



HAL
open science

Génotoxicité des systèmes CRISPR-Cas9.

Grégoire Cullot

► **To cite this version:**

Grégoire Cullot. Génotoxicité des systèmes CRISPR-Cas9.. Médecine humaine et pathologie. Université de Bordeaux, 2019. Français. NNT : 2019BORD0344 . tel-03270870

HAL Id: tel-03270870

<https://theses.hal.science/tel-03270870>

Submitted on 25 Jun 2021

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.

Université de Bordeaux

Année 2019

Thèse n°

THÈSE

pour le

DOCTORAT DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE : Sciences de la Vie et de la Santé

SPÉCIALITÉ : Biologie Cellulaire et Physiopathologie

Présentée et soutenue publiquement le 13 décembre 2019

par

Grégoire Cullot

Né le 24 février 1993 à Dijon (21)

Génotoxicité des systèmes CRISPR-Cas9.

Inserm U1035 - Biothérapie des Maladies Génétiques, Inflammatoires et du Cancer
Equipe « Biothérapie »

Membres du jury :

M. Arveiler, Benoit	PU-PH	Univ. Bordeaux	Président
Mme. Miccio, Annarita	DR	Univ. Paris-Descartes	Rapporteur
M. De Massy, Bernard	DR	Univ. Montpellier	Rapporteur
Mme. Giovannangeli, Carine	DR	Univ. P. et M. Curie	Examineur
Mme. Bedel, Aurélie	MCU-PH	Univ. Bordeaux	Invité
M. Moreau-Gaudry, François	PU-PH	Univ. Bordeaux	Directeur

Titre : Génotoxicité des systèmes CRISPR-Cas9.

Résumé : La thérapie génique est une stratégie thérapeutique prometteuse pour le traitement des maladies monogéniques. Si les premières approches, dites additives, ont reposé sur l'utilisation de vecteurs viraux, une part grandissante se tourne désormais vers l'édition génique. Celle-ci est permise par la mise au point de nouvelles générations d'endonucléases, et en particulier le système CRISPR-Cas9. Moins d'une décennie après sa caractérisation, le système CRISPR-Cas9 a permis de faire passer l'édition génique à un stade clinique. Toutefois, dans le même laps de temps, plusieurs interrogations ont été soulevées vis-à-vis de la génotoxicité pouvant être induite par la Cas9. Une littérature émergente pointe le risque de génotoxicité au site ciblé. Le travail de thèse présenté ici s'inscrit dans cette thématique. La première partie de l'étude a eu pour objectif de décrire la génotoxicité induite par une unique cassure double-brin faite par la Cas9. La caractérisation des effets a été faite à la fois à l'échelle nucléotidique, par le suivi de la balance HDR / InDels, mais également à l'échelle du chromosome. Le suivi de l'intégrité chromosomique a permis de mettre en lumière un nouveau risque de génotoxicité encore non-caractérisé. Un système de détection sensible et spécifique de ce risque a été mis au point pour continuer de le caractériser. Le second objectif a été de répondre aux limites soulevées par la génotoxicité non-voulue, en mettant au point une méthode d'édition génique plus sûre et aussi efficace, *via* l'utilisation d'une unique cassure simple-brin par la Cas9^{D10A}-nickase.

Mots clés : CRISPR/Cas9, Génotoxicité, Edition génique

Title : Genotoxicity of CRISPR-Cas9 systems.

Abstract : Gene therapy is a promising therapeutic strategy for the monogenic diseases treatment. If the first approaches, called additive, have relied on the use of viral vectors, a growing share is now turning to gene editing. Less than a decade after its characterization, the CRISPR-Cas9 system has moved gene editing to a clinical stage. However, in the same period of time, several questions have been raised regarding the genotoxicity that can be induced by Cas9. An emerging literature points to the risk of genotoxicity at the targeted site. The thesis work presented here is part of this theme. The first part of the study aimed to describe the genotoxicity induced by a single double-stranded break made by Cas9. Characterization of the effects was done both at the nucleotide level, by monitoring the HDR / InDels balance, but also at the chromosome scale. The monitoring of chromosomal integrity has brought to light a new risk of genotoxicity that was not characterized. A sensitive and specific detection system for this risk has been developed to further characterize it. The second objective was to address the limitations of unwanted genotoxicity by developing a safer and more efficient gene editing method through the use of a single single-stranded breakage by Cas9^{D10A}-nickase.

Keywords : CRISPR/Cas9, Genotoxicity, Gene editing

INSERM U1035, BMGIC - Univ. Bordeaux, France

Biothérapie des Maladies Génétiques, Inflammatoires et du Cancer.

Equipe Biothérapie

Remerciements

La Recherche se doit d'être peu sujette à l'influence des émotions pour être aussi objective que possible. Dès lors, il serait tentant de voir les acteurs de la recherche de la même manière, froids et objectifs. Une conclusion qui s'avérerait plus qu'inexacte. En effet, la Recherche est avant tout permise par un travail collectif dans lequel chacun œuvre et apporte sa contribution nécessaire. À ce titre, tous ceux ayant participé aux travaux présentés dans ce manuscrit mériteraient de figurer sur la couverture. En guise de réparation, j'essaierai dans ces quelques lignes de mettre en lumière vos participations, tout d'abord pour vous en remercier, mais également pour informer le lecteur que les résultats présentés ici ne sont pas issus d'un travail solitaire.

À chaque thèse, son commencement, et à Bordeaux, cela rime souvent avec concours de l'école doctorale. Ici, je serais toujours reconnaissant de la confiance et de l'aide que m'ont accordé François Moreau-Gaudry et Aurélie Bedel ; rapidement rejoint par Sandrine Dabernat pour compléter ce trio de choc. Dès les premières réunions de préparations, j'ai ressenti une bienveillance. Plus largement, cette bienveillance, qualité commune à toutes les personnes ayant participé à ce travail, fût présente tout au long de ces trois années.

François, Aurélie, je vous remercie infiniment du temps que vous m'avez consacré, toujours avec une précieuse disponibilité et une écoute chaleureuse. Grâce à vos conseils et votre expertise, vous avez su encadrer et bâtir ce projet de thèse qui n'aurait jamais abouti sans vous. En même temps, savoir vos qualités d'empathies m'a aidé à traverser les moments personnels les plus compliqués ; en sachant me donner du temps quand j'en avais le plus besoin. Bien que je ne puisse ici que parler en mon nom, je pense que l'ensemble des personnes qui travaillent avec vous, et en particulier les étudiants, seront d'accord avec l'affirmation suivante : vous pouvez être fiers de votre façon d'encadrer et de vos qualités humaines.

Et à chaque thèse, une fin, marquée par la rédaction de ce manuscrit et de la soutenance. L'occasion de souligner une fois encore, le temps et le travail investis par François et Aurélie pour me guider dans l'élaboration du manuscrit et me préparer au mieux à la défense de thèse. Cet examen final n'a de sens que par sa soumission à l'expertise des rapporteurs et des membres du jury, dont il faut souligner l'engagement bénévole. Je tiens à remercier Mme. Miccio, Mr. De Massy et Mme. Giovannangeli d'avoir fait une lecture critique de ce manuscrit et pour leurs rapports favorables. L'apport des différents membres du jury a permis à la soutenance d'avoir été un moment scientifiquement riche. Les discussions m'ont permis de me confronter aux limites des résultats présentés ici ; limites qui forment les points de départ des projets suivants. Enfin, je tiens à souligner

et à remercier chaleureusement Mr. Arveiler pour avoir présidé ce jury, et pour avoir, de plus, su maîtriser parfaitement les conditions particulières imposées par la visioconférence.

Je profite de ces lignes pour mettre en valeur l'importance du travail de Jérôme Toutain, Caroline Rooryck (également tutrice de thèse) et de Perrine Pennamen pour le volet cytogénétique des résultats présentés ici, et dont dépend une grande partie de l'originalité de ce projet de thèse.

Et puis, il y a trois années passées aux côtés des membres de l'équipe Biothérapie, et plus largement, de l'unité 1035. Ces années ont été animées et rythmées par une multitude d'histoires et d'interactions, venant égayer chaque journée, dont je ne pourrais malheureusement pas faire entièrement le récit ici. Toutefois, les lignes suivantes tenteront de montrer au lecteur comment chacun y a participé.

Jean-Marc, Sandrine, David et également, Emmanuel, Cécile, au fil des discussions, vous m'avez aidé à élaborer ce projet de thèse, toujours avec une bienveillance qui a fait que chaque réunion était un moment de partage, constructif scientifiquement, plutôt qu'une confrontation contre des avis critiques. Ces bonnes conditions sont précieuses lors de l'apprentissage des métiers de la recherche.

Magalie, Alice, Isabelle LG., Véronique, Isabelle M., Quentin, Jennifer, sans vous, je pense que je ne saurais encore rien faire. À combien de questions avez-vous répondu pour me permettre de sortir de soucis et tracas expérimentaux... Le compte a rapidement été perdu. Vous êtes les véritables catalyseurs du savoir expérimental, et ce, toujours avec gentillesse.

Et lassantes auraient été ces années si votre joie de vivre et votre sympathie, s'exprimant aux travers de sourires, de rires et de petites histoires à partager, n'étaient pas venus les animer. Sur ce point, je salue également Jérôme, Denis, Clément, Catherine, Sandrine F., Anthony, Arnaud et Amélie. Tous, vous avez contribué au fait que ces trois années furent un plaisir.

Un laboratoire sans étudiant serait bien dépeuplé. Certes, on pourrait le déplorer, mais c'est également une aubaine pour découvrir pleins de personnalités et nouer des amitiés. Il y a ceux qui ont été de passage, ceux qui ont été présents tout le long et, bien heureusement, une nouvelle génération pour prendre le relais.

Julie, « c'est pas la fête à Neu-Neu ici » (en fait si, un peu quand même). Toutefois, citer une simple phrase serait trop réducteur, ne rendant pas hommage à ta forte personnalité. Tu m'as beaucoup appris lors de mes débuts et surtout appris à bien manipuler... bien que je dois t'avouer avoir pris certaines libertés par rapport aux enseignements initiaux. Par exemple, nos techniques diffèrent en salle BET. Désolé, mais c'était tentant. Tu as toujours été positive malgré les difficultés inhérentes

à chaque thèse. En plus, d'avoir toujours animé le bureau étudiant, grâce à toi, tu honorais ce bureau par un parfum délicat après un énième décapage de ton bureau. La poussière n'avait pas une chance et préférer se réfugier sur mon bureau, j'étais bien plus délicat avec elle. A l'heure où j'écris ces lignes, je sais que l'on sera amené à se revoir bientôt et j'en suis content.

Florence, indissociable partenaire des galères et, heureusement, de nombreux bons moments passés. C'est un exercice hautement difficile que décrire trois ans d'amitiés en quelques lignes. Combien de fois tu m'as aidé ? Coude à coude serré sous la hotte à passer les cellules, yeux vissés sur le microscope à compter les cellules pour évaluer la toxicité jusqu'à pas d'heure, aliquoter, transfecter, extraire, purifier, analyser ... Et je pense que l'analyse des données NGS est le meilleur exemple de nos moments passés. Toi et moi, face à d'austères fichiers textes dont on devait essayer d'en extraire la substance. On a fait beaucoup de tests et d'essais, parfois sans savoir trop ce qu'on faisait, mais toujours dans la bonne humeur en jouant à Hill Climb Racing. Si seulement /sudo pouvait être aussi efficace dans la vraie vie qu'en ligne de commande. Sans toi, ces années auraient été beaucoup plus dures.

Julian, Samuel, la colonne vertébrale de l'autre bureau étudiant. Je vous remercie d'avoir toujours été d'attaque et d'avoir transmis votre bonne humeur. Par exemple, toi, Samuel, face à la « non-stérilité » de la salle de culture cellulaire, tu as transformé ça en « les 10 commandements de la pièce de culture », affiche qui trône toujours sur l'étuve. Et toi, Julian, débusquer tes tasses de café aux endroits les plus loufoques a toujours été un challenge. Enfin, ce n'est pas tout. Tu as toujours accepté sans rechigner si j'avais besoin d'un coup de main, jonglant entre les responsabilités de l'hôpital et celles du laboratoire. D'ailleurs, tu vas bientôt avoir une casquette bien plus importante, celle de père, et je te souhaite tout le meilleur dans cette belle aventure.

Pour compléter, une véritable filiale a été encadrée par Sandrine. Etienne, Alexandre, Mario, Pascaline, Thomas, vous avez également tous contribué à la vie du laboratoire et je vous en remercie.

Enfin, heureusement que les nouvelles générations sont là pour prendre le relais et cette cuvée s'avère prometteuse. Sandra O., Pauline, Juliette, Sandra B., Claire, Mathilde, je vous souhaite tout le meilleur et j'espère que vous passerez aussi de précieux moments et de belles années.

Puisque j'écris ces lignes après avoir soutenu ma thèse, j'ai la chance d'être déjà réengagé dans un nouveau projet au sein de l'unité 1218. Changement d'unité, mais même couloir. Cette nouvelle aventure n'est permise que par la confiance que Valérie et Béatrice m'accordent et je les remercie de m'avoir donné cette opportunité.

Valérie, Béatrice, au fil de nos discussions, vous m'avez beaucoup fait progressé dans le domaine des CRISPR et également permis d'avoir accès au NGS. Je suis persuadé qu'à vos côtés, et avec Katie, Valérie L., Elodie, Stéphane, Bruno, Amal, Mathilde et Jordan, de nouveaux bons moments sont à venir.

Et puis les amis. Pour m'avoir fait l'honneur de venir à ma soutenance de thèse, sans vraiment tout y comprendre, je tiens à remercier mes amis les plus proches.

Marion, Jass, Jordy, Stéphane, Maxime et Robin (bien que tu n'aies pas pu venir), merci d'avoir été là pour les moments difficiles et ceux importants. Je pense qu'on n'a pas fini de partager des bons moments, pas toujours très fins ni intellectuels, mais qu'importe l'important, c'est de s'amuser. Je ne m'étendrai pas trop, je sais que vous avez la moquerie facile. Je remercie tout de même un peu plus Marion c'est la plus civilisée d'entre vous (et si ça peut semer de la jalousie, tant mieux).

Au cœur des relations professionnelles, amicales et familiales, Solène. S'il est toujours agréable de pouvoir partager déboires et réussites de vie de thésard à une oreille attentive et compréhensive, je pense que l'essentiel n'est pas là. Ces trois dernières années ont été marquées par une série de situations personnelles compliqués et de drames familiaux. Les vivre seul et loin de sa famille est une épreuve que l'on ne peut souhaiter à personne. Je ne te remercierai jamais assez d'avoir été là dans les moments difficiles. Le reste de la vie nous attend, ne perdons pas de temps pour en profiter.

Enfin, la famille. Nadine, Clément, Mathilde et pour toujours dans nos cœurs, Benoît. Ces dernières lignes sont les plus difficiles à écrire tant elles sont encore empreintes d'émotions vives d'une cicatrice encore ouverte. Nous avons toujours été soudés. Comment décrire tant de nos moments passés ensemble en si peu d'espace. Je ne crois pas qu'on puisse.

Table des matières

Remerciements	3
Figures et tableaux	10
Abréviations	12
Contexte scientifique et objectifs	16
Introduction	17
I. Concept et défis de la thérapie génique	17
i. Naissance du concept et définition	17
ii. Une thérapie d'avenir	17
iii. Immunodéficiences primaires et hémoglobinopathies : cibles de la thérapie génique	19
iv. Deux types de thérapies géniques	19
II. Thérapie génique additive	20
i. Les vecteurs viraux, support de la thérapie additive	20
ii. Essais cliniques	22
a. La <i>success story</i> pour la thérapie génique du DICS-ADA	22
b. Des résultats mitigés pour les autres déficits immunitaires primaires	22
iii. Vers des vecteurs viraux plus sûrs	24
a. Des vecteurs γ -rétroviraux améliorés	24
b. Les vecteurs lentiviraux, des vecteurs intégratifs davantage sécurisés	24
iv. Limites de la thérapie additive	27
a. Mutagénèse insertionnelle	27
b. Régulation transcriptionnelle	27
c. Maladies monogéniques dominantes	28
v. Vers l'édition des gènes ?	28
III. Edition génique par la Cas9-nucléase	28
i. Les endonucléases, clé de voûte de l'édition génique	29
a. Des méganucléases aux <i>TALENs</i>	29
b. Découverte des systèmes CRISPR-Cas	33
c. Le système CRISPR-Cas9	35

ii.	Vectorisation du système CRISPR-Cas9	41
a.	Vecteurs AAV	41
b.	Méthodes physiques	45
c.	Vers d'autres procédés de vectorisation innovants ?	47
iii.	Applications précliniques et cliniques basées sur CRISPR-Cas9	49
a.	Édition génique <i>ex vivo</i> : drépanocytose et β -thalassémie	49
b.	Édition génique <i>in vivo</i>	57
c.	Vers d'autres essais cliniques	65
IV.	CRISPR-Cas9, cassure double-brin et génotoxicité	65
i.	Génotoxicité <i>off-target</i>	66
a.	Caractériser le fonctionnement de la RNP Cas9:ARNg	66
b.	Caractériser l'activité <i>off-target</i>	69
c.	Répondre au défi de la spécificité	71
ii.	Génotoxicité <i>on-target</i> – Echelle du nucléotide	80
a.	Des InDels non-voulus à risque ?	80
b.	Connaitre les voies de réparation pour en maîtriser la finalité	81
c.	Détection des cassures double-brin	82
d.	Voies de réparation des cassures double-brin	84
e.	Quelle voie de réparation pour quelle matrice d'ADN exogène ?	96
f.	Régulation des voies de réparation endogènes	99
g.	La balance HDR / InDels.	105
h.	Vers un ratio HDR / InDels parfait ?	107
iii.	Génotoxicité <i>on-target</i> – Echelle du kilobases	107
a.	Des moyens de détections spécifiques	108
b.	Un risque à mesurer	108
iv.	Génotoxicité <i>on-target</i> – Echelle du mégabase et du chromosome	110
a.	Délétion à l'échelle du chromosome	110
b.	Délétion et inversion à l'échelle du mégabase	112
V.	Édition génique par la Cas9^{D10A}-nickase	114
i.	HDR et cassure simple-brin	114
a.	Cassure simple-brin ou cassure double-brin ?	116
b.	Deux voies de réparation	116
ii.	Une solution viable contre la génotoxicité ?	122
a.	Améliorer l'efficacité	122

b.	Evaluer la génotoxicité _____	122
	Résultats _____	123
I.	Objectifs du projet de thèse _____	123
II.	Modèle d'étude : la porphyrie érythropoïétique congénitale _____	123
i.	Une pathologie candidate à la thérapie génique _____	123
ii.	La PEC : un outil d'analyse pour l'édition génique précise _____	124
III.	Première partie : Comparaison de deux procédés d'édition génique _____	125
IV.	Second partie : Edition génique et délétion terminale _____	153
i.	Objectif : Création d'un outil sensible de détection des délétions terminales _____	153
a.	La PEC : un outil de détection des délétions terminales _____	153
b.	Questions scientifiques _____	155
ii.	Mise au point des outils expérimentaux _____	155
a.	Définition de la dose de 5-ALA _____	155
b.	Création d'une paire de lignées isogéniques pour étudier la dépendance à p53 _____	158
iii.	Résultats _____	160
a.	Délétion terminale, effet du statut p53 _____	160
b.	Outils d'édition et risque de délétion terminale _____	165
	Discussion et perspectives _____	167
	Quelles solutions pour éditer les cellules ? _____	167
i.	Edition génique par la Cas9-nucléase _____	167
ii.	Edition génique par la Cas9 ^{D10A} , une solution alternative viable ? _____	174
iii.	Vers d'autres méthodes d'édition génique indépendantes de la CDB ? _____	176
	Bibliographie _____	180

Figures et tableaux

- Figure 1. La thérapie génique, une stratégie thérapeutique en plein essor.
- Figure 2. Deux stratégies pour la thérapie génique additive.
- Figure 3. Vers des vecteurs γ -rétroviraux davantage sécurisés.
- Figure 4. Trois générations d'endonucléases spécifiques.
- Figure 5. Système CRISPR-Cas, une immunité adaptative.
- Figure 7. Structure et fonctionnement du système CRISPR-Cas9.
- Figure 6. Classification des systèmes CRISPR-Cas en deux classes et cinq types.
- Figure 7. Structure et fonctionnement du système CRISPR-Cas9.
- Figure 8. Activité endonucléasique de la RNP Cas9:crRNA:tracrRNA et activation des voies de réparation.
- Figure 9. Contourner la limite d'empaquetage des AAVs.
- Figure 10. Méthodes physiques de transfection.
- Figure 11. Procédés alternatifs de vectorisation.
- Figure 12. Mimer l'induction d'HPFH par une délétion de 13 nt.
- Figure 13. Mimer l'induction d'HPFH par une délétion de 13,6 kb.
- Figure 14. Suppression de l'expression érythroïde-spécifique de BCL11A.
- Figure 15. Correction directe de la mutation p.E6V par édition génique.
- Figure 16. Correction directe *in vivo* de la mutation de la tyrosinémie de type I.
- Figure 17. Correction directe *in vivo* du déficit en ornithine carbamyl transférase.
- Figure 18. Correction *in vivo* par excision d'exon de la myopathie de Duchenne avec deux cassures double-brin.
- Figure 19. Correction *in vivo* par excision d'exon de la myopathie de Duchenne avec une coupure double-brin.
- Figure 20. Correction *in vivo* par excision d'intron de l'amaurose congénitale de Leber type 10.
- Figure 21. Structure et fonctionnement de la RNP SpCas9:ARNg.
- Figure 22. Caractériser l'activité off-target.
- Figure 23. Des systèmes s'inspirant des générations précédentes d'endonucléases.
- Figure 24. Variants haute spécificité, eSpCas9-1.1 et SpCas9-HF1.
- Figure 25. Implication de REC3 dans la spécificité.
- Figure 26. Variant haute spécificité : High-Fidelity Cas9.
- Figure 27. Modification de l'ARNg pour améliorer la spécificité.
- Figure 28. Le complexe MRN et ATM, acteurs précoces de la reconnaissance des CDBs.
- Figure 29. Rôle central de MDC1 dans le recrutement de BRCA1 et 53BP1.
- Figure 30. Voies de réparation des cassures double-brin.
- Figure 31. Voie de la cNHEJ.
- Figure 32. La cNHEJ, un complexe dynamique spatialement.

Figure 33. Voie de réparation par RH.

Figure 34. Voies de réparation par SSA et Alt-EJ.

Figure 35. Système CRIS-PITCh v2.

Figure 36. Système HTTI.

Figure 37. Voies de réparation des CDBs par des ssODNs.

Figure 38. Cycle cellulaire et compétition entre 53BP1 et BRCA1.

Figure 39. Fonctionnement de 53BP1 pour inhiber la résection des brins.

Figure 40. Rôle des CDK et de BRCA1 pour activer la résection des brins

Figure 41. Unique CDB et délétions de plusieurs kilobases.

Figure 42. Unique CDB et délétions de plusieurs centaines de bases.

Figure 43. Double CDBs et délétions à l'échelle du chromosome.

Figure 44. Double CDBs et délétions à l'échelle du chromosome par activité off-target.

Figure 45. Voies de réparation des CSBs par HDR.

Figure 46. Implication du MMR dans l'HDR initiée aux CSBs.

Figure 47. Implication de RAD51 dans l'HDR initiée aux CSBs.

Figure 48. Méthode de détection des délétions terminales par la modification de la fonctionnalité d'UROS.

Figure 49. Mise au point de la dose de 5-ALA.

Figure 50. Création des deux lignées fibroblastiques.

Figure 51. Sensibilité des lignées fibroblastiques au 5-ALA.

Figure 52. Effet de p53 dans la fréquence des délétions terminales.

Figure 53. Validation méthodologique par biologie moléculaire.

Figure 54. Types de lésions et risque de délétions terminales.

Figure 55. Position génomique des gènes d'intérêts cliniques.

Figure 56. Modèle de formation des délétions terminales et des duplications inversées.

Figure 57. Vers d'autres méthodes d'édition génique ?

Table 1. Essais cliniques de thérapie génique additive des CSH utilisant des vecteurs γ -rétroviraux de première génération.

Table 2. Essais cliniques de thérapie génique additive des CSH utilisant des vecteurs SIN- γ -rétroviraux ou SIN-lentiviraux.

Table 3. Avantages et inconvénients des trois formes du système CRISPR-Cas9.

Table 4. Vers la correction d'autres maladies rares avec le système CRISPR-Cas9 par les entreprises leader du marché (CRISPR Therapeutics, Editas Medicine, Intellia Therapeutics).

Abréviations

AAV : *Associated Adeno-Virus*

Abraxas : *BRCA1 A Complex Subunit*

ABRO-1 : *Abraxas brother 1*

ADN : Acide désoxyribonucléique

Alt-EJ : *Alternative End-joining*

ARN : Acide ribonucléique

ARNg : ARN guide

ATM : *Ataxia Telangiectasia Mutated*

BARD1 : *BRCA1 associated RING domain 1*

BCL11A : *B-cell lymphoma/leukemia 11A*

BGH : *Bovine growth hormone polyadenylation*

BLM : *Bloom Syndrome RecQ Like Helicase*

BM1 : *Polycomb complex protein BMI-1*

BRCA1 : *Breast cancer 1*

BRCA2 : *Breast cancer 2*

CARF : *CRISPR-associated Rossmann fold*

CCND2 : *Cycline D2*

CCR5 : Récepteur à C-C chimiokine de type 5

CDB : Cassure double-brin

CFD : *Cutting Frequency Determination*

CHK2 : *Checkpoint Kinase 2*

cHS4 : *Chicken beta-globin 5'-DNase I-Hypersensitive Site 4*

cNHEJ : *Canonical non-homologous end-joining*

CRISPR : *Clustered Regularly Interspaced Short Palindromic Repeats*

crRNA : *CRISPR RNA*

CSH : Cellule souche hématopoïétique

CTD : *C-terminal domain*

CTiP : *CtBP-interacting protein*

dCas9 : *DeadCas9*

DHS : *DNase I hypersensitivity sites*

DICS-ADA : Déficit immunitaire combiné sévère par déficit en adénosine désaminase

DICS-X1 : Déficit immunitaire combiné sévère lié à l'X
DMD : Dystrophin
DMPK : *Dystrophia myotonica protein kinase*
DNA-PK : *DNA-dependent protein kinase*
EF1 α : *Elongation factor 1a*
ERCC1 : *ERCC Excision Repair 1*
EXO1 : Exonuclease 1
Fah : *Fumarylacetoacetate Hydrolase*
FAM175B : *Family With Sequence Similarity 175 Member B*
FANCD2 : *FA Complementation Group D2*
FANCM : *FA Complementation Group M*
GAPDH : D-glycéraldéhyde-3-phosphate déshydrogénase
GFP : *Green fluorescent protein*
GNP : Nanoparticule d'or
GSC : Granulomatose septique chronique
HBB : *Hemoglobin Subunit Beta*
HbF : Hémoglobine fœtale
HBG : *Hemoglobin subunit gamma-1*
HDR : *Homology Directed Repair*
HEK293 : *Human embryonic kidney 293*
HELB : *DNA Helicase B*
hGRK1 : *Human rhodopsin kinase 1*
HPFH : *Hereditary persistence of fetal hemoglobin*
IL2R γ : *Interleukin 2 receptor subunit gamma*
IL1RN : *Interleukin 1 Receptor Antagonist*
InDels : Insertions / Délétions
IRES : *Internal ribosome entry site*
iPSC : *Induced Pluripotent Stem Cell*
ITR : *Inverted Terminal Repeat*
KO : *Knock-out*
LCA10 : *Leber congenital amaurosis type 10*
LMO2 : *LIM Domain Only 2*
LS : *Long-synapse*

LTR : *Long terminal repeat*
MBS : *Métabisulfite de sodium*
MECOM : *MDS1 and EVI1 complex locus protein EVI1*
MDM2 : *Mouse double minute 2 homolog*
MDC1 : *Mediator Of DNA Damage Checkpoint 1*
MLL : *Mixed-Lineage Leukemia*
MMEJ : *Microhomology-mediated end-joining*
MN : *Méganucléase*
MRN : *MRE11, RAD50, NBS1 (Nijmegen Breakage Syndrome 1)*
MYBPC3 : *Myosin Binding Protein C*
NTBC : *2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione*
OGM : *Organisme génétiquement modifié*
OTC : *Déficit en ornithine transcarbamylase*
PALB2 : *Partner and localizer of BRCA2*
PAM : *Protospacer adjacent motif*
PAXX : *Paralog Of XRCC4 and XLF*
PBS : *Phosphate buffered saline*
PGK : *Phosphoglycérate kinase*
PIGA : *Phosphatidylinositol glycan anchor biosynthesis class A*
RAD51 : *RAD51 Recombinase*
RAG : *Recombination activating gene*
rAPOBEC1 : *Recombinant apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 1*
RH : *Recombinaison homologue*
RIF1 : *Replication Timing Regulatory Factor 1*
RNP : *Ribonucléoprotéine*
RPA : *Replication protein A*
RPE1 : *Retinal Pigmented Epithelial 1*
RS-1 : *RAD51 stimulator-1*
RVD : *Repeat Variable Diresidue*
SA : *Splice acceptor*
SD : *Splice donor*
SDSA : *Synthesis-dependent strand annealing*
SHLD : *Shieldin*

SIN : *Self-inactivated*
SMC1 : *Structural Maintenance Of Chromosomes 1A*
SS : *Short Synapse*
SSA : *Single-strand annealing*
ssODN : *Oligonucléotide simple-brin*
SORCS1 : *Sortilin Related VPS10 Domain Containing Receptor 1*
SV40 : *Virus simien 40*
SWA : *Syndrome de Wiskott-Aldrich*
TadA : *tRNA adenosine deaminase*
TALEN : *Transcription activator-like effector nuclease*
TBG : *Human thyroxine binding globulin*
THY1 : *Thy-1 Cell Surface Antigen*
TK : *Thymidine kinase*
TRA : *T Cell Receptor Alpha locus*
tracrRNA : *Trans-activating crRNA*
TRB : *T Cell Receptor Beta locus*
TP53 : *Tumor protein 53*
UGI : *Uracil DNA glycosylase inhibitor*
UROS : *Uroporphyrinogen-III synthase*
VIH : *Virus de l'immunodéficience humaine*
WSTF : *Williams–Beuren syndrome transcription factor kinase*
WWOX : *WW Domain Containing Oxidoreductase*
XLF : *XRCC4-like factor*
XRCC4 : *X-Ray Repair Cross Complementing 4*
ZFN : *Zinc Finger Nuclease*
ZFP : *Zinc Finger Protein*
5-ALA : *Acide 5-aminolévulinique*
53BP1 : *Tumor suppressor p53-binding protein 1*
Ψ : *Signal d'encapsidation*

Contexte scientifique et objectifs

La thérapie génique est une stratégie thérapeutique prometteuse pour le traitement des maladies monogéniques. Si les premières approches, dites additives, ont reposé sur l'utilisation de vecteurs viraux, une part grandissante se tourne désormais vers l'édition génique. Celle-ci est permise par la mise au point de nouvelles générations d'endonucléases, et en particulier le système CRISPR-Cas9. Moins d'une décennie après sa caractérisation, la révolution CRISPR-Cas9 a permis de faire passer l'édition génique à un stade clinique.

Toutefois, dans le même laps de temps, plusieurs interrogations ont été soulevées vis-à-vis de la génotoxicité pouvant être induite par la Cas9. Si l'activité hors-cible de la Cas9 a rapidement été décrite puis maîtrisée par le développement de variants à spécificité augmentée, une littérature émergente pointe le risque de génotoxicité au site ciblé. Le travail de thèse présenté ici s'inscrit dans cette thématique.

La première partie de l'étude a eu pour objectif de décrire la génotoxicité induite par une unique cassure double-brin faite par la Cas9. La caractérisation des effets a été faite à la fois à l'échelle nucléotidique, par le suivi de la balance HDR / InDels, mais également à l'échelle du chromosome. Le suivi de l'intégrité chromosomique a permis de mettre en lumière un nouveau risque de génotoxicité encore non-caractérisé. Un système de détection sensible et spécifique de ce risque a été mis au point pour continuer de le caractériser.

Le second objectif a été de répondre aux limites soulevées par la génotoxicité non-voulue, en mettant au point une méthode d'édition génique plus sûre et aussi efficace, via l'utilisation d'une unique cassure simple-brin par la Cas9D10A -nickase.

Introduction

I. Concept et défis de la thérapie génique

i. Naissance du concept et définition

La naissance du concept de thérapie génique émergea de la communauté scientifique dès les années 1970. Elle est définie par l'*European Medicines Agency* comme une substance active faite de, ou contenant des acides nucléiques recombinants administrés à des personnes en vue de réguler, réparer, remplacer, ajouter ou supprimer une séquence génétique. Cette substance doit avoir son effet thérapeutique, prophylactique ou diagnostique directement en lien avec la séquence qu'elle contient ou du produit de l'expression de celle-ci.

Cependant, de par la complexité technique sous-jacente, « l'accouchement de la thérapie génique fut long et difficile » pour reprendre les termes d'Alain Fischer sur la thérapie génique dans la Revue du Praticien. Du concept aux résultats concrets, il faudra attendre vingt ans pour avoir le premier succès concret de thérapie génique¹. Depuis, la thérapie génique est intimement liée à l'évolution de la génétique et de la biologie moléculaire des cinquante dernières années.

ii. Une thérapie d'avenir

Malgré l'apparition initiale d'effets secondaires aujourd'hui maîtrisés, la thérapie génique est désormais en plein essor avec plus de 2700 essais cliniques décomptés de 1989 à 2018 (*Figure 1A*). Ce chiffre s'explique par au moins 3 facteurs combinés.

La compréhension grandissante des fonctions, des structures et de la régulation du génome humain, ainsi que l'essor des technologies de séquençage sont des piliers dans la réussite de la thérapie génique.

En parallèle, le développement de nouveaux outils servant la thérapie génique, que ce soit des procédés de vectorisation ou des outils de modifications du génome, ont permis l'ouverture de nouvelles perspectives thérapeutiques.

Finalement, alors que l'espoir d'une thérapie génique a longtemps été majoritairement restreint aux maladies monogéniques hématologiques, la thérapie génique a pu opérer un élargissement thématique (*Figure 1B*). Désormais, les cancers, traités *via* l'immunothérapie ou par des virus oncolytiques, les maladies infectieuses, cardio-vasculaires, neurologiques, oculaires ou inflammatoires sont de nouvelles cibles pour la thérapie génique.

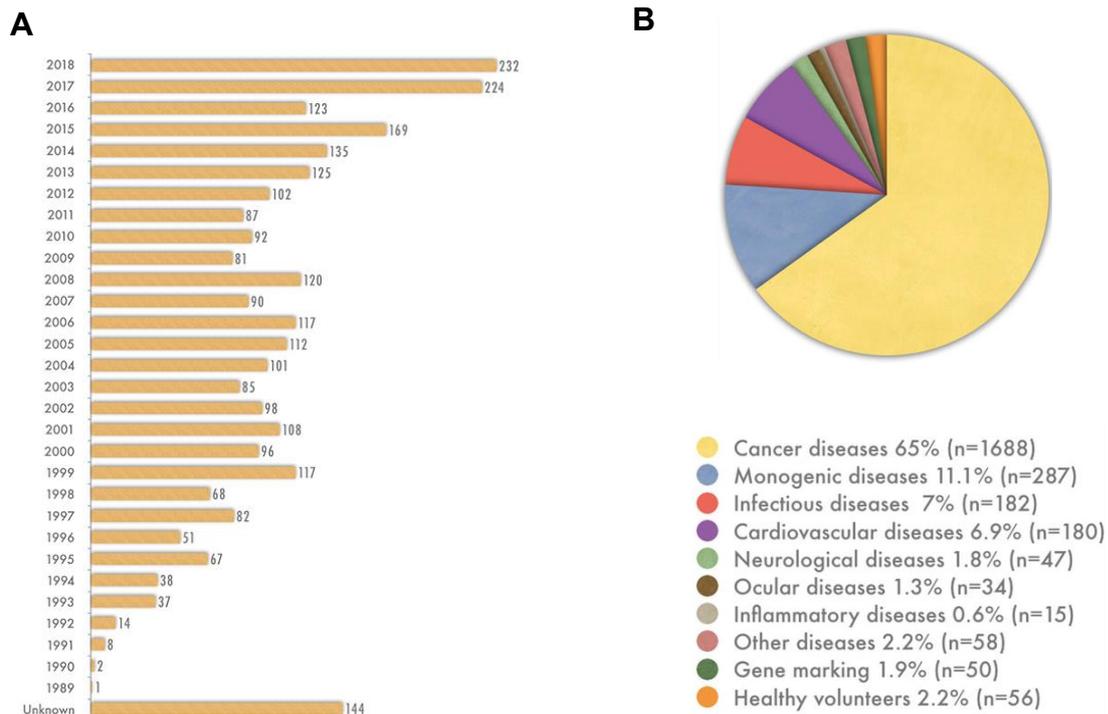


Figure 1. La thérapie génique, une stratégie thérapeutique en plein essor (d'après Wiley J and Sons Ltd.).
(A) Nombre d'essais cliniques de thérapie génique recensés par an entre 1989 et 2018. **(B)** Répartition des essais cliniques selon le type de pathologie traitée.

L'ensemble de ces facteurs ont su inscrire la thérapie génique dans les stratégies thérapeutiques d'avenir. Restant conscient des nombreuses pathologies candidates à une thérapie génique et de la diversité des stratégies thérapeutiques, l'accent sera ici mis sur la thérapie génique des maladies monogéniques et notamment, les immunodéficiences primaires et les hémoglobinopathies.

iii. Immunodéficiences primaires et hémoglobinopathies : cibles de la thérapie génique

Les maladies héréditaires monogéniques entraînant un dysfonctionnement du système hématopoïétique ont rapidement été candidates à la thérapie génique. Les meilleurs exemples sont les syndromes d'immunodéficiences primaires avec les déficits immunitaires combinés sévères par déficit en adénosine désaminase (DICS-ADA) ou lié à l'X (DICS-X1), le syndrome de Wiskott-Aldrich (SWA) et la granulomatose septique chronique (GSC) ainsi que les hémoglobinopathies (drépanocytose, β -thalassémies) et les porphyries.

Toutes ces pathologies n'ont pour seul traitement curatif que la greffe allogénique de moelle osseuse, permettant d'apporter des cellules souches hématopoïétiques (CSH) saines qui reconstitueront un système hématopoïétique fonctionnel. Cependant, les problèmes de disponibilités des greffons HLA-compatibles, les rejets de greffes et la morbidité associée rendent cette solution thérapeutique difficile à mettre en place.

La thérapie génique se propose d'être une stratégie thérapeutique alternative pour palier à cette problématique et profite de deux avantages principaux pour être une candidate robuste.

- La grande connaissance du système hématopoïétique, avec notamment une cellule souche multipotente mieux définie que pour d'autres tissus et la facilité d'accès à ce compartiment constituent un net avantage vis-à-vis des pathologies affectant d'autres tissus.
- La génétique de ces pathologies est bien établie et plus facilement corrigible que des maladies plus méconnues ou multifactorielles.

Profitant de ces deux atouts, la thérapie génique s'est positionnée comme une solution pour les maladies hématologiques notamment par une greffe autologue de CSH génétiquement modifiées.

iv. Deux types de thérapies géniques

Les procédés de thérapie génique peuvent être classés en deux catégories : la thérapie génique additive et l'édition génique.

Alors que la thérapie génique additive vise à introduire une copie saine du gène muté à l'origine de la pathologie, l'édition génique a pour but de corriger directement le gène d'intérêt.

Bien qu'elles suivent le même principe initial de soins par les gènes, ces deux approches distinctes possèdent des méthodologies et des problèmes respectifs.

Les paragraphes suivants seront consacrés à la description du principe de thérapie génique additive, des succès historiques et actuels ainsi qu'aux limites intrinsèques auxquelles les approches d'édition génique tentent de répondre. Ces dernières seront décrites plus en profondeur, l'essentiel des travaux exposés dans cette thèse y étant consacré.

II. Thérapie génique additive

La thérapie génique additive peut être décrite comme l'addition d'une séquence nucléique recombinante dans le génome d'une cellule-cible dont le produit d'expression a un effet thérapeutique. L'incorporation de ce « gène-médicament » peut être envisagée dans un protocole comprenant une modification *ex vivo* des cellules-cibles suivi d'une greffe autologue de ces dernières ou par une approche *in vivo* visant à modifier directement les cellules au sein de l'organisme (*Figure 2*). Dans le cas des pathologies hématologiques, l'accessibilité des cellules souches hématopoïétiques (CSH) et les techniques éprouvées de greffes ont permis de favoriser l'approche *ex vivo*.

Dans les deux types d'approches, la clé de voûte réside dans la capacité de transfert de l'ADN aux cellules cibles ainsi qu'à l'obtention d'une expression stable et maîtrisée.

i. Les vecteurs viraux, support de la thérapie additive

Parmi les différentes méthodes en transfert de gène, l'idée d'utiliser des virus comme vecteurs thérapeutiques s'est rapidement imposée². Les vecteurs viraux sont déficients pour la réplication mais conservent la capacité des virus à transporter du matériel génétique et à reconnaître et infecter des cellules. Ces atouts leur ont permis de s'imposer comme moyen de transfert préférentiel.

Dans le but de corriger des CSH, les vecteurs viraux capables de s'intégrer dans le génome offrent un avantage majeur en évitant la dilution de l'information génétique au cours des divisions cellulaires. Les vecteurs viraux intégratifs ont donc rapidement suscité un grand intérêt pour le traitement des maladies monogéniques pour lesquelles la stabilité dans le temps de l'effet thérapeutique demeure cruciale.

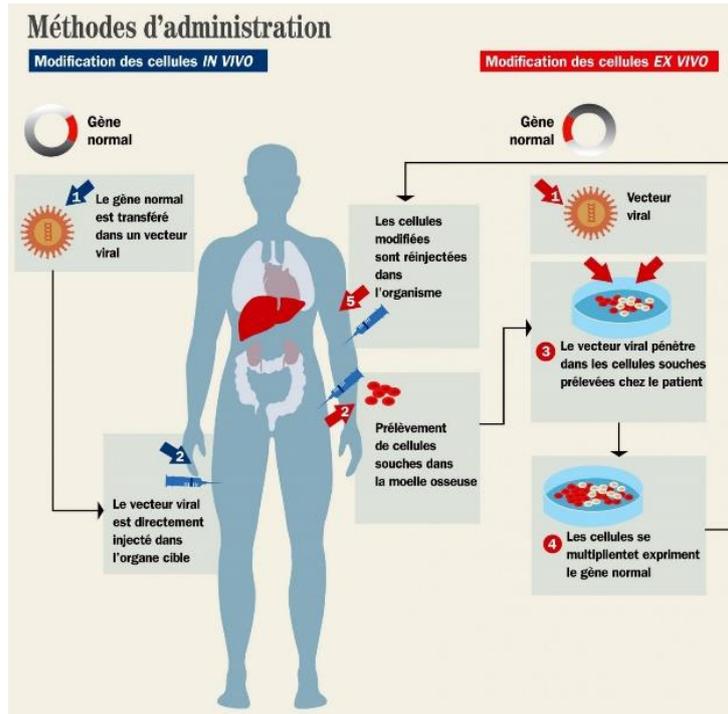


Figure 2. Deux stratégies pour la thérapie génique additive.

Illustration de la thérapie génique avec une approche *in vivo* (gauche) ou *ex vivo* (droite).

ii. Essais cliniques

Les vecteurs γ -rétroviraux ont été employés dès les années 90 lors des premiers essais cliniques de thérapie génique concernant les déficits immunitaires primaires.

a. La *success story* pour la thérapie génique du DICS-ADA

Le premier essai clinique recensé utilisant les vecteurs γ -rétroviraux a eu lieu en 1990 pour le traitement de deux jeunes patientes atteintes de déficits immunitaires combinés sévères par déficit en adénosine désaminase (DICS-ADA)¹. Depuis, cinq essais cliniques ont été conduits pour un total de 55 patients traités à l'aide de vecteurs γ -rétroviraux³ avec une efficacité de plus de 80% sans effets indésirables graves. Cette *success story* a permis l'obtention d'une autorisation de mise sur le marché du Strimvelis, premier médicament de thérapie génique *ex vivo* développé par GlaxoSmithKline en collaboration avec San Raffaele Telethon Institute for gene therapy.

b. Des résultats mitigés pour les autres déficits immunitaires primaires

Comme pour la thérapie génique du DICS-ADA, les essais cliniques concernant le traitement du (DICS-X1), du syndrome Wiskott-Aldrich (SWA) et de la granulomatose septique chronique liée à l'X (GSC-X) ont obtenus des résultats plus que prometteurs.

Toutefois, contrairement au DICS-ADA, des effets indésirables graves liés au caractère oncogénique des vecteurs γ -rétroviraux ont été observés (*Table 1*). Au total, hors patients atteints de DICS-ADA, pour 33 patients traités avec des vecteurs γ -rétroviraux, 18 ont développé des effets indésirables graves dont 4 sont décédés de leucémie^{4,5}.

Ces événements ont mis en exergue le principal écueil des vecteurs γ -rétroviraux : le risque de mutagénèse insertionnelle. Celui-ci provient essentiellement de deux caractéristiques. Les vecteurs γ -rétroviraux sont intégratifs. Cette intégration se fait de façon aléatoire dans le génome, avec une préférence d'intégration pour les séquences promotrices et *enhancers*. De plus, ils possèdent une forte capacité à modifier l'activité transcriptionnelle des gènes avoisinant le site d'intégration grâce aux *enhancers* contenus dans ses *Long Terminal Repeat (LTR)* rétroviraux^{4,6,7}.

Disease	Trial number (phase)	Starting year; site	Vector	Conditioning	Number of patients	SAEs (deaths) ^a
SCID-X1	NA	1999; France	gRV-IL-2R γ	No	10	5 (1)
SCID-X1	NA	2002; UK	gRV-IL-2R γ	No	10	1
WAS	DRKS00000330 (phase I/II)	2006; Germany	gRV-WAS	Yes	10	9 (3)
ADA-SCID	NCT00599781 (phase I/II)	2000; Italy	gRV-ADA	Yes	22	0
ADA-SCID	NCT03478670 (phase IV)	2017; Italy	gRV-ADA (Strimvelis)	Yes	5	0
ADA-SCID	NCT00018018 (phase I/II)	2001; USA	gRV-ADA	Yes (6); no (4)	10	0
ADA-SCID	NCT00794508 (phase II)	2009; USA	gRV-ADA	Yes	10	0
ADA-SCID	NCT01279720	2003; UK	gRV-ADA	Yes	8	0
CGD	NCT00564759	2004; Germany	gRV-CYBB	Yes	2	2 (1)
CGD	NCT00927134	2004; Switzerland	gRV-CYBB	Yes	1	1

ADA, adenosine deaminase; ADA-SCID, ADA severe combined immunodeficiency; CGD, chronic granulomatous disease; gRV, gammaretrovirus; HSPC, haematopoietic stem and progenitor cell; IL-2R γ , interleukin-2 receptor subunit- γ ; LTR, long terminal repeat; NA, not available; SAEs, serious adverse events; SCID-X1, X-linked SCID; WAS, Wiskott-Aldrich syndrome. ^aRelated to the drug product. ^bPersonal communication from A. Aiuti.

Table 1. Essais cliniques de thérapie génique additive des CSH utilisant des vecteurs γ -rétroviraux de première génération. (modifiée d'après Cavazzana et al., Nat. Rev. Drug Discov 2019)

Parmi les effets indésirables graves ayant mené aux décès de patients, 4 sont des leucémies (DICS-X1 et SWA) et l'un est une scépticémie (GSC).

La mise en évidence de ce double défaut a permis d'expliquer les effets délétères obtenus lors des premiers essais cliniques :

- Pour le DICS-X1, les analyses génétiques subséquentes ont montré une mutagenèse insertionnelle entraînant l'activation transcriptionnelle des oncogènes *LMO2*, *CCND2* et *BM1*^{5,8}.
- Pour le SWA, les leucémies observées sont associées à l'expansion clonale avec un vecteur intégré à proximité de *LMO2*, *MDS1* ou *MN1* entre autres⁴.
- Pour la GSC, le premier essai clinique a induit une expansion clonale par activation du locus *MECOM* (*MDS1* and *EVI1* Complex), se traduisant cliniquement par une myélodysplasie⁶. La même expansion clonale a été constatée pour les deux patients du second essai; une transplantation allogénique a alors été faite.

Ce travail de caractérisation a également constitué une première étape vers l'amélioration et la sécurisation des vecteurs viraux, nécessaires pour la réussite de la thérapie génique additive.

iii. Vers des vecteurs viraux plus sûrs

a. Des vecteurs γ -rétroviraux améliorés

Plusieurs modifications des vecteurs γ -rétroviraux ont permis d'en sécuriser leur utilisation^{9,10} (*Figure 3*) :

- Suppression de la région *enhancer*-promotrice U3 des LTR rétroviraux, remplacée par un promoteur interne (SIN-rétrovirus pour *Self-Inactivated* rétrovirus).
- Ajout d'insulateur type *chs4* (*chicken beta-globin 5'-DNase I-Hypersensitive Site 4*), pour isoler son environnement transcriptionnel des régions génomiques juxtaposées.
- Ajout d'un gène suicide permettant de neutraliser les cellules génétiquement modifiées en cas d'apparition d'effets secondaires délétères.

Un vecteur SIN- γ -rétroviral a ainsi été utilisé dans un nouvel essai clinique pour la DICS-X1 incluant 9 patients, sans effet indésirable grave (*Table 2*). Malgré les améliorations, les vecteurs γ -rétroviraux laissent de plus en plus leur place à une nouvelle génération de vecteurs intégratifs issus des lentivirus.

b. Les vecteurs lentiviraux, des vecteurs intégratifs davantage sécurisés

Développés à partir de 1996, les vecteurs lentiviraux, dérivés du VIH, ont su pallier à plusieurs difficultés rencontrées par les vecteurs rétroviraux.

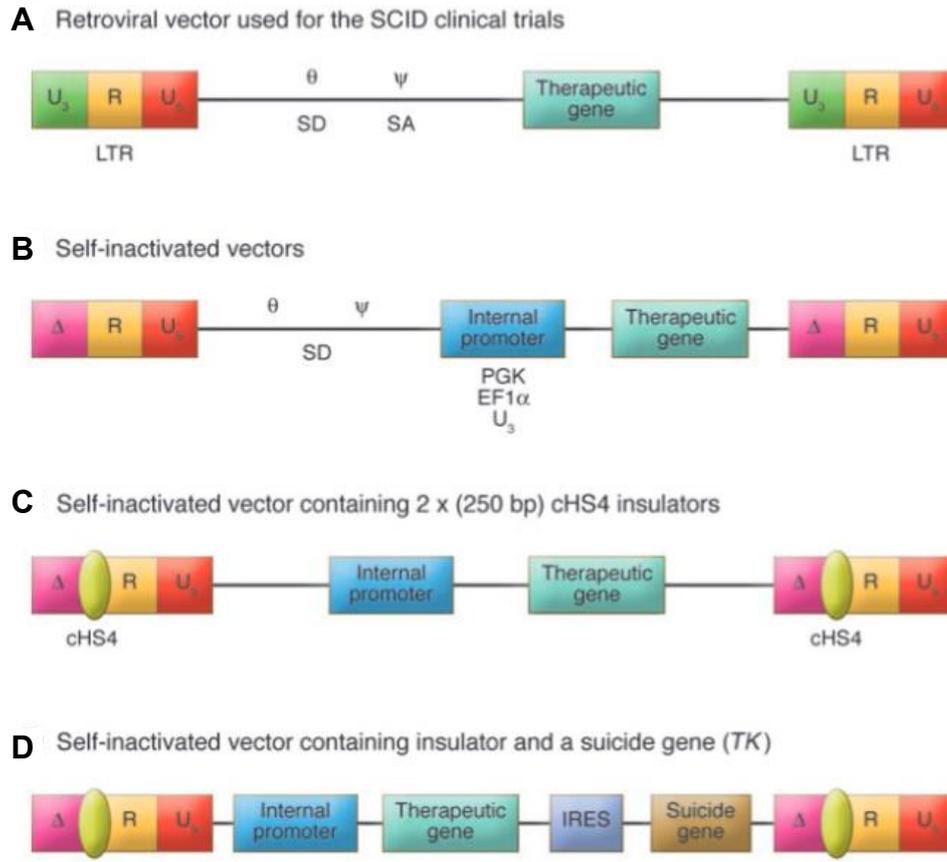


Figure 3. Vers des vecteurs γ -rétroviraux davantage sécurisés (Cavazzana-Calvo et Fischer, *J. Clin. Invest.* 2007)

(A à D) Représentation schématique des générations successives de vecteurs γ -rétroviraux : (A) vecteur γ -rétroviral de première génération, (B) vecteur SIN- γ -rétroviral (suppression région U3 LTR (Δ) et ajout promoteur interne), (C) ajout d'insulateur (cHS4), (D) ajout gène suicide (TK), précédé d'un IRES. La thymidine kinase permet le blocage de la réplication après ajout de Ganciclovir.

SIN, *Self-Inactivated* ; PGK, phosphoglycérate kinase ; EF1 α , *elongation factor 1 α* ; SD, *splice donor* ; SA, *splice acceptor* ; Ψ , signal d'encapsulation ; TK, Thymidine kinase ; IRES, *Internal Ribosome Entry Site*

Disease	Trial number (phase)	Starting year; sites	Vector	Conditioning	Number of patients
Primary immune deficiencies					
SCID-X1	NCT01410019, NCT01129544 (phase I/II)	2010; France, USA	SIN gRV EF1 α Sprom-IL-2R γ	No	9
SCID-X1	NCT01306019 (phase I/II)	2011; USA	SIN LV EF1 α Sprom-IL-2R γ	Yes	8
WAS	NCT01515462 (phase I/II)	2010; Italy	SIN LV WASprom-WAS	Yes	7
WAS	NCT01347346, NCT01347242 (phase I/II)	2011; France, UK	SIN LV WASprom-WAS	Yes	11
WAS	NCT01410825	2011; USA	SIN LV WASprom-WAS	Yes	4
ADA-SCID	NCT01380990, NCT01852071, NCT02999984 (phase I/II)	2012; UK, USA	SIN LV EF1 α Sprom-ADA	Yes	61
CGD	NCT02757911 (phase I/II)	2016; France	SIN LV Chimericprom-CYBB	Yes	2
CGD	NCT01855685 (phase I/II)	2013; UK, Germany, Switzerland	SIN LV Chimericprom-CYBB	Yes	4
CGD	NCT02234934 (phase I/II)	2015; USA	SIN LV Chimericprom-CYBB	Yes	3
ART-SCID	NCT03538899	2018; USA	SIN LV Artprom-DCLRE1C	Yes	NA
Haemoglobinopathies					
β -Thal	LG001 (phase I/II)	2007; France	SIN LV LCR- β prom- β -globin	Yes	2
β -Thal	NCT02151526 or HGB205 (phase I/II)	2013; France	SIN LV LCR- β prom- β -globin	Yes	4
β -Thal	NCT01745120 or HGB204 (phase I/II)	2013; USA, Australia, Thailand	SIN LV LCR- β prom- β -globin	Yes	18
β -Thal	NCT02453477 (phase I/II)	2015; Italy	SIN LV LCR- β prom- β -globin	Yes	9
β -Thal	NCT01639690 (phase I)	2012; USA	SIN LV LCR- β prom- β -globin	Yes	4
SCD	NCT02151526 or HGB205 (phase I/II)	2013; France	SIN LV LCR- β prom- β -globin	Yes	3
SCD	NCT02140554 or HGB206 (phase I)	2014; USA	SIN LV LCR- β prom- β -globin	Yes	9
SCD	NCT02247843 (phase I)	2014; USA	SIN LV LCR- β prom- β -globin	Yes	1
SCD	NCT03282656 (phase I)	2018; USA	SIN LV LCR- β prom-BCL11A shRNA	Yes	1
SCD	NCT02186418-phase I/II	2014; USA, Jamaica	SIN LV LCR- β prom- β -globin	Yes	2
Metabolic disorders					
X-ALD	NA	2006; France	SIN LV MNDprom-ABCD1	Yes	4
X-ALD	NCT01896102 (phase II/III)	2013; USA, UK, France	SIN LV MNDprom-ABCD1 (Lenti-D)	Yes	17
MLD	NCT01560182 (phase I/II)	2010; Italy	SIN LV PGKprom-ARSA	Yes	20
MPS-I	NCT03488394	2018; Italy	SIN LV PGKprom-IDUA	Yes	1
Congenital cytopenia					
Fanconi anaemia	NCT03157804 (phase I/II)	2016; Spain	SIN LV PGKprom-FANCA	No	4
Fanconi anaemia	NCT01331018 (phase I)	2012; USA	SIN LV PGKprom-FANCA	Yes	2

The promoter and cDNA used for each vector are described. ADA, adenosine deaminase; ADA-SCID, ADA severe combined immunodeficiency; ARSA, arylsulfatase A; ART-SCID, Artemis-deficient SCID; Artprom, Artemis promoter; β prom, β -globin promoter; β -thal, β -thalassaemia; CGD, chronic granulomatous disease; Chimericprom, chimeric promoter; EF1 α Sprom, elongation factor 1 α short promoter; gRV, gammaretrovirus; HSPC, haematopoietic stem and progenitor cell; IL-2R γ , interleukin-2 receptor subunit- γ ; LCR, β -globin locus control region; LV, lentivirus; MLD, metachromatic leukodystrophy; MNDprom, myeloproliferative sarcoma virus enhancer, negative control region deleted, dI587rev primer binding site substituted promoter; MPS-I, mucopolysaccharidosis type I; NA, not available; PGKprom, phosphoglycerate kinase 1 promoter; SCID-X1, X-linked SCID; shRNA, short hairpin RNA; SIN, self-inactivating; WAS, Wiskott-Aldrich syndrome; WASprom, WAS promoter; X-ALD, X-linked adrenoleukodystrophy. *Personal communication from A. J. Thrasher. †Personal communication from H. B. Gaspar. ‡M.C., unpublished data.

Table 2. Essais cliniques de thérapie génique additive des CSH utilisant des vecteurs SIN- γ -rétroviraux ou SIN-lentiviraux. (modifiée d'après Cavazzana et al., Nat. Rev. Drug. Discov., 2019).

Un vecteur SIN- γ -rétroviral à promoteur interne a été utilisé pour traiter le DICS-X1 (SIN gRV EF1 α prom, première ligne). Le reste des essais cliniques présentés ont utilisé des vecteurs SIN-lentiviraux (SIN LV).

Bien que relativement proches des rétrovirus en terme de structure et fonctionnalité, les vecteurs lentiviraux possèdent deux avantages majeurs :

- Les lentivecteurs sont capables de franchir activement la membrane nucléaire grâce à un complexe de pré-intégration leur permettant d'infecter des cellules en division mais également quiescentes⁹. Cette caractéristique est particulièrement intéressante lorsque la cellule-cible est la CSH.
- Leur intégration s'effectue préférentiellement dans des régions géniques plutôt que promotrices.

Débarrassés des facteurs de virulences, les vecteurs lentiviraux ont également bénéficié des travaux de sécurisation faits pour les vecteurs γ -rétroviraux pour former une nouvelle génération de SIN-lentivecteurs avec insulateurs et promoteur interne.

Toutes ces évolutions ont permis aux vecteurs lentiviraux d'améliorer le traitement par thérapie génique des syndromes d'immunodéficience primaire, des hémoglobinopathies (β -thalassémies et de drépanocytose), de l'anémie de Fanconi¹¹ et de plusieurs maladies métaboliques (*Table 2*). Au total, plus de 150 patients ont été traités par des vecteurs lentiviraux dans le cadre d'une thérapie génique *ex vivo* visant les CSH sans qu'aucun effet indésirable grave n'ait été décrit³.

iv. Limites de la thérapie additive

La sécurisation des vecteurs viraux intégratifs a conduit à des succès majeurs pour la thérapie additive. Malgré tout, plusieurs phénomènes et limites sont à prendre en considération :

a. Mutagénèse insertionnelle

L'utilisation de vecteurs intégratifs entraîne un risque inhérent de mutagénèse insertionnelle. La première génération de vecteurs γ -rétroviraux en est le meilleur exemple. Bien que plus sécurisés, les SIN-lentivecteurs s'intègrent également dans le génome et peuvent être à l'origine d'expansions clonales¹², de défauts d'épissage dus aux particules s'intégrant dans des régions introniques^{13,14} et conservent un potentiel oncogénique non-nul¹⁵. L'intégration dans des loci, dits « *safe harbor* », pourrait permettre de diminuer le risque oncogénique. Dans cette optique, la technologie des cellules souches pluripotentes induites (iPSCs) pourrait permettre une maîtrise du nombre de copie et du site d'intégration^{16,17}.

b. Régulation transcriptionnelle

Une des autres limites intrinsèques des vecteurs rétroviraux est la taille maximale d'information génétique pouvant être portée par ces derniers. La séquence d'ADN portée par les vecteurs rétroviraux est souvent constituée d'un ADNc du gène d'intérêt couplé à un promoteur constitutif

ou tissu-spécifique. Cependant, cette construction ne permet pas de reconstituer la complexité de la régulation physiologique du gène à son locus endogène³.

c. Maladies monogéniques dominantes

Dans un cadre plus large, le traitement des maladies monogéniques dominantes est un défi pour la thérapie génique additive. Dans cette situation, l'apport d'une copie saine du gène peut ne pas suffire à la résolution de l'état pathologique.

v. Vers l'édition des gènes ?

Pour faire face à ces limites, et grâce au développement de nouveaux outils de « chirurgie » du génome, la thérapie génique par édition génique a connu un regain d'intérêt ces dernières années. La maîtrise de ce procédé permettrait de se défaire de plusieurs verrous techniques et notamment de :

- S'affranchir du risque de mutagenèse insertionnelle par l'utilisation des méthodes de transfection non-intégratives.
- Conserver la régulation physiologique offerte par le locus endogène
- Corriger n'importe quel gène indépendamment de sa taille en offrant une stratégie unique pour tous les « grands gènes ».
- Traiter les maladies monogéniques dominantes. Un exemple récent est la correction du gène *DMPK* dans un modèle murin de dystrophie myotonique de type I¹⁸.

III. Edition génique par la Cas9-nucléase

L'idée d'une correction directe de la mutation causale au gène endogène a accompagné l'émergence du concept de thérapie génique. Cependant la difficulté technique sous-jacente et la faible disponibilité d'outils technologiques ont créé un décalage entre la naissance du concept et sa mise en pratique expérimentale jusqu'à un stade clinique. Dans le cadre des maladies monogéniques, l'idéal poursuivi par l'édition génique est un système « *hit and run* » offrant une correction permanente et sans cicatrice de la mutation grâce à un apport transitoire des outils d'édition du génome.

Le succès de l'édition génique dépend de trois axes principaux et seront décrits dans une première partie. Ces derniers sont :

- Le développement d'endonucléases permettant d'induire des cassures double-brin spécifique dans le génome, avec en particulier le système CRISPR-Cas9.
- La compréhension et la maîtrise des voies de réparation des cassures double-brin

- La mise au point de méthodes de vectorisation non-intégratives.

Moins d'une décennie après l'utilisation du système CRISPR-Cas9 pour l'édition *in vitro* des cellules eucaryotes, l'édition génique est passée à un stade clinique pour plusieurs pathologies monogéniques.

i. Les endonucléases, clé de voûte de l'édition génique

La preuve de faisabilité d'une modification ciblée a été faite par Smithies et al.¹⁹ dès 1985, sans induire de cassure double-brin (CDB). Cependant cette preuve de concept a été associée à une très faible efficacité (10^{-6} à 10^{-9} dans cellules eucaryotes). Rapidement, Choulika et al. (1995)²⁰, Donoho et al. (1998)²¹ et Porteus et Baltimore (2003)²², ont démontré que l'induction ciblée d'une CDB permettait d'augmenter la fréquence de modification d'un facteur 100 à 1000. Ces études marquent les premières utilisations d'endonucléases en vue d'édition génique et ont fait basculer le domaine de l'édition génique dans l'ère de l'ingénierie des endonucléases.

L'édition génique est depuis intimement liée à l'utilisation d'endonucléases spécifiques. Ces dernières sont capables d'induire des cassures double-brin (CDB) de l'ADN de manière site-spécifique. Des années 80 à maintenant, quatre grandes générations d'endonucléases se sont succédées. Si les trois premières (méganucléases, *zinc finger nucleases* (ZFNs) et *Transcription Activator-Like Effectors Nuclease* (TALENs)) ont permis de développer l'édition génique, l'avènement du système CRISPR-Cas9 l'a révolutionnée.

a. Des méganucléases aux TALENs

Comme vu ci-dessus, l'induction d'une CDB ciblée s'est rapidement imposée comme la clé de voûte de la réussite de l'édition génique. Cela a conduit tout un pan de la recherche en génétique à découvrir et développer des endonucléases pour en améliorer l'efficacité et simplifier leur utilisation.

Méganucléases

Les premières études ont utilisé la méganucléase I-Sce et I-Cre (*Figure 4A*). Les méganucléases sont des endonucléases ayant la particularité de posséder un site de restriction de grande taille de 14 à 40 nucléotides. Cette caractéristique leur offre une spécificité de coupure très élevée ; la probabilité de rencontrer plusieurs sites de reconnaissance par génome étant quasi-nulle.

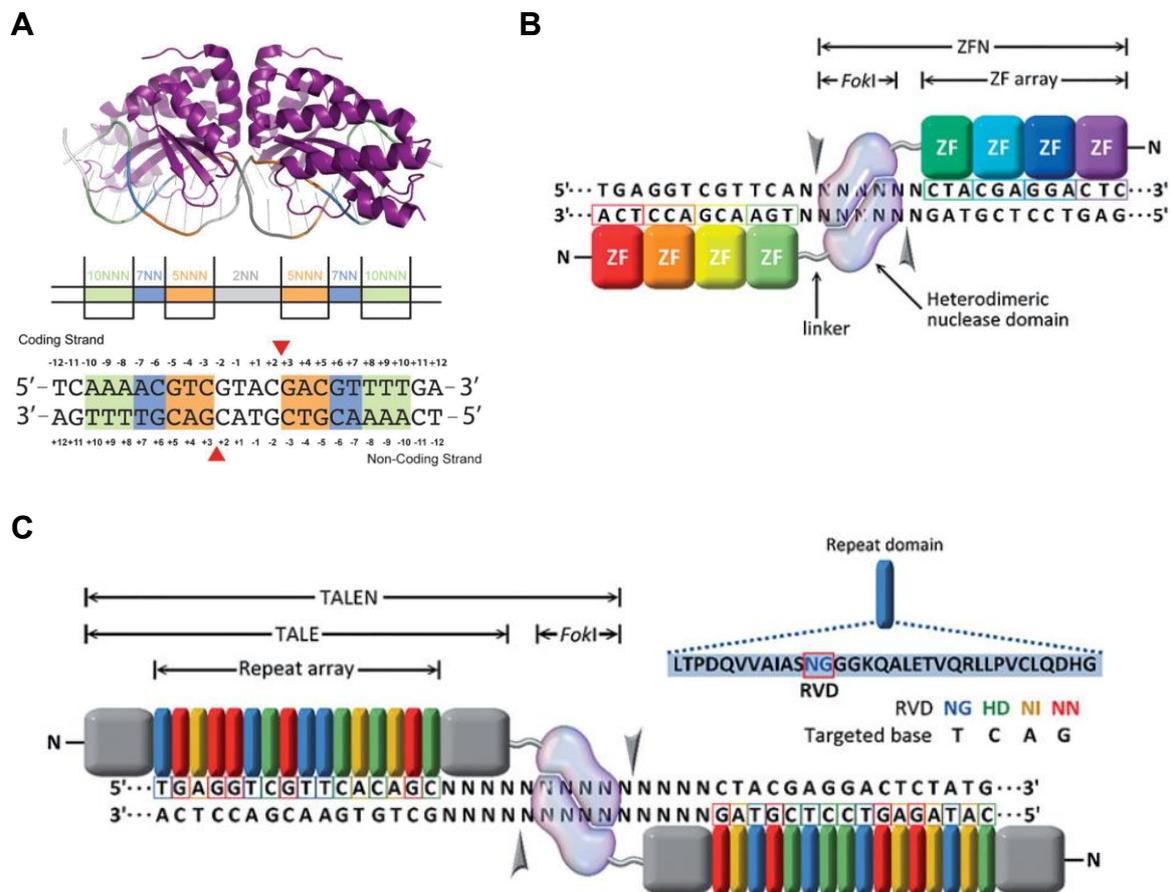


Figure 4. Trois générations d'endonucléases spécifiques.

(A) Représentation de la **méganucléase I-Cre** en contact avec une séquence d'ADN double-brin. La I-Cre fonctionne en homodimère et chaque monomère est capable d'interagir avec 3 régions de la séquence nucléotidique cible : 10NNN (vert), 7NN (bleu), 5NNN (orange). La cassure double-brin est induite au niveau de la région 2NN, indiquée par les triangles rouges. (Prieto et al., *Sci. Rep.*, 2018)

(B) Représentation des **nucléases à doigt de zinc (ZFN pour Zinc Finger Nuclease)**. Les ZFNs sont composées d'une ZFP avec plusieurs domaines en doigt de zinc (ZF array) fusionnée à FokI. Chaque domaine en doigt de zinc reconnaît un triplet nucléotidique. L'hybridation de deux ZFNs à leur séquence cible respective, située de part et d'autre de la coupure et sur chacun des brins, est nécessaire pour permettre la dimérisation et l'activation de l'endonucléase FokI. Le site de coupure est indiqué par les flèches grises. (Lino et al., *Drug. Deliv.*, 2018)

(C) Représentation des **Transcription Activator-Like Effector Nucleases (TALENs)**. Les TALENs sont composés d'un domaine TALE fusionné à FokI. Chaque TALE possède un domaine constitué de répétitions en tandem (repeat array). Chaque repeat possède des acides aminés hypervariables (RVD) situés en position 12 et 13 (encadrée en rouge). Ces deux résidus donnent la spécificité de reconnaissance d'un des 4 nucléotides. Chaque repeat reconnaît donc un unique nucléotide. L'hybridation de deux TALENs à leur séquence cible respective, située de part et d'autre de la coupure et sur chacun des brins, est nécessaire pour permettre la dimérisation et l'activation de l'endonucléase FokI. Le site de coupure est indiqué par les flèches grises. (Lino et al., *Drug. Deliv.*, 2018)

Cependant, chaque méganucléase possède une séquence de reconnaissance spécifique, rendant le nombre de cibles potentielles limité par le répertoire disponible. Des efforts de mutagenèse des méganucléases, pour modifier leur séquence-consensus, et la formation d'hybrides à partir de plusieurs enzymes différentes ont permis d'accroître le nombre de cibles potentielles^{23,24}. Malgré cela, le design d'une méganucléase pour un locus d'intérêt nécessite une ingénierie protéique complexe basée sur une modélisation *in silico* suivi d'une sélection *in vitro*²⁵. Les méganucléases ont notamment été utilisées pour faire une édition génique *in vitro* au locus *RAG1*, impliqué dans le DICS par déficit complet de *RAG1/2*, dans des HEK293. Une correction *in vivo* de CSH murines a été faite dans un modèle de souris déficient pour *Artemis*²⁶.

Nucléases à doigt de zinc

Les premières nucléases à doigts de zinc (ZFN pour *Zinc Finger Nuclease*) ont été synthétisées dès 1996²⁷. Une ZFN est composée d'une protéine à doigt de zinc (ZFP pour *Zinc Finger Protein*), assurant la reconnaissance de la liaison à l'ADN, et d'une endonucléase non-spécifique *FokI* (Figure 4B). La ZFP est constituée de 3 à 6 domaines en doigt de zinc permettant chacun la reconnaissance d'une séquence spécifique de 3 à 4 nucléotides. La mutation d'acides aminés clés dans le doigt de zinc permet de conférer un site de reconnaissance unique à chaque variant²⁷. La fusion en tandem de différents variants permet de créer une ZFP reconnaissant spécifiquement une séquence d'ADN de 9 à 24 nucléotides. L'autre versant des ZFNs est leur activité endonucléasique. Celle-ci est conférée par la fusion de la ZFP au domaine *FokI*. *FokI* est une endonucléase non-spécifique, fonctionnant en dimère. Cette contrainte impose l'utilisation d'une paire de ZFNs, dont l'orientation et l'espacement au locus-cible doit permettre la dimérisation de leur domaine *FokI* respectif pour induire la CDB. Néanmoins, cette obligation permet d'augmenter grandement la spécificité du système. En effet, une ZFN seule ne permet pas d'induire de CDB, réduisant ainsi le risque de CDB aspécifique.

L'étude conduite par Urnov et collègues mérite d'être soulignée et mise en perspective avec la thérapie génique additive²⁸. Le but de cette étude a été de faire la preuve de principe d'une correction par édition génique d'une mutation du gène *IL2R γ* impliqué dans le DICS-X1. Pour ce faire, les auteurs ont utilisé une paire de ZFNs ciblant le gène *IL2RG* ainsi qu'une matrice plasmidique portant une mutation silencieuse. En utilisant cette méthodologie, ils ont obtenu 5,3% de modification génique dans des lymphocytes T CD4⁺. Il est à noter que les ZFNs sont la première technologie à avoir porté l'édition génique au stade clinique²⁹. L'essai clinique a consisté à éteindre l'expression du gène *CCR5*, codant pour le co-récepteur nécessaire à l'entrée du virus d'immunodéficience humaine, espérant mimer le génotype du « patient de Berlin » (mutation $\Delta 32$, insensibilisant au VIH). Pour ce faire, des lymphocytes T CD4⁺ ont été modifiés *ex vivo*, puis une

greffe autologue des cellules modifiées a été effectuée. Les résultats, encourageants, ont montré une diminution, de faible à totale, de l'ARN messager du VIH. Les cellules modifiées étant des cellules différenciées, l'amélioration phénotypique a été transitoire.

Cependant, comme pour les méganucléases, la complexité de leur mise en place limite leur utilisation.

Transcription Activator-Like Effectors Nuclease (TALEN)

Le déchiffrement du code des *Transcription Activator-Like Effector (TALE)*³⁰ et leur conversion en outil génétique a permis de détrôner les ZFNs, en sachant conserver leurs avantages tout en simplifiant leur design. Les TALE sont des facteurs de virulence de *Xanthomonas sp.*, un pathogène de plantes qui modifie le programme transcriptionnel des cellules hôtes en mimant les facteurs de transcription eucaryotes³¹. Les TALEs sont caractérisés par un domaine central composé de répétitions en tandem, de signaux de localisation nucléaire et d'un domaine d'activation de la transcription. Le domaine central est nécessaire à la liaison à l'ADN. Les séquences répétées du domaine central sont composées de 33 à 35 acides aminés, extrêmement conservés au sein des différentes répétitions, à l'exception des résidus aminés situés en position 12 et 13 (*RVD* pour *Repeat Variable Di-residue*). La combinaison d'acides aminés qui composent le RVD définit quel nucléotide peut être reconnu par la séquence répétée. Ainsi, l'ensemble des RVD du domaine central permettent au TALEs de reconnaître une séquence d'ADN spécifique.

Leur conversion en outil génétique par Christian et al. s'est largement inspirée des ZFNs³² (*Figure 4C*). Pour ce faire, les TALEs, débarrassés du domaine d'activation de la transcription, ont été fusionnés par leur extrémité C terminal à l'endonucléase *FokI* aboutissant à la formation de protéines de synthèse, les TALENs. Comme pour les ZFNs, une paire de TALENs correctement espacée et orientée est nécessaire pour induire une CDB.

Ce système d'édition génique a permis l'édition de lignées cellulaires³³, de cellules souches pluripotentes induites humaines³⁴ *in vitro* et également l'édition de cellules *in vivo*³⁵. Cependant, la difficulté de synthèse et de validation des TALENs ont limité la diffusion de ce système au sein de la communauté scientifique³⁶, qui sans le savoir, s'apprêtait à s'emparer du système CRISPR-Cas9.

b. Découverte des systèmes CRISPR-Cas

Un système immunitaire bactérien

A ses débuts, l'étude du système CRISPR-Cas n'était pas prédestinée à venir révolutionner le monde de l'édition génique. Sa description a commencé dès 1987³⁷ par l'identification de courtes répétitions (*repeats*) séparées par des petites séquences d'ADN chez *E. coli*. Ces séquences seront définies plus tard par le terme *Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)*. Sa fonctionnalité n'a pas été immédiatement expliquée. Un premier indice est venu du fait que les séquences courtes (*spacers*) séparant les régions répétées étaient issues de plasmides et de bactériophages viraux. Par la suite, il a été montré que les gènes *cas* associés au locus CRISPR codaient pour des protéines avec des domaines nucléases et hélicases.

La première preuve expérimentale du rôle du système CRISPR-Cas a été apportée en 2007 par l'équipe de Philippe Horvath³⁸. En infectant la bactérie *Streptococcus thermophilus* avec des bactériophages, ils ont pu démontrer que le système CRISPR-Cas servait de système immunitaire adaptatif gardant en mémoire une marque génétique (les *spacers*) des infections passées.

La compréhension du système a été affinée l'année suivante. Dans *E. coli*, les *CRISPR RNAs (crRNAs)*, issus de la transcription et de la maturation du locus *CRISPR*, ont été identifiés comme des « guides » permettant de diriger les protéines Cas contre l'ADN viral³⁹, entraînant sa dégradation directement par les Cas⁴⁰.

Au final, la fonctionnalité du système CRISPR-Cas peut être décrite comme un mécanisme comprenant trois étapes : adaptation, biogénèse des crRNA et inférence avec l'ADN exogène⁴¹ (*Figure 5*).

Enfin l'une des particularités du système CRISPR-Cas est le rôle du *protospacer adjacent motif (PAM)*. Celui-ci est une courte séquence de quelques nucléotides toujours présente sur l'ADN invasif en amont des *protospacer*. Contrairement aux *spacers*, le PAM n'est pas intégré au locus CRISPR. Or, la reconnaissance de ce dernier est nécessaire à la dégradation de l'ADN invasif lors de la phase d'interférence. Ce mécanisme permet de protéger le locus CRISPR de la bactérie vis-à-vis de son système d'interférence.

Une pluralité de systèmes CRISPR-Cas

Preuve de leur intérêt évolutif, au moins un locus CRISPR est retrouvé dans plus de la moitié des génomes bactériens et dans 83% des Archées⁴².

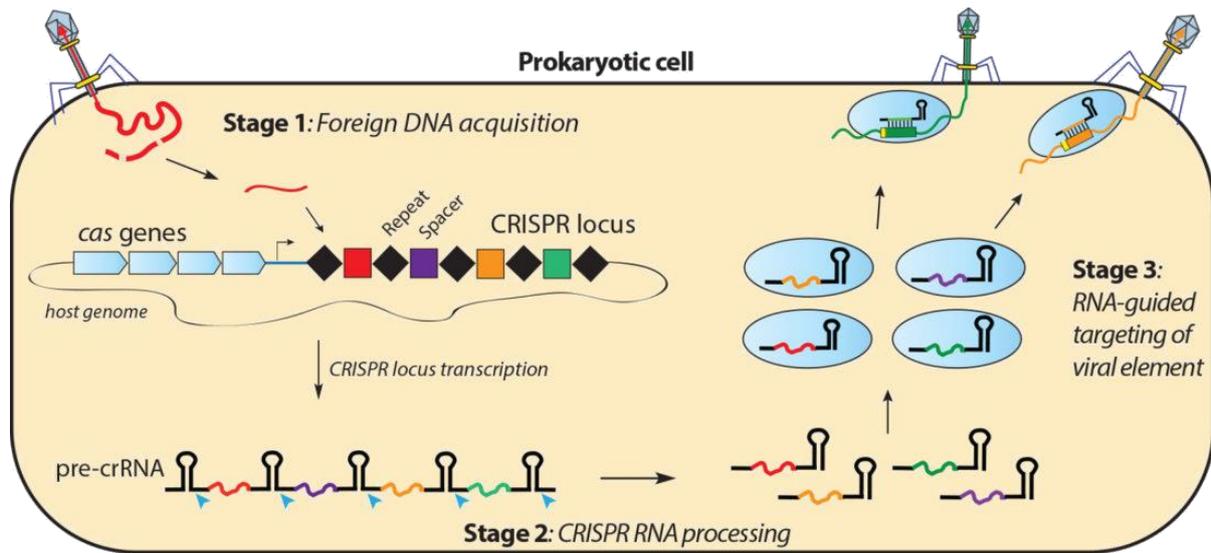


Figure 5. Système CRISPR-Cas, une immunité adaptative. (<http://doudnalab.org/>)

L'immunité adaptative offerte par le système CRISPR-Cas se déroule en 3 étapes successives : adaptation, biogenèse des crRNA, ciblage de l'ADN exogène.

i) L'adaptation permet l'acquisition de nouveaux *spacers* issus d'ADN exogène invasif.

Cette étape débute par l'intervention des protéines Cas, capables de venir prélever un fragment d'ADN viral invasif, appelé *protospacer*. Ces *protospacers* possèdent la particularité d'être juxtaposés à une petite séquence (de 2 à 5 nt), très conservée, appelée *protospacer adjacent motif* (PAM). Ces *protospacers* vont être ensuite intégrés au locus CRISPR pour former un nouveau *spacer* au sein du génome bactérien entre deux séquences répétées (*repeat*) et à l'extrémité *leader-end*. Le PAM n'est pas intégré au locus CRISPR, permettant à la bactérie d'échapper à ses propres endonucléases.

ii) La seconde étape est la biogenèse des crRNAs.

L'action des systèmes CRISPR-Cas est dirigée par les crRNAs. Ces derniers sont issus de la transcription du locus CRISPR, formant un pré-crRNA, et de la maturation du pré-crRNA. Les caractéristiques de taille et de structure secondaire des crRNAs, ainsi que leur maturation sont dépendantes du type de CRISPR-Cas.

iii) Pour finir, les crRNAs s'associent aux protéines Cas pour interférer avec l'ADN invasif.

Les crRNAs s'associent avec une ou plusieurs protéines Cas pour former des RNPs. Les protéines Cas proviennent de l'expression de l'opéron *Cas*. Les RNPs sont capables d'induire des lésions à l'ADN et/ou l'ARN invasif(s) au niveau de séquences complémentaires au *spacer* du crRNA (région colorée des crRNA) (Barrangou Mol Cell 2014). Encore une fois, le nombre de protéines Cas associées aux crRNA, leur fonctionnalité et le type de lésions fait à l'ADN ou l'ARN viral sont dépendants du type de CRISPR.

Toutefois, ses caractéristiques varient fortement d'une espèce à l'autre, que ce soit la longueur des *repeats* (de 23 à 55 nt), le nombre de *spacers* (majoritairement inférieur 50 et jusqu'à 588 pour *Haliangium ochraceum*) leur taille (21-71nt), du PAM associé, mais également le nombre de loci CRISPR au sein d'un même génome (majoritaire 1 à 2 loci et 25 loci putatifs dans *Methanoterris igneus*).

La compréhension plus fine du système CRISPR-Cas et la mise en lumière d'une grande diversité ont mené Makarova et al. à proposer une première classification⁴³, complétée en 2015⁴⁴, basée sur la fonctionnalité des protéine Cas, formant 2 grandes classes (*Figure 6*). En vue d'édition génique, les systèmes de classe 2 possèdent un intérêt notable. Pour la classe 2, l'activité endonucléasique est assurée par une unique protéine, là où pour la classe 1, une coopération entre plusieurs protéines est nécessaire. La vectorisation des systèmes de classe 2 s'avère moins complexe que pour ceux de classe 1.

Les deux classes sont subdivisées en 5 types, identifiés par la présence de gènes *cas* « signatures ». Il est à noter qu'à ces cinq types, un sixième type faisant partie de la classe 2 a été mis en lumière en 2018 et ajouté à la classification⁴⁵. Ce dernier possède la particularité de viser et de dégrader les molécules d'ARN, sans interférer avec l'ADN^{46,47}. Enfin pour chaque type, plusieurs sous-types sont identifiés dû à la présence ou l'absence de gènes *Cas* supplémentaires, non-essentiels, dans l'opéron *Cas*.

Les deux études conduites par Jennifer Doudna et Emmanuelle Charpentier⁴¹ d'une part, et Feng Zhang⁴⁸ d'autre part, ont su propulser le système CRISPR-Cas dans le monde de l'édition génique en utilisant le système CRISPR-Cas9 pour modifier des cellules eucaryotes. Depuis, d'autres systèmes (Cas3, Cas12a, Cas12k) ont également été utilisés pour modifier des cellules eucaryotes et pourraient présenter plusieurs avantages par rapport à la Cas9. Toutefois, le groupe le plus représenté dans la littérature et dans les essais cliniques est issu du système CRISPR-Cas9.

c. Le système CRISPR-Cas9

Le système CRISPR-Cas9 le plus utilisé est issu de *S. pyogenes* (parfois abrégé en SpyCas9 ou SpCas9) et fait partie de la classe 2 - type II – sous-type A⁴⁹. Il est composé d'un opéron de gènes *Cas*, encodant notamment la protéine Cas9, et d'un locus CRISPR particulier, formant des crRNA et des tracrRNA. L'expression de ces deux composantes mène à la formation d'une RNP capable d'induire une cassure double-brin spécifique.

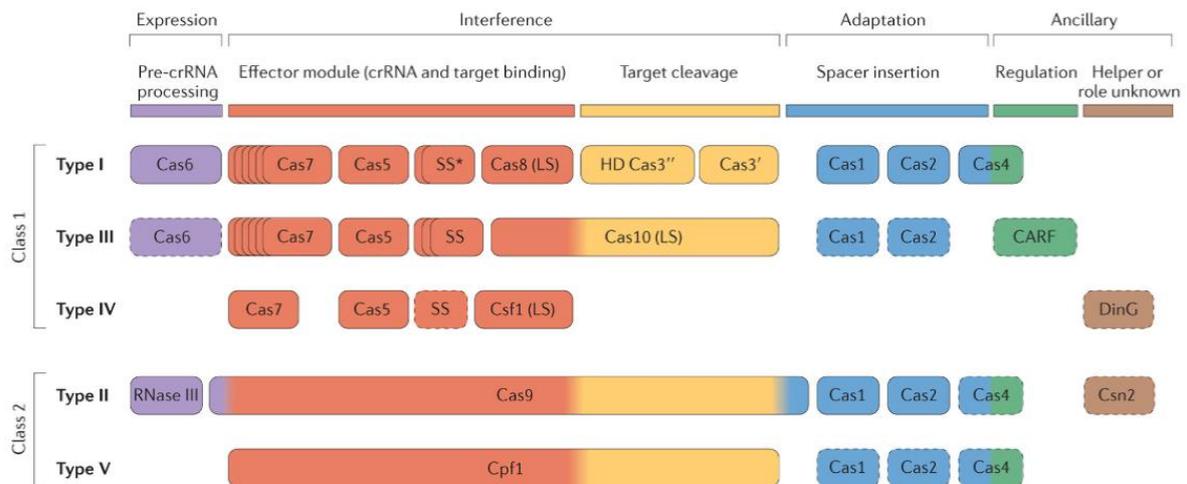


Figure 6. Classification des systèmes CRISPR-Cas en deux classes et cinq types. (Makarova et al., Nat. Rev. Microbiol., 2015)

Le premier niveau de classification définit deux classes. **Les systèmes de classe 1** sont définis par la présence d'un **complexe crRNA- effecteurs à plusieurs sous-unités**. La classe 1 comprend les systèmes de type I et de type III, ainsi qu'un nouveau type putatif encore peu documenté, le type IV. **Les systèmes de classe 2** sont définis par la présence **d'un module effecteur unique couplé au crRNA**. Cette classe comprend les systèmes de type II et le type V.

Le second niveau de classification est la répartition en **type**, chacun possédant un **gène cas « signatures »**. Les gènes **Cas3**, **Cas9** et **Cas10**, **Csf1**, **Cpf1** (appelé également **cas12**) sont les marques respectives des types I à V. Le type IV est encore peu documenté. Le type VI (non-représenté ici), ciblant l'ARN, appartient à la classe 2 et possède le gène **Cas13**.

Les gènes accessoires, non-nécessaires, sont encadrés en pointillés. Leur présence permet de définir les différents **sous-types**, le troisième niveau de classification (non-représenté ici). Par exemple, le système CRISPR-Cas de *Streptococcus pyogenes* et de *Staphylococcus aureus* font partie de la classe 2, type II et sous-type A, caractérisé par la présence de **Csn2** et l'absence de **Cas4**.

LS, large subunit ; SS, small subunit ; CARF, CRISPR-associated Rossmann fold

Opéron *Cas*

L'opéron *Cas* de type II-A est composé de 4 gènes *Cas* (Figure 7A) :

- Les gènes *cas1* et *cas2*, ubiquitaires, et le gène *csn2* (spécifique du sous-type A) participent à l'intégration de nouveau *spacer* pour la phase d'adaptation^{44,50}.
- Le gène signature *cas9* est la véritable spécificité de l'opéron. Faisant partie de la classe 2, ce gène code pour une grosse protéine multidomaine et multifonctionnelle, principale effectrice du système.

Locus CRISPR et biogénèse des crRNA

Le locus CRISPR est similaire aux autres types avec une alternance de *spacer* et de *repeat* avec une extrémité *leader-end*. Toutefois la maturation des pré-crRNA fait appel à des éléments spécifiques du type II et nécessite (Figure 7B) :

- La présence de petits ARNs d'environ 75 nucléotides dont les 25 situés en 5' sont complémentaires aux séquences *repeat* du locus CRISPR, les *tracrRNA* (pour *transactivating-crRNA*). Ces *tracrRNA* sont encodés dans une région en amont de l'opéron *Cas*⁵¹. L'appariement des *tracrRNA* avec les *repeats* du pré-crRNA forme des ARN double-brin. Les structures secondaires du *tracrRNA* permettent de favoriser la liaison avec la Cas9.
- La fixation de la Cas9 au niveau des régions d'ARN double-brin formées par le pré-crRNA et le *tracrRNA*. La Cas9 est nécessaire à l'activité de la RNase III. Ce point illustre le rôle central de la Cas9 dans le fonctionnement du type II puisqu'elle est impliquée à la fois dans la maturation des crRNA et dans l'activité endonucléasique.
- L'activité de la RNase III permet de cliver les double-brin d'ARN et permet de finaliser la formation des crRNA matures. La RNase III est parfois incluse dans la représentation de l'opéron *Cas* bien qu'elle ne constitue pas un gène *Cas* à part entière.

Formation des crRNA:tracrRNA

La maturation du pré-crNA mène à la formation d'une molécule partiellement double-brin et composée du crRNA (de 41 nt) et du *tracrRNAs*.

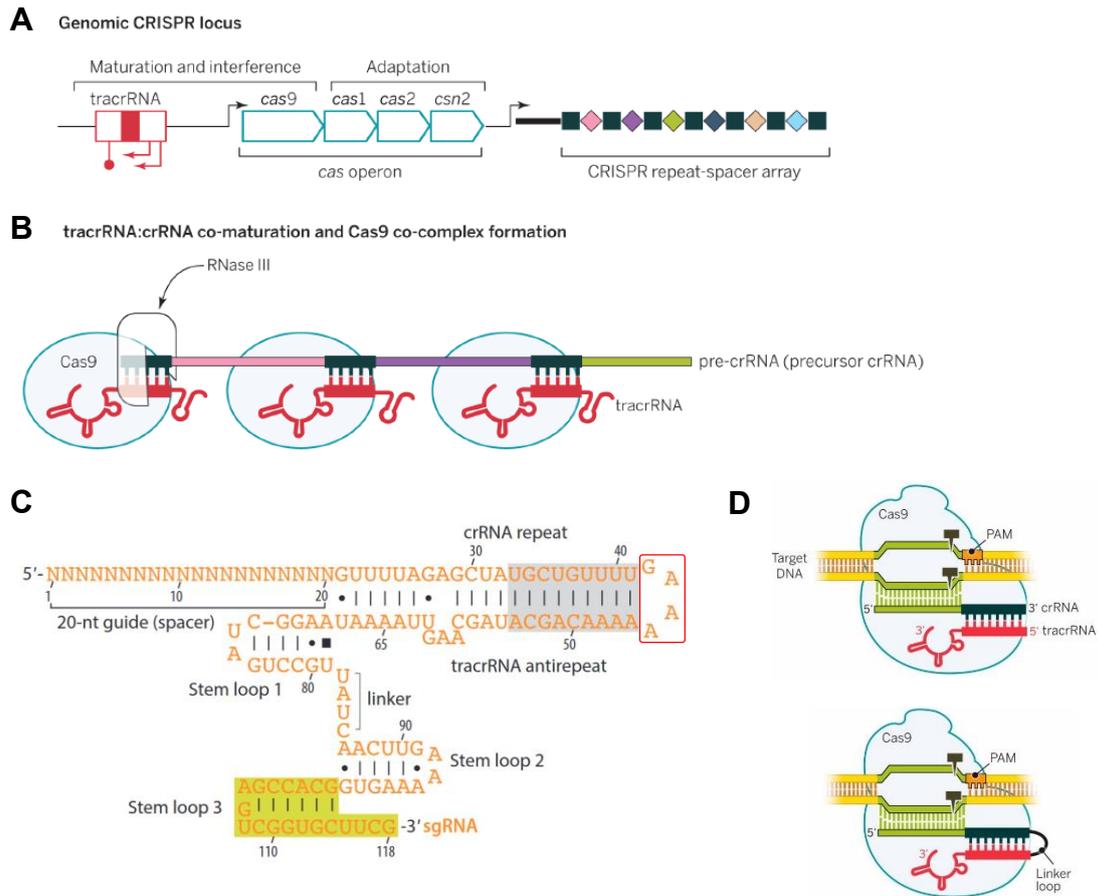


Figure 7. Structure et fonctionnement du système CRISPR-Cas9. (Doudna et Charpentier, *Science*, 2014 et (C) Jiang et Doudna, *Annu. Rev. Biophys.*, 2017)

(A) Représentation du locus CRISPR et de l'opéron Cas du système CRISPR/Cas9 (sous-type A, *Streptococcus pyogenes*). Le locus CRISPR est composé d'une alternance de *spacer* (losanges colorés) et de *repeat* (carrés noirs). L'extrémité *leader-end* est située entre le locus CRISPR et l'opéron Cas. Le locus encodant pour les tracrRNA est représenté en amont de l'opéron Cas.

(B) Maturation du pré-crRNA. Le pré-crRNA est composé de *repeats* (noirs), de *spacers* (rose, violet et vert). Les tracrRNA avec ses structures secondaires sont montrés en rouge. La Cas9 et la RNase III sont nécessaires à la maturation.

(C) Structure détaillée du crRNA:tracrRNA. Le crRNA est composé en 5' des 20 nucléotides du *spacer* et en 3' du *repeat*. La partie 5' du tracrRNA (*tracrRNA anti-repeat*) est complémentaire au *repeat* du crRNA. La partie 3' du tracrRNA est caractérisée par 3 *stem-loops* et un *linker*. Les quatre nucléotides GAAA, encadrés en rouge, permettent de former un *linker loop* (ou *tetraloop*) pour convertir le crRNA:tracrRNA en ARNg. Dans un ARNg, la région grisée et celle en vert clair peuvent être supprimées.

(D) Représentation graphique d'une RNP Cas9:crRNA:tracrRNA (haut) ou RNP Cas9:ARNg (bas) en lien avec une séquence cible.

Le crRNA:tracrRNA peut être décrit en deux parties⁴⁹ (*Figure 7C*):

- Le crRNA est constitué en 5' des 20 nucléotides du *spacer* et en 3' du *repeat* de 21 nt.
- Le tracrRNA est constitué d'un *anti-repeat*, s'hybridant par complémentarité au *repeat* du crRNA, de 3 *stem-loops* successifs et d'un *linker* entre le *stem-loop* 1 et 2.

La molécule crRNA:tracrRNA assure à la fois la reconnaissance de l'ADN cible et la liaison à la Cas9 :

- La séquence de *spacer* du crRNA définit la cible et assure sa reconnaissance.
- La liaison à la Cas9 est permise par le duplex *repeat / anti-repeat* et le *stem-loop* 1 du tracrRNA. Les *stem-loop* 2 et 3 participent à stabiliser la liaison, permettant une efficacité optimale notamment *in vivo*, sans être nécessaires⁴⁹.

En vue d'utiliser ce système comme outil d'édition génique, l'une des premières innovations a été la simplification du complexe crRNA:tracrRNA en une unique molécule d'ARN simple-brin plus courte, l'ARN guide (ARNg), facilitant ainsi sa synthèse et sa vectorisation^{41,48} (*Figure 7D*). La fonctionnalité de l'ARNg s'organise de façon identique, la seule différence étant la présence d'une *tetra-loop* entre le *repeat* et l'*anti-repeat* permettant de fusionner les deux éléments. La structure de l'ARNg permet de conserver les deux fonctionnalités du complexe crRNA:tracrRNA. Les 20 premiers nucléotides en 5' déterminent la séquence-cible tandis que la formation d'une structure secondaire en épingle à l'extrémité 3' permet d'assurer sa liaison à la Cas9.

Fonctionnement de la RNP Cas9:crRNA:tracrRNA

Une fois les crRNA:tracrRNA (ou l'ARNg) formés, ces derniers vont complexer avec la Cas9 pour former une RNP, unité capable d'activité endonucléasique dirigée par ARN. Cette RNP est capable de scanner l'ADN à la recherche d'un PAM (*Figure 8A*). Une fois détectée, la RNP induit une ouverture localisée de l'ADN double-brin pour permettre un test de complémentarité entre l'ARNg (ou le crRNA) et la séquence cible. Si la séquence reconnue est complémentaire, les deux domaines héminucléases de la Cas9, RuvC et HNH, respectivement responsables de la coupure du brin ciblé (complémentaire au crRNA) et du brin non-ciblé, permettent une cassure double-brin (CDB) de l'ADN. Les mécanismes régulateurs régissant le fonctionnement de la Cas9 seront détaillés plus tard dans le manuscrit pour illustrer le lien entre une compréhension fine de la Cas9 et la création de variants améliorés de la Cas9.

La CDB va mener à l'activation des voies de réparation pour permettre de préserver l'intégrité génomique.

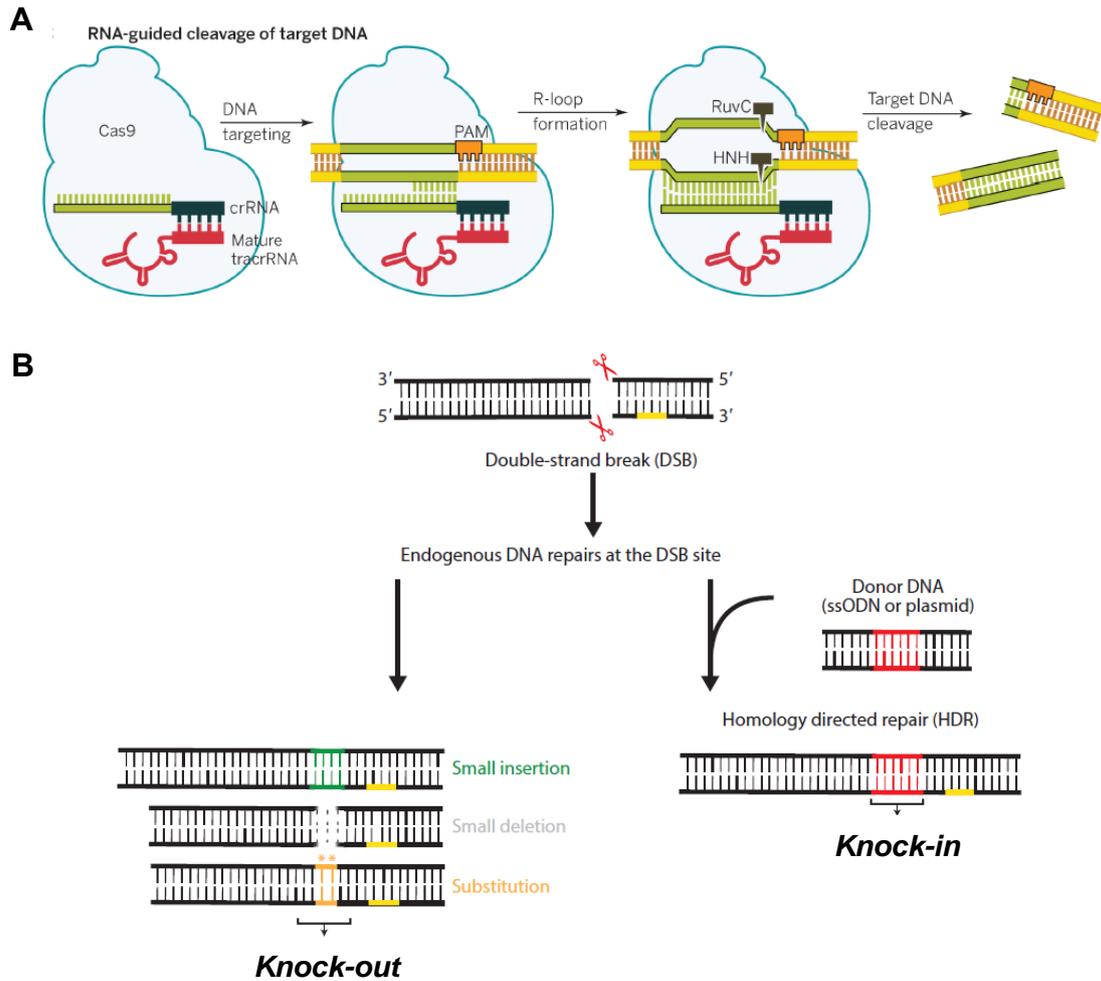


Figure 8. Activité endonucléasique de la RNP Cas9:crRNA:tracrRNA et activation des voies de réparation.

(A) Illustration du fonctionnement de la RNP Cas9:crRNA:tracrRNA. (Doudna et Charpentier, *Science*, 2014)

(B) La cassure double-brin peut être réparée par des voies non-conservatives de l'information génétique menant à des modifications ciblées de quelques paires de base (insertion, délétion, substitution) et donc à l'extinction du gène (*knock-out*). Si une matrice d'ADN exogène homologue à la séquence ciblée est apportée, celle-ci peut servir pour la réparation de la cassure double-brin et permettant l'introduction des mutations d'intérêt (*knock-in*). La matrice peut être un plasmide, un fragment d'ADN double-brin ou un oligonucléotide simple-brin (ssODN) (modifiée d'après Jiang et Doudna, *Annu Rev Biophys*, 2017)

Ces voies peuvent mener à deux catégories d'évènements : un *knock-out*, traduit par la perte de fonctionnalité du gène ciblé, ou du *knock-in* (ou HDR pour *Homology Directed Repair*) permettant l'introduction de mutation d'intérêt via une matrice donneuse apportée lors de l'édition par CRISPR-Cas9 (*Figure 8B*). Ici aussi, les voies de réparation seront détaillées dans une section ci-dessous pour illustrer le lien entre compréhension et amélioration de l'édition génique. Pour résumer, la découverte du système CRISPR-Cas9 a permis d'induire des cassures double-brin de façon site spécifique avec une facilité et une précision inégalée.

ii. Vectorisation du système CRISPR-Cas9

Si la découverte et la caractérisation de nouvelles endonucléases ont suscité un fort engouement, la création de nouveaux moyens de vectorisation pour transporter ces endonucléases a également été un champ actif de recherche et d'innovation.

L'apport des outils d'édition génique aux cellules-cibles, et en particulier aux CSH pour les pathologies hématologiques, peut être fait par de multiples procédés de vectorisation en fonction de la stratégie adoptée, *ex vivo* ou *in vivo*.

Contrairement à la thérapie additive pour laquelle l'intégration est capitale, ici les éléments d'édition peuvent être présents transitoirement dans la cellule. Toutefois, les vecteurs lentiviraux sont encore utilisés dans les études précliniques, leur grande efficacité à transduire les CSH en faisant de très bons outils^{52,53}. Malgré tout, en vue d'une application thérapeutique, la vectorisation idéale du système CRISPR-Cas9 doit i) permettre une transfection optimale des cellules cibles, ii) apporter les composants de façon transitoire et sans cicatrice, et iii) atteindre une forte efficacité d'édition uniquement au niveau de la séquence ciblée.

Pour les études portant sur des applications cliniques du système CRISPR-Cas9, deux procédés sont principalement utilisés : l'électroporation de plasmide ou de RNP Cas9:ARNg dans le cas d'une thérapie *ex vivo*, et les vecteurs issus des virus associés aux adénovirus (*AAV* pour *Adeno-Associated Virus*) pour les thérapies *in vivo*.

a. Vecteurs AAV

Le virus associé à l'adénovirus (AAV) de la famille des parovirus est un virus à ADN linéaire simple-brin. Il est naturellement déficient pour la réplication et a besoin d'un virus *helper*, adénoviral ou herpétique. Ces virus présentent l'avantage de pouvoir infecter une grande variété de types cellulaires (11 sérotypes identifiés), indépendamment du cycle cellulaire⁵⁴.

Les AAV sont non-pathogènes et fréquemment rencontrés chez l'humain, 70% à 80% de la population étant séropositive, mais ces infections sont faiblement immunogènes et ne sont associées à aucune pathologie connue. L'ensemble de ces propriétés font des AAV des outils intéressants pour vectoriser le système CRISPR-Cas9.

Faible capacité d'empaquetage

La principale contrainte des AAV est la capacité d'empaquetage, d'environ 4,7 kb⁵⁵, or les plasmides codant pour la Cas9 et un ARNg font environ 4,2kb. Cela entraîne des difficultés de productions à haut-titre et rend complexe l'insertion d'autres éléments comme des matrices de réparation pour l'HDR, d'autres ARNg ou encore des systèmes rapporteurs ou de sélection des cellules transfectées. Pour pallier à ce problème, plusieurs solutions ont été développées.

La première approche est d'utiliser deux AAVs en co-infection, l'un portant la Cas9, l'autre portant l'ARNg avec, selon les cas, la matrice de réparation et/ou un système rapporteur (*Figure 9A*). Ce système à deux AAV permet d'augmenter artificiellement la taille de la séquence apportée mais cette approche est techniquement plus complexe que l'utilisation d'un unique AAV.

Une seconde approche est de couper en deux la protéine Cas9 ; les parties Cter et Nter étant transportées chacune par un AAV⁵⁶ (*Figure 9B*). Ce système *split-Cas9* est permis par la fusion d'une moitié de protéine intéine aux deux fragments de Cas9, assurant la dimérisation. Une fois regroupé, le domaine intéine est clivé et permet de reconstituer des RNP Cas9:ARNg normales. Encore une fois, la méthodologie est complexifiée par l'utilisation de deux AAVs.

La troisième solution vient de la diversité des systèmes CRISPR-Cas9 (*Figure 9C*). Bien que, la Cas9 de *Streptococcus pyogenes* (SpyCas9) soit à la limite de la capacité d'empaquetage des AAV, des Cas9 « plus petites » ont été caractérisées dans d'autres espèces bactériennes. Le variant le plus connu est la Cas9 de *Staphylococcus aureus* (SaCas9)⁵⁷. Cette dernière fait 70% de la taille de la SpCas9 (soit 3,2 kb) permettant de libérer 1 kb. D'autres variants plus courts ont également été identifiés chez *Streptococcus thermophilus*⁴⁸, *Neisseria meningitidis*⁵⁸, et récemment chez *Campylobacter jejuni* avec une Cas9 de 2,95 kb⁵⁹. Malgré tout, ces variants possèdent des séquences PAM plus grandes et moins spécifiques que la SpyCas9 (5'-NGG-3') ce qui limite *de facto* le nombre de cibles potentielles (SaCas9 5'-NNGRRT-3', StCas9 5'-NNAGAAW-3', NmCas9 5'-NNNNGATT-3', CjCas9 5'-NNNNRYAC-3').

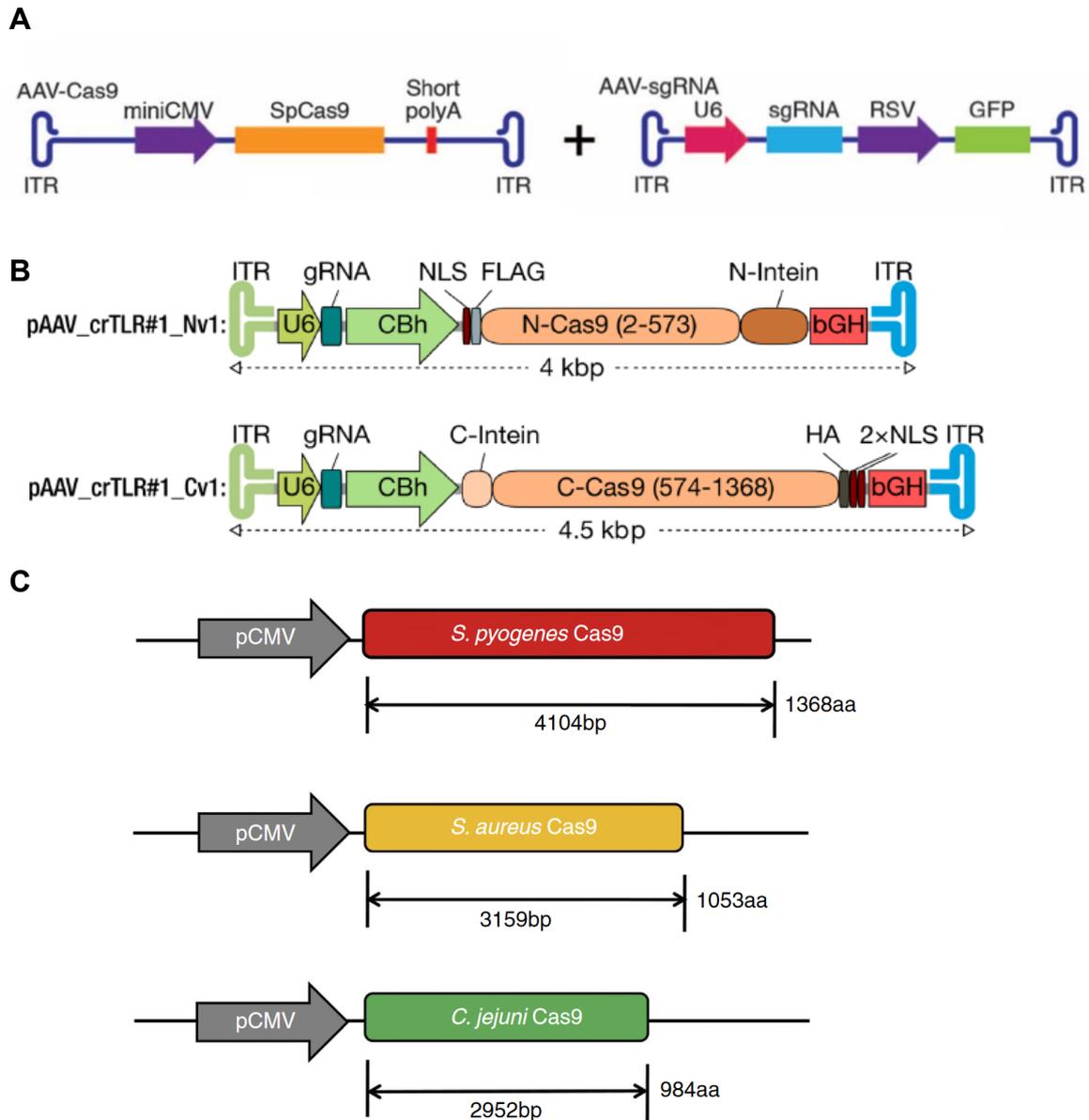


Figure 9. Contourner la limite d’empaquetage des AAVs.

(A) Exemple de système à deux AAVs. La première particule (AAV-Cas9) permet l’expression de la Cas9 de *Streptococcus pyogenes* (SpCas9) sous contrôle d’un promoteur ubiquitaire (miniCMV), la seconde permet l’expression d’un ARNg (sgRNA) et d’un gène rapporteur (*GFP*), chacun sous contrôle d’un promoteur ubiquitaire (*U6* et *RSV*). La cellule-cible doit être transduite par les deux vecteurs pour permettre le fonctionnement du système CRISPR-Cas9. (d’après Long et al., *Science*, 2016)

(B) Système *split-Cas9*. La première particule permet l’expression d’un ARNg et de la partie N-ter de la Cas9 fusionné à la partie N-ter de l’intéine, la seconde contient la partie C-ter de la Cas9 et une séquence C-ter d’intéine capable de s’associer à la partie N-ter. (d’après Truong et al., *NAR*, 2015)

(C) Comparaison de la taille de 3 variants de la Cas9. (d’après Kim et al., *Nat. Commun.*, 2017)

Un risque de réponse immunogène ?

Bien que les vecteurs AAVs soient d'excellents candidats pour vectoriser le système CRISPR-Cas9 *in vivo* à un stade clinique, il est important de souligner que la capsid virale des particules recombinantes est proche du virus sauvage. Puisque les infections par des AAVs sont fréquentes, l'administration de certains sérotypes chez un patient peut entraîner une réponse immunitaire adaptative, conduite notamment par les lymphocytes T mémoire CD8⁺ reconnaissant les protéines de la capsid⁶⁰. Ce phénomène a été observé lors du traitement de deux patients atteints d'hémophilie B après injection intra-hépatique d'AAV de sérotype 2, entraînant la perte du transgène et donc de l'effet thérapeutique⁶¹. D'autre part, une réponse humorale anti-AAV peut également être présente par l'intermédiaire des lymphocytes B mémoires et des anticorps neutralisants. Selon le sérotype, 15 à 60% de la population présente des anticorps neutralisants dirigés contre un type d'AAV⁶².

Un vecteur non-intégratif ?

Le génome viral sauvage est capable de rester à l'état extra-chromosomique ou de s'intégrer dans le génome hôte. Chez l'homme, cette insertion se fait au niveau du site AAVS1 (*AAV-integration-site 1*) situé dans un intron du gène *PPP1R12C* du chromosome 19. Cependant, les AAV recombinants utilisés pour la thérapie génique ont perdu leur capacité d'intégration spécifique mais peuvent rester plusieurs années dans des cellules quiescentes. Toutefois, bien qu'a priori non-intégratives, les méthodes virales restent limitées par le risque intrinsèque de mutagenèse insertionnelle. Une étude caractérisant les sites d'intégration d'AAV1 chez cinq patients traités pour un déficit en lipoprotéine lipase a montré une intégration aléatoire dans le génome nucléaire et des « hotspots » dans le génome mitochondrial avec une fréquence de 1 pour 1000⁶³. Récemment, une étude s'est intéressée au risque d'intégration des AAV dans le cas où les particules apportent le système CRISPR-Cas9 pour induire une CDB⁶⁴. Ce travail fait *in vitro* et *in vivo*, à des loci d'intérêts cliniques et dans des cellules primaires, a montré que l'induction de CDB active fortement l'intégration des AAVs dans le génome, phénomène également observé lors de la correction de la myopathie de Duchenne⁶⁵. Les résultats de ces travaux soulignent l'importance de développer des approches alternatives plus sûres pour répondre à la problématique de l'intégration, notamment pour la thérapie génique *in vivo* pour laquelle la vectorisation est un élément clé. De nombreuses méthodes non-virales, physiques ou basées sur la chimie des polymères ou des nanoparticules sont en développement et pourraient constituer une solution alternative prometteuse pour s'affranchir des risques d'insertion et de réponse immunitaire.

b. Méthodes physiques

Les méthodes physiques sont des méthodes de transfection sans vecteurs qui visent à faire entrer des biomolécules en déformant ou perforant les membranes cellulaires. L'instrumentation nécessaire restreint souvent leur utilisation à des contextes *ex vivo*.

L'électroporation

L'électroporation utilise un champ électrique pulsé pour induire l'ouverture de pores nanométriques dans les membranes cellulaires, permettant aux composés de diffuser au sein de la cellule (*Figure 10A*). Cette approche permet la modification simultanée d'un grand nombre de cellules et se prête particulièrement à la transfection de cellules habituellement réfractaires. L'électroporation permet de transférer le système CRISPR-Cas9 sous forme d'ADN plasmidique, d'ARN ou de RNP ; les avantages et les inconvénients sont résumés (*Table 3*).

Le défaut principal est celui de la cytotoxicité associée aux chocs électriques. Toutefois, selon la forme du système CRISPR-Cas9, le choc électrique nécessaire est variable. La transfection sous forme plasmidique nécessite de franchir à la fois la barrière cytoplasmique et nucléaire. Un choc électrique important est donc nécessaire et entraîne une forte mortalité cellulaire. À l'inverse, le système CRISPR-Cas9 sous forme d'ARN ou de RNP est moins délétère. En effet, l'ARNm Cas9 ou la protéine peuvent porter des signaux de localisation nucléaire. Ainsi, seul le franchissement de la membrane cytoplasmique est nécessaire, autorisant des chocs électriques moindres et une meilleure viabilité cellulaire.

Injection hydrodynamique

Si les méthodes précédentes servent essentiellement pour une approche *ex vivo*, l'injection hydrodynamique est une méthode physique applicable exclusivement *in vivo*. La première preuve de principe de thérapie génique *in vivo* par CRISPR-Cas9 a utilisé cette méthodologie⁶⁶. Ce procédé repose sur l'injection systémique rapide d'un large volume de solution (représentant 8 à 10% du poids totale) contenant les molécules d'intérêts (*Figure 10B*). Le sang étant incompressible, la force de compression engendrée augmente temporairement la perméabilité des cellules endothéliales et parenchymateuses. Les organes cibles sont principalement le foie, les reins, les poumons et les muscles, cardiaques y compris. Si quelques succès ont été rapportés, ce procédé est peu efficace, traumatique, associé à une certaine mortalité, et peut conduire à des dysfonctions cardiaques et hépatiques.

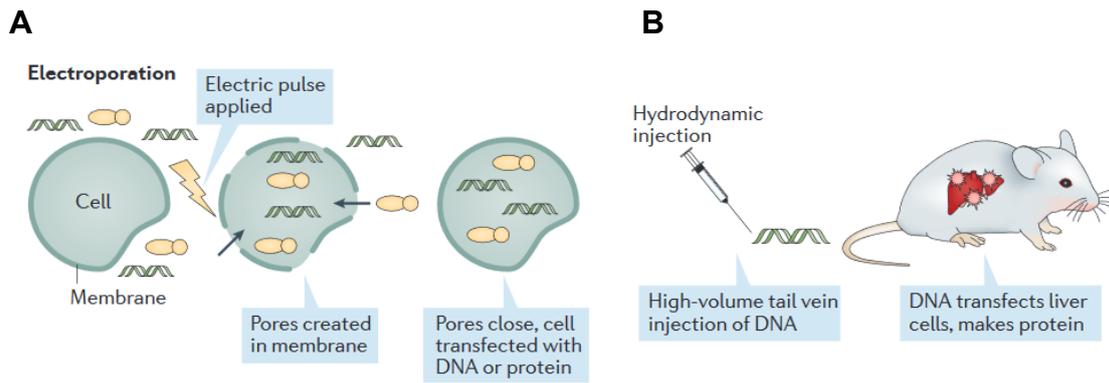


Figure 10. Méthodes physiques de transfection. (d'après Yin et al., Nat. Rev. Drug Discov., 2017)
 (A-B) Représentation schématique de l'électroporation (A) et de l'injection hydrodynamique (B).

	Avantages	Inconvénients
ADN plasmidique	<ul style="list-style-type: none"> - Peu couteux - Cassette de sélection positive 	<ul style="list-style-type: none"> - Risque intégration - Faible efficacité - Doit atteindre le compartiment nucléaire - Doit être transcrit et traduit - Quantité effective non-maitrisée - Expression prolongée - Cytotoxicité
ARNm Cas9 + ARNg	<ul style="list-style-type: none"> - Efficacité transfection améliorée - Compartiment cytosolique - Non-intégratif - Transitoire - Fenêtre d'action courte 	<ul style="list-style-type: none"> - Peu stable - Transcription - Quantité effective non-maitrisée
RNP Cas9:ARNg	<ul style="list-style-type: none"> - Forte efficacité transfection - Compartiment cytosolique - Non-intégratif - Transitoire - Prêt à l'action - Maitrise quantité effective - Fenêtre d'action courte - Viabilité améliorée 	

Table 3. Avantages et inconvénients des trois formes du système CRISPR-Cas9.

c. Vers d'autres procédés de vectorisation innovants ?

Pour l'édition génique, l'évènement d'intégration doit être proscrit autant que possible. Néanmoins, cette particularité constitue plus une opportunité qu'une contrainte et ouvre la porte à de nombreux procédés de vectorisation. Des méthodes pseudo-virales, physiques ou basées sur la chimie des polymères ou des nanoparticules sont en développement et pourraient constituer des solutions alternatives prometteuses pour s'affranchir des risques d'insertion et de réponse immunitaire des approches virales et de la cytotoxicité des chocs électriques lors de l'électroporation. Trois procédés élégants, encore non-utilisés en clinique, sont brièvement décrits ci-dessous.

Une méthode pseudo-virale : les Nanoblades

Les Nanoblades visent à profiter des avantages des vecteurs γ -rétroviraux mais en s'affranchissant du risque de mutagenèse⁶⁷ (Figure 11A). Mangeot et al. ont fusionné la Cas9 à l'extrémité Cter de Gag en séparant les deux composants par un site protéolytique clivable par les protéases du vecteur. Cela permet de créer des particules pseudo-virales contenant des RNP Cas9:ARNg, appelés Nanoblades. Une preuve d'efficacité de cet outil a été faite dans des iPSCs avec un taux d'efficacité avoisinant les 70%. Des Nanoblades « *All-in-one* » peuvent être formés pour permettre d'apporter une matrice de réparation pour stimuler l'HDR.

Une méthode physique : la déformation mécanique

La déformation mécanique vise à induire une ouverture transitoire des membranes plasmiques par des pressions mécaniques⁶⁸ (Figure 11B). Cette approche utilise un dispositif microfluidique dans lequel les cellules passent par un tube dont le diamètre est légèrement inférieur à celui de la cellule sur un segment donné.

Cette conformation entraîne des forces de compression et de cisaillement qui permet l'entrée de macromolécules avoisinantes dans la cellule. La fonctionnalité de lymphocytes T édités sur le gène *PD-1* par des RNPs Cas9:ARNg apportées par électroporation ou par déformation mécanique a été comparée. Si les deux approches possèdent la même efficacité d'édition, la déformation mécanique permet une meilleure préservation *in vivo* de la fonctionnalité des cellules éditées *ex vivo*⁶⁹.

Une méthode chimique : CRISPR-Gold

Les nanoparticules d'or ont déjà plusieurs applications en biologie, notamment pour l'imagerie. Récemment, des particules d'or ont été modifiées pour leur permettre d'interagir avec les RNP Cas9:ARNg et une matrice de réparation pour effectuer de l'HDR (Figure 11C). Cette structure complexe, appelée CRISPR-Gold⁷⁰, permet d'atteindre *in vitro* des niveaux d'HDR dans des cellules primaires myoblastiques plus élevés que par électroporation avec une meilleure viabilité.

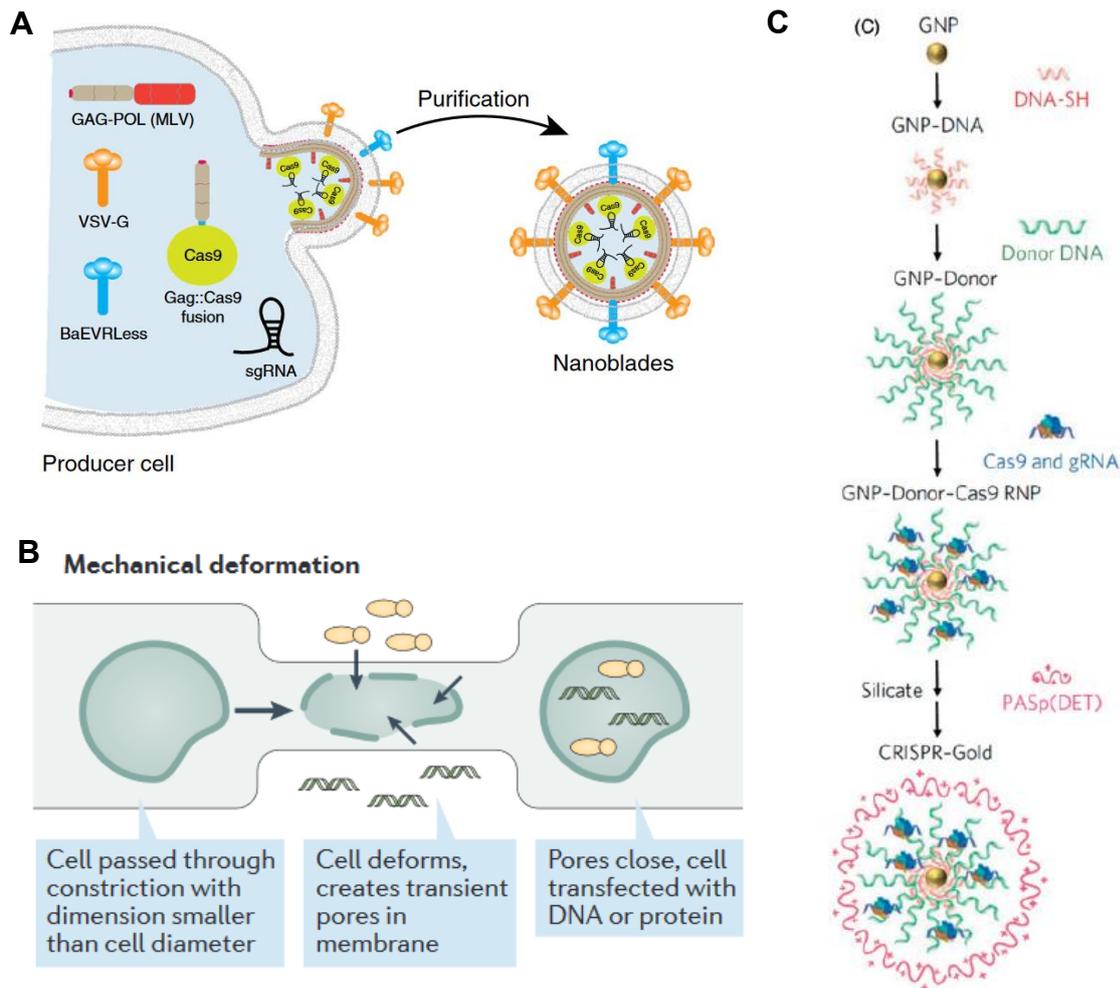


Figure 11. Procédés alternatifs de vectorisation. (d'après Yin et al., *Nat. Rev. Drug Discov.*, 2017 et (C) Lee et al., *Nat Biomed Eng.*, 2017)

(A) Représentation schématique de l'assemblage de Nanoblades.

(B) Illustration de la déformation mécanique.

(C) Représentation du système CRISPR-Gold. Les particules d'or (GNP) sont recouvertes de matrices de réparation (Donor DNA) grâce à un pontage fait par de petits oligonucléotides avec un groupement thiol en 5' (DNA-SH) s'accrochant à la GNP. Les RNPs Cas9:ARNg s'accrochent à cette structure par affinité. Enfin cette structure est recouverte de silicate, anionique, pour permettre la fixation d'un polymère cationique (PASp(DET)), facilitant la sortie des CRISPR-Gold du compartiment endosomal une fois transfectés.

Un niveau similaire d'HDR a été retrouvé *in vivo* après injection intramusculaire dans un modèle murin muté pour la dystrophine. Ces résultats sont très encourageants par leur efficacité et par leur sûreté et d'autres modifications sauront prochainement améliorer ces systèmes.

iii. Applications précliniques et cliniques basées sur CRISPR-Cas9

Le système CRISPR-Cas9 a très rapidement été adopté par le domaine de l'édition génique. Cette adoption a mené à une avancée notable vers la faisabilité d'une stratégie thérapeutique par édition génique des maladies monogéniques. Ces progrès importants sont illustrés par plusieurs études précliniques et cliniques à la fois *ex vivo* et *in vivo*.

a. Édition génique *ex vivo* : drépanocytose et β -thalassémie

Les pathologies associées aux mutations du gène *HBB*, majoritairement drépanocytose et β -thalassémie, sont les meilleurs exemples pour illustrer les possibilités offertes par l'édition génique *ex vivo*. Actuellement, deux stratégies thérapeutiques sont développées : une approche par *knock-out* pour mimer une persistance de l'hémoglobine fœtale (HbF) et une approche par *knock-in* pour faire une correction directe de la mutation causale.

Mimer la persistance héréditaire de l'hémoglobine fœtale.

L'une des stratégies s'inspire du fait que les patients β -thalassémiques ou drépanocytaires co-héritant de mutations causant une persistance héréditaire de l'hémoglobine fœtale (HPFH) présentent des symptômes moins lourds. La gravité est inversement proportionnelle au taux d'HbF, voire asymptomatique si l'expression d'hémoglobine fœtale est suffisamment élevée. Les études présentées ci-dessous ont eu pour objectif d'induire une expression d'hémoglobine fœtale en utilisant la technologie CRISPR-Cas9 en modifiant soit le cluster des gènes de globines soit un *enhancer* du gène *BCL11A*.

Cluster des gènes β -globines

Une première étude s'est concentrée sur une délétion de 13 nucléotides au sein de la séquence promotrice de *HBG1* capable de recruter un répresseur transcriptionnel (BCL11A) et responsable de la transition de γ -globine à une hémoglobine β^{52} (Figure 12A). Pour les patients hétérozygotes pour cette délétion, l'HbF représente 30% de l'hémoglobine totale, contre moins de 1% normalement. L'objectif de l'étude a été d'introduire cette délétion au niveau des gènes responsables de la γ -globine. Des cellules CD34⁺ normales ont été transduites *ex vivo* par des vecteurs lentiviraux permettant l'expression de la Cas9 et d'un ARNg dirigé contre la séquence promotrice d'intérêt. La production de γ -globine est assurée par deux gènes homologues en tandem, *HBG1* et *HBG2*, possédant chacun une séquence reconnue par l'ARNg.

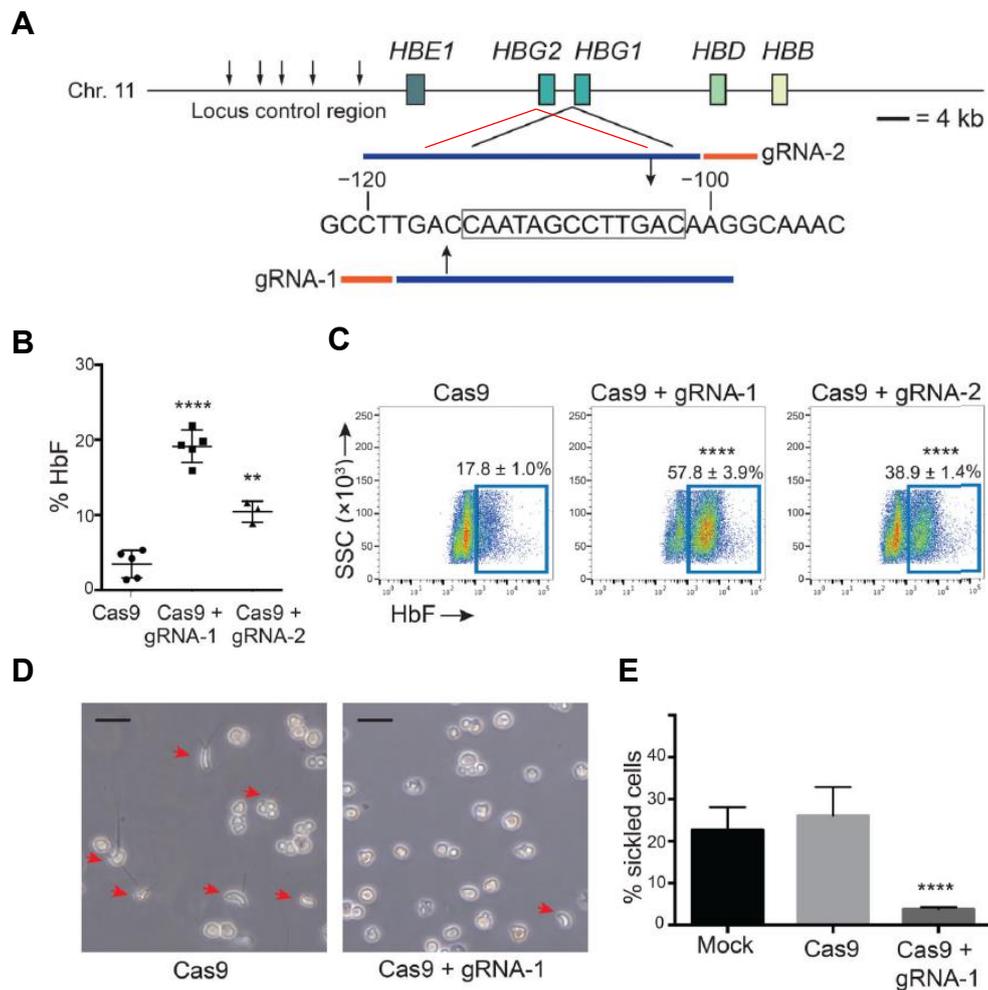


Figure 12. Mimer l'induction d'HPFH par une délétion de 13 nt. (modifiée d'après Traxler et al, *Nat. Med.*, 2016)

(A) Représentation schématique du cluster des β -globines. Les flèches indiquent les sites hypersensibles à la Dnase I dans le LCR (Locus control region). En dessous, vue détaillée d'une région en amont du site d'initiation de la transcription de *HBG1* (traits noirs). Les 13 nucléotides d'intérêt, à l'origine de la HPFH, sont encadrés. Une séquence similaire est retrouvée en amont du gène *HBG2*, indiquée par les traits rouges. Les séquences *spacers* des ARNg (gRNA) sont représentées par une ligne bleue et le PAM (NGG) en orange. Les flèches noires associées indiquent les sites de coupures. L'ARNg1 et l'ARNg2 sont respectivement complémentaires au brin sens et anti-sens.

(B) Quantification par HPLC de l'HbF de $CD34^+$ issues de patients sains, transduites avec un vecteur lentiviral codant pour la Cas9 et l'ARNg1 ou l'ARNg2 et cultivées 21 jours en présence de cytokines érythroïdes-spécifiques. Chaque point représente une expérience indépendante sur des $CD34^+$ issues d'un même donneur.

(C) Illustration représentative et quantification par cytométrie en flux des érythroblastes HbF-positifs pour des $CD34^+$ éditées comme décrit en (B).

(D) Observation microscopique de $CD34^+$ issues de patients drépanocytaires, transduites avec un vecteur lentiviral codant pour la Cas9 et l'ARNg1, différenciées en globules rouges et cultivées en condition hypoxique (2% O_2). Les flèches rouges indiquent les cellules falciformes. La barre d'échelle représente 20 μ m.

(E) Quantification du nombre de cellules falciformes induites par l'hypoxie issus de $CD34^+$ éditées comme en (D).

Deux coupures, distantes de 5,2 kb, sont induites simultanément. Cette délétion de 5,2 kb est détectée dans 20% des clones HUDEP-2 analysés par qPCR, mais n'est pas retrouvée dans des CD34⁺ édités. Il est à noter que l'analyse de cette délétion a été faite sur des CD34⁺ de patients sains édités par électroporation d'un plasmide codant pour la Cas9 et l'ARNg (non-intégratif) et que les résultats ci-dessous ont été obtenus par transduction lentivirale (intégratif). L'amplification d'une région de 431 pb autour des sites de coupures par PCR de CD34⁺ édités montre que la moitié des InDels retrouvés sont représentés par la délétion de 13 nt. Ce taux élevé de délétion particulière est expliqué par la présence de séquence répétée en tandem de 8 nt favorisant une réparation par la voie de l'*Alternative End-Joining* (Alt-EJ). Parmi les cellules transduites, la quantité d'HbF passe de 5% à 20%, traduit par le passage de 18 à 58% de cellules positives pour l'expression d'hémoglobine fœtale (*Figure 12B et C*). Cette étude a été complétée par un travail dans un cadre pathologique. Pour cela, des cellules CD34⁺ de patients drépanocytaires ont été transduites *ex vivo* avec les mêmes outils d'édition, puis différenciées. Une analyse morphologique des cellules différenciées et cultivées en hypoxie montre une diminution du nombre de cellules en faucilles lorsque les cellules sont éditées (*Figure 12D et E*). Une stratégie similaire a été faite dans un modèle de souris transgéniques, portant le cluster des β -globines humains, avec une transduction des CSHs par un vecteur adénoviral *helper-dependant*⁷¹. Dans cette étude, une réactivation de la synthèse d'HbF est observée et l'analyse du locus *HBB* montre majoritairement une délétion d'un nucléotide ; suggérant que la délétion partielle du site de fixation de BCL11A est suffisante pour réactiver l'expression d'HbF. Il est à noter ici que des délétions de 4,9 kb supprimant la région entre les deux sites de coupures sont observées à une fréquence d'environ 2%.

Une solution alternative a été proposée par Antoniani et collègues et repose cette fois-ci sur la cNHEJ pour recréer des larges délétions observées chez des patients atteints d'HPFH⁷². Si l'idée sous-jacente est proche et vise à réactiver la synthèse d'hémoglobine fœtale, la région ciblée est différente. La cartographie des délétions touchant le cluster des globines et retrouvées chez des patients drépanocytaires ou atteints de HPFH a permis de définir une région de 13,6 kb potentiellement intéressante. Cette séquence (*Figure 13A*):

- Est bornée en 5' par la région de Corfu, délétion minimale retrouvée chez des patients β -thalassémiques avec une HbF élevée.
- Est bornée en 3' pour supprimer le premier exon et le promoteur du gène *HBB* tout en laissant intacte la séquence *enhancer* 3' et intronique du même gène.

- Comprend une région minimale de 3,5 kb retrouvée délétée chez 10 des 12 patients HPFH et contenant un site de fixation putatif de BCL11A, facteur transcriptionnel impliqué dans la répression de HbF, une séquence reconnue par le modulateur chromatinien PYR et enfin, plusieurs sites de fixation du facteur de transcription GATA1.

Des CSHs issues de deux patients drépanocytaires ont été nucléofectées avec un plasmide Cas9-GFP et 2 plasmides exprimant les ARNg coupant à chaque extrémité de la région de 13,6 kb. Suite aux deux cassures, la séquence de 13,6 kb est dans certains cas délétée ou inversée (*Figure 13B*). Pour l'un des donneurs, 34,1% de délétions et 28,1% d'inversion sont retrouvées. Les CSHs éditées ont été différenciées en érythrocytes puis analysées sur leur capacité à produire de l'hémoglobine fœtale (*Figure 13C*). Pour le même donneur, la fraction de cellules exprimant de l'hémoglobine est passée de 27% à 63% ; un résultat similaire est obtenu pour le second donneur. Ce résultat s'observe lors de la culture des érythrocytes en hypoxie et se traduit par une amélioration du phénotype drépanocyttaire avec la diminution de 65% à 30% du nombre de cellules en faucilles (*Figure 13D et E*).

Il est à noter que la délétion uniquement de la région de 3,5kb, de la région de Corfu ou une invalidation simple des gènes de β -globines ne permet pas de réactiver une synthèse d'hémoglobine fœtale aussi fortement qu'observée avec la délétion de 13,6kb. Ce résultat peut être expliqué par deux phénomènes synergiques : la délétion des gènes des β -globines et le rapprochement de la séquence *enhancer* 3' et intronique du gène *HBB*, agissant désormais sur les gènes *HBG1* et *HBG2*. De plus, bien que l'étude princeps présentée repose sur la nucléofection du système CRISPR-Cas9 sous forme plasmidique, un travail de comparaison et d'optimisation a été conduit pour améliorer la fréquence de modification des CSHs⁷³. Le meilleur compromis entre cytotoxicité et efficacité a été obtenu à l'aide de RNP Cas9:ARNg. Enfin, un second groupe a conduit des travaux avec une méthodologie similaire, avec suppression totale du gène *HBB* et conservation de la séquence 3' *enhancer*, ont obtenu des résultats concordants avec ceux présentés ci-dessus⁷⁴.

Enhancer du gène *BCL11A*

Une autre stratégie pour corriger les β -hémoglobinopathies consiste à cibler une partie bien précise du gène *BCL11A*, répresseur de l'expression de l'HbF. Le gène *BCL11A* possède dans l'intron 2 une séquence *enhancer* érythroïde-spécifique, reconnue par GATA1 et TAL1 et dont certaines substitutions sont associées à une expression élevée de γ -globine⁷⁵ (*Figure 14A*). La transduction de cellules CD34⁺ de donneurs sains avec un lentivecteur exprimant la Cas9 et un ARNg dirigé contre l'*enhancer* réactive la synthèse de γ -globine et d'HbF⁵³.

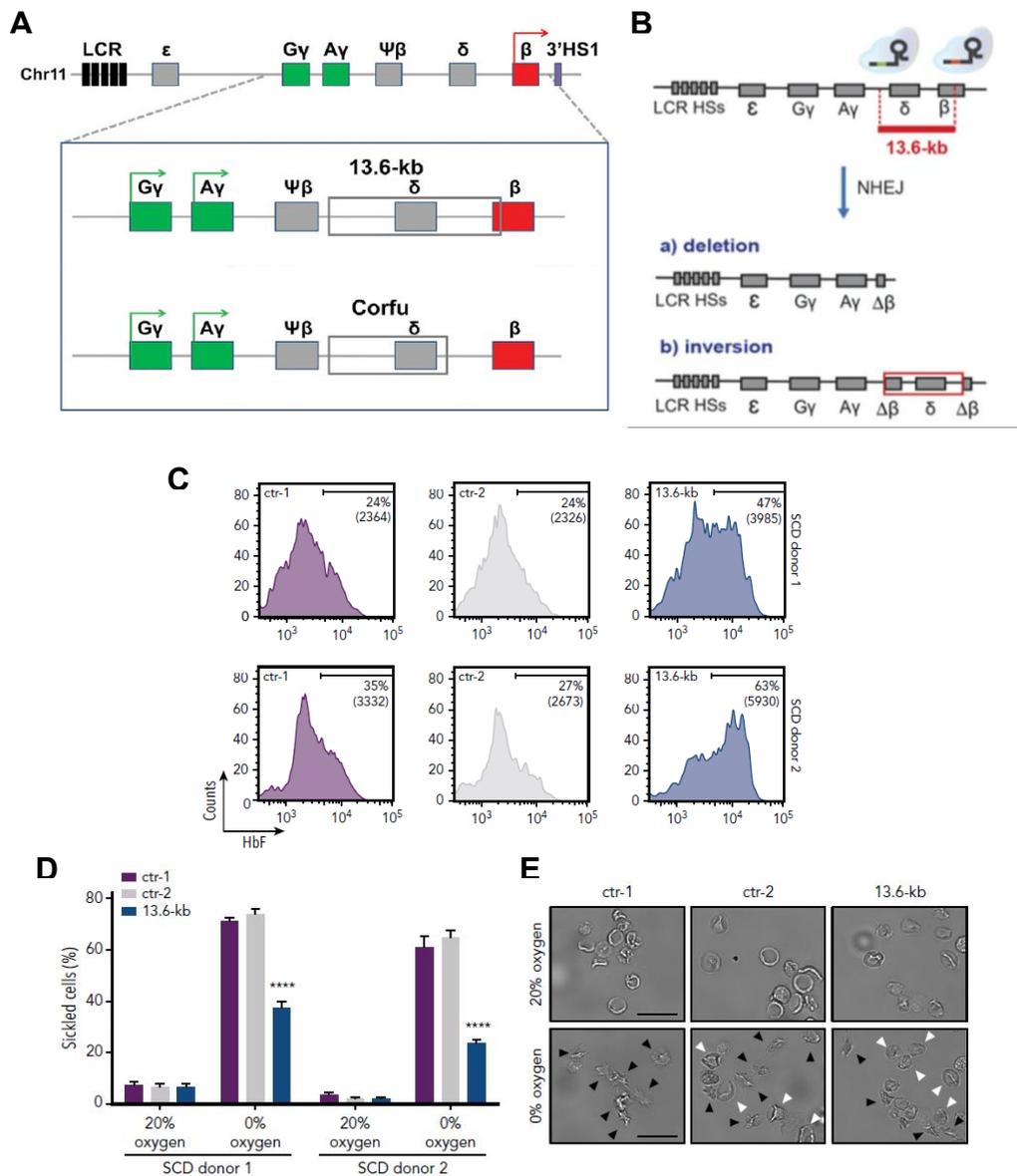


Figure 13. Mimer l'induction d'HPFH par une délétion de 13,6 kb. (Antoniani et al, Blood, 2018 et (A) modifiée d'après Lattanzi et al., Mol. Ther., 2019)

(A) Représentation schématique du cluster des β -globines et vue détaillée avec la séquence de 13,6 kb à exciser. La région de Corfu est montrée pour indication. (B) (Haut) Stratégie d'édition pour exciser la séquence de 13,6 kb (en rouge) par la Cas9 et deux ARNg et (Bas) organisation du cluster des β -globines en cas de délétions ou d'inversions de la séquence de 13,6 kb. (C à E) Des CD34⁺ issues de la moelle osseuse (SCD donateur 1) ou de sang périphérique mobilisé (SCD donateur 2) de patients drépanocytaires ont été transfectées avec un plasmide codant pour une Cas9-GFP et les deux ARNg visant de part et d'autre de la région à exciser. Ces cellules sont ensuite triées par FACS. La fraction GFP-négative constitue le contrôle négatif 2 (ctr-2) et la fraction GFP-positif correspond à la population 13.6-kb. Le contrôle négatif 1 (ctr-1) est issu de CD34⁺ transfectées avec un plasmide Cas9-GFP, sans ARNg, et triées pour la fraction GFP-positif. Les trois populations triées sont ensuite différenciées en globules rouges. (C) Résultat représentatif de l'analyse par FACS de l'expression d'HbF pour les trois populations (ctr-1, ctr-2 et 13.6-kb). (D) Quantification de la proportion de globules rouges falciformes en conditions de normoxie ou d'hypoxie. (E) Images représentatives des globules rouges avant et 60 minutes après la mise en place de l'hypoxie. Les flèches noires et blanches indiquent respectivement les cellules falciformes et non-falciformes. Barre d'échelle 50 μ m.

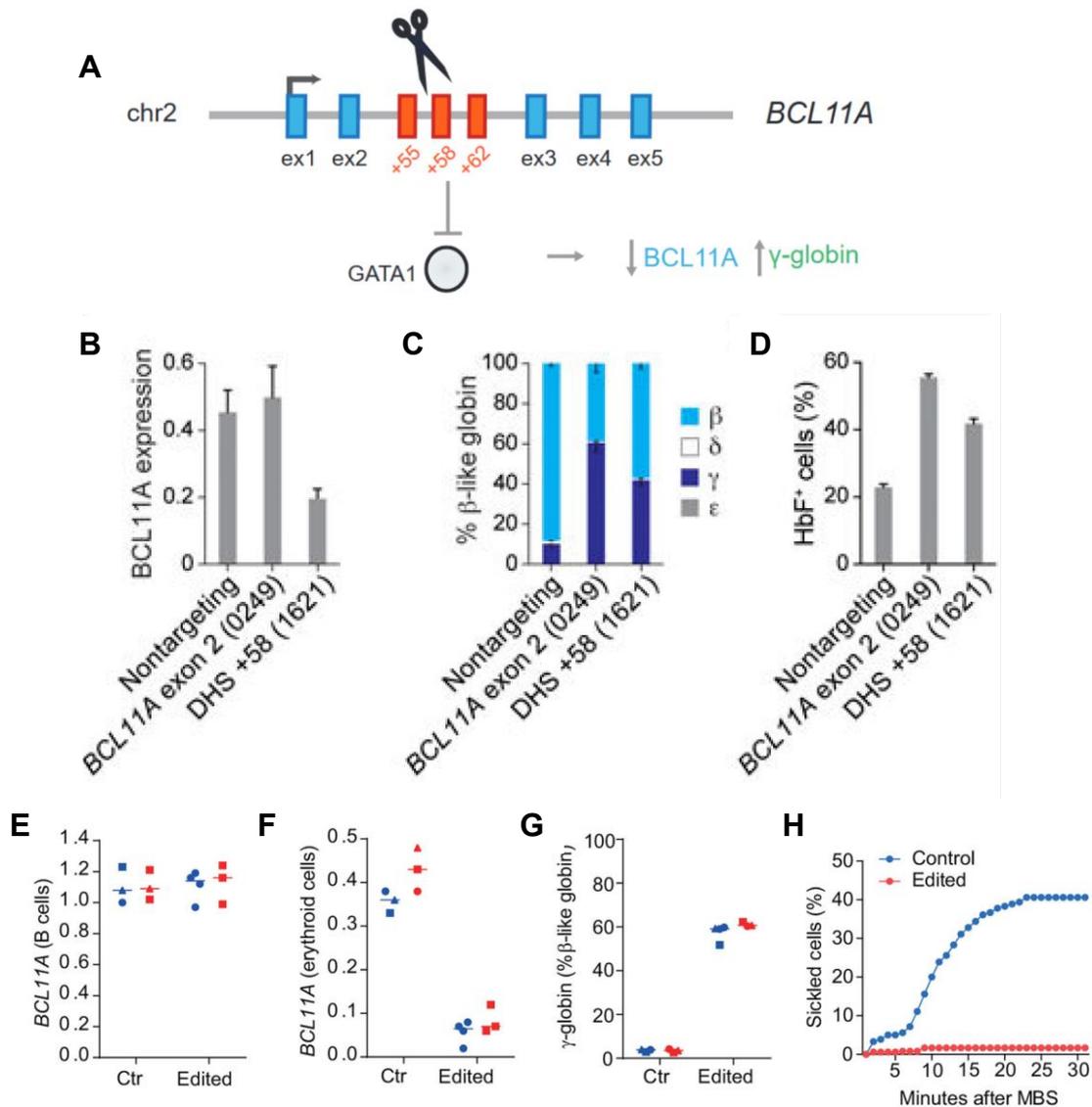


Figure 14. Suppression de l'expression érythroïde-spécifique de BCL11A.

(A) Représentation schématique du gène *BCL11A*. Les sites hypersensibles à la DNase I sont représentés en rectangle orange (+55, +58, +62) et correspondent à une région intronique *enhancer* du *BCL11A*. Une coupure dans le site +58 empêche la fixation de GATA1, activateur transcriptionnel de *BCL11A*. Cela permet une diminution de l'expression de *BCL11A* restreinte au tissu érythroïde et une réactivation de la synthèse d'HbF. (modifiée d'après Magrin et al. *Blood*, 2019)

(B-D) Quantification de l'expression (B) de *BCL11A*, (C) de globines β -like et (D) du nombre de cellules HbF-positives pour des précurseurs érythroïdes issus de CD34⁺ saines, transduites avec un vecteur lentiviral codant pour la Cas9 et un second vecteur portant l'ARNg ciblant l'exon 2 du gène *BCL11A* (*BCL11A* exon2 (0249)), le DHS +58 ou aucune cible (*Nontargeting*). L'expression est normalisée avec celle de la *GAPDH*. (Canver et al., *Nature*, 2015)

(E-H) Des CD34⁺ issues de patients drépanocytaires ont été transfectées *ex vivo* avec une RNP Cas9:ARNg dirigée contre le site de fixation de DHS +58 (*Edited*). Des souris immunodéficientes ont été greffées et leurs moelles osseuses ont été récupérées 16 semaines après greffe. Les cellules humaines ont ensuite été analysées. Quantification par RT-qPCR de l'expression de *BCL11A* (E) dans des cellules B humaines ou (F) dans des cellules érythroïdes humaines. (G) Expression des globines β -like dans des cellules érythroïdes humaines. (H) Quantification du nombre de globules rouges falciformes après différenciation *in vitro* des moelles osseuses de souris greffées avec des CD34⁺ éditées (*Edited*) ou non (*Control*) et traitement au métabisulfite de sodium (MBS) (Wu et al., *Nat. Med.*, 2019)

Ces résultats plus qu'encourageants ont conduit Vertex Pharmaceuticals Incorporated et CRISPR Therapeutics au lancement d'un essai clinique de phase 1/2 utilisant cette stratégie pour des patients drépanocytaires (NCT03745287) ou β -thalassémiques (NCT03655678) (12 patients prévus par essai). Une amélioration du phénotype drépanocytaire et β -thalassémique a été obtenue dans des CSH issues de patients transfectées avec des RNP Cas9:ARNg dirigées contre la séquence *enhancer*⁷⁶. Il est à noter que cette stratégie s'appuie sur une seule cassure double-brin, et non deux comme présenté précédemment.

Correction directe de la mutation causale

La seconde stratégie pour traiter les β -hémoglobinopathies est plus classique de l'édition génique et vise à corriger directement la mutation à l'origine de la pathologie. La drépanocytose se prête particulièrement à cet exercice puisqu'une unique mutation ponctuelle dans le gène *HBB*, p.E6V, est à l'origine de cette pathologie. À l'inverse, la β -thalassémie peut être expliquée par une diversité de mutations rendant plus complexe la mise au point d'une stratégie universelle.

Une étude élégante a été menée par Dever et al. en 2016⁷⁷. Dans cette étude, des cellules CD34⁺ issues de patients drépanocytaires ont été transfectées avec une RNP Cas9:ARNg dirigée contre la mutation E6V et co-transduites avec un AAV recombinant sérotype 6 (rAAV6), capable d'infecter les CSHs. Le rAAV6 permet de délivrer une matrice de 2,6 kb homologue au gène *HBB* et porte plusieurs SNPs dont l'un corrige la mutation p.E6V et cinq forment des mutations silencieuses pour éviter que la séquence soit reconnue par l'ARNg après une correction par HDR (*Figure 15A*). L'analyse au locus montre un fort taux de modification avec environ 50% d'HDR et 40% d'InDels pour uniquement 10% d'allèles non-modifiés (*Figure 15B*).

Une seconde étude s'est affranchie de l'utilisation de vecteur AAV en remplaçant la matrice de correction par un simple ssODN⁷⁸. Un travail systématique d'optimisation des outils d'édition a permis une correction efficace de cellules CD34⁺ issues de patients drépanocytaires. Le criblage de différents ssODNs, de doses et de symétries par rapport à la mutation a permis de définir la combinaison la plus efficace pour faire du *knock-in* (*Figure 15C*) ; un taux d'HDR stable de 28% ayant été obtenu avec 90% de viabilité cellulaire des CD34⁺ post-électroporation. Ici aussi, les évènements de *knock-in* sont combinés à 41% d'InDels et 31% d'allèles intacts (*Figure 15D*). Pour minimiser l'activité hors-cible, un variant de la Cas9 haute spécificité a été utilisé avec une même efficacité sur la cible. Enfin, des cellules CD34⁺ drépanocytaires ont été corrigées et greffées dans des souris NSG. Avant greffe, les taux d'HDR et d'InDels étaient de 29,8% et 42,8% et chutent à 8,8% et 24,5% à seize semaines après greffe, suggérant un potentiel de greffe réduite pour les cellules éditées comparées aux cellules non-éditées.

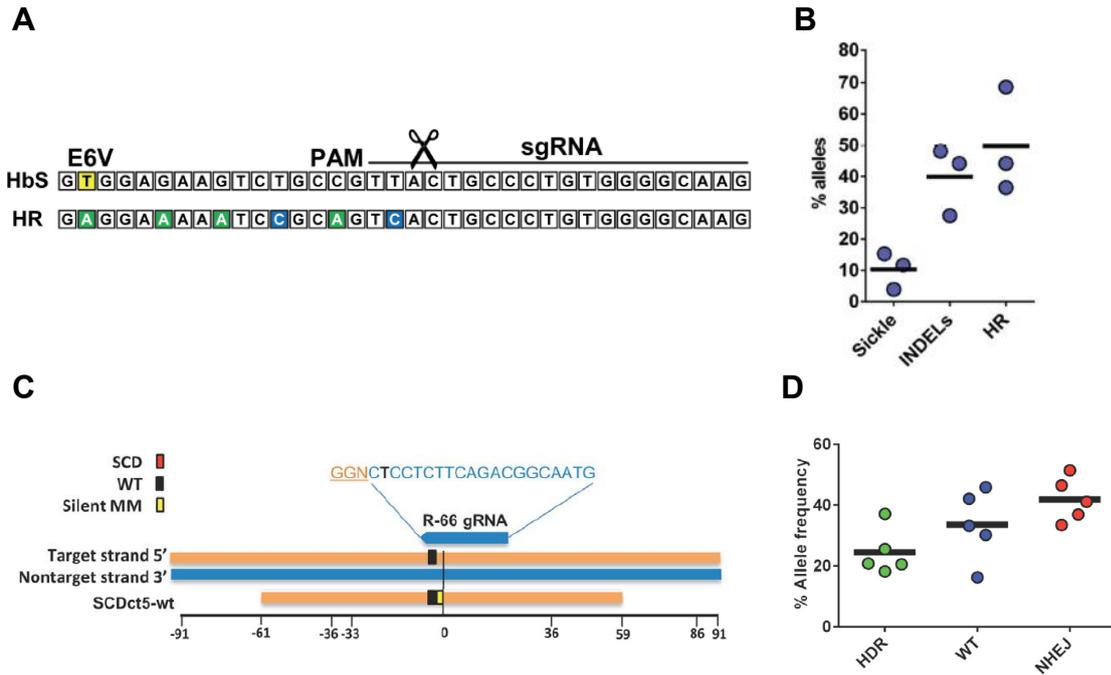


Figure 15. Correction directe de la mutation p.E6V par édition génique.

(A) Représentation schématique de l'exon 1 du gène *HBB* porteur de la mutation p.E6V (jaune) et de la matrice de réparation apportée par le rAAV6 (HR). En plus de celle corrigeant la mutation, la matrice porte cinq mutations silencieuses (colorées en vert ou bleu). (Dever et al., Nature, 2016)

(B) Des CD34⁺ issus des patients drépanocytaires ont été transfectées avec la RNP Cas9:ARNg dirigée contre la mutation puis transduites avec le rAAV6 apportant la matrice de réparation. L'analyse allélique a été faite en séquençant les produits de PCR insérés dans des vecteurs TOPO. 50 à 100 TOPO sont analysés pour chaque réplicat issu de trois patients différents. (Dever et al., Nature, 2016)

(C) Des CD34⁺ issues des patients drépanocytaires ont été transfectées avec la RNP Cas9:ARNg dirigée contre la mutation (R-66 sgRNA) et le SCDct5-wt, un ssODN porteur de la correction de la mutation (WT) et d'une mutation silencieuse juxtaposée (Silent MM). (D) Analyse allélique par ddPCR des CD34⁺ éditées. Chaque réplicat correspond à l'édition de cellules issues de patients drépanocytaires différents. (Park et al., NAR, 2019)

Si cette étude apporte des preuves de faisabilité pour l'édition génique de la mutation de la drépanocytose, elle apporte également un lot d'interrogations vis-à-vis de la sécurité du système CRISPR-Cas9.

b. Édition génique *in vivo*

L'édition génique par le système CRISPR-Cas9 est également une solution prometteuse pour le traitement de pathologie se prêtant à une correction *in vivo*. Pour ces approches, un des éléments clés est la capacité à cibler efficacement les cellules d'intérêts. Si une première preuve de principe a été obtenue par injection hydrodynamique, l'essor de l'édition génique *in vivo* est essentiellement dû aux vecteurs AAV. Comme pour les approches *ex vivo*, des approches reposant soit sur du *knock-in* soit sur du *knock-out* ont été développées.

Correction directe de la mutation : tyrosinémie de type I et déficit en ornithine carbamyl transférase (OTC)

Tyrosinémie de type I

En 2014, une preuve d'édition génique *in vivo* par CRISPR-Cas9 a été obtenue dans un modèle murin de tyrosinémie de type I, porteur d'une mutation homozygote dans le gène *FAH* entraînant un défaut d'épissage⁶⁶ (*Figure 16A*). La déficience en fumaryl acétoacétase entraîne l'accumulation de métabolites toxiques pour le tissu hépatique. L'un des avantages de ce modèle pathologique est que la correction des hépatocytes leur confère un avantage sélectif ; une forte correction initiale n'est ainsi pas nécessaire pour atteindre une amélioration phénotypique. Une injection hydrodynamique a été faite avec une solution contenant un plasmide permettant l'expression de la Cas9 et de l'ARNg visant la mutation et un ssODN de 199 nt, portant la correction de la mutation (*Figure 16B*).

Le marquage immunohistochimique des tissus hépatiques a permis d'estimer la fréquence d'hépatocytes corrigés à 0,40%. Trente jours après, la fréquence allélique de correction atteint 9% avec 26% d'InDels. Si cette étude fait référence de première preuve d'édition génique *in vivo*, une solution alternative à l'injection hydrodynamique pourrait être préférable. Toutefois, les auteurs soulignent que, bien qu'associée à une hépatotoxicité, l'injection hydrodynamique a déjà été utilisée et adaptée pour un essai clinique chez l'homme⁷⁹, et que, si son utilisation permet le traitement d'une maladie monogénique, la balance bénéfice/risque est à considérer. Néanmoins, la faible efficacité de transfection limite largement les possibilités offertes par cette méthodologie.

Déficit en ornithine carbamyl transférase (OTC)

L'utilisation des vecteurs AAV permet d'associer efficacité de transduction et ciblage cellulaire. Une preuve d'édition génique *in vivo* par *knock-in* utilisant le système CRISPR-Cas9 et des vecteurs AAV a été faite dans un modèle de souris déficientes en OTC⁸⁰ (*Figure 17A*).

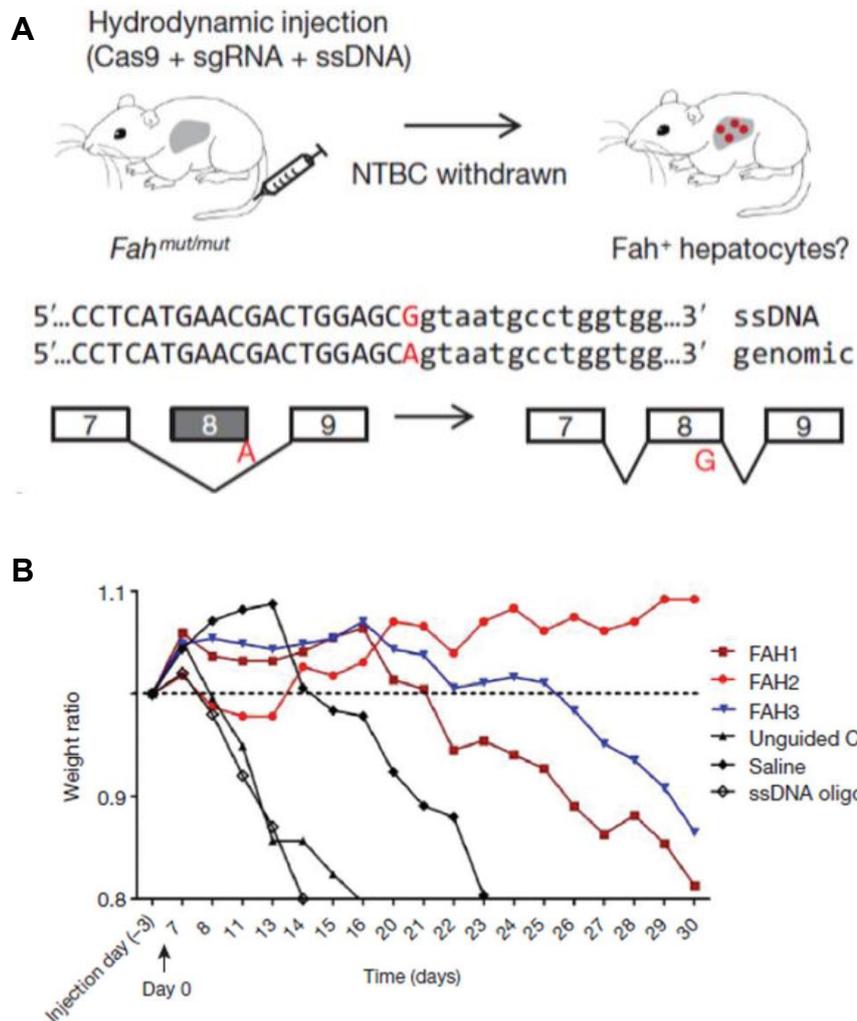


Figure 16. Correction directe *in vivo* de la mutation de la tyrosinémie de type I. (Yin et al., Nat. Biotechnol., 2014)

(A) Représentation schématique de la stratégie thérapeutique pour la correction de la tyrosinémie de type I. Les souris *Fah^{mut/mut}* portent une mutation homozygote G→A au niveau du dernier nucléotide de l'exon 8, entraînant le saut de l'exon 8 pendant l'épissage. Un plasmide codant pour la Cas9, un ARNg dirigé contre la mutation et un ssODN porteur de la correction de la mutation sont apportés au foie par injection hydrodynamique. Le NTBC, inhibiteur pharmacologique de la voie catabolique de la tyrosine, est ensuite retiré de l'eau de boisson et le poids des souris est suivi.

NTBC, 2-(2-nitro-4-trifluorométhylbenzoyl)-1,3-cyclohexanedione

(B) Évolution du poids des souris injectées avec soit une solution saline (*Saline*), le ssODN uniquement (*ssODN oligo*), le ssODN et un plasmide codant pour la Cas9 uniquement (*Unguided Cas9*) ou le ssODN et un plasmide codant pour la Cas9 et l'un des trois ARNg visant la mutation du gène *Fah* (*FAH1*, *FAH2*, *FAH3*). Le jour 0 est marqué par le retrait du NTBC.

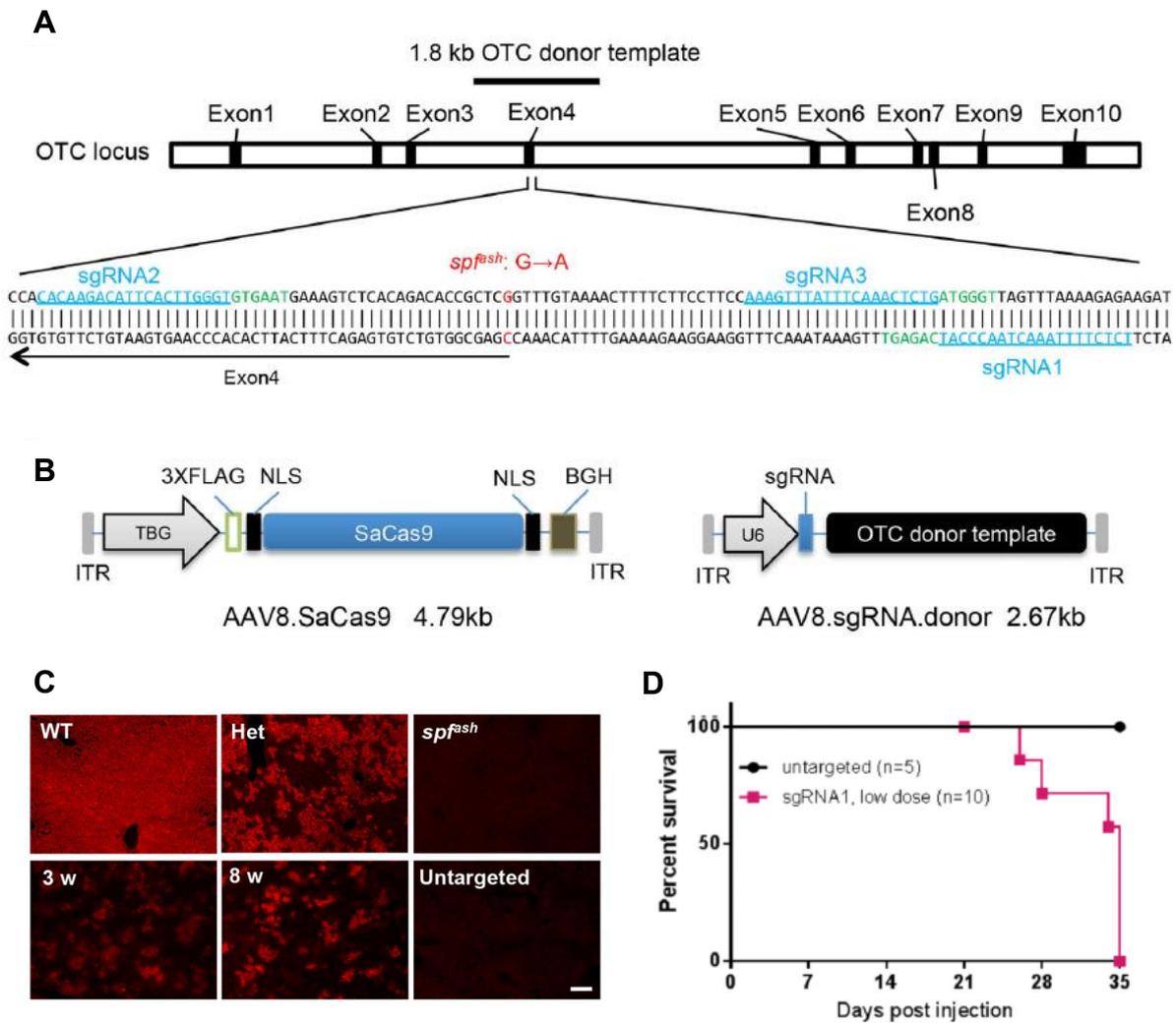


Figure 17. Correction directe *in vivo* du déficit en ornithine carbamyl transférase. (Yang et al., Nat. Biotechnol., 2016)

(A) Représentation schématique du locus *OTC* montrant la mutation *spf^{ash}* située à l'extrémité de l'exon 4 (en rouge), entraînant un défaut d'épissage. La position de la matrice de correction de 1,8 kb est montrée. Trois ARNg permettent de viser à proximité de la mutation. Les *spacers* des ARNg sont en bleu et les PAM de la SaCas9 en vert.

(B) Illustration du système à deux AAVs sérotype 8. ITR, *Inverted Terminal Repeats* ; TBG, *Human thyroxine binding globulin* ; BGH, *bovine growth hormone polyadenylation sequence*.

(C) Marquage immunofluorescent avec des anticorps dirigés contre OTC de tissus hépatiques issus de souris non-traitées sauvages (WT), hétérozygote (HET), homozygote pour la mutation *spf^{ash}* (*spf^{ash}*) ou de souris éditées par injection des deux AAV à deux jours post-natal et sacrifiées à 3 (3 w) ou 8 semaines (8 w). Les souris éditées avec les deux AAVs mais sans ARNg dans la seconde particule servent de contrôles négatifs (Untargeted). Des *clusters* d'hépatocytes OTC-positifs sont observés dans les conditions 3w et 8w.

(D) Courbes de survie des souris adultes traitées avec les deux AAVs, comportant l'ARNg (*sgRNA-1*) ou non (untargeted). Les deux conditions sont traitées avec la même quantité d'AAVs. Des résultats similaires sont obtenus avec une dose élevée d'AAVs.

Pour cela des souris nouveau-nés ont reçu une injection systémique de deux particules AAV sérotype 8 (*Figure 17B*) : l'une permet l'expression de la Cas9 de *Staphylococcus aureus* (SaCas9) (plus petite que la SpyCas9) sous contrôle d'un promoteur hépato-spécifique, la seconde mène à l'expression de l'ARNg et contient une séquence de 1,8kb homologue au gène *OTC*. Un taux d'HDR de 14,3% et de 31% d'InDels est observé à 8 semaines et permet une amélioration phénotypique (*Figure 17C*).

Encouragés par ces résultats, les auteurs ont conduit une approche similaire sur des souris adultes. Cependant, avec la même méthodologie, une fréquence plus élevée d'InDels est observée (44,6%) avec une chute notable du taux d'HDR (0,3%). L'analyse immunohistologique ne permet d'observer que 2,2% de cellules OTC-positives. Pire encore, les souris traitées ont dû être euthanasiées 5 semaines après le traitement (*Figure 17D*). L'analyse du locus *OTC* montre une fraction importante d'allèles portant des délétions dont 6% affectent l'exon situé en amont de la mutation ; cette dernière touchant un site donneur d'épissage. Les auteurs ont supposé que des voies de réparation différentes existent entre les hépatocytes adultes post-mitotiques et ceux de nouveau-nés, capables de divisions cellulaires.

Excision de séquence : Myopathie de Duchenne et amaurose congénitale de Leber

La myopathie de Duchenne et l'amaurose congénitale de Leber possèdent un fond génétique dans lequel la suppression d'un exon peut mener à l'amélioration de la pathologie. Des stratégies d'excision d'exon ou de la séquence intronique pour supprimer la mutation ponctuelle causant la pathologie ont été proposées. Pour les deux pathologies, l'approche thérapeutique repose sur l'utilisation de vecteurs AAV pour délivrer le système CRISPR-Cas9 et sur 2 ARNg permettant de couper de part et d'autre de la séquence à exciser.

Excision d'exon pour la myopathie de Duchenne.

L'édition génique *in vivo* apporte un véritable espoir en vue du traitement de la myopathie de Duchenne. Elle est causée par des mutations ponctuelles, des délétions ou des duplications dans le gène *DMD* entraînant un décalage de cadre de lecture et/ou une perte de l'expression de la protéine. L'idée sous-jacente est de supprimer l'exon muté pour permettre la production d'une protéine tronquée, mais fonctionnelle.

Une première preuve de principe a été obtenue dans un modèle murin porteur d'une mutation non-sens dans l'exon 23 du gène *Dmd* (*souris mdx*)⁸¹ (*Figure 18A*). Pour la correction, un système à deux vecteurs AAV sérotype 9 a été adopté (*Figure 18B*) : le premier exprimant la SaCas9 sous contrôle d'un promoteur CMV et le second permettant l'expression de deux ARNg ciblant dans l'intron 22 et 23 afin d'exciser l'exon 23.

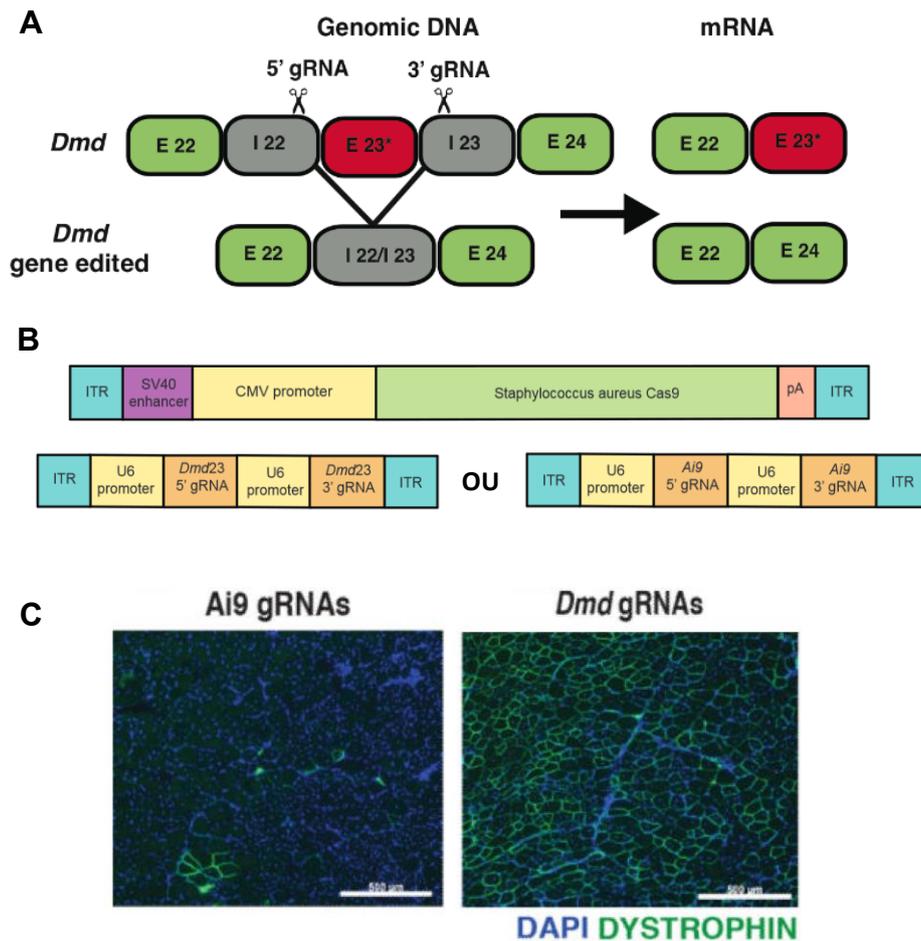


Figure 18. Correction *in vivo* par excision d'exon de la myopathie de Duchenne avec deux cassures double-brin. (modifiée d'après Tabebordbar et al., Science, 2016)

(A) Représentation schématique de la stratégie pour exciser l'exon 23. Les souris *mdx* portent une mutation non-sens dans l'exon 23 qui entraîne la formation d'un ARNm *Dmd* tronqué et instable. L'excision de l'exon 23 par deux ARNg ciblant dans les régions introniques (I22 et I23) permettent de produire un intron de fusion (I 22/23). L'expression du gène *Dmd* édité permet de former un ARNm *Dmd* sans l'exon 23 et mène à la formation d'une protéine toujours hautement fonctionnelle malgré la perte de l'exon.

(B) Illustration du système à deux AAVs sérotype 9. Un système rapporteur avec un locus *Ai9* est intégré à la souris *mdx* (AAV9-*Ai9*, servant dans cette figure de témoin négatif).

(C) Marquage immunofluorescent de la Dystrophine du tibia antérieur de souris *mdx* quatre semaines après avoir été traitées par co-injection intramusculaire de l'AAV9-SaCas9 et (gauche) de l'AAV9-*Ai9* ou (droite) de l'AAV9-*Dmd23*. Barre d'échelle, 500μm.

Les vecteurs AAVs ont été injectés dans le muscle tibial antérieur. Quatre semaines après, l'analyse moléculaire montre une perte à hauteur de 39% de l'exon 23, permettant de détecter la dystrophine à un niveau protéique (*Figure 18C*). Une approche systémique a été faite en injectant en intrapéritonéale les mêmes outils dans des souris nouveau-nés (P3). Une délétion de l'exon 23 a été retrouvée dans plusieurs tissus musculaires avec une efficacité variable de 3 à 18%. Deux autres études, parues également dans Science en 2016, ont obtenu des résultats similaires^{65,82}.

L'exon 50 est l'un des exons le plus fréquemment supprimé dans la myopathie de Duchenne⁸³. Le décalage de cadre de lecture induit peut être rétabli en supprimant l'exon 51 de la protéine. Le *hot spot* de mutation du gène *DMD* se trouve entre l'exon 45 et 55. Un modèle murin Δ Ex50 a été formé en excisant l'exon 50 par deux ARNg. Si la stratégie pour la modélisation est similaire aux travaux présentés ci-dessus, la méthode de correction diffère. L'idée est ici de permettre le saut d'exon en supprimant la séquence *exonic splicing enhancer* (ESE) de l'exon 51⁸³ (*Figure 19A*). Un système à deux AAVs est toujours utilisé et injecté dans le muscle tibial antérieur (*Figure 19B*). L'analyse immunohistologique montre une restauration de l'expression de la dystrophine dans une forte majorité des cellules (*Figure 19C*). Des résultats convaincants ont également été obtenus suite à une injection systémique dans des souris Δ Ex50 à P4. Cette stratégie a également été reproduite dans un modèle canin avec une injection systémique des mêmes deux vecteurs AAVs et conduit à une expression de dystrophine allant de 3 à 90% selon le muscle, permettant une amélioration phénotypique. À la vue de l'ensemble de ces travaux, il est probable qu'un essai clinique ait lieu dans un avenir proche.

Excision d'intron : Amaurose congénitale de Leber

L'amaurose congénitale de Leber type 10 (LCA10) est la cible du premier essai clinique d'édition génique *in vivo* utilisant le système CRISPR-Cas9, débuté en juillet 2019 (NCT03872479). Le traitement repose sur un système baptisé EDIT-101, clin d'œil à la nomenclature anglo-saxonne des cours pour débutants.

La LCA10 est une maladie monogénique récessive du gène *CEP290* et cause une cécité profonde ou totale chez l'enfant. La mutation la plus fréquente est une mutation A>G dans l'intron 26 (c.2991+1655A>G), faisant apparaître un site cryptique d'épissage (*Figure 20A*). Un exon aberrant porteur d'un codon stop apparaît. Ici, la stratégie est similaire à l'excision d'exon, seule la nature de la cible étant différente. Pour supprimer la séquence d'intron contenant la mutation IVS26, un vecteur AAV sérotype 5 permettant l'expression de la SaCas9 sous-contrôle d'un promoteur tissu-spécifique et d'une paire d'ARNg coupant dans l'intron 26 et de part et d'autre de la mutation a été

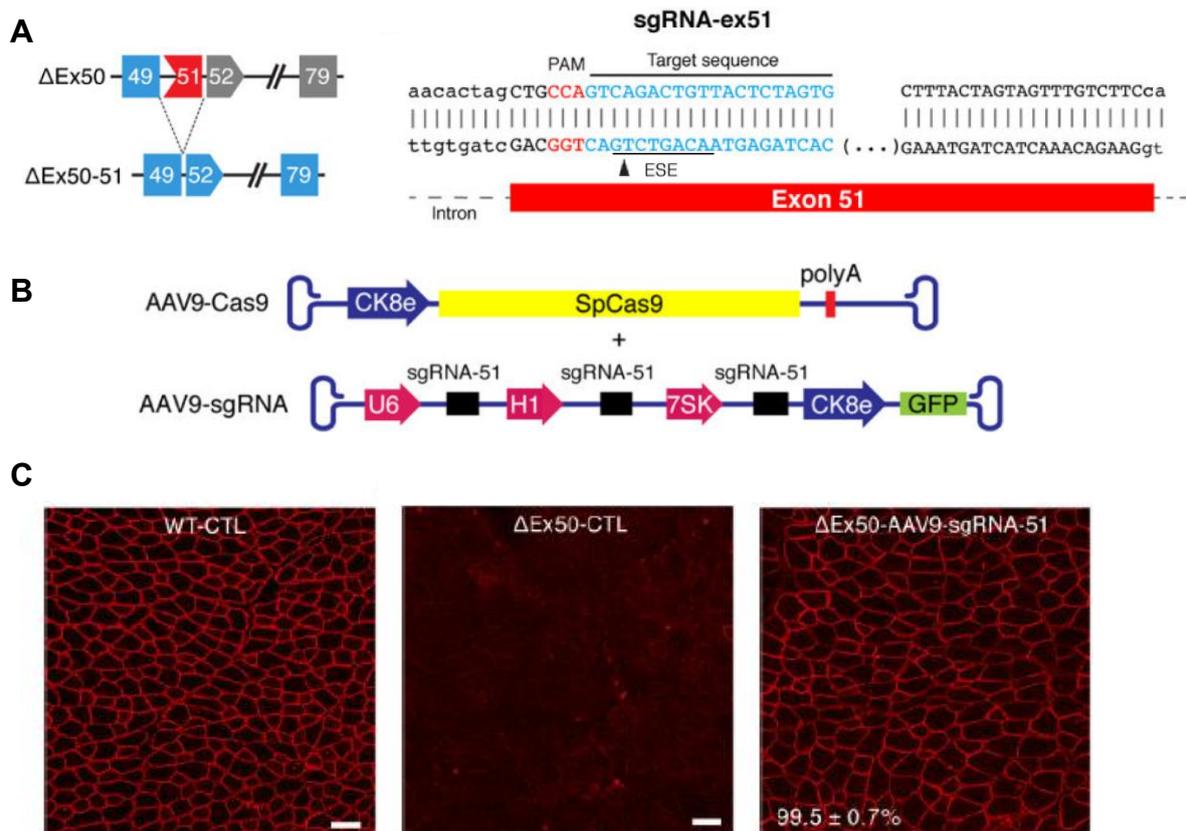


Figure 19. Correction *in vivo* par excision d'exon de la myopathie de Duchenne avec une cassure double-brin. (modifiée d'après Amosii et al., *Sci. Transl. Med.*, 2017)

(A) Représentation schématique de la stratégie pour supprimer l'exon 51 des souris Δ Ex50. La délétion de l'exon 50 souris Δ Ex50 entraîne un décalage du cadre de lecture pouvant être corrigé en supprimant la séquence *exonic splicing enhancer* (ESE) de l'exon 51. Cette suppression entraîne un saut d'exon de l'exon 51 lors de l'épissage et permet de rétablir le cadre de lecture.

(B) Illustration du système à deux AAVs sérotype 9. La première particule (AAV9-Cas9) permet l'expression de la Cas9 de *Streptococcus pyogenes* (SpCas9) sous contrôle d'un promoteur musculaire spécifique (CK8e), la seconde présente trois cassettes d'expression de l'ARNg visant la séquence ESE, promu chacun par un promoteur fort de l'ARN polymérase III et permet l'expression de la GFP sous promoteur musculo-spécifique.

(C) Marquage immunofluorescent de la Dystrophine du tibia antérieur de souris contrôles (WT-CTL), de souris Δ Ex50 (Δ Ex50-CTL), ou de souris Δ Ex50 après avoir été traitées par co-injection intramusculaire (Δ Ex50-AAV9-sgRNA51). Barre d'échelle, 50 μ m.

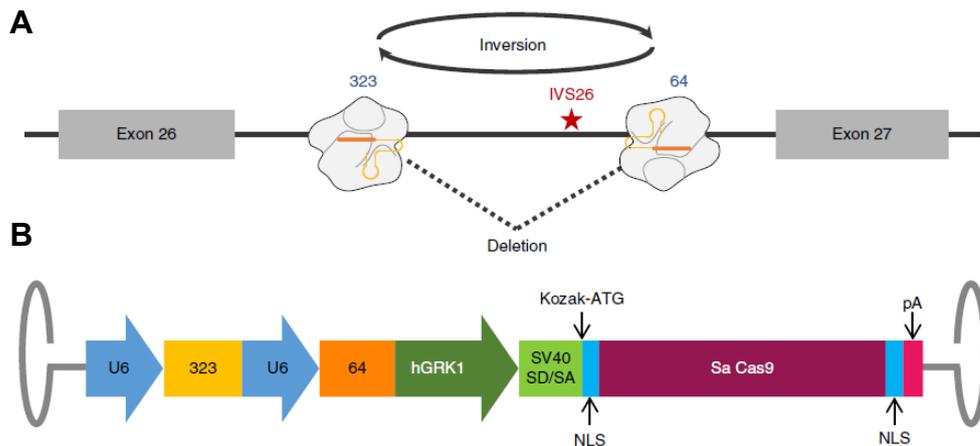


Figure 20. Correction *in vivo* par excision d'intron de l'amaurose congénitale de Leber type 10. (modifiée d'après Maeder et al., Nat. Med., 2019)

(A) Représentation schématique de la stratégie pour supprimer le site cryptique d'épissage IVS26. Deux ARNg ciblant des régions intronique de part et d'autre de la mutation IVS26 permet de supprimer ou d'inverser la région ciblée. Les deux événements mènent à la correction de l'épissage.

(B) Illustration du système avec un unique AAVs sérotype 5.

hGRK1, human G protein-coupled receptor kinase 1

Rare Disease	Companies	Mechanism	Delivery	Notes
α-1-Antitrypsin deficiency	EDIT	NHEJ & HDR	<i>In vivo</i>	By AAV or LNP
	NTLA	n/a	<i>In vivo</i>	By LNP
Amyotrophic lateral sclerosis	CRSP	n/a	<i>In vivo</i>	Collaboration with Target ALS and University of Florida
Autoimmune and inflammatory diseases	NTLA	n/a	<i>Ex vivo</i>	Disease unspecified
β-Hemoglobinopathies (β-thalassemia & sickle cell disease)	CRSP	NHEJ	<i>Ex vivo</i>	CTX001; based on induction of γ -globin
	EDIT	NHEJ & HDR	<i>Ex vivo</i>	By RNP
Cystic fibrosis	NTLA	n/a	<i>Ex vivo</i>	In collaboration with Novartis Institute for Biomedical Research, Inc.
	CRSP	HDR	<i>In vivo</i>	Correction/insertion of CFTR gene
Duchenne muscular dystrophy	EDIT	NHEJ & HDR	<i>In vivo</i>	By AAV or LNP
	CRSP	NHEJ	<i>In vivo</i>	Functional correction of DMD gene by deletion
Friedreich's ataxia	EDIT	NHEJ	<i>In vivo</i>	Correction by small and large deletion; by AAV or LNP
	CRSP	n/a	<i>In vivo</i>	Collaboration with Friedreich's Ataxia Alliance & University of Alabama (Birmingham)
Glycogen storage disease Ia	CRSP	HDR	<i>In vivo</i>	Correction/insertion of G6PC gene
Hemophilia	CRSP	HDR	<i>In vivo</i>	Withdrawn from portfolio Q3/4 2018
Hurler Syndrome	CRSP	HDR	<i>Ex vivo</i>	Correction/insertion of IDUA gene
Inborn errors of metabolism	NTLA	n/a	<i>In vivo</i>	Disease unspecified; by LNP
Leber congenital amaurosis 10	EDIT	NHEJ	<i>In vivo</i>	Correction by small deletion; by local AAV injection
Liver diseases	NTLA	n/a	<i>Ex vivo</i>	Disease unspecified
Primary hyperoxaluria type 1	NTLA	n/a	<i>In vivo</i>	By LNP
Severe combined immunodeficiency	CRSP	HDR	<i>Ex vivo</i>	Withdrawn from portfolio Q3/4 2018
Transthyretin amyloidosis	NTLA	n/a	<i>In vivo</i>	By LNP; collaboration with Regeneron Pharmaceuticals, Inc.
Usher syndrome 2a	EDIT	NHEJ	<i>In vivo</i>	By local AAV injection

Table 4. Vers la correction d'autres maladies rares avec le système CRISPR-Cas9 par les entreprises leader du marché (CRISPR Therapeutics, Editas Medicine, Intellia Therapeutics). (Papasavva et al., Mol. Diagn. Ther., 2019)

mis au point⁸⁴ (Figure 20B). Un modèle murin possédant la séquence humaine du gène *CEP290* avec la mutation *IVS26* a été utilisé. Après injection sous-rétinienne, une fréquence d'édition moyenne de 21% est retrouvée avec un rétablissement phénotypique dépendant de la dose de vecteurs injectés (de 14% à 94%). Ce travail a été complété par l'utilisation de modèle primate (non-humain), dont l'anatomie de la rétine est plus proche de l'humain. Un taux d'édition supérieur à la limite nécessaire à un rétablissement phénotypique a été obtenu. L'ensemble de ces résultats a motivé la conduite du premier essai clinique *in vivo* par CRISPR-Cas9.

c. Vers d'autres essais cliniques

Si le système CRISPR-Cas9 a été naturellement appliqué aux maladies monogéniques avec de nombreux essais cliniques à venir dans les prochaines années (Table 4), actuellement, une part majoritaire des essais cliniques concerne d'autres domaines.

L'immunothérapie anti-cancéreuse en est le meilleur exemple. Le système CRISPR-Cas9 est notamment utilisé pour induire un *knock-out* du gène *PD1* dans des cellules CAR-T pour empêcher leur inactivation. Un essai clinique ambitieux vise à effectuer un *knock-out* multiplex du gène *PD-1* et des gènes *TCR* endogènes, *TRA* et *TRB*, dans des cellules CAR-T⁸⁵ (NCT03545815).

En virologie, le gène *CCR5*, point d'entrée du VIH est également une cible de premier plan. Un essai clinique vise à modifier des CD34⁺ issus de patients atteints du SIDA et d'une hémopathie maligne (NCT03164135).

Pour résumer, moins d'une décennie s'est déroulée entre la preuve d'efficacité du système CRISPR-Cas9 dans des cellules eucaryotes et son utilisation dans plus d'une dizaine d'essais cliniques dans différents domaines. Toutefois, comme toute technologie naissante, cette épopée rapide ne s'est pas faite sans soulever plusieurs questions concernant l'innocuité de la Cas9 et notamment à propos de la génotoxicité potentielle de la cassure double-brin de l'ADN, point commun de toutes les approches présentées ci-dessus.

IV. CRISPR-Cas9, cassure double-brin et génotoxicité

Sont dits génotoxiques, les agents causant directement des lésions à l'ADN et entraînant des mutations. La notion de génotoxicité est donc intimement liée au système CRISPR-Cas9 et sa maîtrise constitue un enjeu majeur en vue d'utiliser l'édition génique comme stratégie thérapeutique sûre. L'objectif est de permettre des conditions d'édition permettant une génotoxicité restreinte aux événements souhaités. Pour y parvenir, les risques génotoxiques et leurs conséquences doivent d'abord être caractérisés puis maîtrisés.

L'un des premiers effets délétères détectés du système CRISPR-Cas9 a été son activité hors-cible (*off-target*). De plus, plusieurs travaux ont également caractérisé certaines formes de génotoxicité non-voulues sur la cible (*on-target*). Cette génotoxicité *on-target* est dépendante du nombre de cassures et peut se restreindre à des anomalies de quelques nucléotides ou toucher des segments d'ADN beaucoup plus importants. Les paragraphes suivants seront consacrés à l'évaluation de la génotoxicité du système CRISPR-Cas9 et aux travaux tentant d'y répondre par l'innovation.

i. Génotoxicité *off-target*

Rapidement après la preuve de son utilisation dans la cellule eucaryote, les défauts de spécificité du système CRISPR-Cas9, lui permettant une activité hors-cible (*off-target*), ont été pointés du doigt. Pour remédier à ce défaut de spécificité, l'un des leviers réside dans la compréhension du fonctionnement de la Cas9. Les observations décrites ci-dessous ont été le terreau d'innovations qui ont mené à l'amélioration de ce système par la création de nouveaux variants.

a. Caractériser le fonctionnement de la RNP Cas9:ARNg

Comprendre le fonctionnement de la Cas9 et de ses mécanismes régulateurs est essentiel. L'induction de la CDB par la Cas9 est une série d'étapes coordonnées dont chacune est finement régulée. La Cas9 est une protéine complexe et multidomaine dont l'activité endonucléasique ciblée nécessite à la fois sa liaison à un ARNg, pour la rendre compétente à la reconnaissance de l'ADN, et des changements conformationnels, pour permettre son activité endonucléasique.

Structure de la Cas9

Au total, sept domaines composent la Cas9⁸⁶ (*Figure 21A*):

- Le lobe de reconnaissance (REC) est composé de trois domaines alpha-hélicoïdaux (REC1, 2 et 3). Ceux-ci sont impliqués dans la régulation de l'interaction entre l'ARNg et la cible ADN.
- Le lobe nucléasique (NUC) porte les domaines héminucléases RuvC et HNH et un domaine C-terminal variable (CTD). Les domaines HNH et RuvC sont respectivement responsables de la coupure du brin ciblé (complémentaire au crRNA) et du brin non-ciblé.
- Les deux lobes sont reliés par un pont riche en arginine (*Figure 14D*).

Formation RNP Cas9:ARNg, premier changement conformationnel

Sous forme d'apo-enzyme (sans ARNg), bien qu'elle soit capable de se lier à l'ADN de façon non-spécifique, et peu affine, la Cas9 est inactive⁸⁷. La spécificité de l'activité endonucléasique est acquise uniquement après la liaison de la Cas9 à un ARNg pour former une RNP.

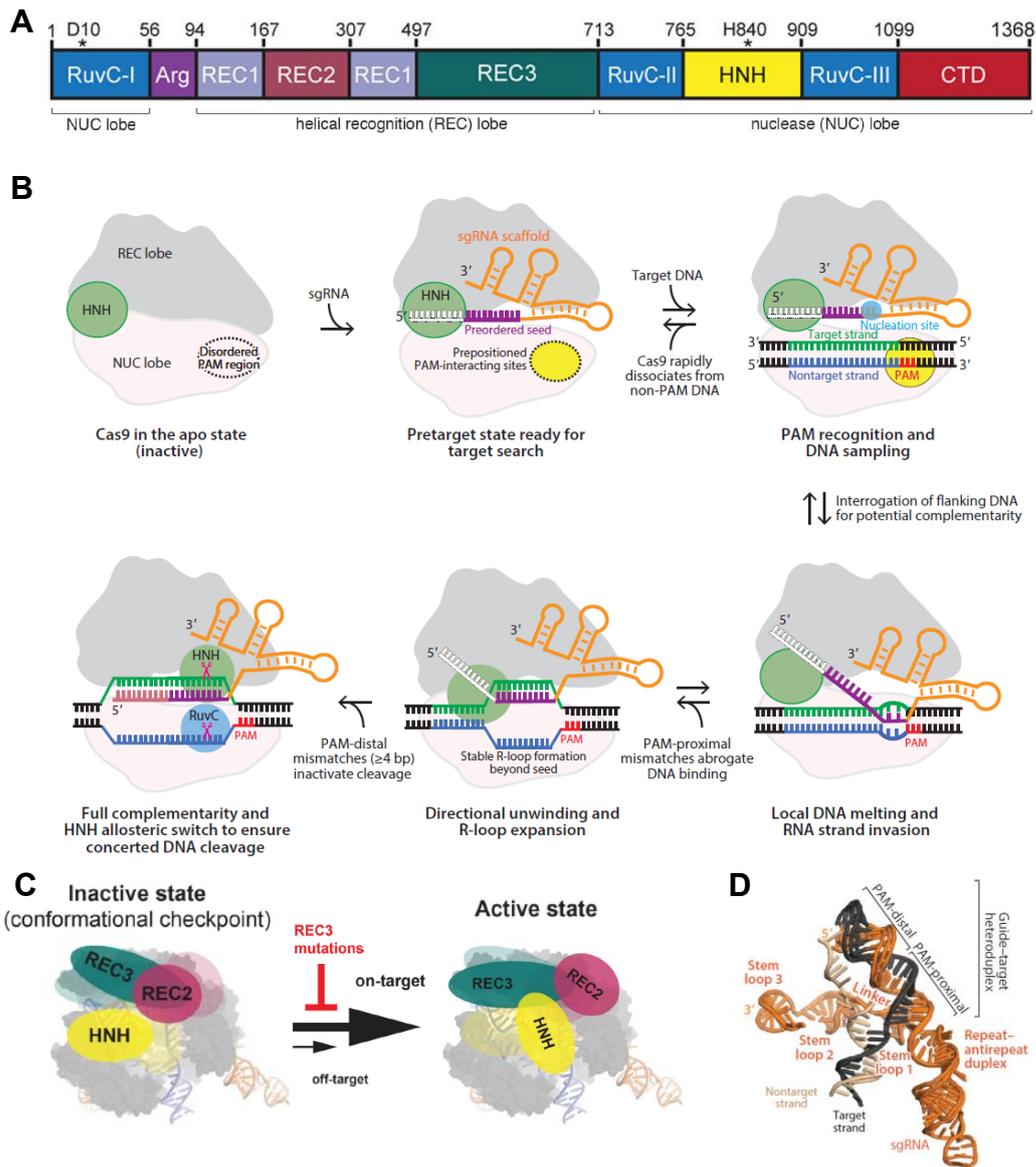


Figure 21. Structure et fonctionnement de la RNP SpCas9:ARNg

(A) Représentation schématique des domaines de la SpCas9. Les résidus D10 et H840 sont essentiels à l'activité respective du domaine RuvC et HNH. (modifiée d'après Chen *et al.*, *Nature*, 2017).

Arg, pont riche en arginine ; **CTD**, *C-terminal domain* ; **NUC**, lobe nucléasique ; **REC**, lobe de reconnaissance

(B) Représentation schématique du fonctionnement de la SpCas9. Les résidus arginines du CTD sont réorganisés (cercle jaune) et les premiers nucléotides du *spacer* de l'ARNg sont pré-organisés en hélice alpha (violet) tandis que la partie distale est non-organisée (blanc). Un second changement conformationnel a lieu si la partie distale (blanche) de l'ARNg est complémentaire du brin ciblé (partie rouge). (C) Ce second changement se traduit par une activation en cascade des domaines REC3, REC2, HNH et enfin RuvC et débouche sur l'induction d'une cassure double-brin. Les mutations dans le domaine REC3 sont à l'origine de Cas9 à spécificité augmentée. (modifiée d'après Jiang et Doudna, *Annu. Rev. Biophys.*, 2017 et (C) Chen *et al.*, *Nature*, 2017)

(D) Analyse cristallographique de RNPs SpCas9:ARNg liées à un fragment d'ADN double-brin cible. La superposition des conformations observées montre que l'hétéroduplexe ADN:ARNg possède une plus grande plasticité dans la région distale au PAM (*PAM-distal*) que proximale (*PAM-proximale*). (Chen *et al.*, *Nature*, 2017)

La comparaison des structures cristallographiques entre l'apo-Cas9 et la RNP Cas9:ARNg montre que la liaison de l'ARNg entraîne un changement conformationnel (*Figure 21B*) ayant deux conséquences principales.

- Les contacts entre la Cas9 et l'ARNg permettent de maintenir les 10 premiers nucléotides du *spacer* situés en 3' dans une position favorable à la formation d'une hélice alpha d'ARN.
- Les résidus arginines du CTD responsables de la reconnaissance du PAM sont repositionnés et permettent à la RNP de devenir compétente pour la reconnaissance spécifique de l'ADN. Pour la Spy.Cas9, ce sont les R1333 et R1335 qui permettent la reconnaissance du PAM 5'-NGG-3'⁴⁹.

Reconnaissance du PAM et test de complémentarité

Une fois la RNP formée et compétente pour la reconnaissance de l'ADN, l'ADN est scanné à la recherche d'une séquence PAM appropriée. La reconnaissance du PAM entraîne une déstabilisation locale de l'ADN. Cette conformation permet l'ouverture du double-brin d'ADN et autorise l'invasion progressive de l'ARN, de l'extrémité proximale au PAM (ou 3' du *spacer*) vers l'extrémité distale (5' du *spacer*), pour former un hétéroduplex ARN:ADN (*Figure 21B*).

Régulation de l'activité endonucléase par REC3, second changement conformationnel

Après la reconnaissance du PAM et la formation de l'hétéroduplex ARN:ADN, l'activité endonucléase peut être activée. L'activité de clivage est plus sélective que la capacité de liaison à l'ADN, au sens où la RNP est capable de se lier à des séquences non-spécifiques sans induire de cassure double-brin. Cette observation suggère qu'il existe un niveau supplémentaire de contrôle entre la reconnaissance de l'ADN et l'activation des domaines nucléasiques.

REC3 est au centre de ce mécanisme de régulation. REC3 agit comme un senseur de la qualité de l'hybridation entre l'ARNg et l'ADN, notamment dans la partie la plus distale au PAM⁸⁶. La présence de mésappariements inhibe le passage de REC3 vers sa forme active. Une fois activé, REC3 agit comme un effecteur allostérique autorisant les changements conformationnels conjugués de REC2 et HNH (*Figure 21C*). Quand REC3 est inactif, REC2 masque le groupement phosphate de la liaison phosphodiester pouvant être clivé par HNH. Le passage de REC3 à l'état actif déclenche un changement de conformation de REC2 puis de HNH. Une fois le domaine HNH repositionné, il stimule l'activité endonucléasique du domaine RuvC, permettant ainsi une coupure synchronisée des deux brins d'ADN.

Bien que sensible aux mésappariements, l'analyse cristallographique montre une plus grande plasticité de l'hétéroduplex dans la région distale au PAM, en lien avec le domaine REC3, que dans sa région proximale, contactant le domaine REC1⁸⁶ (*Figure 21D*). Ces données suggèrent que des mésappariements peuvent être tolérés dans la région distale au PAM malgré le rôle de REC3. Il sera vu que la caractérisation du rôle de REC3 a permis de générer des variants de Cas9 avec une meilleure spécificité.

b. Caractériser l'activité *off-target*

Des mésappariements acceptés

L'activité *off-target* est expliquée par le fait que des mésappariements soient tolérés entre l'ARNg et l'ADN cible et dans la reconnaissance du PAM. L'activité de clivage est affectée par la position et de la nature du mésappariement. D'une manière générale, les mutations affectant une région proche du PAM sont plus sensibles que les mutations affectant la région distale au PAM, pouvant être expliquées par la plasticité différente entre REC1 et REC3. Cette particularité permet de définir deux régions distinctes dans l'ARNg : une *seed sequence*, proximale au PAM et peu tolérante aux mésappariements et une région distale moins sensible aux mésappariements. La *seed sequence* est généralement définie par les 8 à 12 nucléotides proches du PAM (3' du *spacer*), les nucléotides restants formant la région distale⁸⁸.

Nature et nombre de mésappariements

Toutefois, l'étude menée par Hsu et al. permet de nuancer cette règle en montrant que la nature du mésappariement est un facteur tout aussi important⁸⁹. En effet, les mésappariements ARN:ADN rG:dT sont très bien supportés quelle que soit leur position (*Figure 22A*). À l'inverse, la combinaison rC:dC n'est tolérée que dans la région distale. Il est à noter qu'un mésappariement unique peut conduire à des efficacités aussi hautes qu'avec une hybridation parfaite. Le nombre de mésappariement est également un facteur influençant la capacité de clivage. De plus, une activité de clivage est possible même en présence de deux mésappariements, et voire avec trois mésappariements en tandem dans la partie distale au PAM (*Figure 22B*).

Insertion et délétion

Dans la même lignée, Doench et al. ont poursuivi la caractérisation de l'analyse de l'activité *off-target*, en incluant l'impact d'insertion ou de délétion entre la séquence de l'ARNg et l'ADN-cible⁹⁰. Pour cela, des ARNg visant la séquence codante du locus *CD33* ont été formés, ainsi que leurs variants respectifs portant soit un mésappariement, une délétion ou une insertion dans la séquence-cible. Au total, une librairie de 27,897 ARNg a été formée, complétée de 10,618 ARNg servant de témoins négatifs en visant le gène *THY1*.

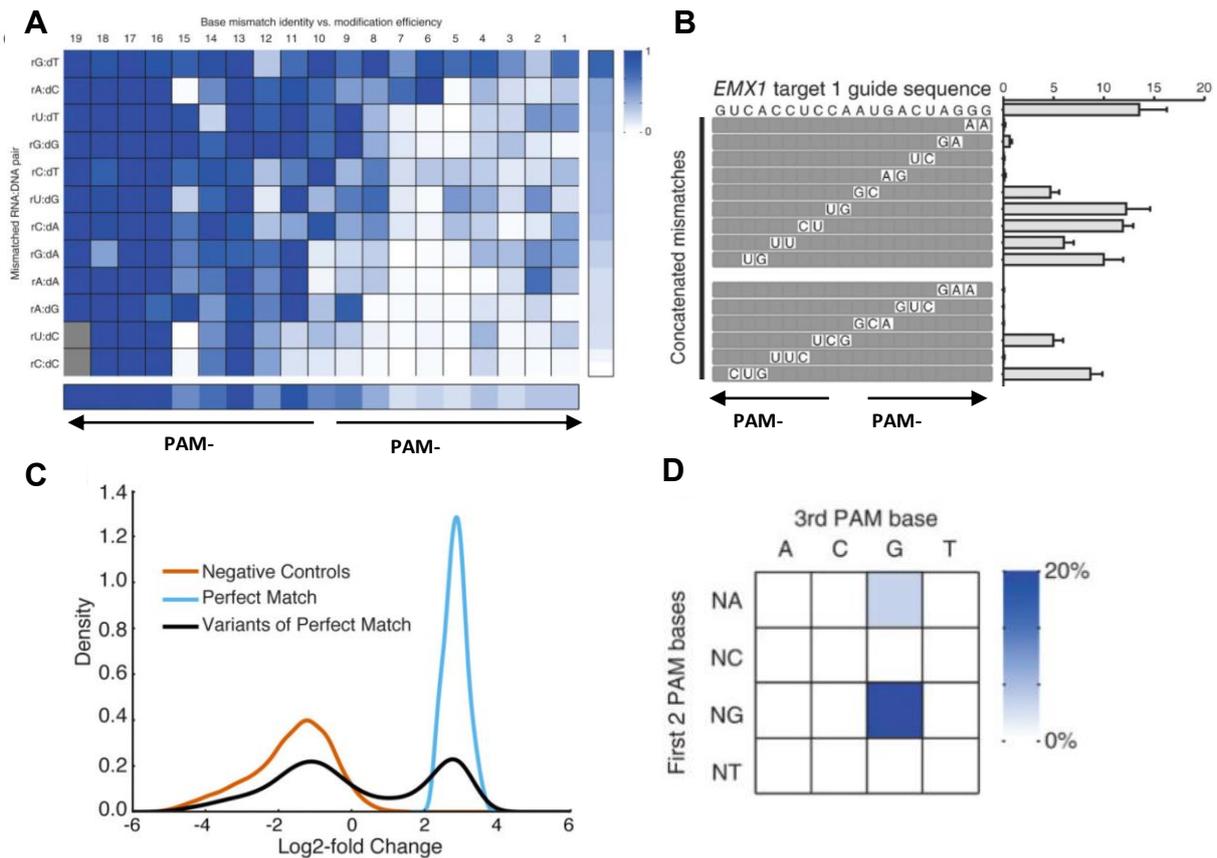


Figure 22. Caractériser l'activité off-target. (modifiée d'après Hsu et al., *Nat Biotechnol*, 2013 et (C) Doench et al., *Nat. Biotechnol.*, 2016)

(A) Heatmap de l'activité de la SpCas9 en présence d'un unique mésappariement. L'activité moyenne pour un type de mésappariement a été calculée respectivement avec les 10 nucléotides proches du PAM (PAM-proximal) (barre à droite). L'activité moyenne pour une position donnée est la moyenne de la colonne correspondante (barre en bas). Le mésappariement rG:dT (première ligne) est très bien supporté.

(B) Activité de la SpCas9 en présence de 2 ou 3 mésappariement consécutifs pour un locus.

(C) Distribution des valeurs \log_2 -fold change, représentative de leur activité de coupure, pour trois population d'ARNg : 65 ARNg parfaitement complémentaires (*Perfect match*) au locus *CD33*, permettant son *knock-out* (contrôle positif), 9,914 ARNg portant une insertion, une délétion ou un mésappariement (*Variant of Perfect Match*) et des ARNg non-complémentaires au locus *CD33*, servant de contrôles négatifs (*Negative Controls*). Une valeur élevée de \log_2 -fold change signifie que l'ARNg a été enrichi dans la population *knock-out* pour *CD33*.

(D) Heatmap de l'activité de la spCas9 en fonction de la séquence du PAM. Pour la SpCas9, le PAM canonique est 5'-NGG-3', toutefois une activité est observée pour le PAM 5'-NAG-3'.

Si 65 ARNg parfaitement complémentaires ont permis une extinction du gène *CD33*, un total de 9,914 variants l'ont également fait (Figure 22C). L'analyse des variants met en lumière que, malgré une insertion ou une délétion dans la séquence du guide ARN, une activité non-nulle peut être retrouvée.

Des PAMs alternatifs reconnus

Enfin, l'étude de la spécificité de la séquence du PAM (5'-NGG-3' pour Spy.Cas9) a permis de montrer qu'une séquence 5'-NAG-3' pouvait être reconnue, bien que l'efficacité soit divisée par cinq comparée à un PAM canonique (Figure 22D). L'utilisation de PAM « alternatifs » (5'-NCG-3' et 5'-NGA-3') a également été rapportée dans une autre étude avec également des efficacités moindres mais non négligeables⁹⁰.

c. Répondre au défi de la spécificité

Pour répondre au manque de spécificité du système CRISPR-Cas9, plusieurs solutions ont été proposées en s'appuyant sur la bio-informatique, en mimant les générations précédentes d'endonucléases ou en développant des ARNg modifiés et des variants de la Cas9.

Une aide bio-informatique

L'une des solutions les plus intuitives est de prêter une attention particulière au choix de la séquence cible en évitant les ARNg susceptibles de reconnaître des séquences hors-cibles. Pour y parvenir plusieurs algorithmes ont été mis au point en se basant sur les observations obtenues dans les études présentées ci-dessus. Un premier algorithme (Hsu-score ou MIT-score) attribuant un score à chaque ARNg selon la position et le nombre de mésappariement a été mis en place en suivant l'étude menée par Hsu et al. facilitant le design d'un ARNg optimal⁸⁹. Un second, le *Cutting Frequency Determination (CFD) score* a été conçu en intégrant les résultats obtenus par Doench et al⁹⁰. En comparant les deux outils, une meilleure courbe ROC est obtenue avec le CFD qu'avec le MIT-score pour l'identification des *off-targets*⁹¹. Ces scores, intégrés sur plusieurs plateformes de design d'ARNg comme CRISPOR⁹¹, constituent une première barrière contre le problème d'activité *off-target*. Le développement d'outils de prédictions est une aire active de recherche. À la croisée entre l'intelligence artificielle et la biologie, un nouvel outil basé sur l'apprentissage automatique (*machine-learning*) a été développé : Elevation⁹². En croisant leurs prédictions avec des banques de données disponibles, Lisgarten et al. ont obtenu une meilleure prédiction que celles utilisant l'algorithme de MIT-score ou le CFD. Il est intéressant de noter que les auteurs ont voulu intégrer un nouveau paramètre dans leur analyse, celui de l'accessibilité à la chromatine.

Cependant, l'impact sur la capacité prédictive a été non-concluant ; la raison évoquée étant une trop faible disponibilité des données. Toutefois, cette tentative illustre la possibilité d'implémenter de nouveaux facteurs pouvant influencer l'activité *off-target* dans l'outil de prédiction.

Si toutes ces approches permettent de faire un choix « optimal », ce choix n'est pas toujours permis, comme c'est le cas en vue d'une correction d'une mutation pathologique. Plusieurs approches ont été proposées pour remédier à ce manque de spécificité modifiant directement la RNP Cas9:ARNg.

Des systèmes s'inspirant des générations précédentes d'endonucléases

Deux approches se sont inspirées des ZFNs et des TALENs pour améliorer le système CRISPR. Les ZFNs et les TALENs reposent sur l'utilisation d'une paire de « guides » apportant chacun une unité de *FokI*.

Système *paired-nickase*

L'approche proposée par Ran et al. réutilise l'idée d'une paire de guides pour augmenter la spécificité du système CRISPR-Cas9⁹³ (Figure 23A). Pour cela, la Cas9 est mutée sur le domaine nucléase RuvC pour former la Cas9^{D10A}-nickase. Celle-ci ne génère plus que des cassures simple-brin (CSBs ou *nicks*). Si la Cas9^{D10A}-nickase est couplée à une paire d'ARN guides ciblant deux loci proches d'une centaine de paires de base, cette approche permet d'induire des « double cassures simple-brin », reconnues comme une CDB par la cellule. Ce système de *paired-nickase* augmente la spécificité par deux avantages. L'utilisation d'une paire de guides impose la reconnaissance d'une grande séquence avec une conformation précise pour autoriser l'induction d'une CDB. Bien que l'activité nickase *off-target* intrinsèque à chaque ARNg reste inchangée, l'utilisation de la Cas9^{D10A} empêche la formation de CDB *off-target*.

Système *deadCas9-FokI*

Suivant la même idée, Guilinger et al. ont proposé d'utiliser la Cas9 comme simple véhicule apportant l'endonucléase *FokI*⁹⁴ (Figure 23B). Pour cela, la Cas9 a été inactivée sur les deux domaines nucléase (*dCas9* pour *deadCas9*) et a été fusionnée à la nucléase *FokI*. Comme précédemment, deux complexes *dCas9-FokI* doivent être dirigés par un couple d'ARNg dans une conformation précise pour permettre la formation d'un dimère de *FokI*, nécessaire à l'induction de la CDB. De la même façon que précédemment, la spécificité de ce système est augmentée par la contrainte de localisation et de reconnaissance de deux séquences complémentaires aux ARN guides. De plus, le système *dCas9-FokI* présente un avantage par rapport au système *paired-nickase*. En effet, en cas de fixation hors-cible d'un des deux composants, l'utilisation d'une nickase entraîne la formation d'une cassure simple-brin alors qu'un monomère de *FokI* ne possède aucune activité nucléasique et laisse donc l'ADN intacte.

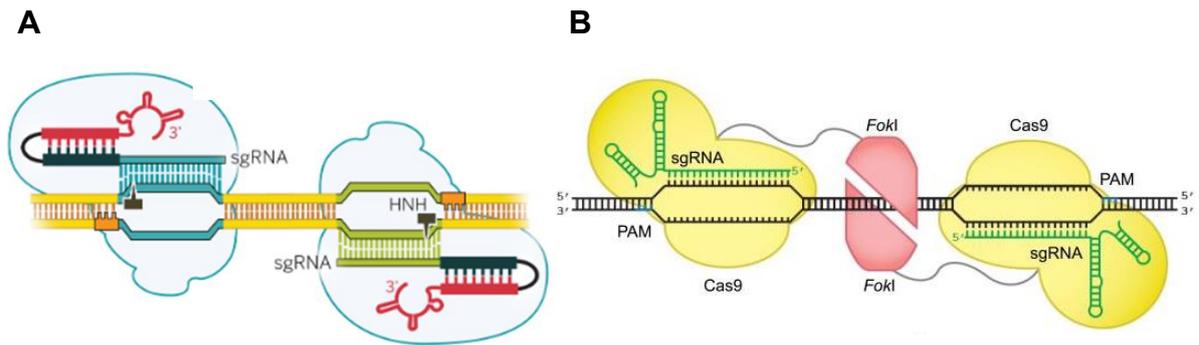


Figure 23. Des systèmes s'inspirant des générations précédentes d'endonucléases.

(A) Système *paired-nickase*. La Cas9^{D10A} (en bleu pâle) (domaine RuvC muté) induit une cassure simple-brin uniquement sur le brin ciblé (flèche noire). (Doudna et Charpentier, *Science*, 2014)

(B) Système *deadCas9-FokI*. La deadCas9 (en jaune) (domaine RuvC et HNH mutés) ne possède plus d'activité nucléasique. (Guilinger et al., *Nat. Biotechnol.*, 2014)

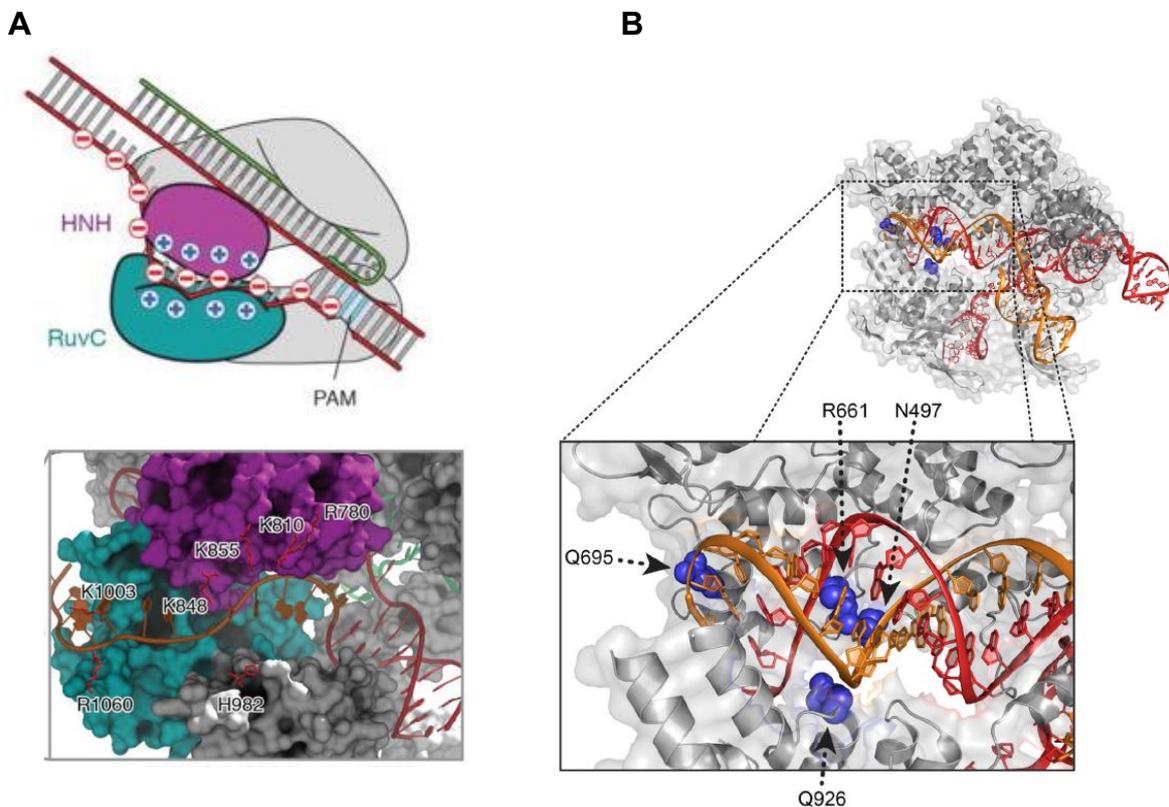


Figure 24. Variants haute spécificité, eSpCas9-1.1 et SpCas9-HF1.

(A) (Haut) Représentation schématique de la structure de la SpCas9 et du sillon chargé positivement en interaction avec l'ADN chargé négativement. (Bas) Vue détaillée des résidus en interaction avec l'ADN. Pour l'eSpCas9-1.1 les résidus K848, K1003 et R1060, situés dans le domaine RuvC ou HNH, ont été mutés en alanine. (Slaymaker et al., *Science*, 2016)

(B) (Haut) Structure cristallographique de la SpCas9:ARNg en interaction avec une cible ADN. (Bas) Vue détaillée des résidus formant des ponts hydrogènes avec le brin ciblé (en bleu). Pour la SpCas9-HF1, les résidus N497, R661, Q695 et Q926 ont été mutés en alanine. (Kleinstiver et al., *Nature*, 2016)

Malgré l'augmentation significative de la spécificité, l'ajout d'éléments supplémentaires pour permettre l'édition complexifie le design et la vectorisation de ces systèmes, limitant leur utilisation. A l'inverse, les stratégies présentées ci-dessous se sont évertuées à augmenter la spécificité du système CRISPR-Cas9, sans ajouter d'autres éléments, en modifiant soit la Cas9 soit la structure de l'ARNg.

Cas9 haute spécificité

Enhanced specificity Cas9 (eCas9-1.1) et High-Fidelity Cas9 (Cas9-HF1)

L'analyse des données cristallographiques de la structure de la Cas9 en complexe avec l'ARNg et l'ADN cible, a permis d'identifier un sillon chargé positivement entre les domaines RuvC et HNH⁸⁸ (*Figure 24A*). Les données semblent indiquer que ce sillon participe à l'ouverture du double-brin d'ADN en séquestrant le brin non-ciblé dans ce sillon chargé positivement et en l'empêchant donc de se réhybrider avec le second brin d'ADN. Cette conformation favorisant l'hybridation de l'ARNg au brin d'ADN ciblé. Partant de ce postulat, deux équipes ont émis l'hypothèse que la neutralisation des résidus chargés positivement pourrait affaiblir la capacité de séparation des brins d'ADN, encourageant la ré-hybridation entre les deux brins d'ADN. Ainsi, l'hybridation entre l'ARNg et le brin ciblé serait moins favorisée et nécessiterait un appariement plus stringent entre l'ARNg et l'ADN cible. Ce sujet d'étude a donné naissance à deux variants de la Cas9, la *enhanced specificity* SpCas9 (eSpCas9-1.1)⁸⁸ (*Figure 24A*) et la Cas9 *High-Fidelity* (SpCas9-HF1)⁹⁵ (*Figure 24B*), mutés respectivement sur 3 et 4 résidus chargés positivement. Ces deux variants ont montré une diminution de l'activité *off-target* tout en conservant une efficacité *on-target* équivalente (*Figure 25D*).

Cependant, une analyse mécanistique par FRET, publiée un an plus tard, a mis en lumière que le gain de spécificité était dû à un mécanisme plus fin⁸⁶. Cette nouvelle étude montre que les variants eSpCas9-1.1 et SpCas9-HF1 (*Figure 25A*) conservent une affinité de liaison à l'ADN similaire à la Cas9 que ce soit pour des cibles entièrement complémentaires ou portant des mésappariements dans la partie distale au PAM (*Figure 25B*). Ce résultat contredit l'hypothèse de départ stipulant que les mutations diminuent la liaison de la RNP à la séquence cible. Par contre, la transition de HNH vers une conformation active est plus sensible aux mésappariements pour les variants eCas-1.1 et Cas9-HF1. Ce résultat suggère que le gain de spécificité est dû à une modification de la régulation de l'activité endonucléasique de la RNP, et non de sa capacité de reconnaissance de l'ADN comme supposé initialement.

C'est cette même étude qui a montré que REC3 est responsable de la transition de HNH vers sa forme active, lui-même régulant le domaine RuvC. Trois des quatre mutations de la SpCas9-HF1 sont situées dans le domaine REC3 ; celles de la eSpCas9-1.1 affectant les domaines RuvC et HNH.

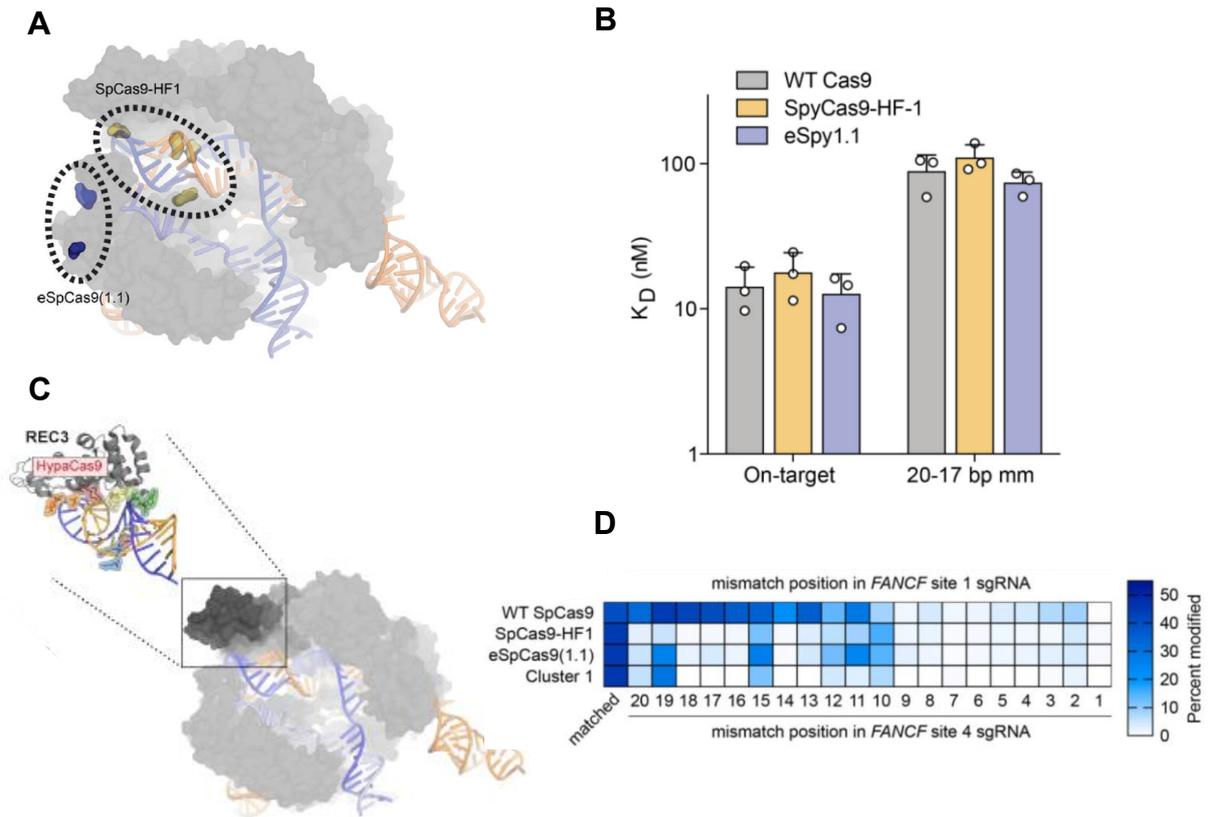


Figure 25. Implication de REC3 dans la spécificité. (modifiée d'après Chen et al. Nature, 2017).

(A) Structure cristallographique de la SpCas9:ARNg en interaction avec une cible ADN. Les mutations de l'eSpCas9-1.1 et de la SpCas9-HF1 sont entourées.

(B) Quantification de la constante de dissociation (K_D) de la SpCas9 WT, d'eSpCas9-1.1 et de SpCas9-HF1 pour une cible d'ADN parfaitement complémentaire (*on-target*) ou avec un mésappariement en position distale du PAM (20-17 bp mm).

(C) Structure cristallographique de la SpCas9:ARNg en interaction avec une cible ADN avec vue détaillée du domaine REC3. Le cluster 1 de REC3 comprend quatre résidus (N692, M694, Q695, H698) en interaction avec l'interface ADN:ARNg. La mutation de ces quatre résidus en alanine forme la Hypa-Cas9.

(D) *Heatmap* de l'activité d'eSpCas9-1.1, de SpCas9-HF1 et d'Hypa-Cas9 (Cluster 1) en présence d'un unique mésappariement.

Hyper-accurate Cas9 (Hypa-Cas9) et Evolved Cas9 (evoCas9)

Profitant des résultats nouvellement acquis, les auteurs ont formé un nouveau variant de la Cas9 en mutant plusieurs acides aminés de REC3 en contact direct avec l'hétéroduplex ADN:ARNg, formant ainsi la Hypa-Cas9 (pour *Hyper accurate-Cas9*)⁸⁶ (Figure 25C). Ce nouveau variant montre une activité sur cible similaire à la Cas9 classique, mais avec une spécificité meilleure et légèrement supérieure à celle d'eSpCas9-1.1 et SpCas9-HF1 (Figure 25D).

Cette étude a largement approfondi la compréhension des mécanismes menant à l'induction de la CDB, et notamment de REC3, et a ouvert la porte à d'autres optimisations jouant sur les changements conformationnels. C'est dans cette lignée que s'inscrit la caractérisation de l'*evolved-Cas9* (evoCas9)⁸⁶. Cette dernière est issue d'une méthode d'évolution dirigée conduite chez la levure. Cette stratégie a mêlé une approche sans *a priori* et rationnelle, au sens où une banque de mutations aléatoires a été réalisée en effectuant une PCR non-fidèle, mais uniquement au niveau du domaine REC3 (et non sur l'ensemble de la Cas9). Le criblage a permis d'isoler un clone portant 4 mutations dans le domaine REC3 (M495V/Y515N/K526E/R661L), dont l'une a été légèrement modifiée pour rétablir l'activité *on-target* (R661L changé par R661Q diminuant l'encombrement stérique). Comme pour la *Hypa-Cas9*, l'*evoCas9* montre une meilleure activité *on-target* que la Cas9-HF1 et moins de *off-target*.

Sniper-Cas9 et High-Fidelity Cas9 (HiFi-Cas9)

Deux autres variants de la Cas9 sont issus d'un criblage de Cas9 mutées dans un système bactérien. Cette fois-ci, aucun *a priori* n'a été pris sur les domaines à muter ; les mutations étant également réparties le long de la séquence codante de la Cas9.

Le premier crible, testant plus de 10 millions de variants, a permis de caractériser la *Sniper-Cas9*⁹⁷, portant 3 mutations ponctuelles affectant les domaines REC1, RuvC, et HNH. La *Sniper-Cas9* possède une activité *on-target* supérieure aux autres variants de la Cas9, cependant cette efficacité est au prix de sa spécificité ; les mésappariements distaux étant tolérés.

Le second crible conduit par Vakulskas et al. a permis la caractérisation d'une Cas9 porteuse d'une unique mutation ponctuelle (R691A) dans le domaine REC3, la *High-Fidelity Cas9* (HiFi-Cas9)⁹⁸. Contrairement aux études précédentes, les paramètres de ce variant ont été décrits lors d'une utilisation sous forme de ribonucléoprotéine (RNP). Ici, la quantification *in vitro* de l'activité *on-target* marque un gain d'efficacité frappant comparé aux autres variants (Figure 26A). Cette efficacité est accompagnée d'une réduction de l'activité *off-target* (Figure 26B). L'étude est complétée par la preuve de l'efficacité et de la spécificité de la HiFi-Cas9 à éditer des cellules CD34⁺ drépanocytaire porteuses de la mutation p.Glu6Val dans le gène *HBB*.

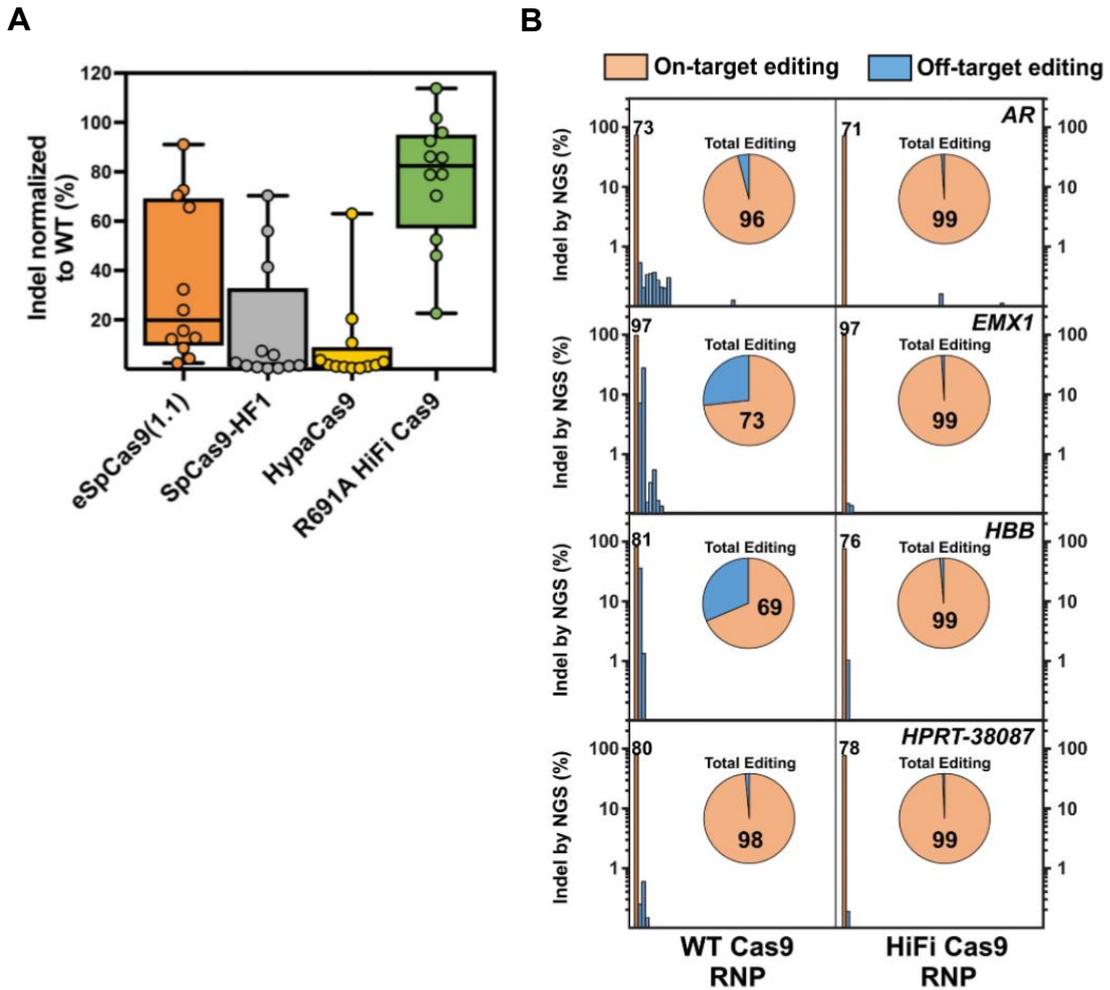


Figure 26. Variant haute spécificité : High-Fidelity Cas9. (modifiée d'après Vakulskas et al., Nat. Med., 2018)

(A) Quantification de l'activité *on-target* de quatre variants de la Cas9 au locus *HPRT*. L'activité est normalisée par rapport à la SpCas9 WT.

(B) Quantification de l'activité *on-target* et *off-target* sur 4 loci de la Cas9 WT et de la HiFi-Cas9. Les cellules HEK293 ont été transfectées avec des RNPs.

Toutes ces études montrent qu'il est possible d'améliorer la spécificité de la Cas9 en modifiant sa structure primaire. Si les mutations ne peuvent pas être directement appliquées à d'autres types de Cas, les méthodologies de criblage sauront certainement fournir des versions plus spécifiques des autres Cas utilisées en vue d'édition génique. Très récemment, une première amélioration de la spécificité de la Cas12a a été faite par l'introduction d'une mutation ponctuelle, appelée Cas12a-HF1⁹⁹.

En parallèle de ces travaux, plusieurs équipes de recherche se sont consacrées à l'amélioration de la spécificité en modifiant l'autre composant du système CRISPR, à savoir l'ARNg.

Modification de l'ARNg

Truncated-ARNg

Pour augmenter la spécificité, une des premières approches s'est intéressée à l'utilité de la partie distale de l'ARNg, tolérante aux mésappariements¹⁰⁰. Fu et collègues ont montré que la troncature de l'ARN guide de 2 à 3 nucléotides à l'extrémité 5' (*truncated-ARNg*), minimisant l'interface ARN:ADN, permettait de réduire l'activité hors-cible (*Figure 27A*). Malgré tout, cette approche présente deux inconvénients : i) l'utilisation d'ARNg tronqués engendre une perte d'efficacité pour certains loci comparée aux ARNg classiques et ii) la réduction de l'ARNg de 20 à 18 nucléotides diminue la taille de la séquence reconnue et augmente donc mécaniquement le nombre de *off-target* potentiels.

Hairpin-ARNg

A l'inverse, Kocak et al. ont émis l'hypothèse que le rajout de structures secondaires à l'extrémité 5' de l'ARNg pourrait créer une barrière énergétique supplémentaire à la formation de la conformation nécessaire à l'activité endonucléase de la Cas9¹⁰¹ (*Figure 27B*). Cette structure supplémentaire en *hairpin* affaiblirait l'interaction entre l'ARNg et l'ADN, rendant le système plus sensible aux mésappariements de séquence. Différentes combinaisons de taille et de structure ont été comparées en utilisant un système rapporteur basé sur le gène *IL1RN*. Une première observation montre que l'augmentation de la taille de l'*hairpin* diminue l'activité *on-target*, plaidant en faveur de l'hypothèse de départ. Un *hairpin* composé d'une boucle de 4 nucléotides et d'une tige de 5 nucléotides permet de garder un bon ratio entre conservation de l'efficacité et augmentation de la spécificité. En effet, l'analyse *off-target* a permis de montrer une diminution globale de l'activité non-spécifique (*Figure 27C*). Malgré tout, l'ajout d'une structure secondaire impacte l'efficacité à certains loci.

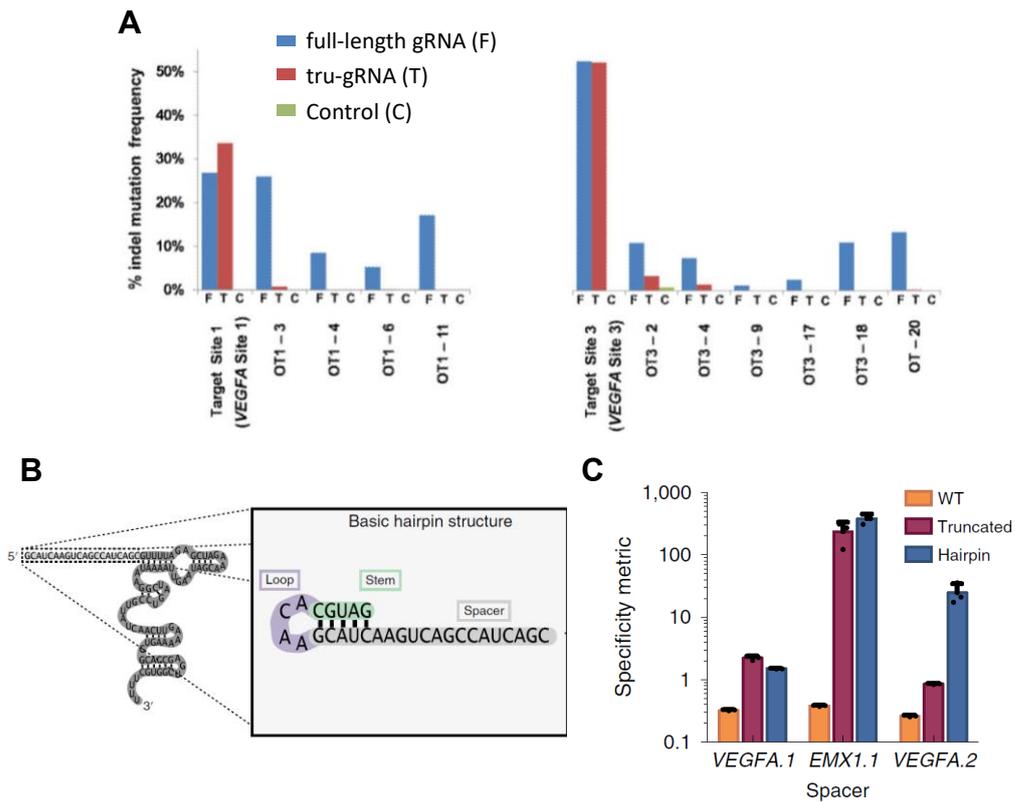


Figure 27. Modification de l'ARNg pour améliorer la spécificité.

(A) Quantification de l'activité *on-target* (*target site*) et *off-target* (OT) sur deux loci du gène *VEGFA* obtenus par transfection d'une RNP Cas9 dirigée par un ARNg de 20 nucléotides (*full-length gRNA*) ou de 18 nucléotides (*tru-gRNA*). Une condition contrôle avec un ARNg ne ciblant pas le locus est ajoutée (*Control*). (modifiée d'après Fu et al., Nat. Biotechnol. 2014)

(B) Représentation schématique de la structure secondaire ajoutée en 5' de l'ARNg. (modifiée d'après Kocak et al., Nat. Biotechnol. 2019)

(C) Quantification du score de spécificité obtenu avec l'ARNg WT, tronqué (*truncated*), ou avec un hairpin en 5' (*hairpin*). Le score de spécificité est défini comme la fréquence *on-target* divisée par la fréquence cumulée des différents *off-target*. (modifiée d'après Kocak et al., Nat. Biotechnol. 2019)

Bien qu'encore imparfaite, les stratégies se basant sur la modification de l'ARNg explorent les mécanismes qui régissent l'interaction entre l'ARN et l'ADN cible, or, malgré quelques spécificités, c'est une étape commune à la plupart des systèmes CRISPR. En ce sens, les travaux faits sur un type peuvent être appliqués à d'autres. Ainsi, Kocak et al. ont appliqué la même méthodologie à la SaCas9 et à la Cas12a et ont obtenu des résultats similaires¹⁰¹.

Des améliorations utilisées ?

A ce jour, ces innovations sont en développement, peu disponibles et ne sont pas souvent utilisées en préclinique ou clinique. Seulement très récemment, la HiFi-Cas9 a été commercialisée par Integrated DNA Technologies Inc. (IDT) sous forme de RNP et est utilisée en préclinique pour une correction directe de la mutation de la drépanocytose⁷⁸. Ainsi, la Cas9 reste encore largement la plus utilisée et sa génotoxicité *off-target* nécessite d'être étudiée.

ii. Génotoxicité *on-target* – Echelle du nucléotide

L'objectif principal de l'édition génique est d'induire des mutations dans une cible définie. Pour la plupart des approches, cela se traduit par des modifications se confinant de quelques paires de base à quelques centaines tout au plus. Les conséquences de la réparation d'une CDB sont généralement de deux types : InDels ou HDR, si une matrice donneuse est apportée. Si pour les stratégies visant à faire du *knock-out*, il est relativement simple d'avoir uniquement des InDels, pour les approches par *knock-in*, faire de l'HDR sans induire en parallèle une part de génotoxicité non-voulue, par l'apparition d'InDels, demeure un réel défi ; la balance entre HDR et InDels devenant un véritable marqueur de la qualité du procédé d'édition génique.

a. Des InDels non-voulus à risque ?

L'étude de Yang et al. visant à corriger le déficit en ornithine carbamyl transférase (OTC) de modèle murin par *knock-in* illustre le risque potentiel des InDels⁸⁰. Lorsque les souris adultes, porteuses d'une mutation ponctuelle, ont été traitées avec les deux AAVs pour permettre la correction de la mutation, l'état des souris a empiré et toutes ont été euthanasiées moins de 5 semaines après le traitement (*Figure 16B*). L'analyse a montré une élévation du taux d'acide orotique urinaire avec une hyperammoniémie, caractéristiques du déficit en OTC, par rapport à des souris non-traitées. Ces observations suggèrent que l'activité résiduelle du cycle de l'urée, présente malgré la mutation ponctuelle, a été perdue par l'apparition d'InDels touchant l'exon situé juste en amont de la mutation à corriger. Cet exemple montre le risque potentiel des InDels pour les pathologies dont les mutations ponctuelles permettent une activité résiduelle. Un second exemple récent des effets non-voulus est donné par l'étude de Park et al. visant à corriger des cellules CD34⁺ drépanocytaires par une stratégie de *knock-in*⁷⁸.

Lors de l'édition de cellules issues de patients avec un complexe RNP Cas9:ARNg dirigé contre le gène *HBB*, une augmentation du pourcentage d'HbF a été observée, et ce en présence ou non et du ssODN porteur de la correction. Les auteurs supposent que ce phénomène pourrait être expliqué par l'introduction biallélique de mutations β -thalassémiques ou par une augmentation opportuniste de la formation d'HbF puisque les chaînes β -globines ne sont plus disponibles. Si les auteurs s'interrogent brièvement sur le bénéfice du maintien d'un niveau d'HbF à un niveau thérapeutique, la question de savoir si le bénéfice de cet effet non-voulu n'est pas dépassé par les risques d'induire une β -thalassémie mineur ou majeur revient rapidement.

Ces deux exemples illustrent le besoin de supprimer les InDels pour les approches par *knock-in* tout en conservant une forte efficacité d'HDR, autrement dit maximiser le ratio HDR/InDels. La réussite de l'édition génique par *knock-in* repose donc sur la maîtrise des produits de réparation et doit passer par la compréhension des voies de réparation et des mécanismes de régulation qui les régissent.

b. Connaître les voies de réparation pour en maîtriser la finalité

Pour la cellule, la CDB est un événement indésirable qui doit être résolu pour maintenir son état physiologique. Des voies de réparation altérées sont notamment associées à plusieurs maladies génétiques, à une prédisposition au cancer, à des troubles du développement et un vieillissement prématuré¹⁰². Pour y répondre, l'évolution a doté les cellules eucaryotes d'un système de reconnaissance des lésions génomiques et d'une batterie de voies de réparation, finement régulées.

Comprendre les voies de réparation et leur fonctionnement a permis le développement de nouveaux outils pour améliorer la balance HDR / Indels. Si les acteurs-clés des voies de réparation principales sont connus depuis un certain temps et ont été les premières cibles d'inhibiteurs pharmacologiques, un intérêt croissant est porté aux acteurs en amont, c'est à-dire, ceux qui régulent l'entrée dans une voie plutôt qu'une autre. Une littérature croissante montre la compétition entre la protéine 53BP1 et BRCA1 pour « s'adjuger » la prise en charge de la CDB. Cette compétition débouche soit sur la résection des brins, stimulés par BRCA1, soit par un maintien des extrémités, par 53BP1. Les voies de réparation subséquentes sont dépendantes de cet événement, expliquant pourquoi ce nœud central de régulation est la cible de nouveaux inhibiteurs pharmacologiques.

Les paragraphes suivants seront consacrés à comprendre comment le système de reconnaissance des cassures double-brin mène rapidement au recrutement de 53BP1 et de BRCA1. Ensuite, les voies de réparation seront décrites.

Un paragraphe sera ajouté pour détailler par quelles voies les matrices exogènes peuvent participer à la réparation des cassures double-brin. Enfin, la compétition entre 53BP1 et BRCA1 pour la résection des brins, ou non, sera discutée.

c. Détection des cassures double-brin

La reconnaissance des CDB fait intervenir trois principaux acteurs du recrutement de 53BP1 et BRCA1, à savoir le complexe MRN, la kinase ATM et MDC1 (*Figure 28*).

Complexe MRN, premier acteur.

Le complexe MRN est le premier intervenant après une CDB. Il est composé de MRE11, RAD50 et NBS1 et co-localise très rapidement avec les CDBs. Le complexe se fixe au centre de la CDB et forme un hétérotétramère qui maintient les deux extrémités. Dans le détail, MRE11 possède deux domaines de liaison à l'ADN tandis que le pontage des deux extrémités est assuré par RAD50 au niveau de superhélices¹⁰³.

Recrutement et activation de la kinase ATM.

La fixation du complexe MRN au centre de la CDB permet le recrutement de la kinase ATM (*ataxia telangiectasia mutated*), autre élément phare de la réparation des CDB. En absence de lésion génomique, la protéine ATM reste sous forme d'homodimère catalytiquement inactif. Cependant, son recrutement au niveau du complexe MRN permet d'induire une série d'autophosphorylations permettant son activation sous forme de monomère.

Une fois activée, la kinase ATM permet une transduction du signal (détaillée ci-dessous) en phosphorylant des centaines de protéines, dont p53 et CHK2, menant à l'arrêt du cycle cellulaire et/ou à l'apoptose. En parallèle, ATM active les voies de réparation par la formation de γ -H2AX.

Formation de γ -H2AX, marque typique des CDBs.

La phosphorylation de H2AX sur la sérine 139 par la kinase ATM a été documentée dès 1998¹⁰⁴. H2AX est un variant des histones H2A représentant 10 à 15% des histones H2A chez les mammifères¹⁰⁵. A l'état normal, les H2AX sont phosphorylées sur la tyrosine 142 par le *Williams-Beuren syndrome transcription factor kinase* (WSTF) et non-phosphorylées sur la sérine 139.

Toutefois, lors d'une CDB, les premières unités d'ATM recrutées par le complexe MRN, phosphorylent la sérine 139 des histones H2AX, pour former la γ -H2AX, tandis que leur tyrosine 142 est déphosphorylée par les phosphatases EYA (*Eye Absent*)¹⁰⁶. Ce double événement permet le recrutement d'une protéine centrale du processus de réparation, MDC1 (*mediator of DNA damage checkpoint protein 1*).

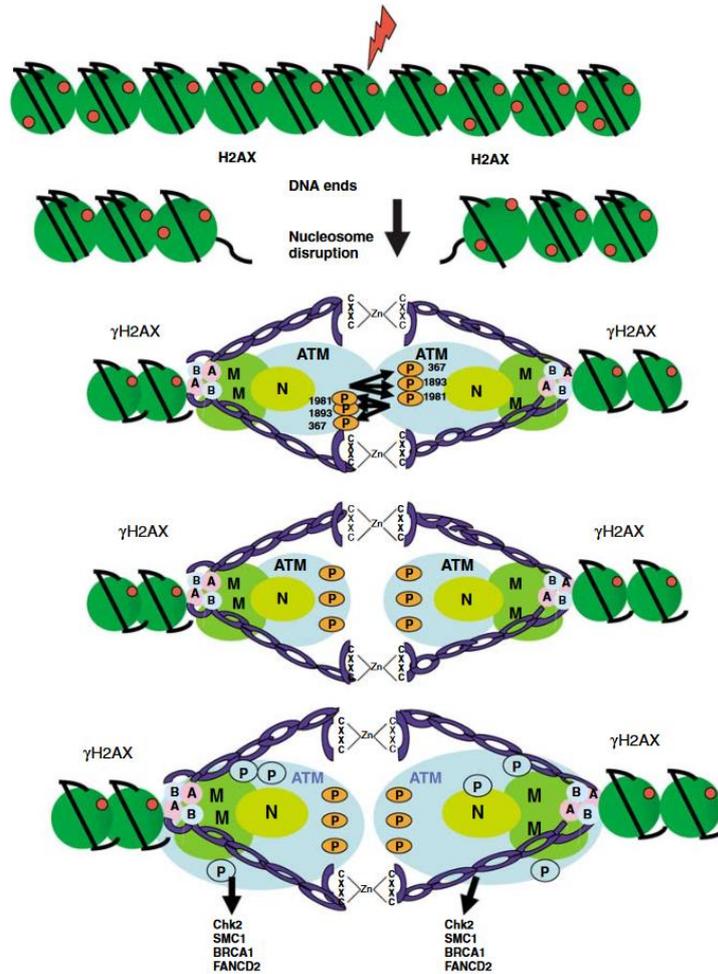


Figure 28. Le complexe MRN et ATM, acteurs précoces de la reconnaissance des CDBs. (*Lavin, Oncogene, 2007*)

Représentation schématique des premières étapes de la reconnaissance des CDBs.

MRN, MRE11, RAD50 et NBS1 ; **ATM**, ataxia-telangiectasia mutated ; **CHK2**, Checkpoint kinase 2 ; **SMC1**, Structural maintenance of chromosomes protein 1 ; **BRCA1**, breast cancer 1 ; **FANCD2**, FA Complementation Group D2

MDC1, un acteur central aux rôles multiples.

MDC1 est constitutivement phosphorylée par la *casein kinase 2*. Cette phosphorylation lui permet d'être en contact, *via* NBS1, avec d'autres complexes MRN que ceux fixés au centre de la CDB. MDC1 est recruté au niveau des γ -H2AX formées. Une fois ancrée au niveau de la CDB, MDC1 possède trois rôles essentiels¹⁰⁵ (*Figure 29*) :

- MDC1 permet l'amplification du signal d'ATM pour former des foci γ -H2AX. MDC1 est en contact avec le complexe MRN, lui-même capable d'activer et d'ancrer ATM. Les protéines ATM propagent la marque γ -H2AX, formant ainsi une boucle de rétrocontrôle positif. MDC1 participe à l'ancrage d'ATM au niveau de la CDB.
- MDC1 participe au recrutement des ubiquitines ligases E3, RNF8 et RNF168. Après avoir été ancrée, ATM phosphoryle MDC1 sur le domaine TQXF. Cette modification permet le recrutement en cascade des ubiquitines ligases RNF8 et RNF168. Ces dernières ubiquitinent les lysines K13 et K15 des histones H2A et les H2A, autorisant la fixation de 53BP1 et BRCA1
- MDC1 participe à la méthylation des histones, *via* MMSET. ATM phosphoryle la méthyltransférase MMSET et permet sa liaison à MDC1. Une fois lié à MDC1, MMSET entraîne la méthylation des histones H4K20. Ces méthylations sont reconnues par la protéine 53BP1.

Une voie commune pour recruter 53BP1 et BRCA1

Ici, l'objectif n'est pas de faire une description exhaustive de tous les acteurs impliqués dans la reconnaissance de la CDB et dans le remodelage chromatinien, mais de montrer que 53BP1 et BRCA1, qui ont des rôles antagonistes par la suite, sont recrutés par des canaux communs. Il est imaginable que favoriser 53BP1 ou BRCA1 par des modulateurs pharmacologiques pour activer préférentiellement l'une ou l'autre des voies de réparation ne pourra être fait que sur des événements en aval de la reconnaissance des CDBs.

d. Voies de réparation des cassures double-brin

Une fois la reconnaissance effectuée et 53BP1 et BRCA1 recrutés, la réparation de la CDB peut débuter. La voie de réparation choisie est dépendante du contexte cellulaire avec en premier lieu la compétition entre BRCA1 et 53BP1. Les voies de réparation peuvent être classées en deux grandes catégories en prenant en compte leur dépendance à l'homologie de séquence¹⁰⁷ (*Figure 30*).

- La voie indépendante de l'homologie, la *canonical non-homologous end-joining* (cNHEJ)

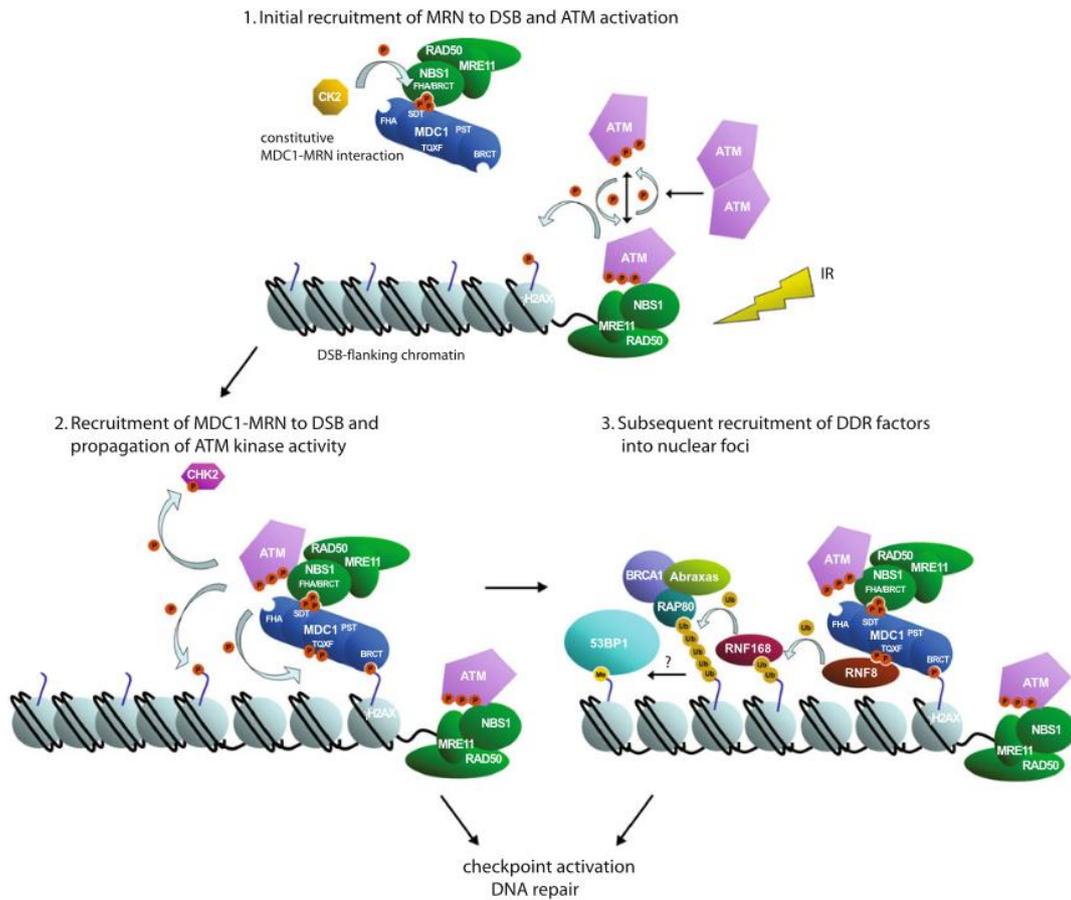


Figure 29. Rôle central de MDC1 dans le recrutement de BRCA1 et 53BP1. (Jungmichel et Stucki, *Chromosoma*, 2010)

Représentation schématique du rôle de MDC1 dans le recrutement de 53BP1 et BRCA1.

MRN, MRE11, RAD50 et NBS1 ; ATM, ataxia-telangiectasia mutated ; CHK2, Checkpoint kinase 2 ; MDC1, Mediator Of DNA Damage Checkpoint 1 ; BRCA1, breast cancer 1 ; Abraxas, BRCA1 A Complex Subunit

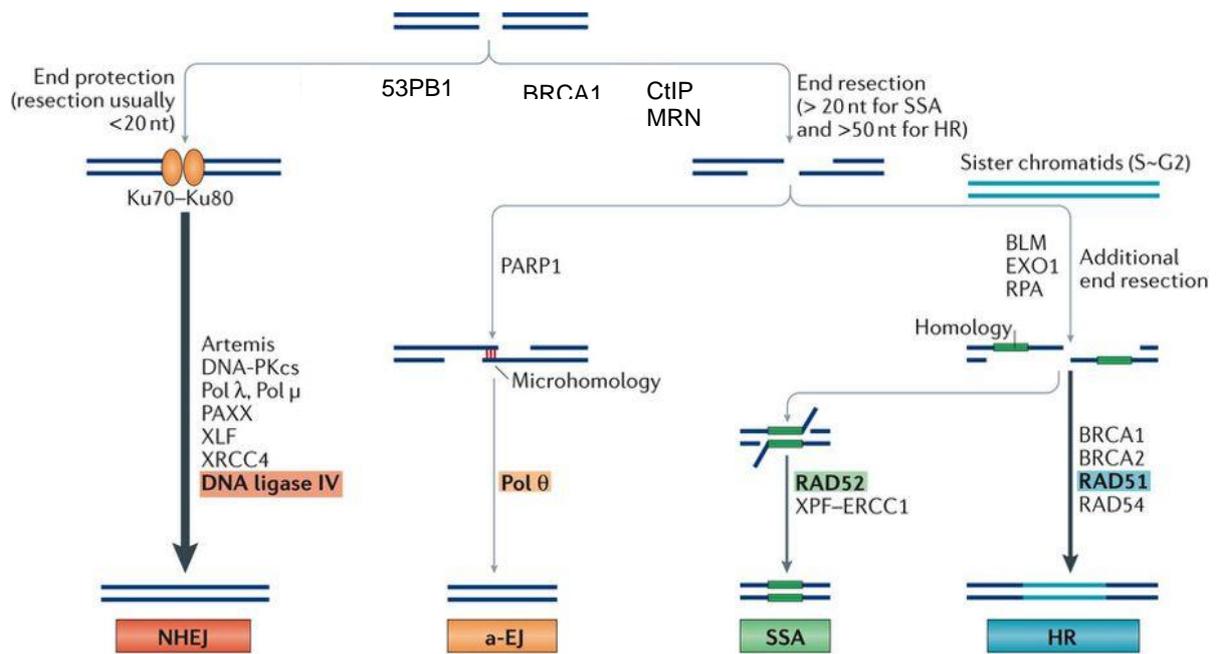


Figure 30. Voies de réparation des cassures double-brin. (modifiée d'après Chang et al., Nat. Rev. Mol. Cell. Biol., 2017)

Représentation schématique des différentes voies de réparation, de leur organisation et des acteurs principaux impliqués. Après la reconnaissance de la CDB, le premier nœud de régulation implique la compétition entre 53BP1 et BRCA1. La prise en charge de la CDB par 53BP1 favorise la cNHEJ, voie indépendante de l'homologie et de la résection. La prise en charge de la CDB par BRCA1 active les voies de l'HDR. Les acteurs caractéristiques de chacune des voies sont surlignés.

- Les voies de réparation dirigées par homologie (HDR pour *homology-directed repair*). Trois sont identifiables : i) la recombinaison homologue (RH), ii) l'*alternative end-joining* (Alt-EJ) et iii) le *single-strand annealing* (SSA).

Cette classification selon la dépendance à l'homologie coïncide avec l'importance de la résection des brins, mécanisme clé pour faire apparaître les régions d'homologies. La cNHEJ, largement indépendante de l'homologie, ne nécessite pas de résection tandis que les voies de l'HDR nécessitent une résection variable, d'extensive pour la RH à très courte pour l'Alt-EJ. C'est sur ce nœud de régulation que 53BP1, anti-résection, et BRCA1, pro-résection, agissent pour favoriser l'une ou l'autre des voies de réparation. Toutefois, ce niveau de régulation sera vu par la suite pour permettre de présenter dans un premier temps les acteurs impliqués dans chacune des voies ; certains étant des cibles de 53BP1 et BRCA1, il est important de détailler leur rôle d'abord.

Classical non-homologous end-joining (cNHEJ)

La cNHEJ est la voie indépendante de l'homologie de séquence. Brièvement, la cNHEJ est une voie de réparation rapide, à haute capacité, qui maintient les deux extrémités de la CDB et les relie pour conserver la structure du chromosome¹⁰⁸. La cNHEJ peut avoir lieu tout au long du cycle cellulaire, exception faite de la mitose. Cette réparation peut se traduire par l'insertion ou la perte de quelques nucléotides (InDels) et peut être non-conservative de l'information génétique (*Figure 31A*).

La cNHEJ se caractérise par trois étapes principales : i) fixation des protéines Ku70-80, ii) modifications des extrémités et iii) ligation (*Figure 31B*).

Fixation des protéines Ku70-80

La première étape est caractérisée par le recrutement d'un hétérodimère Ku70-Ku80 à chacune des extrémités de la cassure double-brin. Les protéines Ku70-Ku80 sont abondantes dans le noyau et possèdent une forte affinité pour les extrémités franches (*blunt ends*) et pour les petites extrémités d'ADN simple-brin. La fixation rapide et avec une forte affinité de cet hétérodimère permet de protéger les extrémités des exonucléases ne faisant pas partie de la cNHEJ¹⁰⁹. Cette structure sert de plate-forme aux protéines nécessaires à la cNHEJ.

Modifications des extrémités

La modification des extrémités est une étape permettant de les rendre compatibles pour la ligation. Certaines CDBs présentant des extrémités franches peuvent être directement liées (*Figure 31B, gauche*).

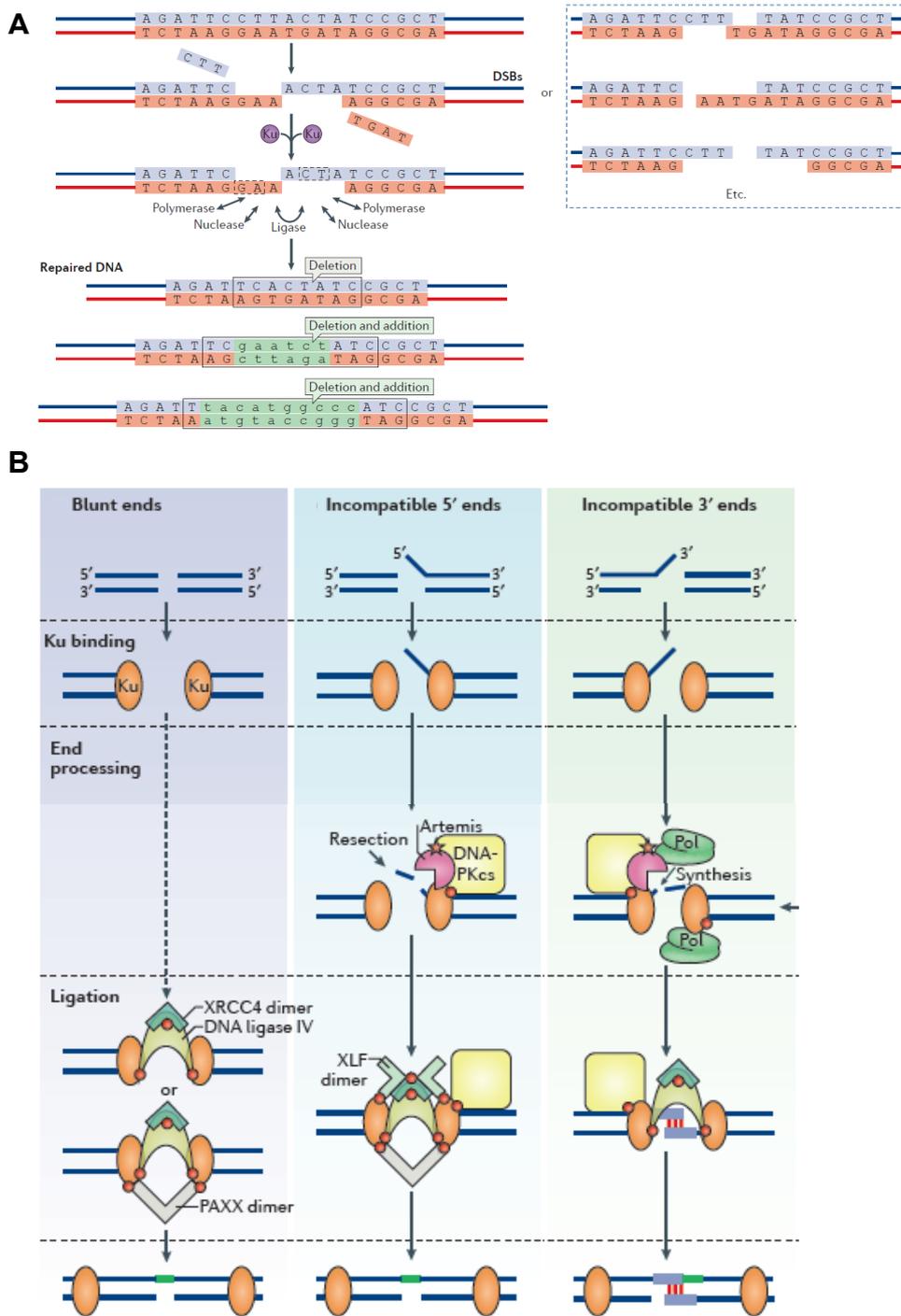


Figure 31. Voie de la cNHEJ. (modifiée d'après Chang et al., Nat. Rev. Mol. Cell. Biol., 2017)

(A) Réparation par la cNHEJ. La ligation de la CDB nécessite l'appariement de quelques nucléotides complémentaires pouvant être présents directement après la CDB (encadrée en pointillé) ou être le fruit de l'activité de polymérase et de nucléases. Des extrémités incompatibles sont montrées en encadré. (B) Correction des extrémités pour les rendre compatibles à la ligation. Cas d'une CDB à extrémités franches (gauche), à extrémités 5' sortantes (centre), à extrémités 3' sortantes (droite).

Toutefois, une grande majorité des CDBs possèdent des extrémités incompatibles empêchant de les relier directement. Les extrémités sont considérées incompatibles quand les segments simple-brin de part et d'autre de la coupure ne permettent aucun appariement (*Figure 31A, encadré*). En effet, bien que considérée comme indépendante de l'homologie, la cNHEJ s'appuie sur des appariements de quelques nucléotides (<4 nt) pour relier les deux extrémités¹⁰⁷. Le rôle de l'endonucléase Artemis et des ADN polymérase μ et λ est de faire apparaître ces régions de micro-homologies pour rendre les extrémités compatibles à la ligation. Une étape de modification est alors nécessaire. Deux cas peuvent être identifiés selon si le segment simple-brin incompatible est 5' sortant ou 3' sortant¹⁰⁷.

Les extrémités non-compatibles 5' sortantes sont prises en charge par le complexe DNA-PK et l'endonucléase Artemis (*Figure 31B, centre*). La DNA-PK est recrutée à l'extrémité non compatible 5' sortante, s'auto-phosphoryle et active Artemis par phosphorylation. Celle-ci clive l'extrémité 5' sortante à la jonction entre simple et double-brin d'ADN. Le résultat est la formation d'une extrémité franche compatible pour une ligation.

La prise en charge des extrémités non-compatibles 3' sortantes est plus complexe et fait intervenir l'ADN polymérase μ et, dans une moindre mesure, le complexe DNA-PK/Artemis (*Figure 31, droite*). L'action du complexe DNA-PKs/Artemis est légèrement différente que précédemment. Ici, la coupure ne fait que raccourcir les extrémités 3' sortantes en générant des extrémités simple-brin de 4 nucléotides. La résolution de l'incompatibilité est surtout due à l'action de la polymérase μ . Cette dernière est capable d'ajouter des nucléotides sur l'extrémité 3' sortante indépendamment d'une matrice modèle. La polymérase μ ajoute quelques nucléotides pour tenter de créer une région de micro-homologie de 1 à 2 nucléotides. De plus, elle est capable d'ajouter un nucléotide sur le brin 3' d'une extrémité franche en utilisant le fragment 3' sortant de l'autre extrémité de la CDB. L'ajout de ce nucléotide permet de créer une région de micro-homologie. Plusieurs cycles itératifs de résection-polymérisation sont parfois nécessaires à la création d'extrémités compatibles.

Il est à noter que la conséquence de l'activité des endonucléases et des polymérase est à l'origine des insertions et délétions (InDels) observées lors de la réparation par la cNHEJ. Une fois les extrémités rendues compatibles par leur appariement au niveau de microhomologie, l'ADN polymérase λ , modèle-dépendante, rajoute les nucléotides manquants avant de procéder à la ligation.

Ligation

Les acteurs majeurs de la ligation sont l'ADN ligase IV et XRCC4. L'ADN ligase IV possède deux domaines de liaison aux protéines Ku70-80 et un domaine de liaison à XRCC4. XRCC4 forme un homodimère qui stimule l'activité de l'ADN ligase IV. Deux autres acteurs, PAXX et XLF, pourraient avoir une activité promotrice de la ligation dans certains cas, toutefois leurs rôles ne sont pas encore clairement définis.

Un mécanisme dynamique spatialement

La cNHEJ est un phénomène hautement dynamique (*Figure 32*). Dans un premier temps, un complexe large est formé à l'aide des protéines Ku70-Ku80 et DNA-PKs permettant de maintenir les deux extrémités d'ADN ensemble. Ensuite, les fragments d'ADN sont rapprochés (par l'action de la DNA-PKs, de XLF, et de XRCC4 associé à LIG4) et alignés pour permettre des cycles de modification des extrémités et des tentatives de ligation. Ce mécanisme est hautement dynamique, ainsi le passage d'une structure à l'autre est réversible tant que la ligation n'est pas effectuée¹¹⁰. Tous les acteurs modifiant les extrémités possèdent la particularité d'intervenir uniquement lorsque les deux extrémités de l'ADN sont proches et alignées, prêtes à être réunies. Cette condition permet de s'assurer que la ligation des brins sera faite dès l'apparition de jonction compatible.

La recombinaison homologue (RH)

La recombinaison homologue (RH) est la voie canonique de l'HDR. Elle est une voie de réparation fidèle et conservative de l'information génétique. Une particularité notable de la RH est d'être restreinte aux phases S/G2 du cycle cellulaire ; la RH utilisant le chromatide sœur intacte comme modèle de réparation. C'est une voie multi-étapes, plus complexes que la cNHEJ, décrit en plusieurs étapes successives¹⁰⁸ (*Figure 33A*) : résection des extrémités 5', formation des filaments nucléoprotéiques et une étape finale suivant soit le modèle de double-jonction Holliday (*Figure 33B*) ou la *synthesis-dependant strand annealing* (SDSA) (*Figure 33C*).

Résection des extrémités 5'

Le complexe MRN assure la reconnaissance de la CDB et l'enclenchement de la réparation par l'activation d'ATM au niveau de la coupure. Si le contexte cellulaire est favorable à la RH, la résection des brins a lieu. Cette résection se produit en deux étapes : une phase initiatrice par le complexe MRN, suivi d'une extension par des exonucléases.

La résection débute avec l'endonucléase MRE11 qui induit une cassure simple-brin de l'ADN à environ 300 paires de bases de la CDB sur les brins portant les extrémités 5'. Les deux fragments formés sont dégradés par l'action exonucléasique 3'-5' de MRE11, soutenue par la protéine CtIP¹¹¹.

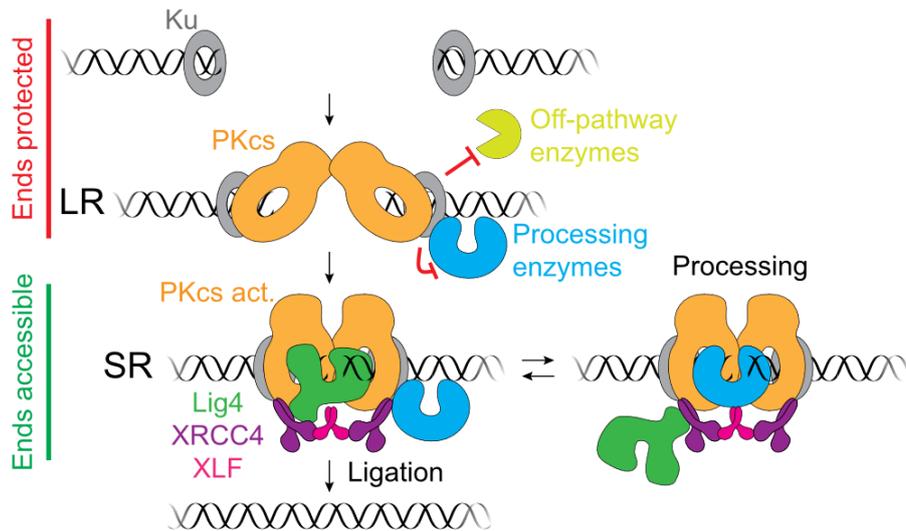


Figure 32. La cNHEJ, un complexe dynamique spatialement. (*Stinson et al., bioRxiv, 2019*)

Différentes conformations au cours de la cNHEJ. La fixation des protéines Ku70-80 (*Ku* en gris) et de la DNA-PKs (*PKcs* en orange) permet de former une grande synapse (LR pour *Long-range*) de laquelle sont exclues les enzymes capables d'interagir avec les extrémités de l'ADN. La taille de la synapse est ensuite réduite (SR pour *Short-range*) et devient accessible aux enzymes de la voie de cNHEJ.

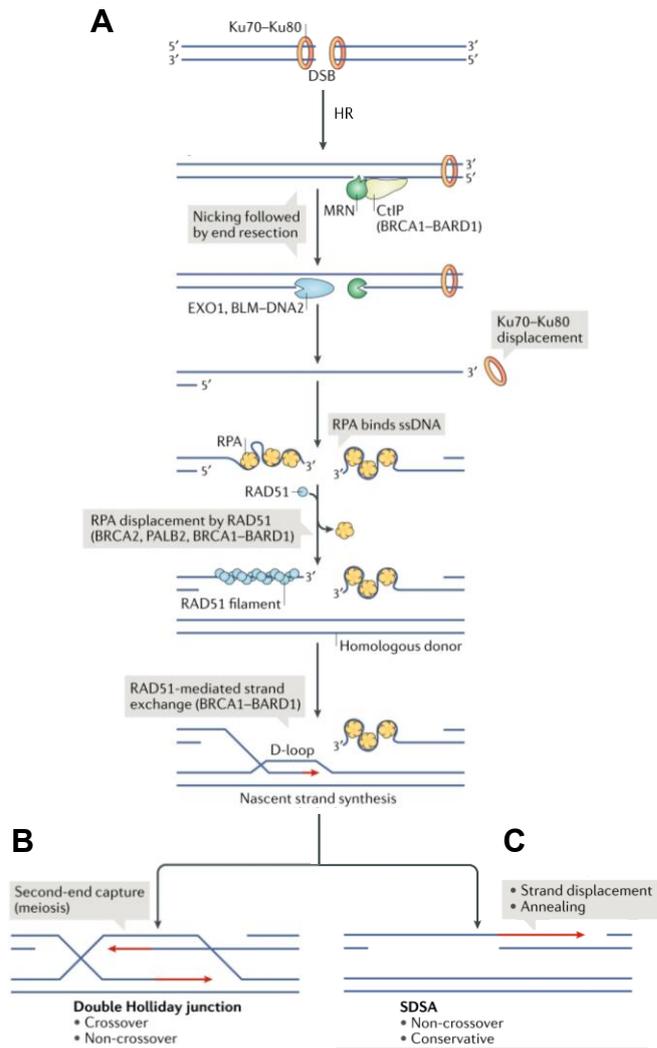


Figure 33. Voie de réparation par RH. (modifiée d'après Scully et al., Nat. Rev. Mol. Cell. Biol, 2019)

(A) Représentation schématique de la réparation par la recombinaison homologue (HR). (B) Modèle de double-jonction Holliday. La RH est terminée par la dissolution ou la résolution des jonctions Holliday. (C) Modèle de SDSA. Le deuxième brin réséqué reste passif.

La dégradation de ces fragments permet de se défaire de l'hétérodimère Ku70-Ku80 au niveau de la CDB. L'étape de résection extensive est assurée par le recrutement de l'exonucléase EXO1, l'hélicase BLM et l'endonucléase DNA2 au point de cassure induit par MRE11. Leur action combinée permet une activité d'exonucléase 5'-3', laissant apparaître deux extrémités 3' simple-brin¹¹².

Formation des filaments nucléoprotéiques

Les extrémités 3' simples brins générées sont rapidement couvertes par plusieurs unités du complexe RPA¹¹³. Ce complexe hétérotrimérique (RPA1, RPA2, RPA3) est abondant dans le noyau et possède une forte affinité pour l'ADN sous forme simple-brin. La fixation de RPA empêche l'interaction du simple-brin avec d'autres éléments susceptibles de s'y hybrider et empêche la formation de structures secondaires.

Les complexes RPA doivent être ensuite remplacés par des protéines RAD51. La formation de filaments nucléoprotéiques RAD51, possédant une activité recombinase, déclenche la recherche de séquences homologues pour poursuivre la RH.

Toutefois, les complexes RPA forment une barrière à l'association de RAD51 à l'extrémité 3' simple-brin. Le chargement des protéines RAD51 est assuré par BRCA2. Cette dernière est au cœur d'un complexe liant l'ADN simple-brin, les monomères RAD51 et le complexe BRCA1-BARD1 par l'intermédiaire de PALB2. L'action de BRCA2 permet le remplacement de RPA par RAD51, formant les filaments nucléoprotéiques.

Recherche d'une séquence homologue

Ces filaments sont des structures dynamiques, cibles de facteurs favorisant ou inhibant leur stabilité. Cette régulation dynamique assure de leur permettre une recherche efficace de séquences homologues, tout en restreignant ses possibilités aux substrats appropriés. Cette recherche est soutenue par le complexe BRCA1-BARD1 en facilitant l'assemblage d'une synapse entre le filament RAD51 et la séquence homologue¹¹⁴.

Un filament RAD51 envahit une molécule d'ADN double-brin pour former une triple hélice d'ADN. Cette structure permet d'effectuer un test de complémentarité entre le brin invasif et la séquence complémentaire du double-brin envahie. Si suffisamment de paires de bases sont appariées, le brin homologue au filament RAD51 est éjecté en formant une *D-loop*. La formation de la *D-loop* est permise par l'hydrolyse de l'ATP porté par RAD51. Cette hydrolyse entraîne un changement conformationnel de RAD51 et la dissociation de RAD51 du brin invasif¹⁰⁸. La première jonction Holliday est formée au carrefour du brin invasif et du brin déplacé.

Une fois la *D-loop* formée, le brin invasif recrute la polymérase δ . Cette dernière allonge le brin invasif en prenant la séquence complémentaire comme modèle de réparation. Ce phénomène est également dynamique et peut faire appel à plusieurs cycles d'invasion-polymérisation-libération. Plusieurs protéines, dont FANCM et BLM, limitent la taille de la séquence pouvant être polymérisée en déstructurant la *D-loop*¹¹⁵. A partir de ce point, deux options sont possibles. Selon le contexte cellulaire, la RH est terminée soit par la formation de double-jonction Holliday soit par la voie du SDSA.

Modèle de double-jonction Holliday

Ici, le deuxième brin réséqué, et couvert de RAD51, recrute le brin déplacé de la *D-loop* et forme une deuxième synapse symétrique à la première. Une polymérase δ est également recrutée et permet la réparation du deuxième brin. La formation de la synapse est à l'origine de la deuxième jonction de Holliday. Cette voie de RH est terminée soit par leur dissolution par l'action du complexe BLM-Topoisomérase 3-RMI1-RMI2¹⁰⁸ soit, le cas échéant, par la résolution des jonctions Holliday par des résolvas. Cette résolution peut être symétrique ou asymétrique et mener à des *crossing-over*¹¹⁶. Cette voie de RH est notamment dominante lors de la recombinaison méiotique.

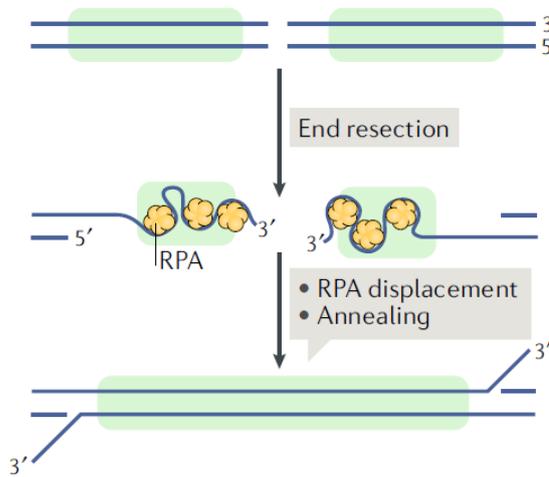
Synthesis-Dependant Strand-Annealing

La voie *synthesis-dependant strand annealing* (*SDSA*) est majoritaire dans les cellules mitotiques. Ici, le deuxième brin réséqué reste passif. Après polymérisation du brin invasif par la polymérase δ , le brin déplacé au sein de la *D-loop* se ré-apparie avec son brin complémentaire permettant la libération progressive du brin néo-synthétisé. Ce dernier rejoint le second brin réséqué par complémentarité de base. Cette hybridation est facilitée par l'action de RECQ5 qui fragilise l'interaction entre RAD51 et le filament d'ADN¹¹⁷. Finalement, la SDSA est terminée par la synthèse du brin manquant et la ligation des extrémités reformées. La SDSA préserve l'intégrité génomique par l'absence de *crossing-over*.

Single-strand annealing (SSA)

Le SSA est une voie alternative d'HDR, RAD51-indépendante, qui intervient lorsqu'une CDB a lieu entre des séquences répétées (*Figure 34A*). Comme la RH, le SSA nécessite la résection bidirectionnelle des brins, initiée par le complexe MRN et complétée par EXO1 et BLM-DNA2. Si des séquences répétées sont présentes de part et d'autre de la CDBs, la résection fait apparaître des séquences complémentaires. Ces dernières s'hybrident et les extrémités 3' sortantes sont dégradées.

A Single-strand annealing



B Alternative end joining

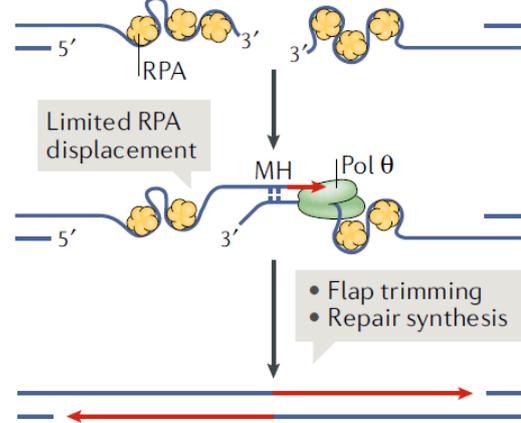


Figure 34. Voies de réparation par SSA et Alt-EJ. (modifiée d'après Scully et al., *Nat. Rev. Mol. Cell. Biol.*, 2019)

(A) Réparation par la *single-strand annealing* (SSA). La résection fait apparaître des séquences répétées homologues (régions vertes) qui s'hybrident. Les complexes RPA, empêchant l'hybridation, doivent être déplacés au préalable. Le SSA convertit deux séquences répétées en une, entraînant une perte d'information génétique.

(B) Réparation par l'*alternative end-joining* (Alt-EJ). Cette voie se traduit par l'hybridation des extrémités 3' sortantes au niveau de régions de micro-homologies (MH), non-couvertes par des complexes RPA. L'ADN polymérase θ possède une activité hélicase lui permettant d'enlever les complexes RPA et une activité polymérase pour synthétiser les fragments d'ADN manquants.

Bennardo et al. ont identifié RAD52 et le complexe ERCC1-XPF comme acteur clé de la SSA, permettant respectivement l'hybridation des régions complémentaires et la suppression des extrémités 3' sortantes excédentaires. Les polymérases impliquées dans la fin du processus ne sont pas identifiées, toutefois la ligation est assurée par l'ADN ligase 1¹¹⁸. Le SSA entraîne la perte de séquences situées entre les régions homologues et la rend non-conservative de l'information génétique. Plusieurs études ont mis en lumière une compétition entre RH et SSA. En effet, Stark et collègues ont montré qu'une invalidation de RAD51 ou BRCA2, pro-RH, entraînait une forte augmentation de l'activité du SSA¹¹⁹. L'intérêt physiologique de la SSA serait d'intervenir en cas de RH défectueuse ou de répondre aux CDB survenant lors des phases S/G2, favorables à la RH et donc à la résection des brins, mais avant la formation du chromatide sœur¹²⁰.

Alternative end-joining (Alt-EJ)

L'*alternative end-joining* est parfois aussi appelée *microhomology-mediated end joining (MMEJ)*. Toutefois, ce nom prête à confusion car la cNHEJ utilise également des régions de micro-homologies de quelques nucléotides. La cNHEJ et l'Alt-EJ s'appuient tous deux sur des séquences de micro-homologie. Toutefois, leur taille varie fortement avec des micro-homologies de 1 à 4 nt pour la cNHEJ contre 4 à 20 nt pour l'Alt-EJ.

L'Alt-EJ est indépendante de RAD51 (RH) et de RAD52 (SSA) et de la machinerie de la cNHEJ. Elle est basée sur l'hybridation de quelques nucléotides complémentaires exposés par la résection des brins (*Figure 34B*). Cette résection peut se confiner à l'étape assurée par le complexe MRN. Celle-ci est suffisante pour faire apparaître des régions de micro-homologie permettant l'hybridation des extrémités 3' simple-brin. L'acteur moléculaire clé est ici la polymérase θ . Dans les eucaryotes supérieurs, l'action de la Pol θ est permise par l'hybridation de régions homologues de 2 à 6 nt et mène à l'insertion de nouveaux nucléotides¹¹⁸. De plus, la polymérase θ possède une activité hélicase, ATP-dépendante, capable de défaire les liaisons entre RPA et l'extrémité simple-brin facilitant l'hybridation entre les régions complémentaires¹²¹. Les extrémités 3' sortantes excédentaires sont supprimées par le complexe ERCC1-XPF comme lors du SSA. Cette voie entraîne des délétions et est donc également non-conservative de l'information.

La part de l'Alt-EJ dans la prise en charge des CDBs est peu caractérisée. Plusieurs études ont montré que son activité est exacerbée lorsque la cNHEJ ou la RH sont défectueuses, bien qu'elle puisse être active dans un état plus physiologique¹²².

e. Quelle voie de réparation pour quelle matrice d'ADN exogène ?

Une partie des stratégies d'édition génique consistent à introduire des mutations d'intérêt *via* l'apport de matrices d'ADN. Toutefois, plusieurs types de matrices peuvent être utilisés et

n'empruntent pas les mêmes voies pour faire du *knock-in*. Comprendre comment les matrices portant les mutations interagissent avec l'ADN et par quelle voie de réparation, permet d'identifier des leviers pour améliorer l'efficacité de *knock-in*.

Matrice plasmidique

Les premières matrices utilisées ont été les plasmides, fragments d'ADN double-brin et circulaires. Ils possèdent des grandes régions homologues à la séquence cible (plusieurs centaines de paires de bases). Ces régions encadrent la cassette portant les mutations à insérer. Grâce à ses bras d'homologie, les matrices plasmidiques sont capables d'utiliser la RH avec résolution des jonctions Holliday pour s'incorporer dans le génome. Cependant, l'efficacité de *knock-in* par RH est faible en raison de la forte activité de la cNHEJ, particulièrement dans les cellules primaires. Toutefois, pour les stratégies le permettant, l'utilisation de cassette de sélection peut offrir la possibilité de sélectionner les cellules correctement éditées.

Pour le système CRISPR-Cas9, deux approches ont été proposées pour augmenter le taux de *knock-in* en passant par des voies autres que la RH. Une première méthode a été développée pour faire du *knock-in via* l'Alt-EJ¹²³. L'astuce du système CRIS-PITCh v2 (CRISPR - *Precise Integration into Target Chromosome*) réside principalement dans le design du plasmide donneur (*Figure 35*). Celui-ci est composé de la cassette d'intérêt à insérer, flanqué de part et d'autre part de courtes séquences d'homologies de 20 nucléotides. Cet ensemble est encadré en 5' et 3' de séquences ciblées par le même ARNg que celui visant le gène d'intérêt. De cette façon, après transfection du plasmide donneur et du plasmide permettant l'expression de la Cas9 et de l'ARNg, la RNP formée coupe à la fois le gène d'intérêt et libère un fragment d'ADN double-brin du plasmide. Ce fragment porte de courtes séquences d'homologies permettant une réparation par l'Alt-EJ. Ce procédé permet d'améliorer jusqu'à 2 fois l'efficacité de *knock-in* dans les cellules primaires humaines en comparaison de ce qui est obtenu avec une simple matrice plasmidique.

De la même manière, un procédé dépendant de la cNHEJ a été mis au point. Le système HITI (*Homology-Independent Targeted Integration*) repose sur l'induction d'une coupure dans un donneur plasmidique et dans le gène d'intérêt avec le même ARN guide¹²⁴ (*Figure 36*). Une fois linéarisée, la matrice donneuse peut s'intégrer au niveau du site génomique de coupure. En l'absence de séquences d'homologie dirigeant le sens d'intégration, celle-ci peut se faire dans un « sens » ou dans l'autre. Toutefois, si l'intégration se fait dans le « mauvais sens », deux séquences de coupures sont reformées et reconnues par le même guide. Ces nouvelles séquences cibles peuvent être recoupées jusqu'à intégrer la cassette dans le sens correct.

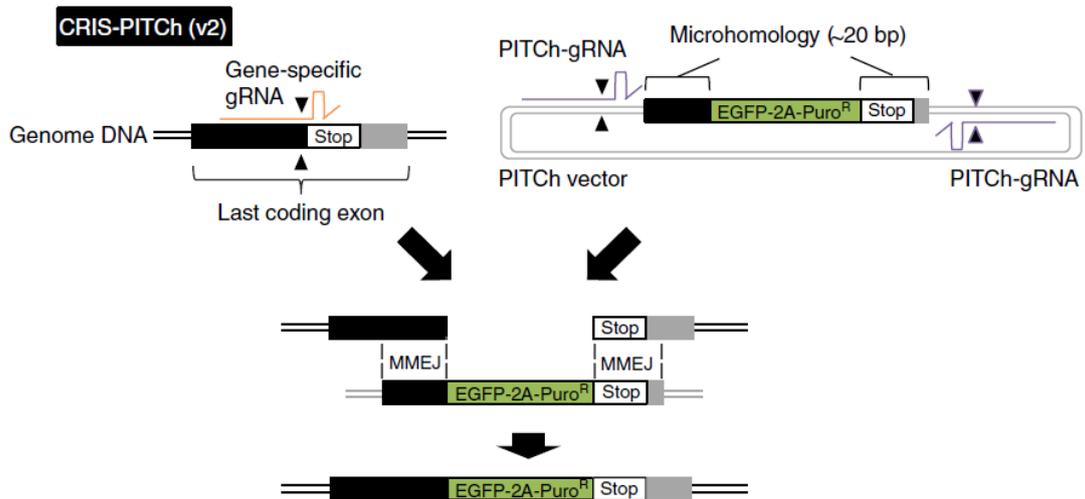


Figure 35. Système CRIS-PITCh v2. (modifiée d'après Sakuma et al., Nat. Protoc., 2016)

Vue d'ensemble du fonctionnement du système CRISPR - Precise Integration into Target Chromosome.

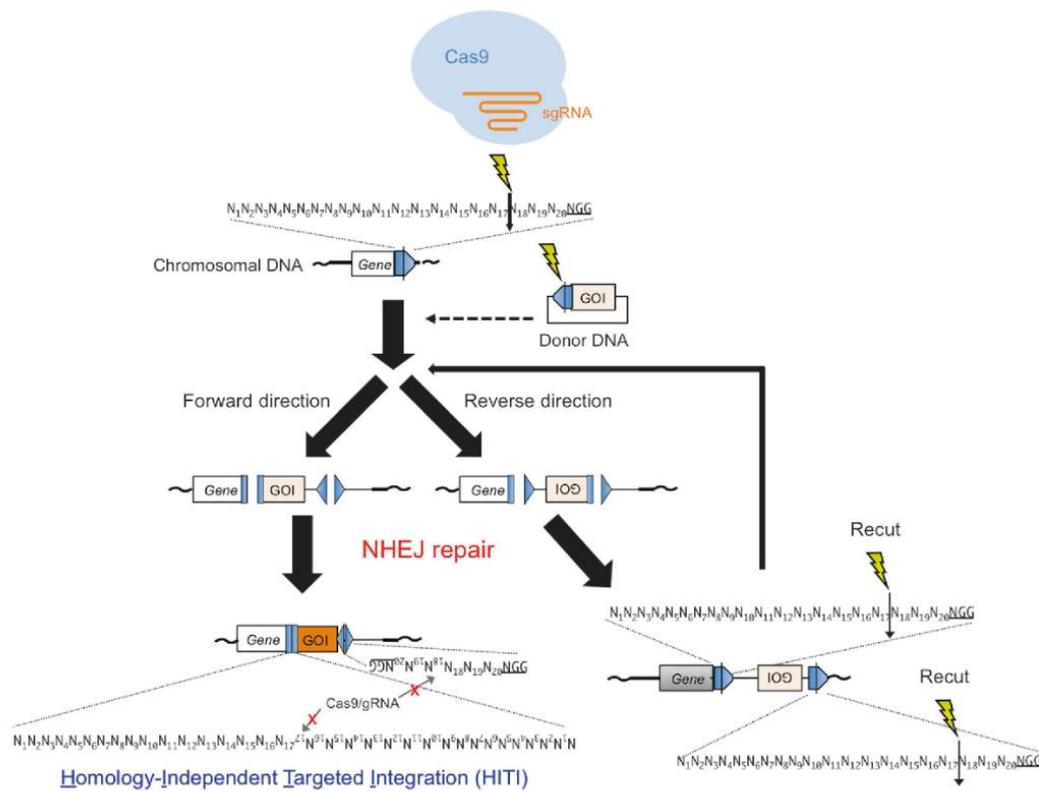


Figure 36. Système HITI. (modifiée d'après Suzuki et Izpisua Belmonte, J. Hum. Genet., 2018)

Vue d'ensemble du fonctionnement du système Homology-Independent Targeted Integration (HITI).

Une des limites de ce système est que si des InDels se forment au niveau des séquences reconnus par l'ARNg, ces derniers empêchent de recouper et bloquent la correction du sens d'intégration. Toutefois, au cours de leur étude, Suzuki et al. ont montré que cette situation est minoritaire. L'atout principal du système HITI est de permettre une édition génique sans avoir recours à une voie de réparation nécessitant une résection des brins. Cette particularité permet de modifier des cellules post-mitotiques (*e.g.* neurones) *in vitro* et également *in vivo* à l'aide d'AAV.

Fragments d'ADN double-brin

Les fragments double-brin d'ADN (par exemple, produits de PCR) constituent une deuxième forme de matrice utilisée. Les systèmes PITCh et HITI sont deux exemples de procédés utilisant un donneur plasmidique réduit sous forme de fragment double-brin avant son intégration. Les fragments double-brin s'intègrent principalement en utilisant la RH avec résolution des jonctions Holliday, et une participation mineure de la voie menant à la dissolution des jonctions Holliday et du *synthesis-dependant strand annealing (SDSA)*^{125,126}, également appelé *Short Fragment Homologous Recombination (SHFR)* dans ce cas particulier¹²⁷.

Oligo-désoxynucléotides

Les oligo-désoxynucléotides simple-brin (ssODNs pour *single-strand oligodesoxynucleotides*) se sont rapidement imposés comme alliés de l'édition génique. Les ssODNs sont peu coûteux à produire, facile à vectoriser et ont montré de meilleurs résultats que leur équivalent double-brin¹²⁸. Le *knock-in* par des ssODNs est possible par deux mécanismes d'HDR : la SDSA et la *single-strand DNA incorporation (ssDI)* (*Figure 37A et B*).

Pour le ssDI, le ssODN profite de l'instabilité au niveau de la cassure pour s'hybrider avec le brin complémentaire en formant un pont enjambant la cassure (*bridge model*). Le ssODN est ensuite lié aux extrémités du brin homologue. La résection des brins inhibe le ssDI en rendant plus complexe le pontage par le ssODN des deux extrémités de la CDB. A l'inverse, la résection favorise l'apparition de régions simple-brin complémentaires au ssODN permettant d'enclencher la voie du SDSA. Lors de l'induction d'une CDB, les profils d'intégration des mutations ont montré une préférence pour la voie SDSA plutôt que ssDI¹²⁶ (*Figure 37C*).

f. Régulation des voies de réparation endogènes

Les paragraphes ci-dessus montrent le lien étroit entre le type de matrice et les voies de réparation. D'une manière générale, le *knock-in* utilise des voies de l'HDR, exception faite du système HITI. Or, les voies de l'HDR sont dépendantes de la résection des brins suite à la reconnaissance de la CDB. L'une des possibilités pour améliorer l'efficacité de *knock-in* est de réussir à forcer la cellule à engager la résection des brins.

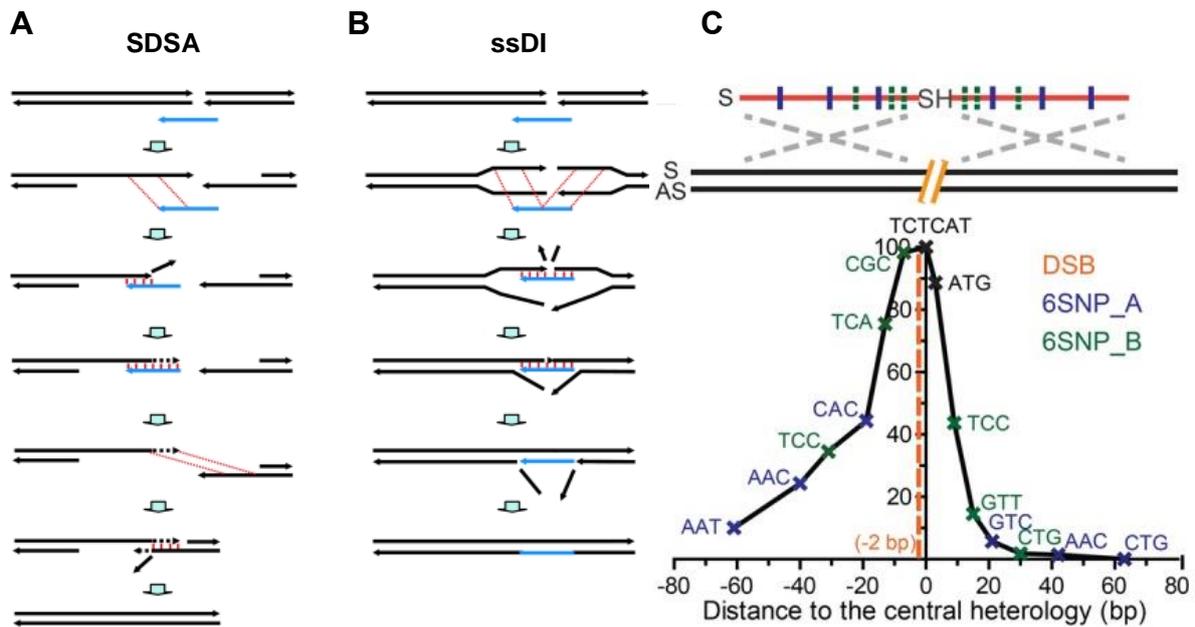


Figure 37. Voies de réparation des CDBs par des ssODNs.

(A-B) Modèles de réparation par des ssODNs soit par la voie du SDSA (A), soit par le voie du *single-strand DNA Incorporation* (ssDI) (B). Le ssODN est représenté en bleu. (Storici et al., *Mol. Cell. Biol.*, 2006)

(C) Fréquence d'insertions de mutations en fonction de leur position par rapport au site de la coupure double-brin. Deux ssODNs portant chacun 6 SNP différents ont été testés (6SNP_A et 6SNP_B). La courbe est asymétrique et tend à différer d'une distribution normale. (modifiée d'après Kan et al., *Genome Res.*, 2017)

Pour y parvenir, il est important de comprendre quels sont les facteurs qui régissent cette décision ; reséquer ou ne pas reséquer, telle est la question. Comme mentionné précédemment, ce choix binaire est régi par la compétition entre deux acteurs clés pour « s'approprier » le site de coupure : 53BP1, anti-résection et favorisant la cNHEJ et BRCA1, pro-résection, stimulant l'HDR.

Les paragraphes suivant montreront le rôle respectif des complexes 53BP1 et BRCA1 et comment le cycle cellulaire permet la transition de l'un à l'autre (*Figure 38*).

p53-Binding protein (53BP1) contre la résection, pour la cNHEJ

Le rôle principal de 53BP1 est d'empêcher la résection de brins pour permettre la cNHEJ. 53BP1 est une grande protéine qui ne possède pas d'activité enzymatique, mais orchestre les voies de réparation par ses nombreux sites de liaisons à d'autres protéines régulatrices¹²⁹. L'un de ses partenaires est la protéine RIF1. Celle-ci est recrutée après la phosphorylation de 53BP1 par ATM. La déplétion de 53BP1¹³⁰ ou de RIF1¹³¹ conduit à une sensibilité aux rayons ionisants due à une cNHEJ défectueuse *via* le blocage de la résection.

Plusieurs études ont permis d'expliquer le mécanisme sous-jacent expliquant cette relation. Un criblage génétique a identifié la protéine MAD2L2 (aussi appelé REV7) comme acteur en aval de RIF1^{132,133}. Récemment, il a été montré que l'effecteur final de 53BP1-RIF1 est le complexe Shieldin, composé des protéines SHLD1, 2 et 3 et de MAD2L2¹³⁴ (*Figure 39A*). Le complexe Shieldin se lie à l'ADN simple-brin et empêche sa résection extensive. La présence du complexe Shieldin empêche également la fixation des protéines nécessaires à la RH, notamment RPA et RAD51. Enfin, le complexe CST (Ctc1, Stn1, Ten1) et la polymérase α sont des effecteurs du complexe 53BP1-RIF1-Shieldin¹³⁵ (*Figure 39B*). Les complexes CST, ressemblant aux complexes RPA, se fixent aux extrémités sortantes et permettent l'action de la polymérase α pour inhiber la résection des brins. Comme pour la 53BP1 et RIF1, la déplétion de CST entraîne une augmentation de la résection des brins.

D'autres acteurs favorisant la cNHEJ

D'autres mécanismes favorisant la cNHEJ sont indépendants de 53BP1. Bien que ces mécanismes puissent avoir un rôle mineur par rapport à 53BP1, il serait intéressant d'étudier l'effet synergique de l'inhibition simultanée de plusieurs voies favorisant la cNHEJ. La protéine WWOX fait partie des acteurs qui soutiennent la cNHEJ. En éteignant son expression, Schrock et al. ont montré que WWOX inhibe la résection des brins en bloquant la liaison entre RAD50 du complexe MRN et BRCA1¹³⁶. L'hélicase B (HELB) est recrutée au CDB par les protéines RPA et régule la résection des brins en effectuant un rétrocontrôle négatif des protéines EXO1 et BLM-DNA2¹³⁷.

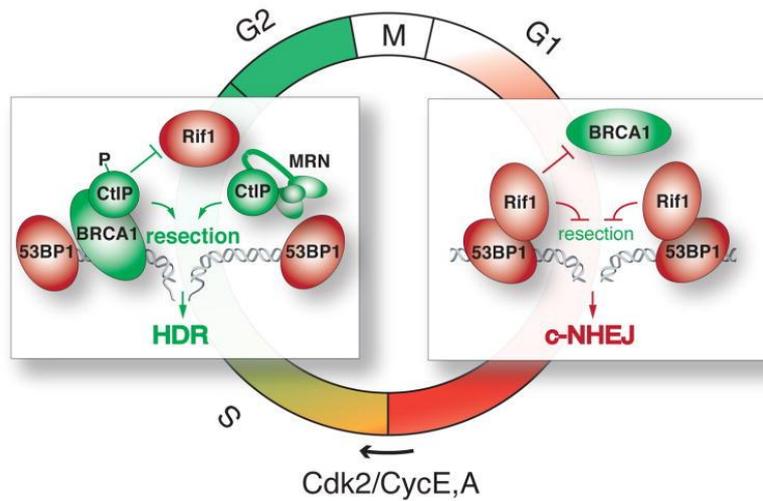


Figure 38. Cycle cellulaire et compétition entre 53BP1 et BRCA1. (modifiée d'après Zimmermann et de Lange, Trends Cell. Biol., 2014)

Représentation graphique de l'exclusion mutuelle de BRCA1 et 53BP1 au niveau de la CDB. En phase G1, le complexe 53BP1-RIF1 empêche l'accumulation de BRCA1. En phase S/G2, le contexte cellulaire est favorable à BRCA1 agissant avec le complexe MRN et CtIP pour permettre la résection des brins et promouvoir l'HDR.

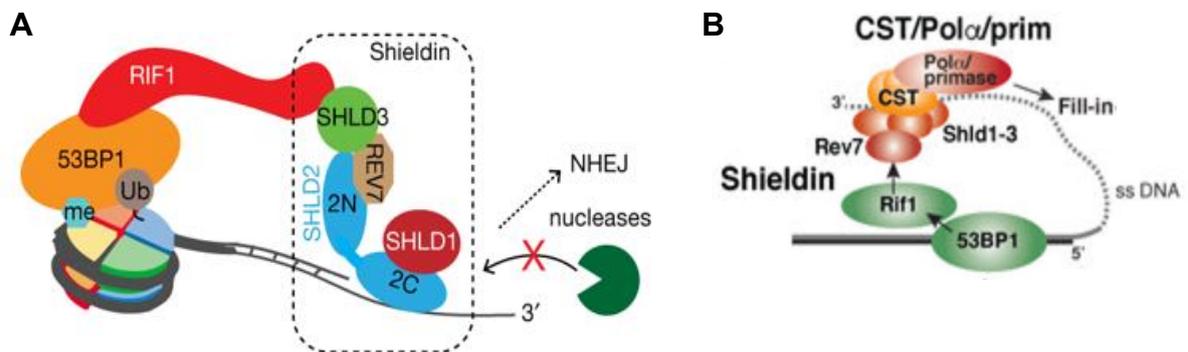


Figure 39. Fonctionnement de 53BP1 pour inhiber la résection des brins. (modifiée d'après Noordermeer et al., Nature, 2018 et Mirman et al., Nature, 2018)

(A-B) Représentation graphique des acteurs situés en aval de 53BP1. (A) RIF1 interagit directement avec 53BP1. RIF1 est en interaction avec le complexe Shieldin, comprenant les protéines SHLD1, SHLD2 et SHLD3 et REV7 (ou MAD2L2). Le complexe Shieldin interagit avec l'ADN simple-brin au niveau d'une CDB et bloque la résection par sa présence. (B) Le complexe CST et la polymérase α sont en lien avec le complexe Shieldin. L'activité polymérase permet de réduire la taille du fragment simple-brin (fill-in).

Phase S/G2 : l'action des CDKs, des cyclines et de BRCA1 en faveur de la résection et de l'HDR

L'avancée du cycle cellulaire dans la phase S/G2 marque un fort changement du contexte cellulaire, favorable à la résection. Ce changement est assuré par l'activité des kinases dépendantes des cyclines (CDK pour *cyclin-dependant kinase*) (*Figure 40A et B*). Ces modulations d'activité vont permettre à la protéine BRCA1 de passer d'un statut réprimé en phase G1 à pro-actif en phase S/G2 pour permettre la résection des brins¹³⁷.

Rôle de CDK2

La CDK2 interagit fortement avec le complexe MRN. CDK2 active NBS1 et s'associe avec la protéine MRE11. La liaison entre CDK2 et MRE11 permet la phosphorylation de CtIP sur la sérine 327¹³⁸. Une fois phosphorylée, CtIP peut interagir avec BRCA1. Le complexe CtIP-BRCA1 stimule fortement la résection des brins¹³⁹.

L'hélicase B est également phosphorylée par la CDK2, ce qui entraîne son export du compartiment nucléaire et une levée de l'inhibition de BLM-DNA2 et EXO1 ; cette dernière étant directement activée après phosphorylation par la CDK2¹⁴⁰.

En parallèle, la CDK2 phosphoryle l'ubiquitine ligase UHRF1. Une fois phosphorylée, elle est recrutée par BRCA1 aux CDBs et poly-ubiquitine RIF1 le dissociant de 53BP1¹⁴¹ (*Figure 40C*). La dissociation de 53BP1 de RIF1 augmente la résection des brins par une levée d'inhibition.

Rôle des cyclines

En parallèle des CDK, les cyclines jouent également un rôle dans la balance en faveur de la RH. La cycline D1 se lie à RAD51 et facilite son recrutement aux sites de coupures¹⁴². La cycline A2 quant à elle stabilise l'ARNm de MRE11, du complexe MRN, favorisant sa traduction et sa concentration nucléaire¹⁴³.

Rôle de BRCA1-BARD1

Le complexe BRCA1-BARD1 possède une double action en faveur de l'HDR. Le complexe BRCA1-BARD1 possède une activité ubiquitine ligase des histones H2A. Cette marque chromatinienne déclenche le recrutement du modulateur de chromatine SMARCAD1 au niveau de nucléosomes, eux même reconnus par 53BP1 (*Figure 40D*). SMARCAD1 permet d'éloigner ou de supprimer les histones fixées par 53BP1. Ce phénomène participe à la levée de l'inhibition de la résection. Le second rôle de BRCA1-BARD1 est d'activer la fonction recombinaise de RAD51, facilitant la formation des filaments nucléoprotéiques nécessaire à la recherche de séquence homologue¹⁴⁴ (*Figure 40E*).

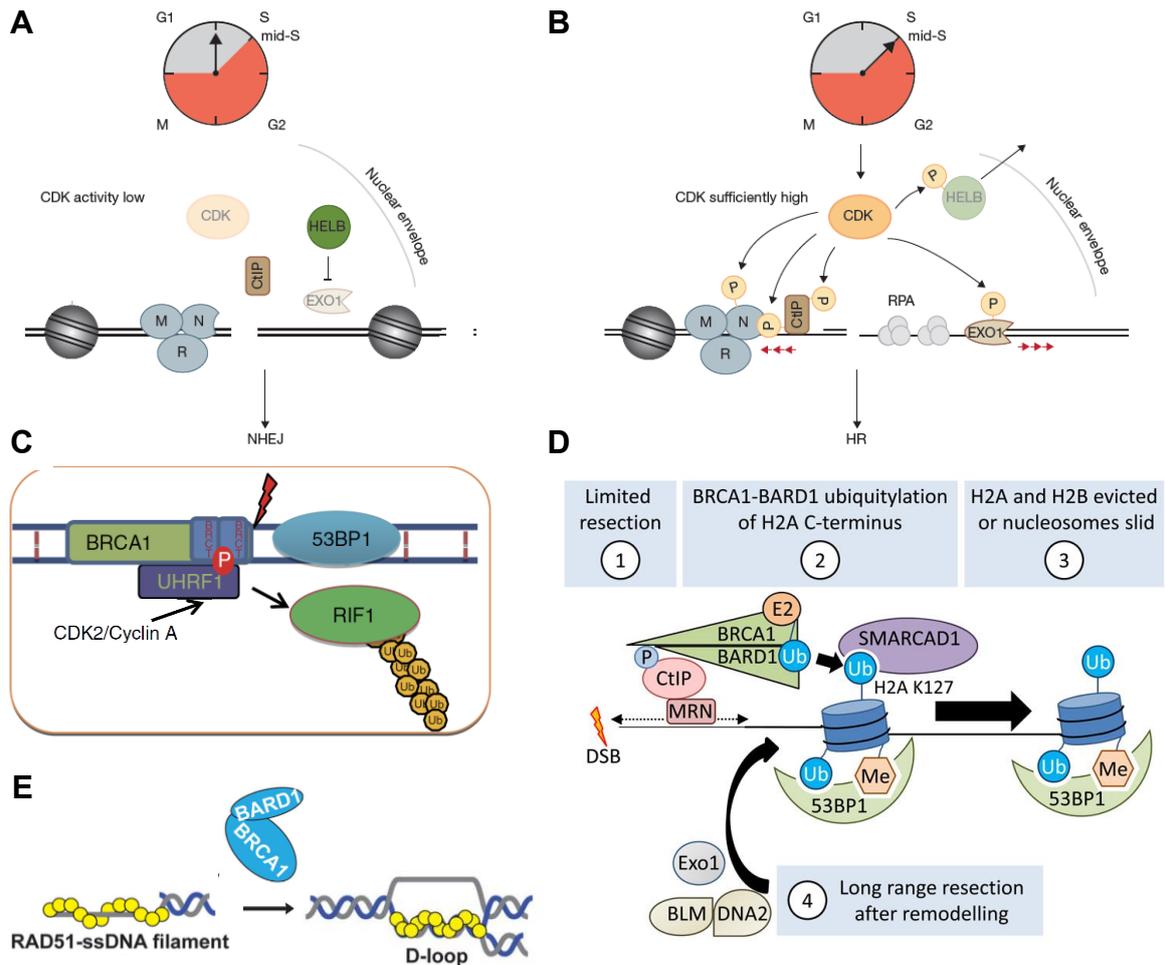


Figure 40. Rôle des CDKs et de BRCA1 pour activer la résection des brins

(A) En phase G1 et début S2, le contexte cellulaire est favorable à la cNHEJ. (B) En phase fin S/G2, les CDKs activent la résection des brins en phosphorylant différents acteurs. (*Hustedt et Durocher, Nat. Cell. Biol., 2016*)

(C) Action de la CDK2 sur UHRF1. (*modifiée d'après Zhang et al., Nat. Commun., 2016*)

(D-E) Rôle du complexe BRCA1-BARD1 *via* SMARCD1 (D) et *via* RAD51 (E). L'action de SMARCD1 permet également de laisser la place pour la phase extensive de résection assurée par EXO1 et BLM-DNA2. ((D) *Densham et al., Nat. Struct. Mol. Biol., 2016* et (E) *Zhao et al., Nature, 2017*)

Tous ces changements de régulation permettent de passer d'un environnement pro-cNHEJ à un contexte en faveur de l'HDR. La dualité entre 53BP1 et BRCA1 montrent que la cNHEJ et l'HDR sont intrinsèquement interdépendants. En vue de *knock-in*, l'activation des voies permettant l'intégration de l'information apportée par la matrice exogène par HDR et l'inhibition des voies non-conservatives pour éviter la formation d'InDels est un enjeu majeur.

g. La balance HDR / InDels.

Les paragraphes ci-dessus montrent la compétition entre les voies de réparation. En se basant sur la connaissance de ces voies, plusieurs stratégies ont été proposées pour tenter d'améliorer la balance HDR/Indels : pharmacologique avec modulation des voies de réparation directement, ou *via* le cycle cellulaire, l'amélioration des matrices donneuses, ou la création de Cas9 de fusion.

Approches pharmacologiques

Plusieurs molécules ont été mises en lumière par leur capacité à augmenter le ratio HDR/Indels.

Inhibiteur de la cNHEJ

- Inhibition de l'ADN ligase IV par le SCR-7¹⁴⁵ ou par la co-expression des protéines E1B55K et E4orf6 de l'adénovirus 4¹⁴⁶
- Inhibition de la DNA-PK par le NU7441¹⁴⁷
- Inhibition de 53BP1 par l'i53¹⁴⁸ ou par l'e18¹⁴⁹.

Activer l'HDR en favorisant la RH

A l'inverse, l'utilisation du RS-1 (pour *RAD51 Stimulator 1*) permet de favoriser l'HDR¹⁵⁰ en facilitant le chargement des protéines RAD51 sur les brins réséqués lors des étapes de RH¹⁵¹.

Activer l'HDR par l'inhibition de p53 ?

Deux études publiées le même jour dans Nature Medicine ont montré l'implication de p53 dans la balance HDR/InDels^{152,153}. Les deux travaux démontrent une inhibition p53-dépendante de la fréquence d'HDR dans des cellules pluripotentes souches. L'inhibition transitoire de p53 par une co-transfection de la protéine MDM2 accroît fortement la fréquence d'HDR.

Si les deux études soulignent la difficulté d'inhiber p53 au vu des risques de réarrangements chromosomiques et d'effets délétères pour la cellule, cela suggère également que les cellules possédant un statut p53 faible sont les plus susceptibles à être modifiées par édition génique. Cette observation peut laisser penser qu'une attention particulière devra être portée au statut p53 des cellules éditées, en particulier pour une perspective clinique.

De nouveaux acteurs en dehors des voies de réparation pour améliorer l'HDR ?

A l'aide d'un screening pharmacologique sur des cellules embryonnaires murines, Yu et collègues ont identifié deux molécules permettant également d'augmenter l'HDR, la Brefeldin A et le L755507¹⁵⁴. Etant respectivement connu comme un agoniste des récepteurs β -3 adrénergique et comme un inhibiteur du transport protéique du réticulum endoplasmique vers l'appareil de Golgi, leur mécanisme d'action en tant qu'activateur d'HDR reste méconnu.

Dans l'ensemble, les modulateurs pharmacologiques permettent une amélioration variable du ratio, poussant certains auteurs à utiliser des « mix » de drogues favorisant la balance HDR/InDels¹⁵⁵. Cependant, les problèmes de doses, de temps d'exposition et la variété des types cellulaires et des loci compliquent leur utilisation.

Cycle cellulaire

Un des moyens d'améliorer la balance HDR/InDels est de positionner les cellules dans une phase du cycle cellulaire favorable à la RH plutôt qu'à la cNHEJ. Lin et al. ont montré qu'une synchronisation des cellules par un traitement préalable à la transfection permettait d'augmenter l'HDR¹⁵⁶. Cette effet a été observé avec l'aphidicoline, bloquant la transition G1/S, et le nocodazole, bloquant l'entrée en mitose. Cependant, l'amélioration du ratio demeure modeste et dépendante du type cellulaire.

Matrices de réparation

Le travail conduit par Richardson et collègues pour améliorer les matrices ssODNs mérite d'être souligné¹²⁸. Leur équipe a étudié l'interaction entre la Cas9 et l'ADN cible et utilisé les résultats pour choisir un oligonucléotide le plus adéquat à faire de l'HDR. Il a été montré qu'après l'induction de la CDB, la Cas9 se dissocie d'abord de l'extrémité 3' du brin non-ciblé (homologue à l'ARNg). Sur la base de cette observation, un oligonucléotide a été designé, en étant complémentaire du brin non-ciblé et asymétrique par rapport au point de cassure. Cette conformation permet de multiplier jusqu'à 6 et 2 fois le taux d'HDR comparé respectivement à une matrice double-brin et un ssODN centré et complémentaire du brin-ciblé.

Formation de Cas9 de fusion

Plusieurs Cas9 de fusion ont été formées afin de parer aux limites des approches pharmacologiques visant les voies de réparation ou le cycle cellulaire. En effet, les protéines de fusion possèdent l'avantage d'avoir une action plus spécifique au niveau de la CDB, et non sur l'ensemble du métabolisme cellulaire.

La fusion de la Cas9 avec la Geminin (Cas9-Gem) entraîne sa dégradation lors de la fin de la mitose et la phase G1¹⁵⁷. De cette façon, la Cas9-Gem n'est active qu'en phase S/G2, favorable à l'HDR. Cette amélioration a permis de diminuer le pourcentage d'InDels tout en conservant un taux d'HDR identique dans des cellules souches pluripotentes humaines induites.

Avec le même principe, Charpentier et al. ont fusionné à la Cas9 le fragment actif de la CtIP, protéine clé de la résection des brins lors de la RH¹⁵⁸. Les auteurs ont également montré une augmentation du taux d'HDR dans des lignées embryonnaires pluripotentes.

Enfin, la Cas9 a récemment été fusionnée avec un dominant-négatif de 53BP1 (Cas9-DN1S)¹⁵⁹. Comme pour les inhibiteurs i53 et e18, la Cas9-DN1S permet à la fois une diminution du pourcentage d'InDels et un taux d'HDR accru dans des lignées cellulaires et des cellules primaires.

h. Vers un ratio HDR / InDels parfait ?

A l'heure actuelle, les stratégies de *knock-in* restent encore un certain défi. Si des progrès majeurs ont été faits avec un taux d'HDR supérieur à 20% dans des cellules CD34⁺ en utilisant une Cas9 sous forme de RNP et un ssODN⁷⁸, la balance HDR/InDels reste souvent inférieure à 1. Le problème des InDels induits par la cNHEJ est intrinsèquement lié à la CDB. Bien qu'un nombre croissant de publications offrent des solutions diverses pour améliorer ce ratio, il est probable que l'amélioration la plus significative vienne d'un changement radical de dogme sur l'édition génique. En effet, si historiquement les endonucléases ont été au centre de l'édition génique, des solutions alternatives, indépendantes de la CDB, montrent des résultats prometteurs vis-à-vis de la balance HDR / InDels ; le dernier chapitre de l'introduction sera consacré à présenter l'une d'elle.

iii. Génotoxicité *on-target* – Echelle du kilobases

La grande majorité des mutations induites par la réparation d'une CDBs sont des InDels de moins de 20 nucléotides, allant rarement jusqu'à quelques centaines de paires de base. Il est important de noter que l'utilisation d'une paire d'ARNg ciblant un segment de plusieurs kilobases permet une délétion efficace de la région^{53,81} ; des inversions étant également retrouvées^{72,84}. La détection des événements induits par la Cas9 s'est largement appuyée sur les techniques de biologie moléculaire permettant de repérer ces changements à l'échelle de la paire de bases, à savoir PCR et séquençage Sanger et/ou haut-débit ; en utilisant des amorces encadrant la région supprimée lorsque deux ARNg sont utilisés. Toutefois, ces techniques n'autorisent que l'analyse d'une zone très restreinte (généralement <1kb) et sont notamment aveugles aux délétions plus grandes que la région amplifiée.

a. Des moyens de détections spécifiques

Une équipe du Wellcome Sanger Institute a émis l'hypothèse que des événements induits à la suite d'une unique CDB pourraient ne pas avoir été détectés, et leur fréquence sous-estimée, à cause du biais d'analyse provoqué par la PCR¹⁶⁰. Des cellules souches embryonnaires murines mâles ont été modifiées en utilisant un système transposon PiggyBac pour leur permettre l'expression de la Cas9 et d'un ARNg dirigé contre le gène *PigA*. Il est situé dans le chromosome X et est donc hémizygote dans le modèle cellulaire utilisé. Les événements induits ont été analysés par une PCR longue-portée permettant d'amplifier une région de 5,7 kb (*Figure 41A*). Pour avoir un tableau complet des événements possibles, les cellules éditées ont été sous-clonées et la région éditée a été amplifiée par PCR en utilisant pour chaque clone des amorces de plus en plus éloignée jusqu'à obtenir un amplifiat. Parmi les allèles amplifiés, ont été observées :

- Des délétions simples allant jusqu'à 9,5 kb.
- L'insertion supposée d'un ARNm rétrotranscrit du gène *Hmgn1* ; 4 exons consécutifs de ce gène ayant été retrouvés au niveau du site édité. (*Figure 41B*)
- Des inversions de séquence jusqu'à quasiment 2 kb (*Figure 41C et D*)
- De grandes insertions de 11 pb à 2,5 kb, recopiant majoritairement des segments du gène *PigA*. Toutefois, pour 5 allèles, l'insertion est homologue à d'autres régions du génome.

Pour confirmer que ces observations ne sont pas dues à des conditions particulières d'édition, cette méthodologie a été reproduite au locus *Cd9* situé dans le chromosome 6, dans une lignée humaine immortalisée (RPE1) et également en remplaçant le système transposon par une nucléofection de la Cas9. Une forte génotoxicité *on-target* a été retrouvée dans chaque condition. Les auteurs concluent leur étude sur une mise en garde sur le fait que ces lésions, souvent mésestimées, pourraient constituer un premier *hit* carcinogénique, notamment lors de l'édition de cellules souches ou progénitrices. L'ensemble de ces résultats souligne le besoin d'une caractérisation complète des événements induits.

b. Un risque à mesurer

Une illustration de ce besoin est donnée par les commentaires ayant fait suite à la publication portant sur la correction par *knock-in* d'une mutation du gène *MYBPC3*, impliquée dans les cardiopathies hypertrophiques, dans des embryons humains¹⁶¹.

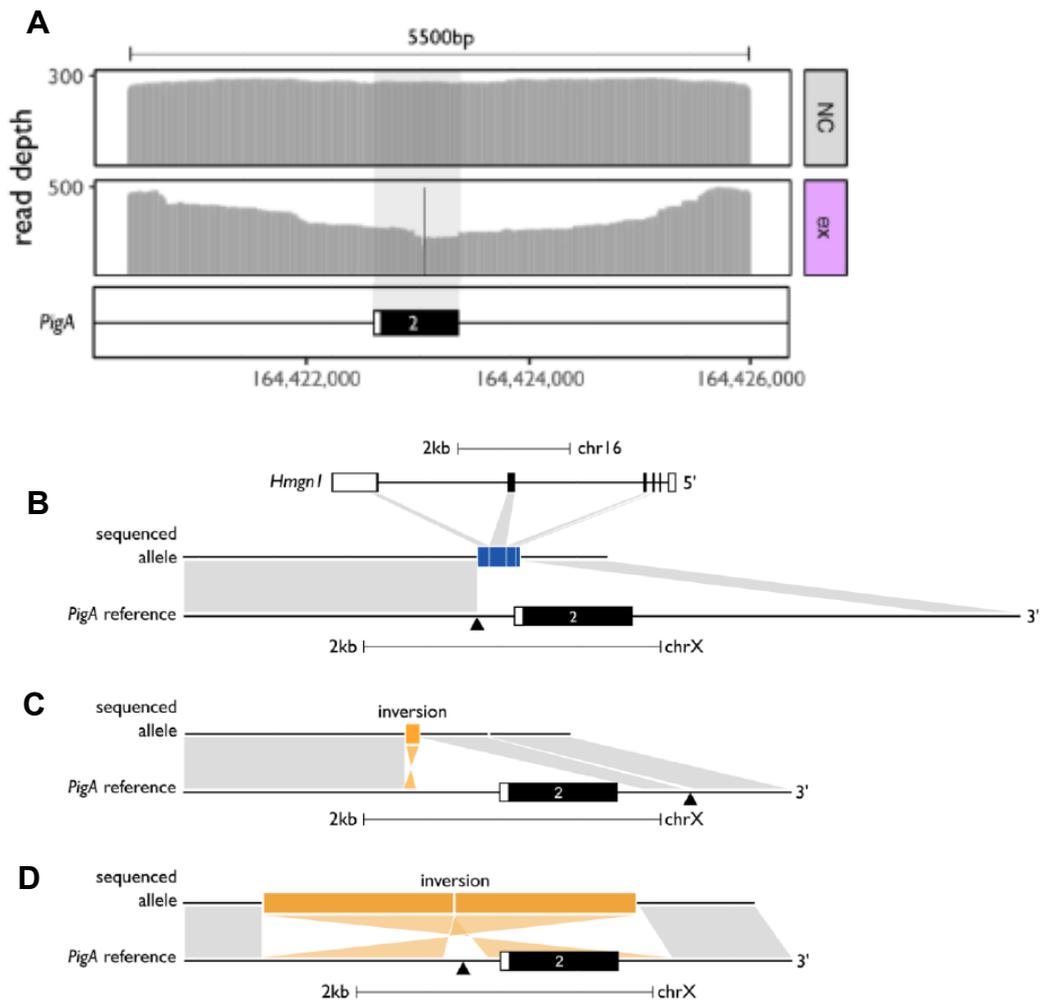


Figure 41. Unique CDB et délétions de plusieurs kilobases. (Kosicki et al., Nat. Biotechnol., 2018)

(A) Quantification de la profondeur des *reads* obtenus par séquençage des produits de PCR *long-range* issus de cellules éditées (*ex* en rose) ou non (NC) au niveau de l'exon 2 du gène *PigA*. La barre verticale noire indique le point de cassure dans la condition *Ex*. La zone grisée correspond à l'exon 2.

(B-D) Exemple d'allèles observés avec, en bas le locus *PigA* de référence et en haut l'allèle séquencé. (B) Grande délétion avec insertion de 4 exons du gène *Hmgn1*, situé sur le chromosome 16. (C) Courte inversion avec grande délétion. (D) Deux grandes inversions avec une délétion au centre.

Des zygotes ont été microinjectés avec la RNP Cas9, avec ou sans ssODNs porteur de la correction, et étonnamment, une correction a été observée dans les deux conditions. L'hypothèse avancée par Ma et al. est qu'un mécanisme particulier par HDR utilise l'allèle maternel (ici, WT) pour corriger l'allèle paternel. Toutefois, le génotype homozygote WT, venant de zygotes considérés comme corrigés, est montré par une PCR amplifiant un fragment de 534 pb.

Adikusuma et al. ont formulé un commentaire en soulignant le fait que l'apparition de ce génotype WT pouvait être également provoqué par des délétions supérieures à 534 pb dans l'allèle paternel, et donc invisibles à une approche par PCR classique¹⁶². Pour appuyer leur commentaire, ils ont reporté les résultats obtenus suite à l'injection de l'ARNm Cas9 dans des zygotes de souris. Si une majorité des zygotes a pu être analysée par une PCR de 600 pb, 30% ont nécessité une PCR de 1,6 kb et 7% de 3,2 kb ; la délétion la plus grande observée étant de 2,3 kb (*Figure 42*). Ces résultats ont pressé Adikusuma et al. à demander une vérification du génotype précis des embryons humains considérés comme corrigés. Toutefois, après vérification avec des régions amplifiées de 493 pb à 20 kb, aucune grande délétion n'a été retrouvée. Cette discussion, publiée dans *Nature* (2018), montre une certaine vigilance vis-à-vis des événements génotoxiques « particuliers » pouvant être induit par la Cas9¹⁶³.

iv. Génotoxicité *on-target* – Echelle du mégabase et du chromosome

Sans en être l'objectif principal, deux études publiées récemment ont mis en lumière le risque de voir apparaître des délétions majeures, de plusieurs dizaines de mégabases, lorsque deux CDB sont induites simultanément dans le même chromosome, de façon volontaire¹⁶⁴ ou involontaire⁷⁸, à cause de l'activité *off-target*.

a. Délétion à l'échelle du chromosome

En 2017, deux études ont profité des atouts du système CRISPR-Cas9 pour supprimer, de façon ciblée, un chromosome entier^{164,165}. Ces résultats obtenus *in vitro* et *in vivo* pourraient être appliqués pour la modélisation et la correction des aneuploïdies.

L'étude d'Adikusuma et al. utilise un modèle ciblant le chromosome Y (la perte de ce dernier n'impactant pas la viabilité de la cellule) dans des cellules embryonnaires murines. Un système rapporteur permet de quantifier la perte de chromosome (*Figure 43A*). Il repose sur le dosage génique de deux gènes situés de part et d'autre du chromosome Y : les gènes *Uba1y* et *Erdr1* respectivement sur le bras court et long. Dans l'ensemble des résultats de l'étude, deux cas sont particulièrement intéressants en vue d'édition génique : celui de deux CDBs encadrant le centromère ou dans le bras long du Chr.Y.

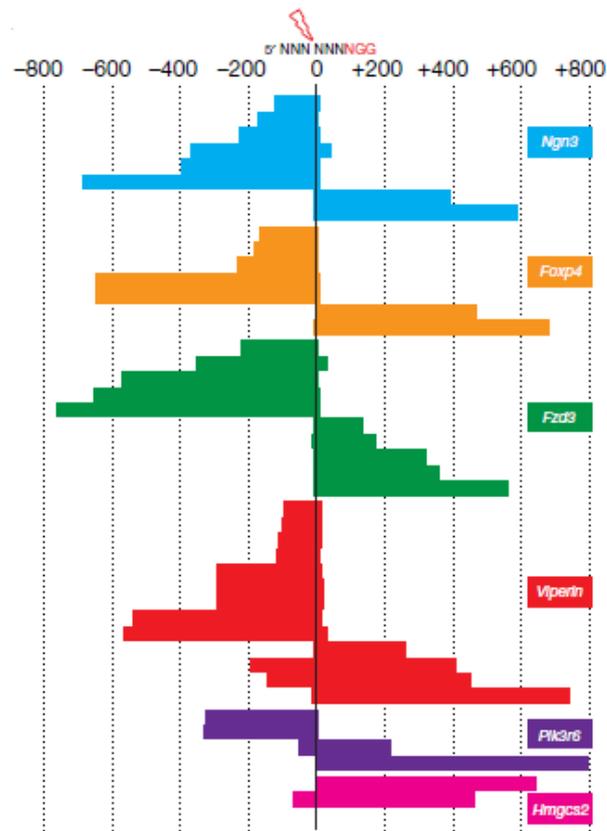


Figure 42. Unique CDB et délétions de plusieurs centaines de bases. (Adikusuma et al., *Nature Brief Communications*, 2018)

Délétions détectées après avoir microinjecté des zygotes de souris avec l'ARNm Cas9 et l'ARNg dirigée contre l'un des 6 loci présentés. Une PCR amplifiant une région de 1,6 kb a été nécessaire. La barre verticale représente le site de coupure de l'ARNg. Un seul locus est édité à la fois.

Dans le premier cas, les deux CDBs encadrent la région centromérique, l'objectif de l'étude étant d'induire la perte du chromosome en supprimant le centromère. Les deux coupures sont distantes de 100 kb. Après les deux coupures péricentromériques, le signal d'*Erd1* et d'*Uba1y* est diminué d'environ 40% suggérant une perte du Chr.Y dans 40% des cellules (*Figure 43B*). La corrélation entre perte de signal des gènes rapporteurs et perte de chromosome a été confirmée par FISH.

Dans le second cas, les deux CDBs sont situées dans le bras long du Chr.Y, à 300 kb du centromère, et sont distantes l'une de l'autre de 100 kb. Après les deux coupures, un déséquilibre génique est observé entre le gène rapporteur situé sur le bras long (*Erd1*) et celui sur le bras court (*Uba1y*). Le marqueur du bras court (*Uba1y*) est équivalent à la condition contrôle (-3%) tandis qu'une perte de signal du marqueur du bras long (*Erd1*) est observée (-27%) (*Figure 43B*). L'hypothèse d'une troncature du chromosome a été émise pour expliquer le différentiel de signal. Toutefois, ce résultat n'a pas été caractérisé plus précisément, étant donné que cette condition ne permet *a priori* pas la perte du chromosome entier, objectif premier de l'étude.

Ces deux cas particuliers de l'étude montrent que des délétions de l'ordre du mégabase, voire du chromosome entier, pourraient être induites par deux coupures. Si ici, déclencher ce type d'évènements était l'objectif et était donc particulièrement surveillé, ce risque pourrait être sous-estimé lorsque l'induction d'une seconde cassure dans le même chromosome est involontaire (*off-target*).

b. Délétion et inversion à l'échelle du mégabase

L'étude récente de Park et al. visant à corriger la mutation de la drépanocytose dans des CD34⁺ illustre ce risque dans un contexte d'édition génique et souligne également l'attention croissante pour ce type d'évènement⁷⁸. L'ARNg utilisé pour viser la mutation d'intérêt situé dans le bras court du Chr.11, possède plusieurs dizaines de sites *off-target*, dont l'un (*OT18*) situé dans le bras long du même chromosome (*Figure 44A*). Malheureusement, bien qu'il y ait 3 mésappariements entre les deux cibles, ce site est fréquemment muté. Deux jours après la transfection de CD34⁺ de patients drépanocytaires avec la RNP Cas9:ARNg dirigée contre la mutation d'intérêt, un taux de 36,7% d'InDels est observé à l'*OT18*. Les deux sites sont distants de 54 Mb, comprenant le centromère du Chr.11. Face à ce constat, un travail de recherche de délétions ou d'inversions de ce segment a été fait à l'aide de PCR dont les amorces encadrent les points de cassures attendus (*Figure 44A*). Le taux d'inversion atteint 3,3% et se maintient entre les trois temps d'analyses (J2, J7 et J12) (*Figure 44B*). Le taux de délétion est de 1,8% à J2 et diminue à 0,2% à J12 (*Figure 44C*) ; cette diminution s'explique par la perte du chromosome provoquée par la perte du centromère et empêchant ainsi l'amplification par PCR.

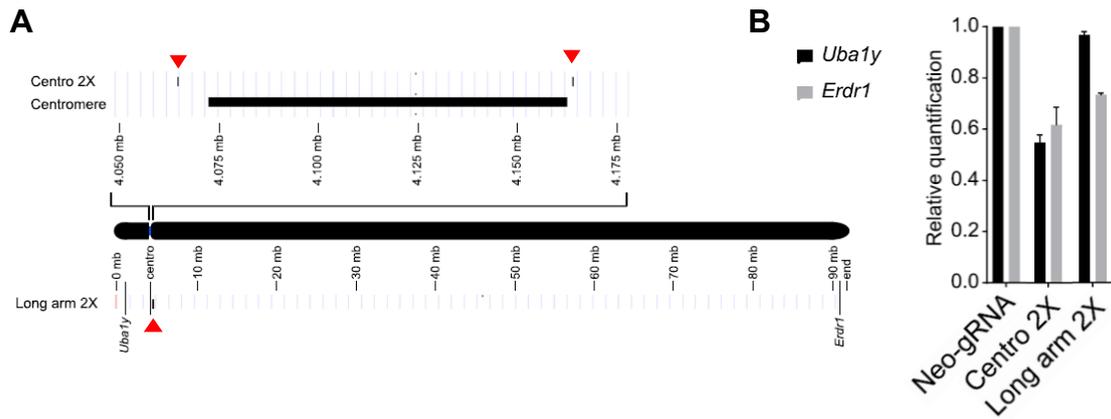


Figure 43. Double CDBs et délétions à l'échelle du chromosome. (modifiée d'après Adikusuma et al. *Mol. Ther.*, 2017)

(A) Localisation des sites de coupures sur le ChrY. Les sites de coupures sont situés de part et d'autre du centromère (Centro 2X) ou dans le bras long (Long arm 2X). Les sites de coupures sont indiqués avec un triangle rouge. Pour le bras long, les deux coupures sont distantes de 100 kb et indiquées par un unique triangle. La position des gènes *Uba1y* (bras court) et *Erdr1* (bras long) est indiquée.

(B) Dosage génique des gènes rapporteurs par qPCR après transfection des cellules souches embryonnaires murines avec un plasmide codant pour la Cas9 et l'ARNg visant les cibles Centro 2X ou Long arm 2X indiquées en (A). *Sox1* a été utilisé comme contrôle interne pour le dosage génique. La transfection d'un ARNg dirigé pour le gène de résistance à la néomycine a été faite pour servir de témoin négatif (*Neo-gRNA*).

Pour remédier à ce problème, le même protocole a été utilisé avec la HiFi-Cas9. Avec ce variant à spécificité augmentée, seulement 0,23% d'InDels sont observés à l'*OT18* avec une même efficacité *on-target*. Dans cette condition, les cellules éditées présentent un taux d'inversions et de délétions similaire aux cellules non-éditées (*Figure 44B et C*).

Ces études soulignent le besoin d'une caractérisation complète des événements induits par la Cas9. Toutefois, celle-ci est souvent partielle, car relativement lourde techniquement, en l'absence d'outils adaptés, et avec *a priori* (e.g. uniquement entre le site *on-target* et le site *OT18*). Concernant les événements à l'échelle du chromosome pouvant être induits par la Cas9, ils n'ont été observés qu'avec deux coupures simultanées dans le même chromosome et la question de savoir si un phénomène comparable est possible suite à une unique coupure reste entière.

Toutefois, l'accumulation de preuves expérimentales entre génotoxicité indésirable et Cas9-nucléase a poussé plusieurs équipes à chercher des solutions alternatives, moins génotoxiques en proposant des approches indépendantes de la cassure double-brin. L'une de ces solutions présentée ci-dessous, repose sur l'utilisation d'une Cas9^{D10A}-nickase, induisant des cassures simple-brin.

V. Edition génique par la Cas9^{D10A}-nickase

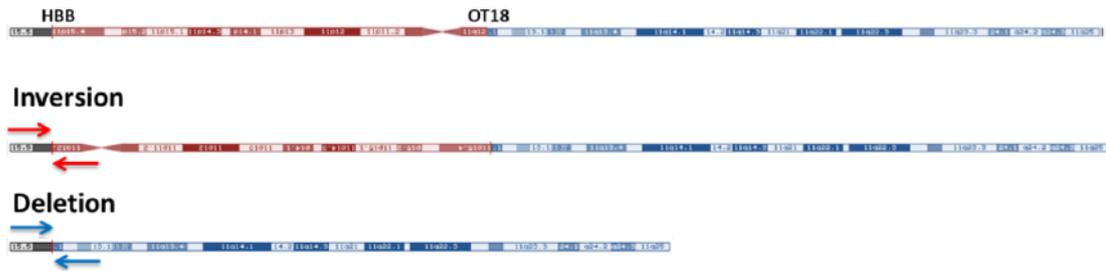
Les cassures simple-brin (CSBs) sont beaucoup plus fréquentes que les CDBs, avec plus de 10 000 occurrences par jour et par cellule¹⁶⁶. Les CSBs sont d'origine diverses : attaque oxydative par des espèces réactives de l'oxygène, activité des ADN topoisomérase I (TOP1) ou encore, produits intermédiaires de la réparation par excision de base (BER) ou de la réparation par excision de ribonucléotide (RER) (lorsque des ribonucléotides sont incorporées dans l'ADN).

Si les origines de CSB sont multiples, leurs voies de réparation sont relativement convergentes permettant la prise en charge de toutes les CSBs avec peu d'acteurs moléculaires. De plus, la correction des CSBs est efficace et peu mutagène en comparaison des CDBs¹⁶⁷⁻¹⁶⁹ ; l'information génétique étant intacte sur le brin non-lésé. En vue d'édition génique, cette méthode d'édition pourrait avoir un intérêt considérable pour lutter contre la génotoxicité indésirable.

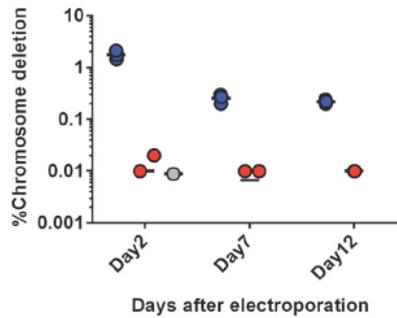
i. HDR et cassure simple-brin

Plusieurs études ont montré qu'il était possible de faire du *knock-in* en induisant une CSB et en apportant simultanément une matrice de réparation.

A Unrearranged Ch11



B



C

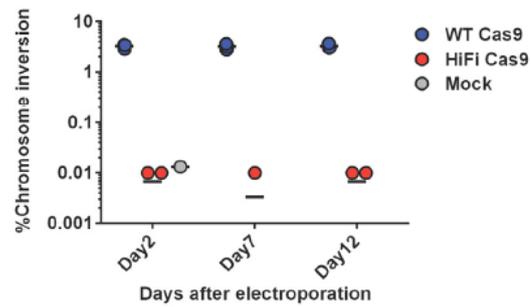


Figure 44. Double CDBs et délétions à l'échelle du chromosome par activité *off-target*. (Park et al., NAR, 2019)

(A) Système de détection par PCR des inversions et des délétions entre le site *on-target*, situé dans le gène *HBB*, et le site *off-target* 18 (OT18), situé dans le bras long du chromosome 11. La région pouvant être inversée ou supprimée est colorée en rouge.

(B-C) Fréquence de délétions (B) ou d'inversions (C) mesurée à 2, 7 et 12 jours après transfection de CD34⁺ par des RNPs dirigées contre le gène *HBB*. La Cas9 WT ou la HiFi-Cas9 ont été utilisées. La quantification est faite par ddPCR.

Ce procédé de *knock-in* a été obtenu avec des méganucléases^{167,170}, des ZFNs¹⁶⁸ et plus récemment avec la Cas9^{D10A}¹⁷¹⁻¹⁷³. Ces études montrent qu'il est possible de faire du *knock-in*, à faible efficacité, en utilisant une stratégie indépendante de la CDB.

a. Cassure simple-brin ou cassure double-brin ?

L'un des premiers doutes à lever concerne la nature exacte de la lésion à l'origine du *knock-in*. La question sous-jacente étant de savoir si le *knock-in* observé est dû à une véritable forme de *knock-in* faisant suite à une CSB ou si la modification fait suite à une CDB. En effet, une activité endonucléasique résiduelle des outils imparfaitement transformés en nickase pourrait persister ou une fraction des CSBs pourrait être transformée en CDBs, notamment lors de la réplication de l'ADN¹⁷⁴. Les observations ci-dessous dessinent une réponse :

- L'inhibition chimique des acteurs de la cNHEJ permet une amélioration du ratio HDR/InDels induit par une CDB, mais n'affecte pas le ratio obtenu par une CSB¹⁷³. Ce comportement différentiel suggère que l'HDR au CSB ne passe pas par sa transformation préalable en CDB.
- L'inhibition des acteurs-clés de la réparation des CSBs permet d'augmenter l'HDR initiée au niveau de CSB. Cette observation a été faite avec l'inhibition pharmacologique de PARP¹⁶⁹, véritable senseur des CSBs, et dans une lignée cancéreuse déficiente pour XRCC1, acteur clé de la réparation¹⁷³.

L'ensemble de ces résultats suggère que ce sont les CSBs échappant à la voie de réparation classique, mais non-transformées en CDB, qui mènent à une forme de *knock-in*.

b. Deux voies de réparation

Cependant, le mécanisme de *knock-in* initié par une CSB est demeuré peu caractérisé jusqu'à l'étude de Davis et Maizels (2014)¹⁷¹ ; la question principale étant de savoir si le *knock-in* se faisait par des voies d'HDR communes à celles déclenchées par une CDBs ou si elle se faisait par une voie alternative. Les auteurs ont étudié le taux d'HDR après transfection de lignées cellulaires humaines avec une Cas9^{D10A}-nickase et une matrice de réparation double (dsDNA) ou simple-brin (ssODN). Lors de cette étude, il a été montré que les CSBs pouvaient emprunter deux voies d'HDR : l'une empruntant la voie du SDSA, la seconde faisant appel à des formes alternatives d'HDR.

HDR RAD51 et BRCA2 dépendantes.

La première voie de réparation caractérisée s'appuie sur une matrice double-brin d'ADN (*Figure 45A*). Elle est dépendante de RAD51 et BRCA2, deux acteurs clés de la RH.

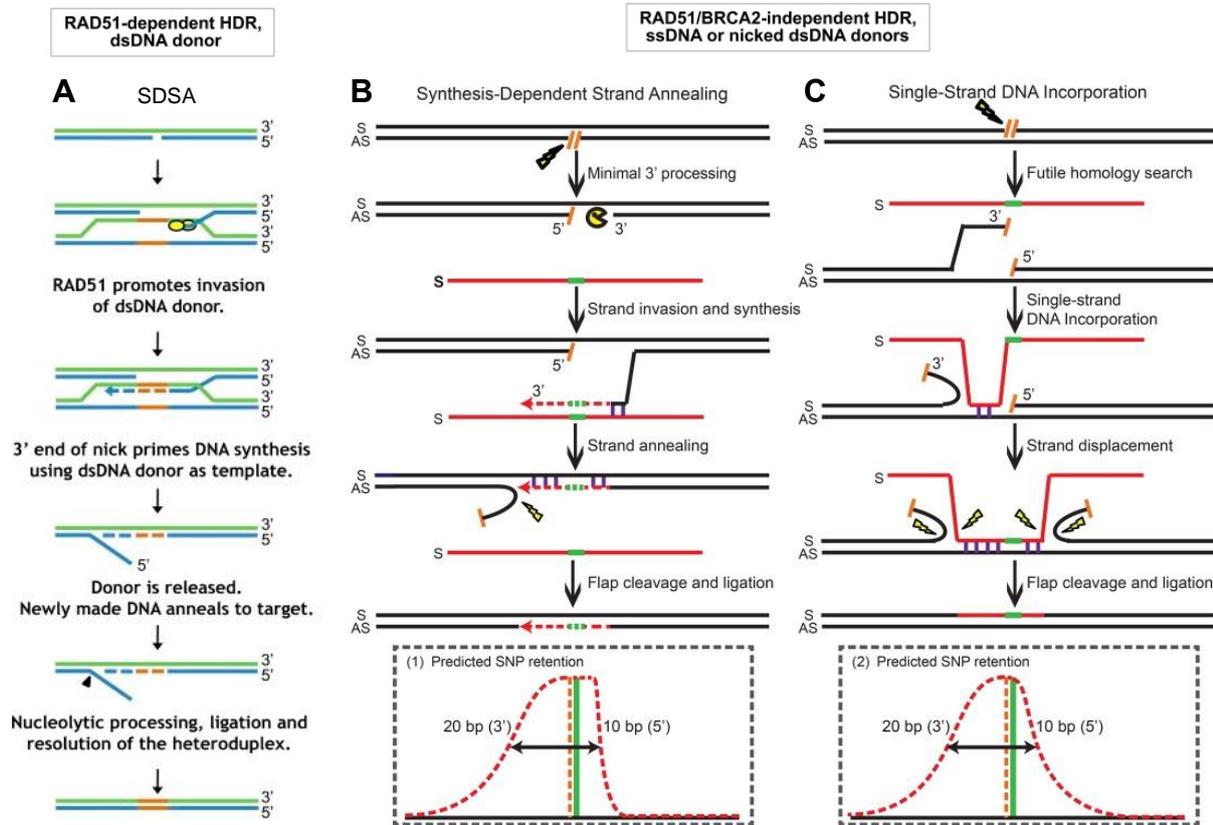


Figure 45. Voies de réparation des CSBs par HDR.

(A) Modèle de réparation des CSBs par HDR avec une matrice double-brin d'ADN. Les étapes sont similaires à la réparation par SDSA lors d'une CDB. (modifiée d'après Davis et Maizels, PNAS, 2014)

(B-C) Modèles de réparation des CSBs par HDR avec un ssODN complémentaire au brin clivé (B) ou au brin intact (C). (B) Le modèle de réparation ressemble à la réparation par SDSA mais est indépendant de RAD51 et BRCA2. Le profil d'intégration des mutations est asymétrique favorisant l'introduction des mutations en 3' du site de coupure. (C) Le modèle de ssDI dans lequel le ssODN complémentaire au brin intact s'hybride au niveau du site de coupure. Le profil d'intégration possède une distribution normale non-centrée sur le site de coupure. (Kan et al., Genome Res., 2017)

Ce modèle concorde avec les observations obtenues dans l'étude de Vriend et al. (ici une matrice plasmidique est utilisée)¹⁷³. En effet, l'inhibition de la RH par le blocage de l'interaction entre BRCA2 et RAD51 par BRC3, entraîne une forte diminution de l'HDR initiée par une CSB et également par une CDB¹⁷³. Ce résultat suggère une voie commune d'HDR utilisant des étapes de la RH résolue par la voie du SDSA.

HDR RAD51 et BRCA2 indépendantes.

La voie alternative d'HDR utilise des ssODNs. Elle est inhibée par RAD51 et BRCA2, mais nécessite BRCA1 et RPA, des caractéristiques partagées par la réparation par SSA¹⁷¹. Cette voie alternative peut être subdivisée en deux. D'une part, le *Single-Strand DNA Incorporation* (ssDI) si le ssODN est complémentaire au brin intact. La seconde arrive lorsque le ssODN est complémentaire au brin clivé. Bien qu'elle partage des caractéristiques de la SSA¹⁷¹, elle est également parfois appelée *Synthesis-Dependant Strand Annealing* (SDSA)¹²⁶. Malgré tout, la notion de synthèse semble être importante pour cette voie et l'appellation de SDSA pourrait être plus adaptée.

Synthesis-Dependant Strand Annealing (SDSA)

Si le ssODN est complémentaire au brin clivé, celui-ci s'hybride et sert de matrice pour la synthèse d'un nouveau brin, permettant l'incorporation des mutations (*Figure 45B*). Dans cette conformation, la « fenêtre » dans laquelle les mutations sont incorporées est asymétrique par rapport au point de cassure. Cette fenêtre asymétrique est concordante avec une HDR par SDSA. La synthèse du nouveau brin se faisant de 5' vers 3', l'extrémité 3' du ssODN sert à l'hybridation et les mutations situées juste après cette région sont les plus susceptibles à être répliquées puis incorporées. Le brin néo-synthétisé se défait du ssODN et se ré-hybride à son brin complémentaire. Si la synthèse dépasse la lésion génomique, la structure résultante est corrigée par la *flap-endonuclease 1*. Après hybridation du brin néo-synthétisé, des hétéroduplexes au niveau des mutations incorporées apparaissent. Ces mésappariements sont corrigés par la voie du MMR, ou séparés par la réplication de l'ADN avant leur résolution. Si le brin porteur des mutations est pris comme modèle de réparation par le MMR, celles-ci sont intégrées définitivement.

Single-Strand DNA Incorporation (ssDI)

Si le ssODN est complémentaire au brin intact, celui-ci s'hybride directement au niveau de la CSB (*Figure 45C*). Ici, la « fenêtre » d'incorporation des mutations est symétrique par rapport au point de cassure, avec un léger décalage en faveur de l'extrémité 3'. Il est supposé qu'un espace est libéré préférentiellement au niveau de l'extrémité 3' de la CSB lors de la recherche d'une séquence homologue.

Ce comportement permettrait également d'expliquer le léger décalage de la fenêtre d'incorporation. Après hybridation par complémentarité, un mésappariement apparaît au niveau des mutations portées par le ssODN. Deux solutions sont possibles pour expliquer l'intégration pérenne des mutations (*Figure 46A*). Soit le ssODN s'incorpore au génome, sert ensuite de modèle à la correction par *mismatch repair* (MMR) et les mutations sont ainsi transmises au brin intact. Soit le ssODN sert d'abord de modèle au MMR transmettant les mutations au brin intact. Le ssODN est ensuite éliminé et la synthèse du brin manquant permet de dupliquer les mutations sans mésappariements.

Impact du *mismatch repair* (MMR)

Puisque les voies d'HDR initiées par des CSBs font apparaître des mésappariements au niveau des mutations introduites, le MMR, en charge de résoudre les mésappariements, pourrait jouer un rôle dans l'HDR initiée à une CSB. Premièrement, la voie du MMR n'est pas nécessaire, puisque de l'HDR peut être obtenue dans des cellules MMR-déficientes¹²⁶. Dans les cellules déficientes pour le MMR, les mésappariements pourraient être résolus après un cycle de division cellulaire. La réplication étant semi-conservative, les cellules filles ne porteront plus de mésappariement. En comparant deux lignées, l'une déficiente l'autre non, il a été observé que le MMR diminue la taille de la fenêtre dans laquelle des mutations peuvent être introduites par la voie du SDSA et diminue par un facteur 5, l'HDR initiée par ssDI¹²⁶ (*Figure 46B*). Au final, le MMR semble donc avoir un rôle inhibiteur de l'HDR en protégeant de l'incorporation de nouvelles mutations.

Ouverture et résection des cassures simple-brin

Les deux mécanismes d'HDR par les ssODN nécessitent une hybridation entre le ssODN et un des brins d'ADN, suggérant qu'une activité de décompaction ou de résection a lieu pour permettre cette hybridation. Une première preuve en faveur de cette hypothèse a été obtenue en utilisant des ssODNs asymétriques par rapport aux points de cassures ; les résultats suggérant une ouverture préférentielle à l'extrémité 3' de la CSB¹⁷¹. Le besoin de séparer les deux brins pourrait également expliquer l'effet bénéfique observé par l'inhibition de RAD51 et BRCA2. La séparation des deux brins au niveau d'une CSB forme des filaments simple-brin, structures reconnues par RPA (*Figure 47*). Comme lors de la RH, les complexes RPA sont remplacés par RAD51 par un processus nécessitant l'intervention de BRCA2. Ce changement facilite la réhybridation des deux brins et permet la réparation de la CSB. À l'inverse, l'inhibition de RAD51 ou de BRCA2 pourrait stabiliser la CSB dans une conformation relâchée, présentant des structures simple-brin accessibles aux ssODNs, favorisant ainsi l'HDR¹⁷⁵.

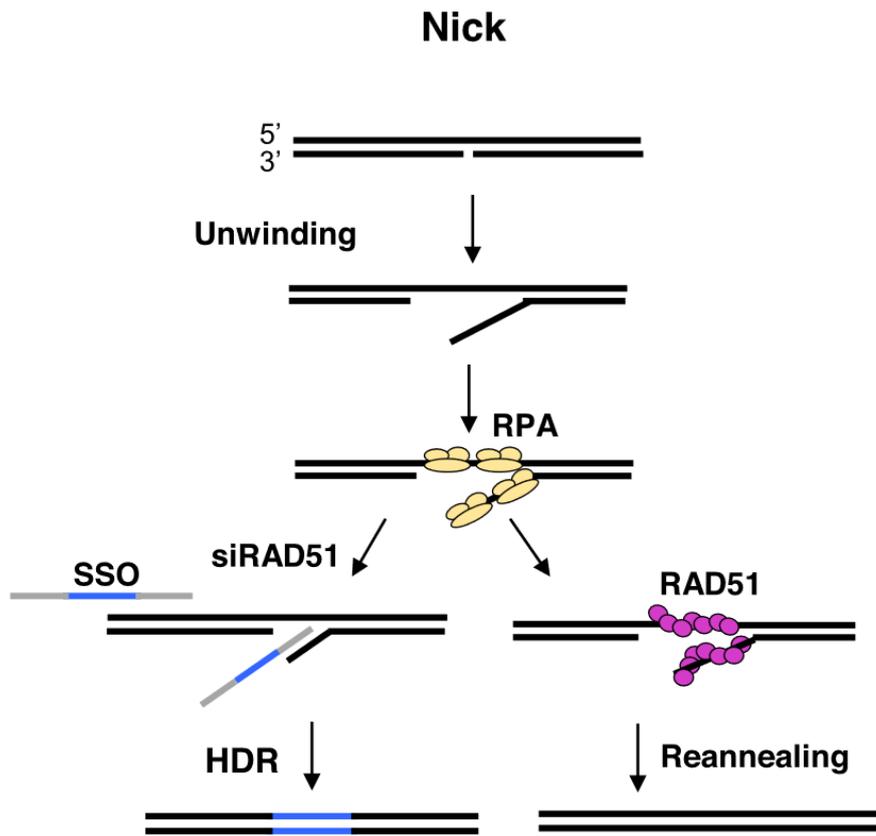


Figure 47. Implication de RAD51 dans l'HDR initiée aux CSBs.

Effet de RAD51. L'ouverture locale au niveau de la CSB forme des structures simple-brin reconnues par les complexes RPA. L'inhibition de RAD51 par un siRNA permet de favoriser l'hybridation du ssODN (SSO) avec le brin clivé. (Davis et Maizels, *Cell. Rep.*, 2016)

Une HDR sans intégration ?

Définir si l'HDR initiée au CSB se traduit ou non par l'intégration physique du ssODN dans le génome peut avoir une grande importance ... mais surtout d'un point de vue législatif. En effet, l'étude conduite par Kan et al renforce les modèles d'HDR par SDSA et ssDI par un travail minutieux¹²⁶. Ce dernier leur permet de confirmer que lors du SDSA, le ssODN n'est pas intégré au génome et sert uniquement de matrice de réparation. Si cette différence peut être anecdotique, elle est fondamentale lors du classement ou non sous l'étiquette d'OGM. L'article conclut qu'une approche utilisant la Cas9^{D10A}-nickase combinée à un ssODN complémentaire au brin clivé permettrait de faire une édition génique sans être considéré comme « organismes génétiquement modifiés ».

ii. Une solution viable contre la génotoxicité ?

a. Améliorer l'efficacité

L'édition génique par la Cas9^{D10A}-nickase pourrait être particulièrement intéressante en vue d'édition génique. Toutefois, l'efficacité obtenue avec ce type d'approche reste faible avec une fréquence d'HDR souvent inférieure à 1%. De plus, il est à noter que les preuves de cette HDR alternative ont principalement été obtenues dans des lignées cellulaires cancéreuses et qu'aucune preuve n'a encore été obtenue dans des cellules souches, notamment hématopoïétiques. L'ensemble de ces raisons peut provoquer une certaine réticence vis-à-vis de sa pertinence dans le but de modélisation pathologique ou dans le cadre d'une stratégie thérapeutique d'édition génique. Malgré tout, encore peu d'études sont consacrées à cette forme d'édition génique, et son optimisation pourrait lui permettre de devenir une alternative viable.

b. Evaluer la génotoxicité

L'un des principaux atouts de l'édition génique par la Cas9^{D10A}-nickase est de s'affranchir de la CDB, évènement délétère pour la cellule. Proposer une approche sans CDB pourrait permettre de se défaire du même coup de la génotoxicité qui lui est associée. Si l'édition génique basée sur la Cas9^{D10A} pour faire de l'HDR *via* un ssODN reste relativement confidentielle, un très fort engouement est porté aux méthodes par édition de bases. Si les premières générations utilisaient une dCas9, les suivantes, plus performantes, reposent également sur la Cas9^{D10A}. Une évaluation prudente et sans *a priori* de la génotoxicité induite par la CSB permettrait de juger de la sûreté de ces approches.

Résultats

I. Objectifs du projet de thèse

L'objectif du projet de thèse a été d'évaluer l'efficacité et la génotoxicité provoquées par la Cas9-nucléase suite à une unique cassure double-brin dans un modèle de maladie héréditaire du globule rouge, la porphyrie erythropoïétique congénitale. Cette évaluation a d'abord été conduite au niveau nucléotidique en quantifiant la balance HDR Indels, marqueur de qualité de l'édition par *knock-in*, puis une évaluation à l'échelle du chromosome a été faite pour caractériser l'impact d'une unique cassure double-brin sur l'intégrité chromosomique. Les résultats obtenus nous ont amenés à développer une approche alternative moins génotoxique, indépendante de la cassure double-brin.

La seconde partie du travail de thèse a été consacrée à développer un outil sensible de détection des délétions terminales par cytométrie en flux. La mise au point de cet outil offre une plateforme d'analyse pour comparer les différentes stratégies d'édition génique en évaluant leur impact sur la génotoxicité *on-target*.

II. Modèle d'étude : la porphyrie érythropoïétique congénitale

i. Une pathologie candidate à la thérapie génique

Dès 2002 avec la création de l'unité E0217 « Pathologie Moléculaire et Thérapie Génique », reconduite en 2007 (U876) sous l'appellation « Transfert de gènes à visée thérapeutique dans les cellules souches », le laboratoire INSERM U1035 (formé en 2011) contribue depuis de nombreuses années au développement de la thérapie génique des formes sévères des porphyries érythropoïétiques^{17,176-179}.

Les porphyries héréditaires correspondent à un ensemble de maladies liées au déficit de l'une des enzymes intervenant dans la biosynthèse de l'hème. Elles sont caractérisées par une accumulation de porphyrines et/ou de leurs précurseurs. Les porphyries sont regroupées en porphyries érythropoïétiques ou porphyries hépatiques selon le tissu dans lequel le trouble du métabolisme prédomine.

Parmi celles-ci, la porphyrie érythropoïétique congénitale (PEC) (ou maladie de Günther) est une maladie autosomique récessive, liée à un déficit profond en uroporphyrinogène III synthase (UROS), la quatrième enzyme de la biosynthèse de l'hème.

A l'état physiologique, l'activité enzymatique d'UROS permet la cyclisation de l'hydroxyméthylbilane (HMB) en uroporphyrinogène de type III (URO III). En absence d'UROS, l'HMB se cyclise spontanément pour former de l'uroporphyrinogène de type I (URO I).

Les deux composés, URO III et URO I peuvent être pris en charge par l'enzyme suivante, l'uroporphyrinogène décarboxylase (UROD), pour former respectivement la coproporphyrinogène de type III (COPRO III) et de type I (COPRO I).

Cependant, si les protoporphyrines de type III sont transformées en protoporphyrines de type IX pour servir à la biosynthèse de l'hème, les protoporphyrines de type I (URO I et COPRO I) sont dans une impasse métabolique, s'accumulent dans la cellule et provoquent une cytotoxicité.

La principale manifestation clinique est une photosensibilité cutanée sévère sur les régions corporelles exposées au soleil, accompagnée chez certains patients, d'une hémolyse nécessitant des transfusions sanguines répétées. Le seul traitement curatif actuel est l'allogreffe de CSH.

Maladie sévère, récessive, hématologique et sans traitement hormis greffe de CSH, la PEC présente tous les arguments pour être candidate à la thérapie génique. Cette maladie a servi de pathologie modèle pour développer des nouveaux outils d'études (modèle murin PEC)¹⁷⁷ et de thérapie génique (vecteurs lentiviraux¹⁷⁶, iPSCs¹⁷) au sein du laboratoire. Combinés à l'arrivée du système CRISPR-Cas9, les travaux actuels se tournent vers une approche par édition génique.

ii. La PEC : un outil d'analyse pour l'édition génique précise

Si l'amélioration des techniques de greffe semble avoir devancé le développement d'outils de thérapie génique sûrs et efficaces, le modèle de la PEC offre malgré tout une formidable plateforme d'analyse pour le développement de l'édition génique dans une cible génique cliniquement pertinente. En effet, la fonctionnalité d'UROS dans les cellules peut être évaluée par cytométrie en flux ; les cellules non-fonctionnelles accumulant des protoporphyrines de type I étant photosensibles (fluorocytes). Cette particularité a permis de suivre l'impact des procédés d'édition génique à la fois au niveau génomique, avec la quantification de la balance HDR/InDels, mais également au niveau fonctionnel avec la quantification systématique des fluorocytes par cytométrie.

III. Première partie : Comparaison de deux procédés d'édition génique

La première partie du projet de thèse a permis de comparer l'efficacité et la génotoxicité de deux méthodes d'édition génique, l'une basée sur la Cas9-nucléase, la seconde basée sur la Cas9^{D10A}-nickase. Dans les deux cas, une unique lésion est induite. Pour se placer dans un modèle de thérapie génique, le modèle de la PEC a été utilisé. Le gène *UROS* a été ciblé et plus précisément, la position c.217T situé dans l'exon 4 d'*UROS* ; la mutation ponctuelle c.217T>C étant associée à un phénotype sévère.

Une évaluation *in vitro* des événements induits par la Cas9-nucléase ou la Cas9^{D10A}-nickase a été faite en suivant quatre principaux paramètres:

- L'activité *off-target*.
- La balance HDR / InDels, témoignant à la fois de l'efficacité et de la génotoxicité *on-target*.
- La fonctionnalité des cellules, en particulier de l'activité d'*UROS*.
- L'intégrité chromosomique par une analyse cytogénétique des conséquences de l'édition génique.

Bien qu'elle possède un profil intéressant en termes de génotoxicité, la principale limite de l'édition génique par la Cas9^{D10A}-nickase est celle de l'efficacité. Pour y pallier, un travail d'optimisation a été conduit pour permettre un taux d'HDR similaire à celui obtenu par la Cas9-nucléase.

Enfin, une preuve de principe de modélisation de pathologie et une de correction d'une mutation ponctuelle ont été faites avec une méthode d'édition génique basée sur la Cas9^{D10A}-nickase, toujours dans le modèle de la PEC.

ARTICLE

<https://doi.org/10.1038/s41467-019-09006-2>

OPEN

CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations

Grégoire Cullot^{1,2}, Julian Boutin^{1,2,3}, Jérôme Toutain⁴, Florence Prat^{1,2}, Perrine Pennamen⁴, Caroline Rooryck⁴, Martin Teichmann^{1,5}, Emilie Rousseau^{1,5}, Isabelle Lamrissi-Garcia^{1,2}, Véronique Guyonnet-Duperat^{2,6}, Alice Bibeyran^{2,6}, Magalie Lalanne^{1,2}, Valérie Prouzet-Mauléon^{1,7}, Béatrice Turcq^{1,7}, Cécile Ged^{1,2,3,8}, Jean-Marc Blouin^{1,2,3,8}, Emmanuel Richard^{1,2,3,8}, Sandrine Dabernat^{1,2,3}, François Moreau-Gaudry^{1,2,3,6,8} & Aurélie Bedel^{1,2,3,8}

CRISPR-Cas9 is a promising technology for genome editing. Here we use Cas9 nuclease-induced double-strand break DNA (DSB) at the *UROS* locus to model and correct congenital erythropoietic porphyria. We demonstrate that homology-directed repair is rare compared with NHEJ pathway leading to on-target indels and causing unwanted dysfunctional protein. Moreover, we describe unexpected chromosomal truncations resulting from only one Cas9 nuclease-induced DSB in cell lines and primary cells by a p53-dependent mechanism. Altogether, these side effects may limit the promising perspectives of the CRISPR-Cas9 nuclease system for disease modeling and gene therapy. We show that the single nickase approach could be safer since it prevents on- and off-target indels and chromosomal truncations. These results demonstrate that the single nickase and not the nuclease approach is preferable, not only for modeling disease but also and more importantly for the safe management of future CRISPR-Cas9-mediated gene therapies.

¹Univ. Bordeaux, 33000 Bordeaux, France. ²INSERM U1035, Biotherapy of genetic diseases, inflammatory disorders and cancers, 33000 Bordeaux, France. ³Biochemistry Laboratory, CHU Bordeaux, 33000 Bordeaux, France. ⁴Medical genetic laboratory, CHU Bordeaux, 33000 Bordeaux, France. ⁵UMR 5320, INSERM U1212, ARNA Laboratory, 33000 Bordeaux, France. ⁶Vectorology Platform, 33000 Bordeaux, France. ⁷INSERM U1218, ACTION, 33000 Bordeaux, France. ⁸Laboratory of excellence, GR-Ex, Imagine institute, 75015 Paris, France. These authors contributed equally: Grégoire Cullot, Julian Boutin, François Moreau-Gaudry, Aurélie Bedel. Correspondence and requests for materials should be addressed to F.M.-G. (email: francois.moreau-gaudry@u-bordeaux.fr)

CRISPR-Cas9 is an RNA-guided DNA endonuclease system targeting a specific genomic sequence complementary to a single-guide RNA (sgRNA) and juxtaposed with a proto-spacer adjacent motif (PAM). This system leads to a DNA double-strand break (DSB) via the RuvC and HNH nuclease domains of the Cas9 enzyme^{1–4}. Most publications report the use of engineered Cas9-nucleases to efficiently induce DSBs at sites of interest^{5–7}. DSBs lead to non-conservative non-homologous end-joining (NHEJ) repair pathway. Insertions or deletions (indels) at the on-target site often cause frameshifts in open reading frames and knockout (KO) genes. CRISPR-Cas9 applications are of particular interest to invalidate genes in the field of human genetics for disease modeling *in vitro* and *in vivo*⁸ and are promising for gene therapy. Sichuan University (China) was the first to submit a trial that consisted in injecting gene-edited cells in a person to evaluate the safety of *PD-1* knockout engineered T cells in treating metastatic non-small cell lung cancer⁹. A prospective phase I trial will start in the USA for patients with melanoma, synovial sarcoma, and multiple myeloma¹⁰. However, the CRISPR-Cas9 approach faces concerns regarding unintended alterations (off-target effect)¹¹. Safety issues regarding genomic instability and chromosomal integrity have not been explored in-depth and could be underestimated. Indeed, CRISPR-Cas9 has already been applied to generate intra-chromosomal translocations to obtain fusion genes such as the *EML4-ALK* oncogene^{12,13} and inter/intra-chromosomal translocations in human HEK293T cells¹⁴. Recently, Adikusuma *et al.* reported frequent large deletions (kilobase-scale) in mouse zygotes after CRISPR-Cas9 cleavage¹⁵. Chromosomal deletions have been described in zebrafish using two DSBs by TALENs (Transcription activator-like effector nucleases) or CRISPR-Cas9 systems¹⁶. Recently, Zuo *et al.* surprisingly demonstrated that multiple Cas9-mediated DNA cleavages on the same human chromosome can eliminate it, thereby offering a new approach for modeling aneuploidy diseases^{16–18}. However, although nuclease is the standard approach for genome editing, it is still unknown whether a part or an entire chromosome could be eliminated after only one DSB in human cells.

Another goal of using CRISPR-Cas9 is the possibility to perform homology-directed repair (HDR) to make precise genome editing (PGE) to insert or correct point mutations. This holds promise for correcting most hereditary diseases due to mutation hotspots. For example, congenital erythropoietic porphyria (CEP [MIM 263700]) is an autosomal recessive disorder characterized by a deficiency in the enzymatic activity of uroporphyrinogen III synthase (UROS; EC 4.2.1.75), the fourth enzyme of the heme biosynthetic pathway (Supplementary Figure 1). c.217 T > C substitution is a hotspot for this disease in almost one-third of all reported CEP cases^{19–21}. The enzymatic defect causes the accumulation of the uroporphyrin I. The lack of UROS causes mutilating dermatological lesions resulting from the release of photocatalytic porphyrins and transfusion-dependent hemolytic anemia with secondary hypersplenism. The precise correction of mutation hotspots, without insertional mutagenesis and with physiological gene regulation, is the holy grail of gene therapy. However, performing HDR is still challenging due to low efficiency and concurrent misrepair of DSBs by the NHEJ. It is still unclear whether large amounts of unwanted on-target indels could ablate the residual function of the original protein, thereby modifying the phenotype of cells and leading to iatrogenic effects.

Because error-prone NHEJ is the main repair pathway of DNA DSB, replacement of DSBs by single-strand break (SSB) could improve PGE by limiting indels. Nicks are the most common forms of DNA damage and elegant proof-of-principle has been provided that single nickase can initiate HDR^{22,23}. Even if

nuclease is still the gold standard system used to edit the genome²⁴, some authors^{25–27} recently proposed that Cas9^{D10A}-nickase could be an alternative to nuclease to reduce NHEJ. However, in this innovative approach, the HDR rate is often under 1%. Hence, knowing whether single nickase is potentially an alternative to nuclease is an area of active investigation. In this work, we characterize allelic outcomes and chromosomal integrity following either nuclease-induced DSB or nickase-induced SSB and monitored the HDR/indel ratio as a quality parameter of PGE (Fig. 1a). Using *UROS* as a target gene provide an easy and quantitative test of UROS function with detection of pathologic type-I porphyrins by flow cytometry. Our findings reveal the globally damaging effects of DSBs on the human genome in cell lines and primary cells in which the p53 tumor suppressor has been inactivated. They also highlight the possibility of using the single nickase approach to dramatically reduce indels and chromosomal terminal deletion while achieving a high HDR rate. This approach is therefore more relevant for testing disease models and for obtaining safer gene therapies.

Results

Profound KO is concurrent to knock-in using Cas9 nuclease.

UROS gene was edited in HEK293T cells by transient expression of Cas9-nuclease, a sgRNA and 181nt-single-stranded oligodeoxynucleotide (ssODN) template. To induce a DSB close to the most frequent mutation in CEP disease (c.217 T > C in *UROS* exon 4, chromosome 10), we designed a sgRNA inducing a DSB near the c.217 position and devised a 181nt-ssODN carrying a silent *SacI* restriction site close to the c.217 T position (Fig. 1b). Since HDR-mediated *SacI* insertion will lead to a modification of the sgRNA seed sequence, this event prevents Cas9-nuclease from re-cutting. We obtained a $21 \pm 2.4\%$ ($n = 5$) HDR rate by RLFP (restriction fragment length polymorphism) (Fig. 2a). TIDER (Tracking of Insertions, DEletions and Recombination events) analysis confirmed the presence of *SacI* (21.8% of alleles) and revealed a high rate of unwanted indels (60.7%) (Supplementary Figure 2a). Next-generation sequencing (NGS) analysis validated the HDR rate (34.4% of alleles) and high level of indels (Fig. 2b left panel). The predominant event is an insertion of one base pair at the cut site (Fig. 2b center and right panels). Only 1.3% of alleles were unmodified, attesting to the extremely high efficiency of Cas9-nuclease. To measure the impact of indels, we carried out an *UROS* enzymatic assay on transfected cells. This indicated a dramatic drop in *UROS* activity ($47.6\% \pm 2.8\%$ compared to non-transfected cells, $n = 3$) (Fig. 2c left panel). *UROS* dysfunction induced the striking appearance of a high level of fluorocytes ($35.7\% \pm 5.1\%$, $n = 3$) (Fig. 2c right panel) due to accumulation of type-I porphyrins in cells. In parallel with on-target analysis, we evaluated the off-target effects of nuclease by quantifying indels in the top 10 off-target sequences predicted by CRISPOR. We observed indels in the first three ranked loci, located in intergenic regions, in accordance with the well-known low specificity of nuclease (Supplementary Table 1). Altogether, using the nuclease approach, we found the simultaneous presence of HDR with unacceptable on-target indels, leading to frequent KO of the *UROS* gene, to a metabolic deficiency, and to off-target side-effects.

Cas9 nuclease induces unwanted chromosomal truncations.

The HDR/indel ratio determined by next generation sequencing (NGS) is often considered as the gold standard to quantify the on-target genotoxicity of DSB. However, it uses PCR so it cannot reveal large chromosomal abnormalities. In view of recent publications describing that multiple DSBs induced by Cas9-nuclease can be used to eliminate an entire chromosome, we wondered

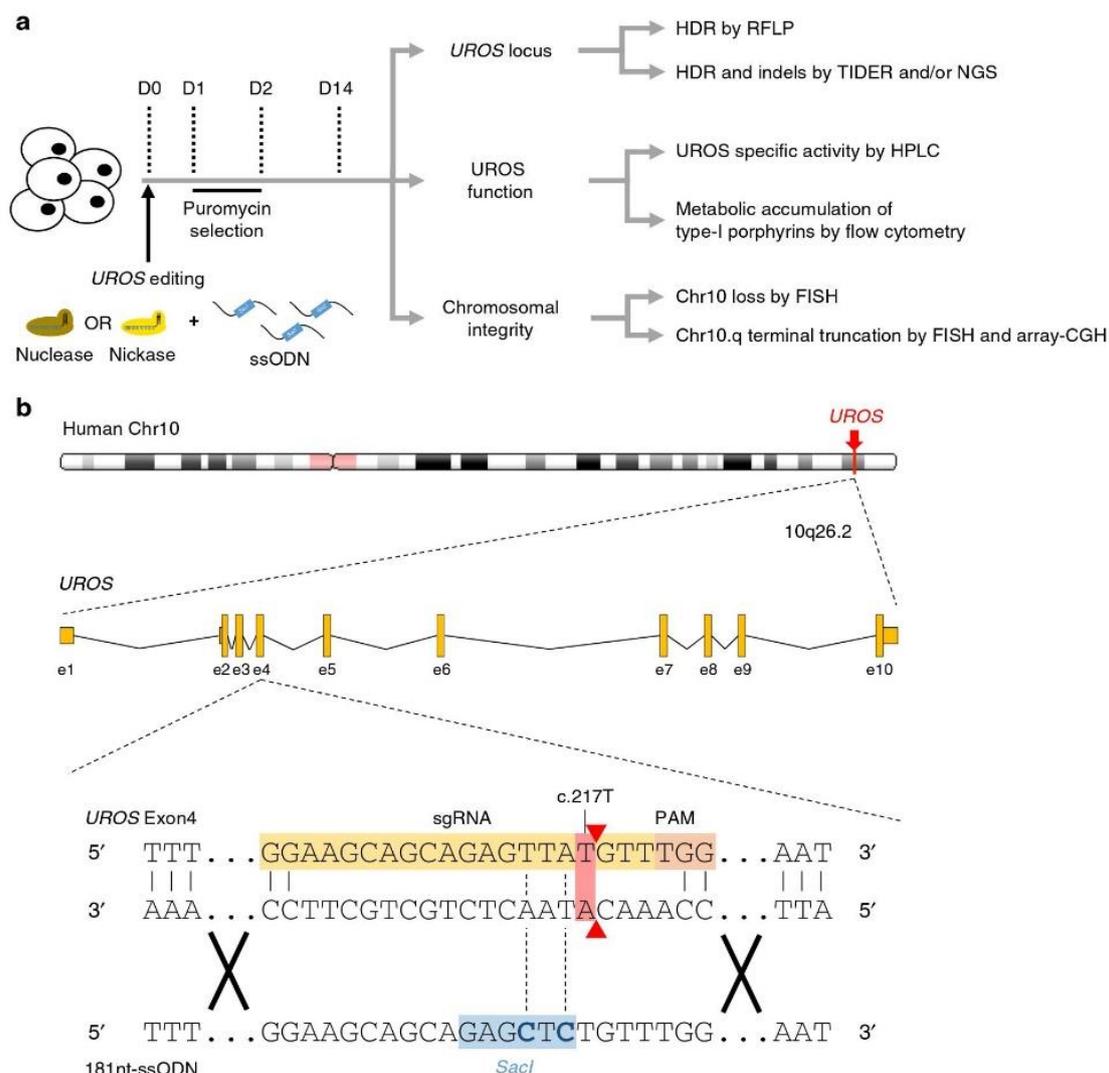


Fig. 1 *UROS* gene editing strategy and workflow analysis. **a** Experimental workflow for *UROS* gene editing and analysis of outcomes. Cells were nucleofected with the 181nt-ssODN template and either with nuclease or nickase followed by puromycin-positive selection. Then, (i) *UROS* locus was characterized by RFLP to quantify HDR and by TIDER or deep sequencing to evaluate indels and to confirm HDR percentage; (ii) *UROS* functionality was assessed by quantifying *UROS*-specific activity and type-I porphyrin accumulation, respectively determined by HPLC and flow cytometry; (iii) Chromosomal integrity was tested for Chr10 loss or Chr10q terminal deletion either by DNA-FISH assay or array-CGH. **b** (Top) Schematic *UROS* locus in chromosome 10 with *UROS* gene overview (middle). (Bottom) Detailed view of exon 4 region and CRISPR-mediated HDR design using a c.217T-targeting sgRNA (highlighted in orange) with adjacent PAM and an 181nt-ssODN carrying a silent *SacI* restriction site (highlighted in blue) close to c.217 T position. Red arrows indicate expected cleavage site using nuclease. Chr chromosome, CGH comparative genomic hybridization, D day, e exon, HDR homology-directed repair, HPLC high performance liquid chromatography, NGS (next-generation sequencing), RFLP (restriction fragment length polymorphism), PAM protospacer adjacent motif, sgRNA single guide RNA, TIDER (tracking of insertions, deletions and recombination events)

whether chromosome 10 (Chr10) could be lost following a single DSB at *UROS* locus. To address this question, we performed fluorescent in situ hybridization (FISH) on HEK293T cells with a specific Chr10 centromere probe. We observed that HEK293T cells had mainly three Chr10 (69%), 25% had four Chr10 and 6% had only two Chr10 (Supplementary Figure 3b). Due to this chromosomal complexity, we analyzed blind a high number of interphase nuclei (at least 599 interphase nuclei per condition). Compared to non-transfected HEK cells, we did not observe any significant chromosomal loss of Chr10 in HEK293T cells transfected with nuclease co-delivered or not with ssODN (Supplementary Figure 3b). To evaluate the genome integrity around the *UROS* locus (locus 10q26.2), we performed another FISH analysis using probes framing *UROS* (a proximal probe at 4.6 Mb upstream *UROS*, labeled in green, and a distal

probe 4.4 Mb downstream *UROS*, labeled in orange). Enlarged mapping of the location of probes is illustrated in Supplementary Figure 3a. Due to the chromosomal complexity of HEK293T cells observed, we decided to focus on cells apparently trisomic for Chr10, i.e., displaying three green signals (3G profile) of the proximal probe. In these cells, we then counted the number of orange signals (distal probe) to evaluate 3 O/3 G or 2 O/3 G profiles, testing the absence or presence of distal deletion at *UROS* locus, respectively. In HEK293T cells transfected with nuclease, we observed a significant 10% increase in cells displaying a 2 O/3 G profile compared to non-transfected (NT) cells (17% of cells transfected with nuclease with a 2 O/3 G profile compared to 7% of NT cells with a 2 O/3 G profile, $p < 0.05$) (Fig. 3a). A similar result was observed in HEK293T cells transfected with nuclease and ssODN (18% of cells with a 2 O/3 G profile). These results

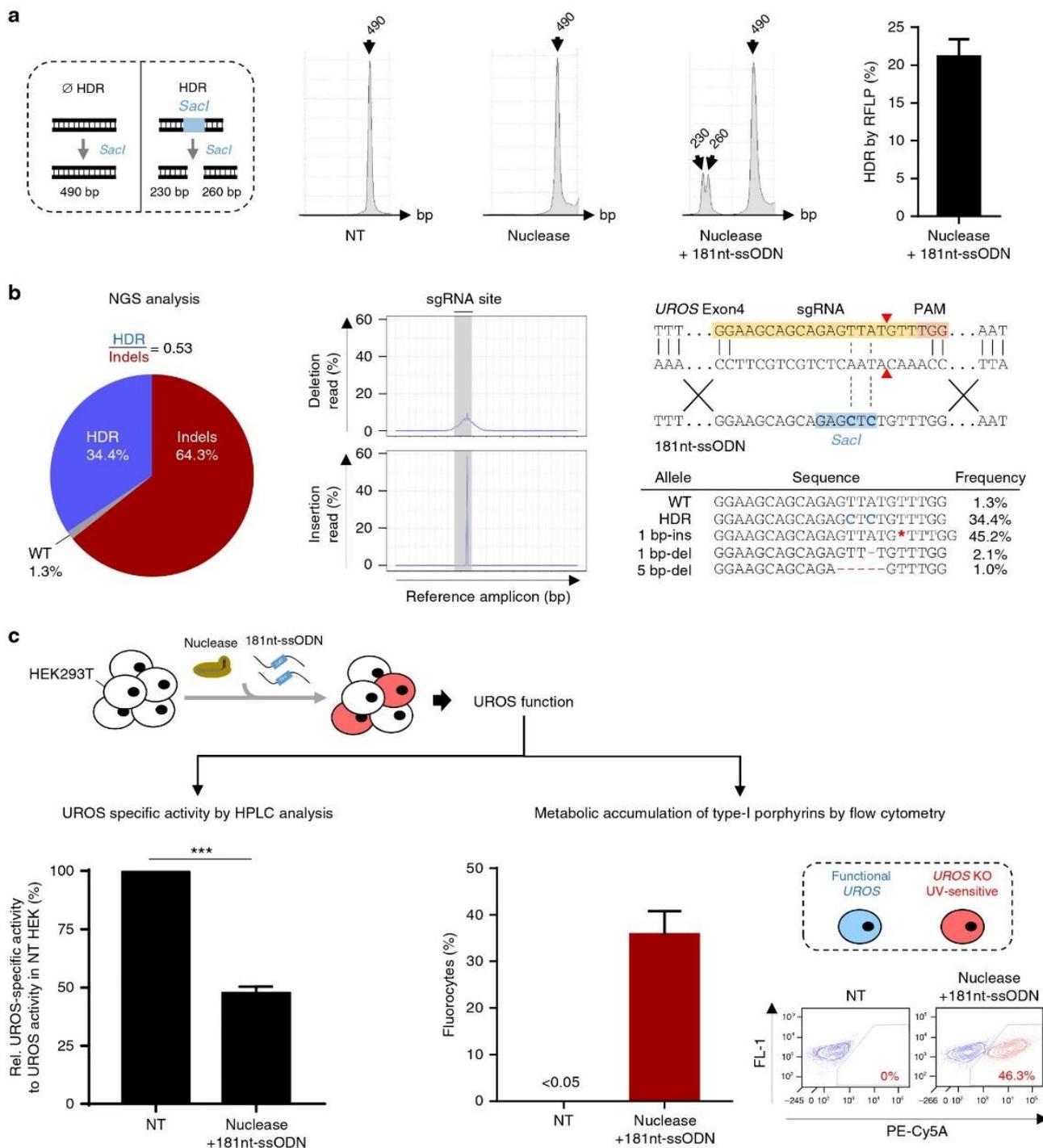
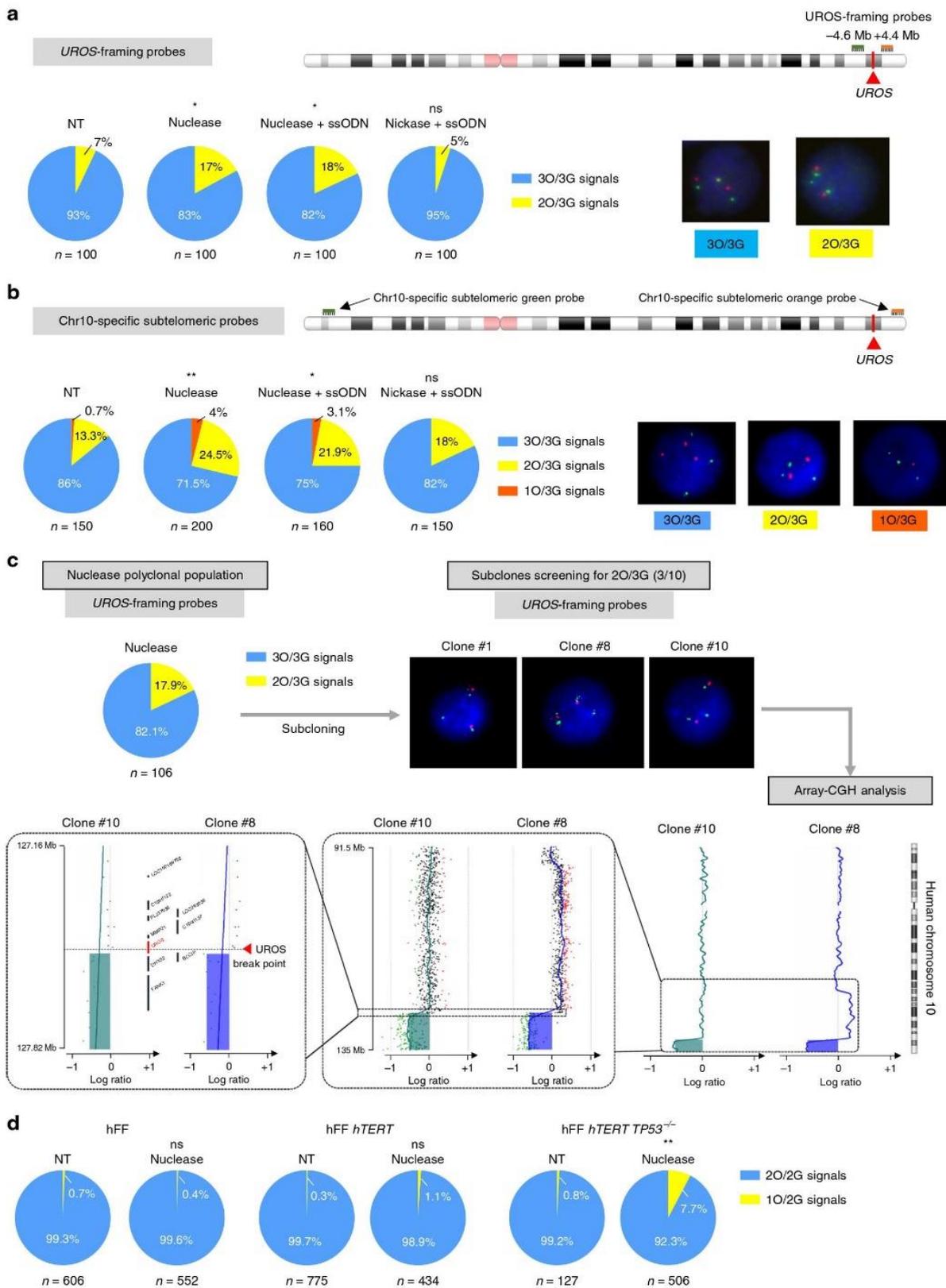


Fig. 2 Nuclease-mediated HDR is associated with predominant indels leading to impaired UROS functionality. **a** (Left) Scheme of *Sacl*I-digested PCR products obtained for alleles with or without HDR. (Center) Illustrative RFLP analysis of non-transfected HEK293T cells (NT) or transfected with nuclease only or co-delivered with the 181nt-ssODN template. (Right) HDR frequency induced by nuclease and a 181nt-ssODN ($n = 5$). **b** (Left) NGS analysis of allelic outcomes and associated HDR/indel ratio following transfection of HEK293T cells with nuclease and 181nt-ssODN. (Center) Frequencies of reads carrying either insertion or deletion. Region spanning sgRNA sequence is highlighted in grey. (Right) Most common observed alleles (with frequencies $\geq 1\%$) aligned on the sgRNA sequence. HDR modification in blue and indels in red. **c** (Left) Quantification of UROS-specific activity from HEK293T cells NT or transfected with nuclease and 181nt-ssODN ($n = 3$). Values are normalized with NT cells. (Right) Fluorocyte frequencies and illustrative flow cytometry results from NT or HEK293T cells transfected with nuclease and a 181nt-ssODN ($n = 5$). Blue and red dots (and associated percentages) depict non-fluorescent cells and fluorocytes, respectively, with type-I porphyrin accumulation. Results are presented as mean \pm SEM. The data are from independent experiments. Statistical significance is inferred on raw data using two-tailed unpaired *t* test for UROS-specific activity. *** $p < 0.001$. For **a**, **c**, source data are provided as a Source data file



could be consistent with a terminal deletion of Chr10 downstream of the *UROS* locus induced by nuclease-mediated DSB.

To confirm this hypothesis, we next hybridized the cells with one Chr10q sub-telomeric orange probe (Supplementary Figure 3a). We confirmed the complex pattern of Chr10 in HEK293T as revealed by a large analysis of a higher percentage

of cells with only two Chr10q probes compared to non-transfected cells (9% in NT versus 22.6% and 20.1% in cells transfected with nuclease only or nuclease with ssODN, respectively, Supplementary Figure 4a). As previously, we then focused on HEK293T cells apparently trisomic for Chr10. To do so, we used another couple of probes: sub-telomeric probes

Fig. 3 Unique DSB impairs chromosomal integrity. **a**, **b** DNA-FISH assay using (a) *UROS*-framing probes (respectively -4.6 Mb upstream and $+4.4$ Mb downstream from *UROS* locus) or (b) Chr10q-specific sub-telomeric probes for NT HEK293T cells or cells transfected with nuclease only, nuclease + ssODN or nickase + ssODN. For the two couples of probes, green (G) and orange (O) fluorescent probes are respectively upstream and downstream of the *UROS* gene. In this way, Chr10q terminal truncation is denoted by loss of orange signals. (Left) quantification of cells with 3 O/3 G, 2 O/3 G or 1 O/3 G signals. (Right) Illustrative DNA-FISH results for HEK293T. Statistical significance is inferred using two-sided chi-square test (versus NT cells). ns, not significant; * $p < 0.05$ ** $p < 0.01$. **c** Array-CGH on truncated clones. First, HEK293T were transfected with nuclease and analyzed by FISH using *UROS*-framing probes. Transfected cells were subcloned and 3 out of 10 clones were identified as 2 O/3 G. Chromosome 10 integrity of clones #8 and #10 was evaluated by array-CGH. Deletion in #10: arr[GRCh37] 10q26.2q26.3(127516127_135404523)x2. Duplication and deletion in clone #8. arr [GRCh37] 10q24.1q26.2(95667790_127496056)x3-4,10q26.2q26.3(127516127_135404523)x2. **d** DNA-FISH assay using *UROS*-framing probes for primary wild-type fresh hFF (human Foreskin Fibroblasts) (hFF), immortalized with *hTERT* (hFF *hTERT*) or *TP53*^{-/-} immortalized fibroblasts (hFF *hTERT* *TP53*^{-/-}), NT or transfected with nuclease. Quantification of 2 O/2 G and 1 O/2 G signals percentages. For (a, b, c), source data are provided as a Source data file

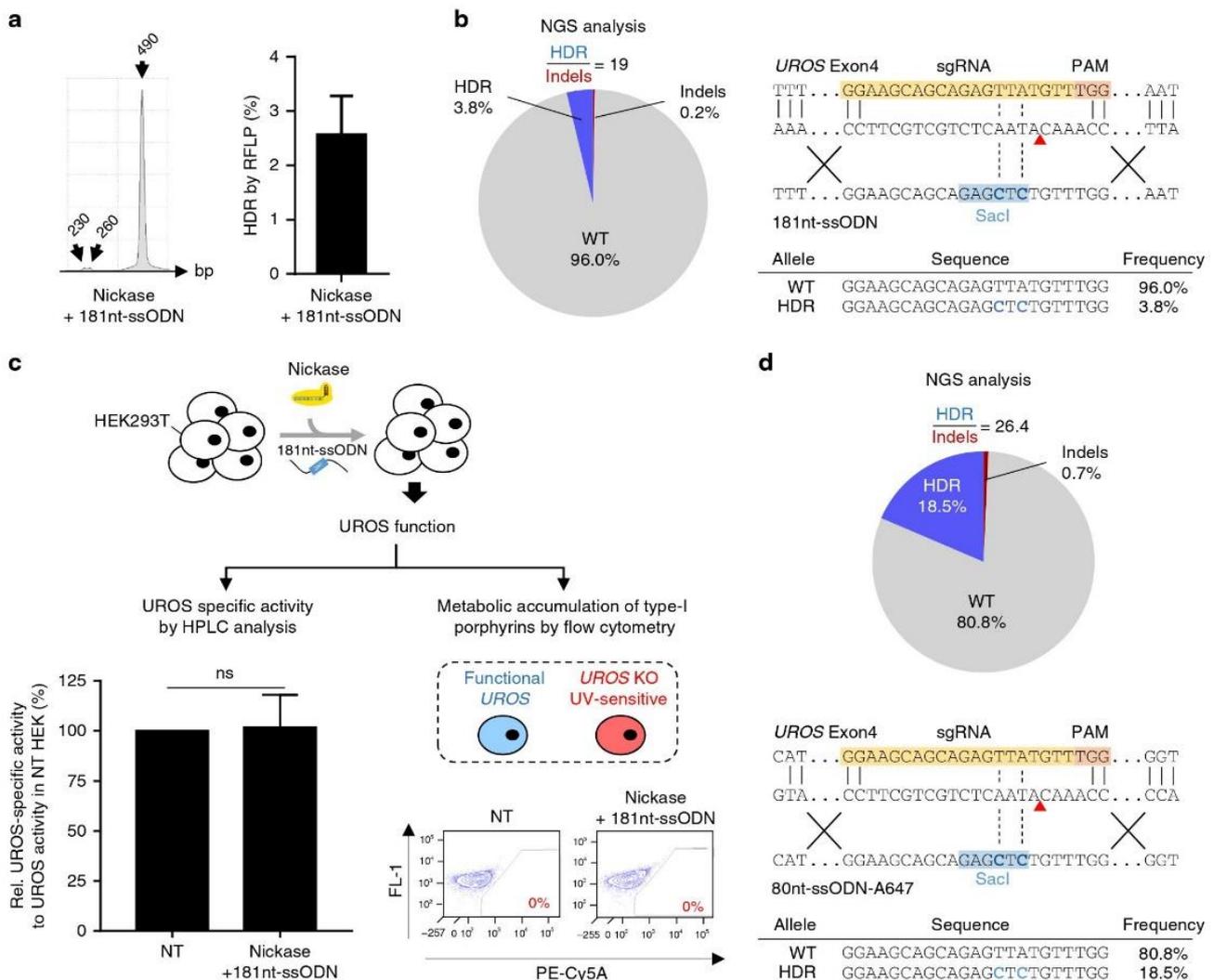


Fig. 4 Single nickase-mediated HDR is associated with minimal indels leading to preserved *UROS* functionality. **a** (Left) Illustrative RFLP analysis and (Right) HDR frequency for HEK293T cells transfected with nickase and a 181nt-ssODN ($n = 4$). **b** (Left) NGS analysis of allelic outcomes and associated HDR/indel ratio following transfection of HEK293T cells with nickase and a 181nt-ssODN. (Right) Most common observed alleles (with frequencies $\geq 1\%$) aligned on the sgRNA sequence. HDR events in blue. **c** (Left) Quantification of *UROS*-specific activity from NT HEK293T cells or transfected with nickase and a 181nt-ssODN ($n = 3$). Values are normalized with NT cells. (Right) Illustrative flow cytometry results from NT cells or transfected with nickase and a 181nt-ssODN ($n = 3$). **d** (Top) NGS analysis of allelic outcomes and associated HDR/indel ratio following transfection of HEK293T cells with nickase and optimized amount of a 80nt-ssODN-A647. (Bottom) Most common observed alleles (with frequencies $\geq 1\%$) aligned on the sgRNA sequence. HDR events in blue. Results are presented as mean \pm SEM. Data are from independent experiments. Statistical significance is inferred on raw data using two-tailed unpaired t-test for *UROS*-specific activity. ns, not significant. For (a, c), source data are provided as a Source data file

at Chr10p-arm (labeled in green) and another one at Chr10q-arm (labeled in orange) (Fig. 3b) and focused on cells with a FISH profile with three green signals (3 G). The percentage of cells with a 3 O/3 G profile significantly decreased from 86% to 71.5 and 75% for nuclease without or with ssODN respectively in favor of cells with only two or one orange signals (from 14 to 28.5% and 25% for nuclease without or with ssODN respectively, $p < 0.01$). The use of sub-telomeric probes therefore allowed us to confirm that Chr10q truncation occurred in about 10% of HEK293T cells transfected with nuclease, without or with ssODN, resulting in a 7.5 Mb loss downstream of *UROS* (Fig. 3b). To check whether megabase-scale deletions could also occur upstream of *UROS* DSB, we compared the 3 O/2 G signal frequency between HEK293T cells non-transfected or transfected with nuclease. We did not observe any difference in 3 O/2 G signal frequency in two independent experiments (3.7 vs. 3.4% in first experiment and 7.6 vs. 8.5% in the second one respectively for NT cells and transfected with nuclease), suggesting that large deletions are mostly unidirectional downstream of *UROS* DSB.

To confirm the loss of 10q arm extremity and map the break point, we picked up 10 single clones of HEK293T cells previously transfected with nuclease. Using *UROS*-framing probes, FISH analysis revealed that 3 out of 10 clones lost one distal probe. We then performed high resolution array-Comparative Genomic Hybridization (CGH) (60 kb resolution) on two positive clones. Results demonstrated chromosomal truncation in both clones starting exactly from *UROS* DSB until the last probe near 10q chromosome telomere (50 kb from telomere) (Fig. 3c). Moreover, we observed in clone #8, upstream of *UROS*, a 31 Mb duplication with 50% of mosaicism. These data highlight the high genomic instability at the DNA break point with megabase-scale deletion or gain.

Altogether, these results strongly suggest that the editing strategy using a single DSB in one chromosome may induce megabase-scale chromosomal damage. To explore whether a similar effect occurs in other cell types, we transfected the K-562 hematopoietic cell line with nuclease and analysed outcomes using *UROS*-framing probes. Similarly, we found that nuclease transfection induced a significant increase in cells with only one signal of the probe distal to *UROS* (+11%), suggesting that chromosomal terminal deletion induced by nuclease-mediated DSB is not cell type-dependent (Supplementary Figure 4b).

HEK293T and K562 cell lines have a very complex and unstable karyotype and are deficient in p53 activity. To explore whether chromosome truncation could occur in primary cells, we repeated same experiments in primary human foreskin fibroblasts (hFF). Deep analysis of FISH assay using *UROS*-framing probes fortunately revealed that the percentage of 1 O/2 G cells does not increase after nuclease in fresh and *hTERT* (human telomerase reverse transcriptase) immortalized fibroblasts (Fig. 3d). Because FISH limit of detection is about 1%, we cannot exclude rare terminal truncations after nuclease transfection. We decided to use *hTERT* immortalized fibroblasts invalidated or not for the *TP53* gene. In contrast to *hTERT* immortalization alone, *TP53* knockout dramatically increased the risk, up to 10-fold (Fig. 3d). We demonstrate the strong involvement of p53 in Cas9 nuclease-induced chromosomal instability and highlight CRISPR-Cas9 nuclease genotoxicity in primary cells with inactivated p53.

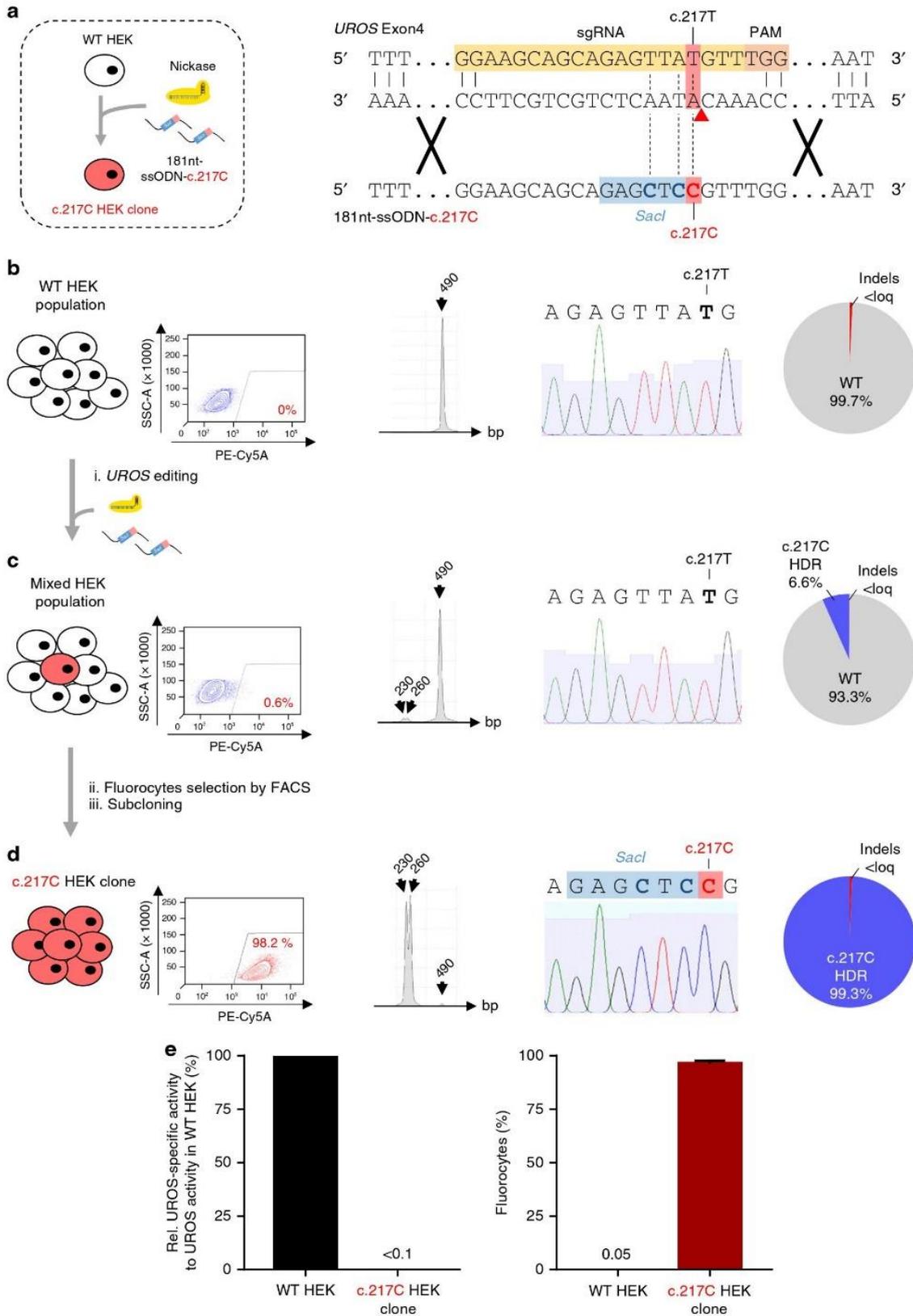
Cas9^{D10A} Nickase use prevents on- and off-target indels. In view of the concerns about the Cas9-nuclease system, we wondered whether it was possible to abolish indels due to NHEJ and terminal chromosomal deletion by switching from DSBs to DNA single-strand breaks (SSBs). We switched to Cas9^{D10A}-nickase (nickase) with the same gRNA and template. We obtained a

modest but reproducible HDR rate as measured by restriction fragment length polymorphism (RFLP) ($2.55\% \pm 0.73$, $n = 4$) (Fig. 4a). PCR allelic analysis confirmed the HDR rate (2/38, 5.3%) and revealed the absence of alleles with on-target indels (0/38, Supplementary Figure 5 left). In parallel, we sorted 34 cellular clones and sequenced them. We observed that cells were either heterozygous (2/34) or not edited, but never with indels (Supplementary Figure 5 right). To increase sensitivity, we performed TIDER (Supplementary Figure 2b) and NGS analysis (Fig. 4b). Analyses found respectively 3.0% and 3.8% of HDR with only 0.2% of on-target indels with nickase compared to 64% with nuclease. This is therefore a very high HDR/indels ratio (19 vs 0.53 with nuclease). In contrast to nuclease, *UROS* enzymatic activity was maintained ($101.8 \pm 16\%$, $n = 3$, Fig. 4c left) without metabolic dysfunction (no fluorocytes, Fig. 4c right), demonstrating that the nickase approach does not impair gene function. Importantly, we did not detect any off-target indels in the predicted top 10 off-target sequences (Supplementary Table 1).

We hypothesized that the ssODN template concentration was a critical limiting factor for HDR efficiency. For this reason, we carried out ssODN dose-scaling and found that HDR frequency reached 5.1% at the highest concentration (2-fold higher with 5 μM vs 0.5 μM , $p < 0.01$, Supplementary Figure 6a). We then retained 5 μM as the optimal ssODN concentration for further experiments. To evaluate whether a shorter ssODN would facilitate ssODN transfection, we designed an 80nt-ssODN template. The HDR rate was similar but not improved (Supplementary Figure 6b). To monitor ssODN transfection efficacy, we used a ssODN coupled to an Alexa-647 fluorochrome (80nt-ssODN-A647). Surprisingly, we observed an unexpected increase in HDR rate without sorting A647-positive cells (+382%, compared to non-optimized condition, Supplementary Figure 6b).

We hypothesized that fluorochrome could stabilize the ssODN template and improve HDR. To confirm this hypothesis, we added 3 locked nucleic acid (LNA) bound in 5' of the same 80nt-ssODN template and observed a similar increase in the HDR rate to the 5' Alexa-647 chemical graft (+325%, compared to non-optimized condition, Supplementary Figure 6b). With all these improvements (high concentration of stabilized ssODN), TIDER analysis validated the high frequency of *SacI* (12.2% of alleles) and with a very low level of indels (1.2%) (Supplementary Figure 2c). NGS confirmed the 18% HDR rate and, most importantly, still with a minimal rate of indels (0.7%, Fig. 4d). Moreover, due to the minimal indels, other alleles were not modified (80.8% are WT vs 1.3% with nuclease).

To test whether single nickase could be used to edit other cell types, we transfected the K-562 cell line with the same optimized tools targeting *UROS* exon 4. We obtained a high rate of HDR ($9.47 \pm 2\%$, $n = 6$) measured by RFLP and confirmed by NGS analysis (8.5%, Supplementary Figure 7a, b). Again, a high HDR/indels ratio up to 14.1 with only 0.6% of indels and no *UROS* dysfunction (absence of fluorocytes, Supplementary Figure 7c) were observed. To confirm that this was not specific to exon 4, we targeted exon 10 of *UROS* near to the frequent c.683 C > T point mutation (p.Thr228Met)¹⁸ and transfected a 75nt-ssODN-A647 template together with the nickase (Supplementary Figure 8a). Similarly, we confirmed the editing by RFLP (Supplementary Figure 8b) without any *UROS* dysfunction (absence of fluorocytes detected by flow cytometry (Supplementary Figure 8c)). These data demonstrate that the nickase strategy can be applied to other cell types and loci. Moreover, unlike nuclease, nickase allows editing and leaves the other alleles non-edited. The absence of indels made it possible to perform iterative editing to reach higher HDR rates. With successive transfections, we dramatically improved the HDR rate at two loci and in different cell types (Supplementary Figure 9).



Cas9^{D10A} nickase use prevents chromosomal truncation. To explore whether SSB can also avoid karyotype abnormalities, we first performed DNA-FISH using a centromeric probe. As expected, no chromosomal loss was observed (Supplementary Figure 3b). Then we used the two *UROS*-framing probes. To be

highlighted, after nickase and ssODN template transfection, the percentage of cells with 2 O/3 G signal was significantly lower than with nuclease ssODN (5 and 17% respectively, Fig. 3a, $p < 0.01$) and similar to non-transfected cells. Sub-telomeric 10q FISH analysis of transfected cells with nickase showed

Fig. 5 Single nickase-mediated gene editing results in c.217 C clone for CEP disease modeling **a** (Left) Scheme of gene editing approach to convert wild-type HEK293T (WT HEK) into homozygous c.217 C HEK clone using nickase and a 181nt-ssODN carrying c.217 C mutation (called 181nt-ssODN-c.217 C). (Right) Detailed view of exon 4 region and CRISPR-mediated HDR design using a c.217T-targeting sgRNA and a 181nt-ssODN-c.217 C carrying c.217 C mutation (red) in addition to silent *SacI* restriction site (blue). Expected cleavage position using nickase is indicated with a red arrow. **b–d** (From left to right) Illustrative flow cytometry results for fluorocyte analysis, representative RFLP analysis, sequence spanning *UROS* exon 4 c.217 position obtained by Sanger sequencing and indels and HDR quantification by TIDER analysis (**b**) for WT HEK, (**c**) for cells transfected with nickase and a 181nt-ssODN-c.217 C (Mixed HEK population) and (**d**) for sorted and subcloned fluorocytes (PE-Cy5A-positive), called c.217 C HEK clone. Loq: limit of quantification. **e** Characterization of c.217 C HEK clone. *UROS* functionality assay with (Left) quantification of *UROS*-specific activity and (Right) fluorocyte frequencies from WT HEK or c.217 C HEK clone. Values for *UROS*-specific activity are normalized against WT HEK. Results are presented as mean \pm SEM. For (**e**), source data are provided as a Source data file

significantly fewer cells with 1 O/3 G or 2 O/3 G (18% vs 25% Fig. 3b) than with nuclease. Importantly, FISH revealed the same pattern as non-transfected cells. Therefore, unlike nuclease, single nickase leads to HDR with only rare indels and no detectable chromosomal truncations using FISH.

CEP modeling with single nickase approach. CRISPR-Cas9 is considered to be a powerful tool to model recessive genetic diseases. However, because of a high rate of indels concurrent to HDR, the probability of obtaining a mutant HEK293T homozygous for *UROS* c.217 T > C with a DSB approach is very low. In the light of our data, we tried a single nickase approach with a template containing the c.217 C mutation (181nt-ssODN-c.217 C) (Fig. 5a). RFLP, Sanger sequencing and TIDER analysis revealed that the editing event had occurred, leading to 6.6% of c.217 C HDR, without detectable indels and the appearance of 0.6% of fluorescent cells detected by flow cytometry (Fig. 5b compared to Fig. 5c). To investigate whether these fluorocytes resulted from the editing of the c.217 C mutation or from indels, we sorted the fluorescent cells. Thirty clonal allelic analyses confirmed that almost all the alleles (29/30) were edited with the c.217 T > C mutation and, importantly, without indels (Supplementary Figure 10). We then subcloned the sorted fluorescent cells and confirmed full editing in one clone with the T > C modification and *SacI* insertion (Fig. 5d). We obtained a fluorescent cell line homozygous for the c.217 C mutation (c.217 C HEK clone) with a strong decrease in *UROS* enzymatic activity (below 1% of the normal level) and therefore with fluorocytes due to the metabolic defect (Fig. 5e).

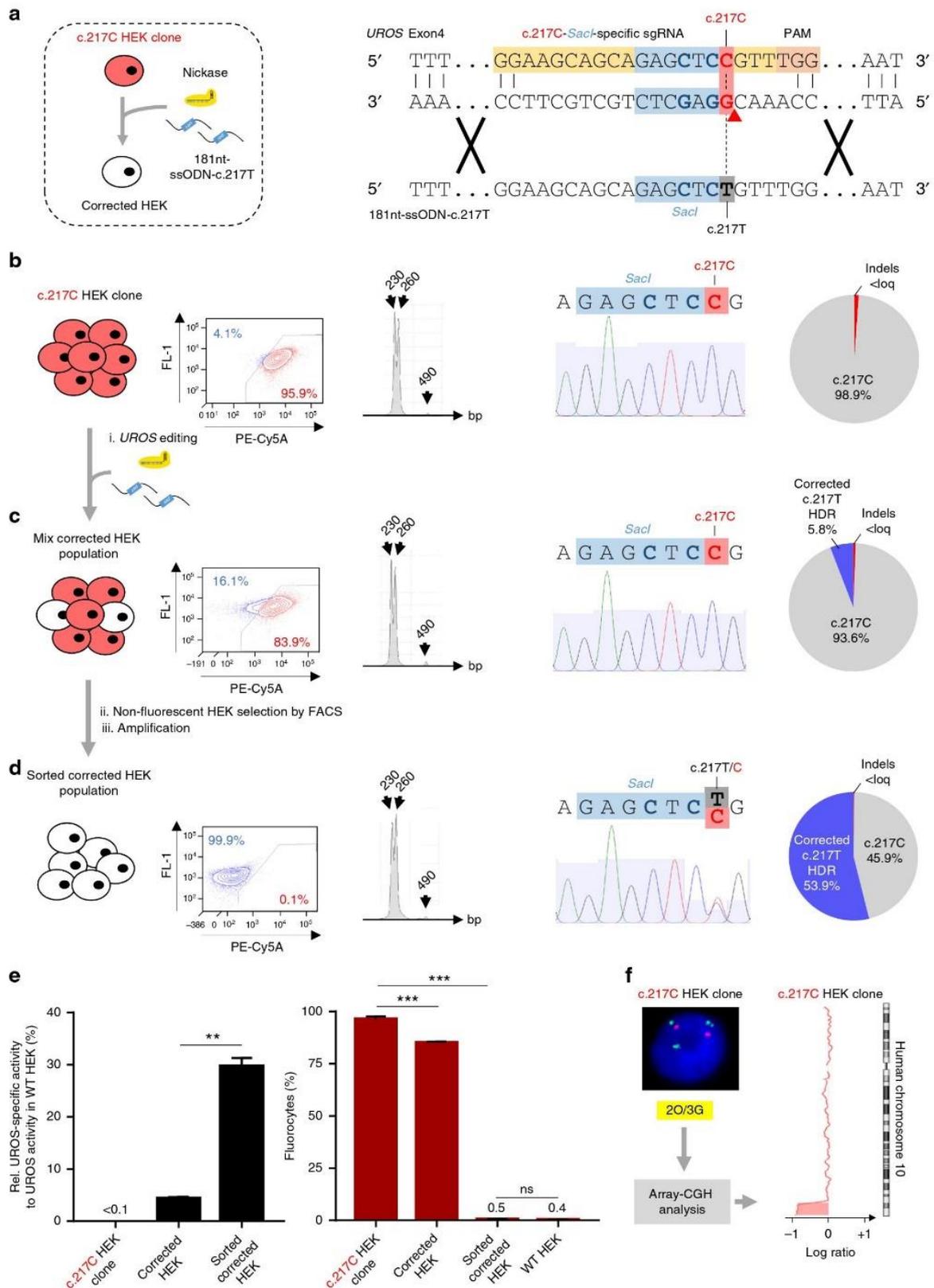
CEP gene correction. To test the ability of the single nickase approach to obtain genetic correction, we designed a novel sgRNA to specifically target the c.217 C mutant and *SacI* (c.217C-*SacI*-specific-sgRNA) (Fig. 6a). Next, we transfected the c.217 C HEK clone (Fig. 6b) with nickase and a 181nt-ssODN-c.217 T carrying the c.217 T correcting T-nucleotide. We obtained 5.8% of c.217 T HDR without indels, resulting in a slight decrease in fluorocyte percentage (Fig. 6c, e) that was correlated with a small rescue of *UROS* enzymatic activity (corrected HEK, 4.37% of WT *UROS* activity, in Fig. 6e). To enrich the corrected cell population, we sorted the non-fluorescent corrected cells (Fig. 6d). The presence of *SacI* as determined by RFLP and sequencing is proof that the non-fluorescent cells originated from cell sorting and not from contamination of the original WT cells (Fig. 6d). Sequencing showed c.217 C/T editing. About one-third of *UROS* activity was recovered in the sorted corrected cells (29.7% of WT *UROS* activity (Fig. 6e left panel), suggesting correction of 1 out of 3 alleles. This correction led to a full metabolic correction with the stable disappearance of porphyrins (at least up to one month) (Fig. 6e right panel). This is in agreement with published results showing that 10% of *UROS* activity is sufficient to avoid porphyrin accumulation *in vitro*²⁸ and *in vivo*²⁹. It is also in

agreement with clinical observations of the absence of phenotypes or minor phenotypes in persons with genotypes inducing residual *UROS* enzymatic activity³⁰. Using the single nickase approach, we obtained homozygous c.217 C CEP modeling (p.Cys73Arg) and a rescue of metabolic function. However, the intensity of Sanger sequencing peaks and TIDER demonstrated equal frequencies between c.217 T HDR and non-corrected c.217 C sequences and suggested that the corrected clone could be disomic and not trisomic for chromosome 10. To ensure that a terminal deletion had not occurred at the *UROS* locus, we performed FISH and array-CGH. Surprisingly, although there was no statistically significant difference in the 2 O/3 G FISH percentage between NT cells and cells transfected with nickase, we observed the loss of the 10q arm extremity starting at the *UROS* locus (Fig. 6f). This demonstrated that even with nickase, this event could rarely appear. We hypothesize that cell sorting based on the highest fluorescence could bias the results by enriching cells with chromosomal truncation (lower *UROS* activity after truncation than with mutation).

Discussion

Additive gene therapies are successful in treating monogenic hematopoietic disorders^{31–33}. However, transgenes semi-randomly integrate the hematopoietic stem cell (HSC) genome leading to a risk of genotoxicity. The aim of iPSC-based gene therapy is to control additive gene therapy by sorting cells with integration in a genomic safe harbor^{29,34}. Editing a gene at its endogenous locus by removing or correcting deleterious mutations rather than adding a new transgene has the potential to improve disease modeling and to solve insertional mutagenesis and non-physiological gene regulation problems^{3,35}. A high-fidelity HDR is a prerequisite for the safe use of CRISPR-Cas9-mediated gene therapy in the future and for obtaining reliably edited cell models of genetic diseases. Off-target modifications by the CRISPR-Cas9 first-generation nuclease have been widely reported^{4,5,36–39}. In our work, we confirm lack of specificity of the nuclease.

Unlike the off-target one, the on-target NHEJ activated in response to DNA DSBs has often been underestimated. Using the CEP model, we show that this is a major limiting factor for PGE. Indeed, flow cytometry analysis revealed an unacceptably high percentage of fluorocytes due to a profound deficiency of *UROS* enzymatic activity. NGS analysis demonstrated the predominant indels on-target caused by the NHEJ following nuclease-mediated gene editing. To improve the HDR/indel ratio, it is possible to insert a positive selection cassette in the donor template. However, that may cause genotoxicity by transgene integration and an immune response against the exogenous selection protein. Other techniques have been proposed: i) induce HDR activation using RAD51 agonist⁴⁰ or cell synchronization to promote HDR during S phase;⁴¹ ii) inhibit the NHEJ by DNA ligase IV^{42,43}, KU70⁴⁴, 53BP1 inhibitors⁴⁵ or downregulation of polymerase θ ^{46,47}. Nevertheless indels remain when using these pharmacological approaches, and inhibition of the genotoxicity DNA repair



pathways has to be evaluated on primary cells. Indeed, using a zinc-finger-mediated DSB approach to target *MLL*, Do et al.⁴⁸ reported that inhibition of the NHEJ by the DNA-PK inhibitor increases chromosomal translocations.

Our study reveals a safety concern regarding the medical perspectives of CRISPR-Cas9 editing. Indeed, using two different

cell lines and immortalized *TP53* KO fibroblasts, we observed a high frequency of chromosomal terminal deletions after only one nuclease-mediated DSB (10% in cancer cell lines and 7.7% in fibroblasts). This truncation is not due to a second cut that destabilizes 10q arm because none of the top-10 off-targets are in chromosome 10. This megabase-scale terminal deletion (7.5 Mb)

Fig. 6 Single nickase-mediated gene editing allows precise genetic and phenotypic correction. **a** (Left) Scheme of gene editing approach to modify the c.217 C HEK clone and turn it into genetically and phenotypically corrected HEK using nickase and a 181nt-ssODN carrying the c.217 T correcting mutation (called 181nt-ssODN-c.217 T). (Right) Detailed view of the c.217 C HEK clone containing c.217 C mutation (red) and *SacI* restriction site (blue). Nickase-mediated HDR design using a c.217C-*SacI*-specific sgRNA and a 181nt-ssODN-c.217 T carrying the c.217 T correcting mutation (grey) in addition to silent *SacI* restriction site (blue). Expected cleavage position using nickase is indicated with a red arrow. **b–d** (From left to right) Illustrative FACS (fluorescent activating cell sorting) results for fluorocyte analysis (PE-Cy5A-positive), representative RFLP analysis, sequence spanning *UROS* exon 4 c.217 position obtained by Sanger sequencing and indels and HDR quantification by TIDER analysis, **(b)** for the c.217 C HEK clone, **(c)** for cells transfected with nickase and 181nt-ssODN-c.217 T (Mix corrected HEK population), and **(d)** for PE-Cy5A-negative HEK293T cells sorted by FACS (called Sorted corrected HEK population). Loq: limit of quantification. **e** *UROS* functionality assays with (Left) quantification of *UROS*-specific activity ($n = 3$) and (Right) fluorocytes frequencies from the c.217 C HEK clone, corrected HEK population and sorted corrected HEK population ($n \geq 3$). Values for *UROS*-specific activity are normalized with WT HEK. Results are presented as mean \pm SEM. Data are from independent experiments. Statistical significance is inferred on raw data using two-tailed unpaired t-test for *UROS*-specific activity and paired one-way ANOVA for fluorocyte frequencies; ** $p < 0.01$ *** $p < 0.001$. **f** Chromosome 10 integrity of c.217 C HEK clone was analyzed by DNA-FISH using *UROS*-framing probes and by array-CGH. arr[GRCh37] 10q26.2q26.3 (127458901_135404523)x2. For **(e)**, source data are provided as a Source data file. ns, not significant

is larger than the kilobase-scale deletions recently described by Adikusuma et al. in mouse zygotes¹⁵. Array-CGH is required to observe these deletions because of the loss of primer sequences prevents PCR amplification. It confirms the full truncation of the 10q arm extremity. This damage caused by DSB could lead to severe outcomes in clinical applications. This truncation is critical not only for *UROS* activity but also for the elimination of multiple downstream genes and telomeres. Given the hypothesis that tumor-suppressor genes or genes involved in DNA repair are lost with the chromosome fragment, this could confer a selective advantage to damaged cells. In our example, 10q terminal deletion eliminated 43 genes (Supplementary Figure 11). Among them, five are proto-oncogenes and seven are tumor suppressors. Alarming, five of the tumor suppressors have been reported in leukemia (*BCCIP*⁴⁹, *PTPRE*⁵⁰, *MGMT*⁵¹, *BNIP3*⁵², *EBF3*⁵³). Therefore, before any clinical applications are envisaged, it will be mandatory to evaluate this oncogenic risk in primary hematopoietic cells to avoid grafting of leukemia-prone cells, as previously observed with additive gene therapy clinical trials using the first generation of oncoretroviral vector⁵⁴. Surprisingly, array-CGH revealed in one clone a mosaic duplication upstream of *UROS* DSB, which testifies to the complexity of DSB repair pathways. This unexpected rearrangement linking duplication with terminal deletion has already been described⁵⁵.

Fortunately, we did not observe any entire loss of chromosome 10. However, this risk has to be considered when two or more cuts by nuclease are needed. Indeed, a recent paper reported that only two cuts at the centromere (but not in the long arm) of chromosome Y induced its loss in mESC¹⁸. Two recent papers in the field of gene therapy recently proposed to excise alpha-globin enhancer⁵⁶ or a long fragment of 13.6kB with a putative gamma-delta intergenic fetal hemoglobin silencer using two cuts framing the locus⁵⁷. In both papers, they obtained deletion or inversion of the targeted region. In light of these data and others^{17,18}, a careful evaluation of chromosomal integrity will be required before these approaches can be considered as entirely safe.

Two recent papers highlight the adverse role of p53 for Cas9-mediated editing^{58,59}. They showed that CRISPR-Cas9 genome editing induced a transient p53-mediated DNA damage response and G1 cell-cycle arrest in immortalized retinal pigment endothelium cells and human pluripotent stem cells, respectively. Inhibition of p53 increases HDR efficiency and cell viability. Edited live cells are selected by their low-functional p53 pathway. This suggest that inhibition of p53 in edited cells could increase cell vulnerability to tumorigenic mutations or chromosomal rearrangements. In our study, we observed a serious side-effect of Cas9 nuclease, i.e. terminal chromosomal deletion in HEK293T and K-562 cell lines. These immortalized cells are known to have

their p53 function inhibited^{60,61}. This probably explains the high HDR rate obtained in our study and their vulnerability to chromosomal extremity losses. Chromosomal rearrangements in primary cells could be promoted depending on their p53 status. We demonstrate the strong involvement of p53 in chromosomal instability-induced CRISPR-Cas9 with a dramatic increase in truncation rate in *TP53* KO primary cells. Taken together, these data highlight the vulnerability of primary cells with p53 dysregulations. It will be critical to evaluate and monitor this risk in hematopoietic stem cells and embryos before pre-clinical applications of CRISPR-Cas9 nuclease.

In our laboratory, we have already used mouse models to demonstrate the efficacy of additive gene therapy in vitro and in vivo for CEP^{29,62,63}. The clinical severity of the disease and the lack of specific treatment, apart from bone marrow (BM) transplantation with an HLA-compatible donor, are strong arguments for gene therapy. DNA sequence damage induced by the NHEJ repair system and karyotype abnormalities using nuclease underline the need to use the CRISPR-Cas9 approach for precise genome editing with caution.

Nicks (SSBs) were long thought to undergo immediate ligation that prevented them from initiating editing. Many believed that nicks could only initiate recombination if they were converted to DSBs. We now have evidence of physiological nick-initiated HDR. For example, once the initial nick is generated, RAG1/2 (Recombination-activating gene) proteins are implicated in VDJ recombination in the adaptive immune system⁶⁴. The proof-of-principle that HDR can be obtained following DNA SSB was first obtained with zinc finger nuclease technology using a DNA donor template with 1–4% efficiency. However, it was 6-fold less efficient than with DSB⁶⁵. The structure of the exogenous repair donor seems to determine the HDR pathway. ssODN have attracted considerable interest as they are cheap to synthesize, easy to use, short-lived and cannot enter the genome. Recent publications^{22,66}, demonstrate that HDR by ssODN may proceed via an alternative repair pathway. It is currently thought that this process depends on whether ssODN is i) complementary to the nick and serves as a matrix to synthesize the edited strand, or ii) complementary to the intact strand of the target and directly fills the nick gap. In both cases, HDR efficiency seems to increase when RAD51 or BRCA2 are inhibited. Interestingly, the base excision repair pathway (BER) but not NHEJ repairs single nicks with high fidelity⁶⁷. The proof of concept with nickase was made more recently²² but still with low efficiency, thus limiting its interest for clinical applications. We demonstrate that a limiting factor for efficient HDR with SSB is the ssODN concentration and stabilization. Moreover, minimal indels allow iterative transfections to reach the same HDR efficiency as nuclease. This promising method that produces high PGE with minimal indels

will provide an important boost to CRISPR-Cas9-mediated gene therapy.

Using CEP disease as a proof of concept, we demonstrate that the single nickase method is the best for homozygous mutation modeling. It allows PGE at the locus with all regulatory elements of the promoters and an ideal physiological enzymatic rescue. Importantly, this approach is devoid of off-target indels and should be preferred to nuclease approaches. Moreover, it strongly reduced genotoxic terminal chromosomal deletion mediated by DSB in two cell lines. The next step will be to evaluate it for PGE in hematopoietic stem/progenitor cells in terms of feasibility and to measure the chromosomal abnormality rate as compared to nuclease. It could be helpful for the development of future clinical trials in hereditary hematopoietic disorders. The current findings pave the way for corrective gene therapy with single nickase, demonstrating it to be a safer alternative genome editing technology with high efficiency and better fidelity. However, truncation may rarely occur even with single nickase, as shown by our monocellular analysis. The low frequency of these rearrangements meant that it was very difficult to detect them in a polyclonal population. Alarmingly, both DSB and SSB may induce complex chromosomal rearrangements. Genome editing without any DNA break such as using CRISPR base editors could be a promising alternative.

Methods

Cell culture and transfection. Human embryonic kidney (HEK) cell line HEK293T (ATCC[®], Manassas, VA, USA) was maintained in Dulbecco's modified Eagle's medium (DMEM), low glucose (1 g/L-1), L-Glutamine (1 g/L-1) and pyruvate (Gibco[®] by Lifetechnologies[™], Carlsbad, CA, USA) supplemented with 10% fetal bovine serum, 100 U/mL penicillin, and 100 µg/mL streptomycin (all from Eurobio, Courtabouef, France). K-562 cell line (ATCC[®]) was maintained in RPMI Medium 1640, L-Glutamine, 25 mM HEPES (Gibco[®]) supplemented with 20% fetal bovine serum, GlutaMAX (Gibco[®]), 100 U/mL penicillin, and 100 µg/mL streptomycin. Human foreskin fibroblasts (HFF) were from a healthy person in accordance with the ethical standards of the committee responsible for human experimentation (Centre Hospitalier Universitaire de Bordeaux). Skin fragments were treated with trypsin-EDTA for 3 h at 37 °C and fibroblasts were isolated. *hTERT* immortalized fibroblasts were from ATCC[®] (CRL 4001, BJ-5ta), invalidated or not for *TP53* by CRISPR-Cas9. Fibroblasts were maintained in the same culture medium as HEK293T. Both cell lines were cultured at 37 °C, 5% CO₂ in a humidified chamber.

Editing tools. Cells were transfected by electroporation using the Nucleofector AMAXA electroporation system (Lonza[®], Bale, Switzerland). In brief, 10⁶ HEK293T (or K-562) cells were nucleofected with 200 ng (or 4 µg) of nuclease- or nickase-containing plasmid and 0.05 µM of specified ssODN. ssODN optimized concentration was used when specified and corresponded to 1.7 µM. Cells were then seeded onto 6-well plates (Corning[®], Tewksbury, MA, USA) and cultured as described above. Transfected cells were then positively selected 24 h after transfection either by addition of 1.25 µg mL⁻¹ puromycin (Sigma-Aldrich, Saint Louis MO, USA) for 36 h or by GFP-positive selection by Fluorescent activating cell sorting (FACS) using BD FACS Aria[®].

For fibroblasts, 4D-Nucleofector system was used with P3 Primary Cell Line 4D-Nucleofector[®] in association with CZ-167 program. In brief, 200,000 hFF were nucleofected with 16.9 µg Cas9 RNP and 5 µM of Alt-R[®] Cas9 Electroporation Enhancer. To form RNP, Cas9 protein was complexed to crRNA:tracrRNA according to the manufacturer's instructions. Then complexes were incubated for 20 min at room temperature before electroporation. Cas9 protein (Alt-R[®] S.p. Cas9 Nuclease V3), crRNA (Alt-R[®] CRISPR-Cas9 crRNA) were purchased from Integrated DNA Technologies.

Nuclease-containing plasmid was a modified version of pSpCas9n-2A-PuroR (pX462) obtained from Addgene (Cambridge, MA, USA) (#48141). First, to switch from Cas9D10A to Cas9, an insert containing the CBh promoter and 253 bp of Cas9 was obtained by *Xba*I and *Bgl*III digestion of pX330 (Addgene, #42230), and then ligated in pX462 digested with the same enzymes. Then, the puromycin-resistance gene (*pac*) was replaced by a human codon-optimized version (PuroR-HO). sgRNA was cloned into pUC19 obtained from Addgene (#50005) using the *Bbs*I restriction site and co-delivered at transfection with the nuclease-containing plasmid.

The nickase-containing plasmid was also a modified version of pX462 with a human codon-optimized version of the puromycin-resistance gene (PuroR-HO). The nickase-containing plasmid used for GFP-positive selection was pSpCas9n(BB)-2A-GFP (PX461), Addgene (#48140). In both nickase-containing plasmids,

sgRNA was directly added using the *Bbs*I restriction site according to the Feng Zhang protocol (Addgene). All sgRNAs were designed using the CHOPCHOP v2 algorithm⁶⁸ (chopchop.cbu.uib.no) and based on a unique sequence with 20 nucleotides. All ssODN templates used in the study were purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA). For 80nt-ssODN-A647 and 75nt-ssODN-A647, an Alexa Fluor[®] 647 was chemically linked at the 5' terminal end to ssODN by NHS Ester link. For 80nt-ssODN-LNA, three Locked Nucleic Acids were added at the 5' terminal end.

Flow cytometry for fluorocyte quantification and sorting. Fluorocyte quantification and sorting for disease modeling and correction was performed by flow cytometry. UV-sensitive type-I porphyrins were excited at 496 nm and the emitted wavelength was approximately 667 nm, detected by the PE-Cy5A PMT channel (FACSCanto, BD, Franklin Lakes, NJ, USA). FL-1 is a control green fluorescent channel used to exclude autofluorescent cells. Cells were sorted by BD FACS Aria[®].

RFLP for HDR quantification. Genomic DNA was extracted using Nucleospin[®] Tissue (Macherey-Nagel, Duren, Germany) according to the manufacturer's protocol. The genomic region flanking *UROS* exon 4 (or exon 10) was amplified by PCR (HotStarTaq Plus DNA polymerase, Qiagen[®], Venlo, Netherlands) with adequate primers (Supplementary Table 2). PCR products were purified with Nucleospin[®] Gel and PCR Clean-up (Macherey-Nagel) and digested with *Sac*I (or *Apal* I for exon 10 *UROS* analysis) restriction enzyme (New England Biolabs, Ipswich, MA, USA) for 1 h at 37 °C. Then, 5 ng digestion products were loaded into the Agilent[®] 2200 TapeStation (Santa Clara, CA, USA) capillary electrophoresis using D1000 ScreenTape and D1000 reagents according to the manufacturer's protocol. Quality control of enzymatic digestion efficiency is included in each assay.

Allelic and cellular clonal analysis. For allelic analysis, purified PCR products were sub-cloned into a "TOPO TA Cloning" vector (Invitrogen, Life Technologies[®] (Carlsbad, CA, USA) according to the manufacturer's protocol. Electro-competent bacteria (Stb3[™]) (Thermo Fisher Scientific, Waltham, Massachusetts, USA) were transformed with the TOPO and colonies were analyzed by Sanger sequencing after PCR. For cellular clonal analysis, transfected cells were isolated by the limiting dilution method and each clone was analyzed by Sanger sequencing after PCR. Sanger sequencing was done on purified PCR products and sequenced by LIGHTRUN (GATC Biotech, Konstanz, Germany).

UROS enzymatic activity and metabolic correction. UROS activity was determined by an enzyme-coupled assay as described previously⁶⁹. Briefly UROS activity was determined by an enzyme-coupled assay. For that, porphobilinogen was first converted to hydroxymethylbilane, the natural substrate for UROS, by hydroxymethylbilane synthase. Then, the uroporphyrinogen reaction products were oxidized to their respective uroporphyrin isomers, which were then resolved and quantitated by reversed-phase high-pressure liquid chromatography. One unit was defined as the amount of enzyme that formed 1 nmol of uroporphyrinogen III per hour at 37 °C.

TIDER analysis and NGS-Deep sequencing for allelic outcomes. Tracking of Insertions, Deletions and Recombination events (TIDER)²⁸ was used to determine HDR and indels frequencies. 490bp PCR product (carrying c.217 T) from non-transfected HEK293T was provided as control chromatogram. For Fig. 6, 490bp PCR product carrying the c.217 C mutation was used as control. Reference chromatogram with either *Sac*I-c.217 C (Fig. 5) or *Sac*I-c.217 T (Fig. 6 and Supplementary Figure 2) were obtained from allelic analysis (Supplementary Figure 5 and Supplementary Figure 10). Defaults parameters were conserved for analysis. The limit of quantification (loq) was defined at 2%.

Genomic DNA was extracted using Nucleospin[®] Tissue (Macherey-Nagel[®]) according to the manufacturer's protocol. The genomic region flanking *UROS* exon 4 was amplified by PCR (KAPA HiFi DNA polymerase, Kapa Biosystems[®], Cape Town, South Africa) with adequate primers (Supplementary Table 2). PCR products were purified with Nucleospin[®] Gel and PCR Clean-up (Macherey-Nagel[®]). To prepare sequencing libraries, the Illumina Nextera XT Kit (Illumina[®], San Diego, California, USA) was used and nested-PCR using Illumina primers was performed on purified PCR products. An Illumina MiSeq instrument (Illumina[®]) was used for high-throughput sequencing. The average depth of each genome analysis was 10,000. Quality of paired-end reads was checked with FastQC (Galaxy, <https://usegalaxy.org/>). Then, data were analyzed using the CRISPR Data Analysis and Visualization (CRISPR-DAV) pipeline⁷⁰.

Cytogenetic examination of chromosome 10. FISH was performed on interphase nuclei of HEK cells or K-562 cell line, with probes targeting the following regions on chromosome 10: centromeric region (XCE 10 probe, labeled in orange) (MetaSystems Probes, Altusheim, Germany), locus 10q26.11 (BAC RP11-79M19 probe, labeled in green) (Empire Genomics, Buffalo, NY, USA), locus 10q26.2 (BAC RP11-31M22 probe, labeled in orange) (Empire Genomics, Buffalo, NY, USA), sub-telomeric regions (p-arm and q-arm sub-telomere probes, respectively

labeled in green and orange) (Cytocell Ltd, Cambridge, UK). Preparations were pre-treated as indicated below. Briefly, the slides were successively immersed in a 2x saline-sodium citrate buffer for 10 min at 37 °C, in a 0.01% pepsin solution for 10 min at 37 °C, in a 1x phosphate-buffered saline (PBS) solution for 5 minutes, in a 3.7% formaldehyde solution for 5 min, and in a 1xPBS solution for 5 minutes. FISH probes and DNA were then co-denatured according to the manufacturers' protocols, and hybridization was performed overnight at 37 °C. The slides were then successively immersed in wash solutions and the nucleic acids were counterstained by 4,6-diamidino-2-phenylindole. The slides were then placed under an Axio Imager2 microscope with an epi-fluorescence source (Carl Zeiss AG, Oberkochen, Germany). The microscope was linked to the Metafer4 software for automated image acquisition and processing (MetaSystems GmbH, Altlußheim, Germany).

Genomic DNA was extracted with the Wizard Kit (Promega Corporation, Madison, USA) following the protocol validated in the laboratory.

Off-target analysis. For sgRNA targeting exon 4 *UROS* locus, the Top 10-predicted off-target sites, identified by CRISPOR software were amplified in genome-edited HEK293T and subjected to Sanger sequencing, followed by comparison to non-transfected cells by TIDER analysis.

Primers used for off-target analysis are in Supplementary Table 2.

Array-CGH. Array-CGH was performed on 8 × 60k oligonucleotide microarrays (Agilent Technologies, CA). DNA was labeled (cyanine 3 or cyanine 5) using the Genomic DNA ULS Labeling Kit from Agilent Technologies and hybridized onto the microarrays according to the manufacturer's instructions (Agilent). Scanning of the microarrays was performed using a G2565CA scanner (Agilent). Data analysis was carried out with Agilent Technologies software, namely Feature Extraction for Cytogenomics V5.0 for the fluorescence ratio calculation and Agilent CytoGenomics 3.0.1.1 to visualize chromosomal imbalances. Deletions and duplications in the heterozygous state were characterized by values of the log₂ ratio of fluorescence intensities (cyanine5/cyanine3) below -0.5 and above +0.3, respectively, with the statistical algorithm ADM2 used at a threshold of 5.

Statistics. Statistical significance was inferred when necessary. Graph Pad Prism 6 software was used for statistical analysis. Results are presented as mean ± SEM (standard error of the mean). The two-tailed unpaired *t* test was done to compare means of two groups. One-way ANOVA, completed with unprotected Fisher's Least Significant Difference test, was used to compare three groups. Two-tailed *Chi*² tests were used to compare percentages. All comparisons are shown with black bars. Null hypothesis was rejected when *p* value < 0.05. **p* < 0.05; ***p* < 0.01; ****p* < 0.001; ns, non-significant.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data Availability

All next-generation sequencing data sets have been deposited in the NCBI database under BioProject accession no. PRJNA521053 with associated BioSample no. SAMN10877893 for HEK293T NT and BioSample no. SAMN10877894, no. SAMN10877895 and no. SAMN10877896 for HEK293T respectively transfected with Nuclease + 181nt-ssODN (Fig. 2b), Nickase + 181nt-ssODN (Fig. 4b) and Nickase + 80nt-ssODN-A647 (Fig. 4d). Reads are available at Sequence Read Archive database (accession number no. SRR8534354, no. SRR8534355, no. SRR8534352 and no. SRR8534353 for HEK293T NT and HEK293T respectively transfected with Nuclease + 181nt-ssODN, Nickase + 181nt-ssODN and Nickase + 80nt-ssODN-A647). The data that support the findings of this study are available from the corresponding author upon reasonable request. Source data for bar charts and pie charts presented in Figs. 2, 3, 4, 6 and 6 as well as Supplementary Figures 3, 4, 6, 7, 8, 9 are provided with the paper.

Received: 28 August 2018 Accepted: 6 February 2019

Published online: 08 March 2019

References

- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
- Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
- Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).
- Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
- Zhang, F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* **29**, 149–153 (2011).
- Yang, Y. et al. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* **34**, 334–338 (2016).
- U.S. National Library of Medicine. *CPD-1 Knockout Engineered T Cells for Metastatic Non-small Cell Lung Cancer*. NCT02793856. <https://clinicaltrials.gov/> 2016.
- Phase I trial of NY-ESO-1 redirected CRISPR edited T cells (NYCE CELLS) engineered to express NY-ESO-1 TCR and gene edited to eliminate endogenous TCR and PD-1, Protocol #1604-1524. NIH OBA RAC Public Review. (2016). <https://clinicaltrials.gov/ct2/show/NCT03399448>.
- Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
- Maddalo, D. et al. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR-Cas9 system. *Nature* **516**, 423–427 (2014).
- Blasco, R. B. et al. Simple and rapid in vivo generation of chromosomal rearrangements using CRISPR-Cas9 technology. *Cell Rep.* **9**, 1219–1227 (2014).
- Choi, P. S. & Meyerson, M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nat. Commun.* **5**, 3728 (2014).
- Adikusuma, F. et al. Large deletions induced by Cas9 cleavage. *Nature* **560**, E8–E9 (2018).
- Xiao, A. et al. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res.* **41**, e141 (2013).
- Zuo, E. et al. CRISPR-Cas9-mediated targeted chromosome elimination. *Genome Biol.* **18**, 224 (2017).
- Adikusuma, F., Williams, N., Grutzner, F., Hughes, J. & Thomas, P. Targeted deletion of an entire chromosome using CRISPR-Cas9. *Mol. Ther.* **25**, 1736–1738 (2017).
- Anderson, K. E., Sassa, S., Bishop, D. F., Desnick, R. J. in: *The Metabolic and Molecular Basis of Inherited Disease* 8 edn., (eds Scriver, C. R., Beaudet, A. L., Sly, W. S.) 2991 (McGraw-Hill, New York, NY, 2001).
- Erwin, A. et al. (eds). *GeneReviews*[®] (University of Washington, Seattle, WA, 1993–2018).
- Ged et al. Congenital erythropoietic porphyria: mutation update and correlations between genotype and phenotype. *Cell. Mol. Biol.* **55**, 53–60 (2009).
- Davis, L. & Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc. Natl Acad. Sci. USA* **111**, E924–E932 (2014).
- Ramirez, C. L. et al. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res.* **40**, 5560–5568 (2012).
- Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
- Chen, J. S. et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
- Gao, Y. et al. Single Cas9 nickase induced generation of NRAMP1 knockin cattle with reduced off-target effects. *Genome Biol.* **18**, 13 (2017).
- Shao, Y. et al. Cas9-nickase-mediated genome editing corrects hereditary tyrosinemia in rats. *J. Biol. Chem.* **293**, 6883–6892 (2018).
- Brinkman, E. K. et al. Easy quantification of template-directed CRISPR-Cas9 editing. *Nucleic Acids Res.* **46**, e58 (2018).
- Bedel, A. et al. Metabolic correction of congenital erythropoietic porphyria with iPSCs free of reprogramming factors. *Am. J. Hum. Genet.* **91**, 109–121 (2012).
- Shady, A. A. et al. Congenital erythropoietic porphyria: identification and expression of eight novel mutations in the uroporphyrinogen III synthase gene. *Br. J. Haematol.* **117**, 980–987 (2002).
- Fischer, A., Notarangelo, L. D., Neven, B., Cavazzana, M. & Puck, J. M. Severe combined immunodeficiencies and related disorders. *Nat. Rev. Dis. Prim.* **1**, 15061 (2015).
- Ribeil, J.-A. et al. Gene Therapy in a Patient with Sickle Cell Disease. *N. Engl. J. Med.* **376**, 848–855 (2017).
- Morris, E. C. et al. Gene therapy for Wiskott-Aldrich syndrome in a severely affected adult. *Blood* **130**, 1327–1335 (2017).
- Papapetrou, E. P. et al. Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 73–78 (2011).

35. Kohn, D. B. & Kuo, C. Y. New frontiers in the therapy of primary immunodeficiency: From gene addition to gene editing. *J. Allergy Clin. Immunol.* **139**, 726–732 (2017).
36. Scott, D. A. & Zhang, F. Implications of human genetic variation in CRISPR-based therapeutic genome editing. *Nat. Med.* **23**, (2017).
37. Wang, Q. & Ui-Tei, K. Computational prediction of CRISPR-Cas9 target sites reveals potential off-target risks in human and mouse. *Methods Mol. Biol.* **1630**, 43–53 (2017).
38. Morgens, D. W. et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* **8**, 15178 (2017).
39. Zischewski, J., Fischer, R. & Bortesi, L. Detection of on-target and off-target mutations generated by CRISPR-Cas9 and other sequence-specific nucleases. *Biotechnol. Adv.* **35**, 95–104 (2017).
40. Song, J. et al. RS-1 enhances CRISPR-Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.* **7**, 10548 (2016).
41. Zhang, J.-P. et al. Efficient precise knockin with a double cut HDR donor after CRISPR-Cas9-mediated double-stranded DNA cleavage. *Genome Biol.* **18**, 35 (2017).
42. Maruyama, T. et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).
43. Hu, Z. et al. Ligase IV inhibitor SCR7 enhances gene editing directed by CRISPR-Cas9 and ssODN in human cancer cells. *Cell Biosci.* **8**, 12 (2018).
44. Chu, V. T. et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).
45. Canny, M. D. et al. Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat. Biotechnol.* **36**, 95–102 (2018).
46. Saito, S., Maeda, R. & Adachi, N. Dual loss of human POLQ and LIG4 abolishes random integration. *Nat. Commun.* **8**, 16112 (2017).
47. Zelensky, A. N., Schimmel, J., Kool, H., Kanaar, R. & Tijsterman, M. Inactivation of Pol θ and C-NHEJ eliminates off-target integration of exogenous DNA. *Nat. Commun.* **8**, 66 (2017).
48. Do, T. U., Ho, B., Shih, S.-J. & Vaughan, A. Zinc Finger Nuclease induced DNA double stranded breaks and rearrangements in MLL. *Mutat. Res.* **740**, 34–42 (2012).
49. Dong, C.-Y. et al. mda-7/IL-24 inhibits the proliferation of hematopoietic malignancies in vitro and in vivo. *Exp. Hematol.* **36**, 938–946 (2008).
50. Kabir, N. N., Rönstrand, L. & Kazi, J. U. Deregulation of protein phosphatase expression in acute myeloid leukemia. *Med. Oncol.* **30**, 517 (2013).
51. Hong, Q. et al. Association between the methylation status of the MGMT promoter in bone marrow specimens and chemotherapy outcomes of patients with acute myeloid leukemia. *Oncol. Lett.* **11**, 2851–2856 (2016).
52. Murai, M. et al. Aberrant DNA methylation associated with silencing BNIP3 gene expression in haematopoietic tumours. *Br. J. Cancer* **92**, 1165–1172 (2005).
53. Tao, Y.-F. et al. Early B-cell factor 3 (EBF3) is a novel tumor suppressor gene with promoter hypermethylation in pediatric acute myeloid leukemia. *J. Exp. Clin. Cancer Res.* **34**, 4 (2015).
54. Hacein-Bey-Abina, S. et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* **118**, 3132–3142 (2008).
55. Rowe, L. R. et al. U-type exchange is the most frequent mechanism for inverted duplication with terminal deletion rearrangements. *J. Med. Genet.* **46**, 694–70 (2009).
56. Mettananda, S. et al. Editing an α -globin enhancer in primary human hematopoietic stem cells as a treatment for β -thalassemia. *Nat. Commun.* **8**, 424 (2017).
57. Antoniani, C. et al. Induction of fetal hemoglobin synthesis by CRISPR-Cas9-mediated editing of the human β -globin locus. *Blood* **131**, 1960–1973 (2018).
58. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **23**, 927–930 (2018).
59. Ihry, R. J. et al. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).
60. Lin, Y.-C. et al. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5**, 4767 (2014).
61. Law, J. C., Ritke, M. K., Yalowich, J. C., Leder, G. H. & Ferrell, R. E. Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. *Leuk. Res.* **17**, 1045–1050 (1993).
62. Richard, E., Robert-Richard, E., Ged, C., Moreau-Gaudry, F. & de Verneuil, H. Erythropoietic porphyrias: animal models and update in gene-based therapies. *Curr. Gene Ther.* **8**, 176–186 (2008).
63. Robert-Richard, E. et al. Effective gene therapy of mice with congenital erythropoietic porphyria is facilitated by a survival advantage of corrected erythroid cells. *Am. J. Hum. Genet.* **82**, 113–124 (2008).
64. Grawunder, U. & Lieber, M. R. A complex of RAG-1 and RAG-2 proteins persists on DNA after single-strand cleavage at V(D)J recombination signal sequences. *Nucleic Acids Res.* **25**, 1375–1382 (1997).
65. Wang, J. et al. Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res.* **22**, 1316–1326 (2012).
66. Davis, L. & Maizels, N. Two Distinct Pathways Support Gene Correction by Single-Stranded Donors at DNA Nicks. *Cell Rep.* **17**, 1872–1881 (2016).
67. Dianov, G. L. & Hübscher, U. Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Res.* **41**, 3483–3490 (2013).
68. Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **44**, W272–W276 (2016).
69. Tsai, S. F., Bishop, D. F. & Desnick, R. J. Coupled-enzyme and direct assays for uroporphyrinogen III synthase activity in human erythrocytes and cultured lymphoblasts. Enzymatic diagnosis of heterozygotes and homozygotes with congenital erythropoietic porphyria. *Anal. Biochem.* **166**, 120–133 (1987).
70. Wang, X. et al. CRISPR-DAV: CRISPR NGS data analysis and visualization pipeline. *Bioinformatics* **33**, 3811–3812 (2017).

Acknowledgements

We thank Dr Feng Zhang for sharing plasmids through Addgene (Cambridge, MA, USA) and Vincent Pitard from the cytometry platform at Bordeaux University. This work was supported by the Association Française contre les Myopathies Grant # 20957, and the Agence de Biomédecine. We thank Florence Lichou (Inserm U1218) and Nicolas Sevenet (Institut Bergonié, Bordeaux, France) for NGS analysis, Julie Bouron for array-CGH analysis (Génétique médicale, CHU Bordeaux, France). Katia Obama and Isabelle Lamy are thanked for administrative assistance and financial management.

Author contributions

G.C., J.B., F.M.G. and A.Be.: design, analysis of data and drafting of manuscript; A.Be., F.M.G.: supervision, funding acquisition, final approval of manuscript; I.L.G.: FACS analysis; V.G.D., A.Bi: vector design and cloning; M.L.: UROS enzymatic assay, M.T. and E.Ro.: TP53 KO immortalized fibroblast line, V.P.M., B.T., C.G., J.M.B., E.Ri. and S.D.: helpful discussion in the field of CRISPR-Cas9, S.D.: Design and proofreading of manuscript, J.T., P.P.: FISH design and experiments, G.C., J.B., F.P.: Performing editing experiments, P.P. and C.R. array-CGH analysis.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-09006-2>.

Competing interests: The authors declare no competing interests.

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

Journal peer review information: *Nature Communications* thanks Hui Yang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

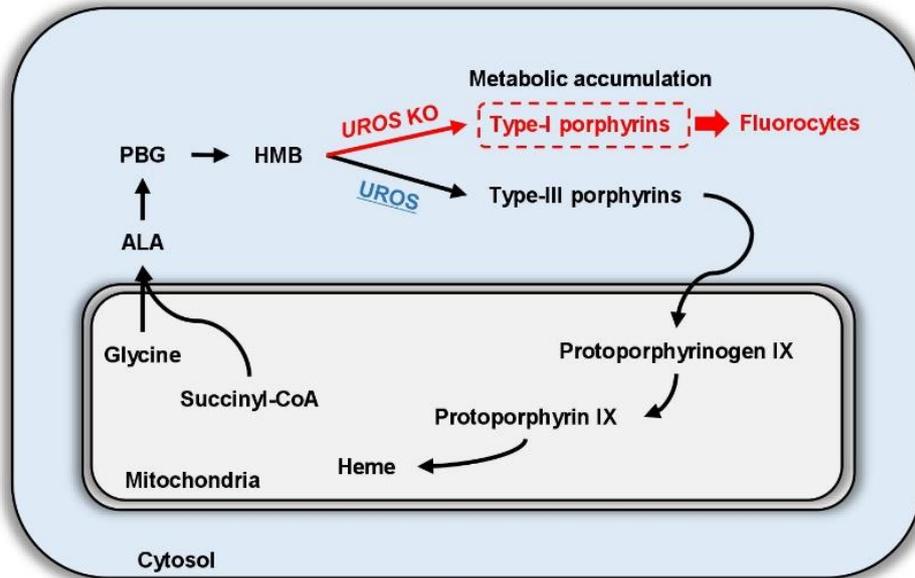
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

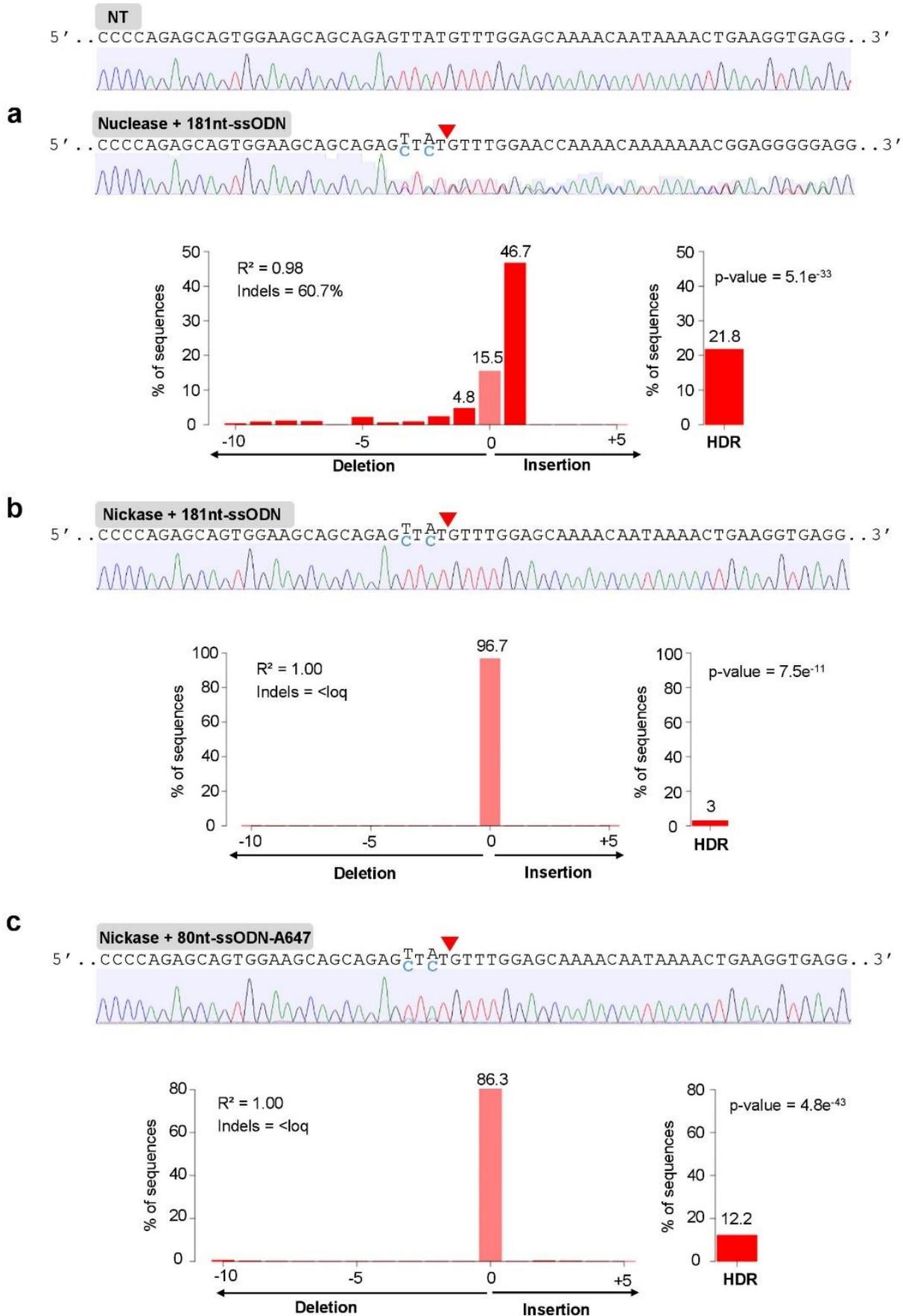
© The Author(s) 2019

Supplementary Figure 1



Supplementary Figure 1: Heme biosynthesis pathways is impaired in congenital erythropoietic porphyria. In physiological conditions, the uroporphyrinogen III synthase (UROS) catalytically cyclizes the hydroxymethylbilane (HMB) into type-III porphyrins leading to heme synthesis. In *UROS* KO cells as in congenital erythropoietic porphyria (CEP), HMB spontaneously cyclizes into type-I porphyrins which enter a metabolic impasse, accumulate and turn the cells into UV-sensitive cells called fluorocytes (PE-Cy5A-positive). CoA, Coenzyme A ; ALA, δ -aminolevulinic acid ; PBG, Porphobilinogen ; HMB, Hydroxymethylbilane.

Supplementary Figure 2

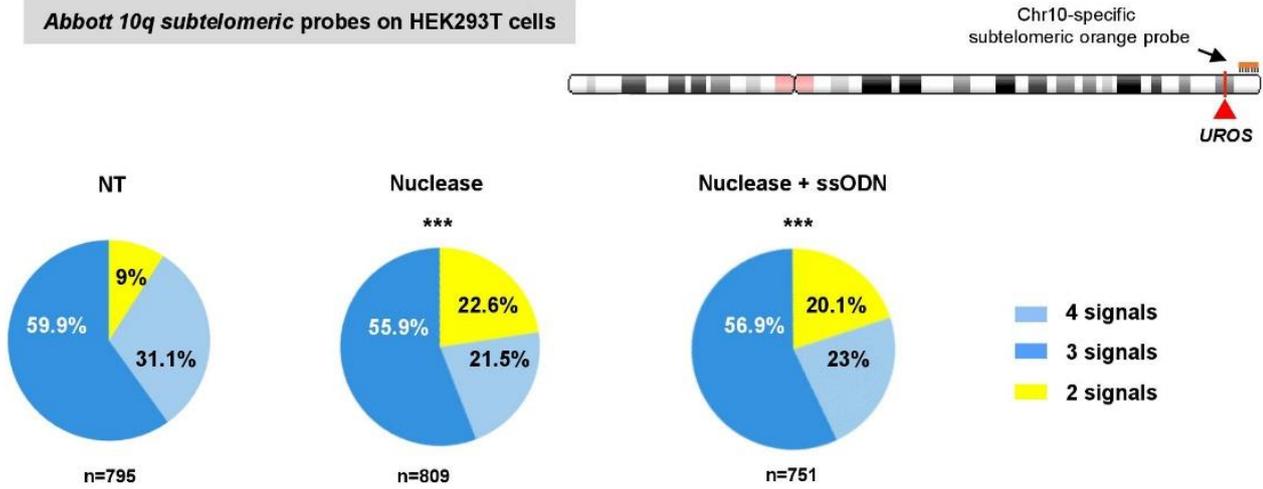


Supplementary Figure 2: TIDER analysis of gene editing. (Top) UROS exon 4 Sanger sequencing of non-transfected (NT) HEK293T used as the reference sequence for TIDER analysis. a,b,c, HDR and indel quantification for cells transfected a, with nuclease and a 181nt-ssODN template, b, with nickase and a 181nt-ssODN template and c, with nickase and optimized amount of a 80nt-ssODN-A647 template. Expected cleavage position is indicated with a red arrow. HDR events in blue.

Supplementary Figure 4

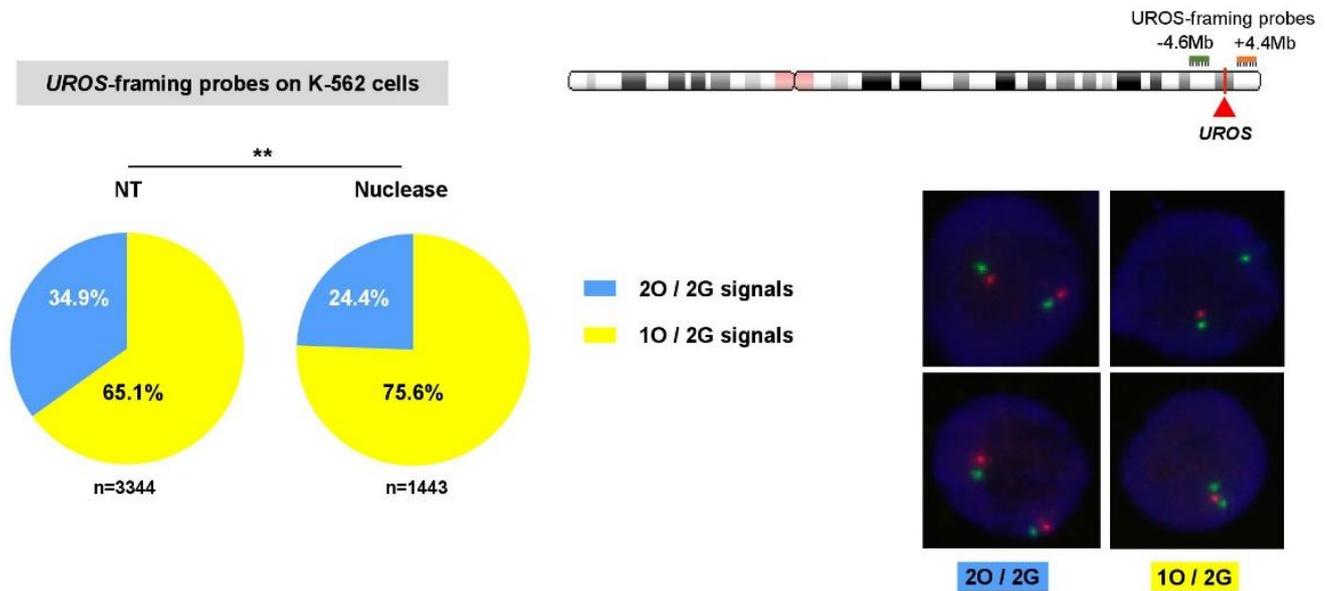
a

Abbott 10q subtelomeric probes on HEK293T cells



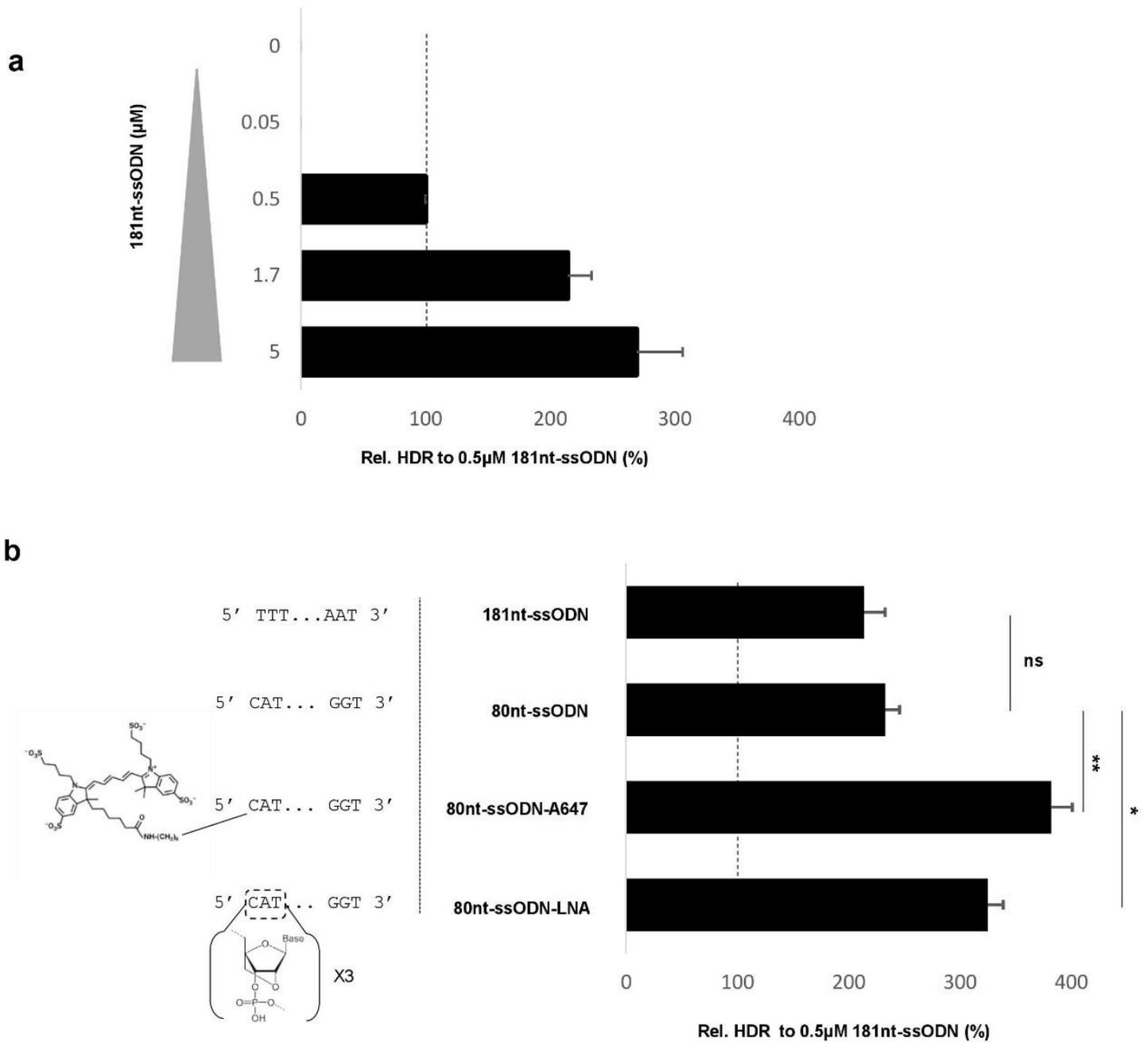
b

UROS-framing probes on K-562 cells



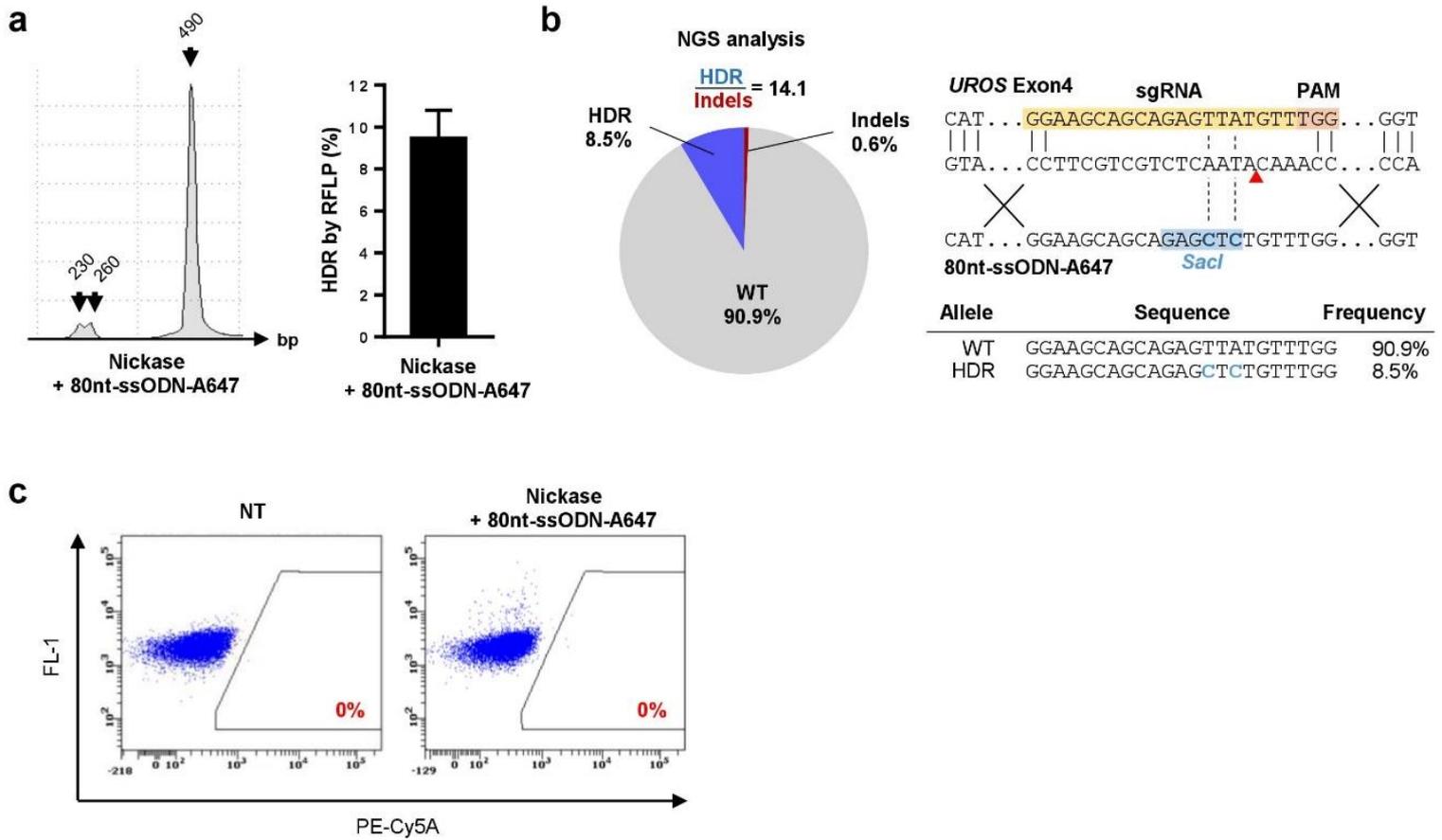
Supplementary Figure 4: Nuclease-mediated gene editing impairs chromosomal integrity in HEK293T and K-562 cells. **a**, Quantification of cells with 2, 3 or 4 orange signals by DNA-FISH assay using Abbott 10q sub-telomeric orange probe in HEK293T cells. **b**, Loss of orange signals is quantified by DNA-FISH assay using *UROS*-framing probes (shown in detailed view of *UROS* locus) for non-transfected K-562 or transfected with nuclease. (Left) Quantification of cells with 2O/2G or 1O/2G profiles. (Right) Illustrative FISH results for K-562 with 2O/2G or 1O/2G profiles. Green and orange fluorescent probes are respectively -4.6Mb upstream and +4.4Mb downstream *UROS* gene. **: $p < 0.01$. For (a) and (b), source data are provided as a Source data file.

Supplementary Figure 6



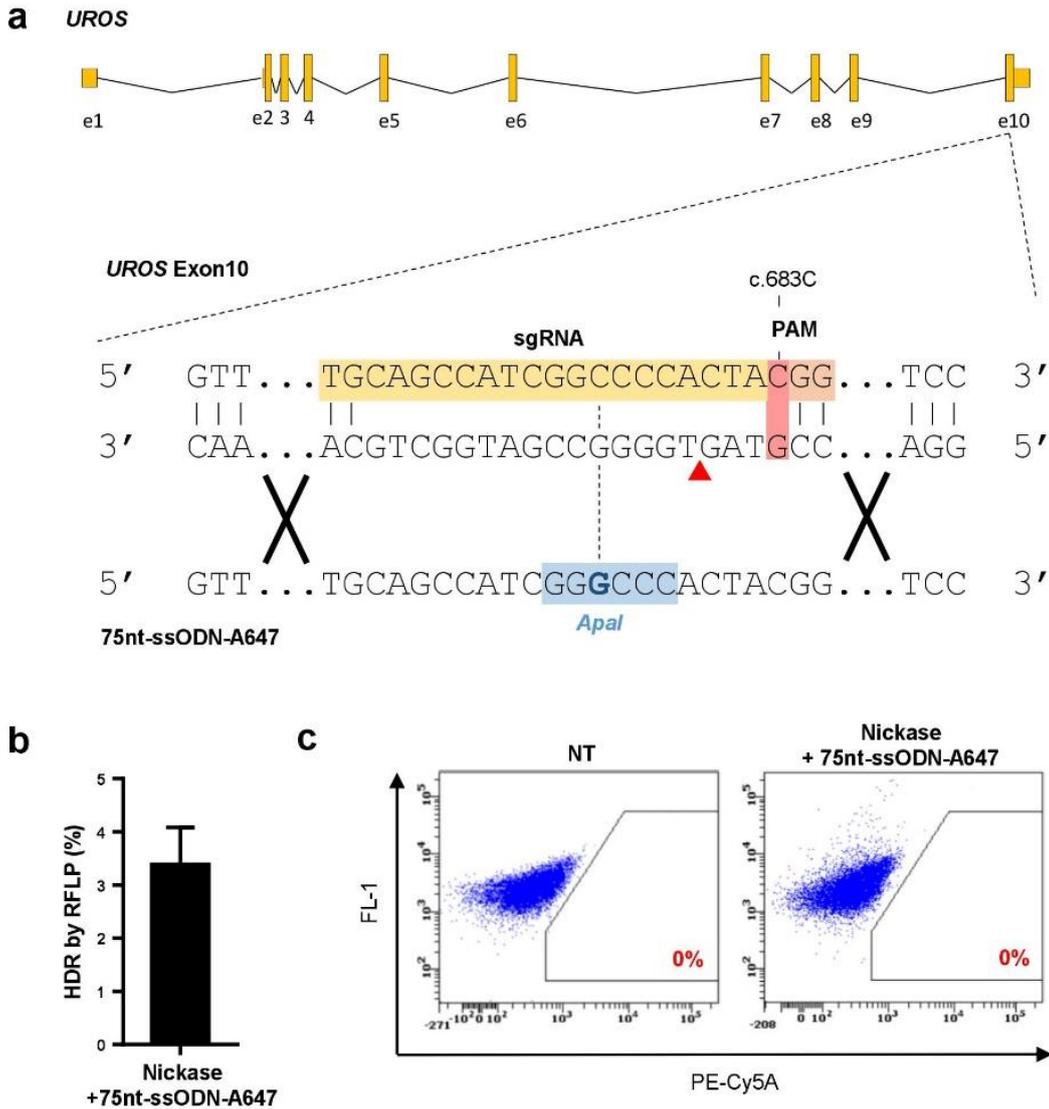
Supplementary Figure 6: Increased ssODN template concentration and 5' terminal modifications improve single nickase-mediated HDR efficiency. **a**, HDR frequencies for HEK293T cells transfected with nickase and co-delivered or not with 181nt-ssODN (0, 0.05, 0.5, 1.7, 5.1 μM) (n=5). **b**, HDR relative frequencies from cells transfected with nickase co-delivered either with 181nt-ssODN, 80nt-ssODN, 80nt-ssODN-A647 or 80nt-ssODN modified at its 5' terminal end with three Locked Nucleic Acid (80nt-ssODN-LNA) (n=5). Values are normalized with 181nt-ssODN condition at 0.5μM. Results are presented as mean ± SEM. Data are from 5 independent experiments. Statistical significance is inferred **a**, using paired one-way ANOVA, **b**, using two-tailed paired t-test to compare 80nt-ssODN vs 181nt-ssODN conditions and paired one-way ANOVA is used to compare 80nt-ssODN vs 80nt-ssODN-A647 and 80nt-ssODN-LNA. ns, not significant; *, p<0.05. For (a) and (b), source data are provided as a Source data file.

Supplementary Figure 7



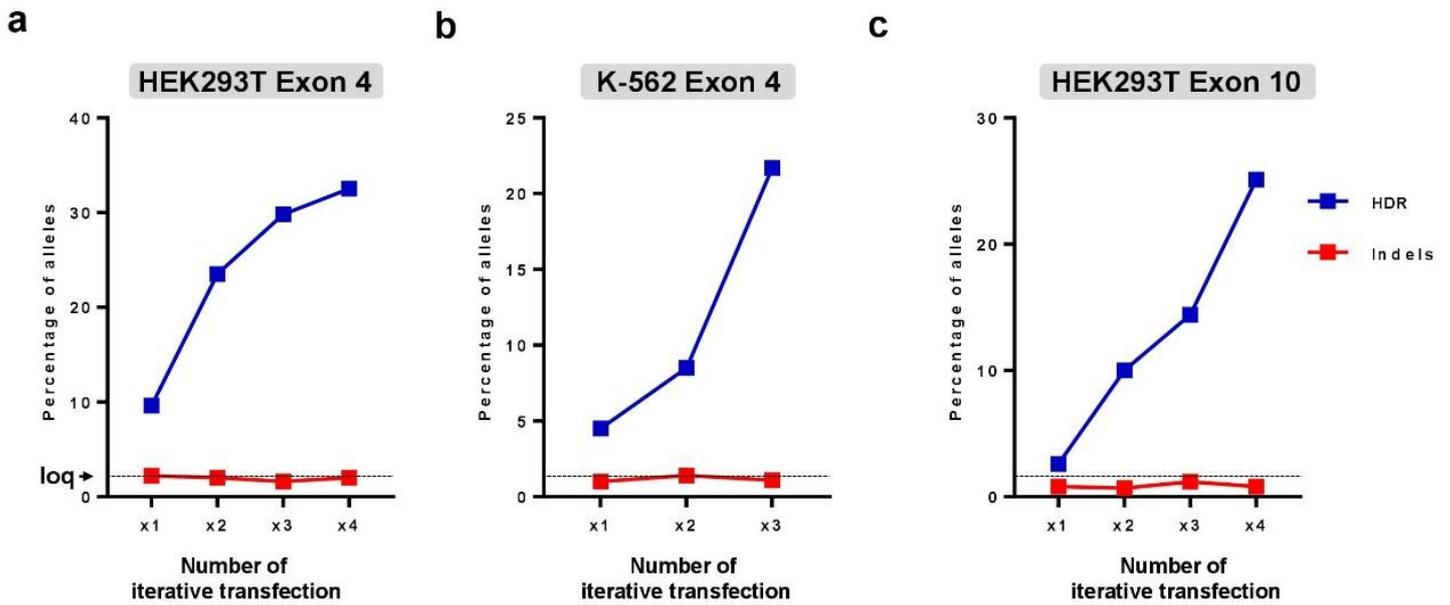
Supplementary Figure 7: Single nickase-mediated approach allows precise genome editing in K-562 cells. **a**, (Left) Illustrative RFLP analysis and (Right) HDR frequency for K-562 cells transfected with nickase and 80nt-ssODN-A647 (n=6). **b**, (Left) NGS analysis of allelic outcomes and associated HDR/indel ratio following transfection of K-562 cells with nickase and 80nt-ssODN-A647. (Right) Most common observed alleles (with frequencies $\geq 1\%$) aligned on the sgRNA sequence. HDR events in blue. **c**, Illustrative FACS results from non-transfected K-562 cells (NT) or K-562 transfected with nickase and 80nt-ssODN-A647. For (a), source data are provided as a Source data file.

Supplementary Figure 8



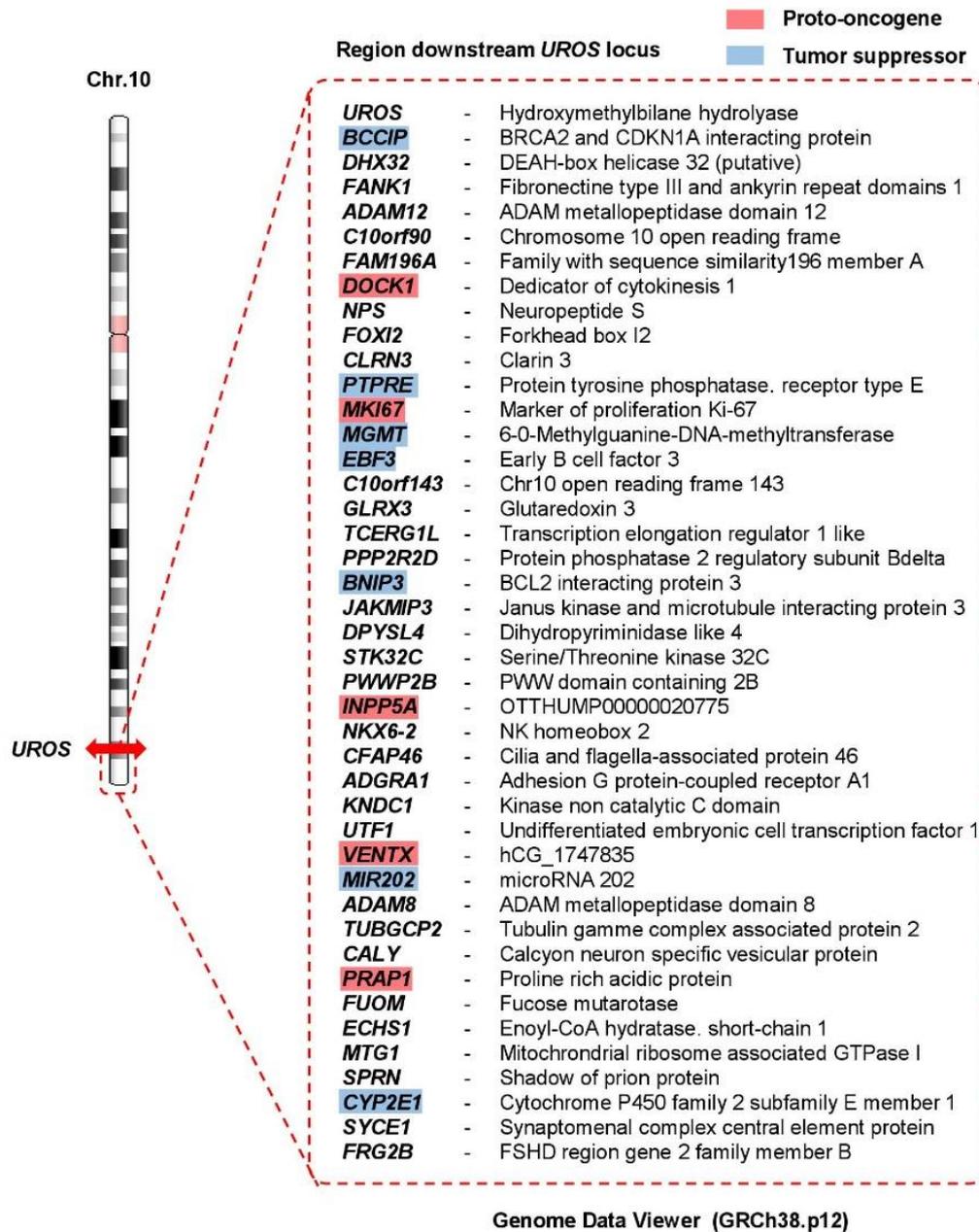
Supplementary Figure 8: Single nickase-mediated approach allows precise genome editing in *UROS* exon 10 locus. **a**, (Top) *UROS* gene overview and (Bottom) detailed view of exon 10 region and CRISPR-mediated HDR design using a c.683C-targeting sgRNA and a 75nt-ssODN-A647 carrying a silent *Apal* restriction site (blue) close to c.683C position and in the sgRNA seed sequence. Expected cleavage position using a nickase is indicated with a red arrow. **b**, HDR frequency for transfected HEK293T cells with nickase and 75nt-ssODN-A647 (n=6). **c**, Illustrative flow cytometry results from non-transfected (NT) HEK293T cells or co-transfected with nickase and 75nt-ssODN-A647. For (b), source data are provided as a Source data file.

Supplementary Figure 9



Supplementary Figure 9: Efficient HDR using iterative rounds of single nickase editing. HDR and indel quantification by TIDER analysis. Editing of **a**, *UROS* exon 4 in HEK293T; **b**, *UROS* exon 4 in K-562 cells and **c**, *UROS* exon 10 in HEK293T after 1 to 4 iterative transfections. HDR and indel rates are respectively presented in blue and red lines. loq: limit of quantification. For (a), (b) and (c), source data are provided as a Source data file.

Supplementary Figure 11



Supplementary Figure 11: List of genes downstream of UROS locus in Chr10. (Left) Schematic UROS locus position on chromosome 10 and (Right) list of 43 genes situated downstream. Genes known to be proto-oncogene or tumor-suppressor are respectively highlighted in red and blue.

Supplementary Table 1

Target	Target sequence	MIT Score	Number of mismatches	Indels (%)		Position (GRCh38)	Strand	Type	Gene
				Nuclease	Nickase				
On- target	GGAAGCAGCAGAGTTATGTTGG	0	0	83	<loq	Chr10:125815054-125815077	+	Exonic	UROS
Off-target 1	GGCAGCAGCAGGGTTATGTTGGG	3.9	2	5	<loq	Chr16:2689598-2689620	+	Intergenic	-
Off-target 2	GGCAGCAGCAGGGTTATGTTGGG	3.9	2	5	<loq	Chr16:2609502- 2609524	+	Intergenic	-
Off-target 3	GAAAGCAGCAGAGTCATGTTGGG	3	2	32	<loq	Chr9:107494173- 107494195	-	Intergenic	-
Off-target 4	AGAAGCATCAGGGTTATGTTAGG	1.4	3	<loq	<loq	Chr4:103916635- 103916657	-	Intergenic	-
Off-target 5	TCAACCACCAGAGTTATGTTGG	1.4	4	<loq	<loq	Chr21:15731392- 15731414	+	Intronic	USP25
Off-target 6	GGGAGCACCCAGAGTGATGTTAGG	0.8	3	<loq	<loq	Chr1:16199632- 16199654	-	Intronic	ARHGEF19
Off-target 7	GTTAGGAGCTGAGTTATGTTGGG	0.7	4	<loq	<loq	Chr7:18471631- 18471653	-	Intronic	HDAC9
Off-target 8	GGCATTAGCTGAGTTATGTTAGG	0.7	4	<loq	<loq	Chr15:58511066- 58511088	-	Intergenic	-
Off-target 9	AAAAGCAGCCAAGTTATGTTGG	0.7	4	< loq	<loq	Chr11:91048325- 91048347	-	Intergenic	-
Off-target 10	CCAATCAGCAGAGTTATGTAAGG	0.7	4	<loq	<loq	Chr8:137332792- 137332814	+	Intergenic	-

Supplementary Table 1: c217T-targeting gRNA off-targets TIDER analysis, using nuclease or single nickase in HEK293T cells. loq: limit of quantification.

Supplementary Table 2

PCR primers sequence (5' to 3')	
UROS Exon 4 Forward	TAGTTCCAGGCACATAGTAAGCAC
UROS Exon 4 Reverse	AGGAGGTGAACAACGAATAGACAG
NGS - UROS Exon 4 Forward	TCTTGTGAAAAGGACGAAACACCGGGGACTCATTTTACCAGCC
NGS - UROS Exon 4 Reverse	TCTACTATTCTTCCCCTGCACTGTTCTGGAA TTTAGTCTCCCAGCA
UROS Exon 10 Forward	GAGAGCGTGGATGCAGA
UROS Exon 10 Reverse	ACAGCAACCATACACTCAG
NGS - UROS Exon 10 Forward	TCTTGTGAAAAGGACGAAACACCGGTGTGCTGAAGCCCTGTTTT
NGS- UROS Exon 10 Reverse	TCTACTATTCTTCCCCTGCACTGTGAGCCTTCCTGATGCCAGT
sgRNA sequence (5' to 3')	
Exon 4 sgRNA Forward	GGAAGCAGCAGAGTTATGTT
Exon 4 sgRNA Reverse	ATAACTCTGCTGCTTCCGGT
Exon 4 c.217C-Sacl- specific sgRNA Forward	GGAAGCAGCAGAGCTCCGTT
Exon 4 c.217C-Sacl- specific sgRNA Reverse	AACGGAGCTCTGCTGCTTCC
Exon 10 sgRNA Forward	TGCAGCCATCGGCCCACTA
Exon 10 sgRNA Reverse	TAGTGGGGCCGATGGCTGCA
HDR template sequence (5' to 3')	
181nt-ssODN	TTTTGGTGTGCAGCTTTCTCATCCTGAAGATTACGGGGGACTCATTTTACTAGCCCCAGAGCAGTG GAAGCAGCAGAGCTCTGTTTGGAGCAAAAACAATAAACTGAAGGTGAGGGTGGGTCTGCTGTCG ATTCCA CTGGACATTTATTTACTCTTATTTCTCCTGCTGGGAGACTAAAT
181nt-ssODN-c.217C	TTTTGGTGTGCAGCTTTCTCATCCTGAAGATTACGGGGGACTCATTTTACTAGCCCCAGAGCAGTG GAAGCAGCAGAGCTCCGTTTGGAGCAAAAACAATAAACTGAAGGTGAGGGTGGGTCTGCTGTC GATTCACCTGGACATTTATTTACTCTTATTTCTCCTGCTGGGAGACTAAAT
80nt-ssODN	CATTTTACCAGCCCCAGAGCAGTGGAAAGCAGCAGAGCTCTGTTTGGAGCAAAAACAATAAACT GAAGGTGAGGGTGGGT
75nt-ssODN	GTTTTCTCTTCTGCTTTATAGTTTGCAGCCATCGGGCCCACTACGGCTCGCGCTGGCCGCCAG GGCCTTCC
PCR primers for off-target analysis (5' to 3' sequence)	
Off-target 1 Forward	AGGTGAAGGTTTTGTGTGCGA
Off-target 1 Reverse	CAGTGAGGAATGAGGGTGGAG
Off-target 2 Forward	AGGTGAAGGTTTTGTGTGCGA
Off-target 2 Reverse	CAGTGAGGAATGAGGGTGGAG
Off-target 3 Forward	ACCCTCTGCTTGTGTAGACT
Off-target 3 Reverse	AGCCTCAGCCTCATCATGTTT
Off-target 4 Forward	ACTGTCCACCA TTTCTCACCA
Off-target 4 Reverse	ATCCTTCTCTTGGCCAGACC
Off-target 5 Forward	GGGCAAAAGAGTAAACCGTGC
Off-target 5 Reverse	CCATGTGCTCTTTCCTAAGCC
Off-target 6 Forward	CCACGTTCAAGCAGGCATAC
Off-target 6 Reverse	GGGTGGGGTAAAGCTCTCTA
Off-target 7 Forward	ATGGGTTAGGTGAAAGCTCAGT
Off-target 7 Reverse	ACTAGTTGCA TCTTCTAGCATC
Off-target 8 Forward	TCAACAA CCGCACCTGGAAG
Off-target 8 Reverse	AACTGCCTGGGCCCTAAC
Off-target 9 Forward	GAATATCTTACTCCCAGCGT
Off-target 9 Reverse	CATCA TTCAGCTTAA TTCAC TTTGT

Supplementary Table 2: Oligonucleotide sequences used for PCR, sgRNA sequence and HDR template.

IV. Second partie : Edition génique et délétion terminale

Les résultats de l'étude présentée ci-dessus illustrent le risque de l'édition génique sur l'intégrité chromosomique. Malgré tout, plusieurs questions demeurent pour caractériser et évaluer ce risque.

En premier lieu, l'une des priorités est d'évaluer ce risque pour des lignées, plus pertinentes que les lignées cancéreuses, dans le cadre d'une stratégie thérapeutique.

De plus, un besoin directement soulevé par l'étude précédente est de confirmer et de quantifier le risque de délétion pouvant être induit par la Cas9^{D10A}-nickase.

Cependant, ces travaux ont également souligné les limites de sensibilité des outils cytogénétiques disponibles pour la détection des délétions terminales. Le développement d'un outil alternatif sensible s'est rapidement imposé comme une nécessité pour pouvoir répondre aux questions ci-dessus.

i. Objectif : Création d'un outil sensible de détection des délétions terminales

a. La PEC : un outil de détection des délétions terminales

L'objectif principal a été de développer un outil de détection sensible en faisant en sorte que la délétion terminale entraîne un phénotype facilement identifiable et quantifiable. Pour permettre cette relation génotype-phénotype, le modèle de la PEC, maladie récessive, a encore une fois été un atout précieux. Dans ce cas, l'accumulation de protoporphyrines de type I est un phénotype détectable par cytométrie en flux et sa survenue est directement dépendante du statut génotypique d'*UROS* : les cellules WT ou hétérozygotes possèdent une activité *UROS* supérieure à 50% rendant les cellules non-fluorescentes tandis que les cellules non-fonctionnelles pour *UROS*, suite à l'inactivation des 2 allèles, le sont (*Figure 48A*).

L'astuce de la méthode de détection des délétions terminales est la suivante : à partir d'un génotype hétérozygote pour *UROS*, avec un allèle normal et un muté (et donc avec des cellules non-fluorescentes), la survenue d'une délétion terminale emportant l'allèle sauvage transformerait le génotype en « homozygote » muté pour *UROS*, par perte d'hétérozygotie, les cellules devenant donc fluorescentes (*Figure 48B*). Pour que la perte de l'allèle sauvage soit due à des délétions terminales, la lésion est faite à un mégabase en amont d'*UROS*, dans le gène *FAM175B*. Si une délétion terminale a lieu lors de la coupure au niveau du gène *FAM175B* et emporte l'allèle WT d'*UROS*, la cellule devient fluorescente. La fréquence d'apparition de fluorocytes permet donc une évaluation sensible du risque de délétion terminale. La délétion terminale de l'allèle portant *UROS* muté n'entraînant pas de changement de phénotype, il est probable que la fréquence réelle de l'évènement soit le double du pourcentage de cellules fluorescentes.

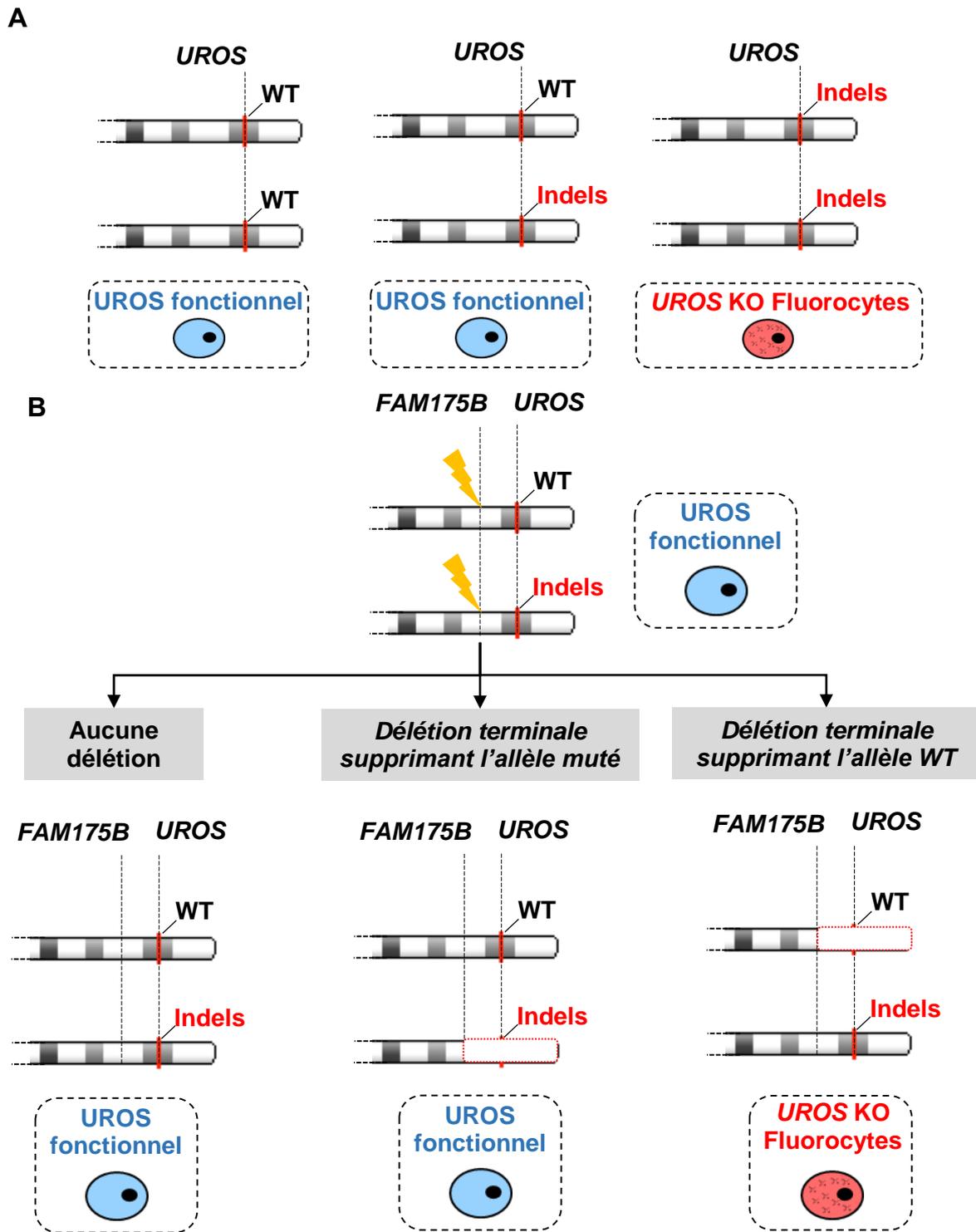


Figure 48 Méthode de détection des délétions terminales par la modification de la fonctionnalité d'UROS

(A) Relation entre le génotype d'UROS et la fluorescence des cellules. Les cellules homozygotes ^{WT/WT} ou hétérozygotes ^{WT/InDels} ne sont pas fluorescentes, à l'inverse, les cellules homozygotes ^{InDels/InDels} le sont (fluorocytes).

(B) (Haut) A partir d'une lignée hétérozygote ^{WT/InDels}, une lésion est induite sur le gène *FAM175B*, situé à 1 Mb en amont d'UROS. (Bas) Les différents cas présentant des délétions terminales ou non sont illustrés. La perte de la copie normale d'UROS par une délétion terminale rend la cellule fluorescente. Les InDels formés au niveau de *FAM175B* ne sont pas représentés dans un souci de clarté.

Ce système nécessite de réaliser une lignée hétérozygote pour *UROS* et de définir les conditions nécessaires à l'accumulation de protoporphyrines de type I. Cette phase de mise au point est décrite ci-dessous. Cette méthodologie présente deux avantages majeurs. Les deux premiers sont ceux de la sensibilité et de l'efficacité de la cytométrie en flux, permettant la détection d'évènements rares tout en conservant une analyse rapide de plusieurs conditions en parallèle avec des centaines de milliers de cellules. Le troisième avantage est la possibilité de trier les cellules fluorescentes pour pouvoir les caractériser davantage par des approches de biologie moléculaire et de cytogénétique ; ce travail permettant également une validation méthodologique du système de détection basée sur la cytométrie.

b. Questions scientifiques

Dans le cadre de notre étude précédente, les délétions terminales ont été observées par FISH uniquement dans des lignées cancéreuses ou dans des fibroblastes immortalisés et invalidés pour p53. La première question a été de savoir si une méthode plus sensible permettrait la détection de délétions terminales dans des lignées diploïdes p53-positives. Pour répondre à cette question, des fibroblastes immortalisés ont été édités sur *UROS* et sous-clonés pour obtenir un clone hétérozygote. Une fraction de cette dernière a ensuite été invalidée pour p53 pour permettre une étude comparative.

La seconde interrogation est de caractériser le risque pour des procédés d'édition génique alternatifs ne faisant pas appel à la Cas9-nucléase, notamment avec la Cas9^{D10A}-nickase ou également avec la deadCas9, catalytiquement inactive. Profitant des deux lignées mises au point précédemment, une comparaison des trois variants a été conduite pour évaluer leur impact sur l'intégrité chromosomique.

ii. Mise au point des outils expérimentaux

Ce travail préparatoire a consisté à définir la dose de 5-ALA rendant détectable les cellules accumulant des protoporphyrines de type I et mettre au point les lignées fibroblastiques.

a. Définition de la dose de 5-ALA

L'analyse de fibroblastes invalidés pour *UROS* ne permet d'obtenir spontanément qu'un faible taux de fluorocytes (0,58% contre 0,06% pour des fibroblastes non-édités). La première étape a été de rendre les cellules non-fonctionnelles pour *UROS* capables d'accumuler suffisamment de protoporphyrines de type I pour permettre leur détection par cytométrie. Une solution a été apportée par l'utilisation de l'acide 5-aminolévulinique (5-ALA). Ce dernier est un précurseur de la

biosynthèse de l'hème. L'exposition des cellules aux 5-ALA stimule la formation de protoporphyrines de type III et IX, si UROS est fonctionnel, ou de type I en son absence.

Toutefois, les trois types de protoporphyrines sont photosensibles. Ainsi, même pour des cellules normales, l'exposition à une concentration excessive de 5-ALA entraîne une présence de protoporphyrines de type III et IX, dernier précurseur menant à l'hème, suffisamment forte pour rendre fluorescente une fraction des cellules traitées.

Des résultats précédemment acquis au laboratoire ont montré qu'une période de clairance faisant suite à l'exposition au 5-ALA permettait aux cellules normales de métaboliser le surplus de protoporphyrines de type III et IX induits tandis que les cellules non-fonctionnelles pour UROS restaient fluorescentes à cause des protoporphyrines de type I accumulés et en impasse métabolique (*Figure 49A*).

Pour définir la dose et le temps de clairance adaptés pour obtenir ce comportement différentiel, les fibroblastes immortalisés par hTERT et invalidés pour p53, utilisés lors de l'étude précédente, ont été transfectés pour invalider la fonction d'UROS. Les outils et les méthodes utilisés sont similaires à ceux présentés dans l'article. Le séquençage montre un pourcentage d'InDels à hauteur de 97%, dont 95% entraînant un décalage du cadre de lecture, témoignant d'une invalidation d'UROS pour cette population. Deux semaines après leur transfection, les cellules éditées ou non ont été exposées à des doses croissantes de 5-ALA, de 0,1 à 1 mM pendant 16h. Les cellules ont ensuite été lavées deux fois au PBS, marquant le début de la période de clairance. Après 6h de clairance, le pourcentage de fluorocytes a été quantifié par cytométrie en flux (*Figure 49B*). L'objectif est de trouver la dose qui permet de rendre les cellules invalidées pour UROS fluorescentes tout en conservant les cellules avec une activité enzymatique UROS non fluorescentes.

Les résultats montrent que, sans 5-ALA, les deux populations présentent un taux faible de fluorocytes (0,06% et 0,58% respectivement pour les fibroblastes non-édités et invalidés pour UROS). Les pourcentages obtenus pour les cellules non-éditées exposées à des doses comprises entre 0,1 et 0,4 mM restent proches à la valeur obtenue sans traitement (comprise entre 0,06 et 0,14%). A partir de 0,5 mM, la quantité de fluorocytes augmente (0,34%), témoin d'une clairance imparfaite. Pour les cellules invalidées pour UROS, le pourcentage de fluorocytes est plus fort dès 0,1mM (7,47%). Un plateau est atteint aux alentours de 94% à partir de 0,5 mM. Les doses 0,3 et 0,4 mM permettent de former respectivement 86% et 92% de fluorocytes. A partir de ces résultats, la dose de 0,3 mM a été retenue. A cette dose, les cellules fonctionnelles ne sont pas fluorescentes et celles invalidées le sont à hauteur de 86%.

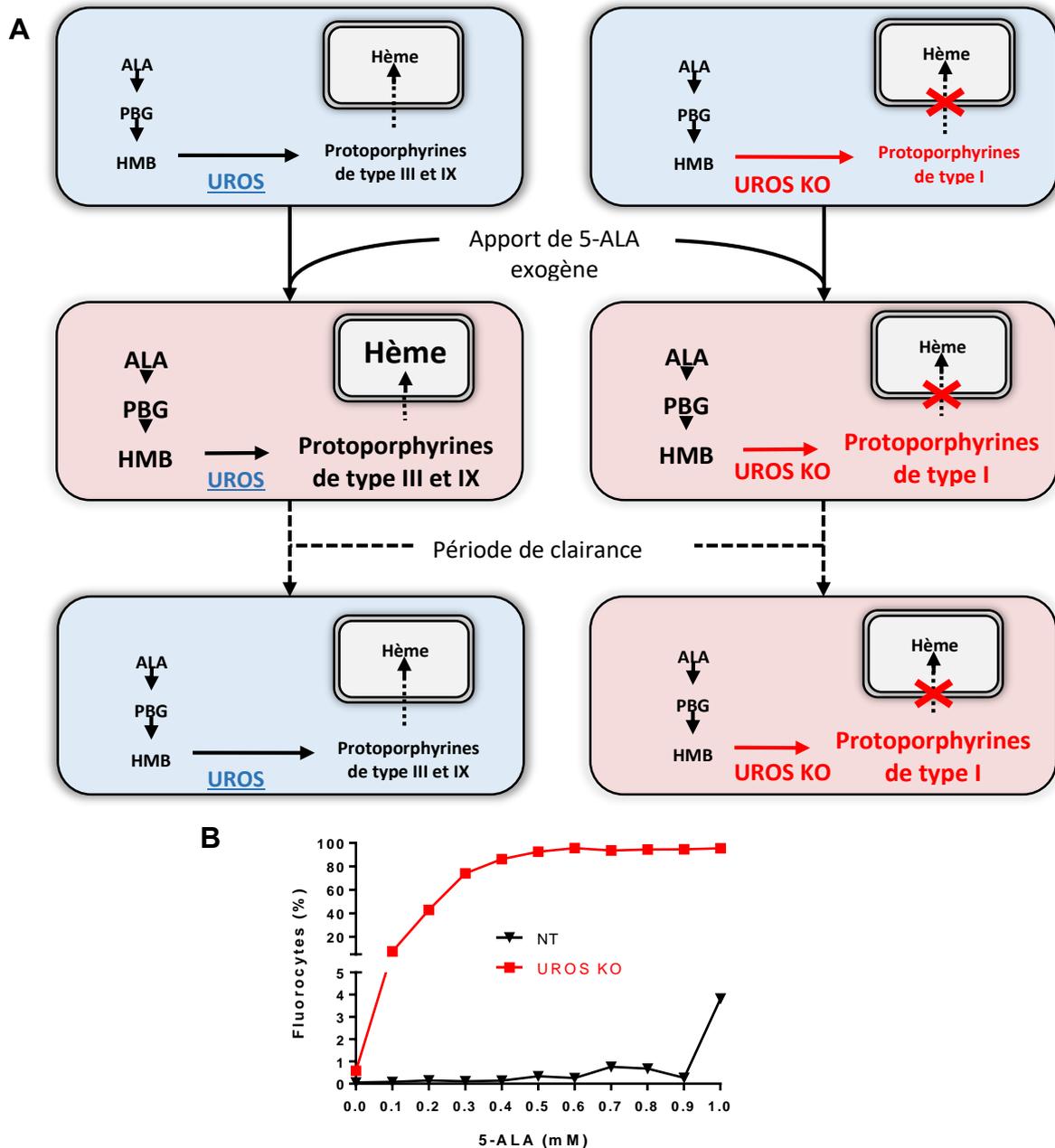


Figure 49. Mise au point de la dose de 5-ALA.

(A) Effet du 5-ALA sur le métabolisme de l'hème pour des fibroblastes fonctionnels pour UROS, avec formation de protoporphyrines de type III et IX (Gauche) ou *knock-out* pour UROS avec formation de protoporphyrines de type I (Droite). (Haut) La quantité de protoporphyrines n'atteint pas un niveau suffisant pour rendre les cellules fluorescentes. (Centre) L'exposition au 5-ALA stimule la biosynthèse de l'hème et augmente la quantité de protoporphyrines à un niveau suffisant pour rendre les cellules fluorescentes, indépendamment du statut d'UROS. (Bas) Après une période de clairance, les cellules fonctionnelles pour UROS métabolisent le surplus de protoporphyrine et redeviennent non-fluorescentes. À l'inverse, les cellules *knock-out* pour UROS ont accumulé des protoporphyrines de type I, en impasse métabolique, et restent fluorescentes.

(B) Quantification par cytométrie en flux du pourcentage de fluorocytes en fonction de la dose de 5-ALA pour des fibroblastes *hTERT*-positifs et invalidés pour p53, transfectés avec des RNPs Cas9:ARNg dirigés contre *UROS* (*UROS KO* en rouge). La condition non-traitée (NT) correspond à des cellules non-transfectées. L'analyse est faite 14 jours après la transfection. Les cellules sont exposées 16h au 5-ALA, lavées deux fois au PBS et une période de clairance de 6h est observée avant l'analyse.

b. Création d'une paire de lignées isogéniques pour étudier la dépendance à p53

Deux lignées de fibroblastes ont été mises au point pour inférer sur la nécessité d'une déficience de p53 dans l'apparition de délétions terminales. La mise en place des lignées s'est faite en deux temps : formation d'un clone hétérozygote pour *UROS* et extinction de p53 pour une fraction des cellules, créant ainsi la deuxième lignée.

Formation d'un clone de fibroblastes immortalisés hétérozygote pour *UROS*

Pour débiter, des fibroblastes immortalisés par hTERT ont été transfectés avec une dose réduite de RNP (4,2µg) composée de la Cas9-nucléase et de l'ARNg visant la position c.217T de l'exon 4 d'*UROS*. Après deux semaines, les cellules ont été exposées au 5-ALA, puis une période de 6h de clairance a été faite avant l'analyse par cytométrie. Parmi les cellules éditées, deux populations sont identifiables avec d'une part 86,2% de fluorocytes et 13,8% de cellules non-fluorescentes (*Figure 50A*). L'objectif étant de trouver un clone hétérozygote pour *UROS* et donc fonctionnel pour *UROS*, des cellules non-fluorescentes ont été triées et passées à l'état monocellulaire. Après expansion clonale, le locus *UROS* de chacun des clones a été analysé. Parmi les huit clones analysés par séquençage, quatre présentent une mutation de l'un des allèles avec l'insertion d'une paire de bases juxtaposée au site de coupure. Un clone hétérozygote possédant ce génotype a été choisi pour la suite de l'étude (*Figure 50B*).

Formation d'une lignée invalidée pour p53 à partir du clone hétérozygote

La seconde étape a été l'invalidation de p53 pour une fraction de cellules issues du clone hétérozygote pour *UROS*. Pour cela, ces dernières ont été transfectées avec des RNP composées de la Cas9-nucléase et d'un ARNg visant l'exon 4 du gène *TP53*. Cinq jours après la transfection, le séquençage montre un décalage de cadre de lecture pour 98% des allèles analysés (*Figure 50C*). Une confirmation de l'invalidation au niveau protéique est en cours de travaux, toutefois les résultats présentés ci-dessous sont en faveur d'une invalidation fonctionnelle effective.

Clone hétérozygote pour *UROS* et sensibilité au 5-ALA.

La dernière étape de mise au point des outils a été de s'assurer que l'hétérozygotie d'*UROS* et le statut de p53 ne modifiaient pas la sensibilité au 5-ALA. Pour cela, quatre manipulations indépendantes de doses-réponses au 5-ALA ont été conduites sur les deux lignées et sur une troisième lignée formée pour servir de contrôle positif. Celle-ci est issue de la transfection du clone hétérozygote pour *UROS* et p53-positifs avec une RNP dirigée contre *UROS* ; cette lignée devenant déficiente pour *UROS*.

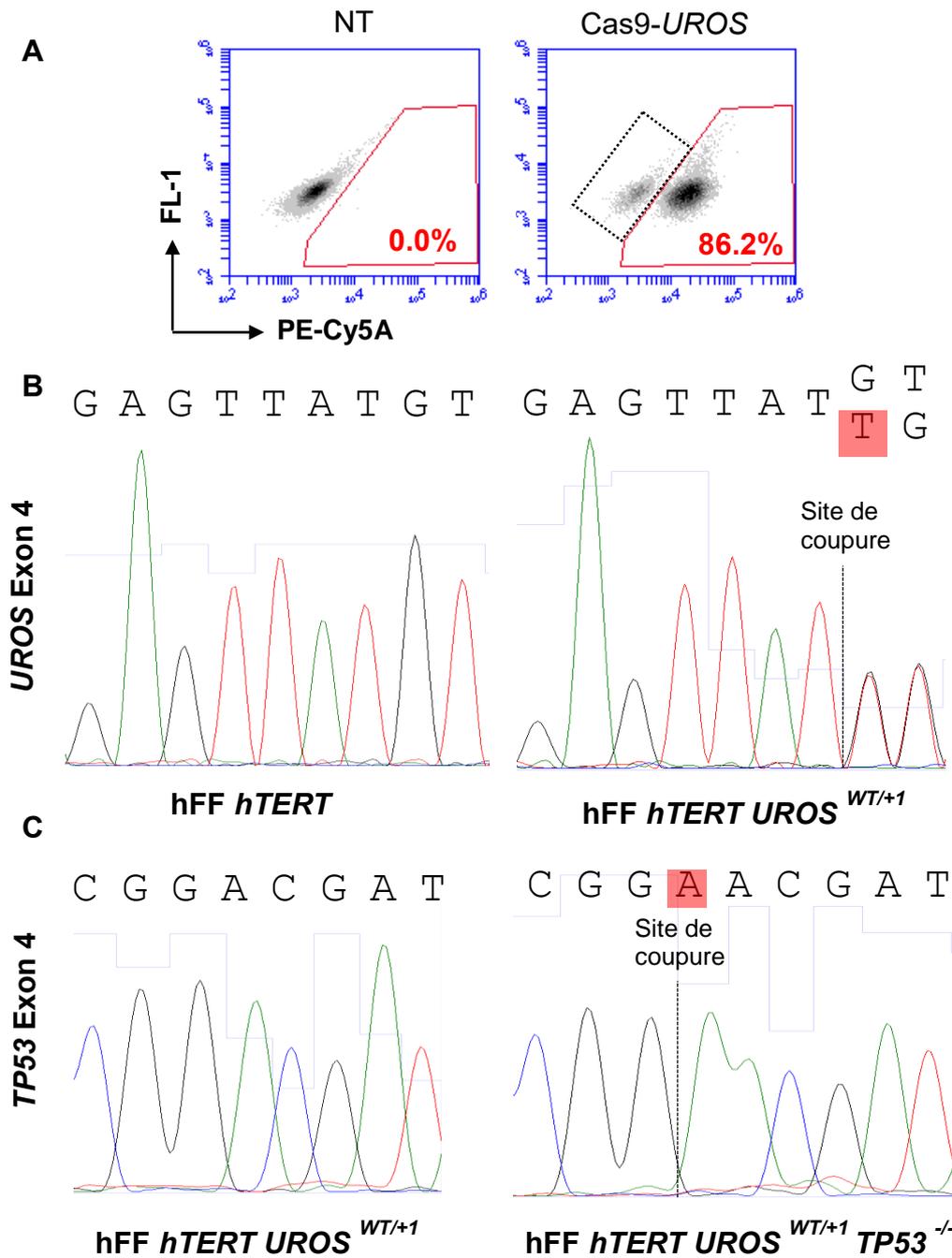


Figure 50. Création des deux lignées fibroblastiques.

(A) Quantification par cytométrie en flux des fluorocytés pour des fibroblastes *hTERT*-positifs (hFF *hTERT*) transfectés avec une RNP Cas9:ARNg dirigée contre *UROS* (Cas9-*UROS*) ou non-transfectés (NT). La fraction encadrée en pointillé a été triée par FACS et passée à l'état monocellulaire.

(B) Séquençage du locus *UROS* de hFF *hTERT* et d'un clone hétérozygote pour *UROS* (hFF *hTERT UROS*^{WT/+1}) issu du tri montré en (A). Le site de coupure de la Cas9:ARNg dirigée contre l'exon 4 d'*UROS* est montré. L'insertion d'une paire de bases est colorée en rouge.

(C) Séquençage du locus *TP53* du clone hFF *hTERT UROS*^{WT/+1} et du même clone transfecté avec une RNP Cas9:ARNg dirigée contre l'exon 4 de *TP53* (hFF *hTERT UROS*^{WT/+1} *TP53*^{-/-}). Le site de coupure de la Cas9:ARNg dirigée contre l'exon 4 de *TP53* est montré. L'insertion d'une paire de bases est colorée en rouge.

Les résultats montrent un comportement similaire à celui observé précédemment (*Figure 51*). A 0,3 mM de 5-ALA, les cellules déficientes pour UROS sont fluorescentes à hauteur de 96,49% contre 0,01% et 0,07% respectivement pour les cellules p53-positives et p53-déficientes avec une activité UROS. L'absence de différence entre un clone hétérozygote et une lignée homozygote normale pour UROS est cohérente avec la physiopathologie associée à UROS. En effet, une activité faible d'UROS (10 % de l'activité spécifique) suffit à rendre la cellule compétente pour la formation de protoporphyrines de type III plutôt que de type I.

iii. Résultats

a. Délétion terminale, effet du statut p53

Une fois les outils mis au point, une première série d'expériences a été menée pour savoir si l'inactivation de p53 est nécessaire pour voir apparaître des délétions terminales. Pour répondre à cette interrogation, les deux lignées hétérozygotes pour *UROS* et possédant un p53 fonctionnel ou non ont été transfectées avec une RNP dirigée contre le gène *FAM175B*, situé à 1 Mb en amont d'*UROS* sur le bras long du chromosome 10.

Le guide visant l'intron 7 de *FAM175B* a été choisi pour sa grande spécificité prédite *in silico* (CFD-score de 97 sur 100); il ne possède que 6 sites prédits avec minimum 3 mésappariements dans la *seed sequence* du guide et 56 sites avec 4 mésappariements également dans la *seed sequence*. Aucun des sites *off-target* avec 3 mésappariements n'est présent sur le même chromosome et seulement 4 cibles *off-target* avec 4 mésappariements sont situées dans le même chromosome. Toutefois, leur CFD-off-target score est très faible (inférieur à 0,03) et aucun site n'est retrouvé entre *FAM175B* et l'extrémité du bras long du Chr.10. En parallèle, un contrôle négatif a été ajouté pour s'assurer de la spécificité de l'effet observé. Ce contrôle négatif est issu de la transfection des deux lignées par une RNP composée d'une Cas9-nucléase dirigée contre le locus *PIGA*, situé sur le chromosome X.

Une efficacité d'édition variable

L'analyse par séquençage de l'édition obtenue aux loci *PIGA* et *FAM175B* montre une forte différence d'efficacité. Sur *PIGA*, une fréquence d'InDels a été quantifiée à 97,5% (n=2) et 98% (n=2) respectivement pour la lignée p53-positive et déficiente. A l'inverse, une édition modérée est retrouvée au locus *FAM175B* avec un taux de modification mesurée à $37 \pm 9,1\%$ (n=4) et $27,6 \pm 11,9\%$ (n=3) respectivement pour la lignée p53-positive et déficiente.

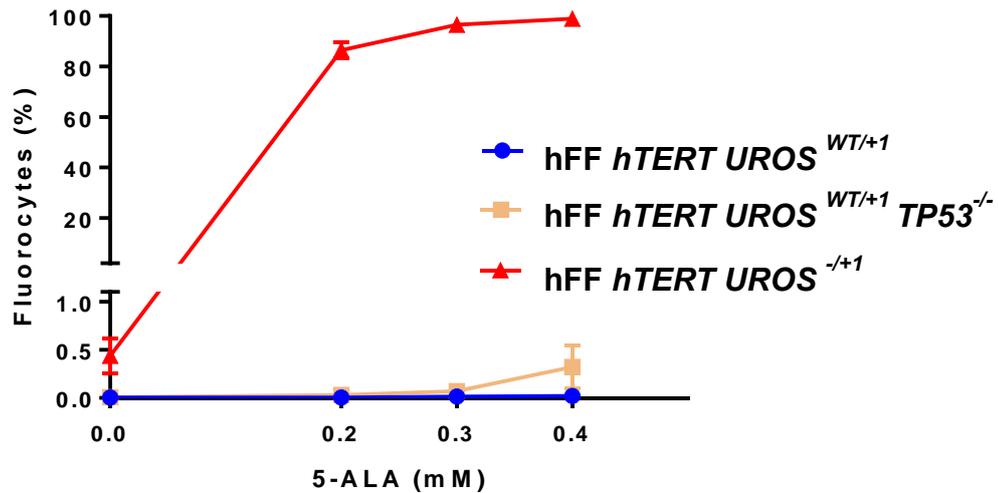


Figure 51. Sensibilité des lignées fibroblastiques au 5-ALA.

Quantification par cytométrie en flux du pourcentage de fluorocytes en fonction de la dose de 5-ALA pour les hFF *hTERT UROS*^{WT/+1} et hFF *hTERT UROS*^{WT/+1} *TP53*^{-/-}. Le contrôle positif de fluorescence est constitué de hFF *hTERT UROS*^{WT/+1} transfectés avec des RNPs Cas9:ARNg dirigées contre *UROS* pour invalider l'allèle WT restant (hFF *hTERT UROS*^{-/+1}). L'analyse est faite 14 jours après la transfection. Les cellules sont exposées 16h au 5-ALA, lavées deux fois au PBS et une période de clairance de 6h est observée avant l'analyse. Les résultats sont présentés avec la moyenne \pm écart-type (n=3, expériences indépendantes).

Effet de p53 dans la fréquence des délétions terminale

Deux semaines après la transfection, les cellules ont été exposées à 0,3mM de 5-ALA, suivi par une période de clairance puis analysées en cytométrie. Les résultats obtenus pour une expérience indépendante sont présentés (*Figure 52A*). Pour la lignée déficiente pour p53, un léger bruit de fond est observé pour les conditions non-transfectées ($0,10 \pm 0,01\%$, n=4) et transfectées avec la Cas9 dirigée contre *PIGA* ($0,12 \pm 0,02\%$, n=3) (*Figure 52B*). Toutefois, une fraction importante de fluorocytes apparait après édition du locus *FAM175B* ($5,56 \pm 0,8\%$, n=4), témoignant de la perte de l'allèle WT d'*UROS*, 1 Mb en aval. Pour la lignée fonctionnelle pour p53, un bruit de fond quasi-nul (0,01%) est observé dans les deux conditions contrôles. Pour les cellules éditées sur *FAM175B*, une fraction non-nulle et supérieure aux contrôles est retrouvée dans chacune des 2 expérimentations indépendantes, avec une moyenne de $0,09 \pm 0,03\%$ de fluorocytes (*Figure 52B*). Ce résultat suggère que des grandes délétions d'au moins 1 Mb sont possibles dans des cellules primaires, alors qu'elles n'avaient pas pu être mises en évidence par le système FISH auparavant. Cet évènement rare pourrait toucher 1/500 cellules.

Validation méthodologique

La fraction de cellules fluorescentes issues de l'édition des fibroblastes *bTERT*-positifs et invalidés pour p53 a été triée par FACS (*Figure 53A*). Le locus *UROS* a été analysé. Le séquençage des cellules éditées non-triées montre que le locus *UROS* est toujours hétérozygote avec un allèle normal et un allèle muté avec une insertion de base (*Figure 53B*). A l'inverse, pour les cellules fluorescentes de la lignée déficientes pour p53, le séquençage du locus *UROS* montre que ces cellules sont homozygotes avec un allèle portant une insertion d'une paire de bases. Ce résultat peut être expliqué par la perte de l'allèle normal et donc une amplification uniquement de l'allèle muté restant, concordant avec l'hypothèse d'une délétion terminale, concordant avec la perte d'activité *UROS* observée. Ces résultats suggèrent que de très grandes délétions (d'au moins 1 Mb) peuvent survenir dans des cellules diploïdes immortalisées, même pour celles ayant une activité p53 normale, ce qui n'avait pas pu être détectée par la technique de FISH.

Cependant, seules des approches de cytogénétique sur la population de fluorocytes permettront de définir si la perte de l'allèle normal est due à des grandes délétions interstitielles ou à des délétions terminales. Toutefois, ce résultat confirme que des pertes de l'ordre du mégabase peuvent être engendrées par une simple CDB et que cette méthodologie de criblage permet la quantification et l'isolement des cellules touchées par des remaniements importants. Enfin, cette validation méthodologique devra être également conduite sur les cellules fluorescentes issues de la lignée fonctionnelle pour p53.

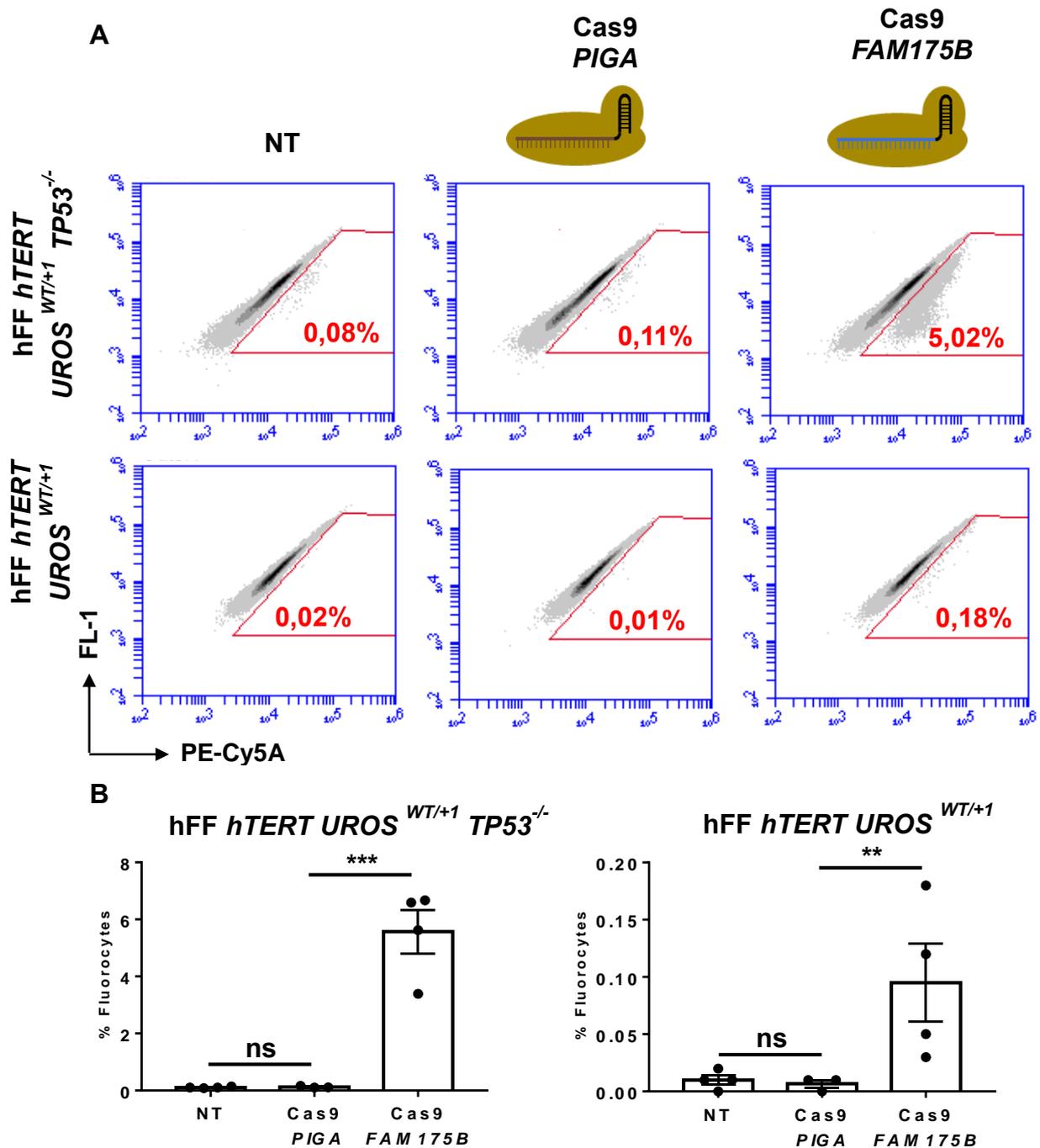


Figure 52. Effet de p53 dans la fréquence des délétions terminales.

(A) Résultats représentatifs et (B) quantification du pourcentage de fluorocytes pour les deux lignées fibroblastiques transfectées avec soit des RNP Cas9:ARNg dirigées contre *PIGA*, situé sur le chromosome X, soit des RNP Cas9:ARNg dirigées contre le locus *FAM175B*. L'analyse est faite 14 jours après la transfection. Les cellules sont exposées 16h au 5-ALA, lavées deux fois au PBS et une période de clairance de 6h est observée avant l'analyse. Les résultats sont présentés avec la moyenne \pm écart-type ($n=4$ et $n=3$ pour la condition Cas9-*PIGA*, expériences indépendantes). Une one-way-ANOVA complétée par un *unprotected Fisher's Least Significant Difference test* ont été utilisés pour inférer sur les différences statistiques entre les paires sélectionnées. ns, $p>0,05$; **, $p<0,01$; ***, $p<0,001$.

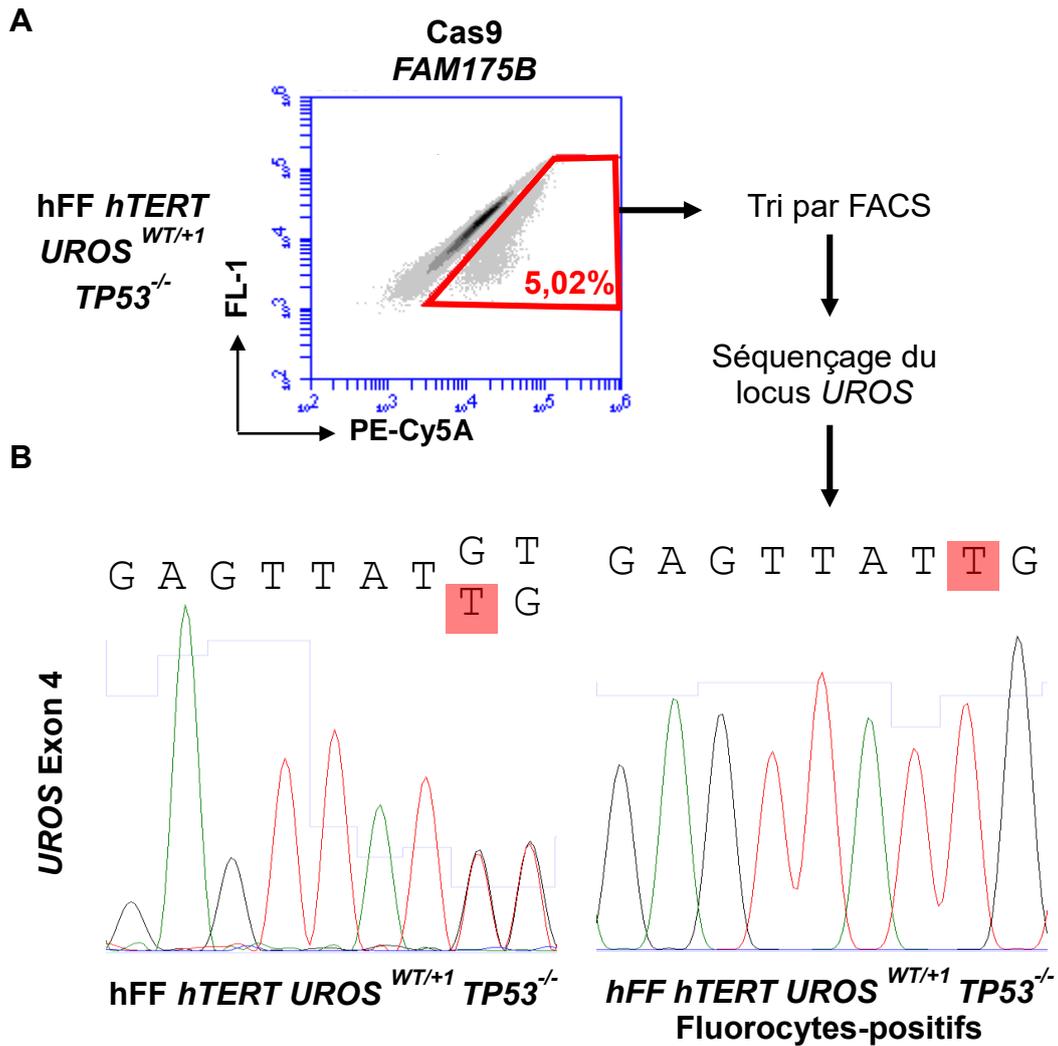


Figure 53. Validation méthodologique par biologie moléculaire.

(A) Les cellules fluorescentes issues de la transfection des hFF *hTERT* *UROS*^{WT/+1} *TP53*^{-/-} avec des RNPs Cas9:ARNg dirigées contre le locus *FAM175B* ont été triées par FACS et le locus *UROS* de cette population a été séquencé.

(B) Analyse du locus *UROS* de hFF *hTERT* *UROS*^{WT/+1} *TP53*^{-/-} non-transfectés (Gauche) et de cellules issues du tri montré en (A) (Droite). La PCR ne permet plus l'amplification de l'allèle WT, concordant avec une perte de cet allèle par très grande délétion. L'insertion est colorée en rouge.

b. Outils d'édition et risque de délétion terminale

Les résultats ci-dessus soulignent le risque d'une édition génique basée sur une coupure double-brin et le besoin d'évaluer également ce risque pour des outils d'édition alternatifs.

L'analyse par FISH d'une population polyclonale de cellules éditées au locus *UROS* avec la Cas9^{D10A}-nickase n'avait pas mis en évidence de délétions terminales, suggérant une instabilité chromosomique moindre avec cet outil, ne provoquant qu'une coupure simple-brin. Cependant, l'analyse du clone modélisant le PEC a montré une délétion terminale à l'édition par Cas9^{D10A}-nickase, suggérant que ces événements étaient probablement rares mais possibles. Une étude a donc été conduite pour comparer le risque de délétions terminales induit soit par la Cas9-nucléase, la Cas9^{D10A}-nickase ou la deadCas9 dans des cellules diploïdes immortalisées. Les deux lignées de fibroblastes, fonctionnelles ou non pour p53 ont été éditées en parallèle sur le locus *FAM175B* avec ces trois variants de Cas9. Une condition non-transfectée et une éditée par la Cas9 au locus *PIGA* ont été faites pour servir de témoin négatif. Les résultats obtenus pour une expérience indépendante sont montrés (*Figure 54A*).

Pour la lignée p53 déficiente, un pourcentage de fluorocytes proche de celui observé précédemment est obtenu pour les cellules éditées par la Cas9 dirigée contre *FAM175B* ($5,66 \pm 0,7\%$ contre $5,56 \pm 0,8\%$ précédemment) (*Figure 54B*). Pour les cellules éditées avec la Cas9^{D10A} et la dCas9, un taux de fluorocytes équivalent aux témoins négatifs est observé avec respectivement $0,16 \pm 0,01\%$ et $0,11 \pm 0,01\%$ de cellules fluorescentes.

Un profil similaire est retrouvé pour la lignée fonctionnel pour p53, avec $0,11 \pm 0,01\%$ de fluorocytes pour les cellules éditées avec la Cas9-nucléase dirigée contre *FAM175B* et seulement $0,03 \pm 0,01$ et $0,01 \pm 0,00\%$ de cellules fluorescentes pour celles édités respectivement avec la Cas9^{D10A} et la dCas9 (*Figure 54B*). Ces résultats suggèrent qu'une édition sans CDB permet une meilleure préservation de l'intégrité chromosomique.

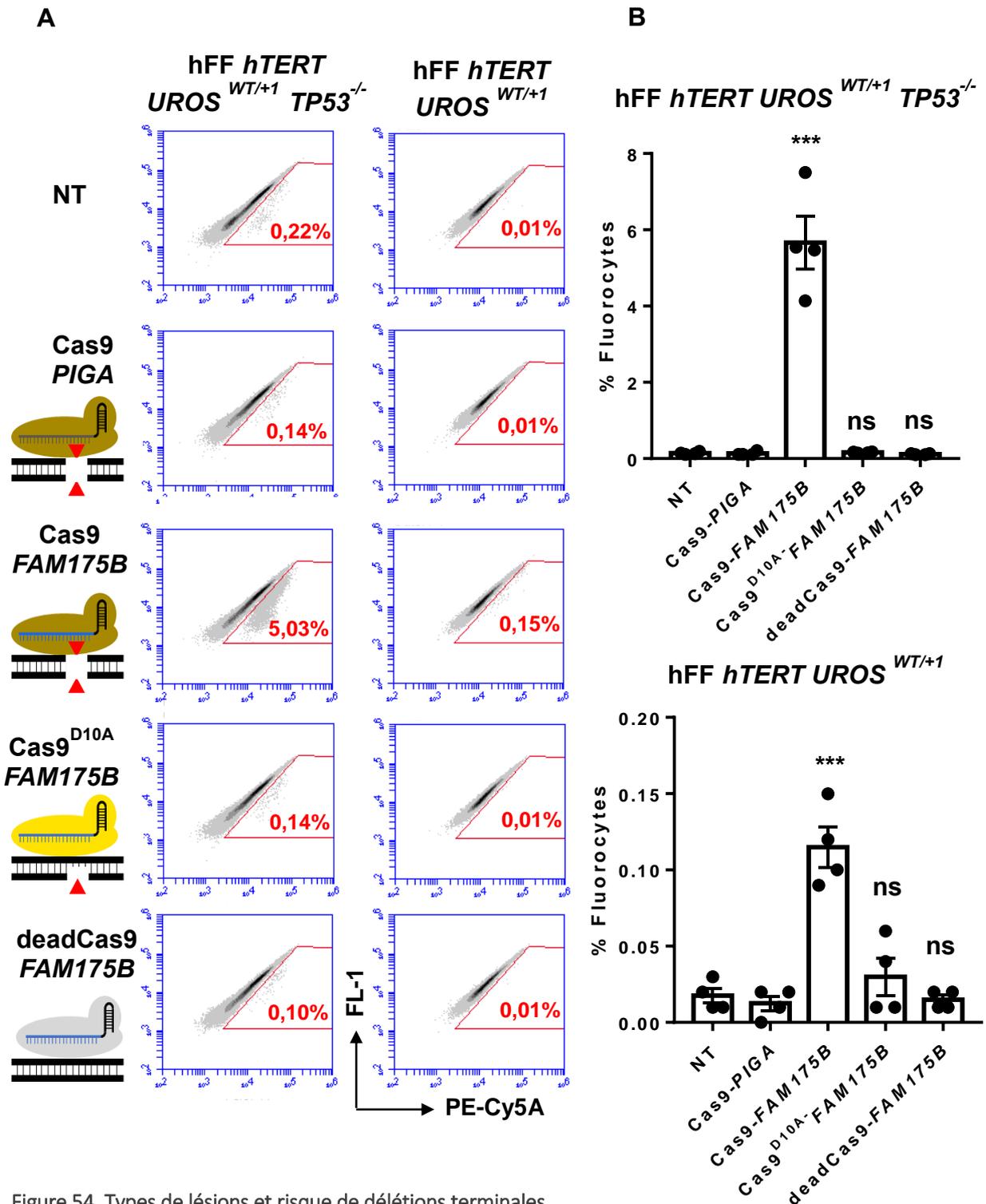


Figure 54. Types de lésions et risque de délétions terminales.

(A) Résultats représentatifs et (B) quantification du pourcentage de fluorocytes pour les deux lignées fibroblastiques transfectées avec soit des RNP Cas9:ARNg dirigées contre *PIGA* (Cas9-*PIGA*) ou contre *FAM175B* (Cas9-*FAM175B*), soit avec des RNP Cas9^{D10A}:ARNg ou deadCas9:ARNg dirigée contre *FAM175B* (respectivement Cas9^{D10A}-*FAM175B* et deadCas9-*FAM175B*). L'analyse est faite 14 jours après la transfection. Les cellules sont exposées 16h au 5-ALA, lavées deux fois au PBS et une période de clairance de 6h est observée avant l'analyse. Les résultats sont présentés avec la moyenne \pm écart-type (n=4, réplicats indépendants). Une one-way-ANOVA complétée par *unprotected Fisher's Least Significant Difference test* ont été utilisé pour inférer sur les différences statistiques par rapport à la condition Cas9-*PIGA*. ns, p>0,05 ; ***, p<0,001.

Discussion et perspectives

Quelles solutions pour éditer les cellules ?

D'un système immunitaire bactérien à un outil au service de la médecine, l'adaptation de la technologie CRISPR-Cas9 s'est faite rapidement. Si cette transformation se traduit par l'arrivée de protocoles cliniques, elle ne s'est pas faite sans soulever plusieurs limites techniques et interrogations vis-à-vis de sa sûreté. Ces questionnements, combinés au développement de méthodes alternatives d'édition génique indépendantes de la CDB, tendent à s'interroger sur le meilleur moyen d'éditer les cellules en vue d'une utilisation à un stade clinique.

i. Edition génique par la Cas9-nucléase

L'édition génique par la Cas9-nucléase semble être à double tranchant. En effet, l'ensemble des résultats obtenus avec cet outil indiquent qu'elle allie efficacité et génotoxicité non-voulue.

L'une des premières limites de la Cas9-nucléase à avoir été décrite est celle de son activité *off-target*. En effet, une forte activité *off-target* a été retrouvée lors de l'édition du locus UROS par la Cas9-nucléase, avec jusqu'à 32% d'InDels pour l'un des sites *off-target*. Cette observation concorde avec les données de la littérature et souligne l'intérêt d'utiliser les variants de la Cas9 à spécificité augmentée. Si la limite de la spécificité a largement été repoussée par l'apparition de ces nouveaux variants, celle de la génotoxicité *on-target* ne semble pas encore maîtrisée.

Dans une stratégie de *knock-in*, l'efficacité de la Cas9 se traduit par une forte capacité à induire de l'HDR. Cette efficacité a été observée ici avec 34,4% d'HDR dans des HEK293T et a été décrite à des niveaux similaires dans des CD34⁺ transfectées par des RNPs⁷⁸. Toutefois, il a été montré que cette efficacité s'accompagnait d'une forte présence d'InDels et menait à un ratio HDR/InDels souvent inférieur à 1. Si des exemples d'effets délétères possibles des InDels ont été présentés dans l'introduction⁶⁶, il est également à noter qu'un ratio inférieur à 1 rend difficile la création de modèle cellulaire homozygote pour une mutation.

Enfin, ce travail a permis de mettre en évidence une forme non-décrite de génotoxicité induite par la Cas9-nucléase : le risque de délétion terminale. Dans un premier temps, l'étude menée par FISH a montré que des délétions terminales de l'ordre du mégabase pouvaient survenir dans des cellules cancéreuses et des fibroblastes immortalisées et déficientes pour p53. Ces délétions terminales se traduisent par la perte d'un fragment de 8 Mb, du site de coupure dans *UROS* à l'extrémité du bras long du Chr.10. Ensuite, la mise en place d'un outil sensible de détection basé sur la cytométrie en flux a permis de montrer que des remaniements de l'ordre du mégabase pouvaient également avoir

lieu dans des fibroblastes immortalisés et fonctionnels pour p53 au niveau du locus *FAM175B* situé à 9Mb de l'extrémité du Chr.10.

Concernant les remaniements dans les fibroblastes fonctionnels pour p53, plusieurs précautions doivent être encore prises et certains points, explorés. Si une première validation de la méthode de détection a été faite par biologie moléculaire en analysant le locus *UROS* des fibroblastes *knock-out* pour p53 et fluorescents, l'étude cytogénétique (CGH ou FISH) reste nécessaire. Elle permettra de savoir si le phénotype fluorescent est dû à des délétions terminales, à de grandes délétions intra-chromosomiques supérieures au mégabase ou à l'occurrence des deux événements. Cette analyse devra également être faite sur les fibroblastes fonctionnels pour p53.

Certaines réserves ont également été émises *a posteriori* sur le choix du locus *FAM175B*. En effet, référencée sous le nom d'*ABRO-1*, une étude a montré que *FAM175B* jouait un rôle de suppresseur de tumeur en participant à la stabilisation de p53 en réponse aux CDBs¹⁸⁰. Une seconde étude a démontré que le *knock-out* de *FAM175B* entraînait une augmentation du nombre de ponts ultrafins d'anaphase (*Ultrafine anaphase bridges*) et de micronucléi¹⁸¹. Ces structures apparaissent lors de défaut de ségrégation des chromosomes pendant la mitose et sont des marqueurs de l'instabilité chromosomique. Ces deux études peuvent soulever des interrogations quant aux résultats obtenus, toutefois certains points peuvent être avancés pour conforter leur robustesse.

Le premier concerne la « temporalité » du *knock-out* de *FAM175B* lors de son édition par la Cas9-nucléase. En effet, les deux études ont utilisé des lignées préalablement invalidées pour *FAM175B* avant d'étudier les effets. Il est possible que les effets de l'invalidation de *FAM175B* ne soient pas immédiats et que l'expression basale, avant la lésion, de *FAM175B* suffise à produire un niveau de protéine permettant une signalisation normale le temps de sa réparation.

Si ce point est théorique, les deux suivants sont expérimentaux. Tout d'abord, il est à noter que des délétions terminales ont été observées au locus *UROS* dans une lignée de fibroblastes déficients pour p53, suggérant une indépendance à la fonction de *FAM175B*. Concernant les fibroblastes fonctionnels pour p53, un résultat préliminaire concordant a été obtenu en suivant le même protocole de détection par cytométrie, mais en ciblant le locus *SORCS1*. Le gène *SORCS1* est également situé en amont d'*UROS* (*Figure 55*). Ici, 0,08% de cellules fluorescentes ont été observées après édition des fibroblastes p53-fonctionnels (contre 0,04% pour la condition Cas9-*PIGA*). Ces résultats encore partiels doivent être confirmés, mais ils suggèrent que les délétions terminales peuvent survenir en ciblant des gènes non-impliqués dans la réparation de l'ADN.

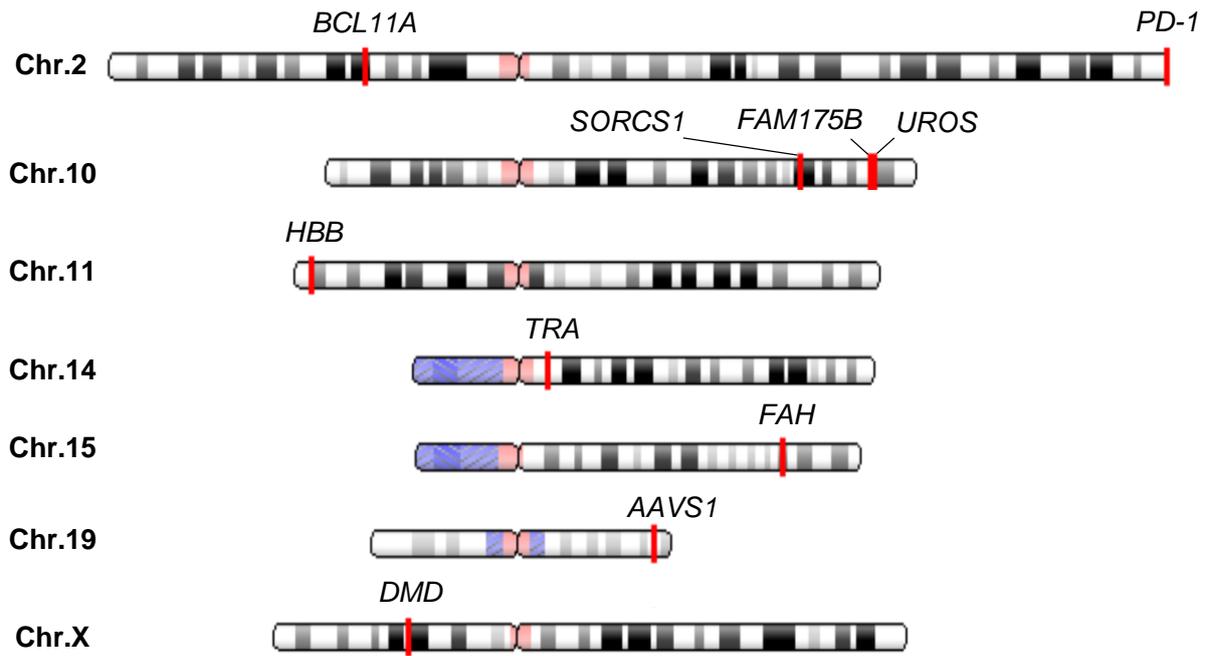


Figure 55. Position génomique des gènes d'intérêts cliniques.

Ces résultats soulignent le besoin de prêter une attention particulière aux délétions terminales, notamment en vue d'une approche clinique. A ce titre, plusieurs loci situés dans des régions subtélomériques méritent d'être soulignés (*Figure 55*) :

- Le gène *PD-1*, cible pour l'immunothérapie anti-cancéreuse, est situé à 300 kb de l'extrémité du bras long du Chr.2,
- Le gène *HBB* est situé à 5,2 Mb de l'extrémité du bras court du Chr. 11.
- *AAVS1*, souvent considéré comme un site *safe-harbor*, est situé à 3,6 Mb de l'extrémité du bras long du Chr.19.

A l'inverse, d'autres gènes d'intérêts cliniques pourraient paraître davantage « protégés » du risque de délétion terminale lors de leur édition par la Cas9-nucléase grâce à leur position plus « péricentromérique ». Pour illustrer, les gènes *TRA*, *BCL11A*, *DMD*, et *FAH* sont respectivement situés à 86, 60, 33 et 21 Mb d'une extrémité de chromosome. L'hypothèse sous-jacente formulée ici serait celle d'une relation entre la distance au télomère et la fréquence de délétion terminale. Toutefois, un résultat préliminaire obtenu lors de l'édition de fibroblastes invalidés pour p53 au locus *SORCS1* pourrait dédire cette hypothèse. En effet, une fréquence supérieure de cellules fluorescentes a été observée après édition du locus *SORCS1* comparée au locus *FAM175B* (5,97% contre 4,11%). Or, *SORCS1* est situé à 27 Mb de l'extrémité du Chr.10 contre 8 Mb pour *FAM175B*. Cette observation suggère que la fréquence de délétions terminales découle d'une équation plus complexe qu'une relation distance-fréquence et incite à une évaluation systématique du risque de délétion terminale quel que soit le locus visé. Malgré tout, les résultats étant préliminaires et en attente d'être reproduits, cette conclusion doit être mise en perspective.

Pour compléter ce travail d'évaluation des risques dans des protocoles à visée clinique, l'une des perspectives immédiates sera de savoir si ces délétions terminales sont également retrouvées dans des cellules d'intérêt clinique, notamment les CSHs et lymphocytes T. De plus, une évaluation de l'impact fonctionnel des délétions terminales, notamment sur leur capacité à conférer un avantage prolifératif, pourrait être précieuse.

Si, parce que les essais cliniques sont pour certains d'ores et déjà lancés, il est important d'avoir une évaluation rapide de ce risque pour les applications cliniques, une meilleure compréhension des mécanismes qui le régissent pourrait permettre de le diminuer, voire de l'abolir.

Un premier élément de réponse apporté par les résultats présentés ici est la dépendance à p53. En effet, l'invalidation de p53 multiplie par 60 la fréquence de délétions terminales. Si l'implication du « gardien du génome » dans la protection de l'intégrité génomique pouvait être attendue, ces

résultats sont à mettre en perspective avec ceux obtenus par les études d'Haapaniemi et al. et Ihry et al.^{152,153}. Ces deux études décrites précédemment ont montré que l'édition génique par *knock-in* induisait un biais de sélection pour les cellules ayant un statut p53 faible ; ces dernières étant plus tolérantes aux modifications géniques. Partant de ce postulat, il a été montré qu'une inhibition transitoire de p53 permettait d'augmenter le taux d'HDR. Suivant ces observations, il sera intéressant d'étudier le statut p53 des cellules fluorescentes issues de l'édition de fibroblastes *a priori* fonctionnels pour p53. En effet, il est plausible que les cellules fluorescentes soient celles qui ont une activité p53 faible, et donc un risque de délétions terminales augmenté. L'ensemble de ces résultats pourrait souligner l'aspect double-tranchant de l'inhibition de p53, augmentant à la fois la génotoxicité voulue et non-désirée.

Il est peu probable que la perte de fragment de chromosome soit un processus physiologique pour la cellule et l'hypothèse peut donc être émise que cet événement découle d'un défaut de reconnaissance ou réparation de la CDB. A ce sujet, l'étude de Do et al. mérite d'être soulignée¹⁸². Dans ce travail, une CDB unique a été induite avec un ZFN ciblant le gène *MLL* dans des K-562, lignées cancéreuses avec un caryotype complexe. Les cellules ont ensuite été marquées par FISH avec deux sondes *break-apart* ; si le locus est intègre, un signal jaune (*merged*) est observé, si l'intégrité est affectée, deux signaux rouges et verts distincts apparaissent. Suite à l'édition, 6,6% des cellules présentent un réarrangement chromosomique contre 2,3% pour les cellules non-éditées. De plus, ce taux augmente à 10,3% lorsque les cellules sont pré-traitées avec du NU7026, un inhibiteur de la DNA-PK, acteur clé de la cNHEJ. De la même façon, dans le papier princeps sur la délétion de chromosome entier, Zuo et al. ont montré que le blocage de la cNHEJ augmentait le taux de délétion¹⁶⁵. Ces résultats montrent le lien entre une déficience dans les voies de réparation et apparition de réarrangement chromosomique. Ce même lien de causalité pourrait sous-tendre l'apparition de délétions terminales. L'utilisation d'inhibiteurs et d'activateurs spécifiques de chacune des voies de réparation pourrait conduire à mieux comprendre la participation de chacune dans cet événement. Ce travail pourrait également permettre d'évaluer l'effet sur le risque de délétions terminales des différents inhibiteurs proposés pour augmenter le taux d'HDR ; l'hypothèse étant ici émise que l'inhibition d'acteurs clés de la réparation mènera à l'augmentation du risque. Sur ce point, cette étude pourra être complétée par l'effet des Cas9 de fusion ayant également pour but d'augmenter l'HDR^{158,159}.

Enfin, si des délétions terminales suite à une CDB n'avaient pas été décrites dans le cadre de l'édition génique, celles-ci ont largement été rapportées dans le cadre de la cytogénétique constitutionnelle et en cancérologie. Dans ce domaine, la délétion terminale du bras court du Chr.17 bien documentée. En effet, cette délétion entraîne la perte d'une copie de *TP53* et est associé

à un phénotype cancéreux agressif¹⁸³. En cytogénétique constitutionnelle, le syndrome du cri du chat (délétion du bras court du Chr.5) en est un exemple connu. Plus rare, mais d'un intérêt particulier ici, c'est le syndrome de délétion du 10q26, traduit dans certains cas par des délétions terminales du bras long du Chr.10 allant jusqu'à 17 Mb¹⁸⁴. Toutefois, si les délétions terminales sont constatées, peu d'études sont consacrées aux mécanismes qui permettent leur apparition. Malgré tout, deux hypothèses pourraient expliquer cet événement. La première est que la cellule crée *de novo* un télomère au niveau d'une CDB lorsque des défauts de régulation lui empêchent de restreindre l'activité télomérase à la réparation des télomères^{185,186}. Cette explication implique une activité télomérase présente dans le métabolisme cellulaire. Si cet impératif pourrait exclure certains types cellulaires différenciés, les cellules les plus souches, particulièrement intéressantes pour la thérapie génique, sont probablement sujettes à ce type de dysfonctionnement. Si cette hypothèse est issue de travaux faits chez la levure, la seconde provient de l'analyse d'une cohorte de 50 patients porteurs de délétion terminale combinée à une duplication¹⁸⁷. Ici, l'idée est que ces événements proviennent d'une réparation anormale de la CDB avec un repliement sur soi du brin réséqué (*fold-back*) (Figure 56). Cette étude donne un modèle théorique élégant expliquant à la fois les délétions terminales et la délétion terminale avec une duplication en amont observée lors de l'analyse par CGH de clones HEK293T. Il est possible que cette réparation anormale fasse intervenir une voie hautement mutagène, la *microhomology-mediated break-induced replication* (MMBIR). Le MMBIR est impliqué dans la réparation des CDBs survenant lors de la réplication de l'ADN et est notamment capable de replier sur lui-même un brin réséqué au niveau de régions homologues (*fold-back*)¹¹⁸. Là où les délétions terminales sont un souci pour l'édition génique, la Cas9-nucléase pourrait permettre de faciliter la création de modèles cellulaires porteurs de délétions terminales et permettre une meilleure connaissance de leur origine et de leurs conséquences.

Enfin, si un nouveau risque a été souligné dans le travail présenté ici, le tableau de la génotoxicité induite par la Cas9-nucléase n'est peut-être pas complet. En effet, un des événements mentionné juste ci-dessus et auquel le système de détection par cytométrie est complètement aveugle est celui de la duplication de séquence. Toutefois, l'analyse par CGH des cellules fluorescentes, porteuses de délétions terminales, pourrait permettre de savoir si elles sont accompagnées ou non de duplications en amont. Le développement de systèmes sensibles de détection spécifiques des duplications pourrait être un atout précieux pour évaluer davantage la sûreté des procédés d'édition génique sur la stabilité chromosomique.

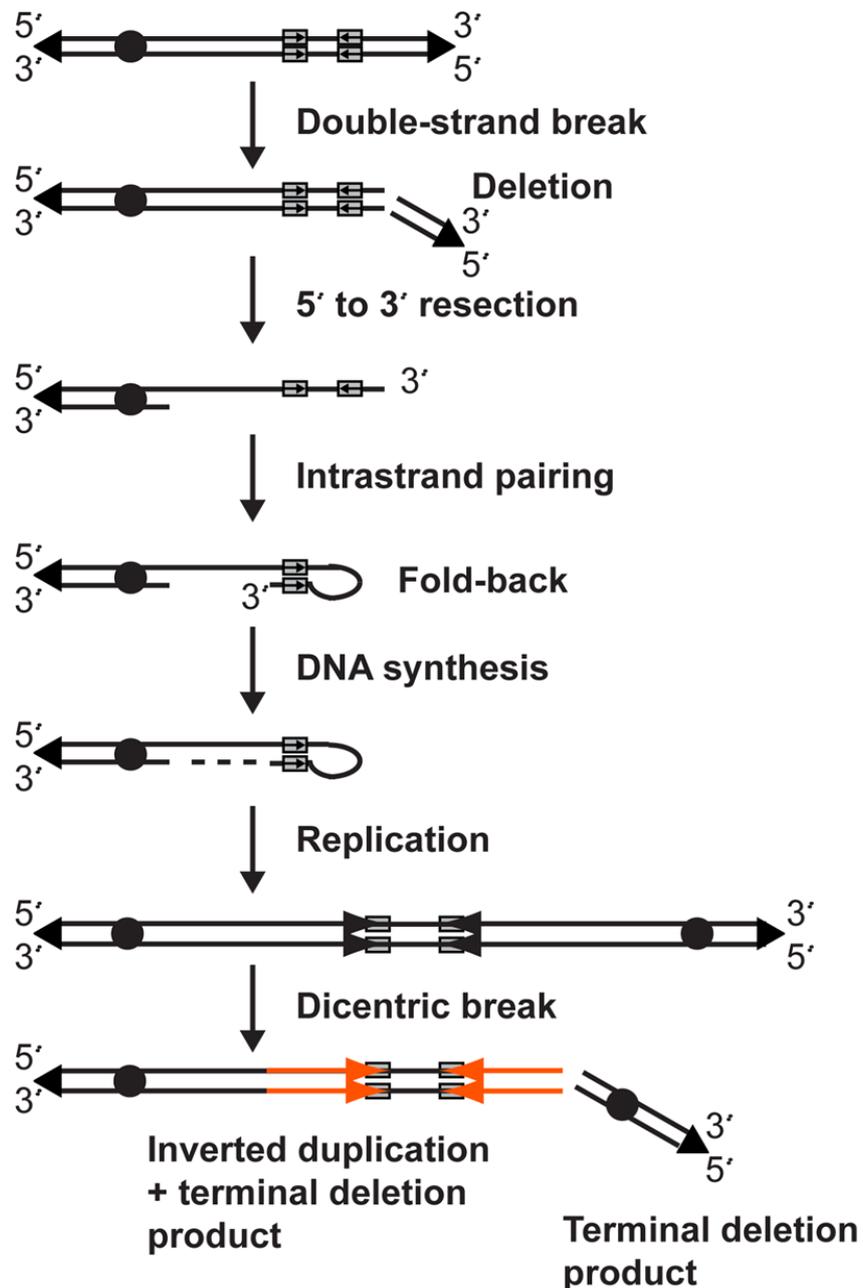


Figure 56. Modèle de formation des délétions terminales et des duplications inversées. (modifiée d'après Hermetz et al. Plos Genet, 2014)

L'élément déclencheur est la survenue d'une CDB. Cette dernière est réséquée pour faire apparaître une extrémité 3' sortantes. Dans certains cas, il est proposé que cette extrémité se replie sur elle-même au niveau de séquences inversées homologues (*fold-back*). Une synthèse d'ADN permet de raccorder l'extrémité repliée et le brin réséqué. Un chromosome sans télomère à l'extrémité clivé apparaît. La réplication de ce chromosome mène à la formation d'un chromosome dicentrique. Ce dernier est instable lors de la division cellulaire et une seconde CDB survient entre les deux centromères. Deux produits sont formés : un chromosome avec une délétion terminale et un second avec une délétion terminale et une duplication inversée. Ces deux structures peuvent être stabilisées par l'ajout *de novo* d'un télomère.

ii. Édition génique par la Cas9^{D10A}, une solution alternative viable ?

L'ensemble des résultats obtenus lors de la caractérisation de la génotoxicité induite par la Cas9-nucléase a souligné le besoin de développer une méthode plus sûre et moins génotoxique. La seconde partie du travail a été d'utiliser et d'améliorer une approche alternative d'édition basée sur la Cas9^{D10A}-nickase.

Comparée à la Cas9-nucléase, la Cas9^{D10A}-nickase permet une édition génique avec une génotoxicité fortement diminuée, ce qui constitue son intérêt et sa limite. À l'inverse de la nucléase, cette approche d'édition génique induit peu d'évènements non-voulus mais est également peu efficace, avant optimisation, pour faire de l'HDR.

En revenant d'abord sur l'intérêt d'une approche sans CDB, le premier constat est celui d'absence de lésion au niveau des sites *off-target*. En effet, bien que l'ARNg possède la même spécificité, les CSBs sont peu mutagènes et le risque de perte d'information génétique aux sites *off-target* est supprimé.

Dans une stratégie de *knock-in*, l'édition avec la Cas9^{D10A}-nickase se traduit par un ratio HDR/InDels élevé, largement supérieur à 1. Ce ratio a été rendu possible par une diminution drastique des InDels formés au site de coupure. Toutefois, le taux d'HDR est également diminué, rendant cette approche difficilement applicable tant pour la création de modèles cellulaires que pour une perspective thérapeutique.

Augmenter le taux d'HDR tout en conservant un ratio HDR/Indels élevé a donc été un enjeu majeur pour faire passer cette approche d'une possibilité théorique à pratique. Une première optimisation a été faite en modifiant la concentration et la chimie des ssODNs. Ce travail a permis de multiplier par 5 le taux d'HDR (de 3,8% à 18,5%) avec un ratio HDR/InDels de 26,4. Une étude plus extensive pourrait être conduite pour perfectionner davantage les matrices pour une édition basée sur la Cas9^{D10A}-nickase.

L'analyse par NGS a montré qu'après un cycle d'édition avec la Cas9^{D10A} et une matrice de réparation, seule une minorité des allèles sont modifiés. À partir de ce constat, l'idée a été d'augmenter artificiellement la fenêtre pendant laquelle l'édition génique pouvait avoir lieu en effectuant des cycles itératifs d'édition génique. En utilisant cette méthodologie, un taux d'HDR équivalent à celui obtenu avec la Cas9-nucléase est atteint, tout en conservant une fréquence d'InDels minimale. Cependant, cette méthodologie, basée sur des électroporations successives, se voit limitée par sa cytotoxicité et par le temps nécessaire à sa mise en place. En effet, chaque électroporation est un évènement traumatique pour la cellule, nécessitant d'espacer dans le temps

deux cycles successifs. Ce procédé dans sa forme actuelle se prête davantage à la création de lignées cellulaires qu'à la modification de cellules primaires non-immortalisées. Toutefois, l'utilisation de nouvelles méthodes de transfection moins cytotoxiques pourrait permettre de palier à cette limite.

Enfin le taux d'HDR initié par la Cas9^{D10A} pourrait être amélioré par la modulation des voies de réparation. L'apport au moment de la transfection d'un siRNA dirigée contre RAD51¹⁷¹ ou du dominant négatif de RAD51, RAD51(K133R)¹⁷¹, permet d'augmenter le taux d'HDR. A partir de ce résultat, une étude s'est intéressé à former une Cas9^{D10A} fusionnée avec RAD51(K133R) pour posséder une action plus localisée qu'une drogue¹⁸⁸. Le knock-out de XRCC1 est également associé à une augmentation du taux d'HDR initié aux CSBs¹⁷³. A terme, une meilleure compréhension de la réparation des CSBs par des ssODNs pourrait permettre d'identifier de nouvelles cibles pharmacologiques pour augmenter l'HDR.

Enfin, l'étude de la génotoxicité de la Cas9^{D10A}-nickase à l'échelle du chromosome a permis de montrer une nette diminution du risque de délétions terminales. Dans la première partie du travail présenté ici, seul un clone HEK293T édité avec la Cas9^{D10A} a été observé alors que l'analyse par FISH d'une population polyclonale n'était pas significativement différente de la population non-éditée. De plus, aucune augmentation du nombre de cellules fluorescentes n'a été constatée par cytométrie après édition des deux lignées fibroblastiques par la Cas9^{D10A}. Toutefois, une réserve sera émise vis-à-vis de l'absence de cellules fluorescentes après édition par la Cas9^{D10A}. En effet, il est difficile de juger de l'efficacité de coupure de la Cas9^{D10A} dans les cellules éditées. Or, seul un faible pourcentage d'InDels est observé après la coupure avec la Cas9-nucléase, pouvant témoigner d'une faible efficacité de coupure ou d'évènements non-détectés par une PCR classique. Si dans le cas de la Cas9-nucléase, une efficacité de coupure sous-optimale se traduirait par une sous-estimation du risque de délétions terminales, il est nécessaire d'avoir une efficacité maximale pour affirmer l'absence de délétions terminales suite à l'utilisation de la Cas9^{D10A}. Les ARNg ont été validées *in vitro* en ciblant des produits de PCR contenant la séquence cible et un travail pour comprendre leur faible efficacité *in cellulo* est en cours.

Malgré tout, l'ensemble de ces résultats suggèrent que les délétions terminales induites par la Cas9^{D10A} sont rares dans les lignées cancéreuses aux caryotypes complexes, et inexistantes, ou suffisamment faibles pour ne pas être détectées, dans des lignées diploïdes immortalisées et fonctionnelles ou non pour p53.

Si l'édition génique avec la Cas9^{D10A} pourrait être une solution alternative à la Cas9, la preuve de son efficacité dans des cellules d'intérêt cliniques reste à faire. En effet, la littérature consacrée à l'édition génique par la Cas9^{D10A} est conduite *in vitro* et utilise principalement des lignées cellulaires

relativement éloignées des cellules d'intérêt clinique. Dans le cadre de cette thèse, les preuves concernant la Cas9^{D10A} ont également été acquises dans des lignées cancéreuses ou fibroblastiques. Toutefois, il est à noter qu'une preuve d'édition a été obtenue dans des cellules souches pluripotentes induites avec un taux de modification de 1%¹⁸⁸. Ce résultat a été permis par l'utilisation de la Cas9^{D10A} fusionnée au dominant négatif de RAD51. Néanmoins, la confirmation de ces résultats sur des cellules primaires, et notamment les CSH, est un enjeu crucial pour déterminer si l'édition génique par la Cas9^{D10A} est une solution alternative viable.

Néanmoins, la Cas9^{D10A} pourrait permettre de pallier à une des limites de l'édition génique par la Cas9 : l'édition par HDR des cellules quiescentes et post-mitotiques. Les cellules post-mitotiques et quiescentes sont réfractaires à l'édition génique par la Cas9-nucléase, s'expliquant par une dépendance des voies de l'HDR au cycle cellulaire^{77,189}. L'une des conséquences est une diminution du potentiel de greffe des cellules CD34⁺ éditées, traduit par une diminution dans le temps de la fraction de cellules corrigées⁷⁸. Cette observation pourrait traduire une édition préférentielle des cellules progénitrices plutôt que véritablement souches.

A l'inverse, la réparation des CSBs étant moins dépendante du cycle cellulaire, il est envisageable qu'un système d'édition basé sur la Cas9^{D10A} puisse autoriser la modification des cellules quiescentes, en assurant un maintien de la correction dans le temps. Dans un cadre plus large, l'édition génique des cellules post-mitotiques pourrait particulièrement intéresser le domaine des neurosciences en offrant un outil d'édition par HDR des neurones.

iii. Vers d'autres méthodes d'édition génique indépendantes de la CDB ?

Quatre solutions alternatives, indépendantes de la CDB, ont récemment été décrites et utilisent de nouvelles méthodes d'édition. De plus, si le passage d'une cassure double-brin à une coupure simple-brin limite l'instabilité chromosomique induite, certaines approches d'édition génique ne reposent sur aucune coupure et s'inscrivent dans la continuité de ce concept.

Le premier système repose sur l'ouverture de l'ADN par l'action de trois dCas9 ciblant des séquences en tandem¹²⁸. Ici, bien qu'aucune coupure ne soit induite, il est supposé que la séparation des brins permet l'hybridation de la matrice simple-sbrin et l'incorporation des mutations d'intérêts. Toutefois, l'efficacité reste faible avec 0,7% d'allèles modifiés dans des lignées cancéreuses et relativement complexe à vectoriser dû à la présence de trois ARNg différents.

Une seconde approche provient de la diversité des systèmes CRISPR/Cas. La protéine Cas12k a été mise en lumière par le groupe de Feng Zhang en juillet 2019¹⁹⁰ (*Figure 57A*).

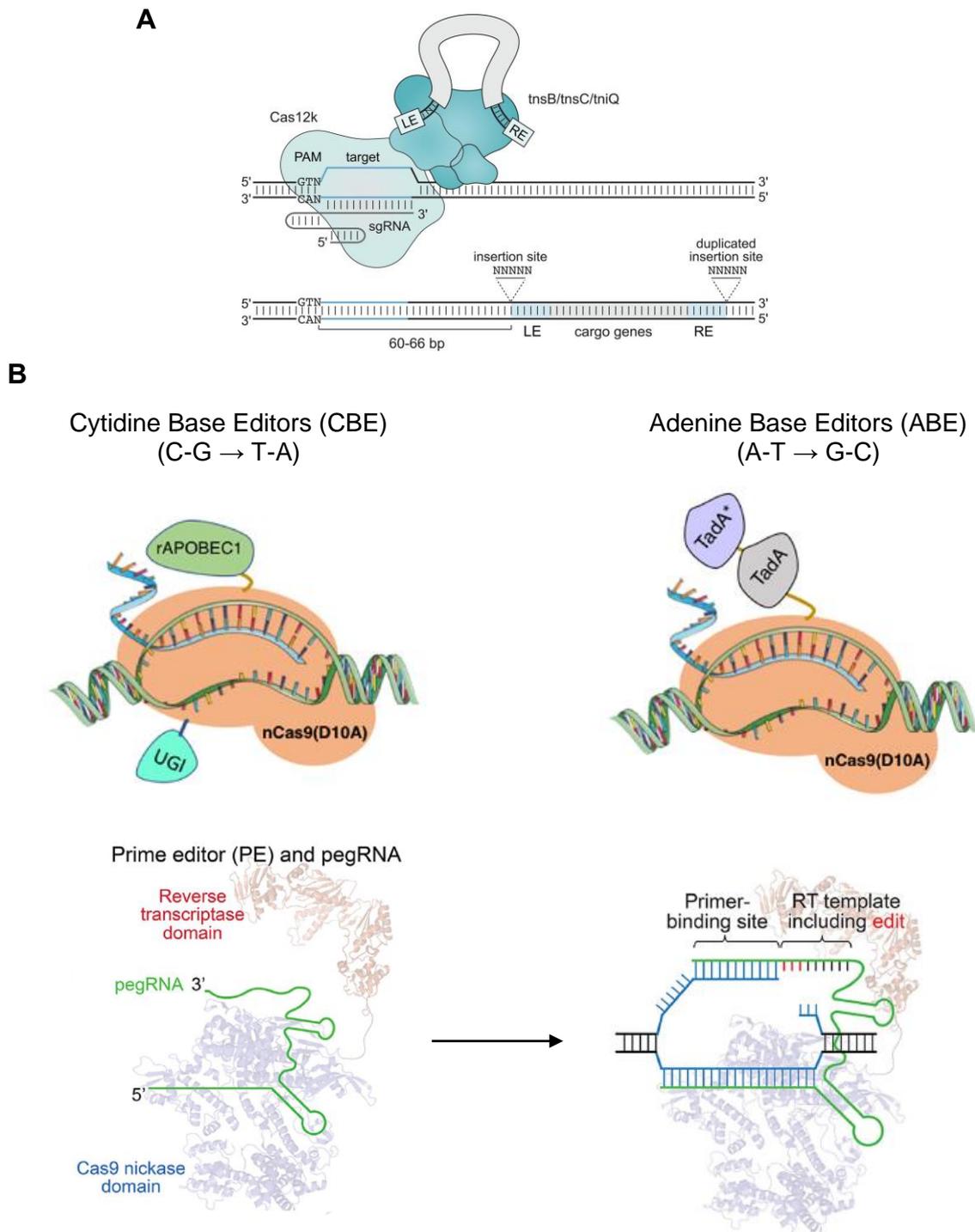


Figure 57. Vers d'autres méthodes d'édition génique ?

(A) Système CAST. La Cas12k est associée à des éléments tnsB/tnsC/tniQ qui font partie de la machinerie des transposons. Ce système permet l'intégration d'une séquence « cargo » jusqu'à 10 kb. (d'après Strecker et al., Science, 2019)

(B) Systèmes par édition de base. La Cas9^{D10A} est fusionnée à une cytidine déaminase (rAPOBEC1) ou à une adénosine déaminase (TadA). Le groupement UGI (uracil DNA glycosylase inhibitor) permet d'améliorer l'efficacité d'édition. (modifiée d'après <https://blog.addgene.org/>)

(C) Système de *prime-editing*. Le *pegRNA* constitue la partie ARN. La partie protéique est constituée de la Cas9^{H840A}-nickase (en bleu) fusionnée à une *reverse transcriptase*. Les mutations d'intérêts portées par l'ARN et à incorporer sont montrées en rouge sur le schéma de droite (modifiée d'après Anzalone et al., Nature, 2019)

Ce système CRISPR-Cas a la particularité de posséder uniquement les éléments minimums nécessaires au ciblage et à la formation d'une *R-loop*, mais la Cas12k est dépourvu d'activité endonucléasique. En parallèle, ce système CRISPR-Cas est associé à des Tn7-like transposons. A la croisée de ces deux observations, il a été supposé que les transposons se servent de la machinerie CRISPR-Cas pour permettre la formation d'une *R-loop* et faciliter leur intégration¹⁹¹. Ce système biologique a été détourné pour servir d'outil d'édition génique. Le complexe CAST (pour *CRISPR-associated transposase*) est formé de la Cas12k, d'un ARN-guide et des protéines du système transposon ainsi que la séquence « cargo » d'intérêt à introduire. Ce complexe permet l'intégration unidirectionnelle de la séquence « cargo » entre 60 à 66 nt en aval du PAM. Les avantages de ce système sont considérables permettant une intégration unidirectionnelle, ciblée et allant jusqu'à 10kb. Cet outil permettrait d'envisager le remplacement entier d'exon porteur de mutation ponctuelle.

L'édition de base peut véritablement marquer le début de l'acte deux de l'édition génique (*Figure 57B*). Les premières générations d'éditeurs de base reposent sur une dCas9 fusionnée avec une enzyme permettant la modification de base, à savoir soit une cytidine déaminase¹⁹² ou une adénine déaminase¹⁹³. Ici, la dCas9 sert uniquement de guide pour amener l'enzyme au niveau des bases à modifier. Bien que des problèmes de spécificité et d'activité hors-cible, inhérents au développement de nouvelle technologie de ce type, aient été décrits, l'édition de base est une nouvelle approche prometteuse pour l'édition génique. Toutefois, les dernières générations d'éditeurs de base utilisent la Cas9^{D10A} plutôt que la dCas9. Si cette transition, d'une absence de lésion à une cassure simple-brin, permet d'améliorer le taux de modifications, il serait intéressant d'évaluer l'impact de ce changement d'un point de vue de la stabilité chromosomique.

Enfin, un nouveau système a été décrit en octobre 2019 par l'équipe de David Liu, pionner de l'édition de base¹⁹⁴ (*Figure 57C*). Le *prime-editing* répond à l'une des limites principales de l'édition de base. En effet, si l'édition de base est limitée aux mutations de type transitions, le *prime editing* autorise également les transversions, avec en premier lieu la transversion T vers A pour permettre la correction de la drépanocytose. Dans ces systèmes les deux composants de la RNP sont fortement modifiés. La partie ARN est toujours composée d'une partie « guide » permettant le ciblage de l'ADN, mais contient désormais également la séquence porteuse de mutations à introduire ; cet ensemble constitue le *prime editing guide RNA* (pegRNA). La partie protéique est composée d'une Cas9-nickase (ici, H840A) fusionnée à une *reverse transcriptase* (RT), chargée de convertir le fragment d'ARN donneur en ADN. Une preuve de principe d'édition a été obtenue avec la correction de la mutation de la drépanocytose dans des lignées cellulaires, avec un ratio

HDR/InDels très élevé. Une preuve a également été faite dans des cellules post-mitotiques avec la modification *in vitro* de neurones.

Ces méthodes d'édition génique pourraient constituer des solutions alternatives pour pallier aux limites de la génotoxicité de la Cas9, tout en conservant, pour certaines, une forte efficacité. Toutefois, ici encore, des travaux devront encore être conduits pour permettre une maîtrise de la spécificité, une amélioration de l'efficacité et avoir une caractérisation complète du tableau de leur génotoxicité, autorisant à terme des applications cliniques prometteuses.

Bibliographie

- 1 Blaese RM, Culver KW, Miller AD, Carter CS, Fleisher T, Clerici M *et al.* T Lymphocyte-Directed Gene Therapy for ADA– SCID: Initial Trial Results After 4 Years. *Science* 1995; **270**: 475–480.
- 2 Rogers S, Pfuderer P. Use of Viruses as Carriers of Added Genetic Information. *Nature* 1968; **219**: 749–751.
- 3 Cavazzana M, Bushman FD, Miccio A, André-Schmutz I, Six E. Gene therapy targeting haematopoietic stem cells for inherited diseases: progress and challenges. *Nat Rev Drug Discov* 2019; **18**: 447–462.
- 4 Braun CJ, Boztug K, Paruzynski A, Witzel M, Schwarzer A, Rothe M *et al.* Gene Therapy for Wiskott-Aldrich Syndrome—Long-Term Efficacy and Genotoxicity. *Science Translational Medicine* 2014; **6**: 227ra33–227ra33.
- 5 Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* 2008; **118**: 3132–3142.
- 6 Stein S, Ott MG, Schultze-Strasser S, Jauch A, Burwinkel B, Kinner A *et al.* Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. *Nat Med* 2010; **16**: 198–204.
- 7 Cattoglio C, Pellin D, Rizzi E, Maruggi G, Corti G, Miselli F *et al.* High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. *Blood* 2010; **116**: 5507–5517.
- 8 Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, Hubank M, Kempinski H *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest* 2008; **118**: 3143–3150.
- 9 Cavazzana-Calvo M, Fischer A. Gene therapy for severe combined immunodeficiency: are we there yet? *J Clin Invest* 2007; **117**: 1456–1465.
- 10 Thornhill SI, Schambach A, Howe SJ, Ulaganathan M, Grassman E, Williams D *et al.* Self-inactivating gammaretroviral vectors for gene therapy of X-linked severe combined immunodeficiency. *Mol Ther* 2008; **16**: 590–598.
- 11 Río P, Navarro S, Wang W, Sánchez-Domínguez R, Pujol RM, Segovia JC *et al.* Successful engraftment of gene-corrected hematopoietic stem cells in non-conditioned patients with Fanconi anemia. *Nat Med* 2019; **25**: 1396–1401.
- 12 Cavazzana-Calvo M, Payen E, Negre O, Wang G, Hehir K, Fusil F *et al.* Transfusion independence and HMGA2 activation after gene therapy of human β -thalassaemia. *Nature* 2010; **467**: 318–322.
- 13 Cesana D, Sgualdino J, Rudilosso L, Merella S, Naldini L, Montini E. Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. *J Clin Invest* 2012; **122**: 1667–1676.

- 14 Moiani A, Paleari Y, Sartori D, Mezzadra R, Miccio A, Cattoglio C *et al.* Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. *J Clin Invest* 2012; **122**: 1653–1666.
- 15 Cesana D, Ranzani M, Volpin M, Bartholomae C, Duros C, Artus A *et al.* Uncovering and Dissecting the Genotoxicity of Self-inactivating Lentiviral Vectors In Vivo. *Mol Ther* 2014; **22**: 774–785.
- 16 Papapetrou EP, Lee G, Malani N, Setty M, Riviere I, Tirunagari LMS *et al.* Genomic safe harbors permit high β -globin transgene expression in thalassemia induced pluripotent stem cells. *Nat Biotechnol* 2011; **29**: 73–78.
- 17 Bedel A, Taillepierre M, Guyonnet-Duperat V, Lippert E, Dubus P, Dabernat S *et al.* Metabolic Correction of Congenital Erythropoietic Porphyria with iPSCs Free of Reprogramming Factors. *Am J Hum Genet* 2012; **91**: 109–121.
- 18 Scrudato ML, Poulard K, Sourd C, Tomé S, Klein AF, Corre G *et al.* Genome Editing of Expanded CTG Repeats within the Human DMPK Gene Reduces Nuclear RNA Foci in the Muscle of DM1 Mice. *Molecular Therapy* 2019; **27**: 1372–1388.
- 19 Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* 1985; **317**: 230–234.
- 20 Choulika A, Perrin A, Dujon B, Nicolas JF. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol Cell Biol* 1995; **15**: 1968–1973.
- 21 Donoho G, Jasin M, Berg P. Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. *Mol Cell Biol* 1998; **18**: 4070–4078.
- 22 Porteus MH, Baltimore D. Chimeric Nucleases Stimulate Gene Targeting in Human Cells. *Science* 2003; **300**: 763–763.
- 23 Arnould S, Chames P, Perez C, Lacroix E, Duclert A, Epinat J-C *et al.* Engineering of Large Numbers of Highly Specific Homing Endonucleases that Induce Recombination on Novel DNA Targets. *Journal of Molecular Biology* 2006; **355**: 443–458.
- 24 Ashworth J, Havranek JJ, Duarte CM, Sussman D, Monnat RJ, Stoddard BL *et al.* Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* 2006; **441**: 656–659.
- 25 Mussolino C, Mlambo T, Cathomen T. Proven and novel strategies for efficient editing of the human genome. *Current Opinion in Pharmacology* 2015; **24**: 105–112.
- 26 Rivière J, Hauer J, Poirot L, Brochet J, Souque P, Mollier K *et al.* Variable correction of Artemis deficiency by I-Sce1-meganuclease-assisted homologous recombination in murine hematopoietic stem cells. *Gene Ther* 2014; **21**: 529–532.
- 27 Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA* 1996; **93**: 1156–1160.

- 28 Urnov FD, Miller JC, Lee Y-L, Beausejour CM, Rock JM, Augustus S *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 2005; **435**: 646–651.
- 29 Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G *et al.* Gene Editing of CCR5 in Autologous CD4 T Cells of Persons Infected with HIV. *N Engl J Med* 2014; **370**: 901–910.
- 30 Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* 2009; **326**: 1509–1512.
- 31 Gu K, Yang B, Tian D, Wu L, Wang D, Sreekala C *et al.* R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 2005; **435**: 1122–1125.
- 32 Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A *et al.* Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* 2010; **186**: 757–761.
- 33 Wienert B, Funnell APW, Norton LJ, Pearson RCM, Wilkinson-White LE, Lester K *et al.* Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. *Nature Communications* 2015; **6**: 7085.
- 34 Ramalingam S, Annaluru N, Kandavelou K, Chandrasegaran S. TALEN-mediated generation and genetic correction of disease-specific human induced pluripotent stem cells. *Curr Gene Ther* 2014; **14**: 461–472.
- 35 Qiu Z, Liu M, Chen Z, Shao Y, Pan H, Wei G *et al.* High-efficiency and heritable gene targeting in mouse by transcription activator-like effector nucleases. *Nucleic Acids Res* 2013; **41**: e120.
- 36 Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014; **346**: 1258096.
- 37 Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology* 1987; **169**: 5429–5433.
- 38 Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007; **315**: 1709–1712.
- 39 Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slikhuis RJH, Snijders APL *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* 2008; **321**: 960–964.
- 40 Marraffini LA, Sontheimer EJ. CRISPR Interference Limits Horizontal Gene Transfer in *Staphylococci* by Targeting DNA. *Science* 2008; **322**: 1843–1845.
- 41 Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012; **337**: 816–821.
- 42 Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 2007; **8**: 172.
- 43 Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P *et al.* Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 2011; **9**: 467–477.

- 44 Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Microbiol* 2015; **13**: 722–736.
- 45 Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR–Cas systems. *Curr Opin Microbiol* 2017; **37**: 67–78.
- 46 Cox DBT, Gootenberg JS, Abudayyeh OO, Franklin B, Kellner MJ, Joung J *et al.* RNA Editing with CRISPR–Cas13. *Science* 2017; **358**: 1019–1027.
- 47 Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ *et al.* RNA targeting with CRISPR–Cas13a. *Nature* 2017; **550**: 280–284.
- 48 Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013; **339**: 819–823.
- 49 Jiang F, Doudna JA. CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics* 2017; **46**: 505–529.
- 50 Saprunauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* 2011; **39**: 9275–9282.
- 51 Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 2011; **471**: 602–607.
- 52 Traxler EA, Yao Y, Wang Y-D, Woodard KJ, Kurita R, Nakamura Y *et al.* A genome-editing strategy to treat β -hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nat Med* 2016; **22**: 987–990.
- 53 Canver MC, Smith EC, Sher F, Pinello L, Sanjana NE, Shalem O *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* 2015; **527**: 192–197.
- 54 Miao CH, Nakai H, Thompson AR, Storm TA, Chiu W, Snyder RO *et al.* Nonrandom transduction of recombinant adeno-associated virus vectors in mouse hepatocytes in vivo: cell cycling does not influence hepatocyte transduction. *J Virol* 2000; **74**: 3793–3803.
- 55 Wu Z, Yang H, Colosi P. Effect of genome size on AAV vector packaging. *Mol Ther* 2010; **18**: 80–86.
- 56 Truong D-JJ, Kühner K, Kühn R, Werfel S, Engelhardt S, Wurst W *et al.* Development of an intein-mediated split–Cas9 system for gene therapy. *Nucleic Acids Res* 2015; **43**: 6450–6458.
- 57 Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ *et al.* In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 2015; **520**: 186–191.
- 58 Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* 2013; **10**: 1116–1121.
- 59 Kim E, Koo T, Park SW, Kim D, Kim K, Cho H-Y *et al.* In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat Commun* 2017; **8**. doi:10.1038/ncomms14500.

- 60 Colella P, Ronzitti G, Mingozzi F. Emerging Issues in AAV-Mediated In Vivo Gene Therapy. *Mol Ther Methods Clin Dev* 2017; **8**: 87–104.
- 61 Mingozzi F, Maus MV, Hui DJ, Sabatino DE, Murphy SL, Rasko JEJ *et al.* CD8(+) T-cell responses to adeno-associated virus capsid in humans. *Nat Med* 2007; **13**: 419–422.
- 62 Vandamme C, Adjali O, Mingozzi F. Unraveling the Complex Story of Immune Responses to AAV Vectors Trial After Trial. *Hum Gene Ther* 2017; **28**: 1061–1074.
- 63 Kaeppel C, Beattie SG, Fronza R, van Logtenstein R, Salmon F, Schmidt S *et al.* A largely random AAV integration profile after LPLD gene therapy. *Nature Medicine* 2013; **19**: 889–891.
- 64 Hanlon KS, Kleinstiver BP, Garcia SP, Zaborowski MP, Volak A, Spirig SE *et al.* High levels of AAV vector integration into CRISPR-induced DNA breaks. *Nat Commun* 2019; **10**: 1–11.
- 65 Nelson CE, Hakim CH, Ousterout DG, Thakore PI, Moreb EA, Rivera RMC *et al.* In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* 2016; **351**: 403–407.
- 66 Yin H, Xue W, Chen S, Bogorad RL, Benedetti E, Grompe M *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol* 2014; **32**: 551–553.
- 67 Mangeot PE, Risson V, Fusil F, Marnef A, Laurent E, Blin J *et al.* Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat Commun* 2019; **10**: 1–15.
- 68 Sharei A, Zoldan J, Adamo A, Sim WY, Cho N, Jackson E *et al.* A vector-free microfluidic platform for intracellular delivery. *Proc Natl Acad Sci USA* 2013; **110**: 2082–2087.
- 69 DiTommaso T, Cole JM, Cassereau L, Buggé JA, Hanson JLS, Bridgen DT *et al.* Cell engineering with microfluidic squeezing preserves functionality of primary immune cells in vivo. *Proc Natl Acad Sci U S A* 2018; **115**: E10907–E10914.
- 70 Lee K, Conboy M, Park HM, Jiang F, Kim HJ, Dewitt MA *et al.* Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat Biomed Eng* 2017; **1**: 889–901.
- 71 Li C, Psatha N, Sova P, Gil S, Wang H, Kim J *et al.* Reactivation of γ -globin in adult β -YAC mice after ex vivo and in vivo hematopoietic stem cell genome editing. *Blood* 2018; **131**: 2915–2928.
- 72 Antoniani C, Meneghini V, Lattanzi A, Felix T, Romano O, Magrin E *et al.* Induction of fetal hemoglobin synthesis by CRISPR/Cas9-mediated editing of the human β -globin locus. *Blood* 2018; **131**: 1960–1973.
- 73 Lattanzi A, Meneghini V, Pavani G, Amor F, Ramadier S, Felix T *et al.* Optimization of CRISPR/Cas9 Delivery to Human Hematopoietic Stem and Progenitor Cells for Therapeutic Genomic Rearrangements. *Molecular Therapy* 2019; **27**: 137–150.
- 74 Ye L, Wang J, Tan Y, Beyer AI, Xie F, Muench MO *et al.* Genome editing using CRISPR-Cas9 to create the HPFH genotype in HSPCs: An approach for treating sickle cell disease and β -thalassemia. *Proc Natl Acad Sci U S A* 2016; **113**: 10661–10665.

- 75 Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C *et al.* An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* 2013; **342**: 253–257.
- 76 Wu Y, Zeng J, Roscoe BP, Liu P, Yao Q, Lazzarotto CR *et al.* Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat Med* 2019; **25**: 776–783.
- 77 Dever DP, Bak RO, Reinisch A, Camarena J, Washington G, Nicolas CE *et al.* CRISPR/Cas9 β -globin gene targeting in human haematopoietic stem cells. *Nature* 2016; **539**: 384–389.
- 78 Park SH, Lee CM, Dever DP, Davis TH, Camarena J, Srifa W *et al.* Highly efficient editing of the β -globin gene in patient-derived hematopoietic stem and progenitor cells to treat sickle cell disease. *Nucleic Acids Res* 2019; **47**: 7955–7972.
- 79 Khorsandi SE, Bachellier P, Weber JC, Greget M, Jaeck D, Zacharoulis D *et al.* Minimally invasive and selective hydrodynamic gene therapy of liver segments in the pig and human. *Cancer Gene Ther* 2008; **15**: 225–230.
- 80 Yang Y, Wang L, Bell P, McMenamin D, He Z, White J *et al.* A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat Biotechnol* 2016; **34**: 334–338.
- 81 Tabebordbar M, Zhu K, Cheng JKW, Chew WL, Widrick JJ, Yan WX *et al.* In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* 2016; **351**: 407–411.
- 82 Long C, Amoasii L, Mireault AA, McAnally JR, Li H, Sanchez-Ortiz E *et al.* Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* 2016; **351**: 400–403.
- 83 Amoasii L, Hildyard JCW, Li H, Sanchez-Ortiz E, Mireault A, Caballero D *et al.* Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science* 2018; **362**: 86–91.
- 84 Maeder ML, Stefanidakis M, Wilson CJ, Baral R, Barrera LA, Bounoutas GS *et al.* Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat Med* 2019; **25**: 229–233.
- 85 Baylis F, McLeod M. First-in-human Phase 1 CRISPR Gene Editing Cancer Trials: Are We Ready? *Curr Gene Ther* 2017; **17**: 309–319.
- 86 Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 2017; **550**: 407–410.
- 87 Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 2014; **507**: 62–67.
- 88 Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science* 2016; **351**: 84–88.
- 89 Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 2013; **31**: 827–832.

- 90 Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 2016; **34**: 184–191.
- 91 Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud J-B *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* 2016; **17**: 148.
- 92 Listgarten J, Weinstein M, Kleinstiver BP, Sousa AA, Joung JK, Crawford J *et al.* Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat Biomed Eng* 2018; **2**: 38–47.
- 93 Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, Trevino A *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 2013; **154**: 1380–1389.
- 94 Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* 2014; **32**: 577–582.
- 95 Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen N, Zheng Z *et al.* High-fidelity CRISPR-Cas9 variants with undetectable genome-wide off-targets. *Nature* 2016; **529**: 490–495.
- 96 Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol* 2018; **36**: 265–271.
- 97 Lee JK, Jeong E, Lee J, Jung M, Shin E, Kim Y *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* 2018; **9**. doi:10.1038/s41467-018-05477-x.
- 98 Vakulskas CA, Dever DP, Rettig GR, Turk R, Jacobi AM, Collingwood MA *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human haematopoietic stem and progenitor cells. *Nat Med* 2018; **24**: 1216–1224.
- 99 Kleinstiver BP, Sousa AA, Walton RT, Tak YE, Hsu JY, Clement K *et al.* Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol* 2019; **37**: 276–282.
- 100 Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 2014; **32**: 279–284.
- 101 Kocak DD, Josephs EA, Bhandarkar V, Adkar SS, Kwon JB, Gersbach CA. Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nat Biotechnol* 2019; **37**: 657–666.
- 102 Ciccia A, Elledge SJ. The DNA Damage Response: Making it safe to play with knives. *Molecular cell* 2010; **40**: 179.
- 103 Lavin MF. ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks. *Oncogene* 2007; **26**: 7749–7758.
- 104 Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 1998; **273**: 5858–5868.

- 105 Jungmichel S, Stucki M. MDC1: The art of keeping things in focus. *Chromosoma* 2010; **119**: 337–349.
- 106 Cook PJ, Ju BG, Telese F, Wang X, Glass CK, Rosenfeld MG. Tyrosine Dephosphorylation of H2AX Modulates Apoptosis and Survival Decisions. *Nature* 2009; **458**: 591–596.
- 107 Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology* 2017; **18**: 495–506.
- 108 Scully R, Panday A, Elango R, Willis NA. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* 2019; : 1–17.
- 109 Stinson BM, Moreno AT, Walter JC, Loparo JJ. A mechanism to minimize errors during non-homologous end joining. *bioRxiv* 2019. doi:10.1101/563197.
- 110 Graham TGW, Walter JC, Loparo JJ. Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. *Mol Cell* 2016; **61**: 850–858.
- 111 Sartori AA, Lukas C, Coates J, Mistrik M, Fu S, Bartek J *et al.* Human CtIP promotes DNA end resection. *Nature* 2007; **450**: 509–514.
- 112 Garcia V, Phelps SEL, Gray S, Neale MJ. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature* 2011; **479**: 241–244.
- 113 San Filippo J, Sung P, Klein H. Mechanism of Eukaryotic Homologous Recombination. *Annual Review of Biochemistry* 2008; **77**: 229–257.
- 114 Zhao W, Steinfeld JB, Liang F, Chen X, Maranon DG, Jian Ma C *et al.* BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing. *Nature* 2017; **550**: 360–365.
- 115 Willis NA, Frock RL, Menghi F, Duffey EE, Panday A, Camacho V *et al.* Mechanism of tandem duplication formation in BRCA1 mutant cells. *Nature* 2017; **551**: 590–595.
- 116 Bizard AH, Hickson ID. The Dissolution of Double Holliday Junctions. *Cold Spring Harb Perspect Biol* 2014; **6**. doi:10.1101/cshperspect.a016477.
- 117 Paliwal S, Kanagaraj R, Sturzenegger A, Burdova K, Janscak P. Human RECQ5 helicase promotes repair of DNA double-strand breaks by synthesis-dependent strand annealing. *Nucleic Acids Res* 2014; **42**: 2380–2390.
- 118 Verma P, Greenberg RA. Noncanonical views of homology-directed DNA repair. *Genes Dev* 2016; **30**: 1138–1154.
- 119 Stark JM, Pierce AJ, Oh J, Pastink A, Jasin M. Genetic Steps of Mammalian Homologous Repair with Distinct Mutagenic Consequences. *Mol Cell Biol* 2004; **24**: 9305–9316.
- 120 Bhargava R, Onyango DO, Stark JM. Regulation of Single Strand Annealing and its role in genome maintenance. *Trends Genet* 2016; **32**: 566–575.

- 121 Mateos-Gomez PA, Kent T, Deng SK, McDevitt S, Kashkina E, Hoang TM *et al.* The helicase domain of Polθ counteracts RPA to promote alt-NHEJ. *Nat Struct Mol Biol* 2017; **24**: 1116–1123.
- 122 Truong LN, Li Y, Shi LZ, Hwang PY-H, He J, Wang H *et al.* 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 USA* 2013; **110**: 7720–7725.
- 123 Sakuma T, Nakade S, Sakane Y, Suzuki K-IT, Yamamoto T. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nature Protocols* 2016; **11**: 118–133.
- 124 Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* 2016; **540**: 144–149.
- 125 Kan Y, Ruis B, Lin S, Hendrickson EA. The Mechanism of Gene Targeting in Human Somatic Cells. *PLoS Genet* 2014; **10**. doi:10.1371/journal.pgen.1004251.
- 126 Kan Y, Ruis B, Takasugi T, Hendrickson EA. Mechanisms of precise genome editing using oligonucleotide donors. *Genome Res* 2017; **27**: 1099–1111.
- 127 Pierandrei S, Luchetti A, Sanchez M, Novelli G, Sangiuolo F, Lucarelli M. The Gene Targeting Approach of Small Fragment Homologous Replacement (SFHR) Alters the Expression Patterns of DNA Repair and Cell Cycle Control Genes. *Mol Ther Nucleic Acids* 2016; **5**: e304.
- 128 Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature Biotechnology* 2016; **34**: 339–344.
- 129 Panier S, Boulton SJ. Double-strand break repair: 53BP1 comes into focus. *Nature Reviews Molecular Cell Biology* 2014; **15**: 7–18.
- 130 Nakamura K, Sakai W, Kawamoto T, Bree RT, Lowndes NF, Takeda S *et al.* Genetic dissection of vertebrate 53BP1: A major role in non-homologous end joining of DNA double strand breaks. *DNA Repair* 2006; **5**: 741–749.
- 131 Chapman JR, Barral P, Vannier J-B, Borel V, Steger M, Tomas-Loba A *et al.* RIF1 Is Essential for 53BP1-Dependent Nonhomologous End Joining and Suppression of DNA Double-Strand Break Resection. *Molecular Cell* 2013; **49**: 858.
- 132 Boersma V, Moatti N, Segura-Bayona S, Peuscher MH, van der Torre J, Wevers BA *et al.* MAD2L2 controls DNA repair at telomeres and DNA breaks by inhibiting 5' end-resection. *Nature* 2015; **521**: 537–540.
- 133 Xu G, Chapman JR, Brandsma I, Yuan J, Mistrik M, Bouwman P *et al.* REV7 counteracts DNA double-strand break resection and affects PARP inhibition. *Nature* 2015; **521**: 541–544.
- 134 Noordermeer SM, Adam S, Setiাপutra D, Barazas M, Pettitt SJ, Ling AK *et al.* The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* 2018; **560**: 117–121.

- 135 Mirman Z, Lotterberger F, Takai H, Kibe T, Gong Y, Takai K *et al.* 53BP1/Rif1/Shieldin counteract DSB resection through CST/Pol α -dependent fill-in. *Nature* 2018; **560**: 112–116.
- 136 Schrock MS, Batar B, Lee J, Druck T, Ferguson B, Cho JH *et al.* Wwox–Brca1 interaction: role in DNA repair pathway choice. *Oncogene* 2017; **36**: 2215–2227.
- 137 Hustedt N, Durocher D. The control of DNA repair by the cell cycle. *Nature Cell Biology* 2017; **19**: 1–9.
- 138 Buis J, Stoneham T, Spehalski E, Ferguson DO. Mre11 regulates CtIP-dependent double strand break repair by interaction with CDK2. *Nat Struct Mol Biol* 2012; **19**: 246–252.
- 139 Cruz-García A, López-Saavedra A, Huertas P. BRCA1 Accelerates CtIP-Mediated DNA-End Resection. *Cell Reports* 2014; **9**: 451–459.
- 140 Tomimatsu N, Mukherjee B, Catherine Hardebeck M, Ilcheva M, Vanessa Camacho C, Louise Harris J *et al.* Phosphorylation of EXO1 by CDKs 1 and 2 regulates DNA end resection and repair pathway choice. *Nat Commun* 2014; **5**: 3561.
- 141 Zhang H, Liu H, Chen Y, Yang X, Wang P, Liu T *et al.* A cell cycle-dependent BRCA1–UHRF1 cascade regulates DNA double-strand break repair pathway choice. *Nat Commun* 2016; **7**. doi:10.1038/ncomms10201.
- 142 Jirawatnotai S, Hu Y, Michowski W, Elias JE, Becks L, Bienvenu F *et al.* A function for cyclin D1 in DNA repair uncovered by interactome analyses in human cancers. *Nature* 2011; **474**: 230–234.
- 143 Kanakkanthara A, Jeganathan KB, Limzerwala JF, Baker DJ, Hamada M, Nam H-J *et al.* Cyclin A2 is an RNA binding protein that controls Mre11 mRNA translation. *Science* 2016; **353**: 1549–1552.
- 144 Zhao W, Steinfeld JB, Liang F, Chen X, Maranon DG, Ma CJ *et al.* Promotion of RAD51-mediated homologous DNA pairing by BRCA1-BARD1. *Nature* 2017; **550**: 360–365.
- 145 Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* 2015; **33**: 538–542.
- 146 Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology* 2015; **33**: 543–548.
- 147 Robert F, Barbeau M, Éthier S, Dostie J, Pelletier J. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med* 2015; **7**. doi:10.1186/s13073-015-0215-6.
- 148 Canny MD, Moatti N, Wan LCK, Fradet-Turcotte A, Krasner D, Mateos-Gomez PA *et al.* Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat Biotechnol* 2018; **36**: 95–102.

- 149 Nambiar TS, Billon P, Diedenhofen G, Hayward SB, Tagliatalata A, Cai K *et al.* Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nat Commun* 2019; **10**: 1–13.
- 150 Song J, Yang D, Xu J, Zhu T, Chen YE, Zhang J. RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat Commun* 2016; **7**: 10548.
- 151 Jayathilaka K, Sheridan SD, Bold TD, Bochenska K, Logan HL, Weichselbaum RR *et al.* A chemical compound that stimulates the human homologous recombination protein RAD51. *Proc Natl Acad Sci U S A* 2008; **105**: 15848–15853.
- 152 Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med* 2018; **24**: 927–930.
- 153 Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, Theriault K *et al.* p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat Med* 2018; **24**: 939–946.
- 154 Yu C, Liu Y, Ma T, Liu K, Xu S, Zhang Y *et al.* Small Molecules Enhance CRISPR Genome Editing in Pluripotent Stem Cells. *Cell Stem Cell* 2015; **16**: 142–147.
- 155 Riesenberger S, Maricic T. Targeting repair pathways with small molecules increases precise genome editing in pluripotent stem cells. *Nat Commun* 2018; **9**: 1–9.
- 156 Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* 2014; **3**: e04766.
- 157 Howden SE, McColl B, Glaser A, Vadolas J, Petrou S, Little MH *et al.* A Cas9 Variant for Efficient Generation of Indel-Free Knockin or Gene-Corrected Human Pluripotent Stem Cells. *Stem Cell Reports* 2016; **7**: 508–517.
- 158 Charpentier M, Khedher AHY, Menoret S, Brion A, Lamribet K, Dardillac E *et al.* CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat Commun* 2018; **9**: 1133.
- 159 Jayavaradhan R, Pillis DM, Goodman M, Zhang F, Zhang Y, Andreassen PR *et al.* CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites. *Nature Communications* 2019; **10**. doi:10.1038/s41467-019-10735-7.
- 160 Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* 2018; **36**: 765–771.
- 161 Ma H, Marti-Gutierrez N, Park S-W, Wu J, Lee Y, Suzuki K *et al.* Correction of a pathogenic gene mutation in human embryos. *Nature* 2017; **548**: 413–419.
- 162 Adikusuma F, Piltz S, Corbett MA, Turvey M, McColl SR, Helbig KJ *et al.* Large deletions induced by Cas9 cleavage. *Nature* 2018; **560**: E8–E9.
- 163 Ma H, Marti-Gutierrez N, Park S-W, Wu J, Hayama T, Darby H *et al.* Ma *et al.* reply. *Nature* 2018; **560**: E10–E23.
- 164 Adikusuma F, Williams N, Grutzner F, Hughes J, Thomas P. Targeted Deletion of an Entire Chromosome Using CRISPR/Cas9. *Mol Ther* 2017; **25**: 1736–1738.

- 165 Zuo E, Huo X, Yao X, Hu X, Sun Y, Yin J *et al.* CRISPR/Cas9-mediated targeted chromosome elimination. *Genome Biol* 2017; **18**. doi:10.1186/s13059-017-1354-4.
- 166 Hossain MdA, Lin Y, Yan S. Single-Strand Break End Resection in Genome Integrity: Mechanism and Regulation by APE2. *Int J Mol Sci* 2018; **19**. doi:10.3390/ijms19082389.
- 167 Davis L, Maizels N. DNA Nicks Promote Efficient and Safe Targeted Gene Correction. *PLoS One* 2011; **6**. doi:10.1371/journal.pone.0023981.
- 168 Ramirez CL, Certo MT, Mussolino C, Goodwin MJ, Cradick TJ, McCaffrey AP *et al.* Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res* 2012; **40**: 5560–5568.
- 169 Metzger MJ, Stoddard BL, Monnat RJ. PARP-mediated Repair, Homologous Recombination, and Back-up Non-Homologous End Joining-Like Repair of Single-Strand Nicks. *DNA Repair (Amst)* 2013; **12**: 529–534.
- 170 McConnell Smith A, Takeuchi R, Pellenz S, Davis L, Maizels N, Monnat RJ *et al.* Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc Natl Acad Sci U S A* 2009; **106**: 5099–5104.
- 171 Davis L, Maizels N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc Natl Acad Sci U S A* 2014; **111**: E924–E932.
- 172 Davis L, Maizels N. Two Distinct Pathways Support Gene Correction by Single-Stranded Donors at DNA Nicks. *Cell Reports* 2016; **17**: 1872–1881.
- 173 Vriend LEM, Prakash R, Chen C-C, Vanoli F, Cavallo F, Zhang Y *et al.* Distinct genetic control of homologous recombination repair of Cas9-induced double-strand breaks, nicks and paired nicks. *Nucleic Acids Res* 2016; **44**: 5204–5217.
- 174 Kuzminov A. Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc Natl Acad Sci USA* 2001; **98**: 8241–8246.
- 175 Maizels N, Davis L. Initiation of homologous recombination at DNA nicks. *Nucleic Acids Research* 2018; **46**: 6962.
- 176 Géronimi F, Richard E, Lamrissi-Garcia I, Lalanne M, Ged C, Redonnet-Vernhet I *et al.* Lentivirus-mediated gene transfer of uroporphyrinogen III synthase fully corrects the porphyric phenotype in human cells. *J Mol Med* 2003; **81**: 310–320.
- 177 Ged C, Mendez M, Robert E, Lalanne M, Lamrissi-Garcia I, Costet P *et al.* A knock-in mouse model of congenital erythropoietic porphyria. *Genomics* 2006; **87**: 84–92.
- 178 Robert-Richard E, Moreau-Gaudry F, Lalanne M, Lamrissi-Garcia I, Cario-André M, Guyonnet-Dupérat V *et al.* Effective gene therapy of mice with congenital erythropoietic porphyria is facilitated by a survival advantage of corrected erythroid cells. *Am J Hum Genet* 2008; **82**: 113–124.
- 179 Robert-Richard E, Lalanne M, Lamrissi-Garcia I, Guyonnet-Duperat V, Richard E, Pitard V *et al.* Modeling of congenital erythropoietic porphyria by RNA interference: a new tool for preclinical gene therapy evaluation. *J Gene Med* 2010; **12**: 637–646.

- 180 Zhang J, Cao M, Dong J, Li C, Xu W, Zhan Y *et al.* ABRO1 suppresses tumourigenesis and regulates the DNA damage response by stabilizing p53. *Nat Commun* 2014; **5**: 1–15.
- 181 Xu S, Wu X, Wu L, Castillo A, Liu J, Atkinson E *et al.* Abro1 maintains genome stability and limits replication stress by protecting replication fork stability. *Genes Dev* 2017; **31**: 1469–1482.
- 182 Do TU, Ho B, Shih S-J, Vaughan A. Zinc Finger Nuclease Induced DNA Double Stranded Breaks and Rearrangements in MLL. *Mutat Res* 2012; **740**: 34–42.
- 183 Liu Y, Chen C, Xu Z, Scuoppo C, Rillahan CD, Gao J *et al.* Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. *Nature* 2016; **531**: 471–475.
- 184 Lin S, Zhou Y, Fang Q, Wu J, Zhang Z, Ji Y *et al.* Chromosome 10q26 deletion syndrome: Two new cases and a review of the literature. *Mol Med Rep* 2016; **14**: 5134–5140.
- 185 Makovets S, Blackburn EH. DNA damage signalling prevents deleterious telomere addition at DNA breaks. *Nat Cell Biol* 2009; **11**: 1383–1386.
- 186 Ouenzar F, Lalonde M, Laprade H, Morin G, Gallardo F, Tremblay-Belzile S *et al.* Cell cycle-dependent spatial segregation of telomerase from sites of DNA damage. *J Cell Biol* 2017; **216**: 2355–2371.
- 187 Hermetz KE, Newman S, Conneely KN, Martin CL, Ballif BC, Shaffer LG *et al.* Large Inverted Duplications in the Human Genome Form via a Fold-Back Mechanism. *PLoS Genetics* 2014; **10**: e1004139.
- 188 Rees HA, Yeh W-H, Liu DR. Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat Commun* 2019; **10**. doi:10.1038/s41467-019-09983-4.
- 189 Genovese P, Schirotti G, Escobar G, Tomaso TD, Firrito C, Calabria A *et al.* Targeted Genome Editing in Human Repopulating Hematopoietic Stem Cells. *Nature* 2014; **510**: 235–240.
- 190 Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin EV *et al.* RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 2019; **365**: 48–53.
- 191 Faure G, Shmakov SA, Yan WX, Cheng DR, Scott DA, Peters JE *et al.* CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat Rev Microbiol* 2019; **17**: 513–525.
- 192 Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 2016; **533**: 420–424.
- 193 Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 2017; **551**: 464–471.
- 194 Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019. doi:10.1038/s41586-019-1711-4.

Titre : Génotoxicité des systèmes CRISPR-Cas9.

Résumé : La thérapie génique est une stratégie thérapeutique prometteuse pour le traitement des maladies monogéniques. Si les premières approches, dites additives, ont reposé sur l'utilisation de vecteurs viraux, une part grandissante se tourne désormais vers l'édition génique. Celle-ci est permise par la mise au point de nouvelles générations d'endonucléases, et en particulier le système CRISPR-Cas9. Moins d'une décennie après sa caractérisation, le système CRISPR-Cas9 a permis de faire passer l'édition génique à un stade clinique. Toutefois, dans le même laps de temps, plusieurs interrogations ont été soulevées vis-à-vis de la génotoxicité pouvant être induite par la Cas9. Une littérature émergente pointe le risque de génotoxicité au site ciblé. Le travail de thèse présenté ici s'inscrit dans cette thématique. La première partie de l'étude a eu pour objectif de décrire la génotoxicité induite par une unique cassure double-brin faite par la Cas9. La caractérisation des effets a été faite à la fois à l'échelle nucléotidique, par le suivi de la balance HDR / InDels, mais également à l'échelle du chromosome. Le suivi de l'intégrité chromosomique a permis de mettre en lumière un nouveau risque de génotoxicité encore non-caractérisé. Un système de détection sensible et spécifique de ce risque a été mis au point pour continuer de le caractériser. Le second objectif a été de répondre aux limites soulevées par la génotoxicité non-voulue, en mettant au point une méthode d'édition génique plus sûre et aussi efficace, *via* l'utilisation d'une unique cassure simple-brin par la Cas9^{D10A}-nickase.

Mots clés : CRISPR/Cas9, Génotoxicité, Edition génique

Title : Genotoxicity of CRISPR-Cas9 systems.

Abstract : Gene therapy is a promising therapeutic strategy for the monogenic diseases treatment. If the first approaches, called additive, have relied on the use of viral vectors, a growing share is now turning to gene editing. Less than a decade after its characterization, the CRISPR-Cas9 system has moved gene editing to a clinical stage. However, in the same period of time, several questions have been raised regarding the genotoxicity that can be induced by Cas9. An emerging literature points to the risk of genotoxicity at the targeted site. The thesis work presented here is part of this theme. The first part of the study aimed to describe the genotoxicity induced by a single double-stranded break made by Cas9. Characterization of the effects was done both at the nucleotide level, by monitoring the HDR / InDels balance, but also at the chromosome scale. The monitoring of chromosomal integrity has brought to light a new risk of genotoxicity that was not characterized. A sensitive and specific detection system for this risk has been developed to further characterize it. The second objective was to address the limitations of unwanted genotoxicity by developing a safer and more efficient gene editing method through the use of a single single-stranded breakage by Cas9^{D10A}-nickase.

Keywords : CRISPR/Cas9, Genotoxicity, Gene editing

INSERM U1035, BMGIC - Univ. Bordeaux, France

Biothérapie des Maladies Génétiques, Inflammatoires et du Cancer.

Equipe Biothérapie