
FICHE NOUVELLES TECHNIQUES

NUCLEASES DIRIGÉES (SDN)

Présentation générale de la technique (Grand public)

Les **nucléases site-spécifiques (SDN pour Site Directed Nuclease)** sont des complexes moléculaires capables d'interagir avec une séquence précise de l'ADN d'un génome (un site) et de provoquer une coupure de la molécule d'ADN au niveau du site d'interaction. Leur spécificité pour une séquence de nucléotides d'ADN provient soit de la nucléase elle-même (cas des méganucléases, nucléases à doigts de zinc ZFN, nucléases de type TALEN), soit de l'apport conjoint d'une protéine nucléase et d'un petit ARN homologue à la séquence sélectionnée qui guide l'interaction de la nucléase avec l'ADN (cas de l'ARN dit guide des nucléases de type CRISPR/Cas) (Fig. 1)

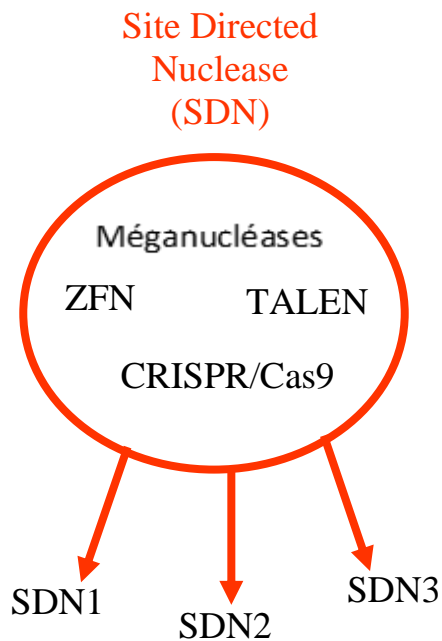


Figure 1 : Les nucléases dirigées (SDN). Dans le cercle rouge sont montrées les types de nucléases site-spécifiques, à savoir les TALEN, ZFN, méganucléases et Cas9 (avec ARN guide). Les flèches indiquent que ces nucléases peuvent être utilisées pour trois types de modification génétique appelés SDN1, SDN2 et SDN3, qui se caractérisent non pas la nucléase utilisée mais par la modification de l'ADN obtenue (voir texte et Fig. 2).

Principe et fonctionnement au niveau cellulaire et moléculaire

Principe général

La technologie de mutagenèse dirigée par les nucléases ciblées¹ repose sur l'interaction spécifique d'une protéine ou d'un complexe moléculaire caractérisé par une activité d'endonucléase avec une séquence nucléotidique d'ADN présélectionnée par l'utilisateur. Le clivage de l'ADN ciblé active une cascade de réparation de type *Non Homologous End Joining* (NHEJ) (Kim et al., 1996) induisant des modifications de l'ADN, ou de type *Homologous Recombination* (HR) si une séquence d'ADN modèle est transférée concomitamment à la nucléase (Moehle et al., 2007). C'est sur ce principe général que reposent les trois techniques actuelles de mutagenèse dirigée utilisant les nucléases de type ZFN, TALEN et CRISPR/Cas9.

ZFN

Les ZFN sont des protéines chimériques composées du domaine de clivage, le plus souvent celui de l'endonucléase bactérienne Fok1, fusionné en amino-terminal à 3 à 6 motifs « en doigt de zinc », ou motifs *Zinc Finger*, de type Cys₂His₂ (Kim et al., 1996).

Le domaine catalytique de l'endonucléase clive l'ADN de manière non spécifique mais son activité dépend d'une homo-dimérisation.

¹ Nucléases ciblées : enzyme qui coupe l'ADN après avoir reconnu une séquence longue, par rapport aux enzymes de restriction qui reconnaissent des séquences le plus souvent comprises entre 4 et 8 nucléotides.

Les motifs « en doigt de zinc » sont des structures protéiques qui interagissent avec un triplet de nucléotides de l'ADN (Duca et al., 2008). La spécificité de l'interaction entre un doigt de zinc et un triplet d'ADN est connue et il est donc possible de définir la structure peptidique requise pour promouvoir l'interaction entre une séquence pré-établie de 6, 9 ou 12 nucléotides et une protéine composée d'une suite de 2, 3 ou 4 motifs *Zinc Finger*. Pour note, un autre motif d'interaction avec l'ADN de type *Drosophila Ubxhomeodomain* a été testé selon le même principe (Kim et al., 1996).

L'expression concomitante de deux ZFN capables de se fixer à l'ADN en reconnaissant des séquences physiquement proches permet la dimérisation des domaines endonucléases et donc la coupure de l'ADN entre les sites de fixation. Cette coupure double brin (DSB, *double strand break*) active le système de réparation de la cellule. Cette réparation se fait par des mécanismes comme le collage des extrémités d'ADN (NHEJ), et est souvent associée à l'introduction de petites mutations (changement de bases, élimination ou addition d'un petit nombre de bases) qui peut mener à l'inactivation de gènes. En présence d'une séquence matrice, montrant une forte similarité de séquence avec la séquence coupée, une recombinaison homologue (RH) peut entraîner la substitution de la séquence coupée par la séquence matrice. Ceci permet d'introduire des changements précis dans la séquence (modification dirigée de la séquence d'un gène) où même d'insérer des fragments d'ADN hétérologues (transgénèse ciblée).

TALEN

Partageant le même objectif que les technologies ZFN et CRISPR, la technologie des TALEN (*Transcriptor Activator Like Effector Nuclease*) repose sur le principe de l'interaction spécifique d'une protéine avec une séquence d'ADN présélectionnée. Le clivage de l'ADN cible par une endonucléase active une cascade de réparation de type *Non Homologous End Joining* (NHEJ) induisant des modifications de l'ADN, ou de type *Homologous Recombination* (HR), si une séquence d'ADN modèle est transférée concomitamment. Il en résulte une modification de l'ADN ciblé, par délétion ou substitution nucléotidique, voire par addition d'une séquence d'ADN (Hockemeyer et al., 2011).

Les TALEN sont des protéines chimériques composées du domaine de clivage d'une endonucléase (par exemple : Fok1) fusionné à des répétitions de séquences de 33 à 35 acides aminés qui dérivent des motifs TALE. Les TALE sont des protéines bactériennes capables d'interagir spécifiquement avec les séquences d'ADN d'une plante et de modifier la transcription de gènes de celle-ci (Mussolino and Cathomen, 2012)(Gaj et al., 2013).

L'endonucléase Fok1 ne clivant l'ADN que sous forme de dimère, son activité repose sur l'expression concomitante de deux TALEN, chacune se fixant de part et d'autre d'une séquence nucléotidique de 12 à 20 paires de bases. La coupure des deux brins d'ADN (DSB, *double strand break*) active le système de réparation de l'ADN de la cellule. Les mécanismes de réparation par « collage » des extrémités d'ADN (NHEJ) permettent l'inactivation de gènes. En présence d'une séquence matrice, montrant une forte similarité de séquence avec la séquence coupée, une recombinaison homologue (HR) peut entraîner la substitution de la séquence coupée par la séquence matrice. Ceci permet d'introduire des

changements précis dans la séquence (modification dirigée de la séquence d'un gène) où même d'insérer des fragments d'ADN hétérologues (transgénèse ciblée).

CRISPR Cas9

Le système CRISPR/Cas (CRISPR : *clustered regularly interspaced short palindromic repeats*) est un mécanisme d'immunité adaptative décrit chez des procaryotes (il existe plus de trois systèmes différents et c'est le *CRISPR/CRISPR-associated Cas system Type II* qui est utilisé). Ce mécanisme repose sur l'expression du locus CRISPR du génome bactérien. Ce locus est constitué de l'insertion de séquences d'ADN provenant du génome d'agents « infectieux » des bactéries et de motifs d'ADN spécialisés. La transcription du locus CRISPR permet la synthèse des CRISPR-RNA ou crRNA, composés d'une séquence de ciblage du génome de l'agent infectieux et de motifs spécifiques. Le crRNA est associé à une seconde molécule, le TracrRNA qui s'hybride au crRNA. Ces deux molécules associées sont reconnues par l'endonuclease bactérienne Cas9 pour former un complexe qui interagit et clive l'ADN double brin du génome de l'agent lors d'une infection. C'est la complémentarité de la séquence du crRNA avec la séquence de l'agent qui donne la spécificité de coupure de la Cas9 qui contribue ainsi à éliminer l'agent infectieux. Cette interaction requiert des motifs particuliers de la séquence cible, dont la séquence PAM (*Protospacer adjacent Motif*) et les éléments d'espacements du crRNA (*spacer*) pour assurer un fonctionnement optimal du complexe (Mali et al., 2013a; Richter et al., 2013). Le fonctionnement du locus CRISPR est assimilé à un système de protection ciblé contre certains agents infectieux.

Les technologies d'ingénierie moléculaire utilisant ce système reposent sur la possibilité de contrôler le ciblage et le fonctionnement de ces composants et de les réunir dans la cellule dont on souhaite modifier le génome. Le crRNA est ainsi modifié pour cibler le complexe vers une séquence génétique choisie par l'utilisateur et exprimé sous la forme ARN dit guide (sgRNA, *small guide RNA*) rassemblant les motifs du TracrRNA et du crRNA en une seule molécule. Techniquement, il est nécessaire que tous les éléments décrits ci-dessus soient présents concomitamment dans la cellule et l'introduction de cet ARN guide est effectué en parallèle avec le transfert d'une unité d'expression de la protéine Cas9. Il existe des outils informatiques permettant le dessin d'ARN guides à disposition de la communauté scientifique (<http://crispr.mit.edu/> - <http://zifit.partners.org/ZiFiT/> - <http://eendb.zfgenetics.org/>).

Dans cette stratégie, l'endonuclease Cas9 recrutée vers cette séquence coupe le double brin d'ADN (DSB, *Double Strand Break*) et active le système de réparation d'ADN de la cellule. Les mécanismes de réparation par « collage » des extrémités d'ADN (NHEJ) permettent l'inactivation de gènes.

En présence d'une séquence matrice, montrant une forte similarité de séquence avec la séquence coupée, une recombinaison homologue (HR) peut entraîner la substitution de la séquence coupée par la séquence matrice. Ceci permet d'introduire des changements précis dans la séquence (modification dirigée de la séquence d'un gène) où même d'insérer des fragments d'ADN hétérologues (transgénèse ciblée).

En pratique, ce système permet :

- L'inactivation de gènes (Cong et al., 2013; Mali et al., 2013b), ce qui permet les études fonctionnelles des gènes ; l'inactivation d'un gène dont la protéine a une fonction non souhaitée (un récepteur à un virus par exemple) ou l'inactivation d'une voie métabolique pour en favoriser une autre par exemple.
- La conversion génique, pour provoquer des échanges d'allèles : introduire un allèle de résistance à une maladie, changer un allèle associé à une maladie contre l'allèle physiologique (Auer and Del Bene, 2014; Hisano et al., 2015; Kimura et al., 2014).
- L'insertion ciblée dans une région génomique précise de fragment d'ADN : transgénèse, cisgénèse ou intragénèse ciblées.

Utilisations possibles

SDN1

Le terme SDN1 désigne les techniques qui produisent une coupure double-brin aléatoire à proximité (quelques bases) d'une séquence cible sur l'ADN, coupure qui est ensuite réparée par les différentes voies naturelles de réparation de la cellule. Chez les végétaux vasculaires, en l'absence d'une matrice d'ADN exogène permettant une réparation homologue, la réparation des cassures double-brins s'effectue principalement par ligation non homologie-dépendante (NHEJ, Non Homologous End Joining) des extrémités d'ADN coupées (Fig.2). Ce rétablissement de la continuité de la molécule d'ADN peut engendrer des délétions et/ou insertions (d'une ou de plusieurs bases) et/ou des mutations ponctuelles (substitutions). Dans certains cas, les extrémités de la coupure peuvent aussi se lier à d'autres extrémités de coupures doubles brins présentes dans le noyau au moment de la réparation, créant ainsi une translocation chromosomique : les mutations obtenues via SDN1 sont, par le type de modification de séquence, similaires à des mutations spontanées qui interviennent au hasard sur le génome d'une plante ou des mutations provoquées par un agent génotoxique extérieur appliqué par l'homme (rayonnements ionisants, agents chimiques,...).

La technique SDN1 est généralement utilisée afin d'inactiver un gène précis puisque la réparation de l'ADN par NHEJ peut conduire à une rupture de l'enchaînement des éléments de l'ADN nécessaires à son expression. Dans certains cas, elle peut au contraire produire un gain de fonction du gène (par exemple en agissant sur les éléments régulateurs de son expression par inactivation d'un site répresseur de transcription) ou produire une modification de la fonction (si la mutation provoque un changement d'acide aminé d'une protéine permettant une modification fonctionnelle). Si le gène lui-même code un répresseur d'une voie (par exemple métabolique ou développementale), la perte de fonction du gène cause une activation de la voie. Pour que cette technique puisse effectivement relever *in fine* du SDN1 l'absence de transgène ou de séquence d'ADN exogène devra être confirmée.

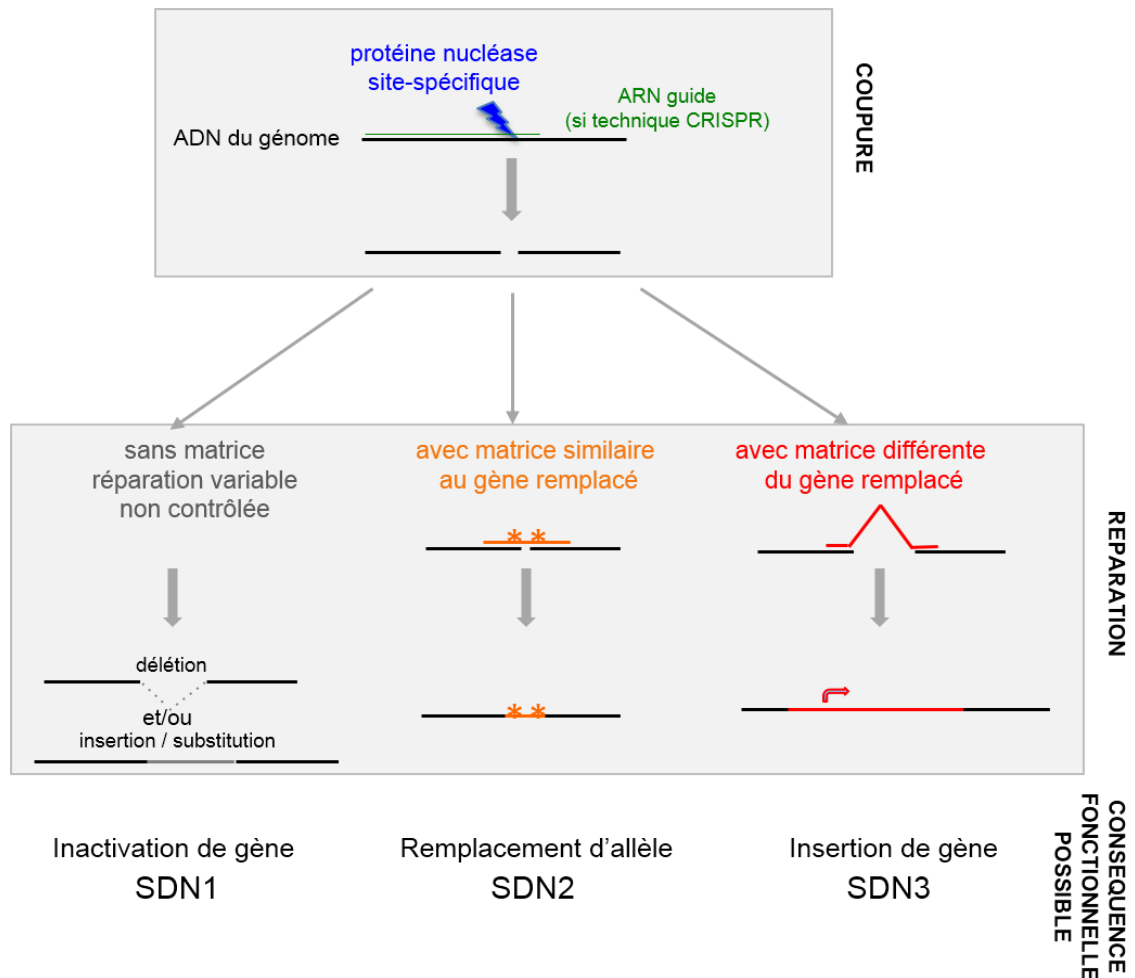


Figure 2 : Représentation schématique des différents types de modifications obtenus par les techniques utilisant des nucléases-spécifiques de sites, de type 1 (SDN1), 2 (SDN2) et 3 (SDN3).

SDN2 et SDN3

Les termes SDN2 et SDN3 désignent les techniques avec lesquelles l'expérimentateur induit une coupure double brin ciblée comme pour SDN1 et cherche en plus à utiliser le mécanisme de réparation pour insérer un élément d'ADN nouveau. L'expérimentateur fournit pour cela une matrice ADN « de réparation » dont les extrémités sont identiques, en séquence, aux deux régions adjacentes du site de coupure choisi. Elle peut être un oligonucléotide synthétique simple brin (longueur de l'ordre de 100 paires de bases) ou de l'ADN double brin (longueur pouvant atteindre plusieurs kilobases). Cette matrice peut être intégrée au génome par la machinerie de réparation de la cellule en utilisant un mécanisme de recombinaison homologue spontanée. Cependant, même en présence de cette matrice, la coupure peut aussi être réparée par NHEJ, comme toute coupure double brin de l'ADN. Ainsi, les techniques SDN2 et SDN3 conduisent à des événements qui doivent être triés par l'expérimentateur parmi d'autres, non recherchés, correspondant à des cassures réparées au hasard de type SDN1.

Les techniques désignées par le terme SDN2 utilisent une matrice ADN de réparation **similaire sur toute sa longueur à la séquence ciblée**, à l'exception **de courtes régions et/ou délétions et/ou insertions (d'une à quelques bases)** situées dans sa zone centrale (Fig. 2). Elles peuvent être utilisées pour obtenir des mutations ponctuelles (une ou plusieurs) **choisies** et des délétions (et insertions de petite taille) **contrôlées**. Un événement de mutation du même type (substitution, insertion, délétion) pourrait aussi être obtenu par mutagenèse aléatoire ou mutation naturelle.

On peut vouloir utiliser SDN2 dans plusieurs cas :

- La mutation est connue dans la diversité génétique naturelle d'une espèce et on souhaite l'introduire directement dans une variété de cette espèce ;
- La mutation est connue après mutagenèse au hasard dans une variété de plante et on souhaite l'introduire dans une autre variété (dans ces deux cas, la mutagenèse dirigée est une alternative aux rétrocroisements répétés avec la variété désirée) ;
- L'allèle du gène modifié n'est pas connu dans l'espèce, mais des données sur d'autres espèces de plante ou des expériences de laboratoire suggèrent que son introduction dans une variété de l'espèce présenterait un intérêt potentiel.
- On souhaite obtenir deux ou quelques mutations physiquement très proches (quelques nucléotides) au sein du même gène, ce qui serait très difficile à obtenir par mutagenèse aléatoire.

Pour que cette technique puisse effectivement relever *in fine* du SDN2 l'absence de transgène ou de séquence d'ADN exogène devra être confirmée.

Les techniques désignées par le terme SDN3 utilisent une **matrice de réparation dont la séquence ne ressemble pas à la séquence ciblée**, sauf aux extrémités utilisées pour promouvoir l'intégration par recombinaison homologue (Fig.2). Elles permettent notamment d'introduire un transgène (longueur de l'ordre de plusieurs kilobases) comme la « transgenèse classique » mais en différent par le fait que le site d'intégration du transgène au sein de l'ADN génomique est choisi par l'expérimentateur.

Multiplexage et obtention simultanée de plusieurs modifications spécifiques de sites

Les techniques relevant des SDN permettent d'effectuer simultanément plusieurs modifications spécifiques de sites différents.

Sur un locus :

Dans le cas de SDN2, par exemple, l'usage d'une « matrice de réparation » facilite une combinaison de plusieurs petites mutations/insertions/délétions choisies par l'expérimentateur sur un même locus. Il est donc possible de produire rapidement de nombreux allèles différents pour un même gène. Obtenir de telles modifications par mutagenèse au hasard ou mutation naturelle est d'occurrence statistiquement très faible et serait donc beaucoup plus fastidieux, voire pratiquement impossible pour le sélectionneur. De même l'utilisation de SDN3 permet l'insertion en un seul locus de plusieurs transgènes. Ceci représente d'une part un gain de temps pour le manipulateur et l'avantage que tous les transgènes seront transmis à la descendance comme un seul locus.

Sur de multiples locus, le multiplexage :

Plus généralement, les techniques SDN peuvent être appliquées simultanément sur plusieurs régions du génome par l'introduction de plusieurs nucléases et/ou ARN guides et de plusieurs matrices de réparation (on parle alors de multiplexage) (Raitskin and Patron, 2016). On obtient ainsi des modifications génétiques contrôlées dans plusieurs gènes ou séquences à la fois. Cela permet en particulier d'inactiver simultanément les différents allèles d'un gène dans des plantes polyploïdes (le blé hexaploïde, par exemple) (Wang et al., 2014), de cibler des familles de gènes, ou des gènes impliqués dans des voies métaboliques ou développementales différentes, ou encore de modifier des gènes très proches physiquement (liés génétiquement), dont il serait difficile d'associer les différents allèles par recombinaison naturelle dans la même plante via des croisements.

Limites entre SDN1, 2 et 3

La classification en SDN1, SDN2 ou SDN3 des diverses stratégies de modification du génome a fait l'objet de discussions. Le groupe s'accorde, dans le cadre de cette réflexion, sur le fait que :

- Le SDN2 se différencie de SDN1 par l'utilisation d'une matrice qui permet la réparation de la cassure causée par la nucléase. Le SDN2 découle de recombinaisons entre un gène cible et une version modifiée de ce gène apportée par une matrice ADN ou oligonucléotidique.
- SDN2 se différencie du SDN3 par le fait que, pour SDN2, le gène dont la modification est recherchée est déjà présent dans l'organisme, que son emplacement n'est pas modifié dans le génome, et que son nombre de copies n'est pas modifié non plus. On obtient par cette technique un allèle nouveau du gène ciblé.
- La limite n'est cependant pas aussi claire entre SDN2 et SDN3 dans le cas d'insertions car il n'est pas toujours possible de juger si une insertion génère un nouveau gène. En effet, de nouvelles séquences d'ADN peuvent conduire à la formation d'un nouvel ARN (même court, comme un petit ARN) ou à de nouvelles régulations d'expression d'un ARN. Le niveau de similarité requis pour distinguer SDN2 de SDN3 devra être défini au cas par cas si cela s'avérait nécessaire sur le plan réglementaire (cf. partie 7.).

Modalités de mise en œuvre

Les différents composants de ces **complexes effecteurs** peuvent être introduits dans une cellule de plante sous diverses formes biochimiques.

- **La nucléase** peut être introduite sous la forme d'un gène, c'est-à-dire sous la forme d'un ADN (qui sera ensuite transcrit en ARN et traduit en protéine par la cellule), d'un ARNm (qui sera traduit en protéine par la cellule) ou directement sous forme de protéine.
- **L'ARN guide de CRISPR** peut être introduit sous forme d'ADN (qui sera transcrit en ARN dans la cellule) ou directement sous forme d'ARN.
- La matrice de réparation, Dans le cas des SDN2/3 (fig. 2) est introduite sous forme d'oligonucléotide ADN simple brin synthétique (comme pour la technique de mutagenèse dirigée par un oligonucléotide (ODM, voir ci-dessous) ou d'ADN double brin. L'ADN double brin peut être obtenu par PCR ou synthèse chimique ou provenir d'une bactérie

hébergeant la matrice sous la forme d'un plasmide. Dans ce dernier cas, l'introduction de la matrice de réparation peut se faire soit par l'exposition des cellules de plante à la bactérie *Agrobacterium* portant le plasmide, soit par l'introduction du plasmide entier ou d'un fragment obtenu par digestion par des enzymes de restriction purifiées à partir de la bactérie d'hébergement du plasmide.

ZFN

La présence des deux protéines ZFN dans le noyau de la cellule que l'on souhaite modifier est nécessaire. La technologie repose actuellement sur le transfert de deux unités d'expression codant chacune une ZFN, sur le transfert des deux ARNm codant cette même paire de ZFN, voire sur le transfert des ZFN sous forme de protéines.

La taille des séquences codantes des ZFN (environ 1150 pb) a permis de tester de nombreux systèmes de vectorisation avec pour simple requis d'éviter le maintien de l'expression de la paire de ZFN dans la cellule afin de prévenir une toxicité liée à une continuelle activité endonucléase. Dans le cas où les ZFN seraient introduits de façon stable par transformation génétique, une fois la mutation générée, ces transgènes ne seraient plus utiles et, dans la plupart de cas, pourraient être éliminés par ségrégation. Cette élimination pourrait être validée par une analyse moléculaire simple et sensible de type séquençage.

TALEN

La présence des deux protéines TALEN dans le noyau de la cellule que l'on souhaite modifier est nécessaire. La technologie repose actuellement sur le transfert de deux unités d'expression codant chacune une TALEN, sur le transfert des deux ARNm codant cette même paire de TALEN ou sur l'injection des deux TALEN sous forme protéique.

La taille des séquences codantes des TALEN (environ 3kb) a permis de tester de nombreux systèmes de vectorisation (voir **tableau 1**) avec pour simple requis d'éviter le maintien de l'expression de la paire de TALEN dans la cellule afin de prévenir une toxicité liée à une activité endonucléase. Dans le cas où les TALEN seraient introduits de façon stable par transformation génétique, une fois la mutation générée, ces transgènes ne seraient plus utiles et, dans la plupart de cas, pourraient être éliminés par ségrégation. Cette élimination pourrait être validée par une analyse moléculaire simple et sensible de type séquençage.

CRISPR-Cas9

La question de la vectorisation est purement technique et fait l'objet de travaux qui visent à favoriser les systèmes transitoires.

La taille de la séquence codant la protéine Cas9 (plus de 1000 acides aminés pour la protéine Cas9 de *S. pyogenes* qui est la plus utilisée) limite l'utilisation des vecteurs de transfert viraux. Les techniques de transfection classiques avec des vecteurs plasmidiques sont favorisées dans l'attente de l'obtention de protéines Cas9 de plus petite taille. L'utilisation de système d'expression ADN de la protéine Cas9

ayant l'inconvénient de promouvoir une expression prolongée de cette protéine pour laquelle une action transitoire est suffisante, des techniques de transfection d'ARNm codant la protéine Cas9 ou l'administration de protéines Cas9 purifiées se sont également développées (Hwang et al., 2013).

En parallèle, il est possible, d'exprimer le sgRNA par un vecteur ADN, ou de les transférer directement sous forme d'ARN avec le système d'expression de la Cas9 choisi. (Cho et al., 2013; Hwang et al., 2013).

Le contrôle de la disparition ou de la modification des séquences ciblées dans les cellules produites est accessible par les techniques de séquençage classique et peut faire l'objet d'un référentiel de qualité.

Pour la transmission du système de modification CRISPR/Cas9 à la descendance cellulaire la méthode d'introduction est importante et fait l'objet de recherche : sous la forme d'ARNm, de sgRNA ou de protéine, il n'y a pas de transmission car la demi-vie de ces molécules est courte. L'utilisation de systèmes d'expression d'ADN, même non intégré, peut nécessiter une validation pour en vérifier la disparition. Les techniques de cette validation sont simples et sensibles.

Avantages et contraintes par rapport aux techniques existantes

ZFN

Le choix des motifs « en doigt de zinc » à utiliser pour construire une ZFN ayant des qualités d'hybridation spécifiques est complexe et repose sur une expertise qui nécessite souvent un recours à une société prestataire de service pour la fabrication des ZFN. Toutefois, des bases de données compilant les informations relatives aux ZFN fonctionnelles peuvent être trouvées sur Internet, de même que des sites proposant les algorithmes pertinents (exemple : <http://zifit.partners.org/ZiFiT/> - <http://eendb.zfgenetics.org/>). Par ailleurs, l'usage montre qu'il n'est pas possible avec cette technique de cibler toutes les séquences nucléiques.

TALEN

L'intérêt du système vient de ce qu'il est possible, par le biais d'une ingénierie protéique, de « programmer » la protéine pour qu'elle reconnaisse une séquence choisie. Cette technique est robuste mais nécessite une mise au point complexe comparée à d'autres techniques.

CRISPR-Cas9

Plusieurs avantages caractérisent ce système. D'une part, le complexe moléculaire chargé de couper l'ADN est composé de molécules dont l'ingénierie est aujourd'hui aisée. L'activité de coupure de l'ADN repose notamment sur l'utilisation d'une protéine unique pour toutes les applications et ne requérant aucune modification particulière liée à la cible d'ADN qu'elle va couper. D'autre part, l'activité ciblée de ce complexe dépend d'un appariement de l'ARN avec un ADN complémentaire (hybridation type Watson-Crick) reposant sur des interactions moléculaires facilement prédictibles permettant une ingénierie facile du mécanisme de ciblage du complexe.

Bien que l'interaction ARN/ADN soit spécifique la coupure dans des sites différents de la cible (qui auraient la même séquence ou une séquence similaire) est possible. Néanmoins, la technique est récente et de nombreuses études ont proposé des méthodes pour en augmenter la spécificité (Cho et al., 2013; Cradick et al., 2013; Fu et al., 2013; Hsu et al., 2013). Il est à signaler que de récentes études de séquençage indiquent que ces coupures non-spécifiques dites « *off-target* » sont rares (Smith et al., 2014; Suzuki et al., 2014; Veres et al., 2014).

En contre-point, de nombreuses stratégies sont développées afin de réduire les coupures non souhaitées dans le génome notamment en se focalisant sur la structure de l'acide nucléique *guide* qui est un élément d'influence majeure sur cette spécificité (Cho et al., 2014), ou sur l'analyse de certains mésappariements qui semblent tolérés (Semenova et al., 2011). Ainsi ont été testés plus de 700 variants de sgRNA en parallèle, permettant la mise au point d'outils informatiques mis à disposition de la communauté scientifique pour mieux choisir les ARN guides (Hsu et al., 2013).

En parallèle, des variants de la protéine Cas9 ont été produits afin de ne générer que des coupures simple brin (nickase), par mutation du domaine nucléase ou par remplacement de la fonction catalytique. La coupure double brin nécessite 2 sites physiquement proches, et donc deux sgRNA différents. Ceci limite significativement (de 50 à 1500 fois) les effets des coupures non souhaitées, qui sont physiologiquement réparées comme le sont les fréquentes ruptures de brin (review : (Bortesi and Fischer, 2015)).

Détection

(Plantes issues de SDN1/SDN2 ne contenant pas ou plus de séquences de nucléases et RNA guide insérées dans le génome ou présentes dans les cellules)

La modification est-elle détectable dans les plantes porteuses de la modification et leurs produits ?

Les techniques SDN1/SDN2 induisent des mutations dans un gène endogène, mais les plantes obtenues n'ont pas de transgène complet ajouté.

Les mutations induites, **si elles sont décrites dans un document d'accompagnement**, peuvent être tracées phénotypiquement (plus ou moins facilement en fonction du type de phénotype) et/ou génétiquement. Avec une documentation de la modification il est donc possible de réaliser des tests de détection génétique à l'aide de techniques de biologie moléculaire facilement réalisables et basées sur la PCR ou des tests de détection phénotypiques (dans le cas de la modification d'un caractère).

Si ces mutations induites ne sont pas décrites, il est théoriquement possible d'observer la mutation par l'utilisation d'une technique de séquençage. Se pose alors la difficulté du choix de la séquence de comparaison. En effet entre deux plantes de la même variété, il existe de très nombreuses différences de séquences et il sera, en pratique, biologiquement impossible de différencier une mutation obtenue par SDN1 ou SDN2 d'une mutation naturelle.

La détection de la modification dans les produits issus d'une plante ayant été engagée dans des processus relevant d'un SDN1 ou d'un SDN2 s'inscrit dans cette même problématique si le produit contient de l'ADN et dans celle évoquée dans le cadre SDN3 pour les produits ne contenant pas d'ADN.

Peut-on identifier la modification et la technique ayant donné lieu à cette modification dans la plante porteuse de la modification et ses produits ?

Avec une documentation de la modification il est possible et techniquement facile d'identifier la modification et éventuellement de l'associer à la technique.

Sans documentation de la modification et/ou de la méthode utilisée, il est difficile de différencier de manière systématique et parfaitement fiable, en l'état actuel, une plante obtenue par SDN1/2 d'une plante obtenue par des techniques de sélection (exploitant la variabilité naturelle ou induite par mutagenèse). Bien que la modification puisse parfois être elle-même identifiable si on la cherche, elle ne pourrait pas à l'heure actuelle être facilement imputée à l'emploi de l'une ou l'autre de ces techniques ou méthodes. Il est possible que des travaux de recherche sur les effets de diverses techniques, y compris de vectorisation, permettent dans le futur d'identifier la technique utilisée. Un faisceau de preuves découlant de la vectorisation et/ou des modifications génomiques et phénotypiques de la plante pourra alors fournir un cadre de décision.

La question d'un profil de mutations hors cibles spécifique dues à l'utilisation de CRISPR et permettant indirectement de remonter à celles-ci a été discutée. Lors d'auditions le GT a abordé cette question avec J.P. Concordet². Dans l'état actuel des connaissances, s'il est envisageable d'entreprendre de déterminer si un tel profil existe, il ne semble pas possible de l'utiliser comme méthode d'identification de la technique utilisée. En effet, un tel profil, s'il existait, dépendrait de la séquence ciblée par les CRISPR. Et s'il venait à être mis en évidence, il n'est pas clair qu'il puisse constituer une preuve univoque de l'origine du variant génétique via une technique CRISPR à l'exclusion de toute autre occurrence spontanée, mais il pourrait participer à un faisceau de preuves recherchées par les autorités compétentes.

Le GT a discuté de la possibilité d'utiliser les séquences PAM (utilisées spécifiquement dans la technologie CRISPR) aux alentours des mutations détectées pour pouvoir discriminer entre une mutation naturelle ou induite (chimiquement ou par radiation) et une mutation produite par l'utilisation de CRISPR. A l'heure actuelle, ces séquences PAM sont caractérisées principalement par la présence d'un dinucléotide GG à l'extrémité 3' de la séquence cible. Dans le cas de SDN1, la nucléase coupant environ 3 bases en amont de la séquence GG, l'extrémité 3' de la séquence PAM doit se trouver à l'intérieur ou à proximité immédiate de la modification obtenue dans le génome d'origine. Cette séquence peut, de plus, se retrouver éloignée ou être perdue du fait même de la mutation induite. De plus, la succession de deux guanines (G) se produit statistiquement de nombreuses fois par gène, donc très fréquemment dans le génome et ne peut être utilisée comme marqueur de la technique CRISPR, à l'exclusion d'une autre. Enfin, il émerge un grand nombre de versions différentes de la nucléase utilisée, n'utilisant plus forcément la séquence GG. Ceci rend pratiquement impossible l'identification de toutes les séquences PAM potentielles. La présence de GG ne fournira donc pas une preuve univoque que l'utilisation de la technique est à l'origine de la mutation observée. De plus, de nouvelles méthodes s'affranchissent de ces motifs PAM (Gao et al., 2016; Kleinstiver et al., 2015).

Le GT a jugé souhaitable (et possible avec l'utilisation de séquençages génomiques) d'obtenir dans les prochaines années plus d'informations sur les éventuelles spécificités des profils de modifications

² Chercheur INSERM U1154, CNRS UMR7196, MNHN auditionné par le GT pour son expertise sur la prévision des coupures hors cibles CRISPR/Cas9

spécifiques des SDN d'une part et de la transformation par *Agrobacterium* ou d'autres méthodes d'introduction des composants SDN, d'autre part.

Le lien entre la modification et la technique utilisée dans le cadre de l'analyse du produit s'inscrit dans cette même problématique si le produit contient de l'ADN et dans celle évoquée dans le cadre SDN3 si les produits ne contiennent pas d'ADN.

Si une détection est possible, est-elle applicable dans le cadre d'une recherche de présence fortuite en situation de mélange plus ou moins complexe au cours du processus de production, de transformation et de distribution ?

S'il n'existe pas de déclaration de la modification, compte tenu de la difficulté d'identification de la technique expliquée plus haut dans le cas d'une plante donnée, il devient encore plus difficile voire totalement impossible dans le cas d'une présence fortuite éventuelle (donc à très faible taux), de déterminer biologiquement de façon formelle l'origine du variant génétique (plante obtenue par SDN1/2/ODM ou variant naturel).

Si la modification génétique est déclarée et si l'objectif est de **détecter la présence** de cette modification, il est alors possible de développer des méthodes d'analyse de présence dans un ensemble (pool) de plantes ou de graines. Les techniques disponibles se basent sur le **séquençage** de la région du génome correspondante ou l'utilisation de marqueurs moléculaires spécifiques de la mutation. Dans le cas particulier de variétés rendues tolérantes à un herbicide, l'utilisation de l'herbicide sur un échantillon du champ mis en culture permet d'identifier la présence du trait. Ces méthodes pourraient alors être mises en place dans le cadre de recherche de présence fortuite.

Si la modification génétique est déclarée et si l'objectif est de **prouver une propriété intellectuelle éventuelle**, la situation est compliquée par le fait que ces mutations peuvent aussi être produites par d'autres méthodes (variants naturels, mutagenèse non dirigée).

La probabilité de co-occurrence des mêmes variants génétiques indépendamment par deux méthodes est liée au nombre et au type de mutations. Il faudrait dans ce cas établir des seuils de probabilité de co-occurrence pour chaque cas, ce qui serait très difficile à mettre en place. Dans un raisonnement similaire, l'environnement génétique autour de la mutation entre autres données pourrait être évalué pour qualifier une plante, mais dans l'état actuel des connaissances la preuve biologique de l'origine du variant resterait difficile à mettre en évidence.

Dans le cas du SDN2, le GT s'est interrogé sur la possibilité offerte par la technique de proposer un « étiquetage moléculaire » en introduisant un certain nombre de **mutations silencieuses**³ revendiquées par l'obtenteur comme marquant son obtention dans le cadre d'un brevet, par exemple. Si une telle étiquette moléculaire était mise en place, elle permettrait une traçabilité des plantes issues de SDN2. Le GT note que de telles mutations silencieuses peuvent avoir des effets non contrôlés, bien que peu probables, par exemple en agissant sur l'épissage, la stabilité ou la vitesse de traduction de l'ARN messager (Sauna and Kimchi-Sarfaty, 2011). Un marquage hors de l'exome si cela est possible serait judicieux.

³ Mutation de la séquence ADN ne changeant pas la séquence d'acides aminés codée par le gène et sans insertion de nouvelles bases dans le génome

Détection et traçabilité des méthodes des plantes issues de SDN3

La modification est-elle détectable dans les plantes porteuses de la modification et leurs produits ?

La technique SDN3 introduit un transgène qui peut être ou non étranger à l'espèce. Le cas est proche des OGM classiques, sauf que l'insertion du transgène en question ne se fait pas entre des pieds de T-DNA d'*Agrobacterium*.

Les modifications induites, **si elles sont décrites dans un document d'accompagnement**, peuvent être tracées génétiquement et éventuellement phénotypiquement (plus ou moins facilement suivant le phénotype).

Avec un document d'accompagnement, il est possible de réaliser des tests de détection génétique à l'aide de techniques de biologie moléculaire de routine (techniques de PCR associées ou non au séquençage de l'ADN), basées sur la description de la modification, ou des tests de détection du phénotype induit par cette modification. La détection phénotypique est plus ou moins aisée selon la fonctionnalité qu'il est nécessaire de détecter.

Sans document d'accompagnement, L'insertion par SDN3 de transgènes pour lesquels aucune information n'est disponible, que ce soit au niveau de la séquence ou de la protéine recombinante produite, ne peut qu'être détectée par screening, et encore seulement si certains des éléments génétiques présents dans la construction utilisée sont listés dans les bases de données disponibles. Une stratégie exhaustive d'analyse de la totalité du génome de la plante porteuse de la modification avec une comparaison, également exhaustive, des résultats de séquençage avec le génome des autres espèces peut être théoriquement entreprise. Cependant, cette stratégie demande un savoir-faire sur l'analyse de séquençage à haut débit et des banques de données avec les génomes de référence et les transgènes couramment utilisés que l'on ne peut pas aujourd'hui considérer comme courant dans un tel objectif. De plus, dans les cas de cisgénèse par SDN3, certains transgènes non préalablement connus pourraient ne pas être détectés à cause de l'imprécision actuelle de l'annotation des génomes en gènes et à cause de la diversité génétique de chaque espèce qui comprend des délétions et insertions naturelles de gènes. En ce qui concerne les produits, si ces produits contiennent de l'ADN, il est possible d'appliquer la même stratégie mais avec les mêmes limites. Si les produits ne contiennent pas d'ADN, la détection de différences que l'on puisse attribuer à une modification du génome de la plante porteuse n'est pas toujours possible avec les techniques et données actuelles, et devra de toute façon être confirmée par une analyse génétique. Dans cette problématique, toute information minime sur des propriétés atypiques relatives à la plante et/ou ses produits peut apporter des indices permettant d'orienter et faciliter les analyses. Ainsi, cette approche est facilitée dans le cas où la séquence exogène est connue par ailleurs (par exemple, toxine Bt) (Holst-Jensen et al., 2016). Cependant, au regard de la complexité des génomes et de leur variabilité, et en prenant en compte les techniques et données les plus récentes, il reste difficile d'assurer la détection d'une modification sans document d'accompagnement.

Peut-on identifier la modification et la technique ayant donné lieu à cette modification dans la plante porteuse de la modification et ses produits ?

Avec un document d'accompagnement, il est possible de mettre en place des techniques simples d'analyse de la séquence d'ADN de la plante, ou du produit s'il contient de l'ADN, pour retrouver ces éléments spécifiques et associer la modification à la technique qui l'a générée. Les produits issus de ces plantes mais ne contenant plus d'ADN ne pourraient pas être identifiés.

Sans documentation permettant de connaître la méthode utilisée et tout en considérant des problèmes cités précédemment lors de l'analyse d'une modification sans information, il est possible de différencier une plante obtenue par SDN3 d'une plante obtenue :

- Par sélection classique par exemple s'il est détecté un transgène exogène connu comme celui codant une toxine Bt.
- Par transgénèse classique obtenue par l'utilisation d'agrobactéries en identifiant les traces de bordures de l'ADN-T. Cependant, celles-ci peuvent ne pas être présentes et il ne sera pas possible dans ce cas de conclure à l'utilisation de SDN3 à l'exception de toute autre méthode de transgénèse (biolistique, agrobactéries, transformation de protoplastes) sans l'aide de SDN.
- Dans le cas de la cisgénèse ou de l'intragenèse obtenue par SDN3, se reporter à ce paragraphe.

Si une détection est possible, est-elle applicable dans le cadre d'une recherche de présence fortuite en situation de mélange plus ou moins complexe au cours du processus de production, de transformation et de distribution ?

Si une description documentaire de la modification existe, il est techniquement et financièrement possible de faire une détection ciblée dans des échantillons prélevés au champ et des matrices alimentaires.

S'il n'existe pas de documentation sur la modification moléculaire et/ou phénotypique, la recherche d'éléments de constructions transgéniques à l'aveugle dans un produit SDN3 peut être réalisée comme dans le cas d'OGMs classiques (voir en 3.1) et en considérant les mêmes limites que celles évoquées plus haut dans le cadre d'une recherche sans information et en ajoutant le problème crucial de la sensibilité de l'analyse. Le fait que la modification ait été réalisée à l'aide de SDNs rend cependant inutilisables les tests événement-spécifiques basés sur l'amplification des sites d'insertion des transgènes. La difficulté d'analyses est accrue si les éléments des constructions utilisées relèvent de l'espèce transformée (cisgénèse, intragenèse). Il est à noter que dans ce cas de recherche de présence fortuite le séquençage de génome entier reste encore inenvisageable pour le contrôle en routine pour des raisons de coût, de délai, dans certains cas de disponibilités des données de référence, et pour des raisons de sensibilité de détection.

Bibliographie

- Auer, T.O., and Del Bene, F. (2014). CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. *Methods San Diego Calif* 69, 142–150.
- Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33, 41–52.
- Cho, S.W., Kim, S., Kim, J.M., and Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 230–232.
- Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.-S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24, 132–141.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823.
- Cradick, T.J., Fine, E.J., Antico, C.J., and Bao, G. (2013). CRISPR/Cas9 systems targeting β -globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* 41, 9584–9592.
- Duca, M., Vekhoff, P., Oussedik, K., Halby, L., and Arimondo, P.B. (2008). The triple helix: 50 years later, the outcome. *Nucleic Acids Res.* 36, 5123–5138.
- Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822–826.
- Gaj, T., Gersbach, C.A., and Barbas, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405.
- Gao, F., Shen, X.Z., Jiang, F., Wu, Y., and Han, C. (2016). DNA-guided genome editing using the *Natronobacterium gregoryi* Argonaute. *Nat. Biotechnol.*
- Hisano, Y., Sakuma, T., Nakade, S., Ohga, R., Ota, S., Okamoto, H., Yamamoto, T., and Kawahara, A. (2015). Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Sci. Rep.* 5, 8841.
- Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.* 29, 731–734.
- Holst-Jensen, A., Spilsberg, B., Arulandhu, A.J., Kok, E., Shi, J., and Zel, J. (2016). Application of whole genome shotgun sequencing for detection and characterization of genetically modified organisms and derived products. *Anal. Bioanal. Chem.* 408, 4595–4614.
- Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832.
- Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Kaini, P., Sander, J.D., Joung, J.K., Peterson, R.T., and Yeh, J.-R.J. (2013). Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. *PLoS ONE* 8, e68708.
- Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1156–1160.

Kimura, Y., Hisano, Y., Kawahara, A., and Higashijima, S. (2014). Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. *Sci. Rep.* *4*, 6545.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P.W., Li, Z., Peterson, R.T., Yeh, J.-R.J., et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* *523*, 481–485.

Mali, P., Esvelt, K.M., and Church, G.M. (2013a). Cas9 as a versatile tool for engineering biology. *Nat. Methods* *10*, 957–963.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-guided human genome engineering via Cas9. *Science* *339*, 823–826.

Moehle, E.A., Moehle, E.A., Rock, J.M., Rock, J.M., Lee, Y.-L., Lee, Y.L., Jouvenot, Y., Jouvenot, Y., DeKolver, R.C., Dekolver, R.C., et al. (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 3055–3060.

Mussolino, C., and Cathomen, T. (2012). TALE nucleases: tailored genome engineering made easy. *Curr. Opin. Biotechnol.* *23*, 644–650.

Raitskin, O., and Patron, N.J. (2016). Multi-gene engineering in plants with RNA-guided Cas9 nuclease. *Curr. Opin. Biotechnol.* *37*, 69–75.

Richter, H., Randau, L., and Plagens, A. (2013). Exploiting CRISPR/Cas: Interference Mechanisms and Applications. *Int. J. Mol. Sci.* *14*, 14518–14531.

Sauna, Z.E., and Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* *12*, 683–691.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 10098–10103.

Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L., et al. (2014). Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* *15*, 12–13.

Suzuki, K., Yu, C., Qu, J., Li, M., Yao, X., Yuan, T., Goebel, A., Tang, S., Ren, R., Aizawa, E., et al. (2014). Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. *Cell Stem Cell* *15*, 31–36.

Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* *15*, 27–30.

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

