



HAL
open science

Évaluation des outils d'édition du génome pour les applications biomédicales : du diagnostic au traitement

Delphine Leclerc

► **To cite this version:**

Delphine Leclerc. Évaluation des outils d'édition du génome pour les applications biomédicales : du diagnostic au traitement. Médecine humaine et pathologie. Université de Rennes, 2023. Français. NNT : 2023URENB025 . tel-04300637

HAL Id: tel-04300637

<https://theses.hal.science/tel-04300637>

Submitted on 22 Nov 2023

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

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

THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES

ECOLE DOCTORALE N° 605
SCIENCE DE LA VIE ET DE LA SANTE
SPECIALITE : GENETIQUE, GENOMIQUE, BIOINFORMATIQUE

Par

Delphine LECLERC

**Évaluation des outils d'édition du génome pour les applications biomédicales :
du diagnostic au traitement**

Thèse présentée et soutenue à Rennes, le 16 Juin 2023

Unité de recherche : INSERM U1242 – Oncogenesis, Stress & Signalling (OSS)

Rapporteurs avant soutenance :

Annarita MICCIO DR INSERM, Institut IMAGINE, Paris
Els VERHOEYEN DR INSERM, Centre Méditerranéen de Médecine Moléculaire, Nice

Composition du Jury :

Président :

Eric OGIER-DENIS DR INSERM, Centre Eugène Marquis, Rennes

Examineurs :

Stéphane BEZIEAU PU-PH, CHU de Nantes, Nantes
Béatrice TURCQ CR CNRS, Université de Bordeaux, Bordeaux
Annarita MICCIO DR INSERM, Institut IMAGINE, Paris
Els VERHOEYEN DR INSERM, Centre Méditerranéen de Médecine Moléculaire, Nice

Dir. de thèse :

Frédéric MOURIAUX PU-PH, CHU de Rennes, Rennes

Co-dir. de thèse :

David GILOT MCU, Université de Rennes, Rennes

REMERCIEMENTS

En premier lieu, je tiens à remercier le Ministère de l'Enseignement Supérieur, de la Recherche, et de l'Innovation d'avoir financé ces trois belles années de thèse. Je suis également très reconnaissante envers les organismes financeurs des multiples projets auxquels j'ai participé : la Ligue nationale contre le cancer (et ses donateurs), AVIESAN Plan Cancer, l'Agence Nationale de la Recherche et le Cancéropôle Grand-Ouest. Votre engagement en faveur de la recherche nous est indispensable pour mener à bien nos projets.

Je remercie chaleureusement la fondation L'Oréal-UNESCO pour m'avoir décerné le prix Jeune Talent France 2022 Pour les Femmes et la Science. J'ai passé une semaine incroyable et riche en émotions en présence de 34 autres jeunes femmes que je salue au passage. Cette expérience fut mémorable et les retombées du prix m'ont offert une visibilité incroyable. Cette reconnaissance me donne aujourd'hui encore plus de force, de confiance et de volonté pour poursuivre ma carrière académique.

Merci aussi à L'Université de Rennes, aux organisateurs et aux financeurs du programme Ma Thèse en 180 secondes, grâce à qui j'ai pu remporter cette année le premier prix du Jury de la finale régionale Bretonne.

Je remercie Dr Annarita MICCIO et Dr Els VERHOEYEN d'avoir évalué bénévolement mon travail en qualité de rapporteurs. Je remercie aussi Pr Stéphane BEZIEAU, Dr Béatrice TURCQ et Dr Eric OGIER-DENIS d'avoir accepté d'être examinateurs de cette thèse.

Monsieur MOURIAUX, merci d'avoir accepté d'être le directeur de cette thèse. Vous avez toujours été de bons conseils avec moi. J'ai pu assister en direct et pour la première fois de ma vie à une opération médicale, au bloc, à vos côtés. Cette expérience me rappelle l'importance de la responsabilité éthique du métier de scientifique, puisque chaque échantillon biologique que je manipule provient d'un patient. Même si mon projet de thèse initial, qui portait sur le mélanome uvéal, n'a pas abouti, vous avez quand même continué à m'encourager dans les nouveaux projets que j'ai entrepris.

David, c'est à toi que reviennent les remerciements les plus importants. Sans toi, rien de tout de travail de thèse n'aurait vu le jour. Ta passion pour la recherche, ta rigueur scientifique et ton professionnalisme ont été pour moi des modèles à suivre, m'encourageant constamment à donner le meilleur de moi-même. Quand je me retrouve face à un nouveau défi scientifique, je me demande ce que tu aurais fait à ma place et je me rappelle toujours à quel point il est important de croire en ses propres idées, de prendre des risques, de garder la motivation et de ne pas abandonner. Tu as su t'adapter parfaitement à mes difficultés pour me faire progresser et tu m'as toujours soutenue lorsque j'ai souhaité postuler à des appels à candidature. Tu t'es engagé jusqu'au bout et même au-delà en soutenant ma candidature en post-doctorat à San Diego. Je peux dorénavant partir en toute confiance pour relever de nouveaux défis. On m'a très souvent répété à quel point j'avais de la chance de travailler sous ta supervision. Merci pour ton engagement, ta confiance et ta bienveillance durant ces trois années.

Parce que, la recherche c'est aussi un travail d'équipe ! Jamais je n'aurai pu initier ces travaux sans l'expertise clinique des médecins et biologistes du CHU de Rennes. Grâce aux liens étroits tissés entre les médecins, les biologistes, et les chercheurs, de beaux projets ont été réalisés : merci Pr Sylvie ODENT, Pr Marc-Antoine BELAUD-ROTUREAU, Pr Sylvie JAILLARD, Dr Cristèle DUBOURG, Dr Laura MARY et Dr Louise GOUJON.

Gregory et Antti, je vous remercie pour votre soutien au quotidien. Vous êtes de très chouettes collègues, très rigoureux tout en étant capables de doser avec une touche d'humour. Cela m'a beaucoup aidé à prendre du recul et à me redonner le sourire, de l'espoir et de l'énergie dans les moments les plus difficiles. Ces derniers mois n'ont pas été des plus simples, et vous avez su incroyablement bien rebondir. Je vous souhaite le meilleur pour la suite, vos projets sont vraiment géniaux et vous êtes sur la très bonne voie pour réussir.

Je salue également mes camarades de paillasse du bâtiment 15 : Panos, Corentin, Matthis, Tanguy et Raffaele. Merci à vous pour votre énergie positive, toutes les petites discussions partagées entre deux manips, et votre aide pour les (multiples) pannes techniques auxquelles j'ai été confrontée.

Merci à tous les membres de l'équipe et à OSS de m'avoir soutenue durant ces 3 années. Cédric, je te remercie de m'avoir initiée à la rédaction scientifique et au processus de révision. J'ai rédigé mon premier article sous ta supervision et j'en suis très satisfaite ! Éric, merci d'avoir appuyé ma candidature en post-doctorat et d'avoir encadré bénévolement ma fin de thèse. Tony et Elodie, merci de m'avoir acceptée dans votre équipe jusqu'à la fin. Je profite de l'occasion pour souligner le travail remarquable des ingénieurs, techniciens et administratifs de l'unité qui assurent tous, à leur échelle, le bon fonctionnement du laboratoire. Votre travail au quotidien est essentiel pour nous protéger, nous apprendre les bons gestes et mener nos recherches dans de bonnes conditions.

J'aimerais remercier également les experts de la technologie de séquençage : Stéphane qui a lu pas moins d'un millier de réactions au séquenceur capillaire, Amandine qui m'a gentiment aidée à charger mes premières puces sur la plate-forme de séquençage haut-débit de l'hôpital et Bénédicte qui a formaté et analysé l'ensemble de mes données brutes.

Un grand merci aussi à Raphaël PINEAU qui m'a formé à l'expérimentation animale, plus particulièrement aux expériences de xénogreffe, tu es un atout majeur pour l'unité !

Et jamais je n'aurai pu achever aussi rapidement ce travail sans l'aide technique précieuse de Laurence et Flavie. Merci d'être restées à mes côtés parfois tard le soir, pour assurer le bon déroulement des expériences.

Depuis mon enfance et jusqu'à mon master, j'ai toujours été la plus réservée de la classe, et cela m'a valu beaucoup de doutes sur mes capacités à entamer une thèse. Durant ces trois années, j'ai pu acquérir de nombreuses compétences en communication et je me suis affirmée tant pour :

- L'encadrement des stagiaires,
- La supervision des techniciens (parfois beaucoup plus âgés que moi),
- Les travaux dirigés devant les étudiants en licence et master,
- La vulgarisation scientifique (grand public, lycéens et journalistes)
- Les présentations orales devant les experts scientifiques.

Je remercie toutes les personnes qui m'ont encouragée dans cette direction pour franchir ce cap, c'est exactement ce dont j'avais besoin pour prendre confiance en moi : David GILOT, Frédéric MOURIAUX, Stéphanie LE BRAS, Laurent RICHARD PARPAILLON, Michèle LESIMPLE, Virginie VALLET, Pascale QUIGNON, Stéphane DESCHAMPS, Cédric MENARD, Julien LE BONHEUR, Sarah KITAR, Valérie SAUVAGE, toute l'équipe *For Women In Science*, Charlotte KAN et Tatiana HAEN. Cette dynamique a également été insufflée par ma passion pour la recherche, aussi je suis convaincue que l'enthousiasme est un moteur pour atteindre des objectifs ambitieux – et que l'on peut tous s'améliorer en trouvant notre moteur.

J'adresse aussi mes remerciements à Thomas DERRIEN et Aurélien SERANDOUR pour avoir gentiment accepté de faire partie de mon comité de thèse. Vos remarques pertinentes et encouragements positifs m'ont également aidé à prendre confiance en moi.

Un salut chaleureux à toute l'équipe du laboratoire de cytogénétique et biologie cellulaire du CHU de Rennes, et surtout aux techniciennes qui ont démarré à nos côtés l'activité EPIQUE : Sylvia, Delphine et Muriel. Un clin d'œil à tous les (nombreux) stagiaires dont j'ai supervisé techniquement le travail : Lucie, Charlotte, Diane, Tiffany, Valentin, Louise, Flavie et Chloé. Vous avez tous été géniaux avec moi, vous m'avez beaucoup apporté sur le plan personnel et professionnel. Je me suis découverte en tant que manageuse, pas si évident alors que je n'y avais jamais été confrontée auparavant.

Un clin d'œil également à tous mes amis de promo qui passent par ici : Priscille, Élise, Pauline, Marine, Fred, Margot, Églantine, Baptiste, Ilona, Aurélia, Manon, Blandine, Clément, Hélène et Justine.

Merci Maxence pour ton soutien et tes conseils très judicieux, je deviens une meilleure personne un peu plus chaque jour grâce à toi.

Pour terminer, je tiens à remercier mes parents, jamais je n'aurai pu arriver à ce niveau sans votre aide. Merci papa d'assurer tous mes déplacements quand je rentre à la maison. Surtout, merci maman, tu me pousses à donner le meilleur de moi-même depuis mon plus jeune âge. Désolée maman si tu n'arrives pas à relire ma thèse parce que « c'est de l'hébreux », cela ne t'enlèvera en rien toute ta volonté bienveillante. J'espère que tu comprendras que pour la suite de ma carrière j'ai besoin d'explorer de nouveaux horizons en dehors de la France. Je me protégerai et redoublerai de vigilance, c'est promis.

TABLE DES MATIERES

REMERCIEMENTS	1
TABLE DES MATIERES	6
LISTE DES FIGURES	8
LISTE DES TABLEAUX	9
LISTE DES ABBREVIATIONS	10
PARTIE I : INTRODUCTION	12
I. Les outils d'édition du génome.....	13
1. Historique : premières démonstrations d'édition ciblée du génome	13
2. Méganucléases et cassures double brins ciblées de l'ADN.....	14
3. Nucléases de l'ADN programmables.....	15
3.1 Nucléases à doigt de zinc.....	15
3.2 Nucléases effectrices de type activateur de transcription	16
4. Émergence des outils CRISPR/Cas.....	18
4.1 Découverte du système CRISPR/Cas chez la bactérie	18
4.2 Vers la création d'un outil d'édition du génome programmable	21
4.3 Réparation des cassures double brins induites par CRISPR/cas9	22
4.4 Augmentation du ratio HDR/NHEJ.....	23
4.5 Coupures en dehors de la cible.....	25
4.6 Autres effets indésirables associés aux cassures double brins de l'ADN	35
5. Réingénierie des outils CRISPR/Cas pour l'édition génétique de précision sans CDB.....	37
6. CRISPR 2.0 : les éditeurs de bases cytosine et adénine	37
6.1 Origine des éditeurs de bases.....	37
6.2 Challenges associés aux éditeurs de bases	41
7. CRISPR 3.0 : le <i>prime editing</i>	49
7.1 Mécanisme du PE	49
7.2 Origine des différentes versions PE	51
7.3 Améliorer l'efficacité du PE	51
7.4 Quantifier les modifications hors de la cible du PE	56
7.5 Modifications non désirées	57
7.6 Nouvelles versions PE	59
8. Transposases et recombinaisons	61
9. Résumé : les outils d'édition du génome	66
10. Autres outils CRISPR/Cas.....	67

II. Applications biomédicales des outils d'édition du génome CRISPR/Cas9	69
1. Edition du génome pour la thérapie génique	69
1.1 Contexte et définition	69
1.2 Limites de la thérapie génique additive	69
1.3 Thérapie génique par édition de l'ADN	75
2. Edition du génome pour l'aide au diagnostique	87
2.1 Interprétation biologique des variants génétiques	87
2.2 Développement tests fonctionnels pour la reclassification des VSI	90
PARTIE II : HYPOTHESES ET OBJECTIFS	93
PARTIE III : RESULTATS	99
RESULTAT 1	100
RESULTAT 2	123
RESULTAT 3	144
RESULTAT 4	172
PARTIE IV : DISCUSSION	205
BIBLIOGRAPHIE	214
ANNEXES	228
ANNEXE I : Mécanismes de réparation des cassures double brin de l'ADN	229
ANNEXE II : Résumés d'articles collaboratifs	232
ANNEXE III : Formations suivies	235
ANNEXE IV : Missions enseignement	239
ANNEXE V : Congrès	242
ANNEXE VI : Récompenses	245

LISTE DES FIGURES

Figure 1 : Mécanisme d'action des ZFNs et TALENs	17
Figure 2 : Fonctionnement du système immunitaire CRISPR/Cas9 de <i>Streptococcus pyogenes</i>	20
Figure 3 : Sites de coupure hors cibles du candidat thérapeutique NTLA-2001.	29
Figure 4 : Nouvelles architectures CRISPR/Cas9 de haute spécificité.	32
Figure 5 : Mécanisme d'action des éditeurs de bases.	40
Figure 6 : Les différents types d'éditeurs de bases.	49
Figure 7 : Mécanisme d'action du <i>prime editing</i>	50
Figure 8 : Efficacité et pureté du PE3 au locus <i>tyr-1</i> dans des embryons de poissons zèbres.	58
Figure 9 : Nouvelles versions PE utilisant des paires de pegRNA.	60
Figure 10 : Mécanisme d'action des tyrosines recombinases.	61
Figure 11 : Preuve de concept et modèle du système ShCAST.	64
Figure 12 : Vue d'ensemble des outils d'édition du génome.	66
Figure 13 : Physiopathologie de l'amyloïdose héréditaire à transthyrétine (ATTR).	77
Figure 14 : Distribution des variants génétiques délétères humains.	78
Figure 15 : Le répresseur transcriptionnel BCL11A contrôle le passage de l'hémoglobine foétale HBF à l'hémoglobine adulte HBA.	81
Figure 16 : Représentation schématique des particules de type virales (VLP) transportant des cargos ARN ou des complexes RNP.	84
Figure 17 : Nombre et type de variants pathogènes atteignables par PE.	86
Figure 18 : Les différentes étapes du diagnostic génétique.	88
Figure 19 : Test fonctionnel basé sur la résistance à la méthylation pour la classification des variants de signification inconnue localisés dans le gène <i>MLH1</i>	91
Figure 20 : Objectifs de la plateforme EPIQUE.	94
Figure 21 : Caractérisation des cellules cancéreuses responsables de l'agressivité du mélanome uvéal.	97
Figure 22 : Thérapie ultra-personnalisée <i>versus</i> thérapie pan-mutation par édition génétique.	208
Figure 23 : Les voies de réparation de l'ADN dans les cellules mammaliennes.	230

LISTE DES TABLEAUX

Tableau 1 : Ingénierie de variants Cas9 haute-fidélité.....	35
Tableau 2 : Taille approximative de la fenêtre d'activité de quelques éditeurs de bases.	42
Tableau 3 : Les différents types d'architectures PE.....	53
Tableau 4 : Principales caractéristiques des vecteurs viraux utilisés pour la thérapie génique.....	70
Tableau 5 : Thérapies géniques acceptées aux Etats-Unis.....	72
Tableau 6 : Candidats thérapeutiques reposant sur la technologie BE.	79

LISTE DES ABBREVIATIONS

A	Adénine
AAV	Virus adénoassocié
ABE	Éditeur de base adénine
ADN	Acide desoxyribonucléique
alt-NHEJ	Jonction des extrémités non homologues alternative
ARN	Acide ribonucléique
ARNi	ARN interférence
ARNm	Acide ribonucléique messenger
ATTR	Amyloïdose à transthyrétine
BER	Réparation par excision de base
C	Cytosine
c-NHEJ	Jonction des extrémités non homologues classique
CAR	Récepteur antigénique chimérique
CAST	<i>CRISPR-associated transposase</i>
CBE	Éditeur de base cytosine
CDB	Cassure double brin
CDS	Séquence d'ADN codante
CRISPR	<i>Clustered regularly interspaced short palindromic repeats</i>
CRISPR/KO	CRISPR <i>knock-out</i>
CRISPRa	CRISPR activateur
crRNA	ARN CRISPR
dCas9	Cas9 déficiente
FACS	Tri cellulaire activé par la fluorescence
FDA	<i>Food and Drug Administration</i>
G	Guanine
HDR	Recombinaison homologue dirigée
HELIX	<i>Homing endonuclease-assisted large-sequence integrating CAST complex</i>
INDELS	Insertions et délétions
IPE	<i>Impure Prime Edit</i>
IV	Injection intraveineuse
Kb	Kilobase
LTR	Longues répétitions terminales
MLH1dn	Dominant négatif du gène <i>MLH1</i>
MMR	Réparation des mésappariements
NHEJ	Jonction des extrémités non homologues
NMD	Voie de dégradation de l'ARNm non sens
PAM	<i>Protospacer adjacent motif</i>
PASTE	<i>Programmable Addition via Site-specific Targeting Elements</i>
pb	Paire de base
PBS	Site de liaison de l'amorce
PE	<i>Prime editing</i>
pegRNA	ARN guide PE
PPE	<i>Pure Prime Edit</i>

pre-crRNA	Précurseur d'ARN CRISPR
RBP	Protéine de liaison à l'ARN
RNAseq	Séquençage de l'ARN
RNP	Complexe ribonucléoprotéique
RT	Rétrotranscriptase
scRNAseq	Séquençage de l'ARN en cellules uniques
sgRNA	<i>Single guide RNA</i> , ARN guide
ShCAST	Système CAST de la cyanobactérie <i>Scytonema hofmanni</i>
SpCas9	Nucléase Cas9 de la bactérie <i>Streptococcus pyogenes</i>
ssDNA	ADN donneur simple brin
T	Thymine
TALE	Domaine effecteur de type activateur de la transcription
TALEN	Nucléase d'effecteur de type activateur de transcription
tracrRNA	ARN CRISPR trans activateur
TTR	Transthyrétine
U	Uracile
UDG	Uracile glycosylase d'ADN
UGI	Inhibiteur de l'uracile glycosylase d'ADN
VLP	<i>Virus-Like Particles</i>
ZF	Domaine à doigt de zinc
ZFN	Nucléase à doigt de zinc

PARTIE I :

INTRODUCTION

I. Les outils d'édition du génome

Au cours de ces dernières années, l'avènement des outils d'édition du génome a révolutionné la recherche biologique et la médecine. Le répertoire de ces outils s'est considérablement enrichi, en proposant des améliorations de plus en plus efficaces et précises. Ainsi, tels de véritables chirurgiens de l'ADN, les chercheurs sont dorénavant capables de modifier l'ADN à façon, c'est-à-dire d'insérer, de retirer ou bien de modifier entre un et plusieurs dizaines voire centaines de nucléotides dans l'ADN de cellules vivantes. De plus, il est désormais possible de programmer la modification désirée au nucléotide près, parfois parmi plusieurs milliards de nucléotides selon les espèces étudiées. Ces nouvelles opportunités ont permis de développer de nombreuses applications, aussi bien en recherche fondamentale, pour mieux comprendre la fonction des gènes (génétique inverse), qu'en recherche appliquée, notamment pour corriger des mutations responsables de maladies génétiques chez l'humain.

1. Historique : premières démonstrations d'édition ciblée du génome

Les expériences de transgénèse ont marqué le début de l'ingénierie des génomes. La transgénèse est définie comme l'introduction délibérée de matériel génétique étranger dans le génome d'un organisme vivant. En 1974, les pionniers Rudolf Jaenisch et Beatrice Mintz ont généré le premier animal transgénique en injectant l'ADN viral du simian virus 40 (SV40) dans des blastocystes murins¹. D'autres expériences phares ont suivi dont celles menées par Jon W. Gordon et Frank H. Ruddle ayant démontré l'aspect héréditaire et donc stable de la transgénèse dans des modèles murins^{2,3}. Cependant, le caractère aléatoire de l'intégration de l'ADN viral étranger constituait une limite majeure à cette technique de transgénèse.

La première preuve de concept de manipulation non aléatoire du génome fut démontrée au milieu des années 1980 par deux équipes américaines indépendantes^{4,5}. En transfectant des cellules mammaliennes avec une cassette contenant respectivement le gène de la globine- β ou bien un fragment du gène Hprt, ces deux équipes ont prouvé que leur cassette pouvait être intégrée dans le génome par mécanisme de recombinaison homologue dirigée (HDR). Cette découverte a été récompensée par le prix Nobel de médecine pour Mario Capecchi, Oliver Smithies et Martin Evans en 2007. Toutefois, la probabilité d'obtenir un événement HDR dans les cellules demeurait très faible.

En décembre 1994, l'équipe du Dr. Maria Jasin démontre qu'il est possible d'introduire des réparations ciblées de l'ADN⁶. En introduisant un site de restriction de 18 paires de bases pour la méganucléase I-SceI dans l'ADN de cellules murines en culture, une coupure double brin site-spécifique a pu être générée dans le génome murin. Ce résultat démontra également que les cassures double brin (CDB) du génome pouvaient être réparées selon deux voies de réparation distinctes : par HDR ou bien plus fréquemment par jonction des extrémités non homologues (NHEJ), résultant de l'apparition aléatoire de petites insertions et délétions (INDEL) à proximité du site de coupure. A l'époque des années 1990, l'utilisation des méganucléases comme premier outil capable d'induire des CDB dites « site-spécifique » a été envisagé.

2. Méganucléases et cassures double brins ciblées de l'ADN

Les méganucléases ont été découvertes pour la première fois en 1985 chez la levure de boulanger *Saccharomyces cerevisiae*⁷, puis finalement identifiées dans l'ensemble des domaines de classification des êtres vivants : bactéries, archées et eucaryotes⁸. Contrairement aux enzymes de restriction bactériennes classiques, les méganucléases reconnaissent un site de restriction de très grande taille (14 à 40 paires de bases). Même si plusieurs centaines de méganucléases existent dans la nature, la probabilité de trouver une enzyme capable de couper à l'endroit désiré est très faible. Le répertoire naturel des méganucléases étant limité, les scientifiques ont tenté de les améliorer afin de générer des enzymes capables de reconnaître de nouveaux sites de restriction. Certains groupes ont généré des méganucléases « chimères », en associant des domaines de liaison à l'ADN issus d'enzymes différentes⁹⁻¹¹. D'autres approches plus rationnelles ont consisté à muter certains résidus préalablement connus pour interagir spécifiquement avec l'ADN (données de cristallographie)¹²⁻¹⁴. L'évolution dirigée des protéines a également permis de générer, par pression de sélection chez la bactérie, une version évoluée de la nucléase I-SceI en 2006¹⁵. Même si ces premières expériences semblent prometteuses, rediriger les méganucléases vers des cibles de l'ADN d'intérêt sans pour autant impacter leur efficacité et leur spécificité s'avère techniquement complexe et long. Au cours des années 2000, les chercheurs ont développé de nouveaux outils possédant des sites de reconnaissance de l'ADN beaucoup plus flexibles et surtout « programmables » par l'expérimentateur, afin que n'importe quelle séquence d'ADN puisse *en théorie* être clivée par une nucléase.

3. Nucléases de l'ADN programmables

3.1 Nucléases à doigt de zinc

Les domaines à doigt de zinc Cys₂-His₂ sont des motifs de liaisons à l'ADN présents dans de nombreux facteurs de transcription eucaryotes^{16,17}. Un motif à doigt de zinc est constitué d'une trentaine d'acides aminés organisés en une structure tridimensionnelle composée de deux feuillets beta et une hélice alpha ($\beta\beta\alpha$) stabilisée par un ion zinc. Des études cristallographiques ont démontrées les acides aminés en position -1, 3 and 6 de l'hélice alpha reconnaissent un triplet de nucléotides spécifique dans l'ADN¹⁸. L'ingénierie moléculaire a permis de générer des domaines à doigts de zinc capables de reconnaître quasiment toutes les combinaisons parmi les 64 triplets de nucléotides existants¹⁹⁻²⁴. Ainsi, en combinant en tandem plusieurs briques élémentaires de domaines à doigts de zinc reconnaissant chacun un triplet de nucléotides spécifique, il est possible de reconnaître, *en principe*, une séquence d'ADN « sur mesure » beaucoup plus longue (assemblage modulaire). C'est en associant plusieurs doigts de zinc avec une endonucléase que les chercheurs ont pu concevoir les nucléases à doigt de zinc (ZFNs) (**figure 1**).

En 1996, l'équipe Américaine du Pr. Srinivasan Chandrasegaran a pour la première fois prouvée que l'ingénierie du génome était « programmable ». En fusionnant au domaine catalytique de l'endonucléase Fok I deux motifs à doigt de zinc, l'équipe a mis en évidence que Fok I était capable de cliver *in vitro* une séquence d'ADN spécifique²⁵. Bien que cette stratégie ait été employée pour inactiver divers gènes²⁶⁻²⁸, elle présente un taux d'échec assez élevé puisqu'elle ne prend pas en compte les effets « contextes-dépendants » de chaque sous-unité à doigt de zinc. L'équipe du Pr. J Keith Joung a précisément évalué ce taux d'échec à 76% après avoir analysé des centaines de combinaisons de trois motifs à doigt de zinc sur 104 séquences cibles différentes en utilisant un système rapporteur par double hybride²⁹. D'autant plus que certaines études ont alerté sur la cytotoxicité des ZFNs, notamment dues aux coupures hors cibles^{30,31}.

Des méthodes de criblage visant à améliorer l'affinité et la spécificité des tandems de doigt de zinc pour l'ADN ont été décrites³²⁻³⁴. Même si ces techniques ont permis d'améliorer les ZFNs et même de les porter jusqu'aux études cliniques, (leader Sangamo Therapeutics, <https://www.sangamo.com/>) (NCT03653247, NCT03432364, NCT03666871), elles nécessitent tout de même une expertise très spécialisée et présentent une efficacité limitée.

3.2 Nucléases effectrices de type activateur de transcription

En 2009, de nouveaux domaines de liaison à l'ADN appelés domaine effecteur de type activateur de la transcription (TALE) ont été décrit chez des bactéries phytopathogènes du genre *Xanthomonas*^{35,36}. Ces facteurs de transcription produit par les bactéries ont la particularité d'activer la transcription de gènes cibles chez l'hôte contaminé afin de promouvoir l'infection.

A la différence des motifs à doigt de zinc, les motifs TALEs ne reconnaissent qu'un seul nucléotide. Chaque motif est constitué de 34 d'acides aminés, dont deux résidus variables en position 12 et 13 qui confèrent une spécificité de liaison à un nucléotide particulier. Les di-résidus variables NI (Asn-Ile), NG (Asn-Gly), HD (His-Asp) et NK (Asn-Lys)/NH (Asn-His) reconnaissent respectivement les nucléotides A, T, C et G³⁷. Sur le même principe que les ZFNs, les scientifiques ont mis en place les nucléases effectrices de type activateur de transcription (TALENs), en fusionnant en tandem plusieurs motifs TALE avec le domaine catalytique de l'endonucléase Fok I (**figure 1**). En comparaison aux ZFNs, il a été montré que les TALENs étaient moins cytotoxiques et que les coupures hors cibles étaient moins fréquentes^{38,39}. Leur principale limite et leur taille : alors qu'un *cDNA* codant pour un ZFN mesure environ 1kb, celui d'un TALEN peut atteindre 3kb, rendant plus fastidieuses les étapes de clonage et de d'introduction des TALENs dans des cellules vivantes⁴⁰.

L'entreprise Cellectis, fondée en 1999 par les français André Choulika et David Sourdive et siégeant à Paris, dispose d'une licence exclusive permettant d'utiliser leur outil TALEN® pour faciliter l'ingénierie des récepteurs antigéniques chimériques des lymphocytes T (CAR-T cells), et possède aujourd'hui un portefeuille de 6 candidats en phase 1 d'essai clinique (<https://www.cellectis.com/en>).

Bien que de nombreux efforts ont permis d'améliorer les outils ZFNs et TALENs, ces technologies restent expérimentalement complexes à mettre en place : elles nécessitent des étapes de re-ingénierie pour chaque nouvelle cible ADN. De plus, il est difficile de prédire le comportement des interactions ADN-protéines, comme démontré par la faible efficacité des ZFNs générés par assemblage modulaire. La conception de ces outils est freinée par l'absence de règles claires, reproductibles et robustes. Avec l'émergence de la technologie CRISPR, beaucoup plus simple à programmer, l'engouement initial lié aux technologies ZFNs et TALENs se dissipa progressivement à partir de 2013.

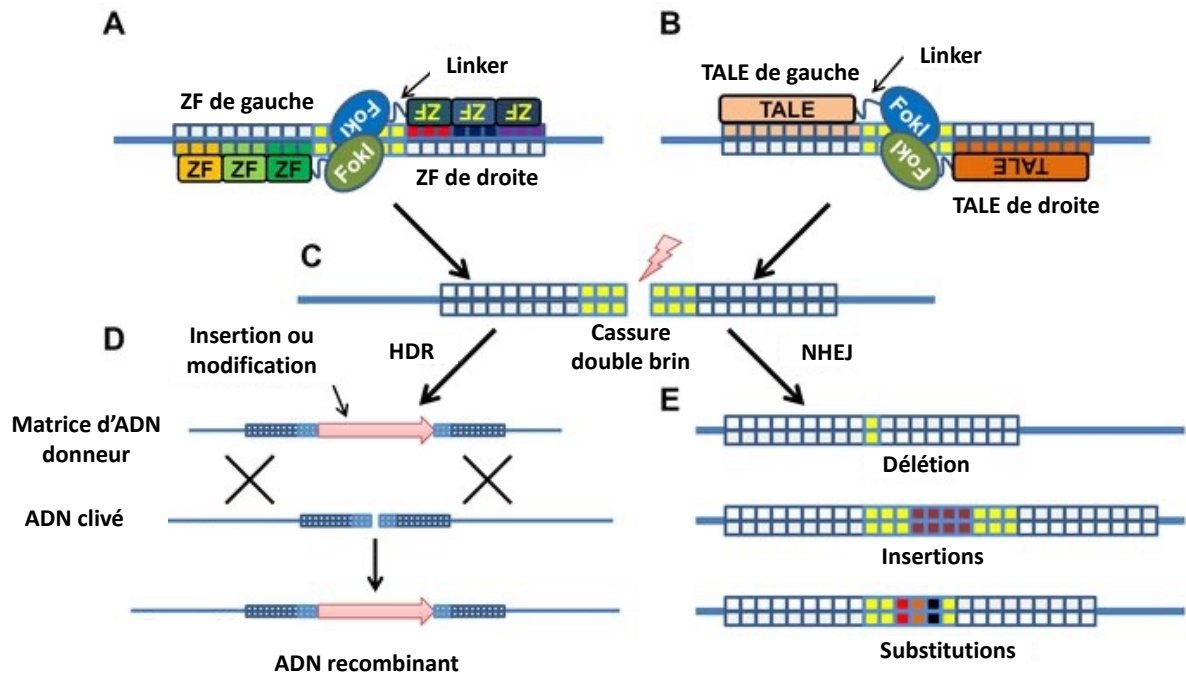


Figure 1 : Mécanisme d'action des ZFNs et TALENs

(A) Représentation schématique d'un dimère ZFN, symbolisé ici par deux monomères adjacents et antiparallèles de trois domaines à doigt de zinc associés en tandem (ZF de gauche et ZF de droite). Chaque sous unité ZF reconnaît un triplet de nucléotide unique illustré par les carrés de différentes couleurs, et chacun des monomères ZF est associé *via* un linker à une sous unité catalytique de la nucléase Fok I. (B) Représentation schématique d'un dimère TALEN, illustré par deux monomères de neuf domaines TALE adjacents et antiparallèles (TALE de gauche et TALE de droite) associés *via* un linker à une sous unité catalytique de la nucléase Fok I. (C) L'assemblage des deux monomères ZFN ou TALEN sur l'ADN permet la dimérisation de l'endonucléase Fok I, engendrant une cassure double brin de l'ADN. Deux voies de réparation permettent de réparer ces cassures afin d'introduire des modifications dans l'ADN : la recombinaison homologe dirigée (HDR) ou bien la jonction des extrémités non homologues (NHEJ). (D) Lorsqu'une matrice d'ADN donneur contenant une modification d'intérêt est fournie aux cellules par l'expérimentateur, celle-ci peut être intégrée au génome par HDR, générant ainsi un ADN recombinant renfermant la modification désirée (événement rare). (E) Plus fréquemment, les deux brins clivés peuvent être réassemblés par jonction des extrémités grâce la voie NHEJ. Ce type de réparation étant imparfait, il aboutit à l'apparition de petites insertions, délétions (INDELs) et substitutions. Cette stratégie est très souvent utilisée pour inactiver des gènes. Figure traduite et adaptée de Young Hoon Sun et al.⁴¹.

4. Émergence des outils CRISPR/Cas

4.1 Découverte du système CRISPR/Cas chez la bactérie

Afin de se défendre contre les bactériophages, les bactéries et les archées ont développé au cours de l'évolution des mécanismes de défense immunitaire très sophistiqués. Parmi ces mécanismes, deux sont spécifiquement consacrés à éliminer l'ADN viral étranger : le système restriction-modification et le système CRISPR/Cas (*Clustered Regularly Interspaced Short Palindromic Repeats*).

Le système de restriction-modification a premièrement été découvert en 1952 et 1953 par Salvatore Luria et Mary Human^{42,43}. Afin de protéger leur génome, les méthyltransférases bactériennes parsèment leur ADN de groupements méthyl. En cas d'infection, l'ADN viral, non méthylé (en général), est alors reconnu comme étranger par la bactérie. Pour se défendre, celle-ci utilise diverses enzymes de restriction sensibles à la méthylation. Ces enzymes reconnaissent une courte séquence d'ADN non méthylé et génèrent une coupure double brin du matériel génétique étranger, stoppant ainsi la réplication du phage^{44,45}. Le système restriction-modification est conservé chez près de 90% des génomes procaryotes⁴⁶. Le 15 février 2023, REBASE (<http://rebase.neb.com>) a répertorié près de 260 000 enzymes de restrictions. Depuis la découverte de ce système, les enzymes de restriction ont été très largement exploitées par les laboratoires en tant qu'outils phares du génie génétique (dans des tubes à essais).

Les séquences répétées inversées du système CRISPR ont été identifiées pour la première fois en 1987⁴⁷. En étudiant la séquence du gène *iap* chez *Escherichia coli*, Y Ishino *et al.* ont identifié un fragment insolite composé de 5 séquences homologues répétées de 29 nucléotides séparées par des séquences *spacer* de 32 nucléotides. Des études comparatives ont démontré que ce type de séquence était conservé chez de nombreuses autres espèces procaryotes et archées⁴⁸⁻⁵⁰. La fonction immunitaire du système CRISPR/Cas a été expérimentalement démontrée en 2007 par Rodolphe Barrangou et Philippe Horvath⁵¹. En infectant une souche sensible de *Streptococcus thermophilus* avec des bactériophages virulents isolés à partir d'échantillons de yaourt, ces chercheurs ont isolé 9 clones bactériens résistants aux bactériophages. Une analyse génétique a révélé que les clones mutants avaient acquis au moins une nouvelle séquence *spacer* d'une trentaine de paires de bases (pb) dans leur génome,

au niveau d'un locus CRISPR, à proximité des gènes Cas. Curieusement, ces séquences *spacers* correspondaient à 100% aux séquences *proto-spacers* retrouvés dans le génome des phages et confèrent aux bactéries un phénotype résistant⁵¹. En cas de réinfection, l'acquisition de ces nouvelles séquences *spacers* permettent aux enzymes Cas de cliver spécifiquement les séquences *proto-spacers* de l'ADN/ARN du phage pour stopper sa réplication (**figure 2**). Très peu de temps après cette découverte, il a été montré que l'activité endonucléase des protéines Cas était guidée par les ARN CRISPR (crRNA), directement transcrits à partir des séquences *spacers*⁵². D'autres travaux ont révélés l'importance des ARN CRISPR trans activateurs (tracrRNA) et motifs *protospacer-adjacent* (PAM) pour le fonctionnement du système CRISPR/Cas^{53,54}.

Selon l'organisation des locus CRISPR et la fonction des différents gènes Cas associés, les systèmes CRISPR/Cas ont été hiérarchisés en deux grandes classes (I et II), incluant six types (I à VI). Diverses enzymes Cas sont impliquées dans les étapes de l'intégration des séquences *spacer* (généralement Cas1 et Cas2), la maturation des crRNA et le clivage de l'ADN ou l'ARN viral étranger. Les systèmes de types I, II et V reconnaissent et clivent de l'ADN, celui de type VI agit sur l'ARN, celui de type III agit à la fois sur l'ARN et l'ADN et pour celui de type IV ce n'est pas très bien décrit⁵⁵⁻⁵⁸.

Le système CRISPR-Cas de type II, dérivé de chez *Streptococcus pyogenes*, est notamment composé de l'enzyme Cas9 (SpCas9) (**figure 2**). Grâce à son activité endonucléase, cette dernière est responsable du clivage double brin de l'ADN viral étranger. Précisément, le domaine HNH de la Cas9 clive le brin d'ADN ciblé par l'ARN alors son domaine RuvC-like va cliver le brin non ciblé⁵⁹. La protéine Cas9 a été jusqu'à présent la plus exploitée pour le développement des outils d'édition du génome.

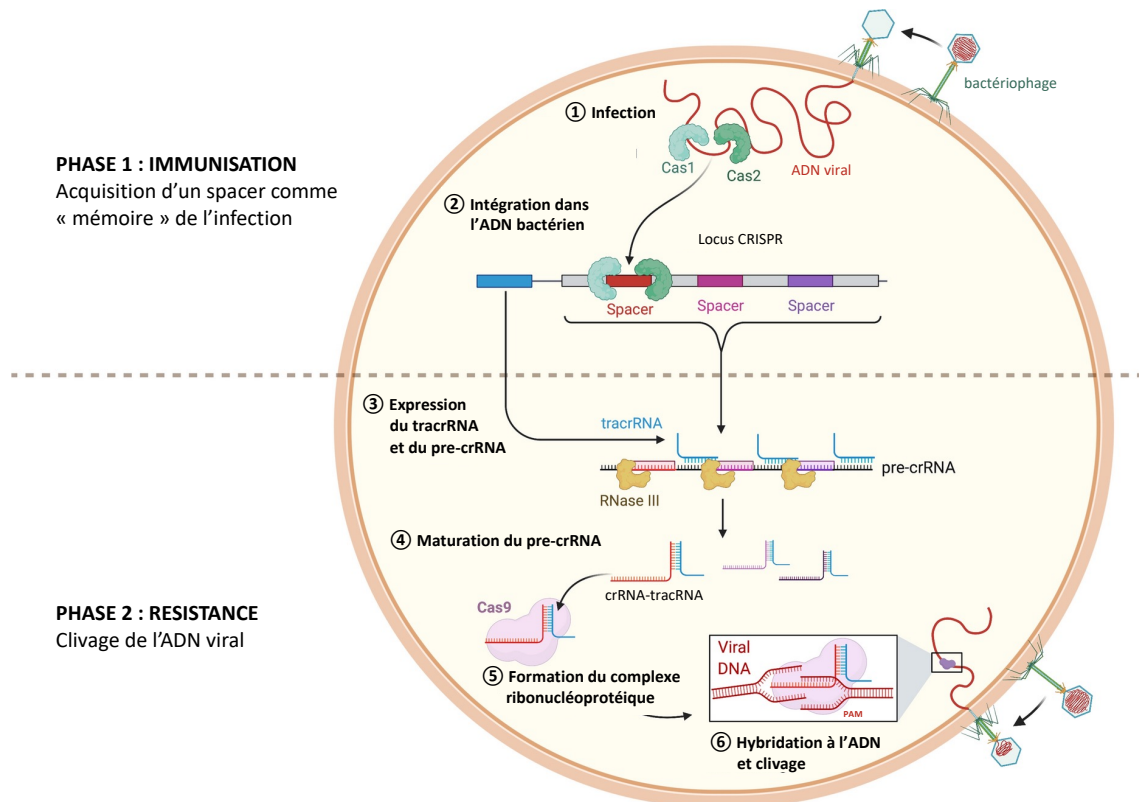


Figure 2 : Fonctionnement du système immunitaire CRISPR/Cas9 de *Streptococcus pyogenes*.

(1) Dès qu'un phage est adsorbé à la surface d'une cellule, celui-ci injecte son ADN viral dans la bactérie puis utilise la machinerie cellulaire de l'hôte pour se répliquer et produire de nouveaux virions. (2) Seulement quelques bactéries vont survivre à l'infection, grâce à l'incorporation d'une séquence virale *spacer* d'une trentaine de nucléotides dans leur génome, au niveau d'un locus CRISPR. Chez la plupart des espèces bactériennes, ce processus est médié par les enzymes Cas1 et Cas 2. (3) En cas de réinfection par le même phage, le locus CRISPR est transcrit et forme un long précurseur ARN (pre-crRNA) composé de plusieurs *spacers* et séparés par des séquences répétées. (4) Le pre-crRNA est mûré par la RNase III pour générer des crRNA (*spacers* individualisés). Un crRNA s'assemble par hybridation complémentaire avec un tracrRNA pour former un complexe crRNA-tracrRNA. (5) Le complexe crRNA-tracrRNA s'associe avec l'endonucléase Cas9 pour former un complexe ribonucléoprotéique et guider la Cas9 jusqu'à sa cible. (6) Après appariement de type Watson-Crick entre le *spacer* bactérien et le *proto-spacer* viral, l'enzyme Cas9 vient catalyser le clivage du matériel génétique viral entrant, stoppant ainsi l'infection : c'est la résistance. A noter que une petite séquence PAM est présente dans le *proto-spacer* viral mais absente dans le locus CRISPR, empêchant ainsi le clivage du locus CRISPR par la Cas9 (auto-immunité). Schéma traduit et adapté de Abozar Ghorbani et al.⁶⁰.

4.2 Vers la création d'un outil d'édition du génome programmable

Cette découverte capitale a incité les scientifiques à reprogrammer le système CRISPR-Cas9 de *Streptococcus pyogenes* pour cliver une séquence d'ADN d'intérêt. Pour fonctionner, l'enzyme Cas9 a besoin d'un tracrRNA et d'un crRNA. Le tracrRNA fonctionne comme un adaptateur pour la Cas9, tandis que le crRNA dirige le complexe vers la séquence virale cible. Pour former un complexe tracrRNA-crRNA stable, l'équipe d'Emmanuelle Charpentier a démontré que l'extrémité 3' du crRNA s'associe avec l'extrémité 5' du tracrRNA⁵³ (**figure 2**). A partir de cette observation, Jinek et al., ont eu l'idée de former en 2012 un ARN chimérique en fusionnant le tracrRNA et le crRNA. Cet ARN chimérique a été baptisé *single guide RNA* (sgRNA ou ARN guide), et a permis de condenser le système CRISPR/Cas9 en 2 éléments seulement : la Cas9 et le sgRNA⁶¹. **Les tests effectués *in vitro* ont démontré qu'en remplaçant juste la séquence d'hybridation à l'ADN dans l'ARN guide - de 20 nucléotides seulement - il était possible de programmer spécifiquement la cassure double brin dans une séquence d'ADN d'intérêt⁶¹.** La seule contrainte étant qu'un motif PAM NGG soit présent dans l'ADN, à proximité de la région ciblée par l'ARN guide⁶². Ce motif sert de « référence » à la Cas9 qui vient cliver sa cible ADN trois nucléotides exactement avant l'extrémité 5' de ce triplet NGG. Ces expériences ont mis en évidence toute l'efficacité, la versatilité et le caractère programmable de ce système dirigé par l'ARN. En effet, contrairement aux interactions ADN-protéines, les interactions ADN-ARN sont prédictibles. D'autant plus que les ARN sont souples, très faciles à concevoir (principe d'appariement Watson-Crick) et à produire.

Ces travaux publiés en juin 2012, très largement cités par la communauté scientifique, ont marqué le point de départ du développement des outils CRISPR/Cas. Les deux auteurs correspondantes, Jennifer A. Doudna et Emmanuelle Charpentier remportèrent le prix Nobel de Chimie en 2020.

CRISPR/Cas9 est rapidement devenu un outil d'édition du génome de référence pour induire des CDB et réparations de l'ADN sur mesure, il est efficace : *in vitro*, chez la bactérie⁶³ et dans des cellules en culture⁶⁴⁻⁶⁶, mais également *in vivo* dans de nombreux organismes modèles en biologie (animaux et végétaux)⁶⁷⁻⁷², et même chez l'homme⁷³ (leaders Intellia therapeutics <https://www.intelliatx.com/> et Editas medicine <https://www.editasmedicine.com/>). En février 2023, pas moins de 36000 articles avec la mention « CRISPR » ou « Cas9 » dans leur titre ou abstract ont été publiés (<https://pubmed.ncbi.nlm.nih.gov/>).

De la même manière que pour les ZFNs et TALENs, ce nouvel outil CRISPR/Cas9 a surtout été utilisé pour inactiver des gènes (CRISPR/KO) et étudier leur fonction, mais aussi pour installer de petites modifications génétiques d'intérêt en exploitant la réparation par HDR.

4.3 Réparation des cassures double brins induites par CRISPR/cas9

Les mécanismes de réparation des CDB les plus fréquents sont la jonction des extrémités non homologues classique (c-NHEJ) et la recombinaison homologe dirigée (HDR). **L'annexe I** résume les mécanismes moléculaires régissant ces voies de réparation, et sont précisément décrits dans ces revues⁷⁴⁻⁷⁷.

Brièvement, la réparation par c-NHEJ est non conservative. Elle est basée sur la re-ligation des deux extrémités lésées. Autrement dit, la séquence d'ADN d'origine n'est pas rétablie et laisse apparaître à la place de petites insertions/délétions autour du site de coupure (INDELs)⁷⁴. Les outils d'édition du génome exploitent la réparation non conservative pour inactiver des gènes cibles. En effet, la plupart du temps, l'apparition d'INDELs dans les séquences exoniques entraîne un décalage du cadre de lecture et l'apparition d'un codon stop prématuré, menant bien souvent à l'activation de la voie de dégradation de l'ARN messager non-sens (NMD)⁷⁸. Bien que l'apparition de ces INDELs ne soit pas contrôlée, plusieurs algorithmes et outils de type *machine learning* permettent de prédire ces modifications^{79,80}.

Lorsqu'un expérimentateur désire intégrer au niveau du site de coupure une modification génétique précise de type substitution ou INDEL ou bien insérer une cassette beaucoup plus grande, il peut concevoir un ADN donneur que la cellule utilisera comme matrice de réparation⁸¹. Le choix du type de matrice de réparation est déterminé selon la taille de la modification génétique souhaitant être intégrée au génome.

Les petites modifications (1 à 50 nucléotides) requièrent un ADN donneur simple brin assez court (ssDNA), inférieur à 200 nucléotides. Cette matrice doit contenir à chacune de ses extrémités une séquence de 30 à 50 nucléotides homologue aux séquences flaquant le site de coupure, appelés bras d'homologie⁸¹⁻⁸³. L'avantage des ssDNA est qu'ils sont très faciles à synthétiser chimiquement.

Lorsque l'expérimentateur souhaite insérer un plus grand fragment (ex : cassette, tag, rapporteur fluorescent, gene de résistance), il peut utiliser un ADN donneur double brin, linéaire ou circulaire, contenant des bras d'homologies entre 400 et 1000 pb^{4,84-86}. Cette matrice sera recopiée puis intégrée au génome par mécanisme de recombinaison homologue dirigée (HDR). Comme elles sont plus longues, les matrices d'ADN double brin sont produites le plus souvent par clonage.

Bien que l'expérimentateur fournisse aux cellules une matrice de réparation, des études ont montrées que les cellules utilisent davantage la réparation de type c-NHEJ au lieu de la voie HDR⁸⁷. Cela conduit à des taux d'édition d'intérêt relativement faible, et peut être expliqué par divers facteurs. Notamment, il a été démontré que la cinétique de réparation par HDR était beaucoup plus lente que celle par NHEJ, de l'ordre de 7h ou plus *versus* une trentaine de minutes⁸⁸.

Diverses méthodes ont donc été développées pour augmenter le ratio HDR/NHEJ et faciliter l'installation de modifications génétiques d'intérêts.

4.4 Augmentation du ratio HDR/NHEJ

Deux techniques complémentaires majeures permettent d'améliorer les rendements HDR : la modification de l'expression ou de l'activité des acteurs spécifiques des voies HDR/NHEJ ou bien l'optimisation des outils CRISPR/Cas9.

Premièrement, il a été démontré qu'en inhibant l'expression de certains acteurs clés du NHEJ (KU70/80 et l'ADN ligase IV), ou, en contraire, en surexprimant certaines protéines de la réparation par HDR (Rad51), il était possible de favoriser significativement les évènements HDR par rapport aux évènements NHEJ *in vitro* et *in vivo*⁸⁹⁻⁹³. Même si ces preuves de concept sont très prometteuses, il semblait important pour les chercheurs de développer des méthodes permettant d'inhiber le NHEJ de manière transitoire, afin de faciliter les applications thérapeutiques des réparations par HDR.

Des méthodes de criblage ont permis d'identifier puis de tester de nouveaux composés chimiques capables d'améliorer les taux d'HDR, telles que les molécules L755507, Brefeldin A, RS-1 et inhibiteurs de l'ADN ligase IV^{94,95}. En 2015, Takeshi Maruyama et al. ont ainsi démontré que la molécule SCR7, qui cible le domaine de liaison à l'ADN

de la ligase IV, permet d'augmenter jusqu'à 19 fois l'efficacité de l'HDR dans des cellules mammaliennes⁹⁶. De manière similaire, Van Trung Chu et al. ont utilisé les protéines E1B55K et E4orf6 dérivées d'un adénovirus qui permettent l'ubiquitination et la dégradation de l'ADN ligase IV⁹⁰ et d'autres ont utilisés les inhibiteurs de DNA-PK KU-0060648 et NU7441⁹⁷.

Comme les réparations HDR ont principalement lieu durant les phases S et G2, une autre stratégie utilisable *in vitro* consiste à synchroniser les cellules en culture et à délivrer au bon moment la matrice de réparation d'ADN^{98,100}. Alexandre Orthwein et al. ont également démontré que la surexpression d'un mutant gain de fonction CtIP (CtIP-T847E) dans des cellules 53BP1 KO permettait d'activer la voie HDR dans les cellule en phase G1⁹⁹.

Deuxièmement, l'optimisation des matrices de réparation et de la Cas9 peut également s'avérer efficace. Comme démontré par l'équipe de Jean-Paul Concordet et Carine Giovannangeli, l'utilisation d'oligonucléotides chimiquement modifiées avec des groupements phosphorothioate améliore les rendements HDR dans des cellules en culture et permettent d'insérer certaines modifications allant jusqu'à 100 paires de bases¹⁰⁰. Natasa Savic et al ont conçu une matrice de réparation directement attachée à la Cas9 par liaison covalente, permettant d'améliorer **jusqu'à 24 fois** l'efficacité de réparation par HDR¹⁰¹. Les études cinétiques de Christopher D Richardson et al. ont quant à elle démontrées que quelques heures après la coupure, la Cas9 libérait dans un premier temps l'extrémité 3' du brin non ciblé par l'ARN guide. Ainsi, en concevant une matrice ssDNA complémentaire au brin libéré en premier par la Cas9 (ssDNA asymétrique), les taux d'HDR ont pu être améliorés jusqu'à 60% dans les cellules humaines¹⁰². Dernier exemple parmi tant d'autres, en 2022, Evelyn Chen et al., ont démontré qu'en fusionnant la Cas9 avec PRDM9, une histone méthyltransférase responsable du dépôt des marques H3K36me3 et H3K4me3, il était possible d'augmenter le ratio HDR:indel de 5 fois, et cela dans plusieurs lignées cellulaires¹⁰³.

Malgré toutes ces stratégies, l'essor de nouvelles études alertant sur les dangers potentiels des CDB et coupures hors-cibles ont amené les scientifiques à prendre du recul sur ces outils (frein pour la thérapie *in vivo* chez l'homme).

4.5 Coupures en dehors de la cible

Bien que la Cas9 soit guidée spécifiquement par un ARN guide de 20 nucléotides, les nombreuses données de la littérature sont formelles : les outils d'édition du génome CRISPR/Cas sont susceptibles de générer des CDB non désirées en dehors de la cible, communément appelées modifications « *off-target* »^{104,105}. Ces modifications surviennent lorsque l'ARN guide se fixe sur une séquence nucléotidique similaire à la séquence cible (contenant entre 1 et 5 mésappariements ou INDELS)^{106,107}. L'activité hors-cible peut générer de petites INDELS^{104,105}, de larges insertions ou délétions¹⁰⁸ voire même des translocations chromosomiques^{108,109}. Plus important encore, dans le domaine pré-clinique et clinique, il est possible que de telles coupures s'immiscent dans des proto-oncogènes ou gènes suppresseurs de tumeurs. On peut aisément concevoir les conséquences préjudiciables associées au gain ou la perte de fonction de tels gènes sur le long terme (développement d'un cancer).

Il est donc très important pour les scientifiques de développer des méthodes sensibles, et efficaces à l'échelle du génome afin de localiser et quantifier précisément ces événements indésirables et faciliter la sélection des outils thérapeutiques les plus sécurisés.

4.5.1 Méthodes pour quantifier les coupures hors cibles

L'identification des sites de coupure hors de la cible peut se faire par deux grandes approches : soit à l'aide d'outils de prédictions bio-informatiques (avec *à priori, in silico*), soit à l'aide de techniques expérimentales pangénomiques basées sur l'activité des nucléases (sans *à priori, in vitro* ou *in cellulo*). Une fois la nucléase ayant opéré sur son substrat, les sites de coupure hors de la cible sont séquencés à haut débit afin de quantifier précisément leur incidence.

Méthodes avec *a priori* – *in silico*

Les outils de prédictions bio-informatiques permettent d'identifier de potentiels sites de coupure hors de la cible selon le degré d'homologie entre la séquence de l'ARN guide et des régions de l'ADN. Typiquement, la Cas9 peut tolérer des mésappariements ADN:ARN de 3 à 6 paires de bases sur les 20^{106,107} et/ou des INDELS de 1 à 5 paires de bases¹⁰⁵ tout en conservant son activité nucléase. Généralement, les mésappariements/INDELS localisés à distance de la séquence PAM sont davantage

tolérés que ceux localisés à proximité du PAM¹⁰⁶. L'équipe de Feng Zhang a montré que le mésappariement d'un seul nucléotide entre l'ARN guide et l'ADN pouvait parfois mener à une efficacité d'édition aussi importante que celle obtenue avec une hybridation parfaite¹⁰⁶. De nombreux outils bio-informatiques ont donc été programmés pour rechercher dans un génome donné ces potentiels sites flanqués par un motif PAM tels que Cas-OFFinder, CHOPCHOP ou CRISPOR¹¹⁰⁻¹¹⁵. Pour chaque site potentiel, des algorithmes calculent des scores de fiabilité (ex : MIT-score¹¹⁴, Cutting Frequency CFD score¹¹⁶) en prenant en compte le nombre de mésappariements ADN:ARN, leur nature et leur position sur l'ARN guide. Même si ces techniques de prédictions sont accessibles à tous et faciles à appréhender, elles demeurent tout de même très peu fiables et présentent un taux de faux positifs très élevé (**figure 3**). De plus, certaines études ont démontré que la SpCas9 est capable de reconnaître, en moindre mesure, certaines PAM alternatives telles que 5'-NCG-3' et 5'-NGA-3' au lieu de 5'-NGG-3'¹¹⁶.

Méthodes sans à priori in vitro

Aux cours des années, plusieurs méthodes visant à évaluer la capacité des nucléases à cliver l'ADN *in vitro*, ont été mises au point puis discutées en détail^{117,118}. La première, appelée digenome-seq, est basée sur la **digestion in vitro** d'un **génom**e complet (digenome) par la Cas9. Les fragments digérés sont ensuite détectés par séquençage du génome complet¹¹⁹⁻¹²¹. C'est probablement l'une des plus chères et moins sensibles à ce jour car elle requiert le séquençage du génome complet avec une couverture suffisante pour détecter les distinguer les événements rares du bruit de fond (environ 4 à 5 millions de reads pour le génome humain)¹²¹.

Par la suite, d'autres méthodes ont été mises au point. Succinctement, le principe des technologies CIRCLE-seq et SITE-seq^{122,123} repose sur la ligation d'adaptateurs de séquençage au site de coupure, lorsque la Cas9 génère une cassure. La présence des adaptateurs permet d'enrichir uniquement les fragments clivés pour faciliter leur détection par séquençage. Dans le cadre du SITE-seq, ces adaptateurs sont biotinylés. Le CHANGE-seq, développé plus récemment, utilise les propriétés de la transposase Tn5 pour générer des fragments d'ADN circularisés qui seront ensuite mis en contact avec la Cas9 et son ARN guide. La coupure de l'ADN circulaire par la Cas9 engendre l'apparition d'un fragment linéaire qui est ensuite enrichi par amplification puis détecté par séquençage haut-débit¹²⁴.

Généralement, ces techniques *in vitro* présentent l'avantage d'être plus sensibles que les techniques *in cellulo* (inférieur à 0.1% *in vitro*)¹¹⁹. Cependant, elles demeurent approximatives compte-tenu du taux élevé de faux positifs (figure 3). Dans le cadre de l'évaluation de la spécificité de la molécule thérapeutique NTLA-2001 (traitement de l'amyloïdose héréditaire à transthyrétine par CRISPR/KO), la technologie SITE-seq a permis d'identifier au total 475 potentiels sites de coupures hors-cible, dont 468 se sont révélés être de faux positifs (>98,5%) après analyse dans les cellules éditées (**figure 3**)⁷³. De plus, ces techniques ne permettent pas de prendre en compte certains critères inhérents à l'environnement intracellulaire, tels que la structure de la chromatine et son accessibilité, les modifications épigénétiques, le type cellulaire ou encore la compétition pour la liaison à l'ADN avec d'autres protéines nucléaires qui affectent l'activité des nucléases¹²⁰.

En 2018, Daesik Kim et al. ont revisité le Digenome-seq en utilisant une méthode d'extraction d'ADN douce permettant de conserver l'ADN chromatinien natif à la place de l'ADN dénaturé sans histone (DIG-seq). Ils ont démontré que la plupart des sites clivés dans l'ADN dénaturé sans histone ne l'étaient pas dans les régions condensées de l'ADN chromatinien natif, confirmant l'idée que l'apparition des coupures hors-cibles dépendent bien de l'état chromatinien¹²⁰. Par conséquent, le nombre de site faux positifs est considérablement réduit avec cette technique.

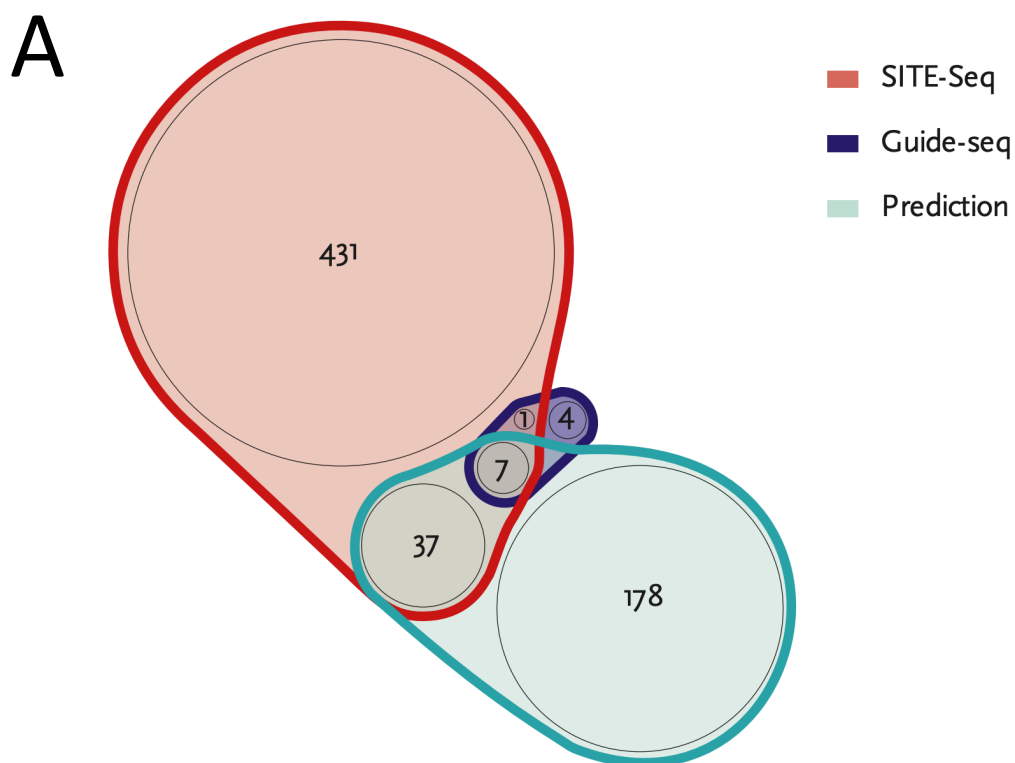
Méthodes sans à priori *in cellulo*

Contrairement aux méthodes *in vitro*, les méthodes *in cellulo* évaluent les contraintes spécifiquement liées au type cellulaire, au contexte intracellulaire, à la concentration des complexes ribonucléoprotéiques (RNP) présents dans le noyau et à leur propriété de diffusions qui sont limités par l'ensemble des constituants du noyau.

La technologie GUIDE-seq est probablement l'une des plus populaires à ce jour^{125,126}. La première étape repose sur la transfection des cellules avec la Cas9, l'ARN guide (ou autre type de nucléases) ainsi qu'un oligodéoxynucléotide double brin de 34 pb (étiquette moléculaire) qui sera intégré à certains site de coupure par un mécanisme dépendant du NHEJ¹²⁷⁻¹³⁰. Lors de la deuxième étape, l'ADN génomique est extrait, fragmenté puis les segments d'ADN ayant intégrés l'étiquette moléculaire sont spécifiquement enrichis par PCR puis détectés par séquençage haut débit. D'un point de vue technique, la sensibilité du GUIDE-seq dépend principalement de la quantité

d'ADN de départ et surtout la profondeur de séquençage utilisée. D'un point de vue biologique, la sensibilité des techniques *in cellulo* est limitée puisque l'apparition d'une coupure est déterminée par nombreux paramètres biologiques fluctuants cités précédemment ainsi que l'efficacité de transfection et la concentration de l'étiquette moléculaire ayant réellement atteint le noyau. D'autant qu'une cassure ne mène pas forcément à l'étiquetage, et par conséquent de véritables sites de coupure hors de la cible sont potentiellement manqués. A noter également que cette technologie n'est pas adaptée pour les cellules résistantes à la transfection et l'oligodéoxynucléotide a déjà été associé à des effets cytotoxiques dans les cellules primaires¹²⁰.

D'autres techniques expérimentales ont été développées, telles que le GUIDE-tag (version améliorée du GUIDE-seq)¹³¹, le DISCOVER-seq¹³², le BLISS¹³³, le BLESS¹³⁴, l'IDLV¹³⁵, l'HTHTS¹³⁶, le ONE-seq¹³⁷, le TAG-seq¹³⁸, et l'ITR-seq¹³⁹.



B

Site description	Annotation	PHH lot 1		PHH lot 2	
		Mean Δ indel (%)	p value	Mean Δ indel (%)	p value
On-target	On-target	92.57 \pm 7.85	0.001	93.50 \pm 0.10	1.91E-07
Off-target 1	Intergenic	7.37 \pm 0.72	0.002	1.43 \pm 0.12	1.99E-04
Off-target 2	Intronic	3.70 \pm 0.10	3.90E-06	1.07 \pm 0.15	4.11E-03
Off-target 3	Intergenic	0.87 \pm 0.06	3.50E-05	0.50 \pm 0.20	0.037
Off-target 4	Intergenic	1.50 \pm 0.69	0.03	0.33 \pm 0.06	0.001
Off-target 5	Intergenic	0.30 \pm 0.10	0.02	0.13 \pm 0.06	0.029
Off-target 6	Intronic	0.27 \pm 0.06	0.01	0.10 \pm 0.00	0.211
Off-target 7	Intergenic	0.27 \pm 0.06	0.01	0.10 \pm 0.00	0.091

Figure 3 : Sites de coupure hors cibles du candidat thérapeutique NTLA-2001.

(A) Le diagramme de Venn illustre les potentiels sites de coupure hors cibles pour l'ARN guide 5'-AAAGGCUGCUGAUGACACCU-3' de NTLA-2001. Ces sites ont été mis en évidence par le logiciel de prédiction *in silico* Cas-OFFinder (vert), le GUIDE-seq (violet) et le SITE-Seq (rouge). Le site réellement ciblé par la Cas9 fait bien parti des 7 candidats identifiés par les 3 méthodes. (B) Les 657 potentiels sites identifiés par ces 3 méthodes ont tous analysés par séquençage d'amplicon à haut débit et 7 locis au total (1,07%) ont été validés comme sites de coupures hors-cible après traitement de deux lots d'hépatocytes primaires humains (PHH) avec des doses saturantes de NTLA-2001. Données provenant de Julian D Gillmore et al.⁷³.

En somme, comme démontré sur la **figure 3**, en combinant différents outils (*in silico*, *in vitro*, *in cellulo*), la probabilité de manquer de potentiels sites de coupures hors de la cible est minimisée. Les technologies *in vitro* telles que le SITE-seq ou le CIRCLE-seq restent tout de même les plus exhaustives et les plus sensibles. En 2018, l'équipe de J. Keith Joung a mis au point la **stratégie « VIVO »** visant à quantifier les effets hors de la cible à partir de modèles animaux¹⁴⁰. La première étape de « **découverte** », exploite les différents outils préalablement décrits (ex : outils *in silico* combiné au CIRCLE-seq). Une deuxième étape de « **confirmation** » des candidats est effectuée par séquençage d'amplicon profond. Cette étape est réalisée *in vivo* (ex : modèle murins), en utilisant comme échantillon les tissus qui ont été exposés aux nucléases et en prenant en compte les facteurs de dose injectée et temps passé après l'injection. La limite de détection de cette stratégie est de l'ordre d'environ ~0.1% pour la détection des INDELS (limite de sensibilité des technologies de séquençage haut-débit)¹⁴⁰.

4.5.2 Stratégies visant à améliorer la spécificité des outils d'ingénierie du génome

De multiples stratégies ont été déployées par les scientifiques afin d'abaisser la fréquence des modifications hors cible.

Optimisation de l'ARN guide

Des études ont démontrées que l'extrémité 5' de l'ARN guide était capable de tolérer certains mésappariements avec l'ADN génomique sans affecter l'activité de la Cas9^{106,141}. Par conséquent, l'équipe de J Keith Joung a émis l'hypothèse que cette extrémité 5' était dispensable pour maintenir l'activité de la Cas9. En utilisant des ARN guides tronqués, avec une taille entre 17 et 18 nucléotides, cette équipe a démontré qu'il était possible de réduire ces événements indésirables de plus de 5000 fois. Cette modification ne semble pas affecter l'efficacité de clivage au site réellement ciblé par l'ARN guide, même si cela reste à démontrer sur un grand nombre de loci¹⁴². Tout de même, nous pouvons souligner que d'un point de vue purement statistique, la diminution de la taille de l'ARN guide, et donc de la séquence d'ADN reconnue, augmente nécessairement la probabilité de retrouver une séquence similaire dans le génome (apparition de nouveaux sites de coupures hors-cibles potentiels).

Charles A. Gersbach et son équipe ont quant à eux démontré que l'ajout d'une structure secondaire de type épingle à cheveux au niveau de l'extrémité 5' de l'ARN guide impactait la Cas9 à adopter la conformation nécessaire pour cliver l'ADN. Cette

structure secondaire génère en effet une barrière stérique et énergétique qui est d'autant plus importante à résoudre lorsque des mésappariements ADN:ARNg sont présents. En moyenne, la spécificité des effecteurs Cas a été améliorée d'environ 55 fois avec cette stratégie¹⁴³.

D'un point de vue thermodynamique, les duplexes ADN-ADN sont moins stable que les duplexes ADN-ARN^{144,145}. L'équipe de Daniel G Anderson a donc supposé qu'un guide chimérique, formé à la fois de désoxyribonucléotides et de ribonucléotides (ADN et ARN), serait moins susceptible de reconnaître des séquences d'ADN en présence de mésappariements. En modulant le nombre et la position des désoxyribonucléotides dans la séquence du guide, cette équipe a démontré que les guides chimériques ADN/ARN permettaient aussi de réduire l'activité hors-cible de la SpCas9 dans des lignées cellulaires¹⁴⁶.

Optimisation de l'architecture des outils CRISPR/Cas

En s'inspirant de l'architecture des ZFNs et des TALENs, les chercheurs ont développé des paires de nickase. En combinant deux ARN guides ciblant deux séquences d'ADN adjacentes, la probabilité de retrouver deux mêmes séquences similaires dans le génome diminue statistiquement. La coupure double brin est générée par les deux coupures simple brin (une coupure sur chaque brin d'ADN apportée par chaque nickase) (**figure 4**). Cette stratégie a été mise au point par l'équipe de Feng Zhang et a effectivement démontrée une réduction de l'apparition d'INDELs au niveau de plusieurs sites de coupures hors-cible¹⁴⁷. Toutefois, l'addition d'un ARN guide supplémentaire augmente nécessairement la probabilité de générer une cassure simple brin (susceptibles elles aussi de générer certaines fois des INDELs). Afin de pallier à ce problème tout en conservant une architecture similaire, deux équipes indépendantes ont publié la même année en utilisant une paire de Cas9 dépourvue d'activité endonucléasique (dCas9). Chaque monomère de dCas9 a été fusionné à un domaine catalytique de la nucléase Fok1^{148,149}. Comme pour les ZFN et les TALENs, l'assemblage de deux monomères de dCas9/ARNguide sur l'ADN permet la dimérisation de l'endonucléase Fok1, engendrant l'apparition d'une CDB de l'ADN (**figure 4**). Même si ces deux stratégies semblent intéressantes, il ne faut pas oublier que cela complexifie indéniablement l'architecture initiale des outils CRISPR/Cas et constitue une barrière pour l'administration de ces outils avec les vecteurs viraux dont la taille du génome est limitée.

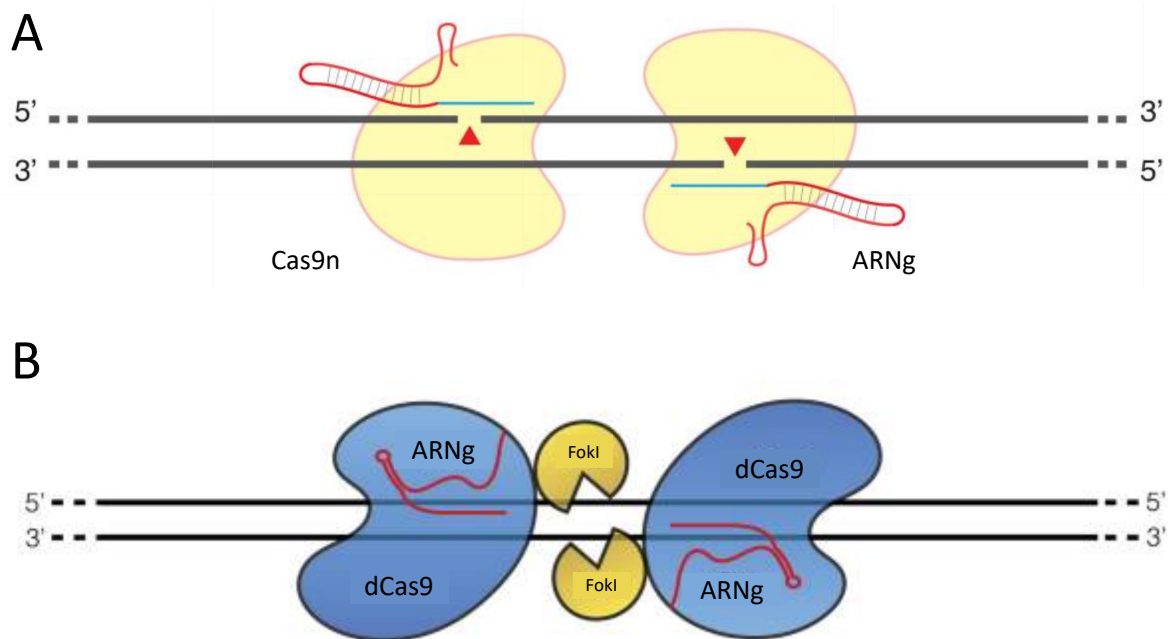


Figure 4 : Nouvelles architectures CRISPR/Cas9 de haute spécificité.

(A) Représentation schématique d'une paire de nickases (Cas9n), composée de deux sous unités adjacentes et antiparallèles de complexe Cas9n/ARNg. Chaque sous unité induit une cassure monocaténaire sur l'ADN représentée par une flèche rouge. Schéma traduit et adapté de F. Ann Ran et al.¹⁴⁷. (B) Représentation schématique d'un dimère dCas9-FokI. Chaque monomère dCas9 est associé *via* un linker à une sous unité catalytique de la nucléase Fok I. L'assemblage des deux monomères sur l'ADN permet la dimérisation de l'endonucléase Fok I, engendrant une cassure double brin de l'ADN. Schéma traduit et adapté de Shengdar Q. Tsai et al.¹⁴⁸.

4.5.3 Ingénierie de variants Cas9 haute-fidélité

Les études cristallographiques du complexe spCas9-ARN guide associé à l'ADN ont révélé la présence d'un sillon chargé positivement entre les domaines HNH et RuvC de la Cas9 (domaines nucléase) et le domaine d'interaction avec la PAM^{150,151}. Ce sillon chargé positivement laisse passer et interagit avec le brin non ciblé par la Cas9, chargé négativement. Sur la base de cette découverte, un groupe de chercheurs a émis l'hypothèse que l'activité de la Cas9 était permise par la séparation des deux brins d'ADN. Selon cette hypothèse, la neutralisation des charges positives de ce sillon permettrait d'affaiblir l'interaction avec le brin non ciblé par la Cas9 (affaiblissement de l'activité hélicase de la Cas9), et donc de rendre plus difficile la séparation des deux brins d'ADN¹⁵². D'un point de vue énergétique, les mésappariements entre l'ARN guide et l'ADN cible seraient donc beaucoup moins tolérés puisqu'ils privilégieraient la réhybridation entre les deux brins d'ADN complémentaires, réduisant ainsi l'apparition des coupures en dehors de la cible¹⁵².

Sur la base de cette hypothèse, cette équipe de chercheurs a entamé une démarche d'ingénierie rationnelle du sillon chargé positivement de la SpCas9. En mutant séquentiellement plusieurs résidus chargés positivement en alanine, Ian M Slaymaker et al. ont identifié le variant eSpCas9(1.1) (K848A/K1003A/R1060A) dont la spécificité a été améliorée (réduction des coupures en dehors de la cible). A titre informatif, en comparaison à la spCas9 sauvage, le variant eSpCas9(1.1) a éliminé 22 des 24 sites de coupure hors de la cible associé aux ARN guides EMX1(1) et VEGFA(5,1) (<0.2% indel)¹⁵².

En partant également des données cristallographiques, Benjamin P Kleinstiver et al. ont aussi supposé que le complexe SpCas9-ARNg pourrait posséder plus d'énergie qu'il n'en faut pour cliver sa cible, et que ce surplus énergétique pouvait être réquisitionné pour également cliver des sites mal appariés avec l'ARN guide. Afin de diminuer les interactions non spécifiques entre la SpCas9 et l'ADN (et donc réduire l'énergie mise en jeu dans cette interaction et augmenter la spécificité de la SpCas9), cette équipe a muté 4 résidus impliqués dans la formation de liaisons hydrogènes avec le phosphosquelette d'ADN pour produire le variant SpCas9-HF1 (N497A, R661A, Q695A, Q926A). La spécificité de ce variant a été évaluée à l'échelle du génome par la technique GUIDE-seq puis confirmée par séquençage d'amplicon profond. **Le variant SpCas9-HF1 permet d'abaisser tous (ou presque) les évènements de**

coupures hors de la cible à des niveaux indétectables et tout en conservant une activité aussi élevée que la SpCas9 sauvage au niveau de la région ciblée¹⁵³.

Deux années après, une autre équipe alerta tout de même la communauté en démontrant que lorsque les variants SpCas9-HF1 et eSpCas9(1.1) étaient délivrés aux cellules HEK293T sous forme de complexe RNP, leur activité au site cible était considérablement réduite en comparaison à la SpCas9 sauvage. Une explication plausible concerne la cinétique de clivage du variant SpCas9-HF1, qui est 100 fois plus lente que la SpCas9 sauvage¹⁵⁴. On peut donc supposer que la cinétique de dégradation des complexes RNP dans les cellules chevauche la cinétique de clivage sur la cible à l'échelle de l'ensemble des cellules. Ces variants ne sont donc pas adaptés pour une utilisation sous forme de complexe RNP.

Cette équipe a alors utilisé un criblage bactérien sans *a priori* afin d'identifier des variants Cas9 susceptibles de conserver une haute activité même lorsqu'ils sont délivrés sous forme de complexe RNP (expression transitoire). Au total, 250 000 clones bactériens ont été testés (5–7 mutations par kilobase). Les candidats présentant une activité hors-cible réduite sans sacrifier l'activité cible ont été retenus dans un premier criblage (875 candidats). Sur ces 875 candidats, un deuxième criblage a été effectué avec un ARN guide différent, et 163 hits ont été retenus (94 mutations). Finalement, l'efficacité et la spécificité des 94 variants ont été systématiquement comparées à la SpCas9 sauvage dans des lignées cellulaires, et le meilleur variant s'est avéré être le R691A, nommé ensuite HiFi Cas9. Comme cette mutation est présente dans le domaine interagissant avec le brin ciblé par l'ARN guide (domaine REC3), les auteurs supposent qu'il est plus difficile pour la Cas9 de passer d'une conformation « inactive » à une conformation « active » en présence de mésappariements. La HiFi Cas9 a été testée sous forme de complexe RNP dans des cellules souches hématopoïétiques CD34+ de donneurs atteints de drépanocytose, et s'est révélée aussi efficace que la SpCas9 sauvage pour corriger la mutation délétère Glu6Val du gène *HBB* et tout en réduisant les coupures hors-cible de 20 fois¹⁵⁵.

De nombreux autres variants ont été décrits et ils sont regroupés au sein du **tableau 1**¹⁵⁶. Plusieurs stratégies complémentaires ont permis d'améliorer la fidélité des outils CRISPR/Cas9, mais leur utilisation préclinique ou clinique reste encore restreinte, la SpCas9 sauvage étant effectivement la plus caractérisée et la plus utilisée actuellement.

Classification	Stratégie	Variants Cas9 à haute fidélité
Ingénierie non rationnelle	Evolution dirigée des protéines (mutagenèse aléatoire)	Sniper-Cas9 (PMID: 30082838), HiFi Cas9 (PMID: 30082871), xCas9 (PMID: 29512652), SpartaCas (PMID: 32298334), efSaCas9 (PMID: 32644995)
Ingénierie rationnelle	Ingénierie des protéines basée sur leur structure et/ou leur fonction	SpCas9nickase (PMID: 23992846), SpCas9-D1135E (PMID: 26098369), eSpCas9 (PMID: 26628643), SpCas9-HF1 (et HF2, -HF3, -HF4) (PMID: 26735016), HypaCas9 (PMID: 28931002), SuperFi-Cas9 (PMID: 35236982), LZ3 Cas9 (PMID: 32187529), eSaCas9 (PMID: 26628643), SaCas9-HF (PMID: 31570596), KKH-SaCas9-SAV1 (and -SAV2) (PMID: 35051997)
	Protéines de fusion	dCas9-Fok1 (PMID: 24770325), miCas9 (PMID: 33247137), Cas9-pDBD (PMID: 26480473), Cas-Cas9 chimeras (PMID: 30451839)
	Séparation des protéines	Split-Cas9 (PMID: 25643054)
Combinaison	Evolution dirigée des protéines couplée à la modélisation structurale	evoCas9 (PMID: 29431739), SaCas9-Q414A (PMID: 31911131)

Tableau 1 : Ingénierie de variants Cas9 haute-fidélité.

Tableau traduit et adapté de Xiaoqiang Huang et al.¹⁵⁶.

4.6 Autres effets indésirables associés aux cassures double brins de l'ADN

En 2020, une étude menée par Oana M. Enache et al. sur plusieurs dizaines de lignées cellulaires cancéreuses a montré que l'introduction de la Cas9 dans ces cellules provoquait l'activation de la voie de signalisation p53, et corrélait avec l'apparition/expansion de mutations perte de fonction p53 dans environ 10% des cas¹⁵⁷. p53 étant considéré comme « le gardien du génome », il n'est pas si surprenant qu'une CDB mène à son activation pour mener à bien la réparation, comme démontré par d'autres¹⁵⁸⁻¹⁶¹. Cependant, ce suppresseur de tumeur étant muté dans près de la moitié des cancers¹⁶², on imagine facilement que de telles mutations pourraient avoir des conséquences irréversibles sur le long terme. Ces conclusions sont à nuancer par le fait qu'une grande partie des expériences ont été réalisées par transduction lentivirale de la Cas9. Une expression transitoire aurait peut-être restreint les effets. De plus, aucun lien de causalité n'a été démontré entre la transduction de la Cas9 et l'apparition de ces mutations.

Des réarrangements chromosomiques peuvent également apparaître à la suite des CDB. En cherchant à réparer une mutation responsable de la porphyrie érythropoïétique congénitale par CRISPR/HDR, l'équipe Bordelaise d'Aurélie Bedel a identifié avec surprise des délétions chromosomiques de l'ordre de 7.5 Megabases dans leurs cellules, apparues exactement au niveau du site de coupure généré par la Cas9 (10% dans des lignées cancéreuses et 7.7% dans des fibroblastes). Ces évènements sont dépendants de l'activation de p53¹⁶³. De même, en utilisant la technologie de séquençage « long read », Michael Kosicki et al. ont mis en évidence des délétions de l'ordre du kilobase (kb) mais également des insertions et des réarrangements génomiques beaucoup plus complexes. Les expériences ont été menées dans des cellules embryonnaires de souris, qui, contrairement aux lignées cancéreuses, disposent d'un caryotype sain et surtout de gènes de réparation à l'ADN non mutés¹⁰⁸. En 2020, dans une étude clinique de phase I (NCT03399448) utilisant l'ingénierie des cellules T, l'inactivation simultanée de 3 gènes *ex vivo* (TRAC, TRBC et PDCD1) a mené à l'apparition de translocations, notamment entre les gènes TRBC1 et TRBC2 et résultant d'une délétion de 9,3 kb. Malgré cela, les cellules T modifiées ont été réinjectées chez 3 patients et la translocation était toujours présente 30 et 170 jours après la réinjection¹⁰⁹.

De nombreuses autres études ont permis d'identifier de tels réarrangements chromosomiques *in vitro* et *in vivo*, dans des modèles murins (délétions, inversions, translocations, intégration de vecteurs, insertion de rétrotransposon)^{164–168}. Des évènements encore plus catastrophiques allant jusqu'à la perte de chromosomes ont également été décrits^{169–171}. On peut facilement émettre l'hypothèse que de tels réarrangements pourraient mener à la sélection de cellules oncogéniques.

L'ensemble de ces études souligne également l'importance des techniques d'analyse de l'ADN à grande échelle (séquençage *long read*, CGH, FISH, caryotype) pour la détection des variants structuraux. Certes, les techniques telles que la PCR classique (amplicon inférieur à 1kb) suivie d'un séquençage short read permettent de très bien quantifier le ratio HDR:indel ainsi que les petites modifications hors-cibles. Cependant elles ne contribuent pas à détecter de tels réarrangements, qui sont très probablement sous-estimés aujourd'hui.

Même si de nombreuses optimisations ont permis de faciliter l'intégration de modifications génétiques précises dans l'ADN, l'abondance des INDELS non

contrôlées et l'existence de réarrangements chromosomiques ne sont pas compatibles pour une application thérapeutique des outils CRISPR *in vivo*. Toutes ces raisons ont amené les scientifiques à repenser ces outils :

Comment améliorer les nucléases pour mettre en place des modifications génétiques d'intérêt avec une grande efficacité, tout en limitant les INDELS aléatoires et les événements indésirables associés aux CDB de l'ADN ?

5. Réingénierie des outils CRISPR/Cas pour l'édition génétique de précision sans CDB

Les nucléases ne peuvent pas directement altérer la séquence de l'ADN ; c'est la CDB suivie des mécanismes de réparation qui sont responsables de l'apparition des modifications génétiques. Alors, pourquoi ne pas créer de nouvelles nucléases Cas9 en les associant avec d'autres enzymes, qui, elles, seraient capables de directement modifier les nucléotides de l'ADN ? Dans ce cas, les CDB de l'ADN deviennent inutiles, et tous les effets secondaires associés supprimés.

6. CRISPR 2.0 : les éditeurs de bases cytosine et adénine

6.1 Origine des éditeurs de bases

Dans la nature, il existe des enzymes appelées déaminases susceptibles de modifier une base nucléique en une autre. Les cytidines déaminases du groupe AID/APOBEC sont capables de déaminer la cytosine (C) pour produire l'uracile (U), s'appariant avec une adénine (A) au lieu d'une guanine (G). Comme les ADN polymérases reconnaissent l'uracile (U) comme étant une thymine (T), ces cytidines déaminases permettent de générer des transitions stables C:G→T:A dans l'ADN ou de changer le code génétique d'un ARN messager (ARNm)¹⁷². Par exemple, l'enzyme AID participe à la diversification du répertoire des immunoglobulines dans l'ADN¹⁷³, et APOBEC1 est capable d'éditer l'ARNm de l'apolipoprotéine B dans l'intestin pour produire l'isoforme apoB48¹⁷⁴.

En 2012, les travaux de Martin Jinek et al. ont démontré qu'en introduisant seulement deux mutations dans la Cas9 (Asp10Ala et His840Ala), il était possible de produire une Cas9 ayant perdu son activité endonucléase (dCas9), tout en conservant sa capacité de liaison à l'ADN dépendante de l'ARN guide⁶¹.

Ainsi, en 2016, Alexis C. Komor et al. ont créé pour la première fois le système système d'éditeur de base cytosine (CBE) en fusionnant la dCas9 avec l'enzyme APOBEC1 qui catalyse directement la conversion d'une base (C) en une autre (U) dans l'ADN (version **BE1**)¹⁷⁵. Bien que le CBE de première génération (BE1) soit très efficace *in vitro* (dans des tubes à essais), cette version semblait très peu efficace dans des cellules en culture (0,8 à 7,7%). Dans l'ADN des cellules, le nucléotide U est considéré comme anormal. Il est rapidement reconnu puis excisé par l'uracile glycosylase d'ADN (UDG) qui active le système de réparation par excision des bases (BER) et restaure le U en C avant l'installation de la modification sur l'autre brin. Alexis C. Komor et al. ont donc amélioré le BE1 en fusionnant à la dCas9 un peptide inhibiteur de l'UDG (UGI) provenant du bactériophage PBS1 de *B. subtilis* (version **BE2**). Le BE2 était 3 fois plus efficace que le BE1 pour médier la conversion de C vers T (efficacité jusqu'à 20% dans des cellules en culture). Pour aller encore plus loin, Alexis C. Komor et al ont émis l'hypothèse qu'une Cas9 nickase (A840H)⁶¹ produisant une coupure monocaténaire sur le brin non édité (contenant le G, opposé au U) pourrait stimuler sa réparation par MMR en utilisant le brin édité comme matrice de réparation. De cette manière, la conversion du G en A sur le brin non édité est facilitée, et donc l'incorporation des modifications désirées U:A puis T:A également¹⁷⁵ (**figure 5**). La troisième version **BE3** a produit jusqu'à 37% d'efficacité dans des culture de cellules, mais a également augmenté la fréquence des INDELS (en moyenne 1%). Suite à ces résultats prometteurs, d'autres modifications ont encore été apportées afin d'améliorer d'autant plus l'efficacité du CBE, optimiser son l'expression, sa translocation nucléaire¹⁷⁶, et sa spécificité¹⁷⁷⁻¹⁸¹.

Sur le même principe, Nicole M. Gaudelli et al., ont eu l'idée de créer l'éditeur de base adénine (ABE). La déamination de l'adénine produit l'inosine, reconnue comme étant une guanine par l'ADN polymérase. Cependant, aucune enzyme existante dans la nature n'est capable de catalyser cette réaction à partir d'un substrat ADN (ARN seulement)¹⁸²⁻¹⁸⁴. Nicole M. Gaudelli et al., ont donc entamé une stratégie d'évolution dirigée des protéines afin que l'enzyme TadA d'*Escherichia coli* soit capable de déaminer les adénines sur l'ADN simple brin. Au total, 7 cycles d'évolution ont suffi pour obtenir des résultats spectaculaires : l'ABE version 7.10 a permis de convertir les duplexes A•T en G•C avec une efficacité d'environ 50% dans les cellules humaines et un taux d'INDELS relativement faible (inférieur à 0,1%)¹⁸⁵ (**figure 5**). D'autres versions

ABE encore plus efficaces telles que l'ABEmax¹⁷⁶, l'ABE8¹⁸⁶, et l'ABE8e¹⁸⁷, ont également été optimisées.

En définitive, les outils CBE et ABE permettent de générer les 4 types de transitions dans l'ADN (C→T, A→G, T→C ou G→A). Ces deux outils sont largement exploités dans le domaine préclinique pour corriger des variants pathogènes responsables de maladies monogéniques¹⁸⁸⁻¹⁹⁰, moduler le niveau d'expression des gènes mutant leur région promotrice¹⁹¹, imoduler l'épissage¹⁹² et inactiver des gènes en générant des codons stop prématurés¹⁹³⁻¹⁹⁵. Cette dernière stratégie présente de quelques avantages en comparaison au CRISPR/KO classique : absence d'effets génotoxiques associés aux CDB, absence d'INDELS incontrôlées et multiples (modification unique et localisée – effet unique) et multiplexage possible sans augmenter les risques de réarrangements chromosomiques¹⁹⁶.

L'efficacité des éditeurs de base a été prouvée a de maintes reprises dans de prestigieux journaux, aussi bien dans des cellules vivantes *in vitro*^{197,198}, que chez les plantes¹⁹⁹⁻²⁰¹ et également chez les animaux tels que la drosophile²⁰² les rongeurs²⁰³⁻²⁰⁶, le poisson zèbre^{207,208}, le xénope²⁰⁹, le cochon²¹⁰, les primates non humains^{211,212} et mêmes les embryons humains^{213,214}. Les essais cliniques en sont à leur balbutiement, mais ils vont très certainement se multiplier dans les mois et années à venir. A noter que des éditeurs de l'adénosine dans l'ARN ont également été mis au point en fusionnant une déaminase avec des oligonucléotides antisens ou bien une enzyme Cas13²¹⁵⁻²¹⁹.

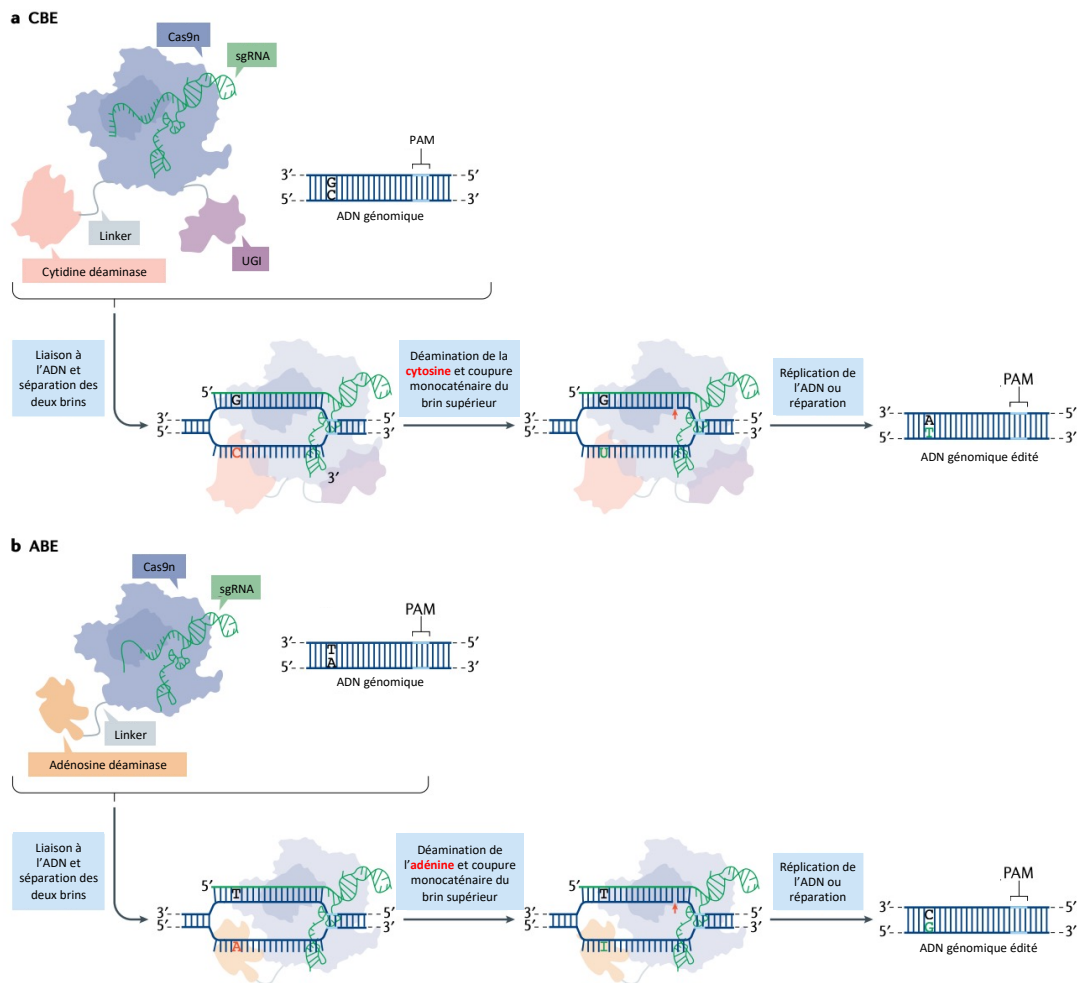


Figure 5 : Mécanisme d'action des éditeurs de bases.

(A) L'éditeur de base cytosine (CBE) catalyse la conversion de duplexes C•G en T•A. Les différents composants du CBE sont annotés dans les bulles rectangulaires. Suite à la liaison du CBE sur l'ADN, la cytidine déaminase va catalyser la conversion des nucléotides cytosine (en rouge) présents sur le brin non ciblé par l'ARN guide en nucléotides uracile (en vert). L'inhibiteur de l'uracile glycosidase (UGI) protège le nucléotide intermédiaire uracile contre son excision par l'uracile glycosylase d'ADN (UDG). La coupure monocaténaire du brin non ciblé par l'ARN guide va stimuler sa réplication ou réparation pour permettre l'installation définitive de la modification. (B) L'éditeur de base adénine (ABE) catalyse la conversion de duplexes A•T en G•C. Les différents composants de l'ABE sont annotés dans les bulles rectangulaires. Suite à la liaison de l'ABE sur l'ADN, l'adénine déaminase va catalyser la conversion des nucléotides adénine (en rouge) en inosine (en vert). La coupure monocaténaire du brin non ciblé par l'ARN guide va stimuler sa réplication ou réparation pour permettre l'installation de la modification. Schéma traduit et adapté de Elizabeth M. Porto et al.²²⁰.

6.2 Challenges associés aux éditeurs de bases

6.2.1 Fenêtre d'activité et bystanders

La principale limite associée aux éditeurs de base est liée à la taille de la fenêtre d'activité de la déaminase. Par exemple, le BE3 est très efficace pour éditer les nucléotides 4 à 8 ciblés par l'ARN guide (en considérant que la séquence PAM est en position 21 à 23) (**tableau 2**). Or, la plupart des applications développées par les scientifiques requièrent la modification d'une base unique, ce qui n'est pas forcément compatible avec cette large fenêtre d'activité. Si l'expérimentateur souhaite convertir un C en T (base cible) mais que plusieurs C (bases adjacentes à la cible) sont présents dans la fenêtre d'édition ou bien juste à proximité, il y a un risque élevé que les cytosines adjacentes à la cible soient également modifiées. Ce type de modification non désirée est appelé *bystander* et est à différencier des modifications hors cible (*off-target*, partie 4.5) qui apparaissent en dehors de la région cible d'intérêt. Lorsqu'un *bystander* surgit dans le cadre de lecture d'un gène et génère une mutation non silencieuse, cela peut avoir des conséquences néfastes : perte d'activité ou d'expression de la protéine voire même apparition d'un codon stop prématuré dans la séquence du gène (plus rarement mutation gain de fonction).

Afin de réduire ces *bystanders*, l'équipe du Pr. David R Liu a développé de nouvelles versions CBE avec des fenêtres d'activité beaucoup plus étroites, typiquement entre 1 et 2 nucléotides seulement (YE2-BE3 et YEE-BE3) au lieu d'une dizaine²²¹ (**tableau 2**). L'équipe du Professeur J Keith Joung a elle démontrée que l'activité du variant APOBEC3A (eA3A-CBE) (**tableau 2**) dépendait très fortement du contexte génétique, avec TCR>TCY>VCN²²². Sangsin Lee et al. ont créé le variant APOBEC3G (A3G-CBE) éditant préférentiellement le second C dans un motif « CC ». Ce nouveau variant CBE a généré des allèles correctement modifiés (sans *bystanders*) avec un pourcentage 6000 fois supérieur par rapport au BE4max²²³.

Editeur de base	Architecture	Fenêtre d'activité et PAM	Référence
BE1	APOBEC1 16 aa Sp dCas9 (D10A, H840A)		Komor et al. Nature
BE2	APOBEC1 16 aa Sp dCas9 (D10A, H840A) 4 aa UGI		Komor et al. Nature
BE3	APOBEC1 16 aa Sp nCas9 (D10A) 4 aa UGI		Komor et al. Nature
HF-BE3	APOBEC1 16 aa HF nCas9 (D10A) 4 aa UGI		Rees et al. Nat. Comm.
BE4, BE4max	APOBEC1 32 aa Sp nCas9 (D10A) 9 aa UGI 9 aa UGI		Koblan et al. Nat. Biotechnol. Komor et al. Sci. Adv.
BE4-GAM	Gam 16 aa APOBEC1 32 aa Sp nCas9 (D10A) 9 aa UGI 9 aa UGI		Komor et al. Sci. Adv.
YE1-BE3	APOBEC1 W60Y R132E 16 aa Sp nCas9 (D10A) 4 aa UGI		Kim et al. Nat. Biotechnol.
EE-BE3	APOBEC1 R132E R132E 16 aa Sp nCas9 (D10A) 4 aa UGI		Kim et al. Nat. Biotechnol.
YE2-BE3	APOBEC1 W60Y R132E 16 aa Sp nCas9 (D10A) 4 aa UGI		Kim et al. Nat. Biotechnol.
YEE-BE3	APOBEC1 W60Y R132E R132E 16 aa Sp nCas9 (D10A) 4 aa UGI		Kim et al. Nat. Biotechnol.
VQR-BE3	APOBEC1 16 aa Sp VQR nCas9 (D10A) 4 aa UGI		Kim et al. Nat. Biotechnol.
VRER-BE3	APOBEC1 16 aa Sp VRER nCas9 (D10A) 4 aa UGI		Kim et al. Nat. Biotechnol.
Sa-BE3	APOBEC1 16 aa Sa nCas9 (D10A) 4 aa UGI		Kim et al. Nat. Biotechnol.
Sa-BE4	APOBEC1 32 aa Sa nCas9 (D10A) 9 aa UGI 9 aa UGI		Komor et al. Sci. Adv.
SaBE4-Gam	Gam 16 aa APOBEC1 32 aa Sa nCas9 (D10A) 9 aa UGI 9 aa UGI		Komor et al. Sci. Adv.
SaKKH-BE3	APOBEC1 16 aa Sa KKH nCas9 (D10A) 4 aa UGI		Kim et al. Nat. Biotechnol.
Cas12a-BE	APOBEC1 16 aa dCas12a 14 aa UGI		Li et al. Nat. Biotechnol.
Target-AID	Sp nCas9 (D10A) 100 aa CDA1 9 aa UGI		Nishida et al. Science
Target-AID-NG	Sp Cas9 (D10A)-NG 100 aa CDA1 9 aa UGI		Nishimasu et al. Science
xBE3	APOBEC1 16 aa xCas9 (D10A) 4 aa UGI		Hu et al. Nature
eA3A-BE3	APOBEC3A N10 16 aa Sp nCas9 (D10A) 4 aa UGI		Gerkhe et al. Nat. Biotechnol.
A3A-BE3	hAPOBEC3A 16 aa Sp nCas9 (D10A) 4 aa UGI		Wang et al. Nat. Biotechnol.
BE-PLUS	10X GCN4 Sp nCas9 (D10A) ScFv rAPOBEC1 UGI		Jiang et al. Cell. Res.
TAM	Sp dCas9 44 aa hAID* P1EX2		Ma et al. Nat. Meth.
CRISPR-X	Sp dCas9 MS2 30 aa hAID* hyperactive mutant		Hess et al. Nat. Meth.
ABE7.9	TadA 32 aa TadA Mutant 32 aa Sp nCas9 (D10A)		Gaudelli et al. Nature
ABE7.10	TadA 32 aa TadA Mutant 32 aa Sp nCas9 (D10A)		Gaudelli et al. Nature
ABE7.10 *	TadA 32 aa TadA Mutant 32 aa Sp nCas9 (D10A)		Ryu et al. Nat. Biotechnol.
xABE	TadA 32 aa TadA Mutant 32 aa Sp n xCas9 (D10A)		Hu et al. Nature
ABESa	TadA 32 aa TadA Mutant 32 aa Sa nCas9 (D10A)		Hua et al. Plant Biotechnol. J.
VQR-ABE	TadA 32 aa TadA Mutant 32 aa Sp VQR nCas9 (D10A)		Hua et al. Plant Biotechnol. J. Yang et al. Prot. Cell
VRER-ABE	TadA 32 aa TadA Mutant 32 aa Sp VRER nCas9 (D10A)		Hua et al. Plant Biotechnol. J.
SaKKH-ABE	TadA 32 aa TadA Mutant 32 aa SaKH nCas9 (D10A)		Hua et al. Plant Biotechnol. J. Yang et al. Prot. Cell

Tableau 2 : Taille approximative de la fenêtre d'activité de quelques éditeurs de bases.

La flèche rouge représente le site de coupure généré par la Cas9 nickase. La fenêtre d'activité des différentes versions CBEs est représentée en vert et celle de l'ABE en rouge (l'intensité de coloration est proportionnelle à l'efficacité d'édition). Les éditeurs susceptibles de convertir un C en G, A ou T (transitions + ransversions) sont indiqués en multicolore. Tableau traduit et adapté de Holly A. Rees et al.²²⁴.

6.2.2 Autre modifications non désirées

D'autres études ont également montré que certaines INDELS apparaissent suite à l'utilisation du CBE, avec une fréquence allant parfois jusqu'à plus de 10% au niveau du site cible²²⁵ (phénomène beaucoup plus rare pour l'ABE, inférieur à 0,1% voir indétectable¹⁸⁵). En effet, lorsque le nucléotide U est reconnu puis excisé, le système BER s'active et génère une coupure transitoire du brin ayant subi la déamination. Si la cellule n'a pas eu le temps de réparer la coupure monocaténaire générée par la Cas9 nickase (sur l'autre brin), alors une CDB transitoire peut apparaître, expliquant probablement l'apparition de certaines INDELS^{225,226}. L'équipe du Pr. David R Liu a démontré qu'il était possible de réduire significativement ces INDELS et améliorer la pureté des transitions C→T en fusionnant la protéine Gam dérivée du bactériophage Mu au BE4 (BE4-Gam)²²⁵. Cette protéine Mu est un orthologue à la protéine eucaryote Ku : Mu se lie à l'extrémité des brins d'ADN ayant subi une CDB et les protège contre la dégradation²²⁷. L'outil BE4-Gam a permis d'abaisser le taux d'INDELS à moins de 1.5% dans les cellules HEK293T²²⁵.

Il a également été démontré que certaines transversions peuvent apparaître en présence du CBE (c'est-à-dire, pour le CBE, C•G → [A•T ou G•C]). Il semble que ces transversions soient le résultat d'une réparation erronée de sites abasiques (lésion de l'ADN provoquée par l'absence de base purique ou pyrimidique), qui surviennent parfois lorsque le nucléotide U est excisé par l'UNG2. L'évolution des BE3 en BE4 (ajout d'une copie UGI) a permis d'abaisser la fréquence de ces transversion²²⁵. Une revue décrivant avec précision les mécanismes régissant l'apparition de ces modifications non désirées est disponible²²⁸.

6.2.3 Périmètre d'action et séquence PAM

Bien que les éditeurs de bases caractérisés par une fenêtre d'activité étroite soient optimaux pour limiter l'apparition des *bystanders*, encore faut-il qu'un motif PAM soit présent à la bonne distance de la base cible dans l'ADN génomique. La séquence et la position de ce motif PAM est dépendante de l'espèce bactérienne d'origine de la Cas9. Par exemple, la Cas9 de *Streptococcus pyogenes* (SpCas9, la plus utilisée) reconnaît une séquence PAM 5'-NGG-3' directement localisée aval de la région reconnue par l'ARN guide (sur le brin non ciblé par l'ARN guide)⁶¹. Bien que la SpCas9 soit très efficace dans les cellules mammaliennes, cette limite réduit considérablement son périmètre d'action dans l'ADN génomique. **Ainsi, seulement environ 26% des**

transitions pathogènes de l'ADN (C→T, A→G, T→C ou G→A) peuvent être corrigés par les éditeurs de bases dérivant de la SpCas9 naturelle²²⁹. Ce constat a poussé les scientifiques à développer de nouvelles Cas9 capables de reconnaître et tolérer un spectre beaucoup plus large de séquences PAM.

La première stratégie déployée par les scientifiques a consisté à explorer la diversité naturelle des enzymes Cas9. Par exemple, la Cas9 provenant de *Staphylococcus aureus* (SauCas9) reconnaît une séquence PAM de consensus 5'-NNGRRT-3' (ou R = A ou G)²³⁰, celle provenant de *Neisseria meningitidis* (NmeCas9) requiert un motif beaucoup plus large 5'-NNNGATT-3'²³¹ et enfin la Cas9 de *Streptococcus canis* (ScCas9) reconnaît un motif plus flexible 5'-NNG-3'²³² (pour une description plus complète des orthologues Cas, voir la revue de Daphne Collias et al.²³³).

La deuxième stratégie repose sur l'évolution dirigée des protéines. La première démonstration a été publiée en 2015 par Benjamin P Kleinstiver et al. L'équipe a généré une bibliothèque de mutants aléatoires SpCas9 centrés sur son domaine d'interaction avec le motif PAM de la SpCas9, préalablement cartographié par cristallographie¹⁵¹. Cette librairie a été transférée dans un système bactérien puis les variants d'intérêts ont été sélectionnés selon leur capacité à reconnaître une nouvelle séquence PAM 5'-NGA-3' pour cliver et donc inactiver un gène toxique (sélection positive basée sur la résistance). Parmi les variants retenus, un deuxième tour de sélection a été effectué afin d'identifier ceux capables de discriminer l'ancien motif 5'-NGG-3'. Ainsi, deux variants ont été identifiés : SpCas9 VQR (D1135V/R1335Q/T1337R) et SpCas9 EQR (D1135E/R1335Q/T1337R)⁶². En 2018 et 2020, l'équipe de David R. Liu a généré de nouveaux variants encore plus flexibles en utilisant une stratégie par évolution continue médiée par le phage, beaucoup plus rapide²³⁴ : xCas9 reconnaît les PAM 5'-NG-3', 5'-GAA-3' et 5'-GAT-3'²²⁹ et SpCas9-NG reconnaît les PAM 5'-NRNH-3' (ou R = A or G et H = A, C or T)²³⁵. Afin de repousser les limites de l'évolution dirigée des protéines, l'équipe de Benjamin P. Kleinstiver a répliqué en sélectionnant **un variant SpCas9 capable de reconnaître quasiment toutes les séquences PAM**. En passant au crible une collection de variants positionnés de manière rationnelle, l'équipe a isolé le mutant SpRY reconnaissant à la fois les PAM 5'-NRN-3' et 5'-NYN-3' (R = A ou G et Y = C ou T), avec tout de même une préférence pour les PAM 5'-NRN-3' (NRN>NYN)²³⁶. **Aujourd'hui, ce variant SpRY est celui qui offre la plus grande flexibilité.** A noter que SpRY a également

tendance à générer plus de coupures hors de la cible en comparaison à la SpCas9 sauvage, phénomène attribuable au spectre beaucoup plus large de séquences PAM reconnues par ce variant²³⁶. Cependant, cette même équipe a également mis au point un variant haute-fidélité (SpCas9-HF1) permettant d'éliminer quasiment tous les évènements de coupures hors cible¹⁵³, qui s'est également montré très spécifique en combinaison avec le variant SpRY²³⁶ (cf. partie 4.5-coupures en dehors de la cible).

Ces nouveaux variants SpCas9 avec PAM élargie ont bien entendu été intégrés aux éditeurs de bases ABE et CBE et ont permis d'améliorer significativement le pourcentage de cibles thérapeutiques accessibles²³⁶ : **au lieu de 26% initialement²²⁹, ces nouveaux éditeurs de bases permettent d'atteindre environ 95% des transitions pathogènes responsables de maladies génétiques décrites dans ClinVar (C→T, A→G, T→C ou G→A)²²⁶.**

6.2.4 Modifications hors de la cible

En plus des *bystanders*, les éditeurs de bases sont susceptibles de générer des déaminations non désirées hors de la cible, communément appelées modifications *off-targets*.

Deux types de modifications hors-cibles ont été décrites pour les éditeurs de bases. Premièrement, celles qui sont dépendantes de la séquence de l'ARN guide, qui surviennent lorsque l'ARN guide se fixe sur une séquence nucléotidique similaire à la séquence cible (cf. partie 4.5-coupures en dehors de la cible)²³⁹. Deuxièmement, celles qui sont indépendantes de la séquence de l'ARN guide. Notamment, l'expression du CBE favorise, à l'échelle du génome, la déamination aléatoire de certains nucléotides facilement accessibles^{177,178}. Comme il s'agit d'un évènement rare (fréquence estimée à 5×10^{-8} par paire de base dans des embryons murins)²²⁶, il est difficile à différencier de la déamination naturelle spontanée dont la fréquence d'apparition dans les cellules est également très faible (100 à 500 cytosines par cellule et par jour, soit entre 3 et 15×10^{-8} déamination par cellule et par jour)²³⁷.

La plupart des techniques de détection de l'activité hors cible des systèmes CRISPR/Cas se basent sur la détection des CDB^{122,123,127}. Or, les éditeurs de bases fonctionnent différemment (incapables de générer des CDB). Des méthodes adaptées ont été mises au point afin d'étudier leur spécificité à l'échelle du génome^{238,239}. Par

exemple, Daesik Kim et al. ont revisité le Digenome-seq (*in vitro*). En incubant un substrat ADN avec le CBE dépourvu du domaine UGI (BE3 Δ UGI), des transitions de C vers U accompagnées de coupures monocaténares sont apparues dans l'ADN. Un cocktail d'enzyme appelé USER (uracil-specific excision reagent) a été appliqué à la solution afin de générer des CDB uniquement au niveau des sites contenant des uraciles²⁴⁰. Les sites de cassures ont ensuite facilement été identifiés par séquençage à haut débit. Une stratégie similaire, appelée EndoV-seq, basée sur la détection des inosines dans l'ADN *in vitro*, a été utilisée pour évaluer la spécificité de l'ABE7.10²⁴¹. En comparant la spécificité de la SpCas9 sauvage à celle des éditeurs de base, ces deux groupes ont démontrés que la plupart des sites hors-cible retrouvés étaient communs aux deux types d'enzymes, mais qu'il existait également des sites exclusivement associés à la SpCas9 ou aux éditeurs de base seuls, démontrant l'importance d'une évaluation indépendante et adaptée pour chaque type d'outil utilisé. Ces deux groupes de chercheurs ont affirmés qu'en général, les déaminases étaient plus spécifiques que les nucléases Cas9²³⁸⁻²⁴⁰. En associant des variants Cas9 haute-fidélité et/ou des ARN guides optimisés avec les éditeurs de base, la spécificité de du BE a été d'autant plus améliorée^{222,240,241}.

Julian Grünewald et al. ont également démontré que l'outil CBE (BE3) présentait une activité hors-cible importante sur l'ARN, atteignant près de la moitié des transcrits²⁴². Cette activité semble être indépendante de la séquence de l'ARN guide. La déaminase APOBEC1 agissant naturellement sur un substrat ARN, il n'est pas si étonnant que cette activité, bien que transitoire, soit détectée. Comme l'enzyme TadA, associée à l'ABE, présente une activité sur l'ARN très faible (environ 0.2%) comparée à APOBEC1²²⁶, l'équipe de Nicole M. Gaudelli a eu l'idée de générer un nouveau variant CBE utilisant la déaminase TadA (CBE-Ts). L'évolution dirigée chez la bactérie a permis de créer dans un premier temps des variants TadA catalysant à la fois les deux types de transversion C \rightarrow T et A \rightarrow G (appelé CABE-Ts). Puis les données cristallographiques ont permis de générer un variant CBE-T, conservant l'activité C \rightarrow T mais dénué de l'activité A \rightarrow G. Ce nouveau variant a permis de réduire considérablement, voire supprimer, l'activité hors cible indépendante de la séquence de l'ARN guide du CBE, à la fois sur l'ADN et l'ARN¹⁸⁰.

6.2.5 Vers les futurs éditeurs de base à transversion

Les outils ABE et CBE étant très efficaces pour générer des transitions (C→T, A→G, T→C ou G→A), d'autres équipes ont tenté d'optimiser de nouveaux éditeurs de base susceptibles de générer des transversions, notamment C→G (**CGBE**)²⁴³⁻²⁴⁵. Brièvement, les équipes de J Keith Joung et Xueli Zhang ont associé la Cas9 nickase avec la cytidine déaminase du rat APOBEC1 et la protéine d'excision de l'uracile N-glycosylase (UNG) dérivée d'*Escherichia coli* pour créer le CGBE1 (C→G). Selon l'hypothèse de travail de ces équipes, l'excision de l'uracile par l'UNG génère un site abasique dans l'ADN qui sera préférentiellement remplacé par un G (mécanisme de réparation/réplication inconnu)^{243,245}. Cette preuve de concept a permis de générer des transversions C→G jusqu'à plus de 60% selon les guides utilisés. Toutefois, un nombre très important de *bystander* C→T et C→A était également présent au site cible (faible pureté). L'équipe de Xueli Zhang a également démontré que chez *Escherichia coli*, la construction AID-nCas9-Ung permettait d'atteindre des éditions C→A (**CABE**) avec une efficacité de d'environ 87%²⁴⁵.

Quelques mois plus tard, l'équipe de David R. Liu a conduit un criblage par CRISPR interférence²⁴⁸ afin d'identifier les gènes de réparation de l'ADN améliorant l'efficacité et la pureté des transversions C→G (**CGBE**). Trois candidats ont été identifiés puis testés en fusion avec le BE4 associé à l'UNG : POLD2, EXO1 et RBMX. Comme les résultats obtenus étaient encourageants mais pas suffisants (pureté légèrement améliorée, d'environ 10% pour certains loci), un travail colossal a été réalisé : 95 architectures CGBE ont été testées sur 8 loci dans les cellules HEK293T et les 10 meilleures constructions ont été testées à grande échelle (10,638 ARN guides) dans des cellules vivantes. Les données de séquençage obtenues ont permis d'entraîner des modèles d'apprentissage automatique (*machine-learning*). Ces modèles sont désormais capables de prédire avec précision l'efficacité et la pureté de l'installation spécifique d'une transversion C→G (R=0,90)²⁴⁸ (<https://www.crisprbehive.design/>). Une autre équipe a également conçu un outil similaire²⁴⁶ (http://www.sunlab.fun:3838/BE_SMART/). Même si les outils CGBE ne sont pas encore optimaux, grâce à ces outils bio-informatiques, il est possible de choisir à l'avance l'ARN guide et la construction CGBE susceptibles d'aboutir au résultat escompté.

Début 2023, l'équipe de Hui Yang a également mis au point l'**AYBE**, permettant de réaliser des transversions $A \rightarrow Y$, ($Y=C$ ou T). En supposant que l'excision de la deoxyinosine produite par l'adénosine déaminase TadA permettrait de diversifier les évènements d'édition, ils ont fusionné l'ABE8e avec la protéine d'excision de l'hypoxanthine N-méthylpurine glycosylase (MPG) humaine. Cette construction a permis d'atteindre des niveaux d'édition $A \rightarrow Y$ jusqu'à 72%. A noter que les évènements $A \rightarrow T$ sont plus fréquents que les $A \rightarrow C$ ²⁴⁷.

Désormais, l'ABE, le CBE, le CGBE, le CABE et l'AYBE permettent de réaliser sur le papier n'importe quel type de conversion des bases : à la fois les transitions et les transversions, élargissant encore un peu plus la boîte à outils CRISPR/Cas (figure 6). A noter que les outils CABE, AYBE et CGBE doivent tout de même être considérablement améliorés (efficacité et spécificité) pour atteindre le niveau de l'ABE et du CBE.

En parallèle de ces conversions de bases de l'ADN, les scientifiques se sont demandées comment introduire très précisément de petite insertions ou délétions dans l'ADN, sans générer de CDB, afin de corriger quasiment n'importe quel variant responsable d'une maladie génétique.

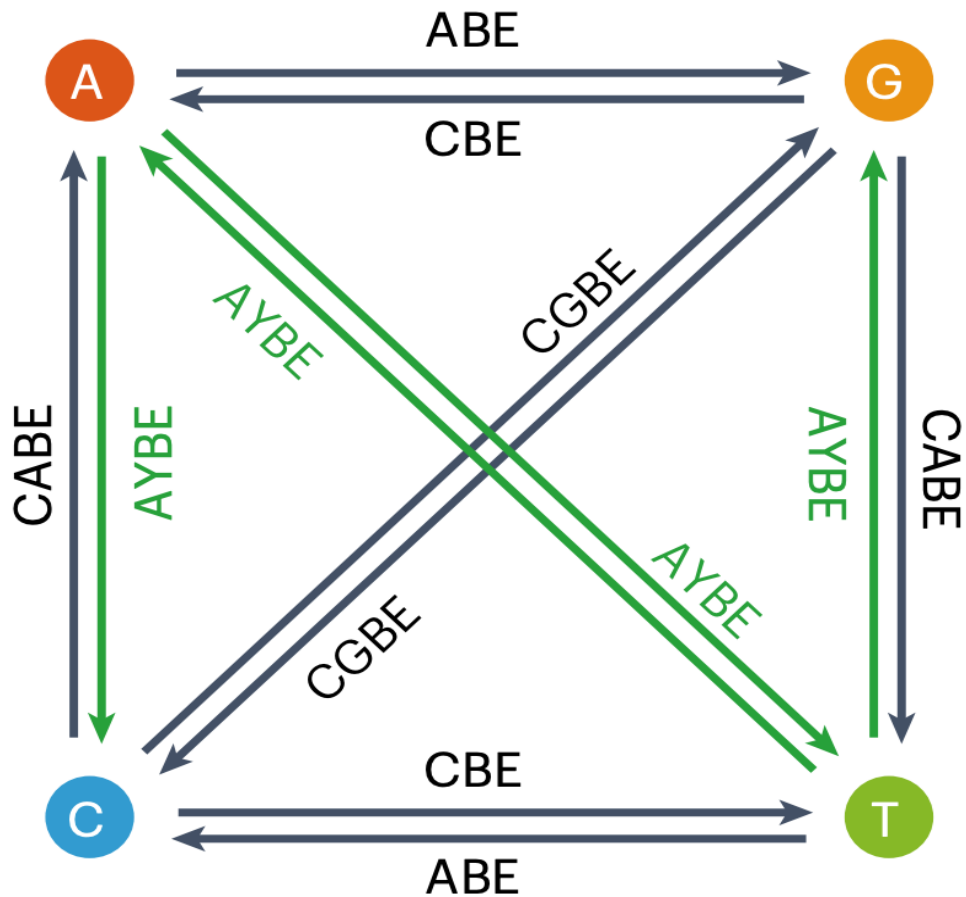


Figure 6 : Les différents types d'éditeurs de bases.

Schéma provenant de Huawei Tong et al.²⁴⁷.

7. CRISPR 3.0 : le *prime editing*

La technologie d'édition du génome *prime editing* (PE) a été publiée en 2019 par Andrew V. Anzalone et al.²⁴⁸. Cette technologie permet d'introduire non seulement les 12 types de conversions de l'ADN, mais également des petites insertions (jusqu'à +80 paires de bases) et délétions (jusqu'à -44 paires de bases) pour réécrire l'ADN à *façon*.

7.1 Mécanisme du PE

L'architecture du PE repose sur la fusion entre une Cas9 nickase (H840A) et une rétrotranscriptase (RT) virale, tous deux associés à un ARN guide PE (pegRNA) mesurant de quelques dizaines à centaines de nucléotides. Le pegRNA comporte 3

régions : la région *spacer* (portion 5'), permettant de guider la Cas9 jusqu'à sa cible ; la région *scaffold* (crRNA, séquence universelle), servant d'adaptateur pour la Cas9, et la région *extension* (portion 3') codant notamment pour l'*édit*, c'est-à-dire la modification génétique souhaitant être introduite dans l'ADN (**figure 7**). La mécanistique du PE est relativement complexe (**figure 7**), et les acteurs intervenant dans l'installation d'une modification génétique d'intérêt n'ont pas encore été tous identifiés.

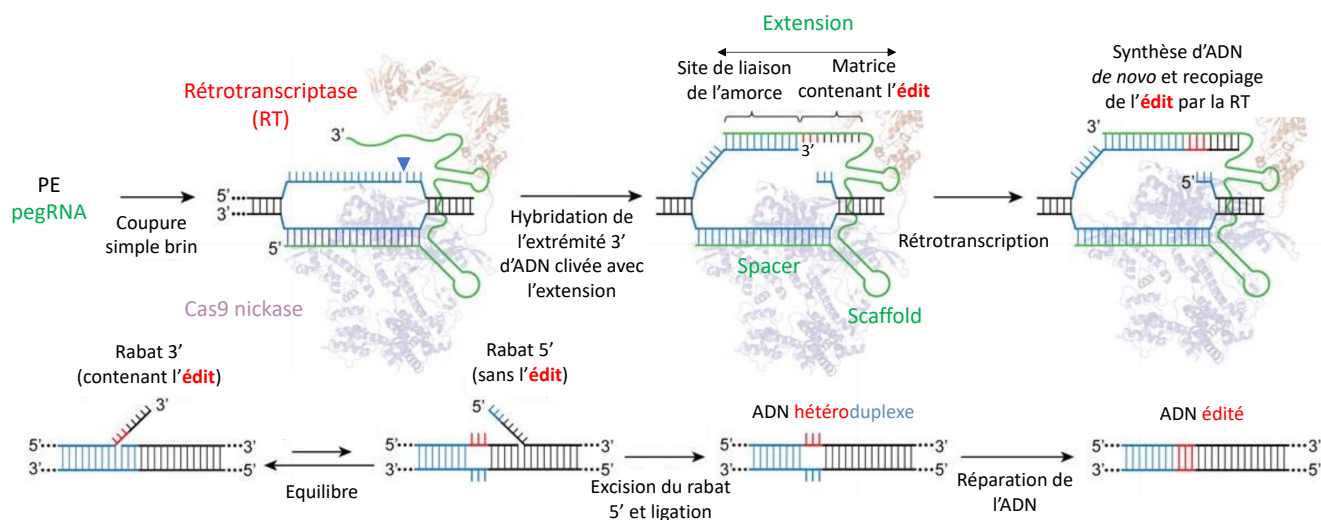


Figure 7 : Mécanisme d'action du *prime editing*.

Le prime editor (PE) est composé d'une Cas9 nickase fusionnée à une reverse transcriptase (RT), tous deux associés à un long ARN guide PE (pegRNA = *spacer* + *scaffold* + *extension*). Suite à la liaison du complexe RNP sur l'ADN, la Cas9 nickase réalise une coupure monocaténaire sur le brin non ciblé par le *spacer*. L'extrémité 3' d'ADN libérée par ce clivage s'hybride au site de liaison à l'amorce présent dans l'*extension* du pegRNA. Cette extrémité est utilisée comme amorce pour initier la réaction de rétrotranscription, qui permet le recopiage de la matrice d'ARN contenant l'*édit*. Suite à cette réaction, un équilibre est obtenu entre les rabats 3' et les rabats 5'. C'est l'excision du rabat 5' par les nucléases FEN1 et EXO1, suivi de la ligation de l'ADN, qui permet de refermer la coupure monocaténaire pour faciliter l'installation de l'édit. L'ADN est alors formé de deux brins non compatibles (ADN hétéroduplexe), dont l'un comporte l'*édit* et l'autre la séquence originale. Deux possibilités permettent de résoudre cet hétéroduplexe d'ADN : la réplication de l'ADN (mécanisme passif) ou la voie de réparation des mésappariements de l'ADN (MMR), à condition que le mésappariement soit de petite taille, c'est-à-dire inférieur à 13 nucléotides s'il s'agit d'INDELS (mécanisme actif). Dans ce cas, soit le brin édité est réparé, soit le brin non édité est réparé, et dans les deux cas le brin d'ADN opposé est utilisé comme matrice de réparation. Schéma traduit et adapté d'Andrew V. Anzalone et al.²⁴⁸.

7.2 Origine des différentes versions PE

Dans leur publication, Andrew V. Anzalone et al. ont successivement mis au point 3 générations PE²⁴⁸. La première génération (**PE1**), composée de la Cas9 nickase fusionnée à la RT du virus de la leucémie murine (M-MLV), s'est révélée très peu efficace dans les cellules HEK293T (inférieur à 5% d'édition en moyenne). Afin d'améliorer cette première génération, 5 mutations ont été introduites dans la rétrotranscriptase M-MLV afin d'en améliorer la thermostabilité, la processivité et l'affinité pour son substrat (D200N, L603W, T330P, T306K, W313F). En moyenne, l'efficacité d'édition a été multipliée par 3 avec cette seconde version (**PE2**). Pour terminer, un ARN guide supplémentaire ciblant le brin non édité (40 à 100 paires de bases en amont du site de coupure du pegRNA) a été ajouté au PE2 pour former la troisième génération (**PE3**). En introduisant une coupure monocaténaire supplémentaire sur le brin non édité, les auteurs suggèrent que ce brin serait préférentiellement réparé par MMR lors de l'étape de résolution de l'hétéroduplexe d'ADN, ce qui faciliterait l'incorporation de l'*édit* (comme pour les éditeurs de base). Le système PE3 a permis d'atteindre des niveaux d'édition de 20 à 50% en moyenne. Toutefois la fréquence des INDELS indésirées (*byproduct*) a augmenté et atteint typiquement une fréquence de 1-10%²⁴⁸. Ces pourcentages sont à relativiser car ils ont été obtenus dans les cellules HEK293T dont les mécanismes de réparation sont altérés. Ces mêmes expériences dans d'autres modèles donnent des résultats globalement décevants. Cette méthode demeure difficile à utiliser avec succès car les règles dictant la conception des pegRNAs sont encore floues.

7.3 Améliorer l'efficacité du PE

7.3.1 Agir sur les voies de réparation de l'ADN

Afin d'identifier les déterminants majeurs affectants le rendements PE, deux équipes indépendantes ont réalisé un criblage génétique^{249,250}. Ces deux équipes ont identifié le même candidat : l'inhibition de la voie MMR, et plus particulièrement le gène *MLH1*, augmente indéniablement l'efficacité du PE. J. Ferreira da Silva et al. ont même démontré que les protéines de la voie de réparation MMR se localisaient au niveau des régions d'ADN ciblées par le PE pour contrecarrer l'installation de l'*édit*²⁵⁰. Au premier abord, cela peut sembler paradoxal puisque la voie de réparation MMR s'avère être impliquée dans la résolution de l'ADN hétéroduplexe. En fait, ces deux équipes suggèrent qu'une fois la réaction de rétrotranscription terminée, cette voie de

réparation pourrait préférentiellement rejeter l'hybridation du rabat 3' (contenant l'édit) sur le brin non édité ou bien stimuler sa dégradation sélective afin de faciliter le rétablissement de la séquence originale.

Par conséquent, Peter J. Chen et al. ont démontré que la co-transfection d'un plasmide codant pour un dominant négatif de MLH1 (MLH1dn) avec le système PE était efficace pour inhiber transitoirement la réparation par MMR et améliorer le rendement du système PE de 2 à 7,7 fois en moyenne dans des cellules en culture. De plus, le ratio édit/INDEL (pureté) a été amélioré d'environ 3,4 fois²⁴⁹. Il est intéressant de noter que dans certaines lignées cellulaires cancéreuses, le système MMR est déjà inactif. Par exemple, la lignée cellulaire HEK293T présente une hyperméthylation au promoteur du gène *MLH1*. Ces différences peuvent expliquer certaines divergences d'efficacité du PE selon le type de cellules utilisées^{248,249}.

De plus, il existerait certaines classes d'*édits* beaucoup moins vulnérables à l'activité MMR, pouvant être intégrées beaucoup plus efficacement dans l'ADN. En effet, la voie MMR reconnaît seulement les hétéroduplexes de petites tailles (inférieurs à une dizaine de nucléotides). Une stratégie intéressante consisterait à ajouter des mutations silencieuses autour de l'édit afin d'augmenter la complexité des mésappariements avec le brin non édité et compromettre la reconnaissance de l'hétéroduplexe par la voie MMR²⁴⁹. De plus, cela permet de réduire l'auto complémentarité entre le *spacer* et l'*extension*.

7.3.2 Optimiser l'architecture du PE

Pour installer une modification génétique, le PE doit tout d'abord s'exprimer dans la cellule, être transporté dans le noyau, s'associer avec le *pegRNA*, atteindre sa cible *via* une PAM puis occasionner la coupure simple brin avant d'initier la réaction de rétrotranscription. De nombreux groupes ont proposé plusieurs stratégies (**tableau 3**) pour améliorer l'ensemble de ces étapes.

Construction	Architecture	Description	Refs.
PE1	NLS ^{SV40} Sp Cas9 (H840A) 33 aa MMLV RT NLS ^{SV40}	PE original, PAM NGG	PMID: 31634902
PE2	NLS ^{SV40} Sp Cas9 (H840A) 33 aa MMLV RT mutant* NLS ^{SV40}	Amélioration de l'activité de la RT MMLV	PMID: 31634902
PE2*	NLS ^{-Myc} NLS ^{SV40} Sp Cas9 (H840A) 33 aa MMLV RT mutant* vNLS ^{SV40} NLS ^{SV40}	Meilleure localisation nucléaire	PMID: 33837189
PEmax	NLS ^{SV40} Sp Cas9 (R221K N394K H840A) 8 aa NLS ^{SV40} 8 aa MMLV RT mutant* NLS ^{SV40} NLS ^{-Myc}	Meilleure localisation nucléaire, expression et nick	PMID: 34653350
CMP-PE1-V1	NLS ^{SV40} HN1 Sp Cas9 (H840A) H1G 33 aa MMLV RT mutant* NLS ^{SV40}	Meilleure accessibilité à la chromatine	PMID: 34082781
hyPE2	NLS ^{SV40} Sp Cas9 (H840A) 16 aa Rad51 DBD 33 aa MMLV RT mutant* NLS ^{SV40}	Meilleure activité	PMID: 34556671
IN-PE	IGF1p NFATC2IPp 33 aa NLS ^{SV40} Sp Cas9 (H840A) 33 aa MMLV RT mutant* NLS ^{SV40}	Meilleure expression	PMID: 35717416
ePPE	NLS ^{SV40} Sp Cas9 (H840A) 16aa NC NLS ^{SV40} 32 aa MMLV RT mutant* ΔRNaseH NLS ^{SV40}	Meilleure activité chez les plantes	PMID: 35332341
PE-P3	NLS ^{SV40} MMLV RT mutant* 33 aa Sp Cas9 (H840A) NLS ^{SV40}	Meilleure activité chez les plantes	PMID: 34949802
PE2 ^{ΔRh}	NLS ^{SV40} Sp Cas9 (H840A) 33 aa MMLV RT mutant* ΔRNaseH NLS ^{SV40}	PE plus petit	PMID: 35332341, PMID: 35294257
sPE	NLS ^{SV40} Sp Cas9 (H840A) NLS ^{SV40} NLS ^{SV40} MMLV RT mutant* NLS ^{SV40}	Cas9 nickase et RT séparées	PMID: 35379962
nCas9 and MCP-RT	NLS ^{SV40} Sp Cas9 (H840A) NLS ^{SV40} MS2CP MMLV RT mutant* NLS ^{SV40}	Cas9 nickase et RT séparées	PMID: 35379962
PE4/PE5	NLS ^{SV40} Sp Cas9 (H840A) 33 aa MMLV RT mutant* NLS ^{SV40} MLH1dn	Meilleure résolution du PE	PMID: 34653350
PE2-VQR	NLS ^{SV40} Sp VQR Cas9 (H840A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NGA	PMID: 33636398
PE2-VRQR	NLS ^{SV40} Sp VRQR Cas9 (H840A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NGA	PMID: 33636398
PE2-VRER	NLS ^{SV40} Sp VRER Cas9 (H840A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NGCG	PMID: 33636398
PE2-NG	NLS ^{SV40} Sp NG Cas9 (H840A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NGN	PMID: 33636398
PE2-SpG	NLS ^{SV40} SpG Cas9 (H840A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NGN	PMID: 33636398
PE2-SpRY	NLS ^{SV40} SpRY Cas9 (H840A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NRN	PMID: 33636398
SaPE2	NLS ^{SV40} Sa Cas9 (N580A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NNGRRT PE plus petit	PMID: 34653350, PMID: 33837189
SaPE2*	NLS ^{-Myc} NLS ^{SV40} SaKKH Cas9 (N580A) 33 aa MMLV RT mutant* vNLS ^{SV40} NLS ^{SV40}	PAM NNGRRT Meilleure localisation nucléaire	PMID: 33837189
Sa ^{KKH} PE2	NLS ^{SV40} SaKKH Cas9 (N580A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NNNRRT PE plus petit	PMID: 33837189, PMID: 35294257
Sa ^{KKH} PE2*	NLS ^{-Myc} NLS ^{SV40} SaKKH Cas9 (N580A) 33 aa MMLV RT mutant* vNLS ^{SV40} NLS ^{SV40}	PAM NNNRRT Meilleure localisation nucléaire	PMID: 33837189
SauriCas9-PE	NLS ^{SV40} Sauri Cas9 (N588A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NNGG PE plus petit	PMID: 35294257
CjCas9-PE	NLS ^{SV40} Cj Cas9 (H559A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NNNRYAC PE plus petit	PMID: 35294257
FnCas9-PE	NLS ^{SV40} Fn Cas9 (H969A) 33 aa MMLV RT mutant* NLS ^{SV40}	PAM NGG Clive l'ADN à 6/7 nucléotides du PAM	PMID: 35410288

Tableau 3 : Les différents types d'architectures PE.

Tableau traduit de Peter J. Chen et al.²⁵¹.

Par exemple, Peter J. Chen et al., ont amélioré le PE2 en optimisant les codons de la RT, en ajoutant plusieurs signaux de localisation nucléaire et en produisant deux mutations dans la Cas9 (R221K & N394K) permettant d'améliorer son activité nucléase. Cette nouvelle architecture (**PEmax**), a devancé les versions précédentes PE2 et PE2*, et constitue à ce jour probablement l'une des meilleures constructions pour s'initier au PE dans des cellules humaines²⁵¹.

Une limite importante associée au PE est également sa taille très imposante (peut atteindre plus de 6kb), alors que la Cas9 seule ne mesure que 4kb environ. Lorsque des plasmides sont utilisés, cette contrainte peut affecter l'efficacité de transfection des cellules. Afin de remédier à ce problème, plusieurs groupes se sont concentrés à réduire la taille du PE, notamment en supprimant le domaine RNaseH de la RT (**PE2^{ARNH}**) ou en utilisant des variants Cas9 de petites tailles (SaCas9, SaCas9KKH) associés à d'autres motifs PAM^{252,253}. Au contraire, d'autres ont cherché à séparer la RT de la Cas9 pour leur offrir plus de flexibilité et surtout permettre l'encapsulation du PE dans des vecteurs viraux adéno-associés (AAV) pour faciliter son administration *in vivo*²⁵⁴⁻²⁵⁶ à des fins thérapeutiques.

Plusieurs équipes ont également démontré que l'efficacité du PE pouvait être améliorée en fusionnant au PE certains domaines de liaison à l'ADN, tel que celui de Rad51 (**hyPE2**)²⁵⁷, ou bien en favorisant le recrutement de facteurs de transcription pour faciliter l'ouverture de la chromatine (**PE+MS2-p65**)²⁵⁸. En utilisant une approche de criblage à haut débit, Minja Velimirovic et al. ont évalué l'effet de plus de 12 000 peptides de 85 acides aminés fusionnés à la Cas9, et ont identifié deux peptides très performants permettant d'améliorer l'efficacité du PE à travers plusieurs lignées cellulaires et loci (**IN-PE**)²⁵⁹.

7.3.3 Optimiser la conception des *pegRNA*

Un autre déterminant majeur de l'efficacité du PE repose sur la conception des *pegRNA*. Le *pegRNA* s'articule en trois régions : la région *spacer* (portion 5'), la région *scaffold* (*crRNA*, séquence universelle), et la région *extension* (portion 3') (**figure 7**). La région *extension* est elle-même composée du site de liaison de l'amorce (PBS) et de la matrice de rétrotranscription (*RT template*), contenant l'*édit* (**figure 7**). Pour une séquence *spacer* donnée de 20 nucléotides, une multitude de combinaisons de séquences *extension* est possible. En effet, il est possible de faire varier la taille et la

position du *PBS* et du *RT template*. Plusieurs outils en ligne d'aide à la conception des pegRNA sont disponibles (<https://primedesign.pinellolab.partners.org/>, <http://www.rgenome.net/pe-designer/>, <http://easy-prime.cc/>, <https://fgcz-shiny.uzh.ch/PnBDesigner/>, <http://pegfinder.sidichenlab.org/>)^{260–264}. Les données de la littérature démontrent qu'il faut tester un nombre très important de constructions, puisque seulement une petite fraction de ces constructions va permettre d'obtenir des rendements PE satisfaisants.

Pour réduire ce travail expérimental long et fastidieux, Hui Kwon Kim et al. ont testé 54,836 paires de pegRNA dans les cellules HEK293T (environ 2000 séquences cibles x 24 combinaisons de *PBS* et *RT templates* par séquence cible + 6,800 autres pegRNA), afin d'identifier par bio-informatique les combinaisons associées à un bon rendement PE²⁶⁵.

A l'issu de ce travail, plusieurs recommandations ont été établies :

- 1) Privilégier une *PBS* de 13 nucléotides et un *RT template* de 12 nucléotides
- 2) Privilégier une séquence *PBS* avec une teneur en GC élevée
- 3) Si la taille du *RT template* est ≤ 12 nucléotides, privilégier le positionnement d'une guanine en dernière position
- 4) Privilégier une modification génétique incluant la suppression du motif PAM

Cette équipe a également créé un modèle de calcul informatique basé sur l'apprentissage profond (DeepPE) prédisant l'efficacité du PE2 pour un pegRNA donné ($R \approx 0,70$)²⁶⁵ (<http://deepcrispr.info/DeepPE/>), mais cet outil ne fonctionne que pour les pegRNA codant pour les transversions C→G. Dans une étude plus récente d'avril 2023, un outil similaire encore plus complet a été mis au point²⁶⁶ (<http://deepcrispr.info/DeepPrime/>)

7.3.4 Optimiser l'expression et la stabilité des pegRNA

Dans la grande majorité des constructions plasmidiques, les ARN guides et pegRNA sont transcrits dans les cellules à partir d'un promoteur U6, dépendant de l'ARN polymérase III (promoteur fort), dont le signal de terminaison de la transcription correspond à une extension de 5 thymines ou plus. Or, la région *scaffold* des pegRNA comporte une suite de 4 thymines, pouvant être considérée comme un signal putatif de terminaison de la transcription. Quelques équipes ont démontré que l'interruption

de ce motif améliorerait l'expression des pegRNA^{267,268}. Lors de la conception du *spacer* et de l'*extension*, il est également important de vérifier l'absence de tels signaux de terminaison.

Un autre critère important concerne la stabilité des pegRNA. L'extrémité 3' du pegRNA, bien que très importante pour initier la réaction de rétrotranscription, est fortement susceptible d'être dégradée par les ribonucléases intracellulaires. Pour y remédier, plusieurs groupes indépendants ont montré que l'ajout de structures secondaires telles que les motifs evopreQ1, mpknot et G-quadruplexe à l'extrémité 3' du pegRNA améliorerait significativement le rendement du PE^{267,269,270}.

7.4 Quantifier les modifications hors de la cible du PE

Contrairement aux autres outils CRISPR/Cas, l'équipe de David R. Liu affirme qu'il est rare que le PE induise des modifications hors de la cible, puisque trois appariements distincts de type « Watson-Crick », agissant comme des points de contrôles, sont nécessaires avant d'aboutir à une édition. Premièrement, la région *spacer* du pegRNA doit s'hybrider spécifiquement à sa cible pour que la Cas9 induise une coupure monocaténaire. Deuxièmement, l'extrémité 3' du brin d'ADN clivé a besoin de s'apparier avec le PBS afin d'initier la réaction de polymérisation. Troisièmement, le rabat 3' nouvellement synthétisé doit impérativement s'hybrider avec le brin d'ADN intact pour que la modification génétique puisse s'installer définitivement dans l'ADN²⁵¹.

Cette affirmation doit être nuancée : il ne suffit pas de passer les 3 points de contrôles pour qu'un évènement hors-cible apparaisse. En effet, dès le premier évènement (cassure monocaténaire) généré par la Cas9 nickase H840A, une activité hors cible dépendante de la séquence du *spacer* peut être détectée. D'ailleurs, Do Yon Kim et al. ont développé et testé la méthode nDigenome-seq pour localiser les coupures monocaténaires induites par la Cas9 nickase H840A *in vitro* (étape d'exploration). Les modifications hors-cible ont ensuite été quantifiées par séquençage d'amplicon profond dans des cellules HEK293T (étape de validation). Sur 9 pegRNA testés dans les cellules HEK293T, l'équipe a tout de même mis en évidence des évènements hors cibles dans 5 régions de l'ADN (fréquence variant de 0.1 à 1.9%)²⁷¹. Deux équipes ont également démontré que les mésappariements présents dans la région PBS sont

tolérés (entre 1 et 4 mésappariements pour un PBS de 15 nucléotides), surtout si ces mésappariements sont localisés à l'extrémité 3' du PBS (à distance du *RT template*). Ainsi, un PBS de petite taille (11–13 nt) permettrait d'augmenter la spécificité du PE^{271,272}.

Bien que de nombreux groupes affirment n'avoir détecté aucune activité hors cible avec le PE, et cela dans plusieurs modèles (plantes²⁷², organoïdes²⁷³, souris²⁷⁴), ceux-ci ont en fait utilisé la méthode du séquençage du génome complet, impliquant une faible profondeur de lecture. Cette méthode est très peu sensible comparée aux méthodes basées sur l'enrichissement tel que le Digenome-seq par exemple.

Comme il n'existe pas encore de méthode de référence sensible permettant de mettre en évidence l'activité hors cible globale du PE, il est encore trop tôt pour tirer des conclusions définitives, mais il est certain que la spécificité du PE peut être améliorée avec les variants Cas9 haute-fidélité telles que la eSpCas9 et la Sniper Cas9²⁷¹. Il serait intéressant de mettre en place des variants RT haute-fidélité, capable d'initier la réaction de polymérisation seulement si l'amorce est parfaitement appariée au PBS. Afin de quantifier l'activité hors-cible dépendante de l'activité RT, on pourrait imaginer ajouter dans le *RT template* un TAG de quelques nucléotides qui pourrait être enrichi par PCR puis détecté par séquençage.

7.5 Modifications non désirées

Même si l'efficacité et la pureté du PE sont globalement meilleures que les rendements obtenus par la technologie CRISPR/HDR, des modifications non désirées au site cible, très diverses, restent tout de même présentes (**figure 8**). Ces modifications non désirées, appelées *byproducts* dans la littérature, correspondent la plupart du temps à des INDELS (pouvant parfois atteindre plusieurs dizaines de paires de bases) et sont localisées à proximité du site de coupure médié par la Cas9. Moins fréquemment, il est également possible de retrouver une partie de la séquence *scaffold* intégrée au génome, lorsque celle-ci subit l'étape de rétrotranscription (**figure 8**). Les résultats de la figure 8 montrent que trois types d'éditations peuvent être identifiés au site cible : les allèles comportant uniquement la modification désirée (PPE = *Pure Prime Edit*), les allèles comportant à la fois la modification désirée et des *byproduct* (IPE = *Impure Prime Edit*) et les allèles composés uniquement de *byproduct*. Dans cet exemple, la

fréquence des PPE (1.10%), qui est déjà très faible, est même nettement en dessous de la fréquence des IPE+PPE (1.79%).

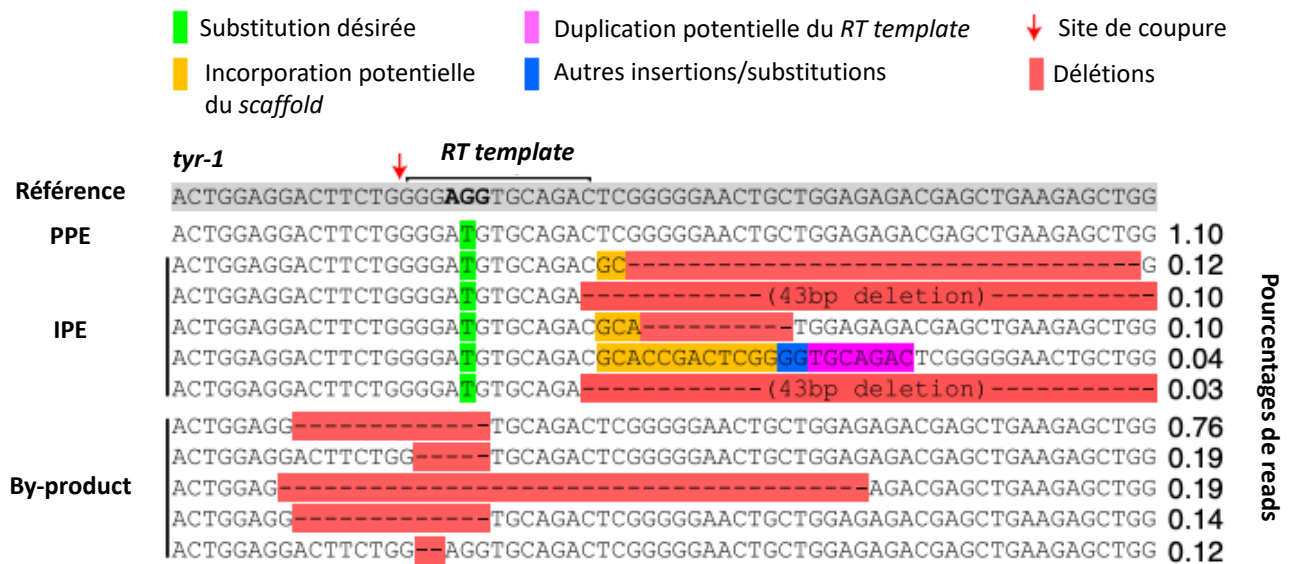


Figure 8 : Efficacité et pureté du PE3 au locus *tyr-1* dans des embryons de poissons zèbres.

Les complexes RNP PE:pegRNA:ngRNA (PE3) ont été injectés dans des embryons de poissons zèbre (stade 1 cellule) afin d'installer une substitution G→T dans le locus *tyr-1*. Après 1 jour d'incubation, 5-10 embryons ont été regroupés pour en extraire l'ADN et l'efficacité d'édition a été évaluée par séquençage d'amplicon centré sur la région ciblée par le pegRNA. Les données de la figure représentent le pourcentage des différents allèles obtenus à l'issue du séquençage. La séquence de référence non éditée est indiquée en gris et le motif PAM en gras. PPE = *Pure Prime Edit* ; IPE = *Impure Prime Edit*. Résultats traduits et adaptés de Karl Petri et al.²⁷⁵.

Afin de réduire la fréquence de ces *byproducts*, l'équipe de David R. Liu suggère d'utiliser la méthode PE2/PE4 au lieu du PE3/PE5. Si ce n'est pas possible car les taux d'édition sont trop faibles, une variante au PE3, appelé **PE3b**, consiste à concevoir un ARN guide s'hybridant spécifiquement sur le brin édité (à condition qu'une séquence PAM soit présente au bon endroit). De cette manière, cette deuxième coupure monocaténaire a lieu seulement si l'hétéroduplexe d'ADN est formé, ce qui permet d'abaisser en moyenne de 13 fois (0.74%) la fréquence des INDELS (versus 1-10% pour le PE3 classique)²⁴⁸.

Depuis sa création, le PE a rapidement été identifié comme une innovation majeure à l'édition à des gènes chez de nombreux modèles et organismes : bactéries²⁷⁶, drosophiles²⁷⁷, poissons zèbres²⁷⁵, plantes^{278,279}, cellules mammaliennes^{248,280}, souris²⁸¹⁻²⁸³ et lapins²⁸⁴. Son utilisation a permis de corriger certains variants génétiques^{281,282}, créer de nouveaux modèles pour l'étude du cancer^{275,285}, marquer des protéines endogènes avec des étiquettes moléculaires pour faciliter leur détection et purification^{248,267} et également de reconstruire des lignages cellulaires²⁸⁶. Bien que la pureté et l'efficacité du PE nécessitent d'être améliorées pour une utilisation clinique, la constante amélioration du système devrait accélérer la compréhension des déterminants du PE afin d'élargir sa portée et résoudre certaines de ses limites.

7.6 Nouvelles versions PE

Seulement quelques mois après la publication du PE *traditionnel* en 2019, d'autres versions utilisant des paires de pegRNA espacées de plusieurs dizaines ou centaines de nucléotides ont été développées, telles que HOPE (homologous 3' extension mediated prime editor)²⁸⁷, PEDAR (PE-Cas9-based deletion and repair)²⁸⁸, twin PE²⁸⁹, GRAND²⁹⁰, PRIME-Del²⁹¹, PETI (prime editor nuclease-mediated translocation and inversion)²⁹² (**figure 9**). Une description détaillée de ces outils est disponible dans cette revue²⁵¹. En comparaison au PE *traditionnel*, ces outils permettent de mettre en place des insertions et délétions de très grandes tailles. Par exemple, le PRIME-Del permet d'installer des délétions jusqu'à 10kb avec une efficacité de 1 à 30%²⁹¹ et le GRAND d'installer des insertions de 250pb avec plus de 28% d'efficacité²⁹⁰. Néanmoins, l'efficacité d'édition semble diminuer lorsque la taille de l'insertion ou de la délétion augmente.

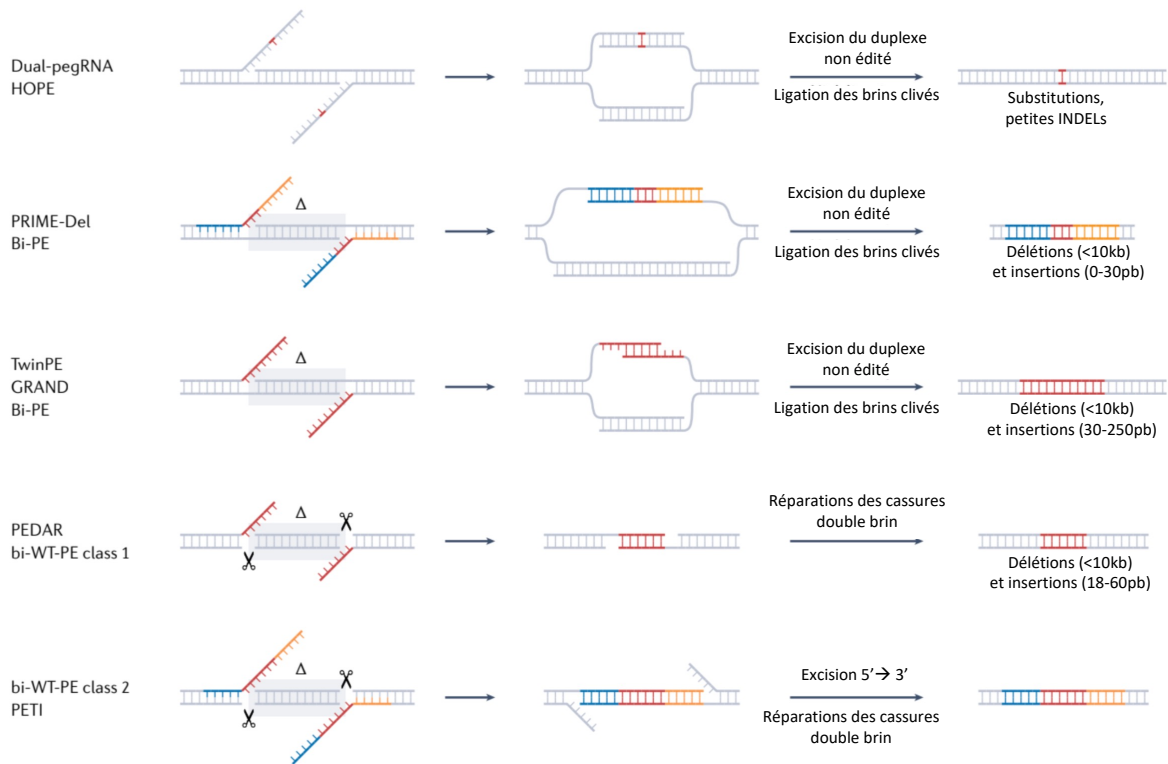


Figure 9 : Nouvelles versions PE utilisant des paires de pegRNA.

L'utilisation de paires de pegRNA permet de générer deux rabats 3' (contenant la modification désirée) complémentaires l'un de l'autre. L'hybridation de ces deux rabats 3' génère un intermédiaire contenant d'une part le duplexe renfermant la modification d'intérêt, et d'autre part le duplexe contenant l'ADN génomique initial, non édité. L'excision du duplexe d'ADN initial (non édité) permet l'installation définitive de la modification d'intérêt. Les nucléotides schématisés en orange et bleus sont homologues à l'ADN génomique initial alors que les rouges représentent la modification génétique d'intérêt (séquence hétérologue). Le delta (Δ) désigne les nucléotides qui seront délétés pour être remplacés par la modification génétique d'intérêt. Schéma traduit et adapté de Peter J. Chen et al.²⁵¹.

8. Transposases et recombinases

Récemment, un dernier type d'outil exploitant les propriétés des transposases et recombinases a été décrit et résumé dans ces revues^{226,293}. Les recombinases catalysent des réactions de recombinaison site spécifique dans l'ADN et permettent notamment de générer des insertions ou inversions (**figure 10**) (ex : sites loxP pour la recombinase Cre). Deux grandes classes de recombinases ont été décrites selon le type d'acide aminé présent dans le domaine catalytique : la famille des sérines recombinases, médiant des réactions irréversibles, et la famille des tyrosines recombinases, régissant des réactions réversibles ou irréversibles selon le type d'enzymes.

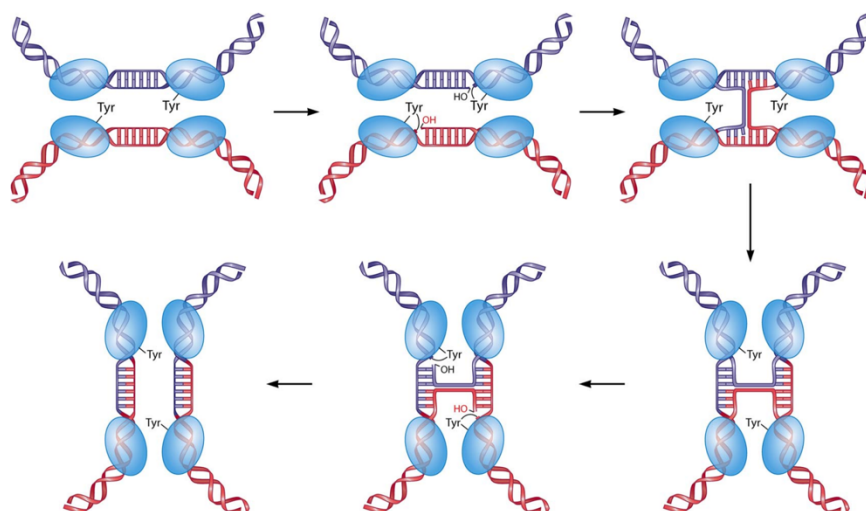


Figure 10 : Mécanisme d'action des tyrosines recombinases.

Classiquement, quatre monomères recombinases, représentés en bleu, se lient spécifiquement à l'extrémité de sites cibles partageant une homologie de séquence (deux monomères à l'extrémité de chaque site cible) et s'associent pour former un complexe pré-synaptique. Les deux monomères actifs (représentés par la tyrosine) génèrent une première cassure monocaténaire à l'intérieur des deux sites permettant l'échange des brins d'ADN. Un changement conformationnel permet d'activer les deux autres monomères qui viennent cliver le second brin d'ADN pour terminer l'échange des brins. Une dernière étape de ligation permet d'obtenir le produit final de recombinaison. Schéma provenant de Lara Rajeev et al.²⁹⁴.

L'exploration de la diversité naturelle des recombinaisons, de même que l'évolution dirigée des protéines a permis d'élargir l'éventail de sites cible reconnus recombinaisons, ouvrant ainsi la voie à l'exploration de nouveaux sites thérapeutiques dans le génome humain²⁹⁵⁻²⁹⁷. Felix Lansing et al. ont d'ailleurs mis au point le système de recombinaison RecF8 permettant de corriger une inversion de 140 kb dans le gène codant le factor VIII de la coagulation, fréquemment retrouvée chez les patients souffrants d'hémophilie sévère. Ce système a atteint une efficacité de correction de 12% dans des cellules dérivées d'un patient, permettant de restaurer l'expression du gène²⁹⁸. Toutefois, comme pour les méganucléases, reprogrammer les sites cibles des recombinaisons s'avère long et fastidieux. Afin de diriger la recombinaison vers davantage de sites, l'équipe de David R. Liu s'est intéressée à fusionner la dCas9 avec la recombinaison Ginβ (site de recombinaison relaxé), mais l'efficacité s'est révélée modeste *in vitro*²⁹⁹. Dans une autre étude, Eleonora I. Ioannidi et al. **ont combiné la technologie PE et la recombinaison site-spécifique pour mettre au point le système PASTE** (*Programmable Addition via Site-specific Targeting Elements*). La technologie PE a tout d'abord permis d'installer les sites de reconnaissance spécifiques de la recombinaison, permettant ensuite à celle-ci de catalyser **l'intégration d'un segment d'ADN allant jusqu'à 36kb à l'endroit désiré, avec des rendements allant de 10 à 50% dans des lignées cellulaires et des cellules primaires humaines**. Les auteurs ont même directement fusionné la recombinaison au PE³⁰⁰. Le système **PASTE** s'avère complexe à mettre en œuvre mais les propriétés des recombinaisons demeurent très intéressantes pour médier des réactions d'insertions, de délétions, de remplacements et d'inversions de séquences. L'accélération de la recherche dans ce domaine permettrait d'atteindre la correction de nouveaux variants génétiques.

De même, les transposases sont des enzymes catalysant le déplacement d'éléments génétiques dans le génome³⁰¹. Ces éléments sont appelés transposons (quelques centaines de pb à plusieurs kb). Brièvement, plusieurs groupes indépendants ont mis en évidence que certains locus CRISPR/Cas ayant perdu leur domaine codant pour l'activité nucléase étaient associés aux gènes transposases. Les études mécanistiques ont démontré d'une part que ces locus agissaient comme des éléments génétiques mobiles (transposons), et d'autre part que les domaines Cas permettent de guider la transposition de manière site-spécifique^{226,302}.

L'équipe de Feng Zhang a ainsi développé le système CAST (*CRISPR-associated transposase*) en utilisant 3 plasmides : **un plasmide helper** comportant les éléments essentiels à la transposition (transposases TniQ, TnsC et TnsB, Cas12K, ARN guide), **un plasmide donneur** comportant le transposon (gène de résistance à la kanamycine flanqué par des extrémités droites et gauches reconnues par les transposases) et un **plasmide cible** comportant le site de liaison à l'ARN guide (**figure 11**). Après transfection des plasmides dans des cellules bactériennes, cette preuve de concept a démontré qu'il est possible de programmer l'insertion site spécifique du transposon kanamycine dans le plasmide cible, avec une efficacité d'environ 12%. La caractérisation du variant CAST de la cyanobactérie *Scytonema hofmanni* (**ShCAST**) a démontré que le site d'insertion du transposon était localisé entre 60 et 66 paires de bases en aval de la séquence PAM et que la taille de l'insertion (du transposon) pouvait facilement atteindre plus de 10kb (**figure 11**). Toutefois, la fréquence des événements d'insertions hors de la cible demeurent très fréquents, notamment à proximité des gènes fortement exprimés chez *E.coli* (ex : gènes ribosomaux) et des événements de cointégration indésirables ont été rapportés par d'autres équipes (duplication du cargo et même insertion du plasmide cible)^{226,302}. Pour remédier à ces problèmes, au début de l'année 2023, l'équipe de Benjamin P. Kleinstiver a fait évoluer le système ShCAST pour créer le système **ShHELIX** (**H**oming **E**ndonuclease-assisted **L**arge-sequence **I**ntegrating **C**AST-comple**X**). Brièvement, la transposase TnsB a été fusionnée avec le variant nickase nAnil (K227M) provenant d'*Aspergillus nidulans*, permettant alors de passer d'un système de transposition répliatif (« copier-coller ») à conservatif (« couper-coller »). En testant ce système chez *E.coli*, les événements de cointégration ont été abaissés de 46.31% à 2.97% (facteur 16), l'efficacité d'insertion au site cible (génom de *E.coli*) a été améliorée de 54.4% à 88.4% (facteur 1.6) et les insertions en dehors de la cible ont été réduites significativement³⁰³. Tout de même, le transfert du système HELIX des cellules procaryotes aux cellules eucaryotes n'a toujours pas été démontré³⁰³.

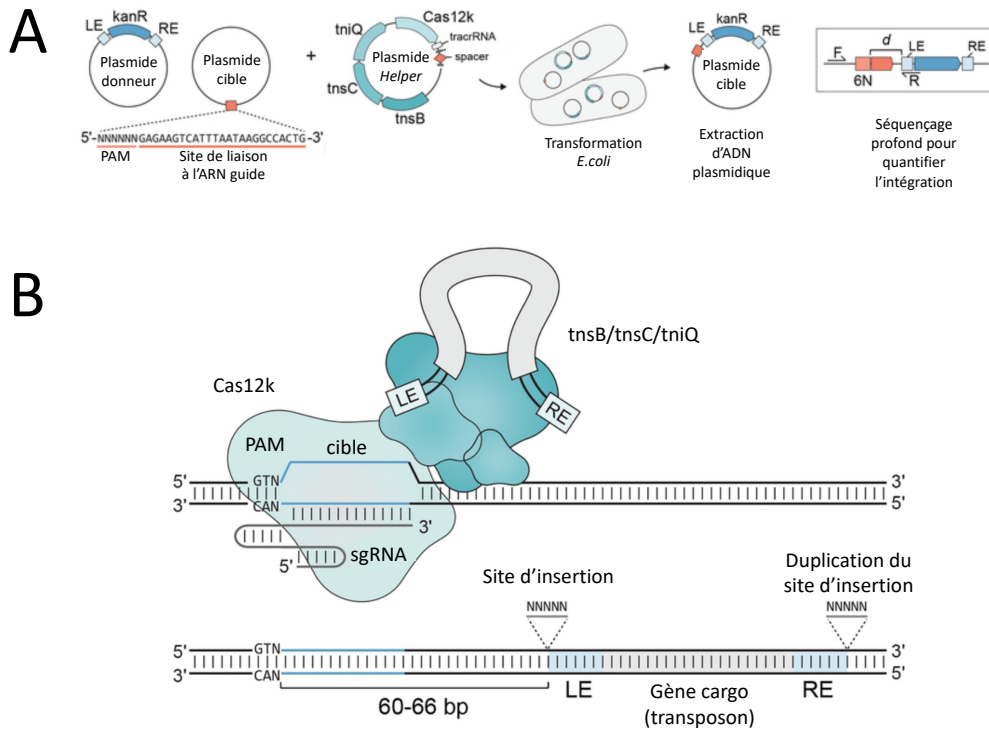


Figure 11 : Preuve de concept et modèle du système ShCAST.

(A) Preuve de concept pour tester l'activité du système ShCAST dans la bactérie *Escherichia coli*. Trois plasmides ont été utilisés puis transformés dans *E.coli* : un plasmide donneur contenant gène de résistance à la kanamycine (transposon) flanqué par des extrémités droites et gauches reconnues par les transposases (RE et LE), un plasmide *helper* comportant les éléments essentiels à la transposition (transposases TniQ, TnsC et TnsB, Cas12K, ARN guide), et un plasmide cible. Quelques heures après la transformation bactérienne, l'efficacité de transposition a été quantifiée par séquençage profond du site d'insertion du plasmide cible. (B) Représentation schématique du fonctionnement du système ShCAST, composée de la **Cas12k** (PAM 5'-GTN-3') associée à un ARN guide, des protéines de transpositions TniQ, TnsC et TnsB et du transposon ADN flanqué par les extrémités LE et RE. L'insertion du transposon se produit entre 60 et 66 paires de bases en aval de la séquence PAM. Schémas traduits et adaptés de Jonathan Strecker et al.³⁰⁴.

En somme, les recombinaisons et les transposases paraissent moins dépendantes des mécanismes de réparations de l'ADN et du cycle cellulaire que les PE, expliquant les rares mutations indésirables au site cible. De plus, l'expérimentateur n'a pas besoin de tester un large éventail de *pegRNA*, et la modification (insertion/inversion/délétion...) peut atteindre plusieurs kb.

Des efforts considérables demeurent nécessaires pour mieux caractériser les recombinaisons et les transposases. Il est temps d'améliorer leur efficacité *in vitro* et *in vivo* et de mettre au point des techniques fiables permettant de quantifier leur rendement et leur activité hors cible pangénomique. Parmi les pistes prometteuses, l'évolution dirigée et rationnelle des protéines couplée aux nouvelles méthodes d'apprentissage profond devraient repousser les limites associées aux recombinaisons et transposases. Cette nouvelle génération d'outil demeure très attrayante. A l'avenir, elle permettrait de **faciliter l'ingénierie cellulaire** (ex : CAR-T cells), **programmer l'insertion d'exons thérapeutiques dans un intron précédent un exon défectueux**, ou programmer **l'insertion de transgènes thérapeutiques dans des loci « sans danger »**, c'est-à-dire à distance des régions régulatrices, des proto-oncogènes et des gènes suppresseurs de tumeurs.

9. Résumé : les outils d'édition du génome

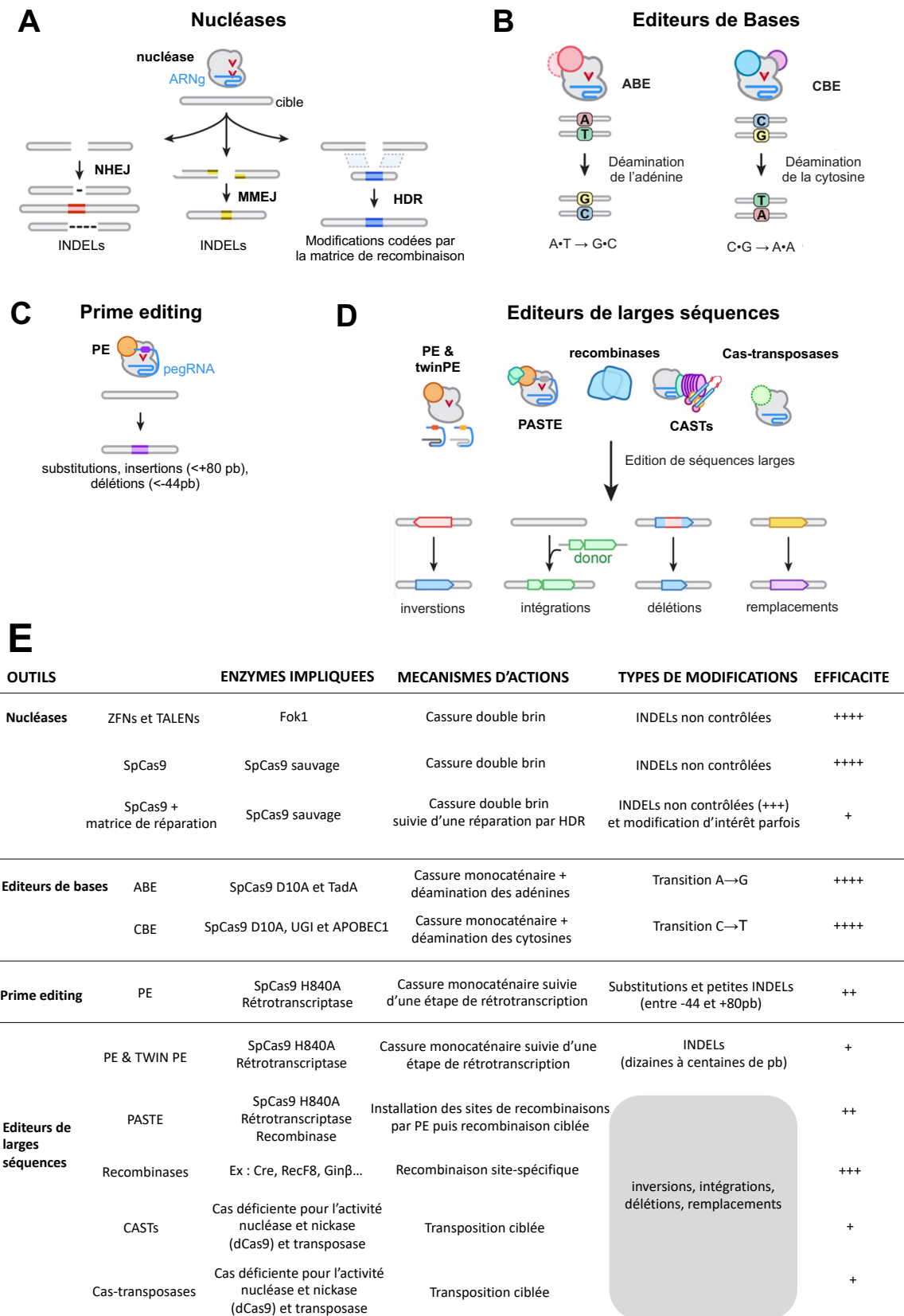


Figure 12 : Vue d'ensemble des outils d'édition du génome.

Les schémas provenant des panels A-D ont été traduits et adaptés d'après Conor J. Tou et al.³⁰².

10. Autres outils CRISPR/Cas

En plus de l'édition génétique, les outils CRISPR ont été exploités pour moduler l'expression des gènes, altérer les marques épigénétiques (et à un plus haut niveau remodeler la structure de la chromatine), imager la dynamique des chromosomes/de la chromatine dans les cellules vivantes et même cibler la molécule d'ARN. Dans cette partie, je ne vais présenter que les outils permettant la modulation de l'expression des gènes. Pour une description plus détaillée, les revues suivantes sont disponibles^{305,306}.

Régulation de la transcription :

La transcription est initiée par le regroupement et la fixation de facteurs de transcription aux niveaux des régions promotrices des gènes, favorisant le recrutement de l'ARN polymérase. Il s'agit de la première étape permettant la régulation de l'expression d'un gène. Afin de guider un facteur de transcription – activateur ou répresseur – vers une région promotrice particulière, les scientifiques ont conçu respectivement les systèmes CRISPR activateurs et CRISPR répresseurs³⁰⁷⁻³¹⁰. Brièvement, ces systèmes fonctionnent avec une Cas9 catalytiquement inactive (dCas9) associée à divers domaines effecteurs de la transcription (ex : VP64, p65, KRAB). Cette association peut être réalisée par fusion directe avec la dCas9 (mais taille limitée)^{307,308}, ou bien indirecte. Il est possible d'associer à la dCas9 un adaptateur composé d'un réseau de peptides favorisant au niveau endogène le recrutement de plusieurs copies d'un facteur de transcription d'intérêt (ex : SunTag), ou bien d'ajouter dans le crRNA des aptamères MS2, pouvant être reconnus par des protéines MS2 elles-mêmes fusionnées à d'autres effecteurs de la transcription³⁰⁹. Ce système est aujourd'hui très bien maîtrisé par la communauté scientifique pour réguler le niveau d'expression de multiples gènes. Contrairement aux méthodes de surexpression ou d'interférence classiques (plasmides, ARNi), la régulation de l'expression du gène est physiologique. De nombreuses maladies peuvent s'accompagner d'une expression anormale des gènes (ex : facteurs inflammatoires dans les maladies chroniques), et réguler leur expression pourrait constituer stratégie thérapeutique d'avenir³¹¹. En comparaison aux outils d'édition des gènes, les systèmes CRISPR activateurs et répresseurs sont réversibles et donc les potentiels effets « hors de la cible » peuvent être mieux maîtrisés.

Le développement de ces différents outils a considérablement élargi les possibilités offertes par la boîte à outils CRISPR, ouvrant la voie à de nouvelles perspectives pour les scientifiques, tant sur le plan fondamental que sur le plan appliqué. Dans la prochaine partie, nous verrons quelles sont les principales applications biomédicales développées grâce aux outils d'édition du génome CRISPR/Cas9 (diagnostique et thérapie), en mettant en évidence leurs avantages et limites.

II. Applications biomédicales des outils d'édition du génome CRISPR/Cas9

1. Edition du génome pour la thérapie génique

1.1 Contexte et définition

Actuellement, plus de 5000 maladies monogéniques ont été décrites et affectent au moins 250 millions de personnes dans le monde³¹². Malheureusement, seulement environ 5% de ces maladies disposent aujourd'hui de traitements efficaces. En agissant sur la cause génétique de ces pathologies, c'est-à-dire la source du problème, la thérapie génique permet dorénavant d'offrir de nouvelles perspectives thérapeutiques pour ces patients.

Il existe actuellement deux types de thérapie génique : la **thérapie génique additive** et l'**édition génique**. La **thérapie génique additive** repose sur l'introduction d'une séquence d'ADN thérapeutique dans le noyau ou le génome d'une cellule dont le produit d'expression permet de pallier au gène défectueux responsable de la pathologie ou de faciliter la destruction de cellules pathologiques (CAR-T cells). L'**édition génétique** consiste à corriger directement, dans l'ADN, les mutations à l'origine de la pathologie ou bien à introduire dans l'ADN de nouveaux variants génétiques protecteurs. Même si l'objectif final de ces deux types de thérapie est le même, chacune d'entre elles est associée à des limites intrinsèques qui seront résumées dans les prochaines parties.

1.2 Limites de la thérapie génique additive

1.2.1 Vecteurs

L'étape clé de la thérapie génique additive consiste à délivrer le transgène thérapeutique dans les cellules/tissus cibles de l'hôte. Cette étape est réalisée par un vecteur, qui peut être d'origine virale ou non.

Brièvement, les vecteurs viraux comportent une enveloppe, qui contient des protéines de surfaces déterminant le tropisme cellulaire/tissulaire ; le transgène thérapeutique

d'intérêt ; et une cassette régulatrice contrôlant l'expression du transgène thérapeutique (*enhancer*, promoteur) ainsi que les éléments nécessaires à l'encapsidation et l'éventuelle intégration du transgène thérapeutique dans le génome de l'hôte. Actuellement, quatre grands types de vecteurs viraux ont été développés pour la thérapie génique : les lentivirus, les γ -rétrovirus, les adénoviruses et les virus adeno-associiés (AAV) (**tableau 4**) :



	ADENOVIRUS	AAV	γ -RETROVIRUS	LENTIVIRUS
TAILLE	≈ 90-100nm	≈ 25nm	≈ 80-100nm	≈ 80-100nm
GENOME	ADN double brin	ADN simple brin	ARN simple brin	ARN simple brin
TAILLE DU CARGO	≈ 8kb - 36kb	≈ 4.7kb	10kb	8kb
TRANSDUCTION	Cellules en division et quiescentes	Cellules en division et quiescentes	Cellules en division	Cellules en division et quiescentes
EFFICACITE DE TRANSDUCTION	Elevée	Modérée	Modérée	Modérée
INTEGRATION	Vecteur non intégratif	Vecteur non intégratif	Vecteur intégratif	Vecteur intégratif
EXPRESSION	Transitoire	Transitoire ou stable	Stable	Stable
NIVEAU DE SECURITE	L2	L1	L2	L2
IMMUNOGENICITE	Elevée	Faible	Modérée-élevée	Modérée-élevée
STRATEGIE POUR LA THERAPIE GENIQUE	<i>In vivo</i>	<i>In vivo</i>	<i>Ex vivo</i>	<i>Ex vivo</i>

Tableau 4 : Principales caractéristiques des vecteurs viraux utilisés pour la thérapie génique.

Tableau traduit d'après <https://biovia.com/>.

Selon le type de protocole utilisé, ces vecteurs peuvent être délivrés aux cellules soit *in vivo*, directement dans l'organisme, soit *ex vivo*. Dans ce cas, les cellules modifiées seront ensuite réinjectées chez le patient (greffe autologue). L'approche *ex vivo* est couramment utilisée pour le traitement des hémoglobinopathies, en raison de l'accès facilité aux cellules souches hématopoïétiques et à la maîtrise des protocoles de greffes de moëlle hématopoïétique.

Les vecteurs AAV sont les vecteurs viraux de premiers choix pour la thérapie génique *in vivo*, en raison de leur faible immunogénicité à l'égard des patients. Généralement, ils persistent dans le noyau des cellules sous forme épisomale, permettant une expression prolongée du transgène dans les cellules post-mitotiques. Toutefois, de plus en plus d'études rapportent des événements d'intégration génomique^{164,313}. Il existe 12 sérotypes naturels AAV (AAV1 à AAV12) et des centaines de variants génétiques présentant chacun un tropisme particulier, et permettant de cibler la plupart des tissus humains^{314,315}.

Les vecteurs γ -rétroviraux dérivés du virus de la leucémie murine et les vecteurs lentiviraux dérivés du virus de l'immunodéficience humaine ont largement été déployés pour le transfert de gènes *ex vivo*. Contrairement aux AAV, ces vecteurs sont intégratifs et permettent une expression stable du transgène dans les cellules mitotiques (cellules souches hématopoïétiques et progéniteurs par exemple).

Grâce aux progrès en nanotechnologies, des vecteurs non viraux formulés à partir de particules lipidiques chargées, liposomes, polymères, nanoparticules inorganiques conjugués à des peptides/lipides de fusion membranaire ont pu être développés³¹⁶⁻³¹⁸. Contrairement aux vecteurs viraux, les vecteurs non viraux sont peu immunogènes (dosages répétés possibles), peu chers, faciles à produire, et peuvent transporter des cargos thérapeutiques de taille importante. Néanmoins, une fois injectés dans la circulation, ces particules présentent des difficultés à franchir les multiples barrières biologiques (barrière sanguine, barrière extracellulaire, barrière cellulaire, barrière nucléaire) et ont naturellement tendance à s'accumuler au niveau du foie. Afin de les rendre encore plus attractifs, les scientifiques proposent de nouvelles formulations pour les diriger spécifiquement vers un type cellulaire ou tissu particulier, limiter leur accumulation hépatique, et améliorer leur efficacité d'internalisation^{319,320}.

1.2.2 Essais cliniques

Entre l'année 2010 et 2020, pas moins de 1908 essais cliniques de thérapie génique ont été recensés³²¹ et aujourd'hui 12 candidats ont été acceptés par l'agence *Food and Drug Administration* (FDA) aux États-Unis (**Tableau 5**).

	PRODUIT	COMPAGNIE	APPROBATION	INDICATION	TYPE DE THERAPIE	VOIE D'ADMINISTRATION	VECTEUR
Traitements anti-cancéreux	ABECMA	Celgene Corporation	FDA, mars 2021	myélome multiple récidivant ou réfractaire après au moins 4 traitements antérieurs	<i>Ex vivo</i> Transduction de cellules T autologues (CAR T cells)	IV	Vecteur lentiviral codant pour un récepteur chimérique à l'antigène ciblant BCMA
	ADSTILADRIN	Ferring Pharmaceuticals	FDA, déc. 2022	cancer de la vessie non invasif musculaire à haut risque à bacille de Calmette-Guérin avec carcinome <i>in situ</i>	<i>In vivo</i>	instillation intra-vésicale (vessie)	Vecteur adénoviral de sérotype 5 codant pour interféron alfa-2b (IFNα2b)
	BREYANZI	Juno Therapeutics	FDA, juin 2022	Lymphome à cellules B récidivant ou réfractaire à au moins 2 thérapies antérieures	<i>Ex vivo</i> Transduction de cellules T autologues (CAR T cells)	IV	Vecteur lentiviral codant pour un récepteur tronqué au facteur de croissance EGFR et un récepteur chimérique à l'antigène ciblant CD19
	CARVYKTI	Janssen Biotech	FDA, fév. 2022	Myélome multiple réfractaire ou récidivant après au moins 4 lignes antérieures de thérapies	<i>Ex vivo</i> Transduction de cellules T autologues (CAR T cells)	IV	Vecteur lentiviral codant pour un récepteur chimérique à l'antigène ciblant BCMA
	KYMRIAH	Novartis Pharmaceuticals Corporation	FDA, aout 2017	Lymphome folliculaire réfractaire ou récidivant après au moins deux thérapies antérieures	<i>Ex vivo</i> Transduction de cellules T autologues (CAR T cells)	IV	Vecteur lentiviral codant pour un récepteur chimérique à l'antigène ciblant CD19
	TECARTUS	Kite Pharma	FDA, Juil. 2020	Lymphome à cellule du manteau ou précurseur de la leucémie aigue lymphoblastique réfractaire ou récidivant	<i>Ex vivo</i> Transduction de cellules T autologues (CAR T cells)	IV	Vecteur gamma-rétroviral codant pour un récepteur chimérique à l'antigène ciblant CD19
Traitements maladies monogéniques	YESCARTA	Kite Pharma	FDA, oct. 2017	Lymphome à grande cellules B réfractaire ou récidivant à la chimiothérapie	<i>Ex vivo</i> Transduction de cellules T autologues (CAR T cells)	IV	Vecteur rétroviral codant pour un récepteur chimérique à l'antigène ciblant CD19
	HEMGENIX	CSL Behring	FDA, nov. 2022	Hémophilie B avec déficience congénitale en facteur IX de la coagulation	<i>In vivo</i>	IV	Vecteur viral adéno-associé de sérotype 5 (AAV5) codant pour un variant gain de fonction humain naturel (R338L) du facteur IX de la coagulation
	LUXTURNA	Spark Therapeutics	FDA, déc. 2017	Dystrophie rétinienne avec mutation bi-allélique du gène RPE65	<i>In vivo</i>	Intra-oculaire	Vecteur viral adéno-associé de sérotype 2 (AAV2) codant pour la protéine humaine RPE65
	SKYSONA	Bluebird bio	FDA, sept. 2022	Adrénoleucodystrophie liée au gène ABCD1	<i>Ex vivo</i> Transduction de cellules CD34+ autologues(HSC)	IV	Vecteur lentiviral codant pour la protéine ALDP
	ZYNTEGLO	Bluebird bio	FDA, aout 2022	bêta- thalassémie nécessitant des transfusions régulières	<i>Ex vivo</i> Transduction de cellules CD34+ autologues(HSC)	IV	Vecteur lentiviral BB305 codant pour la globine-β ^{A-T87Q}
	ZOLGENSMA	Novartis Gene Therapies	FDA, mai 2019	Amyotrophie musculaire spinale infantile liée au gène SMN1	<i>In vivo</i>	IV	Vecteur viral adéno-associé de sérotype 9 (AAV9) codant pour la protéine SMN

Tableau 5 : Thérapies géniques acceptées aux Etats-Unis.

FDA : Food and Drug Administration, CAR-T cells : Chimeric Antigen Receptor lymphocyte T cells. IV : injection intraveineuse.

1.2.3 Limites associés aux produits de thérapie géniques

Mutagenèse insertionnelle

L'intégration aléatoire du transgène thérapeutique dans l'ADN des cellules est l'un des principaux risques de la thérapie génique additive. Les vecteurs rétroviraux peuvent notamment insérer leur matériel dans les régions ouvertes de chromatine, au risque de déréguler l'expression des gènes se trouvant à proximité. Ce type d'effet secondaire a déjà été observé par le passé, au cours de deux essais cliniques conduits en France et au Royaume-Uni, respectivement en 1999 et 2002. Au total 20 patients atteints du déficit immunitaire combiné sévère lié à l'X ont été traités par thérapie génique *ex vivo* (cellules CD34+) avec des vecteurs rétroviraux codant pour le récepteur à l'interleukine 2 (IL2RG)³²². Même si les résultats de la thérapie se sont avérés spectaculaires (18/20 patients presque ou totalement guéris 20 ans plus tard), 6 d'entre eux ont développé des leucémies entre 2.5 ans et 15 ans après la greffe, et l'un d'entre eux est décédé des suites de ce cancer³²³⁻³²⁵. Cela s'explique par l'intégration de certaines copies du transgène thérapeutique à proximité du promoteur du proto-oncogène LMO2. De plus, les longues répétitions terminales (séquences LTR) portées par les vecteurs rétroviraux peuvent agir comme des *enhancers* transcriptionnels. Depuis, des vecteurs rétroviraux plus sophistiqués et moins dangereux ont été mis au point et utilisés dans de nouveaux essais cliniques³²⁶. Notamment, la présence d'une délétion dans l'une des séquences LTR des vecteurs rétroviraux a permis d'inactiver leur activité *enhancer*³²⁷.

Immunogénicité et toxicité

Malgré le succès notable des thérapies géniques basées sur les vecteurs AAV, un des obstacles principal porte sur la réponse immunitaire des patients face à ce type de vecteur. Bien que les vecteurs AAV soient décrits comme peu immunogènes, l'injection d'un titre viral élevé peut malheureusement mener à l'activation des cellules B et T, impactant l'efficacité de la thérapie (rejet des cellules transduites)^{328,329}. Les traitements immunosuppresseurs ont montré une diminution de la réponse immunitaire, mais pas chez tous les patients³³⁰. Ces observations ont incité les cliniciens à mettre en place la titration systématique des anticorps dirigés contre les capsides virales avant l'administration du vecteur de thérapie génique. Plusieurs groupes ont prouvé que des patients possédaient une immunité préexistante contre certains sérotypes AAV, ce qui a conduit à l'exclusion de jusqu'à 50% des patients des protocoles cliniques^{331,332}. Pour contrer ce problème, des groupes de chercheurs se

sont concentrés à isoler de nouveaux variants de protéines de la capsidie résistants aux anticorps neutralisants^{333,334}. L'utilisation de ces nouveaux variants AAV reste encore limitée dans le domaine clinique.

De même, une réponse immunitaire contre le transgène thérapeutique peut susciter des risques de réactions auto-immunitaires, affectant négativement la réponse thérapeutique et menant à l'élimination progressive des cellules transduites. L'expression prolongée du transgène thérapeutique pourrait favoriser l'apparition de cet effet secondaire, d'autant plus si celui-ci code pour une protéine sécrétée. D'ailleurs, certaines études précliniques et cliniques ont reporté une activité cytotoxique médiée par les lymphocytes T à l'encontre des peptides transgéniques et même la présence d'anticorps dirigés contre les protéines thérapeutiques³³⁵⁻³³⁷. Les lymphocytes T régulateurs jouent un rôle primordial dans le maintien de la tolérance immunitaire. Afin de maintenir la tolérance à l'égard du transgène thérapeutique, certains groupes suggèrent d'induire *in vivo* des lymphocytes T régulateurs spécifique aux antigènes d'intérêt grâce à un traitement combiné, ou bien d'injecter au patient des lymphocytes T régulateurs autologues préalablement amplifiés *ex vivo*³³⁸.

Dernièrement, un accident tragique a mené au décès de trois enfants traités pour la myopathie myotubulaire liée à l'X par un vecteur AAV8 (dose élevée : 3×10^{14} copies du génome/kg). Les données biologiques laissent supposer que deux des patients ont succombé en raison d'une hépatotoxicité, tandis que le dernier semble être décédé suite à des saignements gastrointestinaux^{339,340} (NCT03199469). Même si aucun lien n'a pu être établi entre les vecteurs AAV8 et le décès des patients, des études précédentes suggèrent que les vecteurs AAV sont capables d'induire des désordres hépatiques³⁴¹.

Taille du transgène

La taille du génome des vecteurs de type AAV est limitée, ceux-ci ne peuvent encapsider que des ADN inférieurs à 4,7kb en moyenne, séquence répétées inversées, séquence de polyadénylation et promoteur compris. Cela revient à une séquence d'ADN codante (CDS) du transgène thérapeutique ne pouvant pas excéder 3.5kb. Même si plusieurs groupes ont montré qu'il était possible de combiner plusieurs AAV pour transporter différentes parties d'un grand gène qui sont réassemblés par la

suite dans les cellules hôtes (au niveau ADN, ARN ou protéique), l'efficacité de cette approche reste très réduite (système AAV-double)³⁴²⁻³⁴⁵.

Tropisme limité pour certains tissus/cellules

Certains tissus ou cellules demeurent résistants ou peu permissifs à la transduction par les vecteurs AAV *in vivo*. Afin de rediriger le tropisme des vecteurs AAV vers des cibles d'intérêt, les chercheurs ont entamé des démarches d'ingénierie moléculaire des protéines de la capside³⁴⁶. D'autres ont développé des bibliothèques de peptides d'affichage AAV aléatoires, permettant de sélectionner de nouveaux variants capables de transduire les cellules de la barrière hémato-encéphalique³⁴⁷, les cellules endothéliales coronaires³⁴⁸, ou encore les cellules pulmonaires³⁴⁹. A l'avenir, ces nouveaux variants pourraient permettre d'étendre le spectre thérapeutique des vecteurs AAV.

1.3 Thérapie génique par édition de l'ADN

1.3.1 Rationnel

Même si la thérapie génique additive permet, *en théorie*, de traiter les pathologies à transmission autosomique récessive, **cette stratégie n'est pas applicable aux maladies à transmission autosomique dominante**³⁵⁰. Plus largement, cela concerne toutes les maladies génétiques où l'apport d'une copie saine du gène atteint ne permet pas d'améliorer la condition du patient (ex : maladies à expansion de nucléotides, maladies avec mutation gain de fonction, gène toxique). **Face à cette limite, les outils d'édition du génome apportent une nouvelle opportunité, puisqu'ils permettent 1) de corriger la mutation causale (ou d'apporter un nouveau variant/locus protecteur) 2) d'atteindre tous les types de gènes sans limite de taille 3) de préserver l'expression physiologique du gène ciblé (promoteur endogène)**

Depuis quelques années, les outils d'édition du génome ont conquis les stratégies de thérapies géniques, comme illustré par les nombreuses preuves de concept et essais cliniques en cours.

1.3.2 Essais précliniques et cliniques

CRISPR/Cas9 classique pour l'inactivation de gènes toxiques

Le tout premier essai clinique humain utilisant la technologie CRISPR/Cas9 *in vivo* a été réalisé sur des patients atteints d'amyloïdose à transthyrétine (ATTR) (NCT04601051). Cette pathologie d'origine hépatique est causée par des défauts génétiques dans le gène *transthyrétine* (*TTR*) et concerne pas moins de 50 000 personnes dans le monde (**figure 13**). La protéine TTR, principalement sécrétée par le foie, est présente dans le sang et le liquide céphalorachidien. Elle s'organise sous forme de tétramère pour transporter la vitamine A et la thyroxine. Lorsque cette protéine est mutée, celle-ci devient anormale, favorisant la synthèse d'oligomères TTR mal repliés. Ces oligomères peuvent s'assembler pour générer des polymères extracellulaires insolubles et toxiques appelés dépôts amyloïdes (**figure 13**).

Les premiers résultats de phase 1 obtenus sur 6 patients ont été publiés le 27 juin 2021 dans la revue *The New England Journal of Medicine*⁷³. La stratégie thérapeutique consiste à inactiver le gène de la TTR en injectant (IV) aux patients des nanoparticules lipidiques refermant l'ARNm codant pour la Cas9 et l'ARN guide dirigé contre le gène TTR (candidat NTLA-2001). **Seulement 28 jours après l'injection, Julian D. Gillmore et al. ont montré une réduction de la concentration sérique en TTR de 52% à 87% chez les patients, selon la dose injectée.** Ces premières données semblent très encourageantes, et aujourd'hui 72 patients ont été recrutés pour poursuivre l'essai clinique. Bien entendu, les patients devront être suivis pendant plusieurs années afin de suivre l'évolution du traitement sur le moyen et long terme (effets inconnus). Les CDB médiées par CRISPR/Cas9 étant tout de même associés à de nombreux effets génotoxiques, les scientifiques se sont rapidement tournés vers les outils ABE et CBE pour l'inactivation de gènes toxiques.

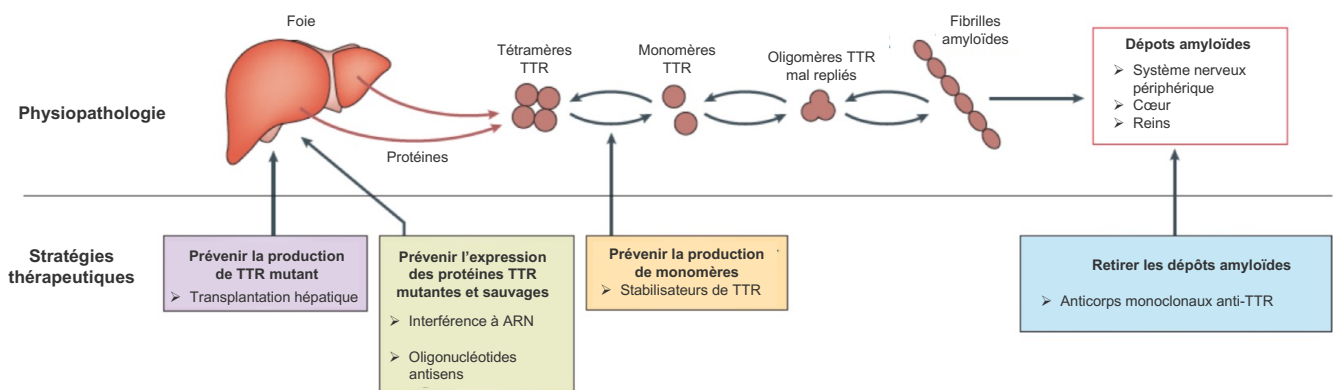


Figure 13 : Physiopathologie de l'amyloïdose héréditaire à transthyréine (ATTR).

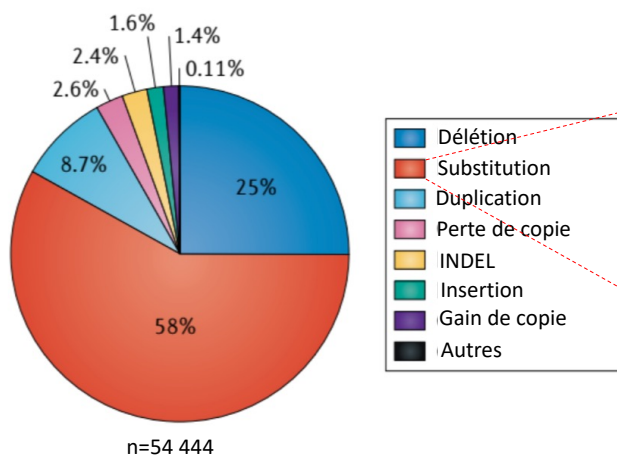
Les dépôts amyloïdes formés par la TTR mutée peuvent se déposer dans divers organes (cœur, reins, tractus gastro-intestinal, système nerveux périphérique) et sont responsables de multiples symptômes (faiblesse musculaire, perte d'autonomie, troubles cardiaques) menant au décès du patient en une dizaine d'années. Les principales stratégies thérapeutiques visent à lutter contre la production des dépôts amyloïdes. Schéma traduits et adaptés de David Adams et al.³⁵¹.

Éditeurs de bases BE : les chiffres

Les éditeurs de bases ABE et CBE (les plus efficaces) permettent de générer des transitions dans l'ADN (A•T→G•C pour l'ABE ou bien C•G→T•A pour le CBE). Ces transitions représentent environ 61% des modifications génétiques nécessaires pour reverser les substitutions pathogènes associés aux maladies génétiques humaines selon la base de données ClinVar en 2018 (**figure 14**). Initialement, les outils BE pouvaient prendre en charge seulement 26% de ces transitions pathogènes de l'ADN (limite liée à la disponibilité d'une séquence PAM à proximité du nucléotide à corriger). Grâce au développement de nouveaux variant Cas9 reconnaissant un spectre beaucoup plus large de séquences PAM (ex : SpRY), le BE peut désormais atteindre environ 95% des transitions pathogènes, soit environ 33,6% de l'ensemble des variants génétiques délétères décrits dans Clinvar²²⁶. **En d'autres termes, en théorie, une personne sur trois atteinte d'une maladie génétique pourrait potentiellement être traitée par BE, offrant des perspectives thérapeutiques inespérées pour les patients.**

La liste des principaux essais en cours sont présentés dans le **tableau 6**. Quelques exemples seront décrits dans les parties suivantes.

A Variants humains associés aux maladies génétiques



B Mutations nécessaires pour reverser une substitution délétère

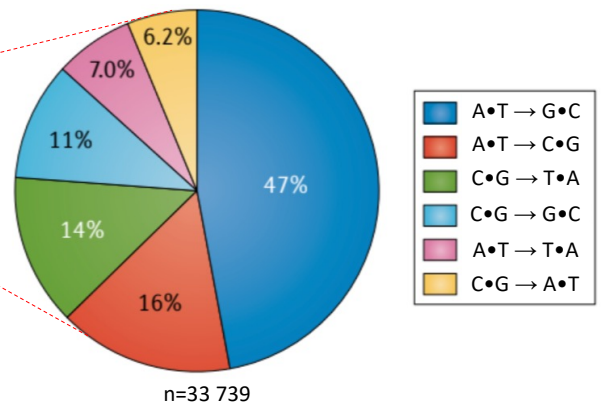


Figure 14 : Distribution des variants génétiques délétères humains.

(A) Classification des variants délétères humains recensés dans ClinVar le 29 Mai 2018. (B) Classification des différents types de mutations nécessaires pour reverser une substitution délétère (58% des variants délétères sont des substitutions). Les transitions sont chiffrées en blanc (14%+47% = 61%) et les transversions en noir. Graphiques traduits et adaptés de Holly A. Rees et al.²²⁴.

PRODUIT	COMPAGNIE	STATUT	CIBLE	TYPE
BE CAR-7	Univ College Londres (laboratoire du Pr Waseem Qasim)	Phase I, Mai 2022 NCT05397184	Gènes CD7, CD52 et TCR Leucémie à cellules T	<i>Ex vivo</i> Transduction de cellules T autologues (CAR T cells)
VERVE-101	Verve Therapeutics	Phase I, Mai 2022 NCT05398029	Gène PCSK9 Hypercholestérolémie familiale et maladies cardiovasculaires	<i>In vivo</i>
VERVE-201	Verve Therapeutics	Investigations Pré-cliniques <i>IND Enabling</i>	Gène ANGPTL3 Hypercholestérolémie familiale et maladies cardiovasculaires	<i>In vivo</i>
BEAM-101	Beam Therapeutics	Phases I/II, Juillet 2022 NCT05456880	Activation des gènes HBG1 et HBG2 Drépanocytose et Bêta-thalassémies	<i>Ex vivo</i> Transduction de cellules CD34+ autologues (HSC)
BEAM-102	Beam Therapeutics	Investigations Pré-cliniques <i>IND Enabling</i>	Gène HBB Drépanocytose	<i>Ex vivo</i> Transduction de cellules CD34+ autologues (HSC)
BEAM-201	Beam Therapeutics	Investigations Pré-cliniques <i>IND Enabling</i>	Gène CD7 (et d'autres gènes non communiqués) Leucémies aiguës lymphoblastiques à cellules T et leucémies aiguës myéloïdes CD7+	<i>Ex vivo</i> Transduction de cellules T autologues (CAR T cells)
BEAM-301	Beam Therapeutics	Investigations Pré-cliniques <i>IND Enabling</i>	Gène G6PC Glycogénose de type I	<i>In vivo</i>

Tableau 6 : Candidats thérapeutiques reposant sur la technologie BE.

Éditeurs de bases BE pour le traitement des hémoglobinopathies

La prise en charge de la drépanocytose par BE est certainement l'une des plus prometteuse. Avec environ 310 000 nouveaux cas chaque année, la drépanocytose est la maladie génétique la plus fréquente dans le monde³⁵². Il s'agit d'une maladie à transmission autosomique récessive caractérisée par la présence d'une hémoglobine mutée drépanocytaire (hémoglobine S, $\alpha_2\beta^S_2$) conférant une forme en faucille aux globules rouges adultes. Cette hémoglobine S résulte d'une mutation faux-sens unique et ponctuelle présente dans le gène codant pour la globine- β (GAG→GTG ; glutamate→valine). Bien que l'allogreffe de moëlle hématopoïétique puisse traiter la maladie en apportant au patient des cellules souches hématopoïétiques saines, encore faut-il trouver un donneur avec un profil immunitaire compatible avec celui du patient. La thérapie génique additive a démontré son efficacité chez certains patients, mais l'expression relativement faible du transgène thérapeutique ne permet pas de compenser totalement l'effet délétère de l'allèle drépanocytaire³⁵³.

Bien que l'ABE ne puisse pas directement corriger l'allèle drépanocytaire (GAG→GTG ; glutamate→valine), cet outil peut convertir le variant délétère en un autre variant non pathogène rare, appelé variant Makassar (GTG→GCG ; valine→alanine). Seize semaines après la greffe de cellules CD34+ éditées dans des modèles murins, l'équipe de David R. Liu a réussi à produire 68% d'hémoglobine fonctionnelle et de réduire de cinq fois le nombre de cellules drépanocytaires falciformes en utilisant cette approche³⁵⁴. Cette stratégie a donné naissance au candidat thérapeutique Beam-102, qui attend de recevoir l'autorisation par la FDA pour démarrer les essais cliniques (<https://beamt.com/>).

La seconde approche consiste à **réactiver par BE l'expression de la globine- γ** (globine foëtale codée par les gènes *HBG1* et *HBG2* formant l'hémoglobine F, **figure 15**), **soit en mutant son répresseur transcriptionnel BCL11A³⁵⁵ soit en mutant les séquences promotrices sur lesquelles ce répresseur se fixe^{186,356}.** Le candidat thérapeutique reposant sur cette stratégie, Beam-101, est aujourd'hui en phase 1 et 2 d'essai clinique. Les études précliniques suggèrent que Beam-101 permet d'atteindre jusqu'à 90% d'édition dans les cellules CD34+, de restaurer l'expression de l'hémoglobine F et d'abaisser celle de l'hémoglobine S : près de 60% de l'hémoglobine totale est représentée par l'hémoglobine F et moins de 40% par l'hémoglobine S (<https://beamt.com/>).

Bien que ces deux stratégies ne permettent pas de corriger directement le variant délétère, l'installation du variant génétique protecteur Makassar ou bien la restauration de l'expression du gène de la globine- γ permettent tous les deux de pallier à l'allèle drépanocytaire. Plusieurs revues décrivant les autres essais cliniques en cours sont disponibles^{357–359}.

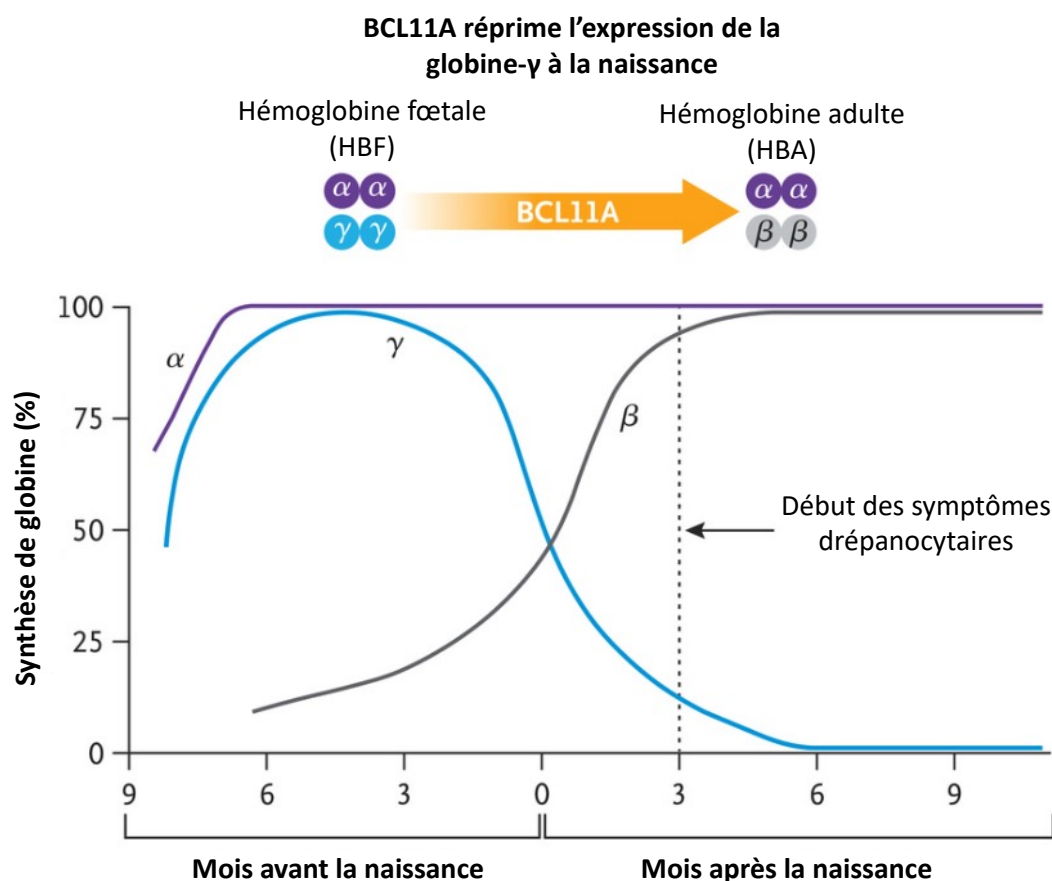


Figure 15 : Le répresseur transcriptionnel BCL11A contrôle le passage de l'hémoglobine fœtale HBF à l'hémoglobine adulte HBA.

Quelques semaines avant la naissance, le facteur BCL11A commence à réprimer l'expression du gène de la globine- γ . L'hémoglobine, constituée de 4 sous unités de globine, passe alors d'une forme fœtale HBF ($\alpha_2\gamma_2$) à une forme adulte HBA ($\alpha_2\beta_2$). La forme HBA devient prédominante juste après la naissance, puis remplace totalement la forme HBF au cours des 6 premiers mois. Les patients atteints de drépanocytose sont incapables de produire une globine- β fonctionnelle (hémoglobine S, $\alpha_2\beta^S_2$), leurs symptômes commencent à apparaître environ 3 mois après la naissance. Afin de restaurer une hémoglobine fonctionnelle chez ces patients, une stratégie consiste à lever le verrou transcriptionnel exercé par BCL11A sur la globine- γ , soit en mutant les régions promotrices sur lesquelles il se fixe, soit en inactivant BCL11A dans les cellules souches hématopoïétiques. Schéma traduit et adapté de Haydar Frangoul et al.³⁶⁰.

Éditeurs de bases BE pour le traitement des maladies infectieuses

Un autre pan de la recherche clinique s'intéresse également aux maladies infectieuses virales. Il est possible d'utiliser la technologie BE pour 1) altérer les protéines de l'hôte utilisées par le virus ou pour 2) inactiver directement le génome viral.

Par exemple, Friederike Knipping et al. ont montré que l'inactivation des corécepteurs CCR5 or CXCR4 humains, qui sont nécessaires à l'entrée du virus de l'immunodéficience humaine (VIH) dans les cellules T CD4+, permettait de protéger les cellules contre la transduction des vecteurs viraux avec une efficacité atteignant jusqu'à 88% *in vitro*³⁶¹. A l'avenir, on pourrait imaginer modifier les cellules souches hématopoïétiques *ex vivo* pour rendre les lymphocytes T du patient résistants à l'infection par le VIH. Dans une autre étude, Hao Zhou et al. ont utilisé le CBE pour générer un codon stop dans le gène S du virus de l'hépatite B. Cette stratégie a permis d'abaisser drastiquement l'expression des protéines d'enveloppes du virus *in vitro*³⁶².

1.3.3 Challenge : méthodes de vectorisation *in vivo* des éditeurs de bases

Plusieurs études ont démontré que la séquence génétique codant pour les éditeurs de bases pouvait facilement être encapsidée dans divers vecteurs viraux (système AAV-double^{189,363–365}, lentivirus³⁶⁶, adénovirus^{206,367}) pour être délivrés efficacement dans divers modèles animaux. **Cependant, l'expression prolongée du BE dans les cellules (ADN épisomal ou intégratif) augmente nécessairement la probabilité de générer 1) une activité hors-cible et 2) une réponse immunitaire dirigée contre la Cas9. Idéalement, des vecteurs délivrant le BE sous forme ARN ou bien sous forme de complexe RNP (expression transitoire) permettraient de réduire ce type de risque.**

Comme les vecteur non viraux ont initialement été développé pour délivrer des siARN, plusieurs groupes ont amélioré leur formulation afin qu'ils transportent l'ARN guide et l'ARNm codant pour la Cas9/le BE^{211,368–372}. Même si ces vecteurs permettent effectivement une expression transitoire des agents d'édition du génome, leur utilisation pré-clinique et clinique est principalement limitée aux maladies hépatiques.

En s'inspirant des travaux de l'équipe d'Emiliano P Ricci³⁷³, l'équipe du Pr. David R. Liu a mis au point en 2022 une plateforme d'ingénierie de particules de types virales

(***Virus-Like Particles*** = **VLP**) pour le transport du BE. Les VLP sont des particules dérivant des rétrovirus dénués d'activité d'infectieuse. A la place de transporter leur matériel génétique, les VLP transportent divers ARN ou protéines, réduisant les risques associés à l'intégration stable du génome viral dans les cellules cibles³⁵⁰. **En fusionnant le BE à la protéine structurale gag via un domaine linker clivable, l'équipe du Pr. David R. Liu a élégamment montré que les VLP étaient capables de transporter des cargos BE sous forme de complexe RNP (figure 16).** Les BE-VLP ont été administrés *in vivo* chez la souris et ont permis d'atteindre environ 20%-30% d'édition dans les cellules RPE murines (injection intra-oculaire) et 63% d'édition dans les cellules hépatiques (injection systémique rétro-orbitaire)³⁷⁴, soit presque aussi bien qu'avec des vecteurs AAV.

De plus, il est possible de moduler à façon le tropisme des VLP en modifiant la composition des protéines de l'enveloppe. Récemment, Jennifer Hamilton et al. ont prouvé qu'en ajoutant des fragments d'anticorps anti-CD19 à la surface de l'enveloppe virale (scFv anti-CD19 fusionné au domaine transmembranaire CD8 α), les VLP ciblaient quasiment-exclusivement les cellules exprimant l'antigène CD19³⁷⁵. Après avoir testé leur vecteur chez la souris, aucune activité d'édition n'a été détectée au niveau du foie, suggérant une réduction de l'hépatotoxicité et surtout de la dose nécessaire pour obtenir l'effet thérapeutique attendu. A l'avenir, cette sélectivité de transduction permettrait peut-être de directement modifier, *in vivo*, les cellules souches hématopoïétiques (CD34+) pour le traitement des hémoglobinopathies ou bien les lymphocytes T pour le traitement du cancer (CD4+/CD8+), évitant toutes les étapes de ponction, tri, modification et amplification *ex vivo* des cellules puis greffe autologue chez le patient.

Le seul inconvénient des VLP actuels est leur immunogénicité modérée/élevée (structure protéique dérivée de celle des rétrovirus). A l'avenir, il serait intéressant de développer des VLP avec des protéines de surface peu immunogènes (ex : variants de capsides AAV résistants aux anticorps neutralisants, mais plus difficile à réaliser à cause de la capside de petite taille et très rigide).

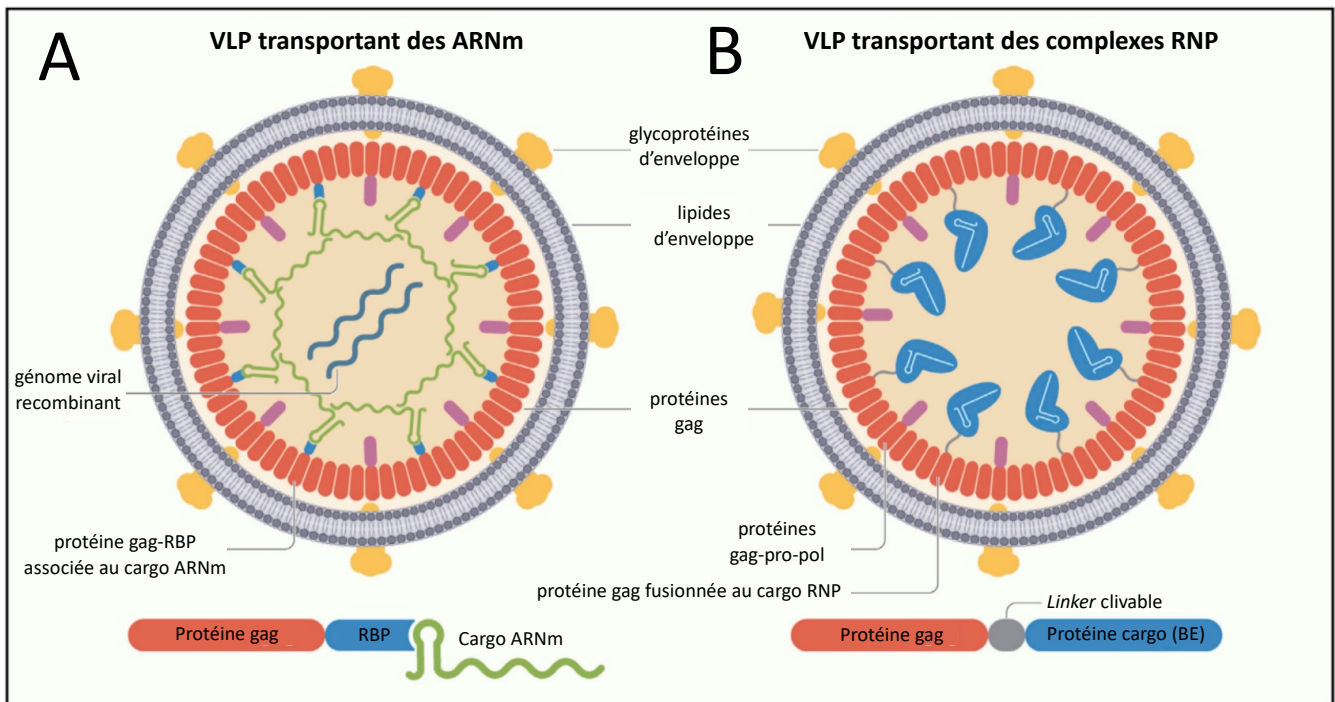


Figure 16 : Représentation schématique des particules de type virales (VLP) transportant des cargos ARN ou des complexes RNP.

Les VLP sont composées d'une pseudo-enveloppe virale externe (lipides et glycoprotéines), d'une matrice et d'une capsid interne (protéines de structures gag et polyprotéines gag-pro-pol), de diverses enzymes nécessaires à la maturation des particules virales et éventuellement d'un génome recombinant incapable de s'intégrer dans l'ADN de l'hôte infecté. **(A)** VLP transportant des cargos ARNm. La fusion de la protéine gag avec une protéine de liaison à l'ARN (ex : protéine MS2) agit comme un appât et permet l'encapsulation des ARNm contenant l'aptamère spécifiquement reconnu par la protéine de liaison à l'ARN (ex : aptamère MS2). **(B)** VLP transportant des cargos RNP. La fusion directe de la protéine gag avec la Cas9 (BE) *via* un linker clivable permet l'encapsulation de la protéine Cas9, elle-même associée à l'ARN guide. Le complexe RNP est relargué suite à l'action d'une protéase virale puis est ensuite libéré dans les cellules transduites. VLP = *Virus Like Particle*, RBP = protéine de liaison à l'ARN, complexe RNP = complexe ribonucléoprotéique, BE = éditeur de base. Schéma traduit et adapté de Aditya Raguram et al.³⁵⁰.

1.3.4 Futures directions

Thérapie génique par PE

Étant dorénavant le plus polyvalent des outils d'édition du génome, **en théorie, le PE pourrait corriger environ 90% des variants pathogènes responsables de maladies génétiques humaines**, c'est-à-dire ceux qui correspondent à des substitutions d'un seul nucléotide ou des petites INDELS (<une douzaine de paires de bases)²⁵¹ (**figure 17**). Même si aucun essai clinique associé au PE n'a encore démarré, de nombreux articles ont d'ores et déjà démontré son efficacité pré-clinique chez la souris pour la correction de divers variants pathogènes responsables de : l'amaurose congénitale de Leber, la tyrosinémie héréditaire, la beta-thalassémie, la phénylcétonurie, le déficit en alpha-1-antitrypsine ou encore la drépanocytose^{252,253,281,282,376}. Un des défis actuels consiste à développer de nouvelles méthodes pour faciliter l'administration du PE *in vivo* (sous forme ADN, ARN ou de complexe RNP). Actuellement, un leader privé détient le monopole de la technologie PE (<https://primemedicine.com/>).

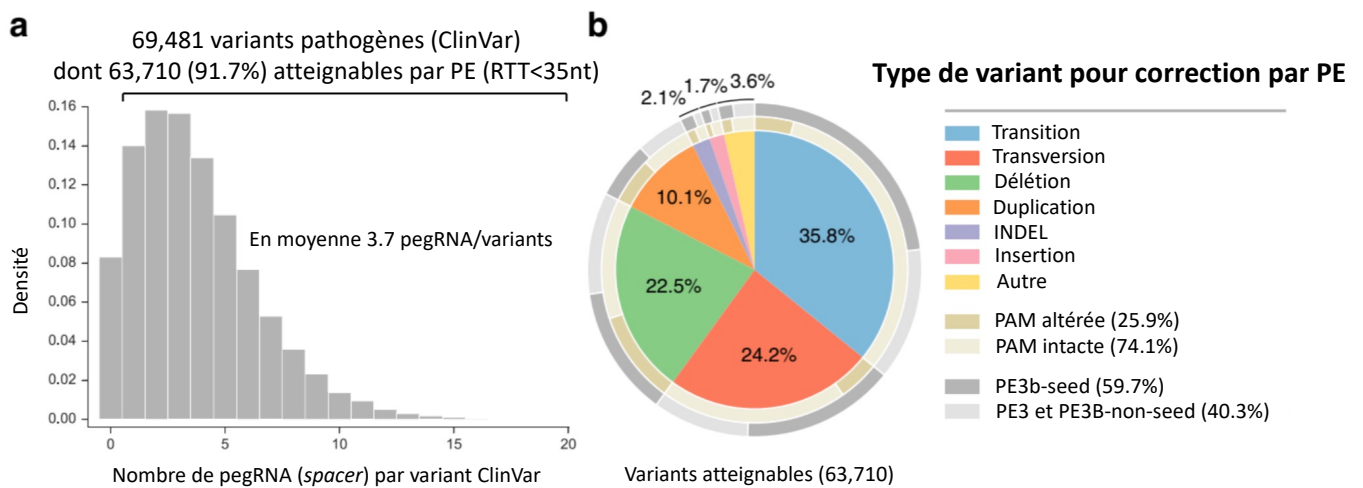


Figure 17 : Nombre et type de variants pathogènes atteignables par PE.

(A) La technologie PE permet d'atteindre, *en théorie*, 91,7% variants génétiques pathogènes humains recensés dans ClinVar. Les auteurs ont utilisé l'outil *PrimeDesign* pour la conception automatique de *pegRNA* et ont dénombré que 63 710 des 69 481 variants ClinVar pouvaient être reversés par au moins un *pegRNA* (91,7%) et qu'en moyenne, 3,7 *pegRNA* pouvaient être conçus par variant (RTT<35nt). (B) Classification par type des 63 710 variants ClinVar atteignables par PE. L'anneau interne beige illustre le pourcentage variant atteignables pour lesquels au moins un *pegRNA* mène à une altération de la séquence PAM (beige foncé). L'anneau externe gris illustre le pourcentage de variants atteignables pour lesquels au moins un *ngRNA* permet d'utiliser la stratégie PE3b-seed (gris foncé). Données traduites et adaptées de Jonathan Y. Hsu et al.²⁶⁰.

2. Edition du génome pour l'aide au diagnostique

2.1 Interprétation biologique des variants génétiques

Depuis l'avènement des méthodes de séquençage à haut-débit, l'enrichissement des connaissances sur le génome humain n'a cessé de croître. La réduction drastique des coûts liés au séquençage du génome humain a progressivement incité les laboratoires de génétiques médicales à installer dans leur service des plateformes de séquençage et de bio-informatique pour faciliter le diagnostic des maladies génétiques. Selon les indications, l'analyse du génome peut se faire à plusieurs échelles : exploration du génome complet, de l'exome, du méthylome, de panels de gènes associés à une pathologie particulière, et même du transcriptome.

Une multitude de nouvelles altérations génétiques associées à des pathologies ont été identifiées. **Néanmoins, la question de l'interprétation du rôle biologique de ces variants s'est rapidement posée. En effet, dans de nombreux cas, les praticiens hospitaliers ne peuvent que constater une corrélation sans avoir de preuve formelle établissant la relation causale entre ces altérations et les pathologies en question.**

En 2013 puis 2015, un groupe de travail composé de représentants Américains du *College of Medical Genetics and Genomics*, de l'*Association for Molecular Pathology* et du *College of American Pathologists* a établi des recommandations pour la classification des ces variants génétiques³⁷⁷ :

- Variants bénins (classe 1)
- Variants probablement bénins (classe 2)
- Variants de signification inconnue (VSI) (classe 3)
- Variants probablement pathogènes (classe 4)
- Variants pathogènes (classe 5)

Cette classification se base sur plusieurs critères : type de variant génétique, localisation, fréquence du variant dans la population générale, connaissances accumulées dans les bases de données, conservation entre les espèces, algorithmes de prédictions *in silico*, études de ségrégation familiales, et données fonctionnelles présentes dans la littérature³⁷⁷. **L'identification d'un variant pathogène ou probablement pathogène permet est indispensable pour établir un diagnostic**

généétique. Celui-ci conditionne la prise en charge adaptée du patient et des membres de sa famille porteurs de l'altération génétique (cas apparentés) : conseils génétiques, surveillance médicale, traitements, médecine personnalisée (ex : thérapie génique, thérapie ciblée pour le cancer), orientation vers des protocoles d'essais cliniques... (figure 18).

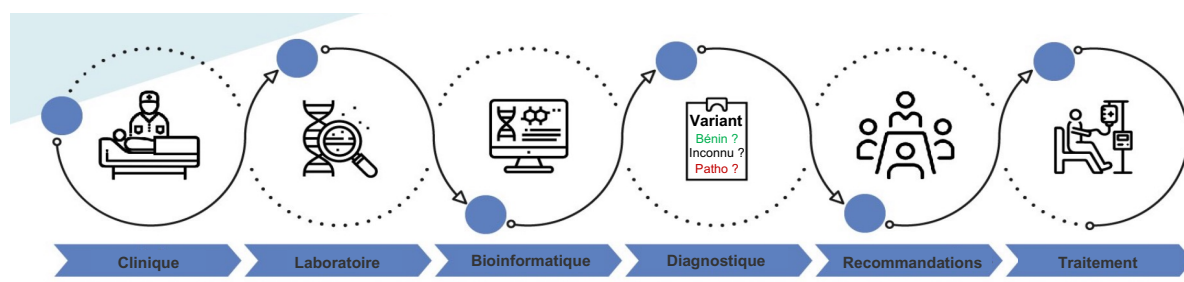


Figure 18 : Les différentes étapes du diagnostic génétique.

Schéma traduit et adapté d'après <https://mdhs.unimelb.edu.au/>.

Les variants de signification inconnue (VSI) sont tellement rares voire uniques dans la population qu'il est très difficile de collecter des informations pour les classer. Ils constituent un verrou au diagnostic génétique. Dans la base de données ClinVar, le 15 Avril 2023, 999 691 VSI ont été référencés, dont 820 125 correspondent à des variants faux-sens. Afin de démontrer leur caractère délétère ou bénin, **une étape de validation fonctionnelle est nécessaire**, mais cette étape demeure très rare en routine car elle requiert beaucoup de temps ainsi qu'une expertise technique particulière.

La validation fonctionnelle consiste à recréer dans un modèle (souris, drosophile, poisson zèbre, cellules...) le variant biologique et à évaluer sa fonction *via* divers tests biologiques. Le modèle cellulaire (culture cellulaire) est généralement très apprécié en raison de son faible coût, de sa facilité de manipulation, de sa rapidité à obtenir des résultats et de sa durabilité (cryoconservation possible).

Actuellement, plusieurs techniques permettent de tester ces VSI :

→ Le **minigène**, qui consiste à évaluer les VSI localisés dans des sites consensus d'épissages introniques ou exoniques. Le variant à tester est inclus dans un plasmide de surexpression contenant un gène rapporteur d'épissage (généralement trois exons et deux introns) qui est transfecté dans des cellules humaines. Une analyse par PCR quantitative ciblée sur le transcrit produit permet d'évaluer l'efficacité d'épissage. Même si cette technique a permis de valider la pathogénicité de plusieurs centaines de VSI³⁷⁸, elle est longue à mettre en place et ne permet pas de reconstituer le contexte génétique associé à chaque VSI, suggérant l'apparition potentielle de résultats faux-positifs ou faux-négatifs. L'implémentation de la technique de séquençage ciblé de l'ARN (RNAseq) dans les laboratoires de génétique permet dorénavant d'évaluer directement l'impact des VSI sur l'épissage.

→ La **surexpression du gène muté via un plasmide**. L'inconvénient est que la surexpression n'est pas physiologique (10 à 50 fois). De plus, le produit du gène surexprimé muté peut entrer en compétition avec la version endogène, rendant l'interprétation parfois complexe. La surexpression du gène muté s'accompagne de divers tests fonctionnels post-transfection, rigoureusement choisis selon la fonction physiologique du gène testé. Par exemple, si le gène testé est un facteur de transcription, il peut être intéressant d'évaluer sa localisation sub-cellulaire et de mesurer le niveau d'expression de ses gènes cibles par PCR quantitative.

Avec l'arrivée des outils d'édition du génome, il est désormais possible de directement réécrire les VSI dans l'ADN de modèles d'intérêt, pour en évaluer leur fonction. Un exemple sera décrit dans la partie suivante.

2.2 Développement tests fonctionnels pour la reclassification des VSI

Environ 5% des cancers diagnostiqués sont liés à une altération génétique qui prédispose au cancer. Le syndrome de Lynch est la forme la plus fréquente de prédisposition au cancer colorectal. Ce syndrome, à transmission autosomique dominante, est causé par des mutations constitutionnelles localisées dans les gènes appartenant au système MMR. Dans environ 75% des cas, ces mutations perte de fonction affectent les gènes *MLH1* et *MSH2*. Environ 30% des variants génétiques identifiées dans ces gènes sont des VSI³⁷⁹.

Le système MMR est indispensable pour résoudre les mésappariements de l'ADN. Ce dernier participe également à la signalisation apoptotique en cas d'exposition à des lésions de type méthylations sur l'ADN³⁸⁰. Chez les patients atteints du syndrome de Lynch, leur le système MMR est déficient. Les cellules deviennent tolérantes à de type de lésion et échappent donc à la signalisation apoptotique. Dans ce contexte, Delphine Bouvet et al (Centre de Recherche Saint-Antoine, Paris) ont mis en place un test fonctionnel *in vitro* basé sur la tolérance à la méthylation. En présence d'un agent de méthylation de l'ADN, seules les cellules avec un système MMR intact entrent en apoptose (**figure 19**). Au total, 88 variants faux-sens ont été criblés, dont 48 VSI (cohorte test) et 40 variants déjà classés bénins ou pathologiques (cohorte de validation). Le test développé a permis de classer correctement 39 des 40 variants de la cohorte de validation, démontrant la robustesse de cette approche. Sur les 48 VSI de la cohorte test, 12 ont été classés comme pathogènes et 28 comme bénins à l'issu du test. Les 8 derniers ont montré un phénotype intermédiaire. Les clinico-biologiques recueillies chez les patients et dans diverses bases de données ont permis de corroborer ses résultats.

Afin de faciliter la reclassification des VSI, on pourrait imaginer, à l'avenir, que ce genre de test soit intégré en routine dans les laboratoires de diagnostic génétique.

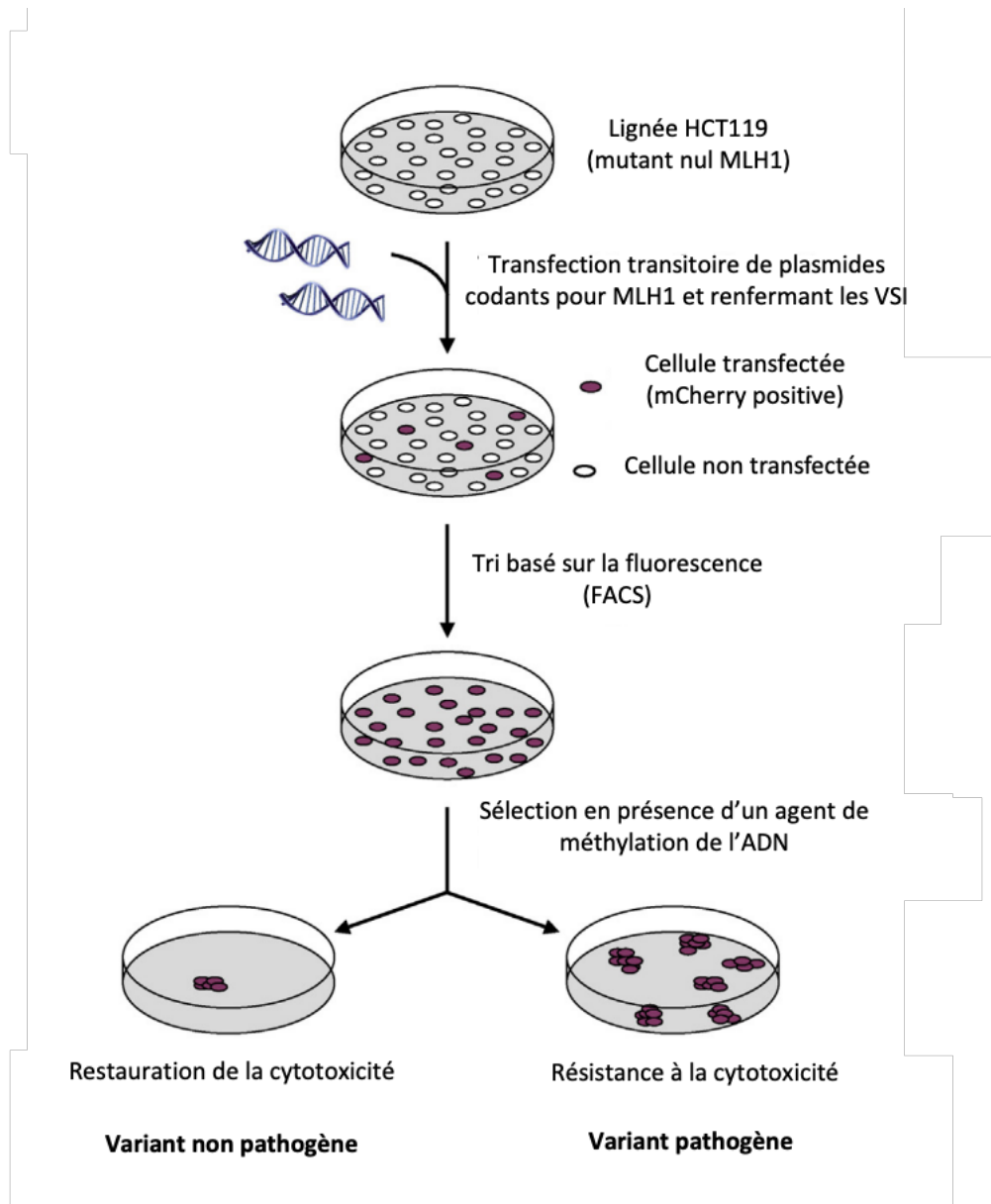


Figure 19 : Test fonctionnel basé sur la résistance à la méthylation pour la classification des variants de signification inconnue localisés dans le gène *MLH1*.

Les auteurs ont utilisé une lignée cellulaire naturellement déficiente pour le gène *MLH1* (mutant nul). Ils ont transfecté dans ces cellules un plasmide d'expression codant pour MLH1 refermant le variant à tester. Ce plasmide code également pour un gène rapporteur fluorescent (mCherry). Les cellules transfectées ont été recueillies par tri cellulaire activé par la fluorescence (FACS ; cellules mCherry positives) puis mises en présence de l'agent de méthylation de l'ADN (N-Methyl-N'-Nitro-N-Nitrosoguanidine) pendant quelques heures. Les cellules survivantes ont continué à proliférer pendant une dizaine de jours. Un test au cristal violet a permis d'évaluer le nombre de colonies résistantes. VSI = Variant de signification inconnue ; FACS = tri cellulaire activé par la fluorescence. Schéma traduit et adapté de Delphine Bouvet et al.³⁷⁹.

PARTIE II :

HYPOTHESES ET

OBJECTIFS

Grâce à l'émergence de la technologie CRISPR/Cas9 et aux multiples améliorations apportées à ces outils, la médecine est rentrée dans une nouvelle ère thérapeutique : celle de la chirurgie du génome. Fort de l'expérience acquise par l'équipe en biologie moléculaire, nous avons développé au sein du CHU de Rennes une plate-forme pour l'Exploration fonctionnelle du rôle Pathologique de variants génétiques rares (activité EPIQUE), localisée au laboratoire de cytogénétique et biologie cellulaire.

Cette nouvelle activité hospitalière, nous permet de **concevoir, tester et valider la spécificité de nouveaux traitements, *in vitro*, pour corriger les mutations responsables de pathologies génétiques rares**. Par la même occasion, nous pouvons utiliser les modèles cellulaires pour **créer des variants génétiques de signification inconnue (VSI) et tester leur rôle biologique (figure 20)**. Cette étape s'inscrit directement dans l'activité diagnostique du laboratoire. En reclassant les VSI en variant pathogène ou bénin, nous sommes en mesure d'apporter aux patients un diagnostic de certitude indispensable à leurs prises en charge ultérieure.

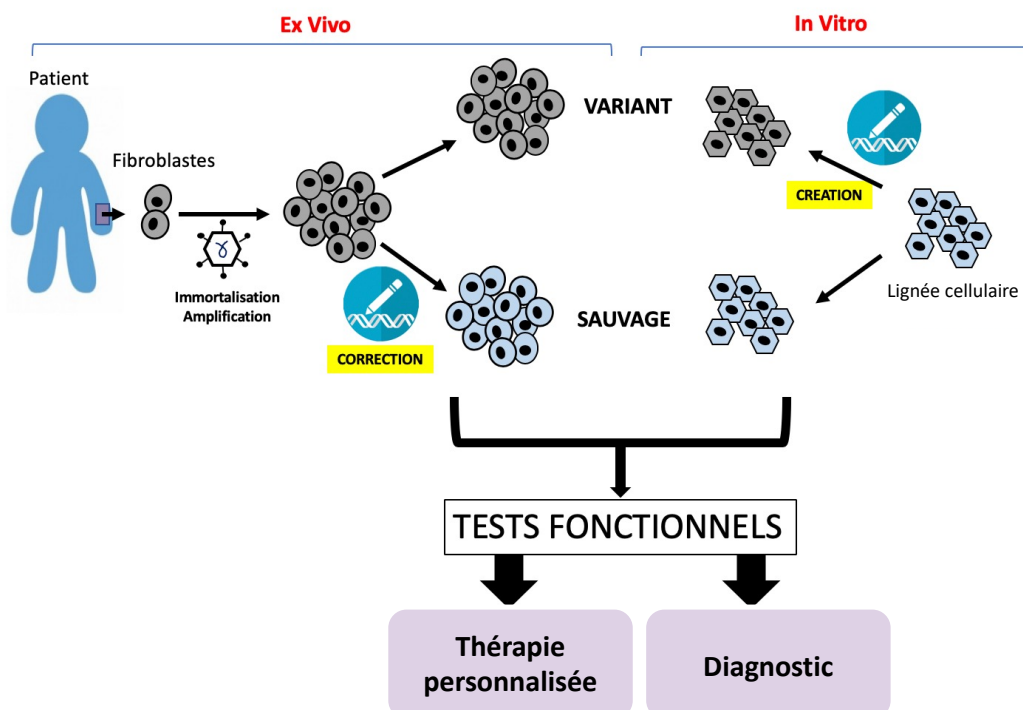


Figure 20 : Objectifs de la plateforme EPIQUE.

L'objectif principal de ma thèse a été de d'importer au laboratoire la technologie **BE** pour corriger un variant faux-sens de classe 4 dans les fibroblastes d'un patient atteint de gangliosidose à GM1 de type II (NM_000404.2(*GLB1*):c.907G>A). En reversant simplement le nucléotide A en G, j'ai pu restaurer une activité thérapeutique de l'enzyme défectueuse responsable de la pathologie. La spécificité du BE a été évaluée dans plus de 40 sites de l'ADN et n'a pas permis de détecter d'activité hors de la cible. Cette preuve de concept suggère qu'à l'avenir, le BE pourrait constituer une stratégie thérapeutique alternative à la thérapie génique additive, actuellement en cours d'évaluation clinique pour cette maladie. De plus, ce variant initialement rangé en classe 4 (probablement pathogène) a pu être reclassifié en variant de classe 5 (pathogène). Ces résultats ont été publiés au début de l'année 2023 dans la revue *The CRISPR journal* (première auteur) et sont présentés dans la première partie de la section résultats.

En parallèle à cet objectif principal, j'ai évalué d'autres outils d'édition du génome pour de tester le rôle de deux VSI directement mis en évidence en routine par le service de cytogénétique et biologie cellulaire du CHU de Rennes.

Le premier variant, identifié par le Pr Sylvie Jaillard, a été séquencé chez une patiente souffrant d'insuffisance ovarienne débutante. Ce variant faux-sens étant localisé dans le gène de l'*AMH*, codant pour une hormone clé intervenant dans le maintien de la réserve ovarienne, son rôle causatif était hautement probable (NM_000479.5:c.35T>G). Afin de reproduire ce variant dans un modèle cellulaire, j'ai importé au laboratoire la technologie **prime editing** et j'ai mesuré le niveau d'expression du gène *AMH* muté par western blot. Les données préliminaires démontrent une accumulation de l'isoforme non mature de l'hormone dans les cellules. Des expériences complémentaires sont en cours pour évaluer son activité. Les données sont présentées en section 2 de la partie résultats.

Le deuxième variant a été identifié par le Dr Laura Mary chez un fœtus atteint d'anomalies du développement sexuel. Celui-ci est localisé dans le gène *PBX1* (NM_002585.3:c.320G>A) qui code pour un facteur de transcription ubiquitaire, notamment impliqué dans le développement embryonnaire. A l'époque, je n'avais pas encore mis au point les protocoles pour le *base* et le *prime editing*. Le gène *PBX1* a donc été inactivé par **CRISPR/KO**, puis un plasmide de surexpression renfermant le

variant génétique a été transfecté dans les cellules. Malgré des résultats mitigés ne permettant pas de conclure sur la pathogénicité de ce variant, la récurrence des substitutions localisés dans ce codon chez les patients présentant des phénotypes similaires suggèrent un effet potentiellement délétère. Cette étude a été publiée le 17 janvier 2023 dans la revue *Genes* (2^{ème} auteur). L'article est présenté en section 3 de la partie résultats.

Au cours de ma thèse, j'ai aussi utilisé la technologie d'activation transcriptionnelle CRISPR (**CRISPRa**) dans le cadre de la révision d'un article scientifique. Dans cette étude, un crible gain-de-fonction CRISPRa a permis d'identifier un groupe de gènes (*SMAD3*, *SLC9A5*, *BIRC3*) impliqué dans la résistance des cellules de mélanome à un inhibiteur de BRAF. Ce crible a été réalisé à l'aide d'une bibliothèque composée de 70 290 ARN guides ciblant 23 430 gènes codants (au moins 3 ARN guides par gènes). J'ai montré que, selon la position de l'ARN guide sur la région promotrice du gène *SMAD3*, son niveau de transactivation pouvait varier. L'article est paru en mars 2021 dans la revue *Embo molecular medicine* (4^{ème} auteur/15). Il est présenté dans la dernière partie de la section résultats.

J'ai également participé à la rédaction deux revues scientifiques. La première porte sur l'activation de voie de signalisation AhR dans le cancer et des stratégies thérapeutiques envisageables. Elle a été publiée en juin 2021 dans la revue *Current Opinion in Immunology* (co-premier auteur). Cette revue s'inscrit dans une étude expérimentale menée au laboratoire visant à identifier les mécanismes de résistances aux inhibiteurs de BRAF. Dans cette étude, j'ai utilisé la technologie **CRISPR/KO** pour inactiver un gène de résistance et j'ai suivi la croissance tumorale des cellules dans un modèle de xénogreffe murin (données non présentées – projet en cours).

La deuxième revue traite du rôle des ARN circulaires dans les cancers hépatiques, et a été acceptée en novembre 2021 dans la revue *JHEP Reports* (co-premier auteur).

Le projet sur lequel j'ai travaillé au cours de mon stage de master a également été publié en octobre 2022 dans la revue *Life Science Alliance* (5^{ème} auteur/12). Cet article montre que les appariements non-canoniques entre le microARN miR-16 et un groupe d'ARN altère son activité suppresseur de tumeur dans le mélanome uvéal. Une signature transcriptomique associée à l'activité miR-16 a été établie et a permis de

discriminer le niveau de prédire la survie des patients. Le résumé de ces trois articles a été inséré dans la partie **annexe II** de cette thèse.

Initialement, mon projet de thèse portait sur le sujet suivant : caractérisation des cellules cancéreuses responsables de l'agressivité du mélanome uvéal.

Malheureusement, je n'ai pas pu poursuivre ce projet pour des raisons techniques. J'avais prévu de mettre en culture des cellules primaires de mélanome uvéal (banque du Pr. Frédéric Mouriaux) pour les étiqueter avec un barcode moléculaire avant de les xéno greffer chez la souris (**figure 21**). Malgré de nombreuses tentatives au cours de ma première année de thèse, les cellules primaires de mélanome uvéal n'ont pas supporté pas la culture *in vitro* dans mes conditions. Cela explique pourquoi, dans la littérature, personne n'a jamais réussi à cultiver de telles cellules. Par conséquent, je me suis focalisée sur ces nouveaux projets basés sur l'utilisation des outils d'édition du génome.

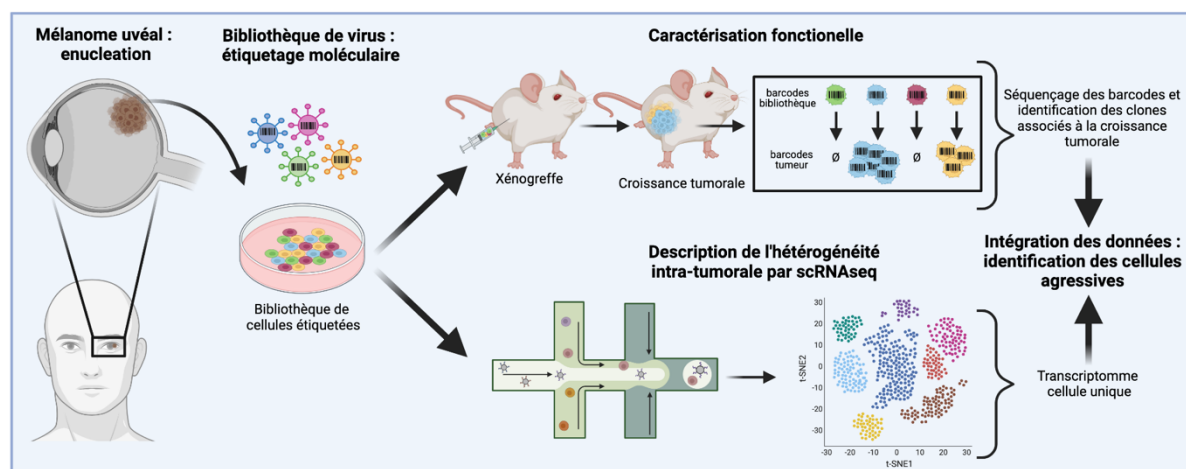


Figure 21 : Caractérisation des cellules cancéreuses responsables de l'agressivité du mélanome uvéal.

Désormais, il est possible de suivre l'évolution clonale et le devenir des cellules *in vivo* grâce aux outils d'étiquetage moléculaire. En couplant cette approche fonctionnelle chez la souris (xéno greffe) avec la technologie de séquençage de l'ARN en cellules uniques (scRNAseq) en amont, il est possible de d'identifier la signature transcriptomique de la ou les sou(s) population(s) cellulaire(s) présentes dans la tumeur primaire et responsable(s) de l'apparition de métastases hépatiques.

PARTIE III :

RESULTATS

RESULTAT 1

L'édition du génome permet de corriger une transition pathogène *in vitro* chez un patient atteint de gangliosidose à GM1 de type II

AVANT PROPOS

Ce premier projet a été réalisé au cours de ma deuxième année de thèse. Il a été déposé comme un preprint sur le site *BioRxiv* le 17 Avril 2022, où il a été téléchargé plus de 400 fois, puis publié dans la revue *The CRISPR Journal* le 6 février 2023.

INTRODUCTION DE L'ARTICLE

La gangliosidose est une pathologie génétique à transmission autosomique récessive touchant environ 1 personne pour 100 000 à 200 000 naissances. Cette pathologie est causée par la présence de mutations bi-alléliques au sein du gène *GLB1*, codant pour l'enzyme bêta-galactosidase, étant notamment responsable du clivage des résidus beta-galactosyl des gangliosides. La gangliosidose se traduit par une accumulation de substrats toxiques, les gangliosides GM1, au sein des lysosomes neuronaux. Malheureusement, cela provoque une neurodégénération à l'origine de troubles cognitifs sévères apparaissant entre la naissance et l'âge adulte selon la gravité (type I à III). Aujourd'hui, il n'existe aucun traitement curatif pour ces patients. Trois thérapies géniques sont en cours d'évaluation clinique (Lysogene, LYS-GM101, NCT04273269 ; PASSAGE BIO, PBGM01, NCT04713475 ; NHGRI, AAV9-GLB1, NCT03952637) et reposent sur l'addition d'une copie exogène de l'ADNc *GLB1* par un vecteur viral adéno-associé (AAV).

Dans cette étude, j'ai montré, à partir de fibroblastes dérivés d'un patient, que l'outil ABE est efficace à plus de 50% pour corriger un variant délétère localisé dans le gène *GLB1* et restaurer une activité thérapeutique bêta-galactosidase. Dans nos conditions expérimentales, aucune activité déaminase hors-cible significative n'a été détectée sur plus de 40 sites génomiques testés et sélectionnés à l'aide d'un outil bio-informatique. Une modification non désirée de type *bystander* a été observée à proximité du variant corrigé, mais les prédictions *in silico* ne suggèrent aucune conséquence sur l'activité bêta-galactosidase. Cette preuve de concept suggère que le BE pourrait être une alternative à la thérapie génique additive pour le traitement de la gangliosidose à GM1.

CONTRIBUTIONS

Grâce à une collaboration menée avec les médecins et biologistes des CHU de Rennes, Toulouse et Lyon, j'ai pu accéder à des échantillons de biopsie de peau d'une famille (trio) dont l'enfant a été diagnostiqué pour la gangliosidose à GM1 en 2019. A partir de ce matériel biologique, j'ai mis au point et mené toutes les expériences scientifiques « de paillasse », et j'ai analysé, mis en forme et interprété l'ensemble des résultats sous la supervision du Dr. David Gilot. Les analyses bio-informatiques et prédictions *in silico* ont été réalisés par Bénédicte Nouyou, Xavier Guillory et Leif A Eriksson. J'ai également supervisé le Dr. Louise Goujon qui a participé à ce projet pour l'obtention de son diplôme d'études spécialisées de génétique médicale.



ORIGINAL RESEARCH

Gene Editing Corrects *In Vitro* a G > A *GLB1* Transition from a GM1 Gangliosidosis Patient

Delphine Leclerc,¹ Louise Goujon,² Sylvie Jaillard,^{3,4} Bénédicte Nouyou,⁴ Laurence Cluzeau,⁴ Léna Damaj,⁵ Christèle Dubourg,^{6,7} Amandine Etcheverry,⁶ Thierry Levade,⁸ Roseline Froissart,⁹ Stéphane Dréano,⁷ Xavier Guillory,^{1,10} Leif A Eriksson,¹¹ Erika Launay,⁴ Frédéric Mouriaux,^{1,12} Marc-Antoine Belaud-Rotureau,^{3,4} Sylvie Odent,^{2,7} and David Gilot^{1,4,*}

Abstract

Ganglioside-monosialic acid (GM1) gangliosidosis, a rare autosomal recessive disorder, is frequently caused by deleterious single nucleotide variants (SNVs) in *GLB1* gene. These variants result in reduced β -galactosidase (β -gal) activity, leading to neurodegeneration associated with premature death. Currently, no effective therapy for GM1 gangliosidosis is available. Three ongoing clinical trials aim to deliver a functional copy of the *GLB1* gene to stop disease progression. In this study, we show that 41% of *GLB1* pathogenic SNVs can be replaced by adenine base editors (ABEs). Our results demonstrate that ABE efficiently corrects the pathogenic allele in patient-derived fibroblasts, restoring therapeutic levels of β -gal activity. Off-target DNA analysis did not detect off-target editing activity in treated patient's cells, except a bystander edit without consequences on β -gal activity based on 3D structure bioinformatics predictions. Altogether, our results suggest that gene editing might be an alternative strategy to cure GM1 gangliosidosis.

Introduction

Over the last two decades, high-throughput sequencing (HTS) technology has accelerated the identification of genetic alterations involved in human diseases, thus allowing many patients to benefit from an accurate diagnosis. Despite this historical breakthrough, a high number of diagnosed patients are currently awaiting personalized therapy. The development of curative gene therapies for these patients is indispensable.¹ According to ClinVar, ~54,500 human pathogenic genetic variants have been associated with diseases and more than 58% are single nucleotide variants (SNVs) (<https://www.ncbi.nlm.nih.gov/clinvar/>). This observation suggests that gene editing strategies aimed at correcting the root cause of disease could be an alternative strategy to cure disease instead of classical gene therapies aimed at bringing a novel functional copy of the altered genes.^{2,3}

Nowadays, gene editing approaches such as base and prime editing (BE and PE) enable specific restoration of a functional gene by correcting pathogenic variants.³ BE has been successfully used to cure multiple diseases in different models and species.^{4–8} To date, the most advanced application in humans is to cure sickle cell anemia through *ex vivo* targeting of hematopoietic stem and progenitor cells from patients.⁵ BE also shows promise in limiting the deleterious effect of progerin (SNV altering the *Lamin A* mRNA splicing), as recently demonstrated in a murine progeria model,⁸ and to lower cholesterol in serum of primates by turning off the *PCSK9* gene.^{7,9}

Although this innovative gene editing approach is promising, it is important to keep in mind the editing efficiency, the potential off-target effects, and the challenge in reaching all targeted cells *in vivo*.^{10,11} To date, BE has the best

¹INSERM U1242, OSS, Univ Rennes, Rennes, France; ²CHU Rennes, Service de Génétique Clinique, Centre de Référence Maladies Rares CLAD-Ouest, FHU GenOMEDS, ERN ITHACA, Hôpital Sud, Rennes, France; ³INSERM, EHESP, IRSET-UMR_S, 1085, Université Rennes 1, Rennes, France; ⁴Service de Cytogénétique et Biologie Cellulaire, CHU Rennes, Rennes, France; ⁵Department of Pediatrics, Competence Center of Inherited Metabolic Disorders, Rennes Hospital, Rennes, France; ⁶Laboratoire de Génétique Moléculaire et Génomique, Centre Hospitalier Universitaire de Rennes, Rennes, France; ⁷Univ Rennes, CNRS, IGDR (Institut de Génétique et Développement de Rennes), UMR 6290, ERL U1305, Rennes, France; ⁸Laboratoire de Biochimie, CHU de Toulouse, Pôle biologie, Institut Fédératif de Biologie, Toulouse, France; ⁹CHU Lyon HCL, LBMM5-Service Biochimie et Biologie Moléculaire, UM Pathologies Héritaires du Métabolisme et du Globule Rouge, Bron, France; ¹⁰Univ Rennes, CNRS, ISCR (Institut des Sciences Chimiques de Rennes)-UMR 6226, Rennes, France; ¹¹Department of Chemistry and Molecular Biology, University of Gothenburg, Göteborg, Sweden; and ¹²Department of Ophthalmology, CHU Rennes, Univ Rennes, Rennes, France.

*Address correspondence to: David Gilot, INSERM U1242, OSS, Univ Rennes, avenue Bataille Flandres-Dunkerque, Rennes 35000, France, E-mail: david.gilot@univ-rennes1.fr

efficiency rate compared to PE or CRISPR-Cas9 coupled to a donor. However, only C-to-T (CBE) and A-to-G (ABE) BE methods are fully validated for gene therapy.^{12,13}

Moreover, CBE and adenine base editor (ABE) deaminases are active on a wide editing window. Indeed, the maximum deaminase activity extends from nucleotide 4 to 6 from the 5' end of the sgRNA depending on the BE version. This point suggests that proximal C or A nucleotides should be edited into T or G within the editing window, depending on the genomic context.^{14,15} Consequently, two kinds of editing events might occur with BEs: the expected on-target edit and the unwanted on-target one(s) called bystander edit(s). Although these bystander events are sometimes unavoidable, plenty of strategies have been proposed and published to improve BE window specificity and reduce their occurrence.^{16,17}

To facilitate the design of sgRNAs for BE applications, several improvements have been implemented such as less stringent protospacer adjacent motif (PAM) than the canonical 5'-NGG-3' (such as NRN and NYN PAM; SpRY Cas enzyme) and high-fidelity (HF) BE Cas enzymes.^{18,19} So the latest generation of BEs is characterized by a high efficiency rate of editing and a low level of off-target edits since these Cas9s cleave only one DNA strand (nickase) and are sometimes HF, making them exploitable for gene therapy.²⁰

In humans, CBE and ABE should, respectively, correct, *in theory*, 14% and 47% of *GLB1* pathogenic variants.³ In this study, we investigated whether BE may constitute an alternative strategy to cure GM1 gangliosidosis. Indeed, there is currently no effective therapy, and only supportive treatments can be offered.²¹ GM1 gangliosidosis is an autosomal recessive lysosomal storage disorder estimated to occur in 1 in 100,000 to 200,000 newborns.²² There are four different types of GM1 gangliosidosis based on the age when symptoms first appear and the severity of disease progression. Defects in the *GLB1* gene (coding the β -galactosidase [β -gal]) cause impaired enzyme activity leading to the toxic accumulation of gangliosides and neurodegeneration that presents as cognitive impairment, paralysis, and early death.

Three ongoing clinical trials (NCT03952637, NCT04273269, and NCT04713475) aim to deliver a functional copy of the *GLB1* gene to slow down or stop disease progression (classical gene therapy), but cannot reverse the damage already caused by the disease. For two clinical trials, the *GLB1* cDNA is carried by adeno-associated viral (AAV) vector injected into the cisterna magna to reach the neuronal cells through the cerebrospinal fluid and to break down GM1 ganglioside. Preclinical studies clearly demonstrated the ability of AAV to transport genes into neuronal cells, suggesting that these vectors should also bring both BE and sgRNA.^{23–25}

Transgenes transferred by this type of vectors are transcriptionally active and are maintained as an extra-chromosomal (episomal) form in the transduced cells for a long-term period.²⁴ However, the preclinical trials have been mainly performed in short lifespan animals such as mice and felines, making it difficult to evaluate the long-term beneficial effect required for patients.^{24,26–29} These studies mainly demonstrated a reduction of lysosomal storage in the central nervous system and an extended lifespan for these animal models.

To the best of our knowledge, the long-term expression of *GLB1* in large animal models through AAV vectors has not been investigated. In other words, it is unknown how long the therapeutic effect will last following the initial establishment of transgene expression, although animal and emerging human data show that the expression can be maintained for at least 10 years in muscles.³⁰ The durability of therapeutic response is key to long-term treatment success, especially since immune responses to AAV vectors may prevent re-dosing with the same therapy.³¹

In this study, we show that 82% of genetic alterations of *GLB1* gene are SNVs. Among the pathogenic variants, 41% could be targeted by ABE. As a proof-of-concept experiment, we designed and validated *in vitro* an ABE strategy for a young patient with GM1 gangliosidosis. Based on the ABE efficiency and the quantification of off-target edits, our data suggest that gene editing is an alternative strategy to cure GM1 gangliosidosis.

Methods

Editorial policies and ethical considerations

Written informed consent was obtained from the patient and his parents. All procedures were in accordance with the ethical standards of the Ethics Committee of Rennes University Hospital and the French law (CCTIRS Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé). Peripheral blood samples (for DNA extraction) and skin biopsies (for functional analyses and personalized gene therapy development) were collected from all the participants.

Exome sequencing and bioinformatics pipeline

Trio exome sequencing was performed at Rennes Hospital University (Molecular Genetics and Genomics Laboratory) using standard pipeline as previously described.³² Identified pathogenic variants were described according to the Human Genome Variation Society nomenclature guidelines (<https://varnomen.hgvs.org/>).

Patient-derived fibroblasts: primary culture and immortalization

The fresh skin punch biopsies were dilacerated with scalpels under a sterile area and then transferred into T25 flasks containing 5 mL of AmnioMAX C-100 (Gibco) medium. After 1–2 weeks, explant growing cells were washed with Dulbecco's phosphate-buffered saline (DPBS; Lonza), trypsinized with TrypLE™ Express Enzyme (1×; Gibco), and passaged in T75 flasks for fibroblast amplification. Primary cells were then immortalized by SV40 T antigen lentiviral transduction. Lentiviral particle production was performed in HEK293T cell line by following Trono Lab recommendations (<http://tronolab.epfl.ch>).

Briefly, psPAX2 (Addgene plasmid ID #12260), pVSVG (Addgene plasmid ID #8454), and pLOX-TtagiresTK plasmids (Addgene plasmid ID #12246) were transfected into HEK293T cells with lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's instructions. Cell supernatant containing lentiviral particles was collected after 2 days, centrifuged, and filtered (0.45 μm) to eliminate HEK293T-derived cell debris. Fibroblasts in primary culture (P6 well plate; Falcon®) were infected with 500 μL of this infectious media in the presence of 8 μg/mL polybrene (Sigma) to achieve viral infection.

Cell culture

Immortalized patient-derived fibroblasts and HEK293T (ATCC) cells were cultured, respectively, in RPMI-1640 medium (Lonza) or Dulbecco's modified Eagle's medium (Gibco), each supplemented with 10% fetal bovine serum (Gibco), 1% glutamine (Lonza), and 1% penicillin-streptomycin (Sigma). When confluent, cells were washed with DPBS (Lonza), digested with TrypLE Express enzyme (1×; Gibco), and subsequently passaged at 1:10 ratio. All cells were maintained at 37°C in a humidified incubator with 5% CO₂. Cell lines were routinely tested for mycoplasma infection with Plasmotest Mycoplasma Detection Kit (Invivogen) and negative results were obtained.

DNA extraction, polymerase chain reaction, and Sanger sequencing

DNA was extracted from frozen cell pellets with the Nucleospin tissue DNA extraction kit (Macherey-Nagel). DNA concentration was evaluated using the NanoDrop1000 spectrophotometer (Thermo Fisher Scientific). Polymerase chain reaction (PCR) was performed from 50 ng DNA with the Phusion™ High-Fidelity DNA Polymerase kit (Thermo Fisher Scientific). The list of primers is presented in Supplementary Table S1. Samples

were subjected to thermal cycling as follows: 98°C for 30 s (initial denaturation step), 98°C for 10 s, primers' annealing temperature for 15 s, 72°C for 30 s (35 cycles), and 72°C for 5 min (final elongation step).

Amplicon size and purity were checked on 1.5% agarose gel electrophoresis before Sanger's reaction. Sanger sequencing was performed directly on 1 μL of PCR reaction in the presence of 0.5 μM primer and the Big Dye Terminator V3.1 (Applied Biosystems™). The following thermocycler program was used: 96°C for 5 min (initial denaturation step), then 96°C for 1 min, primers' annealing temperature for 1 min, and 62°C for 2 min (30 cycles). Sequence products were purified on Sephadex G50 beads (GE Healthcare) and directly loaded onto 3130xl Genetic Analyzer capillary electrophoresis laser coupled system (ABI PRISM).

RNA extraction, reverse transcription, and quantitative PCR

RNA extraction was performed from frozen tissue in culture plates (P6 well plate; Falcon) by using the NucleoSpin RNA kit (Macherey-Nagel). RNA concentration was measured with NanoDrop1000 spectrophotometer (Thermo Fisher Scientific). Reverse transcription was performed on 500 ng total RNA with the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). Quantitative PCR (qPCR) was performed in 384-well plates on 2.5 ng of reverse-transcribed cDNA using the SYBR Green PCR Master Mix (Applied Biosystems) in the presence of 1 μM of forward and 1 μM of reverse primers, using the QuantStudio 5 quantitative PCR equipment (Applied Biosystems). The primer sequences used are available in Supplementary Table S1. Raw data were extracted with QuantStudio Design and Analysis Software (Applied Biosystems). Relative gene expression compared to control conditions was calculated by using the 2^{-ΔΔCt} method. *GAPDH* was used as a housekeeping gene for normalization.

Protein extraction

Cells were lysed and proteins were extracted from fresh fibroblast cells with RIPA buffer (Thermo Fisher Scientific) supplemented with protease and phosphatase inhibitor mini-tablet (Pierce, Thermo Fisher Scientific). Protein quantification was performed with Pierce BCA protein assay kit (Thermo Fisher Scientific) and absorbance was measured with spark microplate reader (TECAN) at 562 nm.

β-Gal enzymatic activity measurement

β-Gal activity was evaluated using a fluorogenic method initially described by Ho and O'Brien.³³ Briefly, 20 μL of

fresh cell lysates were incubated in a white 96-well microplate (Grenier Bio-One) with 100 μ L of buffer solution (pH 4.3) containing 0.5 mM of 4-methylumbelliferyl- β -D-galactopyranoside substrate (Sigma). A kinetic analysis was performed at 37°C for 1–2 h. Every 4 min, the fluorogenic substrate was excited at 366 nm and fluorescence emission, proportional to enzyme activity, was measured at 442 nm with a spark microplate reader (TECAN). Enzyme activity was assessed from the slope of the fluorescence= f (time) curve and normalized per microgram of protein.

Western blot

Protein samples were reduced and denatured with lithium dodecyl sulfate and reducing agent for 10 min at 70°C. Then, 20 μ g of each sample was loaded in 4–12% Bis-Tris Gel (NuPAGE; Invitrogen) and migration was done with following parameters: 1 h 30 min, 200 V, and 400 mA. Proteins were then transferred onto a nitrocellulose membrane (Invitrogen™ iBlot™ Transfer Stack) with iBlot transfer device system (Invitrogen).

After transfer and incubation with a blocking solution (1 h at room temperature, TBS-Tween 0.1%; 5% bovine serum albumin; Eurobio), the membrane was probed with the primary antibody overnight at 4°C (rabbit anti- β -galactosidase antibody, 15518-1-AP Protein Tech, dilution 1/1000; and mouse anti- α -tubulin, dilution 1/10,000; T6199 Sigma-Aldrich) and then for 1 h at room temperature with the horseradish peroxidase (HRP)-linked secondary antibody (HRP-linked anti-rabbit IgG, dilution 1/2000; #7074 Santa Cruz; and HRP-linked anti-mouse IgG, dilution 1/2000; #7076 Santa Cruz). Signal was detected with Amersham ECL select substrate (Cytiva) under a luminescent image analyzer (ImageQuant LAS 4000). Uncropped western blots are available in Supplementary Figure S1.

Cloning

Single guide RNAs were designed according to the recommendations of Walton et al,¹⁹ Supplementary Table S1. Designed oligonucleotides were integrated into BPK1520 backbone (Addgene plasmid ID #65777) by golden gate assembly as previously described.³⁴ Plasmids were transformed into NEB® Stable competent *Escherichia coli* C3040H (New England BioLabs) by following the manufacturer's recommendations. Bacteria were seeded onto ampicillin (100 μ g/mL) Lysogeny broth (LB) agar plates and incubated overnight at 37°C. Isolated colonies were then amplified overnight at 37°C with LB growth medium and ampicillin (100 μ g/mL) under constant agitation. Plasmid DNA was purified

with NucleoBond® Xtra Maxi kit (Macherey Nagel) according to the manufacturer's instructions. Constructs were verified by Sanger sequencing. The plasmid encoding the sgRNA number 8 used in this study is available on Addgene (BPK1520-sgRNA GLB1; Addgene #184378).

ABE transfection

Immortalized fibroblasts were seeded in 6-well plates (Falcon) (500,000 cells/well). Approximately 24 h after seeding (~60% confluency), cells were transfected with 11.55 μ L of Viafect reagent (Promega) according to the manufacturer's protocols with 2200 ng of ABEmax(7.10)-SpRY-P2A-EGFP plasmid (Addgene plasmid ID #140003) and 1100 ng of sgRNA plasmid (Addgene plasmid ID #65777). The medium was replaced 6 and 24 h after transfection to eliminate dead cells. Forty-eight hours after transfection, cells were washed with DPBS (Lonza), digested with TrypLE Express enzyme (Gibco), and resuspended into RPMI-1640 medium (Lonza). GFP positive cells were sorted by fluorescence-activated cell sorting (FACS, FACSaria FUSION; Becton Dickinson). Sorted cells were either lysed for DNA extraction and HTS or individualized by limit dilution for clonal amplification (96-well plates, 1 cell/well; Falcon).

Approximately 3 weeks after the limit dilution, pure clones were amplified in 12-well plates (Falcon). To easily identify the edited cells, β -galactosidase assays were performed and gene editing was confirmed by Sanger sequencing.

siRNA transfection

Immortalized fibroblasts were seeded in 6-well plates (500,000 cells/well; Falcon). Approximately 24 h after seeding (~60% confluency), cells were transfected with 50 nm siRNA CTRL or siRNA *GLB1* (IDT, sequence available in Supplementary Table S1) with lipofectamine RNAimax (Thermo Fisher Scientific) according to the manufacturer's instructions. Forty-eight hours after transfection, cells were stored at –80°C and proteins were then extracted.

Off-target identification (CRISPOR)

Off-target sites were predicted bioinformatically by using the CRISPOR online tool (<http://crispor.tefor.net>).³⁵ Putative off-target sites were ranked according to cutting frequency determination (CFD) score³⁶ and top 10 off-target hits were selected for deep sequencing (amplicons) (data are available in Supplementary Table S3). In addition, putative exonic off-target sites were checked by exome sequencing.

Amplicon HTS

Genomic regions of interest were amplified by PCR (Phusion High-Fidelity DNA Polymerase; Thermo Fisher Scientific or KAPA2G Robust HotStart PCR; Sigma) from genomic DNA samples. Amplicons were purified with AMPure XP magnetic beads (Beckman Coulter). The library was prepared with the SureSelect XT HS2 DNA kit (Agilent). We followed the manufacturer's recommendations (except that initial DNA fragmentation, hybridization, and capture steps were skipped). Pooled libraries were sequenced on a MiSeq instrument (Illumina) with a 2×250 bp paired-end running method (MiSeq Reagent Nano Kit v2 500 cycles; Illumina). The flow cell was loaded with 5 pM pooled libraries containing 5% PhiX control V3 (Illumina). Raw sequencing data were demultiplexed with Bcl2Fastq software (v2.19; Illumina) (data are available in Supplementary Table S3).

CRISPResso2

FastQ files were submitted to CRISPResso2³⁷ for precise editing quantification. We used the following parameters: `-wc = -10`, `-q = 30`, `--min-bp-quality-or-N = 30`, `--conversion-nuc-from = A`, and `--conversion-nuc-to = G`. The matrix “selected-nucleotide-percentage-table-around-sgRNA” generated by CRISPRESSO was used to calculate the potential editing percentage. For both on-target and off-target data, the line with nucleotide “N” (*Q* score <30) was removed with a Python script (Supplementary Fig. S2) to avoid potential editing underestimation (data are available in Supplementary Table S4).

Three-dimensional structure prediction

Maestro Software (Schrödinger Release 2021-3; Schrödinger, LLC, New York, NY, 2019) was used for visualization, basic molecular modeling steps (protein preparation, energy minimization, etc.), as well as for the molecular dynamic (MD) simulation trajectory using Desmond.³⁸ Crystal structure of the GLB1 protein in complex with galactose (PDB ID: 3THC) served as the starting point for molecular modeling. The protein preparation wizard was used to prepare the crystal structure as follows: hydrogen atoms were added and possible metal binding states generated.

The protonation and tautomeric states of Asp, Glu, Arg, Lys, and His were adjusted to match a pH of 4.3 and possible orientations of Asn and Gln residues were generated. Hydrogen bond sampling with adjustment of active site water molecule orientation was performed using PROPKA. Water molecules with fewer than three hydrogen bonds to nonwater molecules were deleted. Finally, the protein-ligand complexes were subjected to

geometry refinements using the OPLS4 force field³⁹ in restrained minimizations. The S302G mutant was then generated using the “Residue and Loop mutation” tool included with maestro, followed by an energy minimization of the whole structure in implicit water.

The GLB1 wild-type (WT) and S302G structures were then prepared for MD with the system builder panel using the following parameters: each system was solvated with TIP3P water⁴⁰ under periodic boundary conditions with the minimum distance between any atom in the solute and the edge of the periodic box being 10.0 Å; Na⁺/Cl⁻ counterions were added as appropriate for the neutralization of the system; 0.154 M NaCl. MD simulations for each system were performed with Desmond as follows: 3×100 ns total time; OPLS4 force field; 100 ps trajectory recording intervals; system energy set to 1.2, NPT ensemble class; 300.0K; 1.01325 bar; and model relaxed before simulation. Quality control of our simulations was done using the “Simulation Quality Analysis” tool as well as the “Simulation Interactions Diagram” tool.

Statistical analysis

All data are expressed as the mean – standard error of the mean of at least three individual experiments, excepting the HTS experiments. Data were analyzed by Student's *t*-test using GraphPad Prism v8.0.1 (GraphPad Software, San Diego, CA). *p*-Values were deemed to be statistically significant as follows: **p* < 0.05; ***p* < 0.01; ****p* < 0.001; *****p* < 0.0001. All raw data are available in Supplementary Table S5.

Results

We investigated here whether genome edition could be a credible approach to treat GM1 gangliosidosis by correcting *GLB1* SNVs.

GLB1 genetic variants

First, we analyzed the genetic alterations of the *GLB1* gene found in ClinVar to characterize the origin of GM1 gangliosidosis (Fig. 1 and Supplementary Table S2). The *GLB1* variants are almost exclusively germline (>95%) SNVs (82%), the majority of which are missense (60%) and some frameshift (11%) (Fig. 1A–C). It is interesting to note that the *GLB1* non-coding sequence is also impacted (untranslated region and splice sites; ~20%).

Next, we focused on the 93 known pathogenic variants (class 5 corresponding to 17% of all variants, *N* = 532) to estimate the ability of BE and PE to correct them (Fig. 1D, E). Importantly, PE, ABE, and CBE should, in theory, correct up to 100, 41, and 15%, respectively,

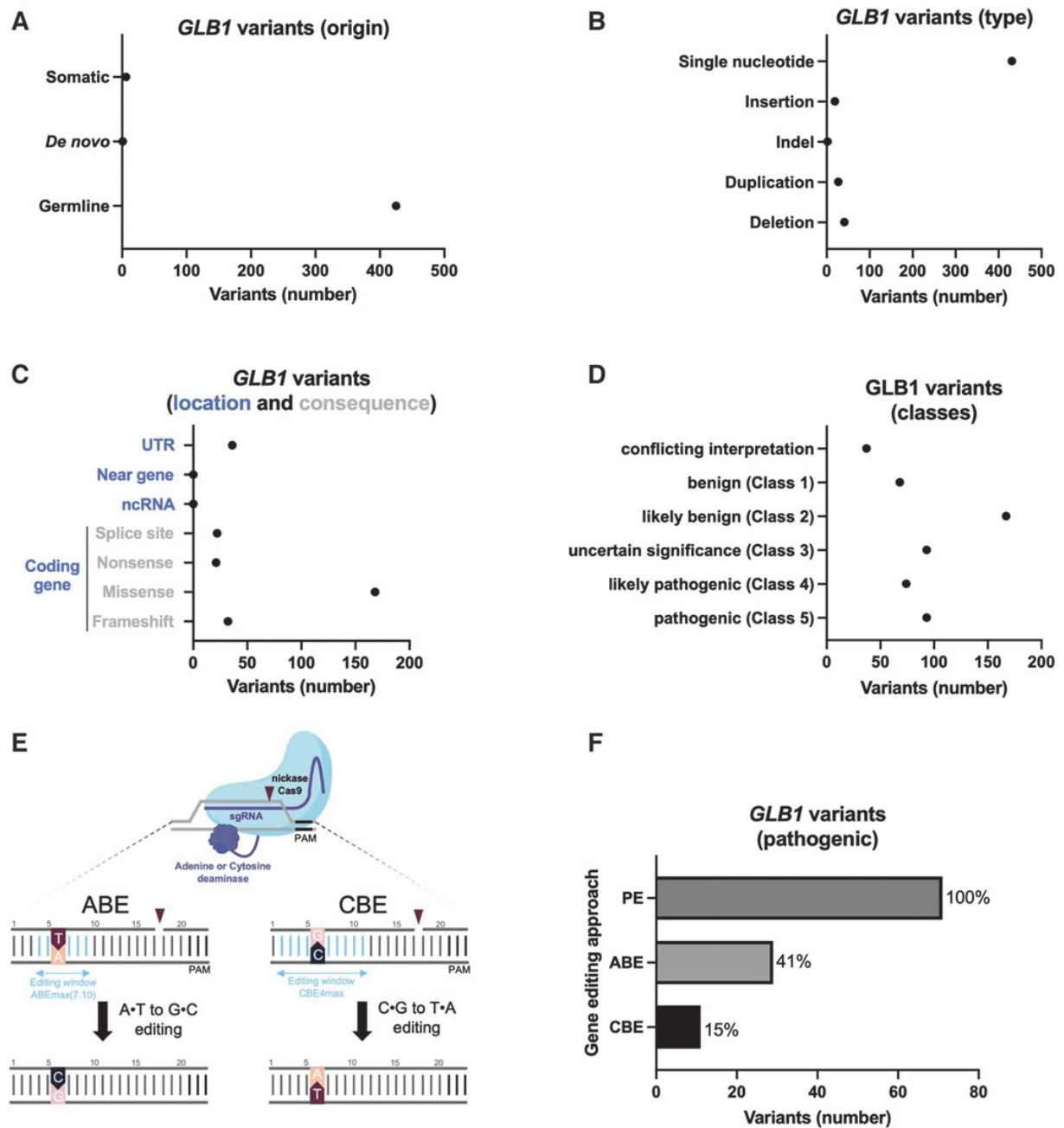


FIG. 1. Genetic alterations of *GLB1*. **(A)** Origin of the *GLB1* variants (somatic, *de novo*, or germline), **(B)** types of *GLB1* variants, **(C)** location and consequence of *GLB1* variants, **(D)** number of *GLB1* variants in the different classes, **(E)** scheme recapitulating base editor effects (ABE and CBE), **(F)** number of *GLB1* variants (class 5) that might be edited by PE, ABE, or CBE. ABE, adenine base editor; CBE, cytosine base editor; PE, prime editing.

of *GLB1* pathogenic variants described in ClinVar (Fig. 1F). Of course, the genomic context of each variant must be taken on a case-by-case basis to consider compatibility with editing window and PAM of specific editors. Since BE is currently the most efficient approach for gene editing in humans,³ we evaluated the ability of ABE to correct a missense variant associated with GM1 gangliosidosis as a proof-of-concept experiment. We selected a patient with a missense variant targetable by ABE (Fig. 2).

GLB1 characteristics of the patient

The patient is the second child of nonconsanguineous parents (Fig. 2A). He has one healthy brother. Pregnancy and delivery were uneventful. He was born at 40 weeks of gestation, and his birth weight was 4015 g. He achieved motor skills with slight delay, as he was able to sit at 9 months, to stand up at 19 months, and to make his first steps at 22 months of age (Fig. 2A). At 30 months of age, he began to lose previously acquired milestones as he could no longer walk independently. Regarding speech development, he could pronounce his first words at the end of his first year of age, but he did not progress further and progressively lost speech during the following months.

At 22 months of age, the parents noticed some episodes of eyelid shaking. The ophthalmologic examination was normal. A first brain magnetic resonance imaging (MRI) was performed at 23 months of age showing normal results. At 3 years of age, generalized epilepsy became obvious. Initial anticonvulsant therapy by Levetiracetam was switched to valproic acid due to side effects. Follow-up cerebral MRI at 3 years of age showed generalized brain atrophy and hypomyelination.

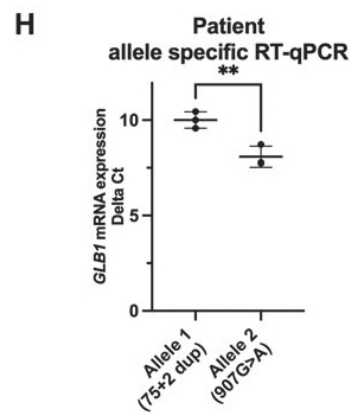
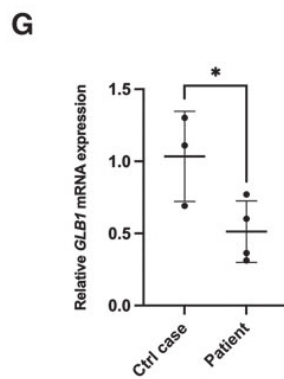
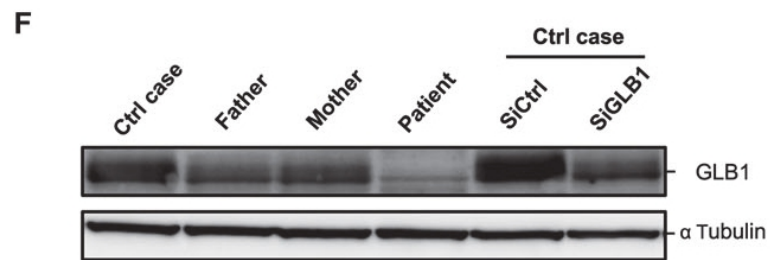
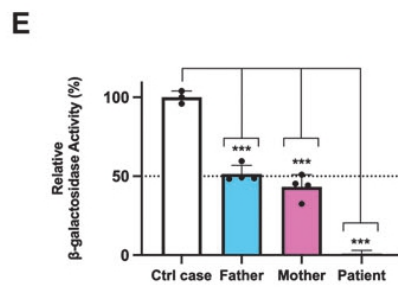
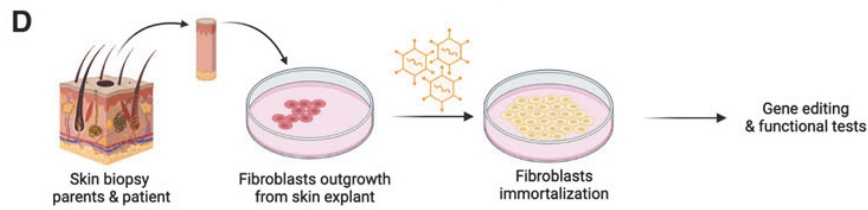
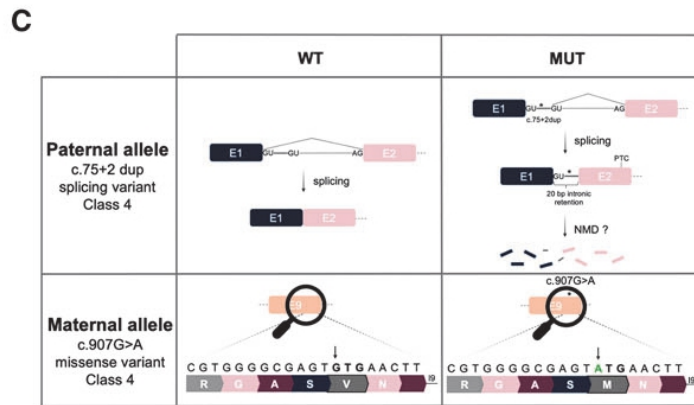
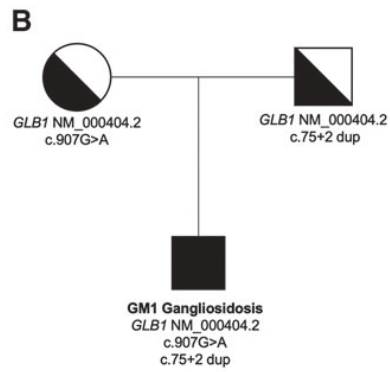
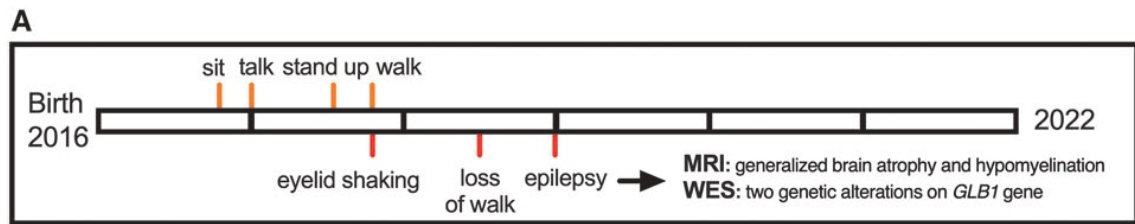
The first genetic investigations, which included screening for Steinert disease, Angelman and Prader Willi syndrome, and CGH array analysis, were normal (data not shown). Trio whole exome sequencing (WES) of the child and his parents identified two heterozygous probable pathogenic variants in the *GLB1* gene (NM_000404.2): c.907G>A (p.Val303Met) and c.75+2 dup, inherited from his mother and father, respectively (Fig. 2B, C). The first variant is a missense variant (never described) and the second is located at a splice site, leading to a 20-nucleotide intron retention in *GLB1* RNA.⁴¹

Since the *GLB1* genetic alterations suggested an altered β -gal activity associated with GM1 gangliosidosis, we evaluated it in white blood cells. Analyses revealed a pathogenic drop of β -gal activity (8 nmol/h/mg of proteins in the patient vs. control: 197 nmol/h/mg of proteins, data not shown). The residual β -gal activity was estimated to be 4% of control. As a control, the neuraminidase activity was similar for the patient and the control sample (data not shown). Thus, the abnormal β -gal activity supported the pathogenicity of the two variants identified in *GLB1* and led to the diagnosis of a late-onset form of infantile GM1 gangliosidosis for this patient.

To confirm this diagnosis, three skin biopsies were performed to generate primary fibroblast cultures (parents and patient) (Fig. 2D). To obtain enough cells, we immortalized these fibroblasts using T antigen from SV40 by a lentiviral transduction.

In accordance with results obtained with Peripheral Blood Mononuclear Cells, we confirmed that patient's fibroblasts displayed an undetectable β -gal activity in our experimental conditions (Fig. 2E). In contrast, the β -gal activity detected in fibroblasts isolated from the parents was reduced by 50% when compared to control

FIG. 2. Characterization of patient-derived fibroblasts. **(A)** Milestones of the patient with GM1 gangliosidosis, **(B)** GM1 gangliosidosis patient pedigree, indicating autosomal recessive inheritance (genealogical tree), **(C)** hypothetical effect of inherited variants on *GLB1* mRNA and amino-acid sequence, **(D)** workflow describing patient-derived fibroblast purification and immortalization using SV40 T antigen (transduced by lentivirus), **(E)** β -galactosidase activity in patient-derived fibroblasts. The measured values of the patient were compared to the values obtained from his parents and a control case (control fibroblasts), $n \geq 3$ biologically independent experiments, each histogram represents the mean \pm SD. **(F)** GLB1 protein levels in the same fibroblasts. Specificity of the GLB1 antibody raised against GLB1 protein was evaluated using siRNA targeting *GLB1* mRNA or a nontargeting siRNA (siCTR). Western blot results are representative of at least three experiments and α -tubulin serves as a loading control, **(G)** expression levels of *GLB1* mRNA evaluated by RT-qPCR in patient-derived fibroblasts or control fibroblasts, **(H)** allele-specific qPCR quantification in patient-derived fibroblasts (Δ Ct). Otherwise indicated, data were subjected to Student's *t*-test and *p*-values were deemed to be statistically significant as follows: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. Raw data are available in Supplementary Table S5. NMD, nonsense-mediated mRNA decay; RT-qPCR, real-time quantitative reverse transcription PCR; SD, standard deviation.



fibroblasts. These results demonstrated that the combination of these two inherited mutations abrogated the β -gal activity and expression (Fig. 2F). Next, we confirmed a reduced amount of *GLB1* mRNA in patient-derived fibroblasts when compared to control fibroblasts (Fig. 2G) and both alleles were detectable (Fig. 2H). Allele 1 (c:75+2dup) has reduced expression compared to allele 2, probably due to nonsense-mediated decay of transcript containing 20 bp of intronic sequence as already suggested.⁴¹ In conclusion, we showed that the combination of these two genetic alterations explains this case of GM1gangliosidosis.

Adenine base editing strategy for GM1 gangliosidosis

To conceive a therapeutic solution for this patient, we selected the ABE approach to correct only one allele; the *GLB1* c.907G>A (p.Val303Met) mutation (maternally inherited), aiming to restore a β -gal activity (50%) as found in his asymptomatic parents (Fig. 3A). The ABE strategy is known to reach a greater efficiency than PE, and ABE is already used in clinical trials.³ Only a PE approach has the potential to repair the other genetic alteration inherited from the father.⁴²

To optimize our chance of designing effective sgRNA for ABE, we selected the human codon-optimized ABE-max(7.10): A-to-G base editor with nSpCas9 SpRY (D10A/L1111R/D1135V/G1218R/E1219F/A1322R/R1335V/T1337R).¹⁹ The main advantage of this Cas9 variant is that the classical NGG PAM evolves to NRN or NYN, with NRN PAM being better recognized than NYN PAM (NRN>NYN). Using this near-PAMless engineered CRISPR-Cas9 variant, we successfully designed and cloned two sgRNAs (Fig. 3A–C). Next, ABE-

max(7.10) and one sgRNA were transfected into immortalized fibroblasts derived from the patient (Fig. 3D).

Transfected cells were selected using the enhanced green fluorescent protein (EGFP) signal encoded by the plasmid pCMV-T7-ABEmax(7.10)-SpCas9-NG-P2A-EGFP (Fig. 3D). Five days after the enrichment of transfected cells (EGFP⁺-sorted cells), the BE efficiency was estimated by DNA sequencing as previously described.^{6,37} CRISPResso2 analyses indicated that A907 was converted into G907 (reference nucleotide) in ~52% of reads from the maternal allele as expected with this BE approach (Fig. 3E). The sgRNA 7 was inefficient (Supplementary Table S4), confirming the requirement to evaluate several sgRNAs for BE.

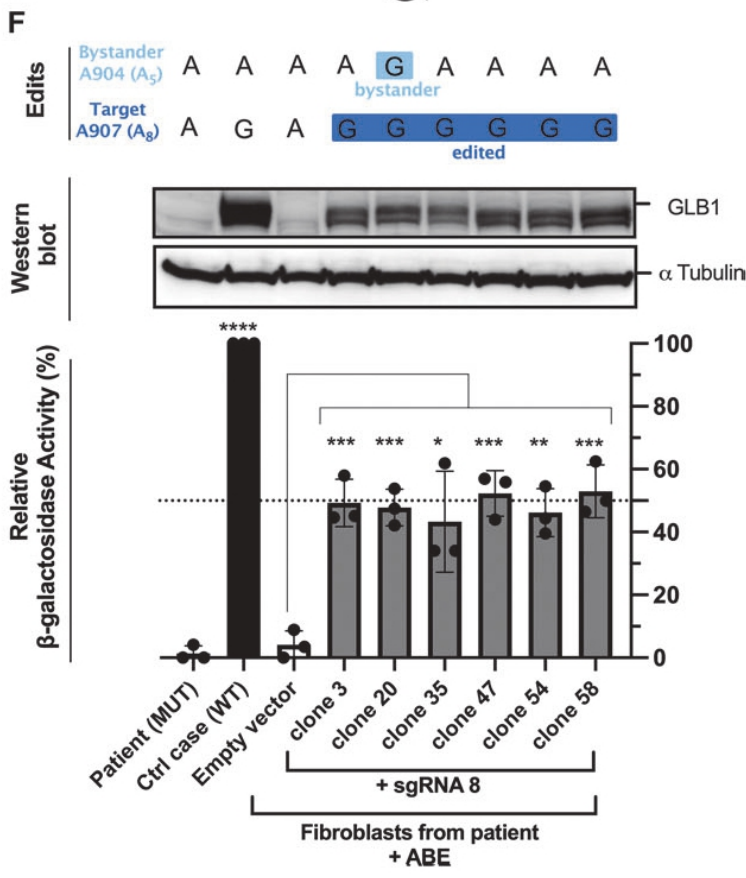
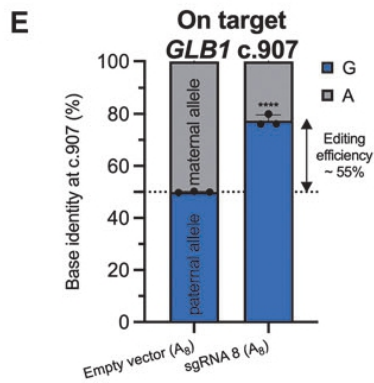
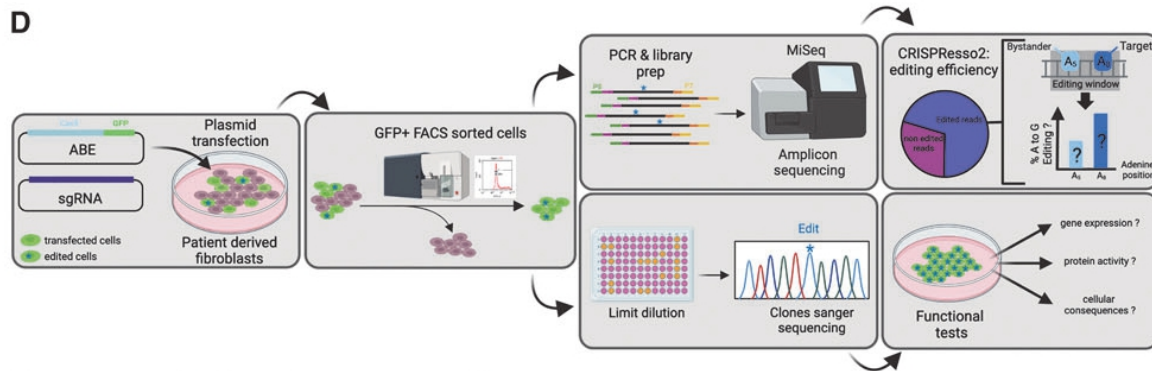
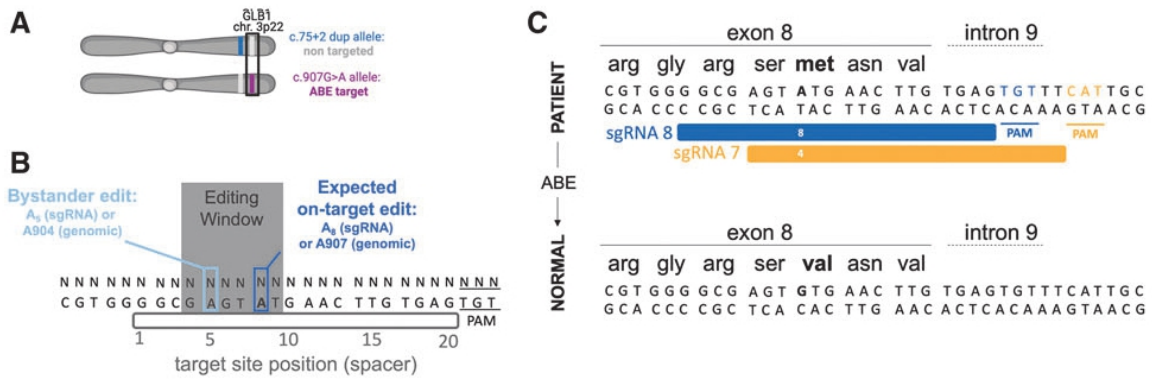
Next, we performed a clonal selection to isolate edited G907 clones (on-target edit). β -Gal activity and β -gal protein levels were assessed (Fig. 3F). For the six clones exposed to the sgRNA8, ABE rescued a therapeutic level of β -gal protein and β -gal activity when compared to donor cells (wt *GLB1*). The restored β -gal activity reached a similar activity to that detected in the fibroblasts isolated from his asymptomatic parents (50% of the normal values) (Fig. 2E).

Window editing of ABE

Closer inspection of our sequencing results indicated that A904 (position A₅ from the 5' sgRNA extremity) is also converted into G904 (~31% of the reads) (Figs. 3F and 4A, B). This result could be explained by the active window of the ABEmax (7.10), which extends from nucleotides N₄ to N₉ (Fig. 3B).¹⁹ In parallel, we did not detect any significant editing event outside this editing window by NGS (raw data Supplementary Table S4).

Since the clone#20 (G907-G904; on-target and bystander edits) displayed a similar β -gal activity than

FIG. 3. Adenine base editing strategy for GM1 gangliosidosis. **(A)** Maternal inherited *GLB1* variant is targetable by ABE, **(B)** scheme explaining the base editor window. Potential bystander edit might be obtained (proximal adenine to the on-target adenine located in the base editor window), **(C)** two sgRNA were designed to correct the pathogenic ATG (Met) into the GTG (Val) on exon 8 of *GLB1* gene using ABE, **(D)** workflow used to evaluate the ABE efficiency and specificity for the patient-derived fibroblasts, **(E)** efficiency of ABE to introduce the on-target edit (A₈→G₈ or 907A→G) using the sgRNA8 in fibroblasts. Results were obtained by amplicon deep sequencing and analyzed by CRISPResso2. Bystander edits were identified. **(F)** Clones isolated after transfection sgRNA8+ABE or sgRNA CTR+ABE were compared to normal fibroblasts or untreated patient-derived fibroblasts. Each clone was characterized for β -galactosidase activity, β -galactosidase protein, and edits. $n=3$ biologically independent experiments, each histogram represents the mean \pm SD. α -Tubulin serves as a loading control. Data were subjected to Student's t -test and p -values were deemed to be statistically significant as follows: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. Raw data are available in Supplementary Table S5. MUT, mutated; WT, wild type.



other edited clones (G907-A904; only on-target edit), we investigated the neutral role of this unwanted edit.

We predicted the 3D structure of GLB1 protein (Supplementary Fig. S3). *In silico* mutation followed by an energy minimization of the whole structure showed minimal impact on the surrounding residues and GLB1 structure ($\text{RMSD}_{\text{all-atoms}} = 0.720 \text{ \AA}$). To study the potential impact of this mutation more in depth, MD simulations were carried out ($3 \times 100 \text{ ns}$) on GLB1 WT and S302G (Supplementary Fig. S3A). Analysis of trajectories did not reveal any significant change in the secondary or tertiary structure of GLB1 (Supplementary Fig. S3B). Altogether, our results strongly suggest that this unwanted on-target event (bystander edit) might have no impact on GM1 gangliosidosis patients.

ABE: off-target detection

Base editors are known to generate bystander edits on surrounding A or C (window editing) (Figs. 3 and 4A). They can have negligible consequence as demonstrated in this study or be avoided by targeting a region without modifiable surrounding bases. However, the off-target edits due to an imperfect sgRNA base pairing to DNA might impair the clinical development of an sgRNA coupled to a base editor.

To uncover genomic loci with the potential for off-target editing, a computational prediction was used (CRISPOR) (Fig. 4A).³⁵ In total, 978 candidate sites were predicted, including only 63 exonic sites (Supplementary Table S3). Since it remains difficult to evaluate all these potential sites, we ranked them according to the CFD score as previously published.⁶ We focused on the top 10 CFD scores (only one is located in an exon; *IDH1* gene, Supplementary Table S3) and off-target effects were investigated by deep sequencing (amplicon-seq). Amplicons were analyzed by CRISPResso2 as previously done for BE efficiency measurement (Fig. 3E).

These samples were stringently quality filtered with a flag for minimum average read quality of 30 (phred33 score=30) to ensure SNP calling was only performed on high-quality reads. In our conditions, no off-target edit was found for the sgRNA8 in patient-derived fibroblasts. We failed to amplify the OT2 region by PCR. Thus, 9 among the top 10 CFD score sites were investi-

gated (Fig. 4C). To reinforce this result, we performed WES (clone #58 and #20 vs. sgRNA control-transfected cells). On the 63 exonic predicted off-target sites, 32 sites were covered by WES. Under our conditions, no off-target edit was found for the sgRNA8 in patient-derived fibroblasts (0/32). In total, 41 potential off-target sites, including those with the highest CFD scores, were analyzed. We did not find any off-target effect using this protocol.

In addition, we quantified indel events around sgRNA from HTS (amplicons) data followed by CRISPResso2 analysis (Supplementary Fig. S4). On average, indel events remain very low ($\approx 0.1\%$) compared to on-target editing, which is in accordance with previous reports (Supplementary Fig. S4A, B).¹⁵ Indel size varies approximately from 1 to 20 bases (Supplementary Fig. S4C). Altogether, our results (*in vitro*) strongly suggest that BE might be useful to cure GM1 gangliosidosis.

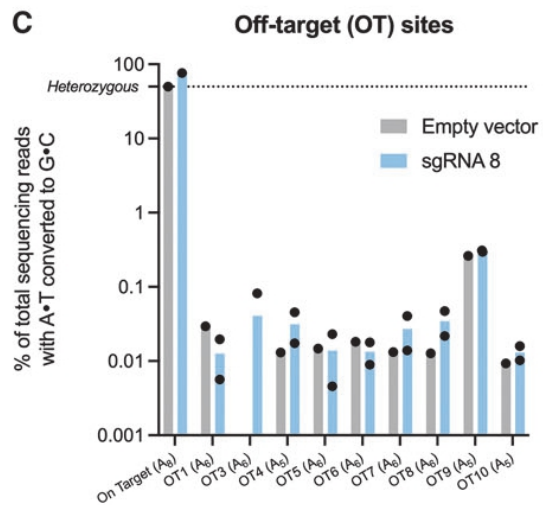
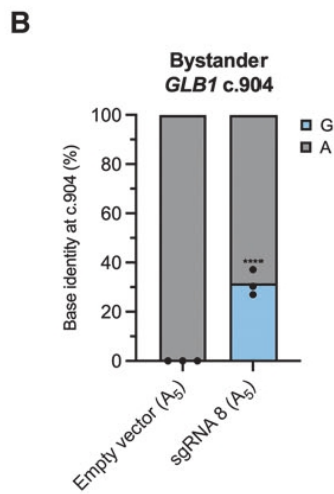
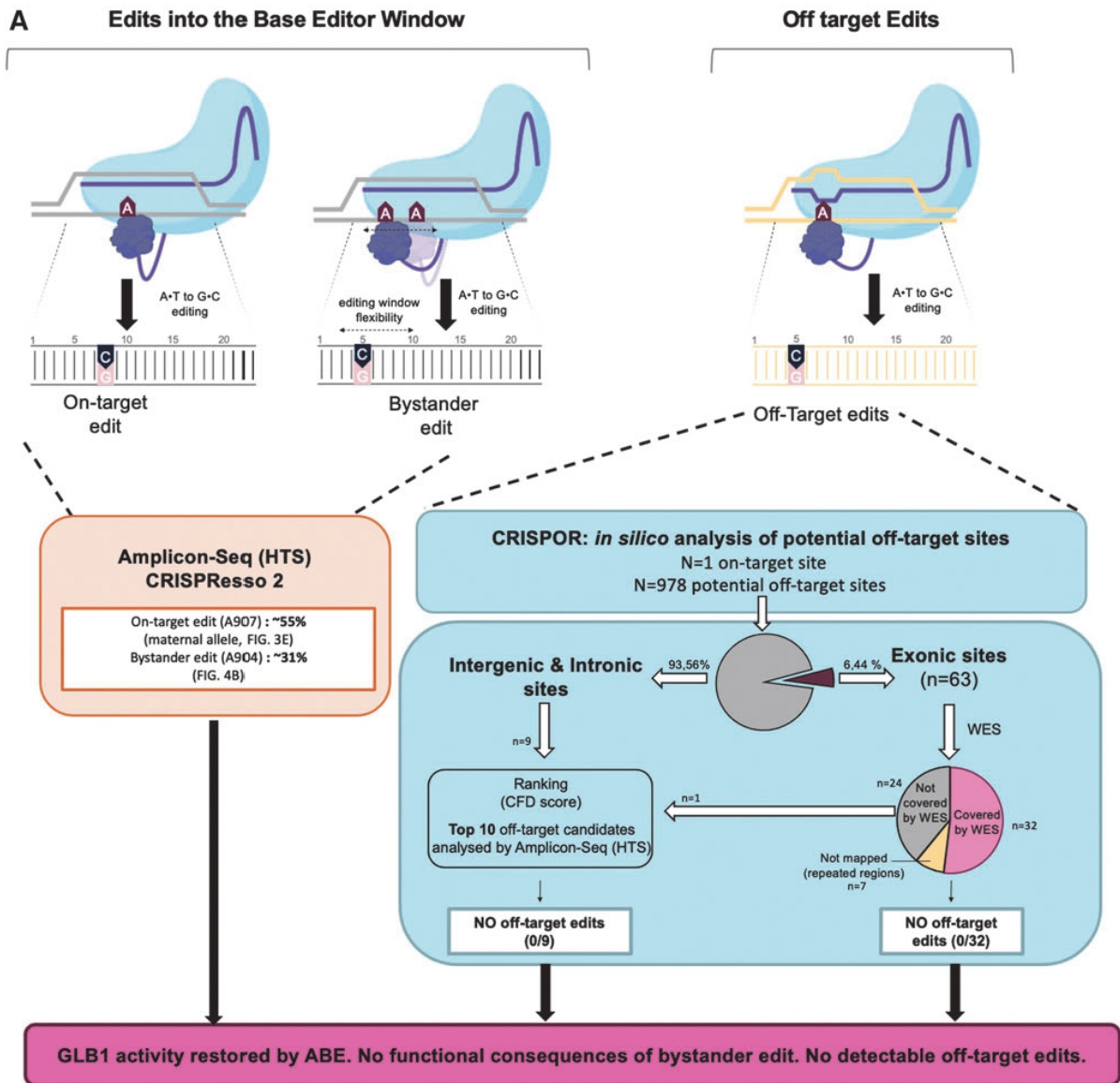
Discussion

Currently, no curative treatment for GM1 gangliosidosis is available. In theory, therapies should slow or stop the progression of this disease.^{21,24} We suspect that the earlier the patient is managed, the greater the therapeutic success will be. The three ongoing clinical trials aim at bringing a healthy *GLB1* coding sequence into the cells of the brain and spinal cord (AAV vector injections into the cisterna magna). To date, it is not possible to evaluate the number of nervous system cells that will be targeted by this classical gene therapy approach (efficiency). Patient follow-up is necessary to determine disease progression and potential adverse effects.

In this study, we provide an *in vitro* proof-of-concept study, strongly suggesting that genome editing can be an alternative strategy for this deleterious disease. To vehiculate the sgRNA and the ABE, the same AAV vector used for ongoing clinical trials^{24,25} might be used for a single exposure. To obtain an optimal effect, the injection (genome editing) should be applied as soon as possible after birth, as already done for another neuronal disease using antisense oligonucleotides.⁴³

However, viral delivery of base editors allows sustained expression in transduced cells, which may increase the frequency of off-target editing.^{44,45} In addition, using

FIG. 4. Specificity assessment of ABE for *GLB1* therapy. **(A)** Scheme recapitulating our workflow to identify the on-target and bystander edits on the *GLB1* gene (base editor window), **(B)** workflow explaining the off-target effects of ABE in patient-derived fibroblasts. Potential off-target edits were predicted by CRISPOR and 41 potential sites were examined by combining whole exome sequencing and amplicon sequencing results. CFD score.^{35,36} **(C)** Analysis of the top 10 off-target predicted sites for sgRNA8+ABE. Data were subjected to Student's *t*-test and *p*-values were deemed to be statistically significant as follows: *****p* < 0.0001. Raw data are available in Supplementary Table S5. CFD, cutting frequency determination.



viral vectors in gene therapy raises the possibility of rare vector integration into the genome of patient's cells, which may promote oncogenesis.^{3,8,46} Recently, another strategy has been published with Virus-like particles (VLPs) efficiently “infecting cells” without carrying viral genetic material. So the safety and efficiency of VLPs render them promising for gene editing approaches, including *in vivo*.^{7,47}

Limit of the study

To treat GM1 gangliosidosis, a classical gene restoration strategy (rescue) is promising due to the lack of hotspot mutations for this lysosome storage disease (Fig. 1), and currently in clinical trials. In other words, the same vector might be useful for all these patients. This rescue approach has been successful in mice and cats using AAV vectors.^{21,24,26–29} It is tempting to postulate that the survival of GM1 patients treated by gene restoration might be long. Thus, the potential drawback of this approach might be a possible decrease of β -gal protein expression level in long term because the AAV vectors are not conceived to integrate the *GLB1* sequence into the patient genome. This limitation has also been discussed in a recent publication.²⁶

On the other hand, gene editing is an alternative strategy to cure GM1 gangliosidosis, by correcting the root cause of disease and probably avoiding repetitive adeno-associated virus injections. However, the main limitation of gene editing to treat such pathologies without hotspots is the need to design and evaluate the efficiency and specificity of a sgRNA per patient. Despite these hurdles, efforts to develop more therapeutics to treat rare and ultra-rare diseases are currently increasing in nonprofit biotechnology companies and academic laboratories.

In August 2022, Cure Rare Disease achieved FDA approval for the first-in-human IND to treat a rare mutation causing Duchenne muscular dystrophy using a tailor-made CRISPR approach. The drug, named CRD-TMH-001, treats promoter and exon 1 mutations on the dystrophin gene in muscle. The therapeutic will upregulate an alternate form (isoform) of the dystrophin protein using CRISPR transactivator with the goal of stabilizing, or potentially reversing, symptom progression of Duchenne muscular dystrophy. The dosing of the drug will occur imminently at University of Massachusetts Chan Medical School (one-time administration).

Nowadays, traditional gene addition therapy seems more adapted for these kinds of diseases. However, the situation is changing rapidly. To date, gene editing is well suited to correct autosomal dominant or dominant-negative pathogenic variants^{26,48,49} and diseases associated with frequent variants. Three BE treatments are

starting a clinical evaluation in humans: VERVE-101 (NCT05398029, Verve Therapeutics) for heterozygous familial hypercholesterolemia (*PCSK9* gene), BEAM-101 (NCT05456880, Beam Therapeutics) for Sickle Cell Disease (*HBB* gene), and BE CAR-7 (base-edited T cells, ISRCTN15323014, Medical Research Council United Kingdom) for T cell leukemia.

Off-target edits might restrain the utility of gene editing for human therapy. In this study, by combining WES and amplicon sequencing for the 10 potential off-target sites ranked by CRISPOR, we did not identify off-target edits in our experimental conditions (0/41 sites). It is important to note that off-target editing independent of the DNA-sgRNA pairing has not been investigated in this study. For a deeper DNA off-target identification, unbiased genome wide, empirical approaches (either biochemical or cell culture based) should be more relevant as some sites may be missed by *in silico* off-target prediction programs.^{50,51}

In addition, whole genome sequencing and RNA editing by base editor must be analyzed for a clinical application, according to the recent FDA press release (Human Gene Therapy Products Incorporating Human Genome Editing). In this study, we performed a proof-of-concept study using patient-derived fibroblasts. Additional experiments are needed to replicate these results in cells derived from additional patients and/or new GM1 gangliosidosis models. Indeed, several GM1 gangliosidosis animal models are available, but the *GLB1* genetic modifications carried by these animals are not compatible for a correction by BE.^{24,26–29} In the future, a GM1 animal model carrying such compatible mutation might be developed to evaluate *in vivo* the gene editing efficiency/feasibility as already done for progeria.⁸

Conclusion

GM1 gangliosidosis is a rare autosomal recessive disorder estimated to occur in 1 in 100,000 to 200,000 newborns. In this study, we showed that 56% of the pathogenic mutations in *GLB1* gene might be targeted by the BE (ABE+CBE) and 100% by the PE. Importantly, only one pathogenic allele needs to be corrected to cure the disease. Moreover, we demonstrated the efficiency and the safety of ABE in patient-derived fibroblasts in accordance with other BE studies.^{8,18} In conclusion, our study strongly suggests that gene editing might be an alternative strategy to cure GM1 gangliosidosis immediately after birth to limit irreversible damage.

Acknowledgments

First, the authors would like to thank the family (parents and son) for their kind support.

We would like to thank their collaborators for helpful advices, protocols, and vectors: Pr. C. Bendavid, Dr. C. Moreau, Pr. D.R. Liu (vectors), Pr. B. Kleinstiver [pCMV-T7-ABEmax(7.10)-SpRY-P2A-EGFP (RTW5025)], the Biosit platforms, especially Dr. P. Gripon for BSL3 and L. Deleume and A. Aimé for FACS analysis, the CRISP'edit platform (Bordeaux), especially Dr. B. Turcq and Dr. V. Prouzet-Mauleon, the head of the Inserm research unit Dr. E. Chevet, the technicians from the Cell Biology service CHU Rennes, and Pr. E. Tucker and Dr. A. Matvere for critical reading of this article.

Authors' Contributions

D.L.: conceptualization, data curation, formal analysis, investigation, visualization, validation, writing—original draft, and writing—review and editing.

L.G.: data curation, investigation, and writing—review and editing.

S.J.: data curation, formal analysis, investigation, project administration, and writing—review and editing.

B.N.: data curation, formal analysis, validation, software, and writing—review and editing.

L.C.: investigation and writing—review and editing.

L.D.: resources and writing—review and editing.

C.D.: data curation, investigation, resources, and writing—review and editing.

A.E.: data curation, investigation, and writing—review and editing.

T.L.: investigation, resources, and writing—review and editing.

R.F.: investigation, resources, writing—review and editing.

S.D.: data curation, investigation, and writing—review and editing.

X.G.: investigation and writing—review and editing.

L.A.E.: investigation and writing—review and editing.

E.L.: data curation, investigation, and writing—review and editing.

F.M.: project administration and writing—review and editing.

M.-A.B.-R.: project administration and writing—review and editing.

S.O.: resources and project administration.

D.G.: conceptualization, data curation, formal analysis, investigation, visualization, validation, writing—original draft, writing—review and editing, and project administration.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

The authors would like to thank their sponsors: ANR ANR-21-CE18-0020, Cancéropôle Grand-Ouest, AVIESAN Plan Cancer, Ministère de l'enseignement Supérieur, de la Recherche et de l'Innovation (PhD fellowship), Grants to LAE from the Swedish Research Council (VR; Grant No. 2019-3684) and the Swedish Cancer Foundation (Grant No. 21-1447-Pj), and allocation of computing time provided by the Swedish National Infrastructure for Computing at supercomputing center C3SE and NSC, in part, funded by the Swedish Research Council through grant agreement no. 2018-05973. XG is grateful to the Fondation ARC pour la recherche sur le cancer for his postdoctoral fellowship (PDF20191209830).

Supplementary Material

Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Table S4
Supplementary Table S5

References

- Saha K, Sontheimer EJ, Brooks PJ, et al. The NIH somatic cell genome editing program. *Nature* 2021;592(7853):195–204; doi: 10.1038/s41586-021-03191-1
- Doudna JA. The promise and challenge of therapeutic genome editing. *Nature* 2020;578(7794):229–236; doi: 10.1038/s41586-020-1978-5
- Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* 2020;38(7):824–844; doi: 10.1038/s41587-020-0561-9
- Newby GA, Liu DR. In vivo somatic cell base editing and prime editing. *Mol Ther* 2021;29(11):3107–3124; doi: 10.1016/j.ymthe.2021.09.002
- Newby GA, Yen JS, Woodard KJ, et al. Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature* 2021;595(7866):295–302; doi: 10.1038/s41586-021-03609-w
- Krishnamurthy S, Traore S, Cooney AL, et al. Functional correction of CFTR mutations in human airway epithelial cells using adenine base editors. *Nucleic Acids Res* 2021;49(18):10558–10572; doi: 10.1093/nar/gkab788
- Banskota S, Raguram A, Suh S, et al. Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. *Cell* 2022;185(2):250–265.e16; doi: 10.1016/j.cell.2021.12.021
- Koblan LW, Erdos MR, Wilson C, et al. In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice. *Nature* 2021;589(7843):608–614; doi: 10.1038/s41586-020-03086-7
- Musunuru K, Chadwick AC, Mizoguchi T, et al. In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* 2021;593(7859):429–434; doi: 10.1038/s41586-021-03534-y
- Kleinstiver BP, Pattanayak V, Prew MS, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 2016;529(7587):490–495; doi: 10.1038/nature16526
- Boutin J, Cappellen D, Rosier J, et al. ON-target adverse events of CRISPR-Cas9 nuclease: More chaotic than expected. *CRISPR J* 2022;5(1):19–30; doi: 10.1089/crispr.2021.0120
- Gaudelli NM, Lam DK, Rees HA, et al. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat Biotechnol* 2020;38(7):892–900; doi: 10.1038/s41587-020-0491-6
- Rees HA, Minella AC, Burnett CA, et al. CRISPR-derived genome editing therapies: Progress from bench to bedside. *Mol Ther* 2021;29(11):3125–3139; doi: 10.1016/j.ymthe.2021.09.027

14. Komor AC, Kim YB, Packer MS, et al. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 2016;533(7603):420–424; doi: 10.1038/nature17946
15. Gaudelli NM, Komor AC, Rees HA, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 2017;551(7681):464–471; doi: 10.1038/nature24644
16. Tan J, Zhang F, Karcher D, et al. Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nat Commun* 2019;10(1):439; doi: 10.1038/s41467-018-08034-8
17. Jia K, Cui Y, Huang S, et al. Phage peptides mediate precision base editing with focused targeting window. *Nat Commun* 2022;13(1):1662; doi: 10.1038/s41467-022-29365-7
18. Rees HA, Komor AC, Yeh W-H, et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat Commun* 2017;8:15790; doi: 10.1038/ncomms15790
19. Walton RT, Christie KA, Whittaker MN, et al. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 2020;368(6488):290–296; doi: 10.1126/science.aba8853
20. Huang TP, Newby GA, Liu DR. Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat Protoc* 2021;16(2):1089–1128; doi: 10.1038/s41596-020-00450-9
21. Nicoli E-R, Annunziata I, d'Azzo A, et al. GM1 gangliosidosis—A mini-review. *Front Genet* 2021;12:734878; doi: 10.3389/fgene.2021.734878
22. Lang FM, Korner P, Harnett M, et al. The natural history of type 1 infantile GM1 gangliosidosis: A literature-based meta-analysis. *Mol Genet Metab* 2020;129(3):228–235; doi: 10.1016/j.ymgme.2019.12.012
23. McCurdy VJ, Johnson AK, Gray-Edwards HL, et al. Sustained normalization of neurological disease after intracranial gene therapy in a feline model. *Sci Transl Med* 2014;6(231):231ra48; doi: 10.1126/scitranslmed.3007733
24. Hinderer C, Nosratabkhsh B, Katz N, et al. A single injection of an optimized adeno-associated viral vector into cerebrospinal fluid corrects neurological disease in a murine model of GM1 gangliosidosis. *Hum Gene Ther* 2020;31(21–22):1169–1177; doi: 10.1089/hum.2018.206
25. Levy JM, Yeh W-H, Pendse N, et al. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat Biomed Eng* 2020;4(1):97–110; doi: 10.1038/s41551-019-0501-5
26. Gross AL, Gray-Edwards HL, Bebout CN, et al. Intravenous delivery of adeno-associated viral gene therapy in feline GM1 gangliosidosis. *Brain* 2022;145(2):655–669; doi: 10.1093/brain/awab309
27. Baek RC, Boekman MLD, Leroy SG, et al. AAV-mediated gene delivery in adult GM1-gangliosidosis mice corrects lysosomal storage in CNS and improves survival. *PLoS One* 2010;5(10):e13468; doi: 10.1371/journal.pone.0013468
28. Weismann CM, Ferreira J, Keeler AM, et al. Systemic AAV9 gene transfer in adult GM1 gangliosidosis mice reduces lysosomal storage in CNS and extends lifespan. *Hum Mol Genet* 2015;24(15):4353–4364; doi: 10.1093/hmg/ddv168
29. Broekman MLD, Baek RC, Comer LA, et al. Complete correction of enzymatic deficiency and neurochemistry in the GM1-gangliosidosis mouse brain by neonatal adeno-associated virus-mediated gene delivery. *Mol Ther* 2007;15(1):30–37; doi: 10.1038/sj.mt.6300004
30. Buchlis G, Podsakoff GM, Radu A, et al. Factor IX expression in skeletal muscle of a severe hemophilia B patient 10 years after AAV-mediated gene transfer. *Blood* 2012;119(13):3038–3041; doi: 10.1182/blood-2011-09-382317
31. Muhuri M, Levy DI, Schulz M, et al. Durability of transgene expression after RAAV gene therapy. *Mol Ther* 2022;30(4):1364–1380; doi: 10.1016/j.jymthe.2022.03.004
32. Mouden C, Dubourg C, Carré W, et al. Complex mode of inheritance in holoprosencephaly revealed by whole exome sequencing. *Clin Genet* 2016;89(6):659–668; doi: 10.1111/cge.12722
33. Ho MW, O'Brien JS. Hurler's syndrome: Deficiency of a specific beta galactosidase isoenzyme. *Science* 1969;165(3893):611–613; doi: 10.1126/science.165.3893.611
34. Kleinstiver BP, Prew MS, Tsai SQ, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 2015;523(7561):481–485; doi: 10.1038/nature14592
35. Haeussler M, Schönig K, Eckert H, 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):148; doi: 10.1186/s13059-016-1012-2
36. Doench JG, Fusi N, Sullender M, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 2016;34(2):184–191; doi: 10.1038/nbt.3437
37. Clement K, Rees H, Canver MC, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* 2019;37(3):224–226; doi: 10.1038/s41587-019-0032-3
38. Bowers KJ, Chow E, Xu H, et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06; Association for Computing Machinery: New York, NY, United States: 2006; doi: 10.1109/SC.2006.54
39. Lu C, Wu C, Ghoreishi D, et al. OPLS4: Improving force field accuracy on challenging regimes of chemical space. *J Chem Theory Comput* 2021;17(7):4291–4300; doi: 10.1021/acs.jctc.1c00302
40. Jorgensen WL, Chandrasekhar J, Madura JD, et al. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79(2):445869; doi: 10.1063/1.445869
41. Morrone A, Morreau H, Zhou XY, et al. Insertion of a T next to the donor splice site of intron 1 causes aberrantly spliced mRNA in a case of infantile GM1-gangliosidosis. *Hum Mutat* 1994;3(2):112–120; doi: 10.1002/humu.1380030205
42. Anzalone AV, Randolph PB, Davis JR, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019;576(7785):149–157; doi: 10.1038/s41586-019-1711-4
43. Finkel RS, Chiriboga CA, Vajsaar J, et al. Treatment of infantile-onset spinal muscular atrophy with nusinersen: A phase 2, open-label, dose-escalation study. *Lancet* 2016;388(10063):3017–3026; doi: 10.1016/S0140-6736(16)31408-8
44. Akcakaya P, Bobbin ML, Guo JA, et al. In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature* 2018;561(7723):416–419; doi: 10.1038/s41586-018-0500-9
45. Yeh W-H, Chiang H, Rees HA, et al. In vivo base editing of post-mitotic sensory cells. *Nat Commun* 2018;9(1):2184; doi: 10.1038/s41467-018-04580-3
46. Chandler RJ, Sands MS, Venditti CP. Recombinant adeno-associated viral integration and genotoxicity: Insights from animal models. *Hum Gene Ther* 2017;28(4):314–322; doi: 10.1089/hum.2017.009
47. Mangeot PE, Risson V, Fusil F, et al. Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat Commun* 2019;10(1):45; doi: 10.1038/s41467-018-07845-z
48. Diakatou M, Manes G, Bocquet B, et al. Genome editing as a treatment for the most prevalent causative genes of autosomal dominant retinitis pigmentosa. *Int J Mol Sci* 2019;20(10):2542; doi: 10.3390/ijms20102542
49. Bakondi B, Lv W, Lu B, et al. In vivo CRISPR/Cas9 gene editing corrects retinal dystrophy in the S334ter-3 rat model of autosomal dominant retinitis pigmentosa. *Mol Ther* 2016;24(3):556–563; doi: 10.1038/mt.2015.220
50. Kim D, Kang BC, Kim JS. Identifying genome-wide off-target sites of CRISPR RNA-guided nucleases and deaminases with Digenome-seq. *Nat Protoc* 2021;16(2):1170–1192; doi: 10.1038/s41596-020-00453-6
51. Lei Z, Meng H, Lv Z, et al. Detect-seq reveals out-of-protospacer editing and target-strand editing by cytosine base editors. *Nat Methods* 2021;18(6):643–651; doi: 10.1038/s41592-021-01172-w

Received: April 19, 2022

Accepted: November 19, 2022

Online Publication Date: January 9, 2023

SUPPLEMENTARY FIGURES

NB: table S1 to S5 are available online
https://www.liebertpub.com/doi/10.1089/crispr.2022.0045?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed

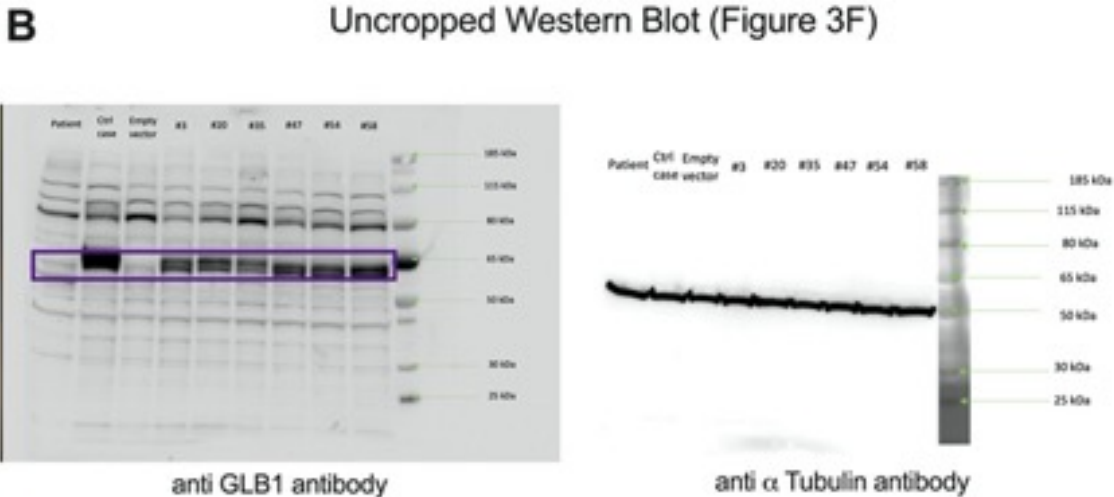
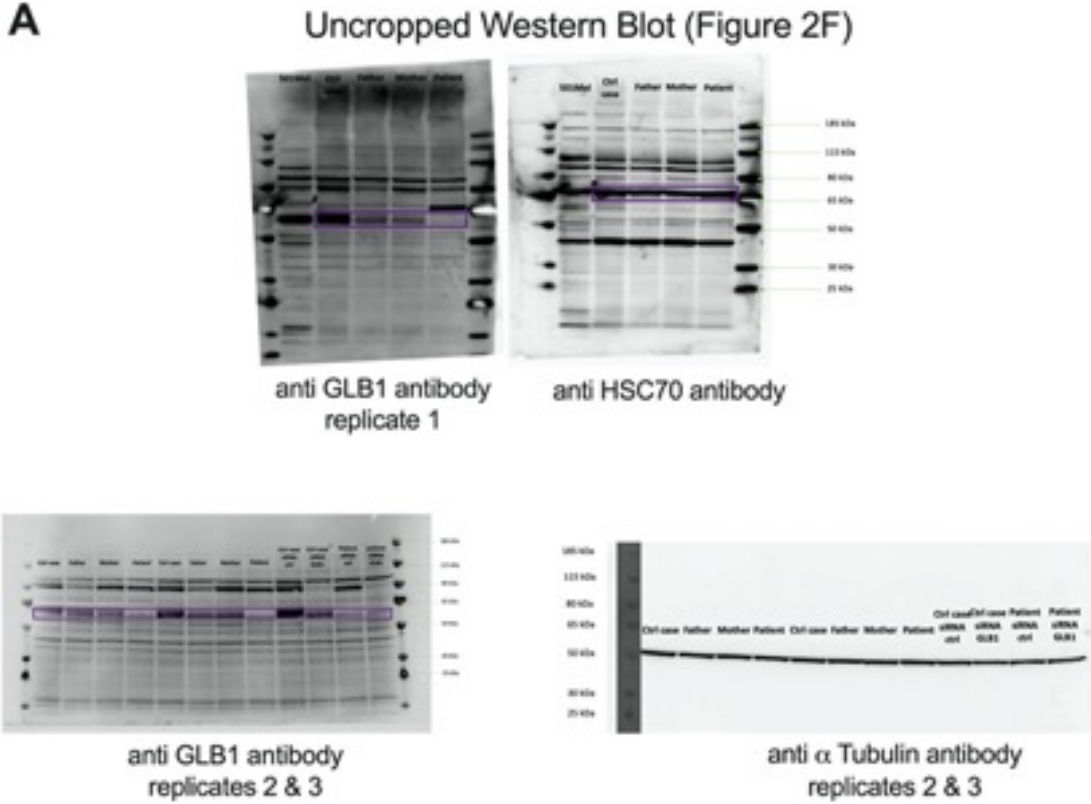


Figure S1: Uncropped western-blots

```

#!/usr/bin/python3
# -*- coding: utf-8 -*-

import argparse
import pandas as pd
from pandas import DataFrame

if __name__ == "__main__":

    # Arguments :
    parser=argparse.ArgumentParser()
    parser.add_argument('--input', help='Input file')
    parser.add_argument('--output', help='Output file')
    args=parser.parse_args()

    # Inputs :
    input_file = args.input
    output_file = args.output

    # Import file into Dataframe
    data = pd.read_csv(input_file, sep="\t")
    df = pd.DataFrame(data)

    # Remove the row begining with 'N'
    df = df.loc[df['Unnamed: 0'] != 'N']

    # Variable 'col' will contain the 1st column with the nucleotides (will be
used for the output file)
    col = df.iloc[:,[0]]

    # Remove the 1st column from the dataframe (used for percentage
calculation)
    del df['Unnamed: 0']

    # Calculation
    pourcentages = df/df.sum()*100

    # Concatenation of the 1st column (nucleotides names) with the dataframe
containing the percentages
    final = pd.concat([col, pourcentages], axis=1)
    print(final)

    # Export to the output file

    export_csv = final.to_csv (output_file, index = None, header=True, encoding=
'utf-8', sep=';')

```

Figure S2: Script (Python) for on-target and off-target analysis

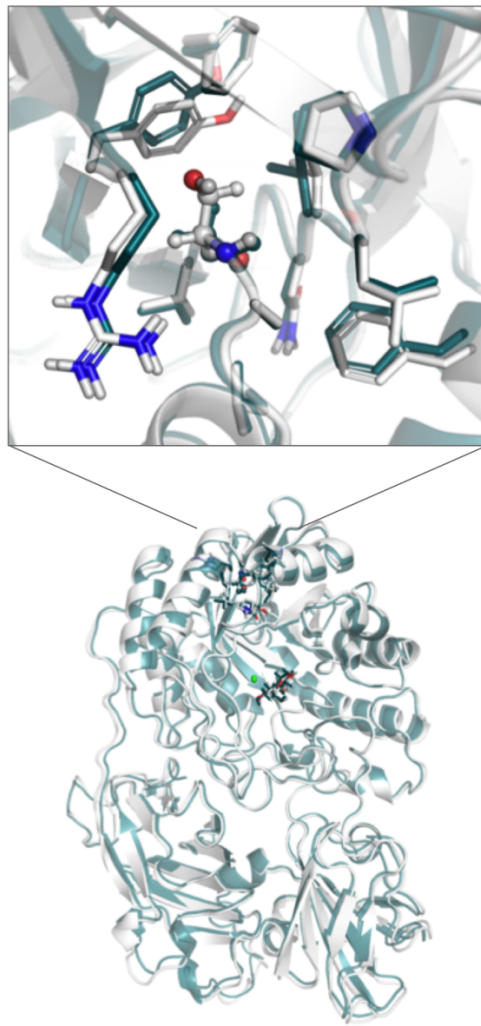
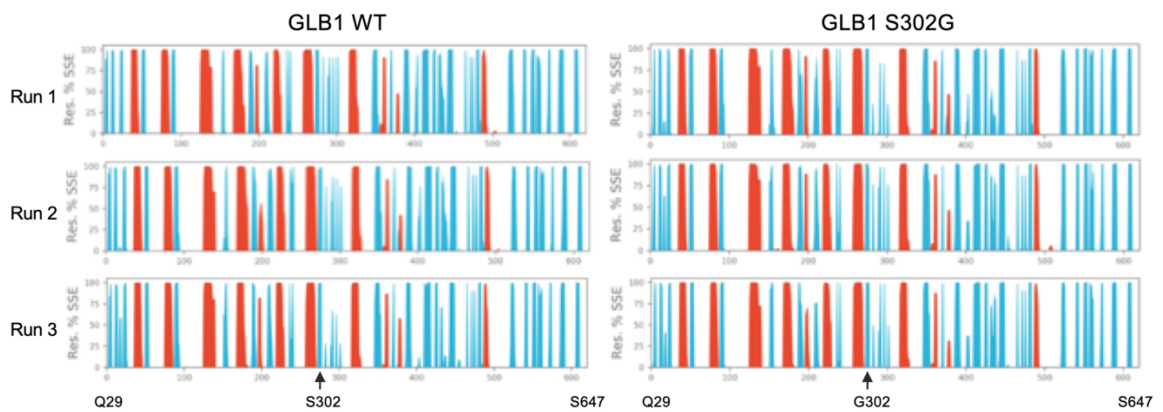
A**B**

Figure S3: ABE bystander edit and 3D structure predictions

(A) molecular dynamics (MD) simulations were carried out (3 x 100 ns) on GLB1 wild type and S302G. (B) Analysis of the trajectories (secondary of tertiary structure of GLB1).

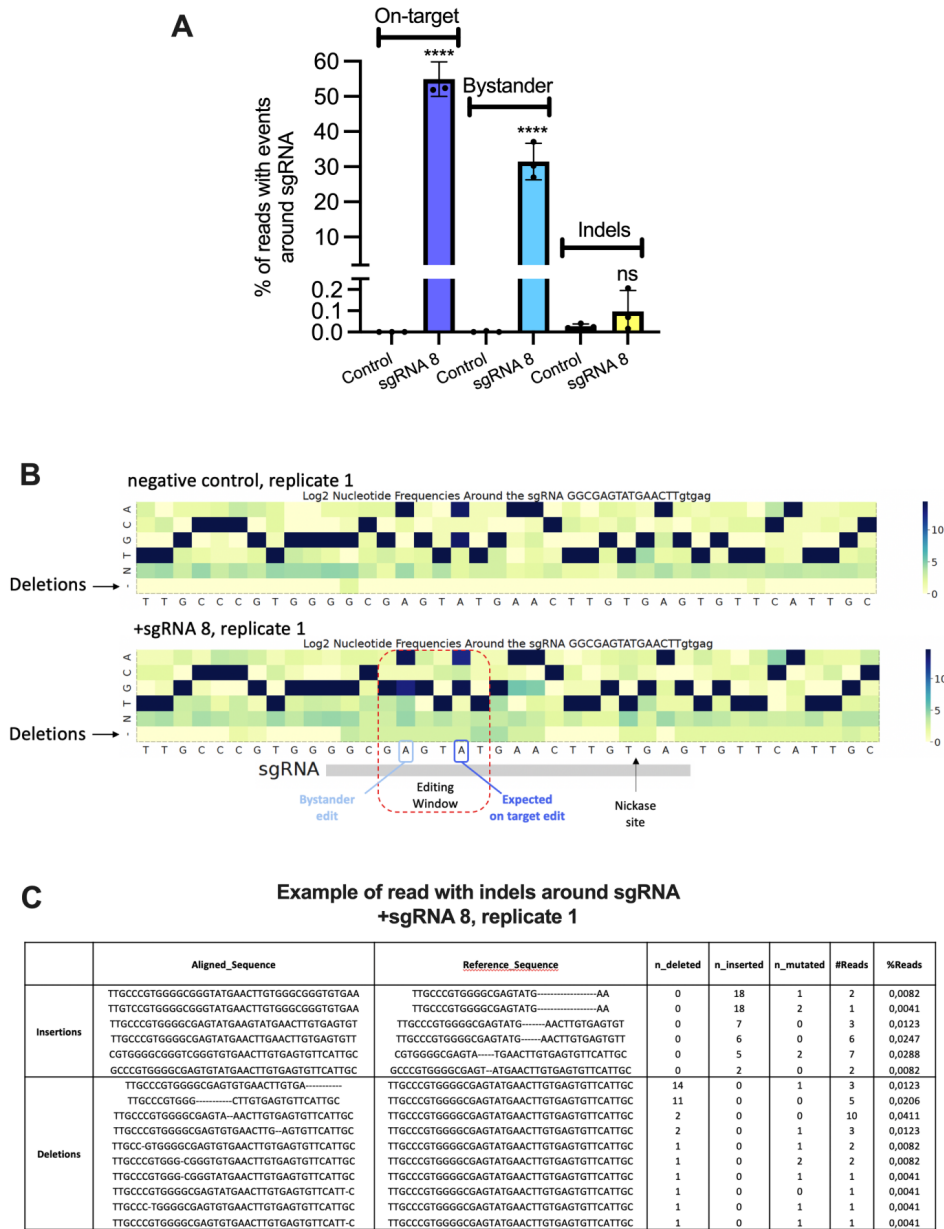


Figure S4: Indels analyses for sgRNA8-ABE strategy

Supplementary Figure 4: Indels analyses for sgRNA8-ABE strategy. (A) Base editing efficiency in patient-derived fibroblasts transfected with sgRNA 8. Results were obtained by amplicon deep-sequencing and analyzed by CRISPResso2, n=3 biologically independent experiments, each histogram represents the mean \pm s.d. Data were subjected to Student's t-test and p-values were deemed to be statistically significant as follows: *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001. (B) Log2 nucleotide frequencies arounds sgRNA 8 and deletions localization. This graph was generated by CRISPResso2 from HTS data. (C) Example of reads with indels around sgRNA. This table was adapted from results generated by CRISPResso2 from HTS data. Raw data are available in Table S5.

CONCLUSION DE L'ARTICLE

Cette preuve de concept réalisée *in vitro* démontre que la simple conversion d'un A en G par BE permet de restaurer une activité thérapeutique de l'enzyme bêta-galactosidase, responsable de la gangliosidose à GM1. Ce variant maternel, initialement décrit comme probablement pathogène, a pu être reclassé en variant de classe 5 (pathogène). A ce jour, le BE demeure, à mon avis, l'outil d'édition du génome le plus efficace et spécifique. Cette technologie très prometteuse sera sans doute largement utilisée en clinique dans un avenir très proche.

Néanmoins, des travaux complémentaires sont nécessaires pour atteindre la phase clinique. **Premièrement**, n'avons pas eu l'opportunité de tester cet outil dans des modèles animaux, car ceux disponibles à ma connaissance ne sont pas porteurs d'altération génétique corrigable par BE. **Deuxièmement**, il est indispensable de délivrer les outils d'édition du génome aux patients sous forme de complexe RNP ou ARN (expression transitoire) pour éviter les réactions immunitaires contre la Cas9 et surtout limiter l'activité hors-cible. Malheureusement, il n'existe pas encore de vecteur « sans danger » respectant ces critères et capables d'atteindre efficacement les cellules neuronales. Les particules lipidiques renfermant l'ARNm BE pourraient *en théorie* être reformulées pour cibler les motoneurones, mais une injection systémique risque d'être insuffisante pour atteindre le système nerveux central³⁸¹. A ma connaissance, l'injection intrathécale de particules lipidiques n'a pas encore été testée chez l'humain³⁸². En revanche, les vecteurs AAV sont efficaces pour transduire les motoneurones après une injection intrathécale (essais cliniques en cours). Des tests pré-cliniques effectués sur six chats atteints de gangliosidose à GM1 ont démontrés que l'administration d'un vecteur AAV9 codant pour la bêta-galactosidase permettait d'améliorer l'espérance de vie des animaux d'un facteur 5 (8 mois versus 3,5 ans), leurs facultés neurologiques et leur qualité de vie³⁸³. Ces résultats suggèrent qu'il est possible de ralentir les effets de la gangliosidose à GM1, chez ces animaux. Les premiers dosages effectués chez les patients inclus dans les 3 essais cliniques en cours (thérapie génique additive) indiquent une amélioration de l'activité bêta-galactosidase et une réduction de l'accumulation de gangliosides à GM1 dans le liquide cébrospinal. L'étude de leurs signes cliniques, sur le long terme, permettrait de mieux apprécier l'efficacité du traitement.

L'avantage de la thérapie génique additive, comparée à l'édition du génome, est que le même traitement peut être administré à tous les patients, indépendamment du type et de la localisation du variant dans le gène atteint. Dans cette étude, nous avons évalué l'outil BE pour corriger un seul variant génétique, alors que 275 variants délétères ont été rapportés pour la gangliosidose à GM1 dans la base de données ClinVar. Toutefois, la plupart du temps, les vecteurs AAV ne s'intègrent pas au génome de l'hôte et persistent dans le noyau cellulaire sous forme épisomale. Nous ne connaissons pas encore la durée de vie de cet ADN épisomal sur le long terme, dans les cellules post-mitotiques neuronales. Contrairement à la thérapie génique additive, l'avantage du BE est qu'une seule infection pourrait suffire pour obtenir des effets stables sur le long terme (modification génétique irréversible – intégrée au génome de l'hôte).

RESULTAT 2

Évaluation fonctionnelle d'un variant homozygote du gène de l'*AMH* identifié chez une patiente souffrant d'insuffisance ovarienne débutante

AVANT PROPOS

Ce projet a été réalisé dans le cadre de la mise en place de l'activité EPIQUE dans le laboratoire de cytogénétique et biologie cellulaire du CHU de Rennes, dirigé par le Pr Marc-Antoine Belaud-Rotureau. Le manuscrit a été soumis en janvier 2023 dans la revue *Journal of Assisted Reproduction and Genetics*. Afin de mieux caractériser les conséquences du variant génétique *AMH* p.(Val12Gly), des expériences complémentaires sont en cours.

INTRODUCTION DE L'ARTICLE

Le Pr Sylvie Jaillard, biologiste au sein du laboratoire de cytogénétique et biologie cellulaire du CHU de Rennes, utilise en routine la technologie de séquençage d'exome afin d'investiguer les causes génétiques de l'insuffisance ovarienne. Un variant de signification inconnue, localisé dans le gène de l'*AMH* chez une patiente présentant une diminution prématurée de sa réserve ovarienne, a retenu son attention. En effet, chez la femme, l'*AMH* est une hormone clé sécrétée par les cellules de la granulosa et agit comme le gardien de la réserve ovarienne en limitant l'entrée en maturation des follicules primordiaux. Des expériences de validations fonctionnelles ont été menées dans l'objectif de reclassifier ce variant (délétère ou bénin). Le variant génétique a été reproduit par la technologie *prime editing* dans un modèle cellulaire, puis l'expression de l'*AMH* a brièvement été évaluée par Western Blot. Les données préliminaires suggèrent une accumulation de la forme non mature de l'*AMH* dans les cellules, expliquant possiblement la faible quantité d'*AMH* circulante chez la patiente et la diminution prématurée de sa réserve ovarienne. Des expériences supplémentaires sont encore nécessaires pour conforter cette hypothèse.

CONTRIBUTION

Dans ce manuscrit, j'ai participé à la conception, à la réalisation et à l'analyse des expériences permettant de reproduire le variant génétique *AMH* p.(Val12Gly) dans les cellules HEK293T (technologie *prime editing*) et de caractériser son impact fonctionnel. Il s'agissait de la première expérience de PE que j'ai réalisé au laboratoire

(faible efficacité d'édition et version PE ancienne, cf. introduction partie 7.2). Ces expériences correspondent à celles présentées en figure 2. Elles ont été menées sous la supervision du Dr David Gilot.

ARTICLE

Functional validation of a rare variant by prime editing: homozygous *AMH* p.(Val12Gly) variant reduces preproAMH processing and promotes diminished ovarian reserve

Sylvie Jaillard^{1,2,*}, Delphine Leclerc³, Diane Bales^{2,4}, Mathilde Domin-Bernhard⁴, Solène Duros⁴, Erika Launay², Linda Akloul⁵, Anna Lokchine², Wilfrid Carré⁶, Alinoé Lavillaureix⁵, Laurence Cluzeau⁷, Elena J. Tucker^{8,9}, Sylvie Odent⁵, David Gilot^{2,3}, Marc-Antoine Belaud-Rotureau^{1,2}

¹ Univ Rennes, CHU Rennes, INSERM, EHESP, IRSET (Institut de Recherche en Santé, Environnement et Travail) – UMR_S 1085, F-35000 Rennes, France

² CHU Rennes, Service de Cytogénétique et Biologie Cellulaire, F-35033 Rennes, France

³ INSERM U1242, COSS, Université de Rennes 1, Rennes, 35042, France

⁴ CHU Rennes, Département de Gynécologie Obstétrique et Reproduction Humaine, F-35033 Rennes, France

⁵ CHU Rennes, Service de Génétique Clinique, CLAD Ouest, F-35033 Rennes, France

⁶ CHU Rennes, Laboratoire de Bioinformatique, Service de Génétique Moléculaire et Génomique, F-35033 Rennes, France

⁷ Université de Rennes 1, UFR Médecine, Histologie, Embryologie et Cytogénétique

⁸ Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, VIC Australia

⁹ Department of Pediatrics, University of Melbourne, Melbourne, VIC Australia

* Corresponding author

Corresponding author: Sylvie Jaillard (ORCID 0000-0002-0593-0261), sylvie.jaillard@chu-rennes.fr

Keywords

AMH, diminished ovarian reserve, premature ovarian insufficiency, variant, prime editing, functional validation

ABSTRACT

Purpose: Ovarian deficiency, including premature ovarian insufficiency (POI) and diminished ovarian reserve (DOR), is a major cause of female infertility. POI is characterized by disorders of the menstrual cycle accompanied by an elevation of serum gonadotropins occurring before the age of 40 years. A decrease in ovarian reserve also occurs with age but is considered pathological if evident before the age of 40, corresponding to DOR. The latter is characterized by a low anti-Müllerian hormone and/or low antral follicle count and/or elevated follicle stimulating hormone but lower than the post-menopausal level. We report here a young adult patient presenting with DOR. Exome sequencing revealed a homozygous variant in the *AMH* gene, considered as a variant of uncertain significance (VUS). The aim of the study was to use functional validation to prove the deleterious impact of this homozygous AMH variant.

Methods: Functional studies using a cellular model have been achieved via the creation of this variant by CRISPR-Cas9 (Prime Editing).

Results: The preproAMH proteolytic maturation is less efficient in homozygous cells than in heterozygous cells, confirming the deleterious nature of this variant at the homozygous state.

Conclusion: These results could explain the low amount of circulating AMH in this patient. Since this variant is present in the general population albeit rare, its homozygous state has to be considered as potential cause of ovarian deficiency.

INTRODUCTION

Ovarian reserve represents the number of available follicles within ovaries. It can be estimated by anti-Müllerian hormone (AMH) levels, follicle stimulating hormone (FSH) levels, and/or antral follicle count (AFC) determined by ultrasounds [1]. Ovarian deficiency, including diminished ovarian reserve (DOR) and premature ovarian insufficiency (POI) is one of the main causes of female infertility. A DOR can be diagnosed by abnormal ovarian reserve testing showing low AMH and/or low AFC and/or elevated FSH but lower than the post-menopausal level, without menstrual irregularities. This condition is mostly the consequence of age, but is considered abnormal if it occurs before the age of 40 (10% of women) [2]. DOR differs as a clinical diagnosis from POI, which occurs in 1% of women under 40 and is characterized by high FSH levels >25 UI/L (post-menopausal level) and perturbation or cessation of the menstrual cycle before the age of 40. It is supposed that DOR and POI can share common genetic defects [3]. For instance, heterozygous variants in GDF9, a member of the transforming growth factor beta (TGF- β) superfamily critical for mammalian ovarian folliculogenesis, have been associated with POI (secondary amenorrhea), DOR, polycystic ovarian syndrome (PCOS) and propensity for dizygotic twins, and biallelic GDF9 variants have been associated with severe POI (primary amenorrhea) [4].

AMH is another member of the TGF- β family consisting of dimeric ligands divided in to three classes (TGF- β s, Bone morphogenetic factors, and Activins). These ligands bind serine/threonine kinase receptors through the extracellular domain, leading to phosphorylation of the type I receptor by the constitutively active type II receptor, thus allowing the activation of intracellular SMAD proteins. AMH has its own dedicated type II receptor, AMH receptor type II (AMHR2) [5]. AMH active protein is encoded by the autosomal AMH gene mapping to 19p13.3 (OMIM*600957) [6]. This gene encodes a preproprotein of 560 amino acids, giving rise to the proAMH (AMH 25-560) after cleavage of the first 24-amino acids corresponding to the signal peptide. A stable non-covalently linked complex (termed herein AMHN,C) is formed after cleavage of the proAMH at position 451/452. AMH homodimerization occurs through the conserved cysteine (C525) to give the disulfide-linked signaling molecule [5]. Circulating AMH contains a mixture of proAMH and AMHN,C, the latter only appearing to activate AMHR2 with AMHC being receptor competent and AMHN increasing its bioactivity. Recently, novel antibodies raised against the C and N terminal parts of human AMH protein have been generated and validated using human samples (human follicular fluid (FF) from small antral follicles and human granulosa cells (GCs)). This revealed new molecular weight isoforms of AMH, suggesting unknown maturation steps and it is proposed that the preproAMH maturation process is yet to be fully elucidated [7].

AMH is synthesized in the gonads of several vertebrate species, in the Sertoli cells of the testes and the ovarian GCs. Circulating levels of AMH are dimorphic, linked with the role of this hormone on gonadal development or function [8]. In females, AMH is a key factor in the regulation and progression

of ovarian follicles, and becomes undetectable after menopause. AMH is produced by the GCs of primary and small antral follicles during fetal life, when Müllerian ducts are already insensitive to AMH [6]. AMH expression is detected in GCs of activated primordial follicles but is absent in follicular stages following FSH-dependent selection [9]. AMH synthesized during the early ovarian follicular stages provides a negative retrocontrol over folliculogenesis, inhibiting follicular recruitment and the stimulating effect of FSH on the follicles. Thus, AMH ensures the protection of the ovarian reserve. Accordingly, a low AMH level is correlated with a low pool of primordial follicles. Circulating AMH is thus considered a useful marker of the ovarian reserve [10]. The role of AMH in fertility has also been proven in murine models. In *Amh* heterozygous and homozygous mutant female mice, decreased circulating AMH is associated with early depletion of primordial follicles and accelerated ovarian reserve depletion [11].

Due to the importance of AMH in female fertility, variants in AMH have been investigated. Heterozygous variants of AMH and AMHR2 with reduced AMH signalling activity *in vitro* have been found in POI women [10]. To date, different scenarios might explain the role of variants affecting the AMH gene. These variants could modify the activity of AMH by modulating the AMH_{N,C} dimerization process but might also alter the preproAMH maturation process, reducing both the release of active AMH dimers and the negative feedback effect of AMH on the primordial follicles.

Here, we identified a homozygous AMH variant, located in the peptide signal region of the protein, in a DOR patient. This rare variant is present in the general population but never in homozygous state in females suggesting the potential for a deleterious role in this case. Accordingly, the circulating levels of AMH and her follicles count were decreased. Thus, we postulated that this variant might affect the AMH protein level and in turn decrease its negative feedback effect on primordial follicles, explaining her DOR. To perform functional experiments, a model harbouring the patient variant in homozygous state was generated using the cutting-edge technology for gene editing, the Prime Editing (PE) technology.

MATERIAL AND METHODS

Participants

The proband was first seen at 27 years old for primary infertility. Menarche occurred at 11 years old and oestroprogestative contraception was started at 16 until 23 years old. Menstrual cycles were regular and short. Genitals and secondary sex characteristics were normal. Hysterosalpingography showed permeable uterine ducts. Ultrasonography examination showed normal uterus and AFC was 9, then ascertained at 7 in the following months, in favour of DOR (AFC<5-7 follicles). Pelvis MRI showed a slight thickening of the right uterosacral ligament, without other endometriosis lesions, and

ovaries with very few follicles. DOR (AMH<0.5-1.1 ng/mL) was confirmed by two hormonal assessments showing AMH: 0.6 and 0.5 ng/mL, FSH: 11.9 and 10.8 IU/L (N: 3-9 IU/L), LH: 6.3 and 5.4 IU/L (N: 2-10 IU/L), and estradiol: 32 and 26 pg/mL (N: 27-161 pg/mL). Autoimmunity testing was negative for ACA/21-OH antibodies, negative for anti TPO antibodies but positive anti TG antibodies along with normal TSH assessment. Karyotype showed 46,XX constitution. *FMR1* screening was normal. The patient was enrolled in oocyte donation program after failed attempt of IVF with insufficient response during ovarian stimulation. During this assisted reproductive management, spontaneous pregnancies occurred at 29 years old (miscarriage) and 9 months later (termination of pregnancy due to a *de novo* fetal 6.85 Mb 8p12p11.23 deletion). A successful pregnancy was finally obtained via oocyte donation. There was no cause of male infertility in her partner. Familial history was unremarkable regarding infertility.

Whole-exome sequencing

Exome capture was performed using the SureSelect Human All Exon V7 kit (Agilent Technologies Inc, Santa Clara, CA, USA) and sequenced on a NextSeq500 instrument (Illumina, San Diego, CA, USA). WES data were processed using C-GeVarA pipeline (Constitutional Genetic Variant Analysis, Bioinformatics Department, Rennes University Hospital) and deposited into SeqR for analysis (<https://seqr.broadinstitute.org/>). We performed two phases of analysis – the first focused on gene priority and the second focused on variant priority as previously described [12]. For gene-centric analysis, we considered the potential pathogenicity of all rare (<0.01 minor allele frequency MAF) coding variants within candidate genes (list available on demand). For variant-centric analysis, we considered the potential pathogenicity of all “high priority” variants in any gene, or “moderate priority” recessive-type variants in any gene. High priority variants include variants affecting essential splice sites, introducing frameshifts or premature stop codons, whereas moderate priority variants include missense variants and in frame codon deletions. Variant pathogenicity was predicted in silico using VarSome (<https://varsome.com>) which gives an overall view on prediction scores including Mutation Taster, Polyphen-2, SIFT, PROVEAN, REVEL, and CADD for example. The conservation of affected residues was assessed by Multiz Alignments of 100 vertebrates (UCSC Genome Browser <https://genome.ucsc.edu/>). MAF and tolerance of the gene to missense and/or loss-of-function variation were evaluated with gnomAD (<https://gnomad.broadinstitute.org/>) which give insight on measures of mutational constraint, LOEUF (“loss-of-function observed/expected upper bound fraction”) and pLI (“probability of being loss-of-function intolerance). Variants of interest were confirmed by targeted sequencing with the Sanger method and are described according to the hg19/GRCh37 version of the genome.

Functional studies

- Cloning

The pegRNA and sgRNA encoded oligonucleotides were respectively integrated into pU6-pegRNA-GG-acceptor backbone (Addgene plasmid #132777) and BPK1520 backbone (Addgene plasmid #65777) by golden gate assembly method which was previously described [13][14]. Guide RNA sequences for PE3 (pegRNA and sgRNA) are listed in Table S1 (Supplementary Information). Plasmids were transformed into NEB® Stable competent *E. coli* C3040H (New England BioLabs, Evry, France) by following manufacturers' recommendations. Bacteria were seeded onto ampicillin (100 µg/mL) LB agar plates and incubated overnight at 37°C. Isolated colonies were then amplified overnight at 37°C with LB growth medium and ampicillin (100 µg/mL) under constant agitation. Plasmid DNA was purified with NucleoBond Xtra Maxi kit (Macherey Nagel, Hœrdt, France) according to manufacturers' instructions. Constructs were verified by Sanger sequencing. The pegRNA and sgRNA encoded plasmids used in this study are available on demand (#132777_AMH_pegRNA_A2; 65777_AMH_sgRNA_A28).

- Cell culture

HEK293T cell line (obtained from ATCC) was maintained in Dulbecco's modified Eagle's medium (DMEM, Gibco, Villebon-sur-Yvette, France) complemented with 10% fetal bovine serum (FBS, Gibco, Villebon-sur-Yvette, France), 1% glutamin (Lonza, Basel, Switzerland) and 1% penicillin-streptomycin (Sigma, St. Quentin Fallavier, France) in a humidified incubator (37°C, 5% CO₂). Cells were negative for mycoplasma (PlasmoTest Mycoplasma Detection Kit, Invivogen, Toulouse, France).

- Transfection

Approximately 110 000 HEK293T cells were seeded in 48 well plates. One day later, when confluency was about 60-70%, cells were transfected (lipofectamine 2000, Invitrogen, Villebon-sur-Yvette, France) with pCMV-PE2-P2A-GFP plasmid (Addgene plasmid #132776), pegRNA-encoded plasmid (Addgene plasmid #132777) and with sgRNA-encoded plasmid (Addgene plasmid #65777) according to PE3 method [13]. Medium was removed 5-6 hours after transfection. Approximately 48h after, transfection efficiency was evaluated under fluorescent microscope (Invitrogen™ Evos™, ThermoScientific, Villebon-sur-Yvette, France) and GFP positives cells were sorted by FACS (FACSaria FUSION, Becton Dickinson) before single-cell seeding (limit dilution in 96 well plates, 1 cell/well). Between 2 and 3 weeks after limit dilution, HEK293T pure clones were amplified in 24 well plates. To identify edited cells, clones were screened by Sanger sequencing as described hereafter.

- DNA Extraction

DNA extraction was performed on confluent wells of P24 plates by adding 180 µL of lysis buffer (10 mM Tris-HCl, pH 7.5, 0.05% SDS) freshly complemented with proteinase K (25 µg/µL). Lysates were incubated 1h30 at 37°C (digestion) followed by 30 min at 80°C (proteinase K inactivation) [13].

- **PCR and Sanger sequencing**

PCR were performed from 1 μ L of DNA lysate using the Phusion™ High-Fidelity DNA Polymerase kit (Thermo Scientific, Villebon-sur-Yvette, France) in the presence of 0.25 μ M of AMH forward primer (5'-GGCAGAGGAGATAGGGGTCT-3') and 0.25 μ M of AMH reverse primer (5'-CCAGTCCAAGTCTTCTCGGA-3'). The thermocycling program was: 98°C for 30 sec (initial denaturation), followed by 35 temperatures cycles with 98°C for 10 sec (denaturation), 60°C for 15 sec (primer annealing) and 72°C for 30 sec (elongation), ended by 72°C for 5 min (final elongation step). A volume of 1 μ L of PCR amplified DNA was then elongated by sanger sequencing using BigDye v3.1 Terminators (Applied Biosystems, Villebon-sur-Yvette, France) following the manufacturers' protocols. We use the same thermal cycler with a different program: 5 minutes at 96° (denaturation) then 30 temperature cycles with 1 minute at 96°C (denaturation), 1 minute at 60°C (primer hybridization) and finally 2 minutes at 62°C (elongation). Sequences were purified with sephadex G50 (GE Healthcare, Vélizy-Villacoublay, France) and migrated on 3130 XL 16 capillaries (Applied Biosystem, Villebon-sur-Yvette, France).

- **Western-blot experiments**

Protein was extracted with RIPA lysis buffer (Pierce, ThermoScientific, Villebon-sur-Yvette, France) supplemented with phosphatase and protease inhibitor (Pierce, ThermoScientific, Villebon-sur-Yvette, France). Protein concentration was assayed by standard colorimetric method (BCA Protein Assay kit, Pierce, ThermoScientific, Villebon-sur-Yvette, France) and absorbance was measured at 562nm on a spark microplate reader (TECAN). For western blot, 14 μ g of protein was denatured 10min at 70°C in the presence of lithium dodecyl sulfate (NuPAGE, ThermoScientific, Villebon-sur-Yvette, France) and reducing agent (NuPAGE). Reduced samples were loaded on 4 to 12% Bis-Tris Gel (NuPAGE). After migration (1h30, 200V, 400mA), samples were transferred on a nitrocellulose membrane with iBlot gel transfer device (Invitrogen, ThermoScientific, Villebon-sur-Yvette, France). Then, the membrane was incubated for 1 hour at room temperature with blocking solution (TBS-Tween 0,1% ; 5% Bovine Serum Albumin, Eurobio, Les Ulis, France). After blocking, primary and secondary antibodies were incubated respectively overnight at 4 degrees and/or 1 hour at room temperature. Antibodies used in this study are the following: AMH primary mouse monoclonal mouse antibody, specific against AMHN (AMH-39/48, Ansh Labs, LLC, Webster, TX, USA, 1:2500) [7], anti-mouse secondary antibody (Cell signalling, Ozyme, Saint-Quentin en Yvelines, France, ref. #7076, 1:2000) and anti HRP-linked mouse beta-Actin antibody (Cell signalling, Ozyme, Saint-Quentin en Yvelines, France, ref. #12262, 1:1000). Antibody-protein complexes were revealed by chemiluminescence method with amersham ECL select substrate (Cytiva, Brumath, France) and signals were detected under luminescent image analyser (ImageQuant LAS 4000).

RESULTS

WES identified homozygous missense variant in the AMH gene

Variant filtration revealed a homozygous candidate variant in the AMH gene (NC_000019.9:g.2249366T>G, NM_000479.5(AMH):c.35T>G, NP_000470.2:p.(Val12Gly), rs149082963) (Supplementary Information Figure S1). This variant is rare in the general population (MAF=0.001422) and in individuals of European background (MAF=0.002613) according to gnomAD v2.1.1, and has never been reported in homozygous state in XX individuals, although is found in homozygous state in one XY individual. This variant leads to the substitution of a valine to glycine in the signal peptide of the AMH preproprotein and is not likely to impact secondary protein structure as these residues share similar properties. This substitution occurs at a position where amino acids with similar properties to valine are tolerated across species. In silico analyses are inconsistent in their predictions as to whether or not the variant is damaging to the protein structure/function. This variant is classified as VUS in ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) and LOVD (<https://databases.lovd.nl/shared/variants/AMH/unique>), and is considered disease-causing in HGMD Pro 2021.4 (<https://my.qiagen.digitalinsights.com/bbp/view/hgmd/pro/all.php>). We interpret this as a variant of uncertain significance (VUS) (class 3) according to the American College of Medical Genetics and Genomics (ACMG) [15].

To confirm the homozygous state and to exclude a microdeletion on the first exon of AMH gene, an array-CGH enriched with probes targeting 130 POI genes was performed (Agilent Technologies, Santa Clara, CA, USA) (Supplementary Information Figure S2). This array contains 8 probes for the AMH gene, including a probe covering the first exon (probe A_16_P34383951: 2,249,329-2,49,373 on chr19 (hg19/GRCh37)). This result confirmed the homozygous state of the variant identified by exome sequencing and confirmed by Sanger sequencing. Parental studies have not been achieved due to lack of parental DNA.

The homozygous p.(Val12Gly) variant is associated with a reduction of AMH level

First, to visualize the VUS position in the AMH preproprotein, we performed structure prediction via the AlphaFold tool, a novel machine learning approach based on physical and biological knowledge about protein structure [16]. The Valine 12 is located in the signal peptide of the AMH preproprotein (Figure 1A). After cleavage of the first 24-amino acids (the signal peptide), the proAMH (AMH 25-560) undergoes further maturation steps by cleavage as recently demonstrated [7].

To elucidate the role of this VUS, functional studies were performed using gene editing workflow to generate a homozygous AMH V12G cell model. To select the best cell model, we firstly evaluated the AMH mRNA expression levels in the publicly available data of single cell RNA-seq data (<https://www.proteinatlas.org>). As expected, AMH mRNA is almost exclusively expressed in GCs (Figure 1B). Since the primary culture of GCs is difficult and requires ovarian or follicular samples

followed by GC purification, we analyzed the AMH expression levels in available cell lines. Cell lines such as HEK293T, K562 and HeLa are usually used for gene editing experiments in order to elucidate VUS function [17]. Based on AMH mRNA levels, we selected the HEK293T cell line to introduce the VUS (Figure 1C). As the predictions need to be functionally validated in cellulo, we developed a workflow to introduce the VUS into the DNA of HEK293T cells by gene editing (Figure 2). We selected the Prime Editing (PE) approach because it requires a single strand break in contrast to the previous CRISPR approach based on a double strand break (Cas9). Thus, the PE Cas9 (Cas9 H840A, nickase) has been shown to generate less off-target edits than wild-type Cas9 (associated with a donor oligonucleotide for DNA repair) [13]. Moreover, the base editor (ABE and CBE) are not able to create this point mutation despite their high gene editing efficiency [18]. We designed the PE guide RNA (pegRNA) using the Prime design tool (<https://github.com/pinellolab/PrimeDesign>) [19] (Figures 2A-B) to target exon 1 of the AMH gene and introduce a C instead of A. In the extension part of the pegRNA, we inserted the C mutation (RT template, Figure 2B). This sequence is introduced into the AMH gene by the reverse transcriptase fused to Cas9 nickase. To increase the gene editing efficiency, we cotransfected a sgRNA promoting a single strand break on the edited (nick-to-pegRNA distance +28). To maximize our chance to obtain a homozygous NC_000019.9:g.2249366T>G clone, we sorted the transfected cells two days after transfection (pegRNA, sgRNA and Cas9 nickase fused to RT). The Cas9-RT plasmid also encoded the GFP protein, allowing this cell sorting (GFP positive cells) (Figure 2C). A clonal selection was next performed from the GFP positive cells. The edit has been confirmed by DNA sequencing (Sanger) (Figure 2D), and the expression level of the preproAMH determined by western-blot. We used the latest antibody raised against the N-terminal of preproAMH, recently validated in patient samples [7]. As expected, a ~65kDa band (corresponding to the preproAMH) was detected in wild-type and mutated HEK293T cells (Figure 2E and (Supplementary Information Figure S3)). Interestingly, preproAMH expression level was higher in mutated clones, suggesting a decrease of the pre-proA-MH proteolytic maturation. The deleterious effect of p.(Val12Gly) (Figure 2F) is in accordance to the low level of circulating AMH form detected in the blood of the patient. blood of the patient.

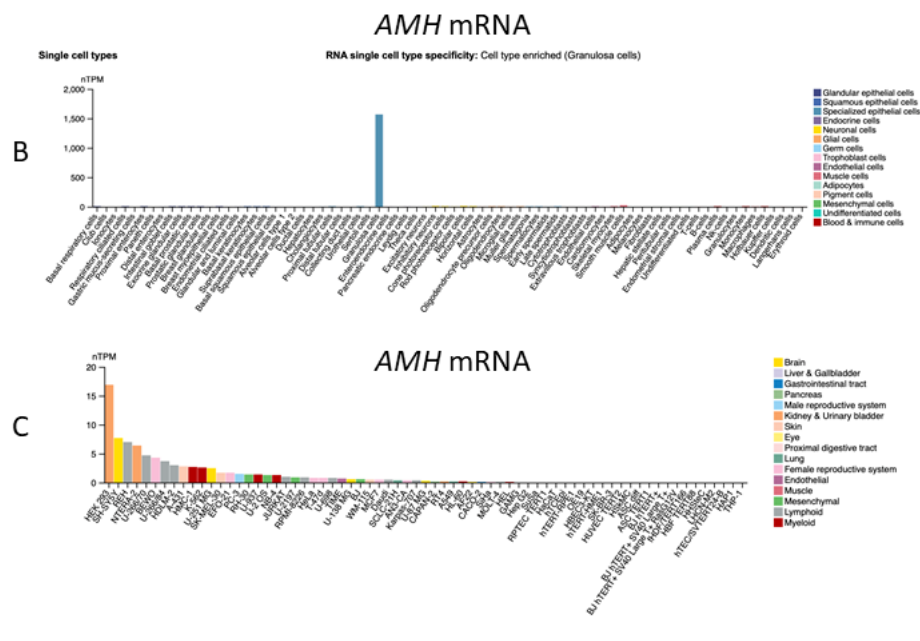
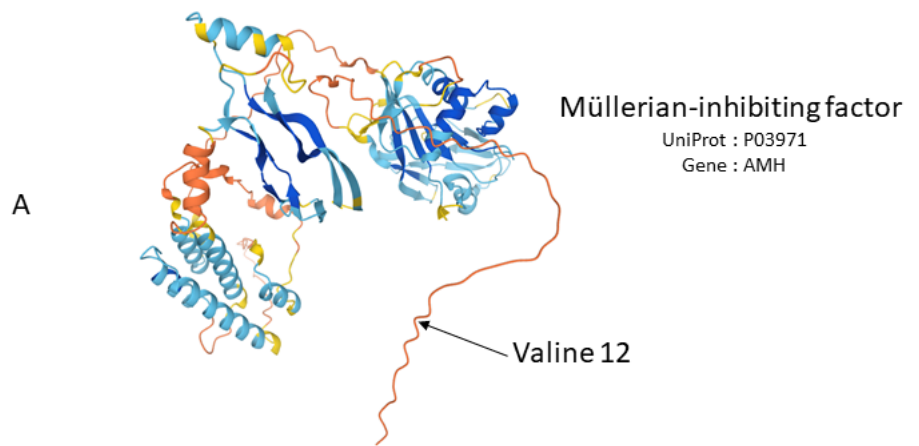


Figure 1: *AMH* mRNA expression and p.(Val12Gly) variant

A: Predicted 3D structure of preproAMH (alias Muellerian-inhibiting factor, UniProt : P03971, Gene: AMH) using AlphaFold [16], (<https://alphafold.ebi.ac.uk>). The Valine 12 is located in the signal peptide of the AMH preproprotein. **B:** *AMH* mRNA expression levels have been obtained from single cell RNA sequencing data available on www.proteinatlas.org. nTPM for normalized Transcripts Per Kilobase Million. The chart shows higher expression in granulosa cells (GCs). **C:** *AMH* mRNA expression levels in a set of commercially cell lines. Data have been obtained from www.proteinatlas.org. nTPM for normalized Transcripts Per Kilobase Million. Cells lines have been ordered based on *AMH* levels. The chart shows higher expression in HEK293T cell line.

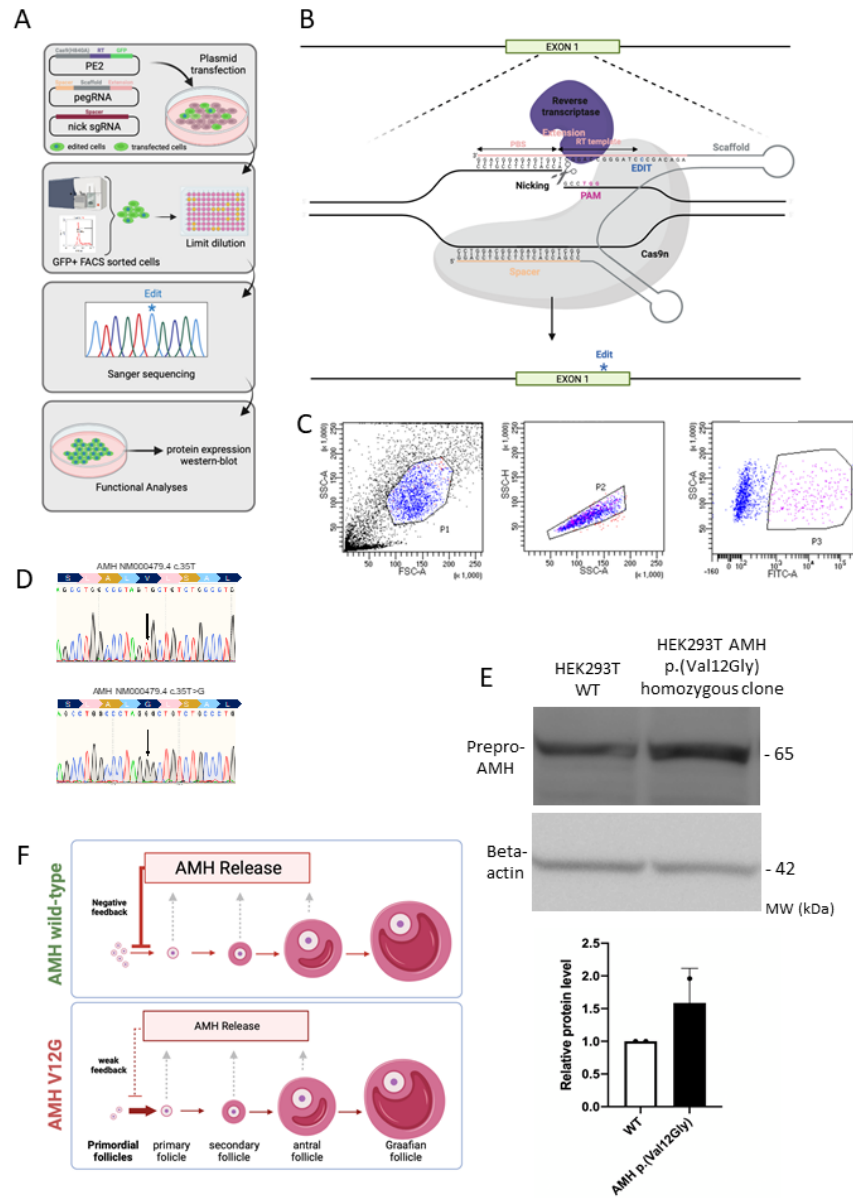


Figure 2: AMH mRNA expression and p.(Val12Gly) variant

A: Workflow to generate p.(Val12Gly) mutated clone from HEK293T cells using Prime Editing. **B:** Scheme illustrating the Prime Editing strategy (pegRNA) and targeted sequences. The pegRNA was designed to target exon 1 of the AMH gene and introduce a C instead of A. Cas9n : Cas9 nickase, PBS : Primer binding site, RT template : Reverse transcriptase template, PAM : Protospacer adjacent motif. **C:** Scatter plots showing the selection of GFP positive cells after transfection. Viable cells have been identified as population 1 (P1). Next, the doublets have been excluded and we only selected the population 2 (P2) to sort the GFP positive cells (channel FITC). **D:** Sanger sequencing illustrations for wild-type (upper arrow) and p.(Val12Gly) AMH clones (lower arrow showing a homozygous mutation). **E:** PreproAMH detection by western-blot experiments. Beta-actin has been used as loading control. Western-blot (upper image) is representative of two experiments. Quantification (lower image) has been performed according to two biological replicates. **F:** Hypothetical Scheme resuming the p.(Val12Gly) deleterious effect for this woman. The negative feedback produced by AMH release is lower due to the lack of secretion of AMH and the low level of circulating AMH form in the blood of the patient.

NC_000019.9 GRCh37	NM_000479.5	NP_000470.3	Phenotype	ACMG	Status	Functional validation	Reference
g.2249171C>G	c..161C>G		PCOS	VUS	Het	Reduced AMH signaling	Gorsic et al., 2019
g.2249263C>G	c..69C>G		PCOS	VUS/LB	Het	Reduced AMH signaling	Gorsic et al., 2019
g.2249272C>T	c..60C>T		PCOS	VUS/LB	Het	Reduced AMH signaling	Gorsic et al., 2019
g.2249366T>G	c.35T>G	p.(Val12Gly)	PMDS PMDS PCOS X-linked hypospadias DOR	VUS LP	Het cp Hom Het Het Hom	No No Reduced AMH signaling No Maturation process alteration	Imbeaud et al., 1994 Mazen et al., 2011 Gorsic et al., 2017 Hughes et al., 2019 Present case
g.2249401G>A	c.70G>A	p.(Ala24Thr)	PCOS	VUS	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2249467C>G	c.136C>G	p.(Pro46Ala)	PCOS	VUS	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2249477G>T	c.146G>T	p.(Ser49Ile)	Disease association (PCOS severity)		Het	Reduced AMH activity	Kevenaar et al., 2008
g.2249603G>A	c.272G>A	p.(Arg91His)	PCOS	VUS	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2249626A>T	c.295A>T	p.(Thr99Ser)	HH (male) PCOS	VUS	Het Het	Reduced AMH secretion Reduced AMH signaling	Malone et al., 2019 Gorsic et al., 2017
g.2250374C>T	c.451C>T	p.(Pro151Ser)	KS (male) PCOS PMDS	VUS	Het Het Het	Reduced AMH secretion Reduced AMH signaling No	Malone et al., 2019 Gorsic et al., 2017 Picard et al., 2017
g.2250389G>A	c.466G>A	p.(Ala156Thr)	PCOS	VUS	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2250476C>G	c.553C>G	p.(Gln185Glu)	PMDS PCOS	VUS	Hom Het	No Reduced AMH signaling	Imbeaud et al., 1994 Gorsic et al., 2017
g.2250480T>C	c.555+2T>C	p.?	PCOS	LP	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2250676G>A	c.581G>A	p.(Arg194His)	PCOS	VUS	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2250897C>A	c.714C>A	p.(Asp238Glu)	KS (female)	VUS	Het	Decreased AMH released	Malone et al., 2019
g.2250973G>C	c.790G>C	p.(Gly264Arg)	POI	VUS	Het	Reduced AMH activity	Alvaro Mercadal et al., 2015
g.2250991C>T	c.808C>T	p.(Pro270Ser)	PCOS	VUS/LB	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2251137C>G	c.864C>G	p.(Asp288Glu)	POI PCOS + Control	VUS	Het Het	Reduced AMH secretion Similar signaling to WT	Alvaro Mercadal et al., 2015 Gorsic et al., 2017
g.2251178G>A	c.905G>A	p.(Arg302Gln)	PMDS PMDS PCOS	VUS	Het cp Het Het	No No Reduced AMH signaling	Hughes et al., 2019 Picard et al., 2017 Gorsic et al., 2017
g.2251327C>T	c.1054C>T	p.(Pro352Ser)	PCOS	VUS	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2251358C>T	c.1084C>T	p.(Pro362Ser)	PCOS	VUS	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2251370C>T	c.1097C>T	p.(Pro366Leu)	PCOS	VUS/LB	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2251388C>T	c.1115C>T	p.(Ala372Val)	PCOS	VUS	Het	Normal AMH signaling	Gorsic et al., 2017
g.2251427C>T	c.1154C>T	p.(Ala385Val)	PCOS	VUS	Het	Reduced AMH signaling	Gorsic et al., 2017
g.2251604G>A	c.1331G>A	p.(Arg444His)	POI	VUS	Het	Reduced AMH activity	Alvaro Mercadal et al., 2015
g.2251666C>T	c.1393C>T	p.(Arg465Cys)	POI	VUS/LP	Het ?	No	ClinVar
g.2251791C>G	c.1518C>G	p.(His506Gln)	PMDS PMDS PCOS	VUS	Het cp Het cp Het	No No Reduced AMH signaling	Imbeaud et al., 1996 Eggers et al., 2016 Gorsic et al., 2017
g.2251829C>T	c.1556C>T	p.(Ala519Val)	PMDS Control	VUS	Het	No Similar signaling to WT	Hughes et al., 2019 Gorsic et al., 2017

Table 1: Variants of interest reported in ClinVar, LOVD, HGMD Pro, and in the literature in women (hg19/GRCh37)

PCOS: polycystic ovarian syndrome

PMDS: persistent Müllerian duct syndrome

HH: hypogonadotropic hypogonadism

HS: Kallman syndrome

POI: premature ovarian insufficiency

DOR: diminished ovarian reserve

VUS: variant of uncertain significance

LB: likely benign

LP: likely pathogenic

Het: heterozygous

Het cp: compound heterozygous

Hom: homozygous

AMH: anti-Müllerian hormone

WT: wild type

DISCUSSION

In women, AMH is expressed by GCs of growing follicles prior to FSH-dependent selection. The protein is involved in the preservation of the primordial follicles stock in women via a negative feedback loop, and inhibits the recruitment and growth of follicles by restricting growth factors and the effect of gonadotropins. The lack of AMH results in a faster depletion of the follicular pool in the ovaries, leading therefore to a diminished ovarian reserve [20].

AMH variants in reproductive conditions

More than 100 genetic alterations of the AMH gene have been reported to date in male and female patients with variability regarding the status (homozygous, compound heterozygous or heterozygous), the associated phenotype (persistent Müllerian duct syndrome (PMDS) as the most frequent, hypogonadotropic hypogonadism, polycystic ovarian syndrome (PCOS), and POI), their classification, and their functional validation. In XY patients with PMDS, a recessive autosomal condition, all types of variants are reported (missense, nonsense, splice, deletions, insertions) [21]. On the contrary, in women, AMH variants have been reported in a heterozygous state with a majority of missense variants (Table 1). It has been proposed that heterozygous variants can have a deleterious effect on dimerization of TGF- β family proteins. Three rare missense variants have been associated to POI with secondary amenorrhea, with one of these inherited from a mother also presenting with POI [10]. Functional studies identified normal production and secretion but impaired function for two of them (p.Gly264Arg, p.Arg444His), and diminished secretion with residual activity for the third one (p.Asp288Glu). The secondary amenorrhea can be explained by the fact that the ovaries can accomplish their function but during a shorter duration [10]. The p.(Arg465Cys) associated with POI in ClinVar is considered likely pathogenic although no functional evidence is provided. Here, we described the first woman carrying the homozygous mutation p.(Val12Gly) in the AMH gene associated with DOR detected through primary infertility assessment and low AMH identification. AMH heterozygous variants have also been reported in PCOS, including the p.(Val12Gly) observed in 6 patients in a cohort of 80 PCOS women [22]. The majority of the reported variants for this condition are in the prodomain region of AMH. They are likely to have an effect on mature protein processing and/or bioactivity as the prodomain, which remains noncovalently associated with the mature domain, influences the activity of the latter. It has been suggested that AMH heterozygous variants have a dominant-negative effect forming heterodimers with wild-type AMH, with only 25% of wild-type AMH dimers having signaling capability in vitro [22]. Impaired AMH could contribute to PCOS phenotype notably through the loss of target gene CYP17 inhibition [23]. Finally, a heterozygous variant has been reported in a woman with Kallmann syndrome inherited from her affected father, impacting AMH signaling which is believed to regulate migration of GnRH neurons toward the brain [24]. These data broaden the spectrum of pathologic conditions that could be linked with AMH variants.

Insight on the p.(Val12gly) variant

p.(Val12Gly) is rare for the individuals of European background population and it has never been found at homozygous state in XX individuals, suggesting that it might be deleterious with an impact on fertility. Interestingly, gnomAD reports one homozygous XY individual bearing this variant, suggesting it could be a mild variant with a residual activity sufficient for male development. Nevertheless, it has also been reported in a homozygous state in XY Egyptian patients with PMDS, suggesting a deleterious role of this variant, due to the impact on a region crucial for translocation of secreted proteins [25]. The p.(Val12Gly) has also been reported in an affected XY individual with PMDS, who was compound heterozygous for this variant and another AMH variant [26] and in a male with X-linked hypospadias bearing a MAMLD1 likely pathogenic variant [27]. To discern a deleterious impact of this variant, we created it using the cutting-edge approach of gene editing. Thus, we avoided the classical overexpression of a cDNA containing the genetic alteration, broadly used for functional studies. Here, using gene editing, we maintained an endogenous expression level of preproAMH, in contrast to levels reached after plasmid transfection. This is a critical point because we evaluated an alteration of the preproAMH maturation and/or the intracellular route. An overexpression of this protein obtained by plasmid transfection might alter these processes and our conclusions. We thus observed an intracellular accumulation of the preproAMH when compared to the normal cells (expressing the wild-type preproAMH). This result suggest that the p.(Val12Gly) variant might alter the preproAMH proteolytic maturation and/or the intracellular route of this peptide. This result is consistent with the position of the variant (located into the signal peptide required for the intracellular route of the preprohormone) [7]. Moreover, a functional study based on overexpression of plasmid encoding NP_000470.2:p.(Val12Gly) demonstrated that this AMH variant is less efficient than AMH wild-type protein to transactivate the TGFBR2 receptor (decrease of ~35% of luciferase activity) [22], although the reason for this altered expression is not explored. Our data suggest that the preproAMH maturation process is altered, likely reducing both the release of active AMH dimers and the negative feedback of AMH on the primordial follicles. Accordingly, we observed a decrease of circulating AMH peptide for this patient.

Limit of the study

We obtained a homozygous clone for p.(Val12Gly) in the HEK293T cells using PE. Although this cell line expresses the highest level of AMH mRNA among the panel of cell lines (Figure 1B), the expression remains low when compared to GC. Unfortunately, it was not possible to collect GCs from this DOR patient due to low response to ovarian stimulation. In addition, we used the most validated antibody targeting the N-terminal part of the preproAMH ((AMH-39/48), Ansh Labs, LLC, Webster, TX, USA) [7].

CONCLUSION

We report the first homozygous AMH variant associated with DOR. Functional validation has been performed using a gene editing approach in order to strengthen the role of this variant in the occurrence of this condition. Our data suggest that the homozygous p.(Val12Gly) variant is associated with an intracellular accumulation of the preproAMH and a reduction of AMH level.

Statements and Declarations

Editorial Policies and Ethical Considerations

Written informed consent was obtained from the patient. All procedures were in accordance with the ethical standards of the Ethics Committee of Rennes University Hospital and the French law (CCTIRS Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé).

Competing interests

The authors declare no conflict of interest.

Data availability statement

There are no linked research data sets for this paper. Data will be made available on request.

Author contribution

SJ, DL, DG and MABR contributed to conception and design of the study. SJ, DG and MABR participated in the project supervision. MDB, SD, LA, SO and ALa were involved in patients' medical care. EL, ALo, LC, and DG contributed to functional work. All authors contributed to the acquisition and/or analysis of data. SJ, DL, DG, EJT and MABR wrote the manuscript. All authors revised it critically and approved the final version.

Acknowledgements

We thank the Bioinformatic department of CHU Rennes (UF Bioinformatique et Génétique Computationnelle, Service de Génétique Moléculaire et Génomique, Pr M. De TAYRAC), Biosit (Cell sorter, CytomeTRI), Établissement Français du Sang (Rennes) for helpful advice, technical assistance and reagents. We would like to thank the patient for taking part in our research.

Funding statement

This work was supported by a CHU Rennes grant (Appel à Projets Innovations 2019 to SJ).

REFERENCES

1. Gleicher N, Kim A, Kushnir V, Weghofer A, Shohat-Tal A, Lazzaroni E, et al. Clinical relevance of combined FSH and AMH observations in infertile women. *J Clin Endocrinol Metab.* 2013;98:2136–45.
2. Pastore LM, Christianson MS, Stelling J, Kearns WG, Segars JH. Reproductive ovarian testing and the alphabet soup of diagnoses: DOR, POI, POF, POR, and FOR. *J Assist Reprod Genet.* 2018;35:17–23.
3. Jaillard S, Sreenivasan R, Beaumont M, Robevska G, Dubourg C, Knarston IM, et al. Analysis of NR5A1 in 142 patients with premature ovarian insufficiency, diminished ovarian reserve, or unexplained infertility. *Maturitas.* 2020;131:78–86.
4. Jaillard S, Bell K, Akoul L, Walton K, McElreavy K, Stocker WA, et al. New insights into the genetic basis of premature ovarian insufficiency: Novel causative variants and candidate genes revealed by genomic sequencing. *Maturitas.* 2020;141:9–19.
5. Hart KN, Stocker WA, Nagykerly NG, Walton KL, Harrison CA, Donahoe PK, et al. Structure of AMH bound to AMHR2 provides insight into a unique signaling pair in the TGF- β family. *Proc Natl Acad Sci U S A.* 2021;118:e2104809118.
6. Brunello FG, Rey RA. AMH and AMHR2 Involvement in Congenital Disorders of Sex Development. *Sex Dev Genet Mol Biol Evol Endocrinol Embryol Pathol Sex Determ Differ.* 2021;1–9.
7. Mamsen LS, Bøtkjær JA, Kristensen SG, Pors SE, Jeppesen JV, Kumar A, et al. High Variability of Molecular Isoforms of AMH in Follicular Fluid and Granulosa Cells From Human Small Antral Follicles. *Front Endocrinol.* 2021;12:617523.
8. McLennan IS, Pankhurst MW. Anti-Müllerian hormone is a gonadal cytokine with two circulating forms and cryptic actions. *J Endocrinol.* 2015;226:R45–57.
9. Bertho S, Neyroud AS, Brun T, Jaillard S, Bonnet F, Ravel C. Anti-Müllerian hormone: A function beyond the Müllerian structures. *Morphol Bull Assoc Anat.* 2021;S1286-0115(21)00251-4.
10. Alvaro Mercadal B, Imbert R, Demeestere I, Gervy C, De Leener A, Englert Y, et al. AMH mutations with reduced in vitro bioactivity are related to premature ovarian insufficiency. *Hum Reprod Oxf Engl.* 2015;30:1196–202.
11. Mullen RD, Ontiveros AE, Moses MM, Behringer RR. AMH and AMHR2 mutations: A spectrum of reproductive phenotypes across vertebrate species. *Dev Biol.* 2019;455:1–9.
12. Tucker EJ, Jaillard S, Grover SR, van den Bergen J, Robevska G, Bell KM, et al. TP63-truncating variants cause isolated premature ovarian insufficiency. *Hum Mutat.* 2019;40:886–92.
13. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature.* 2019;576:149–57.
14. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature.* 2015;523:481–5.
15. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med Off J Am Coll Med Genet.* 2015;17:405–24.
16. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583–9.
17. Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen P-F, et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell.* 2021;184:5635–5652.e29.
18. Saha K, Sontheimer EJ, Brooks PJ, Dwinell MR, Gersbach CA, Liu DR, et al. The NIH Somatic Cell Genome Editing program. *Nature.* 2021;592:195–204.
19. Hsu JY, Grünwald J, Szalay R, Shih J, Anzalone AV, Lam KC, et al. PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *Nat Commun.* 2021;12:1034.
20. Bedenk J, Vrtačnik-Bokal E, Virant-Klun I. The role of anti-Müllerian hormone (AMH) in ovarian disease and infertility. *J Assist Reprod Genet.* 2020;37:89–100.
21. Picard J-Y, Cate RL, Racine C, Josso N. The Persistent Müllerian Duct Syndrome: An Update Based Upon a Personal Experience of 157 Cases. *Sex Dev Genet Mol Biol Evol Endocrinol Embryol Pathol Sex Determ Differ.* 2017;11:109–25.
22. Gorsic LK, Kosova G, Werstein B, Sisk R, Legro RS, Hayes MG, et al. Pathogenic Anti-Müllerian Hormone Variants in Polycystic Ovary Syndrome. *J Clin Endocrinol Metab.* 2017;102:2862–72.
23. Gorsic LK, Dapas M, Legro RS, Hayes MG, Urbanek M. Functional Genetic Variation in the Anti-Müllerian Hormone Pathway in Women With Polycystic Ovary Syndrome. *J Clin Endocrinol Metab.* 2019;104:2855–74.
24. Malone SA, Papadakis GE, Messina A, Mimouni NEH, Trova S, Imbernon M, et al. Defective AMH signaling disrupts GnRH neuron development and function and contributes to hypogonadotropic hypogonadism. *eLife.* 2019;8:e47198.
25. Mazen I, Abdel Hamid MS, El-Gammal M, Aref A, Amr K. AMH gene mutations in two Egyptian families with persistent müllerian duct syndrome. *Sex Dev Genet Mol Biol Evol Endocrinol Embryol Pathol Sex Determ Differ.* 2011;5:277–80.
26. Imbeaud S, Carré-Eusèbe D, Rey R, Belville C, Josso N, Picard JY. Molecular genetics of the persistent müllerian duct syndrome: a study of 19 families. *Hum Mol Genet.* 1994;3:125–31.
27. Hughes LA, McKay-Bounford K, Webb EA, Dasani P, Clokie S, Chandran H, et al. Next generation sequencing (NGS) to improve the diagnosis and management of patients with disorders of sex development (DSD). *Endocr Connect.* 2019;8:100–

SUPPLEMENTARY INFORMATIONS

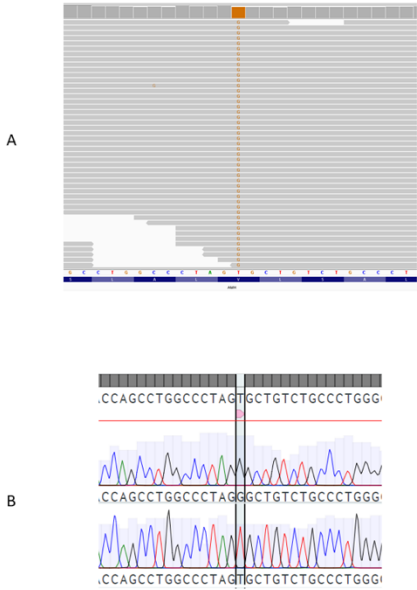


Figure S1: Exome sequencing and Sanger sequencing results

- A:** Integrative Genomics Viewer of the homozygous *AMH* variant
- B:** Sanger validation of the *AMH* homozygous variant. Upper: patient; Lower: control

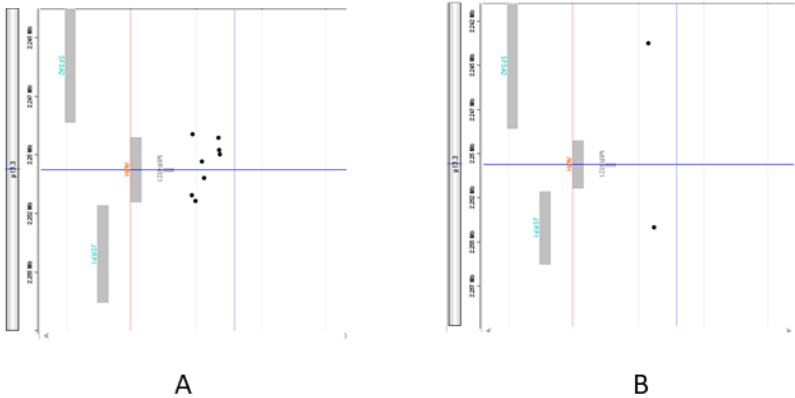


Figure S2: Array-CGH custom Agilent Human Genome CGH microarray 180K

- A:** patient's array profile on *AMH* enriched gene (custom array)
- B:** control profile on *AMH* non-enriched gene (commercial array)

Oligonucleotide array-CGH was performed using a custom Agilent Human Genome CGH microarray 180K (Agilent Technologies, Santa Clara, CA, USA). This microarray is constituted of 180 000 60-mer oligonucleotide probes that span both coding and non-coding regions, and is enriched in 130 genes involved in POI or in ovarian function (list available on demand). Coverage of the human genome is made with a median probe spacing of 13 kb on the backbone and offer a resolution of 3 kb for genes < 10 kb in size, 6 kb for 10-30 kb genes, and 30-50 kb for genes > 30 kb in size. For *AMH* (3 kb), 8 probes ensure covering of the gene whereas no probe is located on this gene on the non-custom 180K array. The experiments were performed according to version 8.0 (December 2019) of the protocol provided by Agilent (Oligonucleotide Array-Based CGH for Genomic DNA Analysis - Enzymatic Labelling for kb Blood, Cells, or Tissues). Reference genomic DNA (gDNA) was the female one provided by Agilent. A graphical overview and analysis of the data were obtained by using the Agilent CytoGenomics software

version 5.2.1.4. Identification of probes with a significant gain or loss was based on the log2 ratio plot deviation from 0 with cut-off values of 0.5 to 1 and -0.5 to -1 respectively. Black dots represent probes with normal ratio.

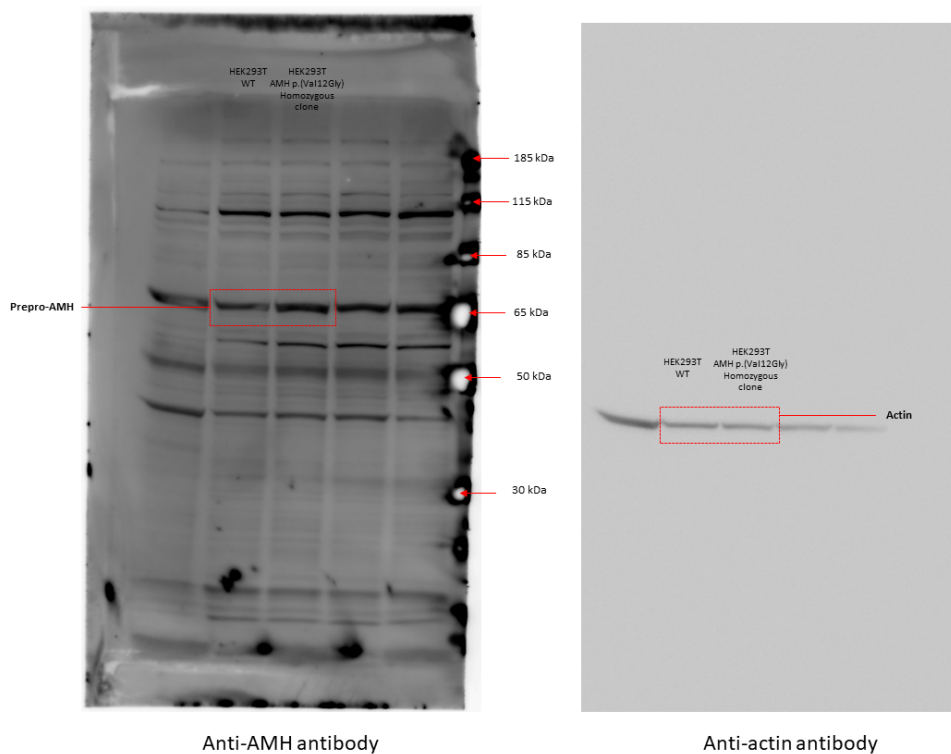


Figure S3: Uncropped western blots

Uncropped pictures of western blot experiments.

pegRNA	WT host plasmid: addgene ID	spacer sequence 5'→3'	Scaffold sequence 5'→3'	Extension sequence 5'→3'
132777 AMH_pegRNA_A2	#132777 BLU-pegRNA-GG-acceptor	GGACTGCTCTCACCAGGC	GTTTATAGAGCTAGAAATAGCAAGTAAAAATANGGCTAGTCCGTATCACTTGAAAAAGTGCCACGAGTGGTGC	AGACAGCCTAGGGCCAGGCTGTGAGAGGCAGG
nick gRNA	WT host plasmid: addgene ID	sgRNA sequence 5'→3'		
65777 AMH_sgRNA_A28	#65777 BPK1520	CAGTCCCCAGCAGAGCCCC		

Table S1: pegRNA and sgRNA sequences

CONCLUSION DE L'ARTICLE

La caractérisation des VSI identifiés par les laboratoires de génétique médicale est indispensable pour assurer au patient une prise en charge adéquate. Les outils d'édition du génome permettent dorénavant de faciliter cette étape. Grâce à la technologie *prime editing*, le variant génétique *AMH* p.(Val12Gly) a été installé dans le modèle cellulaire HEK293T. L'accumulation de la forme non mature de l'AMH dans les cellules HEK293T pourrait probablement expliquer le phénotype de la patiente (diminution prématurée de la réserve ovarienne).

Des expériences complémentaires sont encore nécessaires pour conclure sur la pathogénicité de ce variant génétique. Premièrement, une quantification précise des différents isoformes post-traductionnels de l'AMH pourrait être réalisée par spectrométrie de masse ou test AlphaLISA. Deuxièmement, un autre modèle cellulaire aurait été plus relevant (ex : cellules à granulosa). S'agissant de notre toute première expérience PE, nous avons préféré tester dans un premier temps nos guides dans les cellules HEK293T. Comme le variant est localisé dans la région du *peptide signal* de l'AMH, notre hypothèse est que l'adressage de la protéine au réticulum endoplasmique est altéré. La localisation subcellulaire de l'AMH, son niveau de sécrétion et son activité seront prochainement investigués par le laboratoire.

RESULTAT 3

Évaluation fonctionnelle d'un variant hétérozygote du gène *PBX1* identifié chez un fœtus avec anomalies du développement sexuel

AVANT PROPOS

Ce projet a été réalisé dans le cadre de la mise en place de l'activité EPIQUE dans le laboratoire de cytogénétique et biologie cellulaire du CHU de Rennes, dirigé par le Pr Marc-Antoine Belaud-Rotureau. L'article a été publié le 17 janvier 2023 dans la revue *Genes*.

INTRODUCTION DE L'ARTICLE





Le Dr Laura Mary, médecin généticien au sein du laboratoire de cytogénétique et biologie cellulaire du CHU de Rennes, s'intéresse aux causes génétiques des anomalies du développement sexuel. Une cohorte prospective constituée de 10 patients du CHU de Rennes a été envoyée à la société IntegraGen pour un séquençage d'exome en trio. Dans cette cohorte, un variant génétique apparut *de novo* dans le gène *PBX1* a été identifié chez un patient. *PBX1* est un facteur de transcription ubiquitaire essentiel pour le développement embryonnaire. Des altérations dans ce gène ont déjà été rapportées chez d'autres patients souffrants d'anomalies de développement rénaux. Son rôle dans l'établissement de la gonade bipotentielle chez la souris a déjà été mentionné. Ces arguments laissent supposer que ce variant pourrait être impliqué dans le phénotype du patient. Afin d'évaluer sa fonction, *PBX1* a préalablement été inactivé dans le modèle cellulaire HEK293T, puis un plasmide surexprimant la version mutée du gène a été transfecté dans les cellules. Les résultats obtenus suggèrent un impact du variant sur l'épissage des gènes cibles, mais des études fonctionnelles complémentaires sont nécessaires pour consolider cette observation.

CONTRIBUTION

Dans cette étude, j'ai conçu les ARN guides et j'ai mis au point l'ensemble des protocoles permettant d'inactiver le gène *PBX1*. Les données expérimentales suggèrent que l'inactivation de toutes les copies du gène est délétère pour les cellules. Nous avons seulement obtenu une inactivation partielle du gène (*knockdown*). J'ai formé le Dr. Laura Mary (médecin généticien) à la culture cellulaire, qui a mené les expériences dans le cadre de l'obtention de sa thèse d'université.

Article

Functional Assessment of a New PBX1 Variant in a 46,XY Fetus with Severe Syndromic Difference of Sexual Development through CRISPR-Cas9 Gene Editing

Laura Mary ^{1,2,*} , Delphine Leclerc ³, Audrey Labalme ⁴, Pascale Bellaud ⁵, Séverine Mazaud-Guittot ² , Stéphane Dréano ⁶, Bertrand Evrard ², Antoine Bigand ⁷, Aurélie Cauchoix ⁷, Philippe Loget ⁷, Anna Lokchine ¹, Laurence Cluzeau ¹, David Gilot ³ , Marc-Antoine Belaud-Rotureau ^{1,2} and Sylvie Jaillard ^{1,2} 

- ¹ CHU Rennes, Service de Cytogénétique et Biologie Cellulaire, F-35033 Rennes, France
² Inserm, EHESP, IRSET (Institut de Recherche en Santé, Environnement et Travail)-UMR_S 1085, Université Rennes 1, F-35033 Rennes, France
³ Inserm U1242, Centre de Lutte Contre le Cancer Eugène Marquis, Université de Rennes, F-35033 Rennes, France
⁴ Service de Génétique, Hospices Civils de Lyon, F-69007 Lyon, France
⁵ CNRS, Inserm UMS Biosit, France BioImaging, Core Facility H2P2 Rennes, Université Rennes 1, F-35033 Rennes, France
⁶ CNRS, UMR 6290 IGDR BIOSIT, Molecular Bases of Tumorigenesis, Université Rennes 1, F-35033 Rennes, France
⁷ Service d'Anatomie Pathologique, Hôpital Pontchaillou, CHU Rennes, F-35033 Rennes, France
 * Correspondence: laura.mary@chu-rennes.fr



Citation: Mary, L.; Leclerc, D.; Labalme, A.; Bellaud, P.; Mazaud-Guittot, S.; Dréano, S.; Evrard, B.; Bigand, A.; Cauchoix, A.; Loget, P.; et al. Functional Assessment of a New PBX1 Variant in a 46,XY Fetus with Severe Syndromic Difference of Sexual Development through CRISPR-Cas9 Gene Editing. *Genes* **2023**, *14*, 273. <https://doi.org/10.3390/genes14020273>

Academic Editor: Tomaiuolo Rossella

Received: 19 October 2022

Revised: 16 January 2023

Accepted: 17 January 2023

Published: 20 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Sexual development is a complex process relying on numerous genes. Disruptions in some of these genes are known to cause differences of sexual development (DSDs). Advances in genome sequencing allowed the discovery of new genes implicated in sexual development, such as PBX1. We present here a fetus with a new PBX1 NM_002585.3: c.320G>A,p.(Arg107Gln) variant, presenting with severe DSD along with renal and lung malformations. Using CRISPR-Cas9 gene editing on HEK293T cells, we generated a KD cell line for PBX1. The KD cell line showed reduced proliferation and adhesion properties compared with HEK293T cells. HEK293T and KD cells were then transfected plasmids coding either PBX1 WT or PBX1-320G>A (mutant). WT or mutant PBX1 overexpression rescued cell proliferation in both cell lines. RNA-seq analyses showed less than 30 differentially expressed genes, in ectopic mutant-PBX1-expressing cells compared with WT-PBX1. Among them, *U2AF1*, encoding a splicing factor subunit, is an interesting candidate. Overall, mutant PBX1 seems to have modest effects compared with WT PBX1 in our model. However, the recurrence of PBX1 Arg107 substitution in patients with closely related phenotypes calls for its impact in human diseases. Further functional studies are needed to explore its effects on cellular metabolism.

Keywords: DSD; sexual development; genome editing; RNA-seq; CRISPR-Cas9

1. Introduction

The differences of sexual development (DSDs) are congenital variations characterised by a discrepancy between chromosomal, gonadal, and genital sex development. DSDs are classified in three categories according to the Chicago consensus [1]: sex chromosome DSDs, 46,XY DSDs, and 46,XX DSDs. In 46,XY DSDs, gonadal dysgenesis represents 1:10,000 births [2] and are characterised by disorders of testis formation and maintenance. The establishment of the bipotential gonad and its determination toward testis rely on complex signaling pathways. Briefly, the switch from a bipotential gonad to a testis relies on *SRY*, and its effector, *SOX9*. *NR5A1*, *WT1*, and *GATA4* also directly participate in *SRY* activation [3]. Other genes such as *DHH*, *DMRT1*, or *DAX1*, and more recently *DHX37* and *PPP2R3C*, also play a role in activating pathways leading to testis formation

or repressing pro-ovarian signals [4,5]. Both copy number variants (CNVs) and a single nucleotidic variant (SNVs) have been found to impact these genes or their regulatory elements in patients presenting with isolated or syndromic 46,XY complete or partial gonadal dysgenesis. However, despite growing advances in deciphering the complexity of sex development, almost half of DSD patients does not have a molecular diagnosis [6]. Routine exome or genome sequencing (WES/WGS) combined with functional studies will help in highlighting new causes of DSDs. Focusing on syndromic DSDs may lead to the discovery of new ubiquitous pathways implicated in sexual development.

In this context, *PBX1* appears as a new gene implicated in syndromic DSDs. *PBX1* encodes a homeodomain transcription factor (TF), a protein of the TALE (three amino acid loop extension) TFs family. Its structure is composed of two domains (PBC-A and PBC-B) for dimerization with partner proteins (other TFs, from the TALE family as PKNOX or MEIS proteins, but also various other partners, such as HOX proteins, for example, [7]), and a homeodomain devoid to DNA binding. *PBX1* needs heterodimerization with its partners to change conformation and to migrate to the nucleus. First described in rare forms of acute leukemia through the formation of a fusion transcript with E2A, *PBX1* proved to be a protein implicated in various forms of cancer but also as a key factor of embryonic development [8]. Recently, *PBX1* deletions or SNVs have been implicated in the syndromic congenital anomalies of the kidneys and urinary tract (CAKUTHED), accompanied with various other developmental anomalies [9,10].

Until recently, *PBX1* variants were only reported in CAKUTHED patients. Interestingly, some patients (especially 46,XY individuals with *PBX1* missense mutations) presented with DSDs ranging from cryptorchidism and a micropenis to complete sex reversal, a phenotype that is not fully recalled by murine *PBX1*-KO models [7]. The role of *PBX1* in sexual development is not yet fully understood. In mice, *Pbx1* KO affects gonadal development in both sexes, due to impaired cell proliferation, and XY mice are sex reversed [11,12]. *Pbx1* acts downstream of Sf-1 (encoded by *Nr5a1*) in mouse gonads, but the link between these two proteins remains elusive. Moreover, *Pbx1*-KO animals present with Müllerian agenesis, a phenotype that recalls the uterine variations observed in *PBX1*-mutated patients [7]. Some authors suspected *WNT9B*, *HOXA10*, and *EMX2* were *PBX1* downstream effectors and were responsible for these Müllerian anomalies observed in patients [13,14].

Two mutation hotspots were reported in *PBX1* patients: one between the two dimerization domains (PBC-A and PBC-B) and one in the homeodomain nuclear localization signal, this late hotspot containing two recurring variants, namely p.(Arg234Pro/Gln) and p.(Arg235Gln) (Figure 1). These variants have been explored through functional studies [15,16]. Both studies demonstrated impaired transactivation and dimerization properties in cellular models.

We report here the case of a stillborn fetus with a syndromic DSD phenotype carrying a substitution at *PBX1* Arg107 (NM_002585.3: c.320G>A, NP_002576.1:p.(Arg107Gln)). In this work, we assessed the functional impact of this variant by generating a cellular model knocked-down (KD) for *PBX1* and performing rescue experiments by transfection with *PBX1*-WT or *PBX1*-Arg107Gln coding plasmids.

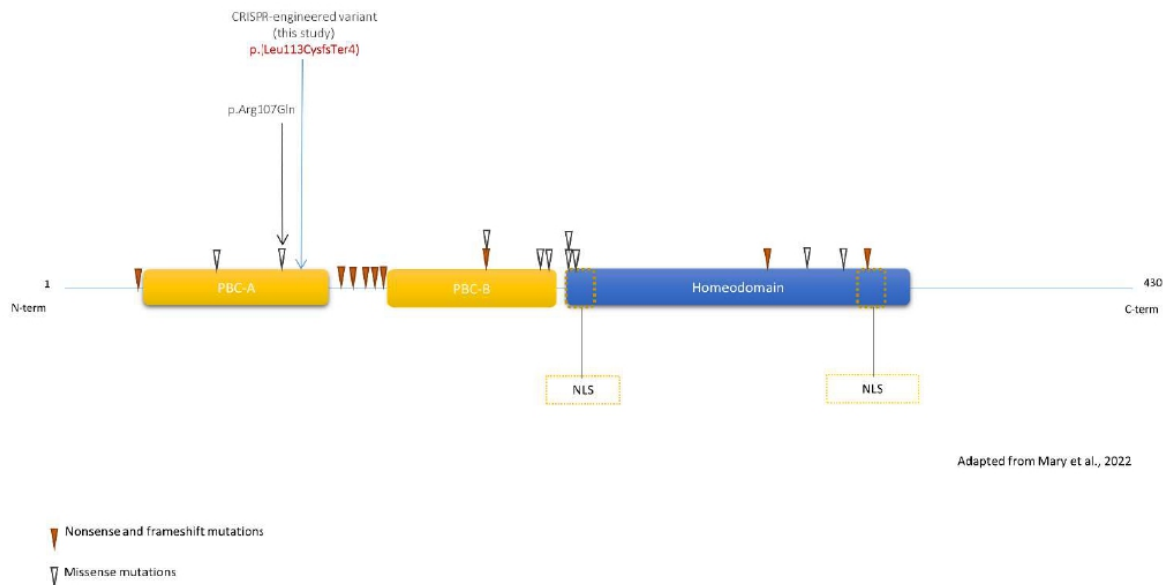


Figure 1. Schematic view of PBX1 with previously published variants. Blue arrow represents the CRISPR-engineered frameshift variant in clone KD2. Isoform: PBX1a - NM_002585.4 - NP_002576.1. All variants impact all PBX1 isoforms, except p.(Thr88Ile)(absent from NP_001340059). Yellow dotted boxes: Nuclear Localization Signal (NLS).

2. Patient and Methods

2.1. Editorial Policies and Ethical Considerations

Written informed consent was obtained from the couple. All procedures were in accordance with the ethical standards of the Ethics Committee of Rennes University Hospital and the French law.

2.2. Patient

The patient was first addressed prenatally at 32 gestation weeks (GW) because of suspected isolated unilateral right kidney agenesis at ultrasound (US) examination. At 39 GW, US showed oligohydramnios and intrauterine growth retardation (IUGR). C-section was decided due to fetal arrhythmia. Apparent death was noticed at expulsion and the child could not be resuscitated. Standardized autopsy procedure was performed, including skeletal X-rays, macroscopic examinations, and histology analyses.

2.3. Cytogenetic Analyses

Chromosomal microarray (CMA) was performed using the Agilent Human Genome CGH microarray 180K, with a genome-wide median probe of 13 kb (Agilent Technologies, Santa Clara, CA, USA). The experiment was performed according to the manufacturer's protocol. Patient's gDNA were obtained from frozen tissues using a Prepito instrument (PerkinElmer Inc., Turku, Finland), or from cell culture as detailed below (Section 2.6). A graphic overview and analysis of the data were obtained using Agilent software. The probes were mapped using the GRCh37 genome version.

FISH analyses were performed on FFPE samples from both gonads using the Vysis CEP(X)/CEP(Y) commercial probe (Abbott, Abbott Park, IL, USA) targeting chromosomes X and Y centromeres, according to the manufacturer's protocol.

2.4. Exome Sequencing and Bioinformatics Pipeline

Trio exome sequencing was performed at Lyon University Hospital. DNA was extracted from parental blood samples using the Hamilton automate machine and from

frozen fetal tissue using a Prepito instrument (PerkinElmer). Exome DNA library was prepared with the MedExome preparation kit (Roche). High-throughput sequencing was performed on a NextSeq500 sequencer (Illumina). The BWA-MEM algorithm v.0.7 (Li & Durbin, 2009) was used to map the reads on the reference genome (GRCh37/hg19). The variant calling and variant predictions was performed according to GATK and FreeBayes best practices using a local pipeline (Papillyon). Variant prioritization was performed according to these predictions and information available on OMIM and ClinVar databases. Databases and in-silico prediction tools aggregators (Varsome, Mobidetails) were used to assess the pathogenicity of the variants.

2.5. Cell Culture

HEK293T (ATCC, further called “HEK cells” in this manuscript) cells were cultured on DMEM (Gibco) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. All cells were maintained at 37 °C in a humidified incubator with 5% CO₂.

2.6. Cell DNA Extraction, PCR and Sanger Sequencing

DNA was extracted from frozen cells pellets with the Nucleospin tissue DNA extraction kit (Macherey-Nagel). PCR was performed from 50 ng DNA with the Phusion™ High-Fidelity DNA Polymerase kit (Thermo Scientific). The list of PCR primers is present in Table S1. Sanger sequencing was performed according to standard protocols after PCR quality check on agarose gel.

2.7. RNA Extraction, Reverse Transcription and Quantitative PCR

RNA extraction was performed from frozen cells in culture plates or on frozen tissue samples (spleen, liver) by using the NucleoSpin RNA kit (Macherey-Nagel). Patient’s RNA quality was assessed using a Bioanalyzer 2100, Eukaryote Total RNA Nano chip (Agilent, Santa Clara, CA, USA). Quantitative PCR (qPCR) was performed after reverse transcription using the High-Capacity cDNA Reverse Transcription kit and the SYBR Green PCR Master Mix (Applied Biosystems), according to the manufacturer’s protocol. The primer sequences used are available in Supplementary Table S1. Raw data were extracted with QuantStudio Design and Analysis Software (Applied Biosystems) and 18S was used as a housekeeping gene for normalization.

2.8. Protein Extraction

Proteins were extracted from HEK cells with RIPA buffer (Thermo Scientific) supplemented with protease and phosphatase inhibitor. Protein quantification was performed with Pierce BCA protein assay kit (Thermo Scientific) and absorbance was measured at 562 nm on a microplate reader.

2.9. Western Blot

Protein samples were loaded in 4 to 12% Bis-Tris Gel (NuPAGE, Invitrogen) after denaturation and reduction according to the manufacturer’s protocol. Migration was performed for 1 h 30, 200 V, 400 mA. After transfer and incubation with a blocking solution (1 h, room temperature, TBS-Tween 0.1%; 5% Bovine Serum Albumin), the membrane was tagged with the primary antibody overnight at 4 °C (rabbit anti-PBX1 antibody, HPA003505-Sigma-Aldrich, dilution 1/500, mouse anti-HSC70, #7298, Santa Cruz, dilution 1/2000) and then 1 h at room temperature with the HRP linked secondary antibody (HRP-linked anti-rabbit IgG, #7074 Santa Cruz, dilution 1/2000; HRP-linked anti-mouse IgG, #7076 Santa Cruz, dilution 1/2000). Uncropped Western blots are available in Supplementary Figure S1.

2.10. Methylene Blue Viability Assays

HEK were seeded in 96-well plates (15,000 cells/well). Approximately 24 h after seeding, cells were used for determining viability threshold for either puromycin treatment, PBX1 plasmid, or siRNA transfections. For puromycin viability assays, cells were treated

with increasing puromycin concentrations ranging from 0.1 ng/ μ L to 16 ng/ μ L in 100 μ L of medium/well. For plasmid viability assays, cells were transfected with either PBX1a-WT or PBX1a-320G>A plasmid concentrations ranging from 0.1 ng/ μ L to 0.5 ng/ μ L in a solution of OptiMEM medium (Gibco) and 2% of Lipofectamine 2000 (ThermoFisher Scientific). For siRNA viability assays, cells were transfected with either control or PBX1 siRNA concentrations ranging from 5 nM to 100 nM in a solution of OptiMEM medium (Gibco) and 2% of Lipofectamine RNAimax (ThermoFisher Scientific). After 2 days of treatment, cells were rinsed, fixed in 95% ethanol for 30 min and dried. Methylene blue was added to the cell pellet for 30 min. The plate was then washed and dried, and the cell pellets were suspended in HCl. Absorbance was measured with microplate reader at 620 nm.

2.11. siRNA Transfection

HEK were seeded in 6-well plates (500,000 cells/well). Approximately 24 h after seeding, cells were transfected with 25 nm siRNA CTRL (#230285603, IDT) or siRNA *PBX1* (IDT, sequence available in Table S1) with Lipofectamine RNAimax (ThermoFisher Scientific) according to the manufacturer's instructions.

2.12. Cloning

Single guide RNAs (sgRNA) were designed according to the recommendations of [17], Table S1. BPK1520 backbone (Addgene plasmid #65777) was used for integration of sgRNAs DNA sequence by golden gate assembly, as previously described [18]. Plasmids were transformed into NEB[®] Stable competent *E. coli* C3040H (New England BioLabs) by following the manufacturers' recommendations. Bacteria were cultured according to [19]. Plasmid DNA was purified with NucleoBond[®] Xtra Maxi kit (Macherey Nagel). Sanger sequencing was performed for construct validation.

2.13. CRISPR-KD

HEK were seeded in 6-well plates (500,000 cells/well). General protocol for CRISPR transfection and selection was described elsewhere [19]. Transfection was performed using a mix of pSpCas9(BB)-2A-Puro plasmid (PX459#62988, Addgene, gift from F. Zhang, [19]), plasmid encoding the sgRNA (see "Cloning" section) and a control plasmid encoding GFP (PMirZIP #25037, Addgene). Three different sgRNA were tested. Transfections with the empty BPK1520 backbone + PMirZIP +/- pSpCas9(BB)-2A-Puro plasmids were used as negative controls. Puromycin was used at a concentration of 8 μ g/mL (determined through methylene blue viability studies, see above) to select Cas9-puro-transfected cells. Limit dilution of transfected cells was performed as described by [19]. Wells containing only one clonal colony after 2–3 weeks of culture were harvested and amplified in 24-well plates. When confluency was reached, DNA from each clone was extracted for Sanger sequencing (see above) to determine the clone's genotype. Clones showing frameshift variations located at the sgRNA's target sequence (Supplementary Figure S2) were amplified in 6-well plates in duplicates, and PBX1 synthesis defect were checked through Western blot (see above). Clones presenting with both frameshift variants and lower PBX1 signals at Western blot compared to a positive control were used for rescue studies and functional analyses (proliferation studies, qPCR, RNA-seq).

2.14. Site-Directed Mutagenesis on PBX1 Plasmid

PBX1a-coding plasmid was obtained from Addgene (*PBX1a*-pCMV1 #21029, Addgene, gift from Corey Largman) and PBX1a exon 3 sequence was checked by Sanger sequencing. Site-directed mutagenesis was performed using Agilent Stratagene QuickChange XL kit (Supplementary Table S1). The reaction product was used to transform XL1-Blue supercompetent cells according to the manufacturer's protocol. Bacteria seeding and amplification, plasmid extraction, and sequencing were performed as described in the "Cloning" section.

2.15. Proliferation Studies

HEK cells and CRISPR-KD-modified cells were seeded in 6-well plates (300,000 cells/well) or in 96-well plates (20,000 cells/well) and transfected with either PMiR-ZIP and PBX1a-WT or PMiR-ZIP and PBX1a-320G>A plasmid as previously described. Untransfected cells were used as negative controls. Cell proliferation was studied after 3 days, either through methylene blue treatment (see above) for 96-well plates (negative controls and transfected cells), or through Malassez counting chamber after digestion with TrypLE™ Express enzyme (1X) and resuspension in DMEM, for 6-well plates (negative controls only). Before cell counting on 6-well plates, pictures at $\times 10$ and $\times 20$ magnification were captured using an Evos M5000 microscope (Invitrogen).

2.16. RNA-seq

RNA-seq was performed at BGI Tech (Shenzhen, China). RNA was extracted from WT HEK or KD clone after 3 days of transfection with either PBX1a-WT or PBX1a-320G>A plasmids. Transcriptome library was performed with the DNB preparation kit (BGI). High-throughput paired-end sequencing was performed on a DNBSEQ-G400 sequencer (BGI). The HISAT2 algorithm [20] was used to map the reads on the reference genome (GRCh38/hg38). Sequencing data filtering was performed using SOAPnuke algorithm [21]. Bowtie2 [22] and RSEM [23] protocol was then used to align the reads to the reference genes and quantify gene expression. Differential gene expression was estimated using the DEGseq method [24] and gene annotation was performed with GO and KEGG enrichment analyses. Differential alternative splicing and fusion gene analyses were studied using rMATS statistical model [25] and Ericscript [26], respectively. Significance thresholds for differentially expressed genes were set as follows: $|\log_2FC| \geq 1$ and False Discovery rate (FDR) ≤ 0.01 according to Poisson distribution and $|\log_2FC| \geq 0$ and q-value ≤ 0.05 according to DEGseq2.

3. Results

3.1. Patient

The patient was a deceased neonate presenting with a complex DSD with prominent genital tubercle, a urethral meatus opening at the base of the genital tubercle, unfused lateral genital folds without a gonad inside, and an absent vaginal opening. Autopsy found a small bicornuate and septated uterus connected to the first tier of a rudimentary vagina, and two pelvic gonads. The right gonad had a testis-like gross appearance. Microscopic analyses found a testis-like gonad with focal dysplasia, the presence of Leydig cells, and an epididymis (Figure 2a). The left gonad was an undifferentiated streak gonad accompanied by both a fallopian tube and an epididymis (Figure 2b). The child also presented with major lung hypoplasia (lung weight/fetal weight < 0.006 for normal values > 0.012 for fetuses > 26 GW) that could not be fully explained by late oligohydramnios, and which led to neonatal death. The autopsy did not confirm the unilateral renal agenesis but highlighted a misplaced right kidney (pelvic) and a normally placed left kidney. Both renal pelvises were anteriorly rotated. Both kidneys were hypoplastic, albeit histologically mature. X-rays highlighted a missing rib pair. Growth and histological examinations were otherwise normal, showing only non-specific hypoxic lesions.

CMA did not reveal any chromosomal imbalances associated with the phenotype and found a XY gonosomal sex. FISH on both gonads highlighted only XY cells, dismissing a gonadal dysgenesis due to XX/XY mosaicism. Trio exome sequencing revealed a heterozygous de novo *PBX1* variant: chr1:g.164761785G>A, NM_002585.3:c.320G>A, p.(Arg107Gln). The variant was predicted to be pathogenic by most of the in silico tools used. The nucleotide position was strongly conserved and variants in this region are not reported in general population databases. No other pathogenic or likely pathogenic variants in mendeliome genes were found.

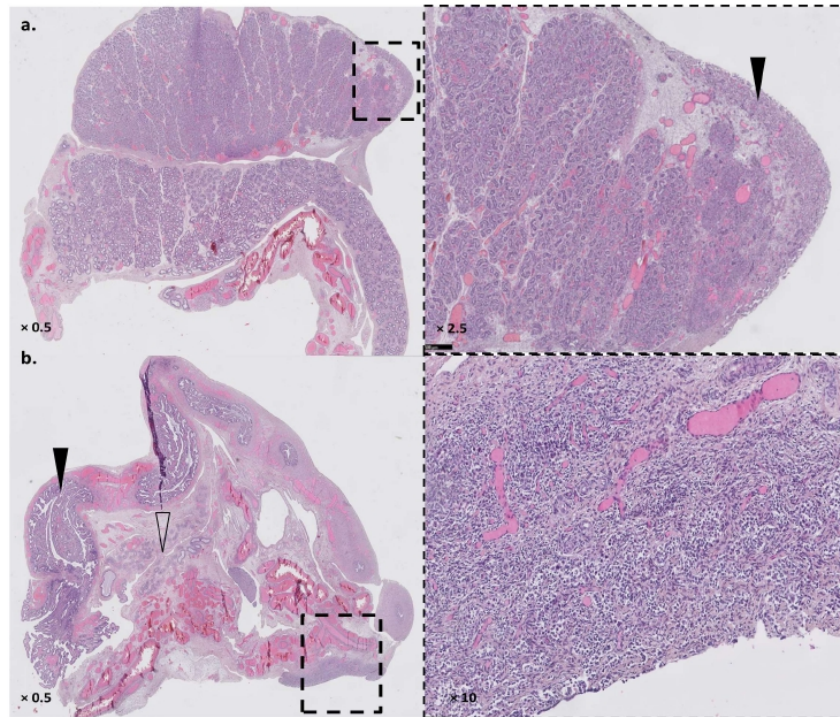


Figure 2. Gonadal phenotype of the fetus. Panel (a): right testis-like gonad with small area of undifferentiated tissue (dashed box, black arrowhead). Panel (b): left streak gonad (dashed box) with both fallopian tube (black arrowhead) and epididymis (white arrowhead). Magnification shows an undifferentiated ovarian-like tissue with rare primordial germ cells and absent albuginea.

3.2. Generation of a Cellular Model Knocked-Down (KD) for *PBX1*, with Impaired Proliferation and Support Adhesion

Due to the delay between the child's death and sample collection, no cultured cells from the patient could be obtained. RNA extractions on multiple frozen tissues (liver, thymus, spleen) only generated degraded RNA that could not be used for further explorations (RNA Integrity Number between 1.9 and 2). We used HEK cell line in order to generate clonal cell lines KO for *PBX1* through CRISPR-KD genome editing. Three single guide RNA (sgRNA_1 to 3) were used in the experiments. After CRISPR editing, limit dilution, and 2–3 weeks of culture, 32, 27, and 13 clones were obtained for each sgRNA, respectively. All were amplified and Western Blot (WB) was performed to determine the edited clones. No complete abolition of *PBX1* expression (i.e., KO clones) was observed (Figure 3a and Supplementary Figure S1). Clones with diminished *PBX1* expression were then Sanger-sequenced at the sgRNA's target sequence to assess their genotype (Figure 3b and Supplementary Figure S2). siRNA against *PBX1* was used to validate the WB results (Figure 3a). One clone (clone KD2) with a clear *PBX1*-diminished signal (Figure 3) was retained for further analyses. PCR-sequencing showed an 8-bp insertion in clone KD2, responsible from a frameshift and a premature stop codon *PBX1* (NM_002585.4):c.337_338insGCCGCAAC, (p.Leu113CysfsTer4), suggesting nonsense-mediated decay (NMD) as a mechanism explaining the KD. Sanger sequencing showed a residual detection of WT *PBX1* exon 3 sequence, suggesting that one of the three *PBX1* alleles (HEK293T are pseudo-triploid) was non-edited. This explained the residual *PBX1* expression at Western Blot.

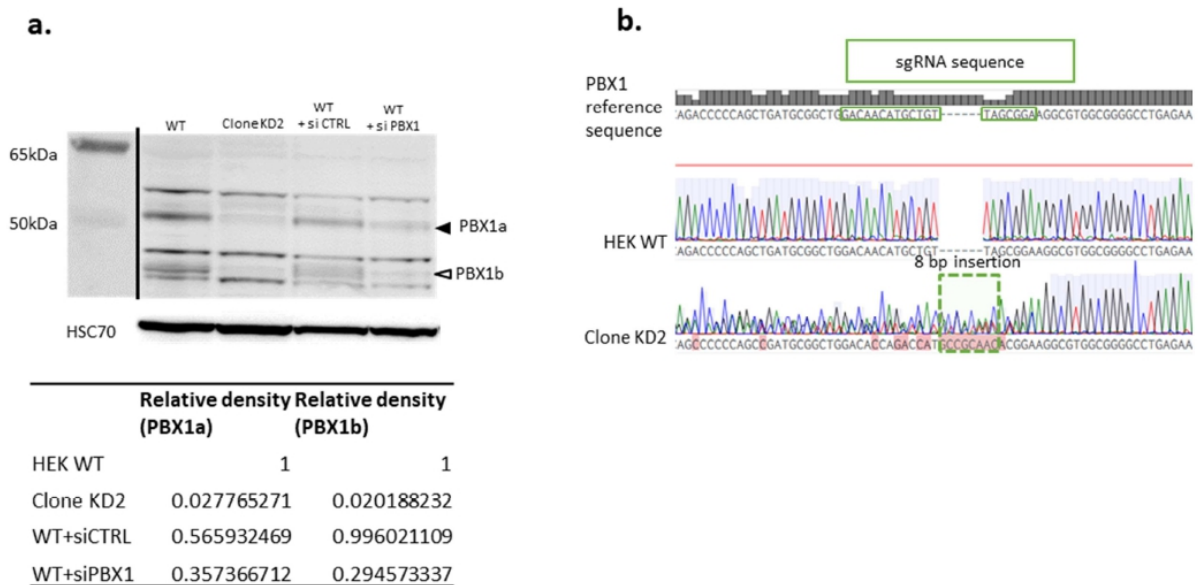


Figure 3. Obtention of a *PBX1* KD clone. Panel (a): Western Blot demonstrating the diminution of both PBX1a (black arrowhead) and PBX1b (white arrowhead) expression in the *PBX1* KD clone (clone KD2). WT HEK cells transfected with control siRNA (WT + si CTRL) or *PBX1* siRNA (WT + si *PBX1*) served as negative and positive controls, respectively. WT: non-transfected HEK cells. Panel (b): Sequence of *PBX1* exon 3 in WT HEK cells (HEK WT) and *PBX1* KD clone (clone KD2) showing an 8-bp insertion (dashed box) at the sgRNA's target sequence (green box) in KD2. Uncropped Western Blot is presented in Supplemental Data (Figure S10).

Compared with WT HEK cells, clone KD2 showed significant proliferation impairment (Figure 4a and Supplementary Figure S3) and grew as colonies instead of growing as a monolayer (Figure 4d, panel (1)). Clones KD2 detached from the flask bottom as soon as 3 days of culture and before reaching confluency, suggesting cell–cell and cell–support adhesion anomalies. aCGH results showed a few differential imbalances between clone KD2 and WT cells (Figure S11).

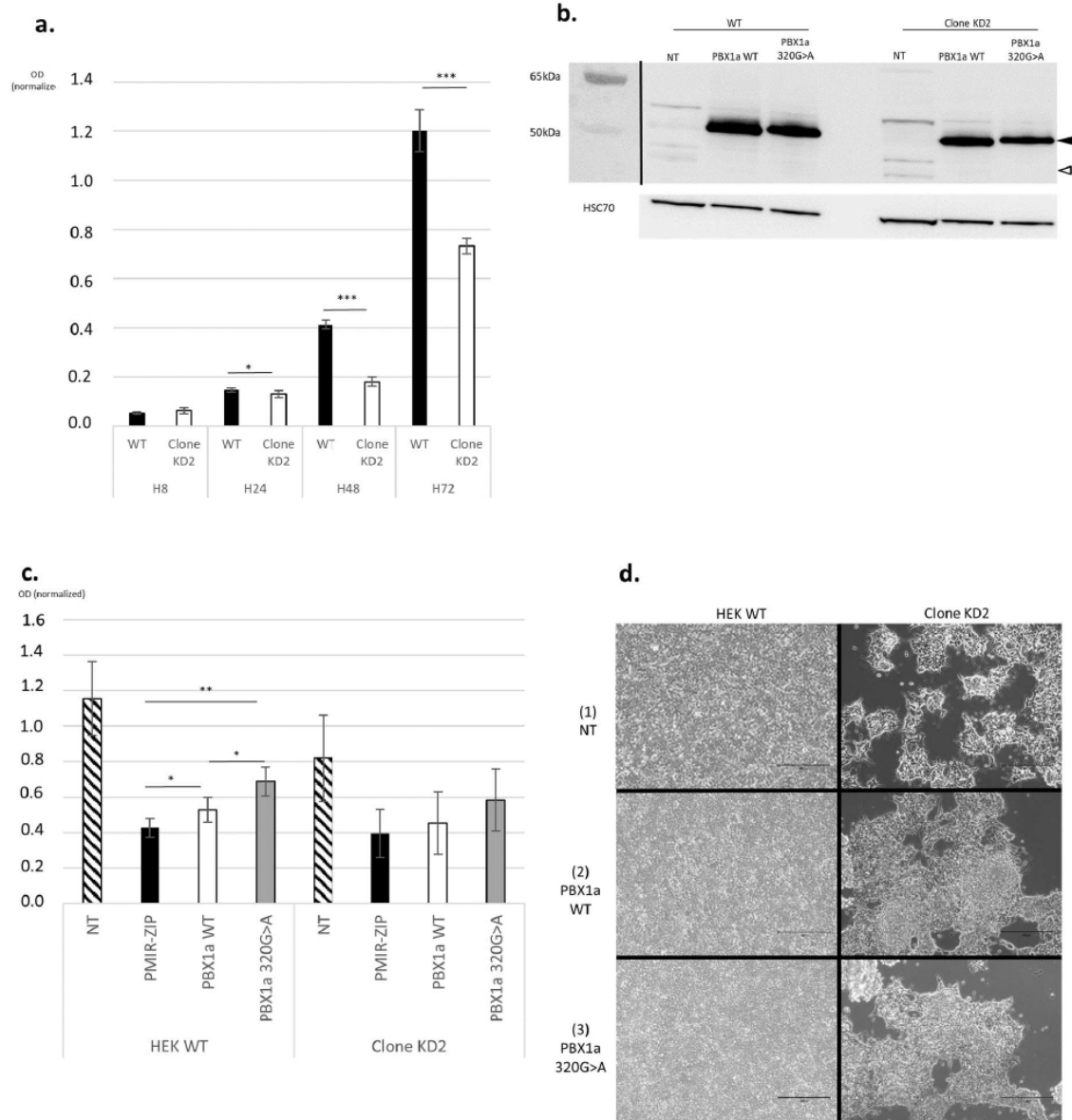


Figure 4. PBX1 partial LoF impairs cell proliferation, not fully rescued by *PBX1a* transfection. (a) *PBX1* LoF impairs cell proliferation (Methylen blue test). OD: optical density. H: hours. Bars represent standard deviations (SD). (b) WB assessing *PBX1a* overexpression (black arrowhead) in transfected HEK WT cells (HEK WT) and KD2 clones. White arrowhead: *PBX1b*. (c) Transfection with either *PBX1a* WT or 320G>A plasmids enhances proliferation in WT HEK cells in comparison with control (PMiR-ZIP, a GFP-encoding plasmid), but not in *PBX1* KD cells, after 3 days of culture. Bars represent SD. (d) *PBX1* KD clones are unable to grow as a monolayer and reach confluency, a phenotype slightly rescued by both WT and mutant *PBX1a* overexpression. For each quantitative experiment, 3 biological replicates were realized. Each biological replicate comprised 3 technical replicates per sample. * = *t* test between 0.01 and 0.05, ** = *t* test between 0.001, and 0.01, *** = *t* test between 0.0001 and 0.001.

3.3. Rescue Experiments by Transfection with *PBX1a*-WT or *PBX1a* 320G>A Plasmids

Both WT HEK (further called WT HEK) cells and clone KD2 cells were transfected with a plasmid encoding either WT *PBX1a* or the *PBX1a* 320G>A (further named “mutant *PBX1a*”) variant of interest.

Transfection efficiency was assessed through WB and both *PBX1*-coding plasmids led to *PBX1* overexpression in all cell types (Figure 4b and Supplementary Figure S10).

3.3.1. *PBX1* Overexpression Enhances Cell Proliferation

Transfection with WT *PBX1a* plasmid significantly enhanced cell proliferation in WT HEK compared with a transfection control (Figure 4c). Transfection of mutant *PBX1a* also significantly increased cell proliferation compared with control. This proliferation enhancement through the mutant *PBX1a* transfection was also significant compared with WT *PBX1a* transfection. A similar pattern of proliferation enhancement was also observed in clone KD2 with both WT and mutant *PBX1a* plasmids, albeit non-significantly (Figure 4c).

3.3.2. Overexpression of *PBX1a* Slightly Rescues KD2 Normal Adhesion

Non-transfected KD2 cells show impaired cell–cell and cell–support adhesion capacities (Figure 4d, panel (1)). Transfection with either WT or mutant *PBX1a* slightly rescued the adhesion phenotype of KD2 cells, with larger and thinner clusters of cells (Figure 4d, panels (2) and (3)). However, early detachment was still observed before confluency. On the contrary, transfection with WT or mutant *PBX1a* plasmid in WT HEK cells did not modify their adhesion phenotype (Figure 4d, panels (2) and (3)).

3.3.3. Targeted RT-qPCR Suggests *MEIS1* as a Potential Disrupted Gene

RT-qPCR was then performed to explore *PBX1* target genes modulations in both cell types. Target genes (*CCND1*, *MEOX1*, *FZD2*, *ZHX2*, *AR*, *NFE2L1*) were chosen according to [27,28] and depending on their expression profile in WT HEK cells (based on Human Protein Atlas data). *PBX1* cofactors *MEIS1* and *PKNOX1* were also studied. For each cell type, WT *PBX1a* plasmid transfection was considered as a control condition. *PBX1* cofactor *MEIS1* was less expressed in WT HEK cells transfected with mutant *PBX1a*, while expression was similar in all transfected KD2 cells. Other gene expressions were not significantly modified (Supplementary Figure S4).

3.4. RNA-seq

To further assess RT-qPCR preliminary results, we performed RNA-seq on WT HEK and clone KD2 transfected with either WT or mutant *PBX1a* plasmids. About 44 million 100 bp-long paired-end reads per sample were generated. The average mapping ratio with reference genome was 96.14%, the average mapping ratio with gene was 78.27%, 16654 mRNA-coding genes were identified.

3.4.1. Both Cell Lines Expressed Multiple *PBX1* Transcripts

Focusing first on *PBX1* expression, we showed that it was expressed at high levels in all four samples, albeit slightly significantly less expressed in KD2 cells compared with WT HEK cells (Supplementary Figure S5). The variation c.320G>A was found in all *PBX1* transcripts in cells transfected with mutant *PBX1a* plasmid (Supplementary Figure S7). Both WT HEK cells and clone KD2 cells expressed a *PBX1a* transcript. Transcript XM_017001395, which was close to, but not identical to, *PBX1a* (use of an alternate exon 9) was also highly expressed. However, the alternate exon 9 has a sequence closely related to *PBX1a* exon 9 and XM_017001395 represents a predicted transcript, which was not assessed through previous functional analyses. Moreover, the alternate exon 9 sequence was absent from our plasmid. This late exon 9 was poorly covered (<20 reads per sample). Thus, it may represent falsely aligned *PBX1a* transcripts. *PBX1b* transcript (NM_001204961) was slightly expressed in both cell lines, albeit significantly less in clone KD2. HEK cells produce at least

four other *PBX1* transcripts (Figure 5, Table 1), including a transcript with an alternative exon 9, which was absent from KD2 cells.

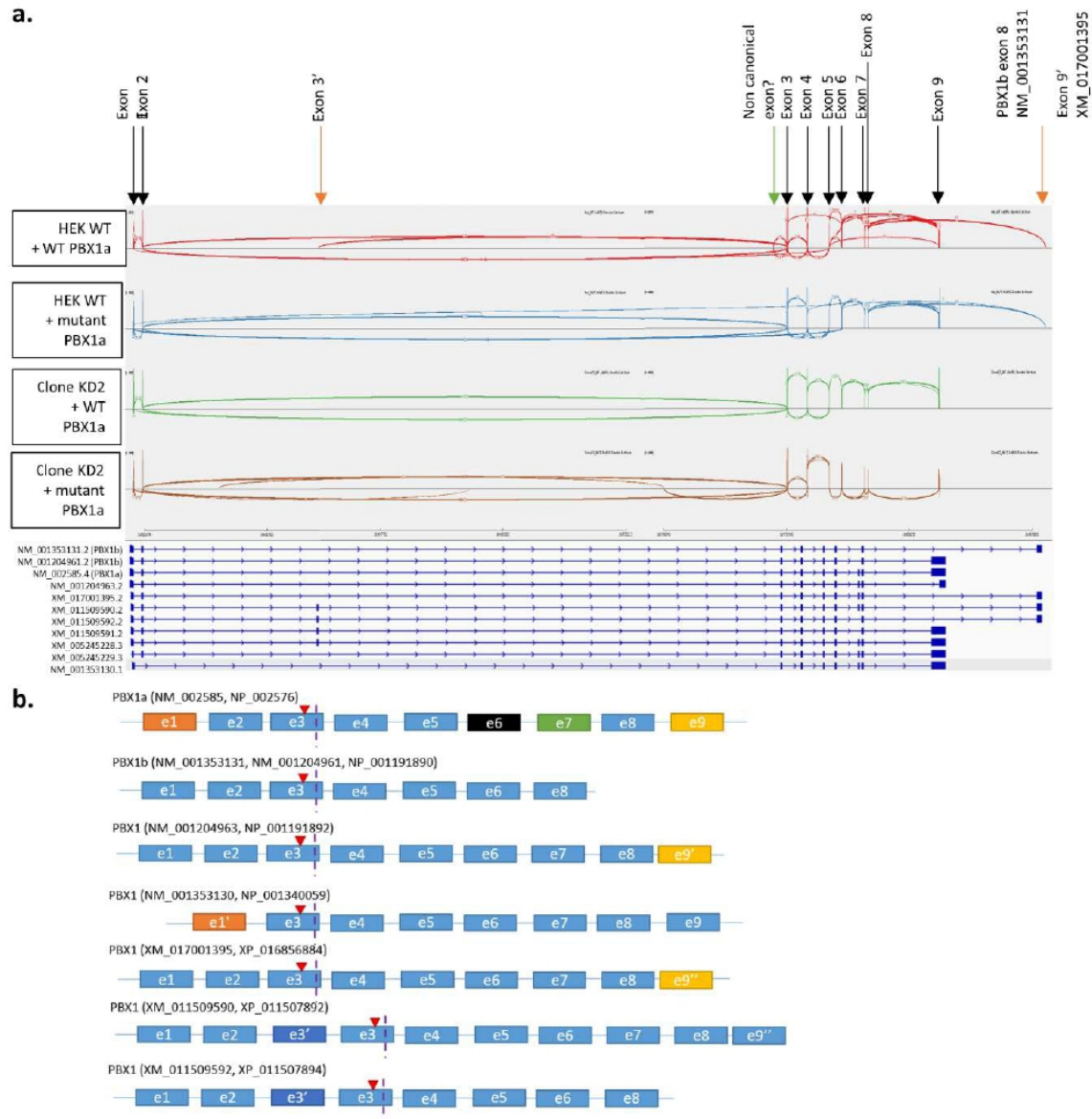


Figure 5. *PBX1* transcripts in HEK cells and clone KD2. (a) Sashimi plot of *PBX1* in HEK cells + WT *PBX1a* (1), HEK cells + *PBX1a*-320G>A (2), clone KD2 + WT *PBX1a* (3), and clone KD2 + *PBX1a*-320G>A. Exon junction coverages < 5 are hidden for lisibility. Black arrows represent exons of *PBX1a* isoform (NM_002585). (b) Schematic view of *PBX1* transcripts. Red triangle represents the location of the c.320G>A mutation. Vertical dotted purple line represents the location of the CRISPR-generated frameshift insertion.

Table 1. *PBX1* isoforms expressed in transfected cells. WT HEK+WT/mutant: WT HEK cells transfected with WT or mutant *PBX1a* plasmid. Clone KD2 + WT/mutant: KD2 cells transfected with WT or mutant *PBX1a* plasmid. TPM: transcript per million.

PBX1 Isoform	WT HEK + WT TPM	WT HEK + Mutant TPM	Clone KD2 + WT TPM	Clone KD2 + Mutant TPM
NM_001204961 = PBX1b	1.60	2.56	0.22	1.10
NM_001204963	NA	NA	NA	NA
NM_001353130	0	0.31	0	0.16
NM_001353131 = PBX1b	1.91	1.79	0	1.07
NM_002585= XM_005245229 = PBX1a	1208.94	976.54	447.82	408.52
XM_005245228	NA	NA	NA	NA
XM_011509590	0	0	0	0.25
XM_011509591	0	0	0.46	0
XM_011509592	1.47	0.18	0.31	0
XM_017001395 (<i>PBX1a</i> ?)	2665.33	2435.52	1155.85	887.08

3.4.2. Gene Expression between WT HEK and KD2 Cell Lines Remains Different despite Being Transfected with the Same Plasmid

RNA-seq results are summarized in Supplementary Table S2. Scatter plots showing DEGs between each comparison group are given in Figure 6.

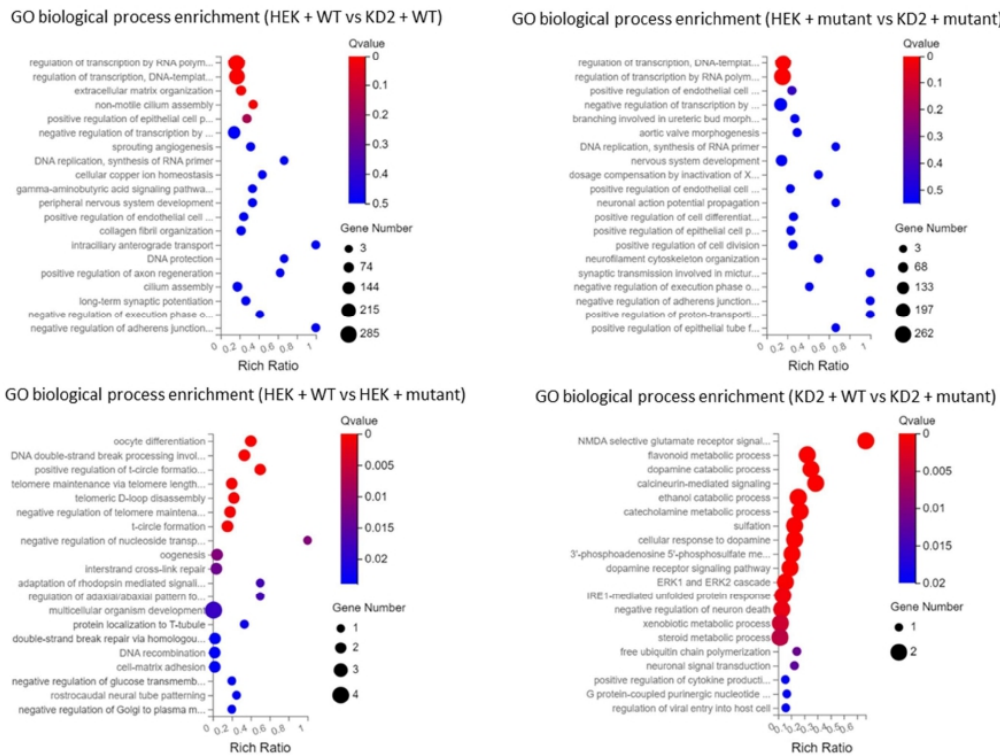


Figure 6. Bubble charts of GO biological processes enrichment in each comparison group. HEK + WT/mutant: WT HEK cells transfected with WT or mutant *PBX1a* plasmid. Clone KD2 + WT/mutant: KD2 cells transfected with WT or mutant *PBX1a* plasmid.

When we compared mutant *PBX1a*-transfected cells to WT *PBX1a*-transfected cells (control vs. case analysis), only one gene (U2 small nuclear RNA auxiliary factor 1, *U2AF1*) was significantly downregulated in mutants (log₂ fold change between case and control: −23.45). The WT-transfected cells and the mutant-transfected showed in fact strong differences on the PCA plot, explaining the fact that only one gene was significantly deregulated when grouping cells in control vs. case analyses (Supplementary Figure S6). *U2AF1* was only expressed in WT *PBX1a*-transfected HEK cells (TPM at 56.94 for WT-transfected WT HEK cells and was null for the other three samples, $FDR^{(\text{Clone KD2_WT/HEK_WT})} = 1.33 \times 10^{-116}$ and $FDR^{(\text{HEK_MUT/HEK_WT})} = 7.46 \times 10^{-120}$).

To assess this result, we first explored the differentially expressed genes (DEGs) between WT HEK cells and KD2 cells transfected with WT *PBX1a* plasmid. We found 2118 DEGs between these two groups (1219 downregulated genes and 899 upregulated genes). The most significantly down-regulated biological processes in KD2 cells ($q\text{-value} \leq 0.05$) according to GO and KEGG classifications were the regulation of transcription (more than 200 genes, including *EMX1*, *GATA4*, *SOX8*, *AR*, *HOXA4*, *HOXA11*, *HOXB5*, and *HOXD4*, *MYC*, and *SMAD* genes) and non-motile cilium assembly (18 genes, including *BBS1*, *CEP290*, and *PCMI*). Upregulated functions included extracellular matrix organization (55 genes, including collagens I to VI subunits, *CCN1* and *ICAM* genes), and plasma membrane repair. DEGs between cell lines transfected with mutant *PBX1a* plasmid were similar (1985 DEGs), but the only significantly downregulated cellular processes were the regulation of transcription (more than 200 downregulated genes in KD2 cells, including all the abovementioned genes and *SOHLH2*). Plasma membrane repair genes also remained upregulated (Figures 6 and 7). Thus, WT HEK and clone KD2 are highly different, a difference in transcriptome that remained whatever the plasmid transfected. Such a difference explains the difficulties to find DEGs in a control vs. case analysis. These differences may in fact reflect the consequences of the CRISPR-engineered frameshift variant in KD2 cells.

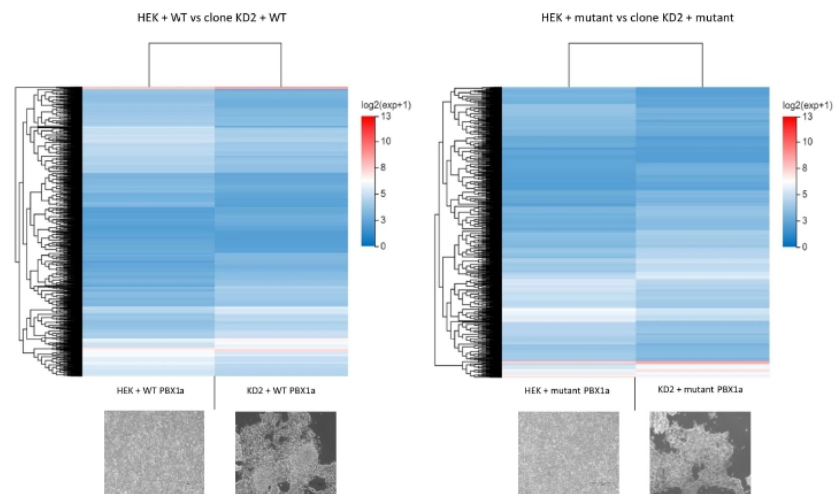


Figure 7. Heatmaps comparing the differential expression between WT HEK and *PBX1* KD cells transfected with similar *PBX1* plasmids. HEK + WT/mutant: WT HEK cells transfected with WT or mutant *PBX1a* plasmid. Clone KD2 + WT/mutant: KD2 cells transfected with WT or mutant *PBX1a* plasmid.

3.4.3. DEGs between WT and Mutant *PBX1a* Transfections Depend on the Cell Line

We then studied DEGs between WT- and mutant-transfected cells without clustering both cell types (Figure 8). We only found a few DEGs in WT vs. mutant experiments. In total, 23 DEGs were identified for HEK cells, implicated in various cellular functions according to GO (oocyte differentiation, DNA double strand break repair, telomere maintenance, regulation of nucleoside transport, Figures 6 and 8). Ten DEGs were identified in KD2 cells,

impacting dopamine and steroid metabolism and cytokine production. In such analysis, only the *ATP5MF-PTCD1* gene was downregulated in both cell types.

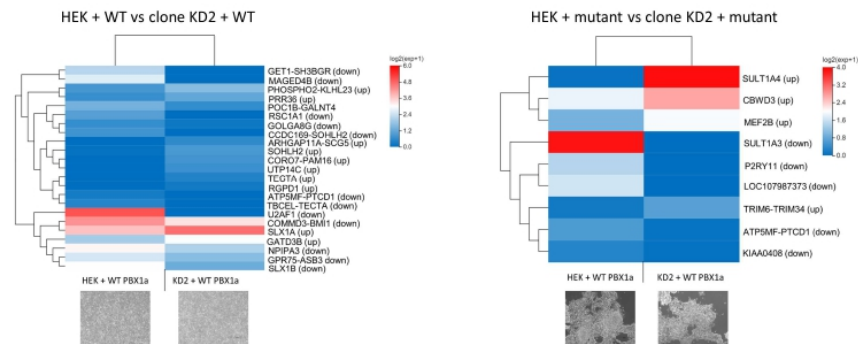


Figure 8. Heatmaps comparing the differential expression between a cell line transfected with WT or mutant PBX1 plasmids. HEK + WT/mutant: WT HEK cells transfected with WT or mutant PBX1a plasmid. Clone KD2 + WT/mutant: KD2 cells transfected with WT or mutant PBX1a plasmid. Up/down: the gene is up/down regulated in cells transfected with mutant plasmid comparing with cells transfected with WT plasmid.

These DEGs differed from preliminary RT-qPCR results. *CCND1* and *AR* were upregulated in WT *PBX1a*-transfected KD2 cells vs. WT *PBX1a*-transfected HEK cells (FDR 8.4×10^{-78} for *CCND1* and 8.4×10^{-81} for *AR*). *AR* was also upregulated in mutant *PBX1a*-transfected KD2 cells vs. WT *PBX1a*-transfected KD2 cells (FDR 1.4×10^{-71}). *MEIS1* expression was not significantly deregulated.

3.4.4. Transfection with Mutant PBX1a May Alter Skipped Exon Splicing Events

Since *U2AF1*, which encodes for a subunit of the splicing factor U2AF that recognizes the pyrimidine-rich tract present at metazoan 3-prime splice sites, was predicted to be downregulated in mutant-transfected cells compared with WT-transfected ones, we then focused on differential alternative splicing events (DASE) among groups. Transfection of mutant plasmid did not significantly modify the splicing events compared with WT plasmid, except in clone KD2, where a diminution of skipped exons events was observed. On the contrary, we observed a large number of DASE when comparing WT HEK cells and clone KD2, with a dramatic diminution of all splicing events in clone KD2 cells compared with WT HEK cells regardless of the plasmid transfected. Most of these late DASE were skipped exons (Supplementary Figure S8).

4. Discussion

4.1. *PBX1* Is a DSD Gene

Recently, Arts et al. [29] reported two stillbirth siblings of both chromosomal sexes presenting with the most severe sides of *PBX1*-linked developmental diseases: *CAKUTHEd*, diaphragmatic eventration, lung hypoplasia, and sex reversal in the 46,XY fetus. Both fetuses carried a new *PBX1* NP_002576.1:p.(Arg107Trp) variant. These phenotypes strongly recall the post-mortem findings observed in our patient, who also carried a similar substitution at the arginine 107. PBC-A variants are rare, and the absence of functional explorations for Arg107 missense variants prompted us to create an in vitro model expressing *PBX1* proteins with Arg107Gln substitution.

4.2. Partial *PBX1* Loss of Function (LoF) Impairs Cell Proliferation

We first aimed at generating a HEK cellular model for *PBX1* through CRISPR-Cas9 editing, but we obtained only a few viable KD clones and no KO clones. The impossibility of generating KO clones suggests that *PBX1* KO is lethal for HEK cells, as has been previously observed in murine models [12]. *PBX1* KD clones presented with a distinct phenotype from

WT cells, with a decreased proliferation rate and a distinct growth phenotype, suggesting abnormal cell adhesion and intercellular communication. KO for *Pbx1* in mouse models are always embryonically lethal and the embryos presented with defects in nephrogenic mesenchyme proliferation [12]. In a large set of cancer types (lymphoblastic leukemia, lung adenocarcinomas and squamous cancers, hepatocellular, and bladder cancers, etc.), *PBX1* amplification promotes cell proliferation and metastasis [30]. The introduction of a missense mutation in *PBX1* is sufficient to avoid tumorigenesis in gastric cancer cells [31]. Since HEK cells are derived from immortalized human embryonic kidney cells, we suggest that complete loss of *PBX1* expression is always lethal for this cell line. *PBX1* partial LoF is sufficient to induce defects in HEK cells proliferation.

4.3. Transfection with *PBX1*-Coding Plasmids Does Not Fully Rescue a *PBX1*-KD Phenotype

Transfection with plasmids containing either WT or 320G>A *PBX1a* isoform led to a significantly increased proliferation rate in WT HEK cells but not in KD cells, despite the fact that both mutant and WT plasmids were expressed at high levels in both cell lines. Cellular adhesion defects were only partially rescued in KD cells, whatever the plasmid transfected. The example of *U2AF1* showed that at least some gene expressions were not rescued by the WT *PBX1a* transfection in clone KD2. At the RNA-level, transfection with the same plasmid led to significant DEGs between WT and KD cells. Thus, contrary to what we expected, transfection of the WT *PBX1* plasmid did not lead to the reappearance of a “WT-like” phenotype in *PBX1*-KD cells, and the expression of a mutant *PBX1* protein did not impair cell proliferation in HEK cells. We suggest some hypotheses to explain these results. First, WT HEK cells expressed higher levels of *PBX1* transcripts contrary to KD cells, including specific isoforms absent from KD cells. *PBX* proteins, including *PBX1*, are sensitive to intracellular concentration [32]. Mutated *PBX1* proteins also have dominant-negative effects and are able to abolish proliferation in a cancer cell line [33]. Diminished intracellular concentration of *PBX1* associated with the production of some truncated *PBX1* proteins in KD clones may explain the absence of full rescue in this cell line. On the contrary, higher *PBX1* concentrations, even mutant proteins, may be sufficient to induce proliferation in WT HEK cells. Overall, only a few DEGs were observed between WT- and mutant-*PBX1a* plasmid transfection experiments, despite the fact that a large amount of *PBX1* proteins was produced in all cell lines. We can also suggest that co-transfection with a plasmid encoding a *PBX1* cofactor (*PKNOX1* for example) would have amplified the effects of *PBX1* plasmid transfection [15]. In fact, *PBX1* is unable to act alone and needs heterodimerization with one (or more) cofactor(s) to reach the nucleus and modify its target genes’ expression. Level-up *PBX1* concentration in cells expressing normal amounts of cofactors may lead to a subset of inactive *PBX1* monomers due to insufficient cofactor concentrations. These findings suggest that the *PBX1*-320G>A variant only has mild effects in HEK cells.

Nevertheless, DEGs between WT HEK cells and KD cells shed the light on interesting *PBX1* target genes. Numerous *HOX* genes, *GATA4*, *SOX8*, *AR*, *MYC*, and *SMAD* genes showed differential expression between WT and KD cell lines. *MYC* is a strong proto-oncogene controlling cell cycle and cell growth [33], its downregulation probably playing a role in the decreased KD cells proliferation capacities. Other DEGs could sum up, at least partially, the phenotype observed in *PBX1*-mutated patients. *HOX* genes, notably *Hoxa11* in mice, are responsible for the establishment and segmentation of the Müllerian ducts, precursors of the uterus and fallopian tubes [34]. Through *GATA4*, *PBX1* controls cardiomyocyte precursors’ proliferation and maturation [35]. *SMAD* proteins, especially *SMAD3* and *SMAD6*, also play a role in the regulation of endocardial cushion transformation [36,37]. *SOX8*, a transcription factor showing redundant functions with *SOX9*, was suggested as responsible for 46,XY DSD [38]. *AR* encodes for the androgen receptor and is responsible for the partial to complete androgen insensitivity in 46,XY patients, a syndrome that may lead to normal-appearing female external genitalia.

Despite the moderate effect of *PBX1a*-320G>A plasmid transfection, some interesting questions arose from our results. *U2AF1*, encoding one of the subunits from the splicing

factor U2AF, was downregulated in mutant-transfected WT HEK cells. This splicing factor plays an important role in targeting 3' splicing sites and in intron removal [39]. Mutations in this gene increases the risk of progression from myelodysplasia to secondary acute myeloid leukemia, through splicing dysregulation [40]. WT HEK cells transfected with the mutant PBX1a plasmid also demonstrated upregulation of *SLX1A* and downregulation of *SLX1B*. These proteins represent the catalytic subunit of the *SLX1-SLX4* structure-specific endonuclease. This complex is required for the repair of specific types of DNA lesions and is critical for cellular responses to replication fork failure [41]. KD cells transfected with *PBX1a* mutant plasmid showed upregulation of a read-through transcript, *TRIM6-TRIM34*, which is considered as a splice variant of *TRIM34*. *TRIM34* localizes to the mitochondria and mediates apoptosis through the mitochondrial pathway [42]. Thus, the expression of *PBX1a-320G>A* transcript may lead to splicing defects (albeit moderately observed in our experiments) and increased genome instability through abnormal resolution of Holliday junctions. *PBX1a-320G>A* transcript may also modify cell proliferation through an increase in *TRIM6-TRIM34* transcript. Finally, most of the DEGs observed between transfected cells are read-through transcripts, i.e., transcripts arising from inefficient 3'-end cleavage of nascent mRNAs, with the production of an mRNA containing two contiguous genes. Such phenomenon can be observed in cells under biological stress such as transfection, but also in several disease contexts including viral infections and renal cancer [43]. Some of these transcripts are known to be biologically active (such as *TRIM6-TRIM34*) but the precise role of such "fusion" transcripts remains unclear for now. Moreover, they are suspected to be overlooked in RNA-seq studies, due to the incorrect sequence assignment to one of the two parent genes [43]. Interestingly, KD cells transfected with mutant *PBX1a* showed *SULT1A3* and *SULT1A4* down- and upregulation respectively, but these two genes share exons with *SLX1A* and *SLX1B*, which were dysregulated in WT HEK cells. These two phenomena (read-through transcripts and exon sharing) raise the question of RNA-seq reliability for the detection of such transcripts. RT-qPCR experiments may be of interest to confirm, at least partly, some RNA-seq results.

4.4. *PBX1* Transfection Gives Fluctuant Results in HEK Cells

Our approaches did not directly study the transactivation effects of the p.(Arg107Gln) mutant nor the subcellular localization of the protein. Previous functional studies targeting another variant in the same cell line (p.(Arg235Gln) introduced in WT HEK cell through plasmid transfection) led to discordant conclusions on the subcellular localization of the resulting protein [15,16]. These works also demonstrated impaired transactivation [15] or heterodimerization [16] properties of *PBX1* mutant proteins. Nevertheless, the transactivation defects were discordant between a *Pbx1*-LoF murine cell line (obtained through CRISPR-Cas9 editing) and HEK cells [15]. Thus, the results observed in both studies and our work highlight the fact that HEK cells may not be the most accurate model to study *PBX1* functions through plasmid transfection. In fact, HEK cells are probably derived from immature neuron-like renal cells rather than renal epithelial cells or fibroblasts [44,45]. HEK cells are pseudotriploid and some authors suggest a high instability of their karyotype after several passages or freezing cycles, especially for HEK 293T cells. Transfection-induced stress is also known to trigger genomic instability in these cell lines [45]. The naturally occurring overexpression of *PBX1* transcripts through *PBX1* genomic amplification was dismissed through aCGH. Genomic instability was highlighted in our cell lines, with WT and KD2 cells presenting with a few different genomic imbalances. These differences may also play a role in the fluctuant transfection results observed, despite the fact they seem to be expected variations in cultured HEK 293T cells. Finally, the dominant-negative effects observed in gastric cancer cells after the introduction of a missense mutation in *PBX1* may also rely on the fact that the mutant gene was introduced through infection techniques [31], leading to a more stable expression. Overall, CRISPR-Cas9 is an efficient tool to generate *PBX1*-KD clones derived from HEK cells, but plasmid transfection may not be the most efficient way to rescue the KD phenotype.

5. Conclusions

PBX1 encodes a ubiquitous transcription factor, which can lead to severe phenotypes when disrupted. We reported here a third patient with a substitution of the Arg107 in one of the dimerization domains of the protein. CRISPR-Cas9 allowed us to generate a *PBX1*-KD clone with a cellular phenotype close to the mesenchymal defects observed in mouse embryos. We were not able to fully rescue the phenotype through WT or mutant *PBX1a* transfection, but the results suggest a disruption of splicing and DNA repair as suspected mechanisms in cells expressing the mutant *PBX1*. Further functional studies focusing on techniques other than plasmid transfection are needed to explore the effects of *PBX1* Arg107 disruption.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14020273/s1>. Figure S1. Uncropped Western Blot of clones obtained after CRISPR-KO editing (a). Clone designed by an arrow was considered closest to a *PBX1* KO (further called clone KD2 in main text). Clones in dashed boxes were re-analysed through WB (b). Arrowheads show *PBX1a* (black) and *PBX1b* (white) expected sizes. Figure S2. Sanger sequencing of clones selected in Figure S1 (dashed boxes, Figure S1a), showing a partially edited *PBX1* exon 3 sequence: at least one *PBX1* allele shows a frameshift, but the WT *PBX1* sequence persists (HEK cells are triploid). Clone 7 is not edited. Clone 25 (sgRNA_2) died before DNA could be extracted for testing. Figure S3. Significant decrease in proliferation capacities of clone KD2 compared to WT HEK cells (WT) after 3 days of culture. Two counting methods were used to avoid cell pellet detachment bias that may occur with methylene blue count. Figure S4. Relative gene expression in WT HEK and KD2 cells lines after transfection. *PBX1* WT: WT *PBX1a* plasmid, *PBX1* 320G>A: mutant *PBX1a* plasmid. *MEOX1* was not conserved for analyses due to a very low level of expression in both cell lines. Figure S5. Scatter plot of differentially expressed genes (DEGs) between samples. *PBX1* is slightly less expressed in clones KD2 compared to WT HEK cells, whatever the transfected plasmid (upside panes, arrow). For a chosen cell type, *PBX1* expression is similar between WT *PBX1a*-transfected and mutant *PBX1a*-transfected cells (downside panes, dashed arrow). HEK+WT/mutant: WT HEK cells transfected with WT or mutant *PBX1a* plasmid. Clone KD2+WT/mutant: KD2 cells transfected with WT or mutant *PBX1a* plasmid. Figure S6. PCA plot of RNA-seq samples. Figure S7. *PBX1*-320G>A-transfected cells express *PBX1*-mutated transcripts. IGV view of RNA sequence from *PBX1* exon 3 in cells transfected with WT *PBX1a*-coding plasmid (WT) or *PBX1a*-320G>A-coding plasmid (mutant). Cells transfected with the mutated *PBX1a* form only express the *PBX1a*-320G>A form at the RNA level. HEK+WT/mutant: WT HEK cells transfected with WT or mutant *PBX1a* plasmid. Clone KD2+WT/mutant: KD2 cells transfected with WT or mutant *PBX1a* plasmid. Figure S8. Differential alternative splicing (as) events among groups. HEK+WT/mutant: WT HEK cells transfected with WT or mutant *PBX1a* plasmid. Clone KD2+WT/mutant: KD2 cells transfected with WT or mutant *PBX1a* plasmid. Skipped Exon (SE), Alternative 5' Splicing Site (A5SS), Alternative 3' Splicing Site (A3SS), Mutually exclusive exons (MXE), Retained Intron (RI). (a) Number (nb) of splicing variations among samples showed a decrease in splicing variations in clone KD2, (b) Due to these large variations in total splicing events among samples, comparison between types of splicing events was performed on relative values using Z-test, with a threshold of 1.96 (α error rate of 5%, two tailed). % age: percentage. Figure S9. Uncropped Western-Blot from Figure 3 showing the *PBX1* KD phenotype of clone KD2. Western Blot demonstrating the diminution of both *PBX1a* (black arrowhead) and *PBX1b* (white arrowhead) expression in *PBX1* KD clone (clone KD2). WT HEK cells transfected with control siRNA (WT + si CTRL) or *PBX1* siRNA (WT + si *PBX1*) served as negative and positive controls, respectively. WT: non-transfected HEK cells. Figure S10. Uncropped Western-Blot from Figure 4 assessing *PBX1a* overexpression after *PBX1a*-coding plasmid transfection. Uncropped WB assessing *PBX1a* overexpression (black arrowhead) in transfected HEK WT cells (HEK WT) and KD2 clones. White arrowhead: *PBX1b*. Figure S11. Comparison of aCGH profiles between WT HEK cells and clone KD2 before plasmid transfection. Table S1. DNA primers, siRNA and CRISPR guides sequences used in this study. Table S2. Summary of RNA-seq results in WT *PBX1a*- and mutant *PBX1a*-transfected cells. DEG: differentially expressed gene, DASE: differential alternative splicing events. Table S3. Relative quantification of *PBX1a* and *PBX1b* expression on Figure S1 Western-Blot.

Author Contributions: L.M.: Investigation, Writing—original draft; D.L., A.L. (Audrey Labalme), P.B., S.M.-G., S.D., B.E., A.B., A.C., P.L., A.L. (Anna Lokchine), L.C. and D.G.: Investigation, Resources; M.-A.B.-R.: Supervision, Writing—Review, and Editing, S.J.: Conceptualization, Supervision, Writing—Review, and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study received financial support from AVIESAN Plan Cancer, Région Bretagne; Agence Nationale de Recherche, University of Rennes 1 and Ministère de la Recherche et de l'Enseignement Supérieur (MESRI, PhD Scholarship).

Institutional Review Board Statement: All procedures were in accordance with the ethical standards of the Ethics Committee of Rennes University Hospital and the French law.

Informed Consent Statement: Written informed consent was obtained from the couple.

Data Availability Statement: The variant found in the patient was recorded in ClinVar database under the number SUB12167949. RNA-seq data were uploaded on the GEO site under the number GSE215967.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hughes, I.A.; Houk, C.; Ahmed, S.F.; Lee, P.A.; LWPES/ESPE Consensus Group. Consensus Statement on Management of Intersex Disorders. *Arch. Dis. Child.* **2006**, *91*, 554–563. [[CrossRef](#)] [[PubMed](#)]
- Lee, P.A.; Nordenström, A.; Houk, C.P.; Ahmed, S.F.; Auchus, R.; Baratz, A.; Baratz Dalke, K.; Liao, L.-M.; Lin-Su, K.; Looijenga, L.H.J.; et al. Global Disorders of Sex Development Update since 2006: Perceptions, Approach and Care. *Horm. Res. Paediatr.* **2016**, *85*, 158–180. [[CrossRef](#)] [[PubMed](#)]
- Okashita, N.; Tachibana, M. Transcriptional Regulation of the Y-Linked Mammalian Testis-Determining Gene SRY. *Sex. Dev.* **2021**, *15*, 351–359. [[CrossRef](#)]
- Baetens, D.; Verdin, H.; De Baere, E.; Cools, M. Update on the Genetics of Differences of Sex Development (DSD). *Best Pract. Res. Clin. Endocrinol. Metab.* **2019**, *33*, 101271. [[CrossRef](#)]
- McElreavey, K.; Bashamboo, A. Monogenic Forms of DSD: An Update. *Horm. Res. Paediatr.* **2021**. [[CrossRef](#)]
- Globa, E.; Zelinska, N.; Shcherbak, Y.; Bignon-Topalovic, J.; Bashamboo, A.; McElreavey, K. Disorders of Sex Development in a Large Ukrainian Cohort: Clinical Diversity and Genetic Findings. *Front. Endocrinol.* **2022**, *13*, 810782. [[CrossRef](#)]
- Mary, L.; Leclerc, D.; Gilot, D.; Belaud-Rotureau, M.-A.; Jaillard, S. The TALE Never Ends: A Comprehensive Overview of the Role of PBX1, a TALE Transcription Factor, in Human Developmental Defects. *Hum. Mutat.* **2022**, *43*, 1125–1148. [[CrossRef](#)]
- Selleri, L.; Zappavigna, V.; Ferretti, E. “Building a Perfect Body”: Control of Vertebrate Organogenesis by PBX-Dependent Regulatory Networks. *Genes Dev.* **2019**, *33*, 258–275. [[CrossRef](#)]
- Heidet, L.; Morinière, V.; Henry, C.; De Tomasi, L.; Reilly, M.L.; Humbert, C.; Alibeu, O.; Fourrage, C.; Bole-Feysot, C.; Nitschké, P.; et al. Targeted Exome Sequencing Identifies PBX1 as Involved in Monogenic Congenital Anomalies of the Kidney and Urinary Tract. *J. Am. Soc. Nephrol.* **2017**, *28*, 2901–2914. [[CrossRef](#)]
- Le Tanno, P.; Breton, J.; Bidart, M.; Satre, V.; Harbuz, R.; Ray, P.F.; Bosson, C.; Dieterich, K.; Jaillard, S.; Odent, S.; et al. PBX1 Haploinsufficiency Leads to Syndromic Congenital Anomalies of the Kidney and Urinary Tract (CAKUT) in Humans. *J. Med. Genet.* **2017**, *54*, 502–510. [[CrossRef](#)]
- Schnabel, C.A.; Selleri, L.; Cleary, M.L. Pbx1 Is Essential for Adrenal Development and Urogenital Differentiation. *Genesis* **2003**, *37*, 123–130. [[CrossRef](#)] [[PubMed](#)]
- Schnabel, C.A.; Godin, R.E.; Cleary, M.L. Pbx1 Regulates Nephrogenesis and Ureteric Branching in the Developing Kidney. *Dev. Biol.* **2003**, *254*, 262–276. [[CrossRef](#)] [[PubMed](#)]
- Ma, W.; Li, Y.; Wang, M.; Li, H.; Su, T.; Li, Y.; Wang, S. Associations of Polymorphisms in WNT9B and PBX1 with Mayer-Rokitansky-Küster-Hauser Syndrome in Chinese Han. *PLoS ONE* **2015**, *10*, e0130202. [[CrossRef](#)]
- Chen, N.; Zhao, S.; Jolly, A.; Wang, L.; Pan, H.; Yuan, J.; Chen, S.; Koch, A.; Ma, C.; Tian, W.; et al. Perturbations of Genes Essential for Müllerian Duct and Wölffian Duct Development in Mayer-Rokitansky-Küster-Hauser Syndrome. *Am. J. Hum. Genet.* **2021**, *108*, 337–345. [[CrossRef](#)] [[PubMed](#)]
- Slavotinek, A.; Risolino, M.; Losa, M.; Cho, M.T.; Monaghan, K.G.; Schneidman-Duhovny, D.; Parisotto, S.; Herkert, J.C.; Stegmann, A.P.A.; Miller, K.; et al. De Novo, Deleterious Sequence Variants That Alter the Transcriptional Activity of the Homeoprotein PBX1 Are Associated with Intellectual Disability and Pleiotropic Developmental Defects. *Hum. Mol. Genet.* **2017**, *26*, 4849–4860. [[CrossRef](#)] [[PubMed](#)]
- Eozenou, C.; Bashamboo, A.; Bignon-Topalovic, J.; Merel, T.; Zwermann, O.; Lourenco, D.; Lottmann, H.; Lichtenauer, U.; Rojo, S.; Beuschlein, F.; et al. The TALE Homeodomain of PBX1 Is Involved in Human Primary Testis-Determination. *Hum. Mutat.* **2019**, *40*, 1071–1076. [[CrossRef](#)] [[PubMed](#)]

17. Walton, R.T.; Christie, K.A.; Whittaker, M.N.; Kleinstiver, B.P. Unconstrained Genome Targeting with Near-PAMless Engineered CRISPR-Cas9 Variants. *Science* **2020**, *368*, 290–296. [[CrossRef](#)]
18. Engler, C.; Marillonnet, S. Golden Gate Cloning. *Methods Mol. Biol.* **2014**, *1116*, 119–131. [[CrossRef](#)]
19. Ran, F.A.; Hsu, P.D.; Wright, J.; Agarwala, V.; Scott, D.A.; Zhang, F. Genome Engineering Using the CRISPR-Cas9 System. *Nat. Protoc.* **2013**, *8*, 2281–2308. [[CrossRef](#)]
20. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A Fast Spliced Aligner with Low Memory Requirements. *Nat. Methods* **2015**, *12*, 357–360. [[CrossRef](#)]
21. Chen, Y.; Chen, Y.; Shi, C.; Huang, Z.; Zhang, Y.; Li, S.; Li, Y.; Ye, J.; Yu, C.; Li, Z.; et al. SOAPnuke: A MapReduce Acceleration-Supported Software for Integrated Quality Control and Preprocessing of High-Throughput Sequencing Data. *Gigascience* **2018**, *7*, gix120. [[CrossRef](#)] [[PubMed](#)]
22. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
23. Li, B.; Dewey, C.N. RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)] [[PubMed](#)]
24. Wang, L.; Feng, Z.; Wang, X.; Wang, X.; Zhang, X. DEGseq: An R Package for Identifying Differentially Expressed Genes from RNA-Seq Data. *Bioinformatics* **2010**, *26*, 136–138. [[CrossRef](#)]
25. Shen, S.; Park, J.W.; Lu, Z.; Lin, L.; Henry, M.D.; Wu, Y.N.; Zhou, Q.; Xing, Y. RMATS: Robust and Flexible Detection of Differential Alternative Splicing from Replicate RNA-Seq Data. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E5593–E5601. [[CrossRef](#)]
26. Benelli, M.; Pescucci, C.; Marseglia, G.; Severgnini, M.; Torricelli, F.; Magi, A. Discovering Chimeric Transcripts in Paired-End RNA-Seq Data by Using EricScript. *Bioinformatics* **2012**, *28*, 3232–3239. [[CrossRef](#)]
27. Thiaville, M.M.; Stoeck, A.; Chen, L.; Wu, R.-C.; Magnani, L.; Oidtmann, J.; Shih, I.-M.; Lupien, M.; Wang, T.-L. Identification of PBX1 Target Genes in Cancer Cells by Global Mapping of PBX1 Binding Sites. *PLoS ONE* **2012**, *7*, e36054. [[CrossRef](#)]
28. Villaescusa, J.C.; Li, B.; Toledo, E.M.; Rivetti di Val Cervo, P.; Yang, S.; Stott, S.R.; Kaiser, K.; Islam, S.; Gyllborg, D.; Laguna-Goya, R.; et al. A PBX1 Transcriptional Network Controls Dopaminergic Neuron Development and Is Impaired in Parkinson's Disease. *EMBO J.* **2016**, *35*, 1963–1978. [[CrossRef](#)]
29. Arts, P.; Garland, J.; Byrne, A.B.; Hardy, T.S.E.; Babic, M.; Feng, J.; Wang, P.; Ha, T.; King-Smith, S.L.; Schreiber, A.W.; et al. Paternal Mosaicism for a Novel PBX1 Mutation Associated with Recurrent Perinatal Death: Phenotypic Expansion of the PBX1-Related Syndrome. *Am. J. Med. Genet. Part A* **2020**, *182*, 1273–1277. [[CrossRef](#)]
30. Blasi, F.; Bruckmann, C.; Penkov, D.; Dardaei, L. A Tale of TALE, PREP1, PBX1, and MEIS1: Interconnections and Competition in Cancer. *Bioessays* **2017**, *39*, 1600245. [[CrossRef](#)]
31. He, C.; Wang, Z.; Zhang, L.; Yang, L.; Li, J.; Chen, X.; Zhang, J.; Chang, Q.; Yu, Y.; Liu, B.; et al. A Hydrophobic Residue in the TALE Homeodomain of PBX1 Promotes Epithelial-to-Mesenchymal Transition of Gastric Carcinoma. *Oncotarget* **2017**, *8*, 46818–46833. [[CrossRef](#)] [[PubMed](#)]
32. Stankunas, K.; Shang, C.; Twu, K.Y.; Kao, S.-C.; Jenkins, N.A.; Copeland, N.G.; Sanyal, M.; Selleri, L.; Cleary, M.L.; Chang, C.-P. Pbx/Meis Deficiencies Demonstrate Multigenetic Origins of Congenital Heart Disease. *Circ. Res.* **2008**, *103*, 702–709. [[CrossRef](#)]
33. Fernandez, L.C.; Errico, M.C.; Bottero, L.; Penkov, D.; Resnati, M.; Blasi, F.; Caré, A. Oncogenic HoxB7 Requires TALE Cofactors and Is Inactivated by a Dominant-Negative Pbx1 Mutant in a Cell-Specific Manner. *Cancer Lett.* **2008**, *266*, 144–155. [[CrossRef](#)]
34. Kyei-Barffour, I.; Margetts, M.; Vash-Margita, A.; Pelosi, E. The Embryological Landscape of Mayer-Rokitansky-Kuster-Hauser Syndrome: Genetics and Environmental Factors. *Yale J. Biol. Med.* **2021**, *94*, 657–672. [[PubMed](#)]
35. Hatzistergos, K.E.; Williams, A.R.; Dykxhoorn, D.; Bellio, M.A.; Yu, W.; Hare, J.M. Tumor Suppressors RB1 and CDKN2a Cooperatively Regulate Cell-Cycle Progression and Differentiation During Cardiomyocyte Development and Repair. *Circ. Res.* **2019**, *124*, 1184–1197. [[CrossRef](#)]
36. Galvin, K.M.; Donovan, M.J.; Lynch, C.A.; Meyer, R.I.; Paul, R.J.; Lorenz, J.N.; Fairchild-Huntress, V.; Dixon, K.L.; Dunmore, J.H.; Gimbrone, M.A.; et al. A Role for Smad6 in Development and Homeostasis of the Cardiovascular System. *Nat. Genet.* **2000**, *24*, 171–174. [[CrossRef](#)]
37. Davis, B.N.; Hilyard, A.C.; Lagna, G.; Hata, A. SMAD Proteins Control DROSHA-Mediated MicroRNA Maturation. *Nature* **2008**, *454*, 56–61. [[CrossRef](#)]
38. Portnoi, M.-F.; Dumargne, M.-C.; Rojo, S.; Witchel, S.F.; Duncan, A.J.; Eozenou, C.; Bignon-Topalovic, J.; Yatsenko, S.A.; Rajkovic, A.; Reyes-Mugica, M.; et al. Mutations Involving the SRY-Related Gene SOX8 Are Associated with a Spectrum of Human Reproductive Anomalies. *Hum. Mol. Genet.* **2018**, *27*, 1228–1240. [[CrossRef](#)]
39. Soares, L.M.M.; Zanier, K.; Mackereth, C.; Sattler, M.; Valcárcel, J. Intron Removal Requires Proofreading of U2AF/3' Splice Site Recognition by DEK. *Science* **2006**, *312*, 1961–1965. [[CrossRef](#)]
40. Zhao, Y.; Cai, W.; Hua, Y.; Yang, X.; Zhou, J. The Biological and Clinical Consequences of RNA Splicing Factor U2AF1 Mutation in Myeloid Malignancies. *Cancers* **2022**, *14*, 4406. [[CrossRef](#)]
41. Svendsen, J.M.; Smogorzewska, A.; Sowa, M.E.; O'Connell, B.C.; Gygi, S.P.; Elledge, S.J.; Harper, J.W. Mammalian BTBD12/SLX4 Assembles a Holliday Junction Resolvase and Is Required for DNA Repair. *Cell* **2009**, *138*, 63–77. [[CrossRef](#)] [[PubMed](#)]
42. An, X.; Ji, B.; Sun, D. TRIM34 Localizes to the Mitochondria and Mediates Apoptosis through the Mitochondrial Pathway in HEK293T Cells. *Heliyon* **2020**, *6*, e03115. [[CrossRef](#)] [[PubMed](#)]
43. Rosa-Mercado, N.A.; Steitz, J.A. Who Let the DoGs out? - Biogenesis of Stress-Induced Readthrough Transcripts. *Trends Biochem. Sci.* **2022**, *47*, 206–217. [[CrossRef](#)] [[PubMed](#)]

44. Shaw, G.; Morse, S.; Ararat, M.; Graham, F.L. Preferential Transformation of Human Neuronal Cells by Human Adenoviruses and the Origin of HEK 293 Cells. *FASEB J.* **2002**, *16*, 869–871. [[CrossRef](#)]
45. Stepanenko, A.A.; Dmitrenko, V.V. HEK293 in Cell Biology and Cancer Research: Phenotype, Karyotype, Tumorigenicity, and Stress-Induced Genome-Phenotype Evolution. *Gene* **2015**, *569*, 182–190. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

SUPPLEMENTARY FIGURES

NB: table S1 to S3 are available online <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9956894/>

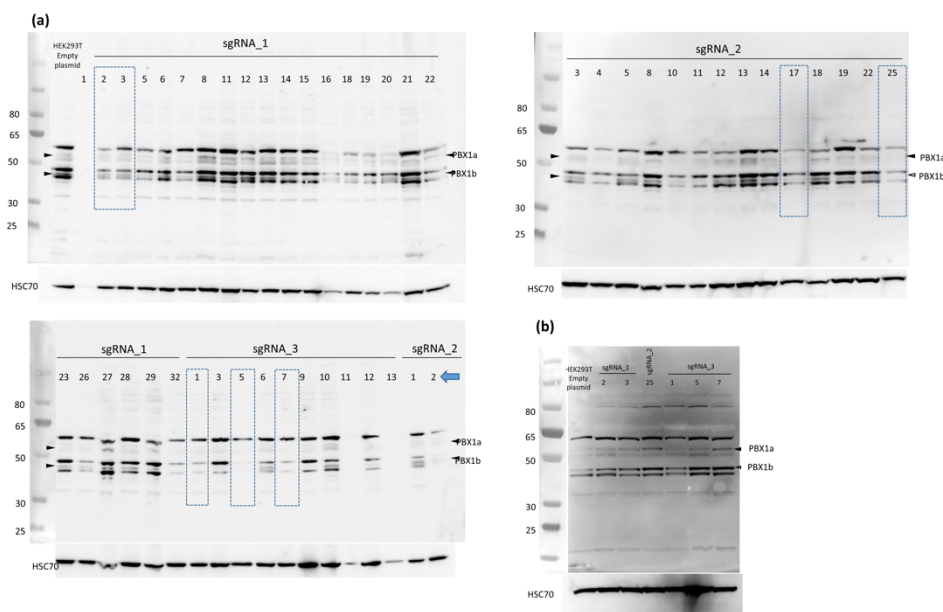


Figure S1. Uncropped Western Blot of clones obtained after CRISPR-KO editing. (a). Clone designed by an arrow was considered closest to a PBX1 KO (further called clone KD2 in main text). Clones in dashed boxes were re-analysed through WB (b). Arrowheads show PBX1a (black) and PBX1b (white) expected sizes.

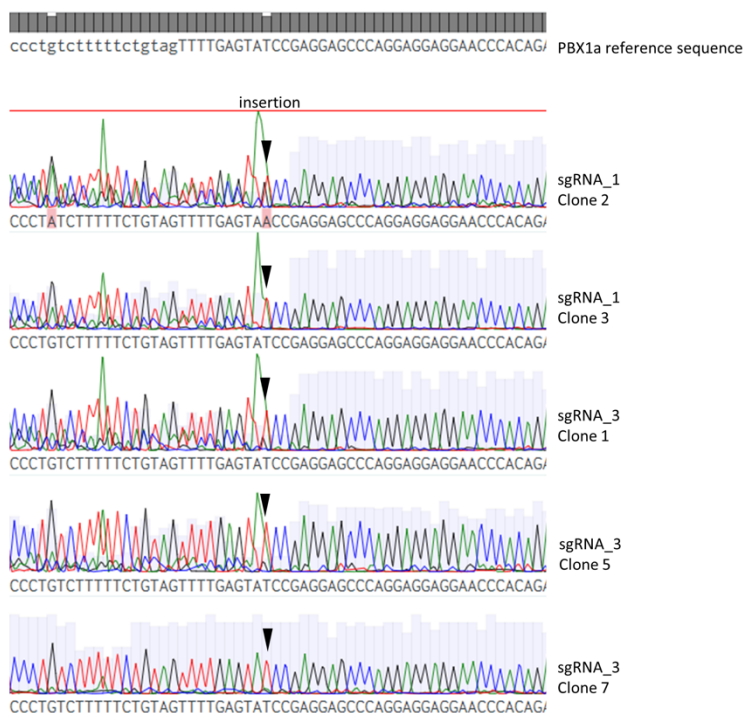


Figure S2: Clones showing frameshift variations located at the sgRNA's target sequence. Sanger sequencing of clones selected in Figure S1 (dashed boxes, Figure S1a), showing a partially edited PBX1 exon 3 sequence: at least one PBX1 allele shows a frameshift, but the WT PBX1 sequence persists (HEK cells are triploid). Clone 7 is not edited. Clone 25 (sgRNA_2) died before DNA could be extracted for testing.

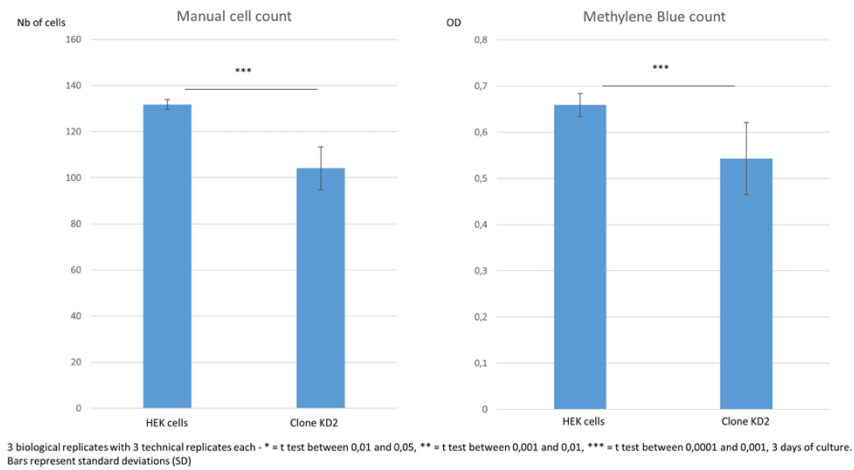


Figure S3: Proliferation's studies. Significant decrease in proliferation. Capacities of clone KD2 compared to WT HEK cells (WT) after 3 days of culture. Two counting methods were used to avoid cell pellet detachment bias that may occur with methylene blue count.

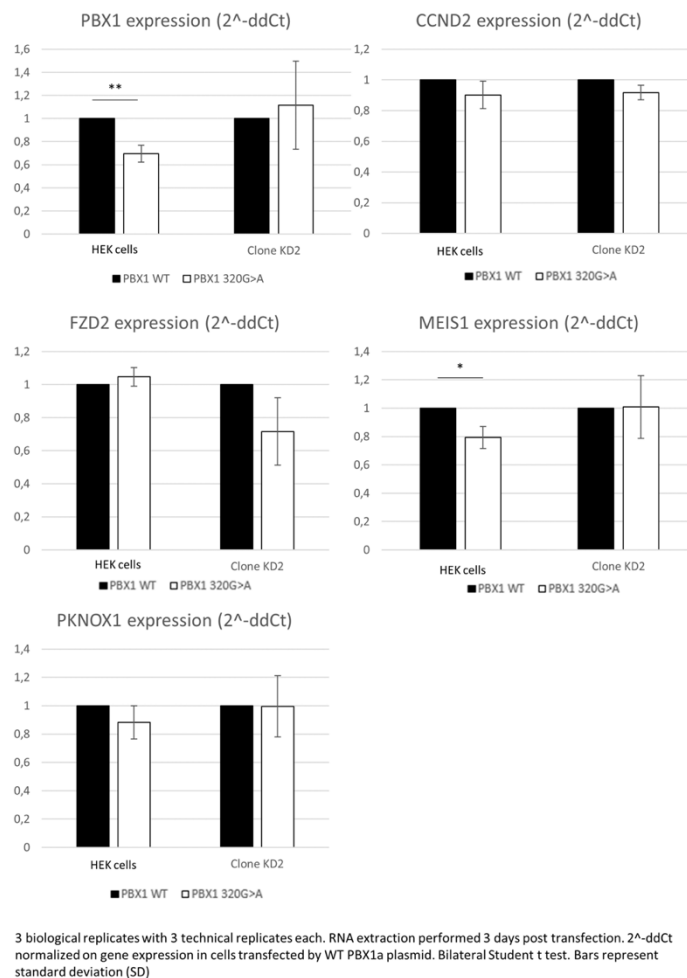


Figure S4. Relative gene expression in WT HEK and KD2 cells lines after transfection. PBX1 WT: WT PBX1a plasmid, PBX1 320G>A: mutant PBX1a plasmid. *MEOX1* was not conserved for analyses due to a very low level of expression in both cell lines.

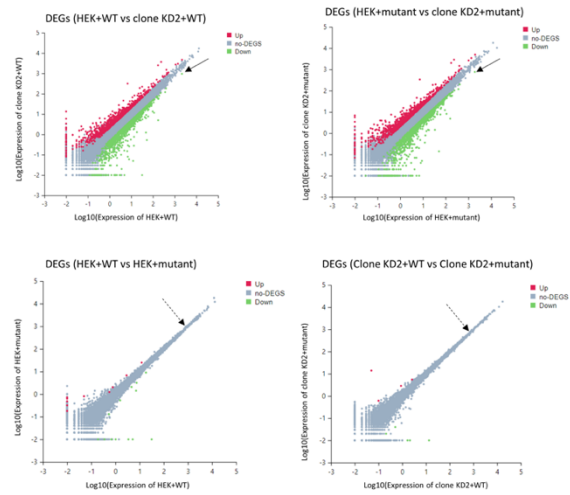


Figure S5. Scatter plot of differentially expressed genes (DEGs) between samples. PBX1 is slightly less expressed in clones KD2 compared to WT HEK cells, whatever the transfected plasmid (upside panes, arrow). For a chosen cell type, PBX1 expression is similar between WT PBX1a-transfected and mutant PBX1a-transfected cells (downside panes, dashed arrow). HEK+WT/mutant: WT HEK cells transfected with WT or mutant PBX1a plasmid. Clone KD2+WT/mutant: KD2 cells transfected with WT or mutant PBX1a plasmid.

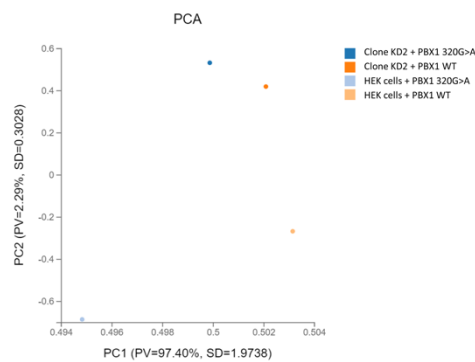


Figure S6. PCA plot of RNA-seq samples.

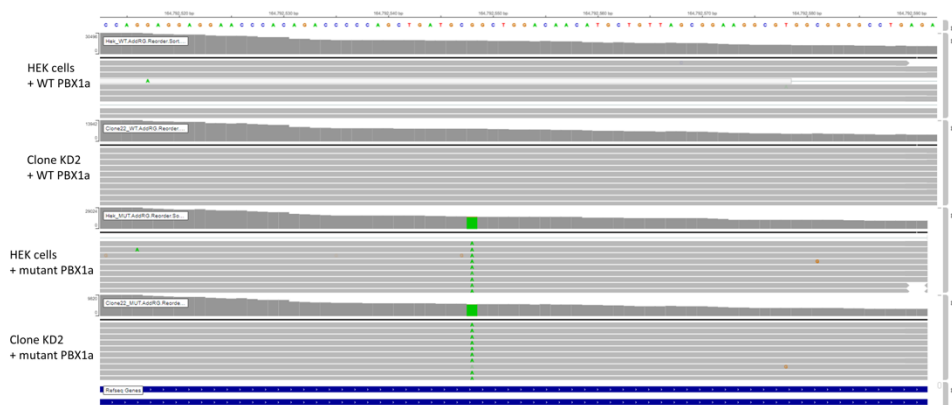


Figure S7. PBX1-320G>A-transfected cells express PBX1-mutated transcripts. IGV view of RNA sequence from PBX1 exon 3 in cells transfected with WT PBX1a-coding plasmid (WT) or PBX1a-320G>A-coding plasmid (mutant). Cells transfected with the mutated PBX1a form only express the PBX1a-320G>A form at the RNA level. HEK+WT/mutant: WT HEK cells transfected with WT or mutant PBX1a plasmid. Clone KD2+WT/mutant: KD2 cells transfected with WT or mutant PBX1a plasmid.

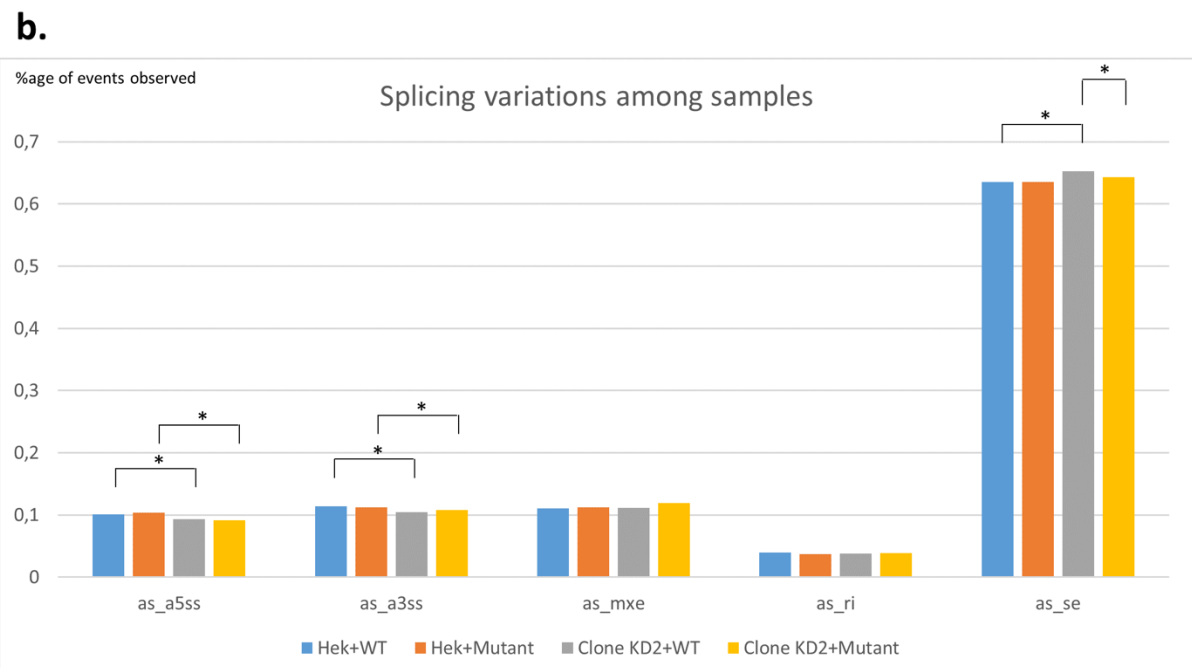
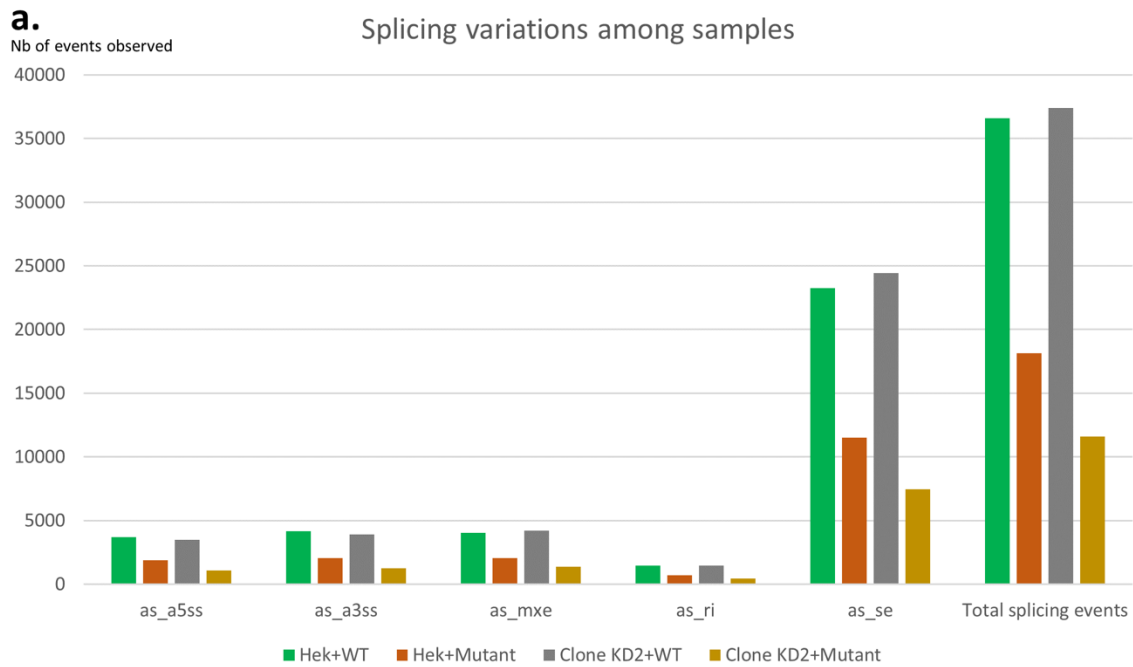


Figure S8: Differential alternative splicing (as) events among groups. HEK+WT/mutant: WT HEK cells transfected with WT or mutant PBX1a plasmid. Clone KD2+WT/mutant: KD2 cells transfected with WT or mutant PBX1a plasmid. Skipped Exon (SE), Alternative 5' Splicing Site (A5SS), Alternative 3' Splicing Site (A3SS), Mutually exclusive exons (MXE), Retained Intron (RI). (a) Number (nb) of splicing variations among samples showed a decrease in splicing variations in clone KD2, (b) Due to these large variations in total splicing events among samples, comparison between types of splicing events was performed on relative values using Z-test, with a threshold of 1.96 (α error rate of 5%, two tailed). % age: percentage.

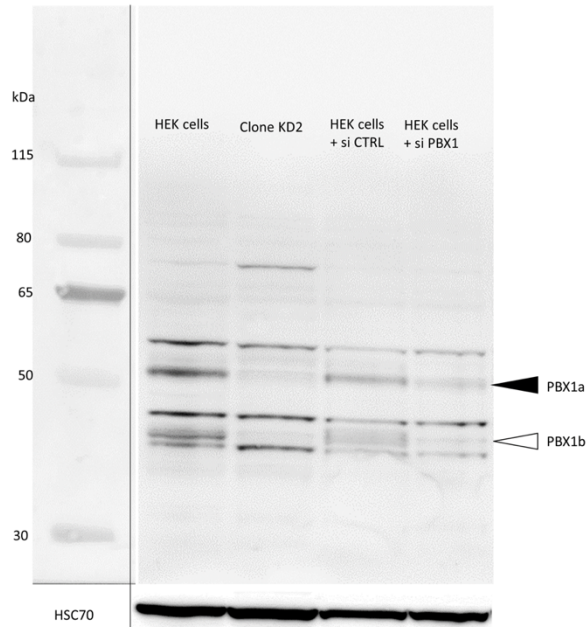


Figure S9. Uncropped Western-Blot from Figure 3 showing the PBX1 KD phenotype of clone KD2. Western Blot demonstrating the diminution of both PBX1a (black arrowhead) and PBX1b (white arrowhead) expression in PBX1 KD clone (clone KD2). WT HEK cells transfected with control siRNA (WT + si CTRL) or PBX1 siRNA (WT + si PBX1) served as negative and positive controls, respectively. WT: non-transfected HEK cells.

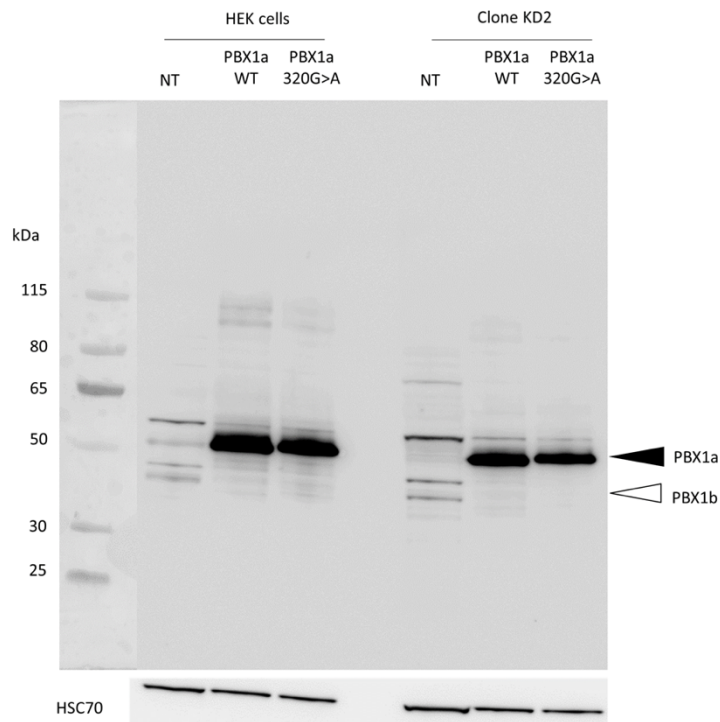


Figure S10. Uncropped Western-Blot from Figure 4 assessing PBX1a overexpression after PBX1a-coding plasmid transfection. Uncropped WB assessing PBX1a overexpression (black arrowhead) in transfected HEK WT cells (HEK WT) and KD2 clones. White arrowhead: PBX1b.

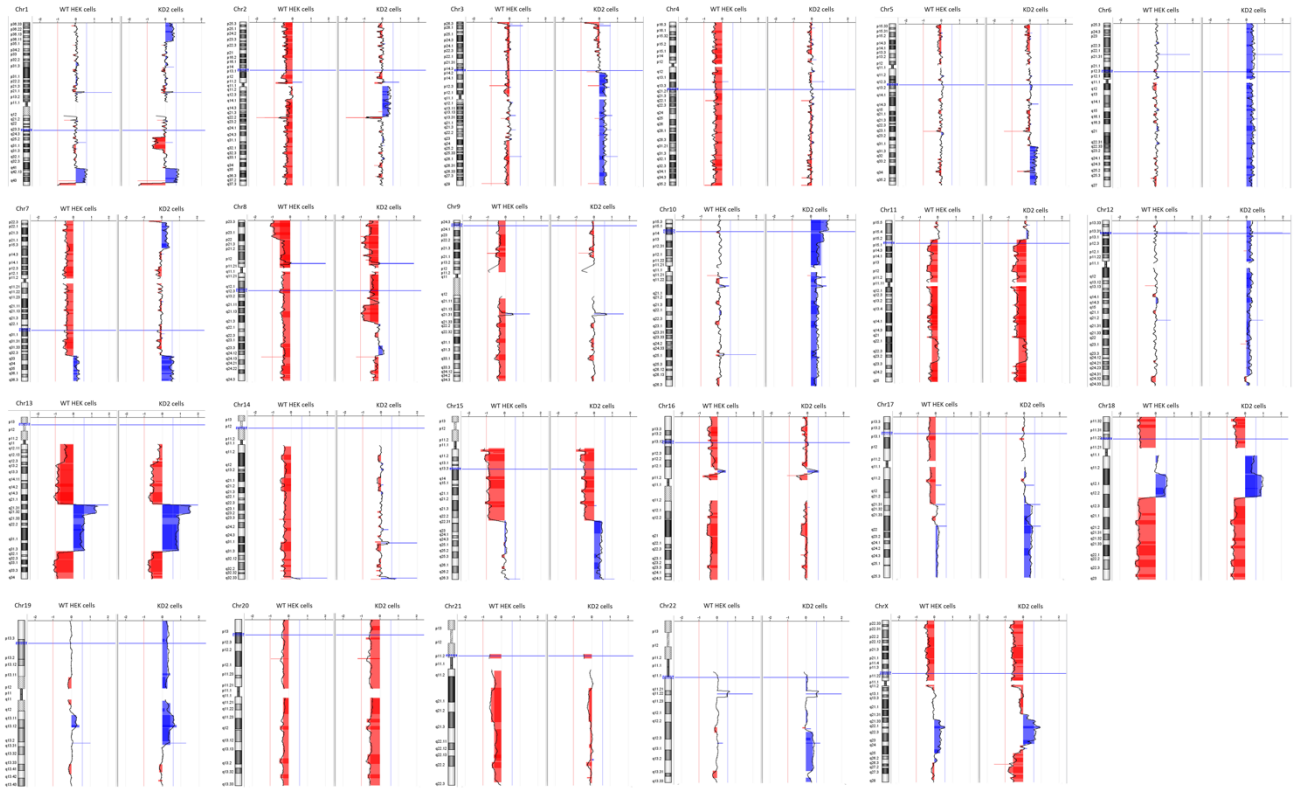


Figure S11. Comparison of aCGH profiles between WT HEK cells and clone KD2 before plasmid transfection.

CONCLUSION DE L'ARTICLE

Dans cette étude, l'impact fonctionnel du variant faux sens *PBX1* c.320G>A, p.(Arg107Gln) a été évalué par surexpression plasmidique dans des cellules HEK293T. Les résultats des tests fonctionnels ne permettent pas de conclure sur la pathogénicité de ce variant avec certitude. Néanmoins, la récurrence des substitutions localisés dans l'Arginine 107 chez les patients présentant des phénotypes similaires suggèrent un effet potentiellement délétère.

En définitive, la surexpression plasmidique n'est pas la meilleure stratégie pour l'étude des facteurs de transcription. En effet, les données de la littérature suggèrent que l'activité de *PBX1* est dose-dépendante, et l'effet délétère du variant peut être masqué par cette surexpression. Au contraire, cette surexpression peut également apparaître comme toxique pour les cellules. D'autant que les cellules ne survivent pas en cas d'inactivation totale du gène par CRISPR/Cas9, et la transfection de plasmide génère un stress aux cellules pouvant altérer la qualité des tests fonctionnels (RNAseq, étude de prolifération cellulaire). La technologie BE aurait été plus adaptée. *PBX1* étant un facteur de transcription à expression ubiquitaire associé à des désordres du développement pléiotropique, il est difficile de tirer des conclusions à partir d'un unique modèle cellulaire. Un autre modèle, plus physiologique, aurait été plus adapté.

RESULTAT 4

Identification de gènes promouvant la plasticité et la résistance du mélanome cutané par criblage CRISPRa

AVANT PROPOS

Cette étude a été portée par le Dr Arthur Gautron, ancien doctorant encadré par David Gilot de 2015 à 2019. Elle a été publiée dans la revue *Embo Molecular Medicine* le 16 Mars 2021. Je débutais ma thèse lors des expériences pour la révision de cet article. J'ai ainsi pu participer à cette étude.

INTRODUCTION DE L'ARTICLE

La résistance des tumeurs aux thérapies ciblées conduit malheureusement à l'échec thérapeutique chez les patients atteints du cancer. Parmi ces mécanismes de résistances, des déterminants génétiques et non génétiques sont mis en jeu. Les déterminants génétiques sont généralement bien caractérisés dans la littérature (mutations gain ou perte de fonction). L'avènement des outils d'analyse de l'expression des gènes en cellule unique (scRNAseq) a montré que les cellules exposées aux traitements subissaient une reprogrammation transcriptomique intense participant à la résistance (déterminants non génétiques). Toutefois, ces observations demeurent très descriptives, et aujourd'hui, les mécanismes sous-jacents mis en jeu ne sont pas bien caractérisés d'un point de vue fonctionnel.

La technologie d'activation transcriptionnelle CRISPR (CRISPRa) est aujourd'hui très bien maîtrisée par la communauté scientifique pour activer l'expression des gènes. En utilisant une bibliothèque d'ARN guides couvrant la plupart des promoteurs de gènes codants humains, il est possible d'activer l'ensemble de ces gènes dans un modèle cellulaire. Typiquement, une cellule reçoit un ARN guide particulier de la bibliothèque, permettant à chaque cellule de surexprimer un gène particulier. Cette bibliothèque de cellules est ensuite soumise à une molécule exerçant une pression de sélection (agent thérapeutique) pour en identifier le mécanisme de résistance. Seules les cellules surexprimant un gène leur permettant de résister à la molécule continueront de proliférer. Cette technique est couramment utilisée pour identifier les mécanismes de résistances transcriptomiques aux traitements anti-cancéreux.

Dans cette étude, Arthur Gautron et al ont utilisé la technique de criblage CRISPRa *in vivo* et *in vitro* dans l'objectif d'identifier, à l'échelle du génome, les gènes codants dont le gain d'expression est associé à la résistance aux inhibiteurs de BRAF dans le mélanome cutané. Ce traitement est utilisé chez les personnes atteintes du mélanome cutané avec une mutation gain de fonction BRAF (V600E).

Les données obtenues ont permis d'identifier trois gènes dont la surexpression est associée à une résistance aux inhibiteurs de BRAF, au mécanisme de rechute et au phénotype mésenchymateux des cellules (dédifférenciation). Notamment, SMAD3, un facteur de transcription de la voie TGF- β , promeut la résistance en stimulant la transcription de nombreux gènes dont *AXL* et *EGFR*. L'expression de *SMAD3* est elle-même favorisée par le facteur de transcription AhR, identifié auparavant par l'équipe comme activé dans les cellules de mélanomes résistantes aux inhibiteurs de BRAF. L'inhibition des voies AhR/TGF- β pourrait constituer une future cible thérapeutique intéressante pour limiter l'acquisition de cellules résistantes et/ou éliminer les « *persisting cells* ».

CONTRIBUTION

J'ai uniquement mené l'expérience présentée dans la **figure EV5**, demandée pour la révision de l'article (*sgRNA targeting four isoforms of SMAD3 and BRAFi resistance*). L'objectif de cette expérience est de montrer que selon le type de promoteur *SMAD3* ciblé par l'ARN guide, le niveau de transactivation du gène peut varier. Ici, les résultats suggèrent que seuls les ARN guides ciblant le bon promoteur permettent d'activer *SMAD3* (isoforme longue). Cette isoforme est la seule à promouvoir la résistance, d'où la nécessité d'étudier indépendamment les ARN guides de chaque promoteur de gènes. Le niveau de transactivation du transcrit peut également varier selon la position de l'ARN guide au promoteur.



CRISPR screens identify tumor-promoting genes conferring melanoma cell plasticity and resistance

Arthur Gautron¹, Laura Bachelot¹, Marc Aubry^{1,2,†} , Delphine Leclerc^{3,†}, Anaïs M Quémener^{1,†}, Sébastien Corre^{1,†} , Florian Rambow^{4,5}, Anaïs Paris¹, Nina Tardif¹, Héloïse M Leclair¹, Oskar Marin-Bejar^{4,5}, Cédric Coulouarn³, Jean-Christophe Marine^{4,5}, Marie-Dominique Galibert^{1,6,*} & David Gilot^{1,‡,**}

Abstract

Most genetic alterations that drive melanoma development and resistance to targeted therapy have been uncovered. In contrast, and despite their increasingly recognized contribution, little is known about the non-genetic mechanisms that drive these processes. Here, we performed *in vivo* gain-of-function CRISPR screens and identified SMAD3, BIRC3, and SLC9A5 as key actors of BRAFi resistance. We show that their expression levels increase during acquisition of BRAFi resistance and remain high in persister cells and during relapse. The upregulation of the SMAD3 transcriptional activity (SMAD3-signature) promotes a mesenchymal-like phenotype and BRAFi resistance by acting as an upstream transcriptional regulator of potent BRAFi-resistance genes such as EGFR and AXL. This SMAD3-signature predicts resistance to both current melanoma therapies in different cohorts. Critically, chemical inhibition of SMAD3 may constitute amenable target for melanoma since it efficiently abrogates persister cells survival. Interestingly, decrease of SMAD3 activity can also be reached by inhibiting the Aryl hydrocarbon Receptor (AhR), another druggable transcriptional factor governing SMAD3 expression level. Our work highlights novel drug vulnerabilities that can be exploited to develop long-lasting antimelanoma therapies.

Keywords Aryl hydrocarbon Receptor; CRISPR-SAM; melanoma; SMAD3; targeted therapy resistance

Subject Categories Cancer; Skin

DOI 10.15252/emmm.202013466 | Received 16 September 2020 | Revised 9 February 2021 | Accepted 11 February 2021 | Published online 16 March 2021

EMBO Mol Med (2021) 13: e13466

Introduction

Identifying molecular cancer drivers is critical for precision oncology. Last year, the cancer genome atlas (TCGA) identified 299 driver genes by focusing on point mutations and small indels across 33 cancer types (Bailey *et al*, 2018). It represents the most comprehensive effort thus far to identify cancer driver mutations. Complementary studies are required to elucidate the role of copy-number variations, genomic fusions, and methylation events in the 33 TCGA projects.

Moreover, there is increasing evidence that non-genetic reprogramming leading to cancer cell dedifferentiation, stemness, invasiveness also contribute to tumor growth and therapy resistance (Puisieux *et al*, 2014; Bai *et al*, 2019). Thus, deciphering the signaling pathways that drive such processes may also lead to innovative cancer therapies. Recent gene expression quantifications performed at single-cell level by single-cell RNA sequencing (scRNA-Seq) demonstrated that cancer cells operate a dedifferentiation process, for instance in glioblastoma and melanoma (Patel *et al*, 2014; Tirosh *et al*, 2016; Rambow *et al*, 2018), promoting tumor growth, stemness, and therapy resistance. This “onco-dedifferentiation” seems to be independent of *de novo* mutations and could offer new targets/strategies to cure cancer. However, these scRNA-Seq studies are mainly descriptive; the tumor growth capability of each gene/RNA is not yet investigated at the genome-scale. Such functional analyses are nowadays feasible using clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 screens (Shalem *et al*, 2015). The majority of the CRISPR-Cas9 screens is based on the inactivation of coding genes but modulation of gene expression is reachable with the CRISPR-Cas9 synergistic activation mediator (SAM) approach (Konermann *et al*, 2015). It corresponds to an engineered protein complex for the transcriptional activation of endogenous genes. Importantly, SAM can further be combined with a human genome-wide library to activate all

1 CNRS, IGDR (Institut de génétique et développement de Rennes)-UMR 6290, Univ Rennes, Rennes, France

2 Plateforme GEH, CNRS, Inserm, BIOSIT - UMS 3480, US_S 018, Univ Rennes, Rennes, France

3 INSERM U1242, Centre Eugène Marquis, Rennes, France

4 Department of Oncology, KU Leuven, Leuven, Belgium

5 VIB Center for Cancer Biology, VIB, Leuven, Belgium

6 Service de Génétique Moléculaire et Génomique, CHU Rennes, Rennes, France

*Corresponding author. Tel: +33 0 223234705; E-mail: mgaliber@univ-rennes1.fr

**Corresponding author (Lead Contact). Tel: +33 0 223234441; E-mail: david.gilot@univ-rennes1.fr

†These authors contributed equally to this work

‡Present address: INSERM U1242, Centre Eugène Marquis, Rennes, France

known coding isoforms from the RefSeq database (23,430 isoforms) for gain-of-function screening without *a priori*. To date, CRISPR screens are mainly performed *in vitro* using cell lines or primary cultures (Meyers *et al*, 2017). A pan-cancer CRISPR-Cas9 knockout screen was performed *in vitro* (324 human cancer cell lines from 30 cancer types) to identify essential genes for cancer cell fitness (defined as genes required for cell growth or viability) and to prioritize candidates for cancer therapeutics (Behan *et al*, 2019). However, because the contribution of the tumor environment in tumor growth is increasingly recognized, it seems important to perform such screens in the relevant patho-physiological context and, for instance, take advantage of animal models.

We selected cutaneous melanoma as a paradigm since novel therapeutic strategies are critically needed (Bai *et al*, 2019). Targeted therapies such as BRAF inhibitors (BRAFi) initially showed great promise in patients with BRAF(V600)-mutated metastatic melanoma. Unfortunately, the vast majority of patients that initially respond to these drugs, almost inevitably develop resistance. Although combination therapies (BRAF and MEK inhibitors) enhance the response and delay relapse, the overall survival remains unsatisfactory highlighting the need of new therapeutic targets (Ascierto *et al*, 2016).

The mechanisms underlying resistance are numerous and probably not mutually exclusive (Sullivan & Flaherty, 2013; Welsh *et al*, 2016; Song *et al*, 2017). Resistance can be driven by a small pre-existing subpopulation, harboring-specific genetic alterations that confer them with resistance to the inhibitors (Wagle *et al*, 2011). Such alterations may also occur *de novo*, during treatment (Welsh *et al*, 2016). In addition, there is increased evidence that non-genetic reprogramming may confer drug-tolerant and/or resistant phenotypes to melanoma cells (Rambow *et al*, 2018; Corre *et al*, 2018; Tsoi *et al*, 2018; Hugo *et al*, 2015a; Talebi *et al*, 2018; Rapino *et al*, 2018; preprint: Marin-Bejar *et al*, 2020). Earlier works demonstrated that phenotype switching from a proliferative to an invasive/mesenchymal-like state is also likely to contribute to therapy resistance (Hoek & Goding, 2010; Konieczkowski *et al*, 2014; Müller *et al*, 2014; Verfaillie *et al*, 2015; Boshuizen *et al*, 2018). Paradoxically, MITF-induced differentiation into a slow cycling, pigment-producing state was also reported to confer tolerance to BRAFi (Müller *et al*, 2014; Smith *et al*, 2016). It therefore seems that various drug-tolerant subpopulations can emerge under therapeutic pressure and that these cells can provide a pool from which resistance develops. Targeting these populations of persister cells is therefore crucial to achieve effective personalized therapies (Nassar & Blanpain, 2016).

Here, we performed unbiased screens to identify genes promoting tumor growth from persister cells and conferring resistance to BRAFi using CRISPR-Cas9 SAM methodology. We demonstrate that, in addition to promote melanoma development, Mothers against decapentaplegic homolog 3 (*SMAD3*), Baculoviral IAP repeat-containing protein 3 (*BIRC3*), and Sodium/hydrogen exchanger 5 (*SLC9A5*) also support relapse since they promote both BRAFi-resistance and tumor growth capability of persister cells. Their expression levels correlated with BRAFi resistance and relapse. Consequently, their inhibition strongly reduced the number of persister cells. Moreover, we demonstrate that the transcription factor AhR governs *SMAD3* expression levels and in turn *SMAD3* drives the expression of a set of genes associated with BRAFi resistance and mesenchymal phenotype. These experiments identify integrated AhR-*SMAD3* signaling as a key driver of melanoma growth and relapse, pointing to a new therapeutic vulnerability in melanoma.

Results

Identification of tumor-promoting genes by *in vivo* gain-of-function CRISPR screen

Since the tumor environment influences, at least in part, the tumor growth capability of cancer cells, we performed an *in vivo* genome-wide CRISPR-Cas9 SAM screen to identify *in vivo* tumor-promoting genes, defined as genes whose expression support tumor growth (in contrast to driver genes bearing a driver mutation such as BRAF (V600E)). To select the most appropriate cellular model, we classified melanoma biopsies from The Cancer Genome Atlas (TCGA) cohort ($n = 458$) in function of differentiation states according to the most recent melanoma profiling data (Fig 1A; Tsoi *et al*, 2018). As anticipated, the vast majority of these tumors, which are almost all drug naive, exhibited a differentiated profile (89%; melanocytic and transitory). We selected the 501Mel cell line since (i) these cells display a melanocytic differentiation state as the majority of diagnosed melanoma, (ii) they harbor the BRAF(V600E) mutation as ~50% of cutaneous melanoma, (iii) they are highly sensitive to BRAFi with an IC_{50} value of 0.45 μM to vemurafenib [PLX4032] (Halaban *et al*, 2010; Corre *et al*, 2018), and importantly (iv) they are unable to generate tumor in nude mice (Ohanna *et al*, 2011). This latter characteristic may allow to identify tumor-promoting genes.

To generate the CRISPR-SAM cell library, we modified the 501Mel cells, to express constitutively defective-Cas9 and the required cofactors for CRISPR-SAM technology (Koneremann *et al*, 2015). These engineered cells were infected with the single-guide RNA (sgRNA) lentivirus library that contained at least three different guides per coding gene (Koneremann *et al*, 2015; Fig 1B). The infection was performed at a multiplicity of infection (MOI) of 0.2 ensuring that only one guide is expressed per infected cell. Infected cells were positively selected using antibiotic selection during 7 days. By DNA sequencing, we observed a normal distribution of the sgRNAs in two cell library replicates (Fig 1C). Only 78 sgRNAs were not detected in our cell library, which validate our protocol and the cell library (>70,100 sgRNAs were detected; Table EV1). Thus, the controls were proper to identify *in vivo* tumor-promoting genes.

The cell library (30×10^6 cells) was fractionated and subcutaneously xenografted in 10 nude mice (3×10^6 cells/mouse) and tumor growth was monitored using caliper over a 5-month period (Fig 1D and Table EV2). As previously demonstrated (Ohanna *et al*, 2011), we confirmed that parental 501Mel are unable to form tumors in nude mice ($n = 6$). In contrast, seven tumors were obtained from the CRISPR-engineered cells xenografted in 10 mice (Fig 1D). The nature of the sgRNAs, their abundance and occurrence across these 7 tumors were determined by DNA-Seq (Fig 1E and F, and Table EV3).

By comparing the most represented genes (sgRNAs) in each tumor (Tum), we identified 3 common genes (Fig 1F). An enrichment of *SMAD3*, *BIRC3*, and *SLC9A5* sgRNAs was found in tumors when compared to their starting abundance (cell library; orange points; Fig 1G), in contrast to EGFR sgRNAs. Thirty-six other genes were recurrently retrieved in the tumors but not in all (Table EV3). Interestingly, *YAP1* which has already been identified as melanoma growth-promoting gene was also found (Table EV3). This supports

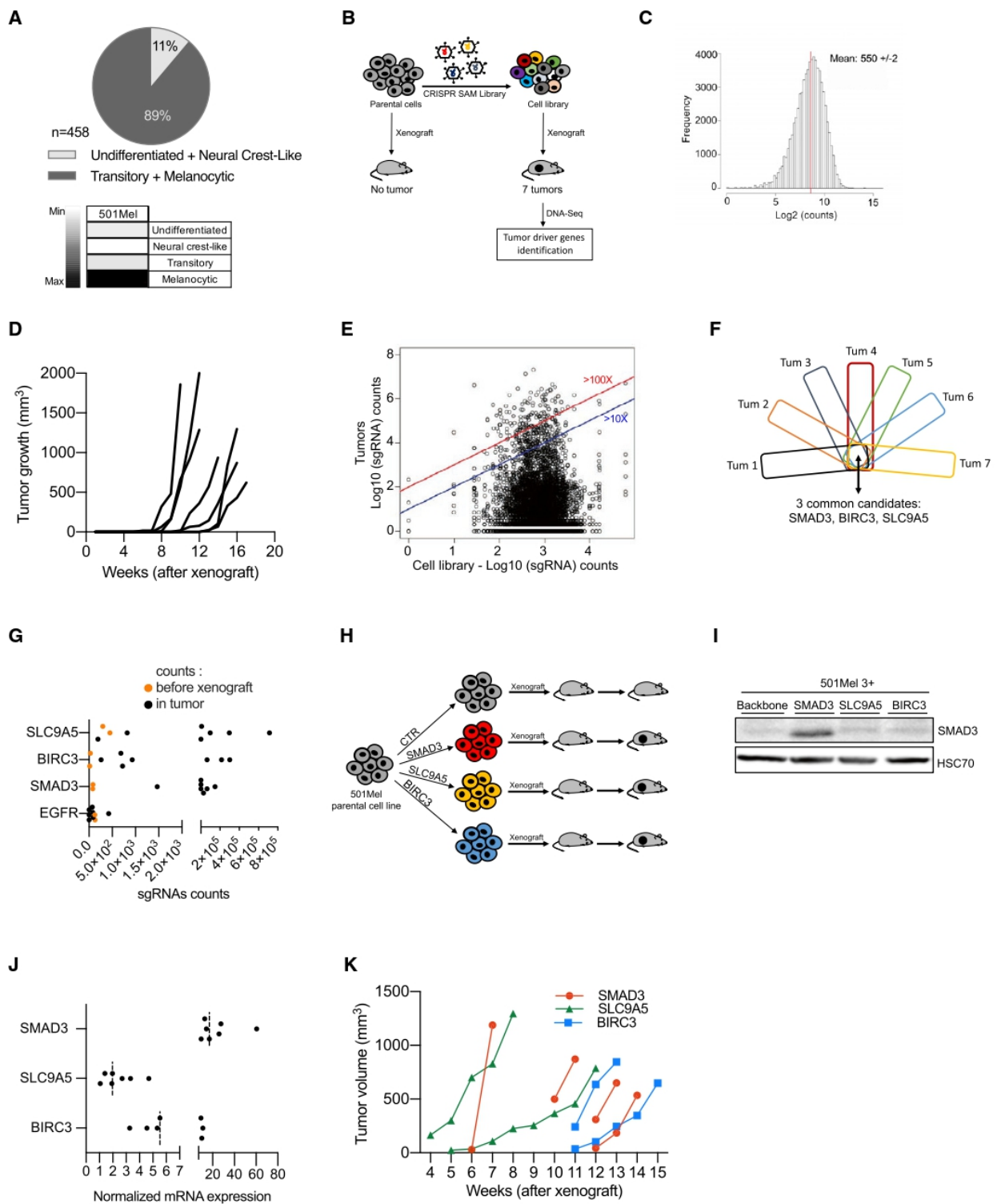


Figure 1.

Figure 1. Identification of tumor-promoting genes by *in vivo* gain-of-function CRISPR screen.

- A Determination of differentiation states of skin cutaneous melanoma (SKCM) biopsies from the TCGA cohort ($n = 458$) according to (Tsoi *et al.*, 2018). The vast majority of these tumors exhibited a differentiated profile (89%; melanocytic and transitory states). The others display a dedifferentiated profile (11%; neural crest-like cells and undifferentiated states). Human melanoma 501Mel cell line is classified as differentiated melanoma cells (melanocytic cells), according to the gene expression profile (four categories were defined by (Tsoi *et al.*, 2018); melanocytic, transitory, neural crest-like, and undifferentiated cells). 501Mel cell line was selected for the CRISPR screens.
- B Workflow depicting the *in vivo* CRISPR-SAM screen to identify tumor-promoting genes. Parental cells and CRISPR-SAM cell library were xenografted on nude mice (3×10^6 cells/mouse, $n = 6$ and $n = 10$, respectively) and tumor growth was monitored during 5 months. Seven tumors were collected and analyzed by DNA-Seq to identify the sgRNAs (Tables EV1 and EV2).
- C sgRNAs distribution in the cell library. sgRNA mean reaches 550 ± 2 .
- D Tumor growth curves for the 7 tumors arising from the CRISPR-SAM-engineered cells xenografted in nude mice as detailed in panel B. (No tumor for the parental cell line (501Mel cells)) (Table EV2).
- E Distribution of sgRNAs in cell library and in the 7 tumors ($\log_{10}(\text{sgRNAs counts})$). Blue and red lines indicated the enrichment ≥ 10 fold or ≥ 100 fold (tumors *vs* cell library *in vitro*) (Table EV1).
- F From the seven tumors, the top hundred genes (enriched) have been selected and common genes are *SMAD3*, *BIRC3*, and *SLC9A5* (Table EV3).
- G sgRNA counts in tumors *versus* in CRISPR-SAM cell library (respectively, black and orange points). Each black point corresponds to one tumor. Two replicates have been shown for CRISPR-SAM cell library (orange points).
- H Workflow depicting the validation step: 501Mel cells overexpressing *SMAD3*, *BIRC3*, or *SLC9A5* (obtained by CRISPR-SAM) were xenografted on nude mice and tumor volume was monitored using caliper. 3×10^6 cells/mouse. $n = 7$, 6, and 6 mice, respectively.
- I *SMAD3* expression levels in 501Mel cells overexpressing the cofactors for CRISPR-SAM approach and the control sgRNA (501Mel 3 + backbone) or the *SMAD3* sgRNA. *SLC9A5* and *BIRC3* sgRNAs are used as controls to show the specificity of the *SMAD3* overexpression. HSC70 serve as loading control.
- J *SMAD3*, *BIRC3*, and *SLC9A5* mRNA expression levels in melanoma cell lines described in H and I. $n = 7$ independent biological experiments. Dashed lines for medians.
- K Tumor growth curves from 501Mel cells overexpressing *SMAD3*, *BIRC3*, or *SLC9A5*.

Data information: Western blot results are representative of at least two independent experiments. Source data and unprocessed original blots are available in Appendix Fig S1 source data. See also Fig EV1.
Source data are available online for this figure.

the robustness of the screen (Lamar *et al.*, 2012; Verfaillie *et al.*, 2015; Hugo *et al.*, 2015b). The majority of the tumor-promoting genes (Table EV4) identified here are not considered as genes required for cell growth or viability (except the essential genes *YAP1*, *SLC25A41*, and *TGIF1*; Behan *et al.*, 2019) and are not frequently altered in melanoma (Fig EV1A). Moreover, the transforming growth factor (TGF)- β pathway seemed well-represented among the tumor-promoting genes (Table EV4). Since our results suggest that a high expression level of these tumor-promoting genes is sufficient to promote melanoma tumor growth, we evaluated the biological consequences of the inhibition of one tumor-promoting gene, *BIRC3*. The chemical inhibitor Birinapant reduced the SKMel28R and Me1402 cell density (Fig EV1B and C), suggesting that the BIRC3 protein is involved in cell proliferation of these melanoma cells as previously demonstrated (Krepler *et al.*, 2013; Vetma *et al.*, 2017).

Next, we examined *in vivo* the ability of these tumor-promoting genes to promote tumor growth by xenografting cell populations overexpressing the sgRNAs individually. We focused on the top 3 tumor-promoting genes: *SMAD3*, *BIRC3*, and *SLC9A5* (Fig 1H–K). We generated three new CRISPR-engineered cell lines, and we evaluated the overexpression levels of these genes by Western blot experiments (*SMAD3*) and RT-qPCR (*SMAD3*, *BIRC3*, and *SLC9A5*) (Fig 1I and J). Finally, we confirmed that they independently foster tumor development (Fig 1K). Altogether, our results demonstrated that *in vivo* CRISPR-SAM screen identifies new tumor-promoting genes, which may constitute amenable target for melanoma.

Genome-wide CRISPR activation screen identifies BRAFi-resistance genes in cutaneous melanoma

BRAFi provokes tumor shrinkage in the vast majority of patients with BRAF(V600)-mutated metastatic melanoma but resistance

almost inevitably occurs (Bai *et al.*, 2019). To examine the genes promoting BRAFi resistance and relapse, we performed an *in cellulo* screen using the same cell library in the presence of BRAFi (Fig 2). Briefly, the CRISPR-SAM 501Mel cell library (40×10^6 cells) was treated for 14 days with BRAFi (2 μM), using either the BRAFi used in clinical practice (vemurafenib), the next-generation inhibitor that is still under investigation in clinical trials (PLX8394), or the solvent (dimethyl sulfoxide (DMSO)) as control (Fig 2A). This procedure allows for the enrichment of sgRNAs (genes) conferring resistance. The nature of the sgRNA present in the resistant population and their abundance was determined by DNA-Seq (Fig 2B). The best hit was the Epidermal growth factor receptor gene (*EGFR*), a well-known BRAFi-resistance gene (Sun *et al.*, 2014; Shaffer *et al.*, 2017). By examining the enrichment of sgRNAs targeting *EGFR* promoter (Fig 2B), we decided to retain genes with at least two sgRNAs among the enriched sgRNAs present in BRAFi-exposed cells (with a false discovery rate, $\text{FDR} < 0.05$) (Tables EV5 and EV6, Appendix Fig S1) since sometimes one of the three sgRNAs designed per gene is not detected or not enriched as observed for *EGFR* and *BIRC3* (Fig 2C). A recent publication confirmed that sgRNAs are not all functional in CRISPR libraries and it could be interesting to increase the number of sgRNAs per target and to cover more TSS per gene (Sanson *et al.*, 2018).

Apart sgRNAs targeting the *EGFR* promoter, the sgRNAs enrichment in BRAFi-exposed cells was unexpectedly weak (Tables EV5 and EV6). Thus, to select the best BRAFi-resistance genes, we examined the gene expression levels of these potential BRAFi-resistance genes identified by CRISPR screen in 12 melanoma cell lines (Fig 2D and E, Appendix Fig S1). We postulated that BRAFi-resistance genes are highly expressed in BRAFi-resistant cells ($n = 6$) as already demonstrated by other approaches for *NRP1*, *AXL*, and *NGFR*. To this end, we confronted CRISPR-SAM candidates (identified in 501Mel cells) to gene expression data from six melanoma cell lines

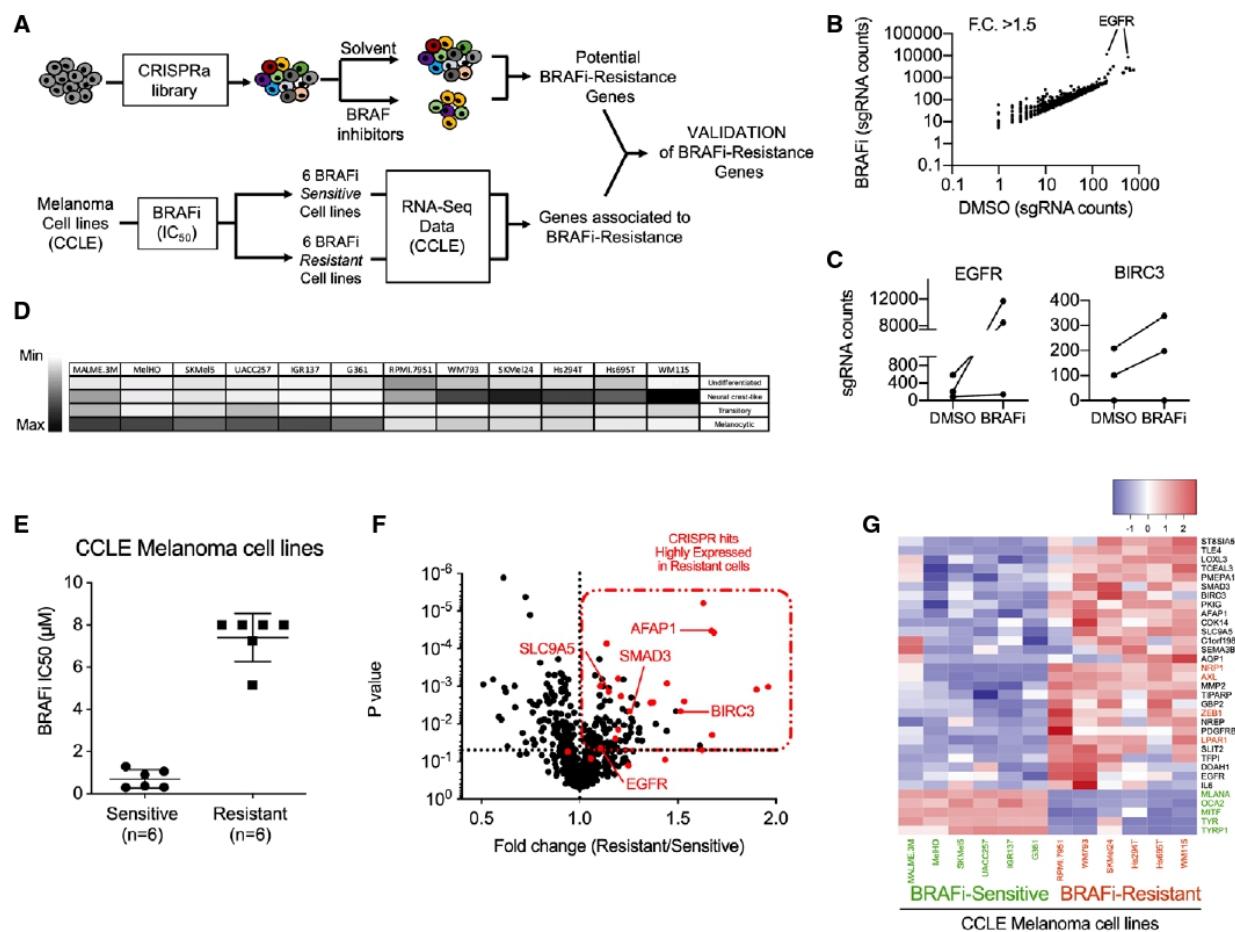


Figure 2. Genome-wide CRISPR activation screen identifies BRAFi-resistance genes in cutaneous melanoma.

A CRISPR-SAM workflow. Cell library was exposed to DMSO (solvent of BRAFi) or BRAFi 2 µM (vemurafenib (Vem) or PLX8394 (PB, Paradox Breaker)) during 14 days. 40 × 10⁵ cells per arm. Experiments were done in duplicate.

B Plot showing sgRNAs detected in BRAFi-resistant cells versus in control cells (cell library exposed to DMSO) with a fold change > 1.5. Raw data are available in Table EV5 (BRAFi for PBV). Appendix Fig S1 depicted the comparison Vem vs. PB.

C sgRNAs counts in BRAFi-resistant cells versus in DMSO-exposed cells for *EGFR* and *BIRC3* (BRAFi for PBV, Appendix Fig S1).

D Determination of differentiation states of 12 human BRAF(V600) melanoma cell lines from CCLE according to (Tsoi et al, 2018).

E BRAF(V600) melanoma cell lines from CCLE were distributed in two groups (Sensitive and "intrinsically" Resistant, n = 6 cell lines per group) according to their BRAFi IC₅₀ (µM, half maximal inhibitory concentrations) (Barretina et al, 2012). Error bars reflect mean ± s.d.

F Volcano plot showing the expression levels of BRAFi-resistant genes (identified in 501Mel by CRISPR-SAM screen) in 12 human melanoma cell lines. The fold change corresponds to the ratio of expression levels found in resistant and sensitive cell lines. In red: selected genes. *EGFR* is considered as positive control and *SMAD3*, *BIRC3*, and *SLC9A5*; selected as favorite genes for the next steps. *SMAD3*, *BIRC3*, *SLC9A5*, and *AFAP1* are BRAFi-resistance genes and potent tumor-promoting genes (Fig 1). Raw data are available in Table EV5 (BRAFi for PBV, Appendix Fig S1).

G Heat map recapitulating the expression levels of the selected hits (red dots in Fig 2F) in BRAFi-resistant and BRAFi-sensitive cell lines. Markers of differentiation (*MITF*, *MLANA*, *OCA2*, *TYRP1*, *TYR*). In red: resistance genes already published (*NRP1*, *AXL*, *ZEB1*, *LPAR1*). Scale corresponds to Z scores.

Source data are available online for this figure.

that were highly resistant to BRAFi according to the Cancer Cell Lines Encyclopedia (CCLE) versus six sensitive cell lines (Barretina et al, 2012; Fig 2E). It is important to note that the CCLE BRAFi-resistant cell lines are therapy-naïve and intrinsically resistant.

We next focused on candidate genes which were both enriched in CRISPR screen and highly expressed in the majority of BRAFi-resistance cell lines (Fig 2F and G). To better evaluate the validity of

our candidate genes identified using this workflow, we added well-established and validated BRAFi-resistant genes (*NRP1*, *AXL*, *ZEB1*, and *LPAR1*) (Müller et al, 2014; Konermann et al, 2015; Rizzolio et al, 2018) and five genes associated with melanoma cell differentiation (*MITF*, *OCA2*, *MLANA*, *TYR*, and *TYRP1*) (Levy et al, 2006). All BRAFi-resistant cell lines presented a dedifferentiated profile as anticipated. Importantly, our candidate genes including *SMAD3*,

SLC9A5, and *BIRC3* (Fig 2G, black color) displayed a similar expression profile than observed for the well-established and validated BRAFi-resistant genes (Fig 2G, red color), strongly suggesting that these genes may also confer BRAFi resistance. *EGFR* and platelet-derived growth factor receptor (*PDGFR*)- β showed high expression only in a few BRAFi-resistant cell lines, as previously observed in patients (Sun *et al*, 2014).

Together, our results confirmed the robustness of the functional *in cellulo* CRISPR-SAM screen.

Validation of BRAFi SAM-selected resistance genes

To evaluate the contribution of the CRISPR-SAM-selected genes in BRAFi resistance, we examined the transcriptome of the differentiated cell line M229 (transitory cell state) at different stages during acquisition of resistance (Fig 3A; Song *et al*, 2017). As described by Song *et al* (2017), cell lines were exposed to chronic exposure to BRAFi and analyzed at different days of treatment (P: parental cells, 2D: two days of treatment, DTP: drug-tolerant persister cells, DTPP: drug-tolerant proliferating persister cells, SDR: single-drug-resistant cell). We observed a sequential upregulation in the expression of BRAFi SAM-selected genes: a group of genes (*MMP2*, *SEMA3B*, *BIRC3*, *TIPARP*, etc) being expressed earlier than a 2nd group (*IL6*, *EGFR*, *AFAP1*, etc). The majority of the candidates were overexpressed while the cells exhibited resistance to a single BRAFi agent (single-drug resistance, SDR). Comparable results were obtained with the M238 melanoma cell line (Figs 3B and EV2A). Importantly, we found that combining the BRAFi with MEKi (DDR) led to comparable upregulation of the BRAFi SAM-selected genes than observed in cells exposed to BRAFi alone (SDR) (Fig 3C). These results indicate that a common gene expression program can confer resistance to inhibitors of MAPK pathway as previously reported (Moriceau *et al*, 2015; Corre *et al*, 2018; Lee *et al*, 2020).

Having shown that combination of BRAFi and MEKi promotes sequential upregulation of BRAFi SAM-selected genes, we monitored their expression levels in an *in vivo* preclinical patient-derived xenograft (PDX) model (Rambow *et al*, 2018). Using scRNA-Seq, our collaborators reported the presence of dedifferentiated drug-tolerant cells exhibiting a neural crest stem cell (NCSC) and invasive profiles at minimal residual disease isolated from the MEL006 PDX model (Rambow *et al*, 2018). Here, we showed that the BRAFi SAM-

selected genes were highly and selectively expressed in both of these cell populations at mRNA level (Fig 3D and E). To reinforce these observations, we performed SMAD3 immunostainings in four BRAFi-mutant PDXs exposed to BRAFi/MEK inhibitors until resistance (recently characterized in (preprint: Marin-Bejar *et al*, 2020), Appendix Fig S2A and B). Immunostainings showed the emergence of SMAD3⁺ cells in Dabrafenib + Trametinib resistant lesions from the MEL003 and MEL006 PDXs in contrast to PDXs characterized by an intrinsic resistance mechanism (MEL007 and MEL037).

Moreover, we found that EGFR-expressing cells sorted from melanoma tumors displayed comparatively high expression levels of the BRAFi SAM-selected genes (Fig EV2B and C). Importantly, these EGFR-positive cells are able to proliferate in the presence of BRAFi and generate BRAFi-resistant colonies (Shaffer *et al*, 2017). In addition, high expression levels of the BRAFi SAM-selected genes have been found in invasive cells when compared to proliferative melanoma cell lines (Verfaillie *et al*, 2015; Fig EV2D). Together, these data confirm the upregulated expression of BRAFi SAM-selected genes in cells shown to contribute to relapse, suggesting their involvement in establishing drug-tolerant and/or resistant phenotypes *in vivo*.

To evaluate the clinical relevance of the SAM-selected BRAFi-resistance genes, we compared their expression levels (median) in two independent BRAFi drug naive/drug-resistant patient cohorts (Fig 3F and G, $n = 21$ (Hugo *et al*, 2015b) and $n = 16$ (Rizos *et al*, 2014) patients, respectively). The expression levels of the selected resistant candidate genes increased during relapse in drug-resistant patients (48 and 31%). Notably, none of these genes have been implicated in recurrent gene-amplification events that are sometimes identified in drug-naive lesions (cBioPortal, TCGA; Gao *et al*, 2013) (Fig EV2E–H). These data indicate that the increase in expression of the SAM-selected genes in BRAFi-resistant cells is likely associated with a (non-genetic) dedifferentiation process of melanoma cells induced by the therapy. Together, these *in vitro* and *in vivo* gene expression analyses strongly support a BRAFi-resistance function for the SAM-selected genes.

BRAFi-resistance genes promote tumor growth

Long-term effect of BRAFi is reduced by the ability of persister cells to resist to BRAFi but also to promote the tumor growth (relapse). Thus, we investigated the capability of the BRAFi-persister cells

Figure 3. Validation of BRAFi SAM-selected resistance genes.

- A Expression levels of our hits in M229 melanoma cells during the BRAFi-resistance acquisition. P: parental cells, 2D: two days of treatment, DTP: drug-tolerant persister cells, DTPP: drug-tolerant proliferating persister cells, SDR: single-drug-resistant cells (BRAFi) (Song *et al*, 2017). Scale corresponds to Z scores.
- B Expression levels of our hits in M238 melanoma cells (Song *et al*, 2017) during the BRAFi-resistance acquisition. Scale corresponds to Z scores.
- C Expression levels of our hits in parental, single-drug-resistant (SDR, BRAFi), or dual drug-resistant (DDR, BRAFi + MEKi) SKMel28 melanoma cell lines (Song *et al*, 2017). Scale corresponds to Z scores.
- D T-distributed Stochastic Neighbor Embedding (t-SNE) plot showing the 4 drug-tolerant states (NCSC (neural crest stem cells), invasive, SMC (starved-like melanoma cells) and pigmented cells) according to single-cell RNA-seq (scRNA-Seq) performed in PDX MEL006 model exposed to BRAFi + MEKi (Rambow *et al*, 2018). Our BRAFi-resistance genes are mainly expressed in NCSC and invasive cells.
- E AUCell score for each drug-tolerant states. **** $P < 0.0001$, Mann–Whitney test.
- F Expression level of our genes (median) in cutaneous melanoma biopsies before the BRAFi treatment (baseline) and during the relapse. Cohort from (Hugo *et al*, 2015b). The expression level of our BRAFi-resistance genes is increased in relapse samples (48%, $n = 21$).
- G Expression level of our genes (median) in cutaneous melanoma biopsies before the BRAFi treatment and during the relapse. Cohort from (Rizos *et al*, 2014). The expression level of our BRAFi-resistance genes is increased in relapse samples (31%, $n = 16$).

Data information: See also Fig EV2.

Source data are available online for this figure.

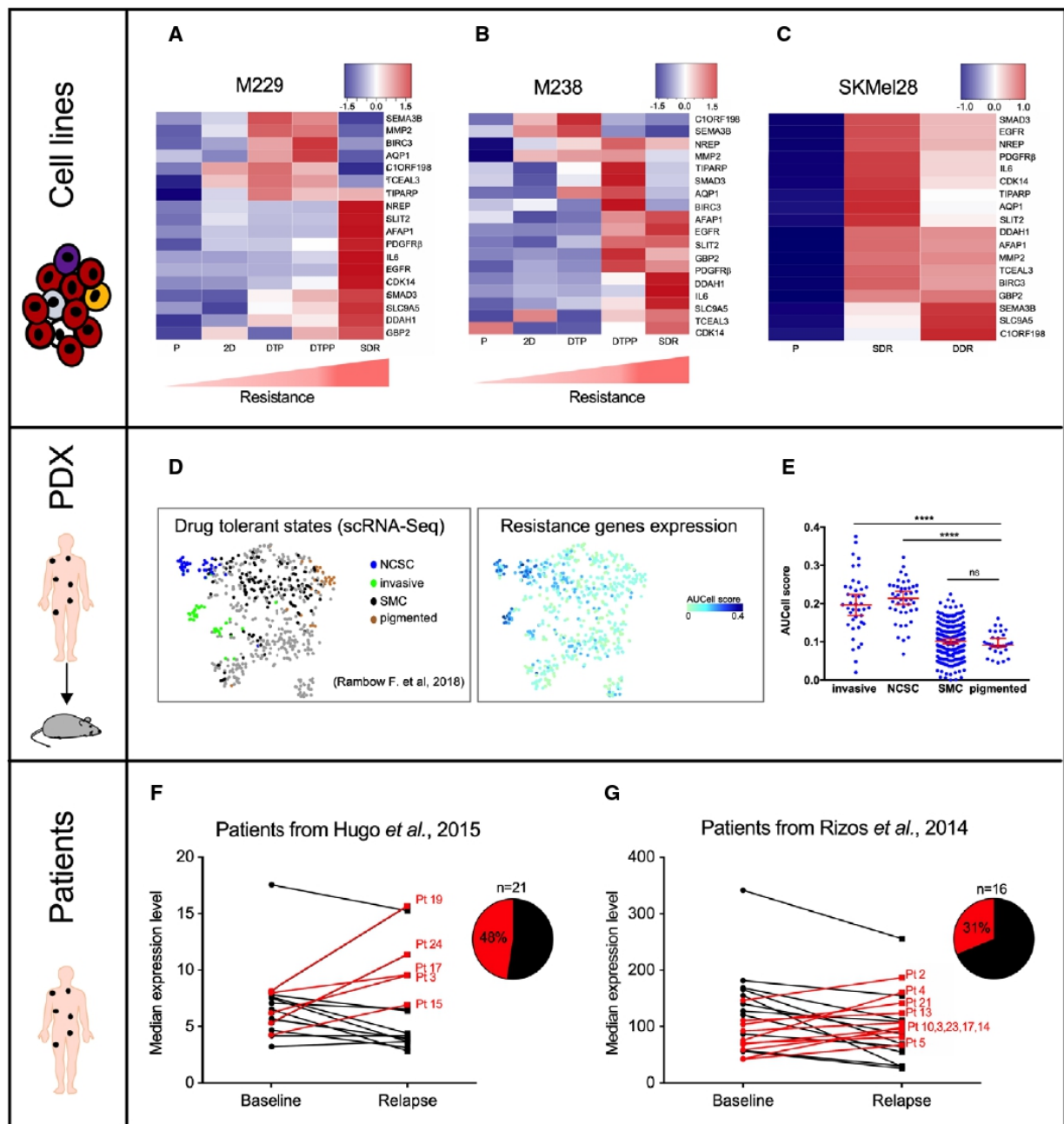


Figure 3.

(obtained from the *in vitro* CRISPR screen, Fig 2) to promote tumor growth into nude mice. The subset of BRAFi-resistant/persister cells (Fig 4A) was engrafted (36×10^6 cells, 3×10^6 /mouse), and tumor growth was monitored (Fig 4B).

These BRAFi-persister cells formed tumors, in contrast to parental 501Mel cells (Ohanna *et al.*, 2011). We determined the nature and abundance of sgRNAs present in each emerging tumor

(Fig 4C and Table EV7), and we detected 97 genes (sgRNAs) detected in all tumors. Interestingly, we recovered the tumor-promoting genes *SMAD3*, *BIRC3*, and *SLC9A5* identified above (Fig 4D). We looked for the enrichment of these sgRNAs in each tumor developed from persister cells (Fig 4E). *EGFR* sgRNA was not frequently enriched in tumors (as previously described for human melanoma tumors (Prahallad *et al.*, 2012; Sun *et al.*, 2014; Shaffer

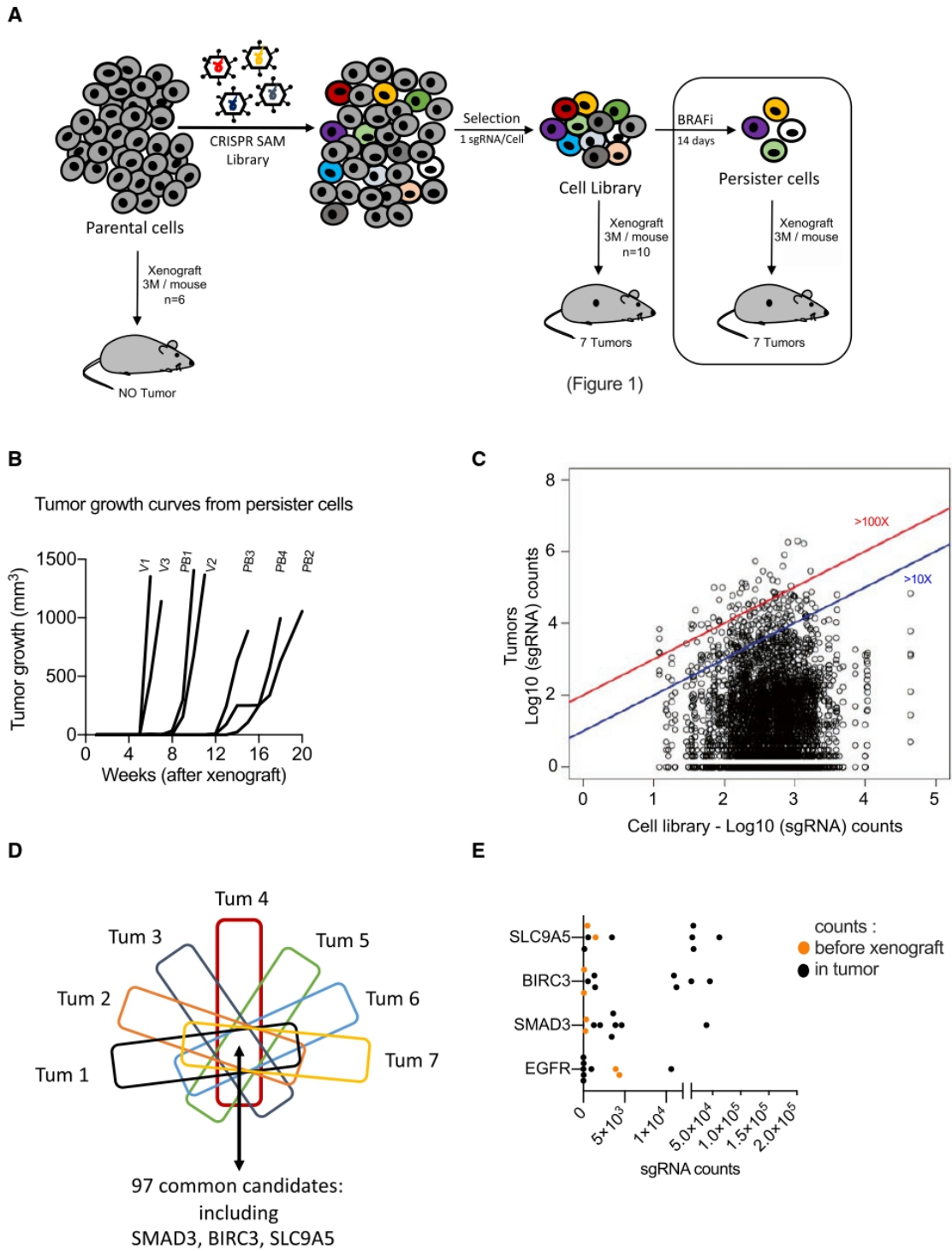


Figure 4.

Figure 4. BRAFi-resistance genes promote tumor growth.

- A Workflow to identify genes involved in tumor growth from BRAFi-persister cells. Cell populations were subcutaneously grafted on nude mice flanks (3×10^6 cells per mouse); Vem-persister cells ($n = 11$ mice) and PLX8394 (PB)-persister cells ($n = 12$ mice) (Table EV2). (BRAFi for PBV, Appendix Fig S1)
- B Tumor growth curves from BRAFi-persister cells (monitored during 5 months; Table EV2). V for Vem-resistant cells and PB for PLX8394-resistant cells.
- C Distribution of sgRNAs in BRAFi-resistant cells (before xenograft) and in the 7 tumors emerging from the BRAFi-resistant cells ($\log_{10}(\text{sgRNAs counts})$). Blue and red lines indicated the enrichment ≥ 10 fold or ≥ 100 fold (tumors versus BRAFi-resistant cells (*in vitro*)). Raw data are available in Table EV7.
- D From the seven tumors arising from BRAFi-resistant cells, the common genes (enriched) have been extracted. Ninety-seven genes including *SMAD3*, *BIRC3*, and *SLC9A5* are detected in all these 7 tumors. Raw data are available in Table EV7.
- E sgRNAs counts in tumors versus sgRNA detected in BRAFi-resistant cell library (*in vitro*) (respectively, black and orange points) for selected candidates. Each black point corresponds to one tumor. *EGFR* was the most potent BRAFi-resistant gene (Fig 2). Two replicates have been shown for CRISPR-SAM cell library (orange points). *SMAD3*, *BIRC3*, and *SLC9A5* were identified as hits in the three screens (Figs 1, 2, and 4).

Source data are available online for this figure.

et al, 2017)) in contrast to *SMAD3*, *BIRC3*, and *SLC9A5* (Fig 4E and Appendix Fig S3). Together this suggests that these 3 genes are potential interesting targets for antimelanoma therapy.

Functional validation of genes involved in BRAFi resistance and relapse

We focused on the transcription factor *SMAD3* as a model gene for monitoring BRAFi resistance and relapse due to its critical function downstream of the TGF β pathway. Although this pathway is known to promote melanoma phenotype switching/dedifferentiation (Sun et al, 2014), to support melanoma growth (Berking et al, 2001) and metastasis, little is known about the role of *SMAD3* in melanoma biology and as a modulator of resistance to targeted therapy.

We confirmed that *SMAD3* mRNA is highly expressed in dedifferentiated cells (Fig 5A–C) and in BRAFi-resistant cells (Fig 2G).

To reinforce the role of *SMAD3* in BRAFi resistance, we showed that gain-of-function of *SMAD3* significantly increases the BRAFi resistance of melanoma cells when compared to different control cells (parental 501Mel cells and the CRISPR-engineered cells: 501Mel cells expressing dCas9 and HSF1-p65-MS2 (named here 501Mel 2+) and the 501Mel 2+ cells expressing a control guide (named 3+ backbone; Fig 5D). In addition, we showed that gain-of-function of *SMAD3* also promotes the three-dimensional (3D) tumor spheroid invasion capability of melanoma cells. Similar results were obtained for *SLC9A5* (Figs 5E and EV3A). These results strongly suggest that a high expression level of *SMAD3* confers BRAFi resistance and invasion capability in melanoma cells.

Thus, we investigated if *SMAD3* impairment re-sensitizes cells to BRAFi inhibitor, using small-interfering RNA (siRNA) (Fig 5F–I). As the dedifferentiation status correlated with BRAFi resistance, we selected two BRAFi-resistant cell lines, SKMel28 BRAFi-resistant cells (SKMel28R, Fig 5F and G; Hugo et al, 2015b) and Me1402 melanoma cells (Fig 5H). In contrast to the SKMel28R, the resistance of which was created by chronic exposure to non-lethal doses of BRAFi, Me1402 cells are intrinsically resistant.

Surprisingly, the single *SMAD3* depletion decreased the cell density in a similar magnitude than BRAFi treatment in these BRAFi-resistant cells (Figs 5G and H, and EV3B). The *SMAD3* depletion did not modify the ERK pathway (Fig 5I) in contrast to the BRAFi. The combo (*SMAD3* depletion and BRAFi (5 μM)) efficiently reduced the number of resistant/persister cells in these two cell lines, suggesting that *SMAD3* is an interesting target to limit resistance to BRAFi. Similar results were obtained by targeting *BIRC3*, *EGFR*, *IL6*, or *AQP1* (Fig EV3C–J).

To transfer this strategy (*SMAD3* inhibition + BRAFi) into clinic, we looked for an efficient inhibitor of *SMAD3*. We identified the chemical inhibitor SIS3 (*SMAD3* inhibitor, *SMAD3i*; Jinnin et al, 2006; Wu et al, 2020). Firstly, we validated the inhibitory efficiency of *SMAD3i* in melanoma cells since it decreased the levels of phospho-*SMAD3* Ser423/425 induced by TGF β (Fig 5J) in accordance with previous studies (Jinnin et al, 2006; Chihara et al, 2017). Since the *SMAD3* regulation by phosphorylation is not fully understood and the phospho-*SMAD3* Ser423/425 status is not strictly correlated with *SMAD3* transcriptional activity (Ooshima et al, 2019), we evaluated the effect of *SMAD3i* using a reporter assay. We demonstrated that *SMAD3i* reduced the transcriptional activity of *SMAD3* in response to TGF β exposure in 4 melanoma cell lines (Fig 5K). To know if the combination (*SMAD3i* + BRAFi) could be broadly used to eradicate persister cells emerging in response to BRAFi treatment, we selected three melanoma cell lines in function of *SMAD3* expression levels (Fig 5L). The differentiated cells (*SMAD3*^{low}) are highly sensitive to BRAFi (decrease of cell density: > 80% at 5 μM BRAFi, Fig 5M) in contrast to dedifferentiated cells (*SMAD3*^{high}), which are highly resistant to BRAFi (decrease of cell density: ~50% at 5 μM BRAFi, Fig 5M).

We showed that *SMAD3i* (SIS3) reduced the cell density of all melanoma cell lines (Fig 5N). These results are in agreement with the *SMAD3* knock-down results (Fig 5G and H). Our experiments also indicated that melanocyte survival is weakly affected by *SMAD3* inhibition (up to 20 μM), in agreement with the non-toxicity of this inhibitor observed *in vivo* (Tang et al, 2017; Wu et al, 2020). The inhibitory effect of *SMAD3i* on melanoma cells seems to be associated with the total *SMAD3* expression levels (Fig 5L). Our results might suggest that these melanoma cells could be “addicted” to *SMAD3* activity.

We finally investigated the interest to combine BRAFi (Vem, alone or in combination with MEKi (Cobi) and *SMAD3i* (SIS3) to eradicate the BRAFi-resistant cell lines (SKMel28R, M229R, and M238R) (Figs 5O and EV3K). We showed that *SMAD3* inhibitor alone or in combination with BRAFi (Vem 5 μM) or BRAFi + MEKi (Cobi 1 μM) might be a promising treatment to reduce the amount of persister cells (melanoma). Together, these results identified *SMAD3* as an amenable target to limit resistance to BRAFi and tumor growth.

The transcription factor AhR drives *SMAD3* expression

Having shown that *SMAD3* expression mediates BRAFi-resistance and tumor growth, we explored the transcriptional program promoting its expression in BRAFi-resistant melanoma cells. We recently

reported that the Aryl hydrocarbon Receptor (AhR), a ligand-dependent transcription factor is an upstream central node regulating the expression of BRAFi-resistance genes and melanoma dedifferentiation (Corre et al, 2018; Leclerc et al, 2021; Paris et al, 2021).

We postulated that AhR may govern SMAD3 expression in BRAFi-resistant cells. We identified three putative canonical binding sites for AhR (XRE for xenobiotic responsive element) in the proximal promoter of SMAD3 (Fig 6A) in accordance with chromatin

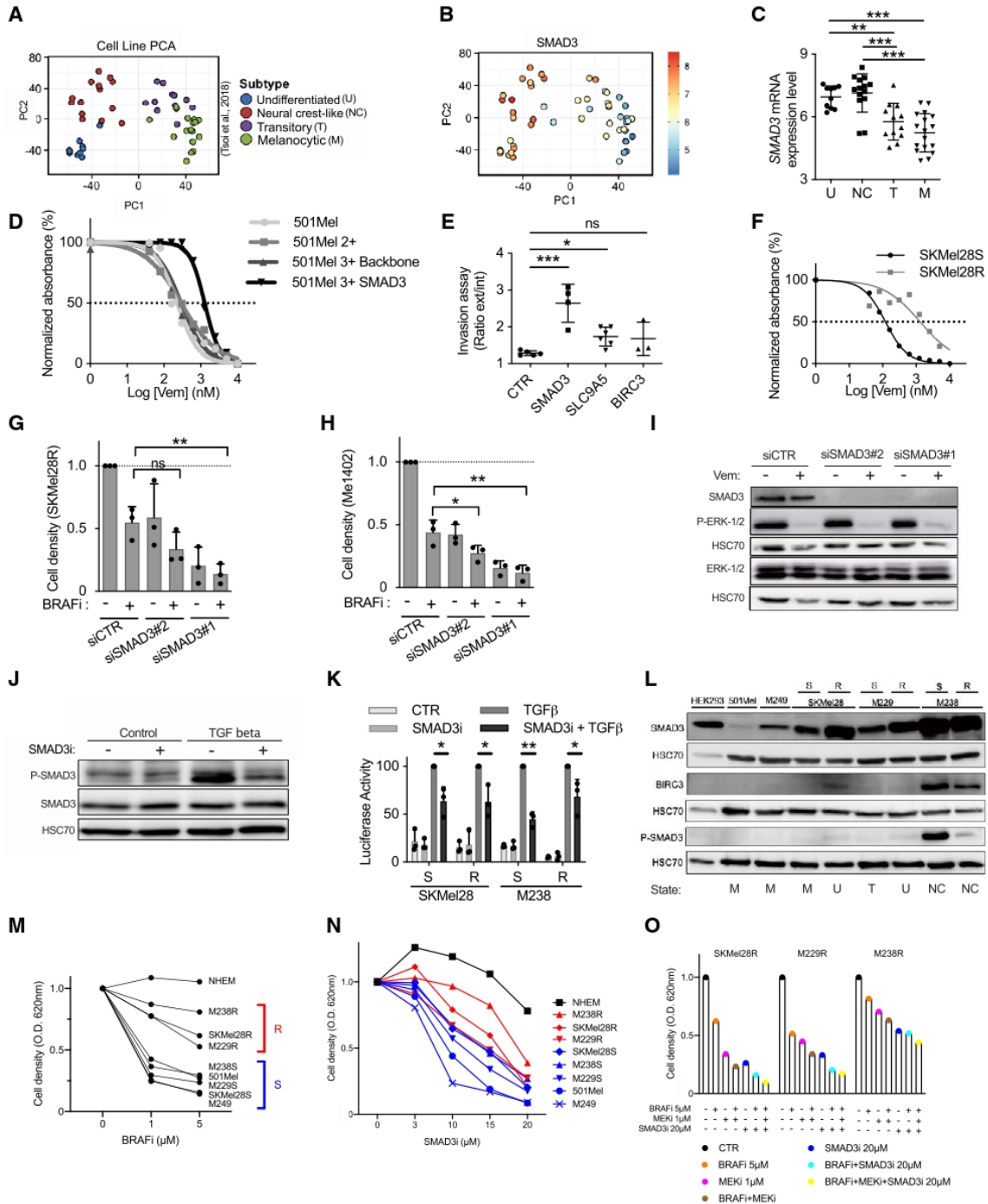


Figure 5.

Figure 5. Functional validation of genes involved in BRAFi resistance and relapse.

- A PCA analysis of melanoma cell lines in function of their dedifferentiation states (generated by the webtool <http://systems.crump.ucla.edu/dediff/index.php>).
- B *SMAD3* expression increases with melanoma cell dedifferentiation. PCA analysis of *SMAD3* expression in melanoma cell lines in function of their dedifferentiation states. Scale: red color corresponds to a high *SMAD3* expression level (unit of the scale: Log₂ FPKM).
- C *SMAD3* expression in these four subtypes of melanoma cells (U, undifferentiated; NC, neural crest-like; T, transitory; M, melanocytic). Number in each group: U = 10, NC = 14, T = 12, M = 17. Error bars reflect mean ± s.d. Multiple comparisons have been done using ordinary one-way ANOVA ***P* < 0.01, ****P* < 0.001.
- D *SMAD3* gain-of-function increases BRAFi resistance. Determination of BRAFi half maximal inhibitory concentrations (IC₅₀ values) for 501Mel cell lines (Log[Vem] (nM)). Parental 501Mel cells (*n* = 5), 501Mel cells expressing dCas9 and HSF1-p65-MS2 (named here 501Mel 2+, *n* = 5), the 501Mel 2 + cells expressing a control guide (backbone, *n* = 5) and the 501Mel 2 + cells overexpressing *SMAD3* (*n* = 3 biologically independent experiments). A representative experiment has been chosen. Dotted line is used to determine the IC₅₀ values.
- E *SMAD3* and *SLC9A5* gain-of-function increases invasion capability of melanoma cells. Invasion assays for engineered cell lines (melanoma spheroids): CTR; 501Mel 2+, *SMAD3*; 501Mel 2 + cells overexpressing *SMAD3*. Two other cell lines overexpressing *SLC9A5* or *BIRC3* have been tested. Explanation for ratio calculation is detailed in Fig EV3A. Results obtained from two biologically independent experiments. (*n* = 5, 4, 6, and 3 spheroids, respectively). Error bars reflect mean ± s.d. Multiple comparisons have been done using ordinary one-way ANOVA, **P* < 0.05, ****P* < 0.001.
- F Characterization of SKMel28 sensitive and resistant cell lines (SKMel28S and SKMel28R). Determination of BRAFi half maximal inhibitory concentrations (IC₅₀ values) (Log[Vem] (nM)). A representative experiment has been chosen among two experiments. Dotted line is used to determine the IC₅₀ values.
- G *SMAD3* depletion (siRNA#1 & #2) decreased cell density and increased BRAFi effect (vemurafenib) on BRAFi-resistant cells (SKMel28R). CTR for non-targeting siRNA DMSO for dimethylsulfoxide (solvent of vemurafenib; BRAFi). *n* = 3 biologically independent experiments. Each histogram represents the mean ± s.d.; Multiple comparisons have been done using ordinary one-way ANOVA, ***P* < 0.01.
- H *SMAD3* depletion (siRNA#1 & #2) decreased cell density and increased BRAFi effect (vemurafenib, BRAFi) on BRAFi-resistant cells (Me1402). CTR for non-targeting siRNA DMSO for dimethylsulfoxide (solvent of vemurafenib). *n* = 3 biologically independent experiments. Each histogram represents the mean ± s.d.; multiple comparisons have been done using ordinary one-way ANOVA, **P* < 0.05, ***P* < 0.01.
- I Validation of *SMAD3* knock-down by Western blot experiments in Me1402 cells exposed or not to vemurafenib (BRAFi (Vem) 5 μM, 2 days). Cells were exposed to BRAFi 24 h after siRNA transfection. Vemurafenib inhibitory effect on mutated BRAF was evaluated by analyzing the phospho-ERK1/2 levels. ERK and HSC70 serves as loading control.
- J Validation of *SMAD3* inhibitor (SIS3, *SMAD3i*). Effect of *SMAD3i* (10 μM) on the level of phospho-*SMAD3* Ser423/425 in response to TGFβ (2 ng/ml, 1 h) (or solvent: HCl 4 mM + Bovine serum albumin 1 mg/ml) in SKMel28R cells (expressing a high endogenous level of *SMAD3* mRNA). Serum starved cells (500,000 per well) were pretreated with *SMAD3i* 10 μM (or control solvent) during 2 h before TGFβ addition. *SMAD3* and HSC70 serve as loading control for Western blot experiments.
- K Inhibitory effect of *SMAD3i* (SIS3) on the *SMAD*-responsive luciferase activity. Vector encodes the Firefly luciferase reporter gene under the control of a minimal (m) CMV promoter and tandem repeats of the *SMAD* Binding Element (SBE). Cells (10,000 per well) were pretreated with *SMAD3i* 10 μM or control solvent during 1.5 h, and next cells were exposed to TGFβ 10 ng/ml (or solvent: HCl 4 mM + Bovine serum albumin 1 mg/ml) for 6 h. *n* = 3 biologically independent experiments. Each histogram represents the mean ± s.d.; Bilateral Student test (with non-equivalent variances); TGFβ vs TGFβ+*SMAD3i*, **P* < 0.05, ***P* < 0.01.
- L *SMAD3*, Phospho-*SMAD3* Ser423/425, and *BIRC3* expression levels in melanoma cell lines. Four subtypes of melanoma cells (U, undifferentiated; NC, neural crest-like; T, transitory; M, melanocytic) have been compared. 501Mel and M249 cells are melanocytic cells in contrast to dedifferentiated BRAFi-resistant cells (R). Three couples of melanoma cell lines (Sensitive (S) and (R)) have been used to illustrate the *SMAD3* and *BIRC3* increases in BRAFi-resistant cell lines. HEK293 cells are used as control (kidney). HSC70 serves as loading control for Western blot experiments. Each antibody has been evaluated on individual membrane (explaining the three HSC70, loading controls).
- M The vemurafenib decreased melanoma cell density. Cell lines have been exposed to Vem (1 or 5 μM, 84 h) to define two groups of cell lines (sensitive (blue) and "resistant" (red) cell lines). Normal human melanocytes (NEHM) have been used to evaluate the effect of Vem on normal cells (mean of 3 independent donors). Representative values (mean of biological triplicates) of *n* = 3 biologically independent experiments.
- N The *SMAD3* inhibitor decreased melanoma cell density. Cell lines have been exposed to *SMAD3i* (0, 3, 10, 15, or 20 μM, 84 h). The cell lines defined as S and R to BRAFi in the item 5 M are indicated in blue and red. Normal human melanocytes (NEHM) have been used to evaluate the effect of *SMAD3i* on normal cells (*n* = 3 independent donors). The *SMAD3i* effect on NHEMs is weak and manageable for these normal cells (NHEM). Representative values (mean of biological triplicates) of *n* = 2 biologically independent experiments.
- O The chemical inhibition of *SMAD3* by SIS3 (*SMAD3i*) improved current therapy effect (BRAFi + MEKi; Vem 5 μM and Cobi 1 μM) on BRAFi-resistant cells (SKMel28R, M229R & M238R). Cells have been treated as detailed for panel M. Representative values (mean of biological triplicates) of *n* = 2 biologically independent experiments.

Data information: Western blot results are representative of at least two independent experiments. Source data and unprocessed original blots are available in Appendix Fig S2 source data. See also Fig EV3.
Source data are available online for this figure.

immunoprecipitation coupled to massively parallel DNA sequencing data (AhR ChIP-Seq) showing AhR binding on *SMAD3* promoter (Lo & Matthews, 2012; Yang et al, 2018). To demonstrate the *SMAD3* induction by AhR, 501Mel cells were exposed to the most potent and well-known AhR ligands (TCDD for 2,3,5,7-tetrachlorodibenzo-dioxin; ITE for 2-(1*H*-Indol-3-ylcarbonyl)-4-thiazolecarboxylic acid methyl ester). These AhR ligands increased *SMAD3* expression, in an AhR-dependent manner (Fig 6B). Comparable results were obtained with a canonical target gene of AhR; the TCDD-induced poly(ADP ribose) polymerase gene (*TIPARP*), supporting the role of AhR in regulating *SMAD3* expression (Fig 6C). The need of an activated AhR-promoting *SMAD3* expression was further confirmed by the use of an AhR antagonist (CH-223191) in SKMel28 cells

(Fig 6D). Long-term chemical inhibition of AhR activity reduced *SMAD3* and *TIPARP* expression levels. In accordance with these results, *SMAD3* expression levels decreased in AhR KO SKMel28 cells (Fig 6E).

Together, our results support the hypothesis that AhR activity drives *SMAD3* expression along with the acquisition of BRAFi resistance.

SMAD3 drives phenotype switching and resistance to melanoma therapies

To explore the mechanism underlying therapy sensitization upon *SMAD3* inhibition, we examined the transcriptional program

regulated by SMAD3 (Fig 7). We hypothesized that the transcription factor SMAD3 may regulate the expression levels of several resistant genes and thereby induce a multifactorial effect, in which multiple drug resistance pathways are activated. We compared the SMAD3 ChIP-Seq (Ramachandran *et al*, 2018) data to BRAFi-resistance gene sets established from three different sources, namely the enclosed screen, the screen performed in A375 (Koneremann *et al*, 2015), and other established BRAFi-resistance genes such as *AXL* or *NRP1* (Fig 7A). We established a list of SMAD3-regulated genes, which includes *SLIT2*, *RUNX2*, *NRP1*, *MMP2*, *JUNB*, *ITGB5*, *AXL*, *AFAP1*,

and *EGFR* (SMAD3-signature). We next showed that basal SMAD3-signature is higher in the three BRAFi-resistant cell lines when compared to parental cell lines (SKMel28S, M229S, and M238S) (Fig 7B). The stimulation of the TGF β -SMAD3 pathway by the recombinant TGF β promoted the SMAD3 signature in these melanoma cell lines (Fig 7C). The inducibility is higher in parental cell lines (S) since the basal expression level of genes forming the SMAD3 signature is weak in parental cells (Fig 7B). Next, we examined the SMAD3 signature in a large panel of melanoma cell lines representative of the distinct differentiation states (U - NC -T - M)

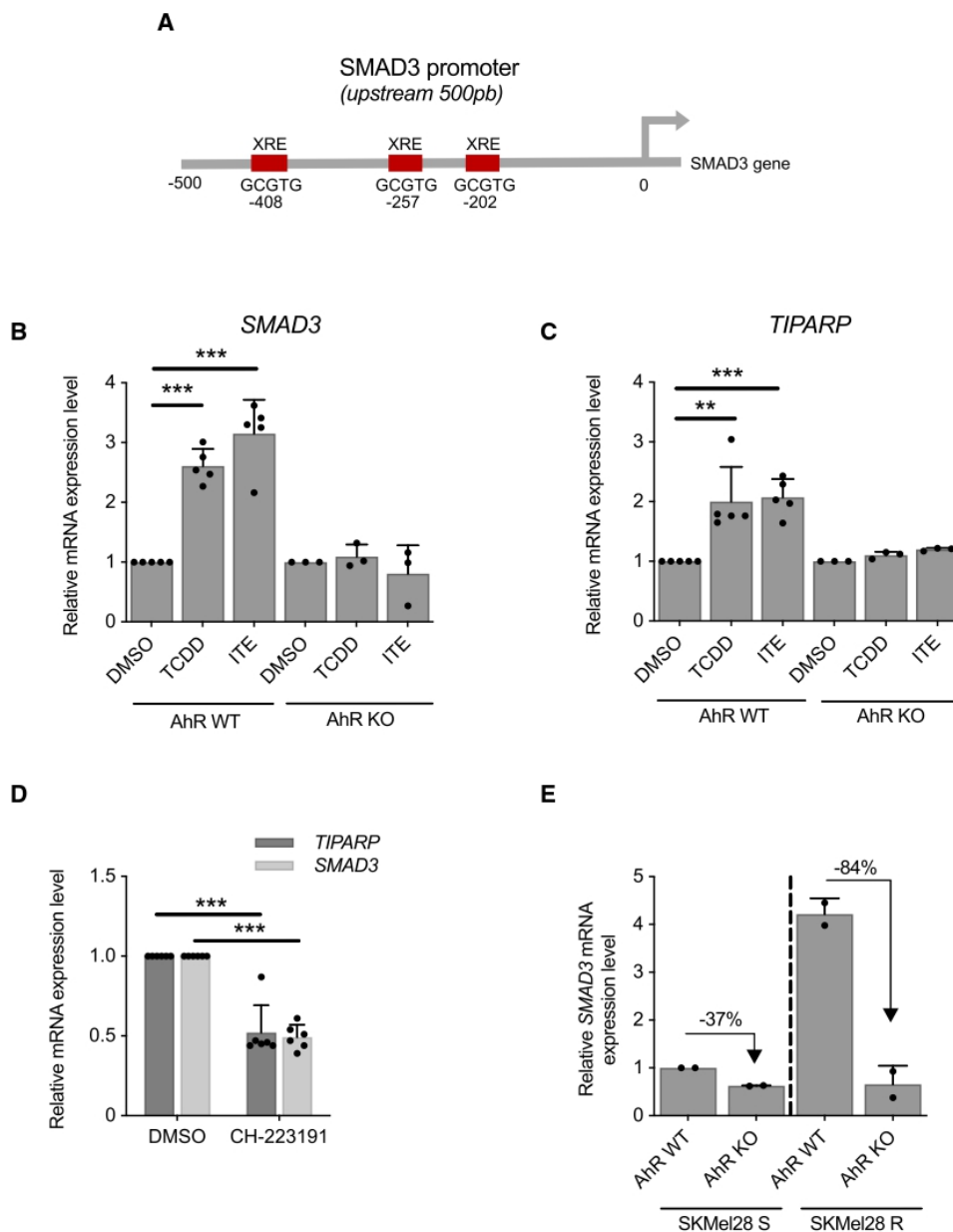


Figure 6.

Figure 6. The Transcription Factor AhR Drives SMAD3 Expression.

- A AhR binding sites (xenobiotic responsive element (XRE); GCCTG) in human SMAD3 proximal promoter.
- B AhR activation by exogenous and endogenous ligands promotes SMAD3 induction. 501Mel cells AhR wild-type or knockout have been exposed to exogenous and endogenous AhR ligands; TCDD (5 nM) or ITE (10 μM) or the solvent (DMSO) during 10 days. $n = 5$ biologically independent experiments for AhR WT cells and $n = 3$ for AhR KO cells. Each histogram represents the mean \pm s.d.; multiple comparisons have been done using ordinary one-way ANOVA, $***P < 0.001$.
- C AhR activation by exogenous and endogenous AhR ligands promotes TIPARP induction. 501Mel cells have been treated as described in B. $n = 5$ biologically independent experiments for AhR WT cells and $n = 3$ for AhR KO cells. Each histogram represents the mean \pm s.d.; multiple comparisons have been done using ordinary one-way ANOVA, $**P < 0.01$, $***P < 0.001$.
- D AhR antagonist (CH-223191) reduces SMAD3 and TIPARP expression levels. SKMel28 cells (AhR wild type) have been exposed to CH-223191 (5 μM) or the solvent (DMSO) during 7 days. $n = 6$ biologically independent experiments. Each histogram represents the mean \pm s.d.; Bilateral Student test (with non-equivalent variances); $***P < 0.001$.
- E Loss of AhR reduces SMAD3 expression levels. SMAD3 expression has been investigated in SKMel28 cells AhR wild-type (WT) or knockout (KO). SKMel28R has been obtained from SKMel28S by chronic exposure to non-lethal doses of BRAFi (Hugo et al, 2015b). R for BRAFi-resistant SKMel28 cells and S for sensitive. $n = 2$ biologically independent experiments. Each histogram represents the mean \pm s.d.

Source data are available online for this figure.

($n = 53$) (Fig 7D) and cutaneous melanoma ($n = 118$, TCGA cohort, tumors not exposed to targeted therapy; Fig 7E). The SMAD3 signature correlated with a dedifferentiation status as suggested above (Figs 2G and 5B, and EV4A). Importantly, a subset of BRAF(V600E) melanoma patients (~20%) expressing the SMAD3-signature was identified (Fig 7E), suggesting that these tumors contained dedifferentiated melanoma cells with potential intrinsic resistance to BRAFi. In addition, SMAD3 signature could be useful to clinicians to propose immune checkpoint immunotherapy to their patients since SMAD3 signature identified anti-PD1 non-responders (prior the selection of the treatment) (Fig 7F). Therefore, these results strongly suggest that the SMAD3-signature may be useful to identify a population of pre-existing drug-resistant cells within drug-naïve lesions. To further illustrate the clinical relevance of our results, we assessed the expression levels of the SMAD3 signature in drug-naïve and BRAFi-resistant patients using publicly available dataset. SMAD3 signature increased in 14/16 patients exposed to BRAFi (baseline vs. relapse, Fig 7G). Altogether, these results indicated that the SMAD3 signature is associated with resistance to both current melanoma therapies.

The analyses of the SMAD3 signature in tumor samples highlighted the tumor heterogeneity (mRNA expression level of genes forming the SMAD3 signature is variable between patients, and all genes forming the SMAD3 signature are not high in each tumor.). As expected, this heterogeneity has been retrieved in melanoma cell lines (Fig 7H). In response to SMAD3 depletion, *AXL* and *EGFR* decreased in the two models in contrast to other genes such as *MMP2*, which decreased in only one model. Altogether, our results indicate that the SMAD3 signature assessment is probably more appropriate than quantification of specific mRNAs such as *EGFR* or *SLIT2* to track pre-existing drug-resistant cells within drug-naïve lesions.

As the shift toward the mesenchymal-like state confers broad resistance to therapies (Redfern et al, 2018), we postulated that the SMAD3 signature may be associated with this particular dedifferentiated phenotype. Comparing the SMAD3 signature with the classical mesenchymal-like signature (Mak et al, 2016) of melanoma (TCGA cohort) highlighted a significant correlation (Figs 7I and EV4A). To further confirm the link between the SMAD3 signature and a mesenchymal state, we searched for similarities with other mesenchymal states identified in two other cancers (glioblastoma [GBM] and hepatoma). As described for the cutaneous melanoma, different differentiation states have been characterized for GBM

(proneural, classical, and mesenchymal GBM) (Jin et al, 2017). We found that the SMAD3 signature overlaps with the mesenchymal GBM signature (Fig 7J). Mesenchymal GBM is the most aggressive GBM usually associated with poor overall survival (Patel et al, 2014; Jin et al, 2017). The SMAD3 signature was also associated with epithelial-mesenchymal transition (EMT) in hepatoma (Fig EV4B and C).

Altogether, these results indicate that the transcription factor SMAD3 and its downstream target genes confer resistance to targeted therapies by promoting a mesenchymal-like phenotype. Our work identifies AhR-SMAD3 axis as a target to overcome therapy resistance of melanoma.

Discussion

Targeted therapy and immunotherapy have greatly improved the prognosis of patients with cancer, but resistance to these treatments restricts the overall survival of patients. Increasing evidence indicates that transcriptomic reprogramming is associated with persister cell emergence (Puisieux et al, 2014; Bai et al, 2019) but the mechanism underlying resistance from this pool of cells remains elusive. Targeting tumor-promoting genes leading reprogramming could therefore constitute an attractive approach to prevent relapse, at least in some specific contexts (Bailey et al, 2018). Here, using a whole genome approach, we searched for pathways that trigger the transcriptional reprogramming of persister cells into drug-resistant cells. Based on CRISPR screen, data mining, and *in vivo* experiments, we identified and validated three genes (*SMAD3*, *BIRC3*, and *SLC9A5*) able to promote both BRAFi-resistance and tumor growth. Our work expands our understanding of the biology of persister cells and highlights novel drug vulnerabilities that can be exploited to develop long-lasting antimelanoma therapies.

Even if CRISPR-SAM screen is a leading-edge genetic tool, several concerns must be considered. As observed for all screening approaches, false positives and false negatives are engendered rendering the validation experiments a crucial step. In this study, we clearly showed that the number of sgRNAs per target is an important parameter. For our best hit, *EGFR*, only two sgRNAs were enriched in BRAFi-treated cells. Thus, it is highly likely that we missed interesting BRAFi-resistance genes (false negatives) due to the number of sgRNA/gene (at least 3 sgRNAs/gene in this library).

A recent publication confirmed that sgRNAs are not all functional in CRISPRa libraries and it could be interesting to increase the number of sgRNAs per target and to cover more TSS per gene (Sanson *et al.*, 2018). Interestingly, the sgRNA library used in our study displays for several genes up to 27 sgRNAs. These sgRNAs target different isoforms (or TSS) of these genes. By examining the sgRNAs targeting *SMAD3*, we found that 2 sgRNA are enriched (BRAFi resistance)

(Fig EV5). These two sgRNAs promote the expression of the longest *SMAD3* isoform; the *SMAD3* mRNA expressed in our model (501Mel cells) (Fig EV5A–E). It is important to note that the nine other sgRNAs targeting *SMAD3* are not able to confer the BRAFi resistance since they target other *SMAD3* isoforms. Based on these observations, we believe that mRNA isoforms identified by RNA sequencing should be considered during the sgRNAs selection for each model.

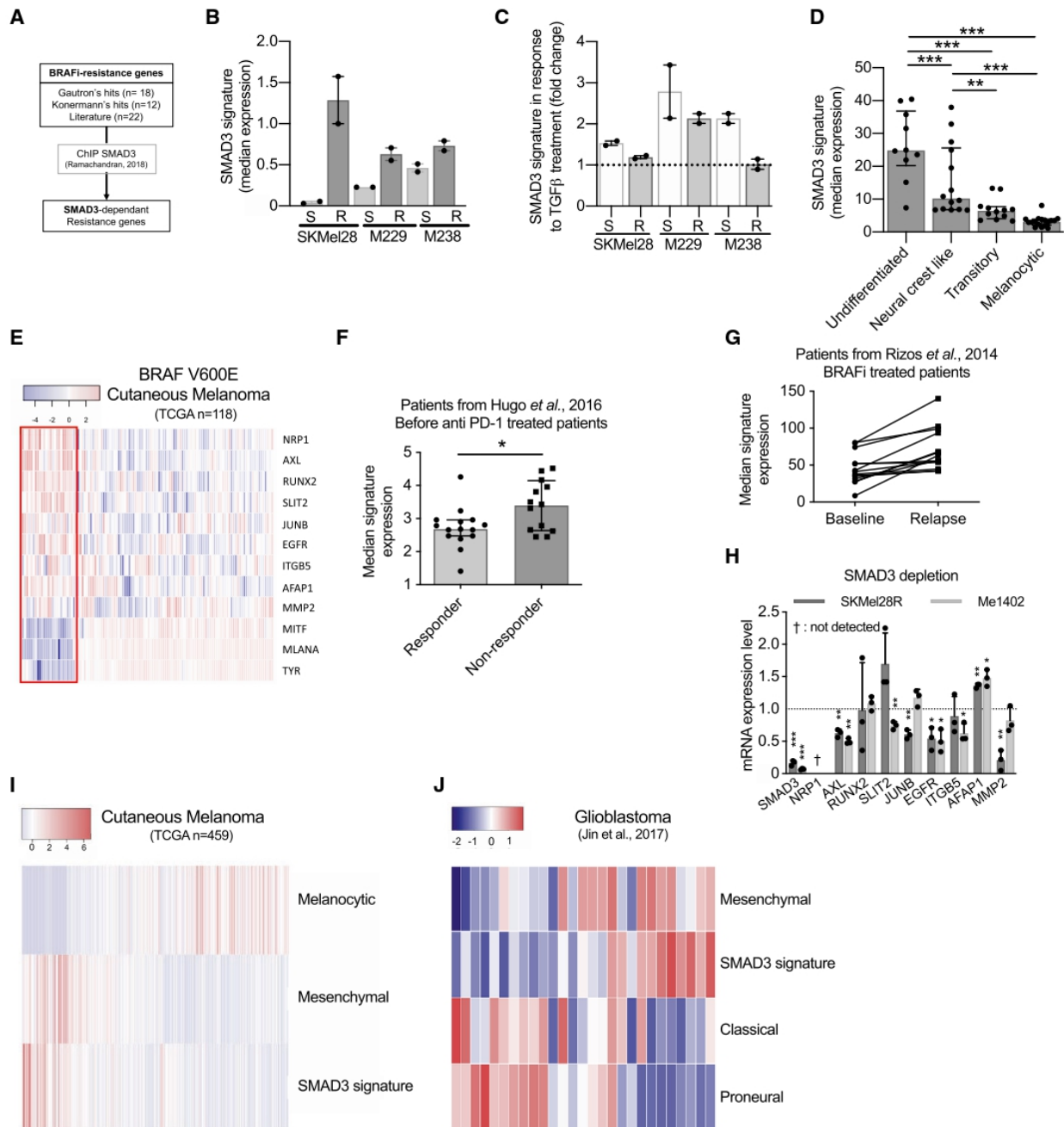


Figure 7.

Figure 7. SMAD3 drives phenotype switching and resistance to melanoma therapies.

- A SMAD3 signature has been established by comparing BRAFi-resistance genes and SMAD3-regulated genes identified by chromatin immunoprecipitation followed by DNA sequencing (ChIP-Seq) (Ramachandran et al, 2018).
- B Basal SMAD3 signature in six melanoma cell lines (S for sensitive to BRAFi and R for Resistant). $n = 2$ biologically independent experiments. Each histogram represents the median with interquartile range.
- C Inducibility of SMAD3-signature in six melanoma cell lines exposed to TGF- β (10 ng/ml, 48 h). Data were normalized to cell lines exposed to solvent (4 mM HCl + 1 mg/ml human BSA). Values obtained with the TGF- β stimulated M238R cell line have been set to 1. $n = 2$ biologically independent experiments. Each histogram represents the median with interquartile range.
- D SMAD3 signature discriminates differentiation states of 53 melanoma cell lines (Tsoi et al, 2018). The SMAD3 signature in four subgroups of melanoma cell lines. Each point represents a cell line ($n = 17, 12, 14,$ and $10,$ respectively, for melanocytic, transitory, neural crest-like, and undifferentiated cell lines); each histogram represents the median with interquartile range; multiple comparisons have been done using ordinary one-way ANOVA; $**P < 0.01,$ $***P < 0.001.$
- E Heat map depicting mRNA levels of SMAD3 signature in BRAF (V600E) non-treated melanoma patients (dataset from TCGA; SKCM, BRAF(V600E) mutated; $n = 118$). Three pigmentation genes (*MITF*, *MLANA*, and *TYR*) have been added to highlight the differentiation states of tumors. ~20% of tumors are considered as dedifferentiated tumors (red box) with a high SMAD3 signature. Scale corresponds to Z scores.
- F SMAD3 signature in pre-treatment biopsies from responders and non-responders to anti-PD-1 treatment (from Hugo's cutaneous melanoma cohort, $n = 15$ and $13,$ respectively) (Hugo et al, 2016). Each histogram represents the median with interquartile range, one-tailed Mann-Whitney test; $*P = 0.0324.$
- G SMAD3 signature in two groups (before BRAFi treatment or during relapse) of V600E patients from Rizos's cohort (Rizos et al, 2014) (14 on 16 patients displayed an upregulation of SMAD3 signature during relapse).
- H SMAD3 depletion decreases expression of SMAD3-regulated genes. SMAD3 knock-down by siRNA decreased mRNA expression of BRAFi-resistance genes in SKMel28R and Me1402. *NRP1* mRNA was not detected in our experimental conditions. $n = 3$ biologically independent experiments. Each histogram represents the mean \pm sd.; Bilateral Student test (with non-equivalent variances) $*P < 0.05,$ $**P < 0.01,$ $***P < 0.001.$ Dotted line highlights the value of 1.
- I SMAD3 signature overlapped with mesenchymal signature in melanoma tumors (TCGA, $n = 459$) (Akbari et al, 2015). Melanocytic signature highlights the differentiation states of tumors (Corre et al, 2018). SMAD3 signature and mesenchymal signature correlate in melanoma tumors (TCGA, $n = 459$) (Mak et al, 2016). Scale corresponds to Z scores.
- J The SMAD3 signature overlaps with the glioblastoma mesenchymal subtype (Jin et al, 2017). Scale corresponds to Z scores.
- Source data are available online for this figure.

By this way, it would be easy to disqualify a part of the sgRNAs of the library targeting weakly expressed isoforms. This attitude could increase the score per isoform. In other words, the use of the z-score calculated per gene may discard interesting candidates. So, an analysis based on RNA isoforms could reduce the false negatives.

The second lesson of this CRISPR screen is the weak sgRNAs enrichment in BRAFi-exposed cells. Except for *EGFR*, the sgRNAs enrichment is about 2. Despite these values, we showed that these candidates are robust as for *SMAD3* (*in vitro*, *in vivo*, and in patients). It is tempting to explain this fact by the duration of treatment (BRAFi exposure) and the dose (2 μ M). By increasing the dose (i.e., 5 μ M), we showed that BRAFi killed more than 90% of 501Mel cells in four days (Fig 5M). This protocol is not achievable because it would induce too many false negatives. The other option consists to increase the duration of BRAFi treatment (and keep a low BRAFi concentration, i.e., 2 μ M). However, a recent publication demonstrated that a long-term BRAFi exposure promotes a dedifferentiation process conferring BRAFi resistance (Tsoi et al, 2018; Bai et al, 2019). This alternative protocol could be perilous by inducing cell resistance to BRAFi independently of the sgRNA expression. Here, we selected a melanoma cell model exhibiting a differentiated profile as the vast majority of metastatic melanoma tumors (89% in the TCGA cohort), a short period of treatment (14 days) and an intermediate dose of BRAFi (2 μ M). In fact, we followed, except the cell line, the protocol established by Feng Zhang's laboratory, who developed the CRISPR-SAM library (Koner-mann et al, 2015). The differentiation status of the cell line could be important since the transactivation mediated by CRISPR-SAM is possible only for active promoter in basal condition and the magnitude of transactivation relies on the basal expression level. Here, we showed that 501Mel cells express low level of *SMAD3* in basal condition and the transactivation obtained by CRISPR-SAM is substantial (Fig 1I). Moreover, SMAD3 is a transcription factor

which promotes the expression of various genes including potent BRAFi-resistance genes (*AXL* and *EGFR*). The robust transactivation of genes encoding a transcription factor, a transporter, or an enzyme is more inclined to be enriched with our protocol, especially if the basal gene expression is low. Here, we identified and validated *in vitro* and *in vivo* the transcription factor SMAD3 and the transporter SLC9A5, validated *in vitro* and *in vivo*.

BRAFi resistance relies, at least in part, on the phenotypic plasticity of melanoma cells (Tsoi et al, 2018; Rambow et al, 2018; Corre et al, 2018). These cells may escape the deleterious effect of drug combinations such as BRAFi + MEKi. Among these cells, those harboring a mesenchymal-like phenotype (usually named invasive cells) display high intrinsic resistance to MAPK therapeutics (Konieczkowski et al, 2014; Müller et al, 2014; Verfaillie et al, 2015; Shaffer et al, 2017). Enrichment in AXL^{high} subpopulation (considered as invasive and mesenchymal-like cells) is a common feature of drug-resistant melanomas (Müller et al, 2014; Konieczkowski et al, 2014). Targeting mesenchymal-like cells using an antibody-drug conjugate, AXL-107-MMAE, showed promising effects in a preclinical model of melanoma (Boshuizen et al, 2018). The emergence of AXL^{high} cells is currently explained by the decrease in *MITF* activity, but the mechanism of resistance to MAPK therapeutics remains unclear. Here, we demonstrate that the AhR-SMAD3 axis governs the expression levels of potent BRAFi-resistance genes, including *AXL*, *EGFR*, and *MMP2*.

The dedifferentiation process, conferring BRAFi resistance, requires transcriptomic reprogramming by transcription factors. Retinoic acid receptor gamma (RXR γ) was identified as a crucial transcription factor that promotes the emergence of the drug-tolerant subpopulation of NCSCs (Rambow et al, 2018). An increase in AXL^{high}-positive cell population was reported following MAPK inhibition in the presence of an RXR γ antagonist. This increase may explain why this co-treatment only delays but does not completely

prevent relapse in PDXs (preprint: Marin-Bejar *et al.*, 2020). This confirms the need to develop strategies that prevent melanoma dedifferentiation during BRAFi treatment. Thus, our data strongly suggest that SMAD3 is a key transcriptional factor involved in the emergence of drug-resistant mesenchymal-like cells in response to MAPK and identify a clinically compatible approach (SMAD3i) that might abrogate such a trajectory. Other transcription factors have been associated to BRAFi resistance such as JUN (Titz *et al.*, 2016) and AhR (Liu *et al.*, 2017; Corre *et al.*, 2018). We recently showed that a sustained AhR activation promotes the dedifferentiation of melanoma cells and the expression of BRAFi-resistance genes (Corre *et al.*, 2018). As proof-of concept, we demonstrated that differentiated and BRAFi-sensitive cells can be directed toward an AhR-dependent resistant program using AhR agonists. To abrogate the deleterious AhR sustained-activation, we identified Resveratrol, a clinically compatible AhR antagonist. Combined with BRAFi, Resveratrol reduces the number of BRAFi-resistant cells and delays relapse. Recently, an independent team confirmed that AhR inhibition is reachable *in vivo* using other AhR antagonists (Kyn 101, Ikena Oncology, or the CH-223191; Campesato *et al.*, 2020) to improve the melanoma therapy. AhR antagonists validated in mice are currently evaluated in clinical trials (NCT04200963, Ikena Oncology) and (NCT04069026, Bayer). In addition, a water-soluble SMAD3 inhibitor has been recently published for the *in vivo* treatment (Wu *et al.*, 2020), suggesting that clinical trials should start soon. Another option overcoming the potential pharmacological caveat would be to use antisense oligonucleotides (ASO) targeting *AhR* or *SMAD3* (Leclerc *et al.*, 2021). We and others demonstrated in melanoma that ASO strategy is feasible *in vivo* by targeting *SAMMSON* mRNA and the lncRNA *TYRP1* (Leucci *et al.*, 2016; Gilot *et al.*, 2017).

In this study, we propose an AhR-SMAD3 impairment as a strategy to overcome melanoma resistance. Recently, conditional deletion of *Smad7*, a negative regulator of TGF- β /SMAD pathway, led to sustained melanoma growth and at the same time promoted metastasis formation (Tuncer *et al.*, 2019), confirming that TGF- β /SMAD pathway is a promising target for melanoma (Javelaud *et al.*, 2007). In addition, Rizos's team further illustrated the link between the TGF- β and melanoma therapy resistance. They showed that TGF- β promotes a dedifferentiation phenotype, which is a common mechanism of resistance to PD-1 inhibitors (Lee *et al.*, 2020).

Several questions remain unsolved. We previously reported that activated AhR reprograms the transcriptome of melanoma cells mediating BRAFi resistance. In this study, we demonstrate that a SMAD3-regulated gene expression program promotes therapy resistance in cutaneous melanoma and EMT. Importantly, *SMAD3* expression levels during resistance acquisition are dependent, at least in part, on AhR. Thus, it would be interesting to precisely define the role of AhR and SMAD3 in the induction of each BRAFi-resistance gene. To date, no physical interaction between AhR and SMAD3 proteins has been reported, suggesting that AhR acts as an upstream regulator of SMAD3 axis. It is noteworthy that the increased expression levels of SMAD3 by AhR expands the possibility of fine tuning gene expression since SMAD3 interacts with SMAD2 but also with JUN, TEADs, and YAP1 (Zhang *et al.*, 1998; Fujii *et al.*, 2012; Piersma *et al.*, 2015). Because these three transcription factors have also been associated with therapies resistance in melanoma (Nallet-Staub *et al.*, 2014; Ramsdale *et al.*, 2015; Verfaillie

et al., 2015; Hugo *et al.*, 2015b), our results suggest that regulation of BRAFi-resistance genes expression is multiparametric and probably more sophisticated than initially thought. Nonetheless, the elucidation of these transcriptional programs and networks governing BRAFi-resistance genes and relapse is important for optimal target selection and the development of rationale and effective combination strategies.

BRAFi resistance may be achieved through the exposure of melanoma cells to TGF- β , demonstrating that transcriptome reprogramming may confer resistance without the need for pre-existing or *de novo* mutations (Viswanathan *et al.*, 2017). The TGF- β pathway promotes a shift toward the mesenchymal state (Antony *et al.*, 2019). The resulting dedifferentiation modifies the expression of the adhesion molecules in the cell, supporting a migratory and invasive behavior. Together, our results strongly indicate that the SMAD3-regulated genes are critical players in melanoma resistance to therapies by promoting an EMT-like process. EMT reversal represents a powerful approach, as it may reduce the invasive behavior of cancer cells and favor re-differentiation, synonymous of a decrease in BRAFi-resistance gene expression (Giannelli *et al.*, 2014; Rodón *et al.*, 2015). By combining anti-EMT drug and targeted therapy such as SMAD3i and BRAFi, we should efficiently reduce amount of persister cells. We anticipate that SMAD3 inhibition should limit the risk of resistance to therapies, since a decrease of expression levels of several BRAFi-resistance genes is obtained with SMAD3i (SIS3). SMAD3 inhibition is expected to be more efficient than inhibitors targeting single downstream targets such as AXL or EGFR. In conclusion, our work highlights novel drug vulnerabilities that can be exploited to develop long-lasting antimelanoma therapies.

Given the plasticity of melanoma cells and the capability of tumor microenvironment to produce TGF- β (Chakravarthy *et al.*, 2018), our work also warrants further investigation of the source of TGF- β as another approach to prevent acquisition of the therapy-resistant mesenchymal phenotype.

Materials and Methods

Reagents

- DMSO—Sigma-Aldrich (D8418)
- BRAF inhibitors: Vemurafenib (PLX4032)—Selleckchem (RG7204); Paradox Breaker (PLX8394)—MedChemExpress (HY-18972)
- SMAD3 inhibitor: SIS3—Santa Cruz Biotechnology (sc-222318)
- MEK inhibitor: Cobimetinib—Selleckchem (GDC-0973)
- 2,3,7,8-TCDD (TCDD)—Sigma-Aldrich (48599)
- 2-(1*H*-Indol-3-ylcarbonyl)-4-thiazolecarboxylic acid methyl ester (ITE)—MedChemExpress (HY-19317)
- CH-223191—Selleckchem (S7711)
- TGF- β recombinant—Santa Cruz Biotechnology (240-B-010)
- BIRC inhibitor: Birinapant—Selleckchem (S7015)

Cell lines and culture conditions

501Mel, Me1402, and HEK293T cell lines were obtained from ATCC. SKMel28 S & R cell lines were obtained from J.C Marine's laboratory at VIB Center for Cancer Biology, VIB, Leuven, Belgium. 501Mel

and SKMel28 AhR knockout cell lines have been established as previously described (Corre *et al.*, 2018). M229S, M229R, M238S, M238R, and M249 were obtained from Thomas Graeber's laboratory at department of Molecular and Medical Pharmacology, University of California, Los Angeles, USA. All melanoma cell lines were grown in humidified air (37°C, 5% CO₂) in RPMI-1640 medium (Gibco BRL, Invitrogen, Paisley, UK) supplemented with 10% fetal bovine serum (FBS) (PAA cell culture company) and 1% penicillin–streptomycin (PS) antibiotics (Gibco, Invitrogen). HEK293T was grown in DMEM (Gibco BRL, Invitrogen, Paisley, UK) supplemented as melanoma cell lines media. SKMel28R, M229R, and M238R are cultivated in presence of 0.1 μM vemurafenib. All cell lines have been routinely tested for mycoplasma contamination (Mycoplasma contamination detection kit; rep-pt1; InvivoGen—San Diego—CA).

CRISPR-SAM screens

A detailed protocol is available as supplementary Methods. Briefly, lentiviral productions have been performed as recommended (<http://tronolab.epfl.ch>), using HEK293T cells, psPAX2, pVSVG, and vectors required for CRISPR-SAM according to Zhang lab (Joung *et al.*, 2017). Infections were performed overnight in presence of 4 or 8 μg of polybrene per ml. All vectors have been provided by Addgene. SgRNA Library was amplified and prepared as described by Zhang Lab (Joung *et al.*, 2017). 501Mel cells were transduced to stably express dCAS-VP64 (cat. no. 61425) and MS2-P65-HSF1 (cat. no. 61426), before to transduce them with SAM sgRNA library (lentiSAMv2, 3-plasmid system, cat. no. 1000000057) at a MOI of 0.2. Infected cells have been selected using antibiotics: Blastidicin (2 μg/ml, 5 days), Hygromycin B (200 μg/ml, 5 days), and Zeocine (600 μg/ml, 5 days). *In vitro* CRISPR-SAM screens (Fig 2) were conducted as described by Zhang laboratory (Joung *et al.*, 2017). The 501Mel cells expressing the sgRNA library were split into three groups: DMSO (solvent for BRAFi), vemurafenib (PLX4032, 2 μM, Selleckchem), and Paradox Breaker (PLX8394, 2 μM, MedChemExpress). After 14 days, resistant cell populations have been amplified. A minimum of 36 × 10⁶ cells has been pellet and further sgRNA enrichment analysis. Cell libraries have been cryopreserved at –80°C for further *in vivo* experiments. The results presented are pooled data from two independent screens. To generate 501Mel cells overexpressing individually *SMAD3*, *BIRC3*, or *SLC9A5*, 501Mel expressing dCAS-VP64 and MS2-P65-HSF1 were transduced to stably express specific sgRNAs (Table EV8). Infected cells have been selected using zeocin (600 μg/ml, 5 days). Manipulations of lentivirus were performed in the biosafety level 3 containment laboratory core facility of the Biology and Health Federative Research Structure of Rennes (Biosit).

Xenograft

Nude mice (4 weeks) were purchased from Janvier Labs and maintained under specific pathogen-free conditions in our accredited animal house (A 35_238_40). The animal study follows the 3R (replace_reduce_refine) framework and has been filed with and approved by the French Government Board (No. 04386.03). Animal welfare is a constant priority: animals were thus euthanized under anesthesia. No blinding was done for animal studies. The animals were randomly allocated into different groups.

To identify tumor-promoting genes *in vivo* (Fig 1D), two cell populations were subcutaneously xenografted on female NMRI nude mice flanks (3 × 10⁶ cells per mouse); 501Mel (6 mice) and 501Mel CRISPR-SAM cell library (10 mice, Table EV2) for Fig 1. For xenograft of other libraries (Fig 4): Vem-persister cells (*n* = 11 mice) and (PLX8394)-persister cells (*n* = 12 mice) (Table EV7). Tumor volume was assessed according to the formula $V = (W(2) \times L)/2$, during 20 weeks. After mice sacrifice, tumors were sampled and conserved at –80°C for further sgRNA enrichment analysis.

For individual validation of tumor-promoting genes (Fig 1H–K): 3 × 10⁶ of 501Mel cells overexpressing either *SMAD3*, *BIRC3*, or *SLC9A5* were subcutaneously xenografted on female NMRI nude mice flanks (6 mice per group). In each group, mice were injected on both flanks: sgRNA 1 on right flank and sgRNA 2 on left flank (see Table EV8 for sgRNA sequences). Tumor growth was assessed as previously described (Gilot *et al.*, 2017) until the endpoint (600 mm³).

In order to identify BRAFi-resistance and tumor growth genes *in vivo* (Fig 4), three cell populations were subcutaneously xenografted on female NMRI nude mice flanks (3 × 10⁶ cells per mouse); 501Mel (6 mice), 501Mel CRISPR-SAM vemurafenib resistant (10 mice), and 501Mel CRISPR-SAM Paradox Breaker resistant (12 mice). Tumor growth was assessed as previously described (Gilot *et al.*, 2017), during 20 weeks. After mice sacrifice, tumors were sampled and conserved at –80°C for further sgRNA enrichment analysis.

These experiments are compliant with all relevant ethical regulations regarding animal research.

sgRNA enrichment analysis

Genomic DNAs from cell pellets (>36.10⁶ cells) and tumors (> 400 mg) were extracted using the Zymo Research Quick-gDNA MidiPrep according to the manufacturer's protocol. PCR amplifications and quality controls have been done as described by Zhang laboratory (Joung *et al.*, 2017).

sgRNA sequencing

Sequencing was performed by the Human & Environmental Genomics core facility of Rennes on a HiSeq 1500 (Rapid SBS kit v2 1x100 cycles, Illumina). Base Calling was performed with Illumina's CASAVA pipeline (Version 1.8).

Bioinformatic analysis of sgRNA and gene hits

Data processing was conducted using the MAGECK v0.5.6 software (Li *et al.*, 2014). Briefly, read counts from different samples are first median-normalized to adjust for the effect of library sizes and read count distributions (mageck count with option: norm-method median). Then, in an approach similar to those used for differential RNA-Seq analysis, the variance of read counts is estimated by sharing information across features and a negative binomial model is used to test whether sgRNA abundance differs significantly between the treatment conditions and the DMSO control. Positively or negatively selected sgRNA are ranked according to adjusted *P*-values (false discovery rate) and gene log fold changes computed with the modified robust ranking aggregation algorithm implemented in MAGECK (mageck test with

options: norm-method median, gene-lfc-method alphamedian, andadjust method fdr).

RNA interference

All siRNAs were transfected at 66 nM using Lipofectamine RNAiMAX (Invitrogen). For survival assay, 4,000 cells were seeded in 96-well plates, in quadruplicates. The following day, cells have been FBS-starved (1% FBS overnight). Thirty-six hours after transfection, cells were exposed to DMSO or BRAFi (vemurafenib (PLX4032); Selleckchem; 1 μ M). After 84 h of treatment, cell density was measured by methylene blue assay, as previously described (Gilot *et al*, 2011). For RNA analysis, 50,000 cells were seeded in 12-well plates. Cells were harvested 48 h after transfection. All siRNAs were purchased from IDT DNA (Table EV8).

Treatment experiments

8,000 (SKMel28R) or 12,000 (Me1402) cells were seeded in 96-well plates, in quadruplicates. For SMAD3i + vemurafenib combination treatment: 5,000 (501Mel) or 8,000 (M249, SKMel28R) cells were seeded in 96-well plates, in triplicates. Cells were exposed, 6 h after seeding, to either DMSO, BRAFi (vemurafenib; 5 μ M), or MEKi (Cobi, 1 μ M) or combination BRAFi (5 μ M) + SMAD3i (0, 3, 10, 15, or 20 μ M or solvent) \pm MEKi (1 μ M) for 4 days. 8,000 (SKMel28R) or 12,000 (Me1402) cells were seeded in 96-well plates, in quadruplicates. For Birinapant treatment (96 h): cells were exposed, 6 h after seeding, to either DMSO or Birinapant (100 nM). Cell density was evaluated using methylene blue assay, as previously described (Gilot *et al*, 2011).

Half maximal inhibitory concentrations (IC₅₀ values)

Cell sensitivity to vemurafenib or SMAD3i (SIS3) has been established by cell density measurement and calculation of the IC₅₀ using GraphPad (PRISM8.0[®]) as previously described (Corre *et al*, 2018). 5,000 (501Mel) or 8,000 cells (other cell lines) were plated and exposed to inhibitor(s) at indicated concentrations for 4 days. Cells have been exposed to inhibitors 6 h after plating.

Melanoma spheroids

Spheroids were prepared using the liquid overlay method. Briefly, 500 μ l of melanoma cells (10,000 cell/ml) were added to a 24-well plate coated with 1.5% agar (Invitrogen). Plates were left to incubate for 72 h; by this time, cells had organized into 3-dimensional (3D) spheroids. Spheroids were then harvested and added into 1 ml of a solution of collagen I (2 mg/ml – Corning) with MEM 1X (Gibco), acetic acid 0.02N and neutralization buffer (HEPES 200 mM pH 7.4; sodium bicarbonate 2.2%; NaOH 0.2 N). The suspension was then added to a 12-well plate coated with 1.5% agar. Normal medium was overlaid on top of the solidified collagen after 2 h of incubation. After 48 h, medium was renewed. Pictures of the invading spheroids were monitored at different times using a Zeiss microscope.

RNA extraction and RT-qPCR expression

Experiments have been done as previously described (Gilot *et al*, 2017). Primers used for RT-qPCR experiments are available in Table EV8.

Western blot experiments

Experiments were performed as previously described (Gilot *et al*, 2017). Membranes were probed with suitable antibodies and signals were detected using the LAS-3000 Imager (Fuji Photo Film). The primary and secondary antibodies are described in Table EV8.

SMAD-luciferase assay

Experiments are based on the SMAD-responsive element: Cignal Lenti SMAD Reporter (luc) (CLS-017L from Qiagen) and as control the Cignal Lenti Negative Control (luc). The SMAD-responsive luciferase vector encodes the Firefly luciferase reporter gene under the control of a minimal (m)CMV promoter and tandem repeats of the SMAD Binding Element (SBE). Melanoma cells line were infected as previously described (Gilot *et al*, 2017), and infected cells were selected using antibiotic selection (puromycin 2 μ g/ml, 7 days). Cells were exposed to TGF β \pm SMAD3i as detailed in Fig 5K legend. Luciferase assays were then performed with a Promega kit according to the manufacturer's instructions. Data were expressed in arbitrary units, relative to the value of luciferase activity levels found in TGF β -exposed cells, arbitrarily set at 100%. Firefly luciferase activity was normalized to protein content using Bicinchoninic Acid Kit from Sigma-Aldrich[®] and measured with using a luminometer (Centro XS3 LB960, Berthold Technologies).

Immunohistochemistry

Tissue samples from representative lesions (preprint: Marin-Bejar *et al*, 2020) were collected and fixed in 4% paraformaldehyde for 24 h and then processed for paraffin embedding (HistoStar[™] Embedding Workstation). Sections of 5 μ m of thickness obtained from the paraffin-embedded tissues (Thermo Scientific Microm HM355S microtome) were mounted on SuperFrost[™] Plus Adhesion slides (Thermo Scientific).

Immunofluorescence with and without tyramide signal amplification

Antibody targeted SMAD3 was used for detecting the following protein: (rabbit, 1:100, Cell Signaling Technologies, #9513, incubation for 30 min at RT).

Furthermore, the Akoya Opal Polaris 7 Color Automation IHC Detection Kit (Akoya, NEL871001KT) was used for the tyramide signal amplification according to the manufacturer's protocol. All dewaxing and staining steps were performed using the Leica Bond RX slide stainer with the standard settings of the Opal 7-color (v5.2 plus) IHC protocol provided by the manufacturer. The steps specific for the staining is as follows: Antigen retrieval was performed for 20 min at 100°C with Bond TM Epitope Retrieval 1 solution (Leica, AR9961). Tissue was blocked for 15 min using a blocking buffer (TBS with 1% BSA (VWR, 22013, bovine serum albumin (BSA), fraction V, biotium (50 g))) with 10% normal goat serum (Invitrogen, 10000C). For introduction of the secondary-HRP, the Envision+/HRP goat anti-Rabbit (Dako Envision + Single Reagents, HRP, Rabbit, Code K4003) was used for the antibody raised in rabbit (SMAD3). The protein SMAD3 was detected using the OPAL 690 reagent.

After the staining in the Leica Bond RX, slides were washed twice for 2 min in TBS-Tween-20 (0.5%) and were then mounted using ProLong Diamond Antifade Mountant (Thermo Fisher, P36961). Numbers of count cells are indicated in Appendix Fig S2 source data file.

In silico analyses

Heat maps were generated with R-packages heatmap3 (Zhao *et al.*, 2014).

Gene Set Enrichment Analysis was performed using the Broad Institute software.

SKCM TCGA expression data were obtained using OncoLNC portal (<http://www.oncolnc.org>; (Anaya, 2016)). Cell state categorization into four differentiation states (Undifferentiated, Neural crest-like, Transitory, Melanocytic) of SKCM TCGA tumors and were performed using expression data of previously established gene sets (Tsoi *et al.*, 2018).

Genomic alterations of SKCM TCGA tumors were analyzed using cBioPortal (<http://www.cbioportal.org>).

CCLC cell lines expression data and IC₅₀ were obtained from GSE36139 and from the original publication (Barretina *et al.*, 2012), respectively.

Expression data of 501Mel were obtained from our previously published RNA-seq (GSE95589) (Gilot *et al.*, 2017).

Expression data of M229, M238, and SKMel28 at different resistance steps were obtained from GSE75313 (Song *et al.*, 2017).

Patient's median expression of our 18 hits in baseline *vs* relapse (BRAFi) is based on expression data from GSE65186 (Hugo *et al.*, 2015b) and GSE50509 (Rizos *et al.*, 2014).

Expression data of invasive *vs* proliferative cell lines were obtained from GSE60666 (Verfaillie *et al.*, 2015).

Analysis of our 18 hits expression in EGFR-positive sorted cells was realized based on GSE97682 (Shaffer *et al.*, 2017).

Analysis of *SMAD3*, *BIRC3*, and *EGFR* expression among 52 cell lines previously categorized as Undifferentiated, Neural crest-like, Transitory, or Melanocytic were performed using <http://systems.crup.ucla.edu/dediff> and GSE80829.

SMAD3 ChIP-Seq data have been obtained from GSE92443 (Ramachandran *et al.*, 2018).

The comparison of the median SMAD3 signature in anti-PD-1 responsive *vs* resistant patients has been performed *via* expression data from GSE78220 (Hugo *et al.*, 2016).

SMAD3 signature median expression was compared to melanocytic one (*MITF*, *OCA2*, *MLANA*, *TYR*, *DCT*) and mesenchymal one (52 genes; (Mak *et al.*, 2016)) *via* expression data from SKCM TCGA obtained from OncoLNC.

The comparison between median expression of SMAD3 signature with classical, proneural, and mesenchymal ones in a cohort of glioblastoma patients was based on expression data from GSE103366 (Jin *et al.*, 2017).

Gene Set Enrichment Analysis on Huh28 ± TGF-β was realized based on GSE102109 (Merdrignac *et al.*, 2018). Gene Set Enrichment Analysis on 3sp cells (mesenchymal) *versus* 3p cells (epithelial) was realized based on GSE26391 (van Zijl *et al.*, 2011).

Single-cell RNA-seq data of a BRAF-mutant PDX model undergoing BRAF and MEK inhibition were retrieved (GSE116237). The 674 single cells were projected into a two-dimensional space using

The paper explained

Problem

Precision medicine has greatly improved the survival of patients with cutaneous melanoma, but relapse limits the long-lasting responses. Relapse is explained by the capability of persister cells to survive despite targeted therapies; however, the molecular mechanism conferring resistance is not fully understood

Results

We performed a genetic screen (CRISPR activation) to identify genes conferring resistance to BRAF inhibitors, promoting survival of persister cells and relapse. The current study finds that genes that enable tumor growth of therapy-naïve melanoma cells are also enriched during acquisition of resistance to BRAFi. SMAD3 targeted genes (SMAD3-signature) could be useful to find subpopulations of pre-existing BRAFi-resistant cells within therapy-naïve lesions. Moreover, chemical inhibition of SMAD3 decreases persister cells population.

Impact

We established in a clinical context that a high expression level of several tumor-promoting genes fuels cancer cell plasticity, therapy resistance, and relapse. By highlighting the role of tumor-promoting genes, this work expands our understanding of the biology underlying tumor growth. Moreover, we identify novel drug vulnerabilities that can be exploited to develop long-lasting anticancer therapies.

t-distributed stochastic neighbor embedding (tSNE) and colored according to their drug-tolerant state (DTC) identity (Rambow *et al.*, 2018). In a second step, the activity of the resistance gene expression signature (18 genes) was quantified per cell using the AUCell algorithm (Aibar *et al.*, 2017) resulting into an AUCell score ($0 < \text{range} < 0.4$), which was used to gradient-color the tSNE plot. Finally, the activity of the resistance gene signature was quantified per DTC population using GraphPad (**** $P < 0.0001$, Mann-Whitney test).

Statistical analyses

Data are presented as mean ± s.d. unless otherwise specified, and differences were considered significant at a *P*-value of less than 0.05. In the box plots, the line within the box is the median, the bottom and top of the box are, respectively, the first and the third quartiles, and the whiskers represent the minimum and maximum of all the data points. Comparisons between groups normalized to a control were carried out using bilateral Student's test (with non-equivalent variances). When at least two factors were compared between two groups, a two-way ANOVA (without adjustment) was used as specified in figure legends. The statistical significance between two independent groups of patient samples was examined using the Mann-Whitney test. All statistical analyses were performed using Prism 8 software (GraphPad, La Jolla, CA, USA) or Microsoft Excel software. All experiments were performed three or more times independently under similar conditions, unless otherwise specified in the figure legends and statistics table containing the raw data (Appendix Fig S2–S3 source data) and Appendix Table S1 (*P*-values).

Data availability

The datasets generated during current study are available: <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8595/>.

Expanded View for this article is available online.

Acknowledgements

The authors thank the Gene Expression and Oncogenesis team from the CNRS UMR6290 especially M. Migault, A. Forestier, and L. Boussemart; the Rennes FHU CAMIn team; and E. Watrin and C. Hitte for providing scientific expertise. The authors acknowledge the SFR Biosit core facilities of Rennes 1 University with the ARCHE animal housing facility, the cell culture L3 facility, and the Human and Environmental Genomics platform for their help and support. This study received financial support from the following: Fondation ARC pour la Recherche (ARC labeled team); Ligue Nationale Contre le Cancer (LNCC); Départements du Grand-Ouest; Région Bretagne; University of Rennes 1; CNRS; SFR Biosit, Association Vaincre le Cancer. Further support was provided by a “Ligue Nationale Contre le Cancer” (LNCC) Grand Ouest fellowship (AG) and from the Région Bretagne (AG), and Fondation ARC pour la Recherche (AG) and the LNCC (AQ) and from French Ministry of Research (NT and DL) and from Institut National contre le Cancer (INCa) (AP). The authors are grateful to Feng Zhang for providing the Human CRISPR 3-plasmid activation pooled library (SAM) (Addgene).

Author contributions

Conceptualization: AG and DG. Methodology: AG, LB, AMQ, MA, and DG. Software: AG, MA, SC, CC, and FR. Formal analysis: AG, LB, AMQ, MA, DL, SC, OM-B, CC, and FR. Investigation: AG, LB, AMQ, AP, MA, DL, SC, AP, HML, OM-B, NT, and DG. Writing-original draft: AG and DG. Writing-review and editing: AG, DG, LB, AMQ, SC, M-DG, FR, MA, CC, and J-CM. Visualization: AG, LB, AMQ, MA, SC, FR, CC, and DG. Supervision: DG. Project Administration: DG and M-DG. Funding: DG, SC, and M-DG.

Conflict of interest

The authors declare that they have no conflict of interest.

For more information

- The Graeber laboratory provides a webtool to investigate the expression profile of your favorite RNAs in 53 melanoma cell lines, classified in four categories based on the differentiation status.
- <https://systems.crump.ucla.edu/dediff/>

References

Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J et al (2017) SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 14: 1083–1086

Akbani R, Akdemir KC, Aksoy BA, Albert M, Ally A, Amin SB, Arachchi H, Arora A, Auman JT, Ayala B et al (2015) Genomic classification of cutaneous melanoma. *Cell* 161: 1681–1696

Anaya J (2016) OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *PeerJ Comput Sci* 2: e67

Antony J, Thiery JP, Huang RY-J (2019) Epithelial-to-mesenchymal transition: Lessons from development, insights into cancer and the potential of EMT-subtype based therapeutic intervention. *Phys Biol* 16(4): 041004

Ascierto PA, McArthur GA, Dréno B, Atkinson V, Liskay G, Di Giacomo AM, Mandalà M, Demidov L, Stroyakovskiy D, Thomas L et al (2016) Cobimetinib combined with vemurafenib in advanced BRAFV600-mutant melanoma (coBRIM): updated efficacy results from a randomised, double-blind, phase 3 trial. *Lancet Oncol* 17: 1248–1260

Bai X, Fisher DE, Flaherty KT (2019) Cell-state dynamics and therapeutic resistance in melanoma from the perspective of MITF and IFN γ pathways. *Nat Rev Clin Oncol* 16: 549–562

Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B et al (2018) Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* 174: 1034–1035

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D et al (2012) The Cancer Cell Line Encyclopedia enables predictive modeling of anticancer drug sensitivity. *Nature* 483: 603–607

Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, Santos R, Rao Y, Sassi F, Pinnelli M et al (2019) Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* 568: 511–516

Berking C, Takemoto R, Schaidler H, Showe L, Satyamoorthy K, Robbins P (2001) Transforming growth factor- β 1 increases survival of human melanoma through stroma remodeling. *Cancer Res* 61: 8306–8316

Boshuizen J, Koopman LA, Krijgsman O, Shahrabi A, Van Den Heuvel EG, Ligtenberg MA, Vredevoogd DW, Kemper K, Kuilman T, Song JY et al (2018) Cooperative targeting of melanoma heterogeneity with an AXL antibody-drug conjugate and BRAF/MEK inhibitors. *Nat Med* 24: 203–212

Campesato LF, Budhu S, Tchaicha J, Weng C-H, Gigoux M, Cohen IJ, Redmond D, Mangarin L, Pourpe S, Liu C et al (2020) Blockade of the AHR restricts a Treg-macrophage suppressive axis induced by L-Kynurenine. *Nat Commun* 11: 4011

Chakravarthy A, Khan L, Bensler NP, Bose P, De Carvalho DD (2018) TGF- β -associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. *Nat Commun* 9: 4692

Chihara Y, Shimoda M, Hori A, Ohara A, Naoi Y, Ikeda J, Kagara N, Tanei T, Shimomura A, Shimazu K et al (2017) A small-molecule inhibitor of SMAD3 attenuates resistance to anti-HER2 drugs in HER2-positive breast cancer cells. *Breast Cancer Res Treat* 166: 55–68

Corre S, Tardif N, Mouchet N, Leclair HM, Boussemart L, Gautron A, Bachelot L, Perrot A, Soshilov A, Rogiers A et al (2018) Sustained activation of the Aryl hydrocarbon Receptor transcription factor promotes resistance to BRAF-inhibitors in melanoma. *Nat Commun* 9: 4775

Fujii M, Toyoda T, Nakanishi H, Yatabe Y, Sato A, Matsudaira Y, Ito H, Murakami H, Kondo Y, Kondo E et al (2012) TGF- β synergizes with defects in the Hippo pathway to stimulate human malignant mesothelioma growth. *J Exp Med* 209: 479–494

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E et al (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6: p11

Giannelli G, Villa E, Lahn M (2014) Transforming growth factor- β as a therapeutic target in hepatocellular carcinoma. *Cancer Res* 74: 1890–1894

Gilot D, Le Meur N, Giudicelli F, Le Vée M, Lagadic-Gossmann D, Théret N, Fardel O (2011) RNAi-based screening identifies kinases interfering with dioxin-mediated up-regulation of CYP1A1 activity. *PLoS One* 6: e18261

Gilot D, Migault M, Bachelot L, Journé F, Rogiers A, Donnou-Fourmet E, Mogha A, Mouchet N, Pinel-Marie ML, Mari B et al (2017) A non-coding function of TYRP1 mRNA promotes melanoma growth. *Nat Cell Biol* 19: 1348–1357

Halaban R, Zhang W, Bacchiocchi A, Cheng E, Parisi F, Ariyan S, Krauthammer M, McCusker JP, Kluger Y, Szoln M (2010) PLX4032, a

- selective BRAFV600Ekinase inhibitor, activates the ERK pathway and enhances cell migration and proliferation of BRAFWTmelanoma cells. *Pigment Cell Melanoma Res* 23: 190–200
- Hoek KS, Goding CR (2010) Cancer stem cells versus phenotype-switching in melanoma. *Pigment Cell Melanoma Res* 23: 746–759
- Hugo W, Shi H, Sun L, Piva M, Song C, Kong X, Moriceau G, Hong A, Dahlman KB, Johnson DB et al (2015a) Non-genomic and immune evolution of melanoma acquiring MAPKi resistance HHS public access. *Cell* 162: 1271–1285
- Hugo W, Shi H, Sun L, Piva M, Song C, Kong X, Moriceau G, Hong A, Dahlman KB, Johnson DB et al (2015b) Non-genomic and immune evolution of melanoma acquiring MAPKi resistance. *Cell* 162: 1271–1285
- Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G et al (2016) Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* 165: 35–44
- Javelaud D, Mohammad KS, McKenna CR, Fournier P, Luciani F, Niewolna M, André J, Delmas V, Larue L, Guise TA et al (2007) Stable overexpression of Smad7 in human melanoma cells impairs bone metastasis. *Cancer Res* 67: 2317–2324
- Jin X, Kim LJY, Wu Q, Wallace LC, Prager BC, Sanvoranart T, Gimple RC, Wang X, Mack SC, Miller TE et al (2017) Targeting glioma stem cells through combined BMI1 and EZH2 inhibition. *Nat Med* 23: 1352–1361
- Jinnin M, Ihn H, Tamaki K (2006) Characterization of SIS3, a novel specific inhibitor of Smad3, and its effect on transforming growth factor-beta1-induced extracellular matrix expression. *Mol Pharmacol* 69: 597–607
- Joung J, Konermann S, Gootenberg JS, Abudayyeh OO, Platt RJ, Brigham MD, Sanjana NE, Zhang F (2017) Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* 12: 828–863
- Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barceña C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H et al (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517: 583–588
- Konieczkowski DJ, Johannessen CM, Abudayyeh O, Kim JW, Cooper ZA, Piris A, Frederick DT, Barzily-Rokni M, Straussman R, Haq R et al (2014) A melanoma cell state distinction influences sensitivity to MAPK pathway inhibitors. *Cancer Discov* 4: 816–827
- Krepler C, Chunduru SK, Halloran MB, He X, Xiao M, Vultur A, Villanueva J, Mitsuchi Y, Neiman EM, Benetatos C et al (2013) The novel SMAC mimetic birinapant exhibits potent activity against human melanoma cells. *Clin Cancer Res* 19: 1784–1794
- Lamar JM, Stern P, Liu H, Schindler JW, Jiang Z-G, Hynes RO (2012) The Hippo pathway target, YAP, promotes metastasis through its TEAD-interaction domain. *Proc Natl Acad Sci* 109: E2441–E2450
- Leclerc D, Staats Pires AC, Guillemin GJ, Gilot D (2021) Detrimental activation of AhR pathway in cancer: an overview of therapeutic strategies. *Curr Opin Immunol* 70: 15–26
- Lee JH, Shklovskaya E, Lim SY, Carlino MS, Menzies AM, Stewart A, Pedersen B, Irvine M, Alavi S, Yang JYH et al (2020) Transcriptional downregulation of MHC class I and melanoma de-differentiation in resistance to PD-1 inhibition. *Nat Commun* 11: 1897
- Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, Wouters J, Radaelli E, Eyckerman S, Leonelli C, Vanderheyden K et al (2016) Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* 531: 518–522
- Levy C, Khaled M, Fisher DE (2006) MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med* 12: 406–414
- Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* 15: 554
- Liu Y, Liang X, Yin X, Lv J, Tang K, Ma J, Ji T, Zhang H, Dong W, Jin X et al (2017) Blockade of IDO-kynurenine-AhR metabolic circuitry abrogates IFN- γ -induced immunologic dormancy of tumor-repopulating cells. *Nat Commun* 8: 15207
- Lo R, Matthews J (2012) High-resolution genome-wide mapping of AHR and ARNT binding sites by ChIP-Seq. *Toxicol Sci* 130: 349–361
- Mak MP, Tong P, Diao L, Cardnell RJ, Gibbons DL, William WN, Skoulidis F, Parra ER, Rodriguez-Canales J, Wistuba II et al (2016) A patient-derived, pan-cancer EMT signature identifies global molecular alterations and immune target enrichment following epithelial-to-mesenchymal transition. *Clin Cancer Res* 22: 609–620
- Marin-Bejar O, Rogiers A, Dewaele M, Femel J, Karras P, Pozniak J, Bervoets G, Van Raemdonck N, Pedri D, Swings T et al (2020) A neural crest stem cell-like state drives nongenetic resistance to targeted therapy in melanoma. *bioRxiv* <https://doi.org/10.1101/2020.12.15.422929> [PREPRINT]
- Merdignac A, Angenard G, Allain C, Petitjean K, Bergeat D, Bellaud P, Fautrel A, Turlin B, Clément B, Dooley S et al (2018) A novel transforming growth factor beta-induced long noncoding RNA promotes an inflammatory microenvironment in human intrahepatic cholangiocarcinoma. *Hepatology* 67: 254–269
- Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, Dharia NV, Montgomery PG, Cowley GS, Pantel S et al (2017) Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* 49: 1779–1784
- Moriceau G, Hugo W, Hong A, Shi H, Kong X, Yu CC, Koya RC, Samatar AA, Khanlou N, Braun J et al (2015) Tunable-combinatorial mechanisms of acquired resistance limit the efficacy of BRAF/MEK cotargeting but result in melanoma drug addiction. *Cancer Cell* 27: 240–256
- Müller J, Krijgsman O, Tsoi J, Robert L, Hugo W, Song C, Kong X, Possik PA, Cornelissen-Steijger PDM, Foppen MHG et al (2014) Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. *Nat Commun* 5: 5712
- Nallet-Staub F, Marsaud V, Li L, Gilbert C, Dodier S, Bataille V, Sudol M, Herlyn M, Mauviel A (2014) Pro-invasive activity of the hippo pathway effectors YAP and TAZ in cutaneous melanoma. *J Invest Dermatol* 134: 123–132
- Nassar D, Blanpain C (2016) Cancer stem cells: basic concepts and therapeutic implications. *Annu Rev Pathol Mech Dis* 11: 47–76
- Ohanna M, Giuliano S, Bonet C, Imbert V, Hofman V, Zangari J, Bille K, Robert C, Bressac-de Paillerets B, Hofman P et al (2011) Senescent cells develop a PARP-1 and nuclear factor- κ B-associated secretome (PNAS). *Genes Dev* 25: 1245–1261
- Ooshima A, Park J, Kim SJ (2019) Phosphorylation status at Smad3 linker region modulates transforming growth factor- β -induced epithelial-mesenchymal transition and cancer progression. *Cancer Sci* 110: 481–488.
- Paris A, Tardif N, Galibert MD, Corre S (2021) AhR and cancer: from gene profiling to targeted therapy. *Int J Mol Sci* 22: 752
- Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL et al (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344: 1396–1401
- Piersma B, Bank RA, Boersema M (2015) Signaling in fibrosis: TGF- β , WNT, and YAP/TAZ converge. *Front Med* 2: 59
- Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, Beijersbergen RL, Bardelli A, Bernards R (2012) Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* 483: 100–103
- Puisieux A, Brabletz T, Caramel J (2014) Oncogenic roles of EMT-inducing transcription factors. *Nat Cell Biol* 16: 488–494

- Ramachandran A, Vizán P, Das D, Chakravarty P, Vogt J, Rogers KW, Müller P, Hinck AP, Sapkota GP, Hill CS (2018) TGF- β uses a novel mode of receptor activation to phosphorylate SMAD1/5 and induce epithelial-to-mesenchymal transition. *Elife* 7: 1–29
- Rambow F, Rogiers A, Marin-Bejar O, Aibar S, Femel J, Dewaele M, Karras P, Brown D, Chang YH, Debiec-Rychter M et al (2018) Toward minimal residual disease-directed therapy in melanoma. *Cell* 174: 843–855
- Ramsdale R, Jorissen RN, Li FZ, Al-Obaidi S, Ward T, Sheppard KE, Bukczynska PE, Young RJ, Boyle SE, Shackleton M et al (2015) The transcription cofactor c-JUN mediates phenotype switching and BRAF inhibitor resistance in melanoma. *Sci Signal* 8: ra82
- Rapino F, Delaunay S, Rambow F, Zhou Z, Tharun L, De Tullio P, Sin O, Shostak K, Schmitz S, Piepers J et al (2018) Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature* 558: 605–609
- Redfern AD, Spalding LJ, Thompson EW (2018) The Kraken Wakes: induced EMT as a driver of tumour aggression and poor outcome. *Clin Exp Metastasis* 35: 285–308
- Rizos H, Menzies AM, Pupo GM, Carlino MS, Fung C, Hyman J, Haydu LE, Mijatov B, Becker TM, Boyd SC et al (2014) BRAF inhibitor resistance mechanisms in metastatic melanoma: spectrum and clinical impact. *Clin Cancer Res* 20: 1965–1977
- Rizzolio S, Cagnoni G, Battistini C, Bonelli S, Isella C, Van Genderachter JA, Bernards R, Di Nicolantonio F, Giordano S, Tamagnone L (2018) Neupilin-1 upregulation elicits adaptive resistance to oncogene-targeted therapies. *J Clin Invest* 128: 3976–3990
- Rodón J, Carducci M, Sepulveda-Sánchez JM, Azaro A, Calvo E, Seoane J, Braña I, Sicart E, Gueorguieva I, Cleverly A et al (2015) Pharmacokinetic, pharmacodynamic and biomarker evaluation of transforming growth factor- β receptor I kinase inhibitor, galunisertib, in phase 1 study in patients with advanced cancer. *Invest New Drugs* 33: 357–370
- Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, Vaimberg EW, Goodale A, Root DE, Piccioni F et al (2018) Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* 9: 5416
- Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, Beqiri M, Sproesser K, Brafford PA, Xiao M et al (2017) Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* 546: 431–435
- Shalem O, Sanjana NE, Zhang F (2015) High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* 16: 299–311
- Smith MP, Brunton H, Rowling EJ, Ferguson J, Arozarena I, Miskolczi Z, Lee JL, Girotti MR, Marais R, Levesque MP et al (2016) Inhibiting drivers of non-mutational drug tolerance is a salvage strategy for targeted melanoma therapy. *Cancer Cell* 29: 270–284
- Song C, Piva M, Sun L, Hong A, Moriceau G, Kong X, Zhang H, Lomeli S, Qian J, Yu CC et al (2017) Recurrent tumor cell-intrinsic and -extrinsic alterations during MAPKi-induced melanoma regression and early adaptation. *Cancer Discov* 7: 1248–1265
- Sullivan RJ, Flaherty KT (2013) Resistance to BRAF-targeted therapy in melanoma. *Eur J Cancer* 49: 1297–1304
- Sun C, Wang L, Huang S, Heynen GJJE, Prahallad A, Robert C, Haanen J, Blank C, Wesseling J, Willems SM et al (2014) Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma. *Nature* 508: 118–122
- Talebi A, Dehairs J, Rambow F, Rogiers A, Nittner D, Derua R, Vanderhoydonc F, Duarte JAG, Bosisio F, Van den Eynde K et al (2018) Sustained SREBP-1-dependent lipogenesis as a key mediator of resistance to BRAF-targeted therapy. *Nat Commun* 9: 2500
- Tang PM-K, Zhou S, Meng X-M, Wang Q-M, Li C-J, Lian G-Y, Huang X-R, Tang Y-J, Guan X-Y, Yan BP-Y et al (2017) Smad3 promotes cancer progression by inhibiting E4BP4-mediated NK cell development. *Nat Commun* 8: 14677
- Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G et al (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352: 189–196
- Titz B, Lomova A, Le A, Hugo W, Kong X, ten Hoeve J, Friedman M, Shi H, Moriceau G, Song C et al (2016) JUN dependency in distinct early and late BRAF inhibition adaptation states of melanoma. *Cell Discov* 2: 16028
- Tsoi J, Robert L, Paraiso K, Galvan C, Sheu KM, Lay J, Wong DJL, Atefi M, Shirazi R, Wang X et al (2018) Multi-stage differentiation defines melanoma subtypes with differential vulnerability to drug-induced iron-dependent oxidative stress. *Cancer Cell* 33: 890–904.e5
- Tuncer E, Calçada RR, Zingg D, Varum S, Cheng P, Freiberger SN, Deng C-X, Kleiter I, Levesque MP, Dummer R et al (2019) SMAD signaling promotes melanoma metastasis independently of phenotype switching. *J Clin Invest* 129: 2702–2716
- Verfaillie A, Imrichova H, Atak ZK, Dewaele M, Rambow F, Hulselmans G, Christiaens V, Svetlichnyy D, Luciani F, Van den Mooter L et al (2015) Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nat Commun* 6: 6683
- Vetma V, Rožanc J, Charles EM, Hellwig CT, Alexopoulos LG, Rehm M (2017) Examining the in vitro efficacy of the IAP antagonist birinapant as a single agent or in combination with dacarbazine to induce melanoma cell death. *Oncol Res Featur Preclin Clin Cancer Ther* 25: 1489–1494
- Viswanathan VS, Ryan MJ, Dhruv HD, Gill S, Eichhoff OM, Seashore-Ludlow B, Kaffenberger SD, Eaton JK, Shimada K, Aguirre AJ et al (2017) Dependency of a therapy-resistant state of cancer cells on a lipid peroxidase pathway. *Nature* 547: 453–457
- Wagle N, Emery C, Berger MF, Davis MJ, Sawyer A, Pochanard P, Kehoe SM, Johannessen CM, Macconail LE, Hahn WC et al (2011) Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J Clin Oncol* 29: 3085–3096
- Welsh SJ, Rizos H, Scolyer RA, Long GV (2016) Resistance to combination BRAF and MEK inhibition in metastatic melanoma: where to next? *Eur J Cancer* 62: 76–85
- Wu N, Lian G, Sheng J, Wu D, Yu X, Lan H, Hu W, Yang Z (2020) Discovery of a novel selective water-soluble SMAD3 inhibitor as an antitumor agent. *Bioorg Med Chem Lett* 30: 127396
- Yang SY, Ahmed S, Satheesh SV, Matthews J (2018) Genome-wide mapping and analysis of aryl hydrocarbon receptor (AHR)- and aryl hydrocarbon receptor repressor (AHRR)-binding sites in human breast cancer cells. *Arch Toxicol* 92: 225–240
- Zhang Y, Feng XH, Derynck R (1998) Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. *Nature* 394: 909–913
- Zhao S, Guo Y, Sheng Q, Shyr Y (2014) Advanced heat map and clustering analysis using heatmap3. *Biomed Res Int* 2014: 986048
- van Zijl F, Mall S, Machat G, Pirker C, Zeillinger R, Weinhaeusel A, Bilban M, Berger W, Mikulits W (2011) A human model of epithelial to mesenchymal transition to monitor drug efficacy in hepatocellular carcinoma progression. *Mol Cancer Ther* 10: 850–860



License: This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Expanded View Figures

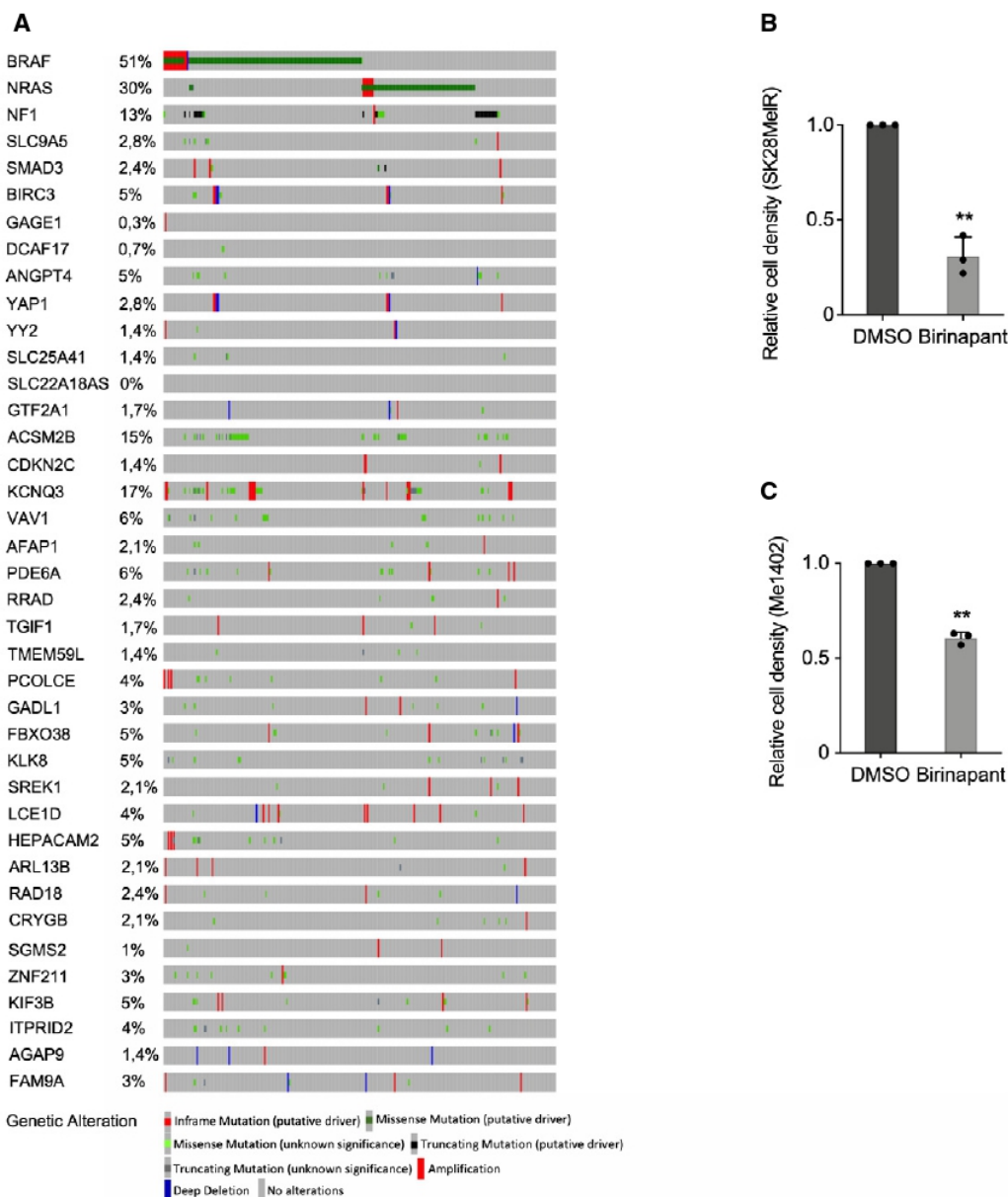


Figure EV1. Tumor-promoting genes are not frequently altered in cutaneous melanoma. Related to Fig 1.

A Genomic alterations and 36 tumor-promoting genes identified in Fig 1. *BRAF*, *NRAS*, and *NF1* have been added to the list of tumor-promoting genes according to melanoma classification (Akbani *et al*, 2015). Analyses were performed using the webtool available at www.cbioportal.org.

B Birinapant, an inhibitor of BIRC proteins (BIRC3 and BIRC2), decreased cell density (SKMel28R cells, Birinapant 100 nM, 96 h).

C Birinapant, an inhibitor of BIRC proteins (BIRC3 and BIRC2), decreased cell density (Me1402 cells, Birinapant 100 nM, 96 h).

Data information: Each histogram represents the mean \pm s.d. from three independent experiments. Bilateral Student test (with non-equivalent variances) $**P < 0.01$. Source data are available online for this figure.

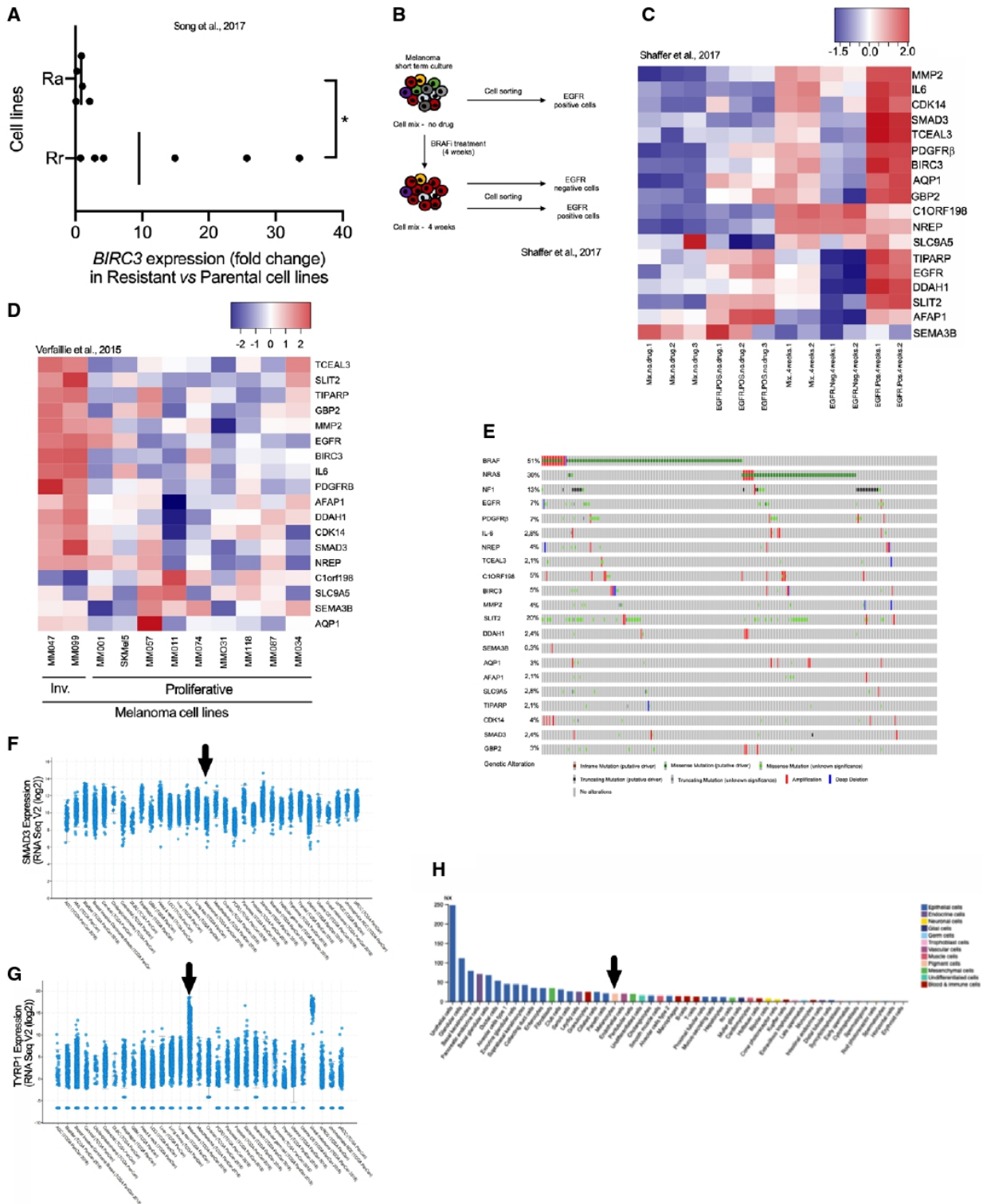


Figure EV2.

Figure EV2. BRAFi-resistance genes are associated with invasive phenotype and resistance to BRAFi *in vitro* and *in vivo*. Related to Fig 3.

- A *BIRC3* expression levels (fold change) in melanoma Rr and Ra cell lines. The fold change has been calculated by comparing *BIRC3* expression levels in resistant/parental cell lines (Song *et al*, 2017). Two groups have been made: Rr ($n = 6$) and Ra ($n = 5$) cell lines; Ra for resistance and MAPK reactivation and Rr for resistance and MAPK redundant. BRAFi resistance in Ra cell lines is due to additional genetic alterations such as BRAF splicing events for M395 SDR. BRAFi resistance in Rr cell lines is due to gene expression reprogramming. Two-tailed Mann–Whitney test, $P = 0.0303$. Vertical bars correspond to the medians.
- B Schematic representation of the workflow. As described in the original paper (Shaffer *et al*, 2017), fresh melanoma cells were obtained by patient tumor dissociation. Cells mix was treated with BRAFi. Cell sorting (EGFR positive and/or negative) was performed at two time points: before treatment (no drug) and after 4 weeks of treatment. Cells “mix” corresponds to the unsorted population. The EGFR-positive cells (exposed to BRAFi) are able to produce colonies.
- C Heat map depicting mRNA expression of BRAFi-resistance genes, using an RNA-Seq dataset obtained from BRAFi-treated melanoma cells (Shaffer *et al*, 2017). Scale corresponds to Z scores.
- D Heat map illustrating the expression levels of BRAFi-resistance genes in invasive (Inv.) and proliferative melanoma cell lines ($n = 2$ and $n = 9$, respectively) (Verfaillie *et al*, 2015). Scale corresponds to Z scores.
- E Genomic alterations in 18 BRAFi-resistance genes identified in Fig 3 (from www.cbioportal.org). *BRAF*, *NRAS*, and *NF1* have been added according to melanoma classification (Akbani *et al*, 2015).
- F *SMAD3* expression levels (RNA Seq V2 (\log_2)) in the TCGA dataset. Picture was downloaded from www.cbioportal.org. Arrow indicates cutaneous melanoma.
- G *TYRP1* expression levels (RNA Seq V2 (\log_2)) in the TCGA data set. Pictures were downloaded from www.cbioportal.org. Arrow indicates cutaneous melanoma. *TYRP1* is highly expressed in drug-naïve tumors; cutaneous and uveal melanoma in contrast to *SMAD3*.
- H *SMAD3* expression and cell types (<https://www.proteinatlas.org/ENSG00000166949-SMAD3/celltype>). The consensus normalized expression (NX) value for *SMAD3* and organ/tissue represents the maximum NX value in the three data sources (HPA, GTEx, and FANTOM5). Black arrow highlights the melanocytes.

Source data are available online for this figure.

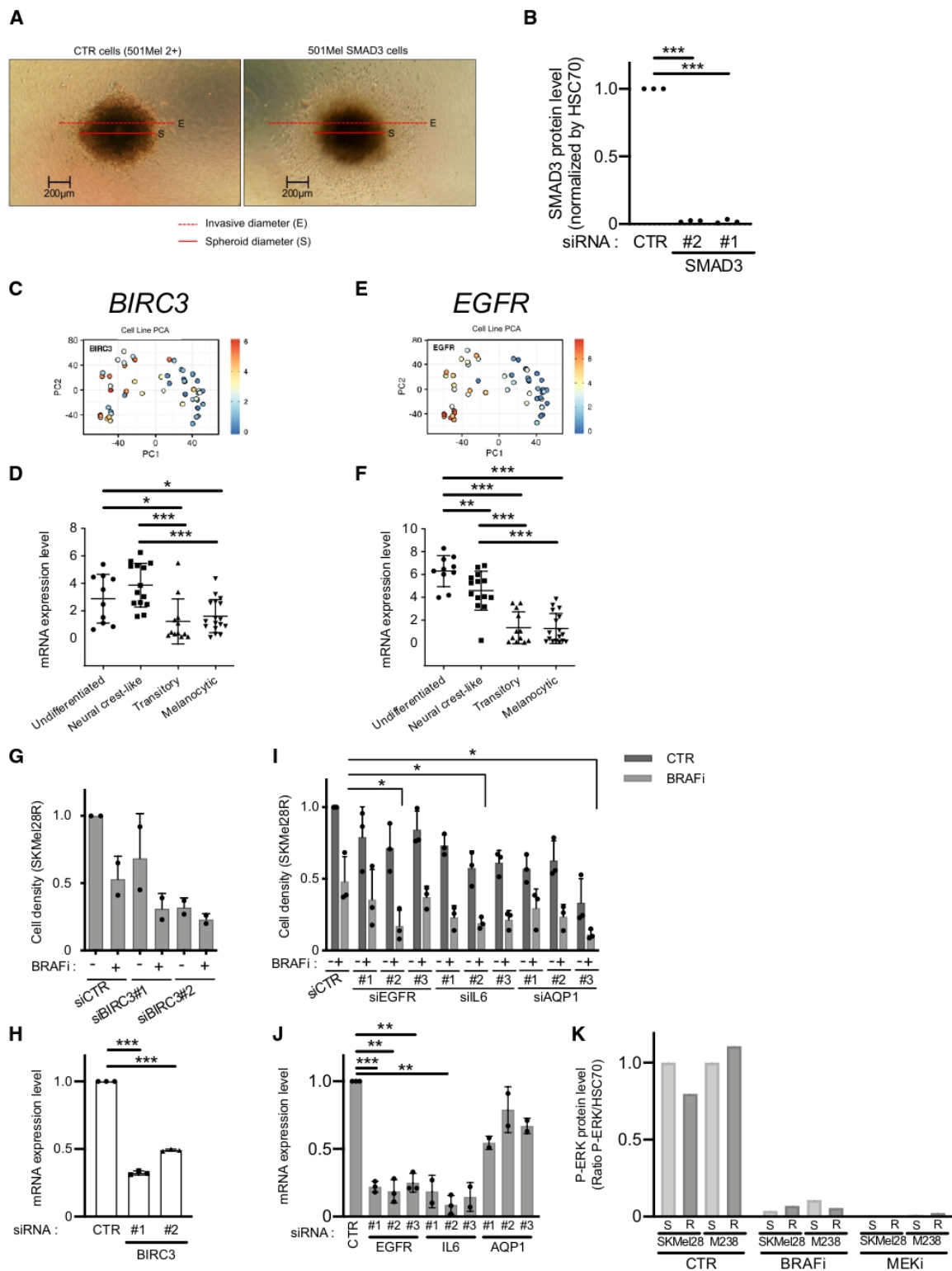


Figure EV3.

Figure EV3. *BIRC3* and *EGFR* are potent BRAFi-resistance genes. Related to Fig 5.

- A Pictures illustrating the ratio (invasive diameter/spheroid diameter; E/S) used in Fig 5E (Invasion assay, melanoma spheroids). Two cell lines are illustrated: CTR cells correspond to 501Mel cells expressing dCas9 and HSF1-p65-MS2 (named here 501Mel 2+) and the SMAD3 cells (overexpressing SMAD3).
- B SMAD3 Knock-down efficiency of two siRNAs targeting *SMAD3* in SKMel28R (Fig 5G). SMAD3 protein expression levels have been quantified in three independent experiments by Western blots. Expression levels have been compared to HSC70. Values obtained with the control siRNA (CTR) have been set at 1. Bilateral Student test (with non-equivalent variances) $***P < 0.001$.
- C PCA analysis of *BIRC3* expression in melanoma cell lines in function of their dedifferentiation states (generated by the webtool <http://systems.crump.ucla.edu/dediff/index.php>). Scale: red color corresponds to a high *BIRC3* expression level (unit of the scale: Log_2 FPKM).
- D *BIRC3* expression in these four subtypes of melanoma cells (undifferentiated = 10; neural crest-like = 14; transitory = 12; melanocytic = 17) (Tsoi et al, 2018). Error bars represents the mean \pm s.d. Multiple comparisons have been done using ordinary one-way ANOVA, $*P < 0.05$, $***P < 0.001$.
- E PCA analysis of *EGFR* expression in melanoma cell lines in function of their dedifferentiation states (unit of the scale: Log_2 FPKM).
- F *EGFR* expression in these four subtypes of melanoma cells (undifferentiated = 10; neural crest-like = 14; transitory = 12; melanocytic = 17). Error bars represents the mean \pm s.d. Multiple comparisons have been done using ordinary one-way ANOVA, $**P < 0.01$, $***P < 0.001$.
- G *BIRC3* depletion (siRNA#1 & #2) reduced cell density and restored BRAFi (vemurafenib) effect on BRAFi-resistant cells (SKMel28R). CTR for non-targeting siRNA. Vem for the BRAFi vemurafenib and DMSO for dimethylsulfoxide (solvent of Vem). $n = 2$ biologically independent experiments. Each histogram represents the mean \pm s.d.
- H *BIRC3* Knock-down efficiency of two siRNAs targeting *BIRC3* in SKMel28R (Fig EV3G). *BIRC3* mRNA expression levels have been quantified in three independent experiments by RT-qPCR. Values obtained with the control siRNA (CTR) have been set at 1. Each histogram represents the mean \pm s.d. $n = 3$ biologically independent experiments. Bilateral Student test (with non-equivalent variances) $***P < 0.001$.
- I Knock-down of *EGFR*, *IL6*, or *AQP1* (siRNA#1, #2, and #3) restored BRAFi (vemurafenib) effect on BRAFi-resistant cells (SKMel28R). $n = 3$ biologically independent experiments are presented. Each histogram represents the mean \pm s.d.. Bilateral Student test (with non-equivalent variances) $*P < 0.05$.
- J Knock-down efficiency of siRNAs targeting *EGFR*, *IL6*, or *AQP1* in SKMel28R (Fig EV3I). mRNA expression levels have been quantified in two or three independent experiments by RT-qPCR. Each histogram represents the mean \pm s.d. $n = 3$ biologically independent experiments for siCTR, siEGFR#1,2,4, IL6#3, and $n = 2$ for the other siRNAs. Values obtained with the control siRNA (CTR) have been set at 1. Bilateral Student test (with non-equivalent variances) $**P < 0.01$, $***P < 0.001$.
- K Efficiency of BRAFi (Vem) and MEKi (Cobi) to inhibit the MAPK signaling pathway (P-ERK) (as used in Fig 5O). Values obtained with the solvent (CTR, here DMSO) have been set at 100%.

Source data are available online for this figure.

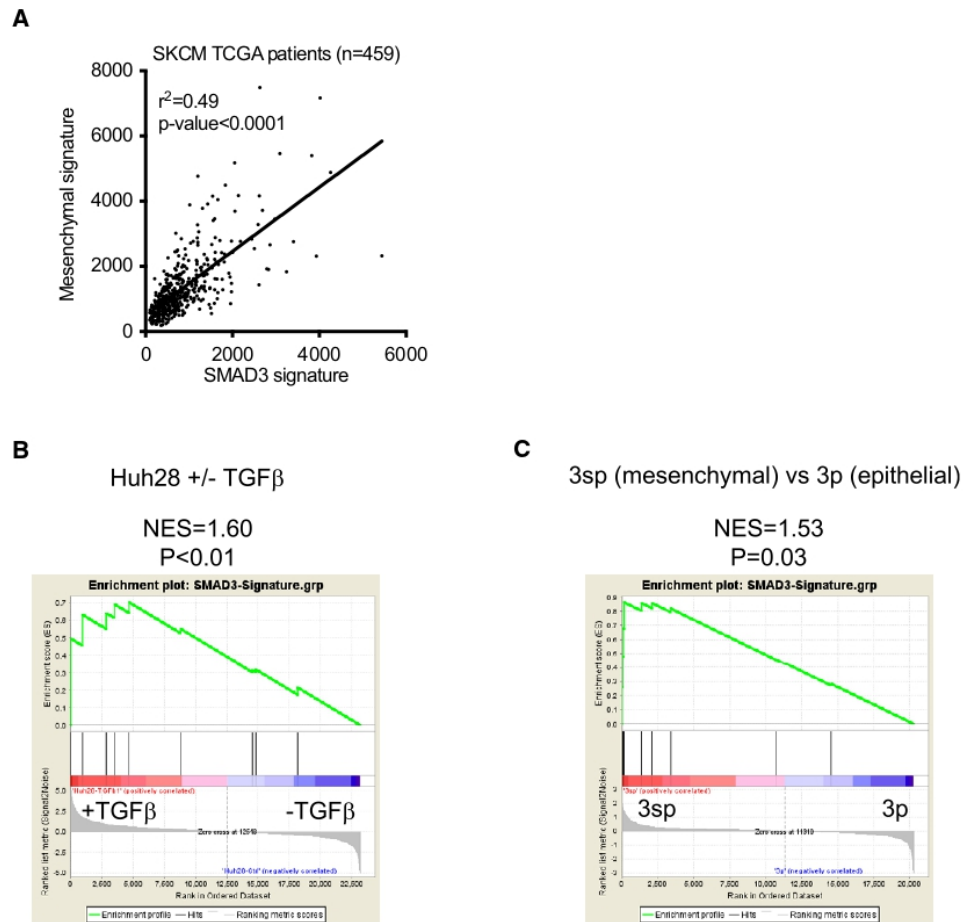


Figure EV4. SMAD3 signature is associated with mesenchymal phenotype in liver cancer cell lines. Related to Fig 7.

- A Plot showing correlation between SMAD3-signature and mesenchymal signature (Mak *et al*, 2016) in the TCGA dataset (SKCM). Pearson correlation test: $P < 0.0001$
- B Gene Set Enrichment Analysis (GSEA) of the SMAD3 signature in TGF- β treated *versus* non-treated Huh-28 cholangiocarcinoma cells (Merdrignac *et al*, 2018).
- C Gene Set Enrichment Analysis of the SMAD3 signature in hepatoma cell lines : mesenchymal (3sp cells) *versus* epithelial (3p cells) (van Zijl *et al*, 2011).

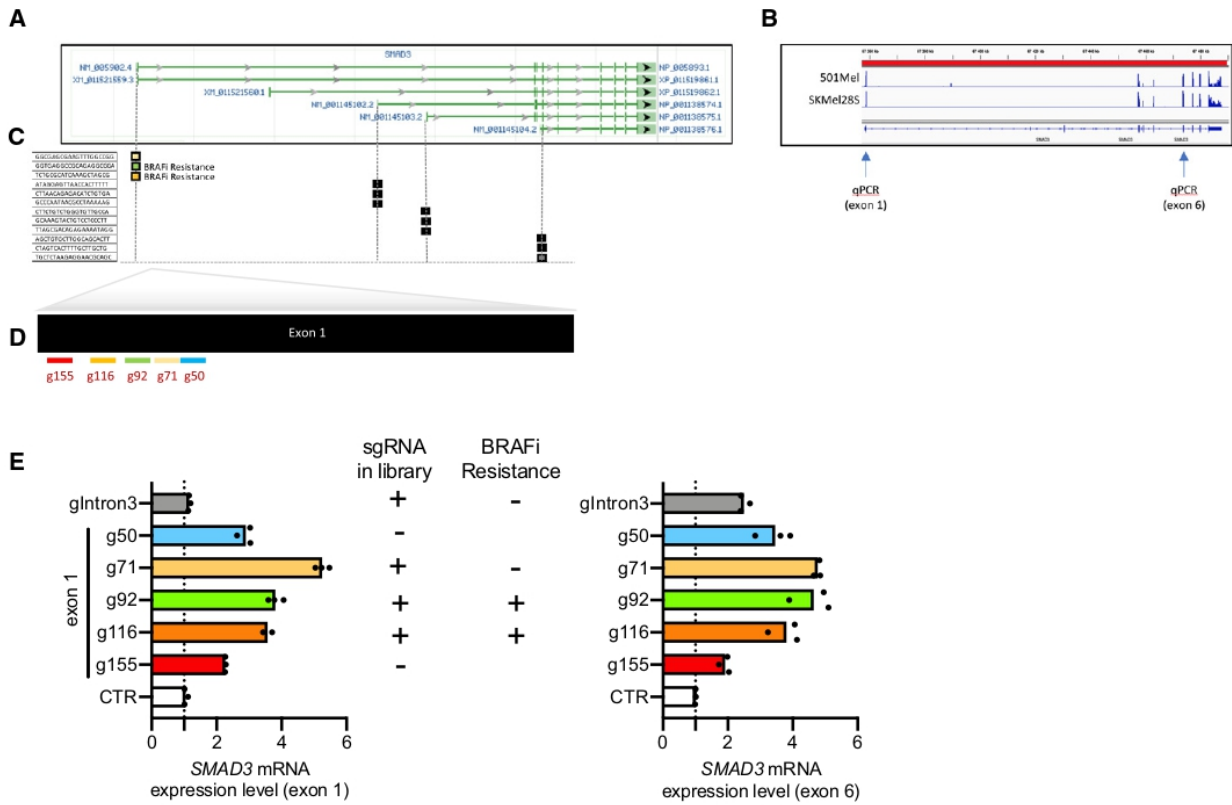


Figure EV5. sgRNA targeting four isoforms of SMAD3 and BRAFi resistance.

A SMAD3 mRNA isoforms according to NCBI. The longest isoform NM_005902.4 encodes SMAD3 protein (NP_005893.1). Three shorter isoforms of SMAD3 mRNA encode shorter SMAD3 proteins.

B Snapshot of SMAD3 mRNA expression in 501Mel and SKMel285 cells using IGV tool. Note that exon 1 is detectable in these melanoma cell lines. Arrows indicate the binding sites of primers for RT-qPCR (exon 1 or 6 for NM_005902.4).

C Scheme to illustrate the binding site of sgRNAs targeting SMAD3 gene. These 12 sgRNAs are found in the CRISPR-SAM library (Konermann et al, 2015). Three sgRNAs per isoform. Only two sgRNA targeting the longest isoform of SMAD3 are associated with BRAFi resistance (Fig 2).

D Ability of six sgRNAs to transactivate SMAD3 expression have been compared. Among the five sgRNAs targeting the first exon, three are found in the library (g71, g92, and g116). We designed g50 and g155 to evaluate the impact of sgRNA position on exon 1. sgRNA gIntron3 has been designed as a control (unable to promote the longest SMAD3 isoform but efficient to transactivate the smallest isoform). These two isoforms contain the exon 6 used for RT-qPCR.

E SMAD3 quantification by RT-qPCR (exon 1 or 6) in HEK293 cells transiently transfected with the three plasmids required for the CRISPR-SAM as previously described (Konermann et al, 2015). CTR for empty plasmid (no sgRNA cloned into lenti sgRNA(MS2)_zeo backbone). RT-qPCR data are representative of two independent experiments. Source data are available in Fig EV5 source data. Each histogram represents the mean of three technical replicates from one representative experiment. No statistical analyses were thus performed. Dotted line highlights the value of 1 (corresponding to the effect obtained by the control sgRNA).

Source data are available online for this figure.

CONCLUSION DE L'ARTICLE

La reprogrammation transcriptomique est un mécanisme utilisé par les cellules tumorales pour résister aux thérapies ciblées anti-cancéreuses. Cette reprogrammation passe généralement par une étape de dédifférenciation cellulaire. Dans cette étude, le criblage CRISPRa (étude sans *a priori*) a permis d'identifier 3 gènes (*SMAD3*, *BIRC3*, et *SLC9A5*) dont la surexpression est associée à une résistance aux inhibiteurs de BRAF dans le mélanome cutané. L'expression de ces 3 gènes a également été retrouvée augmentée dans les données de RNAseq disponibles dans la littérature, montrant que ce mécanisme est très largement retrouvé dans les lignées mais aussi chez les patients en rechute (robustesse du crible).

Les études menées *in vitro* démontrent notamment que l'inhibition de SMAD3 couplée aux inhibiteurs de BRAF réduit drastiquement le nombre de cellules persistantes, à l'origine des rechutes chez les patients.

SMAD3 est un facteur de transcription de la voie TGF- β . Dans ce contexte, l'inhibition de l'axe TGF- β chez les patients ayant développé une résistance aux thérapies ciblées semble, à première vue, pertinent. Dernièrement, plusieurs inhibiteurs au récepteur du TGF- β ont été évalués pour le traitement de divers cancer (carcinome hépatocellulaire, cancer colorectal, cancer du pancréas...). Malheureusement, les résultats des essais cliniques sont mitigés (bénéfices mineurs et/ou effets secondaires graves). Le TGF- β étant une cytokine pléiotropique, agissant à la fois sur les cellules cancéreuses et le microenvironnement tumoral, son inhibition peu en effet entrainer des effets secondaires sévères³⁸⁴. L'inhibition spécifique du facteur de transcription SMAD3 chez les patients ayant développés une résistance aux inhibiteurs de BRAF pourrait représenter une alternative intéressante³⁸⁵. Bien que SMAD2 et SMAD3 soient les deux principaux facteurs de transcription activés par le TGF- β , chacun d'entre eux régule un panel de gènes distincts (fonctions spécifiques)³⁸⁶. Ainsi, au lieu de cibler la voie TGF- β dans sa globalité, inhiber spécifiquement SMAD3 au pourrait constituer une stratégie thérapeutique plus précise.

PARTIE IV :

DISCUSSION

Au cours de ces travaux de thèse, j'ai testé et évalué les performances de quatre outils CRISPR/Cas9 : le CRISPR/KO, le CRISPRa, le BE et le PE. Je pense avoir identifié à la fois leur immense potentiel dans le domaine biomédical, notamment pour le diagnostic génétique et dans le cadre thérapeutique. Malgré la versatilité des technologies CRISPR, il existe encore plusieurs défis majeurs à relever pour faciliter leur utilisation en médecine moléculaire.

Limite liée à l'activité hors-cible

Parmi les outils d'édition du génome décrits dans cette thèse, le BE est certainement l'outil le plus fiable et le plus efficace actuellement pour une application thérapeutique. Contrairement à l'outil CRISPR/Cas9 classique, l'absence de CDB permet de minimiser les risques génotoxiques (INDELS, réarrangements chromosomiques, translocation, activation de p53). Néanmoins, la question des *bystanders* et surtout de l'activité hors cible, qui reste difficile à maîtriser/éradiquer, est importante à prendre en considération. De nombreuses technologies permettent dorénavant de détecter, à l'échelle du génome, cette activité hors-cible. Cependant, quel est l'impact d'une telle activité hors-cible *in vivo* ? Parmi les cellules génétiquement modifiées, y a-t-il des clones minoritaires, susceptibles d'évoluer différemment ? Pour répondre à cette question, il serait intéressant de suivre l'évolution clonale des cellules génétiquement modifiées dans des modèles *in vivo*. Grâce aux technologies d'étiquetage moléculaire, c'est désormais possible³⁸⁷. Le principe consiste à « marquer » individuellement chaque cellule avec une « étiquette » moléculaire unique, par exemple une petite séquence variable d'une dizaine de nucléotides qui est apportée aux cellules par infection lentivirale.

Limite de l'édition du génome pour le traitement des maladies rares

Les essais de thérapie génique actuels se concentrent sur les patients atteints de maladies génétiques récurrentes (rentabilité). Or, il existe de multiples pathologies génétiques rares (>1 personne sur 2000) voir ultra-rares (>1 personne sur 50 000), comme la gangliosidose à GM1, et renfermant parfois des variants génétiques uniques. Comment traiter ces maladies ?

Le point limitant pour un développement généralisé de la thérapie par édition du génome réside dans la conception des ARN guides spécifiques, d'une part pour

chaque maladie génétique mais surtout pour chaque altération génétique. En effet, 275 variants de classe 4 ou 5 localisés dans le gène *GLB1* ont été recensés dans la base de données ClinVar. Contrairement aux maladies génétiques associés à des variants génétiques récurrents (ex : CFTR Δ F508 présent chez près de 70% des patients atteints de mucoviscidose, HBB p.Glu6Val présent chez tous les patients drépanocytaires), cette diversité implique un travail colossal pour optimiser et caractériser une enzyme associée à ARN guide particulier pour chaque mutation.

Dans le cadre de la réglementation actuelle, cela constitue un obstacle à la mise en place des thérapies ultra-personnalisées. Du point de vue des entreprises biotechnologiques à but lucratif, c'est actuellement beaucoup trop long (2/3 ans d'études précliniques) et ce n'est pas rentable (plus de 5 millions de dollars pour un candidat médicament)³⁸⁸. Nous pouvons mettre en perspective ce coût élevé avec celui du séquençage du génome humain, qui s'élevait à plus de 10 000 dollars il y a 10 ans contre seulement 200 dollars aujourd'hui. Nous pouvons émettre que l'hypothèse que les progrès technologiques en matière d'édition du génome devraient suivre une tendance similaire.

De plus, les entreprises biotechnologiques peuvent fabriquer un ARN guide en moins de trois jours et le tester en lignée cellulaire en moins de 2/3 semaines pour notamment évaluer son efficacité et son activité hors de la cible³⁸⁸. Aussi, ces dernières années, de nombreux candidats de thérapie génique ciblant les hémoglobinopathies ont été testés en utilisant le même processus de production, le même type de vecteur et la même plateforme d'administration (cellules CD34+). Ce qui a changé, c'était uniquement le type de transgène thérapeutique délivré aux patients³⁸⁸. Par analogie avec la technologie CRISPR, cela revient juste à changer la séquence de l'ARN guide. Alors, ne serait-il pas possible d'utiliser des véhicules et éditeurs communs pour alléger les réglementations pour faciliter le transfert de la technologie CRISPR chez les patients qui en ont le plus besoin (maladies incurables), dont l'espérance de vie est déjà prédéterminée à la naissance ? D'après le Pr. Fiodor D. Urnov, ce changement de paradigme dans la réglementation préclinique actuelle est nécessaire³⁸⁸ car médecine est aujourd'hui rentrée dans une nouvelle ère thérapeutique : celle de la chirurgie du génome.

Thérapie génique par éditeurs de larges séquences

A l'avenir, une solution intéressante pour remédier à la limite des thérapies ultra-personnalisées serait d'utiliser les éditeurs de larges séquences d'ADN programmables et capables d'effectuer des insertions, inversions, translocations ou délétions à l'échelle du kilobase, sans CDB. Par exemple, l'insertion ou remplacement direct du gène/de l'exon défectueux serait bénéfique à beaucoup plus de patients diagnostiqués avec des variants génétiques différents (traitements « pan-mutation »³⁰², **figure 22**). La séquence génétique thérapeutique pourrait être intégrée dans un locus sans danger choisi par l'expérimentateur. Cette stratégie permettrait également de traiter les pathologies causées par des expansions de nucléotides (maladie de Huntington), duplications (syndrome de duplication MECP2), et translocations (certains cas du syndrome de Down), qui ne peuvent pas être prises en charge par thérapie génique additive classique.

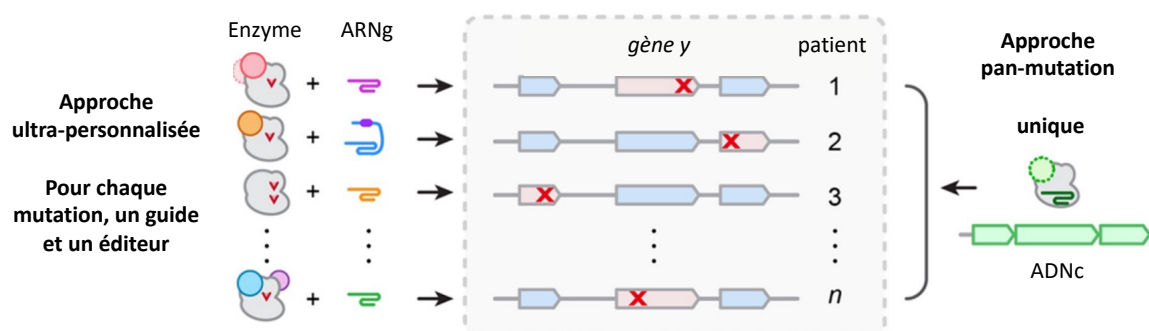


Figure 22 : Thérapie ultra-personnalisée versus thérapie pan-mutation par édition génétique.

Schéma traduit et adapté d'après Conor J. Tou et al.³⁰².

Thérapie génique préventive

Plus largement, la thérapie génique pourrait un jour être utilisée en prévention. Typiquement, si la thérapie est suffisamment sûre et efficace, les patients porteurs de prédispositions génétiques au cancer (cancer du sein héréditaire, syndrome de lynch, polypose adénomateuse familiale), aux maladies neurodégénératives ou maladies cardiovasculaires pourraient en bénéficier. En utilisant la technologie ABE couplée aux vecteurs AAV9, Daniel Reichart et al. ont réussi à corriger la mutation R403Q du gène *Myh6* dans environ 70% des cardiomyocytes ventriculaires murins, prévenant l'apparition de cardiomyopathies chez les rongeurs traités³⁸⁹. Même si d'un point de vue éthique, il est difficilement concevable de traiter des patients « en bonne santé », à long terme, on pourrait imaginer que ce genre de traitement permettrait de diminuer ou retarder le risque d'apparition de la pathologie. Daniel Reichart et al. suggèrent que l'intervention génétique pourrait être destinée uniquement aux patients exprimant des biomarqueurs précoces au développement d'une maladie, c'est-à-dire des patients à risque imminent de développer la pathologie³⁸⁹, mais encore faut-il organiser ce diagnostic génétique.

Dans le cadre du programme de dépistage néonatal Belge, un nouveau projet de recherche intitulé « Baby detect » permet aujourd'hui de dépister les nouveaux nés pour 120 maladies génétiques sévères (<https://babydetect.com/>) (contre 7 maladies en France). Ce test repose sur la technologie de séquençage haut-débit ciblé sur un panel de gènes. L'objectif est de d'accélérer la mise en place du diagnostic précoce pour être le plus efficace dans la prise en charge thérapeutique du patient. Cela signifie qu'un traitement peut débuter avant même la déclaration des premiers symptômes. Même si ce test dépiste uniquement des maladies à développement précoce, il s'agit selon moi d'un premier pas vers la médecine de prévention. La brochure du test indique que le dépistage concerne uniquement une liste de maladies traitables, mais comment faire si des variants de signification inconnu sont identifiés chez ces jeunes patients ? Outre les problèmes éthiques mis-en jeu (informer ou non les parents), quel poids accorder à ces VSI et comment tester leur impact biologique ?

Analyse des VSI dans les laboratoires de diagnostics génétiques

L'arrivée des nouveaux séquenceurs toujours plus performant a permis à l'entreprise Illumina d'annoncer, en 2022, qu'ils factureraient moins de 200 dollars le séquençage d'un génome complet humain. Dans les années à venir, ce prix baissera

probablement. Ce tarif permettra aux cliniciens de prescrire plus d'analyses pour faciliter le diagnostic des maladies génétiques mais également pour faciliter l'exploration des déterminants génétiques de pathologies jusqu'à alors inexplicables (recherche). Par conséquent, le nombre de VSI va fortement s'accroître. Dans ce contexte, il faudra proposer des stratégies pour les tester en prospectif (tests fonctionnels), et éventuellement en routine. Le modèle cellulaire est probablement le facile et le moins cher à mettre en place. Les outils d'édition du génome ayant largement évolués ces dernières années, il devient relativement « facile » d'introduire dans les cellules en culture un variant génétique d'intérêt. De plus, les modèles cellulaires générés pourront être utilisés pour évaluer la réponse des patients aux traitements disponibles. Lors de la mise en place d'un test fonctionnel, il est également essentiel d'inclure des variants de signification connue (perte de fonction, bénin, gain de fonction...) afin de valider son interprétation et sa robustesse. Cela n'a pas été fait dans les articles présentés en résultat 2 et 3, limitant notre capacité à conclure.

L'étude des variants génétiques dans des modèles *in vivo* peut parfois s'avérer nécessaire pour l'étude de pathologies complexes et difficiles à mimer *in vitro* (ex : développement du cancer). En Mai 2023, un modèle de souris génétiquement modifiée exprimant un système PE inducible par la Cre recombinase a été mis au point et a permis d'étudier l'effet d'une mutation somatique Kras associée au cancer du pancréas³⁹⁰.

Méthodes d'enrichissement pour isoler rapidement les clones édités

La technologie PE, bien que peu efficace pour le moment, est la plus exhaustive des technologies CRISPR puisqu'elle permet d'installer à la fois des substitutions et INDELS pouvant atteindre plusieurs dizaines de pb. Son utilisation permettrait de tester un très grand nombre de VSI. Afin d'isoler plus facilement et rapidement les clones correctement édités, nous avons testé au laboratoire une méthode de co-sélection sans marqueur, qui repose sur la modification d'un gène endogène conférant aux cellules éditées une résistance à une drogue toxique³⁹¹. Brièvement, les cellules en culture sont co-transfectées avec le PE et deux pegRNA. Le premier pegRNA est celui d'intérêt et le deuxième pegRNA cible le gène ATP1A1 pour la mise en place d'une mutation conférant aux cellules une résistance à une drogue appelée ouabaine. Comme les événements de co-édition ne sont pas statistiquement indépendants, la sélection des cellules résistantes à la drogue permet par la même occasion d'enrichir

les cellules éditées pour le gène d'intérêt³⁹¹. Dans nos conditions, au laboratoire, cela a permis d'atteindre des taux d'édition de plus de 30% après sélection contre des niveaux indétectables sans sélection (données non montrées). On pourrait même imaginer, à l'avenir, que cette méthode de co-sélection sans marqueur puisse être utilisée pour sélectionner des cellules modifiées en vue d'une thérapie génique *ex vivo*. D'autant que le gène ATP1A1 reste fonctionnel lorsqu'il est muté et résistant à son inhibiteur ouabaïne.

Étude des VSI à grande échelle

La validation d'un seul VSI à la fois demande beaucoup de temps humain. Typiquement, le clonage des ARN guides CRISPR/Cas9 dans un vecteur requiert une semaine. La transfection des cellules avec les guides puis de l'étape de dilution limite peut prendre presque un mois. A cela, il faut ajouter l'étape de criblage des clones afin de sélectionner ceux correctement édités (quelques jours). La mise en place des tests fonctionnels ne peut démarrer qu'après toute ses étapes. Dernièrement, plusieurs groupes indépendants ont montré qu'il était possible d'étudier l'impact de nombreux variants en une seule expérience^{392,393}. Comme preuve de faisabilité, ils ont utilisé la technologie de criblage par BE ou PE. Typiquement, de multiples ARN guides (bibliothèque) sont conçus de tels manières à ce qu'ils recouvrent la totalité de la partie codante du gène étudié (*tiling*). Chaque cellule est ensuite infectée par un lentivirus véhiculant un ARN guide de la bibliothèque, permettant la mise en place d'un variant génétique particulier. A noter que pour faciliter l'étude des variants homozygotes, il est intéressant d'utiliser des cellules haploïdes (ex : cellules HAP1)³⁹⁴ ou bien de préalablement retirer par CRISPR/KO les allèles surnuméraires (ex : les cellules HEK293T sont hypotriploïdes)³⁹³. La bibliothèque de cellule est ensuite soumise au test fonctionnel. Par exemple, Matthew A. Coelho et al. ont évalué les variations localisées dans le gène JAK1 sur leur capacité à induire l'expression des marqueurs de surface MHC1 et PD-L1 en présence d'interféron- γ . Selon leur niveau de fluorescence, les cellules ont été triées par FACS³⁹². Le séquençage des ARN guides intégrés au génome des cellules triées permet d'identifier indirectement les variants génétiques associées à un gain ou une perte de fonction. Avec cette technologie, de nouveaux variants encore non décrits dans la littérature ont même pu être identifiés. A terme, l'analyse à haut débit de variants à travers un gène, une voie de signalisation et même le génome complet permettrait de grandement accélérer la classification des VSI.

Conclusion scientifique et perspectives

L'ingénierie moléculaire est entrée dans une révolution depuis la découverte du système de défense CRISPR/Cas chez les procaryotes. Dorénavant, les scientifiques sont en mesure d'interroger le génome pour mieux comprendre le rôle des gènes dans les modèles biologiques. En recherche appliqué, les outils d'édition du génome permettent également de mieux comprendre les mécanismes de résistance face au cancer (**résultat 4**) et faciliter le diagnostic et le traitement des maladies génétiques (**résultats 1, 2 et 3**). Dorénavant, la médecine est rentrée dans la nouvelle ère de la chirurgie du génome.

En effet, l'avènement de la médecine génomique a permis, ces dernières années, de faciliter le diagnostic de centaines de maladies génétiques. Toutefois, certaines maladies génétiques restent inexplicées, notamment en raison de l'identification de variants de signification inconnue au séquençage, dont le nombre va s'accroître dans les prochaines années. L'exploration fonctionnelle du rôle de ces VSI est indispensable pour la mise en place d'un diagnostic de certitude (variant causal ou non) et la prise en charge du patient. Comme démontré dans les **résultats 2 et 3** de cette thèse, les outils d'édition du génome permettent d'installer ces VSI dans des modèles cellulaires pour en déterminer leurs conséquences.

Le diagnostic de milliers de patients atteints de maladies génétiques jusqu'à alors inexplicées a également créé un déséquilibre par rapport aux thérapies disponibles : un nombre important de patients diagnostiqués se retrouvent en errance thérapeutique. Au cours de ma thèse, j'ai eu l'opportunité de tester les outils d'édition du génome pour corriger un variant génétique délétère. Comme preuve de concept, j'ai utilisé la technologie BE *in vitro*, dans les fibroblastes d'un patient atteint de gangliosidose à GM1 de type II (**résultat 1**).

Avant d'atteindre le domaine clinique, les outils d'édition du génome nécessitent d'être davantage caractérisés et optimisés. Notamment, comme il n'existe pas encore de vecteur sans danger permettant de délivrer avec efficacité les outils d'édition du génome *in vivo* (en dehors du foie), je vais poursuivre mon post-doctorat à San Diego, dans le laboratoire du Dr Shannon M. Miller, ex-doctorante du Pr David R. Liu (<https://www.millerlabresearch.com/>). L'objectif est d'utiliser la technologie d'évolution dirigée des protéines pour créer des vecteurs viraux/non viraux avec des

performances améliorées et capables de transporter des cargos ARN ou protéiques dans ces cellules ou tissus d'intérêts (projets confidentiels).

Conclusion personnelle

Ces trois années m'ont permis d'explorer en détail les principales technologies d'édition du génome et d'acquérir une expertise précieuse. Notre activité suscite un intérêt puisque différents chercheurs et médecins nous ont contacté, avec lesquels j'ai la chance d'interagir pour partager mon expertise. Plusieurs projets collaboratifs sont actuellement en cours.

Mon projet de thèse principal (preuve de concept pour le traitement de la gangliosidose par BE) m'a également permis de remporter le prix Jeune Talent France *For Women in Science* de la fondation l'Oréal Unesco en 2022 et le premier prix du jury de la finale Bretonne « Ma thèse en 180 secondes » en 2023 (**annexe VI**). L'ensemble des formations/ missions enseignements/ congrès auxquels j'ai participé sont indiqués en **annexe III à IV**.

Lauréates 2022 de la fondation l'Oréal UNESCO :

<https://www.fondationloreal.com/fr/nos-programmes-pour-les-femmes-et-la-science-prix-jeunes-talents-nationaux-et-regionaux/jeune>

Vidéo Ma thèse en 180 secondes :

<https://www.youtube.com/watch?v=nCrIpGTyHDk&t=47s>

BIBLIOGRAPHIE

1. Jaenisch, R. & Mintz, B. Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 1250–1254 (1974).
2. Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. & Ruddle, F. H. Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 7380–7384 (1980).
3. Gordon, J. W. & Ruddle, F. H. Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* **214**, 1244–1246 (1981).
4. Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317**, 230–234 (1985).
5. Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503–512 (1987).
6. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994).
7. Jacquier, A. & Dujon, B. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41**, 383–394 (1985).
8. Gimble, F. S. Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol. Lett.* **185**, 99–107 (2000).
9. Epinat, J.-C. *et al.* A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.* **31**, 2952–2962 (2003).
10. Chevalier, B. S. *et al.* Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell* **10**, 895–905 (2002).
11. Silva, G. H., Belfort, M., Wende, W. & Pingoud, A. From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J. Mol. Biol.* **361**, 744–754 (2006).
12. Rosen, L. E. *et al.* Homing endonuclease I-Crel derivatives with novel DNA target specificities. *Nucleic Acids Res.* **34**, 4791–4800 (2006).
13. Arnould, S. *et al.* Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.* **355**, 443–458 (2006).
14. Seligman, L. M. *et al.* Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.* **30**, 3870–3879 (2002).
15. Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.* **128**, 2477–2484 (2006).
16. Miller, J., McLachlan, A. D. & Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**, 1609–1614 (1985).
17. Beerli, R. R. & Barbas, C. F. Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.* **20**, 135–141 (2002).
18. Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809–817 (1991).
19. Segal, D. J., Dreier, B., Beerli, R. R. & Barbas, C. F. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 2758–2763 (1999).
20. Dreier, B., Segal, D. J. & Barbas, C. F. Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J. Mol. Biol.* **303**, 489–502 (2000).
21. Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D. & Barbas, C. F. Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.* **276**, 29466–29478 (2001).
22. Dreier, B. *et al.* Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.* **280**, 35588–35597 (2005).
23. Liu, Q., Xia, Z., Zhong, X. & Case, C. C. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J. Biol. Chem.* **277**, 3850–3856 (2002).
24. Mandell, J. G. & Barbas, C. F. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res.* **34**, W516–523 (2006).
25. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1156–1160 (1996).
26. Kim, H. J., Lee, H. J., Kim, H., Cho, S. W. & Kim, J.-S. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.* **19**, 1279–1288 (2009).
27. Zhu, C. *et al.* Evaluation and application of modularly assembled zinc-finger nucleases in

- zebrafish. *Dev. Camb. Engl.* **138**, 4555–4564 (2011).
28. Xiong, K. *et al.* Targeted editing of goat genome with modular-assembly zinc finger nucleases based on activity prediction by computational molecular modeling. *Mol. Biol. Rep.* **40**, 4251–4256 (2013).
 29. Ramirez, C. L. *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods* **5**, 374–375 (2008).
 30. Pattanayak, V., Ramirez, C. L., Joung, J. K. & Liu, D. R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods* **8**, 765–770 (2011).
 31. Gabriel, R. *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* **29**, 816–823 (2011).
 32. Greisman, H. A. & Pabo, C. O. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657–661 (1997).
 33. Isalan, M., Klug, A. & Choo, Y. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat. Biotechnol.* **19**, 656–660 (2001).
 34. Hurt, J. A., Thibodeau, S. A., Hirsh, A. S., Pabo, C. O. & Joung, J. K. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12271–12276 (2003).
 35. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
 36. Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
 37. Streubel, J., Blücher, C., Landgraf, A. & Boch, J. TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* **30**, 593–595 (2012).
 38. Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* **39**, 9283–9293 (2011).
 39. Mussolino, C. *et al.* TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic Acids Res.* **42**, 6762–6773 (2014).
 40. Gupta, R. M. & Musunuru, K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J. Clin. Invest.* **124**, 4154–4161 (2014).
 41. Sung, Y. H., Baek, I.-J., Seong, J. K., Kim, J. S. & Lee, H.-W. Mouse genetics: catalogue and scissors. *BMB Rep.* **45**, 686–692 (2012).
 42. Luria, S. E. & Human, M. L. A nonhereditary, host-induced variation of bacterial viruses. *J. Bacteriol.* **64**, 557–569 (1952).
 43. Luria, S. E. Host-induced modifications of viruses. *Cold Spring Harb. Symp. Quant. Biol.* **18**, 237–244 (1953).
 44. Wilson, G. G. & Murray, N. E. Restriction and modification systems. *Annu. Rev. Genet.* **25**, 585–627 (1991).
 45. Pingoud, A., Fuxreiter, M., Pingoud, V. & Wende, W. Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci. CMLS* **62**, 685–707 (2005).
 46. Dupuis, M.-É., Villion, M., Magadán, A. H. & Moineau, S. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat. Commun.* **4**, 2087 (2013).
 47. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).
 48. Horvath, P. *et al.* Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1401–1412 (2008).
 49. Jansen, R., Embden, J. D. A. van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
 50. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 7 (2006).
 51. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
 52. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
 53. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
 54. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
 55. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
 56. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev.*

Microbiol. **13**, 722–736 (2015).

57. Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147 (2016).
58. Wang, F. *et al.* Advances in CRISPR-Cas systems for RNA targeting, tracking and editing. *Biotechnol. Adv.* **37**, 708–729 (2019).
59. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29–44 (2016).
60. Ghorbani, A. *et al.* A short overview of CRISPR-Cas technology and its application in viral disease control. *Transgenic Res.* **30**, 221–238 (2021).
61. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
62. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
63. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2579–2586 (2012).
64. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
65. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
66. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
67. DiCarlo, J. E. *et al.* Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336–4343 (2013).
68. Li, D. *et al.* Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 681–683 (2013).
69. Lo, T.-W. *et al.* Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. *Genetics* **195**, 331–348 (2013).
70. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229 (2013).
71. Xie, K. & Yang, Y. RNA-guided genome editing in plants using a CRISPR-Cas system. *Mol. Plant* **6**, 1975–1983 (2013).
72. Feng, Z. *et al.* Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res.* **23**, 1229–1232 (2013).
73. Gillmore, J. D. *et al.* CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *N. Engl. J. Med.* **385**, 493–502 (2021).
74. Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).
75. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).
76. Scully, R., Panday, A., Elango, R. & Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* **20**, 698–714 (2019).
77. Xue, C. & Greene, E. C. DNA Repair Pathway Choices in CRISPR-Cas9-Mediated Genome Editing. *Trends Genet. TIG* **37**, 639–656 (2021).
78. Lykke-Andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* **16**, 665–677 (2015).
79. Allen, F. *et al.* Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4317.
80. Shen, M. W. *et al.* Author Correction: Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **567**, E1–E2 (2019).
81. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
82. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* **8**, 753–755 (2011).
83. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).
84. Ikeda, K. *et al.* Efficient scarless genome editing in human pluripotent stem cells. *Nat. Methods* **15**, 1045–1047 (2018).
85. Hastay, P., Rivera-Pérez, J. & Bradley, A. The length of homology required for gene targeting in embryonic stem cells. *Mol. Cell. Biol.* **11**, 5586–5591 (1991).
86. Zhang, J.-P. *et al.* Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol.* **18**, 35 (2017).
87. Miyaoka, Y. *et al.* Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci. Rep.* **6**, 23549 (2016).
88. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining

- and homologous recombination in human cells. *DNA Repair* **7**, 1765–1771 (2008).
89. Li, G. *et al.* Suppressing Ku70/Ku80 expression elevates homology-directed repair efficiency in primary fibroblasts. *Int. J. Biochem. Cell Biol.* **99**, 154–160 (2018).
 90. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).
 91. Beumer, K. J. *et al.* Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19821–19826 (2008).
 92. Cen, Y., Fiori, A. & Van Dijk, P. Deletion of the DNA Ligase IV Gene in *Candida glabrata* Significantly Increases Gene-Targeting Efficiency. *Eukaryot. Cell* **14**, 783–791 (2015).
 93. Nishizawa-Yokoi, A. *et al.* Suppression of Ku70/80 or Lig4 leads to decreased stable transformation and enhanced homologous recombination in rice. *New Phytol.* **196**, 1048–1059 (2012).
 94. Yu, C. *et al.* Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell* **16**, 142–147 (2015).
 95. Pinder, J., Salsman, J. & Dellaire, G. Nuclear domain ‘knock-in’ screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res.* **43**, 9379–9392 (2015).
 96. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).
 97. Robert, F., Barbeau, M., Éthier, S., Dostie, J. & Pelletier, J. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med.* **7**, 93 (2015).
 98. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014).
 99. Orthwein, A. *et al.* A mechanism for the suppression of homologous recombination in G1 cells. *Nature* **528**, 422–426 (2015).
 100. Renaud, J.-B. *et al.* Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Rep.* **14**, 2263–2272 (2016).
 101. Savic, N. *et al.* Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *eLife* **7**, e33761 (2018).
 102. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
 103. Chen, E. *et al.* Decorating chromatin for enhanced genome editing using CRISPR-Cas9. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2204259119 (2022).
 104. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
 105. Lin, Y. *et al.* CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* **42**, 7473–7485 (2014).
 106. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
 107. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).
 108. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
 109. Stadtmauer, E. A. *et al.* CRISPR-engineered T cells in patients with refractory cancer. *Science* **367**, eaba7365 (2020).
 110. Lin, J. & Wong, K.-C. Off-target predictions in CRISPR-Cas9 gene editing using deep learning. *Bioinforma. Oxf. Engl.* **34**, i656–i663 (2018).
 111. Listgarten, J. *et al.* Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat. Biomed. Eng.* **2**, 38–47 (2018).
 112. Cancellieri, S., Canver, M. C., Bombieri, N., Giugno, R. & Pinello, L. CRISPRitz: rapid, high-throughput and variant-aware in silico off-target site identification for CRISPR genome editing. *Bioinforma. Oxf. Engl.* **36**, 2001–2008 (2020).
 113. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinforma. Oxf. Engl.* **30**, 1473–1475 (2014).
 114. Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17**, 148 (2016).
 115. Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **44**, W272–276 (2016).
 116. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
 117. Tsai, S. Q. & Joung, J. K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat. Rev. Genet.* **17**, 300–312 (2016).

118. Tsai, S. Q. Discovering the Genome-Wide Activity of CRISPR-Cas Nucleases. *ACS Chem. Biol.* **13**, 305–308 (2018).
119. Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–243, 1 p following 243 (2015).
120. Kim, D. & Kim, J.-S. DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA. *Genome Res.* **28**, 1894–1900 (2018).
121. Kim, D., Kang, B.-C. & Kim, J.-S. Identifying genome-wide off-target sites of CRISPR RNA-guided nucleases and deaminases with Digenome-seq. *Nat. Protoc.* **16**, 1170–1192 (2021).
122. Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* **14**, 607–614 (2017).
123. Cameron, P. *et al.* Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat. Methods* **14**, 600–606 (2017).
124. Lazzarotto, C. R. *et al.* CHANGE-seq reveals genetic and epigenetic effects on CRISPR-Cas9 genome-wide activity. *Nat. Biotechnol.* **38**, 1317–1327 (2020).
125. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
126. Tsai, S. Q., Topkar, V. V., Joung, J. K. & Aryee, M. J. Open-source guideseq software for analysis of GUIDE-seq data. *Nat. Biotechnol.* **34**, 483 (2016).
127. Zhang, J.-P. *et al.* Curing hemophilia A by NHEJ-mediated ectopic F8 insertion in the mouse. *Genome Biol.* **20**, 276 (2019).
128. Auer, T. O., Durooure, K., De Cian, A., Concordet, J.-P. & Del Bene, F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* **24**, 142–153 (2014).
129. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
130. He, X. *et al.* Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.* **44**, e85 (2016).
131. Liang, S.-Q. *et al.* Genome-wide detection of CRISPR editing in vivo using GUIDE-tag. *Nat. Commun.* **13**, 437 (2022).
132. Wienert, B. *et al.* Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **364**, 286–289 (2019).
133. Yan, W. X. *et al.* BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, 15058 (2017).
134. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **10**, 361–365 (2013).
135. Wang, X. *et al.* Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* **33**, 175–178 (2015).
136. Chiarle, R. *et al.* Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107–119 (2011).
137. Petri, K. *et al.* Global-scale CRISPR gene editor specificity profiling by ONE-seq identifies population-specific, variant off-target effects. 2021.04.05.438458 Preprint at <https://doi.org/10.1101/2021.04.05.438458> (2021).
138. Huang, H. *et al.* Tag-seq: a convenient and scalable method for genome-wide specificity assessment of CRISPR/Cas nucleases. *Commun. Biol.* **4**, 830 (2021).
139. Breton, C., Clark, P. M., Wang, L., Greig, J. A. & Wilson, J. M. ITR-Seq, a next-generation sequencing assay, identifies genome-wide DNA editing sites in vivo following adeno-associated viral vector-mediated genome editing. *BMC Genomics* **21**, 239 (2020).
140. Akcakaya, P. *et al.* In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature* **561**, 416–419 (2018).
141. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141 (2014).
142. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).
143. Kocak, D. D. *et al.* Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nat. Biotechnol.* **37**, 657–666 (2019).
144. Lesnik, E. A. & Freier, S. M. Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry* **34**, 10807–10815 (1995).
145. Gyi, J. I., Lane, A. N., Conn, G. L. & Brown, T. The orientation and dynamics of the C2'-OH and hydration of RNA and DNA:RNA hybrids. *Nucleic Acids Res.* **26**, 3104–3110 (1998).
146. Yin, H. *et al.* Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat. Chem. Biol.* **14**, 311–316 (2018).
147. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing

- specificity. *Cell* **154**, 1380–1389 (2013).
148. Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32**, 569–576 (2014).
 149. Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32**, 577–582 (2014).
 150. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
 151. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
 152. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
 153. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
 154. Liu, M.-S. *et al.* Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat. Commun.* **11**, 3576 (2020).
 155. Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human haematopoietic stem and progenitor cells. *Nat. Med.* **24**, 1216–1224 (2018).
 156. Huang, X., Yang, D., Zhang, J., Xu, J. & Chen, Y. E. Recent Advances in Improving Gene-Editing Specificity through CRISPR-Cas9 Nuclease Engineering. *Cells* **11**, 2186 (2022).
 157. Enache, O. M. *et al.* Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat. Genet.* **52**, 662–668 (2020).
 158. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
 159. Schirotti, G. *et al.* Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell* **24**, 551-565.e8 (2019).
 160. Ihry, R. J. *et al.* p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).
 161. Álvarez, M. M., Biayna, J. & Supek, F. TP53-dependent toxicity of CRISPR/Cas9 cuts is differential across genomic loci and can confound genetic screening. *Nat. Commun.* **13**, 4520 (2022).
 162. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
 163. Cullot, G. *et al.* CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat. Commun.* **10**, 1136 (2019).
 164. Simpson, B. P., Yrigollen, C. M., Izda, A. & Davidson, B. L. Targeted long-read sequencing captures CRISPR editing and AAV integration outcomes in brain. *Mol. Ther. J. Am. Soc. Gene Ther.* **31**, 760–773 (2023).
 165. Liu, M. *et al.* Global detection of DNA repair outcomes induced by CRISPR-Cas9. *Nucleic Acids Res.* **49**, 8732–8742 (2021).
 166. Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* **33**, 179–186 (2015).
 167. Hanlon, K. S. *et al.* High levels of AAV vector integration into CRISPR-induced DNA breaks. *Nat. Commun.* **10**, 4439 (2019).
 168. Tao, J., Wang, Q., Mendez-Dorantes, C., Burns, K. H. & Chiarle, R. Frequency and mechanisms of LINE-1 retrotransposon insertions at CRISPR/Cas9 sites. *Nat. Commun.* **13**, 3685 (2022).
 169. Papatheanasiou, S. *et al.* Whole chromosome loss and genomic instability in mouse embryos after CRISPR-Cas9 genome editing. *Nat. Commun.* **12**, 5855 (2021).
 170. Zuccaro, M. V. *et al.* Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos. *Cell* **183**, 1650-1664.e15 (2020).
 171. Zuo, E. *et al.* CRISPR/Cas9-mediated targeted chromosome elimination. *Genome Biol.* **18**, (2017).
 172. Conticello, S. G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **9**, 229 (2008).
 173. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
 174. Davidson, N. O. & Shelness, G. S. APOLIPOPROTEIN B: mRNA editing, lipoprotein assembly, and presecretory degradation. *Annu. Rev. Nutr.* **20**, 169–193 (2000).
 175. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
 176. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
 177. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **38**, 620–628 (2020).
 178. Yu, Y. *et al.* Cytosine base editors with minimized unguided DNA and RNA off-target events and

- high on-target activity. *Nat. Commun.* **11**, 2052 (2020).
179. Zuo, E. *et al.* A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nat. Methods* **17**, 600–604 (2020).
180. Lam, D. K. *et al.* Improved cytosine base editors generated from TadA variants. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-022-01611-9.
181. Neugebauer, M. E. *et al.* Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-022-01533-6.
182. Kim, J. *et al.* Structural and kinetic characterization of Escherichia coli TadA, the wobble-specific tRNA deaminase. *Biochemistry* **45**, 6407–6416 (2006).
183. Matthews, M. M. *et al.* Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.* **23**, 426–433 (2016).
184. Gerber, A. P. & Keller, W. An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **286**, 1146–1149 (1999).
185. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
186. Gaudelli, N. M. *et al.* Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* **38**, 892–900 (2020).
187. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
188. Hardouin, G. *et al.* Adenine base editor-mediated correction of the common and severe IVS1-110 (G>A) β -thalassemia mutation. *Blood* **141**, 1169–1179 (2023).
189. Koblan, L. W. *et al.* In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice. *Nature* **589**, 608–614 (2021).
190. Leclerc, D. *et al.* Gene Editing Corrects In Vitro a G>A GLB1 Transition from a GM1 Gangliosidosis Patient. *CRISPR J.* **6**, 17–31 (2023).
191. Antoniou, P. *et al.* Base-editing-mediated dissection of a γ -globin cis-regulatory element for the therapeutic reactivation of fetal hemoglobin expression. *Nat. Commun.* **13**, 6618 (2022).
192. Kluesner, M. G. *et al.* CRISPR-Cas9 cytidine and adenosine base editing of splice-sites mediates highly-efficient disruption of proteins in primary and immortalized cells. *Nat. Commun.* **12**, 2437 (2021).
193. Kim, K. *et al.* Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
194. Kucsu, C. *et al.* CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**, 710–712 (2017).
195. Billon, P. *et al.* CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. *Mol. Cell* **67**, 1068-1079.e4 (2017).
196. Webber, B. R. *et al.* Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nat. Commun.* **10**, 5222 (2019).
197. Yeh, W.-H., Chiang, H., Rees, H. A., Edge, A. S. B. & Liu, D. R. In vivo base editing of post-mitotic sensory cells. *Nat. Commun.* **9**, 2184 (2018).
198. Chen, W. *et al.* CRISPR/Cas9-based Genome Editing in Pseudomonas aeruginosa and Cytidine Deaminase-Mediated Base Editing in Pseudomonas Species. *iScience* **6**, 222–231 (2018).
199. Zong, Y. *et al.* Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
200. Shimatani, Z. *et al.* Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 441–443 (2017).
201. Lu, Y. & Zhu, J.-K. Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. *Mol. Plant* **10**, 523–525 (2017).
202. Marr, E. & Potter, C. J. Base Editing of Somatic Cells Using CRISPR-Cas9 in Drosophila. *CRISPR J.* **4**, 836–845 (2021).
203. Liu, Z. *et al.* Highly efficient RNA-guided base editing in rabbit. *Nat. Commun.* **9**, 2717 (2018).
204. Zafra, M. P. *et al.* Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* **36**, 888–893 (2018).
205. Villiger, L. *et al.* Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nat. Med.* **24**, 1519–1525 (2018).
206. Chadwick, A. C., Wang, X. & Musunuru, K. In Vivo Base Editing of PCSK9 (Proprotein Convertase Subtilisin/Kexin Type 9) as a Therapeutic Alternative to Genome Editing. *Arterioscler. Thromb. Vasc. Biol.* **37**, 1741–1747 (2017).
207. Zhang, Y. *et al.* Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nat. Commun.* **8**, 118 (2017).
208. Tanaka, S. *et al.* In vivo targeted single-nucleotide editing in zebrafish. *Sci. Rep.* **8**, 11423 (2018).
209. Park, D.-S. *et al.* Targeted Base Editing via RNA-Guided Cytidine Deaminases in Xenopus

- laevis Embryos. *Mol. Cells* **40**, 823–827 (2017).
210. Xie, J. *et al.* Efficient base editing for multiple genes and loci in pigs using base editors. *Nat. Commun.* **10**, 2852 (2019).
211. Musunuru, K. *et al.* In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* **593**, 429–434 (2021).
212. Zhang, W. *et al.* Multiplex precise base editing in cynomolgus monkeys. *Nat. Commun.* **11**, 2325 (2020).
213. Li, G. *et al.* Highly efficient and precise base editing in discarded human tripronuclear embryos. *Protein Cell* **8**, 776–779 (2017).
214. Liang, P. *et al.* Correction of β -thalassemia mutant by base editor in human embryos. *Protein Cell* **8**, 811–822 (2017).
215. Montiel-Gonzalez, M. F., Vallecillo-Viejo, I., Yudowski, G. A. & Rosenthal, J. J. C. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18285–18290 (2013).
216. Montiel-González, M. F., Vallecillo-Viejo, I. C. & Rosenthal, J. J. C. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res.* **44**, e157 (2016).
217. Fukuda, M. *et al.* Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. *Sci. Rep.* **7**, 41478 (2017).
218. Vogel, P. *et al.* Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat. Methods* **15**, 535–538 (2018).
219. Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027 (2017).
220. Porto, E. M., Komor, A. C., Slaymaker, I. M. & Yeo, G. W. Base editing: advances and therapeutic opportunities. *Nat. Rev. Drug Discov.* **19**, 839–859 (2020).
221. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
222. Gehrke, J. M. *et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).
223. Lee, S. *et al.* Single C-to-T substitution using engineered APOBEC3G-nCas9 base editors with minimum genome- and transcriptome-wide off-target effects. *Sci. Adv.* **6**, eaba1773 (2020).
224. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
225. Komor, A. C. *et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
226. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
227. d’Adda di Fagagna, F., Weller, G. R., Doherty, A. J. & Jackson, S. P. The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku. *EMBO Rep.* **4**, 47–52 (2003).
228. Gu, S., Bodai, Z., Cowan, Q. T. & Komor, A. C. Base Editors: Expanding the Types of DNA Damage Products Harnessed for Genome Editing. *Gene Genome Ed.* **1**, 100005 (2021).
229. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
230. Friedland, A. E. *et al.* Characterization of *Staphylococcus aureus* Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol.* **16**, 257 (2015).
231. Amrani, N. *et al.* NmeCas9 is an intrinsically high-fidelity genome-editing platform. *Genome Biol.* **19**, 214 (2018).
232. Chatterjee, P., Jakimo, N. & Jacobson, J. M. Minimal PAM specificity of a highly similar SpCas9 ortholog. *Sci. Adv.* **4**, eaau0766 (2018).
233. Collias, D. & Beisel, C. L. CRISPR technologies and the search for the PAM-free nuclease. *Nat. Commun.* **12**, 555 (2021).
234. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **472**, 499–503 (2011).
235. Miller, S. M. *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* **38**, 471 (2020).
236. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
237. Alsøe, L. *et al.* Uracil Accumulation and Mutagenesis Dominated by Cytosine Deamination in CpG Dinucleotides in Mice Lacking UNG and SMUG1. *Sci. Rep.* **7**, 7199 (2017).
238. Kim, D. *et al.* Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017).
239. Liang, P. *et al.* Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat. Commun.* **10**, 67 (2019).
240. Kim, D., Kim, D.-E., Lee, G., Cho, S.-I. & Kim, J.-S. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nat. Biotechnol.* **37**, 430–435 (2019).

241. Rees, H. A. *et al.* Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).
242. Grünewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
243. Kurt, I. C. *et al.* CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2021).
244. Chen, L. *et al.* Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nat. Commun.* **12**, 1384 (2021).
245. Zhao, D. *et al.* Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.* **39**, 35–40 (2021).
246. Yuan, T. *et al.* Optimization of C-to-G base editors with sequence context preference predictable by machine learning methods. *Nat. Commun.* **12**, 4902 (2021).
247. Tong, H. *et al.* Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-022-01595-6.
248. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
249. Chen, P. J. *et al.* Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652.e29 (2021).
250. Ferreira da Silva, J. *et al.* Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. *Nat. Commun.* **13**, 760 (2022).
251. Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome manipulation. *Nat. Rev. Genet.* **24**, 161–177 (2023).
252. Liu, P. *et al.* Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat. Commun.* **12**, 2121 (2021).
253. Böck, D. *et al.* In vivo prime editing of a metabolic liver disease in mice. *Sci. Transl. Med.* **14**, eabl9238 (2022).
254. Liu, B. *et al.* A split prime editor with untethered reverse transcriptase and circular RNA template. *Nat. Biotechnol.* **40**, 1388–1393 (2022).
255. Zheng, C. *et al.* A flexible split prime editor using truncated reverse transcriptase improves dual-AAV delivery in mouse liver. *Mol. Ther.* **30**, 1343–1351 (2022).
256. Gao, Z. *et al.* A truncated reverse transcriptase enhances prime editing by split AAV vectors. *Mol. Ther. J. Am. Soc. Gene Ther.* **30**, 2942–2951 (2022).
257. Song, M. *et al.* Generation of a more efficient prime editor 2 by addition of the Rad51 DNA-binding domain. *Nat. Commun.* **12**, 5617 (2021).
258. Chen, R. *et al.* Enhancement of a prime editing system via optimal recruitment of the pioneer transcription factor P65. *Nat. Commun.* **14**, 257 (2023).
259. Velimirovic, M. *et al.* Peptide fusion improves prime editing efficiency. *Nat. Commun.* **13**, 3512 (2022).
260. Hsu, J. Y. *et al.* PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *Nat. Commun.* **12**, 1034 (2021).
261. Hwang, G.-H. *et al.* PE-Designer and PE-Analyzer: web-based design and analysis tools for CRISPR prime editing. *Nucleic Acids Res.* **49**, W499–W504 (2021).
262. Li, Y., Chen, J., Tsai, S. Q. & Cheng, Y. Easy-Prime: a machine learning-based prime editor design tool. *Genome Biol.* **22**, 235 (2021).
263. Sm, S., Me, K., Ms, S., Z, K. & Je, C. PnB Designer: a web application to design prime and base editor guide RNAs for animals and plants. *BMC Bioinformatics* **22**, (2021).
264. Chow, R. D., Chen, J. S., Shen, J. & Chen, S. A web tool for the design of prime-editing guide RNAs. *Nat. Biomed. Eng.* **5**, 190–194 (2021).
265. Kim, H. K. *et al.* Predicting the efficiency of prime editing guide RNAs in human cells. *Nat. Biotechnol.* **39**, 198–206 (2021).
266. Yu, G. *et al.* Prediction of efficiencies for diverse prime editing systems in multiple cell types. *Cell* **186**, 2256–2272.e23 (2023).
267. Nelson, J. W. *et al.* Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2022).
268. Liu, Y. *et al.* Enhancing prime editing by Csy4-mediated processing of pegRNA. *Cell Res.* **31**, 1134–1136 (2021).
269. Li, X. *et al.* Enhancing prime editing efficiency by modified pegRNA with RNA G-quadruplexes. *J. Mol. Cell Biol.* **14**, mjac022 (2022).
270. Zhang, G. *et al.* Enhancement of prime editing via xrRNA motif-joined pegRNA. *Nat. Commun.* **13**, 1856 (2022).
271. Kim, D. Y., Moon, S. B., Ko, J.-H., Kim, Y.-S. & Kim, D. Unbiased investigation of specificities of prime editing systems in human cells. *Nucleic Acids Res.* **48**, 10576–10589 (2020).
272. Jin, S. *et al.* Genome-wide specificity of prime editors in plants. *Nat. Biotechnol.* **39**, 1292–1299

(2021).

273. Schene, I. F. *et al.* Prime editing for functional repair in patient-derived disease models. *Nat. Commun.* **11**, 5352 (2020).
274. Park, S.-J. *et al.* Targeted mutagenesis in mouse cells and embryos using an enhanced prime editor. *Genome Biol.* **22**, 170 (2021).
275. Petri, K. *et al.* CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol.* **40**, 189–193 (2022).
276. Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T. & Lee, S. Y. A versatile genetic engineering toolkit for *E. coli* based on CRISPR-prime editing. *Nat. Commun.* **12**, 5206 (2021).
277. Bosch, J. A., Birchak, G. & Perrimon, N. Precise genome engineering in *Drosophila* using prime editing. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2021996118 (2021).
278. Perroud, P.-F. *et al.* Prime Editing in the model plant *Physcomitrium patens* and its potential in the tetraploid potato. *Plant Sci. Int. J. Exp. Plant Biol.* **316**, 111162 (2022).
279. Jiang, Y.-Y. *et al.* Prime editing efficiently generates W542L and S621I double mutations in two ALS genes in maize. *Genome Biol.* **21**, 257 (2020).
280. Habib, O., Habib, G., Hwang, G.-H. & Bae, S. Comprehensive analysis of prime editing outcomes in human embryonic stem cells. *Nucleic Acids Res.* **50**, 1187–1197 (2022).
281. Jang, H. *et al.* Application of prime editing to the correction of mutations and phenotypes in adult mice with liver and eye diseases. *Nat. Biomed. Eng.* **6**, 181–194 (2022).
282. Li, C. *et al.* In vivo HSC prime editing rescues Sickle Cell Disease in a mouse model. *Blood* blood.2022018252 (2023) doi:10.1182/blood.2022018252.
283. Qin, H. *et al.* Vision rescue via unconstrained in vivo prime editing in degenerating neural retinas. *J. Exp. Med.* **220**, e20220776 (2023).
284. Qian, Y. *et al.* Efficient and precise generation of Tay-Sachs disease model in rabbit by prime editing system. *Cell Discov.* **7**, 50 (2021).
285. Geurts, M. H. *et al.* Evaluating CRISPR-based prime editing for cancer modeling and CFTR repair in organoids. *Life Sci. Alliance* **4**, e202000940 (2021).
286. Choi, J. *et al.* A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature* **608**, 98 (2022).
287. Zhuang, Y. *et al.* Increasing the efficiency and precision of prime editing with guide RNA pairs. *Nat. Chem. Biol.* **18**, 29–37 (2022).
288. Jiang, T., Zhang, X.-O., Weng, Z. & Xue, W. Deletion and replacement of long genomic sequences using prime editing. *Nat. Biotechnol.* **40**, 227–234 (2022).
289. Anzalone, A. V. *et al.* Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat. Biotechnol.* **40**, 731–740 (2022).
290. Wang, J. *et al.* Efficient targeted insertion of large DNA fragments without DNA donors. *Nat. Methods* **19**, 331–340 (2022).
291. Choi, J. *et al.* Precise genomic deletions using paired prime editing. *Nat. Biotechnol.* **40**, 218–226 (2022).
292. Kweon, J. *et al.* Targeted genomic translocations and inversions generated using a paired prime editing strategy. *Mol. Ther. J. Am. Soc. Gene Ther.* **31**, 249–259 (2023).
293. Cj, T. & Bp, K. Recent Advances in Double-Strand Break-Free Kilobase-Scale Genome Editing Technologies. *Biochemistry* (2022) doi:10.1021/acs.biochem.2c00311.
294. Rajeev, L., Malanowska, K. & Gardner, J. F. Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiol. Mol. Biol. Rev. MMBR* **73**, 300–309 (2009).
295. Karpinski, J. *et al.* Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nat. Biotechnol.* **34**, 401–409 (2016).
296. Lansing, F. *et al.* A heterodimer of evolved designer-recombinases precisely excises a human genomic DNA locus. *Nucleic Acids Res.* **48**, 472–485 (2020).
297. Durrant, M. G. *et al.* Systematic discovery of recombinases for efficient integration of large DNA sequences into the human genome. *Nat. Biotechnol.* **41**, 488–499 (2023).
298. Lansing, F. *et al.* Correction of a Factor VIII genomic inversion with designer-recombinases. *Nat. Commun.* **13**, 422 (2022).
299. Chaikind, B., Bessen, J. L., Thompson, D. B., Hu, J. H. & Liu, D. R. A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic Acids Res.* **44**, 9758–9770 (2016).
300. Ioannidi, E. I. *et al.* Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed integrases. 2021.11.01.466786 Preprint at <https://doi.org/10.1101/2021.11.01.466786> (2021).
301. Aziz, R. K., Breitbart, M. & Edwards, R. A. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res.* **38**, 4207–4217 (2010).
302. Tou, C. J. & Kleinstiver, B. P. Recent Advances in Double-Strand Break-Free Kilobase-Scale Genome Editing Technologies. *Biochemistry* (2022) doi:10.1021/acs.biochem.2c00311.
303. Tou, C. J., Orr, B. & Kleinstiver, B. P. Precise cut-and-paste DNA insertion using engineered

- type V-K CRISPR-associated transposases. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-022-01574-x.
304. Strecker, J. *et al.* RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).
305. M, A. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, (2018).
306. Burmistrz, M., Krakowski, K. & Krawczyk-Balska, A. RNA-Targeting CRISPR-Cas Systems and Their Applications. *Int. J. Mol. Sci.* **21**, 1122 (2020).
307. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).
308. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**, 442–451 (2013).
309. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
310. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
311. Li, T. *et al.* CRISPR/Cas9 therapeutics: progress and prospects. *Signal Transduct. Target. Ther.* **8**, 36 (2023).
312. Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229–236 (2020).
313. Kaeppl, C. *et al.* A largely random AAV integration profile after LPLD gene therapy. *Nat. Med.* **19**, 889–891 (2013).
314. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).
315. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Mol. Ther. J. Am. Soc. Gene Ther.* **16**, 1073–1080 (2008).
316. Yin, H. *et al.* Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.* **15**, 541–555 (2014).
317. Kulkarni, J. A. *et al.* The current landscape of nucleic acid therapeutics. *Nat. Nanotechnol.* **16**, 630–643 (2021).
318. Akinc, A. *et al.* The Onpattro story and the clinical translation of nanomedicines containing nucleic acid-based drugs. *Nat. Nanotechnol.* **14**, 1084–1087 (2019).
319. Wang, X. *et al.* Preparation of selective organ-targeting (SORT) lipid nanoparticles (LNPs) using multiple technical methods for tissue-specific mRNA delivery. *Nat. Protoc.* **18**, 265–291 (2023).
320. Dilliard, S. A., Cheng, Q. & Siegwart, D. J. On the mechanism of tissue-specific mRNA delivery by selective organ targeting nanoparticles. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2109256118 (2021).
321. Arabi, F., Mansouri, V. & Ahmadbeigi, N. Gene therapy clinical trials, where do we go? An overview. *Biomed. Pharmacother. Biomedecine Pharmacother.* **153**, 113324 (2022).
322. Gaspar, H. B. *et al.* Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet Lond. Engl.* **364**, 2181–2187 (2004).
323. Hacein-Bey-Abina, S. *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* **118**, 3132–3142 (2008).
324. Howe, S. J. *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J. Clin. Invest.* **118**, 3143–3150 (2008).
325. Cavazzana, M., Bushman, F. D., Miccio, A., André-Schmutz, I. & Six, E. Gene therapy targeting haematopoietic stem cells for inherited diseases: progress and challenges. *Nat. Rev. Drug Discov.* **18**, 447–462 (2019).
326. Hacein-Bey-Abina, S. *et al.* A Modified γ -Retrovirus Vector for X-Linked Severe Combined Immunodeficiency. *N. Engl. J. Med.* **371**, 1407–1417 (2014).
327. S, Z. *et al.* A self-inactivating lentiviral vector for SCID-X1 gene therapy that does not activate LMO2 expression in human T cells. *Blood* **116**, (2010).
328. Manno, C. S. *et al.* Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat. Med.* **12**, 342–347 (2006).
329. Mingozzi, F. *et al.* CD8(+) T-cell responses to adeno-associated virus capsid in humans. *Nat. Med.* **13**, 419–422 (2007).
330. High, K. A. *et al.* A Phase 1/2 Trial of Investigational Spk-8011 in Hemophilia a Demonstrates Durable Expression and Prevention of Bleeds. *Blood* **132**, 487 (2018).
331. Boutin, S. *et al.* Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors. *Hum. Gene Ther.* **21**, 704–712 (2010).
332. Bulaklak, K. & Gersbach, C. A. The once and future gene therapy. *Nat. Commun.* **11**, 5820 (2020).
333. Tse, L. V. *et al.* Structure-guided evolution of antigenically distinct adeno-associated virus variants for immune evasion. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4812–E4821 (2017).
334. Maheshri, N., Koerber, J. T., Kaspar, B. K. & Schaffer, D. V. Directed evolution of adeno-

- associated virus yields enhanced gene delivery vectors. *Nat. Biotechnol.* **24**, 198–204 (2006).
335. Bittner, R. E. *et al.* Serum antibodies to the deleted dystrophin sequence after cardiac transplantation in a patient with Becker's muscular dystrophy. *N. Engl. J. Med.* **333**, 732–733 (1995).
336. Ferrer, A., Wells, K. E. & Wells, D. J. Immune responses to dystropin: implications for gene therapy of Duchenne muscular dystrophy. *Gene Ther.* **7**, 1439–1446 (2000).
337. Mount, J. D. *et al.* Sustained phenotypic correction of hemophilia B dogs with a factor IX null mutation by liver-directed gene therapy. *Blood* **99**, 2670–2676 (2002).
338. Biswas, M., Kumar, S. R. P., Terhorst, C. & Herzog, R. W. Gene Therapy With Regulatory T Cells: A Beneficial Alliance. *Front. Immunol.* **9**, 554 (2018).
339. Wilson, J. M. & Flotte, T. R. Moving Forward After Two Deaths in a Gene Therapy Trial of Myotubular Myopathy. *Hum. Gene Ther.* **31**, 695–696 (2020).
340. Riyad, J. M. & Weber, T. Intracellular trafficking of adeno-associated virus (AAV) vectors: challenges and future directions. *Gene Ther.* **28**, 683–696 (2021).
341. Mendell, J. R. *et al.* Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *N. Engl. J. Med.* **377**, 1713–1722 (2017).
342. Trapani, I. *et al.* Effective delivery of large genes to the retina by dual AAV vectors. *EMBO Mol. Med.* **6**, 194–211 (2014).
343. Lai, Y. *et al.* Efficient in vivo gene expression by trans-splicing adeno-associated viral vectors. *Nat. Biotechnol.* **23**, 1435–1439 (2005).
344. Nakai, H., Storm, T. A. & Kay, M. A. Increasing the size of rAAV-mediated expression cassettes in vivo by intermolecular joining of two complementary vectors. *Nat. Biotechnol.* **18**, 527–532 (2000).
345. Sun, L., Li, J. & Xiao, X. Overcoming adeno-associated virus vector size limitation through viral DNA heterodimerization. *Nat. Med.* **6**, 599–602 (2000).
346. White, S. J. *et al.* Targeted gene delivery to vascular tissue in vivo by tropism-modified adeno-associated virus vectors. *Circulation* **109**, 513–519 (2004).
347. Körbelin, J. *et al.* A brain microvasculature endothelial cell-specific viral vector with the potential to treat neurovascular and neurological diseases. *EMBO Mol. Med.* **8**, 609–625 (2016).
348. Müller, O. J. *et al.* Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat. Biotechnol.* **21**, 1040–1046 (2003).
349. Körbelin, J. *et al.* Pulmonary Targeting of Adeno-associated Viral Vectors by Next-generation Sequencing-guided Screening of Random Capsid Displayed Peptide Libraries. *Mol. Ther. J. Am. Soc. Gene Ther.* **24**, 1050–1061 (2016).
350. Raguram, A., Banskota, S. & Liu, D. R. Therapeutic in vivo delivery of gene editing agents. *Cell* **185**, 2806–2827 (2022).
351. Adams, D., Koike, H., Slama, M. & Coelho, T. Hereditary transthyretin amyloidosis: a model of medical progress for a fatal disease. *Nat. Rev. Neurol.* **15**, 387–404 (2019).
352. Piel, F. B. *et al.* Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *Lancet Lond. Engl.* **381**, 142–151 (2013).
353. Antoniou, P., Miccio, A. & Brusson, M. Base and Prime Editing Technologies for Blood Disorders. *Front. Genome Ed.* **3**, 618406 (2021).
354. Newby, G. A. *et al.* Base editing of hematopoietic stem cells rescues sickle cell disease in mice. *Nature* **595**, 295–302 (2021).
355. Zeng, J. *et al.* Therapeutic base editing of human hematopoietic stem cells. *Nat. Med.* **26**, 535–541 (2020).
356. Wang, L. *et al.* Reactivation of γ -globin expression through Cas9 or base editor to treat β -hemoglobinopathies. *Cell Res.* **30**, 276–278 (2020).
357. Zarghamian, P., Klermund, J. & Cathomen, T. Clinical genome editing to treat sickle cell disease—A brief update. *Front. Med.* **9**, 1065377 (2022).
358. Germino-Watnick, P. *et al.* Hematopoietic Stem Cell Gene-Addition/Editing Therapy in Sickle Cell Disease. *Cells* **11**, 1843 (2022).
359. Park, S. H. & Bao, G. CRISPR/Cas9 gene editing for curing sickle cell disease. *Transfus. Apher. Sci. Off. J. World Apher. Assoc. Off. J. Eur. Soc. Haemapheresis* **60**, 103060 (2021).
360. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β -Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).
361. Knipping, F. *et al.* Disruption of HIV-1 co-receptors CCR5 and CXCR4 in primary human T cells and hematopoietic stem and progenitor cells using base editing. *Mol. Ther.* **30**, 130–144 (2022).
362. Zhou, H., Wang, X., Steer, C. J., Song, G. & Niu, J. Efficient silencing of hepatitis B virus S gene through CRISPR-mediated base editing. *Hepatol. Commun.* **6**, 1652–1663 (2022).
363. Rothgangl, T. *et al.* In vivo adenine base editing of PCSK9 in macaques reduces LDL cholesterol levels. *Nat. Biotechnol.* **39**, 949–957 (2021).
364. Chemello, F. *et al.* Precise correction of Duchenne muscular dystrophy exon deletion mutations by base and prime editing. *Sci. Adv.* **7**, eabg4910 (2021).
365. Xu, L. *et al.* Efficient precise in vivo base editing in adult dystrophic mice. *Nat. Commun.* **12**,

- 3719 (2021).
366. Suh, S. *et al.* Restoration of visual function in adult mice with an inherited retinal disease via adenine base editing. *Nat. Biomed. Eng.* **5**, 169–178 (2021).
367. Li, C. *et al.* In vivo HSPC gene therapy with base editors allows for efficient reactivation of fetal γ -globin in β -YAC mice. *Blood Adv.* **5**, 1122–1135 (2021).
368. Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat. Biotechnol.* **34**, 328–333 (2016).
369. Yin, H. *et al.* Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat. Biotechnol.* **35**, 1179–1187 (2017).
370. Miller, J. B. *et al.* Non-Viral CRISPR/Cas Gene Editing In Vitro and In Vivo Enabled by Synthetic Nanoparticle Co-Delivery of Cas9 mRNA and sgRNA. *Angew. Chem. Int. Ed Engl.* **56**, 1059–1063 (2017).
371. Jiang, C. *et al.* A non-viral CRISPR/Cas9 delivery system for therapeutically targeting HBV DNA and pscs9 in vivo. *Cell Res.* **27**, 440–443 (2017).
372. Villiger, L. *et al.* In vivo cytidine base editing of hepatocytes without detectable off-target mutations in RNA and DNA. *Nat. Biomed. Eng.* **5**, 179–189 (2021).
373. Mangeot, P. E. *et al.* Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat. Commun.* **10**, 45 (2019).
374. Banskota, S. *et al.* Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. *Cell* **185**, 250-265.e16 (2022).
375. Hamilton, J. R. *et al.* Programmable enveloped delivery vehicles for human genome engineering in vivo. 2022.08.24.505004 Preprint at <https://doi.org/10.1101/2022.08.24.505004> (2023).
376. Zhang, H., Sun, R., Fei, J., Chen, H. & Lu, D. Correction of Beta-Thalassemia IVS-II-654 Mutation in a Mouse Model Using Prime Editing. *Int. J. Mol. Sci.* **23**, 5948 (2022).
377. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **17**, 405–424 (2015).
378. Acedo, A., Hernández-Moro, C., Curiel-García, Á., Díez-Gómez, B. & Velasco, E. A. Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum. Mutat.* **36**, 210–221 (2015).
379. Bouvet, D. *et al.* Methylation Tolerance-Based Functional Assay to Assess Variants of Unknown Significance in the MLH1 and MSH2 Genes and Identify Patients With Lynch Syndrome. *Gastroenterology* **157**, 421–431 (2019).
380. Hawn, M. T. *et al.* Evidence for a connection between the mismatch repair system and the G2 cell cycle checkpoint. *Cancer Res.* **55**, 3721–3725 (1995).
381. Nabhan, J. F. *et al.* Intrathecal delivery of frataxin mRNA encapsulated in lipid nanoparticles to dorsal root ganglia as a potential therapeutic for Friedreich’s ataxia. *Sci. Rep.* **6**, 20019 (2016).
382. Hauck, E. S. & Hecker, J. G. Non-Viral Delivery of RNA Gene Therapy to the Central Nervous System. *Pharmaceutics* **14**, 165 (2022).
383. Gross, A. L. *et al.* Intravenous delivery of adeno-associated viral gene therapy in feline GM1 gangliosidosis. *Brain J. Neurol.* **145**, 655–669 (2022).
384. Teixeira, A. F., Ten Dijke, P. & Zhu, H.-J. On-Target Anti-TGF- β Therapies Are Not Succeeding in Clinical Cancer Treatments: What Are Remaining Challenges? *Front. Cell Dev. Biol.* **8**, 605 (2020).
385. Jinnin, M., Ihn, H. & Tamaki, K. Characterization of SIS3, a Novel Specific Inhibitor of Smad3, and Its Effect on Transforming Growth Factor- β 1-Induced Extracellular Matrix Expression. *Mol. Pharmacol.* **69**, 597–607 (2006).
386. Liu, L. *et al.* Smad2 and Smad3 have differential sensitivity in relaying TGF β signaling and inversely regulate early lineage specification. *Sci. Rep.* **6**, 21602 (2016).
387. Kong, W. *et al.* CellTagging: combinatorial indexing to simultaneously map lineage and identity at single-cell resolution. *Nat. Protoc.* **15**, 750–772 (2020).
388. Urnov, F. D. Imagine CRISPR cures. *Mol. Ther. J. Am. Soc. Gene Ther.* **29**, 3103–3106 (2021).
389. Reichart, D. *et al.* Efficient in vivo genome editing prevents hypertrophic cardiomyopathy in mice. *Nat. Med.* **29**, 412–421 (2023).
390. Ely, Z. A. *et al.* A prime editor mouse to model a broad spectrum of somatic mutations in vivo. *Nat. Biotechnol.* 1–13 (2023) doi:10.1038/s41587-023-01783-y.
391. Levesque, S. *et al.* Marker-free co-selection for successive rounds of prime editing in human cells. *Nat. Commun.* **13**, 5909 (2022).
392. Coelho, M. A. *et al.* Base editing screens map mutations affecting interferon- γ signaling in cancer. *Cancer Cell* **41**, 288-303.e6 (2023).
393. Erwood, S. *et al.* Saturation variant interpretation using CRISPR prime editing. *Nat. Biotechnol.* **40**, 885–895 (2022).
394. Essletzbichler, P. *et al.* Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res.* **24**, 2059–2065 (2014).

395. Clouaire, T. & Legube, G. DNA double strand break repair pathway choice: a chromatin based decision? *Nucl. Acids Res.* **43**, 107–113 (2015).
396. Daley, J. M. & Sung, P. 53BP1, BRCA1, and the Choice between Recombination and End Joining at DNA Double-Strand Breaks. *Mol. Cell. Biol.* **34**, 1380–1388 (2014).
397. Bunting, S. F. *et al.* 53BP1 Inhibits Homologous Recombination in Brca1-Deficient Cells by Blocking Resection of DNA Breaks. *Cell* **141**, 243–254 (2010).
398. Panier, S. & Boulton, S. J. Double-strand break repair: 53BP1 comes into focus. *Nat. Rev. Mol. Cell Biol.* **15**, 7–18 (2014).
399. Trenner, A. & Sartori, A. A. Harnessing DNA Double-Strand Break Repair for Cancer Treatment. *Front. Oncol.* **9**, 1388 (2019).

ANNEXES

ANNEXE I : Mécanismes de réparation des cassures double brin de l'ADN

Les mécanismes de réparation des CDB les plus fréquents sont la jonction des extrémités non homologues classique (c-NHEJ) et la recombinaison homologue dirigée (HDR). Plus rarement, on retrouve également le mécanisme de jonction des extrémités non homologues alternative (alt-NHEJ ou MMEJ) et l'hybridation de l'ADN simple brin (SSA). La **figure 23** résume les mécanismes moléculaires régissant ces voies de réparation, et sont précisément décrits dans ces revues⁷⁴⁻⁷⁷.

Dans les cellules en mitose, les CDB de l'ADN peuvent être causées par diverses substances endogènes ou exogènes, telles que les agents chimiques, les espèces réactives de l'oxygène, ou les radiations ionisantes par exemple. Ces CDB sont considérées comme pathologiques et aléatoires (excepté dans les lymphocytes lors de la recombinaison V(D)J et lors de la méiose)⁷⁴. Comme vu auparavant, il est également possible de générer des CDB de l'ADN programmées avec les outils d'édition du génome.

Lors de l'apparition d'une CDB de l'ADN, de nombreux facteurs vont influencer le type de voie de réparation qui sera utilisée : le type de cassure, la séquence de l'ADN lésé, le contexte chromatinien et surtout le cycle cellulaire^{76,77,395}. Par exemple, la réparation par HDR est active seulement dans les cellules en division. En effet, elle nécessite l'intervention de protéines spécifiquement exprimées en fin de phase S/début G2 du cycle cellulaire ainsi qu'une matrice de réparation intacte (chromatide sœur). Au contraire, la réparation par NHEJ prédomine lors des phases G0 et G1^{76,396}. Cela peut s'expliquer également par l'antagonisme réciproque entre les protéines 53BP1 et BRCA1, qui privilégient respectivement la réparation par NHEJ ou HDR selon la phase du cycle cellulaire (**figure 23**). Lors de la phase G1, la protéine 53BP1 va en effet inhiber la résection de l'ADN lésé et favoriser le recrutement de protéines impliquées dans la réparation NHEJ, tandis que en phase S/G2, la protéine BRCA1 va empêcher le recrutement de 53BP1 et favoriser la résection et l'invasion des brins d'ADN³⁹⁶⁻³⁹⁸.

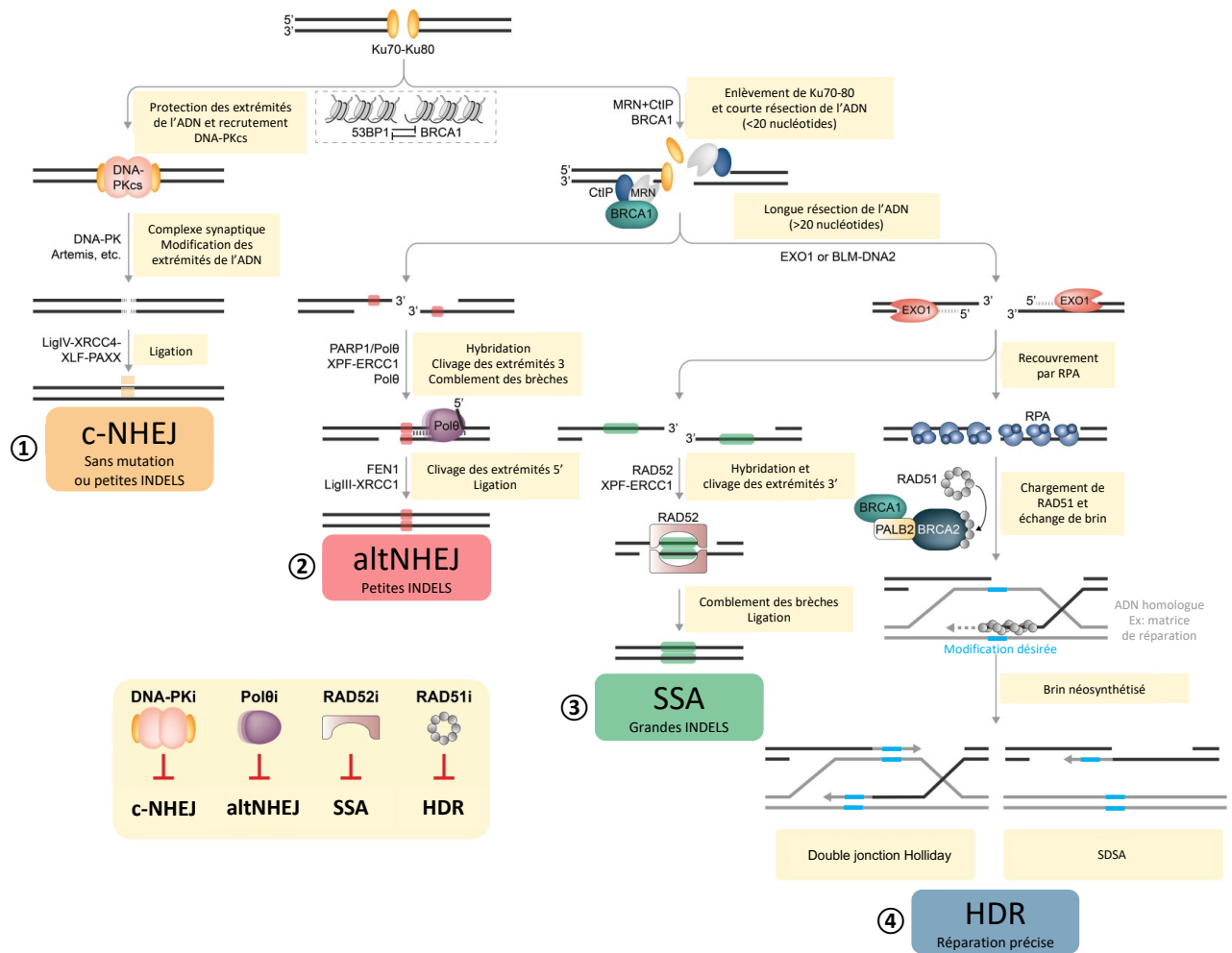


Figure 23 : Les voies de réparation de l'ADN dans les cellules mammaliennes.

(1) La réparation c-NHEJ est initié par la reconnaissance et la protection des extrémités d'ADN coupées par l'hétérodimère Ku70/80. Ce complexe est ensuite stabilisé par le recrutement de la sous unité catalytique de la protéine kinase dépendante de l'ADN (DNA-PKcs). Cette sous unité va relier les deux extrémités d'ADN pour former un complexe synaptique. Dans un second temps, le recrutement de l'ADN ligase IV (LIG4) ainsi que des protéines échafaudage XRCC4, XLF et PAXX vont permettre aux deux extrémités d'ADN de s'aligner et de se rapprocher encore plus. La ligation des deux brins est assurée par LIG4 puis le nouvel ADN réparé est libéré. (2) Lorsque l'hétérodimère Ku70/80 ne protège plus les extrémités de l'ADN, les extrémités libres 5' sont résequées par le complexe MRN (MRE11-RAD50-NBS1), BRCA1 et CtIP. Cela engage l'ADN vers les autres voies de réparation : altNHEJ, SSA ou bien HDR. Lorsque cette résection dite est courte (jusqu'à environ 20 nucléotides), les deux extrémités protubérantes 3' de l'ADN vont s'hybrider (microhomologies, en rouge sur le schéma) à l'aide de la protéine PARP1 afin de rapprocher les deux brins et initier la réparation altNHEJ. PARP1 recrute ensuite l'ADN polymérase θ qui utilise les extrémités 3' de l'ADN comme amorce pour combler les brèches. Les extrémités 3' et 5' additionnelles (rabats) sont ensuite retirées par les nucléases XPF/ERCC1 et FEN1. Pour terminer, les extrémités d'ADN sont réassemblées par le complexe LigIII-XRCC1. (3) Une résection beaucoup plus importante des deux extrémités de l'ADN, par l'exonucléase 5' \rightarrow 3' EXO1, peut ensuite mener à la réparation par SSA ou bien la réparation par HDR. Afin de rapprocher les deux brins

réséqués, la réparation par SSA exploite les homologies de 20 à 25 paires de bases (en vert sur le schéma), stabilisées par RAD52. Les brèches sont ensuite comblées et l'ADN ligué mais les acteurs ne sont pas encore bien connus. **(4)** Lors de la réparation par HDR, les longues extrémités simple brin 3' sont protégées par l'hétérodimère RPA. A l'aide du complexe BRCA1-PALB2-BRCA2, RPA est ensuite remplacé par la recombinaise RAD51 qui permet l'hybridation des extrémités simple brin de l'ADN avec une séquence homologue. Cette séquence homologue peut être une chromatide sœur ou une matrice de réparation d'ADN conçue par l'expérimentateur, contenant une modification génétique d'intérêt. Ainsi, une des extrémités 3' des filaments d'ADN recouverts par RAD51 est utilisée comme amorce pour initier la réaction de polymérisation à partir de cette matrice. La modification génétique d'intérêt est alors recopiée par l'ADN polymérase δ . Le fragment d'ADN nouvellement synthétisé à partir de la matrice de réparation est ensuite réparé par la voie d'hybridation de l'ADN dépendante de la synthèse (SDSA) puis intégré dans l'ADN génomique. Autrement, si la 2^{ème} extrémité 3' de l'ADN doit également être réparée, une double jonction Holliday est formée avec la matrice de réparation. La jonction Holliday est ensuite réparée sans croisement ou avec croisement. Schéma traduit et adapté de A. Trenneret et al.³⁹⁹.

ANNEXE II : Résumés d'articles collaboratifs

Article 1 : 10.1016/j.coi.2020.12.003



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in
Immunology



Detrimental activation of AhR pathway in cancer: an overview of therapeutic strategies

Delphine Leclerc^{1,5}, Ananda Christina Staats Pires^{2,3,5},
Gilles J Guillemain^{2,4,6} and David Gilot^{1,6}

Sustained transcriptional activation of the aryl hydrocarbon receptor (AhR) promotes tumour growth and impairs the immune defence, at least for cutaneous melanoma and glioma. AhR ligands are produced by the tumour microenvironment (TME) and by the tumour itself (intracrine). The recent identification of interleukin-4-induced-1 (IL4I1), a parallel pathway to indoleamine 2,3-dioxygenase 1 (IDO1)/tryptophan 2,3-dioxygenase (TDO), and its ability to generate AhR ligands, confirms that a complete inhibition of AhR ligand production might be difficult to reach. Here, we have focused on recent discoveries explaining the large varieties of AhR ligands and the functional consequences in terms of cancer cell plasticity and consecutive therapy resistance. We also examined therapeutic strategies targeting the AhR signalling pathway and their possible adverse effects. Since the end of 2019, two phase I clinical trials have investigated the ability of the AhR antagonist to 'reset' the immune system and re-sensitize the cancer cells to therapies by preventing their dedifferentiation.

Addresses

¹Inserm U1242, Université de Rennes, France, Centre de lutte contre le cancer Eugène Marquis, Rennes, France

²Neuroinflammation Group, Department of Biomedical Sciences, Centre for Motor Neuron Disease Research, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, NSW, Australia

³Laboratório de Bioenergética e Estresse Oxidativo, Departamento de Bioquímica, Centro de Ciências Biológicas, Universidade Federal de Santa Catarina, Florianópolis, Brazil

Corresponding author: Gilot, David (david.gilot@univ-rennes1.fr)

⁴Pandis.org.

⁵co-first author.

⁶co-last author.

Current Opinion in Immunology 2021, 70:15–26

This review comes from a themed issue on **Special section on indoles as immune regulators**

Edited by **Ursula Grohmann**

<https://doi.org/10.1016/j.coi.2020.12.003>

0952-7915/© 2021 Elsevier Ltd. All rights reserved.

Introduction

The ubiquitous AhR protein belongs to the basic helix-loop-helix/Per-Arnt-Sim (bHLH/PAS) family. AhR

(bHLHe76) is widely distributed in tissues and species [1]. It is found in the fruit fly *Drosophila melanogaster*, the nematode *Caenorhabditis elegans*, the zebrafish *Danio rerio*, the pufferfish *Takifugu rubripes* and other model species, facilitating basic investigations of this bHLH/PAS transcription factor [1,2]. In vertebrates, this receptor has evolved to bind structurally different ligands, including xenobiotics (e.g. aryl hydrocarbons (AH)) [3]. Sustained AhR activation and AhR loss-of-function are deleterious, suggesting that minimal AhR activity is healthy [4*,5,6]. To prevent the deleterious effects of AhR persistent activation, it seems important to consider the diversity of AhR ligands released by bacteria, yeast, fungi, normal and cancer cells, as reviewed in other articles [3,7–10]. Thus, we detailed only the recent findings on AhR ligands and the therapeutic opportunities to short-circuit the AhR signalling pathway.

AhR, an ubiquitous transcription factor, acts as a sensor

A wide variety of compounds coming from different sources can be sensed by AhR. Among them, environmental compounds of the halogenated aromatic hydrocarbons (HAH)/polycyclic aromatic hydrocarbon (PAH) family (i.e. 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD)) have been thoroughly characterized for their toxic side effects, as we will discuss later [11–13]. Others classes of xenobiotics have been shown to activate AhR [8]. In addition, diet, in particular fruit and vegetable containing flavonoids and indole derivatives, represent also an important source of AhR ligands [3,8,9]. Intestinal microflora is also able to generate such ligands: by consuming tryptophan (Trp) coming from diet, they can generate a broad diversity of indole derivatives AhR agonists (i.e. indole-3-acetic acid, tryptamine) [9]. Trp can also be degraded by endogenous pathways, especially the kynurenine pathway, which is responsible for 95% of the Trp degradation [14]. The first step of this pathway consists in the catabolism of tryptophan into the kynurenine agonist by either IDO1, IDO2 or TDO2 enzyme. Kyn can be then converted into downstream metabolites such as kynurenic acid that have also been recently described as an AhR ligand [15**]. Even if AhR agonist are numerous, we still probably have not discovered all of them.

Sustained AhR activation by HAH or PAH, such as TCDD or PAHS, in mice and rats, leads to a broad spectrum of adverse effects, including cancer development and the impairment of endocrine functions and



Emerging roles of circular RNAs in liver cancer

Corentin Louis,^{1,†} Delphine Leclerc,^{1,†} Cédric Coulouarn^{1,*}

Summary

Hepatocellular carcinoma and cholangiocarcinoma are the most common primary liver tumours, whose incidence and associated mortality have increased over recent decades. Liver cancer is often diagnosed late when curative treatments are no longer an option. Characterising new molecular determinants of liver carcinogenesis is crucial for the development of innovative treatments and clinically relevant biomarkers. Recently, circular RNAs (circRNAs) emerged as promising regulatory molecules involved in cancer onset and progression. Mechanistically, circRNAs are mainly known for their ability to sponge and regulate the activity of microRNAs and RNA-binding proteins, although other functions are emerging (e.g. transcriptional and post-transcriptional regulation, protein scaffolding). In liver cancer, circRNAs have been shown to regulate tumour cell proliferation, migration, invasion and cell death resistance. Their roles in regulating angiogenesis, genome instability, immune surveillance and metabolic switching are emerging. Importantly, circRNAs are detected in body fluids. Due to their circular structure, circRNAs are often more stable than mRNAs or miRNAs and could therefore serve as promising biomarkers – quantifiable with high specificity and sensitivity through minimally invasive methods. This review focuses on the role and the clinical relevance of circRNAs in liver cancer, including the development of innovative biomarkers and therapeutic strategies.

© 2021 The Authors. Published by Elsevier B.V. on behalf of European Association for the Study of the Liver (EASL). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Circular RNAs (circRNAs) belong to a family of recently rediscovered RNA molecules. They are produced during the maturation of RNA transcripts. Structurally, circRNAs are covalently closed by a phosphodiester linkage between downstream donor and upstream acceptor RNA splice domains. Mainly considered as splicing errors during the past decade, circRNAs are now accepted as functional RNA molecules. They display tissue- and cell-specific expression patterns and are encoded from thousands of genes.¹ Emerging evidence demonstrates that circRNAs are involved in biological processes contributing to the onset and progression of cancer.^{2–4} In addition, due to their circular structure, circRNAs are resistant to the action of exoribonucleases and therefore exhibit an expanded half-life compared to their parental linear counterparts, allowing detection even when expressed at a low level.^{5,6} It is estimated that exonic circRNAs are very stable in cells, with most species exhibiting a half-life of over 48 hours,⁷ compared to an average half-life of 10 hours for mRNAs.⁶ These properties suggest that circRNAs could represent clinically relevant biomarkers for the management of patients with cancer. This review specifically highlights recent discoveries on the functions and clinical relevance of circRNAs in liver cancer, including hepatocellular carcinoma (HCC) and cholangiocarcinoma (CCA).

CircRNA discovery, biogenesis and function

CircRNA discovery

CircRNAs were first discovered in 1976 on viroid particles.⁸ Then, circular structures were identified by electron microscopy in the cytoplasm of eukaryotic cells, notably in HeLa cells.⁹ Nevertheless, it was only in 1993 that circRNAs (circSry, a 1.3 kb circRNA derived from a single exon of the sex-determining region Y transcript in mice) were molecularly identified by dedicated experiments, including RNase H digestion followed by northern blotting, reverse-transcription PCR and sequencing.¹⁰ Afterward, several circRNAs previously described as “scrambled exon RNAs” were reported, including a circRNA from the rat cytochrome P450 2C24 gene.¹¹ However, no specific biological function was reported for circRNAs at that time.

Next-generation sequencing (NGS) ushered in the beginning of the genome-wide identification of circRNAs. Traditional RNA-sequencing pipelines remove non-conventional reads that do not map with the reference genome, such as reads that arise from gene fusion or those originating from back-splicing events. New bioinformatic algorithms based on back-splice junction overlapping reads recognition made it possible to efficiently detect *de novo* circRNAs and to differentiate them from their

Keywords: circRNA; hepatocellular carcinoma; cholangiocarcinoma; cancer hallmarks; biomarker

Received 9 July 2021; received in revised form 16 November 2021; accepted 18 November 2021; available online 27 November 2021

¹Inserm, Univ Rennes 1, COSS (Chemistry, Oncogenesis Stress Signaling), UMR_S 1242, Centre de Lutte contre le Cancer Eugène Marquis, F-35042, Rennes, France

[†]These authors contributed equally to the work

* Corresponding author.
Address: Inserm UMR_S 1242, “Chemistry, Oncogenesis, stress, Signaling” (COSS), Centre de Lutte contre le Cancer Eugène Marquis, Rue de la Bataille Flandres Dunkerque, Bat D, F-35042 Rennes, France. Tel.: +33-2-2323-3881.
E-mail address: cedric.coulouarn@inserm.fr (C. Coulouarn).



Research Article



Non-canonical miRNA-RNA base-pairing impedes tumor suppressor activity of miR-16

Anaïs M Quéméner¹, Laura Bachelot^{1,*}, Marc Aubry^{2,*}, Stéphane Avner³, Delphine Leclerc^{2,4}, Gilles Salbert³, Florian Cabillic^{5,6}, Didier Decaudin^{7,8}, Bernard Mari⁹, Frédéric Mouriaux^{2,4}, Marie-Dominique Galibert^{1,10}, David Gilot^{1,2}

Uveal melanoma (UM), the most common primary intraocular tumor in adults, has been extensively characterized by omics technologies during the last 5 yr. Despite the discovery of gene signatures, the molecular actors driving cancer aggressiveness are not fully understood, and UM is still associated with very poor overall survival (OS) at the metastatic stage. By defining the miR-16 interactome, we revealed that miR-16 mainly interacts via non-canonical base-pairing to a subset of RNAs, promoting their expression levels. Consequently, the canonical miR-16 activity, involved in the RNA decay of oncogenes, such as *cyclin D3*, is impaired. This non-canonical base-pairing can explain both the derepression of miR-16 targets and the promotion of oncogene expression observed in patients with poor OS in two cohorts. miR-16 activity, assessment using our RNA signature, discriminates the patient's OS as effectively as current methods. To the best of our knowledge, this is the first time that a predictive signature has been composed of genes belonging to the same mechanism (miR-16) in UM. Altogether, our results strongly suggest that UM is a miR-16 disease.

DOI 10.26508/lsa.202201643 | Received 2 August 2022 | Revised 13 September 2022 | Accepted 15 September 2022 | Published online 6 October 2022

Introduction

Uveal melanoma (UM) is the most common primary intraocular tumor in adults. No effective treatment is currently able to counteract UM metastasis (Jager et al, 2020). UM carries mutually exclusive mutations that trigger overactivity of the Gαq pathway (G protein subunit alpha q [GNAQ], G protein subunit alpha 11 [GNAI1], Phospholipase C Beta 4 [PLCB4], or Cysteinyl Leukotriene Receptor 2 [CYSLTR2]) (Robertson et al, 2018).

UM is considered to be a G protein-coupled receptor disease. However, additional genetic events occur, such as BES alterations, BRCA1-associated protein 1 (BAP1), Eukaryotic Translation Initiation Factor 1A X-Linked (EIF1AX) and Splicing Factor 3b Subunit 1 (SF3B1), and recurrent copy number variations (CNVs). UM is usually a diploid tumor with recurrent CNV of whole chromosomes or arms. Among these genome structural variations, monosomy 3 is the most frequent (~50% of cases). 1p loss, 1q gain, 6p gain, 6q loss, 8 gain, 8p loss, and 8q gain complete the UM genomic landscape. Monosomy 3 is clearly associated with a high risk of metastasis (Horsman et al, 1990; Bagger et al, 2014; Robertson et al, 2018; Shain et al, 2019). Although the loss of chromosome 3 induces the loss of *BAP1* (3p21.1), the role of the other genes located on chromosome 3 in tumor aggressiveness is not excluded. Gain of 8q is also common in UM (~50% of cases). It has been suggested that genes on chromosome 8q, such as *MYC*, and *POU Class 5 Homeobox 1 (POU5F1)*, and *Protein Tyrosine Phosphatase 4A3 (PTP4A3)* may explain the poor overall survival (OS) of UM patients (Meir et al, 2007; Durante et al, 2020; Pandiani et al, 2021). *PTP4A3* has been shown to promote migration and invasiveness in an UM cell line, and high *PTP4A3* mRNA levels correlate with poor OSs of patients. It is important to note that the high level of *PTP4A3* mRNA (up-regulated in 66% of UM) is not merely a consequence of 8q gain (Laurent et al, 2011; Duciel et al, 2019). To date, the molecular mechanism promoting high expression levels of *PTP4A3* and *MYC* remains unsolved despite the deleterious consequences.

Although UM is considered to be a G protein-coupled receptor disease with BES alterations and CNVs, at least three other cellular processes seem to be frequently deregulated. These include the YAP pathway which is activated via Hippo-independent activation, splicing activity and translation initiation as a result of alterations in *BAP1*, Serine And Arginine Rich Splicing Factor 2 (*SRSF2*), RNA Binding Motif Protein 10 (*RBM10*) *EIF1AX*, and other genes

¹University of Rennes, Centre National de la Recherche Scientifique (CNRS), Institut de Génétique et Développement de Rennes (IGDR) - UMR 6290, Rennes, France
²INSERM U1242, University of Rennes, Rennes, France ³SPARTE, University of Rennes, CNRS, IGDR - UMR 6290, Rennes, France ⁴Service d'Ophtalmologie, CHU de Rennes, Rennes, France ⁵INSERM U1241, Université Rennes, INRAE, Institut NuMeCan (Nutrition, Metabolisms and Cancer), Rennes, France ⁶Laboratoire de Cytogénétique et Biologie Cellulaire, CHU Rennes, Rennes, France ⁷Laboratory of Preclinical Investigation, Translational Research Department, Institut Curie, PSL Research University, Paris, France ⁸Curie, Department of Medical Oncology, PSL Research University, Paris, France ⁹Fédération Hospitalo Universitaire-OncoAge, CNRS, Institut de Pharmacologie Moléculaire et Cellulaire, Université Côte d'Azur, Valbonne, France ¹⁰CHU Rennes, Service de Génétique Moléculaire et Génomique, Rennes, France

Correspondence: david.gilot@univ-rennes1.fr

*Laura Bachelot and Marc Aubry contributed equally to this work.

ANNEXE III : Formations suivies



ATTESTATION DES FORMATIONS SUIVIES

Mardi 25 avril 2023

LECLERC Delphine

Doctorant(e) à

- **Etablissement** : Université de Rennes
- **Ecole Doctorale** : École doctorale Sciences de la vie et de la santé

Total du nombre d'heures comptabilisées : 82:00

Total autres modes de comptabilisations éventuels : 2.0 Jours

a suivi les formations ci-dessous :

Catégorie :

Nombre d'heures comptabilisées dans la catégorie : 13:00

Autres mode de comptabilisation éventuel :

Sous-catégorie de formation	Intitulé de la formation ou équivalence	Intitulé Session ou demande d'équivalence	Année civile	Comptabilisation (heures et autres)	Type
	Biostatistiques I : Introduction à la biostatistique	DIS-BS-X-320	2021	13:00	Formation

Catégorie : Éthique et intégrité scientifique

Nombre d'heures comptabilisées dans la catégorie : 11:00

Autres mode de comptabilisation éventuel :

Sous-catégorie de formation	Intitulé de la formation ou équivalence	Intitulé Session ou demande d'équivalence	Année civile	Comptabilisation (heures et autres)	Type
	Expérimentation humaine et animale	Session 2021	2021	02:00	Formation
	Ethique de la recherche et intégrité scientifique dans les relations humaines au travail en VISIO	Session du 9 juin 2021 de 10h00 à 12h00	2021	02:00	Formation
	AXE 2 ÉTHIQUE DE LA RECHERCHE ET IS DANS LES DONNEES. 2.7 MALTRAITANCE DES DONNEES DE LA	Session 1	2021	03:00	Formation

Collège doctoral de Bretagne

college-doctoral@doctorat-bretagne.fr

	FRAUDE A L'ERREUR HONNETE				
	Introduction à l'éthique de la recherche et à l'intégrité scientifique	Session du 1er décembre 2020	2020	04:00	Formation

Catégorie : Maîtriser l'information scientifique et technique : outils et méthodes

Nombre d'heures comptabilisées dans la catégorie : 13:30

Autres mode de comptabilisation éventuel :

Sous-catégorie de formation	Intitulé de la formation ou équivalence	Intitulé Session ou demande d'équivalence	Année civile	Comptabilisation (heures et autres)	Type
	Trouver et réutiliser ses données dans le respect du droit	Session du 1er avril 2021	2021	02:00	Formation
	Testez vos connaissances en information scientifique et technique	Testez vos connaissances en information scientifique et technique - 10 juin 2022 (en visio)	2022	02:00	Formation
	How much do you know about scientific information?	How much do you know about scientific information? - 23 mai 2022	2022	02:00	Formation
	Présenter et structurer sa thèse en vue du dépôt électronique en ALL et SHS	Présenter et structurer sa thèse en vue du dépôt électronique en ALL et SHS - 15 mars 2022 (en visio)	2022	02:00	Formation
Publications scientifiques	How to make your research more visible and accessible : publication strategy and ethical aspects	How to make your research more visible and accessible : publication strategy and ethical aspects - 9 mai 2022 - Site: Rennes (Beaulieu)	2022	02:30	Formation
	2.5 LA PROTECTION DES DONNEES A CARACTERE PERSONNEL DANS LE CADRE D'UNE RECHERCHE SCIENTIFIQUE (AXE 2 ÉTHIQUE DE LA RECHERCHE ET INTEGRITE	Session 2020/2021	2021	03:00	Formation

Collège doctoral de Bretagne

college-doctoral@doctorat-bretagne.fr

	SCIENTIFIQUE DANS LES DONNEES)				
--	--------------------------------------	--	--	--	--

Catégorie : Management et communication

Nombre d'heures comptabilisées dans la catégorie : 24:30

Autres mode de comptabilisation éventuel : 2.0 Jours

Sous-catégorie de formation	Intitulé de la formation ou équivalence	Intitulé Session ou demande d'équivalence	Année civile	Comptabilisation (heures et autres)	Type
	Les fondamentaux du management et de la communication	Session 2022	2022	06:00 2.0 Jours	Formation
	Introduction au management des organisations	Session 2022	2022	18:30	Formation

Catégorie : Pédagogie et pratique enseignante

Nombre d'heures comptabilisées dans la catégorie : 20:00

Autres mode de comptabilisation éventuel :

Sous-catégorie de formation	Intitulé de la formation ou équivalence	Intitulé Session ou demande d'équivalence	Année civile	Comptabilisation (heures et autres)	Type
Médiation scientifique	Ma thèse en 180 secondes	2022-2023 - Rennes - Groupe 2	2023	20:00	Formation

ATTESTATION INDIVIDUELLE D'ASSIDUITE EN FORMATION

(L. 6353-1 du code du travail)

Je soussignée Laurence DEFLESSELLE, Directrice Générale représentante légale d'ONIRIS – l'Ecole Nationale Vétérinaire, Agroalimentaire et de l'Alimentation,

Atteste que **Delphine LECLERC**

Inscrit à la formation suivante : **Expérimentation Animale - Fonction de conception ou de réalisation des procédures expérimentales – Module Rongeurs/Lapins**

Référence Oniris : **C21/EAS3**

Date de début : **13/09/2021** Date de fin : **08/10/2021**

D'une durée totale en heures : **57**

Conformément au programme pédagogique préétabli.

→ **A suivi la séquence en FOAD**

D'une durée totale en heures : **15**

Suivi de la réalisation des travaux exigés et des évaluations :

Nature des travaux	Durée estimée en heure	Appréciation
Accès aux ressources pédagogiques documentaires et vidéos avant de répondre au questionnaire d'évaluation	15	A réussi au contrôle des connaissances correspondant

Oniris s'engage à conserver sur tout support approprié et pour une durée d'au moins 5 années à compter de la fin de la formation l'ensemble des pièces justificatives permettant de démontrer la réalité de l'action de FOAD et notamment les justificatifs afférents aux travaux réalisés ainsi que les évaluations y afférentes.

→ **A suivi la séquence à Distance (classe virtuelle)**

D'une durée totale en heures : **32**

A réussi au contrôle des connaissances correspondant.

→ **A suivi la séquence en présentiel**

D'une durée totale en heures : **10**

A réussi au contrôle des connaissances correspondant.



Nantes, le 18 Octobre 2021
Pour la Directrice Générale
Par délégation

Dans votre intérêt, conservez cette attestation.

SANTÉ ET ALIMENTATION AU CŒUR DE LA VIE

ANNEXE IV : Missions enseignement



AVENANT N°01 AU CONTRAT DE RECRUTEMENT EN QUALITÉ DE DOCTORANT CONTRACTUEL

Vu le code de la recherche et notamment son article L412-2 ;

Vu le code de la propriété intellectuelle, et notamment son article R611-14 ;

Vu la loi n°84-16 du 11 janvier 1984 modifiée, portant dispositions statutaires relatives à la fonction publique de l'Etat ;

Vu le décret n°84-431 du 6 juin 1984 modifié fixant les dispositions statutaires communes applicables aux enseignants-chercheurs et portant statut particulier du corps des professeurs des universités et du corps des maîtres de conférences ;

Vu le décret n°86-83 du 17 janvier 1986 modifié, relatif aux dispositions générales applicables aux agents non titulaires de l'Etat ;

Vu le décret n°2000-815 du 25 août 2000 modifié relatif à l'aménagement et à la réduction du temps de travail dans la fonction publique de l'État et dans la Magistrature ;

Vu le décret n°2009-464 du 23 avril 2009 modifié relatif aux doctorants contractuels des établissements publics d'enseignement supérieur ou de recherche ;

Vu l'arrêté du 29 août 2016 fixant le montant de la rémunération du doctorant contractuel ;

Vu l'arrêté du 25 mai 2016 fixant le cadre national de la formation et les modalités conduisant à la délivrance du diplôme national du doctorat ;

Vu la proposition de la direction de l'école doctorale concernée, ayant recueilli les avis favorables du directeur ou de la directrice de thèse et de la direction de l'unité ou de l'équipe de recherche concernée ;

Vu le contrat d'engagement signé entre Madame DELPHINE LECLERC et l'Université de Rennes 1 à compter du 01/10/2020 ;

Entre les soussignés,

Le Président de l'Université de Rennes 1, établissement public à caractère scientifique, culturel et professionnel, d'une part,

ET

Madame DELPHINE LECLERC,
née le 10/06/1997, LONGJUMEAU, Essonne, France
désigné ci-dessous le doctorant contractuel,
d'autre part,

Il a été convenu ce qui suit :

Article 1er : OBJET:

Le présent avenant a pour objet de compléter les missions et la rémunération, initialement prévues dans le contrat doctoral signé entre les parties.

Article 2 : SERVICE CONFIE AU DOCTORANT:

Conformément à l'article 4-bis du contrat précité, le doctorant contractuel pourra accomplir, pendant la durée de son contrat, un service annuel qui comprendra, outre les activités de recherche liées à la préparation de son doctorat, une ou des activités complémentaires parmi celles listées ci-dessous.

- Une mission d'enseignement y compris dans le domaine de la formation continue, pour un service annuel au plus égal au tiers du service annuel d'enseignement de référence des enseignants-chercheurs, défini à l'article 7 du décret du 6 juin 1984 susvisé soit, un maximum annuel de 64h équivalent TD ou TP ;

- Une mission dans le domaine de la diffusion de l'information scientifique et technique et de la valorisation des résultats de la recherche pour une durée annuelle maximale de 32 jours ;

- Une mission d'expertise effectuée dans une entreprise, une collectivité territoriale, une administration, un établissement public, une association ou une fondation pour une durée annuelle maximale de 32 jours.

La durée totale des activités complémentaires aux activités de recherche confiées au doctorant dans le cadre du contrat doctoral ne peut excéder un sixième de la durée annuelle de travail effectif fixée par le décret du 25 août 2000 susvisé.

Article 3 : DESCRIPTION DE LA MISSION:

Unité de recherche du contrat: UMR_S Oncogenesis Stress Signaling - UMR_S 1242

Mission : Enseignement

Objet de la mission : Enseignement en Biologie cellulaire au sein de la composante d'enseignement Sciences de la Vie et de l'Environnement

Nombre d'heures équivalent TD prévu par le présent avenant : 30h éq TD (Taux horaire: 42.86€)

Période : 01/10/2022 au 31/08/2023

Article 4 : RÉMUNÉRATION DE L'ACTIVITÉ COMPLÉMENTAIRE

Rémunération complémentaire mensuelle brute prévue au titre de la présente mission complémentaire: 116.90€.

Cette rémunération est indexée sur l'évolution des rémunérations de la fonction publique.

Article 5 : DURÉE DE L'AVENANT

Le présent avenant prend effet le 01/10/2022 jusqu'au 31/08/2023.

Article 6 : Toute modification éventuelle des dispositions de ce présent avenant fera l'objet d'un avenant complémentaire.

Fait à Rennes, le 10 octobre 2022

L'intéressé(e)

**La Directrice
des Ressources humaines
de l'Université de Rennes 1**

Anne Glidic

DELPHINE LECLERC

**Imputation:
Centre de coût : 912SGEN
Destination : D101**

Unité INSERM U1242 Oncogenesis Stress Signaling

Dr Cédric Ménard

Maître de Conférence des Universités-Praticien Hospitalier en Immunologie
Faculté de Médecine de Rennes (Université Rennes 1)

cedric.menard@chu-rennes.fr

Objet : Attestation d'enseignement

Je soussigné Docteur Cédric MENARD, coordonnateur et responsable des unités d'enseignement de Master 2 « Technologies d'Analyse du Vivant » (TAV) et « Biotechnologies en recherche diagnostique et thérapeutique » (BDT) de la mention « Biologie Santé » du Master « Science, Technologies, Santé » porté par la Faculté de Médecine (Université de Rennes I), atteste que Mademoiselle Delphine LECLERC a bien assuré au cours de l'année universitaire 2022-2023 dans le cadre de l'enseignement de ce Master:

- une séance d'encadrement de Travaux Dirigés (TD) de 3h dans l'UE TAV ;
 - une séance de TD de 3h consacrée à la lecture critique d'article dans l'UE BDT ;
 - un cours magistral de 2 heures intitulé « Potentiels et succès thérapeutiques de l'approche CRISPR » (soit 3 heures équivalent TD) ;
- soit un total de 9h de cours équivalent TD dans notre mention de Master 2.

Fait à la demande de l'intéressé, pour faire valoir ce que de droit.

Fait à Rennes, le 22 novembre 2022

Dr Cédric Ménard



Faculté de Médecine
2 avenue du Prof. Léon Bernard
CS 34317
35043 RENNES CEDEX

TÉL. 33 (0)2 23 23 49 55
FAX 33 (0)2 23 23 49 58
u1236-micmac@univ-rennes1.fr

ANNEXE V : Congrès



Séminaire de Génétique de l'Ouest

9 & 10 Septembre 2021

Domaine de Port aux Rocs
Le Croisic

Attestation de présence

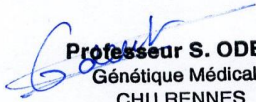
Je soussignée le Pr Sylvie Odent, certifie que : **Mme Delphine LECLERC**

a participé au Séminaire de Génétique de l'Ouest ayant pour thème : **Neurogénétique et communications libres**

Date : 10 Septembre 2021 – 1 jour

Durée de la formation : 03h15

Au Croisic, le 10 Septembre 2021



Professeur S. ODENT
Génétique Médicale
CHU RENNES

Séminaire de Génétique de l'Ouest 2021

COMITE D'ORGANISATION

Pr Sylvie ODENT – CHU de Rennes

Centre de Référence maladies Rares Anomalies du développement CLAD-Ouest

Pôle médico-chirurgical de Pédiatrie et Génétique Clinique CHU de Rennes.

Coordinatrice du réseau GEM-EXCELL - réseau d'excellence de génétique et de génomique du Grand-Ouest

Amandine CHARRETON

Cheffe de projet réseau GEM-EXCELL

Florence PIQUET

Institut de la Mère et de L'enfant - Rennes

11^{èmes}

ASSISES DE GÉNÉTIQUE HUMAINE ET MÉDICALE COUVENT DES JACOBINS, RENNES 1-4 FÉVRIER 2022



www.assises-genetique.org

RENNES



Delphine Leclerc
Centre Eugène Marquis
INSERM U1242
INSERM U1242, OSS
CLCC Eugène Marquis
Rue de la Bataille Flandres Dunkerque
35042 Rennes
France

ATTESTATION DE PRESENCE

La société MCO Congrès atteste que :

Delphine Leclerc

a participé au Congrès : **11èmes Assises de Génétique Humaine et Médicale**

qui s'est déroulé du 01 février 2022 au 04 février 2022

à Rennes - le Couvent des Jacobins - Centre de congrès de Rennes Metropole

Fait à Rennes, le 07 février 2022

11èmes Assises de Génétique Humaine et Médicale

SAS MCO CONGRES

Villa Gaby, 285 Corniche J.F Kennedy
13007 Marseille
Tel. : 04 95 09 38 00 - Fax : 04 95 09 38 01
Siret : 383 730 181 00054 - APE 8230Z
TVA Intra. FR84 383 730 181.

MCO Congrès - SAS au capital de 37 000 euros
SIRET 383 730 181 00054 Marseille - APE 8230Z - TVA INTRACOMMUNAUTAIRE FR84 383 730 181
RCS 383 730 181 RCS Marseille



**CRISPR and
Translational Medicine
congress**

This is to certify that

Delphine LECLERC, PhD student
"Chemistry, Oncogenesis, Stress, Signaling" (COSS)
Inserm U1242
Universite de Rennes 1

has participated at the CRISPR and Translational
Medicine Congress, in Bordeaux
from 31 mars / 1^{er} avril 2022

Congrès CRISPR and Translational Medecine congress
Université de Bordeaux
31 mars/1^{er} avril 2022 Bordeaux (France)

ANNEXE VI : Récompenses



Clichy, le 20 juillet 2022

Envoyé par courrier électronique

Objet : Prix Jeunes Talents France 2022 L'Oréal-UNESCO *Pour les Femmes et la Science*

Chère Madame,

Depuis 24 ans, le programme L'Oréal-UNESCO *Pour les Femmes et la Science* révèle et distingue des femmes scientifiques de tous les continents à différents moments de leur carrière.

Le programme français des Jeunes Talents L'Oréal-UNESCO *Pour les Femmes et la Science* (ci-après le programme) a pour but d'identifier et de récompenser de jeunes chercheuses talentueuses dans les sciences de la vie et de l'environnement, sciences de la matière, mathématiques, informatique et sciences de l'information, sciences de l'ingénieur et technologie, pour la qualité de leurs travaux de recherche, et les encourager à poursuivre une brillante carrière scientifique.

Les dotations sont remises par la Fondation L'Oréal et s'inscrivent dans le cadre d'un partenariat avec la Commission nationale française pour l'UNESCO et l'Académie des sciences. Depuis leur création en 2007, 335 prix Jeunes Talents France ont été remis.

En 2022, vingt-trois doctorantes et douze post-doctorantes ont été sélectionnées.

Par la présente, nous avons le très grand plaisir de vous confirmer que **vous avez été retenue pour faire partie des Jeunes Talents France 2022**, au regard des critères de la sélection du Prix conditionnant l'attribution des dotations.

Nous vous adressons toutes nos plus vives félicitations pour cette récompense.

Pour la bonne forme, nous vous remercions de bien vouloir nous retourner par email, sous huitaine, l'exemplaire des présentes dûment signé, manifestant ainsi votre accord.

Nous vous prions, Chère Madame, d'agréer l'expression de nos salutations distinguées.

Fait à Clichy, le 20 juillet 2022

Madame

Pour la Fondation L'OREAL

Elisa Simonpietri, Pharm. D

Directrice internationale des programmes Science

Secrétaire Exécutive des prix L'Oréal-UNESCO

Pour les Femmes et la Science

Fondation L'Oréal

Signature :

Signature :



Collège doctoral de Bretagne

Rennes, le 26 avril 2023

Finale bretonne Ma thèse en 180s

Attestation

Je soussigné, Valérie Le Cann, responsable administrative adjointe du Collège doctoral de Bretagne, certifie que :

**Madame Delphine LECLERC (doctorante)
Oncogenesis Stress Signalling (OSS-INSERM)
Université de Rennes**

A obtenu le prix du jury de la finale régionale de Ma thèse en 180s à Rennes

Le 14 mars 2023 à l'Espace des sciences aux Champs libres

Fait pour servir et faire valoir ce que de droit,

Valérie Le Cann

Titre : Évaluation des outils d'édition du génome pour les applications biomédicales : du diagnostic au traitement

Mots clés : génétique, édition du génome, CRISPR/Cas9, thérapie génique

Résumé : Ces dernières années, de considérables progrès ont été obtenus en médecine génomique. L'apport de la technologie de séquençage haut-débit dans les laboratoires de génétique médicale a notamment permis de mieux diagnostiquer les pathologies génétiques rares. Malgré cette avancée, un nombre important de patients diagnostiqués se retrouvent en errance thérapeutique. Environ 250 millions de personnes souffrent de pathologies génétiques rares et moins de 5% de ces pathologies disposent aujourd'hui de traitements efficaces. La découverte et l'amélioration constante des outils CRISPR/Cas9 ont ouvert la porte à une nouvelle médecine, la chirurgie du génome. Dans ce contexte, une partie importante des altérations génétiques

responsables de maladies rares peuvent être *en théorie* directement corrigées par cette approche, ouvrant une perspective thérapeutique inespérée pour des patients étant parfois les seuls porteurs de ces altérations génétiques. Au cours de ma thèse, j'ai l'opportunité de développer et d'évaluer les outils d'édition du génome CRISPR/Cas9, d'une part pour concevoir et mesurer l'efficacité de nouveaux traitements *in vitro*, et d'autre part pour installer et décrypter le rôle biologique de variants génétiques de signification inconnue. Désormais, il est possible de corriger ou créer ces variants en recherche fondamentale afin d'aider les cliniciens à conclure sur leur rôle biologique ouvrant la porte à des thérapies personnalisées dans un futur proche.

Title: Evaluation of genome editing tools for biomedical applications: from diagnosis to treatment

Keywords: genetics, gene editing, CRISPR/Cas9, gene therapy

Abstract: Recently, considerable progress has been made in genomic medicine. The use of high-throughput sequencing technology in medical genetics laboratories has made it possible to diagnose rare genetic pathologies. Conversely, a significant number of diagnosed patients are in treatment wandering. No less than 250 million people suffer from rare genetic diseases and less than 5% of these diseases have effective treatments today. Thanks to the emergence and constant improvement of CRISPR/Cas9 tools, medicine has entered a new therapeutic era of genome surgery. In this context, an important part of the genetic alterations responsible for rare

diseases can, *in theory*, be directly corrected by this approach, opening an unexpected therapeutic perspective for patients being sometimes the only carriers of these genetic alterations. During my thesis, I had the opportunity to develop CRISPR/Cas9 gene editing tools, first to design and measure the efficiency of new treatments *in vitro*, and second to install and decipher the biological role of variants of unknown significance. It is now possible to correct or create these variants in basic research in order to help clinicians conclude their biological role, opening the door to personalized therapies in the near future.