

DOMANDA DI INVENZIONE NUMERO	102023000007968
Data Deposito	21/04/2023
Data Pubblicazione	21/10/2024

Classifiche IPC

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
C	12	N	15	86

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
C	12	N	15	90

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
C	12	N	15	11

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
C	12	N	15	85

Titolo

Metodi di editing genomico e costrutti
--

Metodi di editing genomico e costrutti

SETTORE TECNOLOGICO

La presente invenzione riguarda metodi di editing genomico, in particolare si riferisce a un sistema che comprende un acido nucleico donatore comprendente:

- 5 - una sequenza di segnale di degradazione,
- un sito di taglio enzimatico,
- una sequenza di salto ("skipping") ribosomiale,
- una sequenza di DNA esogeno,

10 in cui detto acido nucleico donatore è affiancato all'estremità 5' e 3' da sequenze bersaglio invertite;

 e, opzionalmente, un oligonucleotide complementare alla sequenza bersaglio e/o una nucleasi che riconosce la sequenza bersaglio. L'invenzione si riferisce anche a un metodo di integrazione di una sequenza di DNA esogeno nel genoma di una cellula che comprende il contatto della cellula con l'acido nucleico donatore, un oligonucleotide complementare alla sequenza
15 bersaglio e una nucleasi che riconosce la sequenza bersaglio. L'invenzione riguarda anche i vettori che comprendono il suddetto acido nucleico donatore e/o l'oligonucleotide complementare alla sequenza bersaglio e/o la nucleasi e i suoi relativi usi medici.

STATO DELLA TECNICA

La terapia genica con vettori virali adeno-associati (AAV) è molto promettente per fornire
20 l'espressione a lungo termine di transgeni terapeutici dopo una singola somministrazione. Tuttavia, alcune delle sfide ancora aperte comprendono il contrasto delle mutazioni guadagno di funzione o dell'effetto dominante negativo, che non beneficiano della tradizionale terapia di sostituzione genica. Per superare queste limitazioni, negli ultimi anni l'editing genomico è emerso come un'opzione valida per il trattamento di malattie ereditate con modalità
25 dominante, tra cui le malattie della retina (IRD)(1) . Questo approccio si basa sull'uso di una nucleasi identificata nel sistema batterico, di solito CRISPR/Cas9. Cas9 è una ribonucleoproteina che utilizza una breve sequenza di RNA guida (gRNA) per riconoscere il DNA bersaglio mediante la complementarità delle basi di Watson-Crick. Questa sequenza di DNA bersaglio deve essere adiacente ad una sequenza del motivo adiacente al protospaziatore (PAM) affinché Cas9 si leghi
30 e tagli la sequenza bersaglio del DNA(2). Questo può bloccare la produzione della proteina tossica senza intaccare la copia corretta del gene. Questo approccio è particolarmente utile per il trattamento delle forme dominanti di IRD, come la retinite pigmentosa autosomica dominante

(adRP)(3). La retinite pigmentosa (RP) colpisce 1/3.000 pazienti in tutto il mondo e il 30-40% dei casi ha un'ereditarietà autosomica dominante (AD) (4). Il gene della rodopsina (*RHO*) è quello più comunemente mutato nei pazienti con RP AD (RP4), con la mutazione P23H che è la più comune negli Stati Uniti (5). P23H di *RHO* esercita un effetto di guadagno di funzione tossica, che causa la progressiva degenerazione della retina e la perdita della vista. Per superare gli effetti tossici della *RHO* mal ripiegata, è necessario distruggere l'allele P23H mutante. Pertanto, abbiamo sviluppato una strategia di editing genomico per colpire una forma autosomica dominante di retinite pigmentosa dovuta a una mutazione della *RHO* (rodopsina) P23H prevalente, basata sulla strategia di integrazione mirata omologia-indipendente (HITI) recentemente descritta (6,7). In questo approccio, il sistema CRISPR/Cas9 genera le rotture a doppio filamento (double strand breaks, DB) in un sito specifico del locus guidato da una specifica sequenza di gRNA; le DB risultanti saranno risolte principalmente mediante la via di riparo cellulare che unisce le estremità non omologhe (non-homologous end-joining NHEJ), che è il meccanismo di riparazione predominante nelle cellule terminalmente differenziate come i fotorecettori e, in generale, è attivo in tutte le fasi del ciclo cellulare. HITI sfrutta la via NHEJ per integrare una sequenza esogena (DNA donatore di HITI affiancato da siti bersaglio di gRNA invertiti) in un locus specifico a livello dei DB. Dopo l'integrazione del DNA donatore nell'orientamento desiderato, si verifica la corretta espressione del gene terapeutico dal promotore endogeno.

Inoltre, l'inserimento mediato da HITI di una copia terapeutica del gene di tipo selvatico ("wild-type") ha il potenziale di essere terapeutico indipendentemente dalla specifica mutazione che causa la malattia e potrebbe essere utilizzato per il trattamento di malattie ereditate in modo dominante, sostituendo sia l'allele mutante che quello di tipo selvatico con una copia corretta del gene fornita dal DNA donatore. Ciò eviterebbe le restrizioni della sequenza bersaglio imposte dalla specificità allelica del silenziamento e amplierebbe l'applicabilità della terapia a tutte le mutazioni nello stesso gene. Recentemente abbiamo utilizzato l'HITI per ottenere il silenziamento mirato di *RHO*, indipendentemente dalle mutazioni, seguito dalla sostituzione con una copia sana di *RHO* nel locus *RHO* del topo (7).

Un precedente approccio per l'integrazione di una sequenza di DNA esogeno nel genoma di una cellula basato su HITI è descritto in WO2020079033, qui allegato come riferimento.

Tuttavia, è ancora necessario migliorare la tecnologia HITI.

SOMMARIO DELL'INVENZIONE

Qui gli inventori hanno trovato un approccio HITI sorprendentemente più efficiente che consente la degradazione della proteina tossica e l'espressione della proteina di tipo selvatico. I presenti risultati dimostrano l'efficacia dell'HITI come strategia terapeutica per l'AD RP dovuta a mutazioni RHO in un modello murino umanizzato (8) e, pertanto, potrebbero essere prontamente traslati in un contesto umano.

Pertanto, gli inventori propongono l'approccio HITI come strategia per valutare il potenziale terapeutico in particolare nel locus RHO umano.

È stato valutato un nuovo costrutto per HITI che trasporta una sequenza accettore di splicing per uno splicing efficiente nel sito bersaglio del locus RHO (al posto dei 3 codoni di STOP) seguito da un segnale di degradazione CL1 (9,10) fuso con un sito di taglio della furina attivo per una maggiore degradazione della proteina RHO troncata (11). L'inclusione del segnale di degradazione CL1 promuove una degradazione selettiva delle proteine tronche all'estremità 5' senza influenzare la produzione di proteine di lunghezza completa. CL1 è inoltre fuso con P2A, una sequenza di salto ribosomiale, che favorisce la traduzione della sequenza codificante di RHO. Gli inventori hanno quindi valutato l'efficienza dell'HITI di questo nuovo costrutto in cellule e in topi hRHO-P23H-TagRFP (8) e hanno trovato che, sorprendentemente, i livelli di trascritti di hRHO erano circa 2 volte superiori nelle cellule trasfettate con il donatore di HITI ottimizzato rispetto alle cellule trasfettate con un donatore di HITI precedente, come noto dallo stato dell'arte. Il presente sistema di editing genetico utilizzato nei topi hRHO-P23H-TagRFP ha permesso di migliorare l'efficienza HITI fino al $12 \pm 8\%$ nell'area trasdotta.

DESCRIZIONE DETTAGLIATA DELL'INVENZIONE

Pertanto, è un oggetto dell'invenzione un sistema di editing genetico comprendente:

a) un acido nucleico donatore comprendente:

- una sequenza di segnale di degradazione,
- un sito di taglio enzimatico,
- una sequenza di salto ribosomiale,
- una sequenza di DNA esogeno,

in cui detto acido nucleico donatore è affiancato alle estremità 5' e 3' da sequenze bersaglio invertite; e opzionalmente

b) un oligonucleotide complementare alla sequenza bersaglio (qui anche definito come oligonucleotide complementare) e/o

c) una nucleasi che riconosce la sequenza bersaglio.

È anche un oggetto dell'invenzione un sistema di editing genetico comprendente:

a) un acido nucleico donatore comprendente:

- una sequenza di segnale di degradazione,
- un sito di taglio enzimatico,
- una sequenza di salto ribosomiale,
- una sequenza di DNA esogeno,

in cui detto acido nucleico donatore è affiancato alle estremità 5' e 3' da sequenze bersaglio invertite;

b) un oligonucleotide complementare alla sequenza bersaglio e

c) una nucleasi che riconosce la sequenza bersaglio.

Nel contesto della presente invenzione, l'acido nucleico donatore di preferenza comprende ulteriormente una sequenza accettore di splicing, preferibilmente all'estremità 5' della sequenza di segnale di degradazione.

Pertanto, la presente invenzione fornisce anche un sistema di editing genetico comprendente:

a) un acido nucleico donatore comprendente:

- una sequenza accettore di splicing,
- una sequenza di segnale di degradazione,
- un sito di taglio enzimatico,
- una sequenza di salto ribosomiale,
- una sequenza di DNA esogeno,

in cui detto acido nucleico donatore è affiancato alle estremità 5' e 3' da sequenze bersaglio invertite; e opzionalmente

b) un oligonucleotide complementare alla sequenza bersaglio e/o

c) una nucleasi che riconosce la sequenza bersaglio.

Il sistema di editing genetico dell'invenzione comprende preferibilmente:

a) un acido nucleico donatore comprendente:

- una sequenza accettore di splicing,
- una sequenza di segnale di degradazione,
- un sito di taglio enzimatico,
- una sequenza di salto ribosomiale,
- una sequenza di DNA esogeno,

in cui detto acido nucleico donatore è affiancato alle estremità 5' e 3' da sequenze bersaglio invertite;

b) un oligonucleotide complementare alla sequenza bersaglio e

c) una nucleasi che riconosce la sequenza bersaglio.

- 5 Preferibilmente, la sequenza del segnale di degradazione è la sequenza: CL1, CL2, CL6, CL9, CL10, CL11, CL12, CL15, CL16, SL17, SMN, CIITA, ODc7, ecDHFR, Mini ecDHFR.

Preferibilmente, la sequenza segnale di degradazione si trova in posizione C-terminale e/o destabilizza la sequenza endogena e la indirizza alla degradazione. Preferibilmente, il sito di taglio enzimatico è selezionato dal gruppo costituito da un sito di taglio della furina, un sito di taglio della serina proteasi, un sito di taglio della cisteina proteasi, un sito di taglio della proteasi aspartica, un sito di taglio della metalloproteasi e un sito di taglio della treonina proteasi, e/o questo è attivo e/o ottimizzato.

Preferibilmente, il sito di taglio enzimatico è un sito di taglio della furina, preferibilmente attivo e/o ottimizzato.

- 15 Preferibilmente, la sequenza di salto ribosomiale è una sequenza di salto ribosomiale del Teschovirus-1 2A suino (P2A) o una sequenza di salto ribosomiale del virus Thosera asigna 2A (T2A) o una sequenza E2A o F2A, preferibilmente la sequenza P2A.

Preferibilmente, la sequenza accettore di splicing può comprendere la sequenza nucleotidica (Y)_nNYAG.

- 20 Preferibilmente, la sequenza bersaglio è una sequenza compresa nel gene della rodopsina (Rho), più preferibilmente detto gene Rho presenta una o più mutazioni, come la/le mutazione/i che causano la retinite pigmentosa 4 (RP4 (vedi RHO; OMIM: 180380)), o la retinite pigmentosa 63 (RP63 (vedi OMIM: 614494)).

In alternativa, la sequenza bersaglio è una sequenza compresa in un gene che è mutato in

25 CORD1 (distrofia 1 dei coni-bastoncelli (vedi OMIM: 600624), CORD17 (distrofia 17 dei coni-bastoncelli (vedi OMIM: 615163)), BEST1 (bestrofina-1; malattia di Best; distrofia maculare vitelliforme proteina 2 (vedi OMIM: 607854)), OPA1 (OPA1 GTPasi simile alla dinamina mitocondriale (vedi OMIM: 605290)) o in qualsiasi altro gene mutato in condizioni autosomiche dominanti.

- 30 Preferibilmente, la sequenza bersaglio è compresa in un introne o in un esone del gene, preferibilmente nel primo introne o esone del gene.

Preferibilmente, la sequenza bersaglio è compresa all'interno del:

- primo introne del gene RHO, preferibilmente di origine umana, murina o suina, oppure
- primo esone del gene RHO, preferibilmente umano, di topo o di maiale.

Quando la sequenza bersaglio è compresa all'interno di un esone del gene, la sequenza accettore di splicing non è presente.

- 5 Preferibilmente, la sequenza di DNA esogeno comprende una sequenza codificante (preferibilmente uno o più esoni o suoi frammenti) di una proteina terapeutica, ad esempio la rodopsina, preferibilmente comprende uno o più esoni della rodopsina o suoi frammenti.

Preferibilmente, la sequenza bersaglio è un sito bersaglio di RNA guida (gRNA).

- 10 Preferibilmente, detto oligonucleotide complementare alla sequenza bersaglio è un RNA guida che ibrida con una sequenza bersaglio di un gene o con il suo filamento complementare.

Detto oligonucleotide guida quindi la nucleasi a tagliare all'interno della sequenza bersaglio del gene. Di preferenza, detto RNA guida è adiacente a una sequenza motivo adiacente al protospaziatore (PAM).

- 15 Preferibilmente, detto oligonucleotide complementare alla sequenza di targeting è sotto il controllo di un promotore, preferibilmente un promotore U6,

Preferibilmente, la sequenza bersaglio invertita è una sequenza invertita rispetto a una sequenza bersaglio e/o comprende una sequenza PAM, preferibilmente alla sua estremità 3'.

Preferibilmente, detto acido nucleico donatore comprende ulteriormente uno o più tra:

- 20
- un linker, preferibilmente tra il sito di taglio enzimatico e la sequenza di salto ribosomiale;
 - un'ulteriore sequenza di salto ribosomiale, preferibilmente localizzata all'estremità 3' della sequenza di DNA esogeno;
 - un elemento regolatore post-trascrizionale, preferibilmente localizzato all'estremità 3' della sequenza di DNA esogeno o dell'ulteriore sequenza di salto ribosomiale;
- 25
- una sequenza di terminazione della trascrizione, preferibilmente localizzata all'estremità 3' dell'elemento regolatore post-trascrizionale o all'estremità 3' della sequenza di DNA esogeno o dell'ulteriore sequenza di salto ribosomiale,
- 30
- preferibilmente in cui detto elemento regolatore post-trascrizionale è l'elemento regolatore post-trascrizionale del virus dell'epatite di marmotta (WPRE) e/o detta sequenza di terminazione della trascrizione è una sequenza di segnale di poliadenilazione, preferibilmente il poliA dell'ormone della crescita bovino (BGH polyA) e/o

detta ulteriore sequenza di salto ribosomiale è una sequenza T2A, P2A, E2A, F2A, preferibilmente la sequenza T2A.

Come già menzionato, la sequenza di DNA donatore è affiancata alle estremità 5' e 3' dallo stesso sito bersaglio del gRNA che il gRNA riconosce, ma invertito (ad esempio, un sito bersaglio invertito o una sequenza bersaglio invertita).

Nella presente invenzione, detto acido nucleico donatore (o costruito) comprende preferibilmente:

- una sequenza bersaglio invertita con la sua sequenza del motivo adiacente al protospaziatore (PAM),
- una sequenza accettore di splicing,
- una sequenza di segnale di degradazione, preferibilmente la sequenza CL1,
- un sito di taglio enzimatico, preferibilmente un sito di taglio della furina,
- una sequenza di salto ribosomiale, preferibilmente una sequenza P2A,
- una sequenza di DNA esogeno, preferibilmente uno o più esoni di rodopsina,
- un'ulteriore sequenza di salto ribosomiale, preferibilmente T2A,
- una sequenza di terminazione della trascrizione, e
- un'ulteriore sequenza bersaglio invertita con la sua sequenza del motivo adiacente al protospaziatore (PAM).

Tra il sito di taglio enzimatico e la sequenza di salto ribosomiale può essere presente un linker.

Un elemento regolatore post-trascrizionale può essere presente all'estremità 5' della sequenza di terminazione della trascrizione.

Preferibilmente, detti elementi sono in ordine 5'-3' come elencati ma anche altri ordini possono essere ugualmente idonei.

Preferibilmente, la sequenza di salto ribosomiale comprende o ha essenzialmente una sequenza

che ha almeno l'80% di identità con SEQ ID NO: 1
(GCCACCAACTTCTCCCTGCTGAAGCAGGCCGCGACGTGGAGGAGAACCCCGGCCCC) o con SEQ ID NO: 2

(GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCT) o con una sequenza codificante per SEQ ID NO: 3 (GSG) E G R G S L L T C G D V E E N P G P or

SEQ ID NO: 4 (GSG) A T N F S L L K Q A G D V E E N P G P o suoi frammenti funzionali e/o

la sequenza bersaglio invertita comprende o essenzialmente ha una sequenza che ha almeno il 95% di identità con SEQ ID NO: 5 (ACACCAGGAGACTTGGAACG) o suoi frammenti funzionali e opzionalmente comprende la sequenza SpCas9 PAM (CGG) e/o

l'RNA guida comprende o ha essenzialmente o è codificata da una sequenza che ha almeno il 95% di identità con SEQ ID NO: 5 (ACACCAGGAGACTTGGAACG), o suoi frammenti funzionali e/o l'oligonucleotide complementare alla sequenza bersaglio comprende o ha essenzialmente o è codificata da una sequenza che ha almeno il 95% di identità con SEQ ID NO: 5 (ACACCAGGAGACTTGGAACG) o suoi frammenti funzionali e/o

la sequenza di segnale di degradazione comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con SEQ ID NO: 6 (gcctgcaagaactgggtcagcagcctgagccacttcgtgatccacctg) e/o

il sito di taglio enzimatico comprende o ha essenzialmente una sequenza che ha almeno il 95% di identità con SEQ ID NO: 7 (CGAAAAGAAGA) e/o

il linker comprende o ha essenzialmente una sequenza che ha almeno il 95% di identità con ggaagcgga e/o

la sequenza accettore di splicing comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con SEQ ID NO: 9 (GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGTGT) e/o

la sequenza di DNA esogeno comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con almeno una delle seguenti sequenze: SEQ ID NO: 10

(ATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCCCTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCCTACATGTTTCTGCTGATCGTGCTGGGCTTCCCATCAACTTCTCACGCTCTACGTCACCGTCCAGCACAAGAAGCTGCGCACGCCTCTCAACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTCACAGCACCCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGCCCCACAGGATGCAATTTGGAGGGCTTCTTGCCACCCTGGGCG),

SEQ ID NO: 11

(GTGAAATTGCCCTGTGGTCCTTGGTGGTCCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCATGAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTCACCTGGGTGCATGGCGCTGGCCTGCGCCGCACCCCCACTCGCCGGCTGGTCCAG);

SEQ ID NO: 12

(GTACATCCCCGAGGGCCTGCAGTGCTCGTGGAATCGACTACTACACGCTCAAGCCGGAGGTCAAC
AACGAGTCTTTTGTCTACATGTTCTGTTGCTCCACTTCACCATCCCCATGATTATCATCTTTTCTGCTA
TGGGCAGCTCGTCTTCACCGTCAAGGAG);

SEQ ID NO: 13

- 5 (GCCGCTGCCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTCACCCGCATGGTC
ATCATCATGGTCATCGCTTTCCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTCACCC
ACCAGGGCTCCAACTTCGGTCCCATCTTCATGACCATCCCAGCGTTCTTTGCCAAGAGCGCCGCCATCTA
CAACCTGTCTATATCATGATGAACAAGCAG);

SEQ ID NO: 14

- 10 (TTCCGGAAGTGCATGCTCACCACCATCTGCTGCGGCAAGAACCCACTGGGTGACGATGAGGCCTCTGC
TACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCAGCA) e/o

l'elemento di regolazione post trascrizionale del virus dell'epatite della marmotta (wpre)
comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con SEQ ID NO:
15

- 15 (Taagcttggatccaatcaacctctggattacaaaatttgtaaagattgactggatttcttaactatgttgctccttttacgctatgtgga
tacgctgctttaatgcctttgtatcatgctattgcttcccgatggctttcattttctccttgataaatcctgggtgctgtctctttatgag
gagttgtggccgttgctcaggcaacgtggcgtggtgtgactgtgtttgctgacgcaacccccactgggtggggcattgccaccacctgt
cagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgctgccttggcgctgctggacaggggctc
ggctgttgggcactgacaattccgtggtgtgtcggggaaatcatcgctcctttccttggtgctgcctgtgttgccacctggattctgcgc
20 gggacgtccttctgctacgtcccttcggccctcaatccagcggaacttcttcccgcggtgctgcccggctctgcccgtcttccgcgtct
tcg) e/o

il segnale di poliadenilazione dell'ormone della crescita bovino (BGH pA) comprende o ha
essenzialmente una sequenza che ha almeno l'80% di identità con SEQ ID NO: 16

- (GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCGTGCCTTCCTTGACCCTGG
25 AAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTC
ATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGC
ATGCTGGGGA).

- Un altro oggetto dell'invenzione è un vettore che comprende il sistema di editing genetico come
definito sopra o nel presente documento o l'acido nucleico donatore e/o l'oligonucleotide
30 complementare alla sequenza bersaglio e/o una nucleasi che riconosce la sequenza bersaglio
come definita sopra o nel presente documento.

Il vettore è preferibilmente un vettore virale, selezionato dal gruppo costituito da: vettore adeno-associato (AAV), vettore adenovirale, vettore lentivirale, vettore retrovirale, o un vettore non virale, preferibilmente selezionato tra: un mezzo di rilascio basato su polimeri, particelle, lipidi, peptidi o loro combinazioni, come polimeri cationici, micelle, liposomi, esosomi, microparticelle e nanoparticelle, comprese le nanoparticelle lipidiche (LNP).

Preferibilmente, il vettore comprende anche una sequenza nucleotidica 5'-terminale ripetuta (5'-TR) e una sequenza nucleotidica 3'-terminale ripetuta (3'-TR); preferibilmente, la 5'-TR è una sequenza nucleotidica 5'-terminale invertita ripetuta (5'-ITR) e la 3'-TR è una sequenza nucleotidica 3' terminale invertita ripetuta (3'-ITR).

Preferibilmente gli ITR derivano dallo stesso sierotipo virale o da sierotipi virali diversi. Preferibilmente, il virus è un AAV, preferibilmente del sierotipo 2.

Un ulteriore oggetto dell'invenzione è una cellula ospite che comprende il sistema di editing genetico o il vettore come definito qui o sopra.

Un altro oggetto dell'invenzione è una particella virale che comprende il sistema di editing genetico o un vettore come definito sopra o qui.

Preferibilmente in cui la particella virale comprende le proteine del capsido di un AAV. Di preferenza, la particella virale comprende proteine del capsido di un AAV di un sierotipo selezionato da uno o più del gruppo costituito da AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9 e AAV 10, preferibilmente dal sierotipo AAV2 o AAV8.

Un ulteriore oggetto dell'invenzione è una composizione farmaceutica che comprende uno dei seguenti: un sistema di editing genetico, un vettore, una cellula ospite, una particella virale come definita in precedenza o nel presente documento e un veicolo farmaceuticamente accettabile.

In modo idoneo, un vettore virale come qui definito comprende una particella del vettore virale.

Con il termine "particella del virus" o "particella virale" si intende la forma extracellulare di un virus non patogeno, in particolare un vettore virale, composto da materiale genetico costituito da DNA o RNA circondato da un rivestimento proteico, chiamato capsido, e in alcuni casi da un involucro derivato da porzioni di membrane della cellula ospite comprendente glicoproteine virali.

Come usato nel presente documento, un vettore virale si riferisce anche a una particella virale vettoriale.

I vettori virali compresi nella presente invenzione sono adatti alla terapia genica.

Di preferenza, la particella virale comprende le proteine del capsido di un AAV.

Preferibilmente, la particella virale comprende proteine del capsido di un AAV di un sierotipo selezionato tra uno o più del gruppo costituito da: AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV 10, AAVSH19, AAVPHP.B ; di preferenza dal sierotipo AAV2 o AAV8.

- 5 Un altro oggetto dell'invenzione è un kit che comprende: un sistema di editing genetico, o un vettore, o una cellula ospite conseguente, o una particella virale o una composizione farmaceutica, come definito sopra o nel presente documento, in uno o più contenitori, opzionalmente comprendenti anche istruzioni o materiali di confezionamento che descrivono come somministrare il costrutto di acido nucleico, il vettore, la cellula ospite, la particella virale
10 o la composizione farmaceutica a un paziente.

Ulteriori oggetti dell'invenzione sono il sistema di editing genetico, o un vettore, una cellula ospite conseguente, una particella virale o una composizione farmaceutica, come definito in precedenza o nel presente documento, per uso come medicinale, preferibilmente per uso nel trattamento di una malattia genetica.

- 15 Ulteriori oggetti dell'invenzione sono il sistema di editing genetico, o un vettore, una cellula ospite conformemente, una particella virale o una composizione farmaceutica come definito sopra o qui per l'uso nel trattamento di malattie ereditate con modalità autosomica dominante in cui entrambi gli alleli mutanti e di tipo selvatico sono sostituiti con una copia corretta del gene fornito dal DNA donatore o per l'uso nel trattamento di malattie ereditarie e comuni dovute a
20 guadagno di funzione tossica, preferibilmente dette malattie comprendono la distrofia retinica, preferibilmente la distrofia retinica è selezionata tra la retinite pigmentosa, la distrofia del cono o la distrofia dei coni e dei bastoncelli, la degenerazione maculare, ad es. g. malattia di Stargardt (ELOVL4), Von-Hippel Lindau, retinoblastoma, RP4 (vedi RHO; OMIM: 180380), RP63 (vedi OMIM: 614494), CORD1 (distrofia dei coni e dei bastoncelli 1; vedi OMIM: 600624), CORD17
25 (distrofia dei coni e dei bastoncelli 17; vedi OMIM: 615163), BEST1 (bestrofina-1; malattia di Best; distrofia maculare vitelliforme proteina 2; vedi OMIM : 607854), OPA1 (OPA1 GTPasi mitocondrialesimile alla dinamina; vedi OMIM : 605290), malattie neuronali, epatiche, lipofuscinosi (malattia di Batten e altre), preferibilmente per uso nel trattamento di malattie oculari ereditate con modalità dominante, preferibilmente la degenerazione retinica,
30 preferibilmente retinite pigmentosa, malattie neuronali ed epatiche.

Un altro oggetto dell'invenzione è il sistema di editing genetico o del vettore o del costrutto come definito sopra o nel presente documento per la produzione di particelle virali.

Preferibilmente, la sequenza di salto ribosomiale T2A comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con SEQ ID NO: 2 (GGAAGCGGAGGGCAGGAAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCT) o con una sequenza che codifica per SEQ ID NO: 3 (GSG) E G R G S L L T C G D V E N P G P

5 o loro frammenti funzionali.

Di preferenza, la sequenza di salto ribosomiale P2A comprende o ha essenzialmente una sequenza con almeno l'80% di identità con SEQ ID NO:1 (gccaccaacttctccctgctgaagcagggcgacgtggagaaccccgcccc) o con una sequenza che codifica per SEQ ID NO: 4 (GSG) A T N F S L L K Q A G D V E E N P G P o suoi frammenti funzionali.

10 In una forma di realizzazione preferita, l'oligonucleotide complementare alla sequenza di bersaglio può comprendere o avere essenzialmente o essere codificato da una sequenza avente almeno il 95% di identità con SEQ ID NO: 5 (ACACCAGGAGACTTGGAAACG) o suoi frammenti funzionali.

Preferibilmente, l'acido nucleico donatore comprende ulteriormente un segnale di poliadenilazione, preferibilmente un poliA dell'ormone della crescita bovino.

Preferibilmente, la sequenza bersaglio è una sequenza contenuta nel gene della rodopsina (Rho). Preferibilmente, la sequenza bersaglio è una sequenza compresa nel gene della rodopsina e la sequenza di DNA esogeno (o sequenza di DNA donatore) è una sequenza codificante della proteina rodopsina.

20 Preferibilmente, la sequenza bersaglio è compresa all'interno di:

- il primo esone del gene RHO, preferibilmente di origine umana, murina o suina,
 - il primo introne del gene RHO, preferibilmente umano, di topo o di maiale,
- o loro frammenti funzionali.

Preferibilmente, la sequenza bersaglio è un sito bersaglio di un RNA guida (gRNA) e detto oligonucleotide complementare alla sequenza bersaglio è un RNA guida che ibrida con una sequenza bersaglio di un gene.

Detto RNA guida può comprendere o avere essenzialmente o essere codificato da una sequenza avente almeno il 95% di identità con SEQ ID NO: 5 (ACACCAGGAGACTTGGAAACG) o suoi frammenti funzionali.

30 Detta sequenza di DNA esogeno comprende preferibilmente un gene reporter, preferibilmente detto gene reporter è selezionato tra almeno uno dei seguenti: rosso discosoma (ds-RED),

proteina fluorescente verde (GFP), proteina fluorescente rossa (RFP), luciferasi, β -galattosidasi e β -glucuronidasi.

Detta nucleasi è preferibilmente selezionata tra: una nucleasi CRISPR, una TALEN, una nucleasi DNA-guidata, una meganucleasi e una Nucleasi a zinc finger, preferibilmente tale nucleasi è una nucleasi CRISPR selezionata dal gruppo costituito da: Cas9, Cpf1, Cas12b (C2cl), Cas13a (C2c2), Cas3, Csf1, Cas13b (C2c6) e C2c3 o loro varianti come SaCas9 o VQR-Cas9-HF1.

Detto oligonucleotide complementare, detto acido nucleico donatore, detto polinucleotide codificante la nucleasi sono preferibilmente compresi in un vettore virale o non virale, preferibilmente detto vettore virale è selezionato tra: un virus adeno-associato, un lentivirus, un retrovirus e un adenovirus.

Preferibilmente la cellula è selezionata dal gruppo costituito da: una o più cellule retiniche, preferibilmente cellule ganglionari retiniche, cellule bipolari, cellule amacrine, epitelio pigmentato retinico, cellule orizzontali, cellule dei bastoncelli e dei coni e, preferibilmente, cellule della regione anteriore dell'occhio come epitelio pigmentato dell'iride, epitelio corneale, fibroblasti corneali, linfociti, monociti, neutrofili, eosinofili, basofili, cellule endoteliali, cellule epiteliali, epatociti, osteociti, piastrine, adipociti, cardiomiociti, neuroni, cellule muscolari lisce, cellule muscolari scheletriche, spermatociti, ovociti e cellule del pancreas, cellule staminali pluripotenti indotte (iPScells), cellule staminali, cellule staminali ematopoietiche, cellule staminali progenitrici ematopoietiche, preferibilmente la cellula è una cellula della retina dell'occhio o un epatocita di un soggetto.

In una forma di realizzazione preferita, l'acido nucleico donatore e/o la sequenza accettore di splicing e/o la sequenza del segnale di degradazione e/o il sito di taglio enzimatico e/o la sequenza di salto ribosomiale e/o la sequenza di DNA esogeno e/o le sequenze bersaglio e/o l'oligonucleotide complementare e/o la nucleasi sono come sopra definiti.

Preferibilmente, l'oligonucleotide complementare e/o l'acido nucleico donatore e/o il polinucleotide che codifica la nucleasi sono contenuti in uno o più vettori virali o non virali, preferibilmente un vettore virale selezionato tra: un virus adeno-associato, un retrovirus, un adenovirus e un lentivirus.

Preferibilmente, oggetto dell'invenzione sono le sequenze qui menzionate.

Per acido nucleico donatore si intende generalmente l'acido nucleico che comprende la sequenza esogena che deve essere integrata nel genoma bersaglio.

Tuttavia, può anche essere inteso come comprendente l'oligonucleotide complementare alla sequenza bersaglio.

Nel contesto della presente invenzione, gli elementi della cassetta di DNA donatore e/o gli elementi della cassetta di espressione del gRNA e/o le sequenze del promotore e/o il promotore U6 per l'espressione del gRNA e/o il gRNA e/o il sito di bersaglio del gRNA e/o il Cas9/Cas9-2a-GFP e/o il transgene terapeutico e/o il polyA e/o il T2A e/o il P2A e/o la sequenza accettore di splice e/o il CL1 sono le sequenze rappresentate nelle seguenti sequenze 27, 30, 31, 32, 34 o 62 o nelle sequenze qui descritte.

Preferibilmente, un primo vettore comprende l'acido nucleico donatore e l'oligonucleotide complementare a una sequenza bersaglio e un secondo vettore comprende l'acido nucleico che codifica per la nucleasi che riconosce tale sequenza bersaglio. In alternativa, un primo vettore comprende l'acido nucleico donatore e un secondo vettore comprende l'oligonucleotide complementare a una sequenza bersaglio e l'acido nucleico che codifica per la nucleasi che riconosce tale sequenza bersaglio. Come ulteriore alternativa, vengono forniti tre vettori: un primo vettore che comprende l'acido nucleico donatore, un secondo vettore che comprende l'oligonucleotide complementare a una sequenza bersaglio e un terzo vettore che comprende l'acido nucleico che codifica per la nucleasi che riconosce tale sequenza bersaglio.

Un ulteriore oggetto dell'invenzione è un metodo di integrazione di una sequenza di DNA esogeno nel genoma di una cellula (o in una sequenza di acido nucleico bersaglio in un genoma), preferibilmente di una cellula non in divisione, che comprende il contatto della cellula con:

a) un acido nucleico donatore comprendente:

- una sequenza di segnale di degradazione,
- un sito di taglio enzimatico,
- una sequenza di salto ribosomiale,
- una sequenza di DNA esogeno

dove detto acido nucleico donatore è affiancato alle estremità 5' e 3' da sequenze bersaglio invertite; e opzionalmente

b) un oligonucleotide complementare alla sequenza bersaglio e/o

c) una nucleasi che riconosce la sequenza bersaglio.

Preferibilmente, detto acido nucleico donatore comprende anche una sequenza accettore di splicing, preferibilmente all'estremità 5' della sequenza del segnale di degradazione.

Preferibilmente, il metodo di integrazione di una sequenza di DNA esogeno nel genoma di una cellula (o in una sequenza di acido nucleico bersaglio in un genoma) comprende il contatto della cellula con:

a) un acido nucleico donatore comprendente:

- 5 - una sequenza accettore di splicing,
- una sequenza di segnale di degradazione,
- un sito di taglio enzimatico,
- una sequenza di salto ribosomiale,
- detta sequenza di DNA esogeno

10 in cui detto acido nucleico donatore è affiancato alle estremità 5' e 3' da sequenze bersaglio invertite;

b) un oligonucleotide complementare alla sequenza bersaglio e

c) una nucleasi che riconosce la sequenza bersaglio.

Di preferenza, l'acido nucleico donatore e/o la sequenza del segnale di degradazione e/o il sito di taglio enzimatico e/o il segnale di salto ribosomiale e/o la sequenza di DNA esogeno e/o le sequenze bersaglio e/o l'oligonucleotide complementare e/o la nucleasi sono come definiti sopra o qui.

L'oggetto dell'invenzione è anche un processo di preparazione di una particella di vettore virale che comprende l'introduzione di tali costrutti di DNA in una cellula ospite e l'ottenimento della particella vettoriale virale.

In una forma di realizzazione preferita, l'acido nucleico donatore e/o la sequenza del segnale di degradazione e/o il sito di taglio enzimatico e/o il segnale di salto ribosomiale e/o la sequenza di DNA esogeno e/o le sequenze bersaglio e/o l'oligonucleotide complementare e/o la nucleasi sono quelli definiti sopra.

25 Preferibilmente, l'oligonucleotide complementare e/o l'acido nucleico donatore e/o il polinucleotide che codifica la nucleasi sono contenuti in uno o più vettori virali o non virali, preferibilmente detti vettori virali essendo selezionati tra: un virus adeno-associato, un retrovirus, un adenovirus e un lentivirus; detto vettore non virale è preferibilmente selezionato da un veicolo di rilascio basato su polimeri, particelle, lipidi, peptidi o loro combinazioni, come

30 polimeri cationici, micelle, liposomi, esosomi, microparticelle e nanoparticelle, comprese le nanoparticelle lipidiche (LNP).

Preferibilmente, un primo vettore comprende l'acido nucleico donatore e l'oligonucleotide complementare a una sequenza bersaglio e un secondo vettore comprende l'acido nucleico che codifica per la nucleasi che riconosce detta sequenza bersaglio. In alternativa, un primo vettore comprende l'acido nucleico donatore e un secondo vettore comprende l'oligonucleotide complementare a una sequenza bersaglio e l'acido nucleico che codifica per la nucleasi che riconosce tale sequenza bersaglio. Come ulteriore alternativa, sono forniti tre vettori: un primo vettore che comprende l'acido nucleico donatore, un secondo vettore che comprende l'oligonucleotide complementare a una sequenza bersaglio e un terzo vettore che comprende l'acido nucleico che codifica per la nucleasi che riconosce tale sequenza bersaglio.

Di preferenza, sia la sequenza bersaglio (definita anche sequenza bersaglio) sia la sequenza di acido nucleico bersaglio nel genoma sono riconosciute dalla nucleasi.

In una forma di realizzazione preferita dell'invenzione, la sequenza di acido nucleico bersaglio nel genoma non è più presente una volta che la sequenza di DNA esogeno è stata integrata nel genoma della cellula (preferibilmente una cellula non in divisione) con un orientamento corretto.

In una forma di realizzazione preferita dell'invenzione, il metodo non comprende la modifica dell'identità genetica della linea germinale degli esseri umani.

Preferibilmente, detta sequenza di DNA esogeno comprende un gene reporter, preferibilmente detto gene reporter è selezionato tra almeno uno dei seguenti: rosso discosoma, proteina fluorescente verde (GFP), proteina fluorescente rossa (RFP), luciferasi, β -galattosidasi e β -glucuronidasi.

Nel contesto della presente invenzione, la nucleasi può essere fornita come proteina o come acido nucleico che codifica per detta nucleasi. Detto acido nucleico può essere DNA o RNA, ad esempio può essere l'mRNA di una nucleasi o può essere un cDNA o la sequenza codificante del DNA di una nucleasi o un costrutto di DNA che codifica per la nucleasi. Preferibilmente, detto acido nucleico che codifica per una nucleasi è un costrutto di DNA che comprende un acido nucleico che codifica per Cas9 o spCas9, preferibilmente sotto il controllo di un promotore specifico del tessuto. Detto costrutto può inoltre comprendere un poliA, preferibilmente un poliA sintetico corto (sh polyA). Tutti questi elementi sono ben noti nell'arte e possono avere sequenze nucleotidiche convenzionali.

Preferibilmente, la nucleasi è selezionata tra: una nucleasi CRISPR, una TALEN, una nucleasi guidata dal DNA, una meganucleasi e una nucleasi Zinc Finger; preferibilmente, detta nucleasi

è una nucleasi CRISPR selezionata dal gruppo costituito da: Cas9, Cpf1, Cas12b (C2cl), Cas13a (C2c2), Cas3, Csf1, Cas13b (C2c6) e C2c3 o loro varianti come SaCas9 o VQR-Cas9-HF1.

In genere, detto acido nucleico donatore, detto oligonucleotide complementare alla sequenza bersaglio e detto acido nucleico codificante per la nucleasi sono contenuti in costrutti di DNA.

- 5 Preferibilmente, un primo costrutto di DNA comprende l'acido nucleico donatore e l'oligonucleotide complementare a una sequenza bersaglio e un secondo costrutto di DNA comprende l'acido nucleico che codifica per la nucleasi che riconosce tale sequenza bersaglio. In alternativa, un primo costrutto di DNA comprende l'acido nucleico donatore e un secondo costrutto di DNA comprende l'oligonucleotide complementare a una sequenza bersaglio e
- 10 l'acido nucleico che codifica per la nucleasi che riconosce detta sequenza bersaglio. Come ulteriore alternativa, sono forniti tre costrutti: un primo costrutto comprendente l'acido nucleico donatore, un secondo costrutto comprendente l'oligonucleotide complementare a una sequenza bersaglio e un terzo costrutto comprendente l'acido nucleico codificante per la nucleasi che riconosce detta sequenza bersaglio.

- 15 Anche questi costrutti sono oggetto dell'invenzione.

Preferibilmente, uno o più di detti costrutti di DNA sono contenuti in un vettore, preferibilmente un vettore virale, ancora più preferibilmente un vettore lentivirale o un vettore adeno-associato. In alternativa, tutti o alcuni di detti costrutti di DNA possono essere inseriti in un vettore non virale, ove detto vettore non virale è selezionato tra un mezzo di veicolazione a

20 base di polimeri, particelle, lipidi, peptidi o loro combinazioni, come polimeri cationici, micelle, liposomi, esosomi, microparticelle e nanoparticelle, comprese le nanoparticelle lipidiche (LNP). Anche detti vettori sono oggetto dell'invenzione.

Detto oligonucleotide complementare, detto acido nucleico donatore e detto acido nucleico codificante la nucleasi possono essere contenuti in uno o più vettori virali o non virali, detti

25 vettori virali essendo preferibilmente selezionati tra: un virus adeno-associato, un lentivirus, un retrovirus e un adenovirus. Ciò significa che possono essere nello stesso vettore o in vettori diversi.

Un costrutto comprendente detto acido nucleico donatore e oligonucleotide complementare può comprendere o avere essenzialmente una sequenza avente almeno 80% o almeno 85% o

30 almeno 90% o almeno 95% di identità con SEQ ID N. 32.

Preferibilmente, l'oligonucleotide complementare, l'acido nucleico donatore e un polinucleotide che codifica la nucleasi sono contenuti in un vettore virale o non virale,

preferibilmente detto vettore virale selezionato tra: un virus adeno-associato, un lentivirus, un retrovirus e un adenovirus.

Un altro oggetto dell'invenzione è una cellula ottenibile con il metodo sopra definito, preferibilmente per uso medico e/o per uso nel trattamento di una malattia genetica, preferibilmente per uso nel trattamento di malattie ereditarie autosomiche dominanti in cui entrambi gli alleli mutante e di tipo selvatico sono sostituiti con una copia corretta del gene fornito dal DNA donatore o per uso nel trattamento di malattie ereditarie e comuni dovute a guadagno di funzione tossica, preferibilmente dette malattie comprendono la distrofia retinica, preferibilmente la distrofia retinica è selezionata tra la retinite pigmentosa, la distrofia dei coni o la distrofia dei coni e dei bastoncelli, la degenerazione maculare, ad es. g. malattia di Stargardt (ELOVL4), Von-Hippel Lindau, retinoblastoma, RP4 (vedi RHO; OMIM: 180380), RP63 (vedi OMIM: 614494), CORD1 (distrofia dei coni e dei bastoncelli 1; vedi OMIM: 600624), CORD17 (distrofia dei coni e dei bastoncelli 17; vedi OMIM: 615163), BEST1 (bestrofina-1; malattia di Best; distrofia maculare vitelliforme proteina 2; vedi OMIM : 607854), OPA1 (OPA1 GTPasi mitocondriale simile alla dinamina; vedi OMIM : 605290), malattie neuronali, epatiche, lipofuscinosi (malattia di Batten e altre), preferibilmente per uso nel trattamento di malattie oculari ereditate con modalità dominante, ad esempio la degenerazione retinica. degenerazione retinica, preferibilmente retinite pigmentosa, malattie neuronali ed epatiche.

Preferibilmente, la cellula è selezionata dal gruppo costituito da: una o più cellule retiniche, preferibilmente cellule ganglionari retiniche, cellule bipolari, cellule amacrine, epitelio pigmentato retinico, cellule orizzontali, cellule dei bastoncelli e dei coni, o della regione anteriore dell'occhio, come l'epitelio pigmentato dell'iride, l'epitelio corneale, i fibroblasti corneali, linfociti, monociti, neutrofili, eosinofili, basofili, cellule endoteliali, cellule epiteliali, epatociti, osteociti, piastrine, adipociti, cardiomiociti, neuroni, cellule muscolari lisce, cellule muscolari scheletriche, spermatociti, ovociti e cellule del pancreas, cellule staminali pluripotenti indotte (iPScells), cellule staminali, cellule staminali ematopoietiche, cellule staminali progenitrici ematopoietiche, preferibilmente la cellula è una cellula della retina di un occhio o un epatocita di un soggetto.

Un ulteriore oggetto dell'invenzione è un costrutto (di DNA) che comprende l'acido nucleico donatore e/o l'oligonucleotide complementare a una sequenza bersaglio e/o un acido nucleico che codifica per una nucleasi che riconosce detta sequenza bersaglio, come qui definita.

Nel contesto della presente invenzione, i termini "oligonucleotide complementare a una sequenza bersaglio" o "gRNA" possono comprendere le molecole di DNA che li codificano. Pertanto, le sequenze qui menzionate per l'oligonucleotide complementare a una sequenza bersaglio o per il gRNA possono essere le sequenze che li codificano.

5 In una forma di realizzazione, i metodi dell'invenzione sono *ex-vivo* o *in vitro*. In una forma di realizzazione, nei metodi dell'invenzione la cellula è una cellula isolata da un soggetto o da un paziente.

In una forma di realizzazione alternativa, i metodi dell'invenzione sono *in vivo*.

10 In una forma di realizzazione nei metodi dell'invenzione la cellula è una cellula isolata da un soggetto o da un paziente.

Un altro oggetto dell'invenzione è una composizione, preferibilmente per uso medico, preferibilmente per il trattamento delle malattie/patologie qui menzionate, che comprende

a) un acido nucleico donatore comprendente:

- una sequenza di segnale di degradazione
- 15 - un sito di taglio enzimatico
- una sequenza di salto ribosomiale,
- una sequenza di DNA esogeno

in cui detto acido nucleico donatore è affiancato alle estremità 5' e 3' da sequenze bersaglio invertite; e opzionalmente,

- 20 b) un oligonucleotide complementare alla sequenza bersaglio e/o
- c) una nucleasi che riconosce la sequenza bersaglio.

Preferibilmente, la composizione comprende

a) un acido nucleico donatore che comprende:

- una sequenza di segnale di degradazione,
- 25 - un sito di taglio enzimatico,
- una sequenza di salto ribosomiale,
- detta sequenza di DNA esogeno

dove detto acido nucleico donatore è affiancato alle estremità 5' e 3' da sequenze bersaglio invertite;

- 30 b) un oligonucleotide complementare alla sequenza bersaglio e
- c) una nucleasi che riconosce la sequenza bersaglio.

Preferibilmente, detto acido nucleico donatore comprende anche una sequenza accettore di splicing, preferibilmente all'estremità 5' della sequenza segnale di degradazione.

Pertanto, preferibilmente, la composizione comprende

a) un acido nucleico donatore che comprende:

- 5 - una sequenza accettore di splicing,
- una sequenza di segnale di degradazione,
- un sito di taglio enzimatico,
- una sequenza di salto ribosomiale,
- detta sequenza di DNA esogeno,

10 in cui detto acido nucleico donatore è affiancato alle estremità 5' e 3' da sequenze bersaglio invertite;

 b) un oligonucleotide complementare alla sequenza bersaglio e

 c) una nucleasi che riconosce la sequenza bersaglio.

15 Preferibilmente, sia la sequenza bersaglio (definita anche sequenza bersaglio) sia la sequenza di acido nucleico bersaglio nel genoma sono riconosciute dalla nucleasi.

In una forma di realizzazione preferita dell'invenzione, la sequenza di acido nucleico bersaglio nel genoma non è più presente una volta che la sequenza di DNA esogeno è stata integrata nel genoma della cellula (preferibilmente una cellula non in divisione) con un orientamento corretto.

20 Preferibilmente, gli elementi specificati nell'acido nucleico donatore o nei vettori o nel sistema sono nell'ordine 5'-3' come elencati, ma altri ordini possono essere ugualmente idonei.

In una forma di realizzazione preferita dell'invenzione, le cellule non sono cellule della linea germinale degli esseri umani.

25 Preferibilmente, la sequenza di DNA esogeno corregge una mutazione nel genoma della cellula, preferibilmente una cellula non in divisione.

Le sequenze di bersaglio invertite (o sito di gRNA invertito) nel contesto della presente invenzione sono posizionate a monte e a valle del DNA donatore, che è il costrutto di DNA che viene tagliato e poi integrato nel genoma bersaglio. Le sequenze di bersaglio invertite sono le stesse sequenze riconosciute dall'RNA guida nel locus genomico bersaglio, cioè la sequenza di bersaglio (ad esempio la rodopsina), ma invertite o invertite rispetto alla sequenza genomica. Ciò consente di ottenere un'integrazione monodirezionale.

30

Invertito o inverso può significare che se la sequenza bersaglio ha una specifica sequenza 5'-3', la sequenza bersaglio invertita ha la stessa sequenza ma con orientamento 3'-5'. Pertanto, una sequenza bersaglio invertita può essere complementare all'RNA guida ma invertita. Ciò consente di ottenere un'integrazione monodirezionale, come noto nel metodo HITI. In breve, nel caso in cui il DNA donatore fosse integrato nella direzione opposta, la nucleasi, come Cas9, sarebbe in grado di riconoscere nuovamente il suo sito bersaglio e di tagliarlo. Quando l'integrazione avviene nell'orientamento corretto, la nucleasi non sarebbe più in grado di tagliare il sito bersaglio. Preferibilmente, ciascuna di queste sequenze bersaglio invertite è legata all'estremità 3' a una sequenza motivo adiacente al protospaziatore (PAM). I geni e/o le sequenze di DNA esogeno e/o gli introni e/o gli esoni qui descritti possono essere di qualsiasi origine, preferibilmente di origine umana o murina.

Nella presente invenzione, la sequenza della rodopsina (Rho) è preferibilmente indicata con i seguenti numeri di accesso: umano: AB065668.1, topo: AC142099.3, maiale: AEMK02000087.1.

Nella presente invenzione la sequenza di DNA esogeno può presentare almeno un nucleotide differente rispetto al genoma. Il DNA esogeno può non presentare la/e mutazione/i presenti nel gene che è mirato dall'oligonucleotide complementare a una sequenza bersaglio o dal gRNA.

Nella presente invenzione, RNA guida o gRNA possono essere utilizzati come sinonimi di oligonucleotide a filamento complementare omologo alla sequenza bersaglio o di oligonucleotide complementare o oligonucleotide complementare alla sequenza bersaglio e possono anche riferirsi alla sequenza di DNA che li codifica o alla molecola di DNA corrispondente.

In una forma di realizzazione preferita dell'invenzione, un vettore che comprende IRBP e Cas9 è utilizzato insieme a un secondo vettore che comprende il DNA donatore come definito sopra.

La sequenza di DNA donatore è preferibilmente affiancata alle estremità 3' e 5' dallo stesso sito bersaglio del gRNA che il gRNA riconosce, ma invertito (ad esempio un sito bersaglio invertito).

La cellula ottenibile secondo l'invenzione esprime la sequenza esogena.

Nel contesto della presente invenzione, la nucleasi è preferibilmente presente in un vettore diverso, in particolare quando si utilizzano vettori AAV.

Di preferenza si utilizzano vettori AAV2/8.

Nella presente invenzione, un primo vettore contenente Cas9 o spCas9 è preferibilmente sotto il controllo di un promotore specifico del tessuto, ad esempio un promotore specifico della retina o del fotorecettore, ad esempio Interphotoreceptor Retinoid-Binding Protein (IRBP).

Detto vettore può inoltre comprendere un breve polyA sintetic (sh polyA). Preferibilmente, un secondo vettore comprende la cassetta di espressione del gRNA e il DNA donatore come qui definito. Preferibilmente, il gRNA specifico per la rodopsina è sotto il promotore U6. Preferibilmente, il DNA donatore è affiancato all'estremità 3' e 5' dai siti di bersaglio invertiti del gRNA per la rodopsina, preferibilmente comprendenti il PAM.

Preferibilmente, il secondo vettore può alternativamente comprendere la cassetta di espressione per il gRNA specifico per la rodopsina e il DNA donatore che comprende la sequenza codificante per la rodopsina, come sopra definito.

In una forma di realizzazione preferita, il gene di interesse e l'enzima necessario per l'inserimento specifico nel sito NHEJ sono trasportati da due vettori AAV, in cui a causa delle dimensioni limitate dell'elemento necessario per il processo, possono essere impiegati geni di interesse più grandi. Gli inventori hanno infatti ridotto al minimo le parti strutturali (utilizzando, ad esempio, i siti di inserzione invece dei bracci di omologia), consentendo di inserire nel vettore un cDNA più lungo.

Nel contesto della presente invenzione, l'acido nucleico donatore viene inserito nel gene attraverso una giunzione non omologa.

L'invenzione fornisce anche una composizione farmaceutica che comprende l'acido nucleico come sopra definito o la sequenza nucleotidica come sopra definita o il vettore come sopra definito e diluenti e/o eccipienti e/o veicoli farmaceuticamente accettabili.

Di preferenza la composizione comprende anche un agente terapeutico, preferibilmente selezionato dal gruppo costituito da: terapia enzimatica sostitutiva e terapia con piccole molecole.

Preferibilmente, la composizione farmaceutica viene somministrata attraverso una via selezionata dal gruppo costituito da: fluido spinale cerebrale (CSF), intratecale, parenterale, endovenosa, intralesionale, intraperitoneale, intramuscolare, intratumorale, sottocutanea, intraventricolare, intra cisterna magna, lombare, intracranica, intraspinale, endovenosa, topica, nasale, orale, oculare, sottoretinica o una qualsiasi loro combinazione.

La presente invenzione fornisce anche un vettore che comprende l'acido nucleico o la sequenza nucleotidica di cui sopra per uso medico, in cui detto vettore è somministrato attraverso una via selezionata dal gruppo costituito da: fluido spinale cerebrale (CSF), intratecale, parenterale, endovenosa, intralesionale, intraperitoneale, intramuscolare, intratumorale, sottocutanea, intraventricolare, intra cisterna magna, lombare, intracranica, intraspinale, endovenosa, topica,

nasale, orale, oculare, sottoretinica o qualsiasi loro combinazione. Preferibilmente, il vettore dell'invenzione viene somministrato per via endovenosa, parenterale, oculare, preferibilmente sotto retinica.

Il vettore è preferibilmente un vettore virale, preferibilmente il vettore virale è un vettore lentivirale, un vettore virale adeno-associato, un vettore adenovirale, un vettore retrovirale, un vettore poliovirale, un vettore virale Maloney murino, un vettore virale alfa, un vettore virale del vaiolo, un vettore virale dell'herpes, un vettore virale della vaccinia, un vettore baculovirale o un vettore parvovirale, preferibilmente il virus adeno-associato è AAV2, AAV9, AAV1, AAVSH19, AAVPHP. B, AAV8, AAV6. Nel contesto della presente invenzione, si utilizzano preferibilmente vettori AAV2/8.

Preferibilmente il vettore virale comprende anche una sequenza nucleotidica 5'-terminale ripetuta (5'-TR) e una sequenza nucleotidica 3'-terminale ripetuta (3'-TR), preferibilmente la 5'-TR è una sequenza nucleotidica 5'- terminale ripetuta invertita (5'-ITR) e la 3'-TR è una sequenza nucleotidica 3'- terminale ripetuta invertita (3'-ITR), preferibilmente le ITR derivano dallo stesso sierotipo virale o da sierotipi virali diversi, preferibilmente il virus è un AAV, preferibilmente di sierotipo 2.

In una forma di realizzazione, detto vettore virale che comprende la cassetta di espressione del gRNA e il DNA donatore comprende anche una sequenza 5' terminale ripetuta invertita (ITR), preferibilmente di AAV, preferibilmente localizzata all'estremità 5' del costrutto che comprende la cassetta di espressione del gRNA e il DNA donatore e una sequenza 3' terminale ripetuta invertita (ITR), preferibilmente di AAV, preferibilmente localizzata all'estremità 3' del costrutto che comprende la cassetta di espressione del gRNA e il DNA donatore.

Preferibilmente, detto ITR comprende o ha una sequenza che ha almeno il 95% di identità rispetto a SEQ ID NO 17 o 26.

Preferibilmente detta sequenza nucleotidica è inserita in un vettore, preferibilmente un vettore virale, ancora più preferibilmente un vettore adeno-associato.

Preferibilmente, detto vettore virale che comprende la cassetta di espressione del gRNA e il DNA donatore o il vettore virale dell'invenzione comprende:

- una sequenza AAV 5'-terminale ripetuta invertita (5'-ITR);
- una sequenza bersaglio invertita legata alla sua sequenza del motivo adiacente del propospaziatore (PAM);
- una sequenza accettore di splicing;

- una sequenza di segnale di degradazione,
- un sito di taglio enzimatico
- una sequenza di salto ribosomiale, preferibilmente P2A,
- la sequenza di DNA esogeno, preferibilmente la sequenza codificante del gene della

5 rodopsina

- una sequenza di salto ribosomiale, preferibilmente T2A,
- una sequenza di terminazione della trascrizione;
- un'ulteriore sequenza bersaglio invertita legata alla sequenza del suo motivo adiacente al protospaziatore (PAM);

10 - un oligonucleotide complementare alla sequenza bersaglio, sotto il controllo di un promotore, preferibilmente il promotore U6;

- uno scheletro ("scaffold") di gRNA chimerico e

- una sequenza di AAV 3'-terminale ripetuta invertita (3'-ITR).

In modo idoneo, quando detto vettore non comprende la cassetta di espressione del gRNA,

15 detta cassetta è compresa nel vettore che comprende l'acido nucleico che esprime la nucleasi.

Gli elementi sopra citati possono essere nell'ordine sopra definito, da 5' a 3', ma altri ordini sono ugualmente adatti, come la persona esperta può apprezzare. Il vettore può inoltre comprendere sequenze virali aggiuntive, come ad esempio sequenze di AAV aggiuntive.

Nella presente invenzione "identità di almeno l'80%" significa che l'identità può essere di

20 almeno l'80%, o l'85% o il 90% o il 95% o il 100% rispetto alle sequenze di riferimento. Questo vale per tutte le % di identità menzionate. Nella presente invenzione, "identità di almeno il 95%"

significa che l'identità può essere di almeno il 95%, 96%, 97%, 98%, 99% o 100% rispetto alle sequenze di riferimento. Questo vale per tutte le % di identità menzionate. Nella presente

invenzione, "identità di almeno il 98%" significa che l'identità può essere di almeno il 98%, 99%

25 o 100% rispetto alle sequenze di riferimento. Questo vale per tutte le % di identità menzionate.

Preferibilmente, la % di identità si riferisce all'intera lunghezza della sequenza di riferimento.

Rientrano nella presente invenzione anche le sequenze di acido nucleico derivate dalle sequenze nucleotidiche qui citate, ad esempio frammenti funzionali, mutanti, varianti, derivati, analoghi e sequenze con una % di identità di almeno l'80% con le sequenze qui menzionate.

30 La sequenza codificante della presente invenzione può codificare per una variante del gene, ad esempio può comprendere aggiunte, delezioni o sostituzioni rispetto alla sequenza codificante del gene di tipo selvatico, purché queste varianti della proteina mantengano sostanzialmente la

stessa attività funzionale della proteina originale. La sequenza codificante può anche codificare per un frammento della proteina, purché questo frammento mantenga sostanzialmente la stessa attività funzionale della proteina originale. In modo idoneo, la sequenza codificante può essere codone ottimizzata per l'espressione nell'uomo.

- 5 La presente invenzione comprende anche forme realizzative in cui le sequenze qui citate, ad esempio gli RNA guida (o sequenze gRNA) o i siti di gRNA o le sequenze bersaglio o la sequenza bersaglio invertita o gli oligonucleotidi complementari, hanno un orientamento inverso, cioè dall'estremità 3' all'estremità 5', o hanno una sequenza complementare (anche con orientamento inverso o una sequenza complementare inversa ("reverse complement"))
 10 rispetto alle sequenze menzionate. Anche i detti gRNA o siti di gRNA o le sequenze di bersaglio o sequenze di bersaglio invertite o oligonucleotidi sono oggetto dell'invenzione.

Rientrano nell'invenzione anche gli acidi ribonucleici guida isolati (gRNA) che comprendono o consistono di una sequenza sostanzialmente complementare o che si appaia perfettamente a una sequenza qui descritta (nell'orientamento 5'-3' o nell'orientamento 3'-5') e a sue porzioni
 15 lunghe almeno 15 nucleotidi.

L'acido nucleico donatore nella presente invenzione può comprendere un marcatore per il rilevamento delle proteine, come 3XFLAG, preferibilmente all'estremità 5' della sequenza del segnale di degradazione.

Nel contesto della presente invenzione i frammenti sono preferibilmente lunghi almeno 15
 20 nucleotidi.

Oggetto dell'invenzione è anche un metodo per il trattamento di un soggetto affetto da una delle malattie sopra menzionate, che comprende la somministrazione al soggetto di una quantità efficace del sistema di editing genetico o del vettore o della cellula ospite o della particella virale o della composizione farmaceutica come sopra definita. Di preferenza, oggetto
 25 dell'invenzione sono le sequenze qui menzionate.

L'invenzione sarà ora illustrata mediante esempi non limitativi riferiti alle figure seguenti.

Figura 1: Piattaforma di integrazione mirata indipendente dall'omologia. Schema della strategia di integrazione mirata indipendente dall'omologia in uno specifico locus genomico di interesse. Sono riportati gli elementi importanti del DNA donatore HITI. SA= sequenza accettore di splicing;
 30 3XFLAG= marcatore per il rilevamento della proteina; CL1= segnale di degradazione per la proteina endogena mutata; Fu= sito di taglio ottimizzato per la furina; GSG= linker peptidico; P2A= sequenza di salto ribosomiale derivato da Tescho virus-1 2A suino; T2A= sequenza di salto

ribosomiale derivata da Thosea Asigna Virus 2A; eGFP= sequenza codificante per la proteina fluorescente verde potenziata; WPRE= elementi regolatori post-traslazionali del virus dell'epatite della marmotta; BGH polyA= segnale di poliadenilazione dell'ormone della crescita bovino; U6= cassetta di espressione U6.

- 5 Figura 2: L'integrazione del peptide CL1 aumenta l'efficienza di HITI in cellule HEK 293. (a-b) Schema dei costrutti e schema sperimentale per la valutazione del nuovo donatore HITI. Sono riportati gli elementi importanti del donatore HITI di Optm.HITI. SA= sequenza accettore di splicing; 3XFLAG= marcatore per il rilevamento della proteina; CL1= segnale di degradazione per la rodopsina mutata endogena; Fu= sito di taglio ottimizzato per la furina; GSG= linker peptidico;
- 10 P2A= sequenza di salto ribosomiale derivato da Tescho virus-1 2A suino; Rho= sequenza codificante della rodopsina; T2A= sequenza di salto ribosomiale derivato da Thosea Asigna Virus 2A; eGFP= sequenza codificante della proteina fluorescente verde potenziata; WPRE= elementi regolatori post-traslazionali del virus dell'epatite della marmotta; BGH polyA= segnale di poliadenilazione dell'ormone della crescita bovino; U6= promotore U6 umano. (c) Livelli relativi di trascritti di RHO in cellule trasfettate con il donatore Optm.HITI rispetto al donatore HITI a tre
- 15 codoni di stop tandem (3xSTOP) (IRES/Kozak). Le barre di errore indicano l'errore standard della media. Il test t di Student è stato utilizzato per valutare la significatività, $p < 0,05$.

- Figura 3: Valutazione dell'efficacia di HITI nei fotorecettori di topo. A) I topi eterozigoti sono stati iniettati per via subretinica a 4 settimane di età. Immagini di microscopia a fluorescenza di
- 20 sezioni OCT retiniche di occhi (N=4) trattati con AAV-HITI gRNA; l'occhio controlaterale è stato trattato con AAV-HITI scRNA e usato come controllo negativo. La percentuale (%) di efficienza HITI è riportata di seguito. B) Analisi dell'elettroretinogramma (ERG) eseguita 1 mese dopo il trattamento in occhi trattati con AAV-HITI gRNA e AAV-HITI scRNA iniettati a p7. Per valutare la significatività è stato utilizzato il t-test di Student, $p < 0,01$.

25 Descrizione delle sequenze

p1463_ UMANO RHO HITI Donatore+U6-Scramble RNA (pAAV2.1.InvgRNA.SAS. 3XSTOP.IRES.HRHO.T2A.EGFP.PA.U6.SCR) (Grandezza: 5'ITR to 3'ITR: 3765 bp)

Componenti (da 5' a 3')

- 30 1. RIPETIZIONI TERMINALI INVERTITE ALL'ESTREMITA' 5' (ITRs)
CTGCGCGCTCGCTCGCTCACTGAGGCCGCGCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCC
GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCT (SEQ
ID NO: 17)

- 35 2. Sequenza di AAV aggiuntiva:

TGTAGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCAC TAGT (SEQ ID NO: 18)

3. Siti di gRNA invertiti (5'-3') : ACACCAGGAGACTTGAACG (SEQ ID NO: 5)

4. **SpCas9 PAM 5'-3': CGG**

5. SEQUENZA ACCETTORE DI SPLICING :
GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGTGT (SEQ ID NO: 9)

6. **3x CODONI DI STOP : TAATAAATAATAAATAATAA** (SEQ ID NO :19)

7. SEQUENZA DI INGRESSO RIBOSOMIALE INTERNO (IRES):
TGACAAACTGTACATGCCGTTAACTGTAATTTTTCGTGATTTTTTTGTAG-(SEQ ID NO :20)

8. **Sequenza di cDNA di rodopsina umana:**

8.1 Esone 1 :

ATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCC
CTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCCTACATGTTTCT
GCTGATCGTGCTGGGCTTCCCCATCAACTTCCTCACGCTCTACGTCACCGTCCAGCACAGAAGCTGCG
CACGCCTCTCAACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTCACC
AGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGCCCCACAGGATGCAATTTGGAGGGCTTC
TTTGCCACCCTGGGCG (SEQ ID NO: 10)

8.2 Esone 2 :

GTGAAATTGCCCTGTGGTCCTTGGTGGTCCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCAT
GAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTACCTGGGTGATGGCGCTG
GCCTGCGCCGCACCCCCACTCGCCGGCTGGTCCAG (SEQ ID NO: 11)

8.3 Esone 3 :

GTACATCCCCGAGGGCCTGCAGTGCTCGTGGAATCGACTACTACACGCTCAAGCCGGAGGTCAACA
ACGAGTCTTTTGTCTATCTACATGTTTCGTGGTCCACTTCACCATCCCCATGATTATCATCTTTTCTGCTAT
GGGCAGCTCGTCTTCACCGTCAAGGAG (SEQ ID NO: 12)

8.4 Esone 4 :

GCCGCTGCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTCACCCGCATGGTC
ATCATCATGGTCATCGCTTTCCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTCACC
CACCAGGGCTCCAATTTCGGTCCCATCTTCATGACCATCCAGCGTTCTTTGCCAAGAGCGCCGCCATC
TACAACCCTGTCTATCTATATCATGATGAACAAGCAG (SEQ ID NO: 13)

8.5 Esone 5 :

TTCCGGAAGTGCATGCTCACCACCATCTGCTGCGGCAAGAACCCACTGGGTGACGATGAGGCCTCTGC
TACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCAGCA (SEQ ID NO: 14)

9. Sequenza di salto ribosomiale T2A:

GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCT
(SEQ ID NO: 2)

10. Proteina fluorescente verde potenziata (eGFP) :

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGGCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCACCCTCGTGACCACCCTGACCTACG

GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGGACTTCTTCAAGTCCGCCATGCCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGG
TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCCGACAAG
 5 CAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCG
CCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCACAACCACTACCTG
AGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGT
GACCGCCCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA (SEQ ID NO : 21)

10 11. elemento regolatorio post-trascrizionale del virus dell'epatite di marmotta (wpre):
taagcttggatccaatcaacctctggattacaaaatttgtaaagattgactggattcttaactatgttgctccttttacgctatgtgga
tacgctgctttaatgcctttgtatcatgctattgcttcccgatggctttcattttctcctccttgataaatcctgggtgctgtctctttatga
ggagttgtggccgtgttcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactgggtggggcattgccaccacct
gtcagctcctttccgggactttcgctttccccctccctattggccacggcggaactcatcgccgctgccttgcccgctgctggacagggg
 15 ctcggctgttgggactgacaattccgtggtgtgtcggggaaatcatcgtccttccctggctgctgcctgtgttgccacctggattct
gcgcgggacgtccttctgctacgtcccttcggccctcaatccagcgacgttccctcccgcgccctgctgccggctctgcggcctcttcc
gcgtcttcg (SEQ ID NO: 15)

12. Sequenza non nota/Sequenza di riempimento: AGATCT

20 13. Segnale di poliadenilazione dell'ormone della crescita bovino (BGH pA):
GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCTTGACCCTGG
AAGGTGCCACTCCCCTGTCCTTTCTAATAAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTG
TCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
 25 AGGCATGCTGGGGA (SEQ ID NO: 16)

14. Siti di gRNA invertiti(5'-3') :ACACCAGGAGACTTGGAAACG (SEQ ID NO: 5)

15. SpCas9 PAM 5'-3': CGG

30 16. Sequenza non nota/Sequenza di riempimento:
AAGGGCGATATCCATCACACTGGCGGCGAATTCCCGATTAGGAAAGGGCGAATTCTGCAGATACTAGT
AACGGCCCGCAGTGTGCTGGAATTCAGG (SEQ ID NO: 22)

35 17. Cassetta di espressione U6:
ctgacctgagtttccatgattccttcatatttgcataacgatacaaggctgttagagagataattggaattaatttgactgtaaac
acaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggttagtttgagttttaaattatgttttaaaatggact
atcatatgcttaccgtaacttgaaagtatttcgatttcttgctttatatcttgttgaaaggacgaaacaccgGACTCGCGCG
 40 AGTCGAGGAGgttttagagctagaaatagcaagttaaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcg
gtgctttttgttttagagctagaaatagcaag (SEQ ID NO:23)

18. Sequenza di RNA con nucleotidi disordinati ("scrambled") (5'-3'):
GACTCGCGGAGTCGAGGAG
 (SEQ ID NO:24)

19. Sequenza di AAV aggiuntiva:

CTCGAGCACCTGAATTCTGCAGATATCCATCACACTGGCGGCATCCATCACACTGGCGGCGAATTCCC
GATTAGGAAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCGAATTCCCGATTAGGATCTTCC
TAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAATACTACA (SEQ ID NO: 25)

5 20. ____ Sequenza a ripetizione terminale invertita 3' (3'itr)

aggaacccttagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaggtcgcccgacgccc
gggctttgcccgggcgccctcagtgagcgagcgagcgcgcgag
(SEQ ID NO: 26)

10 **SEQUENZA INTERA (dal 5' al 3') p1463_ UMANO RHO HITI Donatore+U6-
ScramblegRNA**

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCC
GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCCTTGTAG
TTAATGATTAAACCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCAGTAGTA
15 CACCAGGAGACTTGGAAACGCGGGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCA
CAGGTGTTAATAAATAATAAATAAATAATGACAACTGTACATGCCGTTAACTGTAATTTTGCCTGATTT
TTTTGTAGATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTAC
GCAGCCCCCTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCCTACA
TGTTTCTGCTGATCGTGCTGGGCTTCCCCATCAACTTCTCAGCTCTACGTACCCGTCCAGCACAAAGAA
20 GCTGCGCACGCCTCTCAACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCTAGGTGG
CTTCACCAGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGCCACAGGATGCAATTTGGAG
GGCTTCTTTGCCACCCTGGGCGGTGAAATTGCCCTGTGGTCTTGGTGGTCTGGCCATCGAGCGGTA
CGTGGTGGTGTGTAAGCCCATGAGCAACTTCCGTTTCGGGGAGAACCATGCCATCATGGGCGTTGCC
TTCACTGGGTCATGGCGCTGGCCTGCGCCGACCCCCACTCGCCGGCTGGTCCAGGTACATCCCCGA
25 GGGCCTGCAGTGCTCGTGTTGGAATCGACTACTACACGCTCAAGCCGGAGGTCAACAACGAGTCTTTTG
TCATCTACATGTTCTGTTCCACTTCACCATCCCCATGATTATCATCTTTTCTGCTATGGGCAGCTCGTC
TTCAACCGTCAAGGAGGCCGCTGCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAG
GTCACCCGCATGGTCATCATCATGGTCATCGCTTCTCTGATCTGCTGGGTGCCCTACGCCAGCGTGGC
ATTCTACATCTTCACCCACCAGGGCTCCAATTCTGGTCCCATCTTCATGACCATCCCAGCGTTCTTTGCC
30 AAGAGCGCCGCATCTACAACCCTGTCATCTATATCATGATGAACAAGCAGTTCCGGAAGTGCATGCT
CACCACCATCTGCTGCGGCAAGAACCCACTGGGTGACGATGAGGCCTCTGCTACCGTGTCCAAGACGG
AGACGAGCCAGGTGGCACCAGCAGGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGAC
GTCGAGGAGAATCTGGACCTATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGGCCATCC
TGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCGGGCGAGGGCGAGGGCGATGC
35 CACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCCACCT
CGTGACCACCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTT
CTTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC
GACTTCAAGGAGGACGGCAACATCTGGGGCACAAGCTGGAGTACAATAACAGCCACAACGTCTA
40 TATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACG
GCAGCGTGACGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC
GACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCCAACGAGAAGCGCGATCACATGGT
CCTGCTGGAGTTCGTGACCGCCCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAtaagcttgga
tccaatcaacctctggattacaaaatttggaaagattgactggtattcttaactatgttgctccttttacgctatgtggatagcgtcttta
45 atgcctttgtatcatgctattgcttcccgatggctttcattttctcctctgtataaatcctggttgctgtctctttatgaggagttgtggcc
cgttgtcaggcaacgtggcggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttcc
gggactttcgctttccccctccctattgccacggcggaactcatcgccgctgccttgcccgctgctggacaggggctcggtgttgggca
ctgacaattccgtggtgtgtcggggaaatcatcgctctttccttggtgctgcctgtgttggcacctggattctgcgcgggacgtccttct

gctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgcccgtctgcccgtcttcgAGATCTG
 CCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTCCCCCTCCCCGTGCCTTCTTGACCCTGGA
 AGGTGCCACTCCCCTGTCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTC
 ATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG
 5 GCATGCTGGGGAACACCAGGAGACTTGAACGCGGAAGGGCGATATCCATCACACTGGCGGCGAATT
 CCCGATTAGGAAAGGGCGAATTCTGCAGATACTAGTAACGGCCGCCAGTGTGCTGGAATTCAGGctgac
 ctcgagtttcccatgattccttcatatttgcataacgatacaaggctgttagagagataattggaattaattgactgtaaacacaaa
 gatattagtagacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaattatgttttaaaatggactatcata
 tgcttaccgtaacttgaaagtatttgcatttcttggccttatatatcttggtaaaggacgaaacaccgGACTCGCGCGAGTCTG
 10 AGGAGgttttagagctagaaatagcaagttaaaataaggctagtcggttatcaacttgaaaaagtggcaccgagtcggtgctttt
 tgttttagagctagaaatagcaagCTCGAGCACCTGAATTCTGCAGATATCCATCACACTGGCGGCATCCATC
 AACTGGCGGCGAATTCCCATTAGGAAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCGA
 ATTCCCGATTAGGATCTTCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAACT
 ACAaggaacccttagtgatggagtggccactccctctctgcgcgctgctgctcactgaggccggcgaccaaaggctgcccgcag
 15 cccgggctttgcccggcgccctcagtgagcgagcgagcgcgag
 (SEQ ID NO:27)

20 **p1464_ UMANO RHO HITI Donatore+U6-gRNA (Introne1)**
pAAV2.1.InvgRNA.SAS. 3XSTOP.IRES.HRHO.T2A.EGFP.PA.U6.gRNA(Introne1)
 (Grandezza: 5'ITR to 3'ITR: 3765 bp)

Componenti (da 5' a 3')

1. RIPETIZIONI TERMINALI INVERTITE ALL'ESTREMITA' 5' (ITRs)
 CTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTGGGGCGACCTTTGGTCGCCC
 25 GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCT (SEQ
 ID NO: 17)
2. Sequenze di AAV aggiuntive:
 TGTAGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCAC
 30 TAGT (SEQ ID NO: 18)
2. Siti di gRNA invertiti (5'-3') : ACACCAGGAGACTTGAACG (SEQ ID NO: 5)
4. SpCas9 PAM 5': CGG
- 35 5. SEQUENZA ACCETTORE DI SPLICING:
GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTCTCTCCACAGGTGT (SEQ ID NO: 9)
6. 3x CODONI DI STOP : TAATAAATAATAAATAATAA (SEQ ID NO :19)
- 40 7. SEQUENZA DI INGRESSO RIBOSOMIALE INTERNO (IRES):
TGACAAACTGTACATGCCGTTAACTGTAATTTTTCGTGATTTTTTTGTAG (SEQ ID NO :20)
8. Sequenza di cDNA di rodopsina umana:
 8.1 Esone1 :
 45 ATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCC
 CTTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCCTACATGTTTCT
 GCTGATCGTGCTGGGCTTCCCCATCAACTTCTCACGCTCTACGTCACCGTCCAGCACAAGAAGCTGCG

CACGCCTCTCAACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTCACC
AGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGGCCACAGGATGCAATTTGGAGGGCTTC
TTTGCCACCCTGGGCG (SEQ ID NO: 10)

8.2 Esone 2 :

5 GTGAAATTGCCCTGTGGTCTTGGTGGTCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCAT
GAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTCACCTGGGTTCATGGCGCTG
GCCTGCGCCGCACCCCCACTCGCCGGCTGGTCCAG (SEQ ID NO: 11)

8.3 Esone 3 :

10 GTACATCCCCGAGGGCCTGCAGTGCTCGTGTGGAATCGACTACTACACGCTCAAGCCGGAGGTCAACA
ACGAGTCTTTTGTATCTACATGTTTCGTGGTCCACTTCACCATCCCCATGATTATCATCTTTTCTGCTAT
GGGCAGCTCGTCTTCACCGTCAAGGAG (SEQ ID NO: 12)

8.4 Esone 4 :

15 GCCGCTGCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTACCCGCATGGTC
ATCATCATGGTCATCGCTTCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTCACC
CACCAGGGCTCCAATTTCGGTCCCATCTTCATGACCATCCAGCGTTCTTGCCAAGAGCGCCGCCATC
TACAACCCTGTCATCTATATCATGATGAACAAGCAG (SEQ ID NO: 13)

8.5 Esone 5 :

20 TTCCGGAAGTGCATGCTCACCACCATCTGCTGCGGCAAGAACCCACTGGGTGACGATGAGGCCTCTGC
TACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCAGCA (SEQ ID NO: 14)

9. Sequenza di salto ribosomiale T2A:

GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCT
(SEQ ID NO: 2)

10. Proteina fluorescente verde potenziata (eGFP) :

25 ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGGCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACCTGACCTACG
GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCC
30 GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGG
TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGCTATATCATGGCCGACAAG
CAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCG
CCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTG
35 AGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTTCGT
GACCGCCCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA (SEQ ID NO: 21)

11. elemento regolatorio post-trascrizionale del virus dell'epatite di marmotta (wpre):

40 taagcttggatccaatcaacctctggattacaaaattgtgaaagattgactggattcttaactatgttgctccttttacgctatgtga
tacgctgctttaatgcctttgtatcatgctattgcttccgctatggctttcattttctcctcctgtataaatcctgggtgctgtctctttatga
ggagttgtggccggtgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactgggtggggcattgccaccacct
gtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgctgccttgccgctgctggacagggg
ctcggctgttgggactgacaattccgtggtgtgtcggggaaatcatcgtcctttccttggtgctgcctgtgttgccacctggattct
gcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaaccttcttcccggcctgctgcccggctctgcggcctcttcc
45 gcgtcttcg (SEQ ID NO: 15)

12. Sequenza non nota/sequenza di riempimento: AGATCT

13. Segnale di poliadenilazione dell'ormone della crescita bovino (BGH pA):

GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCGTGCCTTCCTTGACCCTGG
AAGGTGCCACTCCCCTGTCCTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTG
TCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGA (SEQ ID NO: 16)

14. RIPETIZIONI TERMINALI INVERTITE ALL'ESTREMITA' 5' (ITRs):

ACACCAGGAGACTTGGAAACG (SEQ ID NO: 5)

15. SpCas9 PAM 5'-3': CGG**16. Sequenza non nota/sequenza di riempimento:**

AAGGGCGATATCCATCACACTGGCGGCGAATTCCCATTAGGAAAGGGCGAATTCTGCAGATACTAGT
AACGGCCGCCAGTGTGCTGGAATTCAGG (SEQ ID NO: 22)

17. Cassetta di espressione U6:

ctgacctcgagtttccatgattcctcatattgcatatacgatacaaggctgtagagagataattggaattaattgactgtaaac
acaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggttagtttcagttttaaaattatgttttaaatggact
atcatatgcttaccgtaacttgaaagtatttcgatttcttgctttatatatcttcttggaaggacgaaacaccgACACCAGGAG
ACTTGGAACGgttttagagctagaaatagcaagttaaaataaggctagtcggttatcaactgaaaaagtggcaccgagtcgg
tgctttttgttttagagctagaaatagcaag (SEQ ID NO:28)

18. Siti di gRNA 5' - 3') :ACACCAGGAGACTTGGAAACG (SEQ ID NO: 5)**19. Sequenza di AAV aggiuntiva:**

CTCGAGCACCTGAATTCTGCAGATATCCATCACACTGGCGGCATCCATCACACTGGCGGCGAATTCC
CGATTAGGAAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCGAATTCCCATAAGGATCTT
CCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAATACTACA (SEQ ID NO: 29)

20. Sequenza a ripetizione terminale invertita 3' (3'itr)

Aggaaccctagtgatggagttggccactccctctctgcgctcgctcgctcactgaggccgggcgaccaaggtcgcccgagcgc
cgggctttgcccgggcccctcagtgagcgagcgagcgcgag (SEQ ID NO: 26)

SEQUENZA INTERA (dal 5' al 3')

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCAGGCGTCGGGCGACCTTTGGTCGCCC
GGCCTCAGTGAGCGAGCGAGCGCGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCT
TGTAGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCAT
AGTACACCAGGAGACTTGGAAACGCGG
GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTCTCTCCACAGGTGTTAATAAATAATAAAT
AATAATGACAACTGTACATGCCGTTAACTGTAATTTTTCGTGATTTTTTTGTAG
ATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCC
CTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCTACATGTTTCT
GCTGATCGTGCTGGGCTTCCCCATCACTTCTCAGCTCTACGTCACCGTCCAGCACAGAAGCTGCG
CACGCCTCTCACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCTAGGTGGCTTCACC
AGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGCCACAGGATGCAATTTGGAGGGCTTC
TTTGCCACCCTGGGCGGTGAAATTGCCCTGTGGTCTTGGTGGTCTGGCCATCGAGCGGTACGTGGT
GGTGTGTAAGCCCATGAGCAACTTCCGCTTCGGGGGAGAACCATGCCATCATGGGCGTTGCCTTCACCT

GGGTCATGGCGCTGGCCTGCGCCGACCCCCACTCGCCGGCTGGTCCAGGTACATCCCCGAGGGCCT
GCAGTGCTCGTGTGGAATCGACTACTACACGCTCAAGCCGGAGGTCAACAACGAGTCTTTTGTATCTA
CATGTTCTGGTCCACTTCACCATCCCCATGATTATCATCTTTTTCTGCTATGGGCAGCTCGTCTTACCG
TCAAGGAGGGCGCTGCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTACCC
5 **GCATGGTCATCATCATGGTCATCGCTTTCCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTAC**
ATCTTACCCACCAGGGCTCCAACCTTCGGTCCCATCTTCATGACCATCCCAGCGTCTTTGCCAAGAGC
GCCGCCATCTACAACCCTGTCATCTATATCATGATGAACAAGCAG
TTCCGGAACTGCATGCTCACCACCATCTGCTGCGGCAAGAACCCACTGGGTGACGATGAGGCCTCTGC
TACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCAGCAGGAAGCGGAGAGGGCAGAGGAAGT
10 **CTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCTATGGTGAGCAAGGGCGAGGAGCTGTTT**
ACCGGGGTGGTGCCATCCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCT
GCCCGTGGCCTGGCCACCCTCGTGACCACCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCT
15 **TCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAA**
CCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC
AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAGTTCAG
GATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCGACCACTACCAGCAGAACCCCCATCG
GCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCC
20 **AACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGG**
ACGAGCTGTACAAGTAA**taagcttggatccaatcaacctctggattacaaaatttgtaaagattgactggtattcttaactat**
gttgctcctttacgctatgttgatacgtgctttaatgcctttgtatcatgctattgcttcccgtatggctttcattttctcctcctgtataa
atcctggttgctgtctctttatgaggagttgtggccgtgtgcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccact
ggttgggcattgccaccacctgtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgctgcc
25 **ttgccgctgctggacaggggctcgctgttgggactgacaattccgtggtgtgtcggggaaatcatcgctcctttccttggtgctcg**
cctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaaccttcttcccgcggcctgct
gccgctctgcggccttctccgctcttcgAGATCT
GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCTTGACCCTGG
AAGGTGCCACTCCCCTGTCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTG
30 **TCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC**
AGGCATGCTGGGGAACACCAGGAGACTTGAACG**CGGA****AAGGGCGATATCCATCACACTGGCGGCGA**
ATTCCCGATTAGGAAAGGGCGAATTCTGCAGATACTAGTAACGGCCGCCAGTGTGCTGGAATTCAGG
ctgacctcgagtttccatgattccttcattttgcataacgatacaaggctgttagagagataattggaattaatttgactgtaa
acaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggttagtttgagttttaaataattggtttaaataatggact
35 **atcatatgcttaccgttaacttgaaagtatttcgatttcttggtttatatatcttggtaaaggacgaaacaccg****ACACCAGGAG**
ACTTGGAACG**gttttagagctagaaatagcaagttaaataaaggctagtcggttatcaacttgaaaaagtggcaccgagtcgg**
tgctttttgttttagagctagaaatagcaagCTCGAGCACTGAATTCTGCAGATATCCATCACACTGGCGGCA
TCCATCACACTGGCGGCGAATTCCCGATTAGGAAAGGGCGAATTCTGCAGATATCCATCACACTGGC
GGCGAATTCCCGATAAGGATCTTCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATC
40 **ATTA****ACTACA****aggaaccctagtgatggagttggcactccctctctgcgcgctcgctcgctcactgaggccggcgaccaaagg**
tcgcccagcgccgggctttgcccggcgccctcagtgagcgagcgagcgcgag
 (SEQ ID NO:30)

1465_ UMANO RHO HITI Donatore+U6-gRNA

(Grandezza: 5'ITR to 3'ITR: 3659 bp)

(pAAV2.1.InvRNA.SAS.KOZAK. HRHO.T2A.EGFP.PA.U6. GRNA)

Componenti (dal 5' al3')

1. RIPETIZIONI TERMINALI INVERTITE AL 5' (ITRs)

CTGCGCGCTCGCTCGCTCACTGAGGCCGCGGCGGCAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCC
GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATCACTAGGGGTTCTCT
(SEQ ID NO: 17)

2. Sequenza di AAV aggiuntiva:

TGTAGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCAC
TAGT (SEQ ID NO: 18)

3. Siti di gRNA invertiti (5'-3') : ACACCAGGAGACTTGGAAACG (SEQ ID NO: 5)

4. SpCas9 PAM 5': CGG

5. SEQUENZA ACCETTORE DI SPLICING :

GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTCTCTCCACAGGTGT (SEQ ID NO: 9)

6. 3x CODONI DI STOP : TAATAAATAATAAATAATAA (SEQ ID NO :19)

7. ----- KOZAK: GCCACC

8. Sequenza di cDNA di rodopsina umana:

8.1 Esone 1 :

ATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCCC
CTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCCTACATGTTTCT
GCTGATCGTGCTGGGCTTCCCCATCAACTTCTCACGCTCTACGTCACCGTCCAGCACAAGAAGCTGCG
CACGCCTCTCAACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTCACC
AGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGCCCCACAGGATGCAATTTGGAGGGCTTC
TTTGCCACCCTGGGCG (SEQ ID NO: 10)

8.2 Esone 2 :

GTGAAATTGCCCTGTGGTCCTTGGTGGTCCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCAT
GAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTCACCTGGGTGATGGCGCTG
GCCTGCGCCGCACCCCCACTCGCCGGCTGGTCCAG (SEQ ID NO: 11)

8.3 Esone 3 :

GTACATCCCCGAGGGCCTGCAGTGCTCGTGGAATCGACTACTACGCTCAAGCCGGAGGTCAACA
ACGAGTCTTTTGTATCTACATGTTTCGTGGTCCACTTCACCATCCCCATGATTATCATCTTTTCTGCTAT
GGGCAGCTCGTCTTCACCGTCAAGGAG (SEQ ID NO: 12)

8.4 Esone 4 :

GCCGCTGCCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTACCCGCATGGTC
ATCATCATGGTCATCGCTTCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTCACC
CACCAGGGCTCCAATTCCGGTCCCATCTTCATGACCATCCCAGCGTTCTTTGCCAAGAGCGCCGCCATC
TACAACCCTGTCATCTATATCATGATGAACAAGCAG (SEQ ID NO: 13)

8.5 Esone 5 :

TTCCGGAAGTGCATGCTCACCACCATCTGCTGCGGCAAGAACCCACTGGGTGACGATGAGGCCTCTGC
TACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCAGCA (SEQ ID NO: 14)

9. Sequenza di salto ribosomiale T2A :

GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCT
(SEQ ID NO: 2)

10. Proteina fluorescente verde potenziata (eGFP):

5 ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGGCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCACCCTCGTGACCACCCTGACCTACG
GCGTGCACTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGG
10 TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCCGACAAG
CAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCG
CCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTG
AGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGT
15 GACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA (SEQ ID NO: 21)

11. elemento regolatorio post-trascrizionale del virus dell'epatite di marmotta (wpre):

20 taagcttgatccaatcaacctctggattacaaaattgtgaaagattgactggatttctaactatgttgctcctttacgctatgtgat
acgctgctttaatgcctttgtatcatgctattgcttccgtagtgctttcattttctcctctgtataaatcctgggtgctgtctttatgagg
agttgtggccgtttgacaggcaacgtggcgtggtgtgactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtc
agctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttggccgctgctggacaggggctcg
gctgttgggcactgacaattccgtggtgtgtgctgggaaatcatcgcttcttcttggtgctgcctgtgttggcacctggattctgcgcg
ggacgtccttctgctacgtcccttcggccctcaatccagcggaccttccctcccgcgctgctgccggctgctggccttctccgcgtctt
cg (SEQ ID NO: 15)

12. Sequenza non nota/sequenza di riempimento: AGATCT

13. Segnale di poliadenilazione dell'ormone della crescita bovino (BGH pA):

30 GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCGTGCCTTCCTTGACCCTGG
AAGGTGCCACTCCCAGTGCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTG
TCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGA (SEQ ID NO: 16)

14. Siti di gRNA invertiti (5'-3'): ACACCAGGAGACTTGAACG (SEQ ID NO: 5)

15. SpCas9 PAM 5'-3': CGG

16. Sequenza non nota/ sequenza di riempimento:

ACTAGTAACGGCCGCCAGTGTGCTGGAATTCAGG (SEQ ID NO: 65)

17. Cassetta di espressione U6:

45 ctgacctcgagtttccatgattccttcatatttgcataacgatacaaggctgttagagagataattggaattaatttgactgtaaac
acaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggttagtttgacgttttaaaattatgttttaaaatggact
atcatatgcttacgttaacttgaaagtatttgcatttcttggtttatatacttgttgaaaggacgaaacaccgACACCAGGAG
ACTTGAAACGgttttagagctagaaatagcaagttaaaataaggctagtccttatcaacttgaaaaagtggcaccgagtcgg
tgctttttgttttagagctagaaatagcaag (SEQ ID NO: 28)

18. sequenza di gRNA diretta all'introne 1 (5'-3'):

ACACCAGGAGACTTGGAAACG (SEQ ID NO: 5)

19. Sequenza di AAV aggiuntiva:

CTCGAGCACCTGAATTCTGCAGATATCCATCACACTGGCGGCATCCATCACACTGGCGGCGAATTCC
CGATTAGGAAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCGAATTCGATAAGGATCTT
CCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAATACTACA (SEQ ID NO: 29)

20. ___ sequenza a ripetizione invertita terminale 3' (3'itr)

Aggaaccctagtgtgaggttgccactccctctctgcgctcgctcgctcactgaggccgggacgacaaaggtcgcccgacgccc
gggctttgcccgggcggcctcagttagcgagcgagcgcgagcag (SEQ ID NO: 26)

Sequenza completa 5-3'p1465_ UMANO RHO HITI Donatore+U6-gRNA

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTGGGCGACCTTTGGTCGCCC
GGCCTCAGTGAGCGAGCGAGCGCGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTT
TGTAGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCAC
TAGTACACCAGGAGACTTGGAAACGCGGGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTC
TCTCCACAGGTGTTAATAAATAATAAATAAAGCCACCATGAATGGCACAGAAGGCCCTAACTTCTA
CGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCCCTTCGAGTACCCACAGTACTACCTGGCTGA
GCCATGGCAGTTCTCCATGCTGGCCGCCTACATGTTTCTGCTGATCGTGCTGGGCTTCCCCATCAACTTC
CTCAGCTCTACGTACCGTCCAGCACAAGAAGCTGCGCACGCCTCTCAACTACATCCTGCTCAACCTA
GCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTACCAGCACCTCTACACCTCTCTGCATGGATACT
TCGTCTTCGGGCCCCACAGGATGCAATTTGGAGGGCTTCTTTGCCACCCTGGGCG
GTGAAATTGCCCTGTGGTCTTGGTGGTCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCAT
GAGCAACTTCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTACCTGGGTGATGGCGCTG
GCCTGCGCCGACCCCCACTCGCCGGCTGGTCCAGGTACATCCCCGAGGGCCTGCAGTGCTCGTGTGG
AATCGACTACTACAGCTCAAGCCGGAGGTCAACAACGAGTCTTTTGTATCTACATGTTCTGTGGTCCA
CTTCACCATCCCCATGATTATCATTTTTTCTGCTATGGGCGAGCTCGTCTTCACCGTCAAGGAGGCCGCT
GCCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTACCCGCATGGTCATCATC
ATGGTCATCGCTTTCCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTCACCCACCAG
GGCTCCAACCTTCGGTCCCATCTTCATGACCATCCAGCGTTCTTTGCCAAGAGCGCCGCCATCTACAAC
CCTGTCATCTATATCATGATGAACAAGCAGTTCCGGAAGTGCATGCTCACCACCATCTGCTGCGGCAA
GAACCCACTGGGTGACGATGAGGCCTCTGCTACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCA
GCAGGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTGAGGAGAATCCTGGACC
TATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACCTGACCTACG
GCGTGAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCG
AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTG
AAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAATAACAACAGCCACAACGTCTATATCATGGCCGACAAGCAG
AAGAACGGCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCATGGTCTGCTGGAGTTCTGTGACCG
CCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAtaagcttgatccaatcaacctctggattacaaaat
tgtgaaagattgactggattcttaactatgttgctcctttacgctatgtggatacgtgctttaatgcctttgtatcatgctattgcttc
cgtatggctttcattttctccttgataaaatcctggttgctgtctctttatgaggagttgtggccggttgctcaggcaacgtggcggtgtgt
gcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttccccctccctatt

gccacggcggaactcatcgccgcctgccttgccgctgctggacaggggctcggtgttgggcaactgacaattccgtggtgttcggg
gaaatcatcgctcttcttggctgctgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatcc
agcggaccttccctcccgccgctgctgcggcctctgcggccttccgcgtcttcgAGATCTGCCTCGACTGTGCCTTCTAG
TTGCCAGCCATCTGTTGTTTGGCCCTCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGTC
CTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAACACCAG
GAGACTTGGAACGCGG
ACTAGTAACGGCCGCCAGTGTGCTGGAATTCAGGctgacctcgagtttcccatgattccttcatatttgcataatcgata
caaggctgttagagagataattggaattaattgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatt
tcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttat
atatcttggtaaaggacgaaacaccgACACCAGGAGACTTGGAACGgttttagagctagaaatagcaagttaaaataag
gctagtccgttatcaacttgaaaaagtggcaccgagtcgggtgctttttgttttagagctagaaatagcaagCTCGAGCACCTGA
ATTCTGCAGATATCCATCACACTGGCGGCATCCATCACACTGGCGGCGAATTCCCGATTAGGAAAGG
GCGAATTCTGCAGATATCCATCACACTGGCGGCGAATTCCCGATAAGGATCTTCTAGAGCATGGCT
ACGTAGATAAGTAGCATGGCGGGTTAATCATTAATACTACAaggaaccctagtgtatggagttggccactccctctctg
cgcgctcgctcgctcactgaggccggcgaccaaaggtcgcccgacggcggttggccggcgccctcagtgagcgagcgagcg
cgag
 (SEQ ID NO:62)

1466_ UMANO RHO HITI Donatore+U6-scrambleRNA
(Grandezza: 5'ITR to 3'ITR: 3659 bp)
(pAAV2.1.InvgrRNA.SAS.KOZAK. HRHO.T2A.EGFP.PA.U6. SCRAMBLE)

Componenti (dal 5' al 3')

1. RIPETIZIONI TERMINALI INVERTITE AL 5' (ITRs)

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTGGGGCGACCTTTGGTCGCCC
 GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCT (SEQ
 ID NO: 17)

2. Sequenze di AAV aggiuntive:

TGTAGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCAC
TAGT (SEQ ID NO: 18)

3. Siti di gRNA invertiti (5'-3') : ACACCAGGAGACTTGGAACG (SEQ ID NO: 5)

4. SpCas9 PAM 5': CGG

5. SEQUENZA ACCETTORE DI SPLICING :

GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGTGT (SEQ ID NO: 9)

6. 3x CODONI DI STOP : TAATAAATAATAAATAATAA (SEQ ID NO :19)

7. KOZAK: GCCACC

8. Sequenza di cDNA di rodopsina umana:

8.1 Esone1:

ATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCC
CTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCCTACATGTTTCT

GCTGATCGTGCTGGGCTTCCCCATCAACTTCCTCACGCTCTACGTCACCGTCCAGCACAAGAAGCTGCG
CACGCCTCTCAACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTCACC
AGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGGCCACAGGATGCAATTTGGAGGGCTTC
TTTGCCACCCTGGGCG (SEQ ID NO: 10)

5 **8.2 Esone 2:**

GTGAAATTGCCCTGTGGTCCTTGGTGGTCCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCAT
GAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTCACCTGGGTTCATGGCGCTG
GCCTGCGCCGCACCCCCACTCGCCGGCTGGTCCAG (SEQ ID NO: 11)

8.3 Esone 3 :

10 GTACATCCCCGAGGGCCTGCAGTGCTCGTGGAATCGACTACTACACGCTCAAGCCGGAGGTCAACA
ACGAGTCTTTTGCATCTACATGTTCTGTGGTCCACTTCACCATCCCCATGATTATCATCTTTTCTGCTAT
GGGCAGCTCGTCTTACCGTCAAGGAG (SEQ ID NO: 12)

8.4 Esone 4 :

15 GCCGCTGCCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTCACCCGCATGGTC
ATCATCATGGTCATCGCTTCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTCACC
CACCAGGGCTCCAATTCCGTCCCATCTTCATGACCATCCAGCGTTCTTTGCCAAGAGCGCCGCCATC
TACAACCCTGTCATCTATATCATGATGAACAAGCAG (SEQ ID NO: 13)

8.5 Esone 5 :

20 TTCCGGAAGTGCATGCTCACCACCATCTGCTGCGGCAAGAACCCTACTGGGTGACGATGAGGCCTCTGC
TACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCAGCA (SEQ ID NO: 14)

9. Sequenza di salto ribosomiale T2A:

GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCT
(SEQ ID NO: 2)

25

10. Proteina fluorescente verde potenziata (eGFP):

30 ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGGCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACCGGCAAGCTGCCCCTGGCCACCCTCGTGACCACCTGACCTACG
GCGTGCACTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCC
GAAGGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGG
TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCATATCATGGCCGACAAG
CAGAAGAACGGCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCG
35 CCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTG
AGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTCTGCTGGAGTTCGT
GACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA (SEQ ID NO: 21)

11. elemento regolatorio post-trascrizionale del virus dell'epatite di marmotta (wpre):

40 taagcttgatccaatcaacctctggattacaaaattgtgaaagattgactggattcttaactatgttgctcctttacgctatgtga
tacgctgctttaatgcctttgtatcatgctattgcttccgctatggctttcattttctcctcctgtataaatcctgggtgctgtctctttatga
ggagttgtggccgtttgtcaggcaacgtggcgtggtgtgactgtgtttgctgacgcaacccccactgggtggggcattgccaccacct
gtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgctgccttgcccgctgctggacagggg
ctcggctgttgggactgacaattccgtggtgtgtcggggaaatcatgctcctttccttggctgctcgctgtgttgccacctggattct
45 gcgcgggacgtccttctgctacgtcccttcggccctcaatccagcgaccttcttcccgcgccctgctgcccgtctgcccgtctgcccctcttc
gcgtcttcg (SEQ ID NO: 15)

12. Sequenza non nota/sequenza di riempimento: AGATCT

13. Segnale di poliadenilazione dell'ormone della crescita bovino (BGH pA):

GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCTGCCTTCCTTGACCCTGG
AAGGTGCCACTCCCCTGCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTG
TCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGA (SEQ ID NO: 16)

14. Siti di gRNA invertiti (5'-3'): ACACCAGGAGACTTGGAACG (SEQ ID NO: 5)

15. SpCas9 PAM 5'-3': CGG

16. Sequenza non nota/sequenza di riempimento:
ACTAGTAACGGCCGCCAGTGTGCTGGAATTCAGG (SEQ ID NO: 65)

17. Cassetta di espressione U6:

ctgacctcgagtttcccatgattccttcatatttgcataacgatacaaggctgtagagagataattggaattaatttgactgtaaac
acaaagatattagtacaaaatacgtgacgtagaaagtaataatttctggtagtttgagttttaaaattatgttttaaatggact
atcatatgcttaccgttaacttgaaagtatttcgatttcttgctttatatacttggtaaaggacgaaacaccg**GACTCGCGCG**
AGTCGAGGAG**g**tttttagagctagaaatagcaagttaaaataaggctagtcggttatcaacttgaaaaagtgccaccgagtcg
gtgctttttgttttagagctagaaatagcaag (SEQ ID NO:23)

18. sequenza di RNA disordinata (5'-3'):

GACTCGCGCGAGTCGAGGAG (SEQ ID NO:24)

19. Sequenza di AAV aggiuntiva:

CTCGAGCACCTGAATTCTGCAGATATCCATCACACTGGCGGCATCCATCACACTGGCGGCGAATTCC
CGATTAGGAAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCGAATTCGATAAGGATCTT
CCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTA ACTACA (SEQ ID NO: 29)

20. sequenza di ripetizione terminale invertita 3' (3'itr):

aggaacccttagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggctgcccgcgcccg
ggctttgcccgggcgccctcagtgagcgagcgagcgcgag (SEQ ID NO: 26)

Sequenza completa (5'-3'):

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCC
GGCCTCAGTGAGCGAGCGAGCGCGCAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTCT
TGTAGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCAC
TAGTACACCAGGAGACTTGGAACG**CGG**GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTC
TCTCCACAGGTGTTAATAAATAATAAATAAAGCCACCATGAATGGCACAGAAGGCCCTAACTTCTA
CGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCCCTTCGAGTACCCACAGTACTACCTGGCTGA
GCCATGGCAGTTCTCATGCTGGCCGCCTACATGTTTCTGCTGATCGTGCTGGGCTTCCCCATCAACTTC
CTCACGCTCTACGTACCGTCCAGCACAAGAAGCTGCGCACGCCTCTCAACTACATCCTGCTCAACCTA
GCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTACCAGCACCTCTACACCTCTCTGCATGGATACT
TCGTCTTCGGGCCACAGGATGCAATTTGGAGGGCTTCTTGCCACCCTGGGCG
GTGAAATTGCCCTGTGGTCCTTGGTGGTCCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCAT
GAGCAACTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTACCTGGGTATGGCGCTG

GCCTGCGCCGCACCCCCACTCGCCGGCTGGTCCAGGTACATCCCCGAGGGCCTGCAGTGCTCGTGTGG
 AATCGACTACTACACGCTCAAGCCGAGGTCAACAACGAGTCTTTTGTCTATCTACATGTTCTGTGGTCCA
 CTTACCATCCCCATGATTATCATCTTTTTCTGCTATGGGCAGCTCGTCTTACCGTCAAGGAGGCCGCT
 5 **GCCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTACCCGCGATGGTCATCATC**
ATGGTCATCGCTTTCCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTACCCACCAG
GGCTCCAACCTTCGGTCCCATCTTCATGACCATCCCAGCGTCTTTGCCAAGAGCGCCGCCATCTACAAC
CCTGTCATCTATATCATGATGAACAAGCAGTTCGGGAAGTGCATGCTCACCACCATCTGCTGCGGCAA
 GAACCCACTGGGTGACGATGAGGCCTCTGCTACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCA
 GCAGGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCTGGACC
 10 **TATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCTGGTCGAGCTGGACGGCGAC**
GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACCGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACCTGACCTACG
GCGTGAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCG
AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTG
 15 **AAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA**
ACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAG
AAGAACGGCATCAAGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCG
 20 **CCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTA****taagcttggatccaatcaacctctggattacaaaat**
ttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgtgctttaatgcctttgtatcatgctattgcttcc
cgtatggcctttcattttctcctcctgtataaatcctggttgctgtctctttatgaggagtgttgcccggtgtgcaggcaacgtggcgtggtgt
gcactgtgtttgctgacgcaacccccactggttggggcattgccaccacgtgcagctcctttccgggactttcgctttccccctccctatt
gccacggcggaactcatcgccgctgccttgccgctgctggacaggggctcggtgttggtgactgacaattccgtggtgtgtcggtg
 25 **gaaatcatcgtcctttccttggtgctgctgcctgtgttgccacctggaattctgcgcgggacgtccttctgctacgtcccttcggccctcaatcc**
agcggaccttccctcccgcgccgtgctgcggcctctgcggccttccgcttctcg**AGATCTGCCTCGACTGTGCCTTCTAG**
TTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCTTCTTACCTGGAAGGTGCCACTCCCACTGTC
CTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAACACCAG
 30 **GAGACTTGGAACGCGG**
ACTAGTAACGGCCGCCAGTGTGCTGGAATTCAGG**ctgacctcgagtttcccatgattccttcatatttgcataatcgata**
caaggctgttagagagataattggaattaattgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatt
tcttgggtagtttcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttgctttat
atatcttgttgaaaggacgaaacaccg**GACTCGCGCAGTCGAGGAG****gttttagagctagaaatagcaagttaaaataag**
 35 **gctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttttgttttagagctagaaatagcaag**
CTCGAGCACCTGAATTCTGCAGATATCCATCACACTGGCGGCATCCATCACACTGGCGGCGAATTCC
CGATTAGGAAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCGAATTCCCGATAAGGATCTT
CCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAACTACA**aggaaccctagtgatg**
gagttggccactccctctctgcgcgctcgctcgctcactgaggccggcgaccaaaggctgcccgcgcccgggctttgccggggcgg
 40 **cctcagtgagcgagcgagcgcgag**
 (SEQ ID NO: 31)

p1501_Ottimizzato RHO HITI Donatore+U6-gRNA

45 (pAAV2.1.InvRNA.SAS.3xFLAG.CL1.Fu.GSG.P2A. HRHO.T2A.EGFP.PA.U6. GRNA)

(Grandezza: 5'ITR to 3'ITR: 3908 bp)

Componenti (dal 5' al 3')

1. RIPETIZIONI TERMINALI INVERTITE ALL'ESTREMITA' 5' (ITRs)

CTGCGCGCTCGCTCGCTACTGAGGCCGCGGGCAAAGCCCGGGCGTCTGGGCGACCTTTGGTCGCCC
GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCATCTAGGGGTTCT (SEQ
ID NO: 17)

2. Sequenza aggiuntiva di AAV:

TGTAGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCAGTAGTCA
ATTGGCGGCCGC (SEQ ID NO: 64)

3. Siti di gRNA invertiti (5') : ACACCAGGAGACTTGGAACG (SEQ ID NO: 5)

4. SpCas9 PAM 5': CGG5. SEQUENZA ACCETTORE DI SPLICING :

GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGTGT (SEQ ID NO: 9)

6. 3XFLAG:

AGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATA
AG (SEQ ID NO:33)

7. sequenza di degradazione cl1: gcctgcaagaactggttcagcagcctgagccacttcgtgatccacctg
(SEQ ID NO: 6)

SITO DI TAGLIO DELLA FURINA OTTIMIZZATO:CGAAAAAGAAGA (SEQ ID NO: 7)

8. sequenza linker gsg: ggaagcgga

9. sequenza di salto ribosomiale p2a: gccaccaacttctcctgtctgaagcaggccggcgacgtggaggagaacccccggcccc
(SEQ ID NO: 1)

10. Sequenza di cDNA di rodopsina umana: :

10.1 Esone1 :

ATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCC
CTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCCTACATGTTTCT
GCTGATCGTGCTGGGCTTCCCCATCAACTTCCTCAGCTCTACGTCACCGTCCAGCACAAGAAGCTGCG
CACGCCTCTCAACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTCACC
AGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGCCCCACAGGATGCAATTTGGAGGGCTTC
TTTGCCACCCTGGGCG (SEQ ID NO: 10)

10.2 Esone 2 :

GTGAAATTGCCCTGTGGTCTTGGTGGTCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCAT
GAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTCACCTGGGTTCATGGCGCTG
GCCTGCGCCGCACCCCCACTCGCCGGCTGGTCCAG (SEQ ID NO: 11)

10.3 Esone 3 :

GTACATCCCCGAGGGCCTGCAGTGCTCGTGTGGAATCGACTACTACACGCTCAAGCCGGAGGTCAACA
ACGAGTCTTTTGTCTATCATGTTTCGTGGTCCACTTCACCATCCCCATGATTATCATCTTTTCTGCTAT
GGGCAGCTCGTCTTCACCGTCAAGGAG (SEQ ID NO: 12)

10.4 Esone 4 :

GCCGCTGCCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTCACCCGCATGGTC
ATCATCATGGTCATCGCTTTCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTCACC
CACCAGGGCTCCAACCTTCGGTCCCATCTTCATGACCATCCCAGCGTTCTTTGCCAAGAGCGCCGCCATC
TACAACCCTGTCTATCTATATCATGATGAACAAGCAG (SEQ ID NO: 13)

5 10.5 Esone 5:

TTCCGGAAGTGCATGCTCACCACCATCTGCTGCGGCAAGAACCCTGAGGCTGACGATGAGGCCTCTGC
TACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCAGCA (SEQ ID NO: 14)

11. Sequenza di salto ribosomiale T2A:

10 GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCT
(SEQ ID NO: 2)

12. Proteina fluorescente verde potenziata (eGFP):

15 ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGTGGTGGCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCACCCTCGTGACCACCCTGACCTACG
GCGTGCACTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGG
TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
20 GCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCCGACAAG
CAGAAGAACGGCATCAAGGTGAAGTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCG
CCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTG
AGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTTCGT
GACCGCCCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA (SEQ ID NO: 21)

25 **13. elemento regolatorio post-trascrizionale del virus dell'epatite di marmotta (wpre):**
Taagcttgatccaatcaacctctggattacaaaatttgtaaagattgactgtattcttaactatgttgctccttttacgctatgtgga
tacgctgctttaatgcctttgtatcatgctattgcttcccgatggctttcattttctcctcctgtataaatcctgggtgctgtctctttatga
ggagttgtggccgtgtgcaggcaacgtggcgtgtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacct
30 gtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacagggg
ctcggtgttgggactgacaattccgtggtgtgtcggggaaatcatcgcttctcctggctgctgcctgtgttgccacctggattct
gcgcgggacgtccttctgctacgtcccttcggccctcaatccagcgaccttcttcccgcgccctgctgcccgtctgcggcctcttc
gcgtcttcg (SEQ ID NO: 15)

35 **14. Sequenza non nota/sequenza di riempimento: AGATCT**

15. Segnale di poliadenilazione dell'ormone della crescita bovino Bovine (BGH pA):

40 GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCTGG
AAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTG
TCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGA (SEQ ID NO: 16)

16. Siti di gRNA invertiti (5'-3'): ACACCAGGAGACTTGGAACG (SEQ ID NO: 5)

45 **17. SpCas9 PAM 5'-3': CGG**

18. Sequenza non nota/sequenza di riempimento:

AAGGGCGATATCCATCACACTGGCGGCGAATTCCCGATTAGGAAAGGGCGAATTCTGCAGATGGTACCCTAGTAACGGCCGCCAGTGTGCTGGAATTCAGG (SEQ ID NO: 63)

19. Cassetta di espressione U6

ctgacctcgagtttcccatgattccttcatatttgcataacgatacaaggctgtagagagataattggaattaatttgactgtaaaca
aaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggttagtttgagttttaaaattatgttttaaaatggact
atcatatgcttaccgtaacttgaaagtatttgcatttcttgctttatatatcttgtgaaaggacgaaacaccgACACCAGGAG
ACTTGGAACGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcgg
tgctttttgttttagagctagaaatagcaag (SEQ ID NO: 28)

20. sequenza di gRNA diretta all'introne 1 (dall'estremità 5' all'estremità 3'):

ACACCAGGAGACTTGGAACG (SEQ ID NO: 5)

21. Sequenza di AAV aggiuntiva:

CTCGAGCACCTGAATTCTGCAGATATCCATCACACTGGCGGCATCCATCACACTGGCGGCGAATTCCCGATTAGGAAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCGAATTCCCGATAAGGATCTTCCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAATACTACA (SEQ ID NO: 29)

22. sequenza terminale a ripetizione invertita 3' (3'itr):

aggaacccttagtgatggagttggccactccctctctgcgcgctcgtcgtcactgaggccgggcgaccaaaggtcgcccgacgccc
gggctttgccgggcgccctcagtgagcgagcgagcgcgag (SEQ ID NO: 26)

**SEQUENZA COMPLETA (5'-3') p1501_Ottimizzata RHO HITI Donatore+U6-gRNA
(pAAV2.1.InvgRNA.SAS.3xFLAG.CL1.Fu.GSG.P2A. HRHO.T2A.EGFP.PA.U6. GRNA)
(Grandezza: 5'ITR to 3'ITR: 3908 bp)**

CTGCGCGCTCGCTCGCTCACTGAGGCCGCGCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTCTGTAGTGAATGATTAAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCACTAGTCAATTGGCGGCCGCACACCAGGAGACTTGGAACGCGGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGTGTAGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGgcttgaagaactggttcagcagcctgagccacttcgtgatccacctgCGAAAAAGAAGAggaagcggagaccaccaacttctccctgctgaagcaggccggcgacgtggaggagaaacccggccc
cATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCCTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCTACATGTTCTGCTGATCGTGCTGGGCTTCCCCATCACTTCTCACGCTCTACGTACCGTCCAGCACAGAAGCTGCGACGCCTCTCACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCTAGGTGGCTTCA
CCAGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGCCACAGGATGCAATTTGGAGGGCTTCTTTGCCACCCTGGGCGGTGAAATTGCCCTGTGGTCTTGGTGGTCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCATGAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTT
CACCTGGGTCTATGGCGCTGGCCTGCGCCGACCCCCACTCGCCGGCTGGTCCAGGTACATCCCCGAGGGCCTGCAGTGCTCGTGTGGAATCGACTACTACAGCTCAAGCCGGAGGTCAACAACGAGTCTTTGT
CATCTACATGTTCTGTGGTCCACTTACCATCCCCATGATTATCATCTTTTCTGCTATGGGCAGCTCGTCTTACCGTCAAGGAGGCCGCTGCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTCACCCGCATGGTCATCATGTCATCGCTTTCCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTACCCACCAGGGCTCCAACCTCGGTCCCATCTTCATGACCATCCAGCGTCTTT

GCCAAGAGCGCCGCGCATCTACAACCCTGTCATCTATATCATGATGAACAAGCAGTTCGGGAAGTGCAT
 TGCTCACCACCATCTGCTGCGGCAAGAACCCACTGGGTGACGATGAGGCCTCTGCTACCGTGTCCAAG
 ACGGAGACGAGCCAGGTGGCACCAGCAGGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCG
 GTGACGTCGAGGAGAATCCTGGACCTATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTG
 5 CCCATCCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
 GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTG
 GCCCCACCTCGTGACCACCTGACCTACGGCGTGACGTGCTTCAGCCGCTACCCCGACCACATGAAGC
 AGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC
 GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
 10 TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAG
 CCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAGTTCAGATCCGCCACA
 ACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCC
 CGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGC
 GCGATCATATGGTCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTAC
 15 AAGTAAtaagccttggatccaatcaacctctggattacaaaattgtgaaagattgactggtattcttaactatgttgctccttttacgc
 tatgtggatacgctgctttaatgcctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgataaatcctggttgctgtct
 ctttatgaggagtgttgcccggtgtgcaggcaacgtggcggtgtgactgtgttgctgacgcaacccccactggtggggcattgcc
 accacctgtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgctgcttgcccgctgctgga
 caggggctcggctgttgggcactgacaattccgtggtgtgtcggggaaatcatcgctcctttccttggtgctgcctgtgttgccacctg
 20 gattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaccttcttcccgcgccctgctgcccggctctgcggcct
 ctcccgctcttcgAGATCTGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGT
 GCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTTCTAATAAAATGAGGAAATTGCATCGC
 ATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGGAGGAT
 TGGGAAGACAATAGCAGGCATGCTGGGGAACACCAGGAGACTTGAACGCGGAAGGGCGATATCC
 25 ATCACACTGGCGGCGAATTCCTGATTAGGAAAGGGCGAATTCCTGCAGATGGTACCACTAGTAACGG
 CCGCCAGTGTGCTGGAATTCAGGctgacctcgagtttcccatgattccttcatatttgcataacgatacaaggctgttag
 agagataattggaattaatttgaactgtaaacacaaagatattagtagcaaaatacgtgacgtagaaagtaataatttcttggttagt
 ttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtgaaagtatttcgatttcttggtttatatacttgtg
 gaaaggacgaaacaccgACACCAGGAGACTTGAACGgtttttagagctagaaatagcaagttaaaataaggctagtccg
 30 ttatcaacttgaaaaagtggcaccgagtcggtgctttttgttttagagctagaaatagcaagCTCGAGCACCTGAATTCTG
 CAGATATCCATCACACTGGCGGCATCCATCACACTGGCGGCGAATTCCCGATTAGGAAAGGGCGAA
 TTCTGCAGATATCCATCACACTGGCGGCGAATTCCCGATAAGGATCTTCTAGAGCATGGCTACGTA
 GATAAGTAGCATGGCGGGTTAATCATTAACTACAaggaaccctagtgtgaggttgccactccctctctgcgcgc
 tcgctcgctcactgaggccggcgaccaaaggctgcccgcgcccgggctttgcccggcgccctcagttagcgagcgagcgcgca
 35 g
 (SEQ ID NO:32)

p1503_Ottimizzato RHO HITI Donatore+U6-Scramble gRNA
 40 (pAAV2.1.InvRNA.SAS.3xFLAG.CL1.Fu.GSG.P2A. HRHO.T2A.EGFP.PA.U6. ScrgRNA)
 (Grandezza: 5'ITR a 3'ITR: 3908 bp)

Componenti (dal 5' al 3')

45 1. RIPETIZIONI TERMINALI INVERTITE ALL'ESTREMITA' 5' (ITRs)
 CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCCGGGCAAAGCCCCGGGCGTGGGGCGACCTTTGGTCGCCC
 GGCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCT (SEQ
 ID NO: 17)

2. Sequenze aggiuntive di AAV:

TGTAGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCAC
TAGTCAATTGGCGGCCGC (SEQ ID NO: 64)

3. Siti di gRNA invertiti (5') : ACACCAGGAGACTTGAACG (SEQ ID NO: 5)

4. SpCas9 PAM 5': CGG

5. SEQUENZA ACCETTORE DI SPLICING :

GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTCTCTCCACAGGTGT (SEQ ID NO: 9)

6. 3XFLAG:

AGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATA
AG (SEQ ID NO:33)

7. segnale di degradazione cl1 gcctgcaagaactggttcagcagcctgagccacttcgtgatccacctg (SEQ ID NO: 6)

SITO DI TAGLIO DELLA FURINA OTTIMIZZATO:

CGAAAAAGAAGA (SEQ ID NO: 7)

8. sequenza linker gsg: ggaagcgga

9. sequenza di salto ribosomiale p2a:

gccaccaacttctccctgctgaagcaggccggcgacgtggaggagaacccccggcccc (SEQ ID NO: 1)

10. Sequenza di cDNA di rodopsina umana :

10.1 Esone 1 :

ATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCC
CTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCCTACATGTTTCT
GCTGATCGTGCTGGGCTTCCCCATCAACTTCCTCAGCTCTACGTCACCGTCCAGCACAAGAAGCTGCG
CACGCCTCTCAACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTACC
AGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGCCCCACAGGATGCAATTTGGAGGGCTTC
TTTGCCACCCTGGGCG (SEQ ID NO: 10)

10.2 Esone 2 :

GTGAAATTGCCCTGTGGTCCTTGGTGGTCCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCAT
GAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTCACCTGGGTCATGGCGCTG
GCCTGCGCCGCACCCCCACTCGCCGGCTGGTCCAG (SEQ ID NO: 11)

10.3 Esone 3 :

GTACATCCCCGAGGGCCTGCAGTGCTCGTGTGGAATCGACTACTACGCTCAAGCCGGAGGTCAACA
ACGAGTCTTTTGTCTATCATGTTTCGTGGTCCACTTCACCATCCCCATGATTATCATCTTTTCTGCTAT
GGGCAGCTCGTCTTCACCGTCAAGGAG (SEQ ID NO: 12)

10.4 Esone 4 :

GCCGCTGCCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTACCCGCATGGTC
ATCATCATGGTCATCGCTTCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTCACC
CACCAGGGCTCCAATTTCGGTCCCATCTTCATGACCATCCAGCGTCTTGGCAAGAGCGCCGCCATC
TACAACCCTGTCATCTATATCATGATGAACAAGCAG (SEQ ID NO: 13)

10.5 Esone 5 :

TTCCGGAAGTGCATGCTCACCACCATCTGCTGCGGCAAGAACCCACTGGGTGACGATGAGGCCTCTGC
TACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCAGCA (SEQ ID NO: 14)

11. Sequenza di salto ribosomiale T2A:

GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCT
 (SEQ ID NO: 2)

12. Proteina fluorescente verde potenziata (eGFP) :

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGTGGTGGCCATCCTGGTCGAGCTGGACGGCGAC
GTAACCGGCCACAAGTTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCACCCTCGTGACCACCCTGACCTACG
GCGTGCACTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGG
TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCATATCATGGCCGACAAG
CAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCG
CCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCCACAACCACTACCTG
AGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTCTGCTGGAGTTCGT
GACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA (SEQ ID NO: 21)

13. elemento regolatorio post-trascrizionale del virus dell'epatite di marmotta (wpre):

Taagcttgatccaatcaacctctggattacaaaatttgtaaagattgactgtattcttaactatgttgctcctttacgctatgtgga
tacgctgctttaatgcctttgtatcatgctattgcttcccgatggctttcattttctcctccttgataaatcctggttgctgtctctttatga
ggagttgtggccggtgtcaggcaacgtggcggtgtgactgtgtttgctgacgcaacccccactggttggggcattgccaccacct
gtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgctgcttgcccgctgtggacagggg
ctcggtgttgggactgacaattccgtggtgtgtcggggaaatcatgctcctttccttggtgctgctcgctgtgttgccacctggattct
gcgcgggagctccttctgctacgtcccttcggccctcaatccagcgacgttcttcccgcgctgctgcccgtctgcccgtctgcccctctcc
gcgtcttcg (SEQ ID NO: 15)

14. Sequenza non nota/sequenza di riempimento: AGATCT

15. Segnale di poliadenilazione dell'ormone della crescita bovino (BGH pA):

GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGG
AAGGTGCCACTCCCACTGTCCTTCTAATAAAAATGAGGAAATTGCATCGATTGTCTGAGTAGGTG
TCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGA (SEQ ID NO: 16)

16. Siti di gRNA invertiti (5'-3') :ACACCAGGAGACTTGGAACG (SEQ ID NO: 5)

17. SpCas9 PAM 5'-3': CGG

18. Sequenza non nota/sequenza di riempimento:

AAGGGCGATATCCATCAGCTGGCGGCGAATTCGATTAGGAAAGGGCGAATTCTGCAGATGGTA
CCACTAGTAACGGCCGCCAGTGTGCTGGAATTCAGG (SEQ ID NO: 63)

19. Cassetta di espressione U6

ctgacctgagtttccatgattccttcatttgcataacgatacaaggctgttagagagataattggaattaatttgactgtaaac
acaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggttagtttgacgttttaaaattatgttttaaaatggact

atcatatgcttaccgtaacttgaaagtatttcgatttcttgctttatatatcttgtggaaggacgaaacaccgGACTCGCGCG
AGTCGAGGAGgttttagagctagaaatagcaagttaaaataaggctagtcggttatcaacttgaaaaagtggcaccgagtcg
gtgctttttgttttagagctagaaatagcaag (SEQ ID NO: 23)

- 5 20. sequenza di RNA disordinata (5'-3'):
GACTCGCGCGAGTCGAGGAG (SEQ ID NO: 24)

21. Sequenza di AAV aggiuntiva:
10 CTCGAGCACCTGAATTCTGCAGATATCCATCACACTGGCGGCATCCATCACACTGGCGGCGAATTCC
CGATTAGGAAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCGAATCCCGATAAGGATCTT
CCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTTAATCATTAACTACA (SEQ ID NO: 29)

22. sequenza terminale a ripetizione invertita 3' (3'itr) :
15 Aggaacccttagtgatggagttggccactccctctctgcgcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcc
cgggctttgcccgggcgccctcagtgagcgagcgagcgcgag (SEQ ID NO: 26)

Sequenza completa (dal 5' al 3') p1503_Ottimizzata RHO HITI Donatore+U6-Scramble gRNA

- 20 CTGCGCGCTCGCTCGCTCACTGAGGCCGCGCGGGCAAAGCCCCGGGCGTGGGCGACCTTTGGTCGCC
CGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCT
TGTAAGTTAATGATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCA
CTAGTCAATTGGCGGCCGCACACCAGGAGACTTGGAAACGCGGATAGGCACCTATTGGTCTTACTGA
CATCCACTTTGCCTTTCTCTCCACAGGTGTAGACTATAAGGACACGACGGAGACTACAAGGATCAT
25 GATATTGATTACAAAGACGATGACGATAAGgcctgcaagaactggttcagcagcctgagccacttcgtgatccacct
gCGAAAAAGAAGAggaagcggagccaccaacttctccctgctgaagcagggcggcagctggaggaggaaccccgcccc
cATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCC
CCTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCTACATGTTTC
TGCTGATCGTGCTGGGCTTCCCCATCAACTTCTCACGCTCTACGTCACCGTCCAGCACAAGAAGCTGC
GCACGCCTCTCAACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTCA
30 CCAGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGCCACAGGATGCAATTTGGAGGGC
TTCTTTGCCACCCTGGGCGGTGAAATTGCCCTGTGGTCTTGGTGGTCTGGCCATCGAGCGGTACG
TGGTGGTGTGTAAGCCCATGAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTT
CACCTGGGTCTATGGCGCTGGCCTGCGCCGACCCCCACTCGCCGGCTGGTCCAGGTACATCCCCGAG
GGCCTGCAGTGCTCGTGTGGAATCGACTACTACACGCTCAAGCCGGAGGTCAACAACGAGTCTTTTGT
35 CATCTACATGTTTCGTGGTCCACTTCACCATCCCCATGATTATCATCTTTTCTGCTATGGGCAGCTCGTC
TTCACCGTCAAGGAGGCCGCTGCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGA
GGTCACCCGCATGGTCATCATCATGGTCATCGCTTCTGATCTGCTGGGTGCCCTACGCCAGCGTG
GCATTCTACATCTTCACCCACCAGGGCTCCAACCTTCGGTCCCATCTTCATGACCATCCAGCGTTCTTT
GCCAAGAGCGCCGCATCTACAACCCTGTCTATATCATGATGAACAAGCAGTTCCGGAACTGCA
40 TGCTCACCACCATCTGCTGCGGCAAGAACCACTGGGTGACGATGAGGCCTCTGCTACCGTGTCCAAG
ACGGAGACGAGCCAGGTGGCACCAGCAGGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCG
GTGACGTCGAGGAGAATCCTGGACCTATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTG
CCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTG
45 GCCACCCTCGTGACCACCCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACATGAAGC
AGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAG

CCACAACGTCATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACA
 ACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCC
 CGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGC
 GCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTAC
 5 AAGTAA
 taagcttggatccaatcaacctctggattacaaaattgtgaaagattgactggatttcttaactatgttgctccttttacgctatgtgga
 tacgctgctttaatgcctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgataaaatcctgggtgctgtctctttatga
 ggagttgtggccggttgcaggcaacgtggcgtggtgtgactgtgtttgctgacgcaacccccactggttggggcattgccaccacct
 gtcagctcctttccgggactttcgtttccccctccctattgccacggcggaactcatcgccgcctgccttgcccgtgctggacagggg
 10 ctgggctgttgggactgacaattccgtggtgttgcggggaaatcatcgtccttccctggctgctcgctgtgttgccacctggattct
 gcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaccttcttcccgcggcctgctgcggcctctgcggcctcttcc
 gcgtcttcgAGATCTGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCTGCCT
 TCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTG
 TCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG
 15 GAAGACAATAGCAGGCATGCTGGGGAACACCAGGAGACTTGGAACGCGGAAGGGCGATATCCATC
 AACTGGCGGCGAATTCCCGATTAGGAAAGGGCGAATTCTGCAGATGGTACCACTAGTAACGGCCG
 CCAGTGTGCTGGAATTCAGGctgacctcgagtttccatgattcctcatattgcatatacgatacaaggctgtagagag
 ataattggaattaatttgactgtaaacacaaagatatagtagacaaatacgtgacgtagaaagtaataatttctgggtagtttga
 gttttaaaattatgttttaaatggactatcatatgcttaccgtaactgaaagtatttcgatttctggctttatatatcttgggaaa
 20 ggacgaaacaccgGACTCGCGCGAGTCGAGGAGgttttagagctagaaatagcaagttaaaataaggctagtcggttat
 caactgaaaaaagtgccagcagtggtgctttttgttttagagctagaaatagcaagCTCGAGCACCTGAATTCTGCAG
 ATATCCATCACACTGGCGGCATCCATCACACTGGCGGCGAATTCCCGATTAGGAAAGGGCGAATTCT
 GCAGATATCCATCACACTGGCGGCGAATTCCCGATAAGGATCTTCTAGAGCATGGCTACGTAGAT
 AAGTAGCATGGCGGGTTAATCATTAACTACAaggaacccttagtgatggagtggccactccctctctgcgctgct
 25 cgctcactgaggccggcgaccaaaggtcgcccgacgcccgggctttgcccgggcgccctcagtgagcgagcgagcgcgag
 (SEQ ID NO:34)

Definizioni

I termini "comprendente", "comprende" e "comprendente di" utilizzati nel presente documento
 30 sono sinonimi di "includente" o "include"; o "contenente" o "contiene", e sono inclusivi o aperti
 e non escludono membri, elementi o fasi aggiuntivi, non citati. I termini "comprendente",
 "comprende" e "comprendente di" includono anche il termine "costituito da".

Nella presente invenzione "identità di almeno l'80%" significa che l'identità può essere di
 almeno l'80%, o l'85% o il 90% o il 95% o il 100% rispetto alle sequenze di riferimento. Questo
 35 vale per tutte le % di identità menzionate. Nella presente invenzione, "identità di almeno il 95%"
 significa che l'identità può essere di almeno il 95%, 96%, 97%, 98%, 99% o 100% rispetto alle
 sequenze di riferimento. Questo vale per tutte le % di identità menzionate. Nella presente
 invenzione, "identità di almeno il 98%" significa che l'identità può essere di almeno il 98%, 99%
 o 100% rispetto alle sequenze di riferimento. Questo vale per tutte le % di identità menzionate.
 40 Preferibilmente, la % di identità si riferisce all'intera lunghezza della sequenza di riferimento.

Rientrano nella presente invenzione anche le sequenze di acido nucleico derivate dalle sequenze nucleotidiche citate, ad esempio frammenti funzionali, mutanti, varianti, derivati, analoghi e sequenze con una % di identità di almeno l'80% con le sequenze citate, nella misura in cui tali frammenti, mutanti, varianti, derivati e analoghi mantengono la funzione della sequenza da cui derivano.

Il termine "funzionale" è inteso come che mantiene la funzione della sequenza da cui derivano. I termini "sistema di editing genetico" e "sistema di editing del genoma" sono equivalenti.

Sequenze di DNA esogene

Le sequenze di DNA esogeno sopra menzionate comprendono un frammento di DNA da incorporare nel DNA genomico di un genoma bersaglio. In alcune forme realizzative, il DNA esogeno comprende almeno una porzione di un gene. Il DNA esogeno può comprendere una sequenza codificante, ad esempio un cDNA relativo a un gene di tipo selvatico o a una sequenza "codone ottimizzata" del fattore che deve essere espresso. In alcuni casi, il DNA esogeno comprende almeno un esone di un gene e/o almeno un introne di un gene. In alcune forme realizzative, il DNA esogeno comprende un elemento potenziatore o un elemento promotore di un gene. In alcune forme realizzative, il DNA esogeno comprende una sequenza discontinua di un gene che comprende una porzione 5' del gene fusa con la porzione 3' del gene. In alcune forme realizzative, il DNA esogeno comprende una sequenza genica di tipo selvatico. In alcune forme realizzative, il DNA esogeno comprende una sequenza genica mutata. In alcune forme realizzative, il DNA esogeno comprende una sequenza genica di tipo selvatico. In alcune forme realizzative, la sequenza di DNA esogeno comprende un gene reporter. In alcune forme realizzative, il gene reporter è selezionato tra almeno uno di: proteina fluorescente verde (GFP), proteina fluorescente rossa (RFP), luciferasi, β -galattosidasi e β -glucuronidasi. In alcune forme realizzative, la sequenza di DNA esogeno comprende un elemento regolatore della trascrizione genica che può comprendere, ad esempio, una sequenza promotrice o una sequenza potenziatrice. In alcune forme realizzative, la sequenza di DNA esogeno comprende uno o più esoni o loro frammenti. In alcune forme realizzative, la sequenza di DNA esogeno comprende uno o più introni o frammenti di essi. In alcune forme realizzative, la sequenza di DNA esogeno comprende almeno una parte della regione non tradotta all'estremità 3' o della regione non tradotta all'estremità 5'. In alcune forme realizzative, la sequenza di DNA esogeno comprende una sequenza di DNA artificiale. In alcune forme realizzative, la sequenza di DNA esogeno comprende una sequenza di localizzazione nucleare e/o una sequenza di esportazione nucleare.

Una sequenza di DNA esogeno, in alcune forme realizzative, comprende un segmento di acido nucleico da integrare in un locus genomico bersaglio. La sequenza di DNA esogeno, in alcune forme realizzative, comprende uno o più polinucleotidi di interesse. In alcune forme realizzative, la sequenza di DNA esogeno comprende una o più cassette di espressione. Tale cassetta di espressione, in alcune forme realizzative, comprende una sequenza di DNA esogeno di interesse, un polinucleotide che codifica un marcatore di selezione e/o un gene reporter e componenti regolatori che influenzano l'espressione. La sequenza di DNA esogeno, in alcune forme realizzative, comprende un acido nucleico genomico. L'acido nucleico genomico deriva da un animale, un topo, un uomo, un non umano, un roditore, un non umano, un ratto, un criceto, un coniglio, un maiale, un bovino, un cervo, una pecora, una capra, un pollo, un gatto, un cane, un furetto, un primate (ad esempio, uistiti, scimmia rhesus), un mammifero domestico o un mammifero agricolo, un'aviana, un batterio, un archeo, un virus o qualsiasi altro organismo di interesse o una loro combinazione. Le sequenze di DNA esogene di qualsiasi dimensione sono integrate in un genoma bersaglio. In alcune forme realizzative, la sequenza di DNA esogeno integrata in un genoma è inferiore a 3, circa 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5, 10, 10,5, 11, 11,5, 12, 12,5, 13, 13,5, 14, 14,5, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 o più di 500 kilobasi (kb) di lunghezza. In alcune forme di realizzazione, la sequenza di DNA esogeno integrata in un genoma ha una lunghezza di almeno 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5, 10, 10,5, 11, 11,5, 12, 12,5, 13, 13,5, 14, 14,5, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 o più di 500 (kb) di lunghezza.

Sequenze di bersaglio

In alcune forme realizzative, il costrutto di bersaglio (che comprende l'acido nucleico donatore affiancato all'estremità 5' e 3' dalle sequenze di bersaglio invertite) comprende almeno due sequenze di bersaglio. Le sequenze di bersaglio sono sequenze di acido nucleico riconosciute e tagliate da una nucleasi. In alcune forme realizzative, la sequenza di bersaglio ha una lunghezza compresa tra circa 9 e circa 12 nucleotidi, tra circa 12 e circa 18 nucleotidi, tra circa 18 e circa 21 nucleotidi, tra circa 21 e circa 40 nucleotidi, tra circa 40 e circa 80 nucleotidi o una qualsiasi combinazione dei sottointervalli (ad esempio, 9-18, 9-21, 9-40 e 9-80 nucleotidi). In alcuni casi, la sequenza di bersaglio comprende un sito di legame per la nucleasi. In alcune forme realizzative, la sequenza di bersaglio comprende un sito di rottura/taglio. In alcuni casi, la sequenza di bersaglio comprende una sequenza di un motivo adiacente al protospaziatore

(PAM). In alcuni casi, la sequenza di acido nucleico bersaglio (ad esempio, il protospaziatore) è di 20 nucleotidi. In alcune forme realizzative, l'acido nucleico bersaglio è inferiore a 20 nucleotidi. In alcuni casi, l'acido nucleico bersaglio è di almeno 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 o più nucleotidi. L'acido nucleico bersaglio, in alcune forme di realizzazione, è al massimo di 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 o più nucleotidi. In alcune forme realizzative, la sequenza di acido nucleico bersaglio è di 16, 17, 18, 19, 20, 21, 22 o 23 basi immediatamente all'estremità 5' del primo nucleotide del PAM. In alcune forme realizzative, la sequenza di acido nucleico bersaglio è 16, 17, 18, 19, 20, 21, 22 o 23 basi poste immediatamente al 3' dell'ultimo nucleotide del PAM. In alcune forme realizzative, la sequenza di acido nucleico bersaglio è 20 basi immediatamente all'estremità 5' del primo nucleotide del PAM. In alcune forme realizzative, la sequenza di acido nucleico bersaglio è 20 basi immediatamente all'estremità 3' dell'ultimo nucleotide del PAM. In alcune forme realizzative, la sequenza di acido nucleico bersaglio si trova all'estremità 5' o all'estremità 3' della PAM. Una sequenza di bersaglio, in alcune forme realizzative, comprende sequenze di acidi nucleici presenti in un acido nucleico bersaglio a cui si lega un segmento di acido nucleico bersaglio di un acido nucleico del filamento complementare. Ad esempio, le sequenze di bersaglio, in alcune forme realizzative, comprendono sequenze a cui un acido nucleico del filamento complementare è progettato per avere un appaiamento di basi. In alcune forme realizzative, una sequenza bersaglio comprende qualsiasi polinucleotide che si trova, ad esempio, nel nucleo o nel citoplasma di una cellula o all'interno di un organello di una cellula, come un mitocondrio o un cloroplasto. Le sequenze di bersaglio prevedono siti di taglio per le nucleasi. In alcuni casi, una sequenza bersaglio è adiacente a siti di taglio per nucleasi. La nucleasi taglia l'acido nucleico, in alcune forme realizzative, in un sito all'interno o all'esterno della sequenza di acido nucleico presente nell'acido nucleico bersaglio a cui si lega la sequenza bersaglio dell'acido nucleico del filamento complementare. Il sito di taglio, in alcune forme realizzative, comprende la posizione di un acido nucleico in cui una nucleasi produce una rottura a singolo filamento o a doppio filamento. Ad esempio, la formazione di un complesso di nucleasi comprendente un acido nucleico a filamento complementare appaiato con una sequenza di riconoscimento della proteasi e complessato con una proteasi determina il taglio di uno o entrambi i filamenti all'interno o in prossimità (ad esempio, entro 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 19, 20, 23, 50 o più coppie di basi) della sequenza di acido nucleico presente in un acido nucleico bersaglio a cui si lega una regione spaziatrice di un acido nucleico a filamento complementare. Il sito di taglio, in alcune forme realizzative, si trova

su un solo filamento o su entrambi i filamenti di un acido nucleico. In alcune forme realizzative, i siti di taglio si trovano nella stessa posizione su entrambi i filamenti dell'acido nucleico (producendo estremità smussate) o in siti diversi su ciascun filamento (producendo estremità sfalsate). Le estremità sfalsate, in alcune forme realizzative, sono ad estremità adesive con protrusioni all'estremità 5' o 3'. Le estremità sfalsate, in alcune forme realizzative, sono prodotte da nucleasi che producono estremità adesive (ad esempio, Cpf I). In alcuni casi, le estremità sfalsate sono prodotte, ad esempio, utilizzando due nucleasi, ciascuna delle quali produce una rottura a singolo filamento in un sito di taglio diverso su ciascun filamento, producendo così una rottura a doppio filamento. Ad esempio, una prima nucleasi crea una rottura a singolo filamento sul primo filamento di DNA a doppio filamento (dsDNA) e una seconda nucleasi crea una rottura a singolo filamento sul secondo filamento di dsDNA in modo da creare sequenze smussate. In alcune forme realizzative, la sequenza di riconoscimento della nucleasi sul primo filamento è separata dalla sequenza di riconoscimento della nucleasi sul secondo filamento da almeno 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500 o 1000 coppie di basi. Il taglio sito-specifico di un acido nucleico bersaglio da parte di una nucleasi, in alcune forme realizzative, avviene in punti determinati dalla complementarità di appaiamento delle basi tra l'acido nucleico del filamento complementare e l'acido nucleico bersaglio. Il taglio sito-specifico di un acido nucleico bersaglio da parte di una proteina nucleasica, in alcune forme realizzative, avviene in punti determinati da un breve motivo, chiamato motivo adiacente al protospaziatore (PAM), nell'acido nucleico bersaglio. Ad esempio, il PAM affianca la sequenza di riconoscimento della nucleasi all'estremità 3' della sequenza di riconoscimento. Ad esempio, il sito di taglio della nucleasi, in alcune forme realizzative, si trova da circa 1 a circa 25, o da circa 2 a circa 5, o da circa 19 a circa 23 coppie di basi (ad esempio, 3 coppie di basi) a monte o a valle della sequenza PAM. In alcuni casi, il sito di taglio della nucleasi è 3 coppie di basi a monte della sequenza PAM.

In alcune forme realizzative, il sito di taglio della nucleasi è di 19 basi sul filamento (+) e di 23 basi sul filamento (-), producendo una protrusione dell'estremità 5' lunga 5 nucleotidi (nt). In alcuni casi, il taglio produce estremità smussate. In alcuni casi, il taglio produce estremità sfalsate o adesive con protrusioni all'estremità 5'. In alcuni casi, il taglio produce estremità sfalsate o adesive con protrusioni di estremità 3'. Gli orologi di varie nucleasi utilizzano sequenze PAM diverse. Ad esempio, diverse proteine Cas, in alcuni casi, riconoscono sequenze PAM diverse. Ad esempio, in *S. pyogenes*, la PAM è una sequenza nell'acido nucleico bersaglio

che comprende la sequenza 5'- XRR-3', dove R è A o G, dove X è un qualsiasi nucleotide e X si trova immediatamente all'estremità 3' della sequenza di acido nucleico bersaglio mirata dalla sequenza spaziatrice. La sequenza PAM di *S. pyogenes* Cas9 (SpyCas9) è 5'- XGG-3', dove X è un qualsiasi nucleotide del DNA ed è immediatamente all'estremità 3' della sequenza di riconoscimento della nucleasi del filamento non complementare del DNA bersaglio. Il PAM di Cpf I è 5'-TTX-3', dove X è un qualsiasi nucleotide del DNA ed è immediatamente all'estremità 5' della sequenza di riconoscimento della nucleasi. Preferibilmente, il complesso Cas9/sgRNA introduce DSB 3 coppie di basi a monte della sequenza PAM nella sequenza genomica bersaglio, dando luogo a due estremità smussate. La stessa sequenza bersaglio di Cas9/sgRNA viene caricata sul DNA donatore nella direzione inversa. I loci genomici bersaglio, così come il DNA donatore, vengono tagliati da Cas9/gRNA e i DNA donatori linearizzati vengono integrati nei siti bersaglio attraverso la via di riparazione DSB NHEJ. Se il DNA donatore è integrato nell'orientamento corretto, le sequenze di giunzione sono protette da un ulteriore taglio da parte di Cas9/gRNA. Se il DNA del donatore si integra con l'orientamento inverso, Cas9/gRNA elimina il DNA del donatore integrato a causa della presenza di siti bersaglio intatti di Cas9/gRNA.

Acidi nucleici a filamento complementare (definiti anche oligonucleotidi complementari)

Un acido nucleico a filamento complementare, ad esempio un oligonucleotide a filamento complementare o un RNA a filamento complementare, si riferisce a un acido nucleico che ibrida con un altro acido nucleico, ad esempio l'acido nucleico bersaglio nel genoma di una cellula. Un acido nucleico a filamento complementare può essere, ad esempio, RNA o DNA. Un acido nucleico a filamento complementare, in alcune forme realizzative, comprende un analogo del nucleotide e/o un nucleotide modificato. L'acido nucleico a filamento complementare, in alcune forme realizzative, è disegnato o progettato per legarsi a una sequenza di acido nucleico in modo sito-specifico. Un acido nucleico a filamento complementare, in alcune forme realizzative, comprende una o più modifiche per fornire all'acido nucleico una caratteristica nuova o migliorata. In alcune forme realizzative, un acido nucleico a filamento complementare comprende un marcatore di affinità dell'acido nucleico e/o un nucleotide sintetico, un analogo del nucleotide sintetico, derivati del nucleotide e/o nucleotidi modificati. L'acido nucleico a filamento complementare, in alcune forme realizzative, comprende una sequenza nucleotidica (ad esempio, uno spaziatore), ad esempio, all'estremità 5' o all'estremità 3' o in prossimità di esse, che ibrida con una sequenza in un acido nucleico bersaglio. In alcune forme realizzative,

lo spaziatore di un acido nucleico a filamento complementare interagisce con un acido nucleico bersaglio in un modo specifico per la sequenza attraverso l'ibridazione (cioè l'appaiamento delle basi). In alcune forme realizzative, la sequenza spaziatrice ibrida con un acido nucleico bersaglio (ad esempio, la sequenza protospaziatrice) che si trova all'estremità 5' o 3' del motivo adiacente al protospaziatore (PAM). In alcune forme realizzative, un acido nucleico a filamento complementare comprende due molecole di acido nucleico separate, definite acido nucleico a doppio filamento complementare. In alcune forme realizzative, un acido nucleico a filamento complementare comprende una singola molecola di acido nucleico, definita acido nucleico a singolo filamento complementare. In alcuni casi, l'acido nucleico a filamento complementare è un singolo acido nucleico a filamento complementare che comprende un crRNA. In alcune forme realizzative, l'acido nucleico a filamento complementare è un acido nucleico a filamento complementare singolo che comprende un costrutto fuso. La regione di bersaglio dell'acido nucleico di un acido nucleico a filamento complementare, in alcune forme realizzative, comprende una sequenza nucleotidica complementare a una sequenza di un acido nucleico bersaglio. La regione di bersaglio dell'acido nucleico, in alcune forme realizzative, comprende la regione dello spaziatore. La sequenza nucleotidica di una regione spaziatrice varia e determina la posizione all'interno dell'acido nucleico bersaglio con cui interagisce l'acido nucleico del filamento complementare. La regione spaziatrice di un acido nucleico a filamento complementare, in alcune forme realizzative, è modificata per ibridarsi con qualsiasi sequenza desiderata all'interno di un acido nucleico bersaglio.

La complementarità è alternativamente perfetta o sostanziale/sufficiente. La complementarità perfetta tra due acidi nucleici significa che i due acidi nucleici formano un duplex in cui ciascuna base del duplex è legata a una base complementare mediante appaiamento di Watson-Crick. Complementarità sostanziale o sufficiente significa che una sequenza in un filamento non è completamente e/o perfettamente complementare a una sequenza in un filamento opposto, ma che si verifica un legame sufficiente tra le basi sui due filamenti per formare un complesso ibrido stabile in un insieme di condizioni di ibridazione (per esempio, concentrazione di sale e temperatura). Tali condizioni possono essere previste utilizzando le sequenze e calcoli matematici standard per prevedere il T_m dei filamenti ibridati, oppure determinando empiricamente il T_m utilizzando metodi di routine. In alcune forme realizzative, la regione bersaglio dell'acido nucleico di un acido nucleico a filamento complementare (ad esempio, la regione dello spaziatore) ha una lunghezza compresa tra 18 e 72 nucleotidi. La regione bersaglio

dell'acido nucleico di un acido nucleico del filamento complementare (ad esempio, la regione dello spaziatore) ha una lunghezza compresa tra circa 12 nucleotidi e circa 100 nucleotidi. Ad esempio, la regione bersaglio dell'acido nucleico di un acido nucleico a filamento complementare (ad es, regione spaziatrice) ha una lunghezza da circa 12 nucleotidi (nt) a circa 80 nt, da circa 12 nt a circa 50 nt, da circa 12 nt a circa 40 nt, da circa 12 nt a circa 30 nt, da circa 12 nt a circa 25 nt, da circa 12 nt a circa 20 nt, da circa 12 nt a circa 19 nt, da circa 12 nt a circa 18 nt, da circa 12 nt a circa 17 nt, da circa 12 nt a circa 16 nt o da circa 12 nt a circa 15 nt. In alternativa, il segmento di DNA targeting ha una lunghezza compresa tra circa 18 nt e circa 20 nt, tra circa 18 nt e circa 25 nt, tra circa 18 nt e circa 30 nt, tra circa 18 nt e circa 35 nt, tra circa 18 nt e circa 40 nt, da circa 18 nt a circa 45 nt, da circa 18 nt a circa 50 nt, da circa 18 nt a circa 60 nt, da circa 18 nt a circa 70 nt, da circa 18 nt a circa 80 nt, da circa 18 nt a circa 90 nt, da circa 18 nt a circa 100 nt, da circa 20 nt a circa 25 nt, da circa 20 nt a circa 30 nt, da circa 20 nt a circa 35 nt, da circa 20 nt a circa 40 nt, da circa 20 nt a circa 45 nt, da circa 20 nt a circa 50 nt, da circa 20 nt a circa 60 nt, da circa 20 nt a circa 70 nt, da circa 20 nt a circa 80 nt, da circa 20 nt a circa 90 nt o da circa 20 nt a circa 100 nt.

In alcune forme di realizzazione, la regione bersaglio dell'acido nucleico di un acido nucleico a filamento complementare (ad esempio, la regione dello spaziatore) è lunga 20 nucleotidi. In alcune forme realizzative, la regione bersaglio dell'acido nucleico di un acido nucleico a filamento complementare (ad esempio, la regione dello spaziatore) è lunga 19 nucleotidi. In alcune forme realizzative, la regione di bersaglio dell'acido nucleico di un acido nucleico a filamento complementare (ad esempio, la regione dello spaziatore) è lunga 18 nucleotidi. In alcune forme di realizzazione, la regione bersaglio dell'acido nucleico di un acido nucleico a filamento complementare (ad esempio, la regione dello spaziatore) è lunga 17 nucleotidi. In alcune forme realizzative, la regione di bersaglio dell'acido nucleico di un acido nucleico a filamento complementare (ad esempio, la regione dello spaziatore) è lunga 16 nucleotidi. In alcune forme di realizzazione, la regione bersaglio dell'acido nucleico di un acido nucleico a filamento complementare (ad esempio, la regione dello spaziatore) è lunga 21 nucleotidi. In alcune forme realizzative, la regione bersaglio dell'acido nucleico di un acido nucleico a filamento complementare (ad esempio, la regione dello spaziatore) è lunga 22 nucleotidi. Una sequenza protospaziatrice, in alcune forme di realizzazione, viene identificata individuando una PAM all'interno di una regione di interesse e selezionando una regione di una dimensione desiderata a monte o a valle della PAM come protospaziatrice. Una sequenza spaziatrice

corrispondente viene progettata determinando la sequenza complementare della regione protospaziatrice. In alcune forme di realizzazione, la sequenza spaziatrice viene identificata mediante un programma informatico (ad esempio, un codice di lettura automatica). Il programma informatico, in alcune forme realizzative, utilizza parametri quali la temperatura di fusione prevista, la formazione della struttura secondaria e la temperatura di appaiamento prevista, l'identità della sequenza, il contesto genomico, l'accessibilità alla cromatina, la % di GC, la frequenza di occorrenza genomica, lo stato di metilazione, la presenza di S Ps e simili. La percentuale di complementarità tra la sequenza di bersaglio dell'acido nucleico (ad esempio, la sequenza dello spaziatore) e la sequenza di riconoscimento della nucleasi all'interno dell'acido nucleico bersaglio (ad esempio, il protospaziatore), in alcune forme realizzative, è almeno del 60%, almeno del 70%, almeno del 75%, almeno dell'80%, almeno dell'85%, almeno del 90%, almeno del 95%, almeno del 97%, almeno del 98%, almeno del 99% o del 100%. La percentuale di complementarità tra la sequenza bersaglio dell'acido nucleico e la sequenza di riconoscimento della nucleasi all'interno dell'acido nucleico bersaglio, in alcune forme realizzative, è almeno del 60% su circa 20 nucleotidi contigui. In alcune forme realizzative, gli acidi nucleici a filamento complementare comprendono modifiche o sequenze che forniscono ulteriori caratteristiche desiderabili aggiuntive (ad esempio, stabilità modificata o regolata, direzionamento subcellulare, tracciamento con un marcatore fluorescente, un sito di legame per una proteina o un complesso proteico e simili).

Esempi di tali modifiche comprendono, ad esempio, un cappuccio all'estremità 5' (ad esempio, un cappuccio di 7-metilguanilato (m7G)); una coda poliadenilata all'estremità 3' (ad esempio, una coda di poli(A) al 3'); una sequenza riboswitch (ad esempio, per consentire una stabilità regolata e/o un'accessibilità regolata da parte di proteine e/o complessi proteici); una sequenza di controllo della stabilità; una sequenza che forma un duplex di dsRNA (ad esempio, una forcina)); una modifica o una sequenza che indirizza l'RNA a una localizzazione subcellulare (ad esempio, nucleo, mitocondri, cloroplasti e simili); una modifica o una sequenza che consente il tracciamento (ad esempio, la coniugazione diretta con una sostanza fluorescente, coniugazione diretta con una molecola fluorescente, coniugazione con un gruppo che facilita il rilevamento fluorescente, una sequenza che consente il rilevamento fluorescente e così via); oppure una modifica o una sequenza che fornisce un sito di legame per le proteine (ad esempio, proteine che agiscono sul DNA, compresi gli attivatori trascrizionali, i repressori trascrizionali, le DNA metil transferasi, le DNA demetilasi, le istone acetiltransferasi, le istone deacetilasi e

combinazioni di queste). Gli acidi nucleici complementari sono forniti in qualsiasi forma, ad esempio sotto forma di RNA, sia come due molecole (ad esempio, crRNA e tracrRNA separati) sia come una molecola (ad esempio, sgRNA). In alcune forme di realizzazione, l'acido nucleico a filamento complementare è fornito sotto forma di un complesso con una proteina nucleasi. In alternativa, l'acido nucleico a filamento complementare è fornito anche sotto forma di DNA che codifica l'RNA. Il DNA che codifica l'acido nucleico a filamento complementare codifica alternativamente un singolo acido nucleico a filamento complementare (ad esempio, sgRNA) o molecole di RNA separate (ad esempio, crRNA e tracrRNA separati). In quest'ultimo caso, il DNA che codifica l'acido nucleico a filamento complementare è fornito come molecole di DNA separate che codificano rispettivamente il crRNA e il tracrRNA. In alcune forme di realizzazione, i DNA che codificano gli acidi nucleici complementari sono integrati in modo stabile nel genoma della cellula e, facoltativamente, sono legati operativamente a un promotore attivo nella cellula. I DNA che codificano gli acidi nucleici a filamento complementare, in alcune forme di realizzazione, sono operativamente collegati a un promotore in un costrutto di espressione. Gli acidi nucleici a filamento complementare sono preparati con qualsiasi metodo adatto. Ad esempio, gli acidi nucleici a filamento complementare sono preparati mediante trascrizione *in vitro* utilizzando, ad esempio, la RNA polimerasi T7. In alcune forme di realizzazione, gli acidi nucleici complementari sono anche molecole prodotte sinteticamente e preparate per sintesi chimica.

Nucleasi

Le nucleasi che riconoscono una sequenza bersaglio sono note a chi è esperto dell'arte e comprendono, ma non solo, le nucleasi a zinc finger (ZFN), le nucleasi effettrici simili ad attivatori di trascrizione (TALEN), le nucleasi a ripetizioni palindromiche corte interspaziate regolarmente raggruppate (CRISPR) e le meganucleasi. Le nucleasi presenti nelle composizioni e utili per i metodi qui descritti sono descritte più dettagliatamente di seguito.

Nucleasi zinc finger (ZFN)

Le "nucleasi zinc finger" o "ZFN" sono una fusione tra il dominio di taglio di Fok I e un dominio di riconoscimento del DNA contenente 3 o più motivi zinc finger. L'eterodimerizzazione in una particolare posizione nel DNA di due singole ZFN con un orientamento e una spaziatura precisa porta a una rottura del DNA a doppio filamento. In alcuni casi, le ZFN fondono un dominio di taglio all'estremità C di ciascun dominio zinc finger. Per consentire ai due domini di taglio di dimerizzare e tagliare il DNA, i due singoli ZFN si legano a filamenti opposti di DNA con le loro

estremità C a una certa distanza. In alcuni casi, le sequenze del linker tra il dominio zinc finger e il dominio di taglio richiedono che l'estremità 5' di ciascun sito di legame sia separato da circa 5-7 bp. Esempi di ZFN che sono utili per la presente invenzione includono, ma non si limitano a, quelli descritti in Urnov et al., *Nature Reviews Genetics*, 2010, 11 :636-646; Gaj et al, *Nat Methods*, 2012, 9(8):805-7; nei brevetti statunitensi N. 6,534,261; 6,607,882; 6,746,838; 6,794,136; 6,824,978; 6,866,997; 6,933, 113; 6,979,539; 7,013,219; 7,030,215; 7,220,719; 7,241,573; 7,241,574; 7,585,849; 7,595,376; 6,903,185; 6,479,626; e U. S. pubblicazione di domanda di brevetto U.S. Nos. 2003/636646. U. S. Nos. 2003/0232410 e 2009/0203140. In alcune forme di realizzazione, uno ZFN è una nickasi a zinc finger che, in alcuni casi, è uno ZFN ingegnerizzato che induce rotture o intaccature sito-specifiche del DNA a singolo filamento. Descrizioni di nickasi a dito di zinco si trovano, ad esempio, in Ramirez et al., *Nucl Acids Res*, 2012, 40(12):5560-8; Kim et al., *Genome Res*, 2012, 22(7): 1327-33.

TALEN

Le "TALEN" o "nucleasi tal-effettrici" sono nucleasi ingegnerizzate simili ad attivatori della trascrizione che contengono un dominio centrale di ripetizioni tandem che legano il DNA, un segnale di localizzazione nucleare e un dominio di attivazione trascrizionale C-terminale. In alcuni casi, una ripetizione tandem legata al DNA comprende 33-35 aminoacidi di lunghezza e contiene due residui aminoacidici ipervariabili nelle posizioni 12 e 13 che riconoscono una o più coppie di basi di DNA specifiche. I TALEN sono prodotti fondendo un dominio di legame al DNA dell'effettore TAL con un dominio di taglio del DNA. Per esempio, una proteina TALE può essere fusa con una nucleasi come un'endonucleasi Fok I di tipo selvatico o mutata o il dominio catalitico di Fok I. Sono state apportate diverse mutazioni a FokI per il suo utilizzo nelle TALEN, che, ad esempio, migliorano la specificità o l'attività di taglio. Tali TALEN sono ingegnerizzati per legare qualsiasi sequenza di DNA desiderata.

Le TALEN sono spesso utilizzate per generare modifiche geniche creando una rottura a doppio filamento in una sequenza di DNA bersaglio, che a sua volta subisce NHEJ o HDR. In alcuni casi, per promuovere l'HDR viene fornito un modello di riparazione del DNA donatore a singolo filamento. Descrizioni dettagliate dei TALEN e del loro utilizzo per l'editing genico si trovano, ad esempio, nei brevetti statunitensi n. 8,440,431; 8,440,432; 8,450,471; 8,586,363; e 8,697,853; Scharenberg et al, *Curr Gene Ther*, 2013, 13(4):291-303; Gaj et al., *Nat Methods*, 2012, 9(8):805-7; Beurdeley et al., *Nat Commun*, 2013, 4: 1762; e Joung and Sander, *Nat Rev Mol Cell Biol*, 2013, 14(1):49-55. DNA

Nucleasi guidate

Le "nucleasi guidate da DNA" sono nucleasi che utilizzano un nucleotide complementare del DNA a singolo filamento per dirigere la nucleasi nel punto corretto del genoma ibridandosi con un altro acido nucleico, ad esempio l'acido nucleico bersaglio nel genoma di una cellula. In alcune forme di realizzazione, la nucleasi guidata dal DNA comprende una nucleasi Argonata. In alcuni casi, la nucleasi guidata da DNA è selezionata tra TtAgo, PfAgo e NgAgo. In alcuni casi, la nucleasi guidata dal DNA è NgAgo.

Meganucleasi

Le "meganucleasi" sono endonucleasi a taglio raro o endonucleasi di inserimento che, in alcune forme di realizzazione, sono altamente specifiche e riconoscono siti bersaglio del DNA di lunghezza compresa tra almeno 12 coppie di basi, ad esempio tra 12 e 40 coppie di basi o tra 12 e 60 coppie di basi.

È qui possibile l'uso di qualsiasi meganucleasi, comprese, ma non solo, I- Scel, I- Scell, I-ScellIII, I- ScelIV, I-SceV, I-SceVI, I-SceVII, I-Ceul, I-CeuAIP, I-Crel, I- CrepsbIP, I- CrepsbIIP, I-CrepsbIIIP, I- CrepsbIVP, I-TIil, I-Ppol, PI-Pspl, F-Scel, F-Scell, F- Suvl, F- Tevl, F-TevII, I-Amal, I-Anil, I-Chul, I-Cmoel, I-Cpal, I-Cpall, I-Csml, I-Cvul, I- CvuAIP, I-Ddil, I-Ddill, I-Dirl, I-Dmol, I-Hmul, I-Hmull, I-HsNIP, I-Llal, I-Msol, I-Naal, I- Nanl, I- NclIP, I-NgrIP, I-Nitl, I-Njal, I-Nsp236IP, I-Pakl, I-PbolIP, I-PcuIP, I-PcuAI, I-PcuVI, I- PgrIP, I-PobIP, I-Porl, I-PorIIP, I-PbpIP, I-SpBetaIP, I-Scal, I-SexIP, I-SnelIP, I-Spoml, I- SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiIP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdelIP, I-Tevl, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I- VinIP, I-ZbilIP, PI-Mtul, PI-MtuHIP, PI-MtuHIIP, PI-Pful, PI-Pfull, PI-Pkol, PI-Pkoll, PI- Rma43812IP, PI-SpBetaIP, PI-Scel, PI-Tful, PI-Tfull, PI-Thyl, PI-TIil, PI-THII, meganucleasi I-Crel, meganucleasi I-Ceul, meganucleasi I-Msol, meganucleasi I-Scel, o qualsiasi sua variante, frammento, mutante o derivato attivo.

CRISPR

Il sistema della nucleasi CRISPR (proteina associata a CRISPR, a ripetizioni palindromiche corte interspaziate regolarmente raggruppate/Cas) è un sistema di nucleasi ingegnerizzato basato su un sistema batterico utilizzato per l'ingegneria genomica. Si basa in parte sulla risposta immunitaria adattativa di molti batteri e archei. Quando un virus o un plasmide invade un batterio, segmenti del DNA dell'invasore vengono convertiti in RNA CRISPR (crRNA) dalla risposta "immunitaria". Il crRNA si associa poi, attraverso una regione di parziale complementarità, a un altro tipo di RNA chiamato tracrRNA per guidare la nucleasi Cas (ad

esempio, Cas9) verso una regione omologa al crRNA nel DNA bersaglio, chiamata "protospaziatore". La nucleasi Cas (ad esempio, Cas9) taglia il DNA per generare estremità smussate in corrispondenza della rottura del doppio filamento nei siti specificati da una sequenza del filamento complementare di 20 nucleotidi contenuta nel trascritto del crRNA. La nucleasi Cas (ad esempio, Cas9), in alcune forme realizzative, richiede sia il crRNA che il tracrRNA per il riconoscimento e il taglio del DNA sito-specifico. Questo sistema è stato ora ingegnerizzato in modo tale che, in alcune forme realizzative, il crRNA e il tracrRNA sono combinati in un'unica molecola (il "singolo RNA guida" o "sgRNA") e la porzione equivalente al crRNA del singolo RNA guida è ingegnerizzata per guidare la nucleasi Cas (ad es, Cas9) per indirizzare la nucleasi verso qualsiasi sequenza desiderata (si veda, ad esempio, Jinek et al. (2012) *Science* 337:816-821; Jinek et al. (2013) *eLife* 2:e00471; Segal (2013) *eLife* 2:e00563). In questo modo, il sistema CRISPR/Cas può essere ingegnerizzato per creare un'interruzione a doppio filamento in un bersaglio desiderato nel genoma di una cellula e sfruttare i meccanismi endogeni della cellula per riparare l'interruzione indotta mediante riparazione diretta per omologia (HDR) o giunzione terminale non omologa (NHEJ). In alcune forme di realizzazione, la nucleasi Cas ha un'attività di taglio del DNA. In alcune forme di realizzazione, la nucleasi Cas dirige il taglio di uno o di entrambi i filamenti in un punto della sequenza di DNA bersaglio. Ad esempio, in alcune forme realizzative, la nucleasi Cas è una nickasi con uno o più domini catalitici inattivati che taglia un singolo filamento di una sequenza di DNA bersaglio. Esempi non limitativi di nucleasi Cas comprendono Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (note anche come Csn1 e Csx12), Cas10, , Cpf1, C2c3, C2c2 e C2c1Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Cpf1, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, loro omologhi, loro varianti, loro mutanti e loro derivati. Esistono tre tipi principali di nucleasi Cas (tipo I, tipo II e tipo III) e 10 sottotipi, tra cui 5 proteine di tipo I, 3 di tipo II e 2 di tipo III (si veda, ad esempio, Hochstrasser e Doudna, *Trends Biochem Sci*, 2015;40(I):58-66).

Le nucleasi Cas di tipo II includono, ma non solo, Cas1, Cas2, Csn2 e Cas9. Queste nucleasi Cas sono note a chi è esperto dell'arte. Ad esempio, la sequenza aminoacidica del polipeptide Cas9 di tipo selvatico di *Streptococcus pyogenes* è riportata, ad esempio, in NCBI Ref. Seq. NP_269215, e la sequenza aminoacidica del polipeptide Cas9 di tipo selvatico di *Streptococcus thermophilus* è riportata, ad esempio, in NCBI Ref. Seq. WP_011681470. Le nucleasi Cas, ad esempio i polipeptidi Cas9, in alcune forme realizzative, sono derivati da una varietà di specie batteriche.

"Cas9" si riferisce a una proteina nucleasi o nickasi a doppio filamento di DNA guidata dall'RNA. La nucleasi Cas9 di tipo selvatico ha due domini funzionali, ad esempio RuvC e HNH, che tagliano diversi filamenti di DNA. Cas9 può indurre rotture a doppio filamento nel DNA genomico (DNA bersaglio) quando entrambi i domini funzionali sono attivi. L'enzima Cas9, in alcune forme realizzative, comprende uno o più domini catalitici di una proteina Cas9 derivata da batteri appartenenti al gruppo costituito da *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor* e *Campylobacter*. In alcune forme di realizzazione, la Cas9 è una proteina di fusione, ad esempio i due domini catalitici sono derivati da specie batteriche diverse. Varianti utili della nucleasi Cas9 comprendono un singolo dominio catalitico inattivo, come un enzima RuvC o HNH o una nickasi. Una nickasi Cas9 ha un solo dominio funzionale attivo e, in alcune forme di realizzazione, taglia solo un filamento del DNA bersaglio, creando così una rottura o un taglio a filamento singolo. In alcune forme di realizzazione, la nucleasi Cas9 mutante avente almeno una mutazione D10A è una Cas9 nickasi. In alcune forme di realizzazione, la nucleasi Cas9 mutante avente almeno una mutazione H840A è una Cas9 nickasi. Altri esempi di mutazioni presenti in una Cas9 nickasi comprendono, senza limitazioni, N854A e N863 A. Una rottura a doppio filamento viene introdotta utilizzando una Cas9 nickasi se vengono utilizzati almeno due RNA bersaglio del DNA che mirano a filamenti opposti di DNA. Una rottura a doppio filamento indotta da una rottura è riparata mediante NHEJ o HDR. Questa strategia di editing genico favorisce l'HDR e diminuisce la frequenza di mutazioni nei siti di DNA fuori dal bersaglio. La nucleasi o nickasi Cas9, in alcuni casi, è ottimizzata per il codone della cellula o dell'organismo bersaglio. In alcune forme di realizzazione, la nucleasi Cas è un polipeptide Cas9 che contiene due mutazioni di silenziamento dei domini della nucleasi RuvC e HNH (D10A e H840A), denominato dCas9. In una forma realizzativa, il polipeptide dCas9 di *Streptococcus pyogenes* comprende almeno una mutazione in posizione D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, A987 o una loro combinazione. Le descrizioni di tali polipeptidi dCas9 e delle loro varianti sono riportate, ad esempio, nella pubblicazione della domanda di brevetto internazionale WO 2013/176772. In alcune forme di realizzazione, l'enzima dCas9 contiene una mutazione in D10, E762, H983 o D986, e una mutazione in H840 o N863. In alcuni casi, l'enzima dCas9 contiene una mutazione D10A o DION. Inoltre, l'enzima dCas9 include in alternativa una mutazione H840A, H840Y o

H840N. In alcune forme di realizzazione, l'enzima dCas9 della presente invenzione comprende le sostituzioni D10A e H840A; D10A e H840Y; D10A e H840N; DION e H840A; DION e H840Y; o DION e H840N. Le sostituzioni sono alternativamente conservative o non conservative per rendere il polipeptide Cas9 cataliticamente inattivo e in grado di legarsi al DNA bersaglio. Per i metodi di editing del genoma, la nucleasi Cas in alcune forme realizzative comprende una proteina di fusione Cas9, come un polipeptide che comprende il dominio catalitico dell'enzima di restrizione di tipo IIS, FokI, legato a dCas9. La proteina di fusione FokI-dCas9 (fCas9) può utilizzare due RNA guida per legarsi a un singolo filamento di DNA bersaglio e generare una rottura a doppio filamento.

10 Veicolazione

I vettori di veicolazione genica della presente invenzione possono essere somministrati a un paziente. Detta somministrazione può essere "*in vivo*" o "*ex vivo*". Un operatore esperto è in grado di determinare i dosaggi appropriati. Il termine "somministrato" comprende la somministrazione mediante tecniche virali o non virali. I meccanismi di somministrazione virale includono, a titolo esemplificativo, vettori adenovirali, vettori virali adeno-associati (AAV), vettori virali erpetici, vettori retrovirali, vettori lentivirali e vettori baculovirali, ecc. I sistemi di veicolazione non virali includono la trasfezione del DNA, come l'elettroporazione, la trasfezione mediata da lipidi, la trasfezione mediata da DNA compattato; liposomi, immunoliposomi, lipofectina, anfifili cationici facciali (CFA) e loro combinazioni. La veicolazione di uno o più geni terapeutici mediante un sistema di vettori secondo la presente invenzione può essere utilizzata da sola o in combinazione con altri trattamenti o componenti del trattamento.

È contemplato l'uso di qualsiasi metodo di somministrazione adatto per la distribuzione delle composizioni della descrizione. I singoli componenti del sistema HITI (ad esempio, la nucleasi e/o la sequenza di DNA esogeno), in alcuni casi, vengono somministrati simultaneamente o separati temporalmente. La scelta del metodo di modificazione genetica dipende dal tipo di cellula da trasformare e/o dalle circostanze in cui avviene la trasformazione (ad esempio, *in vitro*, *ex vivo* o *in vivo*). Una descrizione generale di questi metodi si trova in Ausubel, et al., Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995.

Il termine "contatto con la cellula" comprende tutti i metodi di veicolazione qui descritti. In alcune forme di realizzazione, il metodo come qui descritto prevede il contatto con un DNA bersaglio o l'introduzione in una cellula (o in una popolazione di cellule) di uno o più acidi nucleici comprendenti sequenze nucleotidiche che codificano un acido nucleico a filamento

complementare (ad es., gRNA), un polipeptide modificatore sito-diretto (ad es., proteina Cas) e/o una sequenza di DNA esogeno. Gli acidi nucleici adatti che comprendono sequenze nucleotidiche che codificano un acido nucleico a filamento complementare e/o un polipeptide modificatore sito-diretto comprendono vettori di espressione, dove un vettore di espressione che comprende una sequenza nucleotidica che codifica un acido nucleico a filamento complementare e/o un polipeptide modificatore sito-diretto è un vettore di espressione ricombinante. Esempi non limitativi di metodi di somministrazione o trasformazione includono, ad esempio, l'infezione virale o batteriofagica, la trasfezione, la coniugazione, la fusione di protoplasti, la lipofezione, l'elettroporazione, la precipitazione di fosfato di calcio, la trasfezione mediata da polietileneimina (PEI), la trasfezione mediata da DEAE-destrano, la trasfezione mediata da liposomi, la tecnologia di iniezione di particelle, la precipitazione di fosfato di calcio, la microiniezione diretta e la somministrazione di acido nucleico mediata da nanoparticelle (si veda, ad es, Panyam et., al *Adv Drug Deliv Rev.* 2012 Sep. 13. pii: 50169- 409X(12)00283-9. doi: 10.1016/j .addr.2012.09.023). In alcuni aspetti, la presente invenzione fornisce metodi che comprendono la veicolazione di uno o più polinucleotidi, come ad esempio uno o più vettori qui descritti, uno o più trascritti e/o una o più proteine da essi trascritte, a una cellula ospite. In alcuni aspetti, la descrizione fornisce anche le cellule prodotte con tali metodi e organismi (come animali, piante o funghi) che comprendono o sono prodotti da tali cellule. In alcune forme realizzative, una proteina nucleasi in combinazione con, e opzionalmente complessata con, una sequenza del filamento complementare viene veicolata a una cellula. Per introdurre acidi nucleici in cellule di mammifero o in tessuti bersaglio sono contemplati metodi di trasferimento genico convenzionali, virali e non virali. Tali metodi sono utilizzati per veicolare acidi nucleici che codificano componenti di un sistema HITI a cellule in coltura o in un organismo ospite. I sistemi di veicolazione di vettori non virali comprendono plasmidi di DNA, RNA (ad esempio un trascritto di un vettore qui descritto), acido nucleico nudo e acido nucleico complessato con un veicolo di somministrazione, ad esempio un liposoma. I sistemi di veicolazione virale possono comprendere virus a DNA e RNA, che possono avere genomi sia episomici sia integrati dopo la veicolazione alla cellula. Per una rassegna delle procedure di terapia genica, si veda Anderson, *Science* 256:808-813 (1992); Nabel & Feigner, *TIBTECH* 11 :211-217 (1993); Mitani & Caskey, *TIBTECH* 11 : 162-166 (1993); Dillon. *TIBTECH* 11 : 167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10): 1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(l):31-44 (1995);

Haddada et al., in Current Topics in Microbiology and Immunology Doerfler and Bohm (eds) (1995); e Yu et al., Gene Therapy 1 : 13-26 (1994). I metodi di somministrazione non virale degli acidi nucleici possono comprendere la lipofezione, la nucleofezione, la microiniezione, l'elettroporazione, la biolistica, i virosomi, i liposomi, gli immunoliposomi, i coniugati policationi o lipidi:acidi nucleici, il DNA nudo, i virioni artificiali e l'assorbimento del DNA potenziato da agenti. La lipofezione è descritta, ad esempio, in U.S. Pat. 5,049,386, 4,946,787 e 4,897,355) e i reagenti per la lipofezione sono venduti in commercio (ad esempio, Transfectam.TM. e Lipofectin.TM.). I lipidi cationici e neutri adatti per una lipofezione efficiente dei polinucleotidi con riconoscimento del recettore includono quelli di Feigner, WO 91/17424; WO 91/16024. La veicolazione è contemplata per avvenire nelle cellule (ad esempio, somministrazione *in vitro* o *ex vivo*) o in tessuti bersaglio (ad esempio, somministrazione *in vivo*). La preparazione di complessi lipidi/acidi nucleici, compresi i liposomi mirati come i complessi immunolipidici, è ben nota (si veda, ad esempio, Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028 e 4,946,787). I sistemi virali a base di RNA o DNA sono utilizzati per mirare specifiche cellule dell'organismo e trasportare il carico virale nel nucleo della cellula. I vettori virali sono alternativamente somministrati direttamente (*in vivo*) o utilizzati per trattare le cellule *in vitro*, e le cellule modificate sono facoltativamente somministrate (*ex vivo*). I sistemi basati su virus includono, ma non si limitano a, vettori retrovirali, lentivirus, adenovirali, adenoassociati e vettori del virus herpes simplex per il trasferimento genico. L'integrazione nel genoma dell'ospite, in alcune forme realizzative, avviene con i metodi del trasferimento genico di retrovirus, lentivirus e virus adeno-associati, con conseguente espressione a lungo termine del transgene inserito. In alcune forme di realizzazione si osservano elevate efficienze di trasduzione in molti tipi di cellule e tessuti bersaglio. In alcune forme realizzative, sono utilizzati sistemi a base adenovirale. I sistemi a base adenovirale, in alcune forme di realizzazione, portano all'espressione transitoria del transgene. I vettori a base adenovirale sono in grado di garantire un'elevata efficienza di trasduzione nelle cellule e, in alcuni casi, non richiedono la divisione cellulare. Con i vettori adenovirali è possibile ottenere titoli e livelli di espressione elevati. In alcune forme realizzative, i vettori di virus adeno-associati ("AAV") sono utilizzati per trasdurre le cellule con acidi nucleici bersaglio, ad esempio nella produzione *in vitro* di acidi

nucleici e peptidi e per le procedure di terapia genica *in vivo* ed *ex vivo* (vedere, ad esempio, West et al., Virology 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94: 1351 (1994). La costruzione di vettori AAV ricombinanti è descritta in numerose pubblicazioni, tra cui U.S. Pat. No. 5,173,414; 5 Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81 :6466-6470 (1984); e Samulski et al., J. Virol. 63 :03822-3828 (1989). Le cellule da packaging, in alcune forme realizzative, sono utilizzate per formare particelle virali in grado di infettare una cellula ospite. Tali cellule includono, ma non solo, cellule 293 (ad esempio, per il packaging di adenovirus) e cellule .psi.2 o PA317 (ad esempio, per il 10 packaging di retrovirus). I vettori virali sono prodotti producendo una linea cellulare che incorpora un vettore di acido nucleico in una particella virale. In alcune forme di realizzazione, i vettori contengono le sequenze virali minime necessarie per il packaging e la successiva integrazione in un ospite. In alcuni casi, i vettori contengono altre sequenze virali che vengono sostituite da una cassetta di espressione per i polinucleotidi da esprimere. In alcuni casi, le 15 funzioni virali mancanti sono fornite in trans dalla linea cellulare di packaging. Ad esempio, in alcune forme realizzative, i vettori AAV comprendono sequenze ITR dal genoma AAV che sono necessarie per il packaging e l'integrazione nel genoma dell'ospite. Il DNA virale viene incorporato in una linea cellulare che contiene un plasmide helper che codifica gli altri geni di AAV, ovvero rep e cap, ma che manca delle sequenze ITR. In alternativa, la linea cellulare viene 20 infettata con adenovirus come helper. Il virus helper promuove la replicazione del vettore AAV e l'espressione dei geni di AAV dal plasmide helper. La contaminazione da adenovirus è ridotta, ad esempio, dal trattamento termico, al quale l'adenovirus è più sensibile dell'AAV.

I sierotipi di AAV

Ad oggi, sono state identificate e classificate decine di diverse varianti di AAV (sierotipi) 25 (Srivastava A, Curr Opin Virol. 2016 Dec;21:75-80). Tutti i sierotipi conosciuti possono infettare le cellule di diversi tipi di tessuto. La specificità tissutale è determinata dal sierotipo del capsido e la pseudotipizzazione dei vettori AAV per alterare la loro gamma del tropismo sarà probabilmente importante per il loro utilizzo in terapia. I vettori AAV pseudotipizzati sono quelli che contengono il genoma di un sierotipo di AAV nel capsido di un secondo sierotipo AAV; ad 30 esempio un vettore AAV2/8 contiene il capsido di AAV8 e il genoma di AAV 2 (Auricchio et al. (2001) Hum. Mol. Genet. 10(26):3075-81). Tali vettori sono noti anche come vettori chimerici.

Sierotipo 2

Il sierotipo 2 (AAV2) è quello che finora è stato più ampiamente esaminato. AAV2 presenta un tropismo naturale verso i neuroni, le cellule muscolari lisce vascolari e gli epatociti. Sono stati descritti tre recettori cellulari per AAV2: il proteoglicano eparan solfato (HSPG), l'integrina $\alpha V\beta 5$ e il recettore 1 del fattore di crescita dei fibroblasti (FGFR-1). Il primo funziona come recettore primario, mentre gli ultimi due hanno un'attività di co-recettore e consentono all'AAV di entrare nella cellula mediante endocitosi mediata dal recettore. Questi risultati dello studio sono stati contestati da Qiu, Handa e altri. L'HSPG funziona come recettore primario, anche se la sua abbondanza nella matrice extracellulare può eliminare le particelle di AAV e compromettere l'efficacia dell'infezione.

10 Altri sierotipi

Sebbene l'AAV2 sia il sierotipo più popolare in varie ricerche basate sull'AAV, è stato dimostrato che altri sierotipi possono essere più efficaci come vettori per la veicolazione del gene. Ad esempio, l'AAV6 sembra essere molto più efficace per infettare le cellule epiteliali delle vie aeree, l'AAV7 presenta un tasso di trasduzione molto elevato delle cellule muscolari scheletriche murine (analogamente all'AAV1 e all'AAV5), l'AAV8 è eccellente nella trasduzione di epatociti e fotorecettori, mentre l'AAV1 e l'AAV 5 sono dimostrati efficienti nella veicolazione di geni alle cellule endoteliali vascolari. Nel cervello, la maggior parte dei sierotipi AAV mostra un tropismo neuronale, mentre AAV5 trasduce anche gli astrociti. L'AAV6, un ibrido di AAV1 e AAV2, mostra anche una minore immunogenicità rispetto all'AAV2. I sierotipi possono differire per quanto riguarda i recettori a cui sono legati. Ad esempio, la trasduzione di AAV4 e AAV5 può essere inibita da acidi sialici solubili (di forma diversa per ciascuno di questi sierotipi) e AAV5 ha dimostrato di entrare nelle cellule attraverso il recettore del fattore di crescita derivato dalle piastrine. Nuove varianti di AAV, come i mutanti quadrupli della tirosina o AAV 2/7m8, hanno dimostrato di trasdurre la retina esterna dal vitreo in modelli animali di piccole dimensioni (Dalkara D et al., Sci Transl Med. 2013 Jun 12;5(189):189ra76; Petrs-Silva H et al., Mol Ther. 2011 Feb;19(2):293-301). Un altro mutante AAV denominato ShH10, una variante di AAV6 con un miglior tropismo gliale dopo la somministrazione intravitreale (Klimczak RR et al., PLoS One. 2009 Oct 14;4(10):e7467.). Un ulteriore mutante di AAV con tropismo particolarmente vantaggioso per la retina è l'AAV2 (quad Y-F) (Hickey DG et al., Gene Ther. 2017 Dec;24(12):787-800).

Ai sensi della presente invenzione, una particella virale AAV comprende proteine del capsido di un AAV di un sierotipo selezionato da uno o più del gruppo costituito da AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8 AAV9 e AAV 10, preferibilmente dal sierotipo AAV2 o AAV8.

Qualsiasi vettore compatibile con la cellula ospite può essere utilizzato con i metodi

5 dell'invenzione. Esempi non limitativi di vettori per cellule ospiti eucariotiche includono pXTI, pSG5, pSVK3, pBPV, pMSG e pSVLSV40. In alcune forme realizzative, una sequenza nucleotidica che codifica un acido nucleico a filamento complementare e/o un polipeptide modificatore sito-diretto è operativamente legata a un elemento di controllo, ad esempio un elemento di controllo trascrizionale, come un promotore. L'elemento di controllo trascrizionale è funzionale,

10 in alcune forme realizzative, sia in una cellula eucariotica, ad esempio una cellula di mammifero, sia in una cellula procariotica (ad esempio, una cellula batterica o arcaica). In alcune forme realizzative, una sequenza nucleotidica codificante un acido nucleico a filamento complementare e/o un polipeptide modificatore sito-diretto è operativamente legata a più elementi di controllo che consentono l'espressione della sequenza nucleotidica codificante un

15 acido nucleico a filamento complementare e/o un polipeptide modificatore sito-diretto in cellule procariotiche e/o eucariotiche. A seconda del sistema ospite/vettore utilizzato, nel vettore di espressione può essere utilizzato uno qualsiasi degli elementi di controllo della trascrizione e della traduzione, compresi i promotori costitutivi e inducibili, gli elementi potenziatori della trascrizione, i terminatori della trascrizione, ecc. (e.g., promotore U6, promotore HI, etc.; si veda sopra) (si veda e.g., Bitter et al. (1987) *Methods in Enzymology*, 153 :516-544). In alcune forme realizzative, un acido nucleico a filamento complementare e/o un polipeptide modificatore sito-diretto sono forniti come RNA. In questi casi, l'acido nucleico a filamento complementare e/o l'RNA che codifica il polipeptide modificatore sito-diretto sono prodotti per sintesi chimica diretta o possono essere trascritti *in vitro* da un DNA che codifica

25 l'acido nucleico a filamento complementare. L'acido nucleico a filamento complementare e/o l'RNA che codifica il polipeptide modificatore sito-diretto sono sintetizzati *in vitro* utilizzando un enzima RNA polimerasi (ad esempio, la polimerasi T7, la polimerasi T3, la polimerasi SP6, ecc.) Una volta sintetizzato, l'RNA entra direttamente in contatto con il DNA bersaglio o viene introdotto in una cellula con qualsiasi tecnica adatta all'introduzione di acidi nucleici nelle cellule

30 (ad esempio, microiniezione, elettroporazione, trasfezione, ecc.) I nucleotidi che codificano un acido nucleico a filamento complementare (introdotto come DNA o RNA) e/o un polipeptide modificatore sito-diretto (introdotto come DNA o RNA) e/o una sequenza di DNA esogeno sono

forniti alle cellule mediante un'adeguata tecnica di trasfezione; si veda, ad esempio, Angel e Yanik (2010) PLoS ONE 5(7): el 1756, e i reagenti TransMessenger.RTM. disponibili in commercio di Qiagen, Stemfect.TM. RNA Transfection Kit di Stemgent e TransIT.RTM.-mRNA Transfection Kit di Minis Bio LLC. Gli acidi nucleici che codificano un acido nucleico a filamento complementare e/o un polipeptide modificatore sito-diretto e/o un polipeptide modificatore sito-diretto chimerico e/o una sequenza di DNA esogena possono essere forniti su vettori di DNA. Sono disponibili molti vettori, ad esempio plasmidi, cosmidi, minicircoli, fagi, virus, ecc. utili per trasferire acidi nucleici in cellule bersaglio. I vettori che contengono gli acidi nucleici in alcune forme realizzative sono mantenuti episomicamente, ad esempio come plasmidi, DNA minicircolari, virus come citomegalovirus, adenovirus, ecc. o sono integrati nel genoma della cellula bersaglio, attraverso ricombinazione omologa o integrazione casuale, ad esempio vettori derivati da retrovirus come MMLV, HIV-1 e ALV.

Metodi per modificare il DNA genomico

Sono qui forniti metodi e composizioni per l'integrazione mirata indipendente dall'omologia (HITI) finalizzata ad apportare modifiche all'acido nucleico, come il DNA genomico, compreso il DNA genomico in cellule non in divisione o terminalmente differenziate che non si dividono. Le cellule qui menzionate sono preferibilmente cellule che non si dividono, più preferibilmente cellule terminalmente differenziate o cellule quiescenti. I metodi qui descritti, almeno in alcune forme realizzative, sono indipendenti dall'omologia e utilizzano l'unione delle estremità non omologhe per inserire il DNA esogeno in un DNA bersaglio, come il DNA genomico di una cellula, ad esempio una cellula non in divisione o terminalmente differenziata. In alcune forme di realizzazione, i metodi qui descritti comprendono un metodo di integrazione di una sequenza di DNA esogeno in un genoma di una cellula non in divisione che comprende il porre a contatto la cellula non in divisione con una composizione che comprende un costrutto di bersaglio che comprende la sequenza di DNA esogeno e una sequenza bersaglio, un oligonucleotide complementare alla sequenza bersaglio e una nucleasi, dove la sequenza di DNA esogeno comprende almeno un nucleotide differente rispetto al genoma e la sequenza bersaglio è riconosciuta dalla nucleasi. In alcune forme di realizzazione dei metodi HITI qui illustrati, le sequenze di DNA esogene sono frammenti di DNA contenenti la sequenza desiderata da inserire nel genoma della cellula bersaglio o della cellula ospite. Almeno una parte della sequenza di DNA esogeno ha una sequenza omologa a una parte del genoma della cellula bersaglio o della cellula ospite e almeno una parte della sequenza di DNA esogeno ha una sequenza non omologa

a una parte del genoma della cellula bersaglio o della cellula ospite. Ad esempio, in alcune forme di realizzazione, la sequenza di DNA esogeno può comprendere una porzione di sequenza di DNA genomico di una cellula ospite con una mutazione al suo interno. Pertanto, quando la sequenza di DNA esogeno viene integrata nel genoma della cellula ospite o bersaglio, la mutazione presente nella sequenza di DNA esogeno viene trasportata nel genoma della cellula ospite o bersaglio. In alcune forme di realizzazione dei metodi HITI qui illustrati, la sequenza di DNA esogeno è affiancata da almeno una sequenza bersaglio. In alcune forme di realizzazione, la sequenza di DNA esogeno è affiancata da due sequenze bersaglio. La sequenza bersaglio comprende una sequenza specifica di DNA riconosciuta da almeno una nucleasi. In alcune forme di realizzazione, la sequenza bersaglio è riconosciuta dalla nucleasi in presenza di un oligonucleotide complementare alla sequenza bersaglio. In alcune forme di realizzazione, nei metodi HITI qui illustrati, una sequenza bersaglio comprende una sequenza nucleotidica che viene riconosciuta e tagliata da una nucleasi. Le nucleasi che riconoscono una sequenza bersaglio sono note a chi è esperto del settore e includono, a titolo esemplificativo, le nucleasi a zinc finger (ZFN), le nucleasi transcription activator-like effector (TALEN) e le nucleasi clustered regularly interspaced short palindromic repeats (CRISPR).

Le ZFN, in alcune forme di realizzazione, comprendono un dominio di legame del DNA a zinc finger e un dominio di taglio del DNA, fusi insieme per creare una nucleasi specifica per la sequenza. Le TALEN, in alcune forme di realizzazione, comprendono un dominio di legame del DNA dell'effettore TAL e un dominio di taglio del DNA, fusi insieme per creare una nucleasi specifica per la sequenza. Le nucleasi CRISPR, in alcune forme di realizzazione, sono nucleasi naturali che riconoscono sequenze di DNA omologhe alle ripetizioni palindromiche brevi regolarmente intercalate, comunemente presenti nel DNA procariotico. Le nucleasi CRISPR includono, ma non sono limitate a, Cas9 Cpf1, C2c3, C2c2 e C2c1. Convenzionalmente, una Cas 9 della presente invenzione è una variante con attività al di fuori del bersaglio come SpCas9 D10A (Ran, F.A., et al., Genome engineering using the CRISPR-Cas9 system. Nat Protoc, 2013. 8(11): p. 2281-2308.

(con inattivazione dell'attività di clivaggio del dominio RuvC), SpCas9 N863A (Ran, F.A., et al., Genome engineering using the CRISPR-Cas9 system. Nat Protoc, 2013. 8(11): p. 2281-2308)

(inattivazione dell'attività di taglio del dominio HNH), SpCas9-HF1 (Kleinstiver, B.P., et al., High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature, 2016. 529(7587): p. 490-5) (Riduzione dell'energia di legame di Cas9 mediante ingegneria

proteica), eSpCas9 (laymaker, I.M., et al., Rationally engineered Cas9 nucleases with improved specificity. *Science*, 2016. 351(6268): p. 84-8) (Riduzione della carica positiva di Cas9), EvoCas9 (Asini, A., et al., A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol*, 2018. 36(3): p. 265-271) (mutagenesi del dominio REC3), KamiCas9 (Merienne, N., et al., The Self-Inactivating KamiCas9 System for the Editing of CNS Disease Genes. *Cell Rep*, 2017. 20(12): p. 2980-2991) (Knockout di Cas9 dopo l'espressione)

I metodi HITI qui illustrati, in alcune forme di realizzazione, sono in grado di introdurre mutazioni in un genoma ospite o in un genoma bersaglio e di riparare mutazioni in un genoma ospite o in un genoma bersaglio. Le mutazioni o le sequenze di tipo selvatico, in alcune forme di realizzazione dei metodi qui descritti, si trovano nella sequenza di DNA esogeno da inserire nel genoma ospite o nel genoma bersaglio. Le mutazioni sono note a chi è esperto del settore e comprendono cambiamenti di una singola coppia di basi o mutazioni puntiformi, inserzioni e delezioni. In alcune forme di realizzazione, un cambiamento di una singola coppia di basi risulta in una mutazione missenso che crea un codone che codifica un amminoacido diverso nell'mRNA trascritto rispetto alla sequenza di tipo selvatico. In alcune forme di realizzazione, un cambiamento in una singola coppia di basi determina una mutazione nonsense che codifica per un codone di stop nell'mRNA trascritto. In alcune forme di realizzazione, un codone di stop nell'RNA trascritto determina un troncamento precoce di una proteina tradotta dall'mRNA. In alcune forme di realizzazione, un cambiamento di una singola coppia di basi dà luogo a una mutazione silente che non comporta alcun cambiamento negli amminoacidi codificati da un mRNA trascritto dal genoma ospite o dal genoma bersaglio.

In alcune forme di realizzazione, una mutazione silente si trova in un introne. In alcune forme di realizzazione, una mutazione silente si trova in un esone e crea un codone che codifica per lo stesso amminoacido della sequenza di tipo selvatico. In alcune forme di realizzazione, una mutazione silente si trova in un promotore, in un potenziatore, in un 5' UTR, in un 3' UTR o in un'altra regione non codificante del genoma ospite o del genoma bersaglio. In alcune forme di realizzazione, una mutazione silente provoca uno splicing aberrante di un trascritto di mRNA. In alcune forme di realizzazione, una mutazione silente interrompe una sequenza donatrice o accettore di splicing dell'RNA. In alcune forme di realizzazione, una mutazione silente provoca un'esportazione aberrante dell'RNA. In alcune forme di realizzazione, una mutazione silente provoca una traduzione aberrante o ridotta di un mRNA. In alcune forme di realizzazione, una mutazione silente provoca una trascrizione aberrante o ridotta di un RNA. In alcune forme di

realizzazione le mutazioni comprendono inserzioni nel genoma dell'ospite o nel genoma bersaglio. In alcune forme di realizzazione, le inserzioni comprendono un numero specifico di nucleotidi compreso tra 1 e 4.700 coppie di basi, ad esempio 1-10, 5-20, 15-30, 20-50, 40-80, 50-100, 100-1000, 500-2000, 1000- 4.700 coppie di basi. In alcune forme di realizzazione, il metodo comprende l'eliminazione di almeno un gene, o di un suo frammento, dal genoma ospite o dal genoma bersaglio. In alcune forme di realizzazione, il metodo comprende l'introduzione di un gene esogeno (qui definito anche come sequenza di DNA esogeno o gene di interesse), o di un suo frammento, nel genoma ospite o nel genoma bersaglio. In alcune forme di realizzazione, il metodo comprende la sostituzione di un gene mutato, o di un suo frammento, nel genoma ospite o nel genoma bersaglio con un gene di tipo selvatico, o un suo frammento. In alcune forme di realizzazione, il gene ospite viene silenziato e sostituito da un gene di tipo selvatico o da una sua sequenza codificante. In alcune forme di realizzazione, il metodo modifica almeno un nucleotide del genoma dell'ospite o del genoma bersaglio con conseguente aumento dell'espressione di un gene. In alcune forme di realizzazione, il metodo modifica almeno un nucleotide di un genoma ospite o di un genoma bersaglio con conseguente diminuita espressione di un gene. In alcune forme di realizzazione, il metodo introduce un promotore esogeno nel genoma ospite o nel genoma bersaglio, con conseguente alterata espressione di un gene. In alcune forme di realizzazione, il promotore è un promotore inducibile. I metodi HITI qui illustrati hanno una maggiore capacità di apportare modifiche al DNA genomico in cellule non in divisione. Le cellule non in divisione includono, ma non sono limitate a: cellule retiniche, preferibilmente cellule gangliari retiniche, cellule bipolari, cellule amacrine, epitelio pigmentato retinico, cellule orizzontali, cellule dei bastoncelli e dei coni o cellule della regione anteriore dell'occhio, come l'epitelio pigmentato dell'iride, l'epitelio corneale, i fibroblasti corneali, cellule del sistema nervoso centrale, compresi neuroni, oligodendrociti, microglia e cellule ependimali; cellule dei trasduttori sensoriali; cellule dei neuroni autonomi; cellule di supporto degli organi di senso e dei neuroni periferici; cellule della retina, tra cui fotorecettori, bastoncelli e coni; cellule del rene, tra cui cellule parietali, podociti del glomerulo, cellule del bordo a spazzola del tubulo prossimale, cellule del segmento sottile dell'ansa di Henle, cellule del tubulo distale, cellule del dotto collettore; cellule di derivazione ematopoietica, tra cui linfociti, monociti, neutrofili, eosinofili, basofili, trombociti; cellule del fegato, tra cui epatociti, cellule stellate, cellule di Kupffer e cellule endoteliali epatiche; cellule endocrine pancreatiche, comprese le cellule alfa, beta, delta, gamma ed epsilon; cellule dell'epitelio respiratorio, comprese le cellule

ciliate, le cellule basali, le cellule del calice e le cellule alveolari, cellule germinali, tra cui l'oogonio/oocita, lo spermatide, lo spermatocita, la cellula spermatogoniale e lo spermatozoo; cellule dell'osso, tra cui osteociti, osteoclasti e osteoblasti; cellule del cuore, tra cui cardiomiociti e cellule pacemaker cardiache; cellule follicolari della tiroide; cellule del tratto digestivo superiore, tra cui cellule sierose, cellule mucose e papille gustative; cellule dello stomaco, tra cui cellule parietali, cellule del capo, cellule enteroendocrine; cellule endoteliali, cellule epiteliali, adipociti, cellule del midollo osseo, cellule dell'orecchio interno, cellule del derma, cellule muscolari lisce e cellule muscolari scheletriche. In alcune forme di realizzazione, i metodi HITI qui descritti forniscono un metodo per apportare modifiche al DNA genomico in cellule in divisione, in cui il metodo ha un'efficienza superiore rispetto ai metodi precedenti divulgati nell'arte. Le cellule in divisione includono, ma non si limitano a, cellule staminali ematopoietiche, cellule staminali mesenchimali, cellule staminali neurali, cellule staminali epatiche, cellule satelliti muscolari, cellule dell'epidermide, cellule gliali e astrociti. In alcune forme di realizzazione, il costrutto di bersaglio, gli oligonucleotidi del filamento complementare e/o un polinucleotide che codifica la nucleasi per i metodi HITI qui descritti sono introdotti nella cellula bersaglio o nella cellula ospite da un virus. In alcuni casi, i virus infettano la cellula bersaglio ed esprimono il costrutto di bersaglio, gli oligonucleotidi del filamento complementare e la nucleasi, consentendo al DNA esogeno del costrutto di bersaglio di integrarsi nel genoma dell'ospite. In alcune forme di realizzazione, il virus comprende un virus sendai, un retrovirus, un lentivirus, un baculovirus, un adenovirus o un virus adeno-associato. In alcuni casi il virus è un virus pseudotipato. In alcune forme di realizzazione, il costrutto di bersaglio, gli oligonucleotidi del filamento complementare e/o un polinucleotide che codifica la nucleasi per i metodi HITI qui descritti sono introdotti nella cellula bersaglio o nella cellula ospite con un metodo di veicolazione del gene non virale. I metodi di veicolazione genica non virale, in alcune forme di realizzazione, veicolano il materiale genetico (compresi DNA, RNA e proteine) nella cellula bersaglio ed esprimono il costrutto di bersaglio, gli oligonucleotidi a filamento complementare e la nucleasi, consentendo al DNA esogeno del costrutto di bersaglio di integrarsi nel genoma dell'ospite. In alcune forme di realizzazione, il metodo non virale comprende un reagente di trasfezione (comprese le nanoparticelle) per DNA mRNA o proteine, o l'elettroporazione.

Metodi di trattamento di malattie

Sono inoltre qui forniti metodi e composizioni per il trattamento di malattie, come le malattie genetiche. Le malattie genetiche sono quelle causate da mutazioni nel DNA ereditato. In alcune forme di realizzazione, le malattie genetiche sono causate da mutazioni nel DNA genomico. Le mutazioni genetiche sono note a chi è esperto del settore e comprendono cambiamenti di una

5 singola coppia di basi o mutazioni puntiformi, inserzioni e delezioni. In alcune forme di realizzazione, i metodi qui forniti comprendono un metodo per trattare una malattia genetica in un soggetto che ne ha bisogno, dove la malattia genetica deriva da un gene mutato che ha almeno un nucleotide cambiato rispetto a un gene di tipo selvatico, in cui il metodo comprende il contatto di almeno una cellula del soggetto con una composizione che comprende un

10 costruito bersaglio che comprende una sequenza di DNA omologa al gene di tipo selvatico e una sequenza di bersaglio, un oligonucleotide complementare alla sequenza bersaglio e una nucleasi, in cui la sequenza bersaglio è riconosciuta dalla nucleasi in modo tale che il gene mutato, o un suo frammento, venga sostituito con il gene di tipo selvatico, o un suo frammento. Le malattie genetiche che sono trattate con i metodi qui illustrati comprendono, ma non sono

15 limitate a, le malattie ereditarie autosomiche dominanti in cui sia l'allele mutante che quello di tipo selvatico sono sostituiti con una copia corretta del gene fornito dal DNA donatore o le malattie ereditarie e comuni dovute al guadagno di funzione tossica, preferibilmente dette malattie comprendono la distrofia retinica, preferibilmente la distrofia retinica è selezionata da: retinite pigmentosa, distrofia del cono o distrofia dei coni e dei bastoncelli, degenerazione

20 maculare, ad es. malattia di Stargardt (ELOVL4), Von-Hippel Lindau, retinoblastoma, RP4 (vedi RHO; OMIM: 180380), RP63 (vedi OMIM: 614494), CORD1 (distrofia cono-radicolare 1; vedi OMIM: 600624), CORD17 (distrofia dei coni e dei bastoncelli 17; vedi OMIM: 615163), BEST1 (bestrofina-1; malattia di Best; proteina 2 della distrofia maculare vitelliforme; vedi OMIM : 607854), OPA1 (OPA1 dynamin like GTPase mitocondriale; vedi OMIM: 605290), malattie

25 neuronali, epatiche, lipofuscinosi (malattia di Batten e altre), preferibilmente per il trattamento di malattie oculari ereditate in modo dominante, ad esempio la degenerazione retinica. degenerazione retinica, preferibilmente retinite pigmentosa, malattie neuronali ed epatiche. Le malattie retiniche che possono essere trattate nella presente invenzione sono, ad esempio, la retinite pigmentosa (dovuta a mutazioni nei geni RHO, AIPL1, IMPDH1, RDS, PDE6B o altri), la

30 distrofia dei coni e dei bastoncelli (CRX), la malattia di Stargardt (ELOVL4), Von-Hippel Lindau e il retinoblastoma.

I metodi di trattamento delle malattie genetiche qui illustrati utilizzano preferibilmente sequenze di DNA esogene che comprendono almeno una parte di una sequenza di DNA di tipo selvatico che corrisponde alla sequenza di DNA del gene mutato, in modo che nel metodo la sequenza di DNA mutato venga sostituita con la sequenza di DNA di tipo selvatico.

5 I termini "a", "un" o "il" come utilizzati nel presente documento non comprendono solo aspetti con un membro, ma comprendono anche aspetti con più membri. Ad esempio, le forme singolari "a", "un" e "il" comprendono referenti plurali, a meno che il contesto non indichi chiaramente il contrario. Così, ad esempio, il riferimento a "una cellula" comprende una pluralità di tali cellule e il riferimento all'agente" include il riferimento a uno o più agenti noti
 10 agli esperti del settore, e così via. Il termine "editing del genoma" si riferisce a un tipo di ingegneria genetica in cui il DNA viene inserito, sostituito o rimosso da un DNA bersaglio, ad esempio il genoma di una cellula, utilizzando una o più nucleasi e/o nickasi. Le nucleasi creano rotture a doppio filamento (DSB) specifiche in posizioni desiderate del genoma e sfruttano i meccanismi endogeni della cellula per riparare la rottura indotta mediante la giunzione delle
 15 estremità non omologhe (NHEJ). Le nickasi creano specifiche rotture a singolo filamento specifiche nei punti desiderati del genoma. In un esempio non limitativo, si possono usare due nickasi per creare due rotture a singolo filamento su filamenti opposti di un DNA bersaglio, generando così un'estremità smussata o adesiva. Qualsiasi nucleasi adatta può essere introdotta in una cellula per indurre l'editing del genoma di una sequenza di DNA bersaglio,
 20 comprese, ma non solo, le nucleasi associate alla proteina CRISPR (Cas), le nucleasi a zinc finger (ZFN), le nucleasi effettrici simili agli attivatori della trascrizione (TALEN), le meganucleasi, altre endo- o eso-nucleasi, le loro varianti, i loro frammenti e le loro combinazioni. Il termine "non-homologous end joining" o "NHEJ" si riferisce a un percorso che ripara le rotture del DNA a doppio filamento in cui le estremità della rottura sono direttamente legate senza la necessità di
 25 uno stampo omologo. I termini "polinucleotide", "oligonucleotide", "acido nucleico", "nucleotide" e "molecola di acido nucleico" possono essere usati in modo intercambiabile per indicare gli acidi desossiribonucleici (DNA), gli acidi ribonucleici (RNA) e i loro polimeri in forma singola, doppia o multifilare. Il termine comprende, ma non solo, DNA o RNA a singolo, doppio o multi filamento, DNA genomico, cDNA, ibridi DNA-RNA o un polimero comprendente basi
 30 puriniche e/o pirimidiniche o altre basi nucleotidiche naturali, chimicamente modificate, biochimicamente modificate, non naturali, sintetiche o derivatizzate. Sono comprese anche le modifiche, come la metilazione e/o il capping, e le forme non modificate del polinucleotide.

Più in particolare, i termini "polinucleotide", "oligonucleotide", "acido nucleico" e "molecola di acido nucleico" includono i polidesossiribonucleotidi (contenenti 2-deossi-D-ribosio), i poliribonucleotidi (contenenti D-ribosio), qualsiasi altro tipo di polinucleotide che sia un N- o C-glicoside di una base purinica o pirimidinica, e altri polimeri contenenti strutture non nucleotidiche, ad esempio poliammide (ad es. g., acidi nucleici peptidici (PNA)) e polimorfolino (disponibile in commercio presso Anti-Virals, Inc., Corvallis, Oreg., con il nome di Neugene) e altri polimeri sintetici di acido nucleico specifico per la sequenza, a condizione che i polimeri contengano nucleobasi in una configurazione che consenta l'appaiamento e l'impilamento delle basi, come avviene nel DNA e nell'RNA. In alcune forme di realizzazione, un acido nucleico può comprendere una miscela di DNA, RNA e loro analoghi. Se non specificamente limitato, il termine comprende gli acidi nucleici contenenti analoghi noti di nucleotidi naturali che hanno proprietà di legame simili a quelle dell'acido nucleico di riferimento e sono metabolizzati in modo simile ai nucleotidi presenti in natura. Se non diversamente indicato, una particolare sequenza di acido nucleico comprende anche implicitamente varianti modificate in modo conservativo (ad esempio, sostituzioni degenerate di codoni), alleli, ortologhi, polimorfismi a singolo nucleotide (SNP) e sequenze complementari, oltre alla sequenza esplicitamente indicata. In particolare, le sostituzioni degenerate del codone possono essere ottenute generando sequenze in cui la terza posizione di uno o più codoni selezionati (o di tutti) è sostituita da residui di basi miste e/o deossinosina. Il termine acido nucleico è usato in modo intercambiabile con gene, cDNA e mRNA codificato da un gene. Il termine "gene" o "sequenza nucleotidica codificante un polipeptide" indica il segmento di DNA coinvolto nella produzione di una catena polipeptidica. Il segmento di DNA può comprendere le regioni che precedono e seguono la regione codificante (leader e trailer) coinvolte nella trascrizione/traduzione del prodotto genico e nella regolazione della trascrizione/traduzione, nonché le sequenze che intervengono (introni) tra i singoli segmenti codificanti (esoni). I termini "polipeptide", "peptide" e "proteina" sono qui utilizzati in modo intercambiabile per indicare un polimero di residui di aminoacidi. I termini si applicano a polimeri di aminoacidi in cui uno o più residui di aminoacidi sono un mimetico chimico artificiale di un corrispondente aminoacido presente in natura, così come a polimeri di aminoacidi presenti in natura e a polimeri di aminoacidi non presenti in natura. Come usato in questo documento, i termini comprendono catene di aminoacidi di qualsiasi lunghezza, comprese le proteine complete, in cui i residui di aminoacidi sono legati da legami peptidici covalenti. Un "vettore di espressione ricombinante" è un

costrutto di acido nucleico, generato in modo ricombinante o sintetico, con una serie di elementi di acido nucleico specificati che consentono la trascrizione di una particolare sequenza polinucleotidica in una cellula ospite.

Un vettore di espressione può essere parte di un plasmide, di un genoma virale o di un frammento di acido nucleico. In genere, un vettore di espressione comprende un polinucleotide da trascrivere, collegato operativamente a un promotore. In questo contesto, per "operativamente legato" si intendono due o più elementi genetici, come una sequenza polinucleotidica codificante e un promotore, posti in posizioni relative che consentono il corretto funzionamento biologico degli elementi, come il promotore che dirige la trascrizione della sequenza codificante. Il termine "promotore" viene qui utilizzato per indicare una serie di sequenze di controllo dell'acido nucleico che dirigono la trascrizione di un acido nucleico. Come usato qui, un promotore comprende sequenze di acido nucleico necessarie vicino al sito di inizio della trascrizione, come, nel caso di un promotore di tipo polimerasi II, un elemento TATA. Un promotore può anche comprendere elementi potenziatori o repressori distali, che possono essere situati fino a diverse migliaia di coppie di basi dal sito di inizio della trascrizione. Altri elementi che possono essere presenti in un vettore di espressione comprendono quelli che potenziano la trascrizione (ad esempio, i potenziatori) e che terminano la trascrizione (ad esempio, i terminatori), nonché quelli che conferiscono una certa affinità di legame o antigenicità alla proteina ricombinante prodotta dal vettore di espressione. Il termine "polimorfismo a singolo nucleotide" o "SNP" si riferisce a un cambiamento di un singolo nucleotide in un polinucleotide, anche all'interno di un allele. Ciò può comprendere la sostituzione di un nucleotide con un altro, così come la delezione o l'inserimento di un singolo nucleotide. In genere, gli SNP sono marcatori biallelici, ma possono esistere anche marcatori tri- e tetra-allelici. A titolo di esempio non limitativo, una molecola di acido nucleico che comprende l'SNP A\C può includere una C o una A nella posizione polimorfica. I termini "soggetto", "paziente" e "individuo" sono qui utilizzati in modo intercambiabile per includere un essere umano o un animale. Ad esempio, il soggetto animale può essere un mammifero, un primate (ad esempio, una scimmia), un animale da allevamento (ad esempio, un cavallo, una mucca, una pecora, un maiale o una capra), un animale da compagnia (ad esempio, un cane, un gatto), un animale da laboratorio (ad esempio, un topo, un ratto, una cavia, un uccello), un animale di importanza veterinaria o un animale di importanza economica. Come usato nel presente documento, il termine "somministrare" include la somministrazione orale, il contatto topico, la

somministrazione come supposta, la somministrazione endovenosa, intraperitoneale, intramuscolare, intralesionale, intratecale, intranasale o sottocutanea a un soggetto. La somministrazione avviene per qualsiasi via, compresa quella parenterale e transmucosa (ad esempio, buccale, sublinguale, palatale, gengivale, nasale, vaginale, rettale o transdermica). La somministrazione parenterale comprende, ad esempio, quella endovenosa, intramuscolare, intra-arteriolare, intradermica, sottocutanea, intraperitoneale, intraventricolare e intracranica. Altre modalità di somministrazione includono, ma non solo, l'uso di formulazioni liposomiali, l'infusione endovenosa, i cerotti transdermici, ecc.

Il termine "trattamento" si riferisce a un approccio per ottenere risultati benefici o desiderati, compresi, ma non limitati a, un beneficio terapeutico e/o un beneficio profilattico. Per beneficio terapeutico si intende qualsiasi miglioramento o effetto terapeuticamente rilevante su una o più malattie, condizioni o sintomi in trattamento. Per beneficio profilattico, si intende che le composizioni possono essere somministrate a un soggetto a rischio di sviluppare una particolare malattia, condizione o sintomo, o a un soggetto che riferisce uno o più sintomi fisiologici di una malattia, anche se la malattia, la condizione o il sintomo non si sono ancora manifestati. Il termine "quantità efficace" o "quantità sufficiente" si riferisce alla quantità di un agente (ad esempio, nucleasi del DNA, ecc.) che è sufficiente per ottenere risultati benefici o desiderati. La quantità terapeuticamente efficace può variare a seconda di uno o più fattori: il soggetto e la condizione patologica da trattare, il peso e l'età del soggetto, la gravità della condizione patologica, la modalità di somministrazione e simili, che possono essere facilmente determinati da una persona di comune competenza nell'arte. La quantità specifica può variare in base a uno o più dei seguenti fattori: il particolare agente scelto, il tipo di cellula bersaglio, la localizzazione della cellula bersaglio nel soggetto, il regime di dosaggio da seguire, l'eventuale somministrazione in combinazione con altri composti, i tempi di somministrazione e il sistema fisico di somministrazione in cui viene trasportato.

Il termine "veicolo farmaceuticamente accettabile" si riferisce a una sostanza che favorisce la somministrazione di un agente (ad esempio, una nucleasi del DNA, ecc.) a una cellula, a un organismo o a un soggetto. Per "vettore farmaceuticamente accettabile" si intende un vettore o un eccipiente che può essere compreso in una composizione o formulazione e che non causa effetti tossicologici avversi significativi sul paziente. Esempi non limitativi di veicolo farmaceuticamente accettabile comprendono acqua, NaCl, soluzioni saline normali, Ringer lattato, saccarosio normale, glucosio normale, leganti, riempitivi, disintegranti, lubrificanti,

rivestimenti, edulcoranti, aromi e coloranti e simili. Un esperto dell'arte riconosce che altri supporti farmaceutici sono utili per la presente invenzione.

Il termine "circa" in relazione a un valore numerico di riferimento può comprendere un intervallo di valori più o meno del 10% rispetto a tale valore. Ad esempio, la quantità "circa 10" comprende valori compresi tra 9 e 11, inclusi i numeri di riferimento di 9, 10 e 11. Il termine "circa" in relazione a un valore numerico di riferimento può anche includere un intervallo di valori più o meno pari al 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% o 1% rispetto a tale valore.

Come usato nel presente documento, il termine "derivati" si riferisce anche a polinucleotidi/proteine più o meno lunghi e/o aventi, ad esempio, una percentuale di identità di almeno 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, più preferibilmente di almeno 99% con le sequenze qui descritte. Nella presente invenzione "identità di almeno il 70%" significa che l'identità può essere di almeno il 70%, o il 75%, o l'80%, o l'85% o il 90% o il 95% o il 100% rispetto alle sequenze di riferimento. Questo vale per tutte le % di identità menzionate. Preferibilmente, la % di identità si riferisce all'intera lunghezza della sequenza di riferimento. Il derivato dell'invenzione comprende anche "mutanti funzionali" dei polipeptidi o dei polinucleotidi, che sono polipeptidi o polinucleotidi che possono essere generati mutando uno o più aminoacidi o nucleotidi nelle loro sequenze e che mantengono la loro attività. Nella presente invenzione, per "funzionale" si intende, ad esempio, "che mantiene la propria attività". Rientrano nell'ambito dell'invenzione anche i polinucleotidi che presentano le stesse sequenze nucleotidiche di un polinucleotide qui esemplificato, ad eccezione di sostituzioni, aggiunte o delezioni nucleotidiche all'interno della sequenza del polinucleotide, purché questi polinucleotidi varianti mantengono sostanzialmente la stessa attività funzionale rilevante dei polinucleotidi qui specificamente esemplificati (ad es, codificano una proteina con la stessa sequenza aminoacidica o la stessa attività funzionale codificata dal polinucleotide esemplificato). Pertanto, si deve intendere che i polinucleotidi qui descritti comprendono mutanti, derivati, varianti e frammenti, come discusso in precedenza, delle sequenze specificamente esemplificate. La presente invenzione contempla anche molecole polinucleotidiche con sequenze sufficientemente omologhe alle sequenze polinucleotidiche dell'invenzione, in modo da consentire l'ibridazione con tali sequenze in condizioni rigorose e metodi standard (Maniatis, T. et al, 1982).

Peptidi 2A auto-taglianti

I peptidi 2A sono peptidi lunghi 18-22 aa che possono indurre il taglio delle proteine ricombinanti nella cellula. I peptidi 2A derivano dalla regione 2A del genoma del virus.

Quattro membri della famiglia dei peptidi 2A sono frequentemente utilizzati nella ricerca scientifica. Si tratta di P2A, E2A, F2A e T2A. F2A deriva dal virus 18 dell'afta epizootica; E2A deriva dal virus A della rinite equina; P2A deriva dal teschovirus-1 2A dei suini; T2A deriva dal virus 2 dell'asigna thosea. Detti peptidi comprendono o consistono di preferenza nelle seguenti sequenze.

Peptide	Sequenza
T2A	(GSG) E G R G S L L T C G D V E E N P G P (SEQ ID NO:3)
P2A	(GSG) A T N F S L L K Q A G D V E E N P G P (SEQ ID NO:4)
E2A	(GSG) Q C T N Y A L L K L A G D V E S N P G P (SEQ ID NO:35)
F2A	(GSG) V K Q T L N F D L L K L A G D V E S N P G P (SEQ ID NO:36)

Qualsiasi sequenza di salto ribosomiale può essere utilizzata ai sensi della presente invenzione.

Una preferita è T2A o P2A.

Sequenze accettore di splicing

Lo splicing dell'RNA è una forma di elaborazione dell'RNA in cui un trascritto di RNA messaggero precursore (pre-mRNA) appena prodotto viene trasformato in un RNA messaggero maturo (mRNA). Durante lo splicing, gli introni (regioni non codificanti) vengono rimossi e gli esoni (regioni codificanti) vengono uniti.

All'interno degli introni, per lo splicing sono necessari un sito donatore (estremità 5' dell'introne), un sito di diramazione (vicino all'estremità 3' dell'introne) e un sito accettore (estremità 3' dell'introne). Il sito donatore di splicing comprende una sequenza quasi invariante GU all'estremità 5' dell'introne, all'interno di una regione più ampia e meno altamente conservata. Il sito accettore di splicing all'estremità 3' dell'introne termina l'introne con una sequenza quasi invariante AG. A monte (5'-ward) dell'AG si trova una regione ricca di pirimidine (C e U), o tratto polipirimidinico. A monte del tratto polipirimidinico si trova il punto di diramazione.

Una "sequenza accettore di splicing" è una sequenza nucleotidica che può funzionare come sito accettore all'estremità 3' dell'introne. Le sequenze di consenso e le frequenze delle regioni dei siti di splicing umani sono descritte in Ma, S.L., et al., 2015. PLoS One, 10(6), p.e0130729.

In particolare, la sequenza accettore di splicing può comprendere la sequenza nucleotidica (Y)nNYAG, dove n è 10-20, o una variante con un'identità di sequenza di almeno il 90% o almeno il 95%. In particolare, la sequenza accettore di splicing può comprendere la sequenza (Y)nNCAG, dove n è 10-20, o una variante con un'identità di sequenza di almeno il 90% o almeno il 95%.

Sequenza del segnale di degradazione

Le sequenze del segnale di degradazione sono preferibilmente la sequenza di CL1, CL2, CL6, CL9, CL10, CL11, CL12, CL15, CL16, SL17, SMN, CIITA, ODC7, ecDHFR, Mini ecDHFR. Tali sequenze comprendono o consistono preferibilmente nelle sequenze sotto riportate o nelle sequenze che codificano la sequenza sotto riportata.

CL2	SLISLPLPTRVKFSSLLLIRIMKIITMTFPKKLRS	SEQ ID NO:37
CL6	FYYPIWFARVLLVHYQ	SEQ ID NO:38
CL9	SNPFSSLFGASLLIDSVSLKSNWDTSSSSCLISFFSSVMFSSTTRS	SEQ ID NO:39
CL10	CRQRFSCHLTASYPQSTVTPFLAFLRRDFFFLRHNSAD	SEQ ID NO:40
CL11	GAPHVVLDFELRITNPLSHIQSVSLQITLIFCSLPSLILSKFLQV	SEQ ID NO:41
CL12	NTPLFSKSFSTTCGVAKKTLILAQISSLFLLSSNIAV	SEQ ID NO:42
CL15	PTVKNSPKIFCLSSSPYLAFNLEYLSLRIFSTLSKCSNTLLTSLS	SEQ ID NO:4
CL16	SNQLKRLWLWLLEVRSFDRTLRRPWIHLP	SEQ ID NO:44

SL17	SISFVIRSHASIRMGASNDFFHKLYFTKCLTSVILSKFLIHLLLRSTPRV	SEQ ID NO:45
ecDHFR	MISLIAALAVDYVIGMENAMPWNLPADLAWFKRNTLNKPVIMGRHT WESIGRPLPGRKNI ILSSQPSTDDRVTWVKSVDIAACGDVPEIMVIGGGRVIEQFLPKAQK LYLTHIDAEVE GDTHFPDYEPDDWESVFSEFHDADAQNSHSYCFEILERR	SEQ ID NO:46
Mini ecDHFR	MISLIAALAVDYVIGMENAMPWNLPADLAWFKRNTLNKPVIMGRHT WESIGRPLPGRKNI ILSSQPSTDDRVTWVKSVDIAACGDVPEIMVIGGGRVIEQFLP	SEQ ID NO:47
CIITA	RPGSTSPFAPSATDLPSMPEPALTSR	SEQ ID NO:48
SMN	YMSGYHTGYMEMLA	SEQ ID NO:49
ODC7	MSCAQES	SEQ ID NO:50

Per le sequenze da CL2 a SL 17 si veda Gilon, T., Chomsky, O. e Kulka, R.G. (1998), Degradation signals for ubiquitin system proteolysis in *Saccharomyces cerevisiae*. The EMBO Journal, 17: 2759-2766, <https://doi.org/10.1093/emboj/17.10.2759> (qui incorporato come riferimento)

5 mentre per le sequenze da ecDHFR a ODC7 si veda Tornabene P, Trapani I, Centrulo M, Marrocco E, Minopoli R, Lupo M, Iodice C, Gesualdo C, Simonelli F, Surace EM, Auricchio A. L'inclusione di un degrone riduce i livelli di inteine indesiderate dopo il trans-splicing proteico mediato da AAV nella retina. Mol Ther Methods Clin Dev. 2021 Oct 19;23:448-459. doi: 10.1016/j.omtm.2021.10.004. PMID: 34786437; PMCID: PMC8571531 (qui incorporato come
10 riferimento).

Elementi regolatori

Il costrutto dell'invenzione può comprendere uno o più elementi regolatori che possono agire a livello pre- o post-trascrizionale. Uno o più elementi regolatori possono facilitare l'espressione nelle cellule dell'invenzione.

Un "elemento regolatore" è una qualsiasi sequenza nucleotidica che facilita l'espressione di un polipeptide, ad esempio agisce per aumentare l'espressione di un trascritto o per migliorare la stabilità dell'mRNA. Gli elementi regolatori adatti comprendono, ad esempio, promotori, elementi potenziatori, elementi regolatori post-trascrizionali e siti di poliadenilazione.

- 5 L'invenzione riguarda anche i costrutti che possono comprendere elementi regolatori funzionali nella cellula ospite in cui il vettore che comprende il costrutto deve essere espresso. Una persona di comune abilità nell'arte può selezionare gli elementi regolatori da utilizzare in cellule ospiti appropriate, per esempio, cellule ospiti umane o di mammifero. Gli elementi regolatori comprendono, ad esempio, promotori, sequenze di terminazione della trascrizione, sequenze
10 di terminazione della traduzione, potenziatori, peptidi di segnale, segnali di degradazione ed elementi di poliadenilazione.

- Un costrutto dell'invenzione può facoltativamente contenere una sequenza di terminazione della trascrizione, una sequenza di terminazione della traduzione, una sequenza di peptidi di segnale, siti interni di ingresso al ribosoma (IRES), elementi potenziatori e/o elementi regolatori
15 post-trascrizionali come l'elemento regolatore post-trascrizionale (WPRE) del virus dell'epatite di marmotta (WHV). Le regioni di terminazione della trascrizione possono essere tipicamente ottenute dalla regione non tradotta nell'estremità 3' di una sequenza genica eucariotica o virale. Le sequenze di terminazione della trascrizione possono essere posizionate a valle di una sequenza codificante per garantire una terminazione efficiente. Nel sistema dell'invenzione è
20 tipicamente compreso un sito di terminazione della trascrizione.

Elementi regolatori post-trascrizionali

- I costrutti di acido nucleico della presente invenzione possono comprendere elementi regolatori post-trascrizionali. In particolare, la sequenza codificante la proteina è operativamente legata a uno o più elementi regolatori post-trascrizionali che possono migliorare l'espressione genica.
25 Il costrutto della presente invenzione può comprendere un elemento regolatore post-trascrizionale (WPRE) del virus dell'epatite della marmotta. La sequenza codificante dell'OAT è opportunamente legata a un WPRE.

- Le sequenze WPRE adatte sono ben note a chi è esperto del settore (si veda, ad esempio, Zufferey et al. (1999) Journal of Virology 73: 2886-2892; Zanta-Boussif et al. (2009) Gene
30 Therapy 16: 605-619). In genere, il WPRE è un WPRE di tipo selvatico o un WPRE mutante. Ad esempio, la WPRE può essere mutata per annullare la traduzione della proteina X del virus

dell'epatite della marmotta (WHX), ad esempio mutando il codone di inizio della traduzione della ORF di WHX.

Esempi

MATERIALI & METODI

5 Plasmidi effettori di Cas9 per lo screening dei gRNA

I gRNA sono stati progettati manualmente considerando la posizione della mutazione P23H bersaglio. I plasmidi effettori sono stati generati seguendo il protocollo del laboratorio di Zhang. In breve, i gRNA sono stati generati come oligonucleotidi di DNA a singolo filamento Fwd e Rev, che sono stati poi fosforilati e appaiati utilizzando la polinucleotide chinasi T4 e il T4 Ligation Buffer (NEB). Il corretto appaiamento del gRNA è stato confermato in un gel di acrilammide al 10 50% colorato con SYBR Gold Nucleic Acid Stain (ThermoFisher). I gRNA sono stati successivamente clonati nel plasmide px458 (#48138; Addgene) utilizzando i siti *Bbs* I.

Produzione dei plasmidi del vettore AAV

I plasmidi utilizzati per la produzione del vettore AAV sono derivati dal plasmide pAAV2.1 che 15 contiene gli ITR del sierotipo 2 di AAV. In particolare, gli inventori hanno utilizzato un plasmide pAAV2.1 generato dal nostro gruppo per una precedente pubblicazione (4). La sequenza esatta è riportata nel file di sequenza come riferimento.

Produzione e caratterizzazione del vettore AAV

I vettori AAV sono stati prodotti da Innovavector SRL mediante tripla trasfezione di cellule 20 HEK293 seguita da due cicli di purificazione con CsCl₂. Per ogni preparazione virale, i titoli fisici (GC/mL) sono stati determinati calcolando la media dei titoli ottenuti dall'analisi dot-blot e dalla quantificazione PCR con TaqMan (Applied Biosystems, Carlsbad, CA, USA). Le sonde utilizzate per le analisi dot-blot e PCR sono state progettate per essere analizzate con il promotore IRBP per il vettore pAAV2.1-IRBP-SpCas9-spA e con la regione bGHpA per i vettori di DNA donatori. 25 La lunghezza delle sonde variava tra 200 e 700 bp(12).

Cultura e trasfezione di cellule HEK293

Le cellule HEK293 sono state mantenute in DMEM contenente il 10% di siero fetale bovino (FBS) e 2 mM di L-glutamina (Gibco, Thermo Fisher Scientific, Waltham, MA, USA). Le cellule sono state piastrate in piastre a 6 pozzetti (1*10⁶ cellule/pozzetto) e trasfettate 16 ore dopo con i 30 plasmidi che codificano per Cas9, un plasmide modello che consiste nell'esone1, nell'introne1 e nell'esone2 della sequenza della rodopsina umana guidata da un promotore del citomegalovirus (CMV) e i diversi DNA donatori, utilizzando il metodo del fosfato di calcio (da 1 a 2 mg/1*10⁶

cellule); il terreno è stato sostituito 4 ore dopo. La quantità massima di materiale trasfettato è stato di 3 μ g. In tutti i casi, la quantità di DNA plasmidico è stata equilibrata tra i pozzetti, utilizzando un vettore vuoto quando necessario.

Istologia e microscopia a luce e fluorescenza

5 Per valutare l'espressione di eGFP dopo HITI *in vitro*, le cellule HEK293, piastrate in pozzetti da 6 a una densità di 1×10^6 , sono state trasfettate come precedentemente descritto. Settantadue ore dopo la trasfezione, le cellule sono state analizzate con un Apotome (Carl Zeiss, Oberkochen, Germania) dotato di software ZEN (Carl Zeiss) e utilizzando le impostazioni di eccitazione e rilevamento appropriate per EGFP.

10 Analisi citofluorimetrica

Le cellule HEK293, piastrate in piastre a 6 pozzetti, sono state lavate una volta con PBS, staccate con tripsina allo 0,05% di EDTA (Thermo Fisher Scientific, Waltham, MA USA), lavate due volte con PBS e risospese in una soluzione di selezione contenente PBS, 5% FBS e 2,5 mM EDTA. Le cellule sono state analizzate su un BD FACS ARIA III (BD Biosciences, San Jose, CA, USA) dotato
15 del software BD FACSDiva (BD Biosciences) utilizzando le impostazioni di eccitazione e rilevamento appropriate per EGFP. Il valore soglia per il rilevamento della fluorescenza è stato impostato su cellule non trasfettate e sono state analizzate almeno 10.000 cellule/campione. Almeno 50.000 cellule GFP+ sono state selezionate e utilizzate per l'estrazione del DNA.

Modelli animali

20 I topi sono stati ospitati presso lo stabulario TIGEM (Pozzuoli, Italia) e mantenuti in un ciclo di alternanza luce/buio di 12 ore. I topi hRHO-P23H-TagRFP(8) (indicati come hRHO-P23H) sono stati gentilmente forniti dal Prof. Theodore Wensel. I topi sono stati mantenuti incrociando femmine e maschi omozigoti. Gli animali eterozigoti sperimentali sono stati generati incrociando topi omozigoti P23H con topi C57BL/6. Il genotipo dei topi è stato confermato
25 mediante analisi PCR sul DNA genomico (estratto dalla punta della falange del topo). I topi omozigoti hanno presentato un prodotto di PCR di 975 bp, mentre i topi eterozigoti hanno presentato un prodotto di 975 bp e uno di 195 bp. I topi di tipo selvatico hanno presentato solo un prodotto PCR di 195 bp. Gli inneschi utilizzati per l'amplificazione PCR sono descritti nella tabella 2 come segue:

30

Nome degli Inneschi (primer)	Sequenza (5'-3')	SEQ ID NO:

P23H-RFP_GENO FP	GTTCCGGAAGTGCATGCTCACCAC	51
P23H-RFP_GENO RP	CCCACCAGGAGCAGCGCC	52

Tabella 1: Descrizione degli inneschi (primers) utilizzati nella PCR per rilevare i genotipi dei topi sopra descritti.

Iniezione sottoretinica di vettori AAV nei topi hRHO-P23H

Questo studio è stato condotto in conformità con l'Association for Research in Vision and Ophthalmology (ARVO) Statement for the Use of Animals in Ophthalmic and Vision Research e con il regolamento del Ministero della Salute italiano per le procedure sugli animali (numero di autorizzazione del Ministero della Salute: 252/2022-PR).

I topi (di 1 o 4 settimane) sono stati anestetizzati con un'iniezione intraperitoneale di 2 mL/100g di peso corporeo di ketamina/xylazina, quindi i vettori AAV2/8 sono stati somministrati per via subretinica attraverso un approccio trans-sclerale trans-coroidale, come descritto da Liang et al.(13,14). Gli occhi sono stati iniettati con 1 μ L di soluzione vettoriale. La dose di AAV2/8 (GC/occhio) era compresa tra $1,5 \cdot 10^9$ e $2,5 \cdot 10^9$ GC di ciascun vettore/occhio; pertanto, la co-iniezione ha comportato un massimo di $3 \cdot 10^9$ GC/occhio.

Registrazioni elettrofisiologiche

Per le analisi elettroretinografiche, i topi hRHO-P23H sono stati adattati al buio per 3 ore. I topi sono stati anestetizzati e posizionati in un apparato stereotassico, sotto una fioca luce rossa. Le pupille sono state dilatate con una goccia di tropicamide allo 0,5% (Visufarma, Roma, Italia) e la temperatura corporea è stata mantenuta a 37,5 gradi.

I lampi di luce sono stati generati da uno stimolatore Ganzfeld (CSO, Costruzione Strumenti Oftalmici, Firenze, Italia). I segnali elettrofisiologici sono stati registrati attraverso elettrodi a piastra d'oro inseriti sotto le palpebre inferiori a contatto con la cornea. Gli elettrodi di ciascun occhio erano riferiti a un elettrodo ad ago inserito sottocute a livello della corrispondente regione frontale. I diversi elettrodi erano collegati a un amplificatore a due canali. Dopo aver completato le risposte ottenute in condizioni di adattamento al buio (scotopico), la sessione di registrazione è proseguita con lo scopo di sezionare la via del cono che media la risposta alla luce (fotopica). Per minimizzare il rumore, le diverse risposte evocate dalla luce sono state mediate per ogni fase di luminanza. La risposta scotopica massima dei bastoncelli e dei coni è stata misurata in condizioni di buio (scotopica) con due flash di 0,7 Hz e un'intensità luminosa di 20 cd s/m², mentre le risposte fotopiche dei coni sono state isolate in condizioni di luce con

una luce bianca di fondo continua di 50 cd s/m², con 10 flash di 0,7 Hz e un'intensità luminosa di 20 cd s/m².

Criosezioni di retina e imaging a fluorescenza

Per valutare l'espressione di eGFP nella retina dopo HITI in sezioni istologiche, ai topi hRHO-P23H sono stati iniettati sottoretinalmente i vettori AAV IRBP-Cas9 e DNA donatore. Un mese dopo, i topi sono stati sacrificati e gli occhi sono stati fissati in paraformaldeide al 4% per una notte e infiltrati con saccarosio al 30% per una notte; la cornea e il cristallino sono stati quindi sezionati e i globi oculari sono stati incorporati in un composto a temperatura di taglio ottimale (matrice O.C.T.; Kaltek, Padova, Italia). Le criosezioni retiniche seriali di dieci micrometri di spessore sono state tagliate lungo il meridiano orizzontale, distribuite progressivamente su vetrini e montate con Vectashield con DAPI (Vector Lab, Peterborough, Regno Unito). Quindi, le criosezioni sono state analizzate al microscopio confocale LSM-700 (Carl Zeiss, Oberkochen, Germania), utilizzando le impostazioni di eccitazione e rilevazione appropriate rispettivamente per eGFP, RFP e DAPI. Per valutare l'efficienza dell'HITI nelle criosezioni retiniche di topo dopo la somministrazione di AAV, è stata selezionata l'area maggiormente trasdotta di tre sezioni/occhio, acquisita a 40 ingrandimenti e poi analizzata con il software ImageJ (<http://rsbweb.nih.gov/ij/>). Per ogni immagine sono state contate manualmente e con il plugin ITCN di ImageJ almeno 500 PR, identificate dalla colorazione DAPI. Le PR con segnale compatibile con l'espressione di eGFP sono state identificate in modo inequivocabile in base alla loro forma osservata negli z-stack delle sezioni analizzate e alla presenza di segmenti esterni eGFP+ve.

Estrazione del DNA

I campioni (cellule HEK293 selezionate GFP+, tessuto retinico) sono stati lisati in un tampone di lisi commerciale (GeneArt™ Genomic Cleavage Detection Kit, Invitrogen, Carlsbad, California, Stati Uniti) o in un tampone di lisi convenzionale per l'estrazione del DNA dai tessuti (400mM NaCl, 1% SDS, 20mM TRIS-CL (pH 8.0), 5mM EDTA (pH 8.0)). I tamponi di lisi sono stati integrati con proteinasi K, che è stata inattivata dopo una lisi di 15 minuti a 80 gradi. Da 50 a 200ng di DNA sono stati utilizzati per l'amplificazione PCR della regione che comprende il sito bersaglio di Cas9 (il primo introne di RHO) dal plasmide pCMV-hRHO (Esone1-Introne1-Esone2) o dal genoma del topo, rispettivamente. Gli inneschi (primer) utilizzati sono riportati nella Tabella 1:

Nome dell' Innesco (Primer)	Sequenza (5'-3')	SEQ ID NO:
Indel FW	TTGGAAGCCCGCATCTATC	53
Indel REV	GCCACATCCCTAAATGAGTC	54

Tabella 2: Gli inneschi hRHO-P23H-Indel hanno prodotto prodotti di PCR di 461.

Caratterizzazione della giunzione HITI

Amplificazione della giunzione tramite PCR

- 5 Il DNA estratto dalla retina è stato utilizzato per l'amplificazione PCR delle giunzioni HITI. Sono state amplificate sia le giunzioni all'estremità 5' che all'estremità 3' dell'integrazione. Per la giunzione all'estremità 5', gli inventori hanno utilizzato un innesco forward che riconosce la regione a valle del primo introne del gene hRHO prima del sito di taglio e un innesco reverse che riconosce il sito accettore di splicing - 3XFLAG sul DNA donatore. Per la giunzione
10 all'estremità 3' gli inventori hanno progettato un innesco forward che riconosce la sequenza polyA di bGH del DNA donatore e un innesco reverse che riconosce la sequenza all'interno dell'introne 1 di RHO umano dopo il sito di taglio.

Nome dell' Innesco (Primer)	Sequenza (5'-3')	SEQ ID NO:
Intron1 SeqFP2	CTCTCAGCCCCTGTCCTCAG	55
NMD-HITI_5JnRP(short)	CGTGGTCCTTATAGTCTACACCTGT	56
3'Int Junc FP_pA	GCCAGCCATCTGTTGTTTGC	57
3'Int Jn RP (HITI)	CCTGCCTCAGTTTCTCTCTGTTA	58

15 **Tabella 3: Inneschi utilizzati per amplificare le giunzioni alle estremità 5' e 3' dopo HITI nel locus RHO**

Estrazione di RNA ed espressione di hRHO

- L'RNA totale è stato estratto con RNeasy MiniKit (QIAGEN) da cellule HEK293 selezionate EGFP+/DsRed- e EGFP+/DsRed+. L'RNA (5-15 ng) è stato utilizzato come stampo per la RT-qPCR One-Step (NEB, Massachusetts, USA) secondo le istruzioni del produttore utilizzando il
20 LightCycler 96 (Roche Molecular Systems, Inc.). I livelli di espressione di hRHO sono stati normalizzati rispetto al corrispondente gene costitutivo (ACTB). L'analisi della quantificazione

relativa è stata effettuata con il metodo 2(- $\Delta\Delta C_t$). Gli inneschi utilizzati per l'amplificazione qPCR in tempo reale sono riportati nella Tabella 2:

Nome dell'Innesco (Primer)	Sequenza (5'-3')	SEQ ID NO:
RThRHO_Fw	ATCATGGTCATCGCTTTCCT	59
RThRHO_Rv	TCATGAAGATGGGACCGAAG	60
FwhActinQPCR	GGGAGAAGATGACCCAGATC	61
RvhActinQPCR	GGATAGCACAGCCTGGATAG	8

Tabella 2: inneschi utilizzati nella real-time qPCR per rilevare trascritti di hRHO

5 Risultati

L'ottimizzazione del donatore HITI determina una maggiore efficienza *in vitro*

Per aumentare l'efficienza di HITI, gli inventori hanno progettato un costrutto HITI ottimizzato (optm.HITI) sostituendo i tre codoni di stop in tandem e i siti di inizio traduzione (IRES/Kozak) con il segnale di degradazione CL1 marcato con 3XFLAG; questo segnale è fuso con un sito di taglio attivo della furina per una maggiore degradazione della proteina RHO tronca. Questo è stato preceduto da una sequenza accettore di splicing a monte per un efficiente splicing nel sito mirato del locus RHO. Dopo il taglio da parte di CRISPR-Cas9 e l'integrazione nel giusto orientamento, il DNA donatore sostituirà la sequenza RHO endogena nel locus genomico. Poiché il DNA donatore è una sequenza di coding senza promotore (cds), verrà espresso solo dopo una corretta integrazione dal promotore endogeno (Figura 2a). Oltre alla cds RHO, il donatore HITI porta anche la cds eGFP, per cui le cellule che esprimono eGFP ci permetteranno di determinare l'efficienza dell'integrazione. A tale scopo, le cellule HEK 293 sono state trasfettate con

- i) il plasmide Cas9 sotto il controllo di un promotore CMV
- ii) un plasmide modello che codifica l'esone 1, l'introne 1 e l'esone 2 di RHO umano guidato da un promotore CMV e privo del segnale di poli-adenilazione come descritto in precedenza e
- iii) il plasmide donatore HITI di nuova concezione che consiste nella cassetta di espressione U6 comprendente l'RNA guida (gRNA) per il primo introne o una sequenza di RNA disordinata.

Settantadue ore dopo la trasfezione, le cellule sono state analizzate al microscopio a fluorescenza con filtri di eccitazione ed emissione appropriati (per rilevare le cellule positive alla eGFP) e raccolte per l'analisi quantitativa dei trascritti RHO mediante qPCR.

L'inclusione del segnale di degradazione CL1 determina una degradazione selettiva della proteina RHO endogena troncata all'estremità 5' senza influenzare la produzione di proteine a lunghezza completa (Figura 2b). Questo segnale di degradazione è ulteriormente fuso con P2A(12), una sequenza di salto ribosomiale, che favorisce la traduzione della sequenza codificante di RHO fusa con una proteina reporter eGFP tramite T2A e seguita da WPRE e dalla sequenza poli-A dell'ormone della crescita bovino (BGH). Gli inventori hanno osservato che i livelli di trascrizione di hRHO erano circa 2,2 volte più alti nelle cellule trasfettate con il nuovo donatore HITI ottimizzato rispetto alle cellule trasfettate con il donatore classico, 3X STOP (Figura 2c) e ciò è correlato a una maggiore intensità di fluorescenza delle cellule trasfettate con il donatore HITI ottimizzato rispetto alle cellule con il donatore HITI classico, 3XSTOP.

Valutazione dell'efficienza HITI in vivo

Gli inventori hanno valutato l'efficienza dell'HITI in un modello murino knock-in P23H della retinite pigmentosa autosomica dominante (RP4), recentemente descritto, in cui l'allele RHO endogeno è stato sostituito da un RHO umano marcato con una proteina fluorescente rossa (RFP) che porta la mutazione P23H (hRHO-P23H-tagRFP)(11).

Gli inventori hanno effettuato iniezioni sottoretiniche in topi eterozigoti hRHO-P23H-tagRFP a 4 settimane di età, con due diversi vettori AAV8, uno che codifica per la nucleasi Sp. Cas9 sotto il controllo del promotore specifico del fotorecettore, Interphotoreceptor Retinoid-Binding Protein (IRBP), e un secondo AAV che trasporta il DNA donatore HITI (che trasporta sia RHO che GFP per riconoscere i fotorecettori dove si è verificata l'integrazione) a una dose di $1,5 \times 10^9$ di ciascun vettore/occhio. L'occhio controlaterale è servito da controllo. Gli animali sono stati sacrificati un mese dopo il trattamento e gli occhi sono stati prelevati per ulteriori analisi. L'efficienza dell'HITI è stata valutata mediante microscopia a fluorescenza su sezioni retiniche OCT, contando il numero di cellule positive alla GFP nello strato nucleare esterno (ONL) rispetto ai nuclei colorati con DAPI. Gli inventori hanno osservato un'efficienza HITI fino al $12 \pm 8\%$ nell'area trasdotta ($n=4$; Figura 3a). Inoltre, gli inventori hanno anche valutato l'integrazione di HITI nel sito bersaglio eseguendo l'amplificazione HITI PCR di entrambe le giunzioni all'estremità 5' e 3' tra il donatore HITI e il locus endogeno. L'analisi di sequenziamento Sanger ha dimostrato che l'HITI si è verificato con precisione.

In prospettiva, gli inventori intendono valutare il miglioramento del fenotipo retinico (elettroretinogramma ERG, acuità visiva e morfologia) a scadenze avanzate e i potenziali eventi di editing fuori bersaglio in topi sia eterozigoti che omozigoti hRHO-P23H-tagRFP. I dati preliminari nei topi omozigoti hRHO-P23H-tagRFP iniettati per via subretinica a p7 (1 settimana di età) con AAV-HITI gRNA (HITI ottimizzato) mostrano una migliore risposta ERG rispetto agli occhi trattati con AAV-HITI scRNA a 1 mese dal trattamento (Figura 3b).

Bibliografia

1. Trapani I, Auricchio A. Seeing the Light after 25 Years of Retinal Gene Therapy. Trends Mol Med [Internet]. 2018 Aug 1;24(8):669–81. Available from: <https://doi.org/10.1016/j.molmed.2018.06.006>
2. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (1979) [Internet]. 2013 Feb 15;339(6121):819–23. Available from: <https://doi.org/10.1126/science.1231143>
3. Yanik M, Müller B, Song F, Gall J, Wagner F, Wende W, et al. In vivo genome editing as a potential treatment strategy for inherited retinal dystrophies. Prog Retin Eye Res [Internet]. 2017;56:1–18. Available from: <https://www.sciencedirect.com/science/article/pii/S1350946216300441>
4. Hartong DT, Berson EL, Dryja TP. Retinitis pigmentosa. The Lancet [Internet]. 2006 Nov 18;368(9549):1795–809. Available from: [https://doi.org/10.1016/S0140-6736\(06\)69740-7](https://doi.org/10.1016/S0140-6736(06)69740-7)
5. Kaushal S, Khorana HG. Structure and Function in Rhodopsin. 7. Point Mutations Associated with Autosomal Dominant Retinitis Pigmentosa. Biochemistry [Internet]. 1994 May 1;33(20):6121–8. Available from: <https://doi.org/10.1021/bi00186a011>
6. Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature [Internet]. 2016;540(7631):144–9. Available from: <https://doi.org/10.1038/nature20565>
7. Tornabene P, Ferla R, Llado-Santaeularia M, Centrulo M, Dell'Anno M, Esposito F, et al. Therapeutic homology-independent targeted integration in retina and liver. Nat Commun [Internet]. 2022;13(1):1963. Available from: <https://doi.org/10.1038/s41467-022-29550-8>
8. Robichaux MA, Nguyen V, Chan F, Kailasam L, He F, Wilson JH, et al. Subcellular localization of mutant P23H rhodopsin in an RFP fusion knock-in mouse model of retinitis

pigmentosa. *Dis Model Mech* [Internet]. 2022 May 6;15(5):dmm049336. Available from: <https://doi.org/10.1242/dmm.049336>

9. Trapani I, Toriello E, de Simone S, Colella P, Iodice C, Polishchuk E v, et al. Improved dual AAV vectors with reduced expression of truncated proteins are safe and effective in the retina of a mouse model of Stargardt disease. *Hum Mol Genet* [Internet]. 2015 Dec 1;24(23):6811–25. Available from: <https://doi.org/10.1093/hmg/ddv386>

10. Gilon T, Chomsky O, Kulka RG. Degradation signals for ubiquitin system proteolysis in *Saccharomyces cerevisiae*. *EMBO J* [Internet]. 1998 May 15;17(10):2759–66. Available from: <https://doi.org/10.1093/emboj/17.10.2759>

11. Thomas G. Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nat Rev Mol Cell Biol* [Internet]. 2002 Oct;3(10):753–66. Available from: <https://pubmed.ncbi.nlm.nih.gov/12360192>

12. Doria M, Ferrara A, Auricchio A. AAV2/8 Vectors Purified from Culture Medium with a Simple and Rapid Protocol Transduce Murine Liver, Muscle, and Retina Efficiently. *Hum Gene Ther Methods* [Internet]. 2013 Sep 12;24(6):392–8. Available from: <https://doi.org/10.1089/hgtb.2013.155>

13. Liang FQ, Dejneka NS, Cohen DR, Krasnoperova N v., Lem J, Maguire AM, et al. AAV-Mediated Delivery of Ciliary Neurotrophic Factor Prolongs Photoreceptor Survival in the Rhodopsin Knockout Mouse. *Molecular Therapy*. 2001 Feb 1;3(2):241–8.

14. Liang FQ, Anand V, Maguire AM, Bennett J. Intraocular Delivery of Recombinant Virus. In: Rakoczy PE, editor. *Vision Research Protocols* [Internet]. Totowa, NJ: Humana Press; 2001. p. 125–39. Available from: <https://doi.org/10.1385/1-59259-085-3:125>

RIVENDICAZIONI

1. Un sistema di editing genetico comprendente:

a) un acido nucleico donatore comprendente:

- una sequenza di segnale di degradazione,
- un sito di taglio enzimatico,
- una sequenza di salto ribosomiale,
- una sequenza di DNA esogeno,

in cui detto acido nucleico donatore è affiancato al 5' e 3' da sequenze bersaglio invertite;

b) un oligonucleotide complementare alla sequenza bersaglio e

c) una nucleasi che riconosce la sequenza bersaglio.

2. Il sistema di editing genetico della rivendicazione 1, in cui detto acido nucleico donatore comprende anche una sequenza accettore di splicing, preferibilmente al 5' della sequenza di segnale di degradazione.

3. Il sistema di editing genetico della rivendicazione 1 o 2, in cui la sequenza del segnale di degradazione è CL1 e/o in cui il sito di taglio enzimatico è un sito di taglio della furina, preferibilmente attivo e/o ottimizzato e/o in cui la sequenza di salto ribosomiale è una sequenza di salto ribosomiale del Teschovirus-1 suino 2A (P2A).

4. Il sistema di editing genetico secondo una qualsiasi delle rivendicazioni precedenti, in cui la sequenza bersaglio è una sequenza compresa nel gene della rodopsina (Rho),

preferibilmente dove la sequenza bersaglio è compresa all'interno di:

- il primo introne del gene RHO, preferibilmente di origine umana, murina o suina,
- il primo esone del gene RHO, preferibilmente di origine umana, murina o di maiale,

e/o dove la sequenza di DNA esogeno comprende una sequenza codificante una proteina terapeutica, ad esempio la rodopsina, preferibilmente essa comprende uno o più esoni della rodopsina o suoi frammenti,

e/o dove la sequenza bersaglio è un sito bersaglio di un RNA guida (gRNA),

e/o dove detto oligonucleotide complementare alla sequenza bersaglio è un RNA guida che ibrida con una sequenza bersaglio di un gene o con il suo filamento complementare,

e/o detto oligonucleotide complementare alla sequenza bersaglio è sotto il controllo di un promotore, preferibilmente un promotore U6,

e/o dove le sequenze di bersaglio invertite sono una sequenza invertita rispetto a una sequenza bersaglio e/o comprendono una sequenza PAM, preferibilmente al suo 3'.

5. Il sistema di editing genetico secondo una qualsiasi delle rivendicazioni precedenti in cui detto acido nucleico donatore comprende uno o più di:

- un linker, preferibilmente tra il sito di taglio enzimatico e la sequenza di salto ribosomiale;

5 - un'ulteriore sequenza di salto ribosomiale, preferibilmente localizzata al 3' della sequenza di DNA esogeno;

- un elemento regolatore post-trascrizionale, preferibilmente localizzato al 3' della sequenza di DNA esogeno o dell'ulteriore sequenza di salto ribosomiale;

10 - una sequenza di terminazione della trascrizione, preferibilmente localizzata al 3' dell'elemento regolatore post-trascrizionale o al 3' della sequenza di DNA esogeno o dell'ulteriore sequenza di salto ribosomiale,

preferibilmente dove detto elemento regolatore post-trascrizionale è l'elemento regolatore post-trascrizionale del virus dell'epatite di marmotta (WPRE) e/o detta sequenza di terminazione della trascrizione è una sequenza di segnale di poliadenilazione, preferibilmente la polyA dell'ormone della crescita bovino (BGH polyA) e/o detta ulteriore sequenza di salto ribosomiale è la sequenza T2A.

15

6. Il sistema di editing genetico secondo una qualsiasi delle precedenti rivendicazioni in cui:

a) la sequenza di salto ribosomiale comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con SEQ ID NO: 1

20 (GCCACCAACTTCTCCCTGCTGAAGCAGGCCGCGACGTGGAGGAGAACCCCGGCCCC) o con

SEQ ID NO: 2

(GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGA CCT) o con una sequenza che codifica per SEQ ID NO: 3 (GSG) E G R G S L L T C G D V E E

N P G P o con una sequenza che codifica per SEQ ID NO: 4 (GSG) A T N F S L L K Q A G D

25 V E E N P G P o suoi frammenti funzionali e/o

b) la sequenza bersaglio invertita comprende o ha essenzialmente una sequenza che ha almeno il 95% di identità con la SEQ ID NO: 5 (ACACCAGGAGACTTGGAAACG) o suoi frammenti funzionali e opzionalmente comprende la sequenza SpCas9 PAM (CGG) e/o

c) l'RNA guida comprende o ha essenzialmente una sequenza che ha almeno il 95% di identità con la SEQ ID NO: 5 (ACACCAGGAGACTTGGAAACG), o suoi frammenti funzionali

30

d) l'oligonucleotide complementare alla sequenza bersaglio comprende o ha essenzialmente una sequenza che ha almeno il 95% di identità con la SEQ ID NO: 5 (ACACCAGGAGACTTGAACG) o suoi frammenti funzionali e/o

e) la sequenza di segnale di degradazione comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con SEQ ID NO: 6 (gcctgcaagaactggttcagcagcctgagccacttcgtgatccacctg) e/o

f) il sito di taglio enzimatico comprende o ha essenzialmente una sequenza che ha almeno il 95% di identità con SEQ ID NO: 7 (CGAAAAAGAAGA) e/o

g) il linker comprende o ha essenzialmente una sequenza che ha almeno il 95% di identità con ggaagcgga e/o

h) la sequenza accettore di splicing comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con SEQ ID NO: 9 (GATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGTGT) e/o

i) la sequenza di DNA esogeno comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con almeno una delle seguenti sequenze: SEQ ID NO: 10

(ATGAATGGCACAGAAGGCCCTAACTTCTACGTGCCCTTCTCCAATGCGACGGGTGTGGTACGCAGCCC
CTTCGAGTACCCACAGTACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTGGCCGCCTACATGTTTCT
GCTGATCGTGCTGGGCTTCCCCATCAACTTCCTCACGCTCTACGTCACCGTCCAGCACAAGAAGCTGCG
CACGCCTCTCAACTACATCCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGGTCCTAGGTGGCTTCACC
AGCACCTCTACACCTCTCTGCATGGATACTTCGTCTTCGGGCCCCACAGGATGCAATTTGGAGGGCTTC
TTTGCCACCCTGGGCG),

SEQ ID NO: 11

(GTGAAATTGCCCTGTGGTCCTTGGTGGTCCTGGCCATCGAGCGGTACGTGGTGGTGTGTAAGCCCAT
GAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGGCGTTGCCTTCACCTGGGTGCATGGCGCTGG
CCTGCGCCGCACCCCCACTCGCCGGCTGGTCCAG);

SEQ ID NO: 12

(GTACATCCCCGAGGGCCTGCAGTGCTCGTGTGGAATCGACTACTACACGCTCAAGCCGGAGGTCAAC
AACGAGTCTTTTGTCTATCTACATGTTTCGTGGTCCACTTCACCATCCCCATGATTATCATCTTTTCTGCTA
TGGGCAGCTCGTCTTCACCGTCAAGGAG);

SEQ ID NO: 13

(GCCGCTGCCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTCACCCGCATGGTC
ATCATCATGGTCATCGCTTTCCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCATTCTACATCTTCACCC

ACCAGGGCTCCAACTTCGGTCCCATCTTCATGACCATCCCAGCGTTCTTTGCCAAGAGCGCCGCCATCTA
CAACCCTGTCATCTATATCATGATGAACAAGCAG);

SEQ ID NO: 14

(TTCCGGAAGTGCATGCTCACCACCATCTGCTGCGGCAAGAACCCACTGGGTGACGATGAGGCCTCTGC

5 TACCGTGTCCAAGACGGAGACGAGCCAGGTGGCACCAGCA) and/or

j) L'elemento di regolazione post-trascrizionale del virus dell'epatite di marmotta (wpre) comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con SEQ ID NO: 15

(Taagcttggatccaatcaacctctggattacaaaattgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtgga
10 tacgctgctttaatgcctttgtatcatgctattgcttcccgatggctttcattttctcctccttgataaatcctgggtgctgtctctttatgag
gagttgtggccggtgtcaggcaacgtggcggtgtgactgtgtttgctgacgcaacccccactgggtggggcattgccaccacgtg
cagctcctttccgggactttcgctttccccctcctattgccacggcggaactcatcgccgctgccttggcgctgctggacaggggctc
ggctgttgggcactgacaattccgtggtgtgtcggggaaatcatcgctcctttccttggctgctgcctgtgttgccacctggattctgcgc
gggacgtccttctgctacgtcccttcggccctcaatccagcggaccttcttcccggcgctgctgccggctctgcggcctcttccgcgtct
15 tcg) e/o

k) il segnale di poliadenilazione dell'ormone della crescita bovino (BGH pA) comprende o ha essenzialmente una sequenza che ha almeno l'80% di identità con SEQ ID NO: 16

(GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCGTGCCTTCCTTGACCCTGG
AAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTC
20 ATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGC
ATGCTGGGGA).

7. Un vettore che comprende il sistema di editing genetico secondo una qualsiasi delle rivendicazioni 1-6 o l'acido nucleico donatore e/o l'oligonucleotide complementare alla sequenza bersaglio e/o una nucleasi che riconosce la sequenza bersaglio come definita in una
25 qualsiasi delle rivendicazioni 1-6, preferibilmente dove il vettore è un vettore virale, preferibilmente selezionato dal gruppo costituito da: vettore adeno-associato (AAV), vettore adenovirale, vettore lentivirale, vettore retrovirale, o un vettore non virale, preferibilmente selezionato tra un veicolo di rilascio basato su polimeri, particelle, lipidi, peptidi o loro combinazioni, come polimeri cationici, micelle, liposomi, esosomi, microparticelle e
30 nanoparticelle, comprese le nanoparticelle lipidiche (LNP),
preferibilmente il vettore virale comprende anche una sequenza nucleotidica di ripetizione 5'-terminale (5'-TR) e una sequenza nucleotidica di ripetizione 3'-terminale (3'-TR), preferibilmente

il 5'-TR è una sequenza nucleotidica di ripetizione invertita al 5' (5'-ITR) e il 3'-TR è una sequenza nucleotidica di ripetizione invertita al 3' (3'-ITR), preferibilmente gli ITR derivano dallo stesso sierotipo virale o da sierotipi virali diversi, preferibilmente il virus è un AAV, preferibilmente del sierotipo 2.

- 5 8. Una cellula ospite che comprende il sistema di editing genetico secondo una qualsiasi delle rivendicazioni 1-6 o il vettore secondo la rivendicazione 7.
9. Una particella virale che comprende il sistema di editing genetico secondo una qualsiasi delle rivendicazioni 1-6 o un vettore secondo la rivendicazione 7, preferibilmente dove la particella virale comprende proteine del capsido di un AAV, preferibilmente dove la particella virale comprende proteine del capsido di un AAV di un sierotipo selezionato da uno o più del gruppo costituito da AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8 AAV9 e AAV 10, preferibilmente dal sierotipo AAV2 o AAV8.
- 10 10. Una composizione farmaceutica che comprende uno dei seguenti: sistema di editing genetico secondo una qualsiasi delle rivendicazioni 1-6, o un vettore secondo la rivendicazione 7, una cellula ospite secondo la rivendicazione 8, una particella virale secondo la rivendicazione 9 e un vettore farmaceuticamente accettabile.
- 15 11. Il sistema di editing genetico secondo una qualsiasi delle rivendicazioni 1-6, o un vettore secondo la rivendicazione 7, una cellula ospite secondo la rivendicazione 8, una particella virale secondo la rivendicazione 9 o una composizione farmaceutica secondo la rivendicazione 10 per l'uso come medicinale, preferibilmente per uso nel trattamento di una malattia genetica e/o per uso nel trattamento di malattie ereditarie autosomiche dominanti in cui entrambi gli alleli mutanti e wildtype sono sostituiti con una copia corretta del gene fornito dal DNA donatore o per uso nel trattamento di malattie ereditarie e comuni dovute a guadagno di funzione tossica, preferibilmente dette malattie comprendono la distrofia retinica, preferibilmente la distrofia retinica è selezionata tra retinite pigmentosa, distrofia del cono o distrofia cono-bastoncello, degenerazione maculare, ad es. g. malattia di Stargardt (ELOVL4), Von-Hippel Lindau, retinoblastoma, RP4 (vedi RHO; OMIM: 180380), RP63 (vedi OMIM: 614494), CORD1 (distrofia dei cono-bastoncelli 1; vedi OMIM: 600624), CORD17 (distrofia dei cono-bastoncelli 17; vedi OMIM: 615163), BEST1 (bestrofina-1; malattia di Best; proteina 2 della distrofia maculare vitelliforme; vedi OMIM : 607854), OPA1 (GTPasi simile alla dinamina like mitocondriale OPA1; vedi OMIM : 605290), malattie neuronali, epatiche, lipofuscinosi (malattia di Batten e altre),
- 25
- 30

preferibilmente per uso nel trattamento di patologie oculari ereditate in modo dominante, ad es. degenerazione retinica, preferibilmente retinite pigmentosa, malattie neuronali ed epatiche.

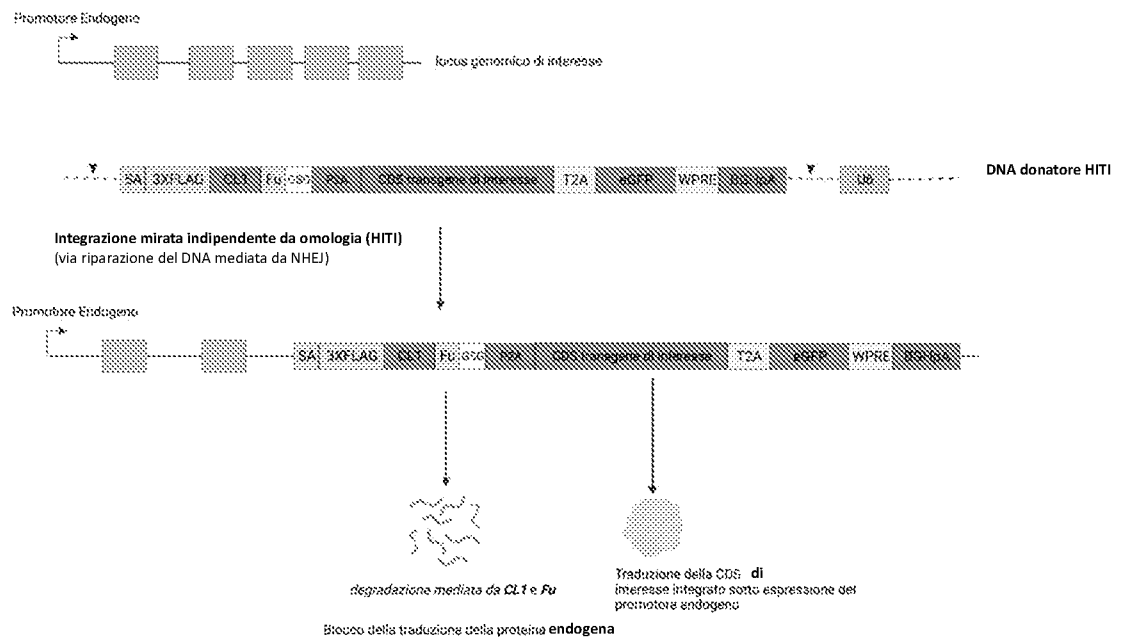
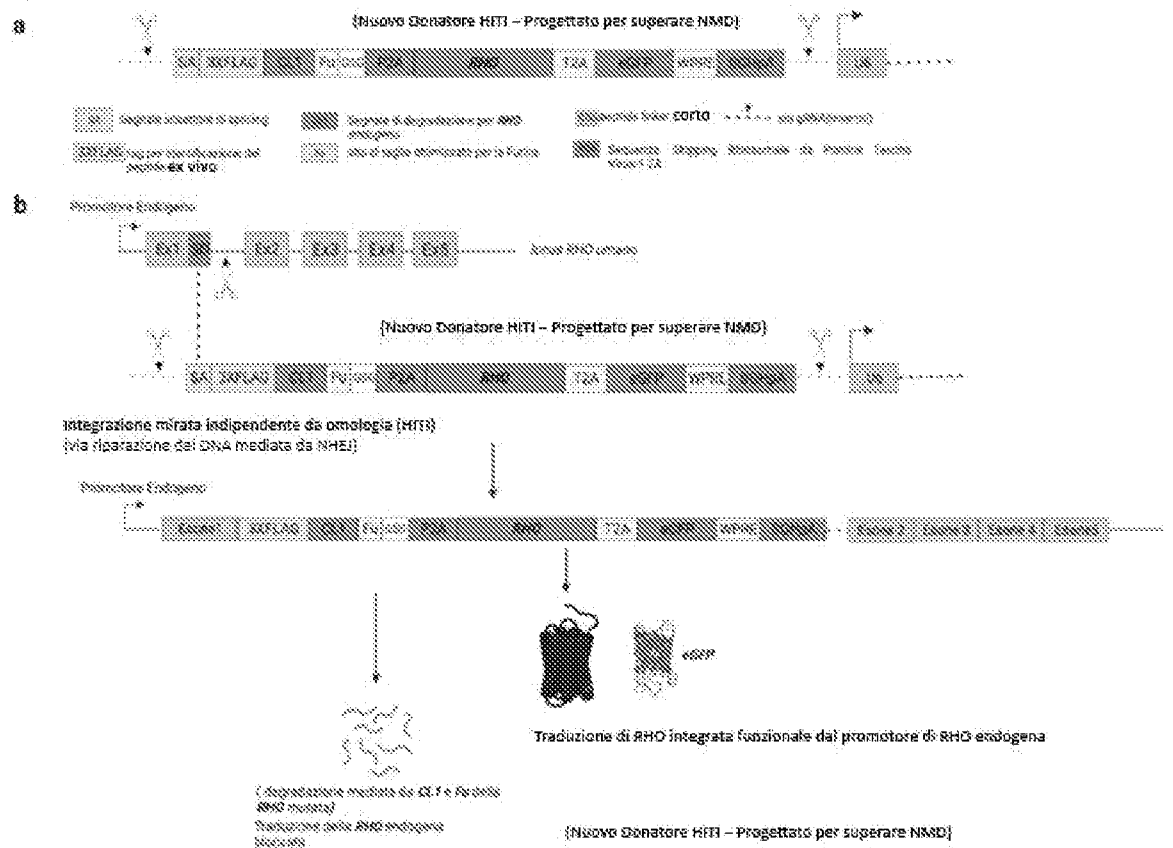
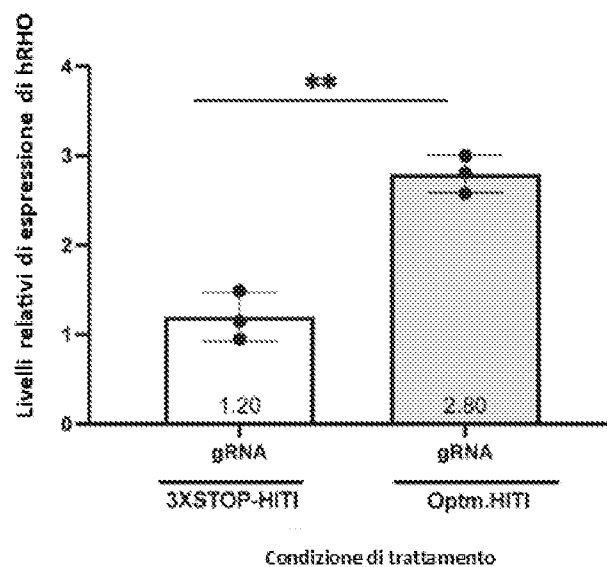


Figura 1

Figura 2



C



a

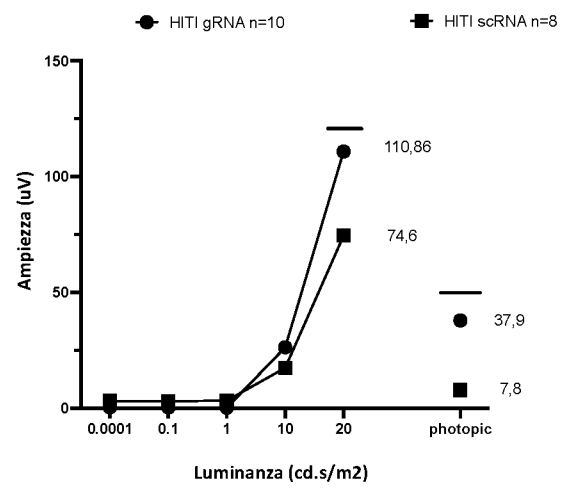
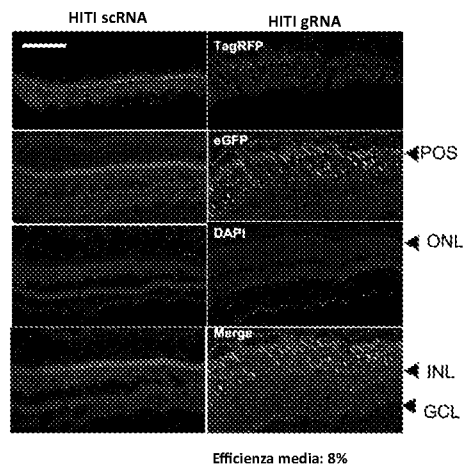


Fig. 3