



Développement d'outils moléculaires pour faciliter l'ingénierie des génomes

Mémoire

Alexandre Raymond-Fleury

Maitrise en Médecine Moléculaire

Maitre ès Sciences (M.Sc.)

Québec, Canada

© Alexandre Raymond-Fleury, 2018

RÉSUMÉ

Malgré les progrès récents associés à l'utilisation des nucléases d'ingénierie du génome, certaines problématiques persistent. Par exemple, l'isolement des cellules génétiquement modifiées est un travail long et parfois laborieux. De plus, la spécificité des nucléases est aussi un point souvent mal caractérisé. Cette dernière peut avoir des conséquences importantes *in vitro*, mais particulièrement *in vivo*. Le but de mon projet de maîtrise était de développer des techniques permettant d'adresser ces deux problématiques. Ici, nous avons créé une cassette de sélection auto-excisable qui exploite le processus de réparation par appariement simple brin (SSA) et qui permet une isolation simple de clones modifiés sans laisser de « cicatrices » dans l'ADN. Le procédé se base sur une sélection à la puromycine, une contre sélection au ganciclovir et une excision de la cassette de résistance via l'utilisation de nucléases thermosensibles. Nous avons démontré la faisabilité de cette approche en intégrant la protéine fluorescente EGFP au locus *AAVS1* et en insérant une étiquette de purification au locus *EZH2*. Ce locus code la protéine *Enhancer of Zeste Homolog 2 (EZH2)*. Dans un deuxième temps, nous avons évalué la possibilité d'utiliser deux paires de nickases pour créer deux cassures simple brin (SSB) et ainsi induire indirectement une cassure double brin dans l'ADN. Cette approche permet de diminuer de façon importante la possibilité de créer des mutations à des sites non-ciblés. Nous avons déterminé que deux SSBs doivent avoir lieu sur des brins opposés et à proximité afin de stimuler la modification génique voulue. Toutefois, ces paires de nickases ont une efficacité plus faible que les nucléases clivant au même endroit. Les nouveaux outils décrits dans ce mémoire permettront éventuellement de simplifier l'ingénierie du génome en diminuant la charge de travail nécessaire à l'enrichissement des cellules modifiées génétiquement et en augmentant la spécificité des nucléases, ce qui permet de réduire les probabilités de mutations à des loci hors cible.

ABSTRACT

Despite recent advances in the use of engineered nucleases, some issues remain. For example, isolation of genetically engineered cells is a long and sometimes laborious process. Moreover, the specificity of nucleases is often poorly characterized. The latter may have important consequences *in vitro*, but particularly *in vivo*. We aimed to develop techniques for addressing these two problems. Here we have created a self-excisable selection cassette that exploits the single-strand annealing (SSA) repair process and allows for simple isolation of modified clones and scarless genome editing. The method is based on puromycin selection, ganciclovir counter-selection and excision of the resistance cassette via the use of heat-sensitive nucleases. We demonstrate the feasibility of this approach by integrating the fluorescent protein EGFP at the *AAVS1* locus and by inserting a purification tag at the *EZH2* locus. This locus encodes the protein *Enhancer of Zeste Homolog 2 (EZH2)*. Secondly, we evaluated the possibility of using two pairs of nickases to create two single-strand breaks that could be repaired as a double strand break. As a proof of concept, we created two single-strand breaks at a distance of 513 bp at the *CCR5* locus. When they take place on opposite strands, the latter can recreate the effect of a double-strand break, but with a lower efficiency than the nucleases cleaving at the same site. We determined that the double-strand break effect created by two single-strand breaks increases as the distance between the two single-strand breaks decreases. The new tools described here make it possible to simplify genome engineering by reducing the workload necessary for the enrichment of genetically modified cells and by increasing the specificity of the nucleases by decreasing the potential for off-target mutagenesis.

TABLE DES MATIÈRES

| | |
|--|-----|
| RÉSUMÉ | iii |
| ABSTRACT | iv |
| TABLE DES MATIÈRES | v |
| LISTE DES TABLEAUX..... | ix |
| LISTE DES FIGURES..... | x |
| LISTE DES ABRÉVIATIONS ET SIGLES | xii |
| REMERCIEMENTS | xv |
| INTRODUCTION | 1 |
| I.1.0 Réparation de l'ADN | 1 |
| I.1.1 Cassures simple brin..... | 1 |
| I.1.1.1 Réparation par excision de base..... | 3 |
| I.1.1.2 Réparation par excision de nucléotides..... | 3 |
| I.1.1.3 Réparation de mésappariement..... | 4 |
| I.1.2 Cassures double brin..... | 5 |
| I.1.2.1 Point de contrôle et réponse aux dommages à l'ADN..... | 5 |
| I.1.2.2 Jonction des extrémités non homologues | 8 |
| I.1.2.2.1 Cascade de réparation | 8 |
| I.1.2.3 Réparation dirigée par séquences homologues | 10 |
| I.1.2.3.1 Cascade de réparation HDR..... | 10 |
| I.1.2.3.1.1 Cascade de réparation de la double jonction de Holliday..... | 12 |
| I.1.2.3.1.2 Cascade de réparation du BIR | 12 |
| I.1.2.3.1.3 Cascade de réparation SDSA..... | 13 |
| I.1.2.4 Réparation par appariement simple brin..... | 13 |

| | |
|---|----|
| I.1.2.4.1 Cascade de réparation SSA..... | 14 |
| I.2.0 Les nucléases d'ingénierie du génome | 15 |
| I.2.1 Les nucléases à doigt de zinc | 15 |
| I.2.1.1 Domaine de reconnaissance de l'ADN | 15 |
| I.2.1.2 Fonction nucléase | 16 |
| I.2.1.2.1 Augmenter la spécificité des nucléases | 17 |
| I.2.1.3 Optimisation du système des ZFN..... | 17 |
| I.2.1.4 Les ZFN et l'édition de génome | 19 |
| I.2.1.5 Limites des ZFN | 20 |
| I.2.2 Les nucléases effectrices de type activateur de transcription | 20 |
| I.2.2.1 Liaison à l'ADN..... | 20 |
| I.2.2.2 Fonction nucléase..... | 22 |
| I.2.2.3 Les TALEN et l'édition du génome..... | 22 |
| I.2.2.4 Limites associées à l'utilisation des TALEN..... | 23 |
| I.2.3 Le système CRISPR/Cas | 23 |
| I.2.3.1 Liaison à l'ADN et formation d'une cassure double brin..... | 24 |
| I.2.3.1.1 Le protospacer adjacent motif..... | 26 |
| I.2.3.1.2 La séquence seed | 27 |
| I.2.3.2 Fonction nucléase..... | 27 |
| I.2.3.3 Variantes du système CRISPR/Cas | 27 |
| I.2.3.4 Limites du système CRISPR..... | 30 |
| I.3.0 L'édition de génome | 31 |
| I.3.1 L'utilisation d'un havre génomique sécuritaire | 31 |
| I.3.2 Techniques utilisées pour la modification génique..... | 32 |
| I.3.2.1 Les transposases..... | 33 |

| | |
|---|-----------|
| I.3.2.1.1 La transposase Sleeping Beauty | 33 |
| I.3.2.1.1.2 Mécanisme de transposition d'ADN par Sleeping Beauty | 33 |
| I.3.2.1.1.3 Ingénierie du génome avec la transposase Sleeping Beauty..... | 33 |
| I.3.2.1.1.4 Limites | 34 |
| I.3.2.1.2 La transposase piggyBac | 34 |
| I.3.2.1.2.1 Mécanisme de transposition d'ADN par piggyBac | 35 |
| I.3.2.1.2.2 Ingénierie du génome avec la transposase piggyBac..... | 35 |
| I.3.2.2 Le système Cre/LoxP | 36 |
| I.3.2.2.1 Mécanisme de recombinaison du système Cre/LoxP | 37 |
| I.3.3 Modification du génome avec les nucléases d'ingénierie | 37 |
| I.3.3.1 Invalidation de gène..... | 38 |
| I.3.3.2 Correction génique..... | 39 |
| I.3.3.3 Ajout de matériel génétique..... | 39 |
| I.3.3.4 Essais cliniques en cours avec les nucléases d'ingénierie | 39 |
| I.3.4 Limites des techniques présentement utilisées | 40 |
| I.3.4.1 Insertion aléatoire | 40 |
| I.3.4.2 Traces des modifications géniques | 40 |
| I.3.4.3 Limites de l'édition génique dans un GSH..... | 40 |
| I.3.4.4 Limites de la modification génique par HDR | 41 |
| PROBLÉMATIQUE ET OBJECTIFS DES TRAVAUX | 42 |
| CHAPITRE 2 : Développement et validation d'une cassette auto-excisable pour la création de modifications géniques sans traces..... | 43 |
| II.2.1 Matériel et méthodes | 44 |
| II.2.2 Résultats | 50 |
| II.2.2.1 Élaboration d'une stratégie pour créer une cassette auto-excisable | 50 |

| | |
|--|----|
| II.2.2.2 Intégration de la protéine fluorescente EGFP au locus <i>AAVSI</i> | 53 |
| II.2.2.3 Introduction d'une étiquette FLAG au locus <i>EZH2</i> | 55 |
| II.2.3 Discussion | 58 |
| II.2.4 Conclusion..... | 60 |
| CHAPITRE 3 : Utilisation des doubles nickases comme outils spécifiques de modification génique..... | 61 |
| III.3.1 Matériel et méthodes | 62 |
| III.3.2 Résultats | 65 |
| III.3.2.1 Une faible distance entre deux paires de nickases augmente leur efficacité. 65 | |
| III.3.2.2 Utilisées seules, les nickases n'induisent pas la formation d'indels | 68 |
| III.3.3 Discussion | 69 |
| III.3.4 Conclusion | 71 |
| Discussion intégrative et perspectives | 72 |
| Conclusion | 75 |
| Références | 76 |

LISTE DES TABLEAUX

CHAPITRE 2

| | |
|--|----|
| Tableau 1 : Séquence des oligonucléotides utilisés pour l'assemblage de la cassette auto-excisable eGFP..... | 49 |
| Tableau 2 : Séquence des oligonucléotides utilisés pour l'assemblage de la cassette auto-excisable FLAG_EZH2..... | 49 |
| Tableau 3 : Séquence des gBlock utilisés pour l'assemblage des cassettes auto-excisable..... | 50 |
| Tableau 4 : Séquence des oligonucléotides utilisés pour les PCR de géotypage et de séquençage. | 50 |

CHAPITRE 3

| | |
|---|----|
| Tableau 5 : Séquence des oligonucléotides utilisés pour l'assemblage de la cassette auto-excisable SSA_GFP. | 64 |
| Tableau 6 : Conditions de transfection de cellules K562..... | 64 |
| Tableau 7 : Séquence des oligonucléotides utilisés pour les PCR de criblage et de séquençage. | 65 |

LISTE DES FIGURES

Introduction

| | |
|---|----|
| Figure 1 : Représentation des 3 voies de réparation d'une cassure simple brin | 2 |
| Figure 2 : Représentation schématique de l'activation du point de contrôle du dommage à l'ADN | 7 |
| Figure 3 : Représentation schématique de la réparation d'une cassure double brin par le mécanisme de NHEJ | 9 |
| Figure 4 : Schéma de la réparation par HDR d'une cassure double brin | 11 |
| Figure 5 : Représentation schématique de la réparation par SSA | 14 |
| Figure 6 : Schéma de ZFN de type sauvage et de leur possibilité d'association menant à de potentiels clivages hors cible..... | 18 |
| Figure 7 : Schématisation du concept de ZFN nickase. | 19 |
| Figure 8 : Schéma d'un TALE..... | 21 |
| Figure 9 : Fréquence du nucléotide reconnu selon le RVD dans la répétition du TALE | 22 |
| Figure 10 : Schématisation des différences entre les sous-unités multiples des systèmes CRISPR de classe 1 versus l'unique sous-unité des systèmes CRISPR de classe 2... | 24 |
| Figure 11 : Représentation schématique du clivage de l'ADN par le système CRISPR/Cas9..... | 25 |
| Figure 12 : Représentation des diverses séquences PAM reconnues par différentes nucléases | 26 |
| Figure 13 : Régulation de l'expression endogène avec les nucléases d'ingénierie du génome | 29 |
| Figure 14 : Représentation schématique de l'utilisation combinée des nucléases d'ingénierie et de la transposase PB..... | 36 |
| Figure 15 : Schéma des différentes modifications génétiques pouvant être apportées à des cellules par l'utilisation des nucléases d'ingénierie du génome | 38 |
| | |
| <u>Chapitre 2</u> | |
| Figure 16 : Représentation schématique de la stratégie pour développer une cassette auto-excisable..... | 51 |

| | |
|---|----|
| Figure 17 : Schéma expérimental permettant la modification génique via l'utilisation de la cassette auto-excisable..... | 52 |
| Figure 18 : Insertion de la protéine verte EGFP au locus <i>AAVS1</i> à l'aide d'une cassette de sélection auto-excisable..... | 54 |
| Figure 19 : Insertion d'une séquence FLAG au locus <i>EZH2</i> en utilisant une cassette de sélection auto-excisable | 56 |
| Figure 20 : Étude de populations clonales isolées avec contre-sélection au ganciclovir en milieu semi-solide lors de l'étiquetage avec le 3xFLAG_2xStrep au locus <i>EZH2</i> endogène. | 58 |

Chapitre 3

| | |
|---|----|
| Figure 21 : Représentation schématique des ZFN et TALEN à hétérodimères obligatoires en version nucléase ou nickase..... | 64 |
| Figure 22 : Représentation schématique de l'utilisation de deux paires de TALEN nickase pour recréer une cassure double brin au locus EGFP | 67 |
| Figure 23 : Représentation schématique de l'utilisation de deux paires de nickases pour recréer une cassure double brin au locus <i>CCR5</i> | 69 |

LISTE DES ABRÉVIATIONS ET SIGLES

| | |
|---------------|--|
| 4-OH-T | 4-OH-tamoxifène |
| 53BP1 | <i>P53-binding protein 1</i> |
| AAVS1 | <i>Adeno-associated virus site 1</i> |
| AD | Domaine d'activation |
| ADNdb | ADN double brin |
| ADNsb | ADN simple brin |
| APE1 | <i>Apurinic apyrimidinic endonuclease-1</i> |
| ATM | <i>Ataxia telangiectasia-mutated</i> |
| ATR | <i>Ataxia telangiectasia and Rad3 related</i> |
| ATRIP | <i>ATR interacting protein</i> |
| BAC | Chromosome artificiel bactérien |
| BANF1 | <i>Barrier-to-autointegration factor</i> |
| BER | Réparation par excision de base |
| BIR | <i>Break induced repair</i> |
| BLM | <i>Bloom syndrome protein</i> |
| BRCA1/2 | <i>Breast cancer 1/2</i> |
| BRCT | <i>C-terminal BRCA1</i> |
| CAS | CRISPR associated protein |
| Cas | <i>CRISPR-associated</i> |
| Cas9 | <i>Cas protein 9</i> |
| CCR5 | Récepteur C-C chimiokine de type 5 |
| CETN2 | <i>Centrin-2</i> |
| CHK | <i>Checkpoint kinase</i> |
| CPF1 | <i>CRISPR from Prevotella and Francisella 1</i> |
| CRISPR | <i>Clustered regularly interspaced palindromic repeats</i> |
| CRISPRa | <i>CRISPR activation</i> |
| CRISPRi | <i>CRISPR interference</i> |
| crRNA | <i>CRISPR RNA</i> |
| CSA/B | <i>Cockayne syndrome group A/B protein aussi ERCC8 / 6</i> |
| dCas9 | <i>Dead Cas9</i> |
| DDB1-2 | <i>DNA damage-binding protein 1-2</i> |
| DDC2 | <i>DNA damage chekpoint 2</i> |
| dHJ | <i>Double Holliday Junction</i> |
| <i>D-loop</i> | Boucle de déplacement |
| DNA-PKcs | <i>DNA-dependant protein kinase catalytic subunit</i> |
| DR | Répétitions directes |
| DSB | Cassure double brin |
| ERCC1 | <i>ERCC excision repair 1</i> |
| ER-LBD | Domaine de liaison du ligand au récepteur à l'œstrogène |

| | |
|-----------------------|--|
| eSpCas9 | <i>Enhanced specificity SpCas9</i> |
| EZH2 | <i>Enhancer of Zeste Homolog 2</i> |
| FEN1 | <i>Flap endonuclease 1</i> |
| FokI-ts | FokI thermosensible |
| GGR | Réparation globale du génome |
| gRNA | ARN guide |
| GSH | <i>Genome safe harbor</i> ou havre génomique sécuritaire |
| H2AX | <i>H2A histone family, member X</i> |
| HDR | Recombinaison homologue |
| indels | Petites insertions et délétions |
| IR | Répétitions inversées |
| ITR | Répétitions inverses de terminaison |
| Mec1 | <i>Serine/threonine-protein kinase MEC</i> |
| MLH1 | <i>MutL protein homolog 1</i> |
| MMR | Réparation par mésappariement |
| MPG | <i>N-methyl purine-DNA glycosylase</i> |
| MSH | <i>MutS protein homolog</i> |
| MSH2/6 | <i>MutS protein homolog 2/6</i> |
| Mx1 | <i>MAX Interactor 1, Dimerization Protein</i> |
| NBS1 | Nibrine |
| NEIL1 | <i>Nei Endonuclease VIII- like protein 1</i> |
| NER | Réparation par excision de nucléotide |
| NHEJ | Jonction d'extrémités non homologues |
| NLS | Signal de localisation nucléaire |
| NTH1 | Endonuclease III- like protein 1 |
| OGG1 | 8-oxoguanine glycosylase |
| PAM | <i>Protospacer adjacent motif</i> |
| PARP1/2 | <i>Poly ADP-ribose polymerase 1</i> |
| PB | PiggyBac |
| PBx | PiggyBak à excision seulement |
| PCNA | <i>Proliferating cell nuclear antigen</i> |
| PMS2 | <i>PMS1 homolog 2</i> |
| PNKP | <i>Polynucleotide Kinase 3'-Phosphatase</i> |
| POLX | ADN polymérase X |
| POL β | ADN polymérase bêta |
| POL δ/ϵ | ADN polymérase delta/epsilon |
| PPP1R12C | Protéine phosphatase 1, sous unité régulatrice 12C |
| RFC | <i>Replication factor C</i> |
| RFN | <i>RNA-guided FokI nucleases</i> |
| RPA | <i>Replication protein A</i> |
| rtTA | Protéine activatrice de la tétracycline |

| | |
|-------------|---|
| RVD | <i>Repeat variable diresidue</i> |
| SaCas9 | <i>Staphylococcus aureus Cas9</i> |
| SB | <i>Sleeping Beauty</i> |
| SDSA | <i>Synthesis dependant strand annealing</i> |
| site AP | Site apurinique / apyrimidinique |
| SpCas9 | <i>Streptococcus pyogenes Cas9</i> |
| SpCas9-HF1 | <i>High fidelity-1 SpCas9</i> |
| SpCas9-HF1 | <i>High fidelity-1 SpCas9</i> |
| SSA | Appariement simple-brin |
| StCas9 | <i>Streptococcus thermophilus Cas9</i> |
| TALE | <i>Trans activating-like effector</i> ou effecteur de type trans-activateur |
| TALEN | Nucléase à effecteur de type trans-activateur |
| TCR | Récepteur des cellules T |
| TCR | Réparation couplée à la transcription |
| TDP1 | <i>Tyrosyl-DNA phosphodiesterase 1</i> |
| TdT | <i>Terminal deoxynucleotidyltransferase</i> |
| Tel1 | <i>Serine/threonine-protein kinase tell</i> |
| tetO | Opéron tétracycline |
| Tet-off | <i>Tetracycline-off</i> |
| Tet-on | <i>Tetracycline-on</i> |
| TFIIH | <i>Transcription factor II H</i> |
| tracrRNA | <i>Trans-activating RNA</i> |
| UDG | <i>Uracil DNA glycosylase</i> |
| VIH | Virus de l'immunodéficience humaine |
| XLF | <i>XRC-like factor</i> |
| XP | Xeroderma pigmentosum |
| XPA/C/D/G/H | <i>XP, complementation group A/C/D/G/H</i> |
| XRCC1-4 | <i>X-ray repair cross-complementing protein 1-4</i> |
| ZFN | Nucléase à doigt de zinc |
| ZFP | Protéine à doigt de zinc |

REMERCIEMENTS

Je tiens à remercier tous ceux qui m'ont côtoyé lors de ma formation à la maîtrise en Médecine Moléculaire.

Merci Yannick, mon directeur de maîtrise, d'avoir fait de moi un meilleur scientifique, et de m'avoir poussé à me surpasser. Je n'aurais jamais cru pouvoir réaliser, et tenter, autant de choses en si peu de temps. Merci de ne pas m'avoir laissé manquer de lecture de chevet pour les dernières années. Merci à Caroline et Sophie, pour le support tout au long de mes travaux. Merci d'avoir répondu à toutes mes questions, de vous être occupé de nous dans le laboratoire et de toujours avoir été là comme support moral pour mes présentations importantes. Merci Jérémy pour avoir fait passé le temps pendant les attentes de ligation et pour toutes ces heures passées en dehors du labo. Je veux également à remercier tous les gens sur l'étage du laboratoire qui m'ont donné des conseils et qui ont fait de cet étage un lieu de travail agréable.

Je tiens aussi à remercier tous ces gens qui m'ont supporté en dehors du laboratoire. À mes parents Guy et Nathalie, pour avoir continué à me supporter dans mes études, pour votre foi en moi et pour vous être intéressés à mes projets. À mon frère Nicolas, qui était toujours là pour moi lorsque j'avais besoin de prendre une pause. À ma sœur Annick, pour être venu me chercher et qu'on parte en promenade de vélo ou de patin à roues alignées.

Merci Joanie d'être là pour moi. Merci de supporter mes heures de travail farfelues, et d'avoir relu les millions de versions de mes présentations. Merci de me divertir et de partager ma vie. Merci tout simplement de toujours être là.

INTRODUCTION

I.1.0 Réparation de l'ADN

L'ADN des cellules eucaryotes contient toute l'information nécessaire au bon fonctionnement de l'organisme. Un peu plus de 1 % du génome humain est constitué de gènes codants. Des mutations dans ces gènes ou dans leurs séquences régulatrices peuvent avoir des conséquences de taille; elles peuvent mener à l'invalidation complète du gène ou à sa transcription incontrôlée, deux facteurs impliqués dans certaines pathologies comme le cancer [1, 2].

I.1.1 Cassures simple brin

Les cassures simple brin de l'ADN sont extrêmement communes, elles surviennent à une fréquence de milliers par cellules par jours. Ces cassures sont majoritairement causées par l'attaque de certains métabolites intracellulaires et par la détérioration naturelle de l'ADN [3]. Heureusement, comme la cellule possède deux brins d'ADN, elle utilise le brin non endommagé comme gabarit de réparation [4]. Différents mécanismes, présentés dans la figure 1, peuvent être utilisés par les cellules pour effectuer la réparation d'une cassure simple brin : i) la réparation par excision de base (BER), ii) la réparation par excision de nucléotides (NER) ou iii) la réparation par mésappariement (MMR).

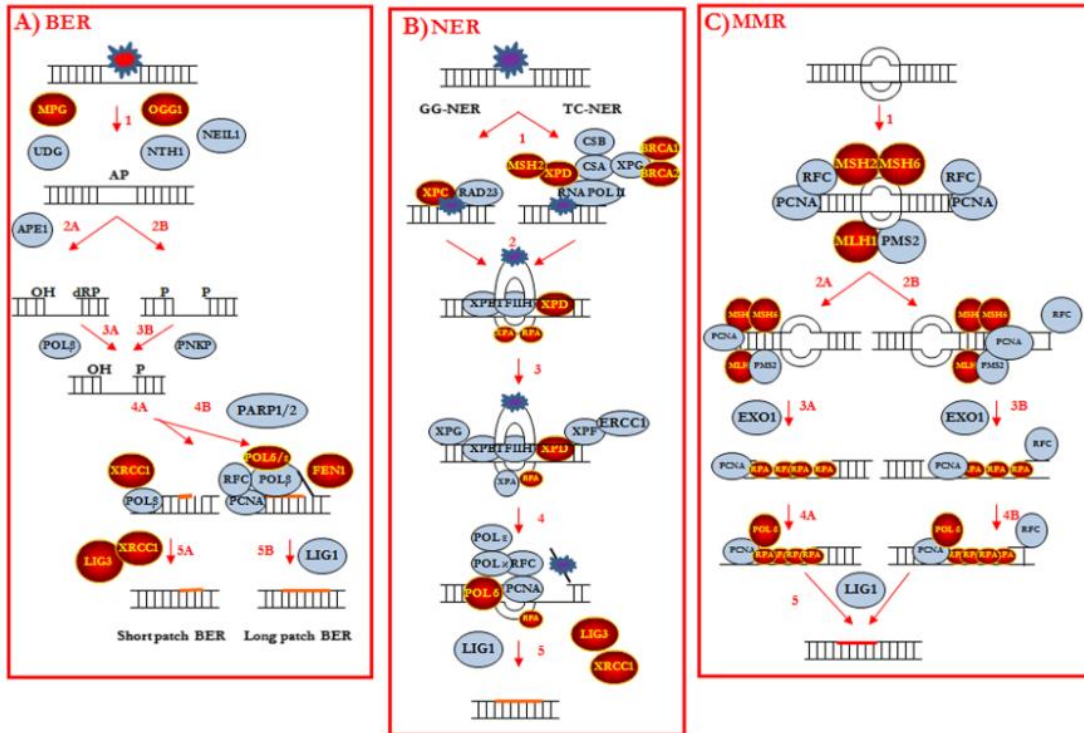


Figure 1 : Représentation des 3 voies de réparation d'une cassure simple brin. Tirée de Esposito et So [5]. Les protéines de réparation de l'ADN pour lesquelles une variation de l'expression ou des polymorphismes ont été observés chez les patients atteints de leucémie myéloïde aigue sont dénotées en rouge. **1A**) La réparation par excision de base retire les bases endommagées de l'ADN. Une base unique (réparation courte, voie A) ou de multiples bases (réparation longue, voie B) peuvent être remplacées. La portion d'ADN réparée est dénotée en rouge. Cette voie comporte 5 étapes : (1) Reconnaissance et excision de la base d'ADN fautive par les ADN glycosylases. (2) Incision du squelette de l'ADN près du site abasique. (3) Traitement des extrémités par l'ADN polymérase β (3A) ou la phosphatase PNKP de façon à générer une extrémité 5' phosphate et une 3' hydroxyle. (4) Synthèse d'ADN. Lors de la réparation courte (4A), la polymérase β, stimulée par XRCC1, synthétise l'ADN. Lors de la réparation longue (4B) la synthèse d'ADN PCNA, RFC et les polymérases β, δ, ε. La structure excédentaire en 5' et résultant de la synthèse d'ADN est excisée par l'endonucléase FEN1. (5) Ligation de l'ADN. Lors de la réparation courte (5A), les extrémités sont liguées par la ligase 3 et XRCC1 tandis que pour la réparation longue (5B), la ligation est réalisée par la ligase 1. **1 B**) Réparation par excision de nucléotides. La séquence réparée est représentée en rouge. Ce type de réparation peut emprunter deux voies : excision de nucléotides génomique globale (GG-NER) qui reconnaît les lésions situées partout dans le génome et la réparation couplée à la transcription (TCR) qui reconnaît seulement les lésions présentes dans les gènes activement transcrits. Ces deux voies diffèrent par les mécanismes impliqués dans la reconnaissance de l'ADN, mais partagent un processus commun pour l'excision des lésions, de la réparation et pour la ligation. Les étapes consistent à (1) Reconnaissance du dommage à l'ADN ; (2) ouverture de l'ADN par les hélicases XPB et XPD ; double clivage aux deux sites de mésappariements par l'endonucléase ERCC1/XPF (4) synthèse d'ADN par les polymérases κ, δ, and ε et (5) ligation par l'ADN ligase 1 lors de la phase S du cycle cellulaire ou la ligase 3/XRCC1 lors de toutes les phases du cycle cellulaire. **1 C**) Réparation de mésappariements. Les mésappariements réparés sont représentés en rouge. La réaction consiste en (1) la reconnaissance du mésappariement par l'hétérodimère MutS-MutL (2) déplacement du complexe MutS-MutL vers PCNA-RFC. Le complexe PCNA-RFC peut être situé en amont (2A) ou en aval (2B) du mésappariement. (3) Excision de l'ADN par Exo1 (4) synthèse par la polymérase δ et (5) ligation par la ligase 1. Abréviations : OGG1 (8-oxoguanine glycosylase), MPG (N-methyl purine-DNA glycosylase), NTH1 (Endonuclease III- like protein 1), NEIL1 (Nei Endonuclease VIII- like protein 1), UDG (Uracil DNA glycosylase), APE1 (Apurinic apyrimidinic endonuclease-1), POL β (ADN polymérase bêta), PNKP (*Polynucleotide Kinase 3'-Phosphatase*),

PARP1/2 (*Poly [ADP-ribose] polymerase 1/2*), XRCC1 (*X-ray repair cross-complementing protein 1*), POL δ/ϵ (ADN polymérase delta/epsilon), FEN1 (*Flap endonuclease 1*), RFC (*replication factor C*), PCNA (*Proliferating cell nuclear antigen*), LIG 1/3 (ADN ligase 1/3), XPA/C/D/G/H (*Xeroderma pigmentosum, complementation group A/C/D/G/H*), MSH2/6 (*MutS protein homolog 2/6*), CSB (ERCC6), CSA/B (*Cockayne syndrome group A/B protein* aussi ERCC8et6, respectivement), RNA POL II (ARN polymérase II), TFIIH (*Transcription factor II H*), RPA (*replication protein A*), POLX (ADN polymérase X), RFC (*replication factor C*), MLH1 (*MutL protein homolog 1*) et PMS2 (*PMS1 homolog 2*).

I.1.1.1 Réparation par excision de base

La réparation par excision de base (BER) permet de réparer un nucléotide ayant été endommagé (ex : oxydation, alkylation ou déamination). Il s'agit d'un mécanisme de réparation activé par différentes ADN glycosylases, reconnaissant chacune une lésion spécifique [6]. L'ADN glycosylase retire d'abord la base endommagée, ce qui crée un site abasique (site AP). Ensuite, une AP-endonucléase clive le squelette phosphodiester, ce qui résulte en une cassure simple brin de la double-hélice d'ADN. L'ADN polymérase β remplace par la suite le nucléotide manquant en utilisant le brin opposé comme modèle [7]. Finalement, le complexe XRCC1 : ADN ligase III α termine la réparation en rejoignant le nouveau nucléotide ajouté au squelette phosphodiester [8].

I.1.1.2 Réparation par excision de nucléotides

La réparation par excision de nucléotide (NER) peut se produire via deux mécanismes différents. Le premier mécanisme est la réparation globale du génome (GGR). Le deuxième mécanisme est la réparation couplée à la transcription (TCR). Le premier est actif en tout temps tandis que le dernier ne s'effectue que lorsque l'ADN est activement transcrit [9].

Le mécanisme GGR, ou GG-NER dans la figure 1, sert à maintenir l'intégrité du génome. Une lésion qui entraîne une distorsion importante dans l'ADN est reconnue par le complexe formé des protéines Xeroderma pigmentosum, complementation group C (XPC) : RAD23B : Centrin-2 (CETN2) [10]. Pour les plus petites distorsions de la double-hélice d'ADN, les protéines *DNA damage-binding protein 2* (DDB2) et *DNA*

damage-binding protein 1 (DDB1) vont former un complexe et créer une plus grosse distorsion qui peut alors être détectée par le complexe XPC-RAD23B-CETN2 [11, 12]. Le complexe XPC : RAD23B : CETN2 recrute par la suite le complexe multiprotéique *Transcription factor II H* (TFIIH) et les endonucléases *XP complementation group F* (XPF) : *ERCC excision repair 1* (ERCC1) et *XP complementation group G* (XPG) [13, 14]. TFIIH ouvre la double-hélice d'ADN, puis XPF et XPG coupent le squelette phosphodiester de l'ADN [15]. L'ADN polymérase δ ou ϵ va ensuite synthétiser le brin d'ADN et refermer la cassure [16].

I.1.1.3 Réparation de mésappariement

La réparation de mésappariement (MMR) est un mécanisme de réparation visant à corriger les mésappariements de nucléotide qui peuvent survenir lors de la réplication de l'ADN. Lors de la détection d'un mésappariement, *MutS protein homolog* (MSH)2 est recruté au site et s'associe avec MSH6 et MSH3 pour former les complexes MutS α et MutS β respectivement [17]. MutS α et MutS β sont des ATPases qui détectent et initient la réparation par MMR. MutS α détecte les mésappariements d'un seul nucléotide et les petites insertions et délétions (indels) de 1 ou 2 nucléotides, alors que MutS β détecte les plus longues [18]. Après que la lésion ait été enlevée, MutS α ou MutS β recrute *proliferating cell nuclear antigen* (PCNA), qui va s'installer à la jonction entre l'ADN simple brin (ADNsb) et l'ADN double brin (ADNdb) et participe au recrutement de complexes protéiques [17]. Différents complexes protéiques sont nécessaires à la réparation de mésappariement comme hMutL α formé du complexe entre *human MutL homolog* (hMLH)1-*human PMS1 protein homolog 2* (hPMS2), hMutL β formé du complexe entre hMLH1-hPMS1 et MutL γ formé du complexe hMLH1-hMLH3 [19]. L'ADN manquant ADN polymérase δ , avec l'aide de la protéine PCNA, et l'ADN ligase I ont été suggérées comme étant responsables de la synthèse de l'ADN manquant et de la ligation du brin nouvellement synthétisé [20-22]. Ce type de réparation prévient la génération d'une mutation qui deviendrait permanente dans la cellule.

I.1.2 Cassures double brin

La présence d'une cassure double brin (DSB) dans la cellule est particulièrement grave, car elle peut mener à un réarrangement génomique. Sous la menace d'instabilité génomique, la cellule doit être en mesure de reconnaître ce type de dommage rapidement. La voie de réparation des dommages à l'ADN a comme fonction principale la reconnaissance rapide et la réparation précise des DSB [23]. Deux mécanismes principaux entrent en jeu lors de la réparation de DSB i) la jonction d'extrémités non homologues (NHEJ) et ii) la recombinaison homologue (HDR) [24, 25]. Il existe toutefois un troisième mécanisme nommé réparation par appariement simple brin (SSA) qui partage des aspects de la réparation par NHEJ et par HDR. De façon intéressante, les étapes initiales de ces processus sont similaires. Le but premier étant de reconnaître et de stabiliser la DSB.

I.1.2.1 Point de contrôle et réponse aux dommages à l'ADN

Les points de contrôle ont d'abord été décrits comme étant un retard lors de la transition du cycle cellulaire permettant la réparation [26]. Toutefois, ces derniers sont maintenant liés à différentes fonctions, comme la transcription des gènes de réparation de l'ADN [27]. Parmi les différents points de contrôle, ceux impliqués dans l'évaluation de l'intégrité génomique jouent un rôle crucial dans la progression du cycle cellulaire. Le point de contrôle du dommage à l'ADN consiste en une cascade de phosphorylation complexe qui permet la reconnaissance de DSB et l'activation de kinases responsables de la transduction du signal. Parmi celles-ci, on compte les kinases *ataxia telangiectasia-mutated* (ATM) et *ataxia telangiectasia and Rad3-related* (ATR), qui avec *DNA-dependent protein kinase catalytic subunit* (DNA-PKcs), jouent un rôle central dans l'activation du point de contrôle via l'activation des kinases *checkpoint kinase* (CHK) 1 et CHK2 responsables de la transmission du signal [28-30]. L'ensemble du mécanisme est illustré à la Figure 2 .

Lorsqu'elle survient, la DSB est rapidement reconnue par l'interaction des extrémités d'ADN coupées avec le complexe MRN composé des protéines Mre11, Rad50 et la nibrine (Nbs1) [31]. Dans ce complexe, Mre11 possède une activité endonucléase permettant l'alignement des deux bouts coupés, tandis que Rad50 lie

l'ADN de façon ATP dépendante [32]. MRN recrute ensuite ATM, qui à son tour phosphoryle l'histone *H2A histone family, member X* (H2AX) et stimule la dégradation d'un des brins d'ADN menant à la formation d'ADN simple brin [33-36]. L'ADN simple brin est couvert par la protéine RPA. C'est cette dernière qui permet, en collaboration avec *ATR interacting protein* (ATRIP) et le complexe RAD9 : RAD1 : HUS1, l'activation d'ATR [37]. Un aspect important de la réponse aux dommages à l'ADN est le relâchement des histones. Ce dernier est nécessaire pour permettre à la machinerie de réparation d'atteindre le site de cassure. Le complexe NuA4 est formé de 16 sous-unités et possède des fonctions d'acétyltransférase et d'hélicase [38]. NuA4 acétyle l'histone 4, réduisant ainsi les interactions histone-histones et permettant l'accès à la chromatine [38]. Différents processus ont été suggérés pour influencer le type de réparation qui sera utilisé par la suite, mais ces derniers demeurent mal compris. Le type de résection des brins d'ADN semble être un des facteurs qui influence le type de réparation. Il a été suggéré qu'une résection courte des brins d'ADN au site de coupure favoriserait la réparation par NHEJ tandis qu'une plus longue favoriserait celle par HDR [39]. De façon intéressante, les facteurs *breast cancer 1* (BRCA1), *p53-binding protein 1* (53BP1) et *Poly ADP-ribose polymerase 1* (PARP1), précédemment associés au processus de choix du type de réparation, jouent tous un rôle dans la résection des extrémités de la DSB [40, 41].

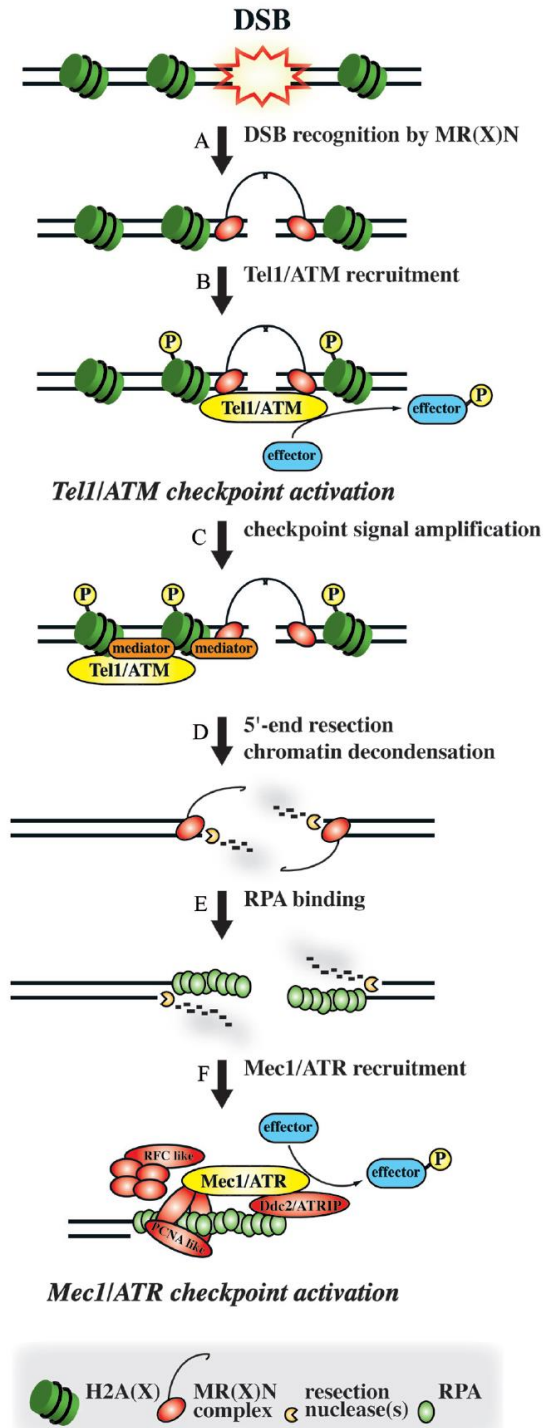


Figure 2 : Représentation schématique de l'activation du point de contrôle du dommage à l'ADN. Tiré de Pardo et al., [27]. L'étape A est la reconnaissance de la DSB, suivi par l'étape B qui est le recrutement d'ATM. Après l'activation du point de contrôle (étape C), il y a résection de l'ADN et décondensation de la chromatine (étape D). La protéine RPA se lie ensuite aux simple-brins d'ADN nouvellement formés (étape E), qui recrute ensuite Mec1/ATR (étape finale). Abréviations : DSB (Double strand break), MR(X)N complexe de Mre11, Rad50 et NBS1 (nibrine pour l'homme, Xrs2 dans la levure) , Tel1 (*Serine/threonine-protein kinase tel1*), Mec1 (*Serine/threonine-protein kinase MEC1*), Ddc2 (*DNA damage checkpoint 2*), ATRIP (*ATR interacting protein*), H2A(X) (*H2A histone family, member X*).

I.1.2.2 Jonction des extrémités non homologues

La réparation par NHEJ comprend 3 étapes : i) la résection de l'ADN endommagé, ii) la synthèse d'ADN par la polymérase et iii) la jonction des fragments par la ligase. Les deux extrémités des deux brins d'ADN seront jointes, pouvant mener à de petites insertions ou délétions (indels) à la jonction des deux extrémités. Le NHEJ est le mécanisme le plus souvent utilisé par les cellules humaines lors des phases G1/G0 [42]. Ceci s'explique par l'absence de gabarit pour effectuer la réparation par HDR [43]. Malgré le fait qu'une chromatide sœur soit présente en S/G2 et que la HDR soit un processus résultant en moins d'erreurs que le NHEJ, certaines études ont démontré que le NHEJ demeure tout de même le premier choix de réparation lors de ces phases [44].

I.1.2.2.1 Cascade de réparation

Les premières protéines à intervenir sont les protéines Ku. Un hétérodimère formé de Ku70 et Ku80 s'attache à une extrémité d'ADN double brin [45]. Une fois lié à l'ADN, Ku change de conformation et devient un port d'attache pour la nucléase, la polymérase et la ligase [46]. Au site d'une DSB, on retrouve deux extrémités clivées d'ADN. Il est présumé qu'un complexe Ku70/80 : ADN se retrouve à chacun des bouts de la cassure, permettant ainsi la préparation de chacune des extrémités en vue de la ligation [47]. L'hétérodimère Ku70/80 recrute alors le complexe Artemis : ADN-PKcs, qui possède une activité nucléase et la capacité d'ouvrir les structures d'ADN en épingle à cheveux [48]. Une fois installé sur l'ADN, ADN-PKcs s'autophosphoryle, changeant ainsi sa conformation ce qui permet la liaison des autres protéines impliquées dans la réparation par NHEJ [49, 50]. Le complexe Artemis : ADN-PKcs participe à la réparation en clivant une variété de types d'ADN endommagé [51]. Suite au clivage, le complexe Ku : ADN recrute une des polymérases de la famille pol X contenant un domaine C-terminal BRCA1 (BRCT) pour synthétiser l'ADN manquant à l'extrémité en surplomb. Parmi les quatre polymérases de la famille pol X, pol μ , pol λ et la *terminal deoxynucleotidyltransferase* (TdT) possèdent un domaine BRCT [52]. TdT possède la capacité de synthétiser un nouveau brin d'ADN indépendamment du brin d'ADN complémentaire, alors que pol λ synthétise un nouveau brin d'ADN en

utilisant le brin complémentaire comme gabarit. Finalement, pol μ peut synthétiser l'ADN de façon dépendante et indépendante au brin complémentaire [53]. Lorsque les extrémités d'ADN sont prêtes à être rejointes, le complexe Ku : ADN recrute le complexe *X-ray repair cross-complementing protein 4* (XRCC4) : ADN ligase IV, qui peut rejoindre des extrémités cohésives complémentaires ou des extrémités franches. La liaison de *XRC-like factor* (XLF) à ce complexe permet de rejoindre des extrémités incompatibles [54, 55].

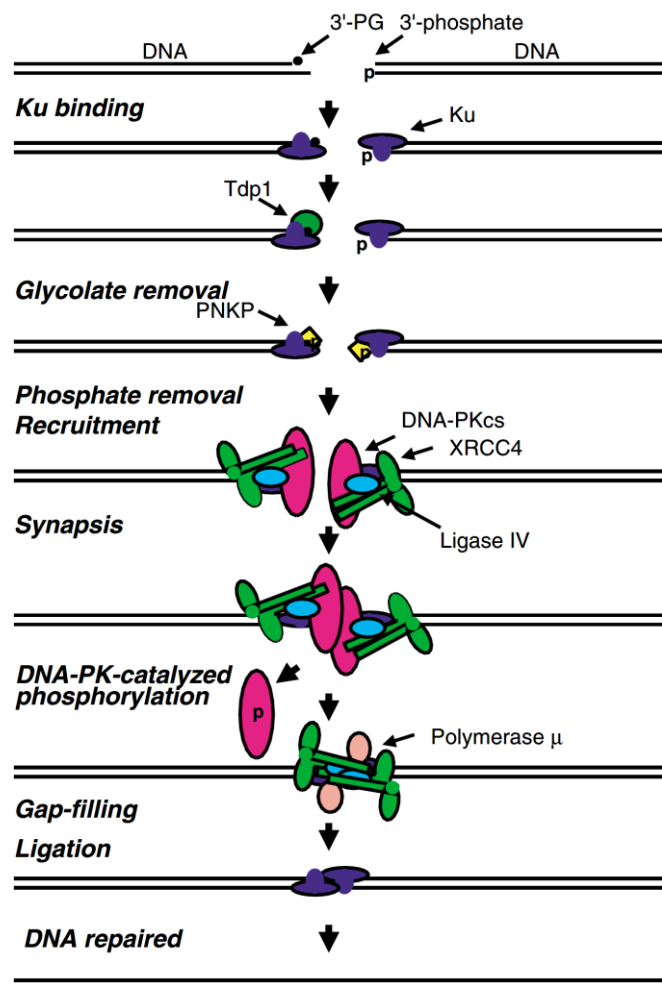


Figure 3 : Représentation schématique de la réparation d'une cassure double brin par le mécanisme de NHEJ. Tirée de Valerie and Povirk [43]. Après la liaison des protéines Ku sur les extrémités d'ADN, TDP1 et PKNP retire le 3'-phosphoglycolate et le 3'-phosphate. Ku recrute ensuite les DNA-PKcs qui aide à enligner les deux extrémités d'ADN. Une fois que les deux complexes de DNA-PKcs se rejoignent, ils s'autophosphorylent ce qui engendre leur départ, ce qui laisse place à l'arrivée de XRCC4

et de l'ADN ligase IV. Abréviations : TDP1 (*Tyrosyl-DNA phosphodiesterase 1*), PKNP (*Polynucleotide Kinase 3'-Phosphatase*), DNA-PKcs (*DNA-dependant protein kinase catalytic subunit*), XRRC4 (*X-ray repair cross-complementing protein 4*).

I.1.2.3 Réparation dirigée par séquences homologues

La HDR comprend 3 étapes : i) l'invasion d'un ADNsb sur un guide de réparation (chromatide sœur ou donneur d'ADN) ii) la synthèse d'ADN par une polymérase et iii) la résolution de l'invasion par une hélicase suite à la ligation des brins nouvellement formés. L'étape initiale est commune aux trois types de mécanisme de HDR existants, soit le *break-induced repair* (BIR), le *synthesis dependant strand annealing* (SDSA) et le *double Holliday Junction* (dHJ). Ce n'est qu'après l'invasion par un ADNsb que ces trois mécanismes opèrent de façons différentes pour terminer la réparation. La réparation par HDR ne peut être utilisée par les cellules humaines que lors de la phase S et la phase G2 [56]. Ceci s'explique par l'accessibilité à une copie du brin d'ADN non-endommagé pour effectuer la réparation pendant ces phases du cycle cellulaire.

I.1.2.3.1 Cascade de réparation HDR

RAD51 se lie à l'ADNsb avec l'aide des 5 paralogues de RAD51, soit Rad51B, Rad51C, Rad51D, XRCC2 et XRCC3. Les paralogues de RAD51 favorisent les interactions entre l'ADNsb et RAD51, ce qui lui permet de remplacer *replication protein A* (RPA) qui couvre l'ADNsb [57]. BRCA2 aide ensuite à l'organisation de RAD51 en microfilament sur l'ADNsb [58]. Subséquemment, RAD54 se lie au microfilament de RAD51 sur l'ADNsb et aide la formation de la boucle de déplacement (*D-loop*, l'invasion de la chromatide sœur / du donneur d'ADN), grâce à son activité ATPase qui lui permet d'ouvrir une double hélice d'ADN [59].

Il est suggéré que l'activité hélicase de RAD54 lui permette de créer de l'espace pour le recrutement des autres protéines au site de l'invasion de l'ADNsb. RAD54 est aussi responsable de l'expulsion de RAD51 de l'ADNsb, permettant ainsi à une ADN polymérase de s'installer [43, 60].

Suite à l'invasion de l'ADNsb et à la formation de la boucle en déplacement, les trois mécanismes de réparation par HDR empruntent différents mécanismes pour poursuivre la réparation.

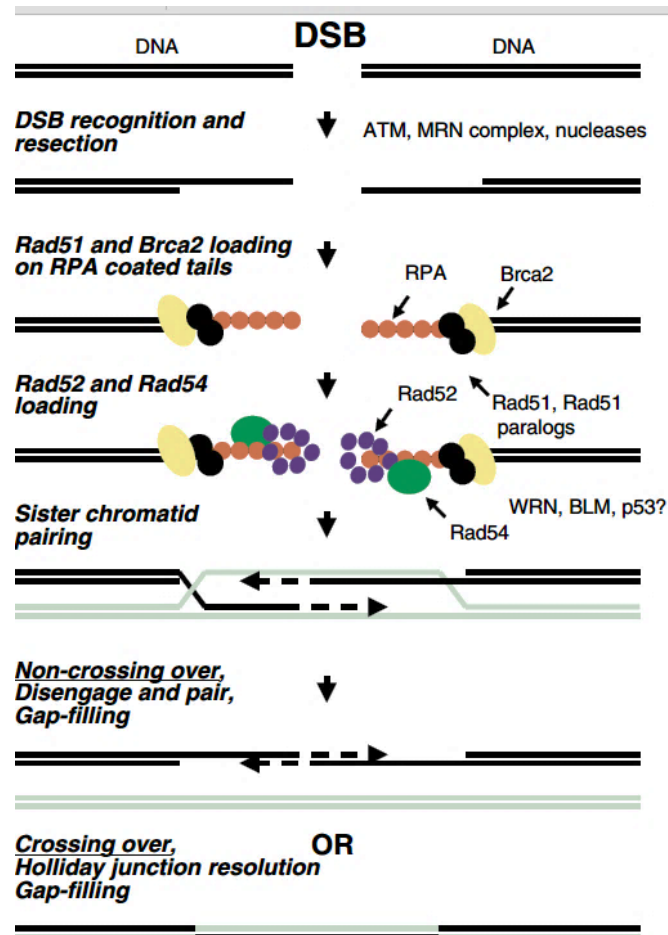


Figure 4 : Schéma de la réparation par HDR d'une cassure double brin. Tirée de Valerie et Povirk [43]. Le mécanisme commence par ATM qui détecte la cassure double-brin et phosphoryle l'histone H2AX, ce qui résulte en le recrutement de BRCA1 et NBS1, qui sont également phosphorylés par ATM. Le complexe MRN, avec l'aide de d'autres nucléases, résectionne ensuite l'ADN en ADNsb. RPA se lie ensuite à cet ADNsb. BRCA2 est recruté par BRCA1 au site de la cassure, et permet le recrutement de RAD51 avec l'aide des paralogs RAD51. RAD51 recrute ensuite RAD52 et RAD54, probablement avec l'aide des protéines BLM, p53 et de l'hélicase WRN. Il y a ensuite deux possibilités d'appariement des chromatides sœurs. La première est sans croisement et s'effectue par simple remplissage de l'ADN manquant. La seconde est un enjambement des chromatides sœurs, aussi nommée résolution d'une jonction de Holliday. Les ADN ligases et ADN polymérases associées au HDR ne sont pas encore connues. Abréviations : HDR (recombinaison homologue), ATM (*Ataxia telangiectasia-mutated*), H2AX (*H2A histone family, member X*), BRCA1 (*Breast cancer 1*), NBS1 (*Nibrine*), MRN (complexe de Mre11, Rad50 et Nbs1), RPA (*replication protein A*), BRCA2 (*breast cancer 2*), BLM (*bloom syndrome protein*), p53 (*tumor protein p53*), hélicase WRN (hélicase Werner).

I.1.2.3.1.1 Cascade de réparation de la double jonction de Holliday

Lors de la réparation par dHJ, deux voies de réparation sont possibles suite à la formation de la première boucle en D et de l'invasion du premier brin : soit l'ADN de l'autre côté de la cassure procède aussi à l'invasion de l'ADN et forme une deuxième boucle de déplacement, soit l'ADN s'apparie à la boucle de déplacement déjà formée. RAD52 est responsable de cet appariement, car il a la capacité de se lier à l'ADN protégé par RPA [61].

Lorsque deux boucles de déplacement sont formées, il en résulte deux jonctions de Holliday qui peuvent être résolues de trois façons. La première résolution se fait par le complexe *Bloom syndrome protein* (BLM) : TopoIII α qui crée deux brins d'ADN sans enjambement chromosomique [62, 63]. Un autre complexe possédant une activité résolvasse a été inféré dans des extraits de cellules humaines, mais son identité demeure inconnue. L'invalidation de RAD51C et de XRCC3 diminue grandement l'efficacité de la résolution des jonctions de Holliday dans les cellules de mammifères, il est donc suggéré que des protéines participant déjà à la réparation par HDR posséderaient cette activité résolvasse [27, 64]. La résolution faite par cette résolvasse inconnue résulte en deux possibilités, soit il y a enjambement des deux brins d'ADN, soit il n'y en a pas.

I.1.2.3.1.2 Cascade de réparation du BIR

Le mécanisme de réparation par BIR est utilisé lorsqu'une cassure double brin survient à une fourche de réplication, résultant en un ADN double brin d'un côté, et simple brin de l'autre côté de la cassure. Un des brins d'ADN du côté double brin va former une boucle d'invasion sur l'ADNsb de l'autre côté de la cassure. Ensuite, la machinerie cellulaire va créer une nouvelle fourche de réplication, et reproduire le brin manquant, [65] jusqu'à ce que le brin nouvellement synthétisé rejoigne la séquence d'ADN présente sur l'ADNsb situé de l'autre côté de la cassure [66]. La synthèse de ce nouveau brin d'ADN requiert la présence de la sous-unité pol32. Il a été suggéré que pol32 permette de créer une nouvelle fourche de réplication sans la présence d'origine de réplication [67].

I.1.2.3.1.3 Cascade de réparation SDSA

Le mécanisme de réparation par SDSA se produit chez les cellules qui se divisent tant par mitose que par méiose et ne résulte pas en l'échange de matériel génétique. Lors de ce processus, il y a synthèse par une polymérase d'un brin d'ADN jusqu'à ce qu'il y ait homologie avec l'ADN situé de l'autre côté de la cassure double brin. Il est supposé que BRCA2 soit responsable de l'appariement du brin nouvellement synthétisé avec l'ADN situé de l'autre côté de la cassure double brin [68]. Ensuite, il est suggéré que la boucle d'invasion est dissoute par l'hélicase BLM [69]. Suite à l'appariement de l'ADNsb et au retrait de la boucle d'invasion, le brin manquant est synthétisé par une polymérase et joint grâce à une ligase.

I.1.2.4 Réparation par appariement simple brin

La réparation par appariement simple brin (SSA) est une voie de réparation unique, puisqu'elle ne nécessite pas de gabarit d'ADN. Le gabarit de chromatide sœur est remplacé par la présence de séquences répétées de chaque côté de la cassure. Lorsque ce type de réparation est utilisé, il y a perte du matériel génétique se situant entre les deux séquences répétées. L'étape initiale de la réparation par SSA est similaire aux autres voies utilisant le HDR et implique la résection, résultant en un surplomb des extrémités 3'. Toutefois, RAD51 n'est pas nécessaire pour la réparation par SSA, ce qui distingue ce mécanisme du HDR. La suite du mécanisme par SSA comporte : i) l'appariement simple brin de deux ADNsb possédant une séquence homologue, ii) la synthèse d'ADN par une polymérase et iii) la jonction des brins nouvellement synthétisés par une ADN ligase. L'ADN situé entre les séquences homologues qui s'apparient est perdu.

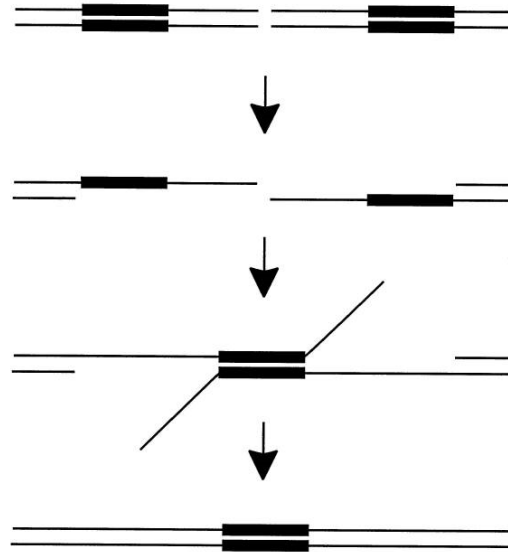


Figure 5 : Représentation schématique de la réparation par SSA. Une cassure double brin à lieu près de deux régions répétées. Formation d'un surplomb en 3'. Appariement des séquences homologues et résolution de la cassure. Tirée de Sugawara et al., [70].

I.1.2.4.1 Cascade de réparation SSA

La première protéine à intervenir lors de la réparation par SSA est CtiP, qui aide à promouvoir la résection de l'ADN double brin 5' de chaque côté de la cassure double brin [71]. Ensuite, RAD52 se lie au surplomb d'ADNsb en 3' et, par contact avec les protéines RPA recouvrant l'ADNsb, aide à l'appariement entre deux séquences identiques situées de chaque côté de la cassure double brin [72, 73]. Une fois les deux ADNsb appariés, le complexe ERCC1 : XPF clive les extrémités flottantes. RAD52 est l'intermédiaire situé entre le complexe ERCC 1 : XPF et l'ADNsb [74]. Lors de la réparation par SSA, la machinerie cellulaire impliquée dans le MMR assure un appariement parfait entre les deux brins d'ADNsb [75]. Il a été démontré que RAD51 et ses paralogues inhibent la réparation par SSA [76]. Remarquablement, une invalidation de RAD51 favorise 90 fois plus le SSA que dans des cellules de mammifères normales [76].

I.2.0 Les nucléases d'ingénierie du génome

Le séquençage du génome humain a mené à l'identification de nombreux gènes et séquences aux fonctions précédemment inconnues [77]. À l'aide de certains outils bio-informatiques, il est maintenant possible de prédire quelle sera la fonction d'un gène. Toutefois, ces analyses permettent d'émettre certaines hypothèses qui doivent être testées *in vitro* et *in vivo* [78]. Ces techniques se basent principalement sur l'inactivation et la surexpression des gènes d'intérêt et l'étude des phénotypes associés. La première nucléase d'ingénierie, la *zinc-finger nuclease* (ZFN), a été mise au point en 1996 par la fusion de protéines à doigt de zinc (ZFP) reconnaissant une séquence spécifique d'ADN avec le domaine nucléase FokI [79]. En 2011, les *trans activator-like effector nuclease* (TALEN) ont été développés, suite à la découverte de nouveaux domaines de liaisons à l'ADN provenant du pathogène *Xanthomonas* pouvant se lier à l'ADN de plante [80-82]. Finalement, le dernier système de nucléase d'ingénierie à avoir été mis au point pour cliver des séquences d'ADN spécifiques est *Clustered Regularly Interspaced Palindromic Repeat* (CRISPR) en 2012 [83]. Les nucléases d'ingénierie permettent la formation de cassures double brin à un endroit spécifique et prédéterminé dans l'ADN [84]. Ces trois types de nucléases d'ingénieries sont le résultat d'une quête visant à obtenir des enzymes pouvant reconnaître de façon spécifique une séquence d'ADN puis la cliver [79]. Le plus gros avantage de ces nucléases d'ingénierie est leur versatilité. Lorsqu'une nouvelle séquence d'ADN doit être ciblée, il ne suffit que de créer un nouveau domaine de liaison à l'ADN. Selon le type de nucléase, ce nouveau domaine de liaison à l'ADN est plus ou moins facile à synthétiser [85-87].

I.2.1 Les nucléases à doigt de zinc

I.2.1.1 Domaine de reconnaissance de l'ADN

Les ZFN sont composées de protéines à doigt de zinc permettant la reconnaissance d'une séquence spécifique d'ADN fusionnées avec la nucléase FokI. Les protéines à doigt de zinc proviennent de domaines de liaison à l'ADN de facteurs

de transcription eucaryotes [88]. Chaque domaine à doigt de zinc cible 3 paires de bases consécutives et des fusions de 3 ou 4 domaines sont assemblées afin de cibler une séquence d'ADN spécifique de 9 ou 12 nucléotides [89]. Les ZFP reconnaissent leur triplet de nucléotide en s'associant sur les bases azotées dans le sillon majeur de la double-hélice d'ADN [90, 91]. Pour créer une ZFN reconnaissant un motif unique dans le génome humain, il est estimé qu'il est nécessaire de cibler une séquence d'environ 16 nucléotides [92]. Ce calcul a été fait en présumant une répartition aléatoire des nucléotides sur une séquence donnée. À partir d'une séquence de 16 nucléotides (soit 4^{16} qui correspond approximativement à 4,29 milliards de nucléotides), la probabilité est donc que cette séquence soit unique [92]. En utilisant 6 ZFP, soit deux ZFN de trois ZFP fonctionnant en dimère, signifie une spécificité de 18 nucléotides, ce qui devrait être suffisant pour que la séquence soit unique dans le génome.

I.2.1.2 Fonction nucléase

La nucléase utilisée en combinaison avec les ZFP pour constituer une ZFN est le domaine nucléique provenant de l'enzyme de restriction FokI [79]. Cette dernière fait partie de la classe IIS des enzymes de restriction qui clivent l'ADN en amont de leur site de reconnaissance [93]. Comme la majorité des enzymes de restrictions du type II (ex : EcoRI et EcoRV), le site catalytique de l'enzyme FokI doit former un dimère pour pouvoir cliver l'ADN. L'enzyme de restriction FokI est du type IIS et clive en dehors de son site de reconnaissance à l'ADN. Chaque monomère de l'enzyme de restriction FokI doit dans un premier temps lier l'ADN, puis interagir avec un second monomère de façon à induire une cassure double brin [94]. Ce processus nécessite la liaison indépendante à l'ADN de deux monomères, et ce, de chaque côté du site de clivage. Pour créer une cassure double brin, deux ZFN sont donc nécessaires dans une orientation et un espacement adéquat [94].

Une nucléase FokI thermosensible a également été développée. Cette nucléase a été développée avec à l'aide d'un système mis au point chez la levure (mettre ref doyon et al 2011) et permettant d'isoler des nucléases mutantes possédant des caractéristiques nouvelles [95]. Avec ce système, des domaines FokI thermosensibles

(FokI-ts) ont été isolés, possédant la particularité d'être actives à la température permissive de 30°C, mais d'être inactive à 37°C.

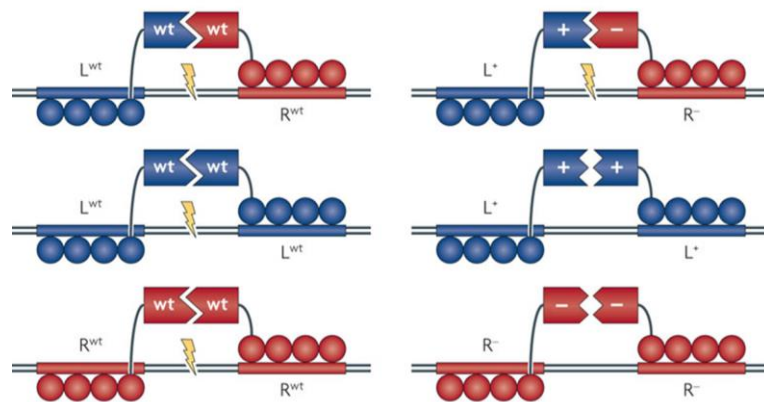
I.2.1.2.1 Augmenter la spécificité des nucléases

La spécificité des nucléases d'ingénierie est une préoccupation importante lors du processus d'optimisation [95-97]. Pour augmenter la spécificité du domaine FokI, des domaines FokI mutants ont été créés afin d'obliger chaque ZFN à n'avoir qu'un seul partenaire d'interaction fonctionnel [95]. Cela empêche chacune des deux ZFN de former un homodimère et de cliver une séquence hors cible.

I.2.1.3 Optimisation du système des ZFN

Un avantage important du système des ZFN est que le domaine de liaison à l'ADN et le domaine nucléase peuvent être modifiés de façon indépendante. Il n'existe présentement aucune modification pouvant être apportée aux ZFN menant à une amélioration de leur liaison à l'ADN. Toutefois, il est possible de modifier la nucléase FokI en modifiant certains acides aminés importants. La nucléase FokI comporte des résidus critiques à sa fonction. Par exemple, l'acide aminé majoritairement responsable de l'activité endonucléase est l'acide aspartique 450 (D450) [98]. Plusieurs autres résidus importants ont aussi été répertoriés tels que D483, R487 ainsi que les acides aminés situés à l'interface de dimérisation [98]. La mutation de certains de ces résidus mène à une forte baisse de l'activité de FokI, soulignant ainsi leur importance [95]. Grâce à la mutagenèse dirigée, il est possible d'isoler des variants de FokI possédant une activité augmentée [99]. En effet, les mutations S418P, F432L, K441E, Q481H, H523Y, N527D, et K559Q mènent à une activité 4 à 5 fois plus élevée. L'étude de ces mutations de façon individuelle a permis de déterminer que S418P and K441E étaient les mutations conférant l'augmentation d'activité observée.

En 2007, Sangamo Biosciences a introduit le concept de ZFN agissant en hétérodimères obligatoires [100]. La mutation des acides aminés à l'interface de dimérisation fait en sorte que les nucléase FokI formant le dimère ne peuvent interagir qu'avec leur partenaire compatible.



Nature Reviews | Genetics

Figure 6 : À droite, schéma de ZFN de type sauvage et de leur possibilité d'association menant à de potentiels clivages hors cible. À gauche, schéma d'hétérodimérisation obligatoire entre deux nucléases dont les interfaces de dimérisation sont complémentaires. Tirée de Urnov et al., [101].

Un exemple de l'application des ZFN en hétérodimères obligatoires est l'utilisation de ZFN possédant des domaines FokI E490K I538K (FokI KK) et FokI Q486E I499L (FokI EL) [100]. Ces modifications permettent d'augmenter la spécificité des nucléases, puisque leurs homodimères ne peuvent s'assembler correctement et cliver l'ADN à un site hors-cible. Malheureusement, l'utilisation de ZFN en hétérodimères obligatoires diminue l'efficacité de clivage au site d'intérêt lorsque comparé à celle des ZFN possédant FokI type sauvage [100]. Malgré tout, il a été démontré qu'il était possible d'améliorer l'activité des ZFN fonctionnant en hétérodimères obligatoires, en ajoutant une mutation supplémentaire à l'interface de dimérisation [95]. Cette dernière permet d'améliorer la dimérisation entre deux hétérodimères, grâce à l'attraction électrostatique entre les nouveaux acides aminés présents à l'interface de dimérisation. Dans ce modèle amélioré, la paire FokI KK + FokI EL est devenue FokI KKR (H537R) + FokI ELD (N496D). Ce variant conserve la spécificité associée avec l'utilisation d'hétérodimères obligatoires, mais possède une activité de clivage similaire à celle de la nucléase FokI type sauvage [95].

Une deuxième paire de FokI à hétérodimères obligatoires a été créée, soit FokI RDD (R487D, N496D et H537R) avec FokI DRR (D483R, N496D et H537R). Cela

permet de créer deux cassures simple brin de façon complètement indépendante dans une même cellule [95].

En utilisant une approche différente, une équipe a incorporé la mutation D450N menant à l'inactivation de son domaine endonucléase [102]. Lorsqu'une ZFN de type sauvage dimérise avec une ZFN D450N, il en résulte un clivage simple brin. Considérant que la réparation d'une cassure simple brin induit peu ou pas d'indels, il est possible de suggérer que l'utilisation de ZFN nickases possède le potentiel de réduire les mutations hors cible [102]. Toutefois, l'induction efficace d'une mutation au site d'intérêt nécessite la création d'une cassure double brin. Pour ce faire, deux ZFN nickases pourraient être utilisées de façon simultanée de façon à créer deux cassures simples brins rapprochés. Peu de données sont présentement disponibles sur l'utilisation et le potentiel des ZFN doubles nickases.

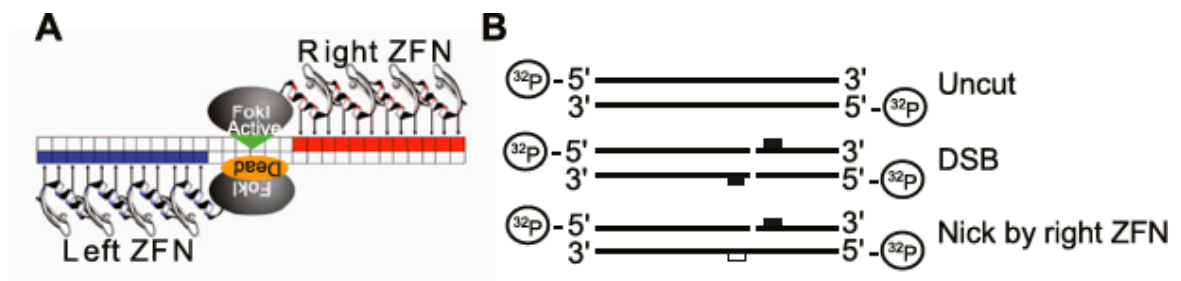


Figure 7 : Schématisation du concept de ZFN nickase. A) Schéma de l'assemblage d'un appariement FokI D450N avec FokI type sauvage permettant une cassure simple brin. B) Type de cassure attendue selon le type de nucléase utilisé. Les boîtes noires représentent la liaison d'une ZFN avec un domaine FokI de type sauvage. La boîte blanche représente la liaison d'une ZFN nickase, soit une ZFN avec le domaine FokI D450N ne pouvant cliver son brin d'ADN. Tirée de Wang et al., [103].

1.2.1.4 Les ZFN et l'édition de génome

La première utilisation des ZFN a ouvert la porte à la possibilité de sonder le génome de façon beaucoup plus rapide que les techniques utilisées précédemment [104]. Cette dernière a aussi facilité le développement de modèles de nombreuses maladies monogéniques [105, 106]. Les ZFN ont été utilisées pour créer des

modifications dans divers organismes, comme le plant de maïs, le poisson-zèbre (*Danio rerio*) et la souris [107-109]. Les ZFN sont également les premières nucléases d'ingénierie à avoir été utilisées en essais cliniques, dans un traitement de patients atteints du VIH [110].

I.2.1.5 Limites des ZFN

Bien que les ZFN aient été les premières nucléases d'ingénierie utilisées, il est important de noter que leur utilisation comporte certaines limites. Les plus grandes limites de l'utilisation des ZFN sont sans contredit le coût de production élevé et la difficulté associée à la synthèse des ZFN [111]. Malgré le fait que la spécificité pour certaines bases azotées des ZFP présentement utilisées soit bien décrite, rien n'indique que la juxtaposition de plusieurs ZFP conserve la spécificité individuelle de chaque ZFP [112]. L'assemblage de plusieurs ZFP permettant l'obtention d'une paire de ZFN spécifique peut donc se révéler laborieux. Il est aussi difficile d'effectuer plus d'une modification dans une lignée cellulaire ou un animal en utilisant les ZFN, puisque la création de deux mutations simultanées implique la synthèse et l'utilisation de 4 ZFN. Finalement, la spécificité des ZFN demeure un enjeu important. Leur obligation d'agir en paire diminue les risques de cassures double brin hors cible, mais ne les élimine pas totalement [106].

I.2.2 Les nucléases effectrices de type activateur de transcription

Les nucléases *trans activating-like effectors (TALE)*, ou TALEN sont la 2^e génération de nucléase d'ingénierie [82]. Tout comme les ZFN, les TALEN reconnaissent l'ADN via une interaction protéine-ADN, entre le domaine TALE et l'ADN et clivent l'ADN grâce à leur fusion à l'endonucléase FokI.

I.2.2.1 Liaison à l'ADN

Le domaine de liaison à l'ADN des TALEN est le TALE. Il a été découvert dans des pathogènes de plante du genre *Xanthomonas Sp.* [113]. Les TALE ont la capacité d'entrer dans le noyau cellulaire des plantes infectées par *Xanthomonas* et de

se lier à l'ADN, sur des séquences effectrices. Ceci permet d'activer la transcription de certains gènes qui sont, en règle générale, favorables à l'infection par cette bactérie [114]. Les TALE sont constitués de répétitions en hélice de 33 à 35 acides aminés [115]. La majorité de ces répétitions sont invariables, sauf pour 2 acides aminés, communément appelés RVD (*repeat variable diresidue*) [115].

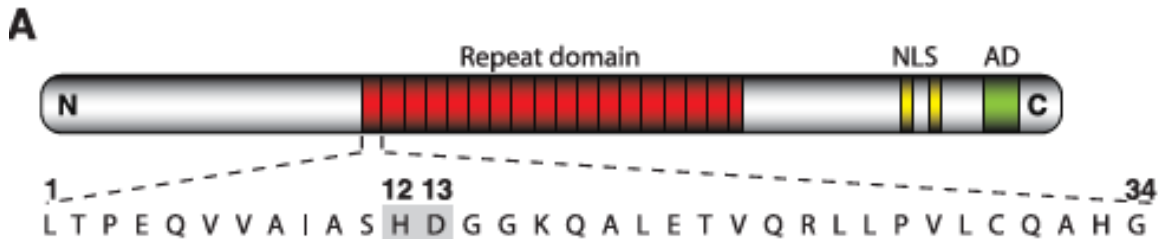


Figure 8 : Schéma d'un TALE. Les résidus 12 et 13 sont les résidus variables dans les répétitions du TALE qui vont reconnaître l'ADN. Abréviations : NLS (Signal de localisation nucléaire), AD (Domaine d'activation). Tirée de Boch et al., 2009 [113].

Selon le RVD présent dans la répétition, un nucléotide différent est reconnu sur l'ADN. En combinant une série de répétitions de TALE possédant des RVD différents, il est possible de reconnaître une séquence spécifique d'ADN [82]. Toutefois, il est important de mentionner que les RVD favorisent la reconnaissance d'un nucléotide spécifique, mais il est aussi possible pour ceux-ci d'en reconnaître un autre (voir figure 9) [116]. Ceci-soulève donc une problématique potentielle de spécificité.

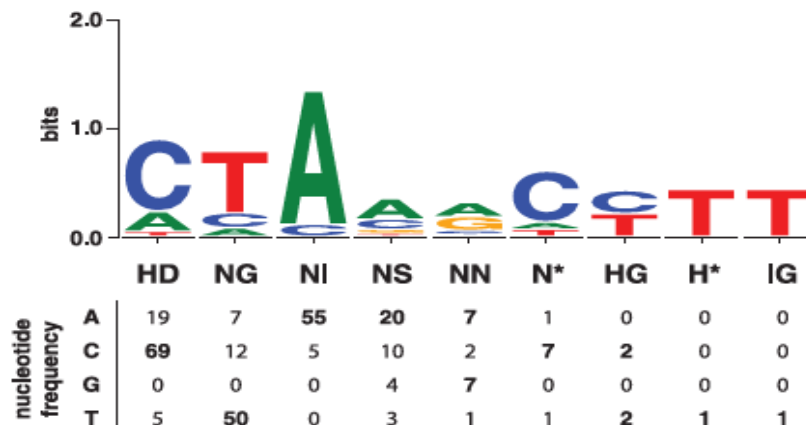


Figure 9 : Fréquence du nucléotide reconnu selon le RVD dans la répétition du TALE. Tirée de Boch et al., [113]. Abréviations : RVD (dirésidu variables répétés), H (histidine), D (acide aspartique), G (glycine), I (isoleucine), S (sérine).

I.2.2.2 Fonction nucléase

Les TALEN clivent l'ADN similairement aux ZFN, en utilisant l'endonucléase FokI [82]. Toutes les modifications et améliorations faites au domaine FokI utilisées pour les ZFN et décrites à la section 2.1.2 peuvent aussi être utilisées pour les TALEN [117].

I.2.2.3 Les TALEN et l'édition du génome

Suite à l'avènement des ZFN, plusieurs études se sont tournées vers l'utilisation de TALEN pour comparer l'efficacité des deux types de nucléases et ont démontré une capacité similaire à induire des cassures double brin [82, 118]. Les premières preuves de concept pour l'utilisation des TALEN ont été réalisées en utilisant deux séquences de TALE bien connues, soit la séquence de AvrBs3 (de *Xanthomonas campestris*) et la séquence de PthXo1 (de *Xanthomonas. oryzae*) [113]. Suite au clonage du monomère de la nucléase FokI avec ces TALE, les TALEN gagnent la capacité de cliver l'ADN double brin lorsqu'utilisées en dimère [80]. De nombreux organismes modèles modifiés génétiquement ont aussi été créés via l'utilisation de TALEN [119, 120]. Certaines études *in vitro* ont démontré le potentiel des TALEN pour la modification génique. Entre autres, il est possible d'invalider le récepteur des cellules T (TCR) dans plus de 80% des lymphocytes isolés du sang humain et même d'intégrer un marqueur fluorescent au site de coupure lorsqu'un donneur de réparation est fourni aux cellules [121, 122]. Ces résultats démontrent un potentiel translationnel important puisque la majorité des études cliniques présentement approuvées effectue la modification des cellules *ex vivo* suivie d'une réimplantation des cellules corrigées.

Il existe quelques études cliniques approuvées qui utilisent les TALEN, dont deux par la compagnie *Collectis*TM utilisant des cellules T génétiquement modifiées avec des récepteurs à antigène chimérique pour traiter deux maladies immunitaires [123] et une

étude approuvée en 2017 en Chine pour le néoplasme cervical lié au virus du papillome humain (Clinical Trial # NCT03057912).

I.2.2.4 Limites associées à l'utilisation des TALEN

Comme leur design et leur mode d'action sont similaires, les ZFN et la TALEN possèdent des limites similaires. Toutefois, les TALEN possèdent une contrainte supplémentaire par rapport aux ZFN : ils doivent préférentiellement reconnaître un T comme premier nucléotide [113]. La production des TALEN se fait à un coût moindre que celui des ZFN [111]. Néanmoins, l'obtention de tous les réactifs de départ afin de créer des domaines TALEN sur mesure est onéreuse, et l'assemblage est une tâche exhaustive à réaliser en laboratoire [86]. Comme c'est le cas pour les ZFN, la réalisation de mutations simultanées est complexe, puisqu'il est nécessaire de synthétiser et d'administrer 2 paires de TALEN. Finalement, la spécificité des TALEN demeure un point important à caractériser, et ce, pour différentes cibles génomiques.

I.2.3 Le système CRISPR/Cas

La 3^e génération de nucléases d'ingénierie du génome est le système CRISPR/CRISPR-associated (Cas). CRISPR/Cas est un système de défense microbien utilisant une nucléase guidée par un ARN pour se défendre contre les éléments mobiles génétiques provenant principalement de phages et de plasmides [124, 125]. Chez les bactéries, sa fonction est similaire au système immunitaire, il sert à défendre l'organisme contre l'invasion par des pathogènes [126]. Le système CRISPR/Cas bactérien est habituellement composé d'une série de séquences répétées espacées par des séquences variables d'ADN [127]. Ces séquences variables sont des fragments d'ADN pathogène qui sont insérés dans le génome bactérien lors d'une infection. Ces séquences seront par la suite transcrites et serviront à guider la nucléase vers l'ADN pathogène lors d'une infection subséquente. Suite à la reconnaissance d'une séquence pathogène, la nucléase induit une cassure double brin qui est souvent fatale pour le pathogène [128]. Le système CRISPR/Cas bactérien est dynamique et s'adapte en intégrant de nombreux fragments d'ADN exogène clivé dans son locus CRISPR. Ceci lui assure une protection contre de nombreux organismes [129].

Il existe deux classes de système CRISPR/Cas. La première comprend les sous-types I, III et IV et la deuxième comporte les sous-types II, V et VI [130, 131]. Le classement des différents systèmes se fait selon l'organisation de leur module effecteur. Ceux de la classe 1 sont présents dans les bactéries et les archéobactéries et ont de 4 à 7 sous-unités protéiques Cas. Ceux de la classe 2 sont plus rares et sont presque exclusifs aux bactéries, leur domaine effecteur est composé d'une seule protéine possédant de multiples domaines fonctionnels [131].

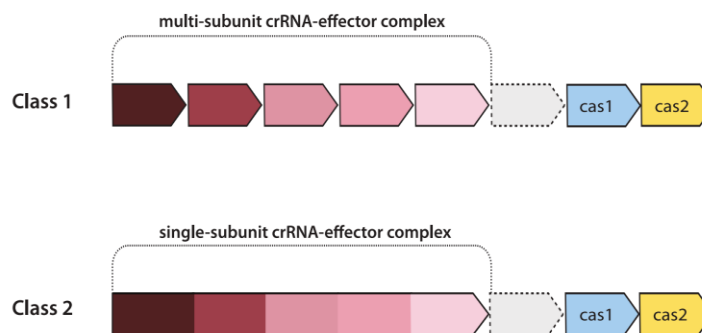


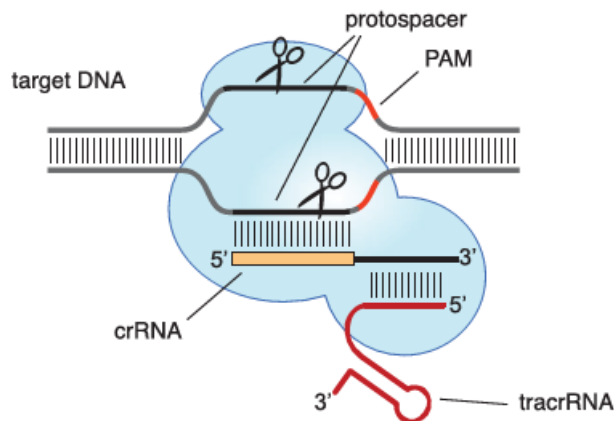
Figure 10 : Schématisation des différences entre les sous-unités multiples des systèmes CRISPR de classe 1 versus l'unique sous-unité des systèmes CRISPR de classe 2. La structure du locus des nucléases de classe 1 est composé de plusieurs complexes distincts tandis que celle des nucléases de classe 2 sont composées d'une seule structure comprenant le complexe effecteur. Tirée de Mohanraju et al [125].

I.2.3.1 Liaison à l'ADN et formation d'une cassure double brin

Le plus connu des systèmes CRISPR/Cas est CRISPR-Cas9, plus particulièrement celui provenant de la bactérie *Streptococcus pyogenes* (SpCas9). SpCas9 est le premier système CRISPR à avoir été programmé en laboratoire pour cliver une séquence d'ADN d'intérêt [83]. L'endonucléase naturelle comporte deux domaines nucléases lui servant à cliver l'ADN et nécessite un crRNA (ARN CRISPR) et un tracrRNA (trans-activating crRNA) [83, 132, 133]. Le crRNA reconnaît un site particulier sur l'ADN exogène tandis que le tracrRNA possède une partie

complémentaire au crRNA et contient une séquence qui forme une boucle permettant le chargement du complexe crRNA/tracrRNA dans la nucléase et la reconnaissance d'un site à cliver. Pour faciliter l'ingénierie des génomes, un ARN chimérique a été développé en laboratoire [83]. Cet ARN chimérique appelé ARN guide (gRNA) remplit les fonctions du crRNA et du tracrRNA afin de simplifier son utilisation [83].

Cas9 programmed by crRNA:tracrRNA duplex



Cas9 programmed by single chimeric RNA

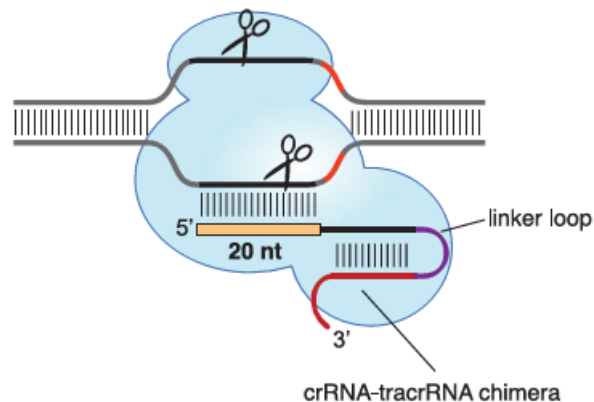


Figure 11 : Représentation schématique du clivage de l'ADN par le système CRISPR/Cas9. Haut : Système CRISPR/Cas bactérien qui nécessite la présence d'un crRNA et d'un tracrRNA pour reconnaître la séquence cible. Bas : gRNA chimérique produit en laboratoire permettant à la nucléase de reconnaître sa cible. Tirée de Jinek et al [83]. Abréviations : PAM (*protospacer adjacent motif*), crRNA (*crispr RNA*), tracrRNA (*transactivating crRNA*).

I.2.3.1.1 Le *protospacer adjacent motif*

Le *protospacer adjacent motif* (PAM) est d'une importance capitale dans la reconnaissance d'une séquence d'ADN par le système CRISPR. En effet, il s'agit d'une courte séquence d'ADN à laquelle la nucléase s'apparie. Cette séquence doit donc se trouver directement en aval de la séquence d'ADN reconnue par le gRNA [134, 135]. Le PAM est une séquence d'ADN reconnu par la nucléase sur l'ADN, et le gRNA utilisé par CRISPR doit reconnaître une séquence adjacente à ce PAM. Le motif PAM diffère selon la souche de la nucléase utilisée, mais est peu variable pour une souche donnée. Le PAM peut restreindre le nombre d'endroits possibles pour cibler une séquence. Toutefois, il s'agit souvent de petites séquences facilement retrouvées dans le génome. Par exemple le PAM de SpCas9, 5'-NGG-3', se retrouve environ à toutes les 10 paires de bases dans le génome [132]. Dans certains cas, la difficulté d'identifier un PAM dans la région d'intérêt peut poser certains problèmes. L'utilisation de nucléases Cas9 provenant de différentes espèces bactériennes peut pallier ce problème puisque ces dernières reconnaissent une séquence PAM différente. Les plus utilisées après SpCas9 sont les nucléases provenant de *S. aureus* (5'-NNGRRT-3'), SaCas9, et *S. thermophilus* (5'-NNAGAW-3') [136].

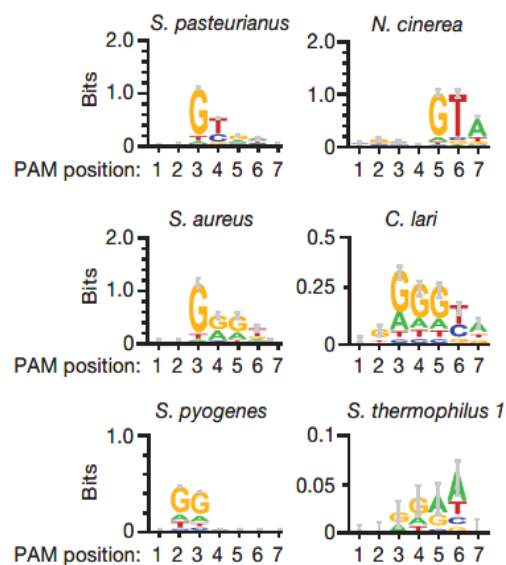


Figure 12 : Représentation des diverses séquences PAM reconnues par différentes nucléases. Tirée de Ran et al [136]. A : adénosine, T : thymine, G : guanine, C : cytosine.

I.2.3.1.2 La séquence *seed*

La reconnaissance de la séquence PAM et le bon appariement des 8 à 12 premiers nucléotides du gRNA sont essentiels pour qu'il y ait génération d'une cassure double brin par la nucléase [137]. Cette courte séquence située en 3' du crRNA nommée *seed* représente les premiers nucléotides du gRNA qui seront hybridés avec la cible d'ADN génomique [137]. Si la séquence *seed* correspond à celle de l'ADN génomique, le gRNA continuera de s'hybrider à sa cible de 3' vers 5' et la nucléase pourra cliver l'ADN [83].

I.2.3.2 Fonction nucléase

Cas9 contient deux domaines possédant une fonction nucléase. Le premier est le domaine HNH, il clive le brin complémentaire. Le second est le domaine RuvC, qui clive le brin non complémentaire [128]. Une fois le PAM reconnu et le gRNA hybridé à sa cible d'ADN génomique, les deux domaines nucléase vont effectuer une coupure franche 3 nucléotides en amont du PAM [128]. Certaines autres nucléases du système clivent l'ADN de façon différente. Par exemple, CPF1 (*CRISPR from Prevotella and Francisella 1*), une nucléase de type V, coupe l'ADN génomique 19 nucléotides en aval du PAM et en laissant un surplomb en 5' [138].

I.2.3.3 Variantes du système CRISPR/Cas

Comme les nucléases du système CRISPR possèdent deux domaines nucléases, il est possible de modifier ces derniers de façon à affecter l'habileté de la protéine à cliver l'ADN. La mutation D10A inactive l'activité nucléase du domaine RuvC, et la mutation H840A enlève l'activité nucléase du domaine HNH. La Cas9 D10A possède donc une activité de nickase tandis que la Cas9 avec les mutations D10A et H840A, aussi nommé *dead Cas9* (dCas9), ne possède aucune activité enzymatique [83, 139]. Pour CPF1, les domaines nucléases sont différents. CPF1 utilise trois résidus nécessaires à l'activité catalytique sur sa séquence cible, soit D917, E1006 et D1255, ils sont respectivement situés dans un domaine nucléase RuvC I, RuvC II et le dernier

résidu est situé dans un domaine du type protéine à doigt de zinc. La mutation D1255A n'enlève pas totalement l'activité de clivage, mais la réduit grandement [138].

Les nucléases désactivées conservent la capacité d'être guidées à n'importe quel endroit du génome par un gRNA spécifique. En liant une Cas9 inactive à l'ADN, il est possible d'interférer avec l'expression d'un gène, aussi appelé *CRISPR interference* (CRISPRi) [140]. En fusionnant une dCas9 avec un activateur de la transcription, il est possible d'induire l'activation d'un gène d'intérêt, aussi appelé *CRISPR activation* (CRISPRa) [141]. Par exemple en liant les domaines activateurs VP64, p65 et rta (l'ensemble de ces activateurs fusionnés est également appelé VPR) à une dCas9, il est possible d'induire l'expression d'un gène ciblé par un gRNA [141]. L'équipe de George Church a également vérifié l'activité de deux autres fusions d'activateur, soit i) SAM : une fusion de VP64 avec un gRNA modifié permettant le recrutement du complexe MCP : p65 : hsf1 et ii) SUNTAG : une fusion de 10 chaînes de peptides GCN4. En combinant l'utilisation d'une librairie de gRNA ciblant plusieurs gènes d'intérêt à un de ces systèmes, il est possible de cribler rapidement la répression ou l'activation de la transcription de plusieurs gènes en même temps [141, 142].

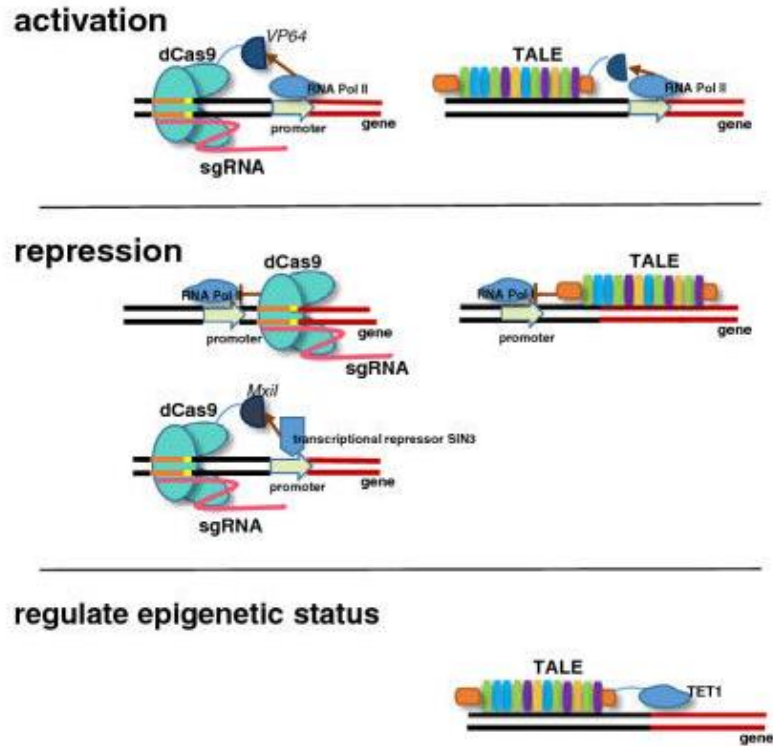


Figure 13 : Régulation de l'expression endogène avec les nucléases d'ingénierie du génome. Tirée de Chen et al., 2014 [143] La protéine Cas9 ou les TALEs peuvent être fusionnées avec des protéines effectrices comme VP64, Mxi1 afin de réguler l'expression d'un gène respectivement à la hausse ou à la baisse. Un TALE fusionné avec un effecteur désacétylant comme TET1 pourrait même aller jusqu'à réguler la disponibilité épigénétique de gène endogène en jouant sur l'acétylation d'histones. Abréviations : TALE (Transcription activator-like effector), VP64 (consiste en 4 copie de VP16 qui est la séquence nucléotidique DALDDFDLDMML de la protéine virale 16 du virus de l'herpès simplex), Mxi1 (MAX Interactor 1, Dimerization Protein), TET1 (Ten-eleven translocation methylcytosine dioxygenase 1).

Pour améliorer la spécificité de Cas9, deux nouvelles versions ont été créées, soit *High fidelity-1 SpCas9* (SpCas9-HF1) [144] et *enhanced specificity SpCas9* (eSpCas9) (1.1) [97]. La création de ces deux versions plus spécifiques se base sur l'observation d'une stabilisation de la Cas9 sur l'ADN génomique via le contact de

quelques-uns de ses acides aminés [145] SpCas9-HF1 est un quadruple variant (N497A/R661A/Q695A/Q926A) réduisant la liaison de Cas9 avec les groupements phosphates présents sur l'ADN cible. Cette version de Cas9 démontre un clivage minimal au site d'intérêt lors qu'un gRNA contenant des mésappariements est utilisé [144]. Le variant eSpCas9, quant à lui, possède trois mutations (K848A/K1003A/R1060A) qui induisent la neutralisation de résidus chargés réduisant ainsi la stabilisation de l'interaction entre la Cas9 et le brin non ciblé. Cette diminution de l'interaction favorise une réhybridation des deux brins complémentaires de l'ADN, nécessitant ainsi une forte liaison entre le gRNA et le brin ciblé pour permettre à Cas9 d'induire une cassure double brin. La présence d'un mésappariement entre la séquence du gRNA et l'ADN ciblé résulte en l'expulsion du gRNA et de Cas9 au profit de la réhybridation des deux brins d'ADN [97].

Le système Cas9 a aussi été modifié pour agir en paire, à l'image des ZFN et des TALEN. Premièrement, il est possible d'utiliser deux Cas9 nickases (Cas9n) comportant la mutation D10A inactivant le domaine RuvC [146]. L'utilisation de deux Cas9n créant deux cassures simple brin rapprochées sur une séquence d'ADN simule une cassure double brin [146]. Deuxièmement, un domaine FokI peut être fusionné à une dCas9 pour former des *RNA-guided FokI nucleases* (RFN) [147]. À l'image des ZFN et des TALEN, les RFN doivent agir en dimère ce qui augmente la spécificité. L'avantage des RFN versus les ZFN et les TALEN est que l'appariement à l'ADN se fait avec un gRNA, ce qui est beaucoup moins coûteux et plus facile à synthétiser que les ZFP ou les TALE.

I.2.3.4 Limites du système CRISPR

Une limite du système CRISPR est l'obligation d'utiliser un gRNA adjacent à une séquence PAM. Il est possible d'outrepasser cet inconvénient en utilisant un système CRISPR possédant un PAM différent. Tout comme pour les autres types de nucléases d'ingénierie, la possibilité de cassures double brin hors cible demeure un problème. Tel que discuté précédemment, les nucléotides 12 à 20 du gRNA sont plus permissifs au mésappariement ce qui peut mener à la reconnaissance de nombreux sites

hors cible. Des études poussées comparant de façon non biaisée les effets hors cibles des différents types de nucléases ainsi que de leur variant seront nécessaires.

I.3.0 L'édition de génome

L'édition de génome est l'ensemble des techniques permettant de modifier le génome d'organismes vivants, soit l'invalidation de gènes dans l'organisme, la correction génique, et l'ajout de gènes dans l'organisme.

La modification peut se faire *in vitro*, *in vivo* ou *ex vivo*. L'édition de génome *in vitro* est utile pour étudier les effets de variants génétiques sur le phénotype cellulaire [148]. L'édition *in vivo* permet de corriger directement dans un organisme vivant une mutation génique causant une maladie [149]. L'édition *ex vivo* consiste en la modification génique en dehors de l'organisme vivant, pour ensuite réintégrer les cellules génétiquement modifiées. Un exemple de traitement *ex vivo* est l'essai clinique pour des patients atteints du virus de l'immunodéficience humaine (VIH) [110]. Des cellules CD4+ ont été prélevées de patients, puis modifiées génétiquement en laboratoire à l'aide de ZFN pour inactiver un co-récepteur (CCR5) permettant l'entrée du virus. Après enrichissement du nombre de cellules modifiées, les patients ont été réinjectés avec leurs cellules modifiées rendues résistantes à l'infection.

I.3.1 L'utilisation d'un havre génomique sécuritaire

L'expression d'ADN inséré de façon aléatoire dans le génome est susceptible à l'action des promoteurs environnants et risque d'être exprimée de façon imprédictible [150, 151]. De plus, l'intégration d'ADN dans la région codante d'un gène important peut enlever une fonction nécessaire à la survie de la cellule par l'inactivation de ce gène, et causer la mort cellulaire [152]. Pour pallier à ce problème, l'intégration d'ADN est maintenant ciblée dans un *genome safe harbour* (GSH), afin de maintenir la régulation cellulaire normale qui est nécessaire au bon fonctionnement de la cellule [153].

Un GSH doit permettre à l'ADN exogène inséré d'être exprimé à des niveaux désirés quel que soit le type cellulaire. Il ne doit aussi pas affecter la régulation

transcriptionnelle de la cellule, que ce soit par la présence de la cassette dans une région importante, ou par la présence d'un nouveau promoteur dans la séquence d'ADN exogène [154]. Deux sites d'intégration dans le génome sont considérés sécuritaires par la communauté scientifique en général pour intégrer de l'ADN exogène et permettre son expression : i) le locus *Adeno-associated virus site 1* (AAVSI) [155], et ii) le site du récepteur C-C chimiokine de type 5 (CCR5).

Le locus *AAVSI* est situé au chromosome 19q13.42, dans le premier intron du gène encodant la sous-unité régulatrice 12C de la protéine phosphatase 1 (PPP1R12C) [156]. Le locus *AAVSI* a été découvert comme étant l'endroit dans le génome où le *Adeno-associated virus* (AAV) s'intègre dans le génome humain [157]. Les AAV ont la particularité de ne s'intégrer spécifiquement qu'au locus *AAVSI* chez l'humain, et résulte en la perte de l'expression de PPP1R12C qui n'a pas d'effet connu chez l'humain. L'intégration d'ADN exogène au site *AAVSI* résulte en une expression répandue dans plusieurs types cellulaires et l'insertion d'un promoteur viral favorisant une forte expression n'entraîne pas d'effet notable sur les gènes présents 200kb en amont ou en aval du site d'insertion [158].

Le locus CCR5 est localisé au chromosome 3p21.31. Le gène encodé est celui du co-récepteur principal du VIH [159]. L'insertion d'une cassette exogène avec un promoteur favorisant une forte expression augmente également l'expression de gène situé en périphérie du site d'intégration [158]. Toutefois, un patient humain porteur d'une mutation homozygote CCR5 $-/-$ est plus susceptible à l'infection par le virus du Nil [160]. Ces deux données expliquent pourquoi CCR5 est un moins bon GSH que *AAVSI*.

I.3.2 Techniques utilisées pour la modification génique

Les nucléases d'ingénierie du génome sont devenues la norme pour effectuer des modifications géniques. Nous allons quand même regarder les techniques classiques qui étaient utilisées avant pour insérer des modifications génétiques. Certaines ont encore une utilité lorsqu'utilisées de pair avec les nucléases d'ingénierie du génome, mais la faible spécificité des techniques classiques comme le système

Cre/LoxP ou les ADN transposases les rendent désuètes pour apporter des modifications génétiques de façons précises.

I.3.2.1 Les transposases

I.3.2.1.1 La transposase *Sleeping Beauty*

La transposase *Sleeping Beauty* (SB) est issue de transposases de la superfamille TC1/mariner qui sont présentes sous formes inactives dans les poissons dues à l'accumulation de mutation [161]. Les transposases de la famille TC1/mariner sont encodées par un gène contenant la transposase flanquée de répétitions inversées (IR) et contenant le motif auquel la transposase peut se lier [162]. Les répétitions inversées mesurent entre 200 et 250 pb et possèdent deux motifs de liaisons de la transposase, aussi appelés répétitions directes (DR).

I.3.2.1.1.1 Mécanisme de transposition d'ADN par *Sleeping Beauty*

Lors de la transposition d'ADN par *Sleeping Beauty* (SB), la transposase commence par se lier à ses répétitions directes sur le transposon. Ensuite, les deux extrémités du transposon sont rapprochées et SB excise le transposon, puis le déplace jusqu'à un locus d'intégration contenant un dinucléotide TA [163]. L'excision du transposon de l'ADN donneur laisse une empreinte de 3 paires de bases et l'insertion du transposon à un dinucléotide TA résulte en la duplication de ce dinucléotide de chaque côté du transposon intégré [161, 164].

I.3.2.1.1.2 Ingénierie du génome avec la transposase *Sleeping Beauty*

Plusieurs propriétés de SB en font un excellent choix pour l'ingénierie de génomes : i) elle peut insérer un transposon ou une séquence exogène dans l'ADN, ii) elle peut insérer son transposon de façon à obtenir une expression à long terme, iii) les techniques de production d'une SB sont peu dispendieuses et iv) l'expression de SB n'est pas immunogène pour l'hôte [163].

Pour utiliser SB afin de modifier le génome d'un organisme, deux vecteurs sont nécessaires. Premièrement, il faut un vecteur encodant le gène de la transposase SB, puis il faut un deuxième vecteur contenant la séquence à être insérée dans le génome flanquée par deux DR reconnues par SB [165, 166].

I.3.2.1.1.3 Limites

L'utilisation de SB comporte certaines limites. La plus importante est l'insertion aléatoire de matériel génétique. La transposase pourrait insérer le fragment dans un suppresseur de tumeur et l'invalider, causant ainsi l'apparition de cancers. Aussi, l'intégration du transposon doit se faire dans un dinucléotide TA et le matériel génétique inséré sera nécessairement entouré par une séquence TA de chaque côté [164]. La modification du génome par transposase ne peut donc pas être utilisée pour créer des SNP ou pour corriger des mutations. Il ne s'agit pas d'un outil de modification du génome qui a des attraits thérapeutiques chez l'humain.

I.3.2.1.2 La transposase piggyBac

La transposase piggyBac (PB) a été isolée de la fausse-arpenteuse du chou (*Trichoplusia ni*) [167]. Comme tous les systèmes de transposon, PB comporte un transposon et une transposase. La transposase PB facilite l'intégration du transposon aux sites 5'-TTAA-3' dans le génome [167]. Le transposon naturel de PB est flanqué de répétitions inverses de terminaison (ITR) de 13pb et d'une seconde paire d'ITR de 19pb chacun. Ces ITR permettent la transposition de la cassette à l'intérieur de ces répétitions dans les insectes seulement [168]. Pour pouvoir transposer une cassette dans des organismes supérieurs, des séquences de terminaison répétées en 5' de 313pb et de 235pb en 3' sont nécessaires [169]. Un des avantages de PB sur les autres ADN transposases est qu'elle ne laisse pas de traces lors de l'excision de son transposon [170].

I.3.2.1.2.1 Mécanisme de transposition d'ADN par piggyBac

Pour exciser son transposon, PB se lie au ITR et coupe l'ADN en laissant des extrémités protubérantes en clivant à l'extrémité du transposon en 3' et en clivant l'extrémité du motif TTAA entourant le transposon en 5' [171]. Ceci crée des extrémités TTAA de chaque côté de la cassure double brin qui pourront être rejointes pour réparer l'ADN. Une fois excisé, le transposon adoptera une structure en épingle grâce aux interactions entre ses extrémités TTAA [172]. Sur la séquence d'ADN où le transposon est prélevé, les deux extrémités protubérantes compatibles (TTAA) sont jointes par PB, sans avoir besoin de la machinerie cellulaire [170]. La transposase peut ensuite procéder à l'insertion du transposon dans un site TTAA situé dans le génome [173]. Le système PB semble favoriser l'intégration de son transposon dans un gène transcrit [174].

I.3.2.1.2.2 Ingénierie du génome avec la transposase piggyBac

La modification du génome par PB se fait avec deux vecteurs. Le premier comporte une cassette contenant la séquence d'intérêt flanquée des deux ITR reconnus par PB. Le deuxième vecteur sert à exprimer le gène de la transposase PB [170]. La transposase PB peut aussi être fournie par transduction d'un ARNm [175]. Cette dernière possède une grande capacité de transposition. En effet, des séquences de plus de 200 kb ont été transposées [176]. Cette technologie peut être utilisée pour effectuer de la transgénèse, de la modification génique et du *gene-trap* [177].

En combinant l'utilisation des nucléases d'ingénierie et de la transposase PB, il est possible de modifier l'ADN sans laisser de traces. En effet, il suffit de produire une cassure double brin au locus d'intérêt avec des nucléases d'ingénierie et de fournir un donneur de réparation comprenant le gène à modifier ainsi qu'une cassette de sélection (ex : puromycine) flanquée par les ITR reconnus par PB. Suite à la transfection, il suffit de sélectionner les cellules ayant intégré le donneur, puis de transférer la transposase PB pour qu'elle procède à l'excision de la cassette de sélection. Cette technique possède le désavantage de nécessiter deux transfections distinctes ce qui peut être une limite importante pour certains types cellulaires. Bien que le fragment excisé ne puisse

pas être réintégré par la PB, il demeure une possibilité que ce dernier soit intégré de façon aléatoire dans le génome. De plus, les ADN transposases sont moins efficaces dans certains types cellulaires pour exciser leur transposon, particulièrement dans les cellules souches embryonnaires [178].

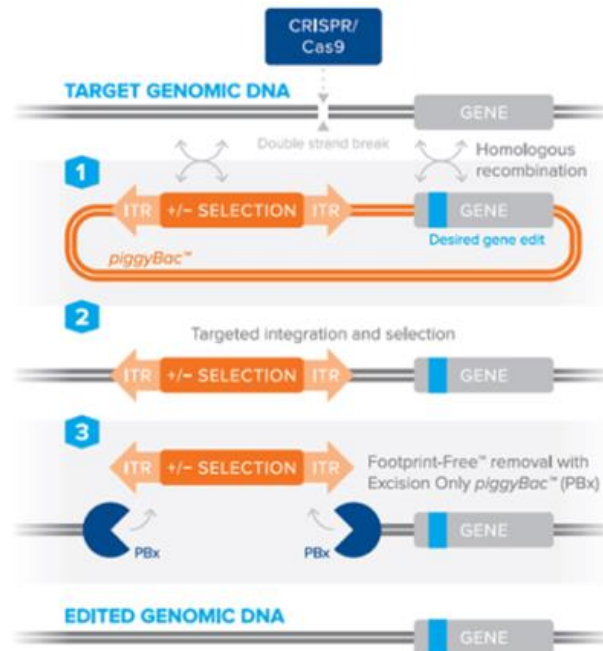


Figure 14 : Représentation schématique de l'utilisation combinée des nucléases d'ingénierie et de la transposase *piggyBac*. Après insertion par recombinaison homologue d'un donneur d'ADN comportant la modification génique souhaitée (marqué en bleu sur l'image) et une cassette de sélection entouré de répétitions inverses de terminaison. Une transposase *piggyBac* à excision seulement (*PBx*) permet ensuite de retirer la cassette de sélection. Image tirée d'Addgene [179].

I.3.2.2 Le système Cre/LoxP

La recombinase Cre est une protéine de 38 kDa encodée par le bactériophage P1 [180]. La Cre reconnaît une séquence de 34 pb dans le bactériophage P1 nommée LoxP et peut recombiner des paires de séquence LoxP [181]. Les sites LoxP consistent en deux séquences palindromiques de 13 pb, avec une séquence centrale de 8 pb :

ATGTATGC [182]. Le système Cre/LoxP est très utilisé pour générer des modèles animaux transgéniques.

I.3.2.2.1 Mécanisme de recombinaison du système Cre/LoxP

Lors de la recombinaison, deux sous-unités de la Cre se lient à chaque séquence LoxP en coopération sur les parties palindromiques. C'est-à-dire qu'une fois que l'une est présente, la deuxième a plus de facilité à lier l'ADN [183]. Lorsqu'un dimère de Cre se trouve sur chacun des sites LoxP, un tétramère est formé via la reconnaissance des hélices A et E [184]. Les sites catalytiques sont maintenant positionnés pour cliver l'ADN. Pour effectuer la recombinaison, le Cre n'a pas besoin de cofacteur ou d'énergie sous forme d'ATP [185]. L'ADN se trouvant entre deux sites LoxP dans la même orientation sera excisé tandis que celui se trouvant entre deux sites LoxP dans des orientations opposées sera inversé [183]. La résultante de la recombinaison est un site LoxP qui demeure dans l'ADN.

I.3.3 Modification du génome avec les nucléases d'ingénierie

Les nucléases d'ingénierie ont apporté un nouveau souffle à l'édition de génome. Contrairement aux systèmes présentés aux sections précédentes, ces nucléases sont construites de façon à cibler une séquence d'intérêt précise, puis à y créer une DSB. La modification de génome se fait en utilisant les voies de réparation cellulaire suite à une DSB. L'avantage par rapport aux techniques antérieures est la possibilité de décider l'endroit dans le génome où la modification génomique est insérée.

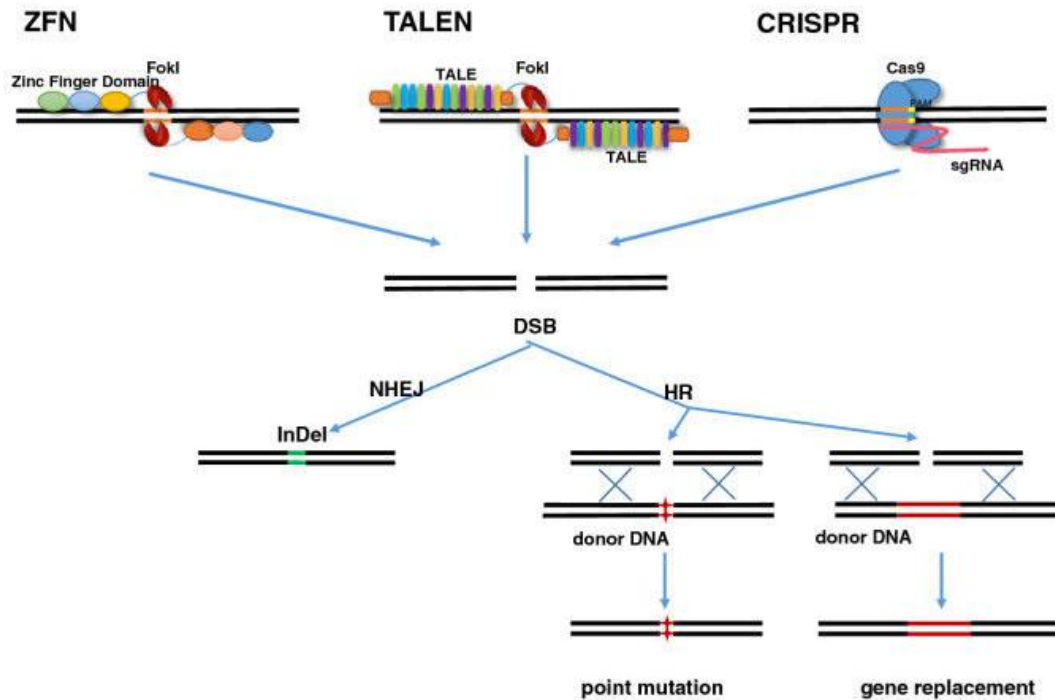


Figure 15 : Les ZFN, TALEN et CRISPR/Cas9 peuvent être utilisés afin d’effectuer des modifications précises dans le génome par l’induction de cassure double-brin à l’ADN qui peuvent être réparé par recombinaison homologue s’il y a présence d’un donneur d’ADN, sinon la cassure sera réparée par NHEJ. Tirée de Chen et al., 2014 [143]. Abréviations : DSB (double strand break), NHEJ (non homologous end-joining), HR (homologous recombination).

I.3.3.1 Invalidation de gène

L’invalidation de gène par les nucléases d’ingénieries se fait en utilisant la voie de réparation d’une cassure double brin par NHEJ. En l’absence de guide de réparation, la cassure double brin est réparé par NHEJ, causant la formation d’indels au site de réparation [106]. En inactivant un gène avec précision, les fonctions d’une protéine peuvent être déterminées [186]. De plus, des modèles animaux d’invalidation génique peuvent être créés en utilisant cette technique. Que ce soit pour étudier la fonction d’un gène d’un point de vue de l’organisme ou pour la création d’un modèle de maladie, il est possible de cibler de façon précise l’endroit où une cassure double brin sera créée [187].

I.3.3.2 Correction génique

La correction génique consiste à modifier une courte séquence, en employant la voie de réparation HDR. Donc, il faut fournir un guide de réparation (ou donneur d'ADN) incluant la correction à insérer dans le gène. Pour être utilisé par la machinerie cellulaire, il faut que le donneur d'ADN ait à chacune de ses extrémités des séquences homologues de l'ADN situé en amont et en aval à la cassure double brin. Ensuite, le donneur d'ADN doit comporter la séquence codante corrigée, qui doit également faire partie du site de reconnaissance à l'ADN de la nucléase utilisée pour cliver l'ADN. L'insertion de la correction génique résulte donc en la perte du motif reconnu par la nucléase. Ensuite, il faut génotyper les cellules et déterminer si la correction est bel et bien présente.

I.3.3.3 Ajout de matériel génétique

L'ajout de matériel génétique profite du mécanisme de HDR pour insérer une séquence d'ADN exogène suite à une cassure double brin. En fournissant un donneur d'ADN comportant à chacune de ses extrémités des séquences homologues de l'ADN situé en amont et en aval à la cassure double brin, la séquence désirée peut être insérée à un endroit précis dans le génome (voir figure 15).

I.3.3.4 Essais cliniques en cours avec les nucléases d'ingénierie

Les ZFN sont les premières nucléases d'ingénierie à avoir été utilisées pour un essai clinique chez l'humain [110]. Les porteurs homozygotes de la mutation CCR5-delta32 ont une résistance accrue à l'infection par le VIH-1 [188]. L'essai clinique consistait à prélever des lymphocytes de patients, à isoler les lymphocytes CD4+ pour ensuite les modifier génétiquement avec une paire de ZFN ciblant le gène de CCR5, afin de désactiver ce dernier [189].

Les TALEN ont été utilisées dans la création de cellules T modifiées pour combattre la leucémie [190]. Ces cellules modifiées génétiquement ont permis la rémission de la leucémie pour deux enfants en Angleterre [191].

CRISPR a été utilisé par une équipe de recherche en Chine pour modifier *ex vivo* les cellules d'un patient avec un cancer du poumon [192]. Après avoir enrichi les cellules modifiées génétiquement, ces dernières ont été injectées au patient. Un premier essai clinique a été accepté aux États-Unis pour l'utilisation de CRISPR afin de modifier les cellules du système immunitaire afin de les rendre plus agressives contre les cellules cancéreuses [193]. Ces essais cliniques impliquent tous une manipulation *ex vivo* des cellules, où les cellules sont génétiquement modifiées, puis enrichies pour en avoir une quantité suffisante à injecter au patient.

I.3.4 Limites des techniques présentement utilisées

I.3.4.1 Insertion aléatoire

Parmi les techniques mentionnées ci-haut, l'utilisation de SB, et de PB utilise une intégration aléatoire. Bien que SB et PB puissent être fusionnées à un domaine de liaison à l'ADN pour favoriser l'insertion de leur transposon à un site précis, rien ne les empêche d'insérer de l'ADN aléatoirement à d'autres régions génomiques [194]. L'intégration d'ADN de façon aléatoire dans le génome n'est pas souhaitable, car cela peut déréguler des fonctions cellulaires vitales (activer un oncogène, inactiver un suppresseur de tumeurs) ou simplement être inséré dans une partie du génome peu transcrite [150, 152].

I.3.4.2 Traces des modifications géniques

La correction génique dans une région intronique d'un gène doit se faire sans laisser de traces. Le système Cre/LoxP laisse une séquence LoxP après la modification génique, alors il ne peut pas être utilisé dans une séquence intronique si l'intégrité de la séquence intronique est importante. La transposase de SB laisse également des traces lors du retrait de son transposon.

I.3.4.3 Limites de l'édition génique dans un GSH

Les gènes insérés dans un GSH n'ont pas la régulation native de la copie endogène. De plus, pour des maladies alléliques dominantes, l'intégration d'une copie saine du gène ne règle pas le problème [195].

I.3.4.4 Limites de la modification génique par HDR

Premièrement, la fréquence de réparation d'une cassure double brin par HDR est très faible comparativement au NHEJ [196]. La façon la plus simple de vérifier que les cellules ont incorporé le donneur d'ADN par HDR suite à une cassure double brin est d'inclure un gène de résistance à un antibiotique ou une étiquette d'identification comme la protéine fluorescente verte (GFP), ou l'étiquette 3xFLAG [197, 198]. Toutefois, l'insertion de cette séquence supplémentaire d'ADN exogène n'est pas souhaitable lorsque l'on cherche à créer une lignée modèle, ou lorsque l'on essaie de corriger un gène contenant une mutation.

Deuxièmement, la réparation par HDR nécessite la présence d'un donneur d'ADN que la cellule peut utiliser comme guide de réparation. Certains gènes sont trop longs pour pouvoir être livré en entier dans les cellules. Par exemple, dans la correction génique de la dystrophie musculaire de Duchenne, le gène à corriger est celui de la dystrophine qui mesure 2,4Mb [199]. On ne peut donc pas livrer une copie du gène corrigé dans les cellules. Pour contourner cette limite, une partie du gène à corriger peut être livrée.

PROBLÉMATIQUE ET OBJECTIFS DES TRAVAUX

Problématique

La modification du génome est maintenant accessible à grande échelle. Une large gamme d'outils est présentement disponible, mais plusieurs lacunes demeurent, principalement au point de vue de la spécificité et des modifications géniques sans traces.

Dans ce contexte, le présent projet de maîtrise visait à :

- 1) Développer des outils d'édition génique faciles à utiliser et qui permettent de donner accès aux techniques d'éditions de génome à un plus grand nombre de laboratoires de recherche.

Objectifs des travaux

- 1) Concevoir et valider un outil permettant la création d'une lignée cellulaire génétiquement modifiée sans laisser de traces.
- 2) Développer une technique permettant de cliver l'ADN avec plus de spécificité

Objectifs spécifiques

- 1.a) Élaborer une cassette de sélection auto-excisable.
- 1.b) Valider l'efficacité de la cassette de sélection auto-excisable.
- 2.a) Déterminer si l'utilisation de doubles nickases peut induire une cassure double brin.
- 2.b) Évaluer l'impact de la distance entre deux cassures simple brin permettant une réparation par NHEJ.

CHAPITRE 2 : DÉVELOPPEMENT ET VALIDATION D'UNE CASSETTE AUTO-EXCISABLE POUR LA CRÉATION DE MODIFICATIONS GÉNIQUES SANS TRACES.

II.2.1 Matériel et méthodes

Clonage

Digestions, ligations, transformations

Les digestions des différents fragments ont été réalisées avec les enzymes décrites ci-bas. Pour chacune des digestions, le tampon suggéré par le manufacturier a été utilisé pendant l'incubation. Les incubations ont été réalisées à 37°C pendant 1 heure. Suite à la digestion, les fragments ont été séparés par électrophorèse sur gel d'agarose 0.8% TAE 1X pour 30-60 minutes. Les bandes d'intérêt ont été excisées du gel et purifiées sur colonnes (Biobasic).

Les réactions de ligation ont été réalisées avec la T4 DNA ligase pendant une heure à température pièce selon les instructions du manufacturier (New England Biolabs). Les réactions de *Gibson cloning* ont été réalisées selon les instructions du manufacturier. Brièvement, le fragment à insérer par Gibson cloning a été amplifié par PCR et le vecteur a été linéarisé. Des aliquotes de 300ng d'insert et 100ng de vecteur ont été incubés 30 minutes à 50°C avec les réactifs du Gibson Assembly Master Mix (NEB #e2611). L'ADN circulaire ainsi formé a ensuite été transformé dans des bactéries compétentes. Deux µL de ligation ou 2 µL de réaction de *Gibson cloning* dilué ¼ dans de l'eau distillée ont été transformés dans des bactéries DH5alpha compétentes. L'ADN a été ajouté à 50 µL de bactéries compétentes, puis incubé 30 minutes sur glace. Les bactéries ont été soumises à un choc thermique à 42°C pendant 30 secondes, puis retournées sur glace pour 5 minutes. Neuf cent cinquante µL de milieu LB (Sigma-Aldrich) ont été ajoutés et les bactéries ont été incubées 1 heure à 37°C avec agitation à 220 rpm. Cent µL de bactéries ont étéensemencées sur une gélose de milieu LB + ampicilline (BioShop). Ensuite, le reste des bactéries ont été centrifugées à 2000 rpm pendant 5 minutes, puis 800 µL du surnageant a été retiré, le culot a été resuspendu dans les 100 µL restant et étalées sur une gélose de milieu LB contenant l'agent de sélection approprié.

Clonage de la cassette auto-excisable AAVS1_SSA_eGFP

Le promoteur hPGK1 a d'abord été amplifié à partir du vecteur

AAVS1_hPGK1_Puro_FS_v2 en utilisant des amorces PCR comprenant les sites de restriction HindIII (Tableau 1, NEB). Le plasmide AAVS1_hPGK1_Puro_FS_v2 a par la suite été digéré HindIII de façon à retirer l'ensemble des cassettes d'expression se trouvant entre les bras d'homologie, puis joint avec le fragment PCR hPGK1. À ce vecteur, un gBlock contenant une partie de la séquence de la protéine fluorescente eGFP, un site de liaison pour la nucléase thermosensible et le promoteur SFFV ont été ajoutés en sous-clonant avec les enzymes de restriction SbfI et NcoI (NEB). La résultante de ces clonages est le plasmide AAVS1_hPGK1_eGF_BS_SFFV qui comprend des bras d'homologie à AAVS1, un promoteur hPGK1, une partie de la séquence codant pour eGFP, un site de liaison pour la nucléase thermosensible et le promoteur SFFV. Le promoteur hPGK1 permettra l'expression de la protéine eGFP suite à la réparation par SSA tandis que le promoteur SFFV permettra l'expression de la nucléase thermosensible.

En parallèle, la séquence codant pour DEST4 (d4), un peptide de type PEST qui permet de diminuer la demi-vie des protéines auxquelles il est fusionné, a été amplifiée par PCR à partir du plasmide CMV-DsRed-IRES-d4EGFP (Table 1, Addgene #41943). Le fragment obtenu a été cloné dans le plasmide 8267_N2A_ts1_2A_Puro_delta_tk (Sangamo Therapeutics) par *Gibson cloning* (NEB). La résultante est l'ajout du domaine d4 au gène de la thymidine kinase de façon à diminuer sa stabilité.

Le fragment Puro_delta_tkd4 a par la suite été sous-cloné dans le plasmide pGF_T2A_BS_ts2_F2A_Puro_BGH_BS_FP (Sangamo Therapeutics) en utilisant les enzymes de restriction BsiwI et AscI (NEB). Le plasmide résultant, pGF_T2A_BS_ts2_F2A_Puro_delta_tkd4_BS_FP, comprend deux sites de liaison (BS) à la nucléase thermosensible (ts2), les gènes de sélection puromycine et thymidine kinase, le tout flanqué à chaque extrémité d'une partie de la séquence de la protéine fluorescente eGFP.

La nucléase thermosensible, les gènes de résistances, le site de liaison pour la nucléase ainsi que la fin de la séquence de eGFP ont été sous-clonés dans le vecteur AAVS1_hPGK1_eGF_BS_SFFV en utilisant les enzymes de restriction Afl2 et Acc65i (NEB). La résultante de ce clonage, AAVS1_hPGK1_eGF_BS_SFFV_

ts2_F2A_Puro_delta_tkd4_BS_FP, est la version finale de la cassette eGFP pour l'intégration au locus *AAVS1*.

Clonage de la cassette auto-excisable EZH2_SSA_TAG_EZH2

Les étiquettes de purification 3xFLAG_2xStrep et le bras d'homologie gauche du locus *EZH2* ont été amplifiés par PCR (Tableau 2) sur le vecteur *EZH2_N_FS_v2_Donor* (Addgene # 79901). Le fragment obtenu a été sous cloné avec les enzymes de restriction *NdeI* et *SbfI* (NEB) dans le vecteur *AAVS1_hPGK1_eGF_BS_SFFV_ts2_F2A_Puro_delta_tkd4_BS_FP* préparé ci-haut. Le vecteur résultant de ce clonage, *EZH2_IHA_3xFlag_2xStrep_BS_SFFV_ts2_F2A_Puro_delta_tkd4_BS_FP* comprend la cassette de sélection, le bras gauche d'homologie à *EZH2* ainsi que les étiquettes de purification 3xFLAG_2xStrep. Considérant que la cassette de sélection au locus *EZH2* vise l'ajout d'une étiquette de sélection pour la purification de la protéine dans son contexte natif, le promoteur *PGK1* n'a pas été inclus dans la cassette, puisque l'expression d'*EZH2* étiqueté sera sous le contrôle du promoteur *EZH2* natif. En tenant compte la petite taille des étiquettes de purification 3xFLAG_2xStrep, une stratégie utilisant une duplication des étiquettes pour favoriser la réparation par SSA a été utilisée, versus une stratégie où seulement une partie de la séquence est répétée.

Dans un deuxième temps, un gBlock comprenant une partie de la séquence du gène de la thymidine kinase, un site de liaison pour la nucléase thermosensible et les étiquettes de purification 3xFLAG_2xStrep a été ajouté au vecteur par clonage en utilisant les enzymes de restrictions *BsrGI* et *EcoRI* (NEB). Le vecteur résultant, *EZH2_IHA_3xFlag_2xStrep_BS_SFFV_ts2_F2A_Puro_delta_tk_BS_3xFlag_2xStrep*, comprend la deuxième copie des étiquettes de purification.

Finalement, le bras d'homologie gauche du locus *EZH2* a été amplifié sur le vecteur *EZH2_N_FS_v2_Donor* (Tableau 2, Addgene # 79901) et cloné par *Gibson cloning* dans le vecteur issu de l'étape précédente. La résultante de ce clonage,

EZH2_3xFlag_2xStrep_BS_SFFV_ts2_F2A_Puro_delta_tk_BS_3xFlag_2xStrep, est la version finale de la cassette Tag pour l'intégration au locus *EZH2*.

ZFN et TALEN

De façon à permettre l'intégration de la cassette, le locus *AAVSI* a été ciblé en utilisant une ZFN décrite précédemment {Hockemeyer, 2009 #4418}. Le locus *EZH2* a quant à lui été ciblé avec un TALEN (Addgene #36774 et #36776) décrit précédemment [86].

Lignées cellulaires, transfection et sélection

Les cellules K562 ont été obtenues de l'ATCC (CCL-243) et maintenues à 37 °C sous 5% CO₂ dans du milieu RPMI supplémenté avec 10% de sérum foetal bovin (FBS), 1% de pénicilline-streptomycine et du GlutaMAX. Les lignées cellulaires ont été testées pour l'absence de mycoplasme.

Les cellules (1×10^6 /transfection) ont été transfectées en utilisant l'Amaya 4D-Nucleofector (Lonza), selon les recommandations du manufacturier. Les cellules ont été transfectées avec 4 µg de donneur, 1 µg de ZFN ciblant le locus *AAVSI* ou 2.5 µg de TALEN ciblant le locus *EZH2*.

Trois jours post-transfection, les cellules ont été mises en sélection avec la puromycine (0.5 ng/mL), et ce pendant 10 jours. La sélection a été effectuée en milieu liquide pour obtenir une population hétérogène de cellules résistantes ou en milieu semi-solide de façon à obtenir des populations clonales. Les cellules résistantes à la puromycine ont par la suite été induites à 30 °C pendant 48 heures pour permettre l'activation de la nucléase thermosensible, puis remises 7 jours à 37 °C pour leur permettre de reprendre une croissance normale. Les cellules ayant excisé la cassette correctement ont été contre-sélectionnées avec le ganciclovir (1000 ng/mL) pendant 10 jours. La sélection au ganciclovir a été réalisée en milieu liquide pour obtenir une population hétérogène de cellules résistantes ou en milieu semi-solide de façon

à obtenir des populations clonales. Les cellules contre-sélectionnées au ganciclovir ont été utilisées pour les analyses subséquentes.

PCR out-out

De façon à analyser le type de réparation étant survenu suite à l'excision de la cassette, un PCR de type *out-out* a été réalisé. L'ADN génomique de 2.5×10^5 cellules a été extrait en utilisant 250 μ l de *QuickExtract DNA extraction solution* (Illumina), selon les instructions du fabricant. L'ADN génomique extrait a été soumis à 30 cycles de PCR pour amplifier le locus *AAVS1* et à 35 cycles de PCR pour amplifier le locus *EZH2* en utilisant les amorces décrites au tableau 4. Les amplicons ont par la suite été déposés sur gel d'agarose 1 % et soumis à une électrophorèse dans du tampon TAE 1x.

Analyses de cytométrie de flux

La fréquence des cellules exprimant EGFP a été déterminée avec un cytomètre BD SORP II, avec un laser à 488nm, un filtre 525/25 et un bec (*nozzle*) de 70 μ m. Le trieur BD SORP FACSaria II a été utilisé afin de trier et récolter les cellules.

Immunobuvardage

Des extraits totaux ont été préparés à partir de $\sim 1 \times 10^6$ cellules par addition de 2x Laemmli buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycérol, 0.004% bleu de bromophénol, 0.125M Tris HCl, ajuster le pH à 6.8) suivi d'une incubation à 95 °C pendant 10 minutes. Pour chaque échantillon, un volume de 10 μ L d'extraits totaux ont été déposés sur un gel d'acrylamide 8 % dans du tampon Tris-Glycine 1x. Les protéines ont été transférées sur une membrane de nitrocellulose, bloquées avec 5% de poudre de lait diluée dans du PBS-Tween 0.01% et incubés avec leurs anticorps primaires pour toute la nuit à 4 °C. Les anticorps primaires suivants ont été utilisés : Anti-FLAG M2 (Sigma, A8592) et anti- α -tubuline (DM1A, sc-32293). Les anticorps secondaires (anti-mouse) dilués 1 : 10000 dans du PBS-

Tween 0.01% ont été incubés 1 heure à température pièce. Les immunobuvardages ont été révélés par chimiluminescence (GE) à l'aide d'un Chemidoc (Biorad).

Topo cloning et séquençage

De façon à déterminer la fréquence des allèles réparés par NHEJ et SSA, les produits des PCR *out-out* ont été TOPO clonés (Zero blunt TOPO PCR cloning kit, Invitrogen), puis transformés dans des bactéries DH5alpha compétentes. Un total de 22 colonies bactériennes a été repiquées. Les plasmides ont été purifiés sur colonne (Biobasic), puis 9 ont été séquencés (Plateforme de séquençage et de génotypage du Centre de Recherche du CHUL) en utilisant les amorces M13 sens et M13 anti-sens.

Tableau 1 : Séquence des oligonucléotides utilisés pour l'assemblage de la cassette auto-excisable eGFP.

| Cible | Amorce sens | Amorce anti-sens |
|---|---|--|
| <i>Gibson cloning</i> sur DEST4 (d4) | GGAGATGGGGGAGGCTAACAGCCAT GGCTTCCCGCCG | GCGGGTTTAAACGGGCCCGGCGCGCCT CACTACACATTGATCCTAGC |
| hPGK1 | /5Phos/CTTCGAACCACGGGGTTGGGG TTGCGC | /5Phos/CTTAAGTGCATGGTACCTGATCC TGCAGGCTAGGGTTCGACCCTGGGGAGA GAGGTCGGTG |
| SFFV | AAACCTCGAGCACCGGTGTAACGCCA TTTTGC | GGCGGGTACCTTCCTCTTCTTCTTGGGT CCCATGGTGGCCGGGCGACTCAGTCTG TCG |

Tableau 2 : Séquence des oligonucléotides utilisés pour l'assemblage de la cassette auto-excisable FLAG_EZH2.

| Cible | Amorce sens | Amorce anti-sens |
|---------------------------------|--|--|
| Donneur_EZH2 _N_FS (étape a) | CAGGGTTATTGTCTCATGAGCGG | AGCGCTCCTGCAGGGGCGCCCTTTTCG AACTGGG |
| Donneur_EZH2 _N_FS (étape c) | CCAGTTCGAAAAGGGCGCCGGCCAG ACTGGGAAG | GCCGATTCATTAATGCAGAATTCTCGA GTCACAATGACTTTAGAAACA |

Tableau 3 : Séquence des gBlock utilisés pour l'assemblage des cassettes auto-excisable.

| Nom | Séquence |
|----------|--|
| SSA_GFP | cctgcaggagcgccttttgcagtttatcataaactgcaaagagatctgttaaacctcgcagcaccgggtaacgccatttt gcaaggcatggaaaaataccaaaccaagaatagagaagttcagatcaaggcgggtacatgaaatagctaaccgttggg ccaaacaggatatctgcggtgagcagtttcggccccggccggggccaagaacagatggtcaccgcagtttcggcccc ggccccgaggccaagaacagatggtccccagatatggcccaaccctcagcagtttctaagaccatcagatgtttccagg ctcccccaaggacctgaaatgacctgcgccttattgaattaaccaatcagcctgcttctcgttctgttcgcgcgcttctgc tccccgagctctataaaagagctcacaaccctcactcggcgcgccagtcctccgacagactgagtcgccggccaccat gggacccaagaagaagaggaag |
| SSA_EZH2 | ccaacggcgacctgtacaacgtgtttgctgggcttgacgtcttggccaaacgcctccgctccatgcacgtctttatcc tggattacgaccaatcgcccgcggctgcgggacgcctctgcaactacctccgggatggtccagaccacgtca ccacccccggctccataccgacgatctgcacctggcgcgcacgtttcccgggagatggggaggtaaacgctttg cagtttatcataaactgcaaagactagtggcggcggagagggcagaggaagtcttctaactgcggtgacgtggagg agaatcccggcctggatccgccggcgactacaaggaccacgacggcgattataaggatcacgacatcgactacaaa gacgacgatgacaaggcggcagcagcgcctgtcccaccctcagtttgagaaggcggaggctctggcggcggaa gaggagatctgcttgagccacccccagttcgaaggcggcgaattctgcattaatgaatcgccaacgcgcggg gagagggcgggttgctattgggcgctcttccgcttctcgtcactgactcgtcgcgtcggctgttcggctgcggc cggtatcagctcactcaaggcggtaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaa aaggccagcaaaaggccaggaaccgtaaaaaggcgcgttgctggcgttttccataggtccgccccctgacgagc atcacaaaaatcgacgctcaagtcaaggtggcgaaccggacaggactataaagataaccaggcgtttccccctggaa gctcctcgtgcgctctcctgttccgacctgcccgttaccggatacctgtccgctttctccctcggga |

Tableau 4 : Séquence des oligonucléotides utilisés pour les PCR de génotypage et de séquençage.

| Cible | Amorce sens | Amorce anti-sens |
|----------------------|----------------------|-----------------------|
| <i>AAVS1 out-out</i> | CGGAACTCTGCCCTCTAACG | CTGGGATACCCCGAAGAGTG |
| <i>EZH2 out-out</i> | GTGGCACAAGAGGCCAAAAT | CGATTGCCATCCTTTCTTTG |
| M13 | GTAAAACGACGGCCAGT | GCGGATAACAATTTACACAGG |

II.2.2 Résultats

II.2.2.1 Élaboration d'une stratégie pour créer une cassette auto-excisable

L'étape initiale était de déterminer la meilleure stratégie pour créer une cassette auto-excisable. En plus de la séquence d'intérêt à insérer, cette dernière devait contenir un gène de résistance pour permettre la sélection des cellules ayant incorporé la cassette. Il devait aussi être possible d'exciser la cassette de façon contrôlée ainsi que de contre-sélectionner pour isoler les cellules ayant excisé correctement la cassette. Pour ce faire, nous

avons décidé d'utiliser la voie de réparation SSA. Lors de la réparation d'une DSB par SSA, le brin d'ADN en 5' de chaque côté de la cassure est digéré pour former un ADN en surplomb 3', puis les séquences identiques qui étaient de chaque côté sont appariées, résultant en l'expulsion du matériel génétique situé entre les deux séquences répétées.

La stratégie finale (Figure 16) comprend une cassette qui en plus des bras d'homologie inclut un gène de résistance à la puromycine, une ZFN thermosensible ainsi que le gène suicide de la thymidine kinase, qui entraîne la mort des cellules lorsque celles-ci sont traitées avec le ganciclovir. Ces éléments sont flanqués par deux sites de liaisons de la nucléase thermosensible, puis en 5' par le début de la séquence à insérer et en 3' par la fin de cette dernière. Il est important de mentionner que, pour qu'il y ait réparation par SSA, la fin de la séquence en 5' doit être répétée dans la partie retrouvée en 3' (Figure 16).

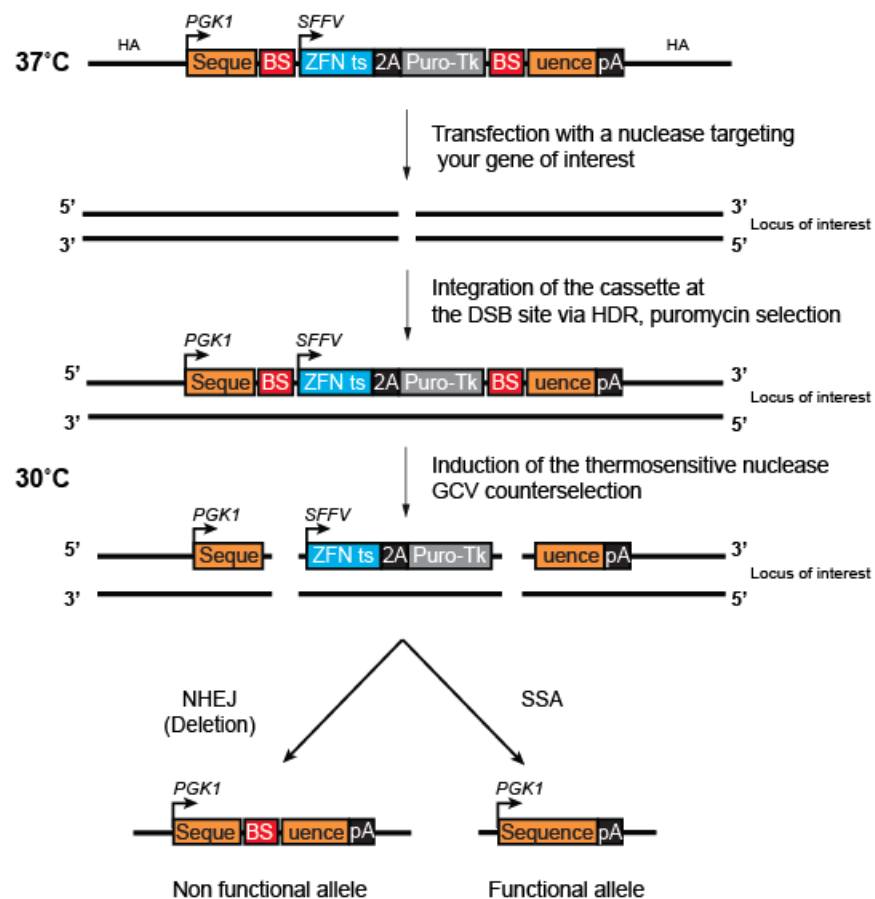


Figure 16: Représentation schématique de l'insertion d'une cassette d'expression exprimant une séquence d'intérêt contenant une cassette de sélection excisable. PGK1 et SFFV sont les

promoteurs du gène *Phosphoglycerate Kinase 1* et du *Spleen focus-forming virus*. BS signifie site de liaison pour la ZFN-ts présente dans la cassette. La séquence 2A est une séquence peptidique permettant l'expression de deux protéines à partir d'un seul ARNm. Puro-TK code pour la protéine chimérique procurant une résistance à la puromycine et sensibilisant la cellule au ganciclovir. La séquence pA est une séquence de poly-adénylation pour aider la production de la séquence d'intérêt une fois le SSA effectué. À noter que des répétitions de la séquence finale sont présentes de chaque côté de la cassette de sélection auto-excisable (représentées par les répétitions des lettres "ue" de chaque côté).

L'utilisation de la cassette est simple et débute par la transfection simultanée d'une nucléase ciblant le locus d'intérêt et de la cassette. Les cellules ayant intégré la cassette au locus d'intérêt sont sélectionnées pendant 10 jours en utilisant la puromycine, puis transférées pendant 48 heures à 30 °C pour permettre l'activation de la nucléase thermosensible (Figure 17). Une fois activée, la nucléase clive la cassette entre les séquences répétées, ce qui permet la réparation par SSA et l'expression de la protéine exogène. Les cellules ayant excisé la cassette sont par la suite remises à 37 °C, puis contre sélectionnées en utilisant le ganciclovir qui tue les cellules possédant encore le gène de la thymidine kinase présent dans la cassette. Il en résulte une population mixte de clones ayant effectués la réparation de la DSB par SSA, pour obtenir un allèle fonctionnel, ou par NHEJ qui résulte en un allèle non fonctionnel (voir figure 16).

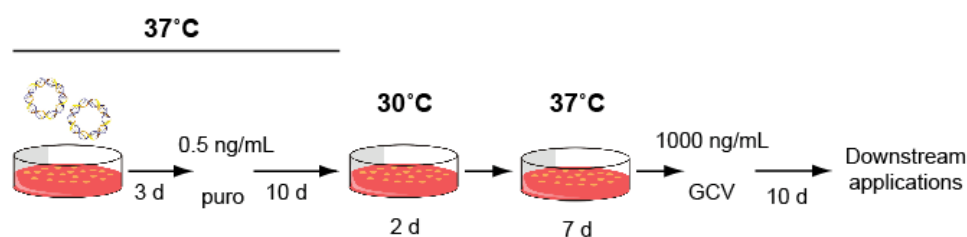


Figure 17 : Schéma expérimental permettant la modification génique via l'utilisation de la cassette auto-excisable.

II.2.2.2 Intégration de la protéine fluorescente EGFP au locus *AAVSI*

Comme première preuve de concept, la protéine fluorescente EGFP a été intégrée au locus *AAVSI*. La résultante de l'induction de la nucléase thermosensible (voir section I.2.1.2) et de la contre-sélection au ganciclovir est une population de cellules vertes réparées par SSA et une population de cellules dont l'allèle EGFP est non fonctionnel puisqu'il a été réparé par NHEJ (Figure 18a). Ici, la présence d'une protéine fluorescente permet de discriminer entre les allèles fonctionnels et non fonctionnels.

Dans un premier temps, l'impact du temps d'induction de la ZFN thermosensible et de la contre-sélection au ganciclovir sur le pourcentage de cellules EGFP⁺ a été quantifié par cytométrie de flux. Dans une population hétérogène (*pool*) de cellules résistantes à la puromycine, une induction de 24h à 30°C sans contre-sélection résulte en 13.6 % de cellules EGFP positives tandis qu'une induction de 48h augmente ce nombre à 19.3 %. Toutefois, lorsqu'une contre-sélection au ganciclovir est appliquée, la population de cellules EGFP⁺ augmente à 18.0 % et 27.7 % respectivement (Figure 18b). Ces résultats soulignent la valeur ajoutée d'une induction plus longue et d'une contre-sélection des cellules induites.

Des résultats similaires ont été obtenus lors de l'induction de clones isolés suite à la sélection à la puromycine. Chez ces populations clonales, le pourcentage de cellules induites qui sont EGFP⁺ est similaire à celui des populations hétérogènes. Malgré tout, le nombre de cellules EGFP⁺ observées suite à la contre-sélection au ganciclovir est plus élevé (69.7 % et 88.1 % respectivement) (Figure 18c). Le pourcentage élevé de cellules EGFP⁺ a été confirmée par la suite par microscopie (Figure 18d).

L'introduction de la séquence EGFP peut aussi être évaluée par PCR de type *out-out*, où il est possible de discriminer entre les allèles ayant réparé par NHEJ et ceux ayant réparé par SSA. En effet, les allèles ayant réparés par NHEJ conservent un site de liaison de la ZFN qui est perdu chez ceux ayant réparé par SSA, menant à un produit PCR plus long (Figure 18e). L'intensité des bandes de PCR amplifiant l'allèle SSA corrèle avec le pourcentage de cellules EGFP positives observées en cytométrie de flux.

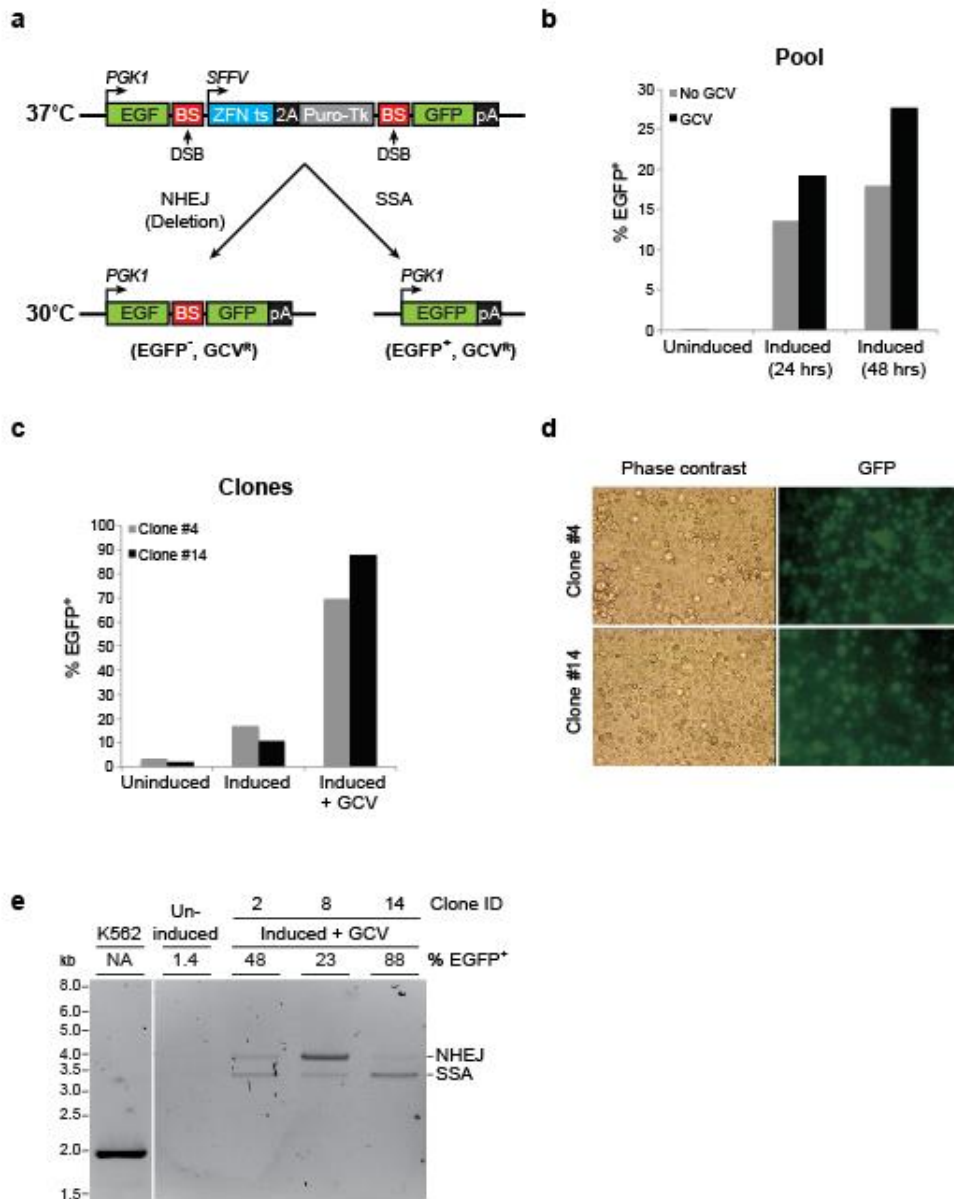


Figure 18 : Intégration de la protéine verte EGFP au locus *AAVS1*. A) Représentation schématique des résultantes possibles suite à l'excision de la cassette EGFP. B) L'induction de la nucléase thermosensible permet de détecter la reconstitution du gène encodant la protéine EGFP et la contre-sélection au ganciclovir augmente la fréquence. C) Cette observation se réplique aussi lors de l'étude de population clonale. D) Observation des

cellules EGFP en contraste de phase et au microscope à fluorescence. E) Amplification par PCR de type out-out des allèles réparés par SSA et NHEJ.

II.2.2.3 Introduction d'une étiquette FLAG au locus *EZH2*

Pour démontrer l'application potentielle de l'utilisation de la cassette excisable, une étiquette FLAG a été introduite au locus *EZH2*. Des cellules K562 ont été transfectées avec une paire de TALEN ciblant l'exon 2 du locus *EZH2* et une cassette de sélection flanquée de 2 parties de la séquence de l'étiquette FLAG et de bras d'homologues au locus *EZH2*. La cassette a été créée de façon à ce que l'étiquette FLAG se retrouve immédiatement après l'ATG codant de la protéine EZH2 qui se situe dans l'exon 2 (Figure 19a). Des clones résistants à la puromycine ont été isolés, induits à 30°C, puis contre-sélectionnés au ganciclovir pendant 10 jours. Suite à son induction, chaque clone isolé en puromycine redevient une population hétérogène puisque chacune des cellules répare les cassures double brin de façon différente. À partir de cette nouvelle population, il est possible d'isoler une deuxième série de clones lors de la contre-sélection au ganciclovir (Figure 19b). Chacun des clones isolés à cette étape aura réparé d'une façon unique ses cassures double brin.

Dans un premier temps, des clones isolés en puromycine ont été induits afin d'étudier l'efficacité du processus d'excision. Suite à un PCR de type *out-out*, il est possible de visualiser l'apparition des allèles réparés par SSA en NHEJ suite à l'induction (Figure 19a,c). Le brin parental non modifié possède une taille de 1070 pb tandis que les allèles ayant réparé par SSA ou NHEJ ont une taille de 1167 pb et 1361 pb respectivement. Nous avons observé que l'induction de la cassure suivie de la réparation génère des populations hétérogènes portants des allèles réparés par SSA ou par NHEJ.

La dynamique de l'enrichissement des différents types de réparation lors de la sélection au ganciclovir fût également étudiée. Le clone #6 sélectionné en puromycine a été induit 48 heures à 30°C et soumis à différent temps de sélection avec 1000 ng/mL de ganciclovir (Figure 19d). Sans contre sélection, une faible proportion des cellules induites a excisé sa cassette correctement via NHEJ ou SSA tandis qu'après 13 jours de sélection, toutes les cellules possédant encore le gène de la thymidine kinase ont été tuées et il est possible de détecter un enrichissement des cellules ayant réparé par SSA et aussi par NHEJ. Ceci

démontre que les cellules provenant d'un même clone empruntent des voies de réparation différentes même si elles ont toutes le même bagage génétique avant la contre sélection. Il est donc essentiel d'établir des populations clonales suite à l'induction et à la contre-sélection. Ceci permet l'étude et la caractérisation des différents évènements de réparation ayant eu lieu dans la population initiale.

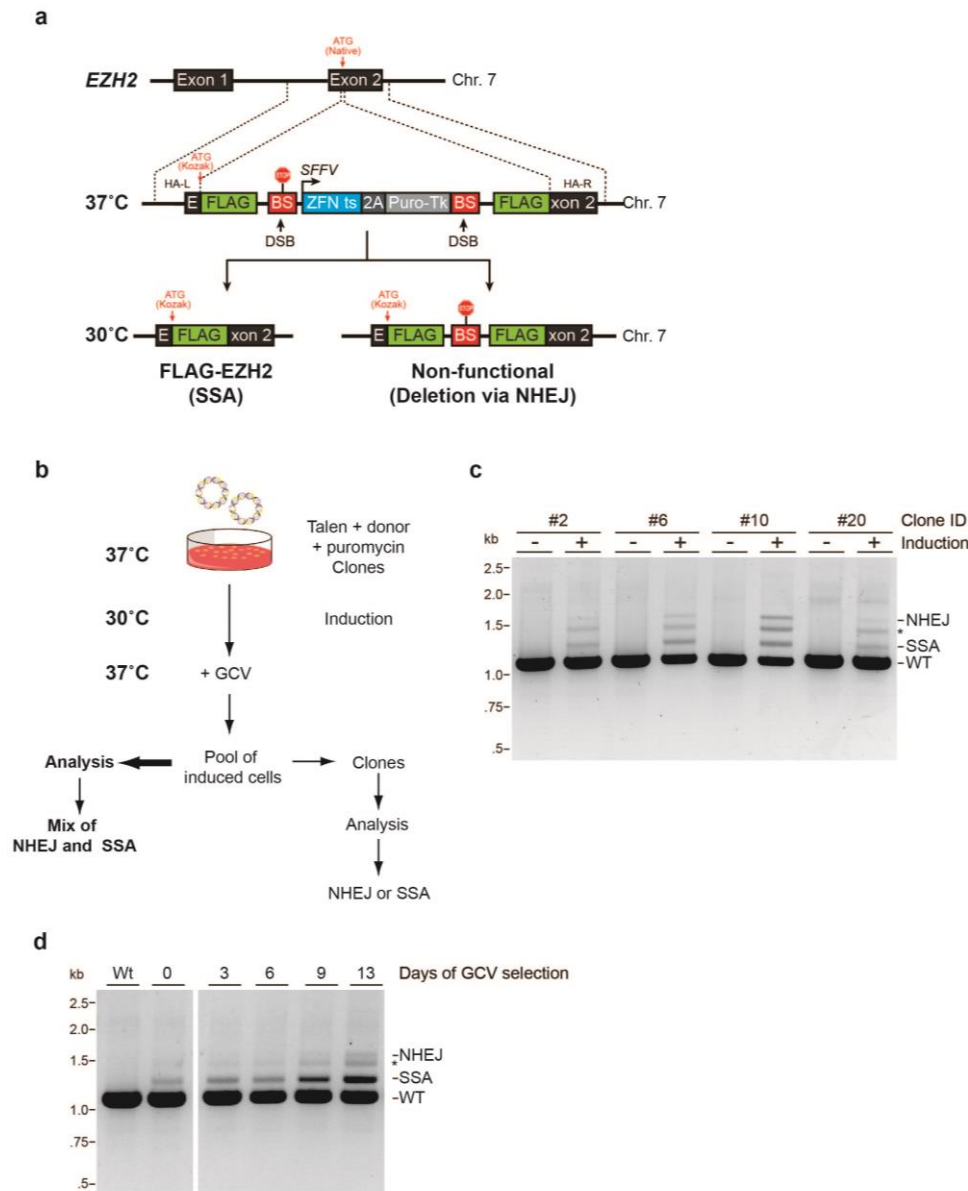


Figure 19 : Insertion d'une séquence FLAG au locus *EZH2* en utilisant une cassette de sélection auto-excisable A) Représentation schématique des résultantes possibles suite à l'excision de la cassette FLAG. B) Schéma expérimental démontrant la production de 2 populations pouvant être étudiées. C) Suite à l'induction de clones cellulaires et à la contre-

sélection en pool de ces clones, il est possible de détecter des fragments PCR correspondant à la réparation de la cassette par NHEJ et par SSA. D) Cinétique de la contre-sélection au ganciclovir d'un pool de cellules suite à l'induction du clone résistant à la puromycine #8. *Artéfact de PCR impossible à TOPO cloner à partir des bandes extraites d'un gel d'agarose sur le clone 10 de la figure 19c.

Pour la suite des analyses, nous avons donc isolé des clones en milieu semi-solide lors de la contre-sélection au ganciclovir (Figure 20a). Des 8 clones analysés, 5 ont réparé via SSA (clones # 2, 3, 6, 10 et 11), 2 via NHEJ (clones # 1 et 5) et un clone (#8) a connu une réparation mixte (Figure 20b). Aucun clone homozygote pour l'insertion du « tag » n'a été observé, puisque la bande sauvage demeure visible dans chacun des échantillons. De façon intéressante, le clone 8 a effectué une réparation mixte. Une hypothèse pour la présence des deux voies de réparations dans un même clone est la présence de 4 allèles du gène *EZH2* dans les cellules K562. Il est probable que ce clone ait effectué la réparation à un allèle par SSA et à un autre allèle par NHEJ. Des immunobuvardages ont été réalisés pour valider la fonctionnalité de l'allèle FLAG chez les différents clones. Tous les clones ayant réparé par SSA possèdent une étiquette fonctionnelle pouvant être reconnue par un anticorps à la taille attendue. Tel que prévu, il est toutefois impossible d'identifier la protéine FLAG dans les clones ayant réparé par NHEJ. Seulement un allèle réparé par SSA permet de produire une protéine FLAG fonctionnelle puisque cette dernière peut être produite chez le clone ayant subi une réparation mixte (voir figure 20). Ceci démontre que la cassette auto-excisable permet de sélectionner et enrichir des clones possédant une modification génique souhaité.

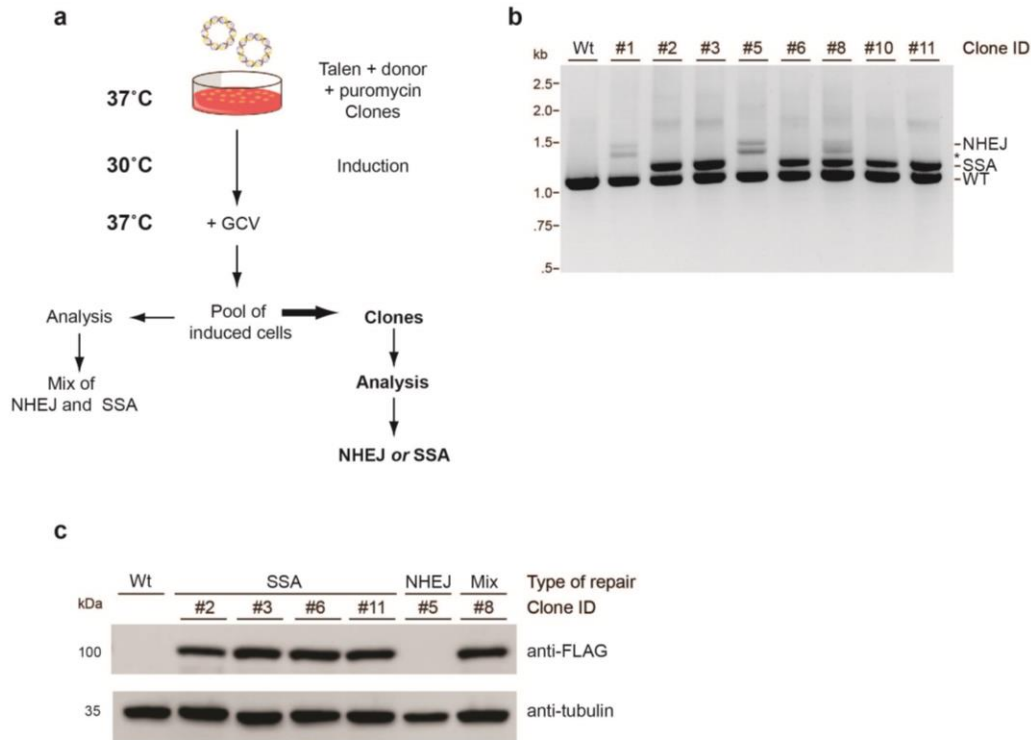


Figure 20: A) Étude de populations clonales isolées en milieu semi-solide lors de la contre-sélection au ganciclovir lors de l'étiquetage du locus *EZH2* endogène avec le 3xFLAG_2xStrep. B) PCR de type *out-out* permettant de déterminer le type de réparation utilisé par chacun des clones isolés. Le clone #6 puroR de la figure 19 a été induit 48h à 30°C puis a été incubé 72h à 37°C pour lui permettre de reprendre un bon rythme de croissance. La contre-sélection au ganciclovir a ensuite été effectuée en matrice semi-solide afin d'obtenir des clones purs au lieu de pool de cellules ayant procédé à la réparation de plusieurs façons. C) Immunobuvardage de l'étiquette FLAG et de la tubuline chez des clones ayant réparé par différents mécanismes. * Bande artéfact du PCR impossible à TOPO cloner.

II.2.3 Discussion

Les expériences menées au cours de ce projet démontrent qu'il est possible de développer une cassette auto-excisable ne laissant aucune trace dans le génome. Elle peut être intégrée sous sa forme inactive à 37°C et son auto excision peut être induite efficacement en incubant les cellules à 30°C. La contre-sélection au ganciclovir nous permet d'isoler une majorité de cellules possédant la modification génique désirée. Nous avons tout d'abord

effectué une preuve de faisabilité en insérant le gène EGFP au locus *AAVS1*. Suite à l'induction de la nucléase thermosensible, les cellules ayant réparé par SSA sont à la fois vertes et résistantes au ganciclovir. Le processus n'est pas parfait, car certaines cellules résistantes au ganciclovir ne sont pas vertes (car la réparation s'est faite par NHEJ plutôt que par SSA). La fraction de ces cellules ayant réparée la DSB par NHEJ varie d'une expérience à l'autre pour des raisons inconnues. Une hypothèse est que la réparation puisse être influencée selon l'étape du cycle cellulaire à laquelle se trouve la cellule, car la résection de l'ADN est influencée selon la phase du cycle cellulaire.

La variabilité semble toutefois réduite lorsque l'induction et la contre-sélection sont faites à partir d'un clone cellulaire résistant à la puromycine plutôt qu'un pool résistant à la puromycine, probablement parce que le *pool* est plus hétérogène au départ.

Nous avons par la suite démontré la fonctionnalité de cette nouvelle approche en insérant une étiquette FLAG fonctionnelle au locus *EZH2*. Suite à l'induction et à la contre-sélection, il est possible d'utiliser des populations hétérogènes ou des clones pour les analyses subséquentes. L'utilisation de clones comporte l'avantage que chaque cellule de la population a effectué le même type de réparation. Lors de nos études, un seul clone semblait avoir effectué une réparation mixte. Ceci est hypothétiquement attribuable à l'intégration de la cassette sur 2 chromosomes (les K562 possèdent 4 copies du chromosome 7, où est localisé *EZH2*). Cette différence peut aussi être attribuée à une erreur de manipulation résultant en la présence de deux populations cellulaires au lieu d'une seule tel que requis pour la préparation de clones. Pour l'étude rapide de certains phénomènes (ex : purification de complexes protéiques), l'utilisation de populations hétérogènes peut être envisageable, mais pour certaines autres applications (ex : étude d'une mutation ponctuelle) il pourrait être préférable de travailler avec une population clonale.

La technique développée ici pourrait s'avérer utile pour les expériences avec les cellules difficiles à cultiver telles les cellules ES et les iPS, puisqu'une seule transfection est nécessaire et que des étapes de sélections sont utilisées pour enrichir les populations d'intérêt. Il est à noter que l'impact des étapes de sélection et d'induction sur la pluripotence devra être évalué.

L'approche développée ici comprend certaines limites. Son efficacité a été démontrée dans un type cellulaire pour deux applications. Les preuves de concept et de fonctionnalité devront être répétées dans d'autres types de cellules couramment utilisées comme les cellules HeLa et même les lignées primaires. Bien que la technique développée compte des étapes de sélection et de contre-sélection permettant l'enrichissement des populations voulues, l'obtention de cellules modifiées repose sur une transfection efficace et la capacité à intégrer la cassette par HDR. Malgré l'efficacité limitée, la nécessité d'isoler des clones puis de les caractériser, l'utilisation de la cassette pourrait s'avérer difficile dans les lignées cellulaires difficiles à transfecter ou qui font peu de réparation par HDR.

Pour améliorer l'efficacité de la cassette de sélection auto-excisable, le gène suicide TK utilisé dans notre cassette pourrait être remplacé par une version optimisée comme TK.007, afin d'améliorer l'efficacité de la contre-sélection au ganciclovir [200].

II.2.4 Conclusion

En conclusion, nous avons développé une cassette auto-excisable permettant l'insertion d'une séquence prédéterminée à un locus d'intérêt. Qu'il s'agisse d'une protéine fluorescente, d'une étiquette de purification ou de l'insertion d'une mutation, cette technique ne laisse aucune cicatrice. Le design de la cassette est versatile et peut être adapté à différents besoins. Davantage d'études seront nécessaires de façon à évaluer l'efficacité de cette technique dans différentes lignées cellulaires, incluant les cellules primaires et les cellules souches.

CHAPITRE 3 : UTILISATION DES DOUBLES NICKASES COMME OUTILS SPÉCIFIQUES DE MODIFICATION GÉNIQUE.

III.3.1 Matériel et méthodes

Lignées cellulaires et transfection

Les cellules K562 ont été obtenues de l'ATCC (CCL-243) et maintenues à 37 °C sous 5% CO₂ dans du milieu RPMI supplémenté avec 10% de sérum fœtal bovin (FBS), 1% de pénicilline-streptomycine et du GlutaMAX (Gibco). Toutes les lignées cellulaires ont été testées pour l'absence de mycoplasme.

Les cellules (1×10^6 /transfection) ont été transfectées en utilisant le nucléofecteur Amaxa 4D (Lonza), selon les recommandations du manufacturier. Les conditions de transfections sont détaillées au tableau 6.

Pour la suite des expériences, une lignée stable de cellules K562 exprimant EGFP de façon constitutive au locus *AAVS1* a été produite en transfectant une ZFN ciblant le locus *AAVS1* déjà décrite [201] ainsi qu'un donneur *AAVS1_EGFP_d4_SA_2A_Puro_FS_v2*. Le donneur a été cloné en amplifiant par PCR la séquence *EGFP_Pest_d4* (Tableau 5) sur le plasmide CMV-DsRed-IRES-d4EGFP (Addgene #41943). Le produit PCR a par la suite été sous cloné dans le vecteur *AAVS1_Puro_FS_v2* en utilisant les enzymes de restrictions *Acc65I* et *NotI* (NEB). Des clones fluorescents ont été isolés en milieu semi-solide contenant de la puromycine, puis amplifiés pour mener à la création d'une lignée stable.

ZFN et TALEN

Des ZFN hétérodimères obligatoires (ELD/KKR) de type sauvage ou possédant la mutation D450N sur le domaine Fok1 ELD et ciblant le locus *CCR5* (*pVax-N2A-3xFLAG-NLS-8266_ELD* et *pVax-N2A-3xFLAG-NLS-8196z_KKR*) ont été utilisés [103]. Des TALEN hétérodimères obligatoires (*pVax_N2A_3xFlag-NLS_101041_DRR* et *pVax_N2A_3xFlag-NLS_101047_RDD*) (RDD/DRR), dérivées de Doyon et al., [95], de type sauvage ou possédant la mutation D450N sur le domaine Fok1 DRR ont aussi été utilisés pour cibler les locus *EGFP* et *CCR5*. La mutation D450N a été insérée sur les domaine

FokI_ELD et FokI_DRR par PCR mutagénique avec *Q5® Site-Directed Mutagenesis Kit Q5* (NEB, #E0554S) (voir Tableau 7). Les nucléases possèdent un épitope FLAG pour permettre leur visualisation par immunobuvardage suite à la transfection.

Essai Surveyor et PCR out-out

L'ADN génomique de 2.5×10^5 cellules a été extrait en utilisant 250 μ l de *QuickExtract DNA extraction solution* (Illumina), selon les instructions du fabricant. Les différents locus ont été amplifiés par 30 cycles de PCR en utilisant les amorces décrites au tableau 7. La formation d'indels a été déterminée en utilisant le *Surveyor mutation detection kit* (Transgenomics), tel que décrit précédemment [202]. Les échantillons traités ont par la suite été séparés sur gel d'acrylamide 10% dans du tampon TBE. La fréquence d'indels a été déterminée en utilisant le logiciel imagelab et la formule suivante:

$$\text{Indel (\%)} = (1 - \text{SQRT}(1 - (\mathbf{b} + \mathbf{c}) / (\mathbf{a} + \mathbf{b} + \mathbf{c}))) \times 100$$

Où **a** est l'intensité intégrée de la bande PCR non digérée et **b** et **c** sont l'intensité intégrée des produits de clivage. SQRT : racine carrée.

De façon à analyser la délétion causée par la réparation par NHEJ via un PCR *out-out*, l'ADN génomique extrait avec la solution d'extraction d'ADN *QuickExtract* a été soumis à 30 cycles de PCR pour amplifier le locus CCR5 en utilisant les amorces décrites au tableau 7. Les amplicons ont par la suite été déposés sur gel d'acrylamide 10% dans du tampon TBE 1x.

Immunobuvardage

Des extraits totaux ont été préparés à partir de $\sim 1 \times 10^6$ cellules par addition de 2x Laemmli buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycérol, 0.004% bleu de bromophénol, 0.125M Tris HCl, ajuster le pH à 6.8) suivi d'une incubation à 95 °C pendant 10 minutes. Pour chaque échantillon, un volume de 10 μ L d'extraits totaux ont été déposés

sur un gel d'acrylamide 8 % dans du tampon Tris-Glycine. Les protéines ont été transférées sur une membrane de nitrocellulose, bloquées avec 5% de poudre de lait diluée dans PBS-Tween et incubées avec leurs anticorps primaires pour toute la nuit à 4 °C. Les anticorps primaires suivants ont été utilisés Anti-FLAG M2 (Sigma, A8592) et anti- α -tubulin (DM1A, sc-32293). L'anticorps secondaire anti-mouse (Cell Signaling, ref : 7076s) et dilués 1 : 10000 dans du PBS-Tween ont été incubés 1 heure à température pièce. Les immunobuvardages ont été révélés par chemiluminescence (GE) à l'aide d'un Chemidoc (Biorad).

Tableau 5 : Séquence des amorces à PCR utilisés pour l'assemblage de la cassette contenant le gène EGFP s'intégrant à *AAVS1*.

| Cible | Amorce sens | Amorce anti-sens |
|-----------------------|---|--|
| CMV-DsRed-IRES-d4EGFP | CAGGGGTACCATGGTGAGCAAGGGC | CTCGGCGGCCGCCTACACATTGATCCT |
| hPGK1 | /5Phos/CTT CGA ACC ACG GGG TTG GGG TTG CGC | /5Phos/CTT AAG TGC ATG GTA CCT GAT CCT GCA GGC TAG GGT CGA CCC TGG GGA GAG AGG TCG GTG |
| SFFV | AAACCTCGAGCACCGGTGTAACGCCA TTTTGC | Ggcggttaccttctcttcttcttgggtcccatggtggccgggc gactcagtctgtc |

Tableau 6 : Conditions de transfection de cellules K562.

| Transfection | Quantité de chaque vecteur |
|--|--|
| Paire de nucléases à CCR5 | 1.25 μ g pVAX_8266_ELD_N2A, pVAX_8196z_KKR_N2A 1.25 μ g pVAX_101047_RDD_N2A pVAX_101041_DRR_N2A |
| Paire de nickases à CCR5 | 1.25 μ g pVAX_8266_ELD_D450N_N2A, 1.25 μ g pVAX_8196z_KKR_N2A 1.25 μ g pVAX_101047_RDD_N2A 1.25 μ g pVAX_101041_DRR_D450N_N2A |
| Lignée stable exprimant EGFP au locus <i>AAVS1</i> | 4 μ g <i>AAVS1</i> _EGFP_D4_SA_2A_Puro_FS_v2 1 μ g de ZFN ciblant le locus <i>AAVS1</i> |
| Paire de nickases rapprochée à EGFP | 2 μ g Tale_2090_KKR, Tale_2091_ELD_D450N 2 μ g Tale_2058_DRR_D450N & Tale_2059_RDD |
| Paire de nickases éloignée à EGFP | 2 μ g Tale_2062_DRR_D450N, Tale_2063_RDD 2 μ g Tale_2090_KKR & Tale_2091_ELD_D450N, |
| Paire de nickases sur le même brin | 2 μ g Tale_2062_DRR_D450N, Tale_2063_RDD 2 μ g Tale_2058_DRR_D450N & Tale_2059_RDD |

Tableau 7 : Séquence des oligonucléotides utilisés pour les PCR de criblage et de séquençage.

| Cible | Amorce sens | Amorce anti-sens |
|------------------------|----------------------------|-----------------------|
| EGFP Cel1 | AGGGGTACCATGGTGAGCAAG | GGGAAGCCATGGCTAAGCTTC |
| CCR5 ZFN Cel1 | AAGATGGATTATCAAGTGTCAAGTCC | CAAAGTCCCCTGGGCG |
| CCR5 TALEN Cel1 | TTAAAAGCCAGGACGGTCAC | TGTAGGGAGCCCAGAAGAGA |
| CCR5 <i>out-out</i> | AAGATGGATTATCAAGTGTCAAGTCC | TGTAGGGAGCCCAGAAGAGA |
| Mutation FokI_D450N | CAGAAAGCCTaACGGCGCCAT* | CTTCCGCCAGGTGCTTT |

* L'adénine en minuscule est la mutation insérée

III.3.2 Résultats

III.3.2.1 Une faible distance entre deux paires de nickases augmente leur efficacité

Nous voulions mettre au point un système permettant d'induire une cassure double brin dans le génome d'un organisme vivant avec plus de spécificité que les systèmes existants actuellement. Il a été démontré avec le système CRISPR/Cas9 que deux cassures simple brin rapprochées résultent en une cassure double brin. Nous avons voulu tester si nous étions capables de faire la même chose, mais en utilisant des domaines FokI orthogonaux, auxquels nous avons inactivé la moitié de chaque hétérodimère obligatoire FokI, avec la mutation D450N, afin de ne créer qu'une cassure simple brin à chaque site où une paire de FokI : FokI_D450N (aussi nommé nickase).

Nous avons donc créé des paires de TALEN nickase dont les domaines FokI de départ étaient les domaines FokI orthogonaux ELD : KKR et DRR : RDD. Dans chacune de ses paires, les domaines FokI ELD et FokI DRR ont été inactivé avec la mutation D450N. Nous avons ensuite vérifié l'efficacité de ces doubles nickases à créer des cassures double brin.

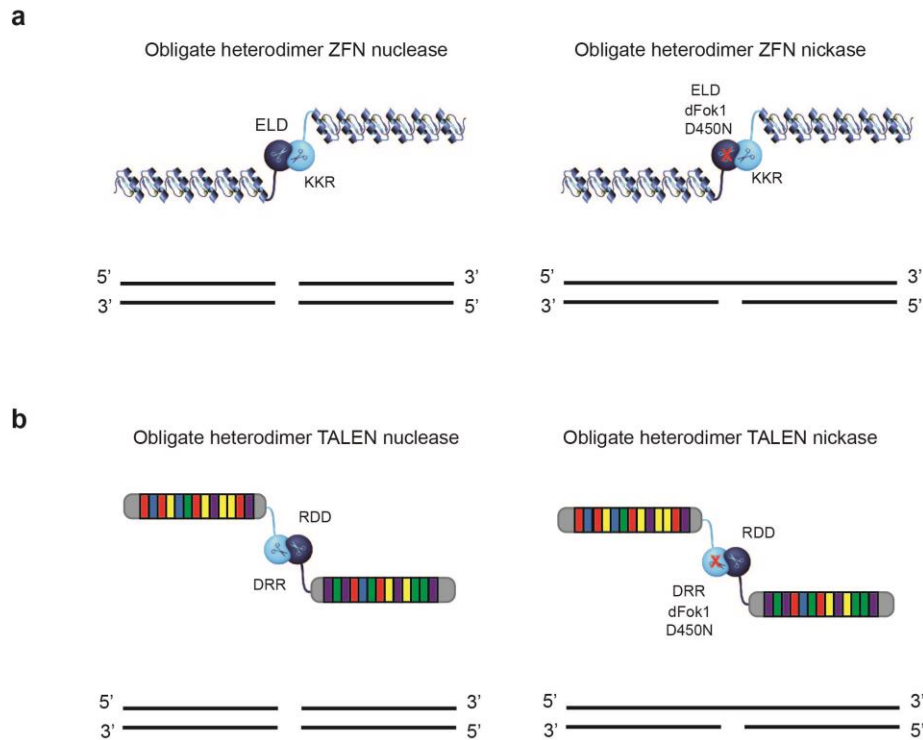


Figure 21 : A) Représentation schématique des ZFN à hétérodimères obligatoires ELD/KKR possédant un domaine Fok1 de type sauvage ou nickase. B) Représentation schématique des TALEN hétérodimères obligatoires RDD/DRR possédant un domaine Fok1 de type sauvage ou nickase. Figurée préparée par Alexandre Raymond-Fleury.

Pour se faire, 3 TALEN nickases ciblant le locus EGFP ont été utilisés (Figure 22). Les sites de clivage de ces dernières se trouvaient à des distances variables les uns des autres (Figure 22a). Les nickases ont été transfectées en paires pour évaluer l'effet de la distance entre les deux sites de clivage sur l'efficacité à générer une cassure double brin réparée par NHEJ, tel qu'évalué par la production d'indels. La transfection de deux nickases clivant le même brin (#30 + #33) ne génère pas de cassure double brin tandis que l'utilisation de deux nickases coupant l'ADN sur les brins opposés produit des indels détectables par l'essai Surveyor (Figure 22b). L'efficacité des doubles nickases à produire des cassures double brins réparés par NHEJ est inversement liée à la distance séparant les deux paires. En effet, une distance de 128 pb entre les deux sites de clivage induit 5% d'indels, mais lorsque cette dernière est réduite à 60 pb, la formation d'indels augmente à 8% (mesuré à partir de

l'intensité de la bande sur un essai Cell1 avec le logiciel Image J, voir figure 22c). Ces changements dans l'efficacité ne semblent pas liés à une modification des niveaux de nucléases transfectés tel qu'évalué par immunobuvardage (Figure 22d). Nous avons donc réussi à recréer une cassure double brin en utilisant un système de doubles nickase orthogonales de façon analogue au système CRISPR.

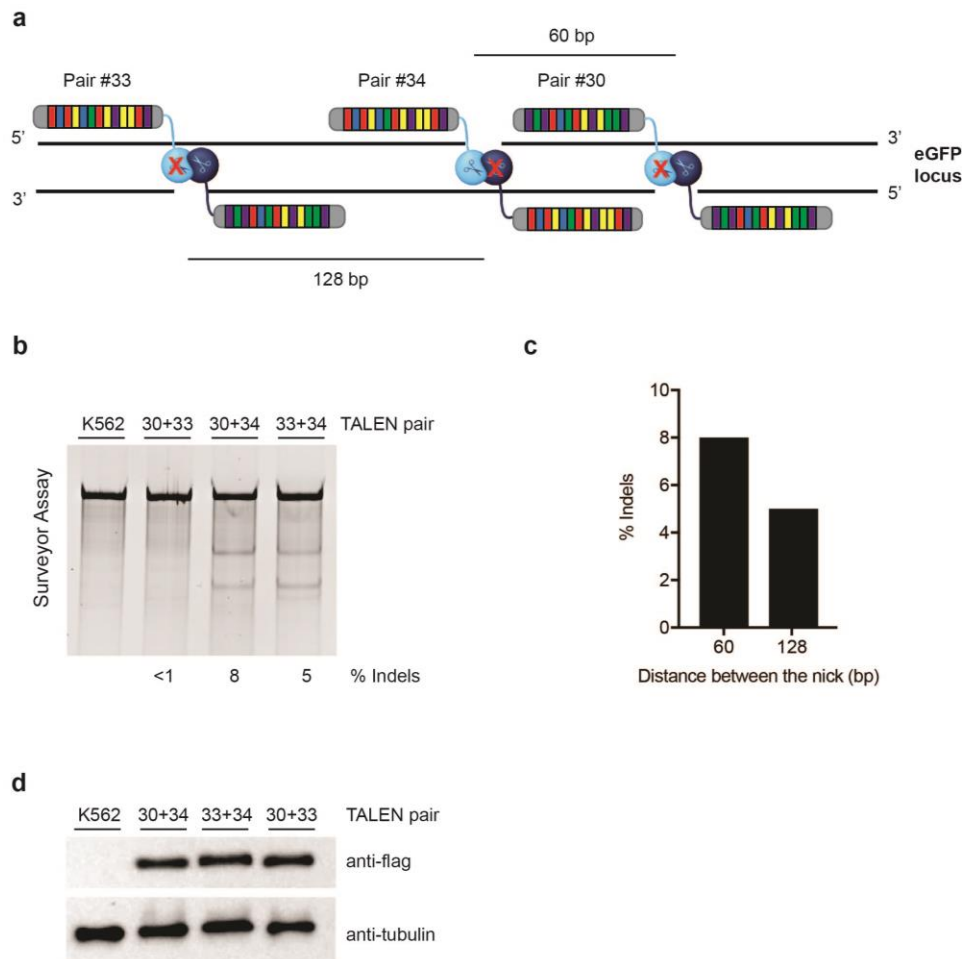


Figure 22 : Représentation schématique de l'utilisation de deux paires de TALEN nickase pour recréer une cassure double brin au locus EGFP. A) différentes paires de TALEN nickase à hétérodimères obligatoires ciblant le locus EGFP ont été utilisées pour déterminer s'il est possible de créer une cassure double brin qui sera réparée par NHEJ. B) Essai Surveyor sur les cellules ayant été transfectées avec différentes nickases. C) Quantification du pourcentage

d'indels selon la distance entre les deux sites de coupure. D) Immunobuvardage contre l'épitope FLAG présent sur les TALEN et contre la tubuline.

III.3.2.2 Utilisées seules, les nickases n'induisent pas la formation d'indels

L'efficacité des nickases seules à induire la formation d'indels a par la suite été investiguée. Des ZFN et TALEN nucléases et nickases ciblant le locus CCR5 ont été transfectés dans des cellules K562. Alors que l'utilisation d'une nucléase induit la formation d'indels, l'utilisation d'une nickase seule ne produit pas d'indels détectables avec l'essai Surveyor, et ce, tant lors de l'utilisation de ZFN que de TALEN (Figure 23a). Lorsqu'utilisés en paire, les deux nickases acquièrent la capacité d'induire une délétion réparée par NHEJ au locus CCR5 tel qu'évalué par un PCR de type *out-out* (Figure 23b). Leur efficacité est toutefois réduite comparativement à l'utilisation de deux nucléases. Ceci démontre que nous pouvons utiliser un système de doubles nickases avec des domaines orthogonaux à un locus endogène afin de recréer une cassure double brin.

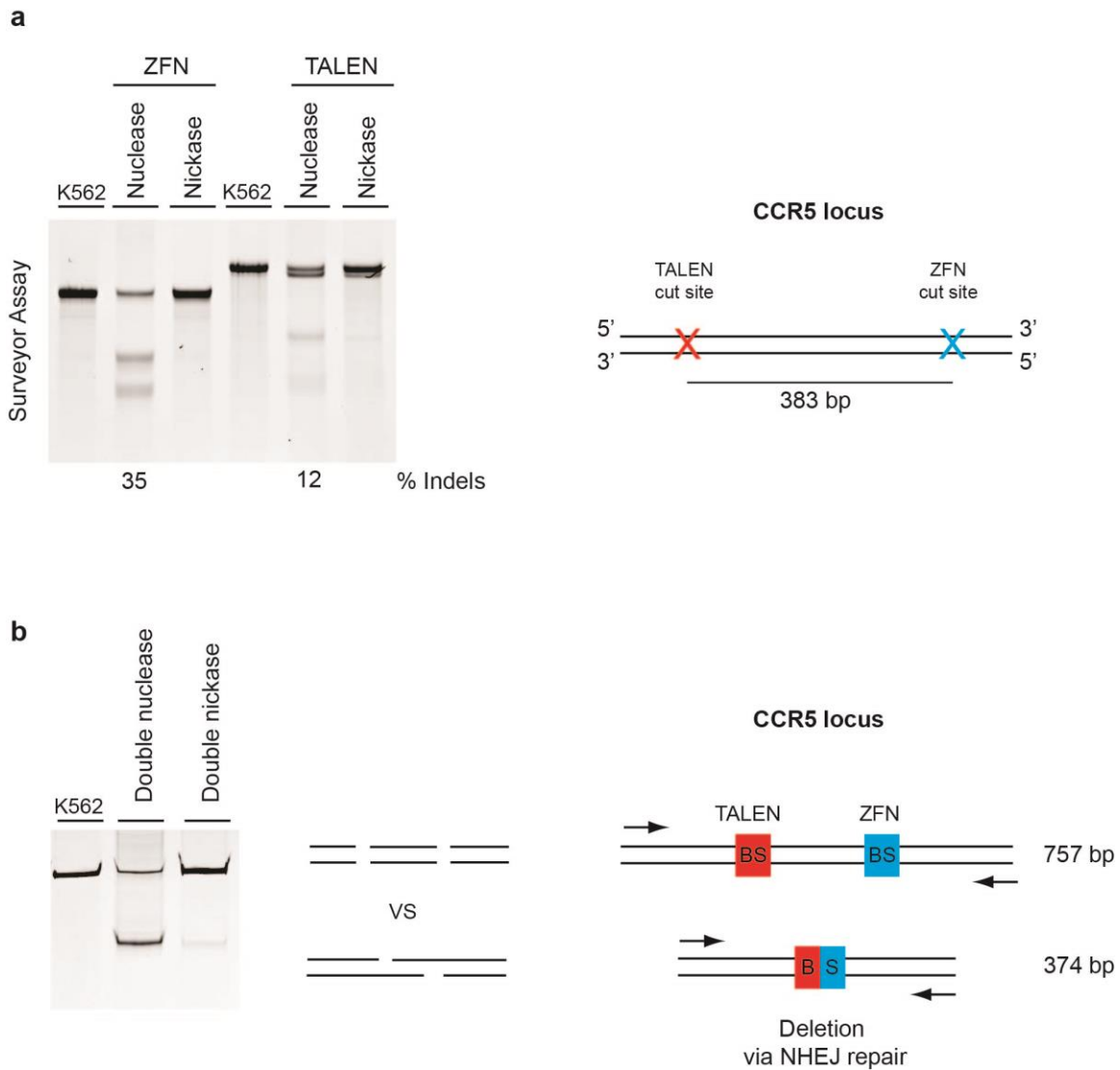


Figure 23: Création d'une cassure double brin au locus CCR5. A) Comparaison de l'utilisation de simples nucléases et de simples nickases par essai Surveyor. B) Comparaison de l'utilisation de doubles nucléases et de doubles nickases par PCR de type *out-out*.

III.3.3 Discussion

La démocratisation de la modification ciblée du génome a ouvert les portes à de nombreuses découvertes et à la création de nombreux modèles cellulaires et animaux. Toutefois, la spécificité demeure un problème relativement peu étudié et l'impact de la formation de cassures double brin hors cible est peu connu. Ici, nous faisons la preuve de concept que deux coupures simple brin réalisées à l'aide d'une paire de ZFNickase et de

TALENickase nickases peuvent être réparées par NHEJ comme une cassure double brin. Dans un premier lieu, nous avons créé des cassures simple brin à différentes distances à un gène rapporteur EGFP. Nous avons observé que plus les cassures simple brin sont rapprochées, plus ces dernières peuvent être réparées par NHEJ. En effet, une diminution de la moitié de la distance entre les deux sites de coupure augmente la fréquence de réparation par NHEJ de près du double (voir Fig 22b). La distance minimale testée ici était de 60 pb entre les deux sites de clivage. Il serait possible de tester des distances plus courtes, mais une distance minimale d'environ 20 pb devrait se trouver entre les deux sites de coupure puisque les modules des TALEN et des ZFN reconnaissent entre 20 et 12 pb d'ADN [80, 203].

La motivation derrière l'utilisation de doubles nickases est de limiter, voire éliminer, les cassures double brin hors cible. Ceci rendrait l'utilisation des nucléases d'ingénierie plus sécuritaire. Nous croyons que l'utilisation de doubles nickases est très spécifique et induirait peu de cassures hors cible. Tout d'abord, chacun des deux modules protéiques doit reconnaître sa séquence cible et son partenaire hétérodimère obligatoire de façon à pouvoir créer une cassure simple brin. Ce processus doit être répété une seconde fois à une distance rapprochée pour créer une seconde cassure simple brin. Le système doit donc reconnaître au total 4 fois 18 paires de bases d'ADN dans la bonne orientation et le bon espacement pour qu'il y ait formation de cassures simple brin réparées par NHEJ. Une fois installé sur l'ADN, chaque domaine FokI doit aussi se lier à son hétérodimère obligatoire pour réussir à couper un brin d'ADN. Les analyses de spécificité n'ont toutefois pas été réalisées. De nombreuses techniques bio-informatiques et expérimentales existent et ces dernières devront être utilisées pour valider la spécificité de notre approche [204-206].

Dans notre preuve de concept, la spécificité vient au prix d'une diminution importante de l'efficacité à former des indels, tel que démontré par nos expériences au locus CCR5. De façon à pallier à cette diminution d'efficacité, notre technique pourrait être couplée à une méthode de co-sélection. En effet, notre équipe a récemment développé une technique de co-sélection basée sur la modification de la pompe sodium-potassium, qui permet de sélectionner les cellules qui ont subi une modification génique en utilisant le glycoside

cardiaque ouabain [207]. Ceci permettrait l'enrichissement d'une population contenant la modification désirée sans augmenter les modifications hors cibles.

Une faible efficacité sans sélection écarte toutefois la possibilité d'utiliser cette technique lors de modifications géniques *in vivo*. Ceci s'explique par le fait qu'aucune technique de sélection des cellules modifiées ne peut être appliquée *in vivo* et par le fait qu'un nombre faible de cellules modifiées n'est souvent pas suffisant pour modifier le phénotype de la maladie. Malgré tout, notre stratégie, couplée à un processus de co-sélection pourrait être parfaitement adaptée à la thérapie génique *ex vivo*, où les cellules malades sont prélevées, modifiées hors du corps, sélectionnées et amplifiées, puis retransférées chez le patient malade [192].

III.3.4 Conclusion

En conclusion, nous avons démontré ici la preuve de concept que deux cassures simple brin, créées par des doubles nickases basées sur ZFN et TALEN ayant des sites de coupure suffisamment proche, peuvent être réparées par NHEJ. Nous estimons que cette approche possède une spécificité augmentée, mais plus d'analyses devront être faites. De façon intéressante, il est possible de fusionner les domaines FokI hétérodimères obligatoires au système CRISPR-Cas dont l'activité nucléase a été abolie (dCas9). Cette dernière devient donc un guide de reconnaissance pour une séquence spécifique d'ADN pour l'enzyme FokI. Il serait intéressant de déterminer si l'utilisation de ce système produit des résultats similaires à ceux obtenus avec les TALEN et les ZFN.

DISCUSSION INTÉGRATIVE ET PERSPECTIVES

Les nucléases d'ingénierie simplifient la création de modifications géniques personnalisées. Malgré sa simplicité apparente, la reprogrammation du système CRISPR demeure une tâche ardue et relativement longue. Cette difficulté est en partie attribuable à l'efficacité variable des différents gRNA pour la même cible et au manque d'outils bio-informatiques capables de prédire l'efficacité des gRNA. Ainsi, la sélection de gRNA actifs, l'identification et l'isolation de clones possédant les modifications génétiques recherchées est un processus long et qui peut être coûteux. Un enjeu important de l'utilisation des nucléases d'ingénierie est la spécificité. En effet, la nucléase doit pouvoir cliver le site d'intérêt sans produire de cassures double brin hors-cible. La caractérisation des coupures hors cibles et l'augmentation de la spécificité des nucléases sont des points critiques pour le développement de la modification du génome comme thérapie.

Les deux projets présentés ici visent à simplifier et améliorer les techniques d'édition du génome, principalement dans les cellules en culture. Les techniques développées adressent les préoccupations liées à la simplification du processus d'édition génique et à la spécificité. La cassette auto-excisable permet d'insérer des mutations précises sans laisser de traces. Les avantages de cette technique incluent la production de lignées cellulaires en une seule transfection et le nombre limité de manipulations nécessaires pour la production des cellules modifiées. L'utilisation des doubles nickases hétérodimères obligatoires nécessite la reconnaissance de 4 séquences d'ADN et l'hétérodimérisation des domaines FokI pour induire deux cassures simple brin. Si ces dernières sont suffisamment rapprochées, elles peuvent être réparées comme une cassure double brin. Dans l'éventualité où une des deux nickase reconnaît une séquence hors cible, une cassure simple brin sera produite, puis réparée sans induire de mutations.

Les preuves de concept pour ces deux techniques ont été réalisées, mais certaines questions demeurent. Entre autres, les expériences présentées ici devront être réalisées dans différentes lignées cellulaires en ciblant divers locus de façon à démontrer la flexibilité des techniques. Certains types cellulaires comme les fibroblastes peuvent être plus difficiles à

transfecter et d'autres, comme les cellules primaires, pourraient ne pas tolérer une exposition à 30 degrés. Il sera aussi intéressant de déterminer la sensibilité à la température de différentes lignées. Par exemple, il pourrait être possible de créer une lignée cellulaire qui possède l'ADN complémentaire de la fumarylacetoacetate hydrolase (FAH), invalidée chez les enfants atteints de tyrosinémie, au locus *AAVSI* de cellules malades [208]. Ce modèle permettrait d'évaluer l'impact de la restauration de l'expression de FAH et d'étudier l'impact de différents promoteurs sur le phénotype de la cellule. Aussi, la capacité des doubles nickases à induire une réparation par HDR devra être investiguée. Lorsque les deux sites de coupure sont suffisamment près l'un de l'autre, la cellule répare un certain pourcentage des cassures par NHEJ, comme s'il s'agissait d'une cassure double brin. L'étape suivante serait de fournir un guide de réparation lors de la transfection de façon à déterminer si les cellules possèdent la capacité de réparer par HDR. Certaines évidences suggèrent qu'une cassure simple brin pourrait être réparée par HDR [209]. Ce type de réparation possède l'avantage de ne pas faire intervenir le processus de NHEJ; soit la réparation est parfaite, soit il y a intégration du fragment.

Si les cassures faites par les doubles nickases peuvent être réparées par HDR, il serait intéressant de combiner les deux méthodes. C'est à dire insérer la cassette auto-excisable dans une cassure créée par les doubles nickases. Ceci permettrait de créer une lignée modifiée qui ne comporte peu ou pas de cassures hors cible. L'efficacité d'intégration de la cassette serait sans doute très faible, mais grâce à la sélection à la puromycine, il serait possible d'amplifier ces événements rares de façon à obtenir une population de cellules ayant intégré la cassette et qui pourront par la suite être induites et contre-sélectionnées. Un point important qui demeure à valider est la spécificité des doubles nickases. Les sites de cassures hors cibles peuvent être prédits bio-informatiquement ou expérimentalement. Les analyses bio-informatiques se basent sur la recherche de séquences dans le génome comportant des séquences semblables. Dans le cas d'une technique qui utilise 4 modules de ZFN en doubles nickases, l'approche à privilégier est l'approche expérimentale. Différentes approches ont été décrites, mais la majorité d'entre elles se basent sur l'insertion de séquences spécifiques aux sites où des cassures double brin ont eu lieu (ex : oligodéoxynucléotides double brin), suivi de séquençage à haut débit [210]. Il est intéressant de noter que malgré tout, une approche

expérimentale pour la caractérisation des cassures hors cible peut ne pas refléter la réalité. Il a été récemment soulevé que les variations interindividuelles présentes dans le génome peuvent avoir un impact sur le nombre de cassures hors cible [211]. En effet, la présence de certains SNP pourrait augmenter ou réduire le nombre de sites hors cible, ce qui souligne l'importance de bien caractériser chacun des patients avant le traitement. Lors de l'utilisation thérapeutique de nucléases d'ingénierie, chaque patient devra sans doute voir son génome séquencé pour assurer plus de sécurité.

Les deux techniques développées ici s'appliquent particulièrement aux cellules en culture, principalement lors de la création de modèles cellulaires. Toutefois, il est possible d'inférer une utilisation pour la thérapie génique. Une fois la flexibilité et la spécificité validées, les techniques présentées dans ce mémoire pourraient être utiles pour la thérapie génique *ex vivo*. Le système de doubles nickases pourrait être utilisé pour créer une cassure double brin très spécifique au locus d'intérêt et pour intégrer la cassette auto-excisable. La modification génique réparatrice serait créée par les différentes étapes de sélection et d'induction, puis les cellules comportant une seule modification seraient réinjectées chez le patient. Il faudrait évidemment que les cellules des patients puissent maintenir leur fonction malgré ces étapes de sélection et de contre sélection. La recherche sur la spécificité et la sélection des cellules modifiées doit bien sûr se faire de pair avec celle sur les techniques d'enrichissement pour que la thérapie génique *ex vivo* puisse connaître du succès.

CONCLUSION

Les travaux présentés dans ce mémoire participeront à simplifier l'utilisation des techniques de modification du génome et à adresser la problématique de spécificité associée à l'utilisation des nucléases d'ingénierie. L'emploi de la cassette auto-excisable permettra de réduire le temps lié au génotypage de clones positifs pour la mutation désirée tandis que l'utilisation de double nickases permettra la création de modifications géniques de façon spécifique. Certaines expériences devront être réalisées pour démontrer la flexibilité des techniques développées, mais ces dernières devraient permettre à un plus grand nombre de laboratoire d'utiliser les nucléases d'ingénierie.

RÉFÉRENCES

1. Ioannidis, J.P., P. Castaldi, and E. Evangelou, A compendium of genome-wide associations for cancer: critical synopsis and reappraisal. *J Natl Cancer Inst*, 2010. 102(12): p. 846-58.
2. Hunter, D.J., et al., A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*, 2007. 39(7): p. 870-4.
3. Caldecott, K.W., Single-strand break repair and genetic disease. *Nat Rev Genet*, 2008. 9(8): p. 619-31.
4. Fortini, P. and E. Dogliotti, Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair (Amst)*, 2007. 6(4): p. 398-409.
5. Esposito, M.T. and C.W. So, DNA damage accumulation and repair defects in acute myeloid leukemia: implications for pathogenesis, disease progression, and chemotherapy resistance. *Chromosoma*, 2014. 123(6): p. 545-61.
6. Hubscher, U., Spadari, S., Villani, G. and Maga, G, DNA polymerases: discovery, characterization and functions in cellular DNA transactions. Vol. 1. 2010, Singapore: World Scientific Publishing Co. Pte. Ltd. 340.
7. Sobol, R.W. and S.H. Wilson, Mammalian DNA beta-polymerase in base excision repair of alkylation damage. *Prog Nucleic Acid Res Mol Biol*, 2001. 68: p. 57-74.
8. Almeida, K.H. and R.W. Sobol, A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification. *DNA Repair (Amst)*, 2007. 6(6): p. 695-711.
9. Atanassov, B., et al., Comparison of the global genomic and transcription-coupled repair rates of different lesions in human cells. *Z Naturforsch C*, 2004. 59(5-6): p. 445-53.
10. Lee, Y.C., et al., The relationships between XPC binding to conformationally diverse DNA adducts and their excision by the human NER system: is there a correlation? *DNA Repair (Amst)*, 2014. 19: p. 55-63.

11. Spivak, G., Nucleotide excision repair in humans. *DNA Repair (Amst)*, 2015. 36: p. 13-8.
12. Scrima, A., et al., Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell*, 2008. 135(7): p. 1213-23.
13. Marteijn, J.A., et al., Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol*, 2014. 15(7): p. 465-81.
14. Fagbemi, A.F., B. Orelli, and O.D. Scharer, Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair (Amst)*, 2011. 10(7): p. 722-9.
15. Lagerwerf, S., et al., DNA damage response and transcription. *DNA Repair (Amst)*, 2011. 10(7): p. 743-50.
16. van Hoffen, A., et al., Nucleotide excision repair and its interplay with transcription. *Toxicology*, 2003. 193(1-2): p. 79-90.
17. Kunkel, T.A. and D.A. Erie, DNA mismatch repair. *Annu Rev Biochem*, 2005. 74: p. 681-710.
18. McCulloch, S.D., L. Gu, and G.M. Li, Bi-directional processing of DNA loops by mismatch repair-dependent and -independent pathways in human cells. *J Biol Chem*, 2003. 278(6): p. 3891-6.
19. Li, G.M., Mechanisms and functions of DNA mismatch repair. *Cell Res*, 2008. 18(1): p. 85-98.
20. Longley, M.J., A.J. Pierce, and P. Modrich, DNA polymerase delta is required for human mismatch repair in vitro. *J Biol Chem*, 1997. 272(16): p. 10917-21.
21. Zhang, Y., et al., Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell*, 2005. 122(5): p. 693-705.
22. Gu, L., et al., ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. *Nucleic Acids Res*, 1998. 26(5): p. 1173-8.
23. Rich, T., R.L. Allen, and A.H. Wyllie, Defying death after DNA damage. *Nature*, 2000. 407(6805): p. 777-83.

24. Burma, S., B.P. Chen, and D.J. Chen, Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair (Amst)*, 2006. 5(9-10): p. 1042-8.
25. Szostak, J.W., et al., The double-strand-break repair model for recombination. *Cell*, 1983. 33(1): p. 25-35.
26. Weinert, T.A. and L.H. Hartwell, The RAD9 gene controls the cell cycle response to DNA damage in *Saccharomyces cerevisiae*. *Science*, 1988. 241(4863): p. 317-22.
27. Pardo, B., B. Gomez-Gonzalez, and A. Aguilera, DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. *Cell Mol Life Sci*, 2009. 66(6): p. 1039-56.
28. Zannini, L., D. Delia, and G. Buscemi, CHK2 kinase in the DNA damage response and beyond. *J Mol Cell Biol*, 2014. 6(6): p. 442-57.
29. Lin, Y.F., et al., DNA-PKcs is required to maintain stability of Chk1 and Claspin for optimal replication stress response. *Nucleic Acids Res*, 2014. 42(7): p. 4463-73.
30. Bartek, J. and J. Lukas, Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell*, 2003. 3(5): p. 421-9.
31. Hopfner, K.P., et al., The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature*, 2002. 418(6897): p. 562-6.
32. Haber, J.E., The many interfaces of Mre11. *Cell*, 1998. 95(5): p. 583-6.
33. Melo, J.A., J. Cohen, and D.P. Toczyski, Two checkpoint complexes are independently recruited to sites of DNA damage in vivo. *Genes Dev*, 2001. 15(21): p. 2809-21.
34. Dubrana, K., et al., The processing of double-strand breaks and binding of single-strand-binding proteins RPA and Rad51 modulate the formation of ATR-kinase foci in yeast. *J Cell Sci*, 2007. 120(Pt 23): p. 4209-20.
35. Lee, J.H. and T.T. Paull, ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science*, 2005. 308(5721): p. 551-4.

36. Grenon, M., C. Gilbert, and N.F. Lowndes, Checkpoint activation in response to double-strand breaks requires the Mre11/Rad50/Xrs2 complex. *Nat Cell Biol*, 2001. 3(9): p. 844-7.
37. Zou, L. and S.J. Elledge, Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science*, 2003. 300(5625): p. 1542-8.
38. Doyon, Y. and J. Cote, The highly conserved and multifunctional NuA4 HAT complex. *Curr Opin Genet Dev*, 2004. 14(2): p. 147-54.
39. Huertas, P. and S.P. Jackson, Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J Biol Chem*, 2009. 284(14): p. 9558-65.
40. Badie, S., et al., BRCA1 and CtIP promote alternative non-homologous end-joining at uncapped telomeres. *EMBO J*, 2015. 34(3): p. 410-24.
41. Hu, Y., et al., PARP1-driven poly-ADP-ribosylation regulates BRCA1 function in homologous recombination-mediated DNA repair. *Cancer Discov*, 2014. 4(12): p. 1430-47.
42. Mao, Z., et al., DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle*, 2008. 7(18): p. 2902-6.
43. Valerie, K. and L.F. Povirk, Regulation and mechanisms of mammalian double-strand break repair. *Oncogene*, 2003. 22(37): p. 5792-812.
44. Shibata, A., et al., Factors determining DNA double-strand break repair pathway choice in G2 phase. *EMBO J*, 2011. 30(6): p. 1079-92.
45. Blier, P.R., et al., Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *J Biol Chem*, 1993. 268(10): p. 7594-601.
46. Lieber, M.R., The mechanism of human nonhomologous DNA end joining. *J Biol Chem*, 2008. 283(1): p. 1-5.
47. Lieber, M.R., The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem*, 2010. 79: p. 181-211.

48. Yoo, S. and W.S. Dynan, Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res*, 1999. 27(24): p. 4679-86.
49. Uematsu, N., et al., Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J Cell Biol*, 2007. 177(2): p. 219-29.
50. Jovanovic, M. and W.S. Dynan, Terminal DNA structure and ATP influence binding parameters of the DNA-dependent protein kinase at an early step prior to DNA synapsis. *Nucleic Acids Res*, 2006. 34(4): p. 1112-20.
51. Yannone, S.M., et al., Coordinate 5' and 3' endonucleolytic trimming of terminally blocked blunt DNA double-strand break ends by Artemis nuclease and DNA-dependent protein kinase. *Nucleic Acids Res*, 2008. 36(10): p. 3354-65.
52. Yamtich, J. and J.B. Sweasy, DNA polymerase family X: function, structure, and cellular roles. *Biochim Biophys Acta*, 2010. 1804(5): p. 1136-50.
53. Gu, J., et al., XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J*, 2007. 26(4): p. 1010-23.
54. Gu, J., et al., Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Res*, 2007. 35(17): p. 5755-62.
55. Tsai, C.J., S.A. Kim, and G. Chu, Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc Natl Acad Sci U S A*, 2007. 104(19): p. 7851-6.
56. Essers, J., et al., Analysis of mouse Rad54 expression and its implications for homologous recombination. *DNA Repair (Amst)*, 2002. 1(10): p. 779-93.
57. Yonetani, Y., et al., Differential and collaborative actions of Rad51 paralog proteins in cellular response to DNA damage. *Nucleic Acids Res*, 2005. 33(14): p. 4544-52.
58. Yang, H., et al., The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction. *Nature*, 2005. 433(7026): p. 653-7.

59. Heyer, W.D., et al., Rad54: the Swiss Army knife of homologous recombination? *Nucleic Acids Res*, 2006. 34(15): p. 4115-25.
60. Solinger, J.A., K. Kiiianitsa, and W.D. Heyer, Rad54, a Swi2/Snf2-like recombinational repair protein, disassembles Rad51:dsDNA filaments. *Mol Cell*, 2002. 10(5): p. 1175-88.
61. Sugiyama, T. and S.C. Kowalczykowski, Rad52 protein associates with replication protein A (RPA)-single-stranded DNA to accelerate Rad51-mediated displacement of RPA and presynaptic complex formation. *J Biol Chem*, 2002. 277(35): p. 31663-72.
62. Li, X. and W.D. Heyer, Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res*, 2008. 18(1): p. 99-113.
63. Plank, J.L., J. Wu, and T.S. Hsieh, Topoisomerase IIIalpha and Bloom's helicase can resolve a mobile double Holliday junction substrate through convergent branch migration. *Proc Natl Acad Sci U S A*, 2006. 103(30): p. 11118-23.
64. Liu, Y., et al., RAD51C is required for Holliday junction processing in mammalian cells. *Science*, 2004. 303(5655): p. 243-6.
65. Bosco, G. and J.E. Haber, Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics*, 1998. 150(3): p. 1037-47.
66. Llorente, B., C.E. Smith, and L.S. Symington, Break-induced replication: what is it and what is it for? *Cell Cycle*, 2008. 7(7): p. 859-64.
67. Lydeard, J.R., et al., Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature*, 2007. 448(7155): p. 820-3.
68. Petalcorin, M.I., et al., CeBRC-2 stimulates D-loop formation by RAD-51 and promotes DNA single-strand annealing. *J Mol Biol*, 2006. 361(2): p. 231-42.
69. Bachrati, C.Z., R.H. Borts, and I.D. Hickson, Mobile D-loops are a preferred substrate for the Bloom's syndrome helicase. *Nucleic Acids Res*, 2006. 34(8): p. 2269-79.

70. Sugawara, N., G. Ira, and J.E. Haber, DNA length dependence of the single-strand annealing pathway and the role of *Saccharomyces cerevisiae* RAD59 in double-strand break repair. *Mol Cell Biol*, 2000. 20(14): p. 5300-9.
71. Bhargava, R., D.O. Onyango, and J.M. Stark, Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet*, 2016. 32(9): p. 566-75.
72. Grimme, J.M., et al., Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nucleic Acids Res*, 2010. 38(9): p. 2917-30.
73. Rothenberg, E., et al., Human Rad52-mediated homology search and annealing occurs by continuous interactions between overlapping nucleoprotein complexes. *Proc Natl Acad Sci U S A*, 2008. 105(51): p. 20274-9.
74. Motycka, T.A., et al., Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. *J Biol Chem*, 2004. 279(14): p. 13634-9.
75. Sugawara, N., et al., Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1. *Proc Natl Acad Sci U S A*, 2004. 101(25): p. 9315-20.
76. Stark, J.M., et al., Genetic steps of mammalian homologous repair with distinct mutagenic consequences. *Mol Cell Biol*, 2004. 24(21): p. 9305-16.
77. Venter, J.C., et al., The sequence of the human genome. *Science*, 2001. 291(5507): p. 1304-51.
78. Edwards, Y.J. and A. Cottage, Bioinformatics methods to predict protein structure and function. A practical approach. *Mol Biotechnol*, 2003. 23(2): p. 139-66.
79. Kim, Y.G., J. Cha, and S. Chandrasegaran, Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A*, 1996. 93(3): p. 1156-60.
80. Miller, J.C., et al., A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, 2011. 29(2): p. 143-8.
81. Moscou, M.J. and A.J. Bogdanove, A simple cipher governs DNA recognition by TAL effectors. *Science*, 2009. 326(5959): p. 1501.

82. Christian, M., et al., Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 2010. 186(2): p. 757-61.
83. Jinek, M., et al., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012. 337(6096): p. 816-21.
84. Chandrasegaran, S. and D. Carroll, Origins of Programmable Nucleases for Genome Engineering. *J Mol Biol*, 2016. 428(5 Pt B): p. 963-89.
85. Rebar, E.J. and C.O. Pabo, Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science*, 1994. 263(5147): p. 671-3.
86. Reyon, D., et al., FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*, 2012. 30(5): p. 460-5.
87. Cho, S.W., et al., Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol*, 2013. 31(3): p. 230-2.
88. Miller, J., A.D. McLachlan, and A. Klug, Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J*, 1985. 4(6): p. 1609-14.
89. Porteus, M.H. and D. Carroll, Gene targeting using zinc finger nucleases. *Nat Biotechnol*, 2005. 23(8): p. 967-73.
90. Sander, J.D., et al., An affinity-based scoring scheme for predicting DNA-binding activities of modularly assembled zinc-finger proteins. *Nucleic Acids Res*, 2009. 37(2): p. 506-15.
91. Dutta, S., et al., A theoretical investigation of DNA dynamics and desolvation kinetics for zinc finger protein Zif268. *BMC Genomics*, 2015. 16 Suppl 12: p. S5.
92. Liu, Q., et al., Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci U S A*, 1997. 94(11): p. 5525-30.
93. Aggarwal, A.K. and D.A. Wah, Novel site-specific DNA endonucleases. *Curr Opin Struct Biol*, 1998. 8(1): p. 19-25.
94. Bitinaite, J., et al., FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A*, 1998. 95(18): p. 10570-5.
95. Doyon, Y., et al., Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods*, 2011. 8(1): p. 74-9.

96. Tsai, S.Q. and J.K. Joung, Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet*, 2016. 17(5): p. 300-12.
97. Slaymaker, I.M., et al., Rationally engineered Cas9 nucleases with improved specificity. *Science*, 2016. 351(6268): p. 84-8.
98. Wah, D.A., et al., Structure of FokI has implications for DNA cleavage. *Proc Natl Acad Sci U S A*, 1998. 95(18): p. 10564-9.
99. Guo, J., T. Gaj, and C.F. Barbas, 3rd, Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J Mol Biol*, 2010. 400(1): p. 96-107.
100. Miller, J.C., et al., An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*, 2007. 25(7): p. 778-85.
101. Urnov, F.D., et al., Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, 2010. 11(9): p. 636-46.
102. Kim, E., et al., Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res*, 2012. 22(7): p. 1327-33.
103. Wang, J., et al., Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res*, 2012. 22(7): p. 1316-26.
104. Porteus, M.H. and D. Baltimore, Chimeric nucleases stimulate gene targeting in human cells. *Science*, 2003. 300(5620): p. 763.
105. Kandavelou, K., et al., Targeted manipulation of mammalian genomes using designed zinc finger nucleases. *Biochem Biophys Res Commun*, 2009. 388(1): p. 56-61.
106. Bibikova, M., et al., Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*, 2002. 161(3): p. 1169-75.
107. Geurts, A.M., et al., Knockout rats via embryo microinjection of zinc-finger nucleases. *Science*, 2009. 325(5939): p. 433.
108. Shukla, V.K., et al., Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature*, 2009. 459(7245): p. 437-41.

109. Doyon, Y., et al., Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol*, 2008. 26(6): p. 702-8.
110. Tebas, P., et al., Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med*, 2014. 370(10): p. 901-10.
111. DeFrancesco, L., Move over ZFNs. *Nat Biotechnol*, 2011. 29(8): p. 681-4.
112. Ramirez, C.L., et al., Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods*, 2008. 5(5): p. 374-5.
113. Boch, J., et al., Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 2009. 326(5959): p. 1509-12.
114. Kay, S. and U. Bonas, How *Xanthomonas* type III effectors manipulate the host plant. *Curr Opin Microbiol*, 2009. 12(1): p. 37-43.
115. Mak, A.N., et al., The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, 2012. 335(6069): p. 716-9.
116. Deng, D., et al., Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, 2012. 335(6069): p. 720-3.
117. Benjamin, R., et al., TALEN gene editing takes aim on HIV. *Hum Genet*, 2016. 135(9): p. 1059-70.
118. Mussolino, C., et al., A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*, 2011. 39(21): p. 9283-93.
119. Ke, Q., et al., TALEN-based generation of a cynomolgus monkey disease model for human microcephaly. *Cell Res*, 2016. 26(9): p. 1048-61.
120. Lee, H.B., et al., An optimized TALEN application for mutagenesis and screening in *Drosophila melanogaster*. *Cell Logist*, 2015. 5(1): p. e1023423.
121. Berdien, B., et al., TALEN-mediated editing of endogenous T-cell receptors facilitates efficient reprogramming of T lymphocytes by lentiviral gene transfer. *Gene Ther*, 2014. 21(6): p. 539-48.
122. Sekine, K., T. Takebe, and H. Taniguchi, Fluorescent labeling and visualization of human induced pluripotent stem cells with the use of transcription activator-like effector nucleases. *Transplant Proc*, 2014. 46(4): p. 1205-7.

123. Juillerat, A., et al., Design of chimeric antigen receptors with integrated controllable transient functions. *Sci Rep*, 2016. 6: p. 18950.
124. Ran, F.A., et al., Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, 2013. 8(11): p. 2281-308.
125. Mohanraju, P., et al., Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science*, 2016. 353(6299): p. aad5147.
126. Barrangou, R. and L.A. Marraffini, CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell*, 2014. 54(2): p. 234-44.
127. Mojica, F.J., et al., Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*, 2005. 60(2): p. 174-82.
128. Gasiunas, G., et al., Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A*, 2012. 109(39): p. E2579-86.
129. Barrangou, R., et al., CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 2007. 315(5819): p. 1709-12.
130. Makarova, K.S., et al., An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*, 2015. 13(11): p. 722-36.
131. Shmakov, S., et al., Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell*, 2015. 60(3): p. 385-97.
132. Deltcheva, E., et al., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 2011. 471(7340): p. 602-7.
133. Makarova, K.S., et al., Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*, 2011. 9(6): p. 467-77.
134. Bolotin, A., et al., Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 2005. 151(Pt 8): p. 2551-61.
135. Deveau, H., et al., Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol*, 2008. 190(4): p. 1390-400.
136. Ran, F.A., et al., In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*, 2015. 520(7546): p. 186-91.

137. Cong, L., et al., Multiplex genome engineering using CRISPR/Cas systems. *Science*, 2013. 339(6121): p. 819-23.
138. Zetsche, B., et al., Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 2015. 163(3): p. 759-71.
139. Qi, L.S., et al., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 2013. 152(5): p. 1173-83.
140. Larson, M.H., et al., CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc*, 2013. 8(11): p. 2180-96.
141. Chavez, A., et al., Comparison of Cas9 activators in multiple species. *Nat Methods*, 2016. 13(7): p. 563-567.
142. Gilbert, L.A., et al., Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, 2014. 159(3): p. 647-61.
143. Chen, L., et al., Advances in genome editing technology and its promising application in evolutionary and ecological studies. *Gigascience*, 2014. 3: p. 24.
144. Kleinstiver, B.P., et al., High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 2016. 529(7587): p. 490-5.
145. Nishimasu, H., et al., Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 2014. 156(5): p. 935-49.
146. Ran, F.A., et al., Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 2013. 154(6): p. 1380-9.
147. Tsai, S.Q., et al., Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*, 2014. 32(6): p. 569-76.
148. Yang, L., et al., CRISPR/Cas9-Directed Genome Editing of Cultured Cells. *Curr Protoc Mol Biol*, 2014. 107: p. 31 1 1-17.
149. Li, H., et al., In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature*, 2011. 475(7355): p. 217-21.
150. Rivella, S. and M. Sadelain, Genetic treatment of severe hemoglobinopathies: the combat against transgene variegation and transgene silencing. *Semin Hematol*, 1998. 35(2): p. 112-25.
151. Karpen, G.H., Position-effect variegation and the new biology of heterochromatin. *Curr Opin Genet Dev*, 1994. 4(2): p. 281-91.

152. Modlich, U., et al., Leukemias following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis. *Blood*, 2005. 105(11): p. 4235-46.
153. Papapetrou, E.P. and A. Schambach, Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Mol Ther*, 2016. 24(4): p. 678-84.
154. DeKolver, R.C., et al., Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res*, 2010. 20(8): p. 1133-42.
155. Samulski, R.J., et al., Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J*, 1991. 10(12): p. 3941-50.
156. Henckaerts, E. and R.M. Linden, Adeno-associated virus: a key to the human genome? *Future Virol*, 2010. 5(5): p. 555-574.
157. Hamilton, H., et al., Adeno-associated virus site-specific integration and AAVS1 disruption. *J Virol*, 2004. 78(15): p. 7874-82.
158. Lombardo, A., et al., Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods*, 2011. 8(10): p. 861-9.
159. Rottman, J.B., et al., Cellular localization of the chemokine receptor CCR5. Correlation to cellular targets of HIV-1 infection. *Am J Pathol*, 1997. 151(5): p. 1341-51.
160. Lim, J.K., et al., CCR5: no longer a "good for nothing" gene--chemokine control of West Nile virus infection. *Trends Immunol*, 2006. 27(7): p. 308-12.
161. Ivics, Z., et al., Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell*, 1997. 91(4): p. 501-10.
162. Plasterk, R.H., Z. Izsvak, and Z. Ivics, Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet*, 1999. 15(8): p. 326-32.
163. Narayanavari, S.A., et al., Sleeping Beauty transposition: from biology to applications. *Crit Rev Biochem Mol Biol*, 2017. 52(1): p. 18-44.
164. Munoz-Lopez, M. and J.L. Garcia-Perez, DNA transposons: nature and applications in genomics. *Curr Genomics*, 2010. 11(2): p. 115-28.

165. Mates, L., et al., Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet*, 2009. 41(6): p. 753-61.
166. Cui, Z., et al., Structure-function analysis of the inverted terminal repeats of the sleeping beauty transposon. *J Mol Biol*, 2002. 318(5): p. 1221-35.
167. Fraser, M.J., et al., Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Mol Biol*, 1996. 5(2): p. 141-51.
168. Cadinanos, J. and A. Bradley, Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res*, 2007. 35(12): p. e87.
169. Li, X., et al., piggyBac internal sequences are necessary for efficient transformation of target genomes. *Insect Mol Biol*, 2005. 14(1): p. 17-30.
170. Wilson, M.H., C.J. Coates, and A.L. George, Jr., PiggyBac transposon-mediated gene transfer in human cells. *Mol Ther*, 2007. 15(1): p. 139-45.
171. Rice, P.A. and T.A. Baker, Comparative architecture of transposase and integrase complexes. *Nat Struct Biol*, 2001. 8(5): p. 302-7.
172. Mitra, R., J. Fain-Thornton, and N.L. Craig, piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J*, 2008. 27(7): p. 1097-109.
173. Fraser, M.J., et al., Assay for movement of Lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology*, 1995. 211(2): p. 397-407.
174. Ding, S., et al., Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*, 2005. 122(3): p. 473-83.
175. Bire, S., et al., Exogenous mRNA delivery and bioavailability in gene transfer mediated by piggyBac transposition. *BMC Biotechnol*, 2013. 13: p. 75.
176. Woodard, L.E. and M.H. Wilson, piggyBac-ing models and new therapeutic strategies. *Trends Biotechnol*, 2015. 33(9): p. 525-33.
177. Bonin, C.P. and R.S. Mann, A piggyBac transposon gene trap for the analysis of gene expression and function in *Drosophila*. *Genetics*, 2004. 167(4): p. 1801-11.

178. Kolacsek, O., et al., Excision efficiency is not strongly coupled to transgenic rate: cell type-dependent transposition efficiency of sleeping beauty and piggyBac DNA transposons. *Hum Gene Ther Methods*, 2014. 25(4): p. 241-52.
179. Singh, A., Goodwin, M., PiggyBac-ing Through the Genome Editing Field. *Addgene's Blog*, 2016.
180. Sternberg, N., et al., Analysis of bacteriophage P1 immunity by using lambda-P1 recombinants constructed in vitro. *Proc Natl Acad Sci U S A*, 1978. 75(11): p. 5594-8.
181. Hoess, R., et al., DNA specificity of the Cre recombinase resides in the 25 kDa carboxyl domain of the protein. *J Mol Biol*, 1990. 216(4): p. 873-82.
182. Hoess, R.H. and K. Abremski, Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. *J Mol Biol*, 1985. 181(3): p. 351-62.
183. Ringrose, L., et al., Comparative kinetic analysis of FLP and cre recombinases: mathematical models for DNA binding and recombination. *J Mol Biol*, 1998. 284(2): p. 363-84.
184. Guo, F., D.N. Gopaul, and G.D. van Duyne, Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, 1997. 389(6646): p. 40-6.
185. Abremski, K. and R. Hoess, Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J Biol Chem*, 1984. 259(3): p. 1509-14.
186. Lillis, A.P., et al., LDL receptor-related protein 1: unique tissue-specific functions revealed by selective gene knockout studies. *Physiol Rev*, 2008. 88(3): p. 887-918.
187. Zhou, X.Y., et al., HFE gene knockout produces mouse model of hereditary hemochromatosis. *Proc Natl Acad Sci U S A*, 1998. 95(5): p. 2492-7.
188. Liu, R., et al., Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell*, 1996. 86(3): p. 367-77.

189. Maier, D.A., et al., Efficient clinical scale gene modification via zinc finger nuclease-targeted disruption of the HIV co-receptor CCR5. *Hum Gene Ther*, 2013. 24(3): p. 245-58.
190. Poirot, L., et al., Multiplex Genome-Edited T-cell Manufacturing Platform for "Off-the-Shelf" Adoptive T-cell Immunotherapies. *Cancer Res*, 2015. 75(18): p. 3853-64.
191. Qasim, W., et al., Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. *Sci Transl Med*, 2017. 9(374).
192. Cyranoski, D., CRISPR gene-editing tested in a person for the first time. *Nature*, 2016. 539(7630): p. 479.
193. Kaiser, J., First proposed human test of CRISPR passes initial safety review. *Science*, 2016.
194. Yant, S.R., et al., Site-directed transposon integration in human cells. *Nucleic Acids Res*, 2007. 35(7): p. e50.
195. Aalto-Setälä, K. and E. Vuorio, Gene therapy of single-gene disorders: preface to the special section. *Ann Med*, 1997. 29(6): p. 549-51.
196. Miyaoka, Y., et al., Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci Rep*, 2016. 6: p. 23549.
197. Dalvai, M., et al., A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. *Cell Rep*, 2015. 13(3): p. 621-33.
198. Dambournet, D., et al., Tagging endogenous loci for live-cell fluorescence imaging and molecule counting using ZFNs, TALENs, and Cas9. *Methods Enzymol*, 2014. 546: p. 139-60.
199. Kenwrick, S., et al., Molecular analysis of the Duchenne muscular dystrophy region using pulsed field gel electrophoresis. *Cell*, 1987. 48(2): p. 351-7.
200. Preuss, E., et al., TK.007: A novel, codon-optimized HSVtk(A168H) mutant for suicide gene therapy. *Hum Gene Ther*, 2010. 21(8): p. 929-41.
201. Hockemeyer, D., et al., Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol*, 2009. 27(9): p. 851-7.

202. Guschin, D.Y., et al., A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol*, 2010. 649: p. 247-56.
203. Sander, J.D., et al., Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods*, 2011. 8(1): p. 67-9.
204. Zischewski, J., R. Fischer, and L. Bortesi, Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol Adv*, 2017. 35(1): p. 95-104.
205. Martin, F., et al., Biased and Unbiased Methods for the Detection of Off-Target Cleavage by CRISPR/Cas9: An Overview. *Int J Mol Sci*, 2016. 17(9).
206. Haeussler, M., et al., Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol*, 2016. 17(1): p. 148.
207. Agudelo, D., et al., Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat Methods*, 2017. 14(6): p. 615-620.
208. Grompe, M., et al., A single mutation of the fumarylacetoacetate hydrolase gene in French Canadians with hereditary tyrosinemia type I. *N Engl J Med*, 1994. 331(6): p. 353-7.
209. Ramirez, C.L., et al., Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res*, 2012. 40(12): p. 5560-8.
210. Tsai, S.Q., et al., GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*, 2015. 33(2): p. 187-197.
211. Scott, D.A. and F. Zhang, Implications of human genetic variation in CRISPR-based therapeutic genome editing. *Nat Med*, 2017.