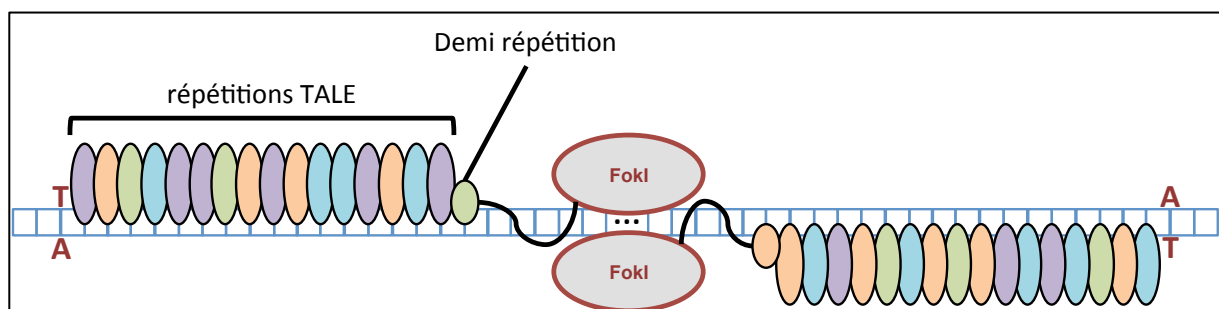


Figure 4. TALE nucléases (TALEN).

Une TALEN est composée de 16 répétitions TALE et d'une demi-répétition, reconnaissant chacune 1 pb de l'ADN, ainsi que du domaine de clivage de l'ADN non spécifique de l'enzyme FokI. Deux TALEN sont nécessaires pour effectuer une cassure double brin de l'ADN cible.

Malgré ces nombreux protocoles, les TALEN peuvent parfois être difficiles à assembler. En particulier, à cause des séquences codant les répétitions TALE pouvant parfois recombiner entre elles. Les TALENs ont néanmoins permis de grandes avancées et ont été beaucoup utilisées dans l'édition du génome. Elles ont permis d'inactiver des gènes dans de nombreux organismes, par exemple la levure, le poisson zèbre, le rat... (Huang et al., 2011; Li et al., 2011).

Cependant, un nouveau venu dans le monde des nucléases artificielles, le système CRISPR/Cas9, a rapidement été adopté grâce à sa simplicité d'utilisation encore plus grande.

1.2.4 Le système CRISPR/Cas9

Le système CRISPR-Cas9 (*Clustered Regularly Interspaced Palindromic Repeats - CRISPR associated protein 9*) est un système de coupure de l'ADN qui a été adapté pour l'édition du génome à partir de protéines découvertes chez des bactéries. Il a été identifié comme étant un système de défense contre virus et plasmides (Deltcheva et al., 2011; Terns and Terns, 2011; Wiedenheft et al., 2012). Différents types de systèmes CRISPR ont été identifiés, plus ou moins complexes. Les types I et III font intervenir de nombreuses protéines Cas pour reconnaître et détruire l'ADN étranger alors que le type II est plus simple puisqu'il ne fait intervenir que trois composants : un crRNA (*CRISPR RNA*), qui va cibler de manière précise une séquence de 20 pb du génome à éliminer, un tracrRNA (*trans-activating crRNA*), qui va s'associer au crRNA et recruter la nucléase Cas9, dernière partie de ce complexe. Cette dernière va couper la séquence cible s'il y a présence de la séquence PAM (*Protospacer Adjacent Motif*) en aval de la séquence de 20 pb reconnue par le crRNA (**Erreur ! Source du renvoi introuvable.**).

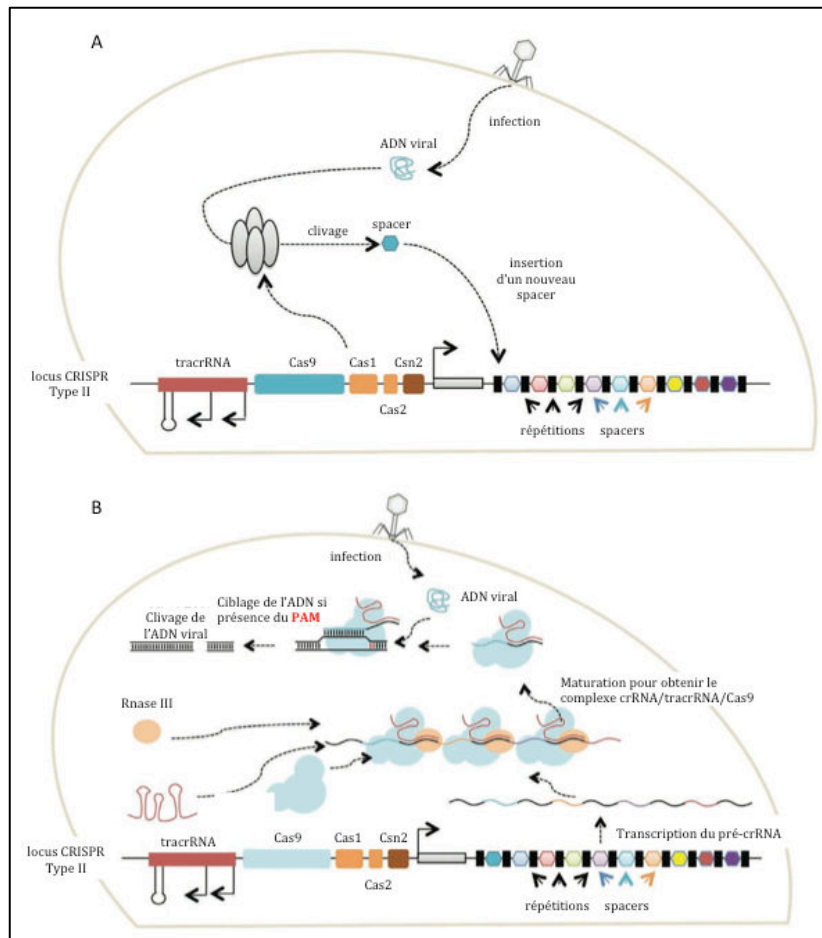


Figure 5. Fonctionnement du système CRISPR/Cas naturel (Adapté de Mali et al., 2013).

(A) Phase d'immunisation. Le système CRISPR intègre des fragments d'ADN exogènes, comme signature de précédentes infections, au locus CRISPR comme « spacers ». (B) Phase d'immunité. La bactérie utilise les informations stockées pour se défendre contre de nouvelles infections. Les spacers intégrés sont transcrits en pré-crRNA qui est maturé pour obtenir les crRNA. Ces derniers vont cibler l'ADN exogène par simple liaison Watson-Crick et recruter par complémentarité le tracrRNA et la nucléase Cas9 pour cliver et dégrader l'ADN exogène s'il y a présence de la séquence PAM.

Son mécanisme a été décrit pour la première fois en 2012 par les équipes d'E. Charpentier et de J. Doudna (Jinek et al., 2012).

Le principe général de défense des bactéries et archées par les systèmes CRISPR naturels est le suivant (Figure 5). Il y a tout d'abord la phase d'immunisation, pendant laquelle un phage infecte la bactérie et y injecte son génome. La bactérie est alors

capable de digérer ce génome en fragments qui subissent différents traitements aboutissant à l'insertion de séquences de 20 pb, appelés spacers, dans le génome de la bactérie, à un endroit précis, situé entre de nombreux motifs de séquences répétées (les *clustered regularly interspaced short palindromic repeats* [CRISPR]) (Deltcheva et al., 2011; Wiedenheft et al., 2012). Vient ensuite la phase d'immunité pendant laquelle les spacers et CRISPR sont transcrits et donnent naissance à un long pre-crRNA, qui va ensuite être digéré pour former plusieurs crRNA. Après association d'un crRNA au tracrRNA, l'hybride crRNA-tracrRNA va s'associer à Cas9. Si la séquence adjacente au PAM est complémentaire, le crRNA va s'y fixer par simple liaison Watson-Crick. Cette fixation va provoquer une modification conformationnelle de la nucléase qui déclenche l'activité nucléase de ses deux domaines catalytiques (HNH et RuvC) clivant chacun un des brins de l'ADN cible (Jinek et al., 2012).

Une fois ses mécanismes moléculaires connus, il a été possible d'adapter le système CRISPR-Cas9 à l'édition du génome. Il suffit tout simplement d'insérer la séquence de 20pb à cibler à la place du *spacer* dans le crRNA.

La protéine Cas9 la plus utilisée reste celle de *Streptococcus pyogenes* dont le PAM est NGG (Deltcheva et al., 2011). Pour simplifier son utilisation, le crRNA et le tracrRNA ont été fusionnés pour obtenir un ARN guide unique d'environ 100 nucléotides, qui reconnaît donc une séquence de 20pb et se lie directement à la protéine Cas9 pour cliver l'ADN cible 3pb avant la séquence PAM. Les extrémités de l'ADN obtenues après coupure sont franches (Figure 6) (Jinek et al., 2012).

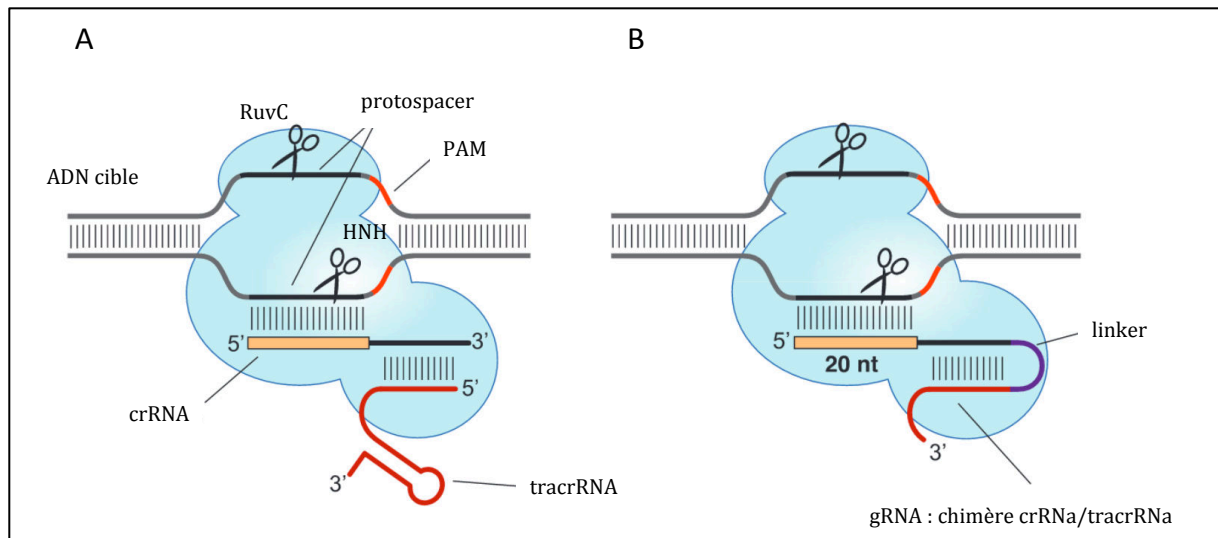


Figure 6. Système CRISPR-Cas9 (Figure adaptée de Jinek et al., 2012).

(A) Système initial découvert chez *S. pyogenes*, le crRNA (CRISPR RNA) se fixe à sa cible par simple liaison Watson-Crick. Le crRNA recrute par complémentarité le tracrRNA (trans-activating RNA) qui recrute lui-même la nucléase Cas9. La nucléase clive ensuite l'ADN, s'il y a présence de la séquence PAM, à l'aide de ses deux domaines nucléases HNH et RuvC. (B) Pour simplifier l'utilisation de ce système le crRNA et le tracrRNA ont été fusionnés pour obtenir un ARN guide unique ou ARNg.

Le système CRISPR-Cas9 est de très loin le plus simple et le plus rapide à mettre en œuvre de toutes les nucléases artificielles. Son efficacité étant souvent très bonne, on assiste à une explosion spectaculaire de cette nouvelle technologie.

De nombreux systèmes CRISPR-Cas9, provenant d'autres bactéries, peuvent être eux aussi utilisés. Il y a celui de *S. aureus* dont le PAM est NNGRRT (ou NNGRR) (Ran et al., 2015), celui de *N. Meningitidis* dont le PAM est NNNNGATT (Hou et al., 2013) ou encore celui de *S. Thermophilus* dont le PAM est NGGNG (Karvelis et al., 2013).

1.2.5 Nouveaux types de nucléases artificielles utilisés

1.2.5.1 Un nouveau système CRISPR : Cpf1

Un nouveau système CRISPR/Cas de type II a récemment été découvert. Il est naturellement composé d'un seul ARN guide d'environ 40nt qui recrute la nucléase Cpf1 (*Clustered Regularly Interspaced Short Palindromic Repeats from Prevotella and Francisella 1*) (Zetsche et al., 2015) (Figure 7). Au contraire des nucléases Cas9 qui induisent une coupure franche, Cpf1 produit majoritairement des extrémités 3' sortantes de 5 nucléotides. Des Cpf1 provenant de différentes bactéries ont été testées en cellules humaines et seulement certaines sont actives, AsCpf1 de *Acidaminococcus* sp, CbCpf1 de *Cachnospiraceae bacterium* (Kleinstiver et al., 2016b) et LbCpf1 de *Lachnospiraceae bacterium* (Zetsche et al., 2015). Ces trois nucléases ont un PAM TTTN en 5' des 23nt de la cible complémentaire à l'ARN.

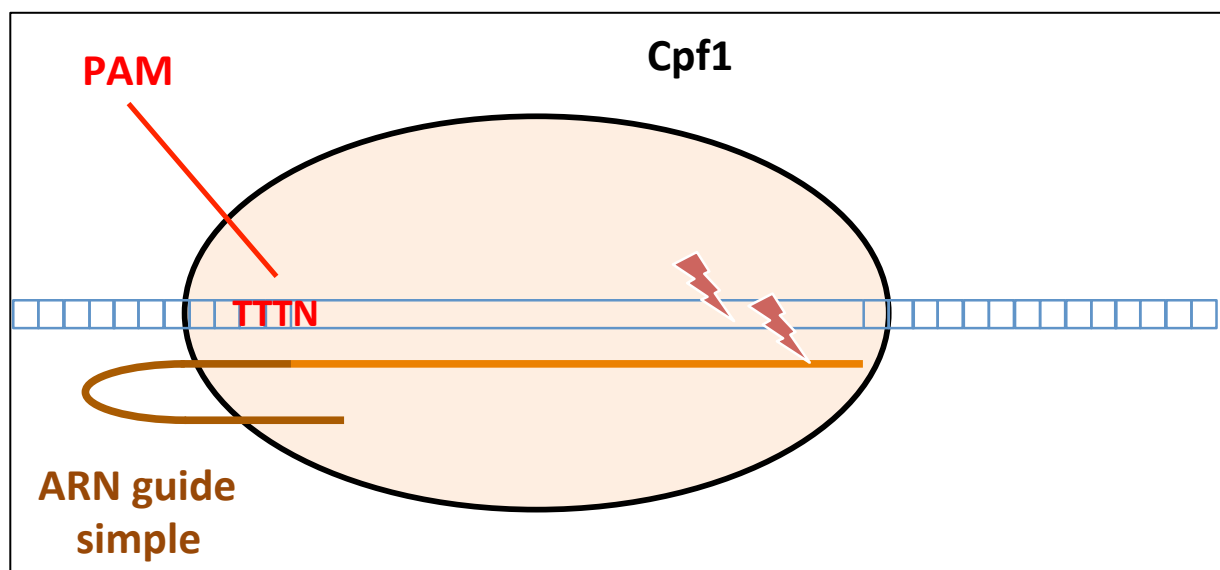


Figure 7. Cpf1 (*Clustered Regularly Interspaced Short Palindromic Repeats from Prevotella and Francisella 1*). Ce système CRISPR est composé d'un ARN guide unique d'environ 40nt et d'une nucléase Cpf1 qui va cliver l'ADN cible en produisant des extrémités 3' sortantes. Pour effectuer une cassure double brin, la présence d'un PAM de type TTTN en amont de la séquence reconnue par l'ARN guide est nécessaire.

1.2.5.2 Nucléase avec ADN guide : Argonaute

Un autre système semble pouvoir être utilisé pour l'édition du génome, la protéine Argonaute. Cette nucléase est bien connue chez les eucaryotes puisque c'est un acteur clé du système ARN interférence. Cependant cette protéine existe aussi chez les procaryotes et semble servir de système immunitaire tout comme le système CRISPR-Cas9. Différentes protéines Argonaute ont été étudiées chez les archées (Swarts et al., 2015), elles utilisent un ADN simple brin phosphorylé en 5' comme guide pour cliver un ADN double brin. Cependant jusqu'ici les protéines Argonautes découvertes n'étaient actives qu'à de hautes températures (95°C pour PfAgo), ne permettant donc pas de les utiliser dans des cellules humaines pour l'édition du génome. La nucléase Argonaute de *Natronobacterium gregoryi*, récemment découverte, fonctionnerait à 37°C et pourrait permettre de modifier des cellules humaines (Gao et al., 2016). Néanmoins, aucun groupe n'a réussi à reproduire ces résultats à ce jour.

1.2.6 Difficultés liées à l'utilisation des nucléases

Les systèmes utilisés pour l'édition du génome sont de plus en plus efficaces et simples d'utilisation. Cependant des difficultés sont à prendre en compte lors de leur utilisation.

1.2.6.1 Le problème de spécificité des nucléases artificielles et l'induction de mutations Off-target

Une des complications potentielles de l'utilisation des nucléases artificielles, pour l'édition du génome, est la fixation de celles-ci à des sites différents de la cible. La coupure de tels sites dits « off-target » par les nucléases peut engendrer des modifications non souhaitées (insertions, délétions voire des réarrangements chromosomiques s'il y a plusieurs coupures simultanées dans le génome). Différents facteurs affectent le clivage de ces sites off-target (Yee, 2016a):

La présence de séquences homologues à la séquence cible dans le génome cible est un facteur très important. Les nucléases artificielles peuvent se fixer malgré la présence de quelques mésappariements avec leur séquence cible (Gabriel et al., 2011; Pattanayak et al., 2013). Et plus il y a de séquences homologues à la séquence cible, plus il y a de risque d'avoir des mutations à ces sites off-target. Pour s'affranchir de ce facteur il est possible d'utiliser des logiciels en ligne qui vont identifier les sites off-target potentiels et permettre de choisir les séquences cibles avec le moins de sites homologues : PROGNOS pour les ZFN et les TALEN (Fine et al., 2014) et de nombreux sites pour les systèmes CRISPR/Cas9, dont le site CRISPOR développé au laboratoire (Haeussler et al., 2016).

Le niveau d'expression de la nucléase est un autre facteur important. Effectivement, un niveau d'expression trop élevé augmente le risque de clivage off-target (Pattanayak et al., 2013). Par exemple, plus de mutations aux sites off-target sont observées dans des lignées cancéreuses, des cellules pour lesquelles le taux de transfection est très, comparé à des ESC (*Embryonic Stem Cell*) pour lesquelles le taux de transfection reste faible (Li et al., 2015; Park et al., 2015; Schwank et al., 2013).

L'accessibilité des sites cibles est un autre facteur qui semble jouer un rôle dans les mutations aux sites off-target. Plus de mutations off-target sont détectées *in vitro* qu'*in vivo*, en ciblant la même séquence avec le système CRISPR/Cas9 (Guilinger et al., 2014; Pattanayak et al., 2013). Cela peut être dû au statut épigénétique des différentes cibles, cependant ce paramètre est difficile à contrôler.

1.2.6.2 Solutions pour augmenter la spécificité des nucléases

Différentes stratégies ont été mises en œuvre pour augmenter la spécificité des nucléases artificielles (Yee, 2016b). Il est maintenant possible, par exemple, d'utiliser des Cas9 nickases. La Cas9 nickase la plus utilisée est le mutant de Cas9 portant la mutation D10A qui inactive le domaine nucléase RuvC. Comme son nom l'indique, cette nucléase effectue un nick dans l'ADN et non pas une cassure double brin. Donc, pour

obtenir une CDB, il est nécessaire d'utiliser deux ARNg avec des séquences cibles proches et situées sur chaque brin (Ran et al., 2013). Cette stratégie permet d'obtenir une CDB seulement si les deux séquences cibles des ARNg sont reconnues, augmentant la spécificité.

Une Cas9 mutée dans les 2 domaines catalytiques, RuvC et HNH, a aussi été fusionnée au domaine de clivage de FokI, utilisé pour les ZFN et les TALEN. Il faudra ici aussi utiliser deux ARNg proches pour avoir une cassure double brin (Tsai et al., 2014).

De nouveaux mutants de Cas9 dits « *High Fidelity* » ou HF ont été développés récemment. La nucléase Cas9-HF1, portant 4 mutations (N497A, R661A, Q695A et Q926A) affectant l'interaction de la nucléase avec l'ADN, permet de diminuer voire d'éliminer les mutations aux sites off-target alors que l'activité au site ciblé ne semble pas diminuée (Kleinstiver et al., 2016a). Un autre mutant, la eSpCas9(1.1) portant 3 mutations (K848A, K1003A et R1060A) permet lui aussi de diminuer voir d'éliminer les mutations aux sites off-target tout en gardant l'activité au site cible (Slaymaker et al., 2016). Cette fois ci, le mutant eSpCas9 diminue plusieurs charges positives impliquées dans l'interaction avec le brin opposé à la séquence cible.

1.2.6.3 L'utilisation des nucléases artificielles in vivo et le mosaïcisme des animaux

Un phénomène apparaît souvent lors de création d'animaux génétiquement modifiés par injection de nucléases dans l'œuf fécondé, c'est le mosaïcisme. Si la cellule injectée se divise rapidement ou que les nucléases tardent à cliver l'ADN, la mutation souhaitée peut avoir lieu à un stade plus avancé du développement de l'animal, et n'affecter qu'une partie des cellules. Dans ce cas, l'organisme entier ne comportera pas la mutation, mais seulement les cellules provenant de la cellule initialement modifiée. Le mosaïcisme génétique des animaux produits peut donner un éventail de phénotypes et compliquer leur étude.

Plusieurs solutions peuvent être envisagées. En premier lieu, il est possible d'utiliser l'animal mosaïque comme géniteur en espérant que ses cellules germinales ont été

modifiées et qu'elles permettront d'avoir une descendance portant la mutation souhaitée. Les autres solutions consistent à diminuer le mosaïcisme en favorisant une coupure rapide de l'ADN. Pour cela il est possible d'injecter directement le complexe ARNg/protéine Cas9 pour éliminer les délais d'expression à partir de plasmides ou d'ARN (Sung et al., 2014; Woo et al., 2015).

1.3 Les systèmes de réparation des CDB

1.3.1 Les cassures double-brin de l'ADN

Les lésions causées par les nucléases artificielles sont des cassures double brin de l'ADN (CDB). La cellule possède la capacité de réparer ce type de lésions grâce à plusieurs systèmes de réparations. En effet, la cellule subit de manière naturelle et tout au long de sa vie, différents dommages remettant en cause l'intégrité de son génome qu'elle doit réparer, dont les CDB. A la suite de CDB, il peut y avoir perte d'informations génétiques ou réarrangements chromosomiques (translocations, etc) pouvant mener à une instabilité génomique et à la mort cellulaire. Cependant des CDB peuvent aussi être nécessaires au bon fonctionnement de l'organisme. Il existe effectivement deux types de CDB programmées ou accidentelles.

Les CDB peuvent faire ainsi partie de processus physiologiques, comme par exemple au cours de la diversification des immunoglobulines (Mehta and Haber, 2014). Dans ce cas, les cassures double brin sont induites par le complexe protéique RAG pour initier la recombinaison V(D)J pour qu'ensuite il y ait échange d'informations génétiques et diversification des récepteurs d'antigènes. Un autre processus physiologique nécessite l'apparition de CDB, la méiose. En effet, lors de la méiose, la nucléase Spo11 crée de nombreuses CDB pour permettre la recombinaison des chromosomes et favoriser l'évolution génétique de l'espèce.

Les CDB peuvent aussi arriver de manière accidentelle, les causes sont alors multiples (UV, RI, chimiothérapies...). Dans ce cas, la cellule doit restaurer l'intégrité de son génome pour éviter toute mutation ou réarrangement de son génome. Ces cassures sont souvent complexes, car les extrémités produites peuvent être endommagées et associées à des pertes d'informations de séquence.

1.3.2 Les systèmes de réparation des cassures double brin

La réparation des CDB est essentielle pour préserver l'intégrité du génome. Pour cela les cellules ont développé des systèmes pour détecter ces atteintes, appelé réponse aux dommages de l'ADN (RDA). Chaque RDA peut se décomposer en plusieurs étapes : détection initiale du dommage à l'ADN, recrutement des facteurs au site de la cassure et enfin la réparation de la lésion. Différents systèmes peuvent réparer les cassures double brins. Les deux principaux sont le NHEJ et la RH, mais d'autres systèmes peuvent être utilisés, en particulier le MMEJ (*Micro-homology Mediated End Joining*) et le SSA (*Single Strand Annealing*).

1.3.2.1 Détection des CDB par la cellule

Dans un premier temps, la CDB est détectée par des protéines appartenant à la famille des PIKK (*Phosphatidylinositol 3-Kinase-like protein Kinase*), ATM (*Ataxia Telangectasia Mutated*) et ATR (*ATM- and Rad3-related*) et par des membres de la famille PARP (*Poly(ADP-ribose) Polymerase*), principalement PARP1 et PARP2 (Harper and Elledge, 2007). ATM et DNA-PKcs (*DNA-dependent Protein Kinase catalytic subunit*) peuvent être directement activés par des agents de dommage à l'ADN, de type irradiation, qui vont créer des CDB (Schreiber et al., 2006).

Dans les secondes suivant la CDB, ATM est activé par autophosphorylation sur la sérine 139 (Kozlov et al., 2011). Ce dernier va phosphoryler les histones H2AX, qui sont alors appelés γ H2AX, permettant ainsi de décondenser l'ADN, étape critique pour l'initiation du processus de réparation (Burma et al., 2001). Les histones γ H2AX sont ensuite reconnus par MDC1 (*Mediator of DNA damage Checkpoint protein 1*) qui est aussi phosphorylé et activé par ATM (Liu et al., 2016). Ce recrutement de MDC1 au site de coupure est nécessaire à l'activation des *check point* de dommage à l'ADN ou encore au recrutement de différentes protéines impliquées directement dans la réparation elle-même, comme MRN (*Mre11-Rad51-Nbs1*), BRCA1 (*Breast Cancer 1*) ou encore 53BP1 (*p53-binding protein 1*) (Liu et al., 2016).

Les protéines PARP1 et PARP2, quant à elles, permettent la poly(ADP-ribosyl)ation principalement d'histones, permettant alors la relaxation de l'ADN sur environ 30nm ce qui augmente l'accès à la CDB des différents acteurs de la réparation, comme BRCA1, Nbs1 (*Nibrin*), hSSB1 (*human Single Strand Binding protein*) ou encore la Ligase 4 (Li and Xu, 2016; Schreiber et al., 2006).

Ces différentes protéines permettent l'accessibilité aux protéines impliquées dans les différents systèmes de réparation des CDB, qui sont décrits plus en détails ci-après.

1.3.2.2 Réparation des cassures double brin par le Non Homologous End Joining

La réparation de la CDB par le système NHEJ est assez simple. Les extrémités sont d'abord protégées et maintenues proche l'une de l'autre, traitées si elles sont complexes et ensuite liguées entre elles pour rétablir l'intégrité du chromosome (Figure 8). Ce système de réparation est actif tout au long de son cycle cellulaire mais semble dominer en phase G1.

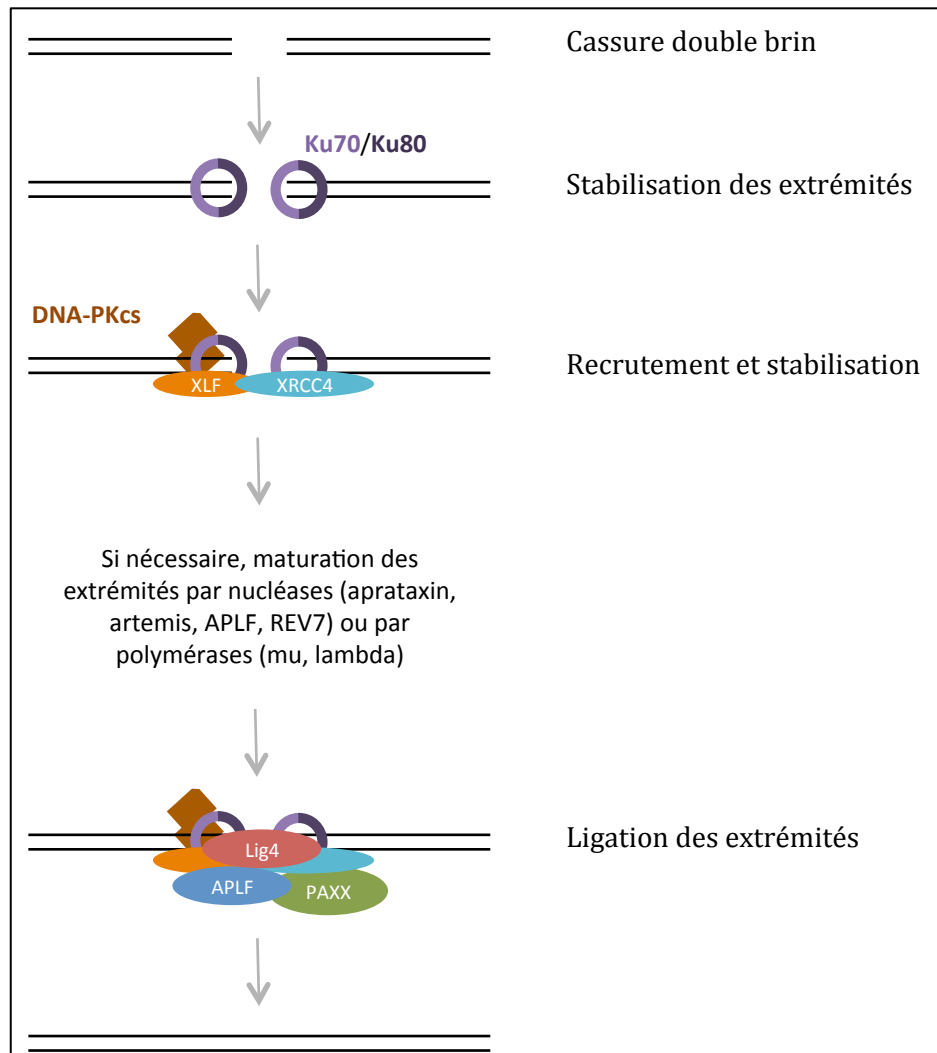


Figure 8. Réparation des cassures double brin par NHEJ.

Les extrémités sont immédiatement fixées par l'hétérodimère Ku70/Ku80 pour être protégées. Elles sont aussi stabilisées par XLF/XRCC4 et il y a recrutement de DNA-PKcs, APLF, PAXX et de la Ligase 4. La ligase 4 effectue pour finir la ligation des extrémités.

1.3.2.2.1 Reconnaissance et stabilisation des extrémités

La réparation par NHEJ est initiée par la fixation de l'hétérodimère Ku70/Ku80 aux extrémités de la cassure (Dyran and Yoo, 1998). Sa structure circulaire suggère que les extrémités de la coupure s'insèrent au centre de ce dimère pour être protégées et rester proches l'une de l'autre dans l'espace, ce qui facilitera la réparation (Davis and Chen,

2013). L'interaction entre l'hétérodimère Ku70/Ku80 et l'ADN est directe et n'est pas séquence-dépendante. L'hétérodimère Ku70/Ku80 permet ensuite le recrutement de différents acteurs de la réparation par NHEJ. Il permet notamment le recrutement des protéines DNA-PKcs, XRCC4 (*X-ray Repair Cross-Complementing Protein 4*), DNA ligase 4, XLF (*XRCC4-Like Factor*) et APLF (*Aprataxin and PNKP Like Factor*) (Iliakis et al., 2015). XRCC4 et XLF forment un filament qui pourrait jouer un rôle dans la stabilisation des extrémités de la coupure en synergie avec l'hétéroduplex Ku70/Ku80.

Le recrutement de la DNA-PKcs à la coupure permet son activation. Ce nouveau complexe Ku70/Ku80/DNA-PKcs est appelé DNA-PK. Il a été montré *in vitro* qu'il phosphorylait plusieurs protéines impliquées dans la réparation par NHEJ, comme Ku70/Ku80, XRCC4, XLF ou encore Artemis, mais l'importance de ces phosphorylations n'a pas encore été validé *in vivo* et son activité réelle n'est pas encore bien déterminée (Jette and Lees-Miller, 2015). Malgré cela deux choses sont certaines, l'activité kinase de DNA-PK est indispensable à la réparation par NHEJ et l'autophosphorylation de DNA-PK est indispensable à sa dissociation des extrémités de la cassure.

1.3.2.2.2 Maturation des extrémités

Si nécessaire, les extrémités de la coupure peuvent être ensuite modifiées (résection, polymérisation, retrait de groupe aux extrémités) pour permettre la ligation. Pour ce faire, différentes protéines semblent être recrutées à la coupure par le binôme Ku/XRCC4 (Davis and Chen, 2013). Par exemple PNKP, une kinase/phosphatase, l'Aprataxin, une nucléotide hydrolase, Artemis et APLF permettent toutes trois la résection des extrémités et enfin, les polymérases Mu et Lambda peuvent aussi intervenir (Rodgers and McVey, 2016). La polymérase Mu, elle, permet une synthèse matrice-dépendante alors que la polymérase Lambda n'a pas besoin de matrice pour polymériser.

Ces différentes interventions sur les extrémités permettant la ligation mais elles peuvent aussi conduire à l'introduction d'erreurs dans la séquence.

1.3.2.2.3 Ligation finale

L'étape finale est la ligation entre les extrémités de la coupure. Cette ligation est effectuée exclusivement par la Ligase 4, qui peut fonctionner seule mais qui peut aussi être stabilisée et stimulée par XRCC4, XLF et possiblement par APLF également (Davis and Chen, 2013; Mari et al., 2006). Si elle est stimulée par XLF, la Ligase pourrait alors effectuer une ligation sur des extrémités non cohésives. Une autre protéine, intervenant au niveau de la ligation finale, a été découverte, PAXX (*PAralog of XRCC4 and XLF*). Cette protéine est un paralogue de XRCC4. Elle interagit avec Ku70/Ku80 et coopère avec XLF pour stimuler la ligation par la Ligase 4 (Xing et al., 2015).

L'ordre d'arrivée des protéines au site de coupure exposé ici reste un modèle. Les études s'accordent sur le fait que l'hétérodimère Ku70/Ku80 arrive le premier mais ensuite rien n'est clairement établi. L'hétérodimère Ku70/Ku80 ne semble pas avoir besoin de la DNA-PKcs pour recruter la ligase (Mari et al., 2006). Une autre étude montre que la DNA-PK est stabilisée et activée par le complexe XRCC4/XLF/Lig4 (Cottarel et al., 2013). Les protéines impliquées dans la réparation par NHEJ semblent coexister et agir selon les besoins et la complexité des extrémités.

1.3.2.3 Réparation des cassures double brin par la Recombinaison Homologue

La recombinaison homologue est un système de réparation plus complexe que le NHEJ. En effet, elle nécessite une étape de résection des extrémités produisant des ADN simple brin qui vont aller à la recherche d'une séquence homologue pour recopier la séquence interrompue au niveau de la CDB (Figure 9). Ce système de réparation est restreint aux phase S et G2 du cycle cellulaire, lorsque le génome est répliqué et que la chromatide sœur est disponible pour servir de matrice de réparation.

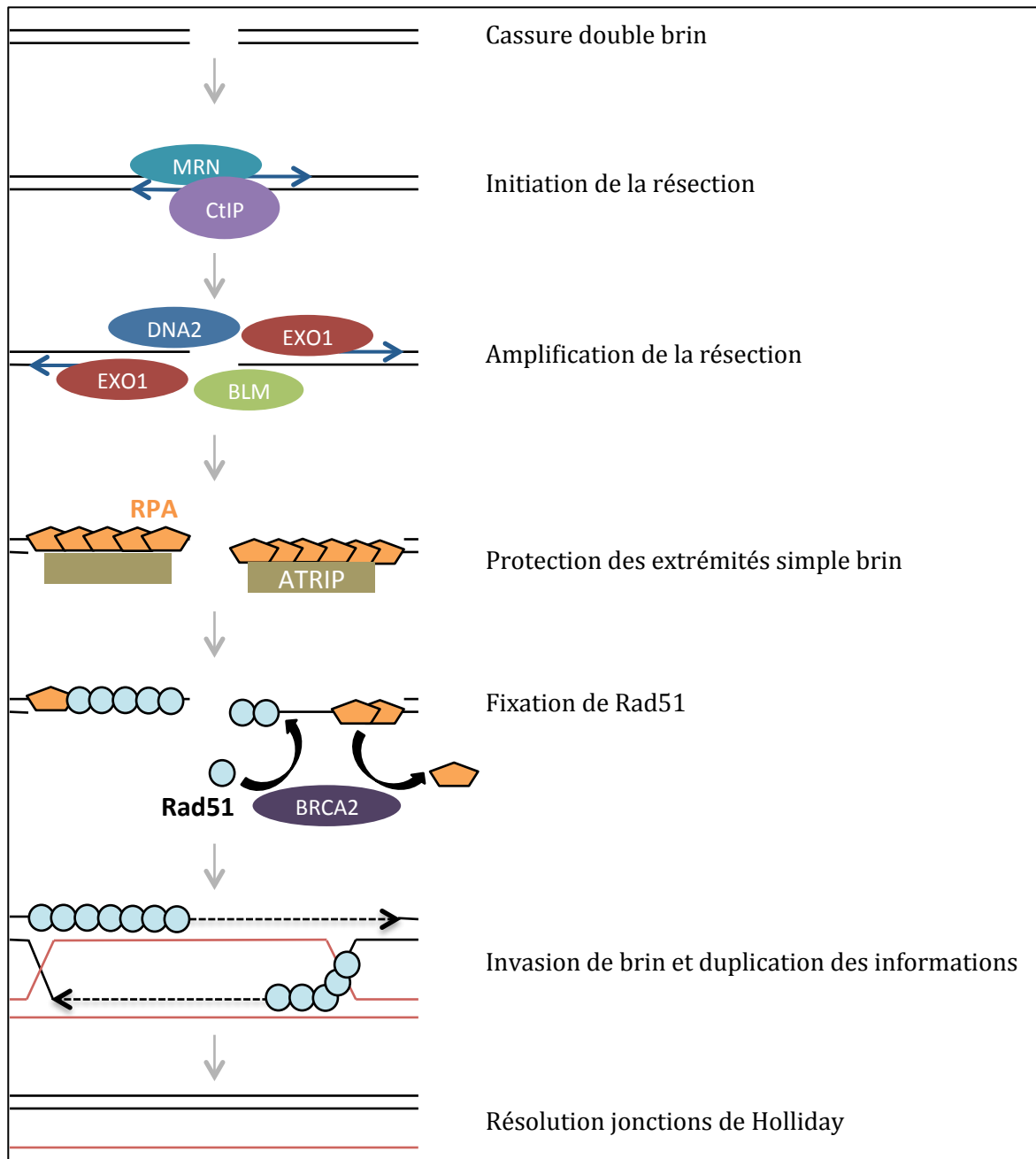


Figure 9. Réparation des cassures double brin par RH.

Les extrémités de la coupure subissent une initiation de la résection par l'action du complexe MRN/CtIP. La résection est ensuite amplifiée par EXO1 et le duo d'hélicases DNA2/BLM pour obtenir plusieurs kb de simple brin. L'ADN simple brin est ensuite fixé par RPA et stabilisé par ATRIP pour garder son intégrité. BRCA2 charge ensuite Rad51 sur l'ADN simple brin pour permettre l'invasion de l'ADN homologue. Les informations sont ensuite copiées pour réparer la cassure double brin. La séparation des ADN matrice et réparé se fait par résolution des jonctions de Holliday.

1.3.2.3.1 Résection des extrémités

La résection est initiée puis accentuée jusqu'à obtenir plusieurs kb de simple brin 3' sortant (Rodgers and McVey, 2016). Les protéines impliquées dans l'initiation et l'expansion de cette résection sont nombreuses (Liu and Huang, 2016). Il y a tout d'abord, le complexe MRN (comportant Mre11, Rad50 et Nbs1) qui une fois fixé à l'ADN, et complexé à CtIP (*C-terminal-binding protein interacting protein*), va permettre l'initiation de la résection. Mre11, le composant central de ce complexe MRN, possède différentes activités enzymatiques dont une activité exonucléase 3'-5' et une activité endonucléase sur l'ADN simple brin. Paradoxalement, la production de l'ADN simple brin 3' sortant, nécessaire à la réparation par RH, fait intervenir une activité exonucléase 5'-3'. Le mécanisme proposé est le suivant, Mre11 effectue un nick dans l'ADN à proximité de la coupure et digère l'ADN via son activité 3'-5' pour obtenir une queue d'ADN simple brin 3' sortante de 50 à 100 nucléotides. La protéine CtIP quant à elle peut jouer deux rôles dans la résection avec implication ou non de son domaine endonucléase (Liu and Huang, 2016; Wang et al., 2014). Lorsque les extrémités de la coupure sont « propres » (comme celles produites par les nucléases artificielles) la résection est dépendante de CtIP mais indépendante de son activité nucléase. Mais lorsque la coupure est plus complexe, après irradiation par exemple, la résection est dépendante de CtIP et de son activité nucléase. Une fois la résection initiée par MRN/CtIP, celle-ci est amplifiée pour obtenir plusieurs kb d'ADN simple brin 3' sortant, nécessaires à la fixation de RPA (*Replication Protein A*) et à la recherche d'homologie. EXO1 (*Exonuclease1*), possédant une activité 5'-3', entre alors en jeu. Cette nucléase préfère les substrats double brin avec une petite extrémité 3' sortante, telle qu'elle est fournie par le complexe MRN/CtIP (Liu and Huang, 2016). Lors de cette amplification de la résection, DNA2 et BLM ont aussi un rôle. BLM (*Bloom*) est une hélicase responsable de la relaxation de l'ADN et DNA2 est à la fois une hélicase et une nucléase. Cependant, l'activité hélicase de DNA2 ne semble pas nécessaire à la résection, contrairement à son activité nucléase (Liu and Huang, 2016).

La résection est donc initiée par le complexe MRN/CtIP et amplifiée par EXO1 et DNA2/BLM pour laisser place à la suite du système de réparation par RH, l'invasion de brin.

Une des protéines clés de la résection, CtIP, a été plus particulièrement étudiée pendant ma thèse. Cette protéine a initialement été décrite comme étant un partenaire de RB1 (retinoblastoma protein 1), CtBP (C-terminal binding protein 1) et BRCA1 ainsi qu'étant un facteur de transcription permettant la régulation de la progression G1/S du cycle cellulaire et un acteur de l'activation du checkpoint G2/M en cas de dommage de l'ADN (Chen et al., 2005; Li et al., 2000; Wong et al., 1998; Yu and Chen, 2004). Ensuite, l'implication directe de cette protéine dans la réparation de l'ADN a été mise en évidence (Sartori et al., 2007). Son rôle principale et le mieux établi est de promouvoir la résection des extrémités des CDB avec l'aide du complexe MRN. Pour se faire, la protéine CtIP est composée de domaines conservés, notamment en Nter et Cter. La partie C-terminale est analogue à son homologue chez la levure, Sae2 (Huertas and Jackson, 2009) et la partie N-terminale contient notamment un domaine d'interaction avec Nbs1 et une domaine de tétramérisation (Davies et al., 2015; Sartori et al., 2007; Yuan and Chen, 2009). Ces différents domaines Nter et Cter sont indispensables à la réparation des CDB.

1.3.2.3.2 Invasion de brin et réparation finale

L'ADN simple brin tout juste formé est recouvert par la protéine RPA (*Replication protein A*) qui va le protéger de la dégradation, de la formation de structures secondaires et le stabiliser pour accueillir Rad51 (Iliakis et al., 2015). Cette structure ADN/RPA est stabilisée par la protéine ATRIP (*ATR-Interacting Protein*) (Ball et al., 2005) et favorisée par la présence de CtIP à la coupure (Huertas and Jackson, 2009).

La formation du filament nucléoprotéique Rad51 fait intervenir différents paralogues de Rad51 (Rad51B, Rad51C, Rad51D, Xrcc2 et Xrcc3) ainsi que BRCA2 (*Breast Cancer 2*), cette dernière servant de plateforme de chargement de Rad51 sur l'ADN en

remplacement de RPA chez les mammifères, rôle joué par la protéine Rad52 chez la levure (Khade and Sugiyama, 2016).

Le filament nucléoprotéique (ADN/Rad51) ainsi formé cherche alors une séquence homologue pour réaliser l'invasion de brin. Les mécanismes de recherche de séquence homologue sont encore mal connus. Une fois la séquence trouvée, il y a formation de la D-loop (*displacement loop*), copie de la séquence par polymérisation à partir de l'extrémités 3', et formation de jonction de Holliday. Cette dernière est résolue par des endonucléases et par l'action complémentaire d'hélicases et de topoisomérases, pour achever la réparation de l'ADN endommagé.

La RH est décrite comme un système de réparation hautement fidèle. Cependant, dans certains cas, elle peut être source de mutations (Guirouilh-Barbat et al., 2014; Rodgers and McVey, 2016). Ces mutations peuvent être dues à la vulnérabilité de l'ADN simple brin, à des erreurs de la polymérase ou encore à des erreurs dues à la présence de séquences répétées, qui peuvent recombiner entre elles (Iliakis et al., 2015).

1.3.2.4 Réparation des cassures double brin par le Micro-homology Mediated End Joining

Le système de réparation par MMEJ (*Micro-homology mediated end joining*) est encore décrit comme étant principalement utilisé lorsque le système NHEJ n'est pas actif et qu'il manque un de ses acteurs principaux (Ku70 ou Ligase4) (Iliakis et al., 2015). C'est pourquoi il est aussi appelé *backup end-joining* ou *alternative end-joining*. Cependant, d'autres études ont montré que le MMEJ est aussi utilisé alors que les systèmes RH et NHEJ sont fonctionnels (Truong et al., 2013) et peut être indispensable pour réparer efficacement les CDB (Thyme and Schier, 2016). Dans cette dernière étude, les auteurs ont montré que le MMEJ est indispensable à la réparation des CDB pendant les premières étapes du développement du poisson zèbre. La réparation des CDB par MMEJ

repose sur une résection modérée des extrémités pour dévoiler de petites homologies (6 à 20 bases) qui vont s'apparier, stabiliser les extrémités et permettre la ligation finale. Lors de ce processus, la protéine PARP1 est capable d'entrer en compétition avec Ku70/Ku80 aux extrémités (Audebert et al., 2004). Elle semble jouer le rôle de plateforme pour recruter les différents acteurs de la réparation par MMEJ. Vient ensuite la résection 5'-3' des extrémités de la CDB par le complexe MRN/CtIP tout comme pendant la RH (Komatsu, 2016; Lee-Theilen et al., 2011; Truong et al., 2013; Wang et al., 2013; Xie et al., 2009). Les protéines Exo1 et BLM/DN2, quant à elles, ne sont pas nécessaires ici pour amplifier la résection (Deng et al., 2014). Les extrémités 3' sortantes non appariées sont éliminées par le complexe XPF-ERCC1 (*Xeroderma Pigmentosum complementation group F – Excision Repair Cross Complementation group 1*) pour permettre aux polymérases, dont la principale est PolQ, d'agir (Mateos-Gomez et al., 2015; McVey and Lee, 2008). Ces enzymes permettent notamment de combler les séquences manquantes après appariement des microhomologies. Vient ensuite la dernière étape de la réparation par MMEJ, la ligation finale. Cette étape est effectuée par la Ligase 3 et la Ligase 1 si la précédente n'est pas disponible (Audebert et al., 2004; Simsek et al., 2011).

1.3.2.5 Réparation des cassures double brin par le Single Strand Annealing

Ce système de réparation est assez proche du MMEJ dans son principe (Bhargava et al., 2016). En effet, il y a nécessité d'avoir résection des extrémités de la CDB pour révéler des régions homologues qui vont s'apparier et causer la perte des informations de séquence entre ces deux répétitions. Alors que le MMEJ repose sur de petites homologies (6 à 20 bases), le SSA requiert de longues homologies (100 bases environ). L'initiation de la résection est dépendante des enzymes CtIP et Rad50, ce dernier faisant partie du complexe MRN, tout comme pour la réparation par RH (Chen et al., 2008; Ivanov et al., 1996; Muñoz et al., 2012; Sartori et al., 2007). Les homologies découvertes peuvent maintenant s'apparier entre elles, avec l'aide de la protéine Rad52. Rad52

favorise en effet la fixation des ADN simple brin entre eux (Khade and Sugiyama, 2016; Mortensen et al., 2009; Rothenberg et al., 2008). ERCC1, complexé à XPF pour former une endonucléase, va ensuite permettre de retirer l'ADN simple brin non apparié pour permettre la ligation finale (Motycka et al., 2004). Pour finir il y a ligation des extrémités, cependant, les ligases spécifiques de la réparation par SSA ne sont pas clairement identifiées.

1.3.2.6 Types d'ADN donneurs utilisés pour l'édition du génome et les approches de Knock-In

Comme expliqué ici, différents systèmes peuvent permettre la réparation des CDB. Dans le cas de l'édition du génome, il est possible d'insérer un ADN donneur que la cellule utilisera comme matrice pour réparer la CDB. Le plus utilisé est un ADN double brin circulaire possédant environ 1kb d'homologie avec la séquence à réparer de chaque côté du transgène à insérer, comme le donneur p84 utilisé lors de ma thèse pour insérer le cDNA de la GFP au locus AAVS1 (DeKolver et al., 2010). Il est aussi possible d'utiliser un ADN double brin du même type mais linéarisé (Murphy, 1998; Reisch and Prather, 2015). L'insertion de ces deux types de donneurs repose sur la réparation par RH grâce aux longues homologies, et permet l'insertion d'un transgène très long allant jusqu'à plusieurs kb.

Il est aussi possible d'utiliser un ADN simple brin plus court, d'environ 100nt, possédant de courtes homologies, 20 à 50nt (Renaud et al., 2016). L'insertion de ce type de donneur repose sur un mécanisme encore mal compris. Il pourrait faire intervenir des étapes similaires au MMEJ ou SSA. De part les limites pratiques de la synthèse d'ADN simple brin, il permet l'insertion de petites séquences exogènes. Différentes approches de KI peuvent être utilisées selon l'objectif à atteindre.

1.3.3 Régulation de la réparation des cassures double brin par NHEJ et RH

Par exemple, en cas de dérèglements, pathologiques par exemple, le choix du système de réparation peut se trouver modifié. Si le phénomène de résection est initié pendant la phase G1 du cycle, la réparation par NHEJ non mutagène est compromise et sera remplacé par le système de réparation par MMEJ, plus mutagène (Bakr et al., 2016).

L'étape la mieux connue du choix du système de réparation est la résection (Li and Xu, 2016). Si la résection des extrémités de la CDB est débutée, la réparation se fera par RH ou par MMEJ et non par NHEJ. La régulation du choix du système de réparation est principalement influencée par le cycle cellulaire et plus précisément par deux protéines 53BP1 et BRCA1.

1.3.3.1 Mécanismes de régulation au cours du cycle cellulaire

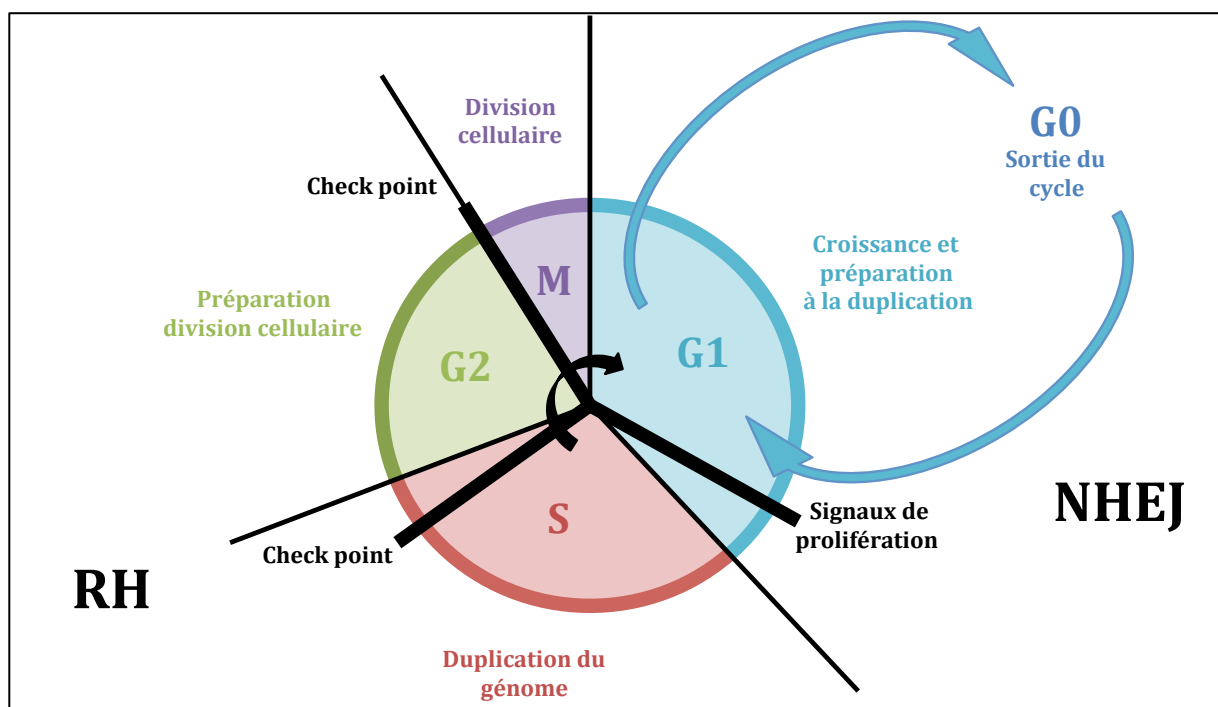


Figure 10. Les 5 phases du cycle cellulaire.

En phase G0 les cellules ne sont pas considérées dans le cycle cellulaire, elle ne sont alors plus en croissance. En phase G1 les cellules sont en croissance et se préparent à dupliquer leur génome. En phase S les cellules dupliquent leur génome. En phase G2 leur génome est dupliqué et les cellules se préparent à se diviser. Puis

en phase M les cellules se divisent pour engendrer 2 cellules filles identiques. Selon l'étape du cycle cellulaire, les CDB sont réparées avec différents systèmes de réparation. Si le NHEJ est utilisé tout au long du cycle cellulaire, la RH peut quant à elle seulement réparer les CDB pendant les phases S et G2.

La réparation peut se faire par NHEJ tout au long du cycle cellulaire mais ce système de réparation semble préférentiellement utilisé pendant la phase G1 alors que la RH, elle, est utilisée pendant les phases S et G2 du cycle cellulaire exclusivement (Figure 10).

Les CDKs ont un rôle majeur dans l'avancée de la cellule dans le cycle cellulaire. Elles permettent, entre autres, l'entrée en phase G2 et M (Trovesi et al., 2013). Elles utilisent aussi leur activité kinase pour réguler de nombreuses protéines impliquées dans les différents systèmes de réparation des CDB, dont les principales sont les suivantes (Cerqueira et al., 2009; Huertas and Jackson, 2009; Komatsu, 2016; Schwertman et al., 2016) : CtIP, Nbs1, EXO1, BRCA2, ATRIP ou encore BRCA1. La phosphorylation par les CDKs de EXO1 et CtIP est spécifique des phases G2 et S (Cerqueira et al., 2009). Les CDK régulent donc positivement les protéines impliquées dans la RH durant les phases S et G2, permettant notamment la résection des extrémités.

Une autre protéine régulée par le cycle cellulaire, ATM, permet la phosphorylation de MDC1 et donc son activation pendant les phases clés pour la RH. MDC1 permet le recrutement au site de la CDB des protéines MRN et BRCA1, protéines clés de la réparation des CDB par la RH (Liu et al., 2016; Wang et al., 2013).

D'autres enzymes, dont l'activité est dépendante du cycle cellulaire, sont aussi importantes ici, les E3 ubiquitine ligases. Parmi elles, il y a RNF138 (*Ring Finger protein 138*) qui ubiquitine CtIP et lui permet ainsi de rester à la CDB pendant les phases S et G2 (Schwertman et al., 2016). Une autre de ses cibles est la protéine Ku80 dont l'ubiquitination va déstabiliser l'hétérodimère Ku70/Ku80 et favoriser l'initiation de la résection pendant les phases S et G2 (Schwertman et al., 2016). Une autre E3 ubiquitine ligase est importante, Rad18, elle est effectivement nécessaire pour le recrutement de BRCA2 et Rad51 au site de coupure via la monoubiquitination de FANCD2 pendant la phase S du cycle (Tripathi et al., 2016). La protéine Rad18 permet aussi d'inhiber le système de réparation par NHEJ en bloquant le recrutement à la coupure de 53BP1 et

RNF168 (Helchowski et al., 2013). Une dernière E3 ubiquitine ligase, APC/C joue un rôle clé dans la régulation de protéines impliquées dans la réparation par RH pendant les S et G2. En effet, pendant ces phases, APC/C fixe Cdh1 puis CtIP pour réguler ce dernier. APC/C/Cdh1 permet notamment de favoriser la dégradation de CtIP et de limiter la résection permettant ainsi un bon chargement de rad51 (Lafranchi et al., 2014).

Pendant les autres phases du cycle cellulaire la kinase ATM joue aussi un rôle. En effet, en dehors de la phase G2, ATM est présente aux extrémités des CDB et permet la phosphorylation et l'activation de 53BP1, un acteur clé dans la régulation des systèmes de réparation comme expliqué dans le paragraphe 6.2.1 (Bakr et al., 2016). Pendant la phase G1 du cycle cellulaire, l'ubiquitination de CtIP par APC/C diminue le niveau de CtIP (Schwertman et al., 2016). Ces différentes activités favorisent donc l'utilisation du NHEJ comme système de réparation des CDB en dehors des phases S et G2.

Les différents types de régulation des protéines impliquée dans les 2 principaux systèmes de réparation des CDB sont résumés dans la Figure 11.

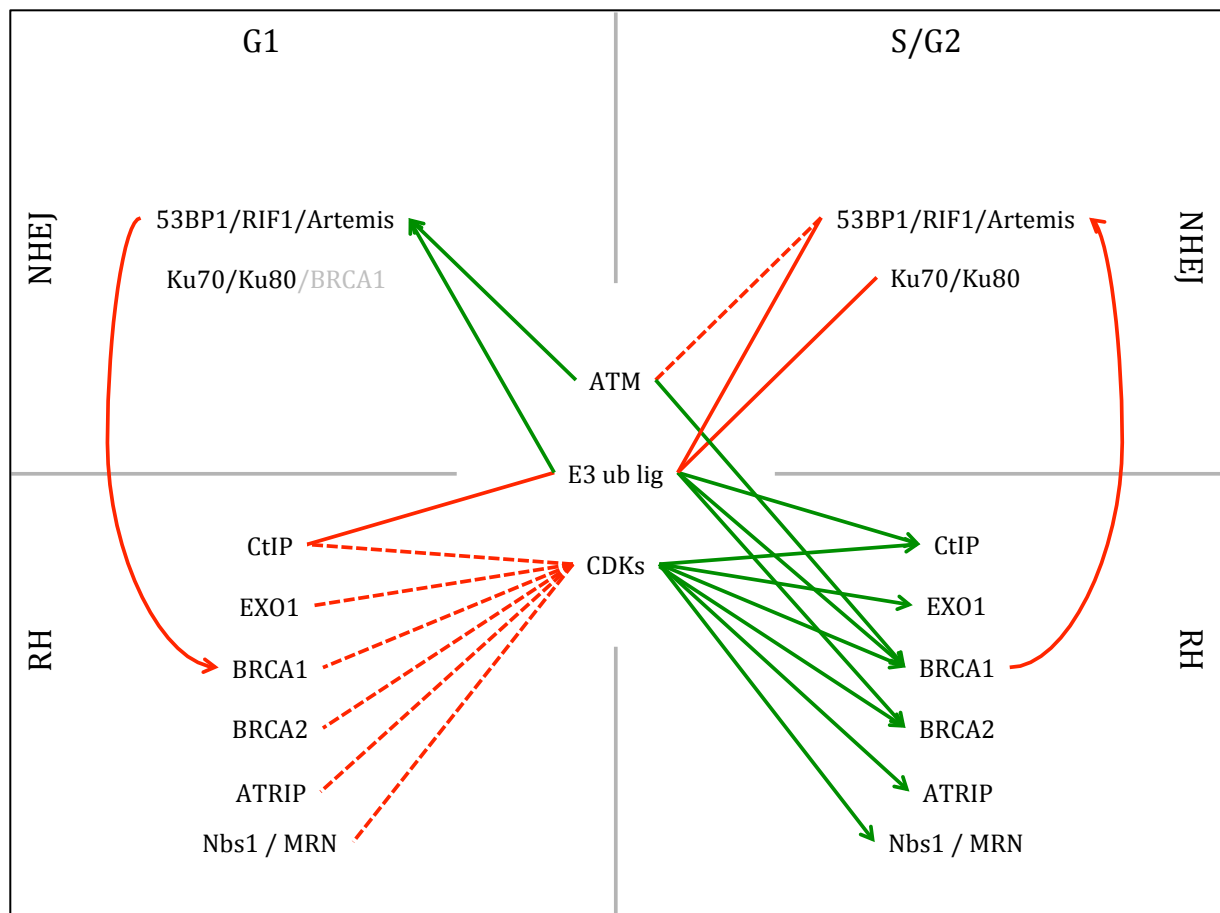


Figure 11. Liste non exhaustive des protéines impliquées dans la régulation des systèmes de réparation selon le cycle cellulaire.

Les différentes enzymes, dépendantes du cycle cellulaire, activent ou inhibent les protéines impliquées dans des systèmes de réparation. En vert sont schématisées les modifications permettant d'activer la protéine cible, en rouge ce sont celles qui inactivent les protéines et en rouges pointillé ce sont les modifications qui, n'ayant pas lieu, n'activent pas les protéines.

1.3.4 53BP1 et BRCA1, deux protéines clés dans la réparation des cassures double brin

Dans la régulation du choix du système de réparation des CDB, 2 protéines antagonistes jouent un rôle clé, 53BP1 et BRCA1, qui favorisent respectivement le NHEJ et la RH (Li and Xu, 2016).

1.3.4.1.1 Implication de la protéine 53BP1 dans le choix du système de réparation

53BP1 a été pour la première fois identifiée comme étant un partenaire de p53, supresseur de tumeur (Iwabuchi et al., 1994). Cette protéine est notamment impliquée dans la régulation de la mitose et lorsqu'elle est absente cause la présence de centromères surnuméraires (Yim et al., 2016). Ce rôle nécessite la régulation de 53BP1 par la kinase Plk1 et la déubiquitinase USP7.

Un autre rôle important de cette protéine, et plus décrit, est son implication dans la régulation de la réparation des CDB. 53BP1 favorise l'utilisation du NHEJ à plusieurs niveaux : en inhibant le phénomène de résection, en recrutant des protéines importantes au NHEJ mais aussi en inhibant directement la RH (Schwertman et al., 2016). Pour cela, cette protéine clé reconnaît les nucléosomes contenant l'histone H4K20me2 ainsi que l'histone H2A ubiquitiné sur sa lysine 15 (H2AK15ub) par RNF168, respectivement via ses domaines TUDOR et UDR (*Ubiquilation-dependant recruitment*) comme observé par la structure tridimensionnelle de 53BP1 fixé à une particule nucléosomique modifiée pour mimer les modifications d'histones précédemment décrites (Fradet-Turcotte et al., 2013; Wilson et al., 2016). Ce recrutement de 53BP1 par la chromatine est contrebalancé par l'action du complexe TIP60 (Jacquet et al., 2016). En effet, si TIP60 acétyle la lysine 15 de l'histone H2A, celui-ci ne peut être ubiquitiné et 53BP1 ne pourra plus s'y fixer. Inversement, si la lysine 15 de l'histone H2A est ubiquitiné par RNF168, TIP60 ne pourra plus s'y fixer. TIP60 entrera aussi en compétition directe avec 53BP1 en se fixant sur l'histone H4K20me2 via son partenaire MBTD1 (Jacquet et al., 2016). L'inhibition de ce nouveau partenaire dans la lignée cellulaire K562 (cellules humaines provenant d'une leucémie myéloïde), permet notamment de diminuer la réparation par RH et d'augmenter celle par NHEJ, par augmentation de la présence de 53BP1 au site de coupure.

Après avoir été recruté à la coupure, 53BP1, si elle est phosphorylée par ATM, active deux effecteurs, RIF1 et PTIP, qui vont ensuite pouvoir activer REV7 et Artemis, deux nucléases impliquées dans la maturation des extrémités complexes pour leur permettre

d'être réparées par NHEJ (Li and Xu, 2016; Schwertman et al., 2016). Le complexe 53BP1/RIF1 atténue aussi la présence de BRCA1 à la CDB pendant la phase G1 bloquant ainsi la résection des extrémités de la coupure et donc la réparation par RH (Escribano-Díaz et al., 2013). En revanche, lors des phases S et G2 du cycle, 53BP1 n'est plus phosphorylé par ATM et sa présence à la CDB est fortement diminuée (Bakr et al., 2016).

1.3.4.1.2 Implication de la protéine BRCA1 dans le choix du système de réparation

La protéine BRCA1 permet l'initiation de la résection directement via son interaction avec CtIP et le complexe MRN. Ce complexe BRCA1/CtIP permet aussi, par compétition de diminuer la présence de 53BP1/RIF1 à la coupure, principalement pendant la phase S du cycle cellulaire (Escribano-Díaz et al., 2013; Li and Xu, 2016; Schwertman et al., 2016). BRCA1 interagit aussi avec MRN et Rad51 pour promouvoir l'invasion de brin (Saha and Davis, 2016). L'implication de BRCA1 dans le système de réparation des CDB par RH n'est plus à prouver, cependant plusieurs études, contradictoires, montrent, ou non, une implication de BRCA1 dans la réparation par NHEJ mais aussi par MMEJ (Bau et al., 2004; Zhong et al., 2002). Ce rôle pourrait dépendre du cycle cellulaire. BRCA1 semble en effet promouvoir le NHEJ et bloquer le MMEJ en phase G1 (Dohrn et al., 2012), puis en phase S/G2, il semble bloquer le NHEJ et promouvoir la RH (Davis et al., 2014).

1.3.4.1.3 Régulation des deux acteurs 53BP1 et BRCA1

53BP1 et BRCA1 sont toutes deux régulées par RNF8 et RNF168, deux E3 ubiquitine ligases dont l'activité est régulée par le cycle cellulaire. D'autres cibles importantes de ces ubiquitine ligases sont les histones proches de la CDB qui vont ainsi permettre le recrutement de 53BP1 (*p53-binding protein 1*) et BRCA1 à la cassure et favoriser leur rétention (Li and Xu, 2016; Schwertman et al., 2016). Ce duo antagoniste est aussi fortement régulé par ATM, en fonction du cycle cellulaire, qui va permettre leur rétention à la cassure et leur activation.

Même si les rôles de ces deux antagonistes, 53BP1 et BRCA1, ne sont pas encore totalement compris, il est certain que ces deux protéines ont un rôle décisif dans le choix

du système de réparation qui sera utilisé et que ce contrôle est dépendant du cycle cellulaire.

1.4 Optimisation des approches d'édition du génome

Depuis l'apparition des nucléases artificielles (ZFN, TALEN et CRISPR-Cas9) il est beaucoup plus simple de modifier précisément le génome de n'importe quelle espèce. Afin d'augmenter cette efficacité et de faciliter l'obtention de mutants, différentes stratégies ont été développées.

1.4.1 Approches développées pour augmenter le taux de mutations

Différentes stratégies ont été mises en place pour tenter d'augmenter le taux de mutations obtenu après coupure par des nucléases. En effet, en réalisant une cassure double brin au sein d'un gène il est possible d'induire un changement du cadre de lecture, pour inactiver ce gène, quand la cellule utilise le système de réparation NHEJ et que des mutations apparaissent.

1.4.1.1 *Choc thermique*

L'une des premières techniques utilisée pour augmenter le taux de mutations après coupure avec des ZFN a été d'effectuer un choc thermique directement après transfection des nucléases (Doyon et al., 2010). Ce choc est effectué en mettant les cellules à 30°C après la transfection. En suivant ce protocole, une augmentation du taux de mutations a été observée. Ceci s'expliquerait par un arrêt du cycle cellulaire et/ou par une accumulation des nucléases dans les cellules. Cette stratégie a seulement été utilisée avec les ZFN, son efficacité avec les TALEN et le système CRISPR/Cas9 n'est pas établie.

1.4.1.2 Coexpression d'une exonucléase, Trex2, avec les nucléases artificielles

Une autre stratégie développée repose sur la coexpression de l'exonucléase Trex2. Cette protéine, fonctionnant en dimère, est une exonucléase 3'-5' non processive retrouvée chez l'homme ainsi que chez la souris, chez qui elle joue un rôle dans le maintien de l'intégrité du génome sans être nécessaire aux différents systèmes de réparation (Chen et al., 2007; Dumitrache et al., 2011). En surexprimant Trex2 avec la nucléase I-SceI, une augmentation du taux de mutations a été observée (Bennardo and Stark, 2010; Certo et al., 2012). Après coupure par I-SceI, Trex2 dégraderait les extrémités avant la réparation et conduisant à des mutations plus fréquentes.

Dans la continuité de cette stratégie, plusieurs autres protéines (Artemis, TDT, Rad2 et Exo1) ont été exprimées avec différents types de nucléases (TALEN, ZFN et I-SceI) et toutes permettent, dans une certaine mesure, d'augmenter le taux de mutations après coupure. Trex2 reste néanmoins la plus efficace (Certo et al., 2012).

1.4.2 Approches développées pour augmenter le taux de recombinaison homologue

Différentes approches ont aussi été développées jusqu'ici pour tenter de favoriser la réparation par RH et donc d'améliorer les possibilités de modification précise et programmée du génome.

1.4.2.1 Synchronisation des cellules en phase S et G2

La première approche développée a été de synchroniser les cellules en phase S ou G2, phases dans lesquelles la RH est active. Le nocodazole, qui bloque les cellules en fin de phase G2, et l'aphidicoline, qui bloque les cellules en phase S, ont été utilisés pour favoriser l'utilisation de la RH (Lin et al., 2014). Dans cette étude, une augmentation entre 55% et 100% du taux d'insertion par homologie d'un ADN simple brin a été observée avec ces deux agents pharmacologiques. Le même type d'approche a été utilisé

dans des cellules souches pluripotentes humaines cette fois ci (Yang et al., 2016). En utilisant de l'ABT-751 ou du nocodazole, jusqu'à 5 fois plus d'insertion ciblée par RH est obtenu dans différentes hPSC tout en gardant leur caractère pluripotent. Ces deux molécules agissent sur les microtubules et bloquent les cellules en phase G2/M. Le même type d'approche a été testée dans différentes levures (*S.Cerevisiae*, *Y. lipolytica*, *A.adeninivorans* etc) : en utilisant l'hydroxyurée pour stopper les cellules en phase S, 3 à 8 fois plus d'insertion ciblée par RH est alors obtenue (Tsakraklides et al., 2015).

Il est donc possible, dans de nombreux systèmes, d'augmenter l'utilisation du système de réparation RH en bloquant les cellules en phase S ou G2. Cependant, ce type d'approche pharmacologique pourrait modifier les propriétés cellulaires.

1.4.2.2 Ciblage de la coupure en phase S et G2 du cycle cellulaire

Une autre stratégie a été développée pour augmenter le taux de RH après coupure par le systèmes CRISPR-Cas9 en faisant de la nucléase Cas9 un substrat du complexe APC/Cdh1 (Gutschner et al., 2016; Howden et al., 2016). Ce complexe permet d'ubiquitiner et de dégrader les protéines substrats en phase M tardive et G1. Cette stratégie ne repose pas sur la synchronisation des cellules mais sur une expression ciblée de la nucléase et donc de la coupure pendant les phases du cycle cellulaire où la RH est utilisée. Le principe est de fusionner le domaine de dégradation de la protéine Geminin à Cas9, pour en faire une cible du complexe APC/Cdh1 (Sakaue-Sawano et al., 2008). La nucléase Cas9-Geminin produite est effectivement présente majoritairement pendant les phase S et G2 et cela semble pouvoir stimuler le taux de réparation par RH (Gutschner et al., 2016). La Cas9-Geminin permet d'obtenir 14% d'insertion par homologie contre 10% avec la Cas9wt. La stratégie est séduisante puisqu'elle ne demande pas de traitement particulier des cellules, il suffit d'utiliser une Cas9 « plus efficace » pour favoriser la RH. Cependant les effets restent assez faibles comparés aux stratégies précédentes et demande à être validés dans plus de types cellulaires. Très récemment cette fusion Cas9-Geminin a notamment été testée dans des iPSCs humaines

et aucune différence au niveau de la RH n'a été observée. Cependant le taux de mutations par NHEJ a diminué, expliqué par la diminution de l'expression de la nucléase en phase G1 (Howden et al., 2016). Cette observation semble logique, en effet, grâce à Geminin la nucléase Cas9 est peu exprimée pendant les phases durant lesquelles le NHEJ est très actif.

1.4.2.3 Modulation des voies de réparation

Une autre approche pour augmenter le taux de RH est de directement moduler les voies de réparation en ciblant les protéines impliquées.

1.4.2.3.1 Inhibition du système de réparation NHEJ

Le NHEJ entrant en compétition avec la RH pour la réparation des CDB, il a été la première cible pour augmenter le taux de RH après coupure par les nucléases. L'idée étant qu'en inhibant la réparation par NHEJ, la cellule va peut être compenser par le système RH.

1.4.2.3.1.1 Inhibition de la ligase 4

En utilisant une lignée de drosophile KO pour la ligase 4, une augmentation de l'utilisation de la RH de 3 à 20 fois après coupure par des ZFN, a été observée, que l'ADN donneur soit linéaire ou circulaire (Beumer et al., 2008). Dans ce système, le taux de mutations est aussi plus faible, indiquant bien une diminution de l'utilisation du système NHEJ. L'effet de l'absence de Lig4 sur l'intégration aléatoire d'un plasmide avec homologies a aussi été testé dans des lignées cellulaires de mammifères, DT40 (poulet) et des NALM6 (humain) (Iizumi et al., 2008). Une augmentation de 2 à 3 fois de l'insertion précise par RH du donneur a été observée. Les auteurs observent en effet une diminution dans ces cellules de l'insertion aléatoire du donneur qui pourrait être attribuée à la réparation par NHEJ. Cette approche a aussi été testée dans des levures de type *C. glabrata* (Cen et al., 2015). De manière naturelle cette levure recombine moins

que *S.cerevisiae* ce qui la rend plus difficile à étudier. Pour tenter de contrer cette déficience de réparation, une lignée de *C. glabrata* Lig4 KO a été développée. Cette dernière permet d'obtenir un ratio insertion ciblée/insertion aléatoire 13 à 35 fois plus élevé sur 3 gènes différents, ADE2, ARG8 et CNA1. Contrairement aux résultats obtenus dans les cellules humaines, l'inhibition du système de réparation NHEJ ne stimule pas ici la RH mais diminue les insertions aléatoires, connues pour être dépendantes du NHEJ. Ces effets contradictoires pourraient être expliqués par le fait que le système de réparation par RH est peu disponible par rapport au NHEJ dans *C. glabrata*. Le fait de bloquer le NHEJ diminuerait ses conséquences et permettrait de faire ressortir les réparations effectuées par le système RH.

Cette stratégie d'inactiver la ligase 4 a même été utilisée chez le riz (Nishizawa-Yokoi et al., 2012). Les auteurs ont alors observé une augmentation de la réparation par RH plus importante dans le riz KO pour la ligase 4 en utilisant un rapporteur portant le site pour l'enzyme I-SceI.

L'utilisation de lignées ou d'animaux KO pour la ligase 4 a prouvé qu'en inhibant le système de réparation NHEJ on pouvait, dans certains cas, stimuler l'utilisation de la RH. L'utilisation de cellules KO pour la ligase 4 est une stratégie parfaite pour établir la preuve de concept, mais ce type de cellule ne peut être utilisé pour créer de nouvelles lignées par édition du génome. En effet, les cellules dont le NHEJ n'est pas fonctionnel vont sans doute compenser ce défaut en utilisant le système MMEJ pour réparer les CDB, système plus mutagène. D'autres stratégies ont donc été développées visant à inhiber le système de réparation par NHEJ de manière transitoire.

La molécule SCR7 a été identifiée par l'équipe de Raghavan comme étant un inhibiteur spécifique de la ligase 4 (Srivastava et al., 2012). Partant des travaux de Chen et al., de 2008 décrivant un inhibiteur (L189) des 3 ligases (Ligase 1, 3 et 4). Srivastava et al. ont utilisé la formule du L189 et la structure du domaine de liaison à l'ADN des ligases pour développer de nouveaux inhibiteurs plus spécifiques. Selon leurs travaux, la molécule

SCR7 permet de cibler spécifiquement la ligase 4 en bloquant son domaine d'interaction à l'ADN. Différents tests montrent une diminution de l'efficacité du NHEJ en présence de SCR7. A la suite de ces travaux, SCR7 a été utilisé comme inhibiteur du NHEJ dans des cellules de mammifères (humaines et murines) pour augmenter le KI par RH après coupure par le système CRISPR-Cas9 (Maruyama et al., 2015). 13 fois plus d'insertion par RH du donneur a ainsi été obtenue. A la suite de ces résultats, de nombreuses équipes ont tenté d'utiliser cette molécule pour augmenter le taux de RH dans leurs cellules, malheureusement les résultats ne semblaient pas reproductibles (Pinder et al., 2015). En mai 2016 a été publié un article remettant en cause les résultats obtenus par l'équipe de Raghavan (Greco et al., 2016). Ils montrent que le protocole donné pour synthétiser SCR7 ne donne pas la molécule avec la structure annoncée. Ils ont ensuite effectué plusieurs tests et montrent que ce pseudo SCR7 inhibe principalement la ligase 1 et non la ligase 4 et qu'il n'y a pas de diminution de la recombinaison par V(D)J, dépendante du NHEJ. Les résultats obtenus par Maruyama et al sur la RH ne sont sans doute pas dus à l'inhibition du NHEJ et restent à mieux comprendre.

Un autre groupe a publié en 2015 des résultats montrant qu'en utilisant des protéines de l'adénovirus humain Ad4, Ad4E1B55K et Ad4E4orf6, le taux de RH après coupure par le système CRISPR/Cas9 peut être augmenté (Chu et al., 2015). A l'aide d'un rapporteur intégré au site AAVS1 dans des cellules HEK293 ils peuvent facilement mesurer le taux de mutations via NHEJ qui aboutit à une expression de RFP, et de réparation par RH qui aboutit à une expression de Venus. Ils observent alors une diminution du NHEJ et une augmentation de la RH, en effet ils passent de 8% de cellules Venus positives à 66%. Ces protéines Ad4E1B55K et Ad4E4orf6 permettent l'ubiquitination et la dégradation via le protéasome de différentes protéines dont la ligase 4 (Forrester et al., 2011). Pour ce faire, E4orf6 recrute Cul5, les elongins B et C, Rbx1 qui vont ensuite recruter E1B55K qui va s'associer à E4orf6 pour induire l'ubiquitination des protéines et leur dégradation par le protéasome (Cheng et al., 2011). En plus de la ligase 4, ces protéines virales permettent aussi de dégrader efficacement Mre11 et p53, pouvant aussi avoir un effet

simultané sur plusieurs systèmes de réparation cellulaires (Cheng et al., 2011; Forrester et al., 2011).

Ces différentes stratégies ont montré qu'en inhibant la ligase 4 il est possible d'obtenir une meilleure efficacité d'édition du génome précise par RH. Cet effet peut être expliqué par la compensation du système NHEJ, non disponible, par la RH. Une autre explication possible pourrait être la mort des cellules engagées dans la voie NHEJ ne pouvant religuer les extrémités de la coupure dû à l'absence de la Ligase 4.

1.4.2.3.1.2 Inhibition de Ku70

Une autre cible spécifique de la réparation par NHEJ est Ku70, protéine qui a l'avantage, contrairement à la Ligase 4, d'intervenir au tout début de la réparation par NHEJ.

Une lignée de cellules embryonnaires de souris a ainsi été modifiée pour contenir dans son génome un rapporteur de type DR-GFP (Pierce et al., 2001). Ce type de rapporteur contient 2 copies inactives de la GFP, dont une contenant le site de clivage de l'enzyme I-SceI. Une fois que I-SceI clive sa cible, il est possible que la CDB soit réparée par recombinaison intramoléculaire, prenant comme matrice la deuxième copie de la GFP et permettant l'expression de la protéine fluorescente. Les auteurs ont ensuite modifié cette lignée pour obtenir des cellules Ku70^{-/-} pour observer l'effet de cette mutation sur la réparation du rapporteur après coupure par I-SceI. Dans ces cellules Ku70 KO, il y a entre 2 et 3 fois plus de cellules GFP positives. Cette approche a aussi été utilisée dans des levures de type *Penicillium digitatum* (Gandía et al., 2016, p. 70). Les auteurs ont ciblé les gènes *chsII* et *chsV* pour y insérer un gène de résistance à l'hygromycine, dans les levures ku70 KO et obtiennent 19 fois plus d'insertion ciblée. Cependant, les levures ku70 KO sont plus sensibles à la chaleur et moins tolérantes au stress de l'ADN. Ce type de mutant augmente effectivement le taux de réparation par RH mais semble grandement affecter les propriétés de ces levures (Cen et al., 2015).

Différentes stratégies ont aussi été développées dans des HCT116 pour tester l'effet d'une diminution transitoire de Ku70 sur le taux de recombinaison homologue (Fattah

et al., 2008). Ne pouvant obtenir un clone Ku70 KO, la protéine étant essentielle pour la survie de ce type cellulaire, ils ont testé la diminution de la protéine Ku70 par différentes approches, lignées Ku70+/-, extinction par siRNA. Dans tous les cas, ils ont observé une augmentation de la recombinaison homologue intramoléculaire entre 3 et 6 fois sur différentes cibles, Ku70, CCR5 et Lig4. Un effet synergique a aussi été observé quand les cellules Ku70+/- étaient en plus traitées par un siRNA dirigé contre Ku70, augmentant alors le taux de recombinaison intramoléculaire jusqu'à 30 fois.

1.4.2.3.2 Stimulation de la modification précise par *Knock-In*

Différentes stratégies ont été développées pour stimuler l'utilisation de la RH comme système de réparation des CDB induites par les nucléases artificielles au cours de l'édition du génome. Différentes protéines impliquées dans la recombinaison homologue ont donc été ciblées, Rad51, qui permet la recherche d'homologie et BRCA1 qui oriente le choix du système de réparation de la CDB vers la RH.

1.4.2.3.2.1 Stimulation du filament de Rad51

L'utilisation de la molécule RS-1 permet d'augmenter le taux de recombinaison homologue après cassure double brin (Song et al., 2016). Cette molécule, identifiée par Jayathilaka et al., permet de stimuler la formation du filament de Rad51 sur un ADN simple brin en le stabilisant (Jayathilaka et al., 2008). Cette particularité permet un effet marqué sur le taux de RH dans des embryons de lapin à différents locus (Rosalike et rbApoAI) en utilisant TALENs et CRISPR-Cas9 comme nucléases (Song et al., 2016). La molécule RS-1 a aussi été testée sur des cellules humaines HEK293A à 2 locus différents, LMNA et PML pour y insérer le cDNA de la protéine fluorescente Clover. Sur les deux gènes le taux de recombinaison homologue est augmenté de 3 à 6 fois grâce à l'utilisation de RS-1 (Pinder et al., 2015).

1.4.2.3.2.2 Stimulation de BRCA1

L'effet de la surexpression de BRCA1 sur le taux de réparation par RH a été testé dans des cellules de la lignée HCC1937, contenant le rapporteur DR-GFP, un taux de RH

intramoléculaire 4 fois plus élevé que dans les cellules contrôle a pu être observé (Dever et al., 2011). Comme vu précédemment, BRCA1 est une protéine clé dans le choix du système de réparation des CDB. Sa surexpression semble privilégier la RH au détriment du NHEJ dans la réparation des CDB induites par les nucléases, cela reste à confirmer.

1.4.2.3.3 Stimulation de l'insertion de mutations ponctuelles et précises

Pour finir, une approche utilisant la nucléase Cas9 et une cytosine déaminase, APOBEC1, a été développée pour permettre un changement de base de l'ADN sans cassure double brin (Komor et al., 2016). APOBEC1 a été fusionnée en Nter de la forme mutante de la protéine Cas9, dépourvue de son activité nucléase. Cela permet une conversion des bases C en U de manière ciblée. Cette fusion permet d'obtenir jusqu'à 8% de modification précise contre 0,5% avec une approche classique, avec la Cas9 et un donneur par RH. Cette approche est intéressante et permet d'obtenir des mutations ponctuelles à un taux raisonnable, cependant, cela peut induire des mutations non souhaitées et ne permet pas d'autres mutations, en dehors de convertir des bases C en U.

1.4.3 Bilan

Plusieurs stratégies, plus ou moins efficaces, ont été développées pour augmenter soit, le taux de mutations, soit, le taux de RH, engendrés par la coupure par une nucléase artificielle. Ces différentes stratégies possèdent des avantages et des inconvénients dont il faut tenir compte. L'utilisation de lignées ou d'organisme KO pour une protéine (ligase 4 ou Ku70) est une approche très efficace puisque la protéine cible est complètement éliminée, mais cette approche peut entraîner de nombreux problèmes. Le système de réparation par NHEJ étant affecté, il est possible que les cellules soient sensibles aux dommages de l'ADN et qu'elles accumulent avec le temps de nombreuses mutations non souhaitées. Si l'utilisation d'un système KO est indispensable il serait alors préférable d'utiliser un KO réversible. Les différents traitements pharmacologiques, par interférence ARN ou par surexpression de protéines adénovirales permettent d'atténuer

les problèmes précédemment énoncés puisqu'ici l'inhibition ou la stimulation de la protéine d'intérêt est transitoire. Cependant, Les traitements peuvent modifier le comportement de la cellule, par exemple en modifiant la pluripotence des cellules souches. Il faut donc tester différentes conditions et trouver la plus efficace pour le type cellulaire d'intérêt. L'approche développée pour augmenter le taux de mutations par surexpression de Trex2 est intéressante mais pourrait aussi engendrer des effets non souhaités, la surexpression d'une exonucléase comme Trex2 pourrait être toxique pour la cellule et causer des mutations *off-target*.

2 Projet de thèse et résultats

Mon projet de thèse a eu pour objectif d'optimiser les techniques d'édition du génome en augmentant le taux de mutations ou le taux d'insertion d'un transgène par homologie après coupure par des nucléases artificielles. Les cassures double brin sont alors induites par des TALEN ou par le système CRISPR/Cas9 dans différentes lignées cellulaires.

Les objectifs de ma thèse sont :

- I) Augmenter le taux de KI induit par les nucléases artificielles en surexprimant, ou fusionnant à la nucléase Cas9, CtIP, protéine impliquée directement dans le système de réparation par RH.
- II) Augmenter le taux de mutations induit par les nucléases artificielles en surexprimant, ou fusionnant à la nucléase Cas9, l'exonucléase Trex2.
- III) Produire et développer des nucléases TALE à façon comme outils de modification du génome.

2.1 Article : Stimulation du taux d'insertion d'un ADN exogène en fusionnant la partie Nter de la protéine CtIP à la nucléase cas9.

Dans cette partie, je me suis intéressée à stimuler l'insertion d'un donneur exogène par RH après coupure par le système CRISPR/Cas9. Le recrutement de la protéine CtIP au site de coupure par les nucléases artificielles a montré des résultats intéressants. J'ai pu définir un domaine minimum suffisant pour stimuler le KI en fusion à la protéine Cas9. Ce domaine est composé d'un domaine de tétramérisation de CtIP indispensable à son effet sur la réparation de la cassure double brin et de plusieurs sites de phosphorylation par CDK, eux aussi indispensables.

2.2 Stimulation du taux de mutations en fusionnant l'exonucléase Trex2 à la nucléase Cas9

Comme présenté en introduction, différentes approches ont été développées pour augmenter le taux de mutations à la suite d'une coupure par une nucléase artificielle. Ce type d'approche peut permettre d'augmenter le taux de KO obtenu après utilisation de nucléases artificielles, que ce soit dans une lignée cellulaire ou *in vivo*.

Les stratégies qui ont attiré notre attention, et qui nous ont permis de développer ce projet, sont celles basées sur l'exonucléase Trex2 (Bennardo and Stark, 2010; Certo et al., 2012). En surexpression ou fusionnée à l'enzyme I-SceI, ou en surexpression avec les nucléases ZFN et TALEN, cette exonucléase permet d'augmenter le taux de mutations au site de coupure. J'ai testé cette stratégie avec le système CRISPR/Cas9, en surexprimant l'exonucléase Trex2 ou en la fusionnant en Cter de Cas9, afin de cibler son effet au site de coupure. Comme contrôle, le mutant Trex2(H188A) qui n'a plus d'activité nucléase, est utilisé en parallèle de Trex2 (Perrino et al., 2005).

2.2.1 Surexpression et fusion de l'exonucléase Trex2 à la nucléase Cas9

Dans un premier temps, j'ai évalué l'impact sur le taux de mutations après coupure par le système CRISPR-Cas9 de l'exonucléase Trex2, en surexprimant cette protéine ou en la fusionnant à la nucléase (Figure 12).

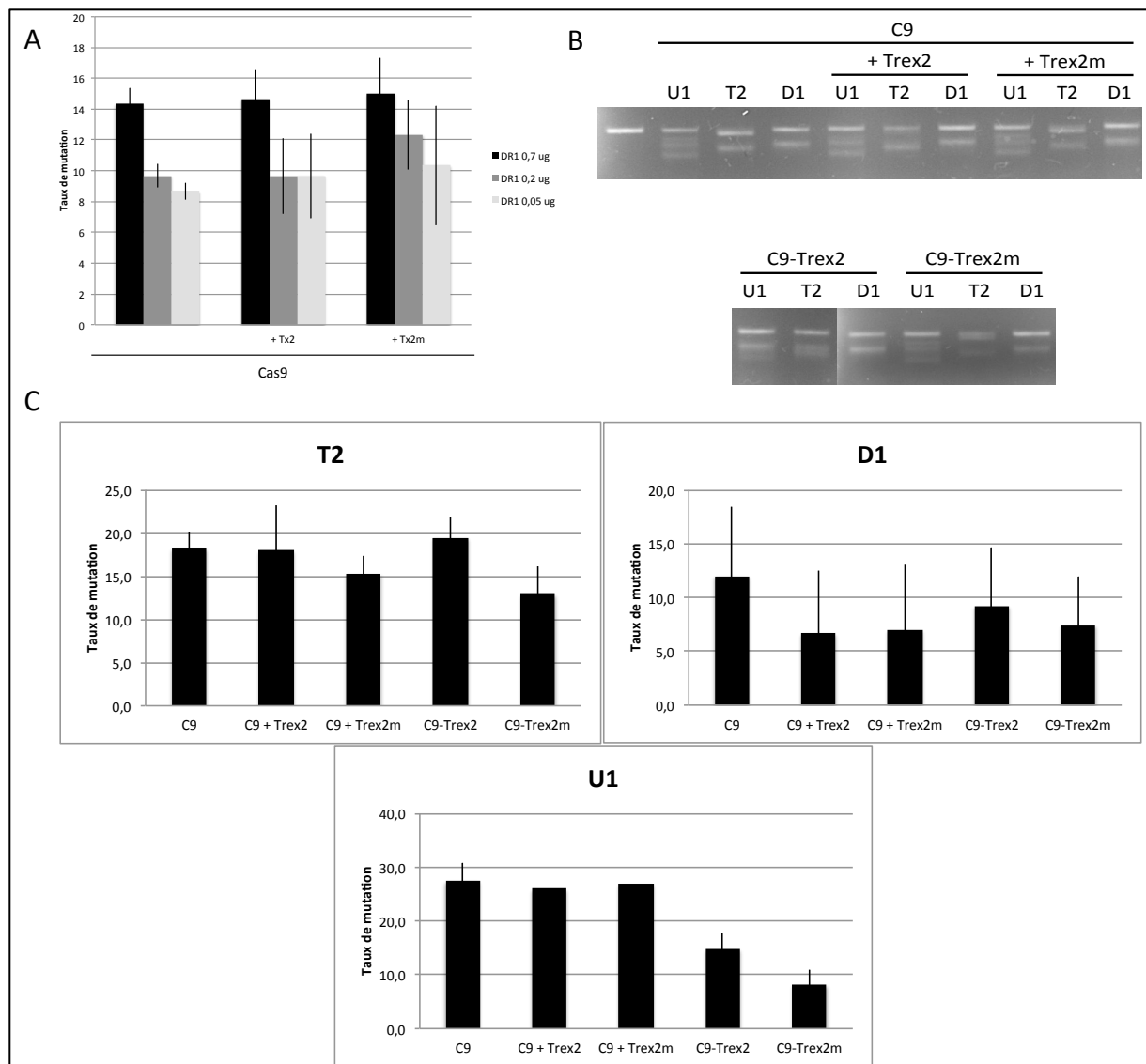


Figure 12. Effet de l'exonuclease Trex2 sur le taux de mutations après coupure par le système CRISPR/Cas9. (A) Taux de mutations observé par test T7E1 après transfection de la nucléase Cas9 wt et de différentes concentrations de l'ARNg DR1 dirigé contre le locus DR-GFP avec ou sans surexpression de Trex2 et Trex2 muté (H118A) dans des cellules de la lignée RG37DR. (B) Exemple de gel de test T7E1 obtenu après transfection de Cas9 wt avec ou sans surexpression de Trex2 et Trex2 muté ou de la nucléase Cas9 fusionnée aux différents Trex2 dans des cellules de la lignée U2OS avec différents ARNg (U1, T2 et D1) dirigés contre le locus AAVS1. (C) Taux de mutations observés par test T7E1 après transfection de Cas9 wt, avec ou sans surexpression de Trex2 ou de son mutant, et de Cas9 fusionnée à Trex2 ou Trex2 muté avec différents ARNg (T2, D1 et U1), dirigés contre le locus AAVS1, dans des cellules de la lignée U2OS.

Aucun changement significatif du taux de mutations n'a été observé lors de l'utilisation de Trex2, que ce soit en surexpression ou en fusion avec la Cas9 wt.

Dans les conditions expérimentales utilisées, l'utilisation de Trex2 en surexpression ou en fusion avec le système CRISPR-Cas9 ne permet pas d'augmenter le taux de mutations au site de coupure.

2.2.2 Recrutement de l'exonucléase Trex2 à la cassure double brin par fusion à la nickase Cas9 (D10A)

J'ai ensuite cherché à augmenter le taux de mutations obtenu avec la Cas9 nickase (mutation D10A) en utilisant 2 ARN guides pour effectuer une CDB. Différents couples d'ARN guides dirigés contre le locus AAVS1 ont été utilisés, permettant de tester les nickases modifiées, fusionnées avec Trex2 ou Trex2(H188A), sur différents substrats. En effet, selon le positionnement des ARNg les nickases vont créer des extrémités 3' ou 5' sortantes. Cette stratégie a été testée dans deux lignées cellulaires différentes, U2OS et RG37DR.

Utilisation de Cas9(D10A)-Trex2 dans les cellules U2OS

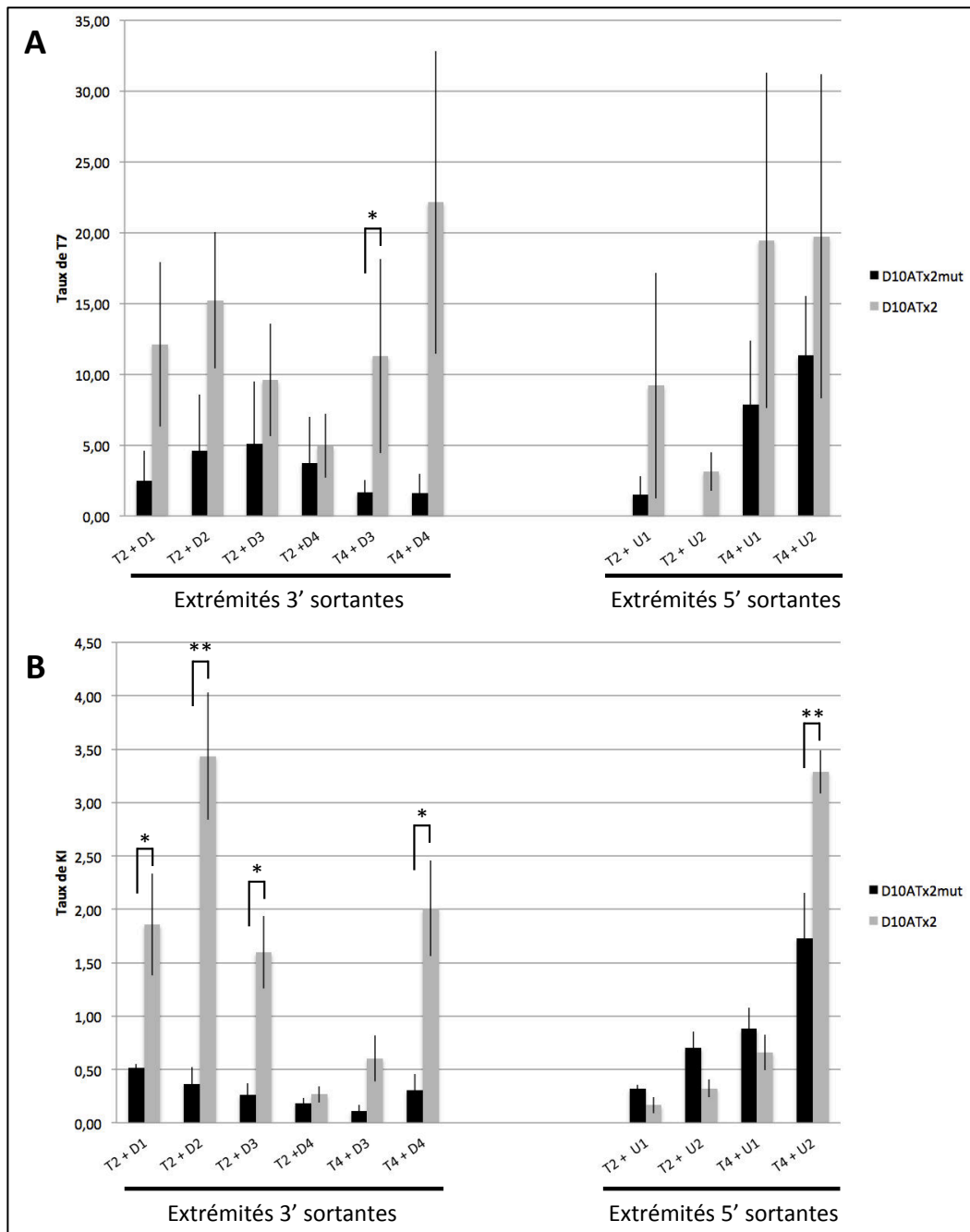


Figure 13. Effet de la fusion de l'exonucléase Trex2 à la nucléase Cas9(D10A) sur le taux de mutations dans les cellules U2OS.

(A) Taux de mutations mesurés par test T7E1. (B) Taux de KI par homologie d'un donneur avec une séquence GFP, obtenu par quantification du taux de GFP par FACS. Afin de déterminer si les moyennes des taux de mutations obtenus avec les différentes nucléases sont différentes un t-test a été utilisé (n = 5).

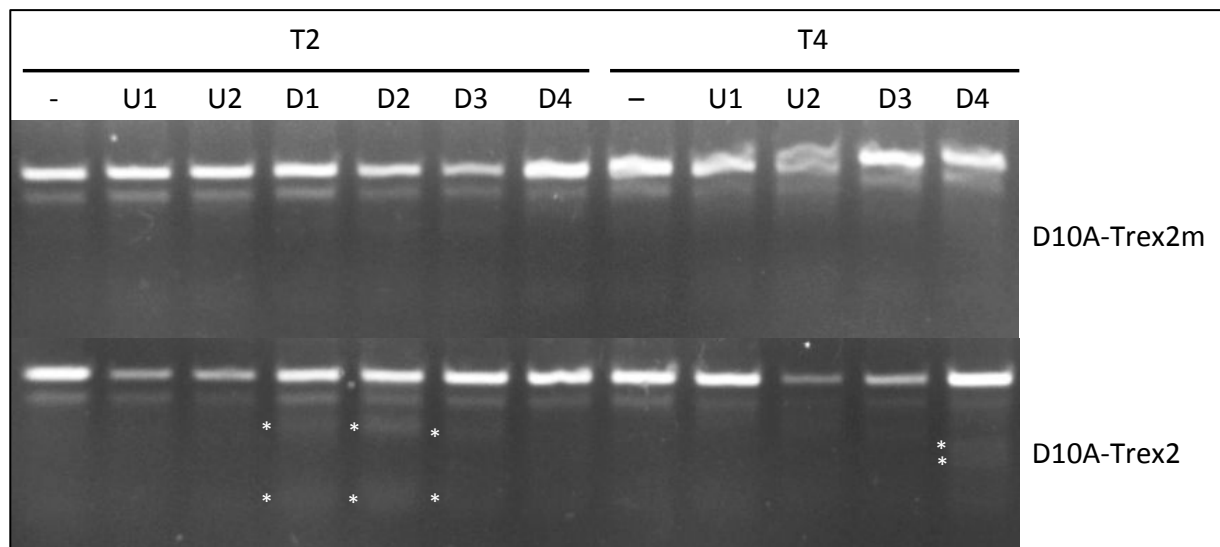


Figure 14. Exemple de gel agarose (2,5%) obtenu après transfection d'U2OS avec Cas9 nickase fusionnée à Trex2 ou Trex2 muté et test T7E1.

Les étoiles blanches représentent les bandes clairement présentes seulement lorsque la Cas9 nickase est fusionnée à Trex2.

Avec l'utilisation de Trex2 fusionnée à la Cas9 nickase, le taux de mutations semble être plus important. Suivant les ARNg utilisés, les augmentations ne sont pas toujours significatives (Figure 13). Lorsque l'on observe les profils de T7E1 sur gel, on observe clairement des changements dus à l'augmentation des mutations après réparation de la CDB (Figure 14).

Le taux d'insertion du donneur GFP quant à lui est clairement modulé avec l'utilisation de l'exonucléase. On note qu'il est plus important pour les couples d'ARNg qui produisent des extrémités 3' sortante que pour les couples qui produisent des extrémités 5' sortantes après coupure.

Utilisation de Cas9(D10A)-Trex2 dans les cellules RG37DR

Compte tenu des résultats encourageants obtenus dans les U2OS, j'ai testé tester la Cas9(D10A)-Trex2 dans une seconde lignée cellulaire, les fibroblastes RG37DR. Dans ces cellules, dans les conditions usuelles avec la nickase Cas9 classique tout comme avec la nickase fusionnée avec Trex2 muté, il est difficile de détecter les mutations induites, avec le test T7E1.

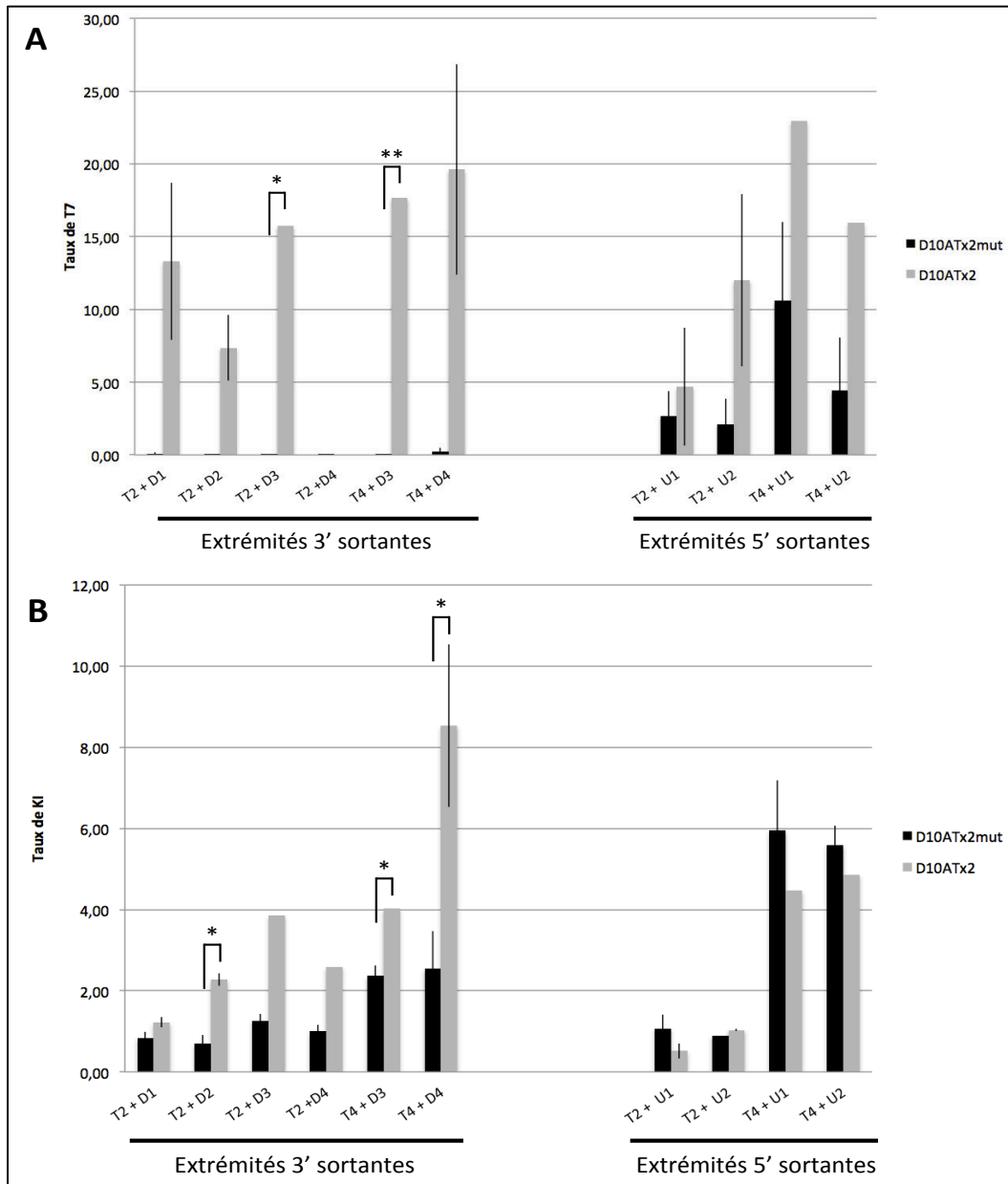


Figure 15. Effet de l'exonuclease Trex2 sur le taux de mutations après coupure par le système CRISPR/Cas9(D10A) dans les cellules RG37DR.

(A) Taux de mutations obtenu par test T7E1. (B) Taux de KI par homologie obtenu par quantification du taux de GFP par FACS. Afin de déterminer si les moyennes des taux de mutations obtenus avec les différentes nucléases sont différentes un t-test a été utilisé (n = 5).

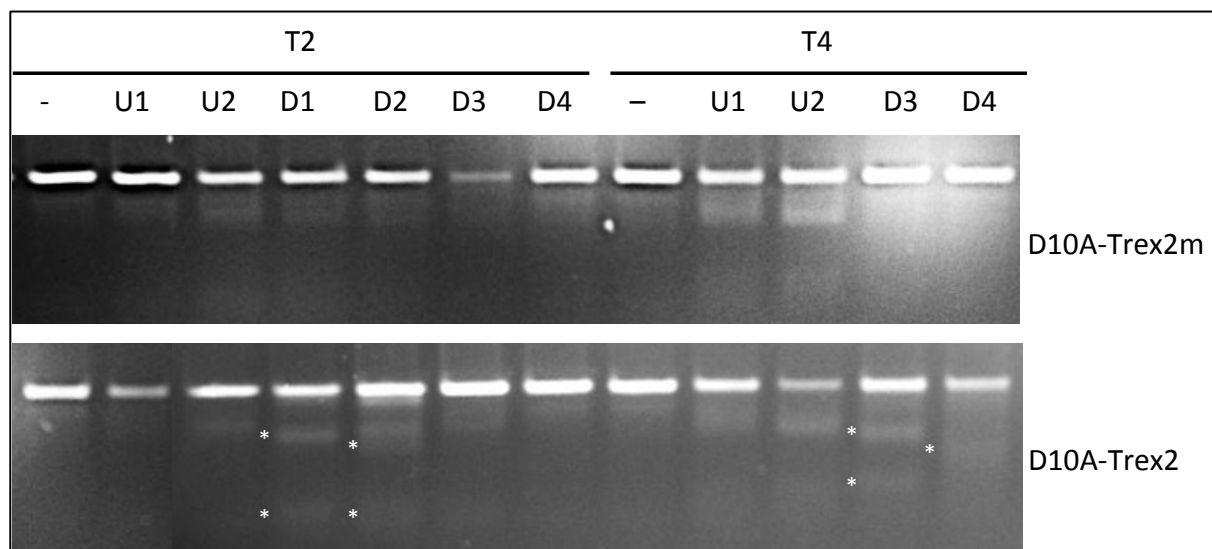


Figure 16. Exemple de gel de test T7E1 obtenu après transfection de RG37DR avec Cas9 nickase fusionnée à Trex2 ou Trex2 muté.

Les étoiles blanches représentent les bandes clairement présentes seulement lorsque la Cas9 nickase est fusionnée à Trex2.

On observe ici le même type de résultats que dans les U2OS, le taux de mutations est clairement augmenté, surtout pour les couples d'ARN guide formant des extrémités 3' sortantes après coupure (Figure 15 et Figure 16). Les niveaux d'augmentation sont plus marqués puisqu'avec l'utilisation de la Cas9(D10A)-Trex2m il est difficile d'observer un taux de mutations par test T7E1 dans ces cellules. Le taux d'insertion de la GFP par homologie est lui aussi augmenté avec l'utilisation de l'exonucléase Trex2.

2.2.3 Conclusion et discussion

J'ai donc observé qu'une fois fusionnée à la nucléase Cas9 wt ou en surexpression simple avec cette nucléase, l'exonucléase Trex2 ne permettait pas d'augmenter le taux de mutations après réparation, alors qu'avec l'utilisation de Cas9 nickases (D10A) j'obtiens une augmentation du taux de mutations et aussi d'insertion par homologie d'un donneur exogène codant pour la GFP. Ce résultat semblait assez surprenant comparé aux

résultats déjà obtenus avec Trex2 (Bennardo and Stark, 2010; Certo et al., 2012). Les auteurs obtenaient alors des changements avec l'utilisation de ZFN, de TALEN et avec I-SceI, alors qu'avec la Cas9 wt cette approche ne semble pas fonctionner.

Que ce soit dans les U2OS ou dans les RG37DR, l'utilisation de Trex2 fusionnée à Cas9 (D10A) dirigée contre le site AAVS1 permet d'augmenter le taux de mutations et de KI principalement quand les ARN guides utilisés sont disposés de sorte que les nickases créent des extrémités 3' sortantes.

Cette nouvelle nickase est testée en ce moment en parallèle de la Cas9(D10A) classique, sans fusion. Les premiers résultats semblent montrer que le fait de fusionner Trex2(H188A) à la nickase réduit considérablement les mutations induites et le taux de KI. Alors que quand on fusionne Trex2wt à la nickase on obtient des taux de mutation et de KI plus élevés qu'avec la Cas9(D10A) classique. Cette manipulation doit être reproduite afin de valider ces résultats.

2.2.3.1 L'exonucléase Trex2 dégrade préférentiellement les extrémités 3' sortantes

Il a été montré que l'exonucléase Trex2 dégrade plus efficacement les substrats avec 1 à 3 nucléotides 3' sortants en comparaison avec des substrats aux extrémités franches (Mazur and Perrino, 2001). Cette observation permet d'expliquer pourquoi la surexpression de Trex2 ou sa fusion à la nucléase Cas9 wt ne change pas le taux de mutations puisque Cas9 laisse des extrémités franches, contrairement aux Cas9 nickases, aux TALEN, aux ZFN et à I-SceI. Cette observation permet aussi d'expliquer pourquoi la nickase Cas9 fusionnée à Trex2 induit plus de mutations quand les paires de guides utilisés donnent des extrémités 3' sortantes.

Il a aussi été montré que Trex2 peut dégrader l'ADN double brin en permettant l'ouverture de la double hélice afin d'agir sur son substrat favori, l'ADN simple brin (de Silva et al., 2009). Cette particularité n'est sans doute pas applicable avec la fusion à Cas9wt car Trex2 est bloquée dans l'espace, mais en surexpression cela aurait pu fonctionner contrairement à ce qui a été observé. Cette capacité à ouvrir l'ADN double

brin ne devait pas être assez puissante, dans le contexte cellulaire les extrémités sont peut être protégées par des protéines telles que Ku ou la Cas9 elle-même.

Peut-on forcer le système et réussir à augmenter le taux de mutations avec la Cas9 wt ?

La protéine Trex2 fonctionne en dimère pour dégrader l'ADN (de Silva et al., 2009). Peut être qu'en surexprimant Trex2 en parallèle de la Cas9 wt fusionnée à Trex2 on pourrait augmenter le taux de mutations même avec des extrémités franches. Il a été montré que dans certains types cellulaires, comme les HEK293, le taux de Trex2 endogène est faible, ce qui pourrait être limitant pour la dégradation des extrémités franches (Chen et al., 2007). Il serait intéressant de surexprimer Trex2 en parallèle de la Cas9-Trex2 pour compenser l'absence de Trex2 endogène. Si c'est le taux de Trex2 endogène qui est limitant on pourra peut être observer une augmentation du taux de mutations à la coupure même avec l'utilisation de la Cas9 wt qui crée des extrémités franches.

2.2.3.2 L'utilisation de l'exonucléase Trex2 modifie aussi le taux d'insertion d'un ADN exogène

Le fait que le taux d'insertion du donneur GFP par homologie soit lui aussi augmenté dans la plupart des cas avec l'utilisation de la Cas9 nickase fusionnée à Trex2 est inattendu. En effet, l'insertion par RH nécessite des extrémités simple brin 3' sortantes alors que l'exonucléase Trex2 dégrade cet ADN simple brin. Cependant il est possible d'imaginer que le fait de dégrader les quelques nucléotides sortants permette d'inhiber la réparation par NHEJ non mutagène en supprimant les extrémités cohésives de la coupure. Le NHEJ non mutagène étant impossible après cette dégradation, la CDB peut donc être réparée par NHEJ mutagène (augmentation observée du taux de mutations) ou alors par RH qui n'était jusqu'alors pas favorisé avec ces extrémités cohésives plus stables, aboutissant alors directement à une augmentation du taux d'insertion du donneur GFP.

2.2.3.3 Conclusion

L'utilisation de l'exonucléase Trex2 permet d'augmenter le taux de mutations ainsi que le taux d'insertion d'un ADN exogène par homologie au site de coupure lorsqu'on utilise deux Cas9 nickases mais pas avec la Cas9 wt. Ce système est plutôt simple à utiliser puisqu'il ne nécessite pas de cotransfection d'un autre plasmide. Cependant, contrairement à la Cas9 wt il faut ici utiliser deux ARNg.

Il serait d'ailleurs intéressant de voir si l'utilisation de Cas9(D10A)-Trex2 augmente le taux de mutations lorsqu'il y a juste un nick et donc voir s'il y a un effet aux sites off target.

La Cas9(D10A)-Trex2 pourrait représenter une alternative intéressante à la Cas9(D10A) classique, elle permettrait d'augmenter le taux de mutations obtenu après coupure et pourrait faciliter la création de cellules ou d'organismes KO.

2.3 Assemblage de TALEN pour l'édition du génome

Au début de ma thèse, j'ai travaillé sur la production de TALEN. J'ai participé à la mise en place de cette production, à son optimisation ainsi qu'à des essais pour son automatisation. L'assemblage des TALEN étant difficile, de nombreuses équipes se sont tournées vers des structures spécialisées pour leur assemblage, comme TACGENE, la plateforme au sein de laquelle j'ai travaillé pendant ma thèse. Je vais vous présenter ici la technique d'assemblage des répétitions TALE qui a été mise en place ainsi que plusieurs collaborations qui ont eu lieu.

2.3.1 Méthode d'assemblage des répétitions TALE

La technique utilisée au laboratoire pour l'assemblage des répétitions TALE repose sur la méthode du *Unit assembly* qui a été optimisée et améliorée (Huang et al., 2014).

2.3.1.1 *Unit assembly*

L'assemblage par *Unit Assembly* repose sur l'utilisation de deux enzymes de restriction dont les extrémités sont compatibles, NheI et SpeI (Huang et al., 2011) (Figure 17). Il y a à disposition 4 plasmides différents contenant chacun une répétition TALE entourée par ces deux sites de restriction pour pouvoir les extraire et les assembler une à une. Un premier plasmide est linéarisé par digestion avec une de ces enzymes (NheI ou SpeI) et on y introduit une autre répétition extraite cette fois par les deux enzymes (Figure 18). Le plasmide ainsi assemblé contient maintenant 2 répétitions. En reproduisant cette manipulation plusieurs fois, on obtient un plasmide contenant le nombre de répétitions TALE souhaité. Les répétitions TALE sont ensuite extraites et insérées dans un plasmide

contenant le domaine de clivage de l'enzyme FokI. C'est cette technique du *Unit Assembly* qui a été améliorée et qui est utilisée pour la production de TALENs au sein de notre laboratoire.

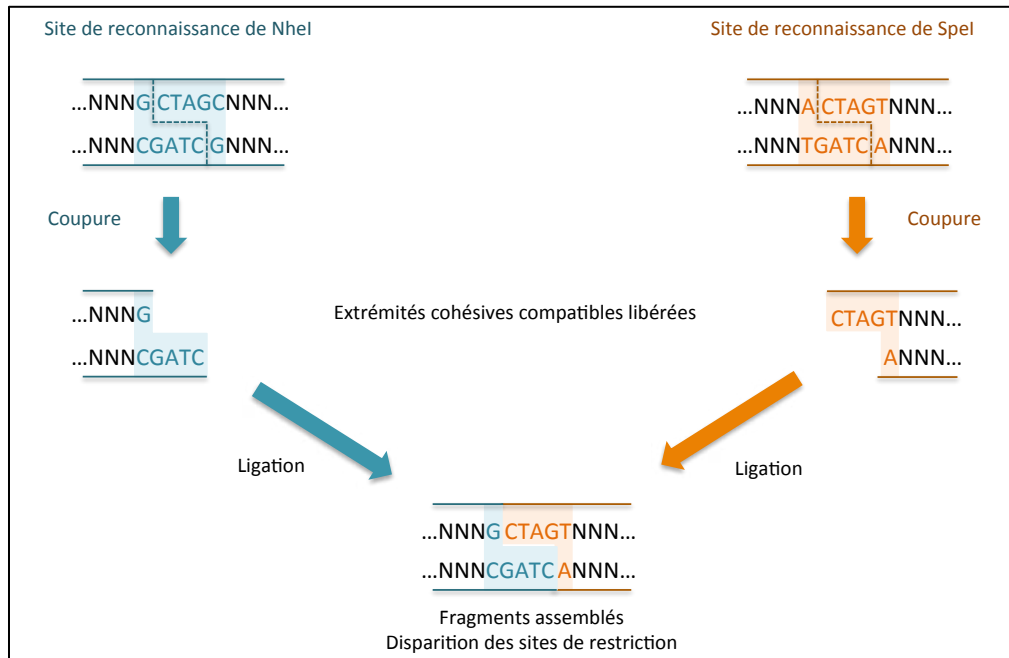


Figure 17. Assemblage des extrémités des enzymes compatibles NheI et SpeI. Après digestion, les extrémités produites par les enzymes sont compatibles et permettent une ligation faisant disparaître les sites de restriction de ces enzymes.

2.3.1.2 Protocole optimisé

Afin de faciliter sa mise en oeuvre, ce protocole de *unit assembly* a été optimisé pour permettre l'assemblage de nombreuses TALEN en parallèle et en seulement 10 jours.

Dans un premier temps, toutes les combinaisons possibles de 4 répétitions TALE ont été assemblées dans un plasmide pSk duo, appelées tétramères. Il suffit donc maintenant d'assembler 4 tétramères au lieu de 16 répétitions indépendantes (Figure 18) (Annexe 1 et 2). Cela permet de gagner du temps et d'avoir beaucoup moins de plasmides à manipuler en parallèle.

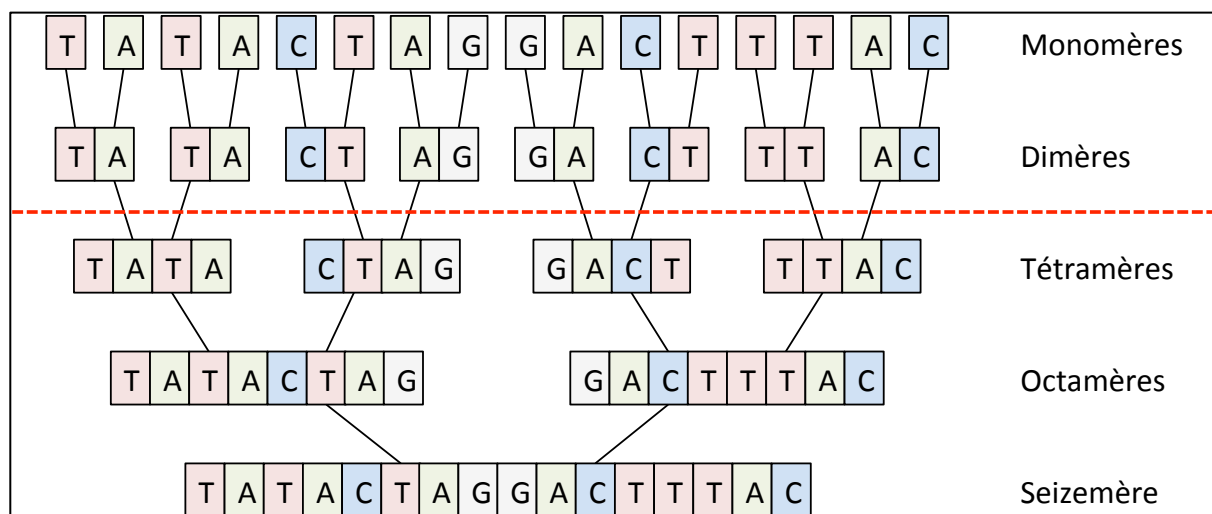


Figure 18. Etapes d'assemblage des répétitions avec le Unit Assembly.

En rouge est indiquée l'étape à laquelle notre optimisation du protocole nous permet de commencer.

2.3.1.2.1 Assemblage des répétitions

Les étapes d'assemblage des TALEN sont une suite de digestions/ligations dont les différentes digestions à réaliser sont résumés dans la Figure 19.

Nom de la digestion	SN	SS	Contrôle d'assemblage	Retrait répétitions	Contrôle int	Retrait linker 1	Retrait linker 2	Contrôle N
Objectif	Assembler les répétitions	Assembler les répétitions	Contrôle les répétitions assemblées	Extraire les répétitions assemblées	Contrôle l'insertion des répétitions	Retirer le linker d'orientation	Linéariser les vecteurs contenant le linker d'orientation	Contrôler l'absence de linker d'orientation et la présence des répétitions
Plasmide à digérer	Sk duo	Sk duo	Sk duo	Sk duo	CS2 nucléase	CS2 nucléase	CS2 nucléase	CS2 nucléase
Enzymes	ScaI, NheI, XhoI, Sall	ScaI, SpeI, SacII, EagI	SacI, HindIII	SpeI, XhoI	HindIII, XhoI	BbsI, NheI	XhoI	HindIII, XhoI, EagI

Figure 19. Digestions nécessaires au protocole optimisé d'assemblage des répétitions TALE.

Les digestions SN (par ScaI et NheI) et SS (par ScaI et SpeI) permettent d'assembler les répétitions TALE entre elles pour former le domaine de reconnaissance de l'ADN. Les

enzymes de restriction NheI et SpeI libèrent des extrémités compatibles permettant une ligation faisant disparaître les deux sites de restriction (Figure 18).

Les tétramères 1 et 3 sont digérés par la digestion SN et les tétramères 2 et 4 par la digestion SS, permettant ainsi de les assembler simplement par ligation comme expliqué dans la Figure 23. Pour chaque plasmide la répétition d'intérêt ainsi qu'une partie de la résistance à l'Ampicilline sont libérées, l'autre partie du plasmide est digérée par d'autres enzymes avec des sites de restriction proches des sites NheI et SpeI, libérant ainsi de petits fragments qui sont éliminés au cours de la purification sur colonne du produit de digestion. Cette dernière étape empêche la reconstitution du plasmide initial.

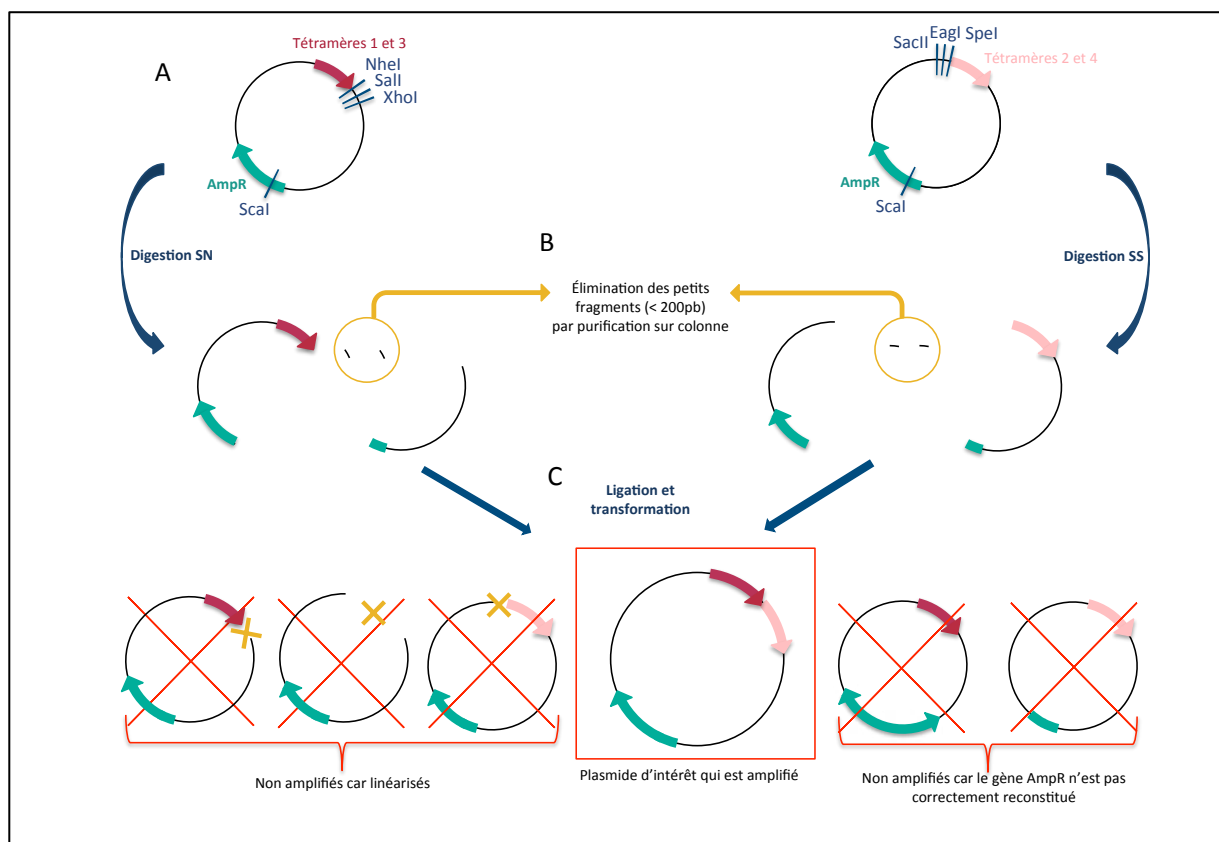


Figure 20. Principe d'assemblage des répétitions TALE.

Les différents plasmides pSk duo sont digérés de manière à libérer la répétition d'intérêt et à éliminer les fragments permettant de reproduire le plasmide du début après ligation.

Une fois les tétramères 1 et 2 assemblés ainsi que les tétramères 3 et 4 on obtient deux octamères (8 répétitions TALE). Ces derniers seront assemblés exactement de la même manière que les tétramères afin d'obtenir le domaine TALE complet de 16 répétitions. Entre chaque étape d'assemblage des répétitions, il est possible d'effectuer une digestion contrôle. Cette digestion par les enzymes *SacI* et *HindIII* permet d'extraire les répétitions assemblées et, après migration sur gel d'agarose, de déterminer facilement le nombre de répétitions assemblées puisqu'une répétition fait environ 100pb.

2.3.1.2.2 Assemblage orienté des TALEN

Une fois les 16 répétitions TALE assemblées, elles sont extraites du plasmide d'assemblage pSk duo par les enzymes *SpeI* et *XhoI*, et insérées dans un plasmide pCS2 contenant les extrémités Nter et Cter de la protéine TALE naturelle, la demi répétition souhaitée, et le domaine nucléase de l'enzyme *FokI* (Figure 21) (Annexe 3). Pour permettre une insertion orientée, un linker a été ajouté. Ce dernier peut être retiré facilement en utilisant les enzymes *BbsI* et *NheI*.

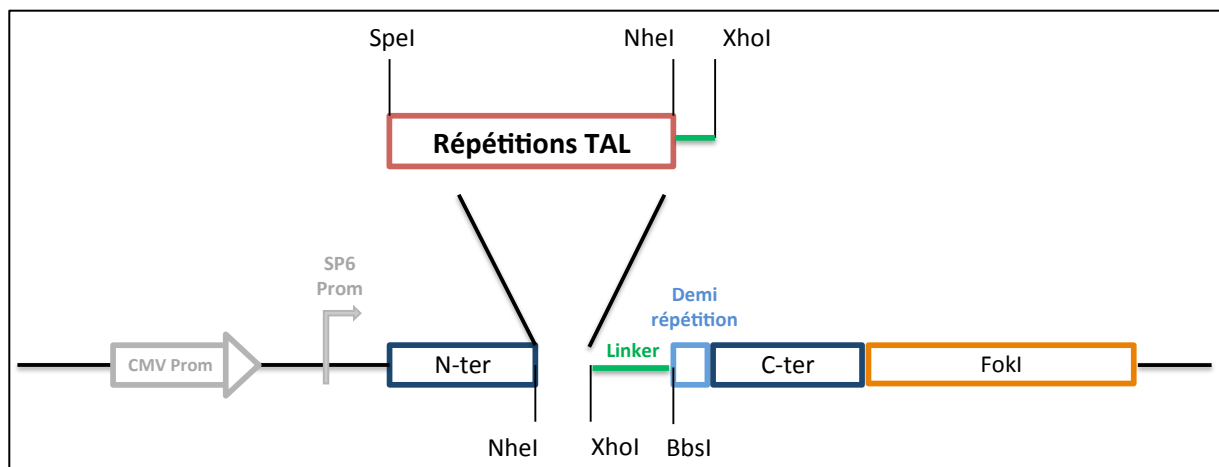


Figure 21. Principe d'assemblage d'une TALEN.

En insérant les répétitions TALE dans le vecteur pCS2 contenant les domaines Nter et Cter de la protéine TALE naturelle ainsi que le domaine nucléase de l'enzyme *FokI* on obtient une TALEN.

2.3.2 Collaborations

Depuis la mise au point de ce protocole de production de TALEN, environ 240 couples de TALEN ont été assemblés pour de nombreuses collaborations dont certaines ont donné lieu à des publications. Deux de ces travaux, que je vais détailler ici, ont permis une étude comparative des nucléases TALEN et de type CRISPR/Cas9.

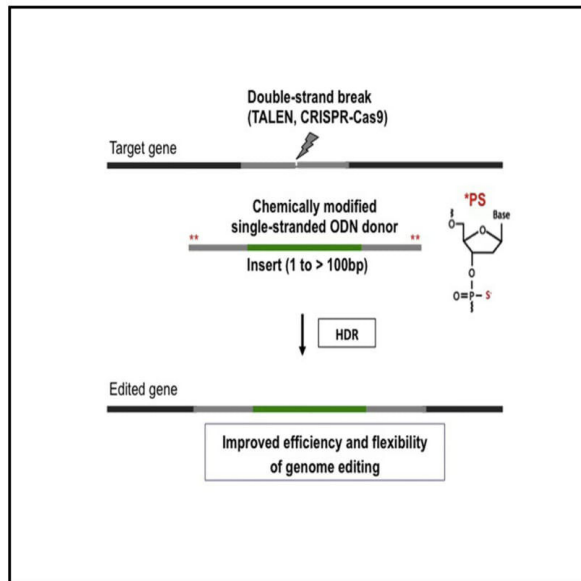
2.3.2.1 Article : *Edition du génome à l'aide d'oligonucléotides modifiés*

Au sein de mon équipe, une approche de modification de cellules humaines ou en rat a été développée, utilisant des oligonucléotides simple brin modifiés aux extrémités à l'aide de phosphorothioate. Cette stratégie permet d'insérer une petite séquence (site de restriction, étiquette flag ou site lox) de manière efficace et précise en utilisant de petites homologie (20 à 50 b). L'intégration du donneur dépendrait alors d'un système de réparation proche du SSA. Lors de ce projet différentes nucléases artificielles que j'ai produites ont été utilisées (TALEN ou CRISPR/Cas9). En particulier, nous avons pu facilement généraliser les résultats obtenus sur un premier site cible communément utilisé, le locus AAVS1, confirmant ainsi que les bénéfices obtenus en utilisant des oligonucléotides modifiés ne sont pas dépendant des séquences ciblées ou de la nature des extrémités d'ADN produites.

Cell Reports

Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases

Graphical Abstract



Authors

Jean-Baptiste Renaud, Charlotte Boix, Marine Charpentier, ..., Ignacio Anegón, Jean-Paul Concordet, Carine Giovannangeli

Correspondence

jpconcordet@mnhn.fr (J.-P.C.),
carine.giovannangeli@mnhn.fr (C.G.)

In Brief

Renaud et al. find that using chemically modified single-stranded oligonucleotide donors with CRISPR-Cas9 or TALEN nucleases enhances genome editing efficiency and flexibility in cell culture, rat, and mouse.

Highlights

- More-efficient genome editing with chemically modified oligonucleotide donors
- Greater flexibility with modified donors, allowing insertions more than 100 bp long
- Efficient mutant isolation in U2OS and RPE1 cell lines, as well as in rat and mouse
- Homozygous loxP site insertion at the mouse ROSA locus



Renaud et al., 2016, Cell Reports 14, 2263–2272
March 8, 2016 ©2016 The Authors
<http://dx.doi.org/10.1016/j.celrep.2016.02.018>

CellPress

Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases

Jean-Baptiste Renaud,¹ Charlotte Boix,¹ Marine Charpentier,¹ Anne De Cian,¹ Julien Cochenne,¹ Evelyne Duvernois-Berthet,¹ Loïc Perrouault,¹ Laurent Tesson,^{2,3} Joanne Edouard,⁵ Reynald Thinar,^{2,3} Yacine Cherifi,⁴ Séverine Menoret,^{2,3} Sandra Fontanière,⁴ Noémie de Crozé,⁵ Alexandre Fraichard,⁴ Frédéric Sohm,⁵ Ignacio Anegón,^{2,3} Jean-Paul Concordet,^{1,*} and Carine Giovannangeli^{1,*}

¹INSERM U1154, CNRS UMR7196, Museum National d'Histoire Naturelle, Paris 75005, France

²INSERM U1064, CHU de Nantes, Nantes 44093, France

³Platform Rat Transgenesis Immunophenomic, CNRS UMS3556, Nantes 44093, France

⁴genOway, Lyon 69007, France

⁵Amagen, CNRS UMS 3504, INRA UMS 1374, Institut de Neurobiologie A. Fessard, Gif-sur-Yvette 91198, France

*Correspondence: jpcconcordet@mnhn.fr (J.-P.C.), carine.giovannangeli@mnhn.fr (C.G.)

<http://dx.doi.org/10.1016/j.celrep.2016.02.018>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

SUMMARY

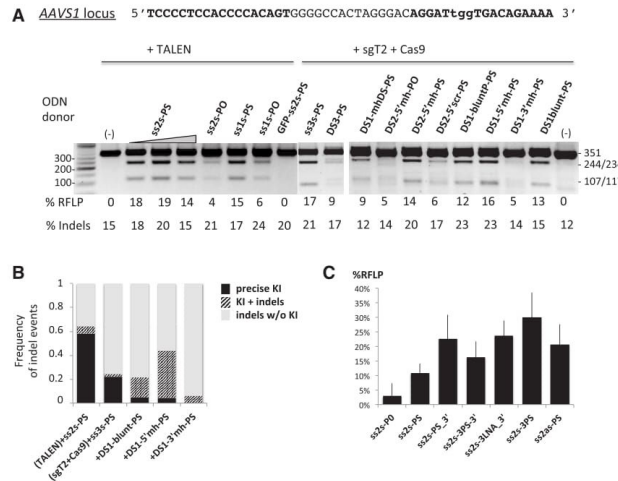
Genome editing has now been reported in many systems using TALEN and CRISPR-Cas9 nucleases. Precise mutations can be introduced during homology-directed repair with donor DNA carrying the wanted sequence edit, but efficiency is usually lower than for gene knockout and optimal strategies have not been extensively investigated. Here, we show that using phosphorothioate-modified oligonucleotides strongly enhances genome editing efficiency of single-stranded oligonucleotide donors in cultured cells. In addition, it provides better design flexibility, allowing insertions more than 100 bp long. Despite previous reports of phosphorothioate-modified oligonucleotide toxicity, clones of edited cells are readily isolated and targeted sequence insertions are achieved in rats and mice with very high frequency, allowing for homozygous loxP site insertion at the mouse ROSA locus in particular. Finally, when detected, imprecise knockin events exhibit indels that are asymmetrically positioned, consistent with genome editing taking place by two steps of single-strand annealing.

INTRODUCTION

Pioneer studies with I-SceI and zinc finger nucleases were used to establish the main principles of genome editing with sequence-specific nucleases (Porteus and Carroll, 2005). TALENs and even more spectacularly CRISPR-Cas9 have now hugely facilitated the design of sequence-specific nucleases to the target locus of interest, which is no longer the limiting step in genome editing (Hsu et al., 2014; Kim and Kim, 2014). When a double-strand break takes place, DNA repair usually proceeds

with fidelity to the original sequence (Bétermier et al., 2014). However, when sequence-specific nucleases are used, mutations are introduced at high rates, possibly due to repeated cleavage and repair cycles until mutations disrupt cleavage site recognition. The DNA double-strand break repair pathways involved are thought to be mainly classical non-homologous end joining (NHEJ) and possibly alternative end joining (altEJ or MMEJ; Bae et al., 2014) that both mediate DNA ligation after processing of chromosomal ends. Other well-characterized pathways for DNA double-strand break repair are homology directed, for instance, homologous recombination using the sister chromatid during cell replication. For purposes of genome editing, donor DNA can be introduced into cells in order to program the sequence modification resulting from homology-directed repair. Several types of donor DNA have been used including plasmid DNA and synthetic oligonucleotides (Carroll and Beumer, 2014). Plasmid DNA with 1 kbp homology arms to sequence flanking the cleavage site is commonly used in gene-targeting experiments in mammalian cells. However, using CRISPR-Cas9 nucleases, it was found that homology arms longer than 1 kbp could stimulate gene-targeting efficiency by 2- to 4-fold (Chu et al., 2015), suggesting that specific optimization may be beneficial. Recently, DNA plasmids with very short homology arms, around 10 bp long, were shown to drive targeted insertion, likely by a MMEJ-based mechanism (Nakade et al., 2014). Single-stranded oligonucleotide (ssODN) donors have also been used successfully in many experimental systems, including cultured cells (Chen et al., 2011; Ran et al., 2013) and embryo injections (Beumer et al., 2013b; Harel et al., 2015; Paix et al., 2014; Shao et al., 2014; Tan et al., 2013). Stretches of homology to the sequence flanking the cleavage site are usually included in the ssODN design (Chen et al., 2011). In the mouse and rat, precise insertions and point mutations have been achieved at high efficiencies with ssODN donors (Yang et al., 2014; Yoshimi et al., 2014), whereas in zebrafish, extensive mutations of the ssODN or target site sequence were found in the great majority of insertions (Bedell et al.,





phosphodiester; PS, phosphorothioate. Fifty-base-pair DNA ladder (NEB) was shown. See also Figure S1 for integration activity of dsODNs. (B) ssODNs, but not dsODNs, direct precise sequence insertions. The frequency of the different types of indels, evaluated by deep sequencing, are reported: % indels w/o KI (gray) and KI events that include the precise KI events (black) and KI with additional indels (hatched). (C) Influence of location and nature of the modification in the oligonucleotide on genome editing efficiency. Different parameters were evaluated: the type of chemical modification (LNA [locked nucleic acid]), the number of modified nucleotides (two at both ends or at 3' end in ss2sPS or ss2sPS_3', resp.; three at both ends or at the 3' end in ss2s-3PS or ss2s-3PS_3' and ss2s-3LNA_3', resp.). RFLP rate is shown for different modified oligonucleotides, as mean $\pm$ SD (3–20 independent experiments). For all the modified ssODNs compared to the PO ODN, p values from Mann-Whitney statistical analysis are <0.05 . See also Figure S2 for impact of polarity of the PS-ssODN and of distance between break and gene edit (Figure S2A) and of homology length on activity of PO-ODNs (Figure S2B).

Figure 1. Optimization of Short ODN Donor Design

(A) Integration activity of short ss- and dsODN donors. TALEN target sequences (in bold) and PAM sequence (in lower cases) in an intron of the *PPP1R12C* gene within the chromosome 19 locus are shown. Single-stranded donor oligonucleotides (ssODNs) with short homology arms, ss1s (43-mer) and ss2s (58-mer), and the 58-bp DS3 duplex (formed by hybridization of ss3s with its complementary strand) were used to make a sequence insertion including a PvuII site at the cleavage site. A control ssODN including a PvuII site but lacking homology arms, GFPss2s-PS, was also used. Double-stranded donors (dsODNs) (different versions of DS1 and DS2, as indicated) included an XhoI site (see sequences in Tables S1 and S2 and Figure S1). dsODNs containing microhomologous ends (DS1-mhDS-PS, DS1-5'mh-PS, DS2-5'mh-PS, DS2-5'mh-PO, and DS1-3'mh-PS) or not (DS1-blunt-PS, DS1-bluntP-PS, and DS2-5'scr-PS) were evaluated. RFLP analysis of DNA from U2OS cells treated with TALEN AAVS1 or (sgT2+Cas9) and various ssODNs (at 12 μ g, except ss2s-PS at 6, 9, and 12 μ g) or dsODNs (at 12 μ g) is shown. The corresponding rate of PvuII or XhoI cleavage (% RFLP) and of mutations evaluated in parallel by the T7 endonuclease I assay, T7E1 (indicated as % indels; gel not shown) are reported below the gel. PO, PO,

2012; Gagnon et al., 2014; Hruscha et al., 2013), therefore much limiting the use of ssODN donors in zebrafish.

We chose to more systematically test the potential of chemically modified ssODNs or double-stranded ODNs (dsODNs) for genome editing with sequence-specific nucleases. We investigated several parameters including the nature and location of modified bases, length of sequence insertion, and distance to cleavage site position. Different types of nucleases were tested, both TALENs and CRISPR-Cas9 nucleases, for genome editing in multiple experimental systems, including cultured cells and model organisms. Our results show that phosphorothioate (PS)-modified ssODNs exhibit improved efficiency and flexibility in most experimental systems tested. Interestingly, detailed analysis of gene edits reveals a variable proportion of imprecise sequence insertions, which could be due to different mechanisms of double-strand break repair by ssODNs.

RESULTS

Design and Activity of Short ss- or dsODN Donors

We first evaluated the homology-directed repair (HDR) capacity of ODNs with short homology arms. We tested both ssODNs and dsODNs to generate small insertions, and in each case, we evaluated different designs. ssODNs have been shown to be effective in genome editing, but several questions remain concerning how to optimize their design. Regular phosphodiester (PO)

ssODNs of around 100 nucleotides (nt) long have been successfully used (Chen et al., 2011; Yang et al., 2013). We first tested chemically modified ssODNs for HDR. We started with the common PS modification, known to improve nuclease resistance (Eckstein, 2000). We designed ssODNs with short homology arms, ss1s, with 18 and 19 nt on each side of the break, and ss2s or ss3s, with 27 or 28 and 24 nt (see Table S1 for sequences). PS chemical modifications were incorporated at two terminal nucleotides of both 5' and 3' ends. Using AAVS1 as a test locus, we introduced a PvuII restriction site sequence in ssODN (Figure 1A) for easy analysis of sequence insertion by restriction fragment length polymorphism (RFLP). We chose to insert the PvuII site directly into the expected cleavage site for the sequence-specific nucleases used here, AAVS1-specific TALEN with ss1s and ss2s and Cas9 combined with sgT2 guide RNA with ss3s. ssODN donors were co-transfected with AAVS1-specific TALEN or Cas9 and sgT2 guide RNA expression plasmids. Fourteen days post-transfection, the AAVS1 locus was amplified by PCR and donor insertion into the AAVS1 site assayed by PvuII digestion (Figure 1A). PS-modified ssODNs (ss1s-PS and ss2s-PS) directed robust sequence insertion efficiency in the presence of nucleases, whereas the corresponding unmodified ssODNs (ss1s-PO and ss2s-PO) led to only modest integration levels (6% versus 15% and 4% versus 14%, respectively). The knockin (KI) rates using PS-modified ssODNs were in the range of indel rates induced by nucleases at this site as measured by

the T7 endonuclease assay (T7E1), suggesting that the majority of modification events were donor-mediated insertions. A PS-ssODN including a PvuII site but lacking homology arms, ssGFPs2s-PS, did not induce any sequence insertion, as expected for an HDR-based mechanism. Finally, using the duplex version of the ODN donor (DS3) strongly decreased the integration rate (by 2-fold), consistent with a mechanism different from homologous recombination being involved for ssODNs.

Based on results showing that donors with microhomologies to the cleavage site ends can enable efficient integration of DNA fragments by MMEJ (microhomology-mediated end-joining) (Nakade et al., 2014; Orlando et al., 2010), we chose to evaluate the potential of dsODNs with microhomologies for integration at the Cas9 cleavage site directed by guide RNA sgT2. Based on the fact that cleavage sites are processed during repair, generating different types of DNA ends (Dorsett et al., 2014), we tested different types of microhomologous ends, either double-stranded (mhDS), single-stranded 3' overhangs (3'mh), or single-stranded 5' overhangs (5'mh) (see sequences in Figure S1 and Table S2). As above, two PS nucleotide residues were included at 5' and 3' ends of dsODNs and the sequence insertion consisted of an exogenous 40- or 46-bp sequence (DS1 or DS2, respectively) including an XhoI restriction site. Control donor oligonucleotides were used that had either no homology (DS1-blunt-PS and DS1-bluntP-PS; a 5' phosphorylated version of the former) or 8 nt 5' mh predicted not to base pair with the sequence at the cleavage site (DS2-5'scr-PS). These dsODN donors were co-transfected with Cas9 and sgT2 guide RNA expression plasmids. Six days post-transfection, the AAVS1 locus was amplified by PCR and sequence insertion into the AAVS1 site assayed by XhoI digestion (Figure 1A). As observed with ssODN donors, the PS modification strongly enhanced integration of duplex donors (14% compared to 5% for DS2-5'mh-PS and -PO, respectively). Sequence insertion was observed for all the designs tested but at different rates (Figure S1B). In particular, as recently reported, dsODN with blunt ends integrated at frequencies corresponding roughly to 1/2 indel mutation rates (Tsai et al., 2015). Importantly, the duplex with the non-complementary overhangs (DS2-5'scr-PS) also led to XhoI sequence insertion. These results were confirmed at another cleavage site, at the *SOD1* gene (Figure S1C). At this target, blunt and overhang-containing duplexes also led to targeted donor sequence insertion regardless of the presence of microhomologous ends (4% and 9% for DS2-sod5'mh-PS and DS2-sod-3'mh-PS, which had microhomologous ends, compared to 7% and 11% for DS2-5'mh-PS or DS1-blunt-PS, without), indicating that there is no need for microhomologous ends to achieve efficient integration. This finding suggested that imprecise sequence insertions were taking place.

ssODNs, but Not dsODNs, Direct Precise Sequence Insertion

To determine the fidelity of donor sequence insertion, the PCR amplicon from the AAVS1 target was subjected to deep sequencing. For ssODNs, the majority of insertion events were precise (representing more than 90% of total KI events) and only a low frequency of imprecise insertions was detected (indicated as "KI+indels" with hatched bars; representing less than 10%

of total KI events; Figure 1B). With the short duplex donors, sequence information in the microhomology arms was expected to result in precise donor sequence insertion without loss of either chromosomal or donor sequence. However, only few integration events were precise. Poor efficiency of precise insertion could be due to insufficient levels of available complementary "acceptor" ends after target sequence cleavage by Cas9, which is known to generate blunt ends (Jinek et al., 2012). Recent studies using TALENs or CRISPR-Cas9 and DNA plasmid donors containing microhomologies, although demonstrating precise MMEJ-based integration, also exhibited significant amounts of imprecise integration events (Nakade et al., 2014). In contrast, previous studies that showed efficient precise insertion with dsODN donors carrying microhomologies were performed with zinc finger nucleases that generate 5' protruding ends (Orlando et al., 2010).

Chemical Modifications of ssODNs Stimulate Genome Editing

Based on the sequencing data, we next focused on ssODNs donors that were the most efficient for precise sequence insertion. We evaluated a set of ssODNs in order to determine parameters impacting genome editing efficiency. We first performed experiments to assay the importance of the nature and location of chemical modifications. PS modifications were positioned at both ends, as above, or at 3' end only, and we compared efficiency of two to three PS-modified nucleotides. In addition to PS modification, we selected LNA, locked nucleic acid, known to improve ODN chemical and thermal stability (Jepsen and Wengel, 2004). Donor insertion into the AAVS1 site was assayed as described above by RFLP analysis. All the chemically modified donors strongly enhanced donor sequence insertion compared to the PO-ssODN donor (Figure 1C). These data are consistent with a model in which ssODN end modifications enhance intracellular stability, thus enabling increased efficacy of genome editing.

Because chemical modifications, especially PS, have been described to be toxic (Rios et al., 2012), we carefully investigated the long-term survival and proliferation of modified cells. Sequence insertion rates were found to remain unchanged up to 15 days after treatment, whether using an ssODN containing PS modifications or standard PO chemistry.

We designed ssODN donors of variable sense and antisense orientations (s and as). At the AAVS1 locus, the antisense donor, ss2as-PS, used with the AAVS1-specific TALEN, was slightly but significantly more efficient than the sense donor, ss2s-PS (Figure 1C; $p = 0.008$ from Mann-Whitney statistical analysis). However, in all the situations that we have tested (for example: AAVS1 locus in Figure S2A; GFP transgene in Figure S4C), both orientations worked, sometimes with slightly different efficiencies, but no general rule emerged for sense or antisense orientation preference.

Improved Flexibility of Genome Editing Using PS-ssODN Donors

We next investigated the effect of homology length on gene editing in the context of ssODN donors and assayed targeted sequence insertion at the AAVS1 locus (Figure S2B). In previous

work using PO-ssODNs, lengths in the range of 100 nt were common in most applications. We compared the ss1, ss2, and ss95 donors with 37, 51, and 84 nt total homology regions, respectively. As already shown in Figure 1A, short ss1-PO and ss2-PO donors drive limited donor sequence insertion compared to their PS counterparts. Using longer PO donors with an 84-nt total homology ss95-PO can strongly improve the activity compared to 51 nt (22% for ss95-PO compared to 5% for the same concentration of ss2-PO; Figure S2B). This observation was confirmed on *SUFU* gene, extending from a 73- to a 97-nt total homology region (43% for Sufu-1FLAG-L-PO versus 12% for the same concentration of Sufu-1FLAG-PO; Figure S2B). Although unmodified PO-ssODNs are intrinsically less efficient than chemically modified ones, using longer PO donors at high concentrations can compensate for their lower activity, possibly because sufficient homology is preserved from exonuclease degradation to maintain activity. However, lengthening of the homology above 90–100 nt has been described to be detrimental (Yang et al., 2013), likely because it could be prone to secondary structures decreasing the amount of donor effectively available for gene editing. In addition, using long homologous regions limits the size of functional sequences that can be inserted because chemical synthesis allows robust production of ssODNs no longer than 200 nt.

When using ssODN donors for sequence insertion, DNA cleavage is commonly designed to take place less than 10 bp away from the desired insertion site (Yang et al., 2013). In experiments described above, we designed the ssODN donors (ss1, 2, and 3) to make the insertion exactly at the expected cleavage site. We therefore wanted to examine in more detail how the rate of sequence insertion varies with its distance to the cleavage site. We tested a combination of nuclease and ssODN donor pairs (TALEN, sgT2, sgU2, and sgS1 and ss2, ss2L, and ss4; Figure S2A), generating DSBs at various distances from the insertion position. Here, we observed robust efficiency for sequence insertions at distances from 8 to 20 bp (Figure S2A), further establishing the flexibility of ssODN design when using PS modifications.

Efficient Gene Tagging with a 102-bp Insert Using PS-ssODNs

To demonstrate the general applicability of our results, we further tested similar reagents at different genomic loci and cell lines. The optimized ODN design was tested on different genes in Hedgehog signaling, a major signaling pathway in development and oncogenesis. We sought to introduce FLAG-tag coding sequences as a typical application of genome editing that should facilitate functional characterization of proteins for which antibody resources are limited. First, different amounts of PS- and PO-ssODNs were co-transfected in U2OS cells with TALENs targeting the ATG start codon of the *SUFU* protein coding sequence. Interestingly, as shown at the *AAVS1* locus, PS-ssODNs resulted in higher efficiency of FLAG-tag sequence integration (36% for Sufu-1FLAG-PS compared to 12% for -PO for example at the higher dose and 27% versus 1% at the lower dose; Figure 2A). The insertion size is limited by ODN synthesis capacity and length of homology arms. Considering that we can use shorter homology sequences by including PS modifications, we tested whether PS-ssODN would allow making longer

sequence insertions. As shown in Figure 2A, we found that using PS-modified ssODN allowed very efficient insertion of more than 100 bp whereas only limited integration could take place with non-modified ssODN (53% for Sufu-3FLAG-HA-PS versus 3% for -PO), demonstrating a striking advantage of modified compared to non-modified ssODNs. To further generalize our findings, we tested insertion of FLAG-tag coding sequences in a series of genes of the Hedgehog signaling pathway (*SUFU*, *GLI2*, *GLI3*, *MLF1*, *SMO*, and *KIF7* genes) in U2OS cells and in a non-transformed RPE1 human cell line (Figure 2B). Efficient insertion could be induced at all loci tested by co-transfection of modified ssODNs with TALENs targeting the corresponding sequences. In the absence of TALEN co-transfection or when using control oligonucleotide with no sequence homology to the target, we could not detect FLAG-tag coding sequence integration by PCR. We further exploited the use of short modified ssODNs to evaluate the FLAG-tag insertion at two genes simultaneously. Transfection of the corresponding pairs of nucleases and donors gave rise to efficient insertion at both targets (*SUFU* and *GLI2* genes; Figure S2C). When comparing results between the two cell lines, we observed that KI rates were higher in U2OS (KI rates ranging between 20% and 50%) than in RPE1 cells (KI rates ranging from 2% to 12%; Figure 2B), possibly due to differences in nuclease cleavage efficiencies and/or intrinsic differences in activities of DNA repair pathways. Importantly, in RPE1 cells, KI events at all loci tested could only be easily detected with PS- and not PO-ssODNs donors (for example, 5% for Sufu-1FLAG-PS versus 0% for -PO and 4% for Sufu-3FLAG-HA-PS versus 1% for -PO; Figure 2B, lower panel).

Efficient Isolation of Tagged Cell Clones with PS-ssODNs

We next examined whether cells carrying the genome edits of interest could be isolated. Given the high frequencies of genome editing observed (see Supplemental Experimental Procedures), we seeded cells at clonal density and directly isolated and analyzed a series of clones. Targeted genome editing was analyzed at the DNA and protein levels. We could readily isolate clones carrying the integration of tag-coding sequences either at one or at multiple alleles in U2OS and in RPE1 cells (Figure 2C). The rate of KI alleles in the clones analyzed was roughly equivalent to the KI rates measured in the initial cell population, therefore supporting the absence of specific toxicity of PS-ssODNs. Fewer positive clones were isolated in RPE1 cells, as expected from the lower genome editing efficiency compared to U2OS cells. At the protein level, it was possible to detect tagged *SUFU* proteins with anti-FLAG antibody in positive clones from both cell lines, with the expected signal amplification for the clones containing the 3FLAG-tagged Sufu protein compared to 1FLAG-tag. The readily achieved isolation of modified clones showed that treatment of cells with PS-modified ODNs was compatible with long-term expansion of modified cells.

Successful KI in Rat, Mouse, and Zebrafish with PS-ssODNs

We next wanted to examine the potential of modified ssODNs in genome editing of model organisms, both testing the benefit of modified ssODNs compared to PO-ODNs and further checking

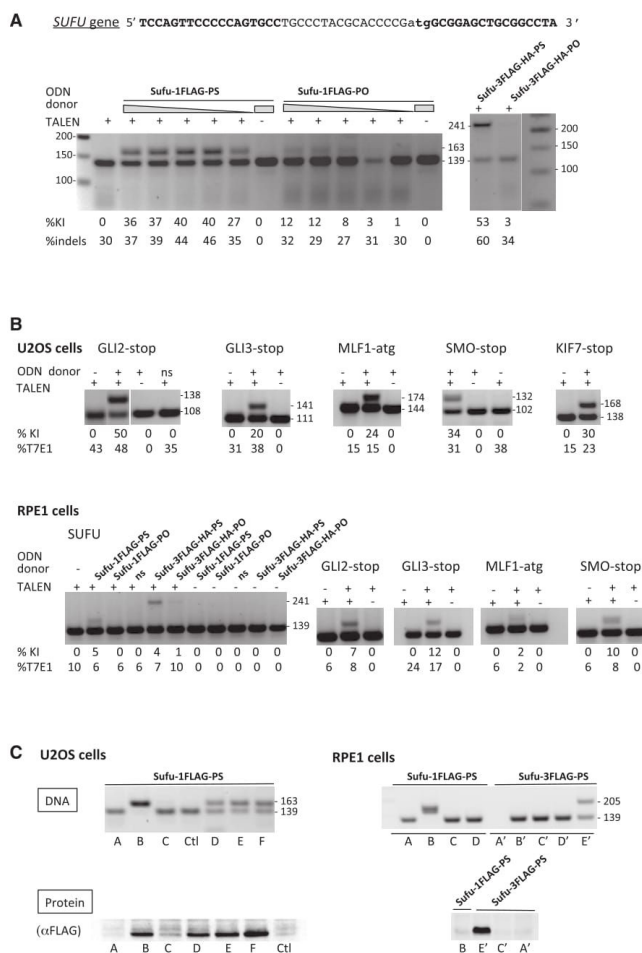


Figure 2. Efficient Tag Sequence Insertion of More Than 100 bp Using Modified ODNs at Genes of the Hedgehog-Signaling Pathway
TALENs and ssODNs were designed to perform insertion of FLAG-tag coding sequences at the ATG or STOP codon positions of the indicated genes in the Hedgehog-signaling pathway. DNA was extracted and PCR amplified to evaluate the efficiency of FLAG-tag sequence insertion and perform the T7E1 assay. % KI indicates the relative abundance of the higher size band corresponding to FLAG-tag sequence insertion relative to the total DNA bands. % indels indicates the mutation rates calculated from the T7E1 assay results and %T7E1 the cleavage rate by T7E1 (gel not shown). Note that, when provided, indels rates were calculated by taking into account the KI population, as detailed in Experimental Procedures.

(A) *SUFU* gene tagging in U2OS cells. TALEN target sequences (in bold) around the start codon (lower cases) of the *SUFU* gene are shown. ssODNs containing insertion of one FLAG-tag sequence (24 bp), *Sufu*-1FLAG-PS, and *Sufu*-1FLAG-PO or 3FLAG-HA sequence (102 bp), *Sufu*-3FLAG-HA-PS, and *Sufu*-3FLAG-HA-PO were used at different doses with TALEN expression plasmid (for *Sufu*-1FLAG-PS and *Sufu*-1FLAG-PO: 15, 12, 9.6, 7.2, and 4.8 μ g or 15 μ g with ssODN alone; 6 μ g for *Sufu*-3FLAG-HA-PO and *Sufu*-1FLAG-PS). Fifty-base-pair DNA ladder is shown on the right.

(B) Targeting of Hedgehog-signaling genes in U2OS and RPE1 cells. PS-ssODNs containing one FLAG-tag coding sequence were used (see Table S1 for sequences). Transfection was performed with TALEN expression plasmids and/or ssODNs (12 μ g), as indicated. ns indicates control ssODN lacking homology arms. For *Smo*-FLAG-PS, we could initially detect products corresponding to potential FLAG-tag sequence integration in the absence of nucleases but realized that this was due to PCR artifacts, and special care was taken to exclude PCR artifacts, which included using DNA extracted 3 weeks after transfection and treating DNA samples with single-strand DNA nucleases before PCR (Figure S3). See also Figure S2C for multiplex tag sequence insertion.

(C) Efficient isolation of clones with FLAG- or 3FLAG-tagged *SUFU* gene from cells treated with PS-modified ssODNs. Cell clones were isolated after transfection with *SUFU*-specific TALEN expression plasmids and the indicated PS-

ssODNs. Uppercase labels indicate specific clones analyzed at both DNA and protein levels; Ctl lane corresponds to untreated cells. Protein expression was analyzed by western blot using anti-FLAG antibody. See also Figure S4 for successful ssODN-mediated KI in rat, mouse, and zebrafish.

for toxicity. In the rat and mouse, it was previously reported that non-modified ssODNs with long homology arms can direct efficient sequence insertion during repair of double-strand breaks induced by TALEN or CRISPR-Cas9 nucleases whereas, in the zebrafish, only modest levels of precise sequence insertion could be achieved (Bedell et al., 2012; Gagnon et al., 2014).

For tests in the rat, we targeted the *Cftr* gene. ssODNs were designed with 45-nt homology arms flanking a single A nt that creates a XbaI restriction site for facile genotyping of *Cftr* mutant rats and that simultaneously disrupts the open reading frame of the *Cftr* gene and target sequence of the guide RNA (to prevent

cleavage of the successfully edited genomic sequence; Figure S4A). The genome editing strategy was first tested in cultured rat C6 glioblastoma cells. In line with results described above, successful insertion, although at very low levels, could be detected when guide RNA and Cas9 expression plasmids were co-transfected with PS-ssODN, but not with PO-ssODN (Figure S4A). PS- or PO-ssODNs were next co-injected with guide RNA and Cas9 mRNA into rat fertilized zygotes. Rat pups born from the injections were genotyped by PCR and XbaI digestion as well as by DNA sequencing. Embryo viability and newborn frequencies were in a similar range with PS- and PO-ssODNs,

Table 1. Efficient Genome Editing with PS-Modified ssODNs in Model Organisms

Target Locus	ssODN Donor	Cas9/sg/ssODN (ng/ μ l)	No. of Newborns	No. of E15	No. of KI (%)	No. of NHEJ (%)
Rat <i>Cftr</i> gene	PS	20/10/15	23	–	6 (26.1)	9 (39)
		50/10/15	41	–	13 ^a (31.7)	22 (53.6)
	PO	20/10/15	49	–	13 ^b (26.5)	31 (63.3)
		50/10/15	65	–	15 ^b + 1 ^c (24.6)	34 (52.3)
Mouse <i>ROSA</i> 26	PS	20/10/25	–	8	3 ^d + 2 ^e (62.5)	4 (50)
	PO	20/10/25	–	21	2 (9.5)	15 (71.4)

(Top) Genome editing with PS- and PO-ssODNs in rat. Two series of injections into rat fertilized oocytes were performed, loading injection pipettes with a solution containing *Cftr* guide RNA, Cas9 mRNA, and *Cftr* PS- or PO-ssODN as indicated (sequence information is provided in Figure S4A). Tail DNA was extracted from newborns, and targeted modification of the *Cftr* gene was examined by PCR and XbaI digestion to quantitate the insertion efficiency (Figure S4A), as well as by PCR and high-resolution DNA capillary electrophoresis to quantitate indel mutations. Quantifications and sequence modifications were confirmed by DNA sequencing of PCR products. Rats that are heterozygotes with one KI and one NHEJ event are included in both KI and NHEJ animal counts indicated in the table. See also Figure S4A.

(Bottom) Genome editing with PS- and PO-ssODNs in mouse. Injection was performed in mouse oocytes with a solution containing *Rosa* guide RNA, Cas9 mRNA, and *Rosa*-PS- or PO-ssODN as indicated (see experimental design in Figure S4B). DNA from E15 embryos was extracted, and insertion of loxP sequence was analyzed by PCR and BglII digestion to evaluate insertion rate (Figure S4B) and confirmed by sequencing of PCR products. Mice that are heterozygotes with one KI and one NHEJ event are included in both KI and NHEJ animal counts indicated in the table. See also Figure S4B.

^aIncluding one homozygous KI animal.

^bIncluding four homozygous KI animals.

^cAll KI events corresponded to precise sequence insertion, except for one KI obtained with the PO-ssODN, which showed a double copy of ODN sequence.

^dIncluding three homozygous KI mice.

^eThe insertion events are precise for all embryos except two: one that had a 137-bp deletion on the 5' side of the insert and the other two point deletions, one on each side of the insert.

indicating that no specific toxicity was associated with PS- compared to PO-ssODNs (Figure S4A). Strikingly, a high proportion of embryos carried the wanted sequence insertion when either type of ssODN was injected: 24.6%–31.7% of newborns were KI-positive, and in several rats, KI was detected on both alleles. All rats with KI events showed precise integration of the ssODN, irrespective of the type of ssODN, with the exception of one rat that showed integration of two ssODN sequences in tandem after injection of PO-ssODN. NHEJ events were frequent in all experimental groups (39%–63.3% of newborns), attesting high efficacy of the guide RNA. In addition, we evaluated the specificity of ssODN integration using the targeted locus amplification (TLA) method (de Vree et al., 2014), which allows to define the genomic integration site of a given exogenous sequence. The integration sites of the ssODN donor were mapped by TLA in a rat that had been found to have precise KI at the *Cftr* locus. TLA showed integration of the ssODN at the *Cftr* locus, as expected, and not at any other genomic site, demonstrating that only specific ssODN integration had taken place.

In the mouse, the ssODN was designed to insert at the *ROSA* locus a 34-bp loxP sequence, a genome edit of high interest to generate alleles for conditional gene inactivation with the Cre-loxP system (Branda and Dymecki, 2004), and a BglII site to facilitate genotyping (Figure S4B). Zygotes injected with PS- or PO-ssODN donors were genotyped at the blastocyst stage (Figure S4B) or at E15 (Table 1). In contrast to the results at the rat *Cftr* locus, co-injection of PS-ssODN led to a much-higher efficiency of sequence insertion (62.5% versus 9.5% with PO-ssODN). Notably, several E15 embryos carried KI at both alleles when using PS-ssODN whereas none was found with PO-ssODN. In both groups, the rates of NHEJ-modified mice were

comparable (50% and 71.4%). As observed in rats, survival rates were equivalent with the PS- and the PO-ssODN donors (Figure S4B). Because the experimental injection setup and initial stages of mouse and rat embryo development are highly similar, it was surprising that the benefit of PS-ssODNs was much more dramatic in the mouse. One possibility is that PS-ssODNs were much more efficient because the sequence edit was much longer in the mouse experiment, a 40-nt insertion compared to 2-nt deletion in the rat, and insertion of long sequences is likely less efficient, making the PO-ssODNs suboptimal.

Finally, we tested PS-ssODNs targeting a *GFP* transgene in zebrafish. After co-injection of PS-ssODN with GFP-specific TALEN proteins, GFP-negative embryos were analyzed individually for sequence modification at the target locus by deep sequencing. As previously reported with PO-ssODNs, only very low levels of precise donor sequence insertion were obtained (Figure 3A). In one embryo (E8), however, a high proportion of precise donor insertion was detected, likely due to insertion at early cleavage stages. Although PS-ssODN did not allow to improve the rate of precise sequence insertion in zebrafish embryos, one embryo exhibited sufficient levels of precise KI to expect germline transmission.

Indels Found in Imprecise KI Are Asymmetrically Positioned

The detailed analysis of sequences corresponding to imprecise KI in zebrafish exhibited some specific characteristics. We observed that in a large majority of cases, the sequence modifications (relative to the expected, precise KI) were located on the 5' side of the 7-bp insert (Figure 3A). Asymmetry was also noted in some imprecise KI events resulting from using PO-ssODNs

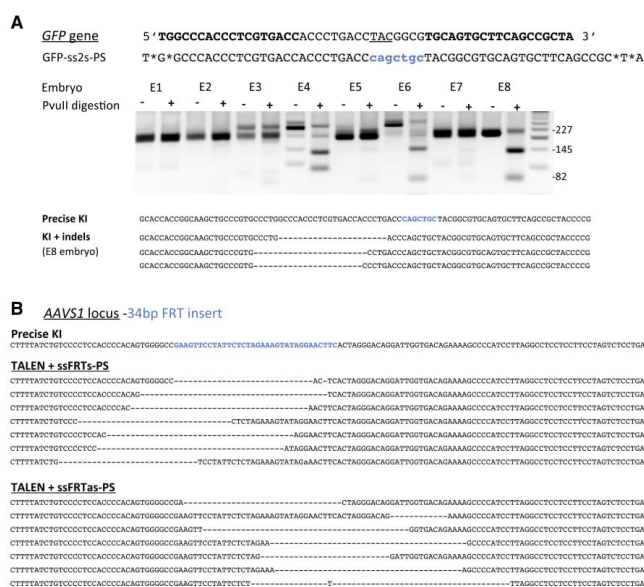


Figure 3. Nature of Imprecise KI Events Using Sequence-Specific Nucleases and ssODNs

(A) Test of genome editing with PS-modified ODNs in zebrafish. (Upper panel) TALEN target sequences (in bold) around the fluorophore-specific codon of the *GFP* gene (underlined). A PS-ssODN, carrying two phosphorothioate linkages at both 5' and 3' ends, was designed to introduce a 7-bp sequence including a PvuII site (in blue lowercase) into *GFP* cDNA, resulting in frame shift and *GFP* inactivation. Preliminary transfection experiments were conducted in cells stably expressing *GFP* to demonstrate that efficient *GFP* inhibition could be induced by co-transfection of *GFP*-specific TALENs together with PS-ODNs and *GFP*-ss2-PS (Figure S4C). (Middle panel) The PS-ssODN was co-injected together with *GFP*-specific TALEN proteins into one-cell-stage zebrafish embryos carrying a ubiquitously expressed *GFP* transgene, and *GFP*-negative transgenic embryos were identified. Single *GFP*-negative embryos (E1–E8) were lysed in PCR buffer and the *GFP* target sequence amplified and PvuII-digested to evaluate the insertion efficiency. (Lower panel) Genotyping was performed by deep sequencing of PCR products. The three most-frequent sequences corresponding to imprecise KI (KI+indels) in E8 embryos are shown with the expected KI sequence (Precise KI) reported on the top as a reference (insert in bold blue).

(B) Examples of imprecise KI events obtained in U2OS cells. The most-frequent imprecise KI products are shown for U2OS cells treated with TALEN targeting the *AAVS1* locus in the presence of the ssFRTs-PS or ssFRTas-PS donors that were designed to introduce a 34-bp-long FRT insert sequence (insert in bold blue). In (A) and (B), imprecise KI sequences correspond to FRT insertion exhibiting additional indels on the 5' side of the insert sequence, referring to the ssODN sequence orientation.

with TALENs in zebrafish (Bedell et al., 2012). We therefore examined in more detail the imprecise insertions detected previously in U2OS cells treated with ssODNs. We analyzed DNA of cells treated with guide RNA sgT2 and ss3s-PS or ss3as-PS, or with TALEN and ss2s-PS or ss2as-PS, and we focused on sequences including indels in addition to the insert sequence 5'-cagctg-3'. We observed that there is an asymmetry in the location of indels: they are mainly located on the 5' side of the edit with ss3s donor as in zebrafish and on the 3' side for the ss3as donor. In other words, referring to the ssODN sequence, indels are always present on the 5' side of the insert in the ssODN sequence (Table 2). In order to further analyze imprecise insertion events, we also sequenced the *AAVS1* target locus after treatment with *AAVS1*-specific TALEN and an ssODN donor designed to introduce a 34-bp-long FRT sequence. In a majority of cases, imprecise insertions exhibited either indels at the *AAVS1* sequence next to the complete insert sequence or partial FRT sequence insertions with deletions of the insert and adjacent *AAVS1* sequence. In both cases, indels were asymmetrically localized, being predominantly on the 5' side (referring to the ODN sequence; Figure 3B; Table 2). Other types of imprecise KI (exhibiting indels on the 3' side or more-complex insertion events) were observed in much-lower proportions. It is interesting to note that, compared to ss2 or ss3 donors, imprecise KI events were slightly more frequent for the FRT donor, which

corresponds to insertion of a longer sequence and could be more error prone. In rat, all KI events obtained with PS-ssODN were precise, whereas in mouse, two out of five KI included indels. The results reported are consistent with a model for ssODN-mediated genome editing dependent on two steps of single-strand annealing, as proposed in yeast (Storici et al., 2006), which is shown in Figure S5 and further discussed below.

DISCUSSION

Here, we demonstrated that chemically modified ssODNs provide several important benefits to the design of ssODN donors for genome editing. Remarkably, we showed that the use of PS-ssODN donors can give rise to much-higher levels of KI, especially for long inserts, both in cells and in rodents. Chemical modifications, PS or LNA, stabilize ssODNs and probably contribute by increasing the effective concentration of ssODN available during DNA repair. Recently, chemically modified guide RNAs were also demonstrated to improve genome editing (Hendel et al., 2015), and chemical modification of different types of linear donors used in genome editing (Sargent et al., 2011) might also be beneficial, as shown here for short ss and dsODNs. PS-ssODNs have been documented to be toxic or inhibit cell proliferation (Rios et al., 2012), but we found here, however, that isolation of cells, mouse embryos, and newborn rats with

Table 2. Asymmetry of Indels Found in Imprecise KI in U2OS Cells

	Total Reads	% Precise KI	% Imprecise KI	
			% KI + 5' Indels	% KI + Other Indels
sgT2+Cas9+ss3s-PS	169,890	11.2	1.1	–
sgT2+Cas9++ss3as-PS	154,707	8.1	1.2	0.1
TALEN+ss2s-PS	13,748	14.7	1.5	–
TALEN+ss2as-PS	32,186	15	1.8	0.6
TALEN+ssFRTs-PS	106,778	20.8	5	0.1
TALEN+ssFRTas-PS	88,714	18.9	3.1	0.1

The percent of DNA target sequences with precise insertion (precise KI) and imprecise insertion exhibiting indels on the 5' side of the insert, referring to the ssODN sequence orientation, (KI + 5' indels) or other types of indels (KI + other indels) were evaluated by deep sequencing and are indicated for different pairs of nucleases and ssODNs, corresponding to either target sequence orientation (s or as). In all cases, the large majority of imprecise KI sequences exhibit additional indels on the 5' side of the insert sequence, except in the case of TALEN+ss2as-PS. In the latter case, additional indels were likely induced by a second step of DNA cleavage by TALENs and NHEJ repair after an initial precise KI event. Indeed, the 7-bp insert resulted in a 22-bp-long spacer between the two TALEN-binding sites that still allowed for DNA cleavage by TALENs (as checked in vitro; not shown) and generation of additional indels after precise KI.

targeted sequence insertion is readily achieved. Concerning the homology length, we showed that as short as 20 nt on each side of the break is sufficient when using PS-ssODNs. Including modified nucleotides in the donor therefore allows to decrease the homology length while maintaining activity and therefore significantly increases design flexibility. The main advantage is to be able to make longer sequence insertions. Here, a sequence insertion of more than 100 bp could be achieved, allowing the insertion of multiple tags that are useful for protein functional analysis. In ssODN design, the distance between the insertion and cleavage sites is also an important parameter that may benefit from ODN chemical modification because, in practice, cleavage sites are limited to sequences adjacent to NGG or NGA PAM motifs of *S. pyogenes* Cas9 and its VQR variant, respectively (Kleinstiver et al., 2015; Zetsche et al., 2015). We found that a distance of 20 bp still allows for very efficient sequence insertion with PS-ssODN.

Concerning the specificity of integration events, we used the recently described approach, TLA (de Vree et al., 2014), to uncover potential off-target insertion sites. Analyzing one KI rat, we observed no insertion of the ssODN outside of the targeted loci, demonstrating specificity. ssODN integration seems to be a highly specific process because it needs both genomic cleavage and homologous regions, allowing stable base pairing near the break.

In addition to ssODNs, we also evaluated short microhomology-containing duplex donors and showed that chemically modified duplexes also exhibit increased insertion activity. However, very poor levels of precise KI were obtained. Enhancement of MMEJ repair pathways or reduction of NHEJ and better understanding of the mechanisms involved in microhomology-based duplex insertion might increase the precision of sequence insertion and reduce NHEJ-dependent sequence modifications. Using the recently described RNA-guided nuclease Cpf1, which generates 5'mh, instead of Cas9 could also improve the precision of insertion events by providing more-appropriate "acceptor" ends (Zetsche et al., 2015).

With ssODN donors, the insertion events are mostly precise in many biological systems, such as human cell lines or rodents, as

reported here. In contrast, in zebrafish, imprecise KI events are predominant. In the latter case, phenotypic screening was performed here with GFP and made it possible to identify embryos with the wanted sequence insertion by screening a limited number of embryos. Phenotypic screening could also be performed at a control pigmentation locus, easily targeted in parallel to the locus of interest by co-injection of specific guide RNA, in order to select for embryos where genome modification efficiently took place and help limit the number of animals to be raised and screened.

ssODN-mediated genome editing is likely not mediated by the classical homologous recombination pathway, which involves double-stranded DNA template during repair (Majumdar et al., 2008; Storici et al., 2006). Based on our analysis of sequences from imprecise KI events, we propose that genome editing with ssODN takes place by two steps of single-strand annealing as proposed in yeast (Figure S5; Storici et al., 2006), and it will be interesting to test the direct contribution of components of the single-strand annealing (SSA) pathway. Better understanding the factors underlying ssODN-based genome editing and cross-regulations between DNA repair pathways may help to stimulate genome editing efficiency in different cellular contexts, as shown in *Drosophila* (Beumer et al., 2013a) or more recently in mouse embryos, where LIG4 inhibition allowed to favor KI events (Chu et al., 2015). In conclusion, we propose here an optimized KI approach based on PS-modified ssODN donors and validate its improved activity in different biological systems, both in cultured cells and in vivo. Such scalable and robust approach should be useful for precise sequence manipulation, expanding the versatility and applicability of genome editing with sequence-specific nucleases.

EXPERIMENTAL PROCEDURES

Plasmid, oligonucleotide, RNA, and TALEN materials are detailed in [Supplemental Information](#).

Cell Transfection

The human cell lines U2OS, RPE1, and rat C6 cells (ATCC) were nucleofected (5×10^5 – 10^6 cells; kit V [Lonza] and Amaxa program X-001) with 2 μ g of each TALEN or 2 μ g of each Cas9 and guide RNA plasmids.

Microinjections into Rat, Mouse, and Zebrafish One-Cell-Stage Embryos

One-cell-stage rat and mouse zygotes were microinjected by standard procedures. For zebrafish injections, *ubi:eGFP* line from Mosimann lab was used (ubiquitous eGFP expression driven by the *ubiquitin* promoter; see details in Supplemental Information).

T7EI Mutation Detection and Calculation

T7 Endonuclease I (T7EI) assays were performed as previously described (Piganeau et al., 2013) using primers in Table S4.

Sequence modification frequencies were estimated as commonly done: indel rate = $1 - (1 - X_c)^{1/2}$ with X_c , rate of cleaved products and if $X_c < 0.15$, indel rate = $X_c/2$. For this calculation, the hypothesis is that all heteroduplexes formed during hybridization are cleaved by T7EI. If there is, in addition to wild-type (WT), another homogenous population (X_1), as in our KI experiments here, then the previous hypothesis is not appropriate and $X_2 = (1 - X_1) - (1 - X_c - [X_1 \times X_1])^{1/2}$ with X_2 , heterogeneous mutated population and indel rate = $X_2 + X_1$. We observed that, when KI rates are higher than T7EI cleavage rates obtained in the absence of donor, the T7EI assay was likely not able to cleave all heteroduplexes, especially the ones with the FLAG sequence insertion (that are abundant when KI rates are high), and even using the corrected formula, we underestimated the mutation rates. It is the reason why we chose in these cases to report the rate of T7EI cleavage and not estimated indel rates.

Evaluation of KI Rates

Integration rates were evaluated by RFLP or PCR. DNA was extracted from transfected cells generally at 6 days after nucleofection. When necessary, nuclease treatment was done before PCR reaction to eliminate residual ODN (see details in Supplemental Information). PCR of the target sites was performed with Phusion polymerase (New England Biolabs). RFLP analysis was performed on 5 μ l of the 25 μ l PCR reaction and run on 2.5% agarose gels. For analysis of tag insertion, small PCR products spanning the insertion site were resolved on 3% agarose gels and %KI quantified by densitometry measurements of the bands using ImageJ.

TLA

TLA and next-generation sequencing using primers within the ssODNs (see Table S5 and Figure S4A) were used to determine the transgenic ssODNs integration sites as previously described (de Vree et al., 2014).

Amplicon Library Preparation, NGS Sequencing, and Sequence Analysis

DNA was isolated from transfected cells (EZNA tissue kits; Omega Biotek) or zebrafish embryos, and 100–150 ng was added to a 50- μ l Phusion polymerase (NEB) reaction. Each sample was assigned a primer set with a unique barcode to enable multiplex sequencing (Table S5). PCR products were purified on a 2% agarose gel, quantified, and pooled into a single sample for sequencing. The single combined sample was treated by the MNHN Genomics Center and sequenced on Ion Torrent PGM. A custom python pipeline was used to count and characterize indels and insert types.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2016.02.018>.

AUTHOR CONTRIBUTIONS

C.G. and J.-P.C. conceived and designed experiments; J.-B.R., C.B., M.C., S.F., A.D.C., L.P., L.T., R.T., S.M., J.E., and N.d.C. performed experiments and did some analyses; J.C. and E.D.-B. wrote the sequence analysis software; C.G., J.-P.C., F.S., Y.C., S.F., A.F., and I.A. analyzed results; and C.G. and J.-P.C. wrote the manuscript with the help of all the authors.

ACKNOWLEDGMENTS

We thank Loïc Ponger and Nicolas Buisine for advice in the development of the software for sequence analysis, Jose Utge and Regis Debruyne for PGM sequencing, Erika Brunet for helpful discussions, genOway's production lab, Wim van Schooten at Cergentis for TLA analysis, and Alain Hilgers for support. This work was supported by ANR Investissement d'Avenir ANR-II-INSB-0014, AFM n°18566, Région Pays de la Loire through Biogenouest, IBISA program, Fondation Progreffe, and ANR LabCom (project "SOURIRAT"). Sequencing experiments were performed by the integrative genomic facility of Nantes and the SSM department at the Museum National d'Histoire Naturelle in Paris. S.F., Y.C., and A.F. are employees of genOway company.

Received: November 4, 2015

Revised: December 16, 2015

Accepted: January 28, 2016

Published: February 25, 2016

REFERENCES

- Bae, S., Kweon, J., Kim, H.S., and Kim, J.S. (2014). Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods* 11, 705–706.
- Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug, R.G., 2nd, Tan, W., Penheiter, S.G., Ma, A.C., Leung, A.Y., et al. (2012). In vivo genome editing using a high-efficiency TALEN system. *Nature* 491, 114–118.
- Bétermier, M., Bertrand, P., and Lopez, B.S. (2014). Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet.* 10, e1004086.
- Beumer, K.J., Trautman, J.K., Christian, M., Dahlem, T.J., Lake, C.M., Hawley, R.S., Grunwald, D.J., Voytas, D.F., and Carroll, D. (2013a). Comparing zinc finger nucleases and transcription activator-like effector nucleases for gene targeting in *Drosophila*. *G3 (Bethesda)* 3, 1717–1725.
- Beumer, K.J., Trautman, J.K., Mukherjee, K., and Carroll, D. (2013b). Donor DNA utilization during gene targeting with zinc-finger nucleases. *G3 (Bethesda)* 3, 657–664.
- Branda, C.S., and Dymecki, S.M. (2004). Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell* 6, 7–28.
- Carroll, D., and Beumer, K.J. (2014). Genome engineering with TALENs and ZFNs: repair pathways and donor design. *Methods* 69, 137–141.
- Chen, F., Pruett-Miller, S.M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., Colingwood, T.N., Frodin, M., and Davis, G.D. (2011). High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* 8, 753–755.
- Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33, 543–548.
- de Vree, P.J., de Wit, E., Yilmaz, M., van de Heijning, M., Klous, P., Verstegen, M.J., Wan, Y., Teunissen, H., Krijger, P.H., Geeven, G., et al. (2014). Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat. Biotechnol.* 32, 1019–1025.
- Dorsett, Y., Zhou, Y., Tubbs, A.T., Chen, B.R., Purman, C., Lee, B.S., George, R., Bredemeyer, A.L., Zhao, J.Y., Soderger, E., et al. (2014). HCoDES reveals chromosomal DNA end structures with single-nucleotide resolution. *Mol. Cell* 56, 808–818.
- Eckstein, F. (2000). Phosphorothioate oligodeoxynucleotides: what is their origin and what is unique about them? *Antisense Nucleic Acid Drug Dev.* 10, 117–121.
- Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., and Schier, A.F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS ONE* 9, e98186.
- Harel, I., Benayoun, B.A., Machado, B., Singh, P.P., Hu, C.K., Pech, M.F., Valenzano, D.R., Zhang, E., Sharp, S.C., Artandi, S.E., and Brunet, A. (2015). A

- platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. *Cell* 160, 1013–1026.
- Hendel, A., Bak, R.O., Clark, J.T., Kennedy, A.B., Ryan, D.E., Roy, S., Steinfeld, I., Lunstad, B.D., Kaiser, R.J., Wilkens, A.B., et al. (2015). Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* 33, 985–989.
- Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C., and Schmid, B. (2013). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development* 140, 4982–4987.
- Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278.
- Jepsen, J.S., and Wengel, J. (2004). LNA-antisense rivals siRNA for gene silencing. *Curr. Opin. Drug Discov. Devel.* 7, 188–194.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821.
- Kim, H., and Kim, J.S. (2014). A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* 15, 321–334.
- Kleistiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R., et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523, 481–485.
- Majumdar, A., Muniandy, P.A., Liu, J., Liu, J.L., Liu, S.T., Cuenoud, B., and Seidman, M.M. (2008). Targeted gene knock in and sequence modulation mediated by a psoralen-linked triplex-forming oligonucleotide. *J. Biol. Chem.* 283, 11244–11252.
- Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., Daimon, T., Sezutsu, H., Yamamoto, T., Sakuma, T., and Suzuki, K.T. (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* 5, 5560.
- Orlando, S.J., Santiago, Y., DeKelver, R.C., Freyvert, Y., Boydston, E.A., Moehle, E.A., Choi, V.M., Gopalan, S.M., Lou, J.F., Li, J., et al. (2010). Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res.* 38, e152.
- Paix, A., Wang, Y., Smith, H.E., Lee, C.Y., Calidas, D., Lu, T., Smith, J., Schmidt, H., Krause, M.W., and Seydoux, G. (2014). Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in *Caenorhabditis elegans*. *Genetics* 198, 1347–1356.
- Piganeau, M., Ghezraoui, H., De Cian, A., Guittat, L., Tomishima, M., Perrouault, L., René, O., Katibah, G.E., Zhang, L., Holmes, M.C., et al. (2013). Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome Res.* 23, 1182–1193.
- Porteus, M.H., and Carroll, D. (2005). Gene targeting using zinc finger nucleases. *Nat. Biotechnol.* 23, 967–973.
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308.
- Rios, X., Briggs, A.W., Christodoulou, D., Gorham, J.M., Seidman, J.G., and Church, G.M. (2012). Stable gene targeting in human cells using single-strand oligonucleotides with modified bases. *PLoS ONE* 7, e36697.
- Sargent, R.G., Kim, S., and Gruenert, D.C. (2011). Oligo/polynucleotide-based gene modification: strategies and therapeutic potential. *Oligonucleotides* 21, 55–75.
- Shao, Y., Guan, Y., Wang, L., Qiu, Z., Liu, M., Chen, Y., Wu, L., Li, Y., Ma, X., Liu, M., and Li, D. (2014). CRISPR/Cas-mediated genome editing in the rat via direct injection of one-cell embryos. *Nat. Protoc.* 9, 2493–2512.
- Storici, F., Snipe, J.R., Chan, G.K., Gordenin, D.A., and Resnick, M.A. (2006). Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. *Mol. Cell. Biol.* 26, 7645–7657.
- Tan, W., Carlson, D.F., Lancto, C.A., Garbe, J.R., Webster, D.A., Hackett, P.B., and Fahrenkrug, S.C. (2013). Efficient nonmeiotic allele introgression in livestock using custom endonucleases. *Proc. Natl. Acad. Sci. USA* 110, 16526–16531.
- Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33, 187–197.
- Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A.W., Rios, X., et al. (2013). Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.* 41, 9049–9061.
- Yang, H., Wang, H., and Jaenisch, R. (2014). Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat. Protoc.* 9, 1956–1968.
- Yoshimi, K., Kaneko, T., Voigt, B., and Mashimo, T. (2014). Allele-specific genome editing and correction of disease-associated phenotypes in rats using the CRISPR-Cas platform. *Nat. Commun.* 5, 4240.
- Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771.

2.3.2.2 Les TALEN, une thérapie pour la dystrophie facio scapulo humérale ?

La dystrophie facio scapulo humérale (FSHD) est une maladie musculaire d'origine génétique pour laquelle il n'existe actuellement pas de traitement curatif. Elle est caractérisée par la perte de marques épigénétiques répressives au sein du macrosatellite composé de répétitions D4Z4 et entraînant une transcription aberrante d'une protéine toxique, DUX4. Cette transcription est possible si la perte de marques répressives est associée à la présence d'un site de polyadénylation à la suite des répétitions D4Z4 permettant la stabilisation de l'ARNm DUX4. L'objectif ici était de cibler et de muter le site de polyadénylation par une approche d'édition du génome à l'aide de TALEN ou du système CRISPR/Cas9 dans des cellules issues de patients FSHD. Dans cet objectif, j'ai assemblé différentes nucléases et testé celles-ci dans les cellules U2OS afin de déterminer l'approche la plus efficace (Figure 22). J'ai ainsi testé différents couples de TALEN ainsi que des ARN guides à utiliser avec les nucléases Cas9 de *S.pyogenes* et *N.meningitis*, dont les séquences PAM sont respectivement NGG et NNNNGATT. La stratégie qui a été retenue est celle utilisant le couple de TALEN 418/419 qui est très efficace (39% de cibles mutées) et idéalement placé pour inactiver le site de polyadénylation puisque ce dernier se trouve dans la séquence spacer situé entre les sites de fixation des deux sous-unités TALEN 418 et 419.

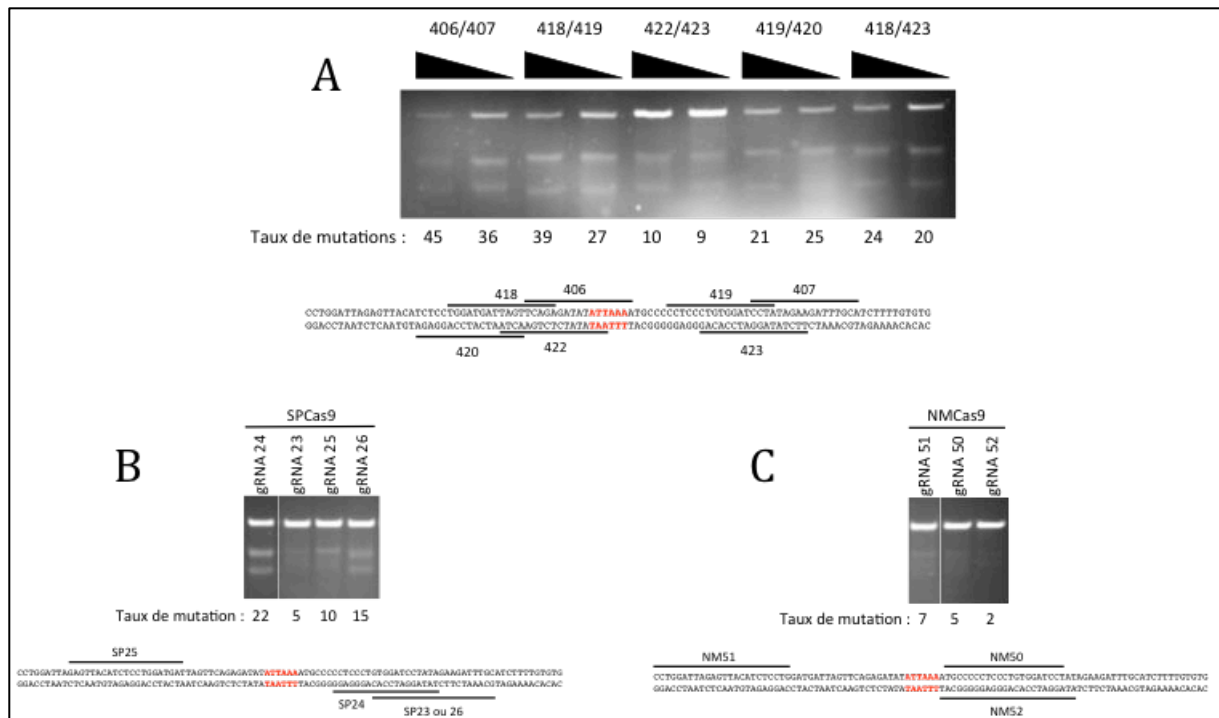


Figure 22. Comparaison de l'efficacité des différentes nucléase au site DUX4 dans les cellules U2OS pour inactiver le site de polyadénylation (en rouge).

(A) Taux de mutations obtenu pour les différents couples de TALEN. Pour chaque couple de TALEN deux doses de nucléases ont été testés, 0,5 et 1 μ g de chaque TALEN. (B) Taux de mutations obtenu pour les différents ARN guides de la Cas9 de *S. pyogenes*. (C) Taux de mutations obtenu pour les différents ARN guides de la Cas9 de *N.meningitidis*.

2.3.2.3 Conclusion

Ces deux projets ont permis d'étudier à la fois l'utilisation des TALEN et des systèmes CRISPR/Cas9 de *S.pyogenes* et *N.meningitidis*. Ces différents systèmes sont efficaces et présentent chacun des avantages. Les différents systèmes CRISPR/Cas9 sont simples à utiliser et à assembler alors que les TALEN peuvent être un peu plus difficiles à assembler. Cependant, comme vu pour le projet DUX4, les TALEN peuvent être plus efficaces et permettre une plus grande amplitude de choix de séquence cible. En effet, ces dernières nécessitant peu de contraintes dans le choix de séquence.

2.3.2.4 Autres publications

J'ai participé au développement d'autres approches d'édition du génome qui ont donné lieu à des publications, listées ci dessous, en particulier chez le rat, en collaboration avec le laboratoire d'I Anegon, et le poisson zèbre, en collaboration avec le laboratoire de F del Bene.

2.3.2.4.1 Article : Edition du génome de rat par Knock-In

J'ai assemblé plusieurs couples de TALEN pour cibler des locus dans le génome de rat (Rosa26 et Hprt1). Ces nucléases artificielles ont été utilisées sous forme d'ARNm par l'équipe d'I Anegon afin d'insérer le cDNA de la GFP par RH à ces locus. Jusqu'à 4% des œufs injectés présentaient un phénotype GFP positif, ces animaux ont pu transmettre cette mutation à leur descendance (Remy et al., 2014) (Annexe 4).

2.3.2.4.2 Article : Edition du génome de poisson zèbre par Knock-Out

A l'aide de TALEN que j'ai assemblé, l'équipe de F Del Bene a inactivé le gène *pkd2l1* dans des poisson zèbre afin d'étudier l'implication du canal exprimé à partir de ce gène dans la réponse aux stimuli de neurones en contact direct avec le liquide cébrospinal (Böhm et al., 2016) (Annexe 5).

2.3.3 Conclusion

Le protocole mis au point au laboratoire permet un rendement important des nucléases TALE avec peu d'erreurs. Les seules erreurs survenant encore ne peuvent être réglés par des optimisations du protocole. Elles sont le plus souvent dues à des recombinaisons entre les répétitions. Ces recombinaisons peuvent être limitées en choisissant des séquences cibles ne comportant pas des blocs de plusieurs nucléotides identiques, ce type de séquence favorise les erreurs.

De nombreuses collaborations ont été développées à l'aide de ce savoir faire en terme de production de TALEN, cependant avec l'arrivée du système CRISPR-Cas9 la demande en TALEN a rapidement diminué et se limite maintenant exclusivement à des applications très spécifiques, comme vu pour le projet DUX4. Une autre application nécessite l'utilisation de TALEN, c'est l'induction de CDB dans le génome mitochondrial (Bacman et al., 2013; Gammage et al., 2016). En effet, jusqu'à présent, seules les TALEN ont pu être dirigées vers les mitochondries à l'aide de signaux de localisation mitochondriale. Dans le cas du système CRISPR/Cas9, la translocation de l'ARN guide semble être une difficulté technique difficile à lever.

3 Discussion et perspectives

Lors de ma thèse j'ai testé différentes approches pour augmenter le taux d'insertion par homologie d'un transgène exprimant la GFP avec l'aide de nucléases artificielles comme les TALEN mais principalement à l'aide du système CRISPR-Cas9.

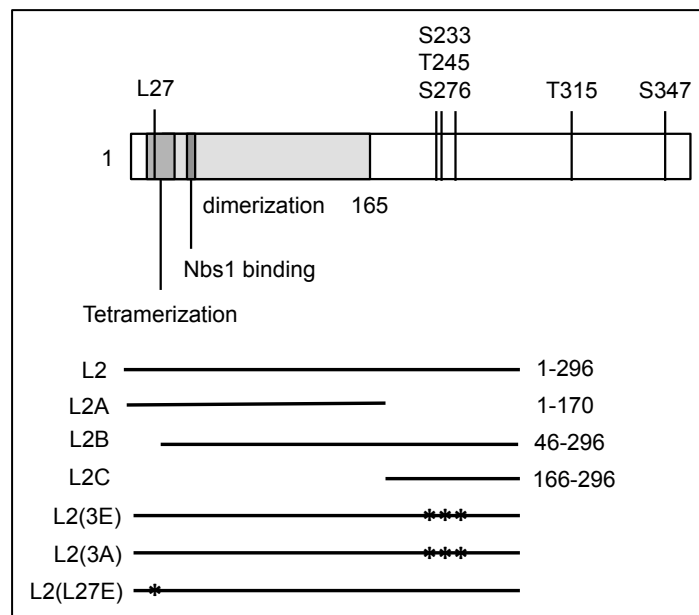


Figure 23. Composition du domaine L2 de la protéine CtIP.

Sont représentés ici le domaine L2 ainsi que les différents mutants testés : Les mutants L2A, L2B et L2C sont des parties du domaine L2, les mutants L2(3E) et (3A) ont leurs sites de phosphorylation par CDK (S233, T245 et S276) mutés pour mimer (E) ou bloquer la phosphorylation (A) et le mutant L2(L27E) ne peut tétramériser avec CtIP.

L'utilisation de la protéine CtIP, impliquée directement dans le système de réparation par RH mais aussi par MMEJ ou encore SSA, a montré des résultats très intéressants. Une fois fusionnée à la Cas9wt notamment, on observe une augmentation du taux d'insertion par homologie de la GFP au site AAVS1 de 2 à 3 fois par rapport à la Cas9 seule (Figure 1-C de l'article). Afin de mieux comprendre la raison de cet effet, la protéine CtIP a été décomposée et chaque fragment a été testé dans les mêmes conditions, en parallèle de la Cas9wt fusionnée à CtIPwt. On observe alors clairement que l'effet observé est dû à la partie Nter de la protéine CtIP (Figure 2-B de l'article). Le plus petit fragment alors

fonctionnel est le fragment L2 composé des 296 premiers acides aminés de la protéine CtIP. Dans cette partie de CtIP se trouvent différents domaines clés : un domaine de fixation à Nbs1, un domaine de dimérisation, un domaine de tétramérisation et différents sites de phosphorylations par CDK ou ATM (Figure 23).

Pour comprendre d'où vient l'effet du domaine L2 sur l'intégration par homologie, différentes mutations du domaine L2 ont été testées (Figure 3 de l'article). On observe alors que le domaine de tétramérisation est important et que les sites de phosphorylations par CDK sont indispensables. Différentes hypothèses pour expliquer pourquoi le domaine L2 permet de favoriser l'intégration par homologie peuvent alors être émises.

3.1.1 Mécanismes d'action possibles du domaine L2

3.1.1.1 Recrutement de CtIP endogène par tétramérisation au domaine L2 ?

Il a en effet été montré que CtIP doit former des tétramères pour être actif et promouvoir la résection des extrémités des CDB et donc la réparation par RH (Davies et al., 2015). Le domaine de tétramérisation de CtIP s'étend de l'aa 18 à l'aa 52 et est donc bien présent sur le domaine L2 de CtIP (Figure 23). Afin de tester l'importance de ce domaine de tétramérisation dans la stimulation du KI, le mutant L27E qui permet d'inhiber la tétramérisation de CtIP et non sa dimérisation (Davies et al., 2015) a été testé (Figure 3-D de l'article). L'utilisation de la Cas9-L2(L27E) permet d'obtenir autant d'insertion par homologie du donneur GFP qu'avec la Cas9-L2, mais induit aussi plus de mutations par NHEJ et la protéine de fusion est bien plus abondante, comme observé en Western blot. Cette nucléase Cas9-L2(L27E) est donc moins efficace que Cas9-L2, montrant ainsi l'importance du domaine de tétramérisation.

Afin de déterminer si l'effet est dû au recrutement de CtIP endogène plusieurs tests peuvent être envisagés. La surexpression de CtIP pourrait potentiellement amplifier

l'effet de la fusion Cas9-L2 sur le KI. Cette possibilité a été testée sur le locus AAVS1 dans des HEK293 (Figure 24). La surexpression de CtIP n'a pas permis d'augmenter l'efficacité de Cas9-L2.

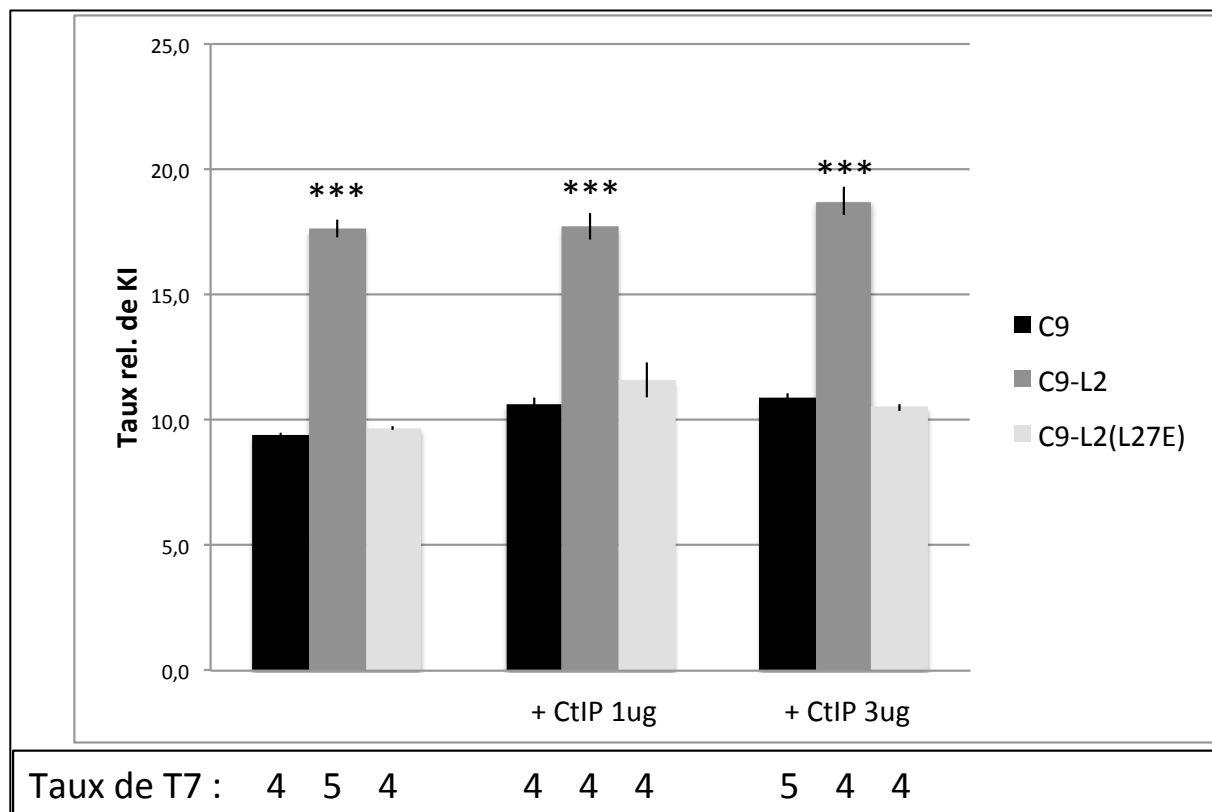


Figure 24. Taux relatif d'insertion par KI d'un donneur GFP exogène au site AAVS1 obtenu dans des HEK293 à l'aide du système CRISPR/Cas9 et de l'ARNg T2.

Différents mutants de la nucléase Cas9 ont été utilisés, fusionnée avec le domaine L2 de CtIP ou avec le domaine L2(L27E) dont la capacité à tétramériser est inactivée, avec ou sans surexpression parallèle de CtIP wt. Afin de déterminer si les moyennes des taux de mutations obtenus avec les différentes nucléases sont différentes un t-test a été utilisé (n = 3).

Ce résultat ne remet pas nécessairement en cause l'hypothèse du recrutement de CtIP par la fusion Cas9-L2. On peut en effet imaginer que CtIP endogène n'est pas limitant ici pour expliquer l'absence d'augmentation de KI après surexpression de CtIP, ou que la surexpression n'a pas fonctionné. Pour vérifier ce point, des western blot pourront être

réalisés. Pour tester l'hypothèse du recrutement de CtIP endogène, d'autres expériences pourraient être faites. Il serait possible d'immunoprécipiter Cas9-L2 après transfection pour voir si CtIP est coimmunoprécipité et inversement. Il serait aussi intéressant d'inhiber CtIP endogène, par interférence ARN par exemple, et de coexprimer avec la Cas9-L2, CtIP wt ou CtIP muté pour son domaine de tétramérisation (mutant L27E) ou encore seulement le domaine L2 de CtIP. On s'attendrait alors à observer une augmentation de l'intégration par homologie seulement avec l'expression de CtIP wt et non pas avec CtIP(L27E) ni L2, pour lesquels il ne devrait plus y avoir de réparation par RH puisque CtIP endogène est éteint.

3.1.1.2 Recrutement d'autres Cas9-L2 par tétramérisation au domaine L2 ?

Le fait que Cas9-L2 puisse recruter d'autres molécules Cas9-L2 par le domaine de tétramérisation du domaine L2 semble plausible cependant son effet sur la réparation par homologie un peu difficile à comprendre. Lors d'une tentative d'expression et de purification de la protéine Cas9-L2, des difficultés ont été rencontrées permettant de mettre en évidence une possible tétramérisation de ces nucléases. En effet, afin d'utiliser cette Cas9-L2 *in vivo* chez le rat notamment, une tentative d'expression et de purification de Cas9-L2 à partir de bactéries *E. Coli* a été faite (résultats non montrés ici). La nucléase modifiée a bien été exprimée dans les bactéries sans mortalité particulière observée, cependant, la purification fût quasiment impossible. Le tag His présent en Cter à la suite du domaine L2 était sans doute inaccessible, peut être dû à la formation de tétramères *via* le domaine L2.

Dans le cas où l'effet est dû au recrutement d'autres Cas9-L2 au site de la CDB, il est assez difficile d'imaginer ce qui pourrait se passer mécaniquement. Est-ce que les autres Cas9 recrutées clivent l'ADN elles aussi ? Difficile à croire, elles ne pourraient pas fixer d'ARNg au même endroit. Est-ce que cela permettrait d'ouvrir encore plus l'ADN autour de la CDB ? Est-ce que cette forte concentration de Cas9-L2 pourrait favoriser la dégradation des extrémités de la coupure par encombrement des extrémités de la CDB ?

Difficile d'imaginer que cet encombrement des extrémités soit stable et que cela n'empêche pas le clivage de la cible.

Afin de vérifier cette hypothèse, il faudrait effectuer une immunoprécipitation de Cas9-L2 après transfection avec le domaine L2 seul. Il serait alors possible de vérifier par Western Blot si le domaine L2 seul est coimmunoprécipité.

3.1.1.3 Recrutement d'autres protéines par le domaine L2 ?

Avec l'utilisation des mutants Cas9-L2(3A) et Cas9-L2(3E) nous avons observé que la phosphorylation de ces sites par CDK est indispensable à l'effet observé sur le taux de KI (Figure 3-C de l'article). La phosphorylation de ces trois sites (S233, T245 et S276) ainsi que S347 et T315 est indispensable à la fixation de Nbs1 sur CtIP (Wang et al., 2013). S'il n'y a pas fixation de Nbs1, CtIP ne peut être phosphorylé par ATM et il ne sera pas actif pour permettre la réparation par RH d'une CDB. Nbs1 interagit avec CtIP via deux domaines distincts, BRCT et FHA. Le domaine BRCT se fixe préférentiellement aux sérines phosphorylées et FHA aux thréonines phosphorylées (Wang et al., 2013). Le domaine L2 de CtIP comporte 3 de ces 5 sites de phosphorylation par CDK importants pour la fixation de Nbs1 (S233, T245 et S276). Est-ce suffisant pour la fixation de Nbs1 ? Pour le vérifier, il faudrait immunoprécipiter Cas9-L2 et ses mutants, Cas9-L2(3A) et Cas9-L3(3E), et voir si on coprécipite aussi Nbs1, et inversement.

Si ces 3 sites sont suffisants pour fixer Nbs1, Cas9-L2 pourrait être phosphorylé par ATM. Cependant, des différents sites cibles d'ATM il n'en reste que deux sur le domaine L2, S231 et T271. La mutation de ces deux en acides aminés non phosphorylables n'a aucune incidence sur la réparation par RH (Wang et al., 2013). Le mutant Cas9-L2(3A) ne bloquerait donc pas l'effet positif sur la réparation par RH en inhibant la phosphorylation par ATM, mais ce point reste à vérifier en mutant les sites de phosphorylation par ATM sur le domaine L2.

Une autre hypothèse serait que si les 3 acides aminés phosphorylés par CDK peuvent fixer Nbs1, celui-ci permette le recrutement au site de coupure du complexe MRN dont il

fait partie. Ce qui pourrait permettre d'initier la résection des extrémités de la coupure et le recrutement des autres acteurs de la réparation par RH. Pour vérifier cette hypothèse il faudrait voir, en plus de Nbs1, Rad50 et Mre11 après immunoprécipitation de Cas9-L2.

Une autre hypothèse pourrait être que ces 3 sites de phosphorylation par CDK participent au recrutement des protéines BLM et EXO1 impliquées dans l'amplification de la résection des extrémités des CDB (Wang et al., 2013).

3.1.1.4 Renouvellement rapide de la nucléase grâce à la présence du domaine L2 ?

Sur tous les Western Blot effectués, la Cas9-L2 est toujours présente en moins grande quantité que la Cas9wt ou que les constructions Cas9 fusionnée à d'autres fragments de CtIP (Figure 3 de l'article). Cette observation semble surprenante quand on sait que le plasmide utilisé est exactement le même et que la quantité transfectée est toujours la même. Est ce que la présence du domaine L2 en Cter de la nucléase Cas9 entraîne sa dégradation plus rapide ?

Au vu de ces résultats, j'ai testé l'effet d'un inhibiteur de protéasome (MG132) sur la quantité de protéines Cas9-L2 produite. Une augmentation de la quantité de Cas-L2 est observée avec des concentrations d'inhibiteur de plus en plus fortes, pour la Cas9wt, le résultat est moins clair (Figure 26). Cependant, avec une augmentation de la concentration d'inhibiteur de protéasome on observe une diminution des mutations au site de coupure par test T7E1 et une légère diminution du taux de KI pour les deux nucléases (Figure 25).

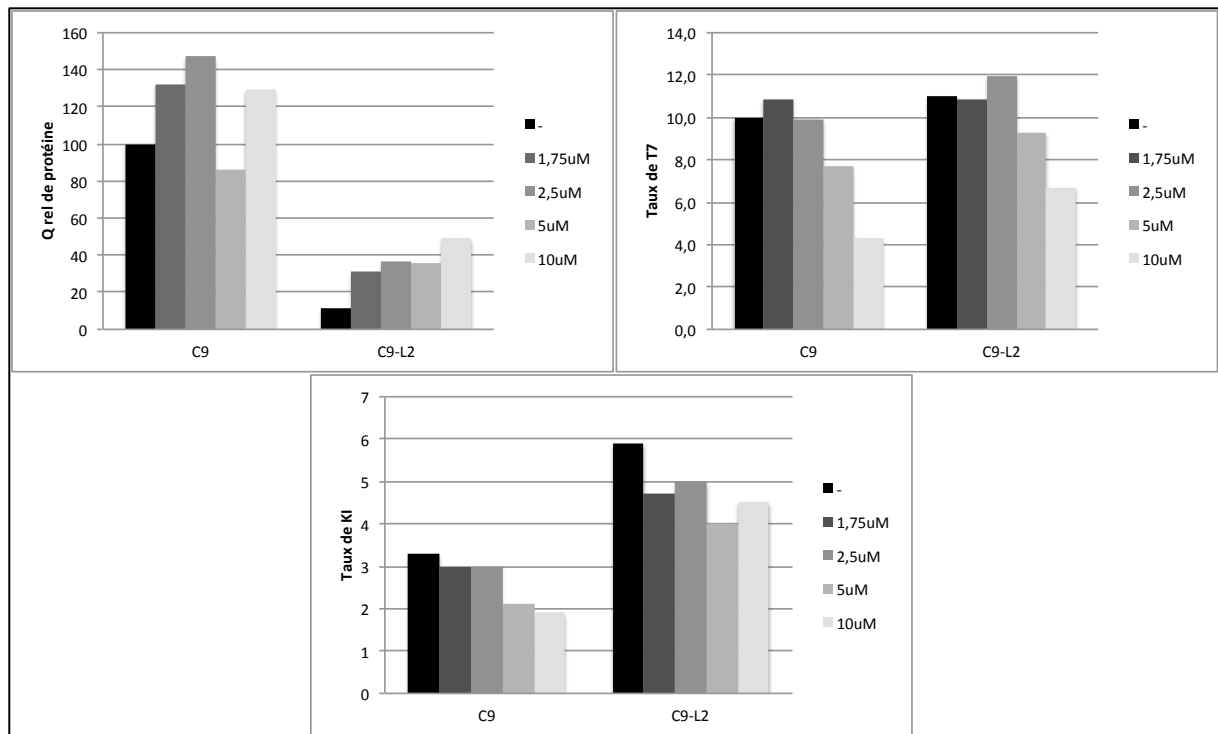


Figure 25. Effet de l'inhibition du protéasome sur l'effet de la Cas9-L2 sur le taux de mutations et le taux d'insertion par homologie d'un donneur exogène exprimant la GFP dans des HEK293.

Les différentes doses d'inhibiteur de protéasome (MG132), appliquées sur les cellules le jour de la transfection et pour 4 jours, modifient la quantité de Cas9-L2 présente sur les Western Blot, le taux de mutations ainsi que le taux de KI.

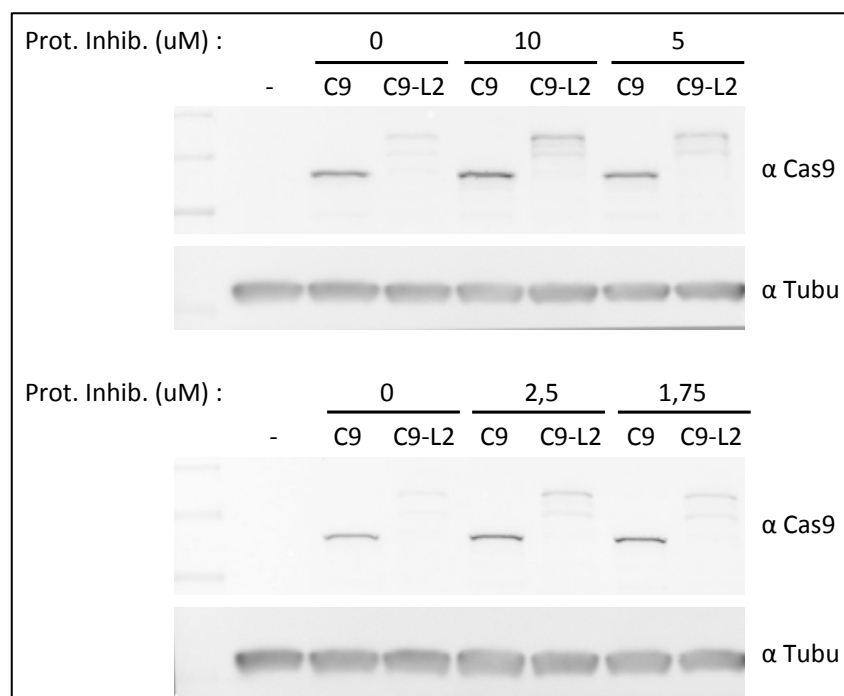


Figure 26. Western blot après transfection dans des HEK293 de la Cas9 wt et de la Cas9-L2 avec inhibition du protéasome par différentes doses de MG132.

La dégradation de la nucléase pourrait jouer un rôle direct ou indirect dans le taux de coupure et le taux d'insertion par KI au site de la CDB, que ce soit avec la Cas9wt ou la Cas9-L2. Mais il est important de reproduire ces résultats préliminaires avant de pouvoir conclure.

Une autre approche pour définir si la dégradation de Cas9-L2 joue un rôle dans l'effet observé sur le taux de KI serait de bloquer l'expression de notre plasmide transfecté (Cycloheximide, inhibiteur de la traduction) et d'observer la vitesse de dégradation de nos différentes nucléases (Cas9wt, Cas9-L2). Cette approche pourrait nous permettre de déterminer clairement si la Cas9-L2 est moins stable que les autres formes.

3.1.1.5 L'effet de Cas9-L2 peut-il être cellule et locus dépendant ?

L'accessibilité de certains locus génomiques peut être plus difficile et empêcher la nucléase de se fixer à sa cible. Par exemple, des nucléases ciblant la même séquence fonctionnent *in vitro* mais pas en cellule (Guilinger et al., 2014; Pattanayak et al., 2013). Cela peut être dû au statut épigénétique des différentes cibles, qui peut différer selon le type cellulaire.

Une étude est en cours pour tester cette nucléase Cas9-L2 dans plusieurs types cellulaires ainsi que sur différents locus (Résultats non montrés ici). Et il semble que cette nouvelle nucléase ne permette pas d'augmenter le taux de KI dans certaines lignées cellulaires, les U2OS ou les RPE1. Comment l'expliquer ? Il est possible que l'effet du domaine L2 soit directement lié à certaines protéines (Nbs1, CtIP, Mre11...) et que l'expression de ces protéines diffère selon la lignée cellulaire utilisée.

3.1.2 Conclusion

L'utilisation de la nucléase Cas9 fusionnée au domaine L2 de la protéine CtIP permet dans de nombreux cas de favoriser l'utilisation de la RH comme système de réparation des CDB. La raison de cet effet n'est pas encore claire. Cependant on peut affirmer que le domaine de tétramérisation est indispensable, il pourrait alors recruter CtIP endogène et favoriser la RH par ce biais (Figure 27). Les sites de phosphorylation par CDK jouent aussi un rôle, par exemple en recrutant d'autres acteurs une fois phosphorylés. En tout cas ce domaine L2, une fois fusionné à la nucléase Cas9 est très intéressant pour les approches d'édition du génome par KI.

Une autre question peut être soulevée, si le domaine L2 influence la réparation par homologie, modifie-t-il la réparation par NHEJ ? Quand on compare les taux de mutations, reflétant l'activité de réparation mutagène par NHEJ ou MMEJ, obtenus avec l'utilisation de la Cas9wt et de la Cas9-L2 il n'y a pas de différence significative (Figure 1-C). Deux hypothèses sont alors à envisager, la première est que l'utilisation de la Cas9-L2 permet d'effectuer plus de CDB qui seront ensuite réparées préférentiellement par RH,

permettant de garder le taux de mutations équivalent à celui obtenu par la Cas9wt. La deuxième hypothèse est que l'utilisation de la Cas9-L2 augmente la résection des extrémités, permettant d'augmenter le taux de RH au détriment de la réparation par NHEJ fidèle tout en augmentant la réparation par NHEJ mutagène, permettant ainsi de garder le même taux de mutations en test T7E1.

Pour conclure, je dirais que ce système est simple d'utilisation puisqu'il s'agit d'une Cas9 plus performante qui peut être utilisée à la place de la Cas9wt. Ce système ne demande pas de traitement préalable ni de transfection d'autres plasmides au contraire des autres approches utilisées à ce jour. Le fragment L2 ne faisant que 196 aa, cette nouvelle nucléase peut être aussi facilement transfectée que la Cas9 wt et aussi vectorisée ou encore produite en protéine pour injection dans des œufs fécondés, après avoir mis au point les conditions de purification. La Cas9-L2 représente une alternative simple à la Cas9wt et permettrait de favoriser l'insertion par homologie d'un donneur exogène. Cette nouvelle Cas9 est actuellement testée *in vivo*, chez le rat et le poisson zèbre ainsi que des des iPSCs.

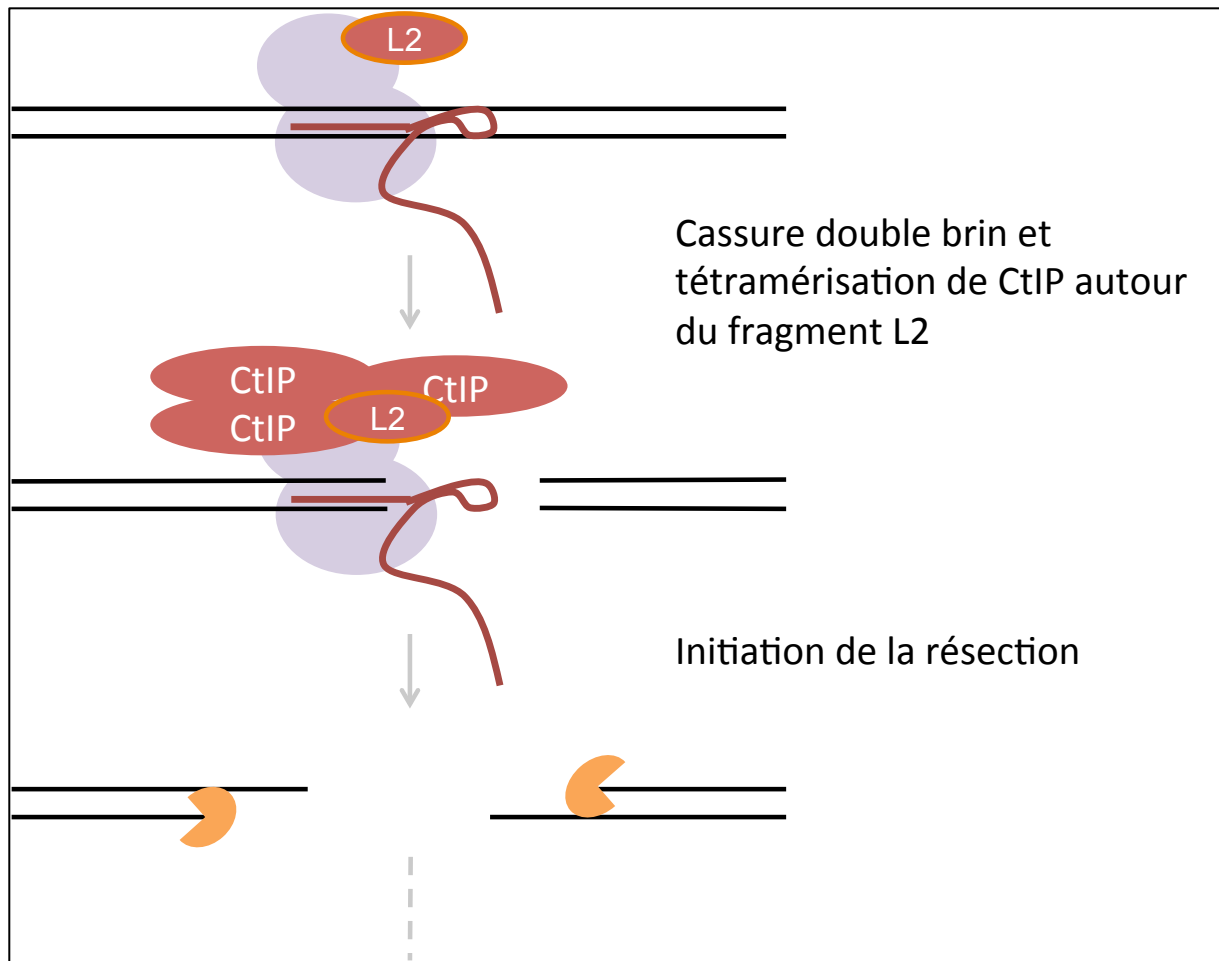


Figure 27. Modèle de recrutement de CtIP par le domaine L2 de CtIP fusionné à la Cas9 wt permettant de favoriser la réparation des CDB par la RH.

3.1.3 Effet d'autres protéines sur le taux de KI

En plus de mieux comprendre pourquoi ce domaine L2 de la protéine CtIP permet d'augmenter la réparation par homologie, il serait intéressant de tester de nouvelles protéines pour augmenter le taux de KI obtenu après coupure par une nucléase. Des protéines de réparation comme Rad18 ou Nbs1 ont été testées, pour favoriser une intégration par RH plus efficace. On peut facilement imaginer d'autres protéines candidates pour renforcer l'effet obtenu ou encore élargir son application à un plus grand nombre de systèmes biologiques, telles que BRCA1, Mre11 ou encore Exo1.

3.1.3.1 Utilisation de la protéine Rad18 pour augmenter le taux de Knock-In

La protéine Rad18 nous a semblé être un candidat intéressant pour augmenter le taux de KI après coupure par les nucléases artificielles. En effet, cette E3 ubiquitine ligase permet le recrutement de BRCA2 et de Rad51 au site de coupure, protéines clé pour la réparation par RH (Tripathi et al., 2016). Rad18 permet aussi d'inhiber la réparation par NHEJ en inhibant le recrutement de 53BP1 et RNF168 au site de cassure (Helchowski et al., 2013).

Cette protéine a été surexprimée dans des U2OS contenant le rapporteur DR-GFP en même temps que l'enzyme I-SceI et elle permet d'augmenter le taux de recombinaison intramoléculaire (Helchowski et al., 2013). L'utilisation de son domaine UBZ seul permet aussi d'augmenter ce taux de recombinaison. Rad18 se fixe aux histones *via* ce domaine UBZ et entre ainsi en compétition avec 53BP1.

C'est pourquoi j'ai testé ces protéines, Rad18 et son domaine UBZ, au site AAVS1 dans des cellules U2OS, RPE1 et RG37DR, en surexpression ou fusionnée à la nucléase Cas9wt afin d'augmenter le taux d'insertion d'un donneur GFP (Figure 28).

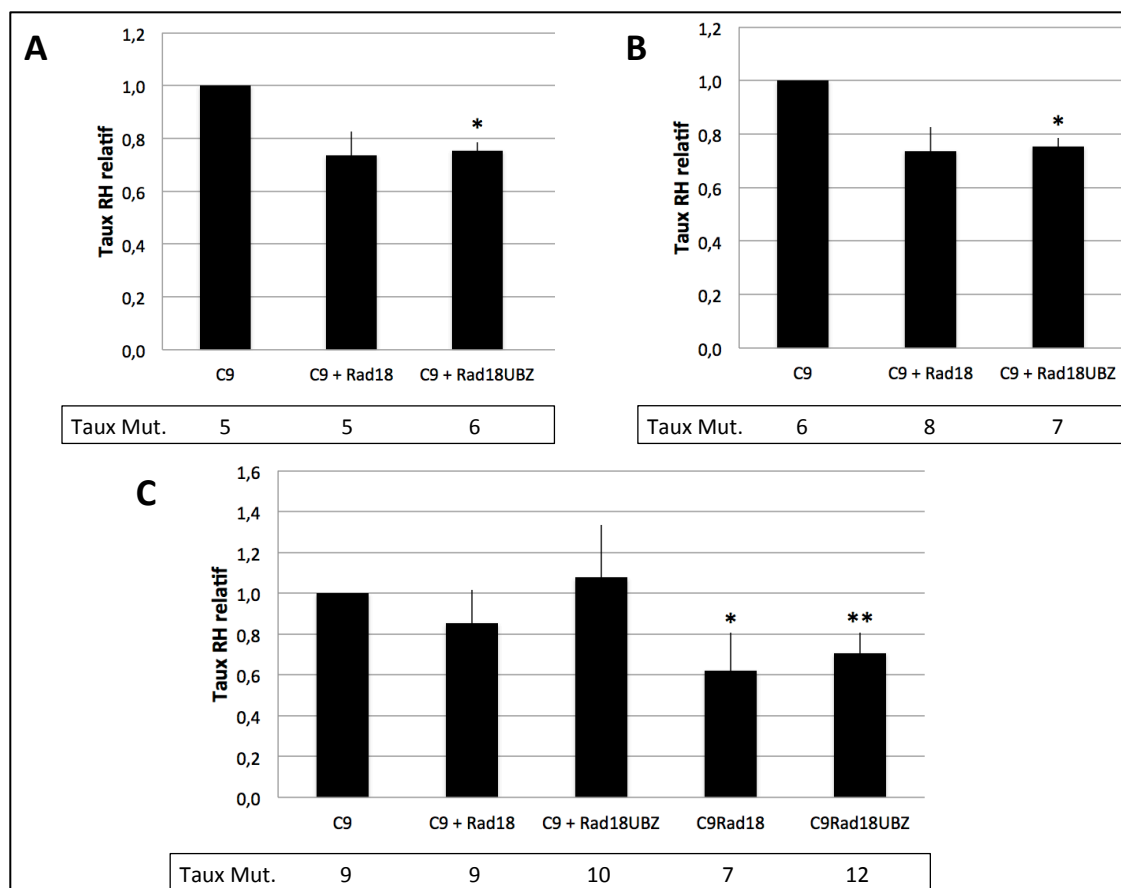


Figure 28. Effet de la surexpression ou du recrutement à la coupure de la protéine Rad18 sur le taux de mutations et le taux de KI.

(A) Effet de la surexpression de Rad18 sur le taux de mutations et de KI par homologie au site AAVS1 dans des U2OS. (B) Effet de la surexpression de rad18 sur le taux de mutations et de KI par homologie au site AAVS1 dans des RPE1. (C) Effet de la surexpression ou de recrutement à la coupure par fusion à la Cas9 de Rad18 sur le taux de mutations et le taux de KI par homologie dans des RG37DR. Afin de déterminer si les moyennes des taux de mutations obtenus avec les différentes nucléases sont différentes un t-test a été utilisé (n = 6).

Aucune augmentation du taux de KI n'a été observée avec l'utilisation de la protéine Rad18 et de son domaine UBD contrairement à ce qui était attendu suite aux résultats obtenus par Helchowski et al. (Helchowski et al., 2013).

Cependant, le système qu'ils ont utilisé, un rapporteur DR-GFP, permet de détecter une recombinaison intramoléculaire. Ici je détecte quant à moi une insertion d'un ADN exogène par recombinaison intermoléculaire. Cette différence de système pourrait expliquer les différences observées. Le fait que le système CRISPR/Cas9 produise des

extrémités franches contrairement à l'enzyme I-SceI pourrait aussi être une explication. En effet, la protéine Rad18 permet le recrutement de BRCA2 et de Rad51 qui nécessitent la présence d'ADN simple brin pour se fixer. Il pourrait être intéressant de tester la surexpression et le recrutement à la coupure de cette protéine en coupant l'ADN à l'aide de TALEN ou de la nickase Cas9(D10A) qui produisent elles des extrémités cohésives. Il serait aussi intéressant de tester cette approche à un autre locus, comme vu précédemment, l'effet pouvant être dépendant du locus et du type cellulaire.

3.1.3.2 Utilisation de la protéine Nbs1 pour augmenter le taux de Knock-In

Nbs1 est un autre candidat intéressant parce qu'il est impliqué dans l'initiation de la résection des extrémités des CDB (Komatsu, 2016). Cette protéine fait en effet partie du complexe MRN, qui, avec CtIP, est impliqué dans les premières étapes de la réparation par RH des CDB.

L'effet sur la RH de la surexpression et du recrutement à la CDB par fusion à la nucléase de Nbs1 a été testé dans des cellules RG37DR et U2OS. Le système CRISPR-Cas9 a alors été utilisé contre le site AAVS1 et le taux d'insertion homologue évalué par l'insertion de la GFP (Figure 29). En plus de la protéine Nbs1, son mutant RRHK (R28A/R43A/H45A/K160A) a été testé en parallèle. Ce dernier permet d'inhiber la fixation de Nbs1 à CtIP, de bloquer sa phosphorylation par ATM et donc, une fois surexprimé dans des cellules dont le Nbs1 endogène a été éteint, de bloquer la réparation par RH et MMEJ (Wang et al., 2013).

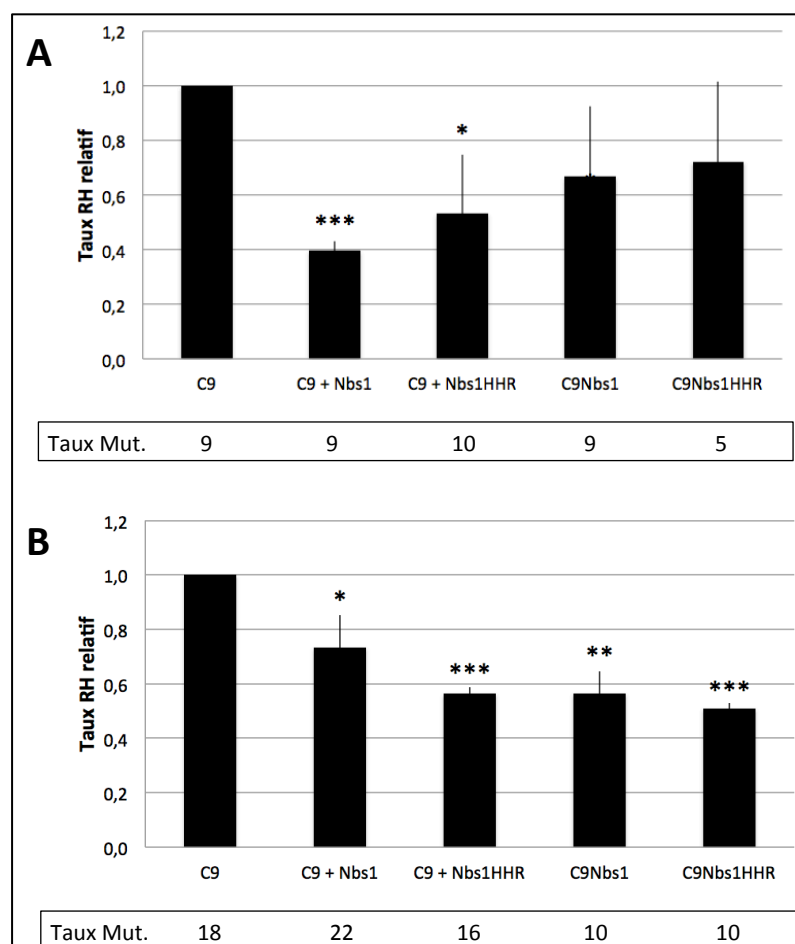


Figure 29. Effet de la surexpression ou du recrutement à la coupure de la protéine Nbs1.

(A) Effet de la surexpression ou de recrutement à la coupure par fusion à la Cas9 de Nbs1 sur le taux de mutations et le taux de KI par homologie dans des RG37DR. (B) Effet de la surexpression ou de recrutement à la coupure par fusion à la Cas9 de Nbs1 sur le taux de mutations et le taux de KI par homologie dans des U2OS. Afin de déterminer si les moyennes des taux de mutations obtenus avec les différentes nucléases sont différentes un t-test a été utilisé (n = 7).

La surexpression de Nbs1 et de son mutant RRHK ne permet pas d'augmenter le taux de KI après utilisation du système CRISPR/Cas9. On observe même une diminution significative du taux de KI.

Cette diminution est surprenante, l'utilisation de Nbs1 bloquerait alors la réparation par insertion par homologie. Pour la surexpression de Nbs1, l'explication pourrait être que Nbs1 séquestre les protéines impliquées dans la réparation par RH (CtIP, Nbs1 ou Mre11) empêchant la CDB d'être réparée par ce système. Cependant, si Nbs1 fixe ces protéines, son recrutement à la CDB devrait permettre leur recrutement aussi, et favoriser la réparation par RH, ce qui n'est pas observé. Peut être que la fusion à la nucléase Cas9 de Nbs1 bloque son activité et/ou ses sites de fixation aux protéines. Il serait intéressant de tester d'autres manières de fusionner Nbs1 à la nucléase, avec un bras de liaison plus long ou en Nter de la protéine afin de peut être libérer des domaines importants.

4 Annexe

4.1.1 Séquence du plasmide pSKduo contenant la répétition HD seulement

gtggcacttttcggggaaatgtgcgcggaacccctatattgtttatttttctaaatacattca
aatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaa
gagtatgagtattcaacattttccgtgtcgccttattcccttttttgcggcattttgccttc
ctgtttttgctcaccagaaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgca
cgagtgggttacatcgaactggatctcaacagcggtgaagatccttgagagttttcgccccga
agaacgttttccaatgatgagcacttttaaaagtctgctatgtggcgcggtattatcccgta
ttgacgcccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgag
tactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgc
tgccataaccatgagtataacactgcgcccaacttacttctgacaacgatcggaggaccga
aggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaa
ccggagctgaatgaagccataccaaacgacgagcggtgacaccacgatgcctgtagcaatggc
aacaacgttgcgcaaactattaactggcgaaactacttactctagcttcccggcaacaattaa
tagactggatggaggcgataaaagtgcaggaccacttctgcgctcggcccttccggctggc
tggtttattgctgataaatctggagccggtgagcggtgggtctcgcggtatcattgcagcact
ggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaacta
tggtgaacgaaatagacagatcgctgagataggtgcctcactgattaagcatttggttaactg
tcagaccaagtttactcatatatacttttagattgatttaaaacttcatttttaatttaaaag
gatctaggtgaagatcctttttgataatctcatgacaaaatcccttaacgtgagttttcgt
tccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctg
cgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcgggtgggttggttgccgga
tcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaata
ctgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctaca
tacctcgctctgctaatacctgttaccagtggctgctgccagtggcgataagtcgtgtcttac
cgggttgactcaagacgatagttaccggataaggcgacggtcgggctgaacgggggggtt
cgtgcacacagcccagcttgagcgaacgacctacaccgaactgagatacctacagcgtgag
ctatgagaaagcgccacgcttcccgaaggagaaaggcggacaggtatccggtaagcggcag
ggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctgggtatctttatagtc

ctgtcgggttttcgccacctctgacttgagcgtcgatTTTTgtgatgctcgtcagggggcg
agcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttt
tgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttg
agtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaa
gcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcag
ctggcacgacagggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagtt
agctcactcattagggcaccacaggctttacactttatgcttcgggctcgtatgttggtgga
attgtgagcgataacaattTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGC
AATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAA
ctagtCATGACggtggcaaacaggctcttgagaccgtccaacgccttctaccagttctctgt
caagcccacggactaaccacagcgcaagttgtagcgattgCTAGCAAGCTTATCGATACCGT
CGACCTCGagggggggcccggtaccgaattcgccctatagtgagtcgtattacgcgcgctca
ctggccgctcgttttacaacgtcgtgactgggaaaaccctggcgttaccgaacttaatcgccct
tgagcacatcccccttttcgccagctggcgtaatagcgaagaggcccgccacgatcgccctt
cccaacagttgcgcagcctgaatggcgaatgggacgcgcctgtagcggcgcatthaagcgcg
gcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcc
tttcgctttcttcccttctttctcgccacgttcgcccggctttccccgtcaagctctaaatc
gggggctcccttttagggttccgattttagtgtttacggcacctcgacccccaaaaaacttgat
taggggtgatgggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacggt
ggagtccacgttctttaatagtggactcttggtccaaactggaacaacactcaaccctatct
cggctctattcttttgatttataagggattttgccgatttcggcctattgggttaaaaaatgag
ctgatttaacaaaaatttaacgcgaatttttaacaaaatattaacgcttacaatttag

4.1.2 Banque des tétramères de répétitions TALE

Tétramère	N°	Tétramère	N°	Tétramère	N°
NCAT	148	TACA	197	TTCC	246
NCCA	149	TACC	198	TTCN	247
NCCC	150	TACN	199	TTCT	248
NCCN	151	TACT	200	TTNA	249
NCCT	152	TANA	201	TTNC	250
NCNA	153	TANC	202	TTNN	251
NCNC	154	TANN	203	TTNT	252
NCNN	155	TANT	204	TTTA	253
NCNT	156	TATA	205	TTTC	254
NCTA	157	TATC	206	TTTN	255
NCTC	158	TATN	207	TTTT	256
NCTN	159	TATT	208		
NCTT	160	TCAA	209		
NNAA	161	TCAC	210		
NNAC	162	TCAN	211		
NNAN	163	TCAT	212		
NNAT	164	TCCA	213		
NNCA	165	TCCC	214		
NNCC	166	TCCN	215		
NNCN	167	TCCT	216		
NNCT	168	TCNA	217		
NNNA	169	TCNC	218		
NNNC	170	TCNN	219		
NNNN	171	TCNT	220		
NNNT	172	TCTA	221		
NNTA	173	TCTC	222		
NNTC	174	TCTN	223		
NNTN	175	TCTT	224		
NNTT	176	TNAA	225		
NTAA	177	TNAC	226		
NTAC	178	TNAN	227		
NTAN	179	TNAT	228		
NTAT	180	TNCA	229		
NTCA	181	TNCC	230		
NTCC	182	TNCN	231		
NTCN	183	TNCT	232		
NTCT	184	TNNA	233		
NTNA	185	TNNC	234		
NTNC	186	TNNN	235		
NTNN	187	TNNT	236		
NTNT	188	TNTA	237		
NTTA	189	TNTC	238		
NTTC	190	TNTN	239		
NTTN	191	TNTT	240		
NTTT	192	TTAA	241		
TAAA	193	TTAC	242		
TAAC	194	TTAN	243		
TAAN	195	TTAT	244		
TAAT	196	TTCA	245		

4.1.3 Séquence du plasmide pCS2 contenant la demi-répétition NG

CGCCATTCTGCCTGGGGACGTCGGAGCAAGCTTGATTTAGGTGACACTATAGAATACAAGCT
ACTTGTTCCTTTTTGCAGGATCTGCCACCATGGCTCCAAAGAAGAAGCGTAAGGTAGACTACA
AAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGGT
ACCGTGGATCTACGCACGCTCGGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGAAGGT
TCGTTTCGACAGTGGCGCAGCACCACGAGGCACTGGTCGGCCATGGGTTTACACACGCGCACA
TCGTTGCGCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCAAGTATCAGGACATG
ATCGCAGCGTTGCCAGAGGCGACACACGAAGCGATCGTTGGCGTCGGCAAACAGTGGTCCGG
CGCACGCGCTCTGGAGGCCTTGCTCACGGTGCGGGAGAGTTGAGAGGTCCACCGTTACAGT
TGGACACAGGCCAACTTCTCAAGATTGCAAAACGTGGCGGCGTGACCGCAGTGAGAGGCAGTG
CATGCATGGCGCAATGCACTGACGGGTGCCCCCTGAACCTGACCCCGGAGCAGGTGGTGGC
CATCGCTAGCAATGGCGGCGGCAGGCCGGCGCTGGAGAGCATTGTTGCCCAGTTATCTCGCC
CTGATCCGGCGTTGGCCGCGTTGACCAACGACCACCTCGTCGCCTTGGCCTGCCTCGGCGGA
CGTCCTGCGCTGGATGCAGTGAAAAAGGGATTGCCGCACGCGCCGGCCTTGATCAAAGAAC
CAATCGCCGTATTCCTGAACGCACATCCCATCGCGTTGCCg gatcccAGCTGGTGAAGAGCG
AGCTGGAGGAGAAGAAGAGCGAGCTGAGACACAAGCTGAAGTACGTGCCCCACGAGTACATC
GAGCTGATCGAGATCGCCAGAAACAGCACCCAGGACAGAATCCTGGAGATGAAGGTGATGGA
GTTCTTCATGAAGGTGTACGGCTACAGAGGCAAGCACCTGGGCGGCAGCAGAAAGCCCCGACG
GCGCCATCTACACCGTGGGCAGCCCCATCGACTACGGCGTGATCGTGGACACCAAGGCCTAC
AGCGGCGGCTACAACCTGCCCATCGGCCAGGCCGACGAGATGCAGAGATACGTGGAGGAGAA
CCAGACCAGAAACAAGCACATCAACCCCAACGAGTGGTGAAGGTGTACCCACGAGCGTGA
CCGAGTTCAAGTTCCTGTTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACC
AGACTGAACCACATCACCAACTGCAACGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGG
CGGCGAGATGATCAAGGCCGGCACCCCTGACCCTGGAGGAGGTGAGAAGAAAGTTCAACAACG
GCGAGATCAACTTCAGAAGCTGATAACTGGAGtctagaACTATAGTGAGTCGTATTACGTAG
ATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAA
AAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAA
TAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGG
AGGTTTTTTAATTCGCGGCCGCGGCCCAATGCATTGGGCCCGGTACGTACCCAGCTTTTGT
TCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTG

AAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCT
GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTTTCCAG
TCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTT
GCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTTCGGCTGC
GGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC
GCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTC
AGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
GTGCGCTCTCCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGG
AAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTTCGTTTCGCT
CCAAGCTGGGCTGTGTGCACGAACCCCCGTTTACGCCCCGACCGCTGCGCCTTATCCGGTAAC
TATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAA
CAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACT
ACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA
AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGT
TTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA
CGGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCA
AAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTTAAATCAATCTAAAGTAT
ATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGA
TCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGG
GAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC
AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTTCGCCAGTT
AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGG
TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGT
GCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTG
TTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATG
CTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGA
GTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTG
CTCATCATTTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATC

CAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCG
TTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGG
AAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTG
TCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCGCGCA
CATTTCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTAAATTCGCGTTAA
ATTTTTGTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAA
TCAAAAGAATAGACCGAGATAGGGTTGAGTGTGTTCAGTTTGGAACAAGAGTCCACTATT
AAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTAC
GTGAACCATCACCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAAC
CCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGA
AGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCG
TAACCACCACCCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTGCGCCATTCAGG
CTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGTCGACCAT
AGCCAATTCAATATGGCGTATATGGACTCATGCCAATTCAATATGGTGGATCTGGACCTGTG
CCAATTCAATATGGCGTATATGGACTCGTGCCAATTCAATATGGTGGATCTGGACCCCAGCC
AATTCAATATGGCGGACTTGGCACCATGCCAATTCAATATGGCGGACTTGGCACTGTGCCAA
CTGGGGAGGGGTCTACTTGGCACGGTGCCAAGTTTGAGGAGGGGTCTTGCCCTGTGCCAAG
TCCGCCATATTGAATTGGCATGGTGCCAATAATGGCGGCCATATTGGCTATATGCCAGGATC
AATATATAGGCAATATCCAATATGGCCCTATGCCAATATGGCTATTGGCCAGGTTCATACT
ATGTATTGGCCCTATGCCATATAGTATTCATATATGGGTTTTCTTATTGACGTAGATAGCC
CCTCCCAATGGGCGGTCCCATATACCATATATGGGGCTTCCTAATACCGCCCATAGCCACTC
CCCCATTGACGTCAATGGTCTCTATATATGGTCTTTCCTATTGACGTCAATATGGGCGGTCTT
ATTGACGTATATGGCGCCTCCCCATTGACGTCAATTACGGTAAATGGCCCGCCTGGCTCAA
TGCCCATTGACGTCAATAGGACCACCCACCATTGACGTCAATGGGATGGCTCATTGCCCAT
CATATCCGTTCTCACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCACTTGGCAGTACA
TCAATATCTATTAATAGTAACTTGGCAAGTACATTACTATTGGAAGGACGCCAGGGTACATT
GGCAGTACTCCCATTGACGTCAATGGCGGTAAATGGCCCGCGATGGCTGCCAAGTACATCCC
CATTGACGTCAATGGGGAGGGGCAATGACGCAAATGGGCGTTCCATTGACGTAAATGGGCGG
TAGGCGTGCCTAATGGGAGGTCTATATAAGCAATGCTCGTTTAGGGAAC

4.1.4 Article : Edition du génome de rat par Knock-In

Efficient gene targeting by homology-directed repair in rat zygotes using TALE nucleases

S  verine Remy,^{1,2,7} Laurent Tesson,^{1,2,7} S  verine Menoret,^{1,2} Claire Usal,^{1,2} Anne De Cian,³ Virginie Thepenier,^{1,2} Reynald Thinard,^{1,2} Daniel Baron,¹ Marine Charpentier,³ Jean-Baptiste Renaud,³ Roland Buelow,⁴ Gregory J. Cost,⁵ Carine Giovannangeli,³ Alexandre Fraichard,⁶ Jean-Paul Concordet,³ and Ignacio Anegon^{1,2}

¹INSERM UMR 1064-ITUN, CHU de Nantes, Nantes F44093, France; ²Platform Rat Transgenesis, Nantes F44093, France; ³INSERM U565, CNRS UMR7196, Mus  um National d'Histoire Naturelle, F75005 Paris, France; ⁴Open Monoclonal Technologies, Palo Alto, California 94303, USA; ⁵Sangamo BioSciences, Richmond, California 94804, USA; ⁶genOway, Lyon F69007, France

The generation of genetically modified animals is important for both research and commercial purposes. The rat is an important model organism that until recently lacked efficient genetic engineering tools. Sequence-specific nucleases, such as ZFNs, TALE nucleases, and CRISPR/Cas9 have allowed the creation of rat knockout models. Genetic engineering by homology-directed repair (HDR) is utilized to create animals expressing transgenes in a controlled way and to introduce precise genetic modifications. We applied TALE nucleases and donor DNA microinjection into zygotes to generate HDR-modified rats with large new sequences introduced into three different loci with high efficiency (0.62%–5.13% of microinjected zygotes). Two of these loci (*Rosa26* and *Hprt1*) are known to allow robust and reproducible transgene expression and were targeted for integration of a GFP expression cassette driven by the CAG promoter. GFP-expressing embryos and four *Rosa26* GFP rat lines analyzed showed strong and widespread GFP expression in most cells of all analyzed tissues. The third targeted locus was *Ighm*, where we performed successful exon exchange of rat exon 2 for the human one. At all three loci we observed HDR only when using linear and not circular donor DNA. Mild hypothermic (30  C) culture of zygotes after microinjection increased HDR efficiency for some loci. Our study demonstrates that TALE nuclease and donor DNA microinjection into rat zygotes results in efficient and reproducible targeted donor integration by HDR. This allowed creation of genetically modified rats in a work-, cost-, and time-effective manner.

[Supplemental material is available for this article.]

Modifying the genome of animals is crucial for an in-depth understanding of physiological and pathophysiological processes as well as for commercial development. Mouse models have traditionally been used in genetics over the past decades, since researchers have in hand all the tools (in particular, embryonic stem [ES] cells) to perform homology-directed repair (HDR) in order to manipulate precisely the genome of this organism (Capecchi 2005). Unfortunately, until recently, HDR-mediated gene targeting was not possible in most other species, including the rat. The rat is a preferred species for studying physiology and certain human pathologies (Jacob 1999, 2010; Jacob and Kwitek 2002). Thus, precise gene targeting technologies would allow many important biomedical questions to be addressed. Despite the recent derivation of rat ES cells (Buehr et al. 2008; Li et al. 2008) to generate knockout (Tong et al. 2010; Yamamoto et al. 2012) and knock-in rats (Meek et al. 2010), rat ES cells are still less robust than mouse ES cells (Zheng et al. 2012). The emergence of engineered nucleases, allowing the rapid and effective generation of genetically modified animals, opens the door to gene targeting in rat but also in other species for which ES cells are not yet available, and provides a faster and more cost-effective strategy compared to the use of ES cells. We and others have shown that zinc finger nucleases (ZFNs) (Geurts et al. 2009; Mashimo et al. 2010), transcription activator-like

effector (TALE) nucleases (Tesson et al. 2011; Tong et al. 2012; Mashimo et al. 2013), meganucleases (Menoret et al. 2013), and the clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein (Cas) system (Li et al. 2013a,b) are highly and reproducibly effective at disrupting endogenous genes in rat zygotes. Mutations introduced are typically small insertions and/or deletions (indels) that result from imprecise repair of nuclease-induced DNA double-strand breaks by the nonhomologous end joining (NHEJ) mechanism. In addition, the presence of donor DNA allows either the insertion of exogenous genetic information or the replacement of an endogenous sequence by the one of interest. Targeting transgene addition or sequence replacement at loci in the mammalian genome by HDR has been demonstrated using ZFNs in mouse (Meyer et al. 2010; Cui et al. 2011; Meyer et al. 2012), rat (Cui et al. 2011), and rabbit (Flisikowska et al. 2011) zygotes. Gene editing in zygotes using TALE nucleases aimed to introduce point mutations (Wang et al. 2013; Wefers et al. 2013; Ponce de Leon et al. 2014) and, only very recently for the first time, an expression cassette in mice (Sommer et al. 2014).

Even if engineered nucleases stimulate HDR events, the frequency of these events remains rare. Finding methods which could increase the cellular activity of these artificial nucleases may en-

⁷These authors contributed equally to this work.

Corresponding author: Ignacio.Anegon@univ-nantes.fr

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.171538.113>.

   2014 Remy et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

hance rates of HDR. It has been reported that accumulation of nuclease proteins induced by a transient hypothermia enhanced the frequency of gene disruption induced both by ZFNs (Doyon et al. 2010) and by TALE nucleases (Miller et al. 2011) in cell lines, suggesting that this strategy could have beneficial effects on the frequency of HDR events in rat zygotes.

In our study, we applied TALE nucleases to achieve the insertion of a large expression cassette or exon exchange through gene targeting by HDR in rat zygotes. For that purpose, we microinjected TALE nuclease pairs designed to target three loci and the corresponding targeting vectors with regions homologous to the nuclease target site. Two targeted loci (*Rosa26* and *Hprt1*) are known to be permissive for transgene expression, and are therefore ideal candidates to integrate a transgene of interest intended to have stable and ubiquitous expression. Indeed, the *Rosa26* locus is commonly used in mice to achieve targeted insertions by HDR using ES cells and has been targeted using ZFNs in mice (Meyer et al. 2010; Hermann et al. 2012). The rat *Rosa26* locus was targeted using ES cells (Kobayashi et al. 2012) but not yet using nucleases. Similarly, the *Hprt* locus is often chosen as a “permissive” locus for targeted integration of transgenes in mouse ES cells (Bronson et al. 1996; Meek et al. 2010). These two loci were chosen to target insertion of a ubiquitous expression cassette encoding GFP, and to assess the potentially beneficial effect of a mild hypothermia treatment on TALE nuclease activity in HDR-mediated transgene integration. The third locus tested was the *Ighm* locus, for which we achieved efficient rat-to-human exon 2 replacement. The results presented demonstrate that TALE nucleases are efficient tools to insert exogenous genetic information or to replace an endogenous sequence with the one of interest.

Results

Activity of *Hprt1.1*, *Hprt1.2*, and *Rosa26* TALE nucleases in cultured rat cells

We designed pairs of TALE nucleases targeting the rat *Hprt1* gene at two different sites (which we refer to as *Hprt1.1* and *Hprt1.2*) within intron 1, or the *Rosa26* locus in intron 1, and examined their cleavage activity in cultured rat C6 cells transfected with plasmid DNA encoding individual TALE nucleases. The cleavage activity of the different TALE nuclease pairs was assessed using the T7 endonuclease assay (Fig. 1). We observed 9%, 15%, and 13% of chromosomes bearing nuclease-induced mutations in rat C6 cells transfected with $2 \times 0.75 \mu\text{g}$ *Hprt1.1*, *Hprt1.2*, or *Rosa26* TALE nuclease pairs, respectively. The frequency of cleavage events increased with higher concentrations ($2 \times 1.5 \mu\text{g}$) of each TALE nuclease pair: 19% for *Hprt1.1*, 21% for *Hprt1.2*, and 17% for *Rosa26*. These results show that all the tested nuclease pairs are active and allowed us to proceed with confidence with experiments in one-cell embryos to achieve targeted knock-in events.

Targeting integration into the rat *Hprt1* and *Rosa26* loci

In order to test whether TALE nucleases could stimulate targeted transgene integration by HDR, we chose to introduce a ubiquitous GFP expression cassette (CAG-GFP-BGHpA) of 3142 bp, flanked by 5' and 3' homology arms of 800 bp each contiguous to the TALE nuclease cleavage point (Figs. 2A, 3A).

TALE nuclease mRNA and linear donor DNA were co-injected into one-cell stage embryos at different concentrations for the TALE nucleases and at 2 ng/ μl for the linear donor DNA (excised ex-

pression cassette and homology arms). As previously described (Meyer et al. 2010), a two-step microinjection procedure was performed: The TALE nucleases mRNA and donor DNA mixture were first injected into the male pronucleus and then into the cytoplasm during withdrawal of the injection pipette. We cultured injected embryos at either 37°C for 3 h or 37°C for 1 h, followed by 30°C for 2 h or 30°C for 3 h to assess the effect of the temperature on the rate of NHEJ and HDR events. Following this period of culture, surviving zygotes were then transferred into recipient females and genotyping was performed on E15 fetuses or on newborns (Table 1).

When zygotes were injected with the three ratios of *Hprt1.1* mRNA/DNA tested, we did not observe significant differences on egg viability (73.5%, 78.5%, and 72.6%), while a mild decrease in their embryonic development was observed at the highest concentration (13.3%, 10.7%, and 8.2%). Concerning the *Hprt1.2* TALE nucleases, increasing concentrations did not significantly affect zygote viability (85.8%, 72.1%, and 78.3%) but did impair birth rate at the highest concentration (26.2% at 5 + 5/2, 20.3% at 10 + 10/2, and 0% at 20 + 20/2 ng/ μl). Finally, for *Rosa26* TALE nucleases, we observed increased toxicity as assayed by zygote viability at higher concentrations (80.8%, 76.7%, and 52.2%); all nuclease concentrations used resulted in similarly lower embryonic development (9.2%, 8.8%, and 10.6%).

The incubation of embryos at 37°C + 30°C or 30°C after microinjection of *Hprt1.2* or *Rosa26* TALE nuclease mRNA did not significantly affect either embryo viability or newborn frequencies.

To identify GFP-expressing animals either due to random integration (RI) and/or HDR, E15 fetuses and pups were first exposed to UV light to assay for GFP expression (Table 1). The positive animals were then confirmed by PCR using primers specific for GFP (data not shown). The strategies used to analyze the donor DNA integration into *Hprt1* and *Rosa26* targeted loci are illustrated in Supplemental Figure S1 and Figures 2A and 3A. The animals harboring HDR events were identified by PCR amplification of the 5' and 3' junctions between the donor sequence and the target locus (Figs. 2B, 3B; Supplemental Fig. S1B; data not shown). The integrity of the 5' and 3' junctions was confirmed by sequencing (Figs. 2D, 3C; data not shown).

No GFP-expressing animals or animals with donor DNA integration were observed when high concentrations (20 + 20 or 50 + 50 ng/ μl) of mRNA encoding *Hprt1.1*, *Hprt1.2*, or *Rosa26* TALE nucleases were injected, probably at least in part due to toxicity (Table 1). With *Hprt1.1*, *Hprt1.2*, or *Rosa26* TALE nucleases at a ratio of 10 + 10/2 ng/ μl mRNA/DNA (37°C condition), the frequency of DNA integration reached 1.02%, 0.63%, and 3.30% of transferred embryos, respectively (Table 1). In these conditions, junction PCR analyses revealed that one *Hprt1.1* (13.1) (Table 2), one *Hprt1.2* (6.1) (Table 2), and two *Rosa26* (2.1 and 3.1) (Fig. 3B) founders harbored an HDR profile, whereas one *Rosa26* was a RI (5.3) (Table 2). Lower concentrations (5 + 5/2 ng/ μl) of *Hprt1.2* and *Rosa26* TALE nucleases were also very efficient at inducing donor DNA integration (3.17% and 1.67% of transferred embryos GFP+, respectively) by HDR (three out of four founders for the *Hprt1.2* locus, e.g., 9.6, 11.9, and 12.2, and two founders for the *Rosa26* locus, e.g., 10.2 and 12.5) (Tables 1, 2).

Interestingly, using *Hprt1.2* TALE nucleases, the HDR frequency of embryos cultured at 37°C for 3 h (0.63 and 2.38% at the dose of 10 + 10/2 and 5 + 5/2, respectively) was increased when embryos were maintained either at 37°C for 1 h + 30°C for 2 h (2.73 and 2.7 at the dose of 10 + 10/2 and 5 + 5/2, respectively) or 30°C for 3 h (2.54 and 5.13 at the dose of 10 + 10/2 and 5 + 5/2, re-

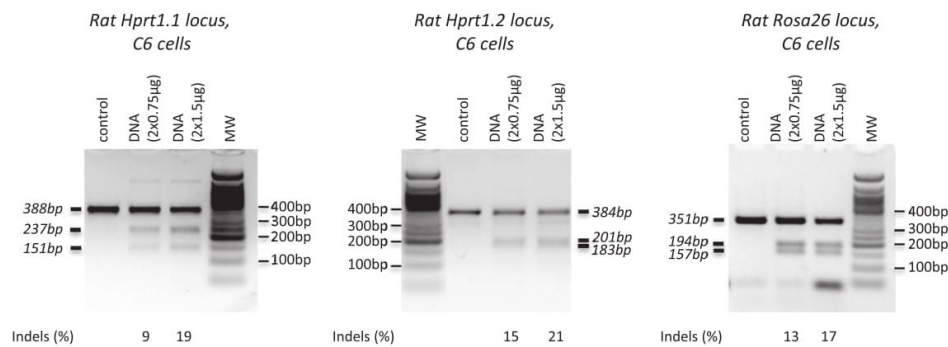


Figure 1. Assay of TALE nucleases for rat *Hprt1.1*, *Hprt1.2*, and *Rosa26* loci. The frequency of TALE nuclease cleavage was determined using T7 endonuclease assay in C6 cells transfected with the indicated amount of rat *Hprt1.1*, *Hprt1.2*, or *Rosa26* TALE nuclease expression vectors. Expression vectors without nucleases serve as negative controls. The expected sizes of undigested and digested fragments are indicated in italics. The efficiency of cleavage is indicated below each gel.

spectively) (Table 1). Transient incubation at a lower temperature also tended to increase the frequency of HDR events in the *Rosa26* locus: 2.78 at 37°C for 1 h + 30°C for 2 h vs. 1.67% at 37°C but 0% for 30°C incubation (Table 1). The analyses of indel events due to NHEJ at *Hprt1.2* and *Rosa26* loci showed that their frequency did not vary significantly among the three temperature conditions (Table 1).

RI of the GFP expression cassette (scored as GFP expression and positive PCRs for *GFP* but with negative 5' and 3' junction PCRs) occurred at frequencies similar to those observed when linearized DNA is microinjected for the generation of transgenic rats (0.67%–2.54%) (Tesson et al. 2005) and without differences among the different experimental conditions. These RI transgenic animals did not contain HDR events, and for *Hprt1.2* TALE nucleases included animals 13.2, 3.2, 5.7, 6.4, 6.5, 7.5, 7.8, 5.1, 9.5, and 10.8 as well as for *Rosa26*, animals 5.3, 7.1, and 19.3 (Table 2).

To confirm targeted integration of donor DNA by HDR further, we performed Southern blot analyses. All of the 20 animals with the *GFP* transgene integrated into the *Hprt1.2* locus as identified using junction PCR analyses (Figs. 2B,D; Table 2; data not shown), harbored a band of 7.6 kb, consistent with the expected integration mediated by HDR (Fig. 2C; data not shown). The profile on the Southern analysis of some of the animals (animals 2.2 and 5.9 in Fig. 2C) showed a single copy of the expression cassette, whereas that of others (animals 1.2 and 3.4 in Fig. 2C) showed the presence of a band of 4.7 kb compatible with the presence of more than one copy in concatemers. None of these embryos showed additional bands, indicating that they did not harbor RI copies of the expression cassette. PCR analyses were performed to determine the configuration of these concatemers. For the embryos injected with *Hprt1.2* TALE nucleases, all the RI and some of the HDR animals bore integrated concatemers (Table 2) in a head-to-tail configuration (Supplemental Fig. S2B) but not in head-to-head or tail-to-tail (data not shown).

At the *Rosa26* locus, the seven animals for which amplified fragments of expected size and sequence were observed both in 5' and 3' junction PCR analyses (Figs. 3B,C; Table 2; data not shown) showed bands of the expected size (9.2 kb and 2.2 kb), consistent with integration at the target locus by HDR (data not shown). Three of these rats had integration of donor concatemers (Supplemental Fig. S2B) in a head-to-tail configuration but not in head-to-head or tail-to-tail (data not shown).

TALE nuclease-induced DNA cleavage was required for efficient targeted donor DNA integration since the microinjection of 2 ng/µl linear donor DNA alone did not result in any HDR positive animals (Table 1), in accordance with a previous publication that showed a very low level of efficacy (<0.1%) for spontaneous HDR in mouse zygotes (Brinster et al. 1989).

We compared the excised linear form of donor DNA vs. the supercoiled form of the plasmid for transgene integration by HDR. As reported in Table 1, of the 25 embryos issued from microinjection of *Hprt1.2* TALE nucleases and circular donor DNA, none integrated donor DNA by HDR, and one of them (circ-4.1) (Table 2) integrated the donor DNA by RI. In comparison, in the same conditions (5 + 5/2 and 30°C incubation), when donor DNA was injected in a linear form we obtained eight GFP+ animals (six by HDR, e.g., 1.2, 2.2, 5.9, 9.7, 10.2, and 10.4, and two by RI, e.g., 6.4 and 6.5) (Tables 1, 2).

In conclusion, targeted transgene integration could be efficiently achieved in all *Hprt1.1*, *Hprt1.2*, and *Rosa26* TALE nuclease injections when the concentration of the TALE nuclease was below 10 ng/µl. The frequency of mutated animals with indel mutations was higher than that of targeted integration by HDR, consistent with NHEJ being the prevalent mechanism for DNA double-strand break repair. In all founders harboring an HDR-mediated transgene insertion except one (*Hprt1.2* 1.1) (Table 2), we found a conserved integrated donor DNA and intact 5' and 3' junctions. Transient diminution of temperature and excised linear vs. circular donor DNA tended to increase the frequency of HDR and thus represent alternative methods for gene targeting HDR in rat zygotes that might be superior for certain loci and nuclease pairs.

Efficient germline transmission and expression of donor sequences targeted into the *Rosa26* locus

Five different *Rosa26* founders with HDR of the CAG-*GFP* expression cassette showed germline transmission of the donor sequence with frequencies between 14% and 46% (Supplemental Table S1). Figure 3D illustrates two 8-d-old GFP+ animals in the offspring of the animal *Rosa26* KI 8.4. Moreover, offspring of each of four *Rosa26* HDR+ founders analyzed presented a similar percentage and mean fluorescence intensity of CD45+ GFP+ cells (Fig. 3E). This expression was comparable in animals with one copy (8.4F1, 9.1F1, and 10.2 F1) or with concatemers (12.5F1) (Fig. 3E). Finally,

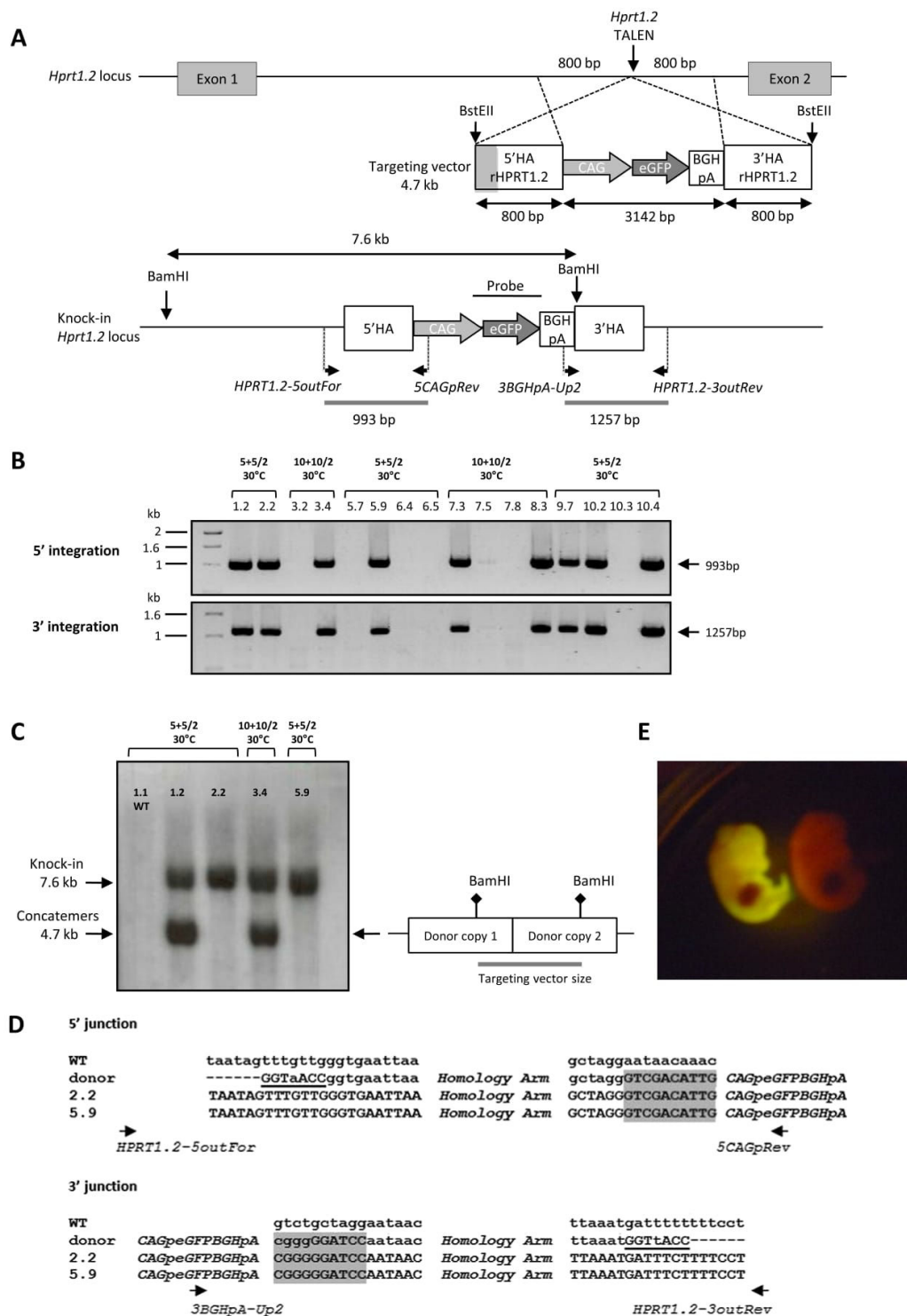


Figure 2. (Legend on next page)

expression of GFP was uniformly detected in most cells of all tissues analyzed, as shown for liver, kidney, and pancreas (Fig. 3D; data not shown). Expression of GFP in embryos generated with *Hprt1.1* and *Hprt1.2* TALE nucleases also showed strong and uniform expression upon in toto analysis (Fig. 2E; data not shown). These data provide evidence that integration of the transgene into the *Rosa26* locus took place early in development, with low mosaicism and high frequency of transmission to the offspring. Additionally, the integrated transgene into the targeted *Rosa26* locus showed high and uniform expression in different lines of transgenic rats, indicating that integration of transgenes into this locus using TALE nucleases is a robust strategy to express transgenes.

Targeted replacement of a rat exon with a human exon at the *Ighm* locus

We aimed to replace an endogenous sequence with an exogenous one since this would allow a wide range of genetic editing applications. As a proof-of-concept model, we applied TALE nucleases targeting the rat *Ighm* locus used in a previous work (Tesson et al. 2011) to obtain an exon exchange directly in rat zygotes. In the previous study, we showed the cleavage activity of this TALE nuclease pair to be 13% in rat S16 cells in vitro and 58% in injected rat zygotes, when delivered as mRNA.

We microinjected sequentially into the pronuclei and cytoplasm of rat zygotes TALE nuclease mRNA directed to sequences in exon 2 of the rat *Ighm* locus, together with donor DNA sequences containing the exon 2 of human *IGHM* (71% homology vs. rat) and homology arms (0.75 and 1.46 kb for 5' and 3' arms, respectively) of rat genomic sequence (Fig. 4A). The homology arms were separated from the TALE nucleases cleavage site by ~150 bp. With this transgene donor, successful exon replacement requires spontaneous cellular trimming of at least one of the 3' single-stranded ends so that the 3' end terminates in a region of homology with the donor. The microinjection statistics are summarized in Table 3. Microinjection with the supercoiled donor plasmid (410 zygotes) or linearized donor sequences (1063 zygotes) resulted in both cases in normal embryo survival (75.1% and 72.7% of microinjected embryos) and normal numbers of newborn animals (22.9% and 15.8% of transferred embryos). Microinjection of TALE nucleases plus the supercoiled plasmid donor DNA did not result in any donor integration by HDR, but two transgenic animals carrying RI could be detected (Tables 3, 4). Microinjection of TALE nucleases plus the linear donor DNA resulted in eight animals (Table 3) with

RI of linear donor DNA (8, 44, 49, 55, 56, 63, 68, and 69) (Table 4) and in one embryo (E3) and two founders (S3 and S4) with bona fide targeted integration by HDR (0.62% of the transferred zygotes) since they showed 5' and 3' flanking PCRs of the expected size and with intact sequences (Figs. 4B,C) as well as the expected 3.2-kb band on the Southern blot (Fig. 4D). Southern blot analysis showed the presence of donor concatemers at the integration site in one founder (S3), whereas the other founder (S4) had one copy (Fig. 4D) and embryo E3 showed concatemers (data not shown). PCR analysis of concatemers showed that they were orientated in a head-to-tail orientation (Supplemental Fig. S2D).

Founder S4 was mated and transmitted the transgene to six out of 17 pups (Supplemental Table S1). mRNA products containing rat and human sequences were detected in the offspring by RT-PCRs of the expected size (Fig. 4E; data not shown) and contained intact rat and human sequences (data not shown). These F1 heterozygous animals were mated to obtain homozygous F2 animals which did not show the presence of wild-type rat CH2 sequences, having both alleles replaced with human sequences (data not shown).

Thus, injection of TALE nucleases mRNAs and linear donor DNA with human sequences resulted in the replacement of rat by human *IGHM* sequences by HDR.

Discussion

ZFNs were used to obtain transgene targeted integration in previous studies in mice (Meyer et al. 2010, 2012; Cui et al. 2011), rats (Cui et al. 2011), and rabbits (Flisikowska et al. 2011). Compared to ZFNs, TALE nucleases can be designed with somewhat fewer positional constraints (Miller et al. 2011), allowing targeting of virtually any sequences, with high activities and very low off-target effects (for review, see Segal and Meckler 2013). TALE nucleases were applied to achieve targeted gene addition of point mutations in rodents (Panda et al. 2013; Wang et al. 2013; Wefers et al. 2013; Ponce de Leon et al. 2014) and (after the submission of this manuscript) in mice using an expression cassette (Sommer et al. 2014).

Here, we used newly generated TALE nucleases for the *Hprt1* and *Rosa26* loci and the previously described TALE nucleases for *Ighm* (Tesson et al. 2011). For the *Hprt1* and *Rosa26* TALE nucleases, we observed that the correlations between in vitro and in vivo efficiency activities are such that *Hprt1.2* > *Rosa26* > *Hprt1.1*. Direct comparison of in vitro efficiencies between *Hprt1* and *Rosa26* TALE

Figure 2. Targeted integration of a GFP cassette into the *Hprt1.2* locus. (A) (Upper) Diagram showing schematic representation of the rat *Hprt1.2* locus, with the site of TALE nuclease action (vertical arrows), and of the targeting vector with the expression cassette CAG-eGFP-BGHpA (3142 bp) and the 5' and 3' homology arms (800 bp each). The homology arms are contiguous to the TALE nucleases' cleavage point. (Gray) The sequence overlap (433 bp) between 3' HARHPRT1.1 (cf. Supplemental Fig. S1) and 5' HARHPRT1.2. BstEII restriction sites are indicated. (A) (Lower) Diagram showing schematic representation of the GFP cassette integration. For flanking PCR analysis, genomic DNA was PCR-amplified with primers situated for the 5' side: upstream of the 5' HA arm (HPRT1.2-5outFor) and in the CAG promoter (SCAGpRev); and for the 3' side: in the BGHpA (3BGHpa-Up2) and downstream from the 3' HA arm (HPRT1.2-3outRev). The position of each primer and the corresponding expected size of PCR products are indicated on the schematic knock-in *Hprt1.2* locus. For Southern blot analysis, genomic DNA was digested with BamHI and was probed with a GFP probe. A unique band at 7.6 kb is predicted for a correct HDR into the *Hprt1.2* locus. (B) Flanking PCR analysis. Gels show the results of analyzing the 5' and the 3' extremities of GFP integration into the *Hprt1.2* locus. A representative panel of 16 animals is illustrated, showing the expected bands of 993 bp using the 5' pair of primers (HPRT1.2-5outFor + SCAGpRev) and of 1257 bp using the 3' pair of primers (3BGHpa-Up2 + HPRT1.2-3outRev). The microinjection conditions in terms of mRNA and DNA concentrations, as well as embryo incubation temperatures, are above each animal. (C) Southern blot analysis following BamHI DNA digestion for the analysis of transgene site-specific integration into the *Hprt1.2* locus using a GFP probe. Four *Hprt1.2*-GFP+ rats with positive junction PCRs showing a unique band at 7.6 kb for a HDR with a single copy of the transgene, or 7.6 kb and 4.7 kb (the size of the transgene) bands when HDR involved transgene concatemers. The diagram at the right explains the expected size of the concatemers once linearized. The absence of additional bands demonstrates that there are no RI integration events in these animals. As a negative control, an offspring rat with no GFP expression and negative for GFP PCR (1.1) showed no bands. (D) Sequence comparison at the 5' and 3' junctions, with wild-type genomic DNA and donor DNA sequences, of two representative embryos (2.2 and 5.9). In donor DNA, the presence of the BstEII site is indicated in 5' and 3' (underlined). The 5' and 3' ends of the expression cassette are colored in gray. (E) Representative E15 *Hprt1.2* HDR-GFP embryos using a Dark Reader Spot Lamp.

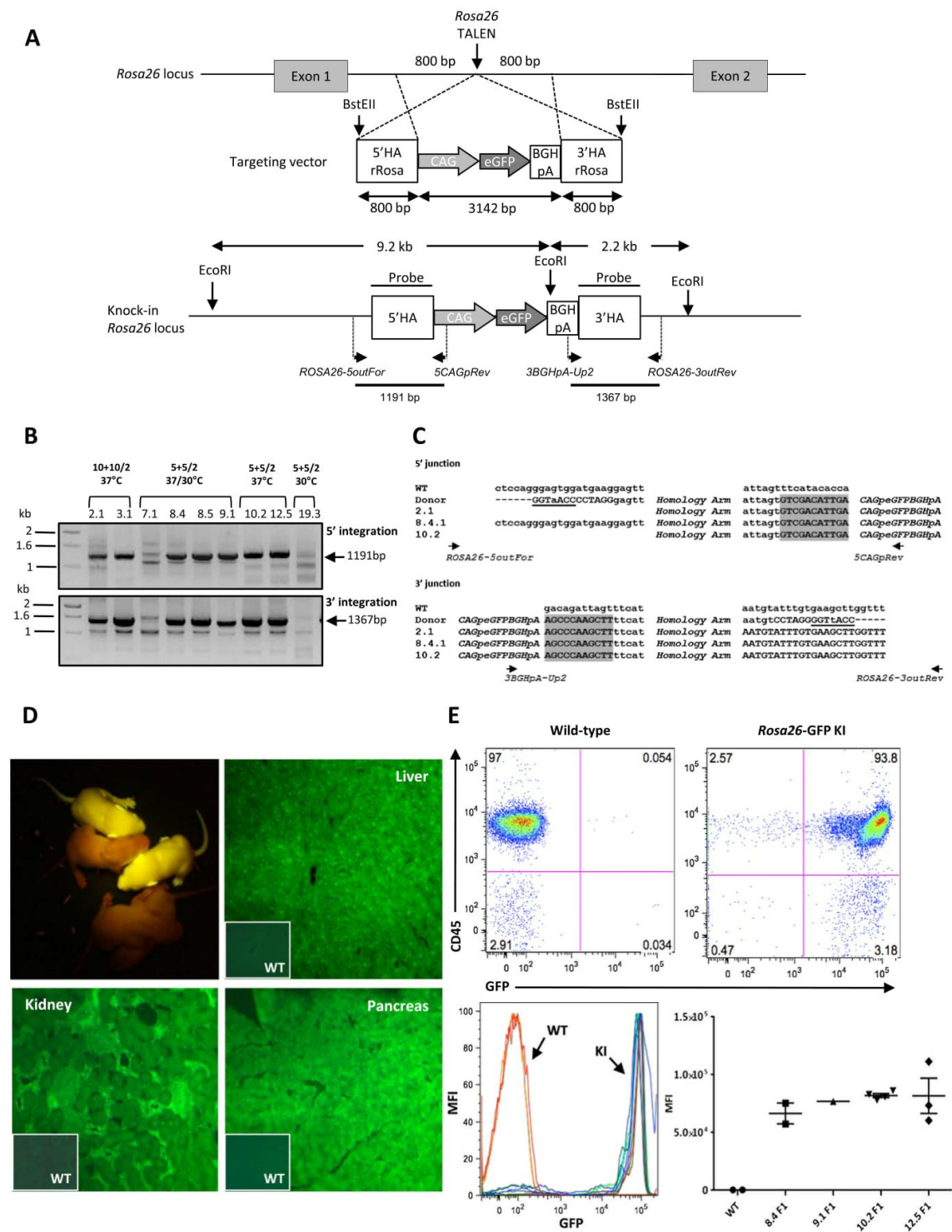


Figure 3. (Legend on next page)

nucleases with those of *Ighm* is difficult since they were transfected in different cell lines.

The efficiency of TALE nuclease-mediated HDR seemed to depend primarily on the efficiency of DNA cleavage by the nuclease and not on the type of locus targeted in terms of expression and chromatin status, since *Hprt1* and *Rosa26* are expressed in the injected zygote, whereas *Ighm* is not. This conclusion is also substantiated by experiments that we performed using a ZFN for the *Ighm* locus with lower nuclease activity (2% in C6 cells and 1.5% of NHEJ following mRNA microinjection) for the exchange of rat by human exon 2 in the *Ighm* locus (same donor DNA as used for the TALE nucleases) that did not result in HDR events (data not shown).

The frequencies of HDR obtained in the present study were between 0.62% and 5.13% of microinjected embryos for the best conditions. Despite differences in the models that make direct comparisons only relative, these frequencies are comparable to the ones for targeted insertion using TALE nucleases and ssODN donors ranging from 1.8% to 6.8% in mouse embryos (Panda et al. 2013; Wefers et al. 2013) but higher than the ones obtained with an expression cassette (2/900, 0.22%) (Sommer et al. 2014). The frequencies obtained in our study are also comparable to the ones observed using ZFN mRNA for HDR in rats (2.4% and 8.3% for two different loci) (Cui et al. 2011) and mice (1.7% and 5% for two different loci) (Meyer et al. 2010; Cui et al. 2011). In this report, as well as in others using ZFNs, TALE nucleases, or CRISPR/Cas (Meyer et al. 2010; Wefers et al. 2013; Auer et al. 2014; Sommer et al. 2014), the NHEJ repair mechanisms occurred always and for each condition at a higher rate than HDR.

A transient cold shock applied to cells lines enhances the cleavage activity of ZFNs (Doyon et al. 2010) and TALE nucleases (Miller et al. 2011), due, at least in part, to an increase in accumulation of nuclease proteins (Doyon et al. 2010), but the effect on HDR was not analyzed. The assessment of this parameter allowed us to show a positive effect of a short or longer incubation at 30°C of injected embryos on the frequency of HDR events but not in the rate of NHEJ, suggesting a preferential repair by the former mechanism. This effect was more accentuated for the *Hprt1.2* locus than for *Rosa26*. Moreover, no or slight effects were observed on viability and embryonic development of rat embryos. Although promising, additional experiments at other loci and in other species are necessary to confirm the generality of these observations.

In HDR strategies using ZFNs or TALE nucleases, donor DNA sequences can be linearized or supercoiled plasmids. In our four HDR insertions (*Hprt1.1*, *Hprt1.2*, *Rosa26*, and *Ighm*), HDR events were observed when donor DNA was delivered under a linear form, and when directly compared in the *Hprt1.2* and *Ighm* loci, the circular form did not generate HDR events. In other studies using linear or circular DNA donors, the outcomes have been very variable. In the only previous study of HDR in rats using nucleases (ZFNs), both circular and linear forms generated HDR events (Cui et al. 2011; Brown et al. 2013). Using ZFNs in mice, two publications showed that both linear and circular DNA generated HDR (Meyer et al. 2010; Meyer et al. 2012), whereas another publication showed that only the linear, but not the circular form, could result in HDR (Hermann et al. 2012). In *Drosophila*, ZFNs or TALE nucleases induced HDR only using supercoiled plasmid donors (Beumer et al. 2008; Katsuyama et al. 2013). In rabbits, the linear form allowed obtention of HDR when using ZFNs (Flisikowska et al. 2011).

Compared to circular DNA, linear DNA has the potential disadvantage of being more prone to RI into the genome (Brinster et al. 1985). Our results show RI of linearized donor DNA (both the GFP expression cassette or human *IGHM* sequences) with frequencies of 0.62% to 1.71% of transferred embryos, close to the frequencies observed when linear DNA is microinjected to generate transgenic rats (Tesson et al. 2005). RI of circular donor DNA was also observed with frequencies of 0.95% and 0.72% for embryos transferred after *Hprt1.2* or *Ighm* TALE nuclease injection, respectively. Therefore, in terms of transgene RI, there is not a significant disadvantage to the use of linear vs. circular donor DNA. Undesirable RI events could be separated from the desired modified allele by breeding (if unlinked to the target locus).

We observed that several of the *Hprt1.1*, *Hprt1.2*, *Rosa26*, and *Ighm* HDR-positive embryos or animals had concatemers. While few studies have analyzed the mechanisms of transgene concatemer formation, one recent one identified NHEJ as the main mechanism of concatamer formation (Dai et al. 2010). It is unclear why concatemers were observed at higher frequencies in the *Hprt1* locus versus the *Rosa26* or *Ighm* loci. Although transgenic animals with RI in concatemers are subject to gene silencing (Garrick et al. 1998), expression of GFP was equivalent in our rats harboring a profile with and without concatemers in the *Rosa26* locus. This similar expression might be due to the insertion in a permissive locus versus RI in sites more prone to methylation. At the same

Figure 3. Targeted integration of a GFP cassette into the *Rosa26* locus. (A) (Upper) Diagram showing schematic representation of the rat *Rosa26* locus, with the site of TALE nuclease action (vertical arrows) and of the targeting vector with the expression cassette (3142 bp) and the 5' and 3' homology arms (800 bp each). The homology arms are contiguous to the TALE nucleases' cleavage point. BstEII restriction sites are indicated. (A) (Lower) Diagram showing schematic representation of the GFP cassette integration. For PCR in/out analysis, genomic DNAs were PCR-amplified with primers situated for the 5' side: upstream of the 5' HA arm (ROSA26-SoutFor) and in the CAG promoter (5CAGpRev); and for the 3' side: in the BGHpA (3BGHpA-Up2) and downstream from the 3' HA arm (ROSA26-3outRev). The position of each primer and the corresponding expected size of PCR products are indicated on the schematic knock-in *Rosa26* locus. For Southern blot analysis, genomic DNA was digested with EcoRI and was probed with the homology arms' probe for *Rosa26*. Two bands at 9.2 kb and 2.2 kb are predicted for a correct HDR into the *Rosa26* locus. (B) Flanking PCR analysis. Gels show the results analyzing the 5' and the 3' extremities of the expression cassette integration into the *Rosa26* locus. A representative panel of nine founders is illustrated, showing expected bands of 1191 bp using the first pair of primers (ROSA26-SoutFor + 5CAGpRev), and of 1367 bp using the second pair of primers (3BGHpA-Up2 + ROSA26-3outRev). The microinjection conditions in terms of mRNA and DNA concentrations as well as embryo incubation temperatures are above each animal. (C) Sequence comparison at the 5' and 3' junctions with wild-type genomic DNA and donor DNA sequences of two representative founders (2.1 and 10.2) and one representative F1 (8.4.1: offspring of founder 8.4). In donor DNA, the presence of the BstEII site is indicated (underlined). The start and the end of the expression cassette integration into the *Rosa26* locus are colored in gray. (D) Two representative 8-d-old *Rosa26* HDR rat pups (F1s of founder 8.4) and two wild-type littermates. GFP expression in adult tissues (liver, kidney, pancreas) of a *Rosa26* HDR rat. Insets show tissues obtained from littermates negative for GFP PCR (wild type, WT). Magnification: $\times 100$. (E) GFP expression in leukocytes from different lines of *Rosa26* HDR rats. FACS analysis of GFP expression in CD45+ leukocytes isolated from peripheral blood from four different lines of *Rosa26* HDR adult rats (F1 generation) and a wild-type littermate. (Upper) FACS patterns obtained from two F1 (an HDR GFP+ and a negative littermate) of founder 8.4. (Lower) Analysis of the level of GFP expression in the peripheral blood of offspring of four *Rosa26* HDR founders using the mean fluorescence intensity of leukocytes. Each point represents one animal and the horizontal bars the mean and standard deviations.

Table 1. Microinjection statistics for the *Hprt1.1*, *Hprt1.2*, and *Rosa26* loci

Target locus	Dose mRNA/DNA (ng/ μ l)	Temperature	No. injected eggs (% viable eggs)	No. E15 (e) or pups (p) (%) ^a	No. of GFP+ animals (%) ^a	No. of RI-positive animals (%) ^a	No. of HDR-positive animals (%) ^a	No. of indel-positive animals (%) ^a
<i>Hprt1.1</i>	50 + 50/2	37°C	113 (73.5)	8e (13.3)	0 (0)	0 (0)	0 (0)	1 (1.67)
	20 + 20/2	37°C	209 (78.5)	17e (10.7)	0 (0)	0 (0)	0 (0)	2 (1.26)
	10 + 10/2	37°C	164 (72.6)	8e (8.2)	1 (1.02)	0 (0)	1 (1.02)	2 (2.04)
<i>Hprt1.2</i>	20 + 20/2	37°C	69 (78.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	10 + 10/2	37°C	258 (72.1)	32e (20.3)	1 (0.63)	0 (0)	1 (0.63)	15 (9.49)
	10 + 10/2	37°C/30°C	151 (72.8)	19e (17.3)	4 (3.64)	1 (0.91)	3 (2.73)	10 (9.09)
	10 + 10/2	30°C	163 (82.8)	22e (18.6)	6 (5.08)	3 (2.54)	3 (2.54)	6 (5.08)
	5 + 5/2	37°C	148 (85.8)	33e (26.2)	4 (3.17)	1 (0.79)	3 (2.38)	17 (13.5)
	5 + 5/2	37°C/30°C	158 (93.7)	42e (28.4)	6 (4.05)	2 (1.35)	4 (2.70)	24 (16.2)
	5 + 5/2	30°C	173 (79.2)	35e (29.9)	8 (6.84)	2 (1.71)	6 (5.13)	14 (12)
	5 + 5/2 ^b	30°C	159 (71.1)	25e (23.8)	1 (0.95)	1 (0.95)	0 (0)	4 (3.81)
	0/2	30°C	110 (80)	38e (43.2)	0 (0)	0 (0)	0 (0)	na
	20 + 20/2	37°C	134 (52.2)	5e (10.6)	0 (0)	0 (0)	0 (0)	1 (2.13)
<i>Rosa26</i>	10 + 10/2	37°C	180 (76.7)	8e (8.8)	3 (3.30)	1 (1.10)	2 (2.20)	3 (3.30)
	5 + 5/2	37°C	161 (80.8)	11p (9.2)	2 (1.67)	0 (0)	2 (1.67)	5 (4.17)
	5 + 5/2	37°C/30°C	158 (68.4)	21p (19.4)	4 (3.70)	1 (0.92)	3 (2.78)	8 (7.41)
	5 + 5/2	30°C	197 (75.6)	47p (31.5)	1 (0.67)	1 (0.67)	0 (0)	6 (4.03)

Rat *Hprt1.1*, *Hprt1.2*, or *Rosa26* TALE nucleases, as mRNA, were injected at different concentrations each in combination with 2 ng/ μ l of donor DNA, both into the cytoplasm and into the male pronucleus. Donor DNA was injected either in a linear form with each of the three TALE nuclease pairs or in a circular form only with rat *Hprt1.2* TALE nucleases (see footnote b, below). Injected eggs were maintained under 5% CO₂ at 37°C for 3 h, 37°C for 1 h, followed by 30°C for 2 h, or 30°C for 3 h until reimplantation. Viability was evaluated after the culture period. Potential toxicity was also assessed by the number of day 15 embryos (E15) or of live pups obtained following the transfer of injected eggs. Percentage of the total transferred is indicated in parentheses. The numbers of E15 or live pups which have integrated the donor DNA sequence either by random integration (RI) or HDR integration (PCR positive both at the 5' and 3' ends) or which have indel mutations (T7 nuclease assay and sequencing) with no HDR are reported in the last four columns. (na) Not applicable.

^aPercentages indicated in parentheses correspond to the percentage of transferred embryos.

^bDonor DNA was injected in a circular form in this condition.

time, expression levels in animals with concatemers were not higher compared to ones with one copy of the transgene, indicating that either some degree of gene silencing has occurred or that some of these extra copies were not intact. In order to improve the efficiency of single-copy transgene integration, concatemer formation may potentially be limited by in situ excision of donor sequences from plasmids. This could be achieved by flanking donor sequences on the injected plasmid by nuclease target sites. Such an approach has been tested in sea urchin and was found to favor HDR in sea urchin embryos injected with ZFNs (Ochiai et al. 2012). The use of a circular donor DNA with a nuclease site for in situ linearization but without any homology arms has also been shown recently to result in high-efficiency homology-independent knock-in of in vitro-transfected cells (Cristea et al. 2013; Maresca et al. 2013) and very recently in zebrafish (albeit concatemers were also observed using this strategy) (Auer et al. 2014). Future studies should compare the efficiency of targeted knock-in by HDR or NHEJ using this in vivo linearized donor DNA approach.

A parameter that can influence HDR following DNA cleavage by nucleases is the distance between the homology arms and the DNA break point. It has been found that the efficiency of gene conversion tracts after double-strand break repair by HDR in cells in vitro rapidly decreases with distance (100 bp) from the nuclease cleavage point (for review, see Johnson and Jasin 2001). In mice microinjected with TALE nucleases, oligonucleotide exchange preferentially occurred in proximity to the double-strand break (Wefers et al. 2013). This is an important point since it may be an obstacle to the exchange of endogenous sequences by new ones. Our results show that the efficiency of HDR using homology arms contiguous to the DNA break, such as in *Hprt1.1*, *Hprt1.2*, and *Rosa26* loci (1.02%, 0.63%, and 2.2%, respectively), were roughly

comparable to that for the *Ighm* locus (0.62%) with homology at ~150 bp from the DNA break point.

Transgenes expression can be influenced by the local environment (position effects) that can lead to transgene silencing or aberrant expression (Milot et al. 1996; Pedram et al. 2006; Gao et al. 2007; Williams et al. 2008). In contrast, targeted integration of DNA sequences into permissive loci, such as *Hprt/Hprt1* (Bronson et al. 1996; Meek et al. 2010) or *Rosa26* (Zambrowicz et al. 1997), by HDR in ES cells is usually chosen for expressing exogenous transgenes. Different ubiquitous promoters have been compared for levels of expression once inserted in the *Rosa26* locus, and CAG was the strongest (Chen et al. 2011). Targeted insertion using nucleases has only been reported in the *Rosa26* locus in mice embryos (Hermann et al. 2012). In accordance with these observations, all *Rosa26* HDR rats analyzed expressed GFP ubiquitously. Thus, integration of other transgenes into the *Hprt1* or *Rosa26* locus via nuclease-stimulated HDR using the donor DNA constructs described in this study should avoid the phenomenon of positional effects and result in transgene expression following the pattern of expression of the promoter used.

This work was done with Sprague-Dawley rats, but other rat strains should also be suitable for HDR-directed genome editing using TALE nucleases, since others have generated knockout rats via NHEJ using TALE nucleases in other strains (Mashimo et al. 2013).

In summary, this report demonstrates the feasibility of TALE nucleases delivered into the zygote to generate rats with targeted, complex, and large transgene insertions. In particular, we targeted transgenes to high-value loci, such as *Rosa26* or *Hprt1*, and replaced DNA sequences in an endogenous locus. This technique is a faster, cheaper, and easier alternative to ES cell manipulation (Dow and Lowe 2012). TALE nucleases now join ZFNs as having effected targeted transgene integration in vivo. CRISPR/Cas9 are even easier to

Table 2. Genotype of all *Hprt1.1*-, *Hprt1.2*-, and *Rosa26*-targeted GFP-positive founders

Target	Donor DNA	ID/Sex	No. of alleles	Allele KI		Status 2nd allele	RI	Concatemers
				5' insertion	3' insertion			
<i>Hprt1.1</i>	GFP	37°C-13.1/F	3	+	+	wt, Δ 2 bp	—	+
<i>Hprt1.2</i>	GFP	37°C-1.1/F	2	+	—	wt	—	+
<i>Hprt1.2</i>	GFP	37°C-6.1/F	2	+	+	Δ 110 bp	—	—
<i>Hprt1.2</i>	GFP	37°C-9.6/F	2	+	+	Δ 41 bp	—	+
<i>Hprt1.2</i>	GFP	37°C-11.9/F	2	+	+	Δ 25 bp	—	—
<i>Hprt1.2</i>	GFP	37°C-12.2/M	Mosaic	+	+	Mosaic	—	+
<i>Hprt1.2</i>	GFP	37°C-13.2/F	2	—	—	wt	+	+
<i>Hprt1.2</i>	GFP	30°C-1.2/F	Mosaic	+	+	Mosaic	—	+
<i>Hprt1.2</i>	GFP	30°C-2.2/F	2	+	+	Δ 9 + Ins 1 bp	—	—
<i>Hprt1.2</i>	GFP	30°C-3.2/M	1	—	—	Large Δ	+	+
<i>Hprt1.2</i>	GFP	30°C-3.4/F	2	+	+	Δ 16 bp	—	+
<i>Hprt1.2</i>	GFP	30°C-5.7/F	2	—	—	wt	+	—
<i>Hprt1.2</i>	GFP	30°C-5.9/F	2	+	+	Large Δ	—	—
<i>Hprt1.2</i>	GFP	30°C-6.4/M	1	—	—	Large Δ	+	+
<i>Hprt1.2</i>	GFP	30°C-6.5/F	2	—	—	Δ 16, Δ 22 + Ins 2 bp	+	+
<i>Hprt1.2</i>	GFP	30°C-7.3/F	2	+	+	Large Δ	—	+
<i>Hprt1.2</i>	GFP	30°C-7.5/F	Mosaic	—	—	Mosaic	+	+
<i>Hprt1.2</i>	GFP	30°C-7.8/F	2	—	—	Large Δ	+	—
<i>Hprt1.2</i>	GFP	30°C-8.3/F	2	+	+	Δ 19 bp	—	+
<i>Hprt1.2</i>	GFP	30°C-9.7/F	2	+	+	wt	—	+
<i>Hprt1.2</i>	GFP	30°C-10.2/F	2	+	+	Δ 29 bp	—	+
<i>Hprt1.2</i>	GFP	30°C-10.4/F	2	+	+	Δ 92 bp	—	+
<i>Hprt1.2</i>	GFP	37-30°C-4.4/F	3	+	+	Δ 19, Δ 12 + Ins2 bp	—	+
<i>Hprt1.2</i>	GFP	37-30°C-4.6/F	2	+	+	Δ 15 bp	—	+
<i>Hprt1.2</i>	GFP	37-30°C-5.1/F	2	—	—	Δ 50 bp	+	+
<i>Hprt1.2</i>	GFP	37-30°C-5.3/F	2	+	+	Large Δ	—	+
<i>Hprt1.2</i>	GFP	37-30°C-7.2/M	Mosaic	+	+	Mosaic	—	+
<i>Hprt1.2</i>	GFP	37-30°C-7.4/F	2	+	+	Δ 10 bp	—	—
<i>Hprt1.2</i>	GFP	37-30°C-9.5/F	2	—	—	wt	+	+
<i>Hprt1.2</i>	GFP	37-30°C-10.2/F	3	+	+	wt, Δ 1 bp	—	+
<i>Hprt1.2</i>	GFP	37-30°C-10.7/F	2	+	+	Δ 6 bp	—	+
<i>Hprt1.2</i>	GFP	37-30°C-10.8/M	Mosaic	—	—	Mosaic	+	+
<i>Hprt1.2</i>	GFP	30°C circ-4.1/M	1	—	—	Large Δ	+	—
<i>Rosa26</i>	GFP	37°C-2.1/F	2	+	+	Δ 17 bp	—	—
<i>Rosa26</i>	GFP	37°C-3.1/M	2	+	+	Δ 11 bp	—	+
<i>Rosa26</i>	GFP	37°C-5.3/F	2	—	—	wt, Δ 1 bp	+	nd
<i>Rosa26</i>	GFP	37-30°C-7.1/F	2	—	—	wt	+	—
<i>Rosa26</i>	GFP	37-30°C-8.4/M	Mosaic	+	+	Mosaic	—	—
<i>Rosa26</i>	GFP	37-30°C-8.5/M	2	+	+	Δ 16 bp	—	—
<i>Rosa26</i>	GFP	37-30°C-9.1/F	2	+	+	Large Δ	—	—
<i>Rosa26</i>	GFP	37°C-10.2/F	2	+	+	Large Δ	—	—
<i>Rosa26</i>	GFP	37°C-12.5/M	2	+	+	Δ 8 bp	—	+
<i>Rosa26</i>	GFP	30°C-19.3/F	2	—	—	wt	+	+

ID refers to each day 15 embryo (E15) or live pup carrying an insertion of the transgene by HDR or RI. The culture conditions after microinjection are indicated for each *Hprt1.1*-, *1.2*-, and *Rosa26*-positive animals. The *Hprt1* locus is in the X chromosome. Some females with three X alleles or males with two X alleles are explained by mosaicism. As examples, female *Hprt1.1* GFP 37°C-13.1/F had one X chromosome with wild-type sequences, a second one with a KI, and a third one with a 2-bp deletion that originated in NHEJ. Others such as *Hprt1.2* GFP 37-30°C-4.4/F had one with a KI and two more with NHEJ mutations. Large Δ refers to either fetuses carrying deletions >300 bp in the PCR used to perform the T7 and sequencing analysis—most of these animals did not show amplification using primers amplifying a 700-bp sequence. (Mosaic) Multiple undefined indels; (nd) not determined; (wt) wild-type allele.

generate and are an important new tool for genome editing, although more work is needed to fully evaluate potential off-target effects.

from Huang et al. (2011) and are described in detail in the Supplemental Material.

Methods

Animals

Sprague-Dawley (SD/Crl) rats were the only strain used and were sourced from the Charles River (L'Arbresle, France).

Design and production of TALE nucleases

TALE nucleases were produced as previously described (Piganeau et al. 2013; Auer et al. 2014) by the unit assembly method adapted

In vitro assay of TALE nucleases

Each subunit of TALE nuclease plasmid was nucleofected into C6 cells (Sigma) and DNA analyzed by PCR with specific primers (Supplemental Table S2), followed by the T7 endonuclease I assay (Menoret et al. 2014), as detailed in the Supplemental Material.

In vitro transcription of TALE nucleases mRNA

TALE nuclease plasmids were in vitro transcribed, polyadenylated, purified, and used as described in detail in the Supplemental Material.

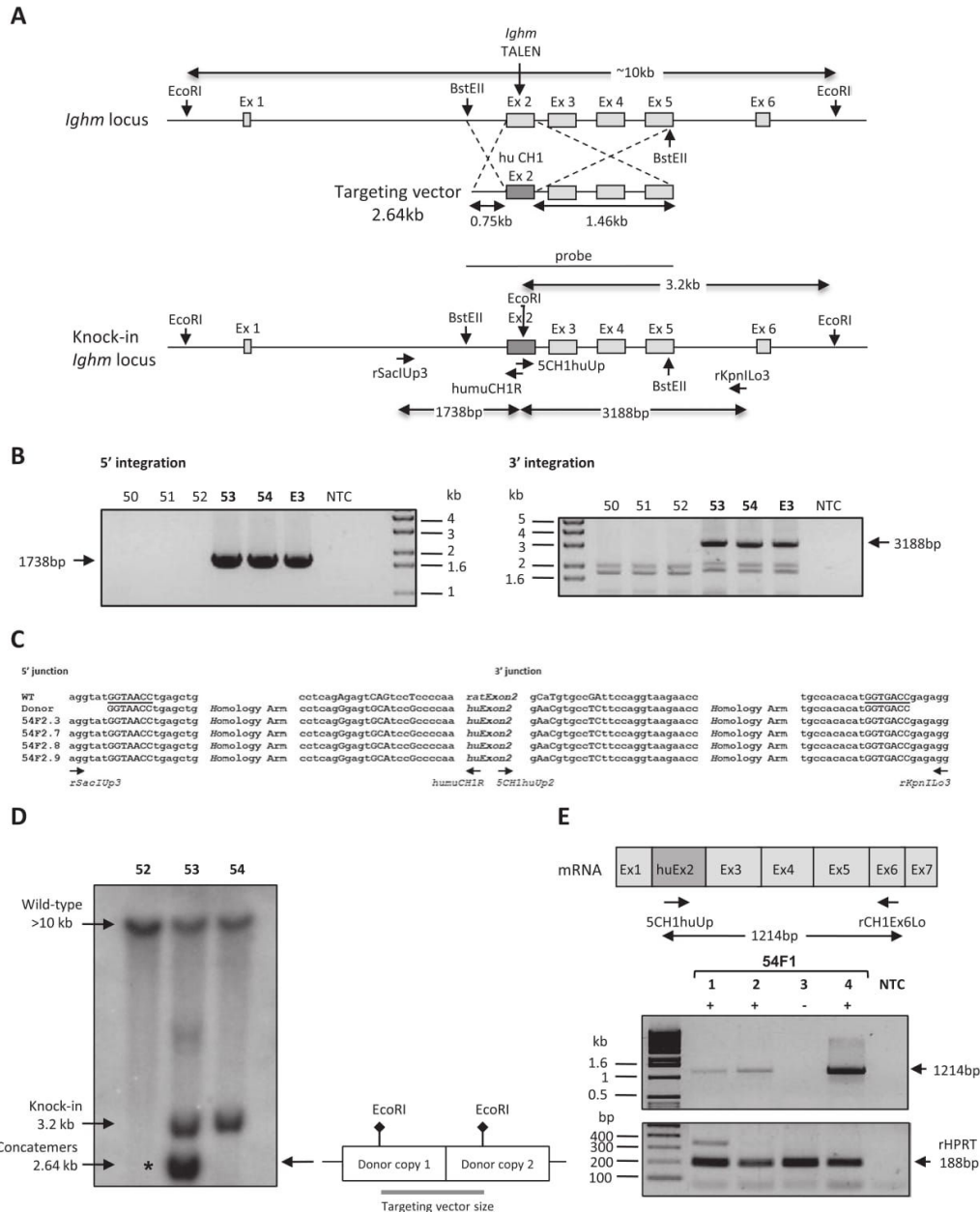


Figure 4. Targeted exon exchange into the rat *Ighm* locus. (A) (Upper) Diagrams showing schematic representations of the rat *Ighm* locus with the site of TALE nuclease action and of the targeting vector containing exon 2 of human *IGHM* flanked by 0.75-kb-long 5' and 1.46-kb-long 3' homology arms. (Lower) Diagram showing the integration by homologous recombination of the donor DNA sequence in the targeted locus, and the position of primers used for 5' and 3' junction PCRs. Integration in the 5' and 3' sides should generate fragments of 1738 bp and 3188 bp, respectively. EcoRI is the restriction enzyme used in Southern blot analyses due to the presence of the EcoRI site in the human but not in the rat *Ighm* sequence. The probe used in Southern blot analyses is the targeting vector. In the case of homologous recombination events, a 3.2-kb band is expected. (B) The left and right gels show the results analyzing, respectively, the 5' and 3' extremity of DNA integration, using the pair of primers indicated in A. Three animals (53, 54, and E3) showed a band of expected size (1738 bp in 5' and 3188 bp in 3' ends). (NTC) No template control. (C) Sequence comparison at the 5' and 3' junctions with wild-type genomic DNA and donor DNA sequences of four representative F2s (54F2.3, 54F2.7, 54F2.8, and 54F2.9: F2 generation from animal 54). Exon exchange was confirmed at the BstEII site in 5' and 3' (underlined). (D) Southern blot analysis of founders for homologous recombination integration. Genomic DNA was digested with EcoRI and 10 µg of DNA were loaded per lane. Blots were probed with the targeting sequence, as indicated in A. Arrows indicate bands of 10 kb and 3.2 kb corresponding to wild-type sequences and HDR sequences, respectively. An asterisk marks the presence of concatemers showing a band at 2.64 kb (the size of the transgene). The diagram at the right explains the expected size of the concatemers once linearized. Two animals (founders 53 and 54) harbor a HDR insertion, whereas no HDR event is observed in the third (animal 52). More than one copy of the donor DNA sequence is observed in animal 53. (E) (Upper) Diagrams showing a schematic representation of the mRNA sequence in HDR animals and the position of primers used for RT-PCR. A band of 1214 bp is expected. (Lower) Electrophoresis gel pictures show the presence of an amplified band of 1214 bp in three animals born from the mating of the founder 54 (HDR) with a wild-type rat. The amplification of rat HPRT serves as a control. (NTC) No template control.

Table 3. Microinjection statistics for the *Ighm* locus

Donor DNA form	Dose mRNA/DNA (ng/μl)	Temperature	No. injected eggs (% viable eggs)	No. E15 (e) or pups (p) (%) ^a	No. of RI-positive animals (%) ^a	No. of HDR-positive animals (%) ^a	No. of indel-positive animals (%) ^a
Linear ^b	5 + 5/10	37°C	1063 (72.7)	5e – 72p (15.8)	8 (1.64)	3 (0.62)	54 (11.1)
Circular ^c	5 + 5/10	37°C	410 (75.1)	64p (22.9)	2 (0.72)	0 (0)	35 (12.5)

Rat *Ighm* TALE nucleases, as mRNA, were injected at a concentration of 10 ng/μl in combination with 10 ng/μl of donor DNA, either in its linear form or in its circular form, both into the cytoplasm and into the male pronucleus. Injected eggs were maintained under 5% CO₂ at 37°C until reimplantation. Viability was evaluated after the culture period. Potential toxicity was also assessed by the number of day 15 embryos (E15) or of live pups obtained following the transfer of injected eggs. The numbers of E15 or live pups which have integrated the donor DNA sequence either by random integration (RI) or HDR integration (PCR positive both at the 5' and 3' ends) or which have indel mutations, are reported in the last three columns.

^aPercentages indicated in parentheses correspond to the percentage of transferred embryos.

^bExcised form.

^cSupercoiled DNA.

Targeting vector construction

Plasmids donor sequences were based on the Brown Norway rat genomic sequence (assembly RGSC_3.4). For the *Hprt1.1*, *Hprt1.2*, and *Rosa26* loci, donor sequences contained a CAG promoter-eGFP cDNA-BGHpA cassette flanked by two 800-bp homologous arms.

For the replacement of rat *Ighm* exon 2, we generated a plasmid containing human *IGHM* exon 2 flanked by rat sequence 5' and 3' homology arms (0.75 kb and 1.46 kb, respectively).

The generation and use of these donor DNA sequences is described in detail in the Supplemental Material.

Microinjection of rat one-cell embryos

Fertilized one-cell-stage embryos were sequentially microinjected into the male pronucleus and into the cytoplasm. One-cell embryo collection and manipulation are described in the Supplemental Material.

Analysis of NHEJ events

Briefly, DNA fragments including the TALE nuclease targeted regions were PCR-amplified with a high-fidelity polymerase (Herculase II fusion polymerase) using specific primers (Supplemental Table S2). Mutations were analyzed using the T7 endonu-

lease I assay (Menoret et al. 2013) and direct sequencing of PCR products.

Analysis of targeted and RI of DNA donor sequences

Donor DNA was amplified using the primer pairs for *GFP*. DNA from GFP+ animals was PCR-amplified with primers situated outside and inside of each extremity of the homology arms as described in detail in the Supplemental Material (Supplemental Table S2). Southern blots were done on genomic DNA digested by *EcoRI* for *Hprt1.1*, *BamHI* for *Hprt1.2*, and *EcoRI* for *Rosa26*.

Analysis of GFP expression

GFP expression was analyzed as described in detail in the Supplemental Material.

Analysis of *Ighm* mRNA

Analysis was performed on total RNA, DNase-treated, and PCR-amplified using primers (Supplemental Table S2) and techniques described in the Supplemental Material section.

Competing interest statement

G.J.C. is a full-time employee of Sangamo BioSciences.

Table 4. Genotypes of all *Ighm*-positive founders analyzed at the target site

Target	Donor DNA	ID/Sex	No. of alleles	Allele KI		Status 2nd allele	RI	Concatemers
				5' insertion	3' insertion			
<i>Ighm</i>	huCH1	E3F	2	+	+	Δ 6 bp	–	+
<i>Ighm</i>	huCH1	8/F	Mosaic	–	–	Mosaic	+	nd
<i>Ighm</i>	huCH1	44/F	2	–	–	Δ 84 bp	+	–
<i>Ighm</i>	huCH1	49/M	2	–	–	Δ 27, Δ 6 bp	+	–
<i>Ighm</i>	huCH1	53/M	2	+	+	Δ 5 bp	–	+
<i>Ighm</i>	huCH1	54/M	2	+	+	Δ 12 bp	–	–
<i>Ighm</i>	huCH1	55/F	2	–	–	Δ 16, Δ 16 bp	+	–
<i>Ighm</i>	huCH1	56/F	2	–	–	Δ 1, Δ 1 bp	+	–
<i>Ighm</i>	huCH1	63/F	2	–	–	Δ 5, Δ 5 bp	+	+
<i>Ighm</i>	huCH1	68/M	Mosaic	–	–	Mosaic	+	+
<i>Ighm</i>	huCH1	69/M	2	–	–	nd	+	+
<i>Ighm</i>	huCH1	Circ-81/F	2	–	–	nd	+	nd
<i>Ighm</i>	huCH1	Circ-85/M	2	–	–	wt	+	nd

Circ-81/F and Circ-85/M are the two animals generated with the circular donor DNA whereas all the other animals were generated with the linear excised form. (Mosaic) Multiple undefined indels; (nd) not determined.

Acknowledgments

Funding was provided by Région Pays de la Loire through Bio-genouest, IBISA program, Fondation Progreffe and TEFOR (Infrastructures d'Avenir of the French government). Sequencing experiments were performed by the integrative genomic facility of Nantes and Gregory J. Cost from Sangamo.

Author contributions: I.A., A.F., L.T., S.R., and S.M. conceived and designed the experiments. M.C., J.B.R., A.D.C., J.P.C., and C.G. performed the design, production, and in vitro assay of *Hprt1* and *Rosa26* TALE nucleases. G.J.C. designed and produced *Ighm* TALE nucleases and edited the manuscript. R.B. generated the *IGHM* DNA construct. L.T. performed in vitro transcription of TALE nucleases, built donor constructs, and carried out PCR and Southern blot analyses. V.T., R.T., and L.T. performed the T7 assay. S.R., S.M., and C.U. performed the microinjections. D.B. performed the statistical analyses. S.R. performed the FACS experiments and microscopy. S.R., L.T., and I.A. wrote the manuscript.

References

- Auer TO, Duroure K, De Cian A, Concordet JP, Del Bene F. 2014. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res* **24**: 142–153.
- Beumer KJ, Trautman JK, Bozas A, Liu JL, Rutter J, Gall JG, Carroll D. 2008. Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc Natl Acad Sci* **105**: 19821–19826.
- Brinster RL, Chen HY, Trumbauer ME, Yagle MK, Palmiter RD. 1985. Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. *Proc Natl Acad Sci* **82**: 4438–4442.
- Brinster RL, Braun RE, Lo D, Avarbock MR, Oram F, Palmiter RD. 1989. Targeted correction of a major histocompatibility class II E α gene by DNA microinjection into mouse eggs. *Proc Natl Acad Sci* **86**: 7087–7091.
- Bronson SK, Plaehn EG, Kluckman KD, Hagaman JR, Maeda N, Smithies O. 1996. Single-copy transgenic mice with chosen-site integration. *Proc Natl Acad Sci* **93**: 9067–9072.
- Brown AJ, Fisher DA, Kouranova E, McCoy A, Forbes K, Wu Y, Henry R, Ji D, Chambers A, Warren J, et al. 2013. Whole-rat conditional gene knockout via genome editing. *Nat Methods* **10**: 638–640.
- Buehr M, Meek S, Blair K, Yang J, Ure J, Silva J, McLay R, Hall J, Ying QL, Smith A. 2008. Capture of authentic embryonic stem cells from rat blastocysts. *Cell* **135**: 1287–1298.
- Capecchi MR. 2005. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* **6**: 507–512.
- Chen CM, Krohn J, Bhattacharya S, Davies B. 2011. A comparison of exogenous promoter activity at the *ROSA26* locus using a PhiC31 integrase mediated cassette exchange approach in mouse ES cells. *PLoS ONE* **6**: e23376.
- Cristea S, Freyvert Y, Santiago Y, Holmes MC, Urnov FD, Gregory PD, Cost GJ. 2013. In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol Bioeng* **110**: 871–880.
- Cui X, Ji D, Fisher DA, Wu Y, Briner DM, Weinstein EJ. 2011. Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat Biotechnol* **29**: 64–67.
- Dai J, Cui X, Zhu Z, Hu W. 2010. Non-homologous end joining plays a key role in transgene concatemer formation in transgenic zebrafish embryos. *Int J Biol Sci* **6**: 756–768.
- Dow LE, Lowe SW. 2012. Life in the fast lane: mammalian disease models in the genomics era. *Cell* **148**: 1099–1109.
- Doyon Y, Choi VM, Xia DF, Vo TD, Gregory PD, Holmes MC. 2010. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nat Methods* **7**: 459–460.
- Flisikowska T, Thorey IS, Offner S, Ros F, Lifke V, Zeitler B, Rottmann O, Vincent A, Zhang L, Jenkins S, et al. 2011. Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc finger nucleases. *PLoS ONE* **6**: e21045.
- Gao Q, Reynolds GE, Innes L, Pedram M, Jones E, Junabi M, Gao DW, Ricoul M, Sabatier L, Van Brocklin H, et al. 2007. Telomeric transgenes are silenced in adult mouse tissues and embryo fibroblasts but are expressed in embryonic stem cells. *Stem Cells* **25**: 3085–3092.
- Garrick D, Fiering S, Martin DJ, Whitelaw E. 1998. Repeat-induced gene silencing in mammals. *Nat Genet* **18**: 56–59.
- Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, Jenkins SS, Wood A, Cui X, Meng X, et al. 2009. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325**: 433.
- Hermann M, Maeder ML, Rector K, Ruiz J, Becher B, Burki K, Khayter C, Aguzzi A, Joung JK, Buch T, et al. 2012. Evaluation of OPEN zinc finger nucleases for direct gene targeting of the *ROSA26* locus in mouse embryos. *PLoS ONE* **7**: e41796.
- Huang P, Xiao A, Zhou M, Zhu Z, Lin S, Zhang B. 2011. Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol* **29**: 699–700.
- Jacob HJ. 1999. Functional genomics and rat models. *Genome Res* **9**: 1013–1016.
- Jacob HJ. 2010. The rat: a model used in biomedical research. *Methods Mol Biol* **597**: 1–11.
- Jacob HJ, Kwitek AE. 2002. Rat genetics: attaching physiology and pharmacology to the genome. *Nat Rev Genet* **3**: 33–42.
- Johnson RD, Jasim M. 2001. Double-strand-break-induced homologous recombination in mammalian cells. *Biochem Soc Trans* **29**: 196–201.
- Katsuyama T, Akamammedov A, Seimiya M, Hess SC, Sievers C, Paro R. 2013. An efficient strategy for TALEN-mediated genome engineering in *Drosophila*. *Nucleic Acids Res* **41**: e163.
- Kobayashi T, Kato-Itoh M, Yamaguchi T, Tamura C, Sanbo M, Hirabayashi M, Nakauchi H. 2012. Identification of rat *Rosa26* locus enables generation of knock-in rat lines ubiquitously expressing *tdTomato*. *Stem Cells Dev* **21**: 2981–2986.
- Li P, Tong C, Mehrian-Shai R, Jia L, Wu N, Yan Y, Maxson RE, Schulze EN, Song H, Hsieh CL, et al. 2008. Germline competent embryonic stem cells derived from rat blastocysts. *Cell* **135**: 1299–1310.
- Li D, Qiu Z, Shao Y, Chen Y, Guan Y, Liu M, Li Y, Gao N, Wang L, Lu X, et al. 2013a. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat Biotechnol* **31**: 681–683.
- Li W, Teng F, Li T, Zhou Q. 2013b. Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat Biotechnol* **31**: 684–686.
- Maresca M, Lin VG, Guo N, Yang Y. 2013. Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated integration through nonhomologous end joining. *Genome Res* **23**: 539–546.
- Mashimo T, Takizawa A, Voigt B, Yoshimi K, Hiai H, Kuramoto T, Serikawa T. 2010. Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PLoS ONE* **5**: e8870.
- Mashimo T, Kaneko T, Sakuma T, Kobayashi J, Kunihiro Y, Voigt B, Yamamoto T, Serikawa T. 2013. Efficient gene targeting by TAL effector nucleases coinjected with exonucleases in zygotes. *Sci Rep* **3**: 1253.
- Meek S, Buehr M, Sutherland L, Thomson A, Mullins JJ, Smith AJ, Burdon T. 2010. Efficient gene targeting by homologous recombination in rat embryonic stem cells. *PLoS ONE* **5**: e14225.
- Menoret S, Fontaniere S, Jantz D, Tesson L, Thinarat R, Remy S, Usal C, Ouisse LH, Fraichard A, Anegón I. 2013. Generation of *Rag1*-knockout immunodeficient rats and mice using engineered meganucleases. *FASEB J* **27**: 703–711.
- Menoret S, Tesson L, Remy S, Usal C, Thepenier V, Thinarat R, Ouisse LH, De Cian A, Giovannangeli C, Concordet JP, et al. 2014. Gene targeting in rats using transcription activator-like effector nucleases. *Methods* doi: 10.1016/j.jmeth.2014.02.027.
- Meyer M, de Angelis MH, Wurst W, Kuhn R. 2010. Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. *Proc Natl Acad Sci* **107**: 15022–15026.
- Meyer M, Ortiz O, Hrabec de Angelis M, Wurst W, Kuhn R. 2012. Modeling disease mutations by gene targeting in one-cell mouse embryos. *Proc Natl Acad Sci* **109**: 9354–9359.
- Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, et al. 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**: 143–148.
- Milot E, Strouboulis J, Trimbom T, Wijgerde M, de Boer E, Langeveld A, Tan-Un K, Vergeer W, Yannoutsos N, Grosveld F, et al. 1996. Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* **87**: 105–114.
- Ochiai H, Sakamoto N, Fujita K, Nishikawa M, Suzuki K, Matsuura S, Miyamoto T, Sakuma T, Shibata T, Yamamoto T. 2012. Zinc-finger nuclease-mediated targeted insertion of reporter genes for quantitative imaging of gene expression in sea urchin embryos. *Proc Natl Acad Sci* **109**: 10915–10920.
- Panda SK, Wefers B, Ortiz O, Floss T, Schmid B, Haass C, Wurst W, Kuhn R. 2013. Highly efficient targeted mutagenesis in mice using TALENs. *Genetics* **195**: 703–713.
- Pedram M, Sprung CN, Gao Q, Lo AW, Reynolds GE, Murnane JP. 2006. Telomere position effect and silencing of transgenes near telomeres in the mouse. *Mol Cell Biol* **26**: 1865–1878.
- Piganeau M, Ghezraoui H, De Cian A, Guittat L, Tomishima M, Perrouault L, Rene O, Katibah GE, Zhang L, Holmes MC, et al. 2013. Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome Res* **23**: 1182–1193.

- Ponce de Leon V, Merillat AM, Tesson L, Anegon I, Hummler E. 2014. Generation of TALEN-mediated GR^{dim} knock-in rats by homologous recombination. *PLoS ONE* **9**: e88146.
- Segal DJ, Meckler JF. 2013. Genome engineering at the dawn of the golden age. *Annu Rev Genomics Hum Genet* **14**: 135–158.
- Sommer D, Peters A, Wirtz T, Mai M, Ackermann J, Thabet Y, Schmidt J, Weighardt H, Wunderlich FT, Degen J, et al. 2014. Efficient genome engineering by targeted homologous recombination in mouse embryos using transcription activator-like effector nucleases. *Nat Commun* **5**: 3045.
- Tesson L, Cozzi J, Menoret S, Remy S, Usal C, Fraichard A, Anegon I. 2005. Transgenic modifications of the rat genome. *Transgenic Res* **14**: 531–546.
- Tesson L, Usal C, Menoret S, Leung E, Niles BJ, Remy S, Santiago Y, Vincent AI, Meng X, Zhang L, et al. 2011. Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol* **29**: 695–696.
- Tong C, Li P, Wu NL, Yan Y, Ying QL. 2010. Production of p53 gene knockout rats by homologous recombination in embryonic stem cells. *Nature* **467**: 211–213.
- Tong C, Huang G, Ashton C, Wu H, Yan H, Ying QL. 2012. Rapid and cost-effective gene targeting in rat embryonic stem cells by TALENs. *J Genet Genomics* **39**: 275–280.
- Wang H, Hu YC, Markoulaki S, Welstead GG, Cheng AW, Shivalila CS, Pyntikova T, Dadon DB, Voytas DF, Bogdanove AJ, et al. 2013. TALEN-mediated editing of the mouse Y chromosome. *Nat Biotechnol* **31**: 530–532.
- Wefers B, Meyer M, Ortiz O, Hrabe de Angelis M, Hansen J, Wurst W, Kuhn R. 2013. Direct production of mouse disease models by embryo microinjection of TALENs and oligodeoxynucleotides. *Proc Natl Acad Sci* **110**: 3782–3787.
- Williams A, Harker N, Ktistaki E, Veiga-Fernandes H, Roderick K, Tolaini M, Norton T, Williams K, Kioussis D. 2008. Position effect variegation and imprinting of transgenes in lymphocytes. *Nucleic Acids Res* **36**: 2320–2329.
- Yamamoto S, Nakata M, Sasada R, Ooshima Y, Yano T, Shinozawa T, Tsukimi Y, Takeyama M, Matsumoto Y, Hashimoto T. 2012. Derivation of rat embryonic stem cells and generation of protease-activated receptor-2 knockout rats. *Transgenic Res* **21**: 743–755.
- Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, Soriano P. 1997. Disruption of overlapping transcripts in the ROSA β geo 26 gene trap strain leads to widespread expression of β -galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci* **94**: 3789–3794.
- Zheng S, Geghman K, Shenoy S, Li C. 2012. Retake the center stage—new development of rat genetics. *J Genet Genomics* **39**: 261–268.

Received December 20, 2013; accepted in revised form April 16, 2014.

4.1.5 Article : Edition du génome de poisson zèbre par Knock-Out

ARTICLE

Received 3 Sep 2015 | Accepted 28 Jan 2016 | Published 7 Mar 2016

DOI: 10.1038/ncomms10866

OPEN

CSF-contacting neurons regulate locomotion by relaying mechanical stimuli to spinal circuits

Urs Lucas Böhm^{1,2,3,4,*}, Andrew Prendergast^{1,2,3,4,*}, Lydia Djenoune^{1,2,3,4,5}, Sophie Nunes Figueiredo^{1,2,3,4}, Johanna Gomez^{1,2,3,4}, Caleb Stokes^{1,2,3,4}, Sonya Kaiser^{1,2,3,4}, Maximilliano Suster^{6,7}, Koichi Kawakami⁷, Marine Charpentier⁵, Jean-Paul Concordet⁵, Jean-Paul Rio^{1,2,3,4}, Filippo Del Bene^{8,9,10} & Claire Wyart^{1,2,3,4}

Throughout vertebrates, cerebrospinal fluid-contacting neurons (CSF-cNs) are ciliated cells surrounding the central canal in the ventral spinal cord. Their contribution to modulate locomotion remains undetermined. Recently, we have shown CSF-cNs modulate locomotion by directly projecting onto the locomotor central pattern generators (CPGs), but the sensory modality these cells convey to spinal circuits and their relevance to innate locomotion remain elusive. Here, we demonstrate *in vivo* that CSF-cNs form an intraspinal mechanosensory organ that detects spinal bending. By performing calcium imaging in moving animals, we show that CSF-cNs respond to both passive and active bending of the spinal cord. In mutants for the channel Pkd2l1, CSF-cNs lose their response to bending and animals show a selective reduction of tail beat frequency, confirming the central role of this feedback loop for optimizing locomotion. Altogether, our study reveals that CSF-cNs constitute a mechanosensory organ operating during locomotion to modulate spinal CPGs.

¹Institut du Cerveau et de la Moelle épinière, Paris 75013, France. ²UPMC Univ. Paris 06, Paris 75005, France. ³Inserm UMR 1127, Paris 75013, France. ⁴CNRS UMR 7225, Paris 75013, France. ⁵Muséum National d'Histoire Naturelle, Paris 75005, France. ⁶Neural Circuits and Behaviour Group, University of Bergen, Bergen 5008, Norway. ⁷Division of Molecular and Developmental Biology, National Institute of Genetics and Department of Genetics, Sokenkai (The Graduate University for Advanced Studies), Mishima, Shizuoka 411-8540, Japan. ⁸Institut Curie, Paris 75005, France. ⁹CNRS UMR 3215, Paris 75005, France. ¹⁰Inserm U 934, Paris 75005, France. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to C.W. (email: claire.wyart@icm-institute.org).

Behavior has long been known to be modulated by the content of cerebrospinal fluid (CSF)^{1,2}. Studies of CSF flow have shown that locomotor defects observed in normal pressure hydrocephalus can be reversed by CSF elimination. More recently, CSF has been shown to relay important clues to migrating neurons during development^{3,4} and to remove metabolites during sleep⁵. However, the mechanism by which molecular and hydrodynamic properties of the CSF change network dynamics is unknown. In the vertebrate spinal cord, a unique and conserved population of GABAergic neurons extends microvilli into the CSF^{6–10}. These neurons, referred to as CSF-cNs (and in *Xenopus* and *Danio rerio* as KA neurons¹¹), are among the most poorly understood components of the vertebrate motor network. We previously showed that CSF-cNs can modulate locomotion¹². Recently, we demonstrated that these neurons perform this modulation by projecting onto essential glutamatergic components of the locomotor central pattern generators¹³ (CPGs). *In vitro* studies show that in the spinal cord, CSF-cNs selectively express the transient receptor potential channel PKD2L1, a molecular component of the response to pH and osmolarity^{14–16}. However, it is unknown whether these neurons *in vivo* relay sensory information onto motor circuits during innate locomotion. Here, we combine calcium imaging in moving zebrafish larvae with quantitative behavioural analysis to investigate CSF-cN activity during locomotion and impact on behavior. We show that CSF-cNs respond to both passive and active bending of the spinal cord *in vivo*. As the PKD2L1 channel is specific to CSF-cNs in the spinal cord, we investigated the role of this channel in the sensory functions of CSF-cNs *in vivo*.

In *pkd2l1* mutants, cellular morphology is unaltered but the response to bending is abolished, suggesting that this channel is necessary for CSF-cN mechanosensory response to bending. Furthermore, the loss of CSF-cN sensory function leads to a reduction of tail beat frequency in *pkd2l1* mutants as well as in animals with impaired synaptic release from CSF-cNs, confirming the importance of this sensory feedback loop in the regulation of locomotion. Together, our results demonstrate that Pkd2l1 orchestrates a feedback loop that relays information on local curvature of the spinal cord to motor circuits in order to regulate the frequency of locomotion.

Results

CSF-cNs exhibit a motile kinocilium and multiple microvilli.

We used a promoter of *pkd2l1* that specifically targets CSF-cNs in the larval spinal cord (Fig. 1a, out of 253 *pkd2l1*⁺ cells from five larvae, CSF-cNs represented 94.2% of total cell counts, see also¹³) to drive the expression of fluorescent proteins (Fig. 1a–e). Previous studies suggested that CSF-cNs extend cilia into the central canal^{6,9,11,17,18} as depicted in Fig. 1c. We co-labeled CSF-cNs with Arl13b–GFP as a marker of the ciliary axoneme in *Tg(βact:arl13-GFP, pkd2l1:Gal4;cmcl2:eGFP, UAS:tagRFP-CAAX)* triple transgenic larvae (Fig. 1d). We observed a single cilium extending from a brush of microvilli (Fig. 1d',d''). To establish whether the CSF-cN cilium was motile, we used APEX-2, a genetically-encoded peroxidase, to label CSF-cNs for electron microscopy (Fig. 1e). Our analysis of the ciliary ultrastructure reveals a 9 + 2 organization, indicating that CSF-cNs extend a single motile cilium into the central canal (Fig. 1e',e''). This was

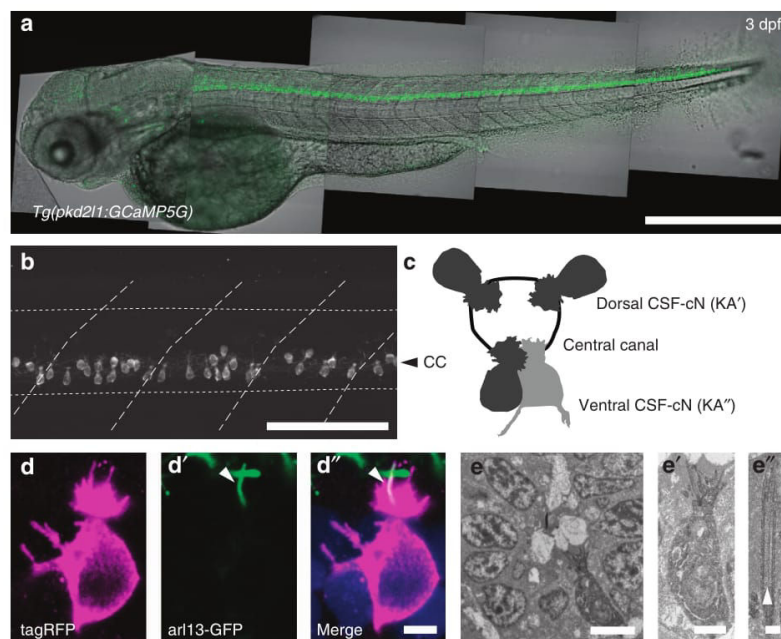


Figure 1 | In the spinal cord, *pkd2l1*⁺ CSF-cNs project an apical extension consisting of one kinocilium and a brush of microvilli into the central canal. (a) Transmitted and fluorescent image showing that the *pkd2l1* promoter in *Tg(pk2l1:GCaMP5G)* transgenic larva drives GCaMP5G expression in CSF-cNs along the entire spinal cord. Scale bar: 500 μ m. (b) A close up from a lateral view in the same transgenic animal shows the morphology of CSF-cNs as elongated cells with an apical extension reaching the central canal. Scale bar: 100 μ m. (c) Schematic depicting dorsal and ventral CSF-cN location around the central canal. (d) Confocal microscopy of 50 μ m sections of the 4 dpf triple transgenic larva *Tg(pk2l1:gal4, UAS:tagRFP-CAAX;cmcl2:GFP, βact:arl13-GFP)* shows that CSF-cNs project one Arl13-GFP + cilium (arrowhead) and multiple microvilli into the CSF. Scale bar: 3 μ m. (e) Transmission electron microscopy in *Tg(pk2l1:Gal4)* larvae injected with UAS:APEX2-tagRFP shows a single ventral CSF-cN reaching the central canal (close up in (e')), scale bar: 2 μ m) and exhibiting a motile 9 + 2 cilium (arrowhead, (e''), scale bar: 250 nm).

confirmed by video recording of CSF-cNs *in vivo* (Supplementary Movie 1) as well as in zebrafish dissociated cell cultures (Supplementary Movie 2). The unusual morphology of CSF-cNs may be described as disorganized hair cells located at the interface with the CSF and extending an ascending ipsilateral axon in the spinal cord.

CSF-cNs show minimal activation during fictive escapes. To investigate the involvement of CSF-cNs in the control of locomotion, we characterized their activity patterns during spinal network activity in paralyzed larvae. We used two-photon calcium imaging to record cellular activity during fictive escape behavior induced by a water jet to the otic vesicle (Fig. 2a). Fictive locomotor output was monitored by recording the ventral nerve root (Fig. 2a)¹⁹. We compared CSF-cN activity to dorsal motor neurons (MNs), which are known to be active during escape behavior (Fig. 2b)²⁰. MNs showed strong calcium transients during fictive escape behavior (mean $\Delta F/F_{\text{MNs}} = 1.2 \pm 0.1$; Fig. 2c top panels and 2d). In contrast, fictive locomotion only led to minute changes in calcium activity in CSF-cNs (mean $\Delta F/F_{\text{CSF-cNs}} = 0.042 \pm 0.007$; Fig. 2c,d bottom panels). To verify that the lack of activity in CSF-cNs was not due to α -bungarotoxin used as a paralytic, we repeated the experiment in immotile *cacnb1^{ts25/ts25}* larvae, where a mutation in the β_{1a} subunit of the dihydropyridine receptor prevents contraction of skeletal muscle²¹. The lack of activation of CSF-cNs during fictive escapes in *cacnb1^{ts25/ts25}* mutants (Supplementary Fig. 1) confirms that CSF-cNs are minimally activated when muscles are not contracting.

Muscle contraction leads to strong activation of CSF-cNs. To test whether CSF-cNs are recruited during active locomotion, we performed calcium imaging in the spinal cord of head-restrained, tail-free animals performing escape behaviours (Fig. 3a). We took advantage of the slow decay of GCaMP5G transients in CSF-cNs ($\tau_{\text{CSF-cNs}} = 1.18 \pm 0.07$ s) relative to tail movements (< 200 ms) to compare calcium signals before and after active tail motion. TagRFP was co-expressed in CSF-cNs with GCaMP5G to track moving cells and to correct for motion artifacts (see Methods section; Fig. 3a' and Supplementary Fig. 2). We simultaneously performed high-speed video recording of the entire larva to quantify the behavior in response to otic vesicle stimulation (Supplementary Movie 3). In contrast to paralyzed larvae, we observed strong activation of CSF-cNs in response to tail bends ($\Delta R/R_{\text{CSF-cNs}} = 0.38 \pm 0.08$). Moreover, after unilateral tail bends, our results revealed a clear activation of dorsal CSF-cNs ipsilateral to the contracting side ($\Delta R/R_{\text{CSF-cNs}} = 0.9 \pm 0.2$; Fig. 3b) whereas dorsal contralateral and ventral cells remained mostly silent ($\Delta R/R_{\text{CSF-cNs}} = 0.20 \pm 0.08$; $\Delta R/R_{\text{CSF-cN}} = 0.10 \pm 0.07$; Fig. 3b). Our data indicate that dorsal CSF-cNs respond to active bending of the spinal cord selectively on the contracting side.

Bending of the tail recruits CSF-cNs in paralyzed animals. To test whether CSF-cNs respond to mechanical stimulation during bending of the spinal cord, we mimicked the mechanical deformation associated with muscle contraction by imposing a local bend on the tail of paralyzed larvae using a mechanical probe. The tail of agarose-embedded larvae was exposed on one side and pressed with a glass probe (Fig. 3c). Here, CSF-cNs

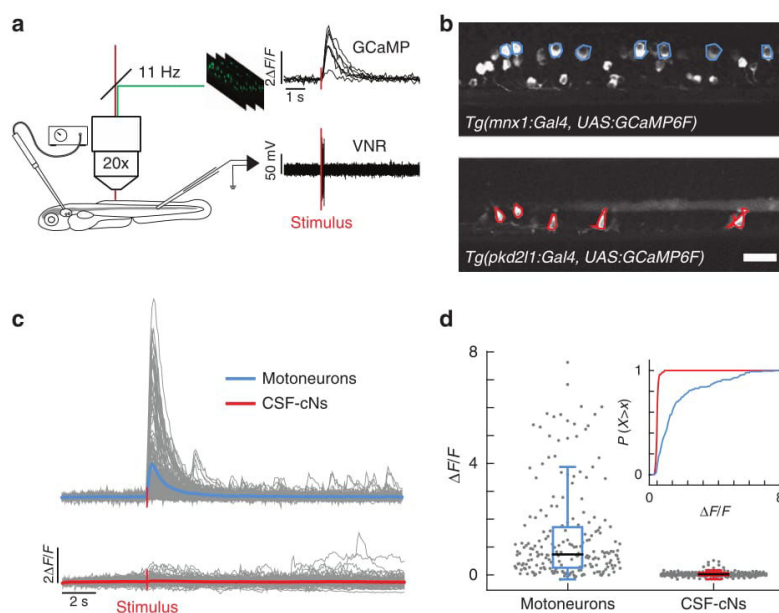


Figure 2 | CSF-cNs are minimally activated during fictive escapes when no muscle contraction occurs. (a) Schematic view of the experimental setup combining 2-photon laser scanning microscope for calcium imaging and electrophysiological recordings of the ventral nerve root. 10 ms-long water jets delivered to the otic vesicle triggered fictive escapes in paralyzed larvae. (b) Lateral view showing expression of GCaMP6f in MNs in the double transgenic larva *Tg(mnx1:Gal4, UAS:GCaMP6f;cryaa:mCherry)* and in CSF-cNs in *Tg(pkcd2l1:gal4, UAS:GCaMP6f;cryaa:mCherry)*. ROIs indicate cells included in the analysis. Only the dorsalmost MNs were analyzed in the *mnx1* line. Scale bar: 20 μm . (c) Typical calcium transients recorded in MNs (blue) and in CSF-cNs (red) during fictive escapes; 'stimulus' indicates when the water jet was triggered, average response in coloured lines. (d) Quantification of calcium transient amplitude in MNs and CSF-cNs (each data point represents one recording from one cell; plots use median as the measure of central tendency; inset is the cumulative histogram of calcium responses). Responses in both populations are greater than baseline (204 MNs from 6 larvae: mean $\Delta F/F = 1.2$, $P < 1.0 \times 10^{-8}$; 192 CSF-cNs from 7 larvae: mean $\Delta F/F = 0.042$, $P = 3.78 \times 10^{-8}$), but CSF-cNs exhibit significantly less activity than MNs ($P = 0.014$).

showed activity in response to mechanical stimulation ($\Delta R/R_{\text{CSF-cNs}} = 1.0 \pm 0.1$; Fig. 3d). The response was restricted to the site of stimulation and its amplitude decreased with increasing distance from the probe (Fig. 3e). Dorsal CSF-cNs ipsilateral to the stimulation site showed the largest responses ($\Delta R/R_{\text{CSF-cNs}} = 1.5 \pm 0.3$; Fig. 3f) while dorsal contralateral and ventral CSF-cNs showed a significant, albeit much smaller, response ($\Delta R/R_{\text{CSF-cNs}} = 0.5 \pm 0.2$; $\Delta R/R_{\text{VCSF-cNs}} = 0.8 \pm 0.1$; Fig. 3f). Results from passive bending further support the hypothesis that CSF-cNs constitute an *in vivo* mechanosensory system locally detecting bending of the spinal cord.

Bending-evoked responses are abolished in *pkd2l* mutants. The restricted expression of PKD2L1 to CSF-cNs in the spinal cord and its conservation across multiple vertebrate species suggest an important role specific to CSF-cNs^{7,15}. To test whether this channel contributes to the mechanical response of CSF-cNs, we used transcription activator-like effector nucleases (TALENs) to generate a *pkd2l1* null mutant (Fig. 4a). CSF-cNs without functional Pkd2l1 show no overall morphological defects in

zebrafish larva and their cilia were still beating *in vitro* ($n=3$). However, the responses of dorsal ipsilateral CSF-cNs to active bending (*pkd2l1*^{icm02/icm02} $\Delta R/R_{\text{CSF-cN}} = 0.07 \pm 0.06$; Fig. 4b) and passive bending (*pkd2l1*^{icm02/icm02} $\Delta R/R_{\text{CSF-cN}} = 0.03 \pm 0.06$; Fig. 4c) abolished in *pkd2l1*^{icm02/icm02} mutants. Our data indicate a crucial role of Pkd2l1 in mediating CSF-cN detection of spinal bending in zebrafish larvae.

***pkd2l1* mutants show reduced tailbeat frequency.** The ultimate output of sensory processing in the central nervous system is behavior. To probe the active contribution of the CSF-cN sensory motor loop to locomotion, we quantified acoustically-induced escape behavior in *pkd2l1*^{icm02/icm02} mutants, where CSF-cN response to bending is impaired (457 escapes from 141 larvae, seven experiments, Fig. 4d). Kinematic analysis revealed a specific decrease of mean tail beat frequency (TBF) in *pkd2l1*^{icm02/icm02} mutants compared to WT siblings (32.8 ± 0.6 Hz versus 35.2 ± 0.7 Hz; Fig. 4d and Supplementary Fig. 3a). We observed a significant effect of trial number on TBF (Fig. 4f), latency, distance, speed and number of oscillations (Supplementary

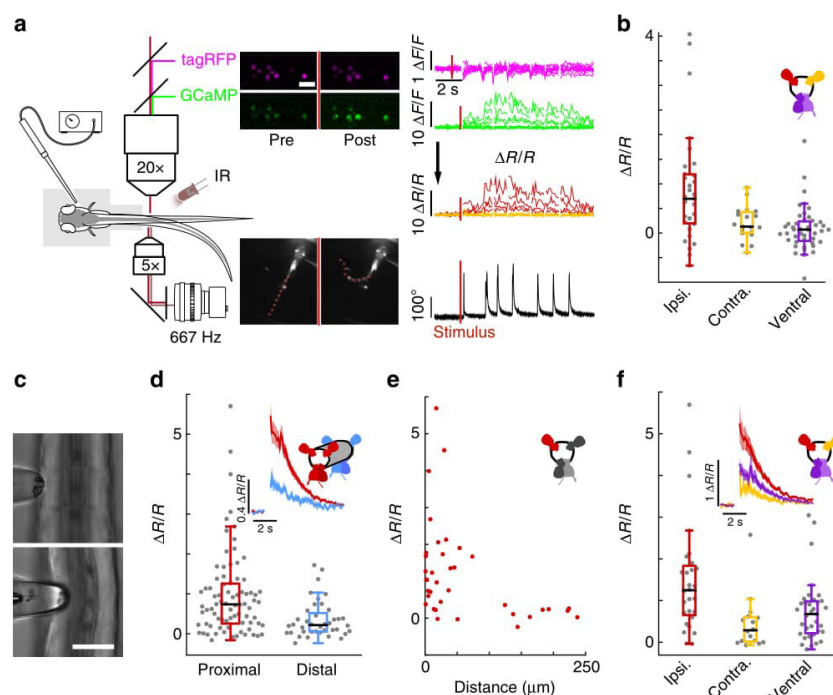


Figure 3 | CSF-cNs respond to active muscle contraction as well as to passive mechanical bending of the spinal cord. (a) Schematic describing 2-photon imaging experiments used to record simultaneously from CSF-cNs expressing tagRFP (magenta) and GCaMP5 (green) in head-embedded *Tg(pk2l1:GCaMP5, pk2l1:tagRFP)* larvae. Infrared illumination combined with high-speed video recording shows unidirectional tail deflections induced by a water jet to the otic vesicle. Sample traces for $\Delta F/F$ of tagRFP and GCaMP shown with tail deflection during escape (note: vertical scale is 10 times larger for GCaMP compared to tagRFP signals). Subtracting the tagRFP signal from the GCaMP signal removed motion artifacts; breaks in the trace arise from frames when cells escaped from the focal plane. Scale bar: 10 μm . (b) Quantification of calcium transient amplitude in response to muscle contraction ($n=11$ larvae) in dorsal CSF-cNs either ipsilateral (red, 31 cells) or contralateral (yellow, 19 cells) or ventral (purple, 44 cells). Only dorsal ipsilateral cells exhibited responses greater than baseline ($P=9.51 \times 10^{-4}$) and all other cell types responded significantly less than dorsal ipsilateral cells (dorsal contralateral: $P=2.43 \times 10^{-3}$, ventral: $P=4.85 \times 10^{-5}$). (c) Passive mechanical stimulation of CSF-cNs in paralyzed larvae ($n=5$) was implemented with mechanical pressure exerted by pushing a glass probe laterally against the fish tail. Scale bar: 50 μm . (d) Response of proximal (<100 μm) and distal (>100 μm) CSF-cNs. Inset: average calcium response of proximal (red) versus distal (blue) CSF-cNs. (e) Response of dorsal ipsilateral (red) CSF-cNs as function of distance from the probe. Inset: Average calcium response of dorsal ipsilateral versus dorsal contralateral and ventral CSF-cNs. All cell types show a response different from 0 (dorsal ipsilateral: $P<1.0 \times 10^{-8}$, dorsal contralateral: $P=5.95 \times 10^{-5}$, ventral: $P<1.0 \times 10^{-8}$) and all other cell types responded significantly less than dorsal ipsilateral cells (dorsal contralateral: $P=1.06 \times 10^{-3}$, ventral: $P=7.22 \times 10^{-3}$).

Fig. 4b) indicating some habituation to stimulus even at an inter trial interval of 2 min. Accounting for this trial effect and using a mixed linear model, we found that TBF was reduced in the *pkd2l1^{icm02/icm02}* mutant across trials (Fig. 4f'). The specific reduction of TBF in larvae lacking CSF-cN sensory response reveals that CSF-cN feedback is needed to maximize locomotor frequency during active locomotion. We recapitulated this effect by silencing the vesicular release of CSF-cNs with Botulinum toxin light chain in *Tg(pkcd2l1:gal4;UAS:BoTxBLC-GFP)* transgenic larvae (36.3 ± 0.7 Hz versus 38.9 ± 0.5 Hz; Supplementary Fig. 4 (ref. 22)). The convergence of effects obtained from these two independent methods for silencing CSF-cNs demonstrates their central role in the modulation of locomotor frequency.

Discussion

The hypothesis that CSF-cNs constitute a mechanosensitive organ was already formulated by Kolmer, although to-date no experiments had formally addressed this question *in vivo*. Here, we report activation of CSF-cNs in response to active as well as

passive spinal bending (Supplementary Fig. 5). In natural conditions, we show here that CSF-cN mechanosensory response shapes innate locomotion. Although a small population of *pkd2l1* expressing neurons was described in dorsal spinal neurons at embryonic stages⁷, this population is undetectable in the larva and therefore unlikely to mediate the behavioural effects presented here (see Supplementary Fig. 1 of reference¹³).

Whether CSF-cNs are the direct mechanoreceptors or whether they receive input from other, unidentified mechanosensitive cells remains to be determined. CSF-cNs may also respond to chemical cues released in the central canal as a response to mechanical stimulation or muscle contraction. Although we cannot exclude a chemical hypothesis, it is unclear how a chemical signal would translate into the lateralized mechanoreponse described here. Ciliary calcium signalling may contribute to the sensory response of CSF-cNs as the PKD2L1 channel contributes to ciliary signalling²³. PKD2L1 positive CSF-cNs were previously shown to be pH sensitive^{14,15,24}, even though acid sensing ion channels (ASICs) carry most of the proton current^{14,25,26}. Our study here focused on the role of Pkd2l1 as, in the spinal cord, the channel is specific to CSF-cNs. However, as recently shown, PKD2L1 can

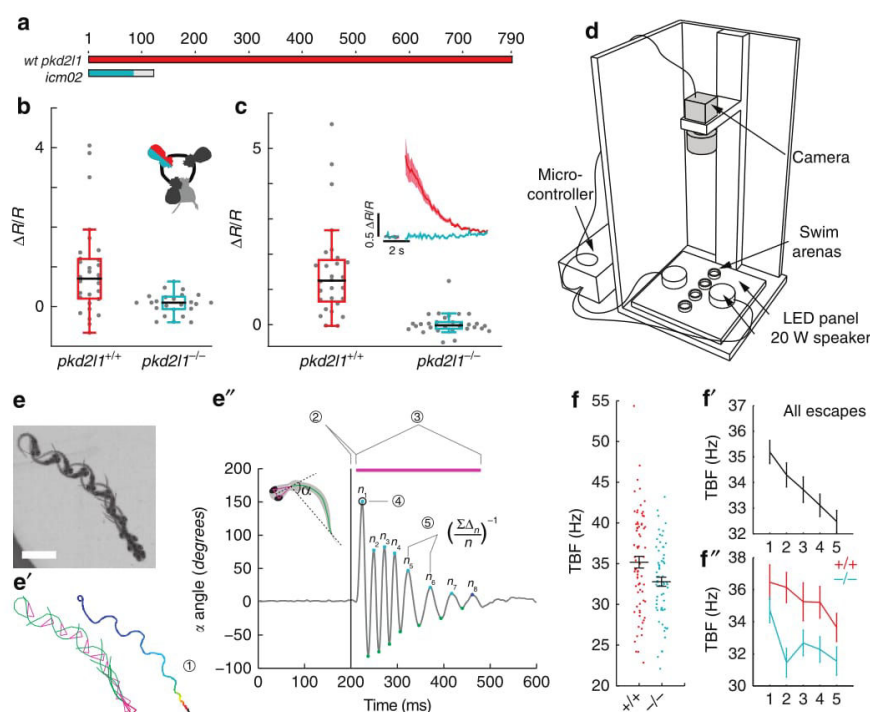


Figure 4 | Mutation of *pkd2l1* abolishes CSF-cN response to active and passive spinal bending and leads to a reduction of swimming frequency during the escape. (a) The TALEN-mediated 8 nucleotide deletion in exon two of *pkd2l1* leads to a frame shift and a premature stop in *pkd2l1^{icm02/icm02}*. (b,c) Calcium transients in dorsal ipsilateral CSF-cNs are abolished in *pkd2l1* mutant after mechanical stimulation during active (b) and passive (c) tail movement (same stimulation paradigms as used in Fig. 3, wildtype data sets for (b,c) are the same as in Fig. 3). Mutants (10 fish, 29 cells) in b show no response different from 0 ($P=0.13$) and are different from wildtypes (11 fish, 31 cells, $P=1.84 \times 10^{-2}$). Mutants (7 fish, 30 cells) in c show no response different from 0 ($P=0.22$) and are different from wildtypes (5 fish, 28 cells, $P=2.79 \times 10^{-4}$). (d) Experimental setup monitoring at high speed the escape response of freely swimming zebrafish larvae isolated in separated swim arenas triggered by 10 ms, 500 Hz acoustic stimuli. Acoustic stimuli were repeated five times per larva with 2 min inter trial interval. (e) Superimposed images showing a typical acoustic escape, (e') tracking corresponding to the same escape, swim distance (1) is derived from swim bladder position. (e'') kinematic analysis relied on the measure of the tail angle over time $\alpha(t)$ enabling to measure latency (2), escape duration (3), C-bend amplitude (4) and number of oscillations and TBF based on detection of subsequent peaks (5). Speed is derived from swim distance divided by duration. Scale bar: 2 mm. (f) Reduction of TBF in the *pkd2l1* mutant. Each point corresponds to a value per larva averaged over multiple trials ($P=0.0091$). (f') TBF decreased across trials in our paradigm ($P=1.7 \times 10^{-5}$). (f'') The reduction of TBF in the *pkd2l1* mutant (blue) compared to wild type is noticeable across trials ($P=0.039$).

trigger large calcium spikes upon changes of extracellular calcium, pH or membrane potential²⁴. This channel in CSF-cNs may therefore act in concert with other channels such as ASICs contributing to mouse and lamprey CSF-cN response *in vitro*^{14,16,27} in order to generate the bending response reported here *in vivo*.

Previously, we demonstrated that spinal CSF-cNs project directly onto locomotor CPGs involved in slow locomotion^{12,13}. During active locomotion, when muscles contract, we show that CSF-cNs can provide a proprioceptive GABAergic feedback that finely tunes the oscillatory frequency of the locomotor CPG. Further studies linking CSF-cNs to components of the fast locomotor CPG, and the escape circuit in particular, will be necessary to establish the neuronal basis for the modulation we report here. Although sensory feedback is not necessary for the oscillations, there are multiple lines of evidence suggesting that mechanosensory feedback can entrain the rhythm^{28,29} and gate locomotor transitions³⁰. The increased frequency of oscillations due to mechanosensory feedback observed here could contribute to the massive reduction of locomotor frequency in fictive compared to active locomotion observed in earlier studies^{31,32}.

Given that the CSF is a rich source of neuromodulators and endocrine signals, it could regulate the spontaneous firing of CSF-cNs via changes in pH or osmolarity via PKD2L1^{14–16,26} or via ATP signalling at the P2X2 receptor¹⁷. Tuning the spontaneous activity of CSF-cNs could modulate spontaneous locomotion in a similar manner as has recently been shown for sensory modulation of arousal locomotion in *C. elegans*³³. Altogether, our data demonstrate a novel role for CSF-cNs in the vertebrate spinal cord and opens new perspectives on a CSF sensory interface regulating locomotion.

Methods

Animal care. Animal handling and procedures were validated by the Institut du Cerveau et de la Moelle épinière, Paris and the French National Ethics Committee (Comité National de Réflexion Ethique sur l'Expérimentation Animale-Ce5/2011/056) in agreement with European Union legislation. Adults were reared at a maximal density of 8 animals per liter in a 14/10 (light/dark) cycle environment. Fish were fed live artemia twice a day and feeding regime was supplemented with solid extracts matching the fish developmental stage (ZM Systems, UK). Larvae were raised at 28.5 °C with a 14/10 day/night light cycle. All experiments were performed at room temperature on 5–6 dpf larvae unless stated otherwise.

Generation of transgenic lines. Transgenic lines used in this study are listed in Extended Data Table 1. To generate the *Tg(pk2l1:Gal4)icm10* transgenic line expressing the Gal4 reporter gene in *pk2l1* expressing cells, we amplified 3.8 kbp of genomic sequence immediately upstream of the predicted ATG site for the zebrafish *pk2l1* gene (ENSDARG00000022503)¹³ as well as 630 bp of highly conserved DNA in intron 2. Both DNA fragments were then subcloned into the pT2KspGFP plasmid using restriction digestion and T4 ligation. The 3.8 kb promoter region was cloned upstream the Gal4 (GFF) sequence in place of the *hsp* promoter, while the intronic sequence was placed after the Gal4 stop codon and before the SV40 poly-A site. For establishing the *Tg(pk2l1:GCaMP5G)icm07*, the same strategy was used by regular cloning with the PT2 vector containing the GCaMP5G construct. *Tg(UAS:GCaMP5G)icm08* was generated by regular cloning in the PT2-14xUAS vector. The transgenic line *Tg(UAS:GCaMP6f, cryaa:mCherry)icm06* was generated by subcloning GCaMP6f³⁴ into pDONR221 and then assembled into the final expression vector in a three-fragment gateway reaction (Invitrogen) using p5E-14xUAS, pME-GCaMP6f, p3E-poly(A) and pDest-cryaa:mCherry. The *Tg(pk2l1:tagRFP)icm17* was generated from *pk2l1:GCaMP5G* by swapping the GCaMP cassette for TagRFP using restriction digest/ligation as above. To generate a GFP reporter plasmid that expresses zebrafish optimized Botulinum Toxin Light Chain (BoTxLC) under Gal4/UAS control, we used an expression vector containing the Gal4-target sequence, 5x UAS, followed by an atg-less GFP. A gateway-compatible vector pT2UASMC5GW-R1R2 (5,419 bp) was used to generate the plasmid pT2UAS:BoTxLC-GFP encoding an in-frame fusion of GFP at the C-terminus of BoTxLC. The generation and validation of the UAS:BoTxLC-GFP line including the electrophysiological characterization that GABA vesicular release is abolished in *Tg(pk2l1:gal4;UAS:BoTxLC-GFP)icm21* transgenic larvae is detailed in a related publication²². The *Tg(mnx1:gal4)icm23* line was based on the injection of the *mnx1* construct previously described³⁵. Stable transgenic lines were established as previously described³⁶. Full complement of transgenic lines generated and used is shown in Supplemental Table 1.

Generation of *pk2l1* mutant. RNA coding for each TALEN monomer (left arm target: 5'-TGAGAAAGGATACAAAAG-3'; right arm target: 5'-TTGAG ATTTCCTTTGACG-3') was synthesized by the mMessage mMachine *in vitro* transcription kit (Invitrogen) and injected into *Tg(pk2l1:GCaMP5G)icm07* embryos at the one-cell stage. Injected embryos were pooled and a crude DNA extract was harvested. A DNA cassette containing the target site was amplified by PCR (primers: 5'-AGGGCAAGAGAATGGCAAGACG-3' and 5'-TGTGTGCTA GGACTGTGGGG-3'), and the resulting band was digested by SacI (Roche) to confirm disruption of the TALEN site. The mutant alleles were cloned by TOPO cloning (Invitrogen) and sequenced to determine the exact nature of the mutations. The *pk2l1^{icm02}* mutant allele is an 8 bp deletion in exon 2 causing a +1 frameshift that terminates in a stop codon within 35 amino acids of the substitution, upstream of the first transmembrane domain of Pkd2l1. All analysis was performed blind to genotype, which was only assayed at the conclusion of each experiment.

Fluorescent immunohistochemistry. Larvae were fixed in 4% PFA + 3% sucrose for 2 h at RT followed by 3X 5 min washes in PBS. After removing the skin, larvae were embedded in 4% low melting agarose and cut in 50 µm sagittal sections. Sections were mounted on a slide and blocked for 1 h in blocking solution (10% NGS, 1% DMSO, 0.5% Triton X100 in 0.1 M PBS). Slides were incubated with primary antibody over night at RT or for two days at 4 °C (1% NGS, 1% DMSO, 0.5% Triton X100 in 0.1 M PBS). After washing three times for 5 min in 0.1 M PBS with 0.5% TritonX100 (PBST), slides were incubated in the dark with the secondary antibody in PBST. After washing again three times for 5 min in PBST, mounting medium was added and slices were imaged on a confocal microscope using a 63X objective (FV-1000, Olympus). Antibodies used here were Anti-GFP (Abcam ab13970, dilution 1:500), anti-TagRFP (Life Technologies R10367, dilution 1:500), Alexa Fluor 568 goat anti-rabbit IgG (Invitrogen A11011, dilution 1:1,000), Alexa Fluor 488 goat anti-chicken IgG (Invitrogen A11039, dilution 1:1,000); all immunohistochemical samples were counterstained with DAPI (Invitrogen D3571, dilution 1:2,000).

Electron microscopy. To label CSF-cNs for electron microscopy, we injected 60 ng µl⁻¹ UAS:APEX2-tagRFP³⁷ into *Tg(pk2l1:Gal4)icm10* embryos at one-cell stage. 2.5 dpf larvae were then selected for CSF-cNs expression and euthanized in 0.2% trisacetic methanesulfonate (MS-222, Sigma-Aldrich, USA) prior to fixation in 2% glutaraldehyde in 100 mM sodium cacodylate. Briefly, whole embryos were shortly rinsed in Na-cacodylate, to which glycine was added, and rinsed again in Na-cacodylate buffer. The peroxidase was further revealed in DAB (0.5 mg ml⁻¹), to which 10 mM hydrogen peroxide was added³⁷. After 5 min, the reaction was stopped by rinsing embryos in Na-cacodylate before postfixation in 2% osmium tetroxide and processed for standard electron microscopy. For sectioning, embryos were oriented in the coronal plane. Stained cells were identified in a widefield microscope and regions of interest serially cut on copper grids to be further observed in a HITACHI 120 kV HT 7700 electron microscope.

Calcium imaging setup. 5–6 dpf larvae were anesthetized in 0.02% MS-222 diluted in fish facility water and then mounted in glass-bottom dishes (MatTek, Ashland, Massachusetts, USA) filled with 1.5% low-melting point agarose. Calcium imaging was performed on a two photon microscope (2p-vivo, Intelligent Imaging Innovations, Inc., Denver) using a 20x objective. Images were acquired at 9 Hz (external pressure and paralyzed fish experiments) or 11 Hz (tail-free experiments).

Experiments on paralyzed larvae: Larvae were paralyzed by injecting 0.5 nl of 0.5 mM α-Bungarotoxin in the ventral axial musculature (Tocris, Bristol, UK) and placed in artificial cerebrospinal fluid solution (ACSF, concentrations in mM: 134 NaCl, 2.9 KCl, 1.2 MgCl₂, 10 HEPES, 10 glucose and 2.1 CaCl₂; 290 mOsm, adjusted to pH 7.7–7.8 with NaOH)³⁸. Ventral nerve root activity was recorded with ~10 µm tip diameter pipettes filled with ACSF. Data was recorded in current clamp mode using a MultiClamp 700 A amplifier (Molecular Devices–Axon Instruments, USA), a Digidata series 1322 A digitizer (Axon Instruments, USA) and pClamp 8.2 software (Axon instruments, USA). 10 ms water jet at 8–15 psi delivered through a pipette located in the vicinity of the otic vesicle induced reliable escapes.

Experiments on tail-free larvae: larvae were embedded dorsally and the tail was freed so that ~1/3 remained fixed in the agar. An 890 nm LED was used to illuminate the fish from the side and behavior was recorded at 675 Hz from the bottom through a 5x microscope objective and imaged on the camera with a 50 mm camera objective. Infrared light from the imaging laser was blocked with a filter located in front of the camera. Behavior was induced as described above. Three trials per imaging position were recorded at ~1 min inter trial intervals and the one with the clearest unilateral escape responses were selected for analysis. For each larva, one dorsal and one ventral Z-position was recorded with the agar boundary of the free tail placed roughly in the middle of the field of view to compromise between tail deflection and motion artifact. After each experiment, the larva was anesthetized with 0.02% MS-222 and a high-resolution z-stack was recorded to identify cell body position.

Passive tail bending experiments: After embedding, roughly half of the larval tail was freed unilaterally to provide access to a blunt 50 µm diameter glass probe. Probe deflections were driven with a mechanotransducer device controlled through LabView software. Tail bend was induced by probe deflection and repeated 10 times.

Calcium imaging data analysis. Regions of interest around CSF-cN cell bodies expressing both GCaMP and tagRFP were manually defined in the software ImageJ. To extract raw fluorescence from the image time series in moving samples (either due to passive mechanical stimulation or active tail movement), the ROIs were shifted with the moving cell using a custom-made tracking algorithm based on cross correlation. In frames where the normalized correlation coefficient fell underneath a given threshold due to the movement artifact, data were omitted for that given frame. This threshold was manually adjusted to minimize tracking errors for each imaging time series. Any $\Delta F/F$ values were calculated as $\Delta F/F = (F(t) - F_0)/(F_0 - F_{bg})$ with F_0 being the average fluorescence of the 10 first frames and F_{bg} being the average fluorescence of the 10 first frames in a region with only background signal. To correct for motion artifacts due to cells moving out of the focal plane or incomplete tracking, we calculated the ratio of the GCaMP and the tagRFP signal ($\Delta R/R$) as $\Delta R/R = (F_{GCaMP}(t) \times F_{0tagRFP}/F_{0GCaMP} \times F_{tagRFP}(t)) - 1$ with F_0 being the average fluorescence of the 10 first frames and both F_0 and $F(t)$ background corrected signals (see Fig. 3a' and extended data Figure S2). Difference in $\Delta R/R$ due to stimulation was then calculated as $\text{difference} = (\Delta R/R)_{\text{post}} - (\Delta R/R)_{\text{pre}}$ with $(\Delta R/R)_{\text{pre}}$ and $(\Delta R/R)_{\text{post}}$ being the average five time points before and after stimulation (external pressure experiments) or the average three time points before and after tail bending (tail free experiments). During motion of the tail, cells can usually not be imaged so between $(\Delta R/R)_{\text{pre}}$ and $(\Delta R/R)_{\text{post}}$ there is a gap of 4 frames (external pressure experiment) or 1 frame (tail free experiment).

Behavior setup and analysis. To assess fish behavior, a video recording chamber was constructed. 6 dpf larvae from *Tg(pkcd2l1:icm02/+,-pkcd2l1:GCaMP5G)* or *Tg(pkcd2l1:Gal4, UAS:BoTx-BLC-GFP)* incrosses were placed in circular swim arenas (Biopatch) atop a plexiglass plate. An Arduino Due software circuit (Arduino) was used to trigger one second-long recordings on a high speed camera (Vieworks) operating at 650 fps. To trigger escape behaviours, a 500 Hz, 10 ms sine wave stimulus was played through a 20 W amplifier (Adafruit) over two speakers (Monacor) at maximum volume 200 ms into the recording. Larvae were allowed to acclimate to the chamber for 5 min prior to the experiment and were allowed to rest 2 min between each stimulus. Each larva was subjected to five trials, the properties of which were averaged within subjects. Larvae were genotyped after data acquisition. Sample sizes were selected by performing a Levene analysis on preliminary data sets with a threshold of 5% difference in means.

Two software packages were used to extract kinematic parameters from high-speed videos of fish during escapes. The first executable, developed in collaboration with R&D Vision, fits a series of eleven points to the fish body—two for each of the eyes, one for the swim bladder, and the remaining eight to the tail. The program identifies these 11 points for each of the 650 frames of video. The second executable consists of a series of MATLAB scripts that derives kinetic parameters from these points. First, the tail angle α was calculated by identifying the midpoint of the two eye points, then calculating a head direction vector based on this point and the center of the swim bladder. A tail direction vector based on the last point on the tail minus the swim bladder was then calculated, and α was calculated as the inverse cosine of these two vector lengths. To remove noise due to tracking error, the raw trace of α was median-filtered, linearly interpolated, and finally smoothed using a smoothing spline method in MATLAB's Curve Fitting Toolbox. Simultaneously, the unprocessed tail angle trace was differentiated to estimate angular velocity, which was smoothed as above. The angular velocity trace was used to define frames of movement—any frame with angular velocity $> 1^\circ/\text{frame}$ is defined as 'in motion'. These frames were then connected into movement bouts using morphological closing with a block size of 10 frames. Detected bouts were then assigned priority weights by their distance from stimulus onset to identify the escape. Non-escape bouts were discarded and not analyzed. Next, local maxima and minima corresponding to bend amplitudes were identified within the escape bout by calculation of regional differences (see <http://www.billauer.co.il/peakdet.html>). These peaks were then marked as sequential (if followed by a peak in the opposite direction, that is, a bend to the other side) or non-sequential. A hemicycle for each peak was calculated as the number of frames from each peak to the next in sequence, regardless of direction. Each peak therefore receives a maximum angle, hemicycle duration, and is flagged as being either positive (in the direction of the C-bend) or negative (in the direction of the counter bend), and either sequential or non-sequential. To improve accuracy, detection of the bout onset was modified from the initial approach, which used a single threshold of the angular velocity. A straight line segment was fit to the ascending phase of the first bend, and a single-term exponential fit was calculated between this segment and a straight horizontal segment at $y=0$ between the stimulus onset and the intercept of the rising phase segment with the y -axis. Escape onset was then defined as the frame at which the resulting exponential fit is greater than a 2° threshold. This method reliably identified a movement onset that was independent of the amplitude of the first bend as well as resistant to noise in the angle trace. From these data, escapes were defined as those movement bouts occurring within 30 ms of the stimulus and in which the first bend is $> 60^\circ$ (that is, a C-bend). Movement bouts failing these inclusion criteria are not analyzed. From the inputs described above we are able to derive latency (escape onset-time of stimulus presentation), bout duration (escape offset-escape onset), bout distance (travel distance of point corresponding to the swim bladder), bout speed (distance/duration), number of oscillations (sum of detected positive and negative peaks/2), C-bend amplitude

(amplitude of first detected peak). Mean tail beat frequency was calculated as the inverse of two times the mean hemicycle of sequential peaks. MATLAB package is available on request; the tail-tracking algorithm is property of R&D Vision.

Zebrafish dissociated cell cultures. Dissociated cells from 4 dpf *Tg(pkcd2l1:GCaMP5G)* and *Tg(pkcd2l1:tagRFP)* larvae were plated acutely in zebrafish ACSF (see Ca^{2+} imaging section) on plastic petri dishes based on a published protocol³⁹. Dissociated CSF-cNs were identified based on their GCaMP5G/tagRFP fluorescence and imaged three hours after plating.

Statistical analysis. All values are mean $\pm$ s.e.m. unless otherwise noted. To account for the fact that all our data are intrinsically nested (multiple measurements of cells and/or repetition of trials within the same fish), we used linear mixed models for all our hypothesis testing⁴⁰. All models were calculated using the freeware R using the 'nlme' package. Detailed analysis scripts are available upon request.

References

- Martin, F. H., Seoane, J. R. & Baile, C. A. Feeding in satiated sheep elicited by intraventricular injections of CSF from fasted sheep. *Life Sci.* **13**, 177–184 (1973).
- Pappenheimer, J. R., Miller, T. B. & Goodrich, C. A. Sleep-promoting effects of cerebrospinal fluid from sleep-deprived goats. *Proc. Natl Acad. Sci. USA* **58**, 513–517 (1967).
- Lehtinen, M. K. *et al.* The cerebrospinal fluid provides a proliferative niche for neural progenitor cells. *Neuron* **69**, 893–905 (2011).
- Sawamoto, K. *et al.* New neurons follow the flow of cerebrospinal fluid in the adult brain. *Science* **311**, 629–632 (2006).
- Xie, L. *et al.* Sleep drives metabolite clearance from the adult brain. *Science* **342**, 373–377 (2013).
- Agduhr, E. Über ein zentrales Sinnesorgan (?) bei den Vertebraten. *Z. Anat. Entwickl. Gesch.* **66**, 223–360 (1922).
- Djenoune, L. *et al.* Investigation of spinal cerebrospinal fluid-contacting neurons expressing PKD2L1: Evidence for a conserved system from fish to primates. *Front. Neuroanat.* **8**, 26 (2014).
- Orts-Del'Immagine, A. *et al.* Morphology, distribution and phenotype of polycystin kidney disease 2-like 1-positive cerebrospinal fluid contacting neurons in the brainstem of adult mice. *PLoS ONE* **9**, e87748 (2014).
- Vigh, B. & Vigh-Teichmann, I. Actual problems of the cerebrospinal fluid-contacting neurons. *Microsc. Res. Tech.* **41**, 57–83 (1998).
- Kolmer, W. Das 'Sagittallorgan' der Wirbeltiere. *Z. Anat. EntwGesch.* **60**, 652–717 (1921).
- Dale, N., Roberts, A., Ottersen, O. P. & Mathisen, J. S. The morphology and distribution of 'Kolmer–Agduhr cells', a class of cerebrospinal-fluid-contacting neurons revealed in the frog embryo spinal cord by GABA immunocytochemistry. *Proc. R. Soc. Lond. B: Biol. Sci.* **232**, 193–203 (1987).
- Wyart, C. *et al.* Optogenetic dissection of a behavioural module in the vertebrate spinal cord. *Nature* **461**, 407–410 (2009).
- Fidelin, K. *et al.* State-dependent modulation of locomotion by GABAergic spinal sensory neurons. *Curr. Biol.* **25**, 3035–3047 (2015).
- Bushman, J. D., Ye, W. & Liman, E. R. A proton current associated with sour taste: Distribution and functional properties. *FASEB J.* **29**, 3014–3026 (2015).
- Huang, A. L. *et al.* The cells and logic for mammalian sour taste detection. *Nature* **442**, 934–938 (2006).
- Orts Del'Immagine, A. *et al.* Properties of subependymal cerebrospinal fluid contacting neurones in the dorsal vagal complex of the mouse brainstem. *J. Physiol. (Lond.)* **590**, 3719–3741 (2012).
- Stoeckel, M.-E. *et al.* Cerebrospinal fluid-contacting neurons in the rat spinal cord, a ?-aminobutyric acidergic system expressing the P2X2 subunit of purinergic receptors, PSA-NCAM, and GAP-43 immunoreactivities: Light and electron microscopic study. *J. Comp. Neurol.* **457**, 159–174 (2003).
- Kolmer, W. Über das Sagittallorgan, ein zentrales Sinnesorgan der Wirbeltiere, insbesondere beim Affen. *Z. Zellforsch.* **13**, 236–248 (1931).
- Masino, M. A. & Fetcho, J. R. Fictive swimming motor patterns in wild type and mutant larval zebrafish. *J. Neurophysiol.* **93**, 3177–3188 (2005).
- McLean, D. L., Fan, J., Higashijima, S.-I., Hale, M. E. & Fetcho, J. R. A topographic map of recruitment in spinal cord. *Nature* **446**, 71–75 (2007).
- Schredelseker, J. *et al.* The beta 1a subunit is essential for the assembly of dihydropyridine-receptor arrays in skeletal muscle. *Proc. Natl Acad. Sci. USA* **102**, 17219–17224 (2005).
- Sternberg, J. R. *et al.* Optimization of a neurotoxin to investigate the contribution of excitatory interneurons to speed modulation *in vivo* (in preparation).
- Delling, M., DeCaen, P. G., Doerner, J. F., Febvay, S. & Clapham, D. E. Primary cilia are specialized calcium signalling organelles. *Nature* **504**, 311–314 (2013).

24. Hu, M., Liu, Y., Wu, J. & Liu, X. Influx-operated Ca^{2+} entry via PKD2-L1 and PKD1-L3 channels facilitates sensory responses to polymodal transient stimuli. *Cell Reports* **13**, 798–811 (2015).
25. Chang, R. B., Waters, H. & Liman, E. R. A proton current drives action potentials in genetically identified sour taste cells. *Proc. Natl Acad. Sci. USA* **107**, 22320–22325 (2010).
26. Orts-Del'Immagine, A. *et al.* A single polycystic kidney disease 2-like 1 channel opening acts as a spike generator in cerebrospinal fluid-contacting neurons of adult mouse brainstem. *Neuropharmacology* **101**, 549–565 (2015).
27. Jalalvand, E., Robertson, B., Wallén, P. & Grillner, S. Ciliated neurons lining the central canal sense both fluid movement and pH through ASIC3. *Nat. Commun.* **7**, 10002 (2016).
28. Grillner, S., McClellan, A. & Perret, C. Entrainment of the spinal pattern generators for swimming by mechano-sensitive elements in the lamprey spinal cord in vitro. *Brain Res.* **217**, 380–386 (1981).
29. Borgmann, A., Hooper, S. L. & Büschges, A. Sensory feedback induced by front-leg stepping entrains the activity of central pattern generators in caudal segments of the stick insect walking system. *J. Neurosci.* **29**, 2972–2983 (2009).
30. Pearson, K. G. Role of sensory feedback in the control of stance duration in walking cats. *Brain Res. Rev.* **57**, 222–227 (2008).
31. Wallén, P. & Williams, T. L. Fictive locomotion in the lamprey spinal cord in vitro compared with swimming in the intact and spinal animal. *J. Physiol. (Lond.)* **347**, 225–239 (1984).
32. Sqalli-Houssaini, Y., Cazalets, J. R., Fabre, J. C. & Clarac, F. A cooling/heating system for use with in vitro preparations: study of temperature effects on newborn rat rhythmic activities. *J. Neurosci. Methods* **39**, 131–139 (1991).
33. Choi, S. *et al.* Sensory neurons arouse *C. elegans* locomotion via both glutamate and neuropeptide release. *PLoS Genet.* **11**, e1005359 (2015).
34. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
35. Zelenchuk, T. A. & Brusés, J. L. In Vivo labeling of zebrafish motor neurons using an mnx1 enhancer and Gal4/UAS. *Genesis* **49**, 546–554 (2011).
36. Fisher, S. *et al.* Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat. Protoc.* **1**, 1297–1305 (2006).
37. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Meth.* **12**, 51–54 (2015).
38. Saint-Amant, L. & Drapeau, P. Whole-cell patch-clamp recordings from identified spinal neurons in the zebrafish embryo. *Methods Cell Sci.* **25**, 59–64 (2003).
39. Joselevitch, C. & Zenisek, D. Imaging exocytosis in retinal bipolar cells with TIRF microscopy. *J. Vis. Exp.* **28**, doi:103791/1305 (2009).
40. Aarts, E., Verhage, M., Veenfliet, J. V., Dolan, C. V. & van der Sluis, S. A solution to dependency: using multilevel analysis to accommodate nested data. *Nat. Neurosci.* **17**, 491–496 (2014).

Acknowledgements

We thank Natalia Maties and Bogdan Buzurin from Animalliance for expert fish care. We thank D. Delpeuch for help with culture, D. Langui for EM imaging, RD Vision for high-speed data acquisition and tracking. We thank Dr Kilborn for customization of 2-photon setup, Vincent Guillemot and Ivan Moszer from the ICM Biostatistics/Bioinformatics core facility for advices on statistics and the imaging facility PICPS from ICM supported by 'Investissements d'avenir' ANR-10-IAIHU-06. This work received support from the École des Neurosciences de Paris Ile-de-France (ENP, Chaire d'excellence for CW and doctoral fellowship for UB), the Fondation pour la Recherche Médicale (FRM), the Bettencourt-Schueller Foundation with the ATIP/Avenir program organized by INSERM and CNRS, the ERC starting grant 'Optoloco' # 311673, the HFSP research grant # RGP0063/2014, the ANR-II-INSB-0014, the ANR-10-IAIHU-06 and a Research-in-Paris fellowship with an EMBO fellowship # ALTF 549–2013 for AP.

Author contributions

U.B. performed calcium imaging and electrophysiology experiments. A.P. and S.K. performed behavioural experiments. A.P. and L.D. generated the mutants with help of M.C. and J.P.C. with TALENs. S.N.F. performed IHC, cultures, recorded beating cilia and contributed to EM data. J.P.R. acquired the EM data with the help of L.D. and J.G. C.S. recorded cilia beating *in vivo*. A.P., S.N.F., M.S., K.K. and F.D.B. generated transgenic animals. U.B., A.P. and C.W. designed experiments, analyzed data and wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Böhm, U. L. *et al.* CSF-contacting neurons regulate locomotion by relaying mechanical stimuli to spinal circuits. *Nat. Commun.* **7**:10866 doi: 10.1038/ncomms10866 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

5 Bibliographie

- Agne, M., Blank, I., Emhardt, A.J., Gäbelein, C.G., Gawlas, F., Gillich, N., Gonschorek, P., Juretschke, T.J., Krämer, S.D., Louis, N., Müller, A., Rudolf, A., Schäfer, L.M., Scheidmann, M.C., Schmunk, L.J., Schwenk, P.M., Stammnitz, M.R., Warmer, P.M., Weber, W., Fischer, A., Kaufmann, B., Wagner, H.J., Radziwill, G., 2014. Modularized CRISPR/dCas9 effector toolkit for target-specific gene regulation. *ACS Synth. Biol.* 3, 986–989. doi:10.1021/sb500035y
- Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J.-C., Stricher, F., Petit, A.-S., Patin, A., Guillier, S., Rolland, S., Prieto, J., Blanco, F.J., Bravo, J., Montoya, G., Serrano, L., Duchateau, P., Pâques, F., 2006. Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.* 355, 443–458. doi:10.1016/j.jmb.2005.10.065
- Audebert, M., Salles, B., Calsou, P., 2004. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J. Biol. Chem.* 279, 55117–55126. doi:10.1074/jbc.M404524200
- Bacman, S.R., Williams, S.L., Pinto, M., Peralta, S., Moraes, C.T., 2013. Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs. *Nat. Med.* 19, 1111–1113. doi:10.1038/nm.3261
- Bakr, A., Köcher, S., Volquardsen, J., Petersen, C., Borgmann, K., Dikomey, E., Rothkamm, K., Mansour, W.Y., 2016. Impaired 53BP1/RIF1 DSB mediated end-protection stimulates CtIP-dependent end resection and switches the repair to PARP1-dependent end joining in G1. *Oncotarget*. doi:10.18632/oncotarget.11023
- Ball, H.L., Myers, J.S., Cortez, D., 2005. ATRIP binding to replication protein A-single-stranded DNA promotes ATR-ATRIP localization but is dispensable for Chk1 phosphorylation. *Mol. Biol. Cell* 16, 2372–2381. doi:10.1091/mbc.E04-11-1006

- Bau, D.-T., Fu, Y.-P., Chen, S.-T., Cheng, T.-C., Yu, J.-C., Wu, P.-E., Shen, C.-Y., 2004. Breast cancer risk and the DNA double-strand break end-joining capacity of nonhomologous end-joining genes are affected by BRCA1. *Cancer Res.* 64, 5013–5019. doi:10.1158/0008-5472.CAN-04-0403
- Bennardo, N., Stark, J.M., 2010. ATM limits incorrect end utilization during non-homologous end joining of multiple chromosome breaks. *PLoS Genet.* 6, e1001194. doi:10.1371/journal.pgen.1001194
- Beumer, K.J., Trautman, J.K., Bozas, A., Liu, J.-L., Rutter, J., Gall, J.G., Carroll, D., 2008. Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc. Natl. Acad. Sci. U. S. A.* 105, 19821–19826. doi:10.1073/pnas.0810475105
- Bhargava, R., Onyango, D.O., Stark, J.M., 2016. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet. TIG* 32, 566–575. doi:10.1016/j.tig.2016.06.007
- Bitinaite, J., Wah, D.A., Aggarwal, A.K., Schildkraut, I., 1998. FokI dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10570–10575.
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., Bonas, U., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi:10.1126/science.1178811
- Bogdanove, A.J., Schornack, S., Lahaye, T., 2010. TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* 13, 394–401. doi:10.1016/j.pbi.2010.04.010
- Böhm, U.L., Prendergast, A., Djenoune, L., Nunes Figueiredo, S., Gomez, J., Stokes, C., Kaiser, S., Suster, M., Kawakami, K., Charpentier, M., Concordet, J.-P., Rio, J.-P., Del Bene, F., Wyart, C., 2016. CSF-contacting neurons regulate locomotion by relaying mechanical stimuli to spinal circuits. *Nat. Commun.* 7, 10866. doi:10.1038/ncomms10866
- Burma, S., Chen, B.P., Murphy, M., Kurimasa, A., Chen, D.J., 2001. ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *J. Biol. Chem.* 276, 42462–42467. doi:10.1074/jbc.C100466200
- Cen, Y., Fiori, A., Van Dijck, P., 2015. Deletion of the DNA Ligase IV Gene in *Candida*

glabrata Significantly Increases Gene-Targeting Efficiency. *Eukaryot. Cell* 14, 783–791. doi:10.1128/EC.00281-14

Cermak, T., Starker, C.G., Voytas, D.F., 2015. Efficient design and assembly of custom TALENs using the Golden Gate platform. *Methods Mol. Biol. Clifton NJ* 1239, 133–159. doi:10.1007/978-1-4939-1862-1_7

Cerqueira, A., Santamaría, D., Martínez-Pastor, B., Cuadrado, M., Fernández-Capetillo, O., Barbacid, M., 2009. Overall Cdk activity modulates the DNA damage response in mammalian cells. *J. Cell Biol.* 187, 773–780. doi:10.1083/jcb.200903033

Certo, M.T., Gwiazda, K.S., Kuhar, R., Sather, B., Curinga, G., Mandt, T., Brault, M., Lambert, A.R., Baxter, S.K., Jacoby, K., Ryu, B.Y., Kiem, H.-P., Gouble, A., Paques, F., Rawlings, D.J., Scharenberg, A.M., 2012. Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nat. Methods* 9, 973–975. doi:10.1038/nmeth.2177

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., Ter-Ovanesyan, D., Braff, J.L., Davidsohn, N., Housden, B.E., Perrimon, N., Weiss, R., Aach, J., Collins, J.J., Church, G.M., 2015. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* 12, 326–328. doi:10.1038/nmeth.3312

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., Huang, B., 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491. doi:10.1016/j.cell.2013.12.001

Chen, L., Nievera, C.J., Lee, A.Y.-L., Wu, X., 2008. Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair. *J. Biol. Chem.* 283, 7713–7720. doi:10.1074/jbc.M710245200

Chen, M.-J., Ma, S.-M., Dumitrache, L.C., Hasty, P., 2007. Biochemical and cellular characteristics of the 3' -> 5' exonuclease TREX2. *Nucleic Acids Res.* 35, 2682–2694. doi:10.1093/nar/gkm151

Chen, P.-L., Liu, F., Cai, S., Lin, X., Li, A., Chen, Y., Gu, B., Lee, E.Y.-H.P., Lee, W.-H., 2005. Inactivation of CtIP leads to early embryonic lethality mediated by G1 restraint and to

tumorigenesis by haploid insufficiency. *Mol. Cell. Biol.* 25, 3535–3542. doi:10.1128/MCB.25.9.3535-3542.2005

Cheng, C.Y., Gilson, T., Dallaire, F., Ketner, G., Branton, P.E., Blanchette, P., 2011. The E4orf6/E1B55K E3 ubiquitin ligase complexes of human adenoviruses exhibit heterogeneity in composition and substrate specificity. *J. Virol.* 85, 765–775. doi:10.1128/JVI.01890-10

Chevalier, B.S., Stoddard, B.L., 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* 29, 3757–3774.

Choulika, A., Perrin, A., Dujon, B., Nicolas, J.F., 1995. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 15, 1968–1973.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., Kühn, R., 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33, 543–548. doi:10.1038/nbt.3198

Cottarel, J., Frit, P., Bombarde, O., Salles, B., Négrel, A., Bernard, S., Jeggo, P.A., Lieber, M.R., Modesti, M., Calsou, P., 2013. A noncatalytic function of the ligation complex during nonhomologous end joining. *J. Cell Biol.* 200, 173–186. doi:10.1083/jcb.201203128

Curtin, S.J., Zhang, F., Sander, J.D., Haun, W.J., Starker, C., Baltes, N.J., Reyon, D., Dahlborg, E.J., Goodwin, M.J., Coffman, A.P., Dobbs, D., Joung, J.K., Voytas, D.F., Stupar, R.M., 2011. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant Physiol.* 156, 466–473. doi:10.1104/pp.111.172981

Davies, O.R., Forment, J.V., Sun, M., Belotserkovskaya, R., Coates, J., Galanty, Y., Demir, M., Morton, C.R., Rzechorzek, N.J., Jackson, S.P., Pellegrini, L., 2015. CtIP tetramer assembly is required for DNA-end resection and repair. *Nat. Struct. Mol. Biol.* 22, 150–157. doi:10.1038/nsmb.2937

Davis, A.J., Chen, D.J., 2013. DNA double strand break repair via non-homologous end-joining. *Transl. Cancer Res.* 2, 130–143. doi:10.3978/j.issn.2218-676X.2013.04.02

Davis, A.J., Chi, L., So, S., Lee, K.-J., Mori, E., Fattah, K., Yang, J., Chen, D.J., 2014. BRCA1 modulates the autophosphorylation status of DNA-PKcs in S phase of the cell cycle.

Nucleic Acids Res. 42, 11487–11501. doi:10.1093/nar/gku824

de Silva, U., Perrino, F.W., Hollis, T., 2009. DNA binding induces active site conformational change in the human TREX2 3'-exonuclease. Nucleic Acids Res. 37, 2411–2417. doi:10.1093/nar/gkp025

DeKelfer, R.C., Choi, V.M., Moehle, E.A., Paschon, D.E., Hockemeyer, D., Meijnsing, S.H., Sancak, Y., Cui, X., Steine, E.J., Miller, J.C., Tam, P., Bartsevich, V.V., Meng, X., Rupniewski, I., Gopalan, S.M., Sun, H.C., Pitz, K.J., Rock, J.M., Zhang, L., Davis, G.D., Rebar, E.J., Cheeseman, I.M., Yamamoto, K.R., Sabatini, D.M., Jaenisch, R., Gregory, P.D., Urnov, F.D., 2010. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Res. 20, 1133–1142. doi:10.1101/gr.106773.110

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607. doi:10.1038/nature09886

Deng, S.K., Gibb, B., de Almeida, M.J., Greene, E.C., Symington, L.S., 2014. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. Nat. Struct. Mol. Biol. 21, 405–412. doi:10.1038/nsmb.2786

Desjarlais, J.R., Berg, J.M., 1992. Toward rules relating zinc finger protein sequences and DNA binding site preferences. Proc. Natl. Acad. Sci. U. S. A. 89, 7345–7349.

Dever, S.M., Golding, S.E., Rosenberg, E., Adams, B.R., Idowu, M.O., Quillin, J.M., Valerie, N., Xu, B., Povirk, L.F., Valerie, K., 2011. Mutations in the BRCT binding site of BRCA1 result in hyper-recombination. Aging 3, 515–532. doi:10.18632/aging.100325

Dohrn, L., Salles, D., Siehler, S.Y., Kaufmann, J., Wiesmüller, L., 2012. BRCA1-mediated repression of mutagenic end-joining of DNA double-strand breaks requires complex formation with BACH1. Biochem. J. 441, 919–926. doi:10.1042/BJ20110314

Doyon, Y., Choi, V.M., Xia, D.F., Vo, T.D., Gregory, P.D., Holmes, M.C., 2010. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. Nat. Methods 7, 459–460. doi:10.1038/nmeth.1456

Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R.,

- Hocking, T.D., Zhang, L., Rebar, E.J., Gregory, P.D., Urnov, F.D., Amacher, S.L., 2008. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.* 26, 702–708. doi:10.1038/nbt1409
- Dujon, B., 1989. Group I introns as mobile genetic elements: facts and mechanistic speculations--a review. *Gene* 82, 91–114.
- Dumitrache, L.C., Hu, L., Son, M.Y., Li, H., Wesevich, A., Scully, R., Stark, J., Hasty, P., 2011. Trex2 enables spontaneous sister chromatid exchanges without facilitating DNA double-strand break repair. *Genetics* 188, 787–797. doi:10.1534/genetics.111.129833
- Dynan, W.S., Yoo, S., 1998. Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic Acids Res.* 26, 1551–1559.
- Epinat, J.-C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Pâques, F., Lacroix, E., 2003. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.* 31, 2952–2962.
- Escribano-Díaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T.F., Tkáč, J., Cook, M.A., Rosebrock, A.P., Munro, M., Canny, M.D., Xu, D., Durocher, D., 2013. A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol. Cell* 49, 872–883. doi:10.1016/j.molcel.2013.01.001
- Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T., Rhodes, D., 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* 366, 483–487. doi:10.1038/366483a0
- Fattah, F.J., Lichter, N.F., Fattah, K.R., Oh, S., Hendrickson, E.A., 2008. Ku70, an essential gene, modulates the frequency of rAAV-mediated gene targeting in human somatic cells. *Proc. Natl. Acad. Sci. U. S. A.* 105, 8703–8708. doi:10.1073/pnas.0712060105
- Fine, E.J., Cradick, T.J., Zhao, C.L., Lin, Y., Bao, G., 2014. An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic Acids Res.* 42, e42. doi:10.1093/nar/gkt1326
- Forrester, N.A., Sedgwick, G.G., Thomas, A., Blackford, A.N., Speiseder, T., Dobner, T., Byrd, P.J., Stewart, G.S., Turnell, A.S., Grand, R.J.A., 2011. Serotype-specific inactivation of

the cellular DNA damage response during adenovirus infection. *J. Virol.* 85, 2201–2211. doi:10.1128/JVI.01748-10

Fradet-Turcotte, A., Canny, M.D., Escibano-Díaz, C., Orthwein, A., Leung, C.C.Y., Huang, H., Landry, M.-C., Kitevski-LeBlanc, J., Noordermeer, S.M., Sicheri, F., Durocher, D., 2013. 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature* 499, 50–54. doi:10.1038/nature12318

Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaepfel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G., Holmes, M.C., Gregory, P.D., Glimm, H., Schmidt, M., Naldini, L., von Kalle, C., 2011. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* 29, 816–823. doi:10.1038/nbt.1948

Gammage, P.A., Gaude, E., Van Haute, L., Rebelo-Guimar, P., Jackson, C.B., Rorbach, J., Pekalski, M.L., Robinson, A.J., Charpentier, M., Concordet, J.-P., Frezza, C., Minczuk, M., 2016. Near-complete elimination of mutant mtDNA by iterative or dynamic dose-controlled treatment with mtZFNs. *Nucleic Acids Res.* doi:10.1093/nar/gkw676

Gandía, M., Xu, S., Font, C., Marcos, J.F., 2016. Disruption of ku70 involved in non-homologous end-joining facilitates homologous recombination but increases temperature sensitivity in the phytopathogenic fungus *Penicillium digitatum*. *Fungal Biol.* 120, 317–323. doi:10.1016/j.funbio.2015.11.001

Gao, F., Shen, X.Z., Jiang, F., Wu, Y., Han, C., 2016. DNA-guided genome editing using the *Natronobacterium gregoryi* Argonaute. *Nat. Biotechnol.* 34, 768–773. doi:10.1038/nbt.3547

Gao, X., Tsang, J.C.H., Gaba, F., Wu, D., Lu, L., Liu, P., 2014. Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Res.* 42, e155. doi:10.1093/nar/gku836

Greco, G.E., Matsumoto, Y., Brooks, R.C., Lu, Z., Lieber, M.R., Tomkinson, A.E., 2016. SCR7 is neither a selective nor a potent inhibitor of human DNA ligase IV. *DNA Repair* 43, 18–23. doi:10.1016/j.dnarep.2016.04.004

Grizot, S., Epinat, J.-C., Thomas, S., Duclert, A., Rolland, S., Pâques, F., Duchateau, P., 2010. Generation of redesigned homing endonucleases comprising DNA-binding domains

derived from two different scaffolds. *Nucleic Acids Res.* 38, 2006–2018. doi:10.1093/nar/gkp1171

Guilinger, J.P., Pattanayak, V., Reyon, D., Tsai, S.Q., Sander, J.D., Joung, J.K., Liu, D.R., 2014. Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat. Methods* 11, 429–435. doi:10.1038/nmeth.2845

Guirouilh-Barbat, J., Lambert, S., Bertrand, P., Lopez, B.S., 2014. Is homologous recombination really an error-free process? *Front. Genet.* 5, 175. doi:10.3389/fgene.2014.00175

Guo, R., Wan, Y., Xu, D., Cui, L., Deng, M., Zhang, G., Jia, R., Zhou, W., Wang, Z., Deng, K., Huang, M., Wang, F., Zhang, Y., 2016. Generation and evaluation of Myostatin knock-out rabbits and goats using CRISPR/Cas9 system. *Sci. Rep.* 6, 29855. doi:10.1038/srep29855

Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G.F., Chin, L., 2016. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Rep.* 14, 1555–1566. doi:10.1016/j.celrep.2016.01.019

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J.-S., Concordet, J.-P., 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 17, 148. doi:10.1186/s13059-016-1012-2

Harper, J.W., Elledge, S.J., 2007. The DNA damage response: ten years after. *Mol. Cell* 28, 739–745. doi:10.1016/j.molcel.2007.11.015

Haun, W., Coffman, A., Clasen, B.M., Demorest, Z.L., Lowy, A., Ray, E., Retterath, A., Stoddard, T., Juillerat, A., Cedrone, F., Mathis, L., Voytas, D.F., Zhang, F., 2014. Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol. J.* 12, 934–940. doi:10.1111/pbi.12201

Helchowski, C.M., Skow, L.F., Roberts, K.H., Chute, C.L., Canman, C.E., 2013. A small ubiquitin binding domain inhibits ubiquitin-dependent protein recruitment to DNA repair foci. *Cell Cycle Georget. Tex* 12, 3749–3758. doi:10.4161/cc.26640

Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.-F., Sontheimer, E.J., Thomson, J.A., 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from

Neisseria meningitidis. *Proc. Natl. Acad. Sci. U. S. A.* 110, 15644–15649. doi:10.1073/pnas.1313587110

Howden, S.E., McColl, B., Glaser, A., Vadolas, J., Petrou, S., Little, M.H., Elefanty, A.G., Stanley, E.G., 2016. A Cas9 Variant for Efficient Generation of Indel-Free Knockin or Gene-Corrected Human Pluripotent Stem Cells. *Stem Cell Rep.* doi:10.1016/j.stemcr.2016.07.001

Hu, J., Lei, Y., Wong, W.-K., Liu, S., Lee, K.-C., He, X., You, W., Zhou, R., Guo, J.-T., Chen, X., Peng, X., Sun, H., Huang, H., Zhao, H., Feng, B., 2014. Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Res.* 42, 4375–4390. doi:10.1093/nar/gku109

Huang, P., Xiao, A., Tong, X., Zu, Y., Wang, Z., Zhang, B., 2014. TALEN construction via “Unit Assembly” method and targeted genome modifications in zebrafish. *Methods San Diego Calif* 69, 67–75. doi:10.1016/j.ymeth.2014.02.010

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., Zhang, B., 2011. Heritable gene targeting in zebrafish using customized TALENs. *Nat. Biotechnol.* 29, 699–700. doi:10.1038/nbt.1939

Huertas, P., Jackson, S.P., 2009. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J. Biol. Chem.* 284, 9558–9565. doi:10.1074/jbc.M808906200

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., Joung, J.K., 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 227–229. doi:10.1038/nbt.2501

Iizumi, S., Kurosawa, A., So, S., Ishii, Y., Chikaraishi, Y., Ishii, A., Koyama, H., Adachi, N., 2008. Impact of non-homologous end-joining deficiency on random and targeted DNA integration: implications for gene targeting. *Nucleic Acids Res.* 36, 6333–6342. doi:10.1093/nar/gkn649

Iliakis, G., Murmann, T., Soni, A., 2015. Alternative end-joining repair pathways are the ultimate backup for abrogated classical non-homologous end-joining and homologous recombination repair: Implications for the formation of chromosome translocations.

Mutat. Res. Genet. Toxicol. Environ. Mutagen. 793, 166–175.
doi:10.1016/j.mrgentox.2015.07.001

Ivanov, E.L., Sugawara, N., Fishman-Lobell, J., Haber, J.E., 1996. Genetic requirements for the single-strand annealing pathway of double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* 142, 693–704.

Iwabuchi, K., Bartel, P.L., Li, B., Marraccino, R., Fields, S., 1994. Two cellular proteins that bind to wild-type but not mutant p53. *Proc. Natl. Acad. Sci. U. S. A.* 91, 6098–6102.

Jacquet, K., Fradet-Turcotte, A., Avvakumov, N., Lambert, J.-P., Roques, C., Pandita, R.K., Paquet, E., Herst, P., Gingras, A.-C., Pandita, T.K., Legube, G., Doyon, Y., Durocher, D., Côté, J., 2016. The TIP60 Complex Regulates Bivalent Chromatin Recognition by 53BP1 through Direct H4K20me Binding and H2AK15 Acetylation. *Mol. Cell* 62, 409–421.
doi:10.1016/j.molcel.2016.03.031

Jayathilaka, K., Sheridan, S.D., Bold, T.D., Bochenska, K., Logan, H.L., Weichselbaum, R.R., Bishop, D.K., Connell, P.P., 2008. A chemical compound that stimulates the human homologous recombination protein RAD51. *Proc. Natl. Acad. Sci. U. S. A.* 105, 15848–15853. doi:10.1073/pnas.0808046105

Jette, N., Lees-Miller, S.P., 2015. The DNA-dependent protein kinase: A multifunctional protein kinase with roles in DNA double strand break repair and mitosis. *Prog. Biophys. Mol. Biol.* 117, 194–205. doi:10.1016/j.pbiomolbio.2014.12.003

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi:10.1126/science.1225829

Kamthan, A., Chaudhuri, A., Kamthan, M., Datta, A., 2016. Genetically modified (GM) crops: milestones and new advances in crop improvement. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* doi:10.1007/s00122-016-2747-6

Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P., Siksnys, V., 2013. crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol.* 10, 841–851. doi:10.4161/rna.24203

Kato, T., Miyata, K., Sonobe, M., Yamashita, S., Tamano, M., Miura, K., Kanai, Y., Miyamoto,

- S., Sakuma, T., Yamamoto, T., Inui, M., Kikusui, T., Asahara, H., Takada, S., 2013. Production of Sry knockout mouse using TALEN via oocyte injection. *Sci. Rep.* 3, 3136. doi:10.1038/srep03136
- Kay, S., Bonas, U., 2009. How *Xanthomonas* type III effectors manipulate the host plant. *Curr. Opin. Microbiol.* 12, 37–43. doi:10.1016/j.mib.2008.12.006
- Khade, N.V., Sugiyama, T., 2016. Roles of C-Terminal Region of Yeast and Human Rad52 in Rad51-Nucleoprotein Filament Formation and ssDNA Annealing. *PloS One* 11, e0158436. doi:10.1371/journal.pone.0158436
- Kim, Y.G., Cha, J., Chandrasegaran, S., 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1156–1160.
- Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., Joung, J.K., 2016a. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495. doi:10.1038/nature16526
- Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Nguyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J., Joung, J.K., 2016b. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* 34, 869–874. doi:10.1038/nbt.3620
- Komatsu, K., 2016. NBS1 and multiple regulations of DNA damage response. *J. Radiat. Res. (Tokyo)* 57 Suppl 1, i11–i17. doi:10.1093/jrr/rrw031
- Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., Liu, D.R., 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi:10.1038/nature17946
- Kozlov, S.V., Graham, M.E., Jakob, B., Tobias, F., Kijas, A.W., Tanuji, M., Chen, P., Robinson, P.J., Taucher-Scholz, G., Suzuki, K., So, S., Chen, D., Lavin, M.F., 2011. Autophosphorylation and ATM activation: additional sites add to the complexity. *J. Biol. Chem.* 286, 9107–9119. doi:10.1074/jbc.M110.204065
- Lafranchi, L., de Boer, H.R., de Vries, E.G.E., Ong, S.-E., Sartori, A.A., van Vugt, M.A.T.M., 2014. APC/C(Cdh1) controls CtIP stability during the cell cycle and in response to DNA damage. *EMBO J.* 33, 2860–2879. doi:10.15252/embj.201489017
- Larcher, T., Lafoux, A., Tesson, L., Remy, S., Thepenier, V., François, V., Le Guiner, C.,

- Goubin, H., Dutilleul, M., Guigand, L., Toumaniantz, G., De Cian, A., Boix, C., Renaud, J.-B., Cherel, Y., Giovannangeli, C., Concordet, J.-P., Anegón, I., Huchet, C., 2014. Characterization of dystrophin deficient rats: a new model for Duchenne muscular dystrophy. *PloS One* 9, e110371. doi:10.1371/journal.pone.0110371
- Lee-Theilen, M., Matthews, A.J., Kelly, D., Zheng, S., Chaudhuri, J., 2011. CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. *Nat. Struct. Mol. Biol.* 18, 75–79. doi:10.1038/nsmb.1942
- Li, J., Xu, X., 2016. DNA double-strand break repair: a tale of pathway choices. *Acta Biochim. Biophys. Sin.* 48, 641–646. doi:10.1093/abbs/gmw045
- Li, L., Wu, L.P., Chandrasegaran, S., 1992. Functional domains in Fok I restriction endonuclease. *Proc. Natl. Acad. Sci. U. S. A.* 89, 4275–4279.
- Li, Q., Barish, S., Okuwa, S., Volkan, P.C., 2015. Examination of Endogenous Rotund Expression and Function in Developing Drosophila Olfactory System Using CRISPR-Cas9-Mediated Protein Tagging. *G3 Bethesda Md* 5, 2809–2816. doi:10.1534/g3.115.021857
- Li, S., Ting, N.S., Zheng, L., Chen, P.L., Ziv, Y., Shiloh, Y., Lee, E.Y., Lee, W.H., 2000. Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response. *Nature* 406, 210–215. doi:10.1038/35018134
- Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P., Yang, B., 2011. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.* 39, 6315–6325. doi:10.1093/nar/gkr188
- Lin, S., Staahl, B.T., Alla, R.K., Doudna, J.A., 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3, e04766. doi:10.7554/eLife.04766
- Liu, T., Huang, J., 2016. DNA End Resection: Facts and Mechanisms. *Genomics Proteomics Bioinformatics* 14, 126–130. doi:10.1016/j.gpb.2016.05.002
- Liu, Y., Li, Y., Lu, X., 2016. Regulators in the DNA damage response. *Arch. Biochem. Biophys.* 594, 18–25. doi:10.1016/j.abb.2016.02.018

- Lv, Q., Yuan, L., Deng, J., Chen, M., Wang, Y., Zeng, J., Li, Z., Lai, L., 2016. Efficient Generation of Myostatin Gene Mutated Rabbit by CRISPR/Cas9. *Sci. Rep.* 6, 25029. doi:10.1038/srep25029
- Mari, P.-O., Florea, B.I., Persengiev, S.P., Verkaik, N.S., Brüggewirth, H.T., Modesti, M., Giglia-Mari, G., Bezstarosti, K., Demmers, J.A.A., Luider, T.M., Houtsmuller, A.B., van Gent, D.C., 2006. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18597–18602. doi:10.1073/pnas.0609061103
- Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., Ploegh, H.L., 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* 33, 538–542. doi:10.1038/nbt.3190
- Mateos-Gomez, P.A., Gong, F., Nair, N., Miller, K.M., Lazzerini-Denchi, E., Sfeir, A., 2015. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* 518, 254–257. doi:10.1038/nature14157
- Mazur, D.J., Perrino, F.W., 2001. Excision of 3' termini by the Trex1 and TREX2 3'-->5' exonucleases. Characterization of the recombinant proteins. *J. Biol. Chem.* 276, 17022–17029. doi:10.1074/jbc.M100623200
- McVey, M., Lee, S.E., 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet. TIG* 24, 529–538. doi:10.1016/j.tig.2008.08.007
- Mehta, A., Haber, J.E., 2014. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* 6, a016428. doi:10.1101/cshperspect.a016428
- Mortensen, U.H., Lisby, M., Rothstein, R., 2009. Rad52. *Curr. Biol. CB* 19, R676–677. doi:10.1016/j.cub.2009.06.001
- Motycka, T.A., Bessho, T., Post, S.M., Sung, P., Tomkinson, A.E., 2004. Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. *J. Biol. Chem.* 279, 13634–13639. doi:10.1074/jbc.M313779200
- Muñoz, M.C., Laulier, C., Gunn, A., Cheng, A., Robbani, D.F., Nussenzweig, A., Stark, J.M.,

2012. RING finger nuclear factor RNF168 is important for defects in homologous recombination caused by loss of the breast cancer susceptibility factor BRCA1. *J. Biol. Chem.* 287, 40618–40628. doi:10.1074/jbc.M112.410951

Murphy, K.C., 1998. Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J. Bacteriol.* 180, 2063–2071.

Nakagawa, Y., Oikawa, F., Mizuno, S., Ohno, H., Yagishita, Y., Satoh, A., Osaki, Y., Takei, K., Kikuchi, T., Han, S.-I., Matsuzaka, T., Iwasaki, H., Kobayashi, K., Yatoh, S., Yahagi, N., Isaka, M., Suzuki, H., Sone, H., Takahashi, S., Yamada, N., Shimano, H., 2016. Hyperlipidemia and hepatitis in liver-specific CREB3L3 knockout mice generated using a one-step CRISPR/Cas9 system. *Sci. Rep.* 6, 27857. doi:10.1038/srep27857

Nishizawa-Yokoi, A., Nonaka, S., Saika, H., Kwon, Y.-I., Osakabe, K., Toki, S., 2012. Suppression of Ku70/80 or Lig4 leads to decreased stable transformation and enhanced homologous recombination in rice. *New Phytol.* 196, 1048–1059. doi:10.1111/j.1469-8137.2012.04350.x

Orr-Weaver, T.L., Szostak, J.W., 1983. Yeast recombination: the association between double-strand gap repair and crossing-over. *Proc. Natl. Acad. Sci. U. S. A.* 80, 4417–4421.

Ousterout, D.G., Kabadi, A.M., Thakore, P.I., Majoros, W.H., Reddy, T.E., Gersbach, C.A., 2015. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat. Commun.* 6, 6244. doi:10.1038/ncomms7244

Park, C.-Y., Kim, D.H., Son, J.S., Sung, J.J., Lee, J., Bae, S., Kim, J.-H., Kim, D.-W., Kim, J.-S., 2015. Functional Correction of Large Factor VIII Gene Chromosomal Inversions in Hemophilia A Patient-Derived iPSCs Using CRISPR-Cas9. *Cell Stem Cell* 17, 213–220. doi:10.1016/j.stem.2015.07.001

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., Liu, D.R., 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 31, 839–843. doi:10.1038/nbt.2673

Pavletich, N.P., Pabo, C.O., 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252, 809–817.

- Perrino, F.W., Harvey, S., McMillin, S., Hollis, T., 2005. The human TREX2 3' -> 5'-exonuclease structure suggests a mechanism for efficient nonprocessive DNA catalysis. *J. Biol. Chem.* 280, 15212–15218. doi:10.1074/jbc.M500108200
- Pierce, A.J., Hu, P., Han, M., Ellis, N., Jasin, M., 2001. Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. *Genes Dev.* 15, 3237–3242. doi:10.1101/gad.946401
- Pinder, J., Salsman, J., Dellaire, G., 2015. Nuclear domain “knock-in” screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res.* 43, 9379–9392. doi:10.1093/nar/gkv993
- Plessis, A., Perrin, A., Haber, J.E., Dujon, B., 1992. Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* 130, 451–460.
- Qi, Y., Li, X., Zhang, Y., Starker, C.G., Baltes, N.J., Zhang, F., Sander, J.D., Reyon, D., Joung, J.K., Voytas, D.F., 2013. Targeted deletion and inversion of tandemly arrayed genes in *Arabidopsis thaliana* using zinc finger nucleases. *G3 Bethesda Md* 3, 1707–1715. doi:10.1534/g3.113.006270
- Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., Koonin, E.V., Sharp, P.A., Zhang, F., 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520, 186–191. doi:10.1038/nature14299
- Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., Zhang, F., 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380–1389. doi:10.1016/j.cell.2013.08.021
- Reisch, C.R., Prather, K.L.J., 2015. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. *Sci. Rep.* 5, 15096. doi:10.1038/srep15096
- Remy, S., Tesson, L., Menoret, S., Usal, C., De Cian, A., Thepenier, V., Thinard, R., Baron, D., Charpentier, M., Renaud, J.-B., Buelow, R., Cost, G.J., Giovannangeli, C., Fraichard, A., Concordet, J.-P., Anegon, I., 2014. Efficient gene targeting by homology-directed repair in

rat zygotes using TALE nucleases. *Genome Res.* 24, 1371–1383. doi:10.1101/gr.171538.113

Renaud, J.-B., Boix, C., Charpentier, M., De Cian, A., Cochennec, J., Duvernois-Berthet, E., Perrouault, L., Tesson, L., Edouard, J., Thinard, R., Cherifi, Y., Menoret, S., Fontanière, S., de Crozé, N., Fraichard, A., Sohm, F., Anegon, I., Concordet, J.-P., Giovannangeli, C., 2016. Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Rep.* 14, 2263–2272. doi:10.1016/j.celrep.2016.02.018

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., Joung, J.K., 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* 30, 460–465. doi:10.1038/nbt.2170

Rodgers, K., McVey, M., 2016. Error-Prone Repair of DNA Double-Strand Breaks. *J. Cell. Physiol.* 231, 15–24. doi:10.1002/jcp.25053

Römer, P., Recht, S., Strauss, T., Elsaesser, J., Schornack, S., Boch, J., Wang, S., Lahaye, T., 2010. Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* 187, 1048–1057. doi:10.1111/j.1469-8137.2010.03217.x

Rothenberg, E., Grimme, J.M., Spies, M., Ha, T., 2008. Human Rad52-mediated homology search and annealing occurs by continuous interactions between overlapping nucleoprotein complexes. *Proc. Natl. Acad. Sci. U. S. A.* 105, 20274–20279. doi:10.1073/pnas.0810317106

Rouet, P., Smih, F., Jasin, M., 1994. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 91, 6064–6068.

Saha, J., Davis, A.J., 2016. Unsolved mystery: the role of BRCA1 in DNA end-joining. *J. Radiat. Res. (Tokyo)* 57 Suppl 1, i18–i24. doi:10.1093/jrr/rrw032

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., Imamura, T., Ogawa, M., Masai, H., Miyawaki, A., 2008. Visualizing spatiotemporal dynamics of multicellular cell-cycle

progression. *Cell* 132, 487–498. doi:10.1016/j.cell.2007.12.033

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., Jackson, S.P., 2007. Human CtIP promotes DNA end resection. *Nature* 450, 509–514. doi:10.1038/nature06337

Schmid-Burgk, J.L., Schmidt, T., Hornung, V., 2015. Ligation-independent cloning (LIC) assembly of TALEN genes. *Methods Mol. Biol. Clifton NJ* 1239, 161–169. doi:10.1007/978-1-4939-1862-1_8

Schreiber, V., Dantzer, F., Ame, J.-C., de Murcia, G., 2006. Poly(ADP-ribose): novel functions for an old molecule. *Nat. Rev. Mol. Cell Biol.* 7, 517–528. doi:10.1038/nrm1963

Schwank, G., Koo, B.-K., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., van der Ent, C.K., Nieuwenhuis, E.E.S., Beekman, J.M., Clevers, H., 2013. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* 13, 653–658. doi:10.1016/j.stem.2013.11.002

Schwertman, P., Bekker-Jensen, S., Mailand, N., 2016. Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nat. Rev. Mol. Cell Biol.* 17, 379–394. doi:10.1038/nrm.2016.58

Simsek, D., Brunet, E., Wong, S.Y.-W., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J., Gregory, P.D., Holmes, M.C., Jasin, M., 2011. DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS Genet.* 7, e1002080. doi:10.1371/journal.pgen.1002080

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., Zhang, F., 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88. doi:10.1126/science.aad5227

Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.-C., Chames, P., Prieto, J., Redondo, P., Blanco, F.J., Bravo, J., Montoya, G., Pâques, F., Duchateau, P., 2006. A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.* 34, e149. doi:10.1093/nar/gkl720

Song, J., Yang, D., Xu, J., Zhu, T., Chen, Y.E., Zhang, J., 2016. RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.* 7, 10548.

doi:10.1038/ncomms10548

Sovová, T., Kerins, G., Demnerová, K., Ovesná, J., 2016. Genome Editing with Engineered Nucleases in Economically Important Animals and Plants: State of the Art in the Research Pipeline. *Curr. Issues Mol. Biol.* 21, 41–62.

Srivastava, M., Nambiar, M., Sharma, S., Karki, S.S., Goldsmith, G., Hegde, M., Kumar, S., Pandey, M., Singh, R.K., Ray, P., Natarajan, R., Kelkar, M., De, A., Choudhary, B., Raghavan, S.C., 2012. An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell* 151, 1474–1487. doi:10.1016/j.cell.2012.11.054

Sugisaki, H., Kanazawa, S., 1981. New restriction endonucleases from *Flavobacterium okeanokoites* (FokI) and *Micrococcus luteus* (MluI). *Gene* 16, 73–78.

Sung, Y.H., Kim, J.M., Kim, H.-T., Lee, J., Jeon, J., Jin, Y., Choi, J.-H., Ban, Y.H., Ha, S.-J., Kim, C.-H., Lee, H.-W., Kim, J.-S., 2014. Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. *Genome Res.* 24, 125–131. doi:10.1101/gr.163394.113

Swarts, D.C., Hegge, J.W., Hinojo, I., Shiimori, M., Ellis, M.A., Dumrongkulraksa, J., Terns, R.M., Terns, M.P., van der Oost, J., 2015. Argonaute of the archaeon *Pyrococcus furiosus* is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Res.* 43, 5120–5129. doi:10.1093/nar/gkv415

Tebas, P., Stein, D., Tang, W.W., Frank, I., Wang, S.Q., Lee, G., Spratt, S.K., Surosky, R.T., Giedlin, M.A., Nichol, G., Holmes, M.C., Gregory, P.D., Ando, D.G., Kalos, M., Collman, R.G., Binder-Scholl, G., Plesa, G., Hwang, W.-T., Levine, B.L., June, C.H., 2014. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* 370, 901–910. doi:10.1056/NEJMoa1300662

Terns, M.P., Terns, R.M., 2011. CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* 14, 321–327. doi:10.1016/j.mib.2011.03.005

Thyme, S.B., Schier, A.F., 2016. Polq-Mediated End Joining Is Essential for Surviving DNA Double-Strand Breaks during Early Zebrafish Development. *Cell Rep.* doi:10.1016/j.celrep.2016.03.072

Tripathi, K., Mani, C., Clark, D.W., Palle, K., 2016. Rad18 is required for functional

interactions between FANCD2, BRCA2, and Rad51 to repair DNA topoisomerase 1-poisons induced lesions and promote fork recovery. *Oncotarget* 7, 12537–12553. doi:10.18632/oncotarget.7247

Trovesi, C., Manfrini, N., Falcettoni, M., Longhese, M.P., 2013. Regulation of the DNA damage response by cyclin-dependent kinases. *J. Mol. Biol.* 425, 4756–4766. doi:10.1016/j.jmb.2013.04.013

Truong, L.N., Li, Y., Shi, L.Z., Hwang, P.Y.-H., He, J., Wang, H., Razavian, N., Berns, M.W., Wu, X., 2013. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 110, 7720–7725. doi:10.1073/pnas.1213431110

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., Joung, J.K., 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* 32, 569–576. doi:10.1038/nbt.2908

Tsakraklides, V., Brevnova, E., Stephanopoulos, G., Shaw, A.J., 2015. Improved Gene Targeting through Cell Cycle Synchronization. *PloS One* 10, e0133434. doi:10.1371/journal.pone.0133434

Urnov, F.D., Miller, J.C., Lee, Y.-L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., Holmes, M.C., 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435, 646–651. doi:10.1038/nature03556

Wang, H., Shi, L.Z., Wong, C.C.L., Han, X., Hwang, P.Y.-H., Truong, L.N., Zhu, Q., Shao, Z., Chen, D.J., Berns, M.W., Yates, J.R., Chen, L., Wu, X., 2013. The interaction of CtIP and Nbs1 connects CDK and ATM to regulate HR-mediated double-strand break repair. *PLoS Genet.* 9, e1003277. doi:10.1371/journal.pgen.1003277

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., Qiu, J.-L., 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi:10.1038/nbt.2969

Wiedenheft, B., Sternberg, S.H., Doudna, J.A., 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338. doi:10.1038/nature10886

Wojtal, D., Kemaladewi, D.U., Malam, Z., Abdullah, S., Wong, T.W.Y., Hyatt, E., Baghestani, Z., Pereira, S., Stavropoulos, J., Mouly, V., Mamchaoui, K., Muntoni, F., Voit, T., Gonorazky, H.D., Dowling, J.J., Wilson, M.D., Mendoza-Londono, R., Ivakine, E.A., Cohn, R.D., 2016. Spell Checking Nature: Versatility of CRISPR/Cas9 for Developing Treatments for Inherited Disorders. *Am. J. Hum. Genet.* 98, 90–101. doi:10.1016/j.ajhg.2015.11.012

Wong, A.K., Ormonde, P.A., Pero, R., Chen, Y., Lian, L., Salada, G., Berry, S., Lawrence, Q., Dayananth, P., Ha, P., Tavtigian, S.V., Teng, D.H., Bartel, P.L., 1998. Characterization of a carboxy-terminal BRCA1 interacting protein. *Oncogene* 17, 2279–2285. doi:10.1038/sj.onc.1202150

Woo, J.W., Kim, J., Kwon, S.I., Corvalán, C., Cho, S.W., Kim, H., Kim, S.-G., Kim, S.-T., Choe, S., Kim, J.-S., 2015. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* 33, 1162–1164. doi:10.1038/nbt.3389

Xie, A., Kwok, A., Scully, R., 2009. Role of mammalian Mre11 in classical and alternative nonhomologous end joining. *Nat. Struct. Mol. Biol.* 16, 814–818. doi:10.1038/nsmb.1640

Xin, J., Yang, H., Fan, N., Zhao, B., Ouyang, Z., Liu, Z., Zhao, Y., Li, X., Song, J., Yang, Y., Zou, Q., Yan, Q., Zeng, Y., Lai, L., 2013. Highly efficient generation of GGTA1 biallelic knockout inbred mini-pigs with TALENs. *PloS One* 8, e84250. doi:10.1371/journal.pone.0084250

Xing, M., Yang, M., Huo, W., Feng, F., Wei, L., Jiang, W., Ning, S., Yan, Z., Li, W., Wang, Q., Hou, M., Dong, C., Guo, R., Gao, G., Ji, J., Zha, S., Lan, L., Liang, H., Xu, D., 2015. Interactome analysis identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway. *Nat. Commun.* 6, 6233. doi:10.1038/ncomms7233

Yang, D., Scavuzzo, M.A., Chmielowiec, J., Sharp, R., Bajic, A., Borowiak, M., 2016. Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Sci. Rep.* 6, 21264. doi:10.1038/srep21264

Yee, J.-K., 2016a. Off-target effects of engineered nucleases. *FEBS J.* doi:10.1111/febs.13760

Yee, J.-K., 2016b. Off-target effects of engineered nucleases. *FEBS J.* doi:10.1111/febs.13760

- Yim, H., Shin, S.-B., Woo, S.U., Lee, P.C.-W., Erikson, R.L., 2016. Plk1-mediated stabilization of 53BP1 through USP7 regulates centrosome positioning to maintain bipolarity. *Oncogene*. doi:10.1038/onc.2016.263
- Yu, X., Chen, J., 2004. DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains. *Mol. Cell. Biol.* 24, 9478–9486. doi:10.1128/MCB.24.21.9478-9486.2004
- Yuan, J., Chen, J., 2009. N terminus of CtIP is critical for homologous recombination-mediated double-strand break repair. *J. Biol. Chem.* 284, 31746–31752. doi:10.1074/jbc.M109.023424
- Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., Koonin, E.V., Zhang, F., 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771. doi:10.1016/j.cell.2015.09.038
- Zhong, Q., Chen, C.-F., Chen, P.-L., Lee, W.-H., 2002. BRCA1 facilitates microhomology-mediated end joining of DNA double strand breaks. *J. Biol. Chem.* 277, 28641–28647. doi:10.1074/jbc.M200748200

Résumé :

L'édition du génome repose sur la création de cassures double brin à un endroit précis du génome à l'aide de nucléases artificielles (ZFN, TALEN, CRISPR/Cas9) et sur les différents systèmes de réparation que la cellule va mettre en place pour réparer ces dommages. Les deux systèmes de réparation principaux sont le NHEJ (Non Homologous End Joining) et la RH (Recombinaison Homologue). Le NHEJ consiste en une ligation directe des extrémités de la coupure pouvant induire de petites insertions ou délétions avant la ligation. Ces mutations, si elles sont introduites dans un exon, vont modifier le cadre de lecture et pouvoir inactiver le gène cible (Knock Out). La RH permet la réparation de la cassure en recopiant les informations présentes sur la chromatide sœur. Si un ADN exogène comportant des homologies avec la séquence à réparer est inséré avec les nucléases artificielles, la cellule peut le prendre comme matrice de réparation, il est ainsi possible d'insérer n'importe quelle mutation ou transgène de manière précise (Knock In). Ici, différentes stratégies ont été développées pour optimiser ces approches d'édition du génome. Le couplage du domaine Nter de la protéine CtIP à la nucléase Cas9 permet d'augmenter le taux d'insertion par homologie d'un transgène au site de coupure. Le couplage de l'exonucléase Trex2 à la nucléase Cas9 nickase permet quant à lui d'augmenter le taux de mutation après coupure. Ces nouvelles approches peuvent être largement utilisées et permettent de faciliter l'édition du génome.

Title :

New genome editing approaches development using artificial nucleases (TALEN and CRISPR/Cas9)

Abstract :

Genome editing relies on the ability of artificial nucleases (TALEN or CRISPR/Cas9 system) to induce double strand break into a precise and unique sequence in a whole genome and on the different DNA repair system. The two major DNA repair systems are NHEJ (Non Homologous End Joining) and HR (Homologous Recombination). NHEJ consists on DNA end direct ligation. This system can lead to deletion or insertion at the cut site. These mutations, when induced in an exon, can induce reading frame change and gene inactivation (Knock out). HR consists on the use of sister chromatid to copy lost information in order to complete the double strand break. If an exogenous DNA with homologies with the targeted DNA is inserted with artificial nucleases, it can be used as a template and can permit to introduce any transgene at the cut site (Knock In). In this work, different strategies were used to optimize genome editing. By fusing Nter part of CtIP to Cas9, the KI rate of an exogenous DNA is increased and by fusing Trex2 exonuclease to Cas9, the mutation rate induced is also increased. These two approaches can be widely used to improve genome editing strategies.

Mots-clés :

Edition du génome – nucléases artificielles – réparation de l'ADN

Keywords :

Genome editing – artificial nucleases – DNA repair