

U001.D1.SM.61E

Traduzione Brevetto Europeo N. 3445388, a nome Vertex Pharmaceuticals Incorporated, dal titolo: “MATERIALI E METODI PER IL TRATTAMENTO DI EMOGLOBINOPATIE”

Descrizione

Campo

La presente descrizione fornisce materiali e metodi per il trattamento di pazienti affetti da emoglobinopatie, sia *ex vivo* che *in vivo*. Inoltre, la presente descrizione fornisce materiali e metodi per la delezione, modulazione o inattivazione di una sequenza di controllo di trascrizione di un gene di linfoma a cellule B 11A (BCL1 1A) in una cellula mediante editing genomico.

Lista di sequenze

Questo brevetto contiene una Lista di sequenze in forma leggibile da computer (nome di file: 160077PCT Sequence Listing; 14.446.299 bytes – file di testo ASCII; creato in data 7 aprile 2017).

Contesto

Emoglobinopatie includono anemie di origine genetica, che danno luogo a una diminuita produzione e/o a una aumentata distruzione di globuli rossi. Questi disturbi includono anche difetti genetici, che danno luogo alla produzione di emoglobine abnormi con una incapacità associata di mantenere la concentrazione di ossigeno. Molti di questi disturbi sono indicati come β -emoglobinopatie a causa della loro incapacità di produrre proteine β -globine normali in quantità sufficienti o incapacità di produrre interamente proteine β -globine normali. Per esempio, β -talassemie risultano da un difetto parziale o

U001.D1.SM.61E

completo nell'espressione del gene della β -globina, portando a una emoglobina adulta carente o assente (HbA). Anemia drepanocitica o falciforme risulta da una mutazione puntiforme nel gene strutturale della β -globina, che porta alla produzione di una emoglobina anormale (HbS) (Atweh, Semin. Hematol. 38(4):367-73 (2001)). Emoglobinopatie danno luogo a una riduzione nella capacità di trasportare ossigeno del sangue, che può portare a sintomi quali stanchezza, stordimento e dispnea, in particolare durante l'esercizio.

Per pazienti ai quali è stata diagnosticata una emoglobinopatia, attualmente sono disponibili soltanto pochi trattamenti sintomatici, come una trasfusione di sangue, per aumentare i livelli di ossigeno nel sangue.

Ingegneria genomica si riferisce alle strategie e tecniche per la modificazione mirata, specifica dell'informazione genetica (genoma) di organismi viventi. L'ingegneria del genoma è un campo di ricerca molto attivo a causa dell'ampia gamma di possibili applicazioni, in particolare nelle aree della salute umana; la correzione di un gene che reca una mutazione dannosa, per esempio, o per esplorare la funzione di un gene. Tecnologie precedenti sviluppate per inserire un transgene in una cellula vivente erano spesso limitate dalla natura casuale dell'inserimento della nuova sequenza nel genoma. Inserzioni casuali nel genoma possono portare a una interruzione della normale regolazione di geni adiacenti, portando a severi effetti indesiderati. Inoltre, tecnologie di integrazione casuali offrono una scarsa riproducibilità, dal momento che non vi è garanzia che la sequenza sarebbe inserita allo stesso posto in due cellule differenti. Strategie recenti di ingegneria del genoma, come ZFN, TALEN, HE e

U001.D1.SM.61E

MegaTAL, consentono di modificare una specifica area del DNA, aumentando quindi la precisione della correzione o dell'inserzione rispetto a tecnologie precedenti. Queste piattaforme più recenti offrono un grado molto superiore di riproducibilità, ma hanno ancora loro limitazioni.

Nonostante sforzi da ricercatori e professionisti medici nel mondo che stanno cercando di affrontare emoglobinopatie, permane ancora un'esigenza critica per lo sviluppo di trattamenti sicuri ed efficaci per emoglobinopatie.

Sommario

La presente descrizione presenta un approccio per affrontare la base genetica di emoglobinopatie. Utilizzando strumenti di ingegneria del genoma per creare cambiamenti permanenti al genoma che possono cancellare, modulare o inattivare una sequenza di controllo di trascrizione del gene BCL11A con un singolo trattamento, la terapia risultante può migliorare gli effetti di emoglobinopatie.

La presente invenzione è definita dalle rivendicazioni allegate.

Sono qui forniti metodi cellulari, *ex vivo* e *in vivo* per creare cambiamenti permanenti nel genoma cancellando, modulando o inattivando una sequenza di controllo di trascrizione del gene BCL11A, che possono venire utilizzati per trattare emoglobinopatie. Sono inoltre qui forniti componenti, kit e composizioni per effettuare tali metodi. Sono anche descritte cellule prodotte mediante tali metodi. Esempi di emoglobinopatie possono essere anemia falciforme e talassemia (α , β , δ , γ , e loro combinazioni).

È qui fornito un metodo per editare un gene di linfoma a cellule B 11A (BCL1

U001.D1.SM.61E

1A) in una cellula umana mediante editing genomico, il metodo comprendendo le fasi di introduzione nella cellula umana di una o più acido desossiribonucleico (DNA) endonucleasi per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB), all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL1 1A, che porta a una delezione, modulazione o inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A. La sequenza di controllo di trascrizione può essere situata all'interno di un secondo introne del gene BCL1 1A. La sequenza di controllo di trascrizione può essere situata all'interno di un sito ipersensibile a DNA +58 (DHS) del gene BCL11A.

È anche descritto nel presente documento un metodo *ex vivo* per il trattamento di un paziente (ad esempio un paziente umano) affetto da una emoglobinopatia, il metodo comprendendo le fasi di: creazione di una cellula staminale pluripotente indotta (iPSC) specifica per il paziente; editing all'interno di o vicino a un gene BCL11A o altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A della iPSC; differenziazione della iPSC con genoma editato in una cellula progenitrice ematopoietica; e impiantazione della cellula progenitrice ematopoietica nel paziente.

La fase di creazione di una cellula staminale pluripotente indotta (iPSC) specifica per paziente può comprendere: l'isolamento di una cellula staminale dal paziente; e l'introduzione di una serie di geni associati a pluripotenza nella cellula somatica per indurre la cellula somatica a diventare una cellula staminale pluripotente. La cellula somatica può essere un fibroblasto. La serie di geni

U001.D1.SM.61E

associati a pluripotenza può essere uno o più dei geni scelti nel gruppo costituito da OCT4, SOX2, KLF4, Lin28, NANOG e cMYC.

La fase di editing all'interno di o vicino a un gene BCL11A o altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A della iPSC può comprendere l'introduzione nella iPSC di una o più acido desossiribonucleico (DNA) endonucleasi per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB) all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che porta a una delezione, modulazione o inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A.

La fase di differenziazione della iPSC con genoma editato in una cellula progenitrice ematopoietica può comprendere uno o più dei seguenti: trattamento con una combinazione di piccole molecole, veicolazione di fattori di trascrizione (per esempio fattori di trascrizione master), o veicolazione di mRNA codificante fattori di trascrizione (per esempio fattori di trascrizione master).

La fase di impiantazione della cellula progenitrice ematopoietica ne paziente può comprendere l'impiantazione della cellula progenitrice ematopoietica nel paziente mediante trapianto, iniezione locale, infusione sistemica, o loro combinazioni.

È anche qui descritto un metodo *ex vivo* per trattare un paziente (per esempio un paziente umano) affetto da una emoglobinopatia, il metodo comprendendo le fasi di: isolamento di una cellula staminale mesenchimale dal paziente; editing all'interno di o vicino a un gene BCL11A o di altra sequenza di DNA che

U001.D1.SM.61E

codifica un elemento regolatorio del gene BCL11A della cellula staminale mesenchimale; differenziazione della cellula staminale mesenchimale editata nel genoma in una cellula progenitrice ematopoietica; e impiantazione della cellula progenitrice ematopoietica nel paziente.

La cellula staminale mesenchimale può venire isolata dal midollo osseo o dal sangue periferico del paziente. La fase di isolamento di una cellula staminale mesenchimale dal paziente può comprendere l'aspirazione di midollo osseo e l'isolamento di cellule mesenchimali utilizzando mezzi di centrifugazione in gradiente di densità.

La fase di editing all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A della cellula staminale mesenchimale può comprendere l'introduzione nella cellula staminale mesenchimale di una o più acido desossiribonucleico (DNS) endonucleasi per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB) all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che porta a una delezione, modulazione o inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A.

La fase di differenziazione della cellula staminale mesenchimale editata nel genoma di una cellula progenitrice ematopoietica può comprendere uno o più dei seguenti: trattamento con una combinazione di piccole molecole, veicolazione di fattori di trascrizione (ad esempio fattori di trascrizione master) o veicolazione di mRNA codificante fattori di trascrizione (per esempio fattori

U001.D1.SM.61E

di trascrizione master).

La fase di impiantazione della cellula progenitrice ematopoietica nel paziente può comprendere l'impiantazione della cellula progenitrice ematopoietica nel paziente mediante trapianto, iniezione locale, infusione sistemica, o loro combinazioni.

È anche qui descritto un metodo *ex vivo* per trattare un paziente (ad esempio un paziente umano) affetto da una emoglobinopatia, il metodo comprendendo le fasi di: isolamento di una cellula progenitrice ematopoietica dal paziente; editing all'interno di o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A della cellula progenitrice ematopoietica; e impiantazione della cellula progenitrice ematopoietica editata nel genoma nel paziente.

Il metodo può inoltre comprendere il trattamento del paziente con un fattore stimolante le colonie granulocitarie (GCSF) prima della fase di isolamento di una cellula progenitrice ematopoietica dal paziente. La fase di trattamento del paziente con un fattore stimolante colonie granulocitarie (GCSF) può venire effettuata in combinazione con Plerixaflor.

La fase di isolamento di una cellula progenitrice ematopoietica dal paziente può comprendere l'isolamento di cellule CD34+.

La fase di editing all'interno di o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A della cellula progenitrice ematopoietica può comprendere l'introduzione nella cellula progenitrice ematopoietica di una o più acido desossiribonucleico (DNS)

U001.D1.SM.61E

endonucleasi per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB) all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che porta a una delezione, modulazione o inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A.

La fase di impiantazione della cellula progenitrice ematopoietica editata nel genoma nel paziente può comprendere l'impiantazione della cellula progenitrice ematopoietica editata nel genoma nel paziente mediante trapianto, iniezione locale, infusione sistemica, o loro combinazioni.

È inoltre qui descritto un metodo *in vivo* per trattare un paziente (ad esempio un paziente umano) affetto da una emoglobinopatia, il metodo comprendendo la fase di editing di un gene BCL11A in una cellula del paziente.

La fase di editing di un gene BCL11A in una cellula del paziente può comprendere l'introduzione nella cellula di una o più acido desossiribonucleico (DNS) endonucleasi per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB) all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che porta a una delezione, modulazione o inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A. La cellula può essere una cellula di midollo osseo, una cellula progenitrice ematopoietica, una cellula CD34+, o loro combinazioni.

Le una o più DNA endonucleasi possono essere endonucleasi Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (note anche come Csn1 e

U001.D1.SM.61E

Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, o Cpf1; un omologo di esse, una ricombinazione della loro molecola presente naturalmente, loro codone-ottimizzato, o loro versioni modificate, e loro combinazioni.

Il metodo può comprendere l'introduzione nella cellula di uno o più polinucleotidi codificanti le una o più DNA endonucleasi. Il metodo può comprendere l'introduzione nella cellula di uno o più acidi ribonucleici (RNA) codificanti le una o più DNA endonucleasi. Gli uno o più polinucleotidi o uno o più RNA possono essere uno o più polinucleotidi modificati o uno o più RNA modificati. Le una o più DNA endonucleasi possono essere una o più proteine o polipeptidi. Le una o più proteine o polipeptidi possono essere affiancate(i) all'estremità N, all'estremità C, o sia all'estremità N che all'estremità C da uno o più segnali di localizzazione nucleare (NLS). Le una o più proteine o gli uno o più polipeptidi possono essere affiancati da due NLS, un NLS situato all'estremità N e il secondo NLS situato all'estremità C. Gli uno o più NLS possono essere un NLS SV40.

Il metodo può inoltre comprendere l'introduzione nella cellula di uno o più acidi ribonucleici guida (gRNA). Gli uno o più gRNA possono essere RNA guida monomolecolari (sgRNA). Gli uno o più gRNA o uno o più sgRNA possono essere uno o più gRNA modificati, uno o più sgRNA modificati, o loro combinazioni. Gli uno o più sgRNA modificati possono comprendere tre residui

U001.D1.SM.61E

di 2'-O-metil-fosforotioato in corrispondenza di o vicino a ciascuna delle loro estremità 5' e 3'. Il sgRNA modificato può essere la sequenza di acido nucleico di SEQ ID NO: 71.959. Le una o più DNA endonucleasi possono essere precomplesate con uno o più gRNA, uno o più sgRNA, o loro combinazioni per formare una o più ribonucleoproteine (RNP). Il rapporto in peso tra sgRNA e DNA endonucleasi nella RNP può essere di 1:1. Il sgRNA può comprendere la sequenza di acido nucleico di SEQ ID NO: 71.959, la DNA endonucleasi può essere una Cas9 di *S. pyogenes* comprendente un NLS SV40 all'estremità N e un NLS SV40 all'estremità C, e il rapporto in peso tra sgRNA e DNA endonucleasi può essere 1:1. Il documento US2015044772, che si riferisce a una proteina Cas di fusione chimerica, descrive la sequenza di acido nucleico di SEQ ID NO: 71.959. Detta sequenza è anche descritta nel documento WO2015183025, che si riferisce a un metodo per analizzare un genotipo utilizzando un sistema CRISPR/Cas.

Il metodo può inoltre comprendere l'introduzione nella cellula di un template donatore polinucleotidico comprendente un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza modificata di controllo di trascrizione.

Il metodo può inoltre comprendere l'introduzione nella cellula di un acido ribonucleico guida (gRNA) e un template donatore polinucleotidico comprendente un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata. Le una o più DNA endonucleasi possono essere una o più endonucleasi Cas9 o Cpf1 che effettuano una rottura a singolo filamento (SSB) o una rottura a doppio filamento (DSB) in

U001.D1.SM.61E

corrispondenza di un locus all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che facilita l'inserimento di una nuova sequenza dal template donatore polinucleotidico nel DNA cromosomico in corrispondenza del locus che porta a una inserzione, modulazione o inattivazione permanente della sequenza di controllo di trascrizione del DNA cromosomico prossimale al locus. Il gRNA può comprendere una sequenza spaziatrice che è complementare a un segmento del locus. Prossimale può significare nucleotidi sia a monte che a valle del locus.

Il metodo può inoltre comprendere l'introduzione nella cellula di uno o più acidi ribonucleici guida (gRNA) e un template donatore polinucleotidico comprendente un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata. Le una o più DNA endonucleasi possono essere una o più endonucleasi Cas9 o Cpf1 che effettuano o creano una coppia di rotture a singolo filamento (SSB) e/o rotture a doppio filamento (DSB), la prima rottura a un locus 5' e la seconda rottura a un locus 3', all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A, che facilita l'inserimento di una nuova sequenza dal template donatore polinucleotidico nel DNA cromosomico tra il locus 5' e il locus 3' che porta a una inserzione, modulazione o inattivazione permanente della sequenza di controllo di trascrizione del DNA cromosomico tra il locus 5' e il locus 3'. Un RNA guida può creare una coppia di SSB o DSB. Il suddetto RNA guida può comprendere una sequenza spaziatrice che è complementare al locus 5' o al locus 3'. In alternativa, il

U001.D1.SM.61E

metodo può comprendere un primo RNA guida e un secondo RNA guida. Il primo RNA guida può comprendere una sequenza spaziatrice che è complementare a un segmento del locus 5' e il secondo RNA guida può comprendere una sequenza spaziatrice che è complementare a un segmento del locus 3'. Il template donatore può essere a singolo o a doppio filamento. La sequenza di controllo di trascrizione modificata può essere situata all'interno di un secondo introne del gene BCL11A. La sequenza di controllo di trascrizione modificata può essere situata all'interno di un sito ipersensibile a DNA +58 (DHS) del gene BCL11A.

Gli uno o due gRNA possono essere uno o due RNA guida monomolecolari (sgRNA). Gli uno o due gRNA o uno o due sgRNA possono essere uno o due gRNA modificati o uno o due sgRNA modificati. Il suddetto sgRNA modificato può comprendere tre residui di 2'-O-metil-fosforotioato a o vicino a ciascuna delle sue estremità 5' e 3'. Il suddetto sgRNA modificato può essere la sequenza di acido nucleico di SEQ ID NO: 71.959. Le una o più endonucleasi Cas9 possono essere pre-complesstate con uno o due gRNA o uno o due sgRNA per formare una o più ribonucleoproteine (RNP). Le una o più endonucleasi Cas9 possono essere affiancate all'estremità N, all'estremità C, o sia all'estremità N che all'estremità C da uno o più segnali di localizzazione nucleare (NLSb). Le una o più endonucleasi Cas9 possono essere affiancate da due NLS, un NLS situato all'estremità N e il secondo NLS situato all'estremità C. Gli uno o più NLS possono essere un NLS SV40. Il rapporto in peso tra sgRNA e Cas9 endonucleasi nella RNP può essere 1:1. Il suddetto sgRNA può comprendere la

U001.D1.SM.61E

sequenza di acido nucleico di SEQ ID NO: 71.959, la endonucleasi Cas9 può essere una Cas9 di *S. pyogenes* comprendente un NLS SV40 all'estremità N e un NLS SV40 all'estremità C, e il rapporto in peso tra sgRNA e endonucleasi Cas9 può essere 1:1.

L'inserzione può essere mediante riparazione diretta dall'omologia (HDR).

SSB, DSB, locus 5' e/o locus 3' possono essere situati all'interno di un secondo introne del gene BCL11A. SSB,

DSB, locus 5' e/o locus 3' possono essere situati all'interno di un sito ipersensibile a DNA +58 (DHS) del gene BCL11A.

Il metodo può inoltre comprendere l'introduzione nella cellula di uno o più acidi ribonucleici guida (gRNA). Le una o più DNA endonucleasi possono essere una o più endonucleasi Cas9 o Cpf1 che effettuano o creano una coppia di rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB), la prima SSB o DSB a un locus 5' e una seconda SSB o DSB a un locus 3', all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che causa una delezione del DNA cromosomico tra il locus 5' e il locus 3' che ha come risultato una delezione, modulazione o inattivazione permanente della sequenza di controllo di trascrizione del DNA cromosomico tra il locus 5' e il locus 3'. Il primo RNA guida può comprendere una sequenza spaziatrice che è complementare a un segmento del locus 5' e il secondo RNA guida comprende una sequenza spaziatrice che è complementare a un segmento del locus 3'. Un RNA guida può creare una coppia di SSB o DSB. Il suddetto RNA guida può comprendere una

U001.D1.SM.61E

sequenza spaziatrice che è complementare al locus 5' o al locus 3'. In alternativa, il metodo può comprendere un primo RNA guida e un secondo RNA guida. Il primo RNA guida può comprendere una sequenza spaziatrice che è complementare a un segmento del locus 5' e il secondo RNA guida può comprendere una sequenza spaziatrice che è complementare a un segmento del locus 3'.

Gli uno o più gRNA possono essere uno o più RNA guida monomolecolari (sgRNA). Gli uno o più gRNA o uno o più sgRNA possono essere uno o più gRNA modificati o uno o più sgRNA modificati. Il suddetto sgRNA modificato può comprendere tre residui di 2'-O-metil-fosforotioato a o vicino a ciascuna delle sue estremità 5' e 3'. Il suddetto sgRNA modificato può essere la sequenza di acido nucleico di SEQ ID NO: 71.959. Le una o più endonucleasi Cas9 possono essere pre-complesstate con uno o più gRNA o uno o più sgRNA per formare una o più ribonucleoproteine (RNP). Le una o più endonucleasi Cas9 possono essere affiancate all'estremità N, all'estremità C, o sia all'estremità N che all'estremità C da uno o più segnali di localizzazione nucleare (NLS). Le una o più endonucleasi Cas9 possono essere affiancate da due NLS, un NLS situato all'estremità N e il secondo NLS situato all'estremità C. Gli uno o più NLS possono essere un NLS SV40. Il rapporto in peso tra sgRNA e endonucleasi Cas9 nella RNP può essere di 1:1. Il suddetto sgRNA può comprendere la sequenza di acido nucleico di SEQ ID NO: 71.959, l'endonucleasi Cas9 può essere una Cas9 di *S. pyogenes* comprendente un NLS SV40 all'estremità N e un NLS SV40 all'estremità C, e il rapporto in peso tra

U001.D1.SM.61E

sgRNA e endonucleasi Cas9 può essere di 1:1.

Il locus 5' e/o il locus 3' possono essere situati all'interno di un secondo introne del gene BCL11A. Il locus 5' e/o il locus 3' possono essere situati all'interno di un sito ipersensibile a DNA +58 (DHS) del gene BCL11A.

Cas9 o Cpf1 mRNA, gRNA, e template donatore possono essere formulati in nanoparticelle lipidiche separate o co-formulati in una nanoparticella lipidica.

Il Cas9 o Cpf1 mRNA può essere formulato in una nanoparticella lipidica, e il gRNA e il template donatore possono essere veicolati alla cellula da un vettore di virus adeno-associato (AAV).

Il Cas9 o Cpf1 mRNA può essere formulato in una nanoparticella lipidica, e il gRNA può essere veicolato alle cellule mediante elettroporazione e template donatore può essere veicolato alla cellula mediante un vettore di virus adeno-associato (AAV).

Le una o più RNP possono essere veicolate alla cellula mediante elettroporazione.

L'editing all'interno o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A può ridurre l'espressione del gene BCL11A.

Il gene BCL11A gene può essere situato sul Cromosoma 2: 60.451.167 – 60.553.567 (Genome Reference Consortium - GRCh38).

Sono inoltre qui previsti uno o più acidi ribonucleici guida (gRNA) per l'editing di un gene BCL11A in una cellula da un paziente affetto da una

U001.D1.SM.61E

emoglobinopatia. Gli uno o più gRNA possono comprendere una sequenza spaziatrice scelta nel gruppo costituito da sequenze di acidi nucleici in SEQ ID NO: 1 – 71.947 della Lista di sequenze. Gli uno o più gRNA possono essere uno o più RNA guida monomolecolari (sgRNA). Gli uno o più gRNA o uno o più sgRNA possono essere uno o più gRNA modificati o uno o più sgRNA modificati. Gli uno o più sgRNA modificati possono comprendere tre residui di 2'-O-metil-fosforotioato a o vicino a ciascuna delle loro estremità 5' e 3'. Gli uno o più sgRNA modificati possono comprendere la sequenza di acido nucleico di SEQ ID NO: 71.959. È inoltre qui previsto un RNA guida monomolecolare (sgRNA) comprendente la sequenza di acido nucleico di SEQ ID NO: 71.959.

Resta sottinteso che le invenzioni descritte nella presente specificazione non sono limitate agli esempi riepilogati in questo Sommario. Vari altri aspetti sono descritti ed esemplificati nel presente documento.

Breve descrizione delle figure

Vari aspetti di materiali e metodi per il trattamento di emoglobinopatie divulgati e descritti nella presente specificazione possono essere meglio compresi facendo riferimento alle figure annesse, nelle quali:

le Figure 1A-C mostrano plasmidi comprendenti un gene codone-ottimizzato per endonucleasi Cas9 di *S. pyogenes*.

La Figura 1A è un plasmide (CTx-1) comprendente un gene codone-ottimizzato per endonucleasi Cas9 di *S. pyogenes* Cas9. Il plasmide CTx-1 comprende anche una sequenza scaffold di gRNA, che include una sequenza spaziatrice di 20 bp dalle sequenze elencate in SEQ ID NO: 1 – 29.482 della

U001.D1.SM.61E

Lista di sequenze.

La Figura 1B è un plasmide (CTx-2) comprendente un differente gene codone-ottimizzato per endonucleasi Cas9 di *S. pyogenes*. Il plasmide CTx-2 comprende anche una sequenza scaffold di gRNA, che include una sequenza spaziatrice di 20 bp dalle sequenze elencate in SEQ ID NO: 1 – 29.482 della Lista di sequenze.

La Figura 1C è un plasmide (CTx-3) comprendente un altro differente gene codone-ottimizzato per endonucleasi Cas9 di *S. pyogenes*. Il plasmide CTx-3 comprende anche una sequenza scaffold di gRNA, che include una sequenza spaziatrice di 20 bp dalle sequenze elencate in SEQ ID NO: 1 – 29.482 della Lista di sequenze.

Le Figure 2A-B illustrano il sistema CRISPR/Cas di tipo II.

La Figura 2A illustra il sistema CRISPR/Cas di tipo II comprendente gRNA.

La Figura 2B illustra il sistema CRISPR/Cas di tipo II comprendente sgRNA.

La Figura 3 mostra il tasso di editing di DNA in cellule staminali e progenitrici ematopoietiche CD34+ (HSPC) e ciascuno dei differenti genotipi di HPFH risultanti.

Le Figure 4A-C mostrano la sovraregolazione di espressione di γ -globina in eritrociti differenziati da HSPC CD34+ umane editate *Bulk* da sangue periferico mobilizzato (mPB).

La Figura 4A illustra ematopoiesi da HSPC CD34+ umane a eritrociti.

U001.D1.SM.61E

La Figura 4B mostra il rapporto di $\gamma/18\text{sRNA}$ per ciascuna della delezione/modificazione.

La Figura 4 mostra il rapporto di γ/α per ciascuna della delezione/modificazione.

Le Figure 5A-B mostrano la sovraregolazione di espressione di γ -globina in eritrociti differenziati da tutte le colonie sottoposte a editing genico da HSPC CD34+.

La Figura 5A mostra il rapporto di mRNA di γ/α globina (%) per ciascuna delle colonie sottoposte a editing genico.

La Figura 5B mostra il rapporto medio di mRNA di γ/α globina (%) per ciascuna delle modificazioni geniche.

La Figura 6 mostra il tasso di Introne di BCL11A (SPY101) di editing di DNA in colonie eritroidi derivate da HSPC CD34+ umane.

Le Figure 7A-B mostrano la correlazione tra genotipo SPY101 e espressione di γ -globina in colonie monocellulari differenziate da HSPC CD34+ di mPB umano sottoposte a editing genico.

La Figura 7A mostra la percentuale di γ -globina a α -globina (HBG/HBA) per le colonie sottoposte a editing genico.

La Figura 7B mostra la percentuale di globine β -simili (HBG/(HBB+HBG)) per ciascuna delle colonie sottoposte a editing genico.

La Figura 8 mostra l'efficacia di editing on-target di numerosi gRNA in cellule CD34+ umane di mPB.

Le Figure 9A-B mostrano il saggio di cattura di ibridi utilizzato per

U001.D1.SM.61E

rilevare editing off-target e risultati generati utilizzando il saggio di cattura di ibridi da HSPC CD34+ di mPB umano.

La Figura 9A mostra una rappresentazione schematica di un saggio di cattura di ibridi utilizzato per rilevare l'attività di editing a potenziali siti off-target.

La Figura 9B mostra l'attività off-target osservata tramite sequenziamento di cattura di ibridi.

Le Figure 10A-B mostrano rapporti di livelli di mRNA di globina misurati in cellule da pazienti con SCD, un paziente affetto da β -talassemia, e donatori sani.

La Figura 10A mostra rapporti di livelli di mRNA di globina misurati in cellule da pazienti con SCD rispetto a donatori sani.

La Figura 10B mostra rapporti di livelli di mRNA di globina misurati in cellule da un paziente affetto da β -talassemia rispetto a donatori sani.

Le Figure 11A-C mostrano la strategia di citometria a flusso utilizzata per rilevare varie popolazioni cellulari sottoposte a editing genico e risultati generati utilizzando la strategia di citometria a flusso.

La Figura 11A mostra sottopopolazioni di HSPC CD34+ umane di mPB, marcatori di superficie associati, e strategia di gating di citometria a flusso.

La Figura 11B mostra una distribuzione simile di tipi di cellule nelle condizioni mock e editate.

La Figura 11C mostra elevate efficienze di editing simili attraverso le sottopopolazioni rispetto a bulk.

U001.D1.SM.61E

La Figura 12 mostra analisi di popolazioni di cellule CD45RA+ umane in topi NSG 8 settimane dopo l'attecchimento o engraftment di HSPC CD34+ di mPB umano. Punti dati rappresentano animali individuali e illustrano la percentuale di cellule vive che erano cellule vive CD45RA+ umane.

La Figura 13 mostra l'efficacia media di editing di un gRNA di SPY101 e proteina Cas9 in HSPC CD34+ di mPB umano a scale di laboratorio e clinicamente rilevanti.

La Figura 14 mostra una panoramica di design di studio di GLP/Tossicologia.

La Figura 15 mostra una panoramica di un approccio sperimentale per analisi di colonie monocellulari e bulk di mRNA di emoglobina e livelli proteici in popolazioni di cellule eritroidi derivate da HSPC CD34+ di mPB umano con editing genico di CRISPR/Cas9.

Le Figure 16A-B mostrano la sovraregolazione di mRNA di γ -globina e proteine in HSPC CD34+ di mPB umano modificate con differenti edit mirati.

La Figura 16A mostra la sovraregolazione di mRNA di γ -globina in HSPC CD34+ di mPB umano differenziate in bulk modificate con differenti edit mirati.

La Figura 16B mostra la sovraregolazione proteica di γ -globina in HSPC CD34+ di mPB umano differenziate in bulk modificate con differenti edit mirati.

La Figura 17 mostra la sovraregolazione media di γ -globina in colonie individuali di HSPC CD34+ di mPB umano differenziate modificate con

U001.D1.SM.61E

differenti edit mirati.

Le Figure 18A-B mostrano una correlazione da genotipo a fenotipo in colonie editate con Target 5 e Target 6 di HSPC CD34+ di mPB umano con differenziazione eritroide.

La Figura 18A include grafici sul lato sinistro che mostrano la percentuale di colonie con ciascun genotipo, e grafici sul lato destro che mostrano la percentuale di colonie con ciascun livello di sovraregolazione di γ -globina (espressa come rapporto di mRNA di globina $\gamma/(\gamma+\beta)$).

La Figura 18B mostra livelli di trascritti di mRNA, per gruppi di colonie con genotipi simili.

La Figura 19 mostra una panoramica di un approccio sperimentale per analisi in bulk di efficienza di editing da DNA genomico, espressione di emoglobina da mRNA, e proteina in popolazioni cellulari con differenziazione eritroide derivate da HSPC CD34+ di mPB umano editate con gene CRISPR/Cas9.

Le Figure 20A-B mostrano la percentuale di editing genico mantenuta per tutta la differenziazione eritroide *ex vivo* di HSPC CD34+ di mPB editate con SPY101 gRNA o SD2 gRNA.

La Figura 20A mostra la percentuale di editing genico mantenuta per tutta la differenziazione eritroide *ex vivo* di HSPC CD34+ di mPB editate con SPY101 gRNA.

La Figura 20B mostra la percentuale di editing genico mantenuta per tutta la differenziazione eritroide *ex vivo* di HSPC CD34+ di mPB editate con

U001.D1.SM.61E

SD2 gRNA.

Le Figure 21A-D mostrano l'aumento in trascritto di γ -globina illustrato come γ/α o $\gamma/(\gamma+\beta)$ in HSPC CD34+ di mPB sottoposte a editing genico i giorni 11 o 15 post-differenziazione eritroide.

La Figura 21A mostra l'aumento in trascritto di γ -globina (γ/α) in HSPC CD34+ di mPB sottoposte a editing genico il giorno 11 post-differenziazione.

La Figura 21B mostra l'aumento in trascritto di γ -globina (γ/α) in HSPC CD34+ di mPB sottoposte a editing genico il giorno 15 post-differenziazione.

La Figura 21C mostra l'aumento in trascritto di γ -globina ($\gamma/(\gamma+\beta)$) in HSPC CD34+ di mPB sottoposte a editing genico il giorno 11 post-differenziazione.

La Figura 21D mostra l'aumento in trascritto di γ -globina ($\gamma/(\gamma+\beta)$) in HSPC CD34+ di mPB sottoposte a editing genico il giorno 15 post-differenziazione.

La Figure 22A-B rappresentano l'analisi di FACS e l'analisi di Intensità di Fluorescenza Mediana (MFI) che mostrano la sovraregolazione di γ -globina in HSPC CD34+ di mPB sottoposte a editing genico il giorno 15 post-differenziazione eritroide.

La Figura 22A è l'analisi di FACS che mostra la sovraregolazione di γ -globina in HSPC CD34+ di mPB sottoposte a editing genico 15 giorni post-differenziazione eritroide.

La Figura 22B è l'analisi di MFI che mostra la sovraregolazione di γ -globina in HSPC CD34+ di mPB sottoposte a editing genico da 4 donatori post-

U001.D1.SM.61E

differenziazione eritroide.

La Figura 23A-D rappresenta dati di cromatografia liquida – spettrometria di massa (LC-MS) in bulk che mostrano la sovraregolazione di γ -globina, illustrata come γ/α o $\gamma/(\gamma+\beta)$ in HSPC CD34+ di mPB sottoposte a editing genico il giorno 15 post-differenziazione eritroide.

La Figura 23A rappresenta dati di cromatografia liquida – spettrometria di massa (LC-MS) in bulk che mostrano la sovraregolazione di γ -globina (γ/α) in HSPC CD34+ di mPB sottoposte a editing genico il giorno 15 post-differenziazione.

La Figura 23B rappresenta dati di cromatografia liquida – spettrometria di massa (LC-MS) in bulk che mostrano la sovraregolazione di γ -globina (γ/α) in HSPC CD34+ di mPB sottoposte a editing genico il giorno 15 post-differenziazione normalizzati a γ -globina (γ/α) in HSPC CD34+ di mPB transfettate con GFP gRNA.

La Figura 23C rappresenta dati di cromatografia liquida – spettrometria di massa (LC-MS) in bulk che mostrano la sovraregolazione di γ -globina ($\gamma/(\gamma+\beta)$) in HSPC CD34+ di mPB sottoposte a editing genico il giorno 15 post-differenziazione.

La Figura 23D rappresenta dati di cromatografia liquida – spettrometria di massa (LC-MS) in bulk che mostrano la sovraregolazione di γ -globina ($\gamma/(\gamma+\beta)$) in HSPC CD34+ di mPB sottoposte a editing genico il giorno 15 post-differenziazione normalizzati a γ -globina (γ/α) in HSPC CD34+ di mPB transfettate con GFP gRNA

U001.D1.SM.61E

La Figura 24 illustra la progettazione di esche o “bait design” di cattura di ibridi.

La Figura 25 mostra un grafico che illustra il potere del metodo di cattura di ibridi per rilevare indel.

La Figura 26 mostra un riepilogo dei dati generati da esperimenti di cattura di ibridi utilizzando SPY101 gRNA.

La Figura 27 mostra un riepilogo dei dati generati da esperimenti di cattura di ibridi utilizzando SD2 gRNA.

La Figura 28 mostra un piano di studio per gli esperimenti di attecchimento o engraftment.

Le Figure 29A-E mostrano dati di analisi del sangue ad interim di 8 settimane per topi non trattati, e topi ai quali sono state iniettate cellule editate con mock, cellule editate con GFP gRNA, cellule editate con SPY101 gRNA o cellule editate con SD2 gRNA.

La Figura 29A mostra dati di sangue ad interim di 8 settimane per topi non trattati (UnTx).

La Figura 29B mostra dati di analisi del sangue ad interim di otto settimane per topi ai quali sono state iniettate cellule editate con mock.

La Figura 29C mostra dati di analisi del sangue ad interim di 8 settimane per topi ai quali sono state iniettate cellule editate con GFP gRNA.

La Figura 29D mostra dati di analisi del sangue ad interim di 8 settimane per topi ai quali sono state iniettate cellule editate con SPY101 gRNA.

La Figura 29E mostra dati di analisi del sangue ad interim di 8 settimane

U001.D1.SM.61E

per topi ai quali sono state iniettate cellule editate con SD2 gRNA.

La Figura 30 mostra dati medi di analisi del sangue ad interim di 8 settimane.

La Figura 31 mostra la percentuale di Indel per HSPC CD34+ di mPB umano elettroporate con vari Cas9 mRNA e SPY101 gRNA (mRNA1-8) a confronto con HSPC CD34+ di mPB umano elettroporate con proteina Cas9 complessata con SPY101 gRNA (un complesso di ribonucleoproteina, RNP).

Le Figure 32A-B mostrano il conteggio cellulare normalizzato e la vitalità cellulare di HSPC CD34+ di mPB umano elettroporate con vari Cas9 mRNA e SPY101 gRNA (mRNA 1-8) a confronto con HSPC CD34+ di mPB umano elettroporate con proteina Cas9 complessata con SPY101 gRNA (RNP).

La Figura 32A mostra il numero di volte di incremento o fold increase nel conteggio cellulare a 48 ore post-elettroporazione per HSPC CD34+ di mPB umano elettroporate con vari Cas9 mRNA e SPY101 gRNA (mRNA 1-8) a paragone con HSPC CD34+ di mPB umano elettroporate con proteina Cas9 complessata con SPY101 gRNA (RNP).

La Figura 32B mostra la vitalità cellulare a 48 ore post-elettroporazione per HSPC CD34+ di mPB umano elettroporate con vari Cas9 mRNA e SPY101 gRNA (mRNA 1-8) a paragone con HSPC CD34+ di mPB umano elettroporate con proteina Cas9 complessata con SPY101 gRNA (RNP).

Le Figure 33A-C mostrano numerosi costrutti di Cas9 RNP utilizzati per ottimizzazione di Cas9 RNP e la percentuale di Indel associata con ciascuno dei costrutti di Cas9 RNP.

U001.D1.SM.61E

La Figura 33A mostra numerosi costrutti di Cas9 RNP.

La Figura 33B mostra la percentuale di Indel per ciascuno dei costrutti di Cas9 RNP utilizzando 1µg di Cas9: 1µg di SPY101 gRNA.

La Figura 33C mostra la percentuale di Indel dei costrutti di Cas9 RNP utilizzando 3 µg di Cas9: 3 µg di SPY101 gRNA.

Le Figure 34A-B mostrano l'efficienza (%) di editing genico (%) per HSPC CD34+ di mPB umano trattate con Cas9 mRNA o proteina Cas9 (Feldan o Aldevron) su scala non clinica e scala clinica.

La Figura 34A mostra l'efficienza (%) di editing genico per HSPC CD34 derivate da mPB umano o midollo osseo (BM) umano trattate con Cas9 mRNA o proteina Cas9 (Feldan o Aldevron) su scala non clinica.

La Figura 34B mostra l'efficienza (%) di editing genico per HSPC CD34+ di mPB umano trattate con proteina Cas9 (Aldevron) su scala clinica.

Le Figure 35A-B mostrano l'efficacia di SPY101 in HSPC CD34+ di mPB umano presentando il rapporto di mRNA di γ/α globina in % e il rapporto di mRNA di $\gamma/(\gamma+\beta)$ globina in % per cellule trattate con Cas9 mRNA e SPY101 gRNA o proteina Cas9 (Feldan o Aldevron) complessata con SPY101 gRNA.

La Figura 35A mostra il rapporto di mRNA di γ/α globina in % per HSPC CD34+ di mPB umano trattate con Cas9 mRNA e SPY101 gRNA o proteina Cas9 (Feldan o Aldevron) complessata con SPY101 gRNA.

La Figura 35B mostra il rapporto di mRNA di $\gamma/(\gamma+\beta)$ globina in % per HSPC CD34+ di mPB umano trattate con Cas9 mRNA e SPY101 gRNA o proteina Cas9 (Feldan o Aldevron) complessata con SPY101 gRNA.

U001.D1.SM.61E

Le Figure 36A-B mostrano l'efficacia di SPY101 in HSPC CD34+ derivate da midollo osseo presentando il rapporto di mRNA di γ/α globina in % e il rapporto di mRNA di $\gamma/(\gamma+\beta)$ globina in % per cellule trattate con proteina Cas9 (Aldevron, tecnicamente ottimizzata vs. non ottimizzata) complessata con SPY101 gRNA.

La Figura 36A mostra il rapporto di mRNA di γ/α globina in % per HSPC CD34+ derivate da midollo osseo trattate con proteina Cas9 complessata con SPY101 gRNA.

La Figura 36B mostra il rapporto di mRNA di $\gamma/(\gamma+\beta)$ globina in % per HSPC CD34+ derivate da midollo osseo trattate con proteina Cas9 complessata con SPY101 gRNA.

Le Figure 37A-B mostrano l'efficacia di SPY101 in campioni di pazienti affetti da SCD e da β -Talassemia.

La Figura 37A mostra la media del rapporto di mRNA di $\gamma/(\gamma+\beta)$ globina in % per cellule con differenziazione eritroide da sei pazienti con SCD e due donatori sani che sono stati trattati con SPY101 gRNA e proteina Cas9. Tutti i valori sono stati sottratti da loro rispettivi campioni di controllo trattati con GFP gRNA e proteina Cas9.

La Figura 37B mostra il rapporto di mRNA di γ/α globina in % per cellule con differenziazione eritroide da un paziente β -Talassemico e due donatori sani che sono stati trattati con SPY101 gRNA e proteina Cas9. Tutti i valori erano sottratti da loro rispettivi campioni di controllo trattati con GFP gRNA e proteina Cas9.

U001.D1.SM.61E

Le Figure 38A-B mostrano il tasso di Introne Bcl11a (SPY101) di editing di DNA quando si utilizza Cas9 mRNA o Cas9 RNP.

La Figura 38A mostra il tasso di Introne BCL11A (SPY101) di editing di DNA quando si utilizza Cas9 mRNA.

La Figura 38B mostra il tasso di Introne BCL11A (SPY101) di editing di DNA quando si utilizza Cas9 RNP.

Le Figure 39A-B mostrano che distruzioni del sito di legame a GATA1 (GBS) causate da SPY101/Cas9 RNP in colonie di singole cellule derivate da HSPC CD34+ di mPB umano con differenziazione eritroide sono collegate a una aumentata espressione di γ -globina.

La Figura 39A mostra il rapporto di mRNA di γ/α globina di colonie editate con SPY101 senza alcuna distruzione di GBS, distruzioni di GBS mono-alleleliche, o distruzioni di GBS bi-alleleliche.

La Figura 39B mostra il rapporto di mRNA di $\gamma/(\gamma+\beta)$ globina di colonie editate con SPY101 senza alcuna distruzione di GBS, distruzioni di GBS mono-alleleliche o distruzioni di GBS bi-alleleliche.

Le Figure 40A-E mostrano una aumentata espressione di γ -globina in HSPC CD34+ di mPB umano editate con SPY101/Cas9 RNP con differenziazione eritroide mediante analisi di citometria a flusso.

La Figura 40A è una analisi di citometria a flusso che mostra l'espressione di α -globina in HSPC CD34+ di mPB umano con differenziazione eritroide editate con SPY101/Cas9 RNP a confronto con espressione di α -globina in HSPC CD34+ di mPB umano con differenziazione eritroide trattate

U001.D1.SM.61E

con GFP gRNA/Cas9 RNP.

La Figura 40B è una analisi di citometria a flusso che mostra l'espressione di β -globina in HSPC CD34+ di mPB umano con differenziazione eritroide editate con SPY101/Cas9 RNP a confronto con espressione di β -globina in HSPC CD34+ di mPB umano con differenziazione eritroide trattate con GFP gRNA/Cas9 RNP.

La Figura 40C è una analisi di citometria a flusso che mostra l'espressione di γ -globina in HSPC CD34+ di mPB umano con differenziazione eritroide editate con SPY101/Cas9 RNP a confronto con espressione di γ -globina in HSPC CD34+ di mPB umano con differenziazione eritroide trattate con GFP gRNA/Cas9 RNP.

La Figura 40D mostra la percentuale di cellule positive a γ -globina in HSPC CD34+ di mPB umano con differenziazione eritroide editate con SPY101/Cas9 RNP a confronto con HSPC CD34+ di mPB umano con differenziazione eritroide trattate con GFP gRNA/Cas9 RNP.

La Figura 40E mostra l'intensità di fluorescenza mediana (MFI) in HSPC CD34+ di mPB umano con differenziazione eritroide editate con SPY101/Cas9 RNP a confronto con HSPC CD34+ di mPB umano con differenziazione eritroide trattate con GFP gRNA/Cas9 RNP.

Breve descrizione della lista di sequenze

SEQ ID NOs: 1 – 29.482 sono sequenze spaziatrici di 20 bp per bersagliamento o targeting all'interno o vicino a un gene BCL11A gene o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una

U001.D1.SM.61E

endonucleasi Cas9 di *S. pyogenes*.

SEQ ID NOs: 29.483 -32.387 sono sequenze spaziatrici di 20 bp per bersagliamento o targeting all'interno o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi Cas9 di *S. aureus*.

SEQ ID NOs: 32.388 – 33.420 sono sequenze spaziatrici di 20 bp per bersagliamento o targeting all'interno o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi Cas9 di *S. thermophilus*.

SEQ ID NOs: 33.421 – 33.851 sono sequenze spaziatrici di 20 bp per bersagliamento o targeting all'interno o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi Cas9 di *T. denticola*.

SEQ ID NOs: 33.852 – 36.731 sono sequenze spaziatrici di 20 bp per bersagliamento o targeting all'interno o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi Cas9 di *N. meningitides*.

SEQ ID NOs: 36.732 – 71.947 sono sequenze spaziatrici di 22 bp per bersagliamento o targeting all'interno o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi Cpf1 di *Acidominococcus*, di *Lachnospiraceae*, e di *Franciscella Novicida*.

SEQ ID NO: 71.948 è un RNA guida (gRNA) campione per una endonucleasi

U001.D1.SM.61E

Cas9 di *S. pyogenes*.

SEQ ID NO: 71.949 mostra una famiglia nota di endonucleasi di inserimento o homing, come classificata dalla sua struttura.

SEQ ID NO: 71.950 è gRNA A (CLO1).

SEQ ID NO: 71.951 è gRNA B (CLO8).

SEQ ID NO: 71.952 è gRNA C (CSO2).

SEQ ID NO: 71.953 è gRNA D (CSO6).

SEQ ID NO: 71.954 è gRNA E (HPFH-15).

SEQ ID NO: 71.955 è gRNA F (HPFH-4).

SEQ ID NO: 71.956 è gRNA G (Kenya02).

SEQ ID NO: 71.957 è gRNA H (Kenya17).

SEQ ID NO: 71.958 è gRNA I (SD2).

SEQ ID NO: 71.959 è gRNA J (SPY101).

SEQ ID NOs: 71.960-71.962 mostrano sequenze di sgRNA campione.

Descrizione dettagliata

Emoglobina fetale

Emoglobina fetale (HbF, $\alpha_2\gamma_2$) è la principale proteina di trasporto di ossigeno in un feto umano e include subunità alfa (α) e gamma (γ). L'espressione di HbF cessa circa 6 mesi dopo la nascita. Emoglobina adulta (HbA, $\alpha_2\beta_2$) è la principale proteina di trasporto di ossigeno in un essere umano dopo circa 34 settimane dalla nascita, e include subunità alfa (α) e beta (β). Dopo 34 settimane una interruzione evolutiva ha come risultato una diminuita trascrizione dei geni di γ -globina e una aumentata trascrizione di geni di β -globina. Dal momento che

U001.D1.SM.61E

molte delle forme di emoglobinopatie sono un risultato dell'incapacità di produrre proteina β -globina normale in quantità sufficienti o dell'incapacità di produrre interamente proteina β -globina normale, l'espressione migliorata di γ -globina (vale a dire HbF) migliorerà la severità della malattia di β -globina.

Linforma a cellule B 11A (BCL11A)

Linforma a cellule B 11A (BCL11A) è un gene situato sul Cromosoma 2 e varia da 60.451.167 a 60.553.567 bp (GRCh38). BCL11A è un fattore di trascrizione "a dito di zinco" o "zinc finger" che reprime emoglobina fetale (HbF) e sottoregola l'espressione di HbF a partire da circa 6 settimane dopo la nascita. Il gene BCL11A contiene 4 esoni, che occupano 102,4 kb di DNA genomico. BCL11A è anche sotto regolazione di trascrizione, comprendendo un dominio di legame in introne 2 per il fattore di trascrizione master GATA-1. Il legame a GATA-1 potenzia l'espressione di BCL11A che, a sua volta, reprime l'espressione di HbF. Introne 2 contiene molteplici siti ipersensibili a DNasi (DHS), inclusi siti indicati come +55, +58, e +62 sulla base della distanza in kilobasi dal sito di inizio di trascrizione. Varie strategie di editing sono discusse sotto per cancellare, modulare o inattivare le sequenze di controllo di trascrizione di BCL11A. SNP presenti naturalmente all'interno di questa regione sono stati associati con una diminuita espressione di BCL11A e aumentati livelli di Hb fetale (Orkin et al. 2013 GWAS study). Questi SNP sono organizzati attorno a 3 siti di ipersensibilità a DNA, +55DHS, +58DHS e +62DHS. Delle tre regioni, la regione +58 DHS sembra essere una regione chiave associata a livelli aumentati di Hb fetale e ospita anche una regione di controllo di trascrizione di

U001.D1.SM.61E

GATA1.

Approccio terapeutico

Giunzione non omologa delle estremità (NHEJ) può essere utilizzata per cancellare segmenti della sequenza di controllo di trascrizione di BCL11A, direttamente o alterando siti donatori o accettori di splicing attraverso il clivaggio da parte di un gRNA che ha come bersaglio più posizioni, o numerosi gRNA.

La sequenza di controllo di trascrizione del gene BCL11A può anche venire modulata o inattivata inserendo un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata. Per esempio, il donatore per la modulazione o inattivazione mediante riparazione diretta da omologia (HDR) contiene la sequenza di controllo di trascrizione modificata del gene BCL11A con piccoli o grandi bracci di omologia fiancheggianti per consentire l'annealing. HDR è essenzialmente un meccanismo privo di errore che utilizza una sequenza di DNA omologa fornita come uno stampo o template durante la riparazione di DSB. Il tasso di riparazione diretta da omologia (HDR) è in funzione della distanza tra la sequenza di controllo di trascrizione e il sito tagliato, quindi la scelta di siti bersaglio sovrapposti o adiacenti è importante. Stampi o template possono includere sequenze extra affiancate dalle regioni omologhe o possono contenere una sequenza che differisce dalla sequenza genomica, consentendo quindi l'editing di sequenza.

In aggiunta alla delezione, modulazione o inattivazione della sequenza di controllo di trascrizione del gene BCL11A mediante NHEJ o HDR, è possibile

U001.D1.SM.61E

una gamma di altre opzioni. Se vi sono delezioni piccole o grandi, può essere introdotto con knock in un cDNA che contiene una sequenza di controllo di trascrizione modificata del gene BCL11A. Un cDNA a piena lunghezza può essere introdotto in qualsiasi “safe harbour” o “porto sicuro” -- vale a dire un punto di inserimento non deleterio che non è il gene BCL11A stesso -- con o senza sequenze regolatorie adatte. Se questo costrutto è inserito con knock in vicino agli elementi regolatori di BCL11A, esso dovrebbe avere un controllo fisiologico simile al gene normale. Due o più (ad esempio un paio di) nucleasi possono essere utilizzate per cancellare regioni di sequenze di controllo, sebbene dovrebbe solitamente venire fornito un donatore per modulare o inattivare la funzione. In questo caso sarebbero forniti due gRNA e una sequenza donatrice.

Sono qui forniti metodi cellulari *ex vivo* e *in vivo* per utilizzare strumenti di ingegneria genomica per creare cambiamenti permanenti nel genoma: 1) modulando o inattivando la sequenza di controllo di trascrizione del gene BCL11A, mediante delezioni che insorgono a causa della via di NHEJ; 2) modulando o inattivando la sequenza di controllo di trascrizione del gene BCL11A gene, mediante HDR; 3) modulando o inattivando la sequenza di controllo di trascrizione del gene BCL11A, mediante delezioni di almeno una porzione della sequenza di controllo di trascrizione e/o introducendo con knock in un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata nel locus genico o un locus “porto sicuro” o “safe harbour”. Tali metodi utilizzano endonucleasi, come nucleasi associate a CRISPR (Cas9, Cpf1 e simili) per cancellare, inserire o editare

U001.D1.SM.61E

permanentemente la sequenza di controllo di trascrizione all'interno o vicino al locus genomico del gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A. In questo modo, esempi riportati nella presente descrizione possono aiutare a cancellare, modulare o inattivare la sequenza di controllo di trascrizione del gene BCL11A con un singolo trattamento o un numero limitato di trattamenti (anziché somministrare potenziali terapie per la durata di vita del paziente).

Sono qui descritti metodi per trattare un paziente affetto da una emoglobinopatia. Un aspetto di tale metodo è una terapia cellulare *ex vivo*. Per esempio, può venire create una cellula staminale pluripotente indotta (iPSC) specifica per paziente. Poi, il DNA cromosomico di queste cellule iPS può venire editato utilizzando i materiali e metodi descritti nel presente documento. Successivamente, iPSC editate nel genoma possono venire differenziate in cellule progenitrici ematopoietiche. Infine, le cellule progenitrici ematopoietiche possono venire impiantate nel paziente.

Un altro aspetto di tale metodo è una terapia cellulare *ex vivo*. Per esempio, può venire isolata dal paziente una cellula staminale mesenchimale, che può venire isolata dal midollo osseo o dal sangue periferico del paziente. Successivamente, il DNA cromosomico di queste cellule staminali mesenchimali può venire editato utilizzando i materiali e metodi descritti nel presente documento. Successivamente, le cellule staminali mesenchimali sottoposte a editing genomico possono venire differenziate in cellule progenitrici ematopoietiche. Infine, queste cellule progenitrici ematopoietiche

U001.D1.SM.61E

possono venire impiantate nel paziente.

Un ulteriore aspetto di tale metodo è una terapia cellulare *ex vivo*. Per esempio, una cellula progenitrice ematopoietica può venire isolata dal paziente. Successivamente, il DNA cromosomico di queste cellule può venire editato utilizzando i materiali e metodi descritti nel presente documento. Infine, le cellule progenitrici ematopoietiche sottoposte a editing genomico possono venire impiantate nel paziente.

Un vantaggio di un approccio di terapia cellulare *ex vivo* è la capacità di condurre una analisi comprensiva del terapeutico prima della somministrazione. Terapeutici a base di nucleasi possono avere lo stesso livello di effetti off-target. Effettuare una correzione genica *ex vivo* permette a un individuo di caratterizzare la popolazione cellulare corretta prima dell'impiantazione. La presente descrizione include il sequenziamento dell'intero genoma delle cellule corrette per assicurare che gli effetti off-target, se presenti, possano essere in posizioni genomiche associate con un rischio minimo per il paziente. Inoltre, popolazioni di cellule specifiche, incluse popolazioni clonali, possono venire isolate prima dell'impiantazione.

Un altro vantaggio di una terapia cellulare *ex vivo* si riferisce alla correzione genetica in iPSC a paragone con altre fonti di cellule primarie. iPSC sono prolifiche, rendendo facile ottenere il grande numero di cellule che saranno richieste per una terapia cellulare. Inoltre, iPSC sono un tipo di cellula ideale per effettuare isolamenti clonali. Questo permette di vagliare la corretta correzione genomica, senza rischiare una diminuzione nella vitalità. Al contrario, altre

U001.D1.SM.61E

cellule primarie sono vitali per soltanto pochi passaggi e difficili da espandere clonalmente. Quindi, la manipolazione di iPSC per il trattamento di una emoglobinopatia può essere molto più facile, e può accorciare la quantità di tempo necessaria per produrre la correzione genetica desiderata.

Per terapia *ex vivo*, un trapianto richiede la depurazione di nicchie di midollo osseo o di HSC di donatore da innestare. Metodi attuali si basano su radioterapia e/o chemioterapia. A causa delle limitazioni che questi impongono, regimi di condizionamento più sicuri sono stati sviluppati e sono ora in sviluppo, come immunodeplezione di cellule di midollo osseo mediante anticorpi o coniugati anticorpi-tossine diretti contro marcatori di superficie di cellule ematopoietiche, per esempio CD117, c-kit e altri. Il successo del trapianto di HSC dipende dall'efficiente ritorno o homing nel midollo osseo, dal successivo attecchimento o engraftment, e dalla ripopolazione del midollo osseo. Il livello di cellule sottoposte a editing genico attecchite è importante, così come la capacità dell'attecchimento o engraftment multilignaggio delle cellule.

Cellule staminali ematopoietiche (HSC) sono un importante bersaglio per terapia genica *ex vivo* poiché forniscono una fonte prolungata delle cellule corrette. Cellule CD34+ trattate sarebbero reintrodotte nel paziente.

Metodi possono anche includere una terapia *in vivo*. DNA cromosomico delle cellule nel paziente è editato utilizzando i materiali e i metodi descritti nel presente documento. Le cellule possono essere cellule di midollo osseo, cellule progenitrici ematopoietiche, o cellule CD34+.

Sebbene cellule ematiche presentino un bersaglio attraente per trattamento e

U001.D1.SM.61E

terapia *ex vivo*, una aumentata efficacia nella veicolazione può permettere una veicolazione *in vivo* diretta alle cellule staminali ematopoietiche (HSC) e/o altri progenitori di cellule B e T, come cellule CD34+. Idealmente, il bersagliamento o targeting e l'editing dovrebbero essere diretti alle cellule pertinenti. Il clivaggio in altre cellule può anche essere impedito mediante l'uso di promotori attivi soltanto in alcune cellule e/o stadi di sviluppo. Promotori supplementari sono inducibili, e pertanto possono venire controllati temporalmente se la nucleasi è veicolata come un plasmide. La quantità di tempo che RNA e proteina veicolati rimangono nella cellula può anche venire adattata utilizzando trattamenti o domini aggiunti per cambiare l'emivita. Un trattamento *in vivo* eliminerebbe un numero di fasi di trattamento, ma un tasso inferiore di veicolazione può richiedere tassi superiori di editing. Un trattamento *in vivo* può eliminare problemi e perdite da trattamento *ex vivo* e attecchimento o engraftment.

Un vantaggio di una terapia genica *in vivo* può essere la facilità di produzione e somministrazione terapeutica. Lo stesso approccio terapeutico e la stessa terapia potranno potenzialmente essere utilizzati per trattare più di un paziente, per esempio un numero di pazienti che condividono lo stesso genotipo o allele o un genotipo o allele simile. Al contrario, la terapia cellulare *ex vivo* tipicamente richiede l'utilizzo di cellule proprie di un paziente, che sono isolate, manipolate e reintrodotte nello stesso paziente.

È inoltre qui descritto un metodo cellulare per editare il gene BCL11A in una cellula mediante editing genomico. Per esempio, una cellula può venire isolata

U001.D1.SM.61E

da un paziente o da un animale. Poi, il DNA cromosomico della cellula può venire editato utilizzando i materiali e metodi descritti nel presente documento.

I metodi qui forniti, indipendentemente dal fatto che si tratti di un metodo cellulare o *ex vivo* o *in vivo*,

possono implicare una combinazione dei seguenti: 1) modulazione o inattivazione della sequenza di controllo di trascrizione del gene BCL11A, mediante delezioni che insorgono a causa della via di NHEJ, 2) modulazione o inattivazione della sequenza di controllo di trascrizione del gene BCL11A, mediante HDR, o 3) modulazione o inattivazione della sequenza di controllo di trascrizione del gene BCL11A, mediante delezione di almeno una porzione della sequenza di controllo di trascrizione e/o introduzione mediante knock-in di gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata nel locus genico o a una posizione eterologa nel genoma (come un sito safe harbour o “porto sicuro”, come AAVS1). Entrambe le strategie di HDR e knock-in utilizzano un DNA stampo o template donatore nella Riparazione diretta da omologia (HDR). HDR in qualsiasi strategia può venire ottenuta producendo una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB) in siti specifici nel genoma utilizzando una o più endonucleasi.

Per esempio, la strategia di NHEJ può implicare la delezione di almeno una porzione della sequenza di controllo di trascrizione del gene BCL11A inducendo una rottura a singolo filamento o una rottura a doppio filamento all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che

U001.D1.SM.61E

codifica un elemento regolatorio del gene BCL11A con una o più endonucleasi CRISPR e un gRNA (per esempio crRNA + tracrRNA, o sgRNA), o due o più rotture a singolo filamento o rotture a doppio filamento all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con due o più endonucleasi CRISPR e due o più sgRNA. Questo approccio può richiedere lo sviluppo e l'ottimizzazione di sgRNA per la sequenza di controllo di trascrizione del gene BCL11A.

Per esempio, la strategia di HDR può implicare la modulazione o l'inattivazione della sequenza di controllo di trascrizione del gene BCL11A inducendo una rottura a singolo filamento o una rottura a doppio filamento all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una o più endonucleasi CRISPR e un gRNA (per esempio, crRNA + tracrRNA, o sgRNA), o due o più rotture a singolo filamento o rotture a doppio filamento all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una o più endonucleasi CRISPR e due o più gRNA, in presenza di un DNA stampo (template) donatore introdotto esogenamente per dirigere la risposta di DSB cellulare a riparazione diretta da omologia (il DNA stampo donatore può essere un breve oligonucleotide a singolo filamento, un breve oligonucleotide a doppio filamento, una lunga molecola di DNA a singolo o a doppio filamento). Questo approccio può richiedere lo sviluppo e l'ottimizzazione di gRNA e molecole di DNA donatore comprendenti un gene BCL11A di tipo selvatico comprendente una sequenza di

U001.D1.SM.61E

controllo di trascrizione modificata.

Per esempio, la strategia di knock-in implica l'introduzione di un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata nel locus del gene BCL11A utilizzando un gRNA (per esempio crRNA + tracrRNA, o sgRNA) o una coppia di gRNA che bersagliano a monte di o nella sequenza di controllo di trascrizione del gene BCL11A, o in un sito safe harbour ("porto sicuro") come AAVS1). Il DNA donatore può essere DNA a singolo o a doppio filamento e comprende un gene BCL11A di tipo selvatico comprendente una sequenza di controllo di trascrizione modificata.

I vantaggi per le strategie menzionate sopra (delezione/modulazione/inattivazione e knock-in) sono simili, comprendendo in linea di principio sia effetti benefici clinici che di laboratorio a breve e a lungo termine.

In aggiunta alle opzioni di editing elencate sopra, Cas9 o proteine simili possono venire utilizzate per bersagliare domini effettori agli stessi siti bersaglio che possono venire identificati per l'editing, o siti bersaglio addizionali all'interno del range del dominio effettore. Una gamma di enzimi di modificazione di cromatina, metilasi o demetilasi possono venire utilizzate per alterare l'espressione del gene bersaglio. Questi tipi di regolazione epigenetica hanno alcuni vantaggi, in particolare poiché sono limitati in possibili effetti off-target.

La regolazione di trascrizione e traduzione implica un numero di differenti

U001.D1.SM.61E

classi di siti che interagiscono con proteine o nucleotidi cellulari. Spesso i siti di legame a DNA di fattori di trascrizione o altre proteine possono venire bersagliati per mutazione o delezione per studiare il ruolo del sito, sebbene essi possano anche essere bersagliati per cambiare l'espressione genica. Siti possono venire aggiunti attraverso giunzione non omologa delle estremità (NHEJ) o editing genomico diretto mediante riparazione diretta da omologia (HDR). Un uso aumentato di sequenziamento genomico, espressione di RNA e studi sull'intero genoma di legame a fattore di trascrizione hanno aumentato la nostra capacità di identificare come i siti portano a regolazione genica evolutiva o temporale. Questi sistemi di controllo possono essere diretti o possono implicare una regolazione cooperativa estensiva che può richiedere l'integrazione di attività da molteplici potenziatori. Fattori di trascrizione tipicamente legano sequenze di DNA degenerate da 6 a 12 bp di lunghezza. Il basso livello di specificità fornito da siti individuali suggerisce che complesse interazioni e regole sono implicate nel legame e nell'esito funzionale. Siti di legame con meno degenerazione possono fornire mezzi di regolazione più semplici. Fattori di trascrizione artificiali possono venire progettati per specificare sequenze più lunghe che hanno sequenze meno simili nel genoma e hanno un potenziale inferiore per clivaggio off-target. Qualsiasi di questi tipi di siti di legame può venire mutato, cancellato o persino creato per consentire cambiamenti nella regolazione o espressione genica (Canver, M.C. et al., Nature (2015)). Fattori di trascrizione di GATA sono una famiglia di fattori di trascrizione caratterizzati dalla loro capacità di legarsi alla sequenza di legame a DNA di GATA. Una

U001.D1.SM.61E

sequenza di legame a GATA è situata nel sito ipersensibile a DNA +58 (DHS) del gene BCL11A.

Un'altra classe di regioni regolatorie geniche aventi queste caratteristiche è rappresentata da siti di legame a microRNA (miRNA). miRNA sono RNA non codificanti che svolgono ruoli chiave nella regolazione genica post-trascrizionale. miRNA possono regolare l'espressione del 30% di tutti i geni codificanti proteine di mammiferi. Un silenziamento genico specifico e potente mediante RNA a doppio filamento (RNAi) è stato scoperto, più piccolo RNA non codificante supplementare (Canver, M.C. et al., Nature (2015)). La classe più grande di RNA non codificanti importanti per silenziamento genico è rappresentata da miRNA. In mammiferi, miRNA sono prima trascritti come lunghi trascritti di RNA, che possono essere unità trascrizionali separate, parte di introni proteici, o altri trascritti. I lunghi trascritti sono chiamati miRNA primari (pri-miRNA), che includono strutture a forcina con accoppiamento di basi in modo imperfetto. Questi pri-miRNA possono essere clivati in uno o più miRNA precursori più corti (pre-miRNA) mediante Microprocessor, un complesso proteico nel nucleo, che implica Drosha.

Pre-miRNA sono brevi stem loop di circa 70 nucleotidi di lunghezza con una sporgenza 3' di due nucleotidi che sono esportati, nei duplex miRNA:miRNA* maturi di 19-25 nucleotidi. Il filamento di miRNA con una stabilità inferiore di appaiamento di basi (il filamento guida) può essere caricato sul complesso di silenziamento indotto da RNA (RISC). Il filamento guida passeggero (segnato con *) può essere funzionale, ma è solitamente degradato. Il miRNA maturo

U001.D1.SM.61E

lega RISC a motivi di sequenze parzialmente complementari in mRNA bersaglio prevalentemente trovati all'interno delle regioni 3' non tradotte (UTR) e induce silenziamento genico post-trascrizionale (Bartel, D.P. Cell 136, 215-233 (2009); Saj, A. & Lai, E.C. Curr Opin Genet Dev 21, 504-510 (2011)).

miRNA possono essere importanti nello sviluppo, nella differenziazione, nel ciclo cellulare e nella crescita, e in virtualmente tutti i pathway biologici in mammiferi e altri organismi multicellulari. miRNA possono anche essere implicati nel controllo del ciclo cellulare, nell'apoptosi e nella differenziazione di cellule staminali, nell'ematopoiesi, nell'ipossia, nello sviluppo muscolare, nella neurogenesi, nella secrezione di insulina, nel metabolismo del colesterolo, nell'invecchiamento, nella replicazione virale e in risposte immunitarie.

Un singolo miRNA può bersagliare centinaia di differenti trascritti di mRNA, mentre un trascritto individuale può essere bersagliato da molti miRNA differenti. Più di 28645 microRNA sono stati annotati nell'ultima versione rilasciata di miRBase (v.21). Alcuni miRNA possono venire codificati da molteplici loci, alcuni dei quali possono venire espressi da cluster co-trascritti in tandem. Le caratteristiche consentono reticoli regolatori complessi con molteplici pathway e controlli di feedback. miRNA possono essere parti integranti di questi feedback e circuiti regolatori e possono aiutare a regolare l'espressione genica mantenendo la produzione di proteine all'interno di limiti (Herranz, H. & Cohen, S.M. Genes Dev 24, 1339-1344 (2010); Posadas, D.M. & Carthew, R.W. Curr Opin Genet Dev 27, 1-6 (2014)).

miRNA possono anche essere importanti in un grande numero di malattie

U001.D1.SM.61E

umane che sono associate con una espressione abnorme di miRNA. Questa associazione sottolinea l'importanza del pathway regolatorio di miRNA. Studi di delezione di miRNA hanno collegato miRNA con la regolazione delle risposte immunitarie (Stern-Ginossar, N. et al., Science 317, 376-381 (2007)).

miRNA hanno anche un forte collegamento con il cancro e possono svolgere un ruolo in differenti tipi di cancro. miRNA sono stati trovati essere sottoregolati in un numero di tumori. miRNA possono essere importanti nella regolazione di pathway correlati a cancro, come controllo del ciclo cellulare e la risposta a danno del DNA, e possono quindi venire utilizzati nella diagnosi e possono venire bersagliati clinicamente. MicroRNA possono regolare delicatamente l'equilibrio di angiogenesi, in modo tale che esperimenti che depletano tutti i microRNA sopprimono angiogenesi tumorale (Chen, S. et al., Genes Dev 28, 1054-1067 (2014)).

Come è stato mostrato per geni codificanti proteine, geni di miRNA possono anche essere soggetti a cambiamenti epigenetici che si verificano con cancro. Molti loci di miRNA possono essere associate con isole di CpG che aumentano la loro opportunità per la regolazione mediante metilazione di DNA (Weber, B., Stresemann, C., Brueckner, B. & Lyko, F. Cell Cycle 6, 1001-1005 (2007)). La maggior parte degli studi hanno utilizzato un trattamento con farmaci di rimodellamento di cromatina per rivelare miRNA silenziati epigeneticamente.

In aggiunta al loro ruolo nel silenziamento di RNA, miRNA possono anche attivare la traduzione (Posadas, D.M. & Carthew, R.W. Curr Opin Genet Dev 27, 1-6 (2014)). La soppressione o knocking out di questi siti può portare a una

U001.D1.SM.61E

espressione diminuita del gene bersagliato, mentre l'introduzione di questi siti può aumentare l'espressione.

miRNA individuali possono essere soppressi con knock out più efficacemente mutando la sequenza "seed" (basi 2-8 del microRNA), che può essere importante per la specificità di legame. Il clivaggio in questa regione, seguito da una riparazione errata da NHEJ, può effettivamente abolire la funzione di miRNA bloccando il legame a siti bersaglio. miRNA potrebbe anche essere inibito da bersagliamento o targeting specifico della speciale regione di ansa adiacente alla sequenza palindroma. Cas9 cataliticamente inattiva può anche venire utilizzata per inibire l'espressione di shRNA (Zhao, Y. et al., Sci Rep 4, 3943 (2014)). In aggiunta al bersagliamento del miRNA, i siti di legame possono anche essere bersagliati e mutati per impedire il silenziamento da miRNA.

Cellule umane

Per migliorare emoglobinopatie, come descritto e illustrato nel presente documento, i principali bersagli o target per editing genico sono cellule umane. Per esempio, nei metodi *ex vivo*, le cellule umane possono essere cellule somatiche, che dopo essere state modificate utilizzando le tecniche come descritte, possono dare origine a cellule progenitrici. Per esempio, nei metodi *in vivo*, le cellule umane possono essere una cellula di midollo osseo, una cellula progenitrice ematopoietica, o una cellula CD34+.

Effettuando l'editing genico in cellule autologhe che sono derivate da e pertanto già completamente abbinate con il paziente che ne ha necessità, è

U001.D1.SM.61E

possibile generare cellule che possono essere reintrodotte in modo sicuro nel paziente, e dare origine efficacemente a una popolazione di cellule che possono essere efficaci nel migliorare una o più condizioni cliniche associate con la malattia del paziente.

Cellule progenitrici (indicate anche come cellule staminali nel presente documento) sono in grado sia di proliferazione che di dare origine a più cellule progenitrici, queste a loro volta avendo la capacità di generare un grande numero di cellule madri che possono a loro volta dare origine a cellule figlie differenziate o differenziabili. Le cellule figlie stesse possono essere indotte a proliferare e produrre una progenie che successivamente si differenzierà in uno o più tipi di cellule mature, mantenendo nel contempo anche una o più cellule con potenziale evolutivo genitoriale. L'espressione "cellula staminale" si riferisce quindi a una cellula con la capacità o il potenziale, in circostanze particolari, di differenziarsi in un fenotipo più specializzato o differenziato, e che mantiene la capacità, in alcune circostanze, di proliferare sostanzialmente senza differenziazione. In un aspetto, l'espressione cellula progenitrice o staminale si riferisce a una cellula madre generalizzata i cui discendenti (progenie) si specializzano, spesso in direzioni differenti, mediante differenziazione, per esempio acquisendo caratteri completamente individuali, come avviene nella diversificazione progressiva di cellule e tessuti embrionali. La differenziazione cellulare è un processo complesso che si manifesta tipicamente attraverso molte divisioni cellulari. Una cellula differenziata può derivare da una cellula multipotente che è essa stessa derivata da una cellula

U001.D1.SM.61E

multipotente, e così via. Mentre ciascuna di queste cellule multipotenti può essere considerata una cellula staminale, la gamma di tipi cellulari a cui ciascuna può dare origine può variare considerevolmente. Alcune cellule differenziate hanno anche la capacità di dare origine a cellule di potenziale evolutivo maggiore. Tale capacità può essere naturale o può essere indotta artificialmente con trattamento con vari fattori. In molti casi biologici, cellule staminali possono anche essere “multipotenti” poiché possono produrre una progenie di più di un tipo cellulare distinto, ma questo non è necessario per la “staminalità”.

L’autorinnovamento può essere un altro aspetto importante della cellula staminale. In teoria, l’autorinnovamento può verificarsi mediante uno di due meccanismi maggiori. Cellule staminali possono dividersi asimmetricamente, una figlia mantenendo lo stato staminale e l’altra figlia esprimendo qualche altra funzione e fenotipo specifici distinti. In alternativa, alcune delle cellule staminali in una popolazione possono dividersi simmetricamente in due staminali, mantenendo così alcune cellule staminali nella popolazione complessiva, mentre altre cellule nella popolazione danno origine soltanto a una progenie differenziata. Generalmente, “cellule progenitrici” hanno un fenotipo cellulare che è più primitivo (vale a dire è a una fase precedente lungo un pathway di sviluppo o una progressione rispetto a una cellula completamente differenziata). Spesso, cellule progenitrici hanno anche un potenziale significativo o proliferativo molto elevato. Cellule progenitrici possono dare origine a molteplici tipi cellulari differenziati distinti o a un singolo tipo cellulare differenziato, a seconda del pathway di sviluppo e dell’ambiente in cui

U001.D1.SM.61E

le cellule si sviluppano e si differenziano.

Nel contesto dell'ontogenia cellulare l'aggettivo "differenziato" o "differenziante" è un termine relativo. Una "cellula differenziata" è una cellula che è progredita ulteriormente nel pathway di sviluppo rispetto alla cellula a cui essa è confrontata. Quindi, cellule staminali possono differenziarsi in cellule precursori ristrette nel lignaggio (come una cellula progenitrice ematopoietica), che a loro volta possono differenziarsi in altri tipi di cellule precursori ulteriormente lungo il pathway (come un precursore ematopoietico), e così fino a una cellula differenziata nello stadio terminale, come un eritrocito, che svolge un ruolo caratteristico in un certo tipo di tessuto, e può o meno mantenere la capacità di proliferare ulteriormente.

L'espressione "cellula progenitrice ematopoietica" si riferisce a cellule di un lignaggio di cellule staminali che danno origine a tutti i tipi di cellule sanguigne, incluse cellule eritroidi (eritrociti o globuli rossi (RBC)), mieloidi (monociti e macrofagi, neutrofili, basofili, eosinofili, megacariociti/piastrine, e cellule dendritiche), e linfoidi (cellule T, cellule B, cellule NK).

Una "cellula del lignaggio eritroide" indica che la cellula contattata è una cellula che subisce eritropoiesi, in modo tale che durante la differenziazione finale forma un eritrocito o globulo rosso. Tale cellule hanno origine da cellule progenitrici ematopoietiche di midollo osseo. All'esposizione a specifici fattori di crescita e altri componenti dell'ambiente ematopoietico, cellule progenitrici ematopoietiche possono maturare attraverso una serie di tipi cellulari di differenziazione intermedi, tutti intermedi del lignaggio eritroide, in RBC.

U001.D1.SM.61E

Quindi, cellule del “lignaggio eritroide” comprendono cellule progenitrici ematopoietiche, rubriblasti, prorubriciti, eritroblasti, metarubriciti, reticolociti, e eritrociti.

La cellula progenitrice ematopoietica può esprimere almeno uno dei seguenti marcatori di superficie cellulare caratteristici di cellule progenitrici ematopoietiche: CD34+, CD59+, Thyl/CD90+, CD381o/-, e C-kit/CDI 17+. In alcuni esempi forniti nel presente documento i progenitori ematopoietici possono essere CD34+.

La cellula progenitrice ematopoietica può essere una cellula staminale di sangue periferico ottenuta dal paziente dopo che il paziente è stato trattato con uno o più fattori come fattore di stimolazione di colonie di granulociti (opzionalmente in combinazione con Plerixaflor). Cellule CD34+ possono essere arricchite utilizzando il sistema di selezione cellulare CliniMACS® Cell Selection System (Miltenyi Biotec). Cellule CD34+ possono venire stimulate in mezzo senza siero (per esempio mezzo CellGrow SCGM, CellGenix) con citochine (per esempio SCF, rhTPO, rhFLT3) prima dell’editing genomico. L’aggiunta di SR1 e dmPGE2 e/o altri fattori è contemplata per migliorare l’attecchimento o engraftment a lungo termine.

Le cellule progenitrici ematopoietiche del lignaggio eritroide possono avere un marcatore di superficie cellulare caratteristico del lignaggio eritroide, come CD71 e Terl 19.

Cellule staminali ematopoietiche (HSC) possono essere un importante bersaglio per terapia genica poiché forniscono una fonte prolungata delle cellule

U001.D1.SM.61E

corrette. HSC danno origine sia a lignaggi mieloidi che linfoidi di cellule ematiche. Cellule ematiche mature hanno una durata di vita finite e devono essere continuamente sostituite durante la vita. Cellule del sangue sono continuamente prodotte dalla proliferazione e differenziazione di una popolazione di HSC pluripotenti che possono essere ricostituite mediante autorinnovamento. Il midollo osseo (BM) è il sito maggiore di ematopoiesi in esseri umani e una buona fonte per cellule staminali e progenitrici ematopoietiche (HSPC). HSPC possono essere presenti in piccoli numeri nel sangue periferico (PB). In alcune indicazioni o trattamenti i loro numeri aumentano. La progenie di HSC matura attraverso gli stadi, generando cellule progenitrici multipotenziali e preposte a lignaggio comprendenti le cellule progenitrici linfoidi che danno origine alle cellule esprimenti BCL 11A. Progenitori di cellule B e T sono le due popolazioni di cellule che richiedono l'attività di BCL11A, così esse potrebbero essere editate agli stadi prima del riarrangiamento, sebbene la correzione di progenitori abbia il vantaggio di continuare ad essere una fonte di cellule corrette. Cellule trattate, come cellule CD34+, dovrebbero essere reintrodotte nel paziente. Il livello di attecchimento o engraftment può essere importante, come lo è la capacità dell'attecchimento o engraftment multilignaggio delle cellule di cellule sottoposte a editing genico dopo infusione di CD34+ *in vivo*.

Cellule staminali pluripotenti indotte

Le cellule umane ingegnerizzate geneticamente descritte nel presente documento possono essere cellule staminali pluripotenti indotte (iPSC). Un

U001.D1.SM.61E

vantaggio dell'utilizzo di iPSC è che le cellule possono essere derivate dallo stesso soggetto a cui le cellule progenitrici devono essere somministrate. Vale a dire, una cellula somatica può essere ottenuta da un individuo, riprogrammata in una cellula staminale pluripotente indotta, e poi ridifferenziata in una cellula progenitrice da somministrare all'individuo (per esempio cellule autologhe). Poiché i progenitori sono essenzialmente derivati da una fonte autologa, il rischio di rigetto di attecchimento o engraftment o risposta allergica può essere ridotto rispetto all'uso di cellule da un altro individuo o gruppo di individui. Inoltre, l'uso di iPSC nega la necessità di cellule ottenute da una fonte embrionale. Quindi, in un aspetto, le cellule staminali utilizzate nei metodi descritti non sono cellule staminali embrionali.

Sebbene la differenziazione sia generalmente irreversibile in contesti fisiologici, numerosi metodi sono stati recentemente sviluppati per riprogrammare cellule somatiche in iPSC. Metodi esemplari sono noti agli esperti nel ramo e sono descritti brevemente qui sotto.

Il termine "riprogrammazione" si riferisce a un processo che altera o inverte lo stato di differenziazione di una cellula differenziata (per esempio una cellula somatica). In altri termini, riprogrammazione si riferisce a un processo per far ritornare la differenziazione di una cellula a un tipo più indifferenziato o più primitivo di cellula. Si noti che collocare molte cellule primarie in coltura può portare a qualche perdita di caratteristiche completamente differenziate. Quindi, la semplice coltura di tali cellule incluse nell'espressione "cellule differenziate" non rende queste cellule non differenziate (per esempio cellule indifferenziate) o

U001.D1.SM.61E

cellule pluripotenti. La transizione di una cellula differenziata a pluripotenza richiede uno stimolo di riprogrammazione oltre gli stimoli ce portano a una parziale perdita del carattere differenziato in coltura. Cellule riprogrammate hanno anche la caratteristica della capacità di passaggi estesi senza perdita di potenziale di crescita, rispetto a genitori cellulari primari, che generalmente hanno la capacità di soltanto un numero limitato di divisioni in coltura.

La cellula da riprogrammare può essere parzialmente o terminalmente differenziata prima della riprogrammazione. La riprogrammazione può comprendere una completa reversione dello stato di differenziazione di una cellula differenziata (per esempio una cellula somatica) a uno stato pluripotente o uno stato multipotente. La riprogrammazione può comprendere una completa o parziale reversione dello stato di differenziazione di una cellula differenziata (per esempio una cellula somatica) a una cellula indifferenziata (per esempio una cellula simile a embrionale). La riprogrammazione può portare all'espressione di particolari geni da parte delle cellule, l'espressione dei quali contribuisce inoltre alla riprogrammazione. In alcuni esempi qui descritti, la riprogrammazione di una cellula differenziata (per esempio una cellula somatica) può far sì che la cellula differenziata assuma uno stato indifferenziato (per esempio è una cellula indifferenziata). Le cellule risultanti sono indicate come "cellule riprogrammate", o "cellule staminali pluripotenti indotte (iPSC o cellule iPS)".

La riprogrammazione può implicare una alterazione, per esempio inversione, di almeno alcuni dei pattern ereditabili di modificazione di acidi nucleici (per

U001.D1.SM.61E

esempio metilazione), condensazione di cromatina, cambiamenti epigenetici, imprinting genomico, etc., che si verificano durante la differenziazione cellulare. La riprogrammazione è distinta dal semplice mantenimento dello stato indifferenziato esistente di una cellula che è già pluripotente o dal mantenimento dello stato inferiore a completamente differenziato esistente di una cellula che è già una cellula multipotente (per esempio una cellula staminale ematopoietica). La riprogrammazione è anche distinta dalla promozione dell'autorinnovamento o della proliferazione di cellule che sono già pluripotenti o multipotenti, sebbene le composizioni e i metodi qui descritti possano anche essere d'uso per tali scopi, in alcuni esempi.

Sono noti nell'arte molti metodi che possono venire utilizzati per generare cellule staminali pluripotenti da cellule somatiche. Qualsiasi metodo di questo tipo che riprogrammi una cellula somatica al fenotipo pluripotente sarebbe appropriato per l'uso nei metodi qui descritti.

Sono state descritte metodologie di riprogrammazione per generare cellule pluripotenti utilizzando combinazioni definite di fattori di trascrizione. Cellule somatiche di topo possono venire convertite in cellule simili a cellule ES con potenziale di sviluppo espanso mediante la trasduzione diretta di Oct4, Sox2, Klf4, e c-Myc; si veda, per esempio, Takahashi e Yamanaka, Cell 126(4): 663-76 (2006). iPSC assomigliano a cellule ES, poiché esse ripristinano la circuitaria trascrizionale associata a pluripotenza e molto del paesaggio epigenetico. In aggiunta, iPSC di topo soddisfano tutti i saggi standard per pluripotenza: specificamente, differenziazione *in vitro* in tipi di cellule dei tre strati germinali,

U001.D1.SM.61E

formazione di teratoma, contributo a chimere, trasmissione di linea germinale [vedere, per esempio, Maherali e Hochedlinger, *Cell Stem Cell*. 3(6):595-605 (2008)], e complementazione tetraploide.

iPSC umane possono essere ottenute utilizzando metodi di trasduzione simili, e il trio di fattori di trascrizione, OCT4, SOX2, e NANOG, è stato stabilito come l'insieme fondamentale di fattori di trascrizione che governano pluripotenza; vedere, per esempio, Budniatzky e Gepstein, *Stem Cells Transl Med*. 3(4):448-57 (2014); Barrett et al., *Stem Cells Trans Med* 3:1-6 sctm.2014-0121 (2014); Focosi et al., *Blood Cancer Journal* 4: e211 (2014); e riferimenti citati all'interno di essi. La produzione di iPSC può essere ottenuta mediante l'introduzione di sequenze di acidi nucleici codificanti geni associati a cellule staminali in un adulto, una cellula somatica, storicamente utilizzando vettori virali.

iPSC possono venire generate o derivate da cellule somatiche differenziate terminalmente, nonché da cellule staminali adulte, o cellule staminali somatiche. Vale a dire, una cellula progenitrice non pluripotente può essere resa pluripotente o multipotente mediante riprogrammazione. In tali casi, può non essere necessario includere così tanti fattori di riprogrammazione quanti necessari per riprogrammare una cellula differenziata terminalmente. Inoltre, la riprogrammazione può essere indotta mediante l'introduzione non virale di fattori di riprogrammazione, per esempio introducendo le proteine stesse, o introducendo acidi nucleici che codificano i fattori di riprogrammazione, o introducendo RNA messaggeri che con la traduzione producono i fattori di

U001.D1.SM.61E

riprogrammazione (vedere, per esempio, Warren et al., *Cell Stem Cell*, 7(5):618-30 (2010). La riprogrammazione può essere ottenuta introducendo una combinazione di acidi nucleici codificanti geni associati a cellule staminali, inclusi, per esempio, Oct-4 (noto anche come Oct-3/4 o Pouf51), Sox1, Sox2, Sox3, Sox 15, Sox 18, NANOG, Klf1, Klf2, Klf4, Klf5, NR5A2, c-Myc, 1-Myc, n-Myc, Rem2, Tert, e LIN28. La riprogrammazione utilizzando i metodi e le composizioni qui descritti può inoltre comprendere l'introduzione di uno o più di Oct-3/4, un membro della famiglia di Sox, un membro della famiglia di Klf, e un membro della famiglia di Myc in una cellula somatica. I metodi e le composizioni qui descritti possono inoltre comprendere l'introduzione di uno o più di ciascuno di Oct-4, Sox2, Nanog, c-MYC e Klf4 per riprogrammazione. Come indicato sopra, il metodo esatto utilizzato per la riprogrammazione non è necessariamente di importanza critica per i metodi e le composizioni qui descritti. Tuttavia, dove cellule differenziate dalle cellule riprogrammate devono essere utilizzate, per esempio, in terapia umana, in un aspetto la riprogrammazione non è effettuata mediante un metodo che altera il genoma. Quindi, in tali esempi, la riprogrammazione può essere ottenuta, per esempio, senza l'uso di vettori virali o plasmidici.

L'efficienza di riprogrammazione (vale a dire il numero di cellule riprogrammate) derivata da una popolazione di cellule di partenza può essere potenziata mediante l'aggiunta di vari agenti, per esempio piccole molecole, come mostrato da Shi et al., *Cell-Stem Cell* 2:525-528 (2008); Huangfu et al., *Nature Biotechnology* 26(7):795-797 (2008) e Marson et al., *Cell-Stem Cell* 3:

U001.D1.SM.61E

132-135 (2008). Quindi, un agente o una combinazione di agenti che potenziano l'efficienza o il tasso di produzione di cellule staminali pluripotenti indotte possono essere utilizzati nella produzione di iPSC specifiche per paziente o specifiche per malattia. Alcuni esempi non limitativi di agenti che potenziano l'efficienza di riprogrammazione includono Wnt solubile, mezzo condizionato con Wnt, BIX-01294 (una G9a istone metiltransferasi), PD0325901 (un inibitore di MEK), inibitori di DNA metiltransferasi, inibitori di istone deacetilasi (HDAC), acido valproico, 5'-azacitidina, dexamethasone, suberoilamide, acido idrossamico (SAHA), vitamina C, e tricostatina (TSA), tra gli altri.

Altri esempi non limitativi di agenti che potenziano la riprogrammazione includono: Suberoilamide acido idrossamico (SAHA (ad esempio MK0683, vorinostat) e altri acidi idrossamici), BML-210, Depudecin (ad esempio (-)-Depudecin), Tossina HC, Nullscript (4-(1,3-Diosso-1H,3H-benzo[de]isochinolin-2-il)-N-idrossibutanamide), Fenilbutirrato (ad esempio fenilbutirrato di sodio) e Acido valproico ((VP A) e altri acidi grassi a catena corta), Scriptaid, Suramin Sodio, Tricostatina A (TSA), Composto APHA 8, Apicidin, Butirrato di sodio, pivaloilossimetil butirrato (Pivanex, AN-9), Trapoxin B, Chlamydocin, Depsipeptide (noto anche come FR901228 o FK228), benzammidi (per esempio CI-994 (per esempio N-acetil dinaline) e MS-27-275), MGCD0103, NVP-LAQ-824, CBHA (acido m-carbssicinnaminico, acido bisidrossamico), JNJ16241199, Tubacin, A-161906, proxamide, oxamflatin, 3-CI-UCHA (ad esempio acido 6-(3-clorofenilureido)caproico idrossamido), AOE (acido 2-ammino-8-osso-9, 10-epossidodecanoico), CHAP31 e CHAP 50. Altri agenti potenziatori di

U001.D1.SM.61E

riprogrammazione includono, per esempio, forme negative dominanti di HDAC (ad esempio forme cataliticamente inattive), inibitori di siRNA di HDAC, e anticorpi che si legano specificamente a HDAC. Tali inibitori sono disponibili, per esempio, da BIOMOL International, Fukasawa, Merck Biosciences, Novartis, Gloucester Pharmaceuticals, Titan Pharmaceuticals, MethylGene, e Sigma Aldrich.

Per confermare l'induzione di cellule staminali pluripotenti per l'uso con i metodi qui descritti, cloni isolati possono venire testati per l'espressione di un marcatore di cellule staminali. Tale espressione in una cellula derivata da una cellula somatica identifica le cellule come cellule staminali pluripotenti indotte. Marcatori di cellule staminali possono essere selezionati dal gruppo non limitativo comprendente SSEA3, SSEA4, CD9, Nanog, Fbx15, Ecatl, Esgl, Eras, Gdf3, Fgf4, Cripto, Daxl, Zpf296, Slc2a3, Rexl, Utfl, e Natl. In un caso, per esempio, una cellula che esprime Oct4 o Nanog è identificata come pluripotente. Metodi per rilevare l'espressione di tali marcatori possono includere, per esempio, RT-PCR e metodi immunologici che rilevano la presenza dei polipeptidi codificati, come Western blot o analisi flussocitometriche. Il rilevamento può implicare non soltanto RT-PCR, ma può anche includere il rilevamento di marcatori proteici. Marcatori intracellulari possono essere meglio identificati tramite RT-PCR, o metodi di rilevamento di proteine come immunocitochimica, mentre marcatori di superficie cellulare sono prontamente identificati, per esempio, mediante immunocitochimica.

Il carattere di cellule staminali pluripotenti di cellule isolate può essere

U001.D1.SM.61E

confermato da test che valutano la capacità delle iPSC di differenziarsi in cellule di ciascuno dei tre strati germinali. Come un esempio, la formazione di teratoma in topi nudi può essere utilizzata per valutare il carattere pluripotente dei cloni isolati. Le cellule possono venire introdotte in topi nudi e istologia e/o immunoistochimica possono venire effettuate su un tumore proveniente dalle cellule. La crescita di un tumore comprendente cellule da tutti i tre strati germinali, per esempio, indica inoltre che le cellule sono cellule staminali pluripotenti.

Creazione di iPSC specifiche per paziente

Una fase dei metodi *ex vivo* della presente descrizione può comportare la creazione di una cellula iPS specifica per paziente, cellule iPS specifiche per paziente, o una linea di cellule iPS specifiche per paziente. Esistono metodi consolidati nell'arte per creare cellule iPS specifiche per paziente, come descritti in Takahashi e Yamanaka 2006; Takahashi, Tanabe *et al.* 2007. Per esempio, la fase di creazione può comprendere: a) l'isolamento di una cellula somatica, come una cellula cutanea o un fibroblasto, dal paziente; e b) l'introduzione di un set di geni associati a pluripotenza nella cellula somatica al fine di indurre la cellula a diventare una cellula staminale pluripotente. Il set di geni associati a pluripotenza può essere rappresentato da uno o più dei geni scelti nel gruppo costituito da OCT4, SOX2, KLF4, Lin28, NANOG, e cMYC.

Esecuzione di una biopsia o aspirato del midollo osseo del paziente

Una biopsia o aspirato è un campione di tessuto o fluido prelevato dal corpo. Esistono molti tipi differenti di biopsie o aspirati. Quasi tutti di essi implicano

U001.D1.SM.61E

l'uso di un utensile aguzzo per rimuovere una piccola quantità di tessuto. Se la biopsia sarà sulla pelle o su altra area sensibile, può prima venire applicato un medicinale anestetizzante. Una biopsia o aspirato può essere effettuata/o secondo uno qualsiasi dei metodi noti nell'arte. Per esempio, in un aspirato di midollo osseo, un grande ago è utilizzato per penetrare nell'osso pelvico per raccogliere midollo osseo.

Isolamento di una cellula staminale mesenchimale

Cellule staminali mesenchimali possono venire isolate secondo qualsiasi metodo noto nell'arte, come dal midollo osseo o dal sangue periferico di un paziente. Per esempio, un aspirato di midollo può essere raccolto in una siringa con eparina. Cellule possono venire lavate e centrifugate su un gradiente di densità Percoll™. Cellule, come cellule sanguigne, cellule epatiche, cellule interstiziali, macrofagi, mastociti, e timociti, possono venire separati utilizzando Percoll™. Le cellule possono venire coltivate in mezzo di Eagle modificato da Dulbecco (DMEM) (basso glucosio) contenente siero bovino fetale (FBS) al 10% (Pittinger MF, Mackay AM, Beck SC et al., Science 1999; 284:143-147).

Trattamento di un paziente con GCSF

Un paziente può opzionalmente venire trattato con un fattore di stimolazione di colonie di granulociti (GCSF) conformemente a qualsiasi metodo noto nell'arte. Il GCSF può essere somministrato in combinazione con Plerixaflor.

Isolamento di una cellula progenitrice ematopoietica da un paziente

Una cellula progenitrice ematopoietica può venire isolata da un paziente mediante qualsiasi metodo noto nell'arte. Cellule CD34+ possono venire

U001.D1.SM.61E

arricchite utilizzando il sistema di selezione cellulare CliniMACS® Cell Selection System (Miltenyi Biotec). Cellule CD34+ possono venire stimulate debolmente in mezzo senza siero (per esempio mezzo CellGrow SCGM, CellGenix) con citochine (per esempio SCF, rhTPO, rhFLT3) prima di editing genomico.

Editing genomico

Editing genomico generalmente si riferisce al processo di modificazione della sequenza nucleotidica di un genoma, preferibilmente in una maniera precisa o predeterminata. Esempi di metodi di editing genomico descritti nel presente documento includono metodi di utilizzo di nucleasi sito-dirette per tagliare acido desossiribonucleico (DNA) in posizioni bersaglio precise nel genoma, creando quindi rotture di DNA a singolo o a doppio filamento in particolari posizioni all'interno del genoma. Tali rotture possono essere e regolarmente sono riparate da processi cellulari endogeni naturali, come riparazione diretta da omologia (HDR) e NHEJ, come esaminato recentemente in Cox et al., Nature Medicine 21(2), 121-31 (2015). Questi due processi principali di riparazione del DNA consistono in una famiglia di pathway alternativi. NHEJ unisce direttamente le estremità del DNA risultanti da una rottura a doppio filamento, talvolta con la perdita o l'aggiunta di sequenza nucleotidica, che può interrompere o potenziare l'espressione genica. HDR utilizza una sequenza omologa, o sequenza donatrice, come uno stampo o template per l'inserimento di una sequenza di DNA definita al punto di rottura. La sequenza omologa può essere nel genoma endogeno, come un cromatidio sorella. In alternativa, il donatore può essere un

U001.D1.SM.61E

acido nucleico esogeno, come un plasmide, un oligonucleotide a singolo filamento, un oligonucleotide a doppio filamento, un oligonucleotide duplex o un virus, che ha regioni di elevata omologia con il locus clivato da nucleasi, ma che può anche contenere sequenza addizionale o cambiamenti di sequenza, incluse delezioni, che possono venire incorporati nel locus bersaglio clivato. Un terzo meccanismo di riparazione può essere la giunzione delle estremità mediata da micro-omologia (MMEJ), nota anche come "NHEJ alternativa", in cui l'esito genetico è simile a NHEJ nel senso che piccole delezioni e inserzioni possono avvenire a livello del sito di clivaggio. MMEJ può fare uso di sequenze omologhe di poche paia di basi che fiancheggiano il sito di rottura del DNA per guidare un esito di riparazione di giunzione delle estremità di DNA maggiormente preferito, e report recenti hanno ulteriormente delucidato il meccanismo molecolare di questo processo; si vedano, per esempio, Cho e Greenberg, *Nature* 518, 174-76 (2015); Kent et al., *Nature Structural and Molecular Biology*, Adv. Online doi:10.1038/nsmb.2961 (2015); Mateos-Gomez et al., *Nature* 518, 254-57 (2015); Ceccaldi et al., *Nature* 528, 258-62 (2015). In alcuni casi può essere possibile predire esiti di riparazione probabili sulla base dell'analisi di potenziali micro-omologie in corrispondenza del sito della rottura del DNA.

Ciascuno di questi meccanismi di editing genomico può essere utilizzato per creare alterazioni genomiche desiderate. Una fase nel processo di editing genomico può consistere nel creare una o due rotture di DNA, quest'ultima come rotture a doppio filamento o come due rotture a singolo filamento, nel

U001.D1.SM.61E

locus bersaglio come vicino al sito di mutazione desiderato. Ciò può essere ottenuto tramite l'uso di polipeptidi sito-diretti, come descritto e illustrato nel presente documento.

Polipeptidi sito-diretti, come una DNA endonucleasi, possono introdurre rotture a doppio filamento o rotture a singolo filamento in acidi nucleici, per esempio DNA genomico. La rottura a doppio filamento può stimolare un pathway di riparazione di DNA endogeno di una cellula (per esempio riparazione dipendente da omologia o giunzione non omologa delle estremità o giunzione non omologa delle estremità alternativa (A-NHEJ) o giunzione delle estremità mediata da micro-omologia). NHEJ può riparare acido nucleico bersaglio clivato senza la necessità di un template omologo. Questo può talvolta avere come risultato piccole delezioni o inserzioni (indel) nell'acido nucleico bersaglio in corrispondenza del sito di clivaggio, e può portare all'interruzione o alterazione di espressione genica. HDR può verificarsi quando un template di riparazione omologo, o donatore, è disponibile. Il template donatore omologo può comprendere sequenze che possono essere omologhe a sequenze fiancheggianti il sito di clivaggio di acido nucleico bersaglio. Il cromatidio sorella può essere utilizzato dalla cellula come template di riparazione. Tuttavia, ai fini dell'editing genomico, il template di riparazione può essere fornito come un acido nucleico esogeno, come un plasmide, oligonucleotide duplex, oligonucleotide a singolo filamento, oligonucleotide a doppio filamento, o acido nucleico virale. Con template donatori esogeni, una sequenza di acido nucleico addizionale (come un transgene) o una modificazione (come un cambiamento di

U001.D1.SM.61E

singola base o di molteplici basi o una delezione) possono venire introdotti tra le regioni fiancheggianti di omologia cosicché la sequenza di acido nucleico addizionale o alterata diventi anche incorporata nel locus bersaglio. MMEJ può portare a un esito genetico che è simile a NHEJ nel senso che piccole delezioni e inserzioni possono avere luogo a livello del sito di clonaggio. MMEJ può utilizzare sequenze omologhe di poche paia di basi fiancheggianti il sito di clivaggio per guidare un esito di riparazione di DNA di giunzione delle estremità preferito. In alcuni casi può essere possibile predire esiti di riparazione probabili sulla base dell'analisi di micro-omologie potenziali nelle regioni bersaglio di nucleasi.

Quindi, in alcuni casi, ricombinazione omologa può venire utilizzata per inserire una sequenza polinucleotidica esogena nel sito di clivaggio di acido nucleico bersaglio. Una sequenza polinucleotidica esogena è chiamata un polinucleotide donatore (o donatore o sequenza donatrice o template donatore polinucleotidico) nel presente documento. Il polinucleotide donatore, una porzione del polinucleotide donatore, una copia del polinucleotide donatore, o una porzione di una copia del polinucleotide donatore può venire inserito/a nel sito di clivaggio di acido nucleico bersaglio. Il polinucleotide donatore può essere una sequenza polinucleotidica esogena, vale a dire una sequenza che non si manifesta naturalmente a livello del sito di clivaggio di acido nucleico bersaglio.

Le modificazioni del DNA bersaglio dovute a NHEJ e/o HDR possono portare a, per esempio, mutazioni, delezioni, alterazioni, integrazioni, correzione genica,

U001.D1.SM.61E

sostituzione genica, tagging genico, inserzione transgenica, delezione nucleotidica, distruzione genica, traslocazioni e/o mutazione genica. I processi di delezione di DNA genomico e integrazione di acido nucleico non nativo in DNA genomico sono esempi di editing genomico.

Sistema di endonucleasi CRISPR

Un locus genomico CRISPR (brevi ripetizioni palindrome raggruppate e separate a intervalli regolari) può essere trovato nei genomi di molti procarioti (per esempio batteri e archei). In procarioti, il locus CRISPR codifica prodotti che funzionano come un tipo di sistema immunitario per aiutare a difendere i procarioti contro invasori estranei, come virus e fagi. Esistono tre stadi di funzione di locus CRISPR: integrazione di nuove sequenze nel locus CRISPR, espressione di CRISPR RNA (crRNA), e silenziamento di acido nucleico invasore estraneo. Sono stati identificati cinque tipi di sistemi di CRISPR (ad esempio Tipo I, Tipo II, Tipo III, Tipo U, e Tipo V).

Un locus CRISPR include un numero di brevi sequenze ripetute indicate come “ripetizioni”. Quando espresse, le ripetizioni possono formare strutture secondarie (per esempio forcine) e/o comprendere sequenze a singolo filamento non strutturate. Le ripetizioni solitamente avvengono in cluster e spesso divergono tra specie. Le ripetizioni sono regolarmente intervallate con sequenze intermedie uniche indicate come “spaziatori”, che danno luogo a una architettura di locus ripetizione-spaziatore-ripetizione. Gli spaziatori sono identici a o hanno una elevata omologia con sequenze invasori estranee note. Una unità spaziatore-ripetizione codifica un crisprRNA (crRNA), che è processato in una forma

U001.D1.SM.61E

matura dell'unità spaziatore-ripetizione. Un crRNA comprende una sequenza "seed" o sequenza spaziatrice che è implicata nel bersagliamento di un acido nucleico bersaglio (nella forma naturalmente presente in procarioti, la sequenza spaziatrice bersaglia l'acido nucleico invasore estraneo). Una sequenza spaziatrice è situate all'estremità 5' o 3' del crRNA.

Un locus CRISPR comprende anche sequenze polinucleotidiche codificanti geni associati a CRISPR (Cas). Geni Cas codificano endonucleasi implicate nella biogenesi e negli stadi di interferenza di funzione di crRNA in procarioti. Alcuni geni Cas comprendono strutture secondarie e/o terziarie omologhe.

Sistemi CRISPR di tipo II

La biogenesi di crRNA in un sistema CRISPR di Tipo II in natura richiede un CRISPR RNA trans-attivante (tracrRNA). Il tracrRNA può essere modificato da RNasi III endogena, e poi si ibrida a una ripetizione di crRNA nell'array pre-crRNA. RNasi III endogena può essere reclutata per clivare il pre-crRNA. crRNA clivati possono essere sottoposti a trimming di esoribonucleasi per produrre la forma matura di crRNA (per esempio trimming di 5'). Il tracrRNA può rimanere ibridato al crRNA, e il tracrRNA e il crRNA si associano con un polipeptide sito-diretto (per esempio Cas9). Il crRNA del complesso crRNA-tracrRNACas9 può guidare il complesso a un acido nucleico bersaglio a cui il crRNA può ibridarsi. L'ibridazione del crRNA all'acido nucleico bersaglio può attivare Cas9 per clivaggio di acido nucleico bersagliato. L'acido nucleico bersaglio in un sistema CRISPR di tipo II è indicato come un motivo adiacente protospaziatore (PAM). In natura il PAM è essenziale per facilitare il legame di

U001.D1.SM.61E

un polipeptide sito-diretto (per esempio Cas9) all'acido nucleico bersaglio. Sistemi di tipo II (indicati anche come Nmeni o CASS4) sono inoltre suddivisi in Tipo II-A (CASS4) e II-B (CASS4a). Jinek et al., Science, 337(6096):816-821 (2012) hanno mostrato che il sistema CRISPR/Cas9 è utile per editing genomico programmabile con RNA e la pubblicazione di domanda di brevetto internazionale n° WO2013/176772 fornisce numerosi esempi e applicazioni del sistema di endonucleasi CRISPR/Cas per editing genico sito-specifico.

Sistemi CRISPR di Tipo V

Sistemi CRISPR di Tipo V hanno numerose differenze importanti da sistemi di Tipo II. Per esempio, Cpf1 è una singola enonucleasi guidata da RNA che, a differenza di sistemi di Tipo II, manca di tracrRNA. Infatti, array di CRISPR associati a Cpf1 possono essere processati in crRNA maturi senza la necessità di un tracrRNA transattivante supplementare. L'array di CRISPR di Tipo V può essere processato in brevi crRNA maturi di 42-44 nucleotidi di lunghezza, ciascun crRNA maturo iniziando con 19 nucleotidi di ripetizione diretta seguiti da 23-25 nucleotidi di sequenza spaziatrice. Al contrario, crRNA maturi in sistemi di Tipo II possono iniziare con 20-24 nucleotidi di sequenza spaziatrice seguiti da circa 22 nucleotidi di ripetizione diretta. Inoltre, Cpf1 può utilizzare un motivo adiacente a protospaziatore ricco di T in modo tale che complessi Cpf1-crRNA clivino efficacemente DNA bersaglio preceduto da un breve PAM ricco di T, che è in contrasto con il PAM ricco di G seguente il DNA bersaglio per sistemi di Tipo II. Quindi, sistemi di Tipo V clivano in un punto che è distante dal PAM, mentre sistemi di tipo II clivano in un punto che è adiacente

U001.D1.SM.61E

al PAM. Inoltre, a differenza di sistemi di Tipo II, Cpf1 cliva DNA tramite una rottura sfalsata a doppio filamento di DNA con una sporgenza 5' di 4 o 5 nucleotidi. Sistemi di Tipo II clivano tramite una rottura smussata a doppio filamento. Similmente a sistemi di Tipo II, Cpf1 contiene un dominio predetto di endonucleasi simile a RuvC, ma manca di un secondo dominio di endonucleasi HNH, che è in contrasto con sistemi di Tipo II.

Geni/polipeptidi Cas e motivi adiacenti protospaziatori

Polipeptidi CRISPR/Cas esemplari includono i polipeptidi Cas9 nella Fig. 1 di Fonfara et al., *Nucleic Acids Research*, 42: 2577-2590 (2014). Il sistema di denominazione di geni CRISPR/Cas ha subito una riscrittura estesa da quando i geni Cas sono stati scoperti. La Figura 5 di Fonfara, *supra*, fornisce sequenze di PAM per i polipeptidi Cas9 da varie specie.

Polipeptidi sito-diretti

Un polipeptide sito-diretto è una nucleasi utilizzata in editing genomico per clivare DNA. La nucleasi o polipeptide sito-diretta(o) può essere somministrata(o) a una cellula o a un paziente come: uno o più polipeptidi, o uno o più mRNA codificanti il polipeptide.

Nel contest di un sistema CRISPR/Cas o CRISPR/Cpf1, il polipeptide sito-diretto può legarsi a un RNA guida che, a sua volta, specifica il sito nel DNA bersaglio a cui il polipeptide è diretto. Nei sistemi CRISPR/Cas o CRISPR/Cpf1 qui descritti il polipeptide sito-diretto può essere una endonucleasi, come una DNA endonucleasi.

Un polipeptide sito-diretto può comprendere una pluralità di domini di

U001.D1.SM.61E

clivaggio di acido nucleico (vale a dire nucleasi). Due o più domini di clivaggio di acido nucleico possono essere collegati insieme tramite un linker. Per esempio, il linker può comprendere un linker flessibile. Linker possono comprendere 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40 o più amminoacidi di lunghezza.

Enzimi Cas9 di tipo selvatico presenti naturalmente comprendono due domini nucleasici, un dominio nucleasico HNH e un dominio RuvC. Qui, "Cas9" sia a Cas9 presente naturalmente che a Cas9 ricombinante. Enzimi Cas9 contemplati nel presente documento possono comprendere un dominio nucleasico HNH o simile a HNH, e/o un dominio nucleasico RuvC o simile a RuvC.

Domini HNH o simili a HNH comprendono un ripiegamento simile a McrA. Domini HNH o simili a HNH comprendono due filamenti β antiparalleli e un'elica α . Domini HNH o simili a HNH comprendono un sito di legame a metallo (per esempio un sito di legame a catione divalente). Domini HNH o simili a HNH possono clivare un filamento di un acido nucleico bersaglio (per esempio il filamento complementare del filamento bersagliato di crRNA).

Domini RuvC o simili a RuvC comprendono un ripiegamento di RNasi H o simile a RNasi H. Domini RuvC/ RNasi H sono implicati in una diversa serie di funzioni basate su acido nucleico compresa l'azione sia su RNA che DNA. Il dominio RNasi H comprende 5 filamenti β circondati da una pluralità di eliche α . Domini RuvC/RNasi H o RuvC/RNasi H-simili comprendono un sito di legame a metallo (per esempio un sito di legame a catione divalente). Domini RuvC/RNasi

U001.D1.SM.61E

H o RuvC/Rnase H-simili possono clivare un filamento di un acido nucleico bersaglio (per esempio il filamento non complementare di un DNA bersaglio a doppio filamento).

Polipeptidi sito-diretti possono introdurre rotture a doppio filamento o rotture a singolo filamento in acidi nucleici, per esempio DNA genomico. La rottura a doppio filamento può stimolare pathway di riparazione di DNA endogeno di una cellula (per esempio riparazione dipendente da omologia o giunzione non omologa delle estremità o giunzione non omologa delle estremità alternativa (A-NHEJ) o giunzione delle estremità mediata da micro-omologia(MMEJ)). NHEJ può riparare acido nucleico bersaglio clivato senza la necessità di un template omologo. Questo può talvolta avere come risultato piccole delezioni o inserzioni (indel) nell'acido nucleico bersaglio in corrispondenza del sito di clivaggio, e può portare all'interruzione o alterazione di espressione genica. HDR può verificarsi quando un template di riparazione omologo, o donatore, è disponibile. Il template donatore omologo può comprendere sequenze che possono essere omologhe a sequenze fiancheggianti il sito di clivaggio di acido nucleico bersaglio. Il cromatidio sorella può essere utilizzato dalla cellula come template di riparazione. Tuttavia, ai fini dell'editing genomico, il template di riparazione può essere fornito come un acido nucleico esogeno, come un plasmide, oligonucleotide duplex, oligonucleotide a singolo filamento, oligonucleotide a doppio filamento, o acido nucleico virale. Con template donatori esogeni, una sequenza di acido nucleico addizionale (come un transgene) o una modificazione (come un cambiamento di singola base o di

U001.D1.SM.61E

molteplici basi o una delezione) possono venire introdotti tra le regioni fiancheggianti di omologia cosicché la sequenza di acido nucleico addizionale o alterata diventi anche incorporata nel locus bersaglio. MMEJ può portare a un esito genetico che è simile a NHEJ nel senso che piccole delezioni e inserzioni possono avere luogo a livello del sito di clonaggio. MMEJ può utilizzare sequenze omologhe di poche paia di basi fiancheggianti il sito di clivaggio per guidare un esito di riparazione di DNA di giunzione delle estremità preferito. In alcuni casi può essere possibile predire esiti di riparazione probabili sulla base dell'analisi di micro-omologie potenziali nelle regioni bersaglio di nucleasi.

Quindi, in alcuni casi, ricombinazione omologa può venire utilizzata per inserire una sequenza polinucleotidica esogena nel sito di clivaggio di acido nucleico bersaglio. Una sequenza polinucleotidica esogena è chiamata un polinucleotide donatore (o donatore o sequenza donatrice o template donatore polinucleotidico) nel presente documento. Il polinucleotide donatore, una porzione del polinucleotide donatore, una copia del polinucleotide donatore, o una porzione di una copia del polinucleotide donatore può venire inserito/a nel sito di clivaggio di acido nucleico bersaglio. Il polinucleotide donatore può essere una sequenza polinucleotidica esogena, vale a dire una sequenza che non si manifesta naturalmente a livello del sito di clivaggio di acido nucleico bersaglio.

Le modificazioni del DNA bersaglio dovute a NHEJ e/o HDR possono portare a, per esempio, mutazioni, delezioni, alterazioni, integrazioni, correzione genica, sostituzione genica, tagging genico, inserzione transgenica, delezione

U001.D1.SM.61E

nucleotidica, distruzione genica, traslocazioni e/o mutazione genica. I processi di delezione di DNA genomico e integrazione di acido nucleico non nativo in DNA genomico sono esempi di editing genomico.

Il polipeptide sito-diretto può comprendere una sequenza amminoacidica avente una identità di sequenza amminoacidica almeno del 10%, almeno del 15%, almeno del 20%, almeno del 30%, almeno del 40%, almeno del 50%, almeno del 60%, almeno del 70%, almeno del 75%, almeno dell'80%, almeno dell'85%, almeno del 90%, almeno del 95%, almeno del 99%, o del 100% rispetto a un polipeptide sito-diretto esemplare di tipo selvatico [per esempio Cas9 da *S. pyogenes*, US2014/0068797 Sequenza ID No. 8 o Sapranaukas et al., Nucleic Acids Res, 39(21): 9275-9282 (2011)], e vari altri polipeptidi sito-diretti. Il polipeptide sito-diretto può comprendere una identità almeno del 70, 75, 80, 85, 90, 95, 97, 99, o 100% a un polipeptide sito-diretto di tipo-selvatico (per esempio Cas9 da *S. pyogenes*, *supra*) su 10 amminoacidi contigui. Il polipeptide sito-diretto può comprendere una identità al massimo del 70, 75, 80, 85, 90, 95, 97, 99, o 100% a un polipeptide sito-diretto di tipo selvatico (per esempio Cas9 da *S. pyogenes*, *supra*) su 10 amminoacidi contigui. Il polipeptide sito-diretto può comprendere una identità almeno del 70, 75, 80, 85, 90, 95, 97, 99, o 100% a un polipeptide sito-diretto di tipo selvatico (per esempio Cas9 da *S. pyogenes*, *supra*) su 10 amminoacidi contigui in un dominio di nucleasi HNH del polipeptide sito-diretto. Il polipeptide sito-diretto può comprendere al massimo una identità del 70, 75, 80, 85, 90, 95, 97, 99, o 100% a un polipeptide sito-diretto di tipo selvatico (per esempio Cas9 da *S. pyogenes*, *supra*) su 10

U001.D1.SM.61E

amminoacidi contigui in un dominio nucleasico HNH del polipeptide sito-diretto. Il polipeptide sito-diretto può comprendere una identità almeno del 70, 75, 80, 85, 90, 95, 97, 99, o 100% a un polipeptide sito-diretto di tipo selvatico (per esempio Cas9 da *S. pyogenes, supra*) su 10 amminoacidi contigui in un dominio nucleasico RuvC del polipeptide sito-diretto. Il polipeptide sito-diretto può comprendere una identità al massimo del 70, 75, 80, 85, 90, 95, 97, 99, o 100% a un polipeptide sito-diretto di tipo selvatico (per esempio Cas9 da *S. pyogenes, supra*) su 10 amminoacidi contigui in un dominio nucleasico RuvC del polipeptide sito-diretto.

Il polipeptide sito-diretto può comprendere una forma modificata di un polipeptide sito-diretto esemplare di tipo selvatico. La forma modificata del polipeptide sito-diretto esemplare di tipo selvatico può comprendere una mutazione che riduce l'attività di clivaggio di acido nucleico del polipeptide sito-diretto. La forma modificata del polipeptide sito-diretto esemplare di tipo selvatico può avere meno del 90%, meno dell'80%, meno del 70%, meno del 60%, meno del 50%, meno del 40%, meno del 30%, meno del 20%, meno del 10%, meno del 5%, o meno dell'1% dell'attività di clivaggio di acido nucleico del polipeptide sito-diretto esemplare di tipo selvatico (per esempio Cas9 da *S. pyogenes, supra*). La forma modificata del polipeptide sito-diretto può non avere alcuna attività sostanziale di clivaggio di acido nucleico. Quando un polipeptide sito-diretto è una forma modificata che non ha alcuna attività sostanziale di clivaggio di acido nucleico, è indicato come "enzimaticamente inattivo".

La forma modificata del polipeptide sito-diretto può comprendere una

U001.D1.SM.61E

mutazione tale da poter indurre una rottura a singolo filamento (SSB) su un acido nucleico bersaglio (per esempio tagliando soltanto uno degli scheletri o backbone di zucchero-fosfato di un acido nucleico bersaglio a doppio filamento). La mutazione può portare a meno del 90%, meno del 80%, meno del 70%, meno del 60%, meno del 50%, meno del 40%, meno del 30%, meno del 20%, meno del 10%, meno del 5%, o meno dell'1% dell'attività di clivaggio di acido nucleico in uno o più della pluralità di domini di acido nucleico del polipeptide sito-diretto di tipo selvatico (per esempio Cas9 da *S. pyogenes*, *supra*). La mutazione può dare luogo a uno o più della pluralità di domini di clivaggio di acido nucleico che mantengono la capacità di clivare il filamento complementare dell'acido nucleico bersaglio, ma riducendo la sua capacità di clivare il filamento non complementare dell'acido nucleico bersaglio. La mutazione può dare luogo a uno o più della pluralità di domini di clivaggio di acido nucleico che mantengono la capacità di clivare il filamento non complementare dell'acido nucleico bersaglio, ma riducendo la sua capacità di clivare il filamento complementare dell'acido nucleico bersaglio. Per esempio, residui nel polipeptide Cas di *S. pyogenes* esemplare di tipo selvatico, come Asp10, His840, Asn854 e Asn856, sono mutati per inattivare uno o più della pluralità di domini di clivaggio di acido nucleico (per esempio domini nucleasici). I residui da mutare possono corrispondere a residui Asp10, His840, Asn854 e Asn856 nel polipeptide Cas di *S. pyogenes* esemplare di tipo selvatico (per esempio come determinato da allineamento di sequenze e/o strutturale). Esempi non limitativi di mutazioni includono D10A, H840A, N854A o N856A.

U001.D1.SM.61E

Un individuo esperto nel ramo riconoscerà che mutazioni differenti da sostituzioni di alanina possono essere adatte.

Una mutazione D10A può essere combinata con una o più mutazioni H840A, N854A, o N856A per produrre un polipeptide sito-diretto sostanzialmente privo di attività di clivaggio di DNA. Una mutazione H840A può essere combinata con una o più mutazioni D10A, N854A, o N856A per produrre un polipeptide sito-diretto sostanzialmente privo di attività di clivaggio di DNA. Una mutazione N854A può essere combinata con una o più mutazioni H840A, D10A, o N856A per produrre un polipeptide sito-diretto sostanzialmente privo di attività di clivaggio di DNA. Una mutazione N856A può essere combinata con una o più mutazioni H840A, N854A, o D10A per produrre un polipeptide sito-diretto sostanzialmente privo di attività di clivaggio di DNA. Polipeptidi sito-diretti che comprendono un dominio nucleasico sostanzialmente inattivo sono indicati come “nickasi”.

Varianti di nickasi di endonucleasi guidate da RNA, per esempio Cas9, possono essere utilizzate per aumentare la specificità di editing genomico mediato da CRISPR. Cas9 di tipo selvatico è tipicamente guidata da un RNA a singola guida progettato per ibridarsi con una sequenza di circa 20 nucleotidi specificata nella sequenza bersaglio (come un locus genomico endogeno). Tuttavia, alcuni appaiamenti errati o mismatch possono essere tollerati tra il RNA guida e il locus bersaglio, riducendo efficacemente la lunghezza di omologia richiesta nel sito bersaglio a, per esempio, appena 13 nt di omologia, e quindi portando a un elevato potenziale per legame e clivaggio di acido nucleico

U001.D1.SM.61E

a doppio filamento mediante il complesso CRISPR/Cas9 altrove nel genoma bersaglio – noto anche come clivaggio “fuori-bersaglio” o off-target. Poiché varianti di nickasi di Cas9 tagliano ciascuna soltanto un filamento, al fine di creare una rottura a doppio filamento è necessario per una coppia di nickasi legarsi in stretta prossimità e su filamenti opposti dell’acido nucleico bersaglio, creando quindi una coppia di nick, che è l’equivalente di una rottura a doppio filamento. Ciò richiede che due RNA guida separati – uno per ciascuna nickasi – debbano legarsi in stretta prossimità e su filamenti opposti dell’acido nucleico bersaglio. Questo requisito essenzialmente raddoppia la lunghezza minima di omologia necessaria perché avvenga la rottura a doppio filamento, riducendo quindi la probabilità che un evento di clivaggio a doppio filamento si verificherà altrove nel genoma, dove i due siti di RNA guida – se esistono – è improbabile che siano sufficientemente vicini tra loro per consentire che si formi la rottura a doppio filamento. Come descritto nell’arte, nickasi possono anche venire utilizzate per promuovere HDR contro NHEJ. HDR può essere utilizzata per introdurre cambiamenti selezionati in siti bersaglio nel genoma attraverso l’uso di specifiche sequenze donatrici che mediano efficacemente i cambiamenti desiderati.

Mutazioni contemplate possono includere sostituzioni, addizioni e delezioni, o qualsiasi combinazione di esse. La mutazione converte l’amminoacido mutato in alanina. La mutazione converte l’amminoacido mutato in un altro amminoacido (per esempio glicina, serina, treonina, cisteina, valina, leucina, isoleucina, metionina, prolina, fenilalanina, tirosina, triptofano, acido aspartico, acido

U001.D1.SM.61E

glutammico, asparagine, glutammina, istidina, lisina o arginina). La mutazione converte l'amminoacido mutato in un amminoacido non naturale (per esempio selenometionina). La mutazione converte l'amminoacido mutato in mimici di amminoacidi (per esempio fosfomimici). La mutazione può essere una mutazione conservativa. Per esempio, la mutazione converte l'amminoacido mutato in amminoacidi che assomigliano nella dimensione, forma, carica, polarità, conformazione e/o rotameri degli amminoacidi mutati (per esempio mutazione cisteina/serina, mutazione lisina/asparagina, mutazione istidina/fenilalanina). La mutazione può causare uno spostamento nella finestra di lettura e/o la creazione di un codone di stop prematuro. Mutazioni possono causare cambiamenti in regioni regolatorie di geni o loci che influiscono sull'espressione di uno o più geni.

Il polipeptide sito-diretto (per esempio polipeptide-diretto variante, mutato, enzimaticamente inattivo e/o condizionalmente enzimaticamente inattivo) può bersagliare acido nucleico. Il polipeptide sito-diretto (per esempio endoribonucleasi variante, mutata, enzimaticamente inattiva e/o condizionalmente enzimaticamente inattiva) può bersagliare DNA. Il polipeptide sito-diretto (per esempio endoribonucleasi variante, mutata, enzimaticamente inattiva e/o condizionalmente enzimaticamente inattiva) può bersagliare RNA.

Il polipeptide sito-diretto può comprendere una o più sequenze non native (per esempio il polipeptide sito-diretto è una proteina di fusione).

Il polipeptide sito-diretto può comprendere una sequenza amminoacidica

U001.D1.SM.61E

comprendente una identità amminoacidica almeno del 15% a una Cas9 da un batterio (per esempio *S. pyogenes*), una proteina di legame a acido nucleico, e due domini di clivaggio di acido nucleico (vale a dire un dominio HNH e un dominio RuvC).

Il polipeptide sito-diretto può comprendere una sequenza amminoacidica comprendente una identità amminoacidica almeno del 15% a una Cas9 da un batterio (per esempio *S. pyogenes*), e due domini di clivaggio di acido nucleico (vale a dire un dominio HNH e un dominio RuvC).

Il polipeptide sito-diretto può comprendere una sequenza amminoacidica comprendente una identità amminoacidica almeno del 15% a una Cas9 da un batterio (per esempio *S. pyogenes*), e due domini di clivaggio di acido nucleico, in cui uno dei o entrambi i domini di clivaggio di acido nucleico comprendono una identità amminoacidica almeno del 50% a un dominio nucleasico da Cas9 da un batterio (per esempio *S. pyogenes*).

Il polipeptide sito-diretto può comprendere una sequenza amminoacidica comprendente una identità amminoacidica almeno del 15% a una Cas9 da un batterio (per esempio *S. pyogenes*), due domini di clivaggio di acido nucleico (vale a dire un dominio HNH e un dominio RuvC), e una sequenza non nativa (per esempio un segnale di localizzazione nucleare) o un linker che collega il polipeptide sito-diretto a una sequenza non nativa.

Il polipeptide sito-diretto può comprendere una sequenza amminoacidica comprendente una identità amminoacidica almeno del 15% a una Cas9 da un batterio (per esempio *S. pyogenes*), due domini di clivaggio di acido nucleico

U001.D1.SM.61E

(vale a dire un dominio HNH e un dominio RuvC), in cui il polipeptide sito-diretto comprende una mutazione in uno dei domini di clivaggio di acido nucleico o entrambi che riduce l'attività di clivaggio dei domini nucleasici almeno del 50%.

Il polipeptide sito-diretto può comprendere una sequenza amminoacidica comprendente una identità amminoacidica almeno del 15% a una Cas9 da un batterio (per esempio *S. pyogenes*), e due domini di clivaggio di acido nucleico (vale a dire un dominio HNH e un dominio RuvC), in cui uno dei domini nucleasici comprende una mutazione di acido aspartico 10, e/o in cui uno dei domini nucleasici può comprendere una mutazione di istidina 840, e in cui la mutazione riduce l'attività di clivaggio del o dei domini almeno del 50%.

Gli uno o più polipeptidi sito-diretti, per esempio DNA endonucleasi, possono comprendere due nickasi che insieme effettuano una rottura a doppio filamento in un locus specifico nel genoma, o quattro nickasi che insieme effettuano o provocano due rotture a doppio filamento in loci specifici nel genoma. In alternativa, un polipeptide sito-diretto, per esempio DNA endonucleasi, può effettuare o provocare una rottura a doppio filamento in un locus specifico nel genoma.

Il polipeptide sito-diretto può essere fiancheggiato all'estremità N, all'estremità C, oppure sia all'estremità N che all'estremità C, da uno o più segnali di localizzazione nucleare (NLS). Per esempio, una endonucleasi Cas9 può essere fiancheggiata da due NLS, un NLS situato all'estremità N e il secondo NLS situato all'estremità C. Il NLS può essere qualsiasi NLS noto

U001.D1.SM.61E

nell'arte, come un SV40 NLS.

Acido nucleico di bersagliamento o targeting genomico

La presente descrizione fornisce un acido nucleico di bersagliamento (targeting) genomico che può dirigere le attività di un polipeptide associato (per esempio un polipeptide sito-diretto) a una specifica sequenza bersaglio all'interno di un acido nucleico bersaglio. L'acido nucleico di bersagliamento o targeting genomico può essere un RNA. Un RNA di bersagliamento o targeting genomico è indicato come "RNA guida" o "gRNA" nel presente documento. Un RNA guida può comprendere almeno una sequenza spaziatrice che si ibrida a una sequenza di acido nucleico bersaglio d'interesse, e una sequenza di ripetizione CRISPR. In sistemi di Tipo II, il gRNA comprende anche un secondo RNA chiamato la sequenza tracrRNA. Nel RNA guida (gRNA) di Tipo II, la sequenza di ripetizione CRISPR e la sequenza di tracrRNA si ibridano l'una all'altra per formare un duplex. Nel RNA guida (gRNA) di Tipo V il crRNA forma un duplex. In entrambi i sistemi il duplex può legarsi a un polipeptide sito-diretto, in modo tale che il RNA guida e il polipeptide sito-diretto formino un complesso. L'acido nucleico di bersagliamento o targeting genomico può fornire specificità di bersaglio al complesso in virtù della sua associazione con il polipeptide sito-diretto. L'acido nucleico di bersagliamento o targeting genomico può quindi dirigere l'attività del polipeptide sito-diretto.

RNA guida esemplari includono le sequenze spaziatrici in SEQ ID NOs: 1 – 71.947 e le sequenze di sgRNA in SEQ ID NOs: 71,950-71.959 della Lista di sequenze. Come compreso dalla persona di ordinaria esperienza nel ramo,

U001.D1.SM.61E

ciascun RNA guida può essere progettato per includere una sequenza spaziatrice complementare alla sua sequenza bersaglio genomica. Per esempio, ciascuna delle sequenze spaziatrici in SEQ ID NOs: 1 – 71.947 della Lista di sequenza può essere collocata in una singola chimera di RNA o un crRNA (insieme a un tracrRNA corrispondente). Si vedano Jinek et al., *Science*, 337, 816-821 (2012) e Deltcheva et al., *Nature*, 471, 602-607 (2011) o la Tabella 1.

L'acido nucleico di bersagliamento o targeting genomico può essere un RNA guida a doppia molecola. L'acido nucleico di bersagliamento o targeting genomico può essere un RNA guida monomolecolare.

Un RNA guida a doppia molecola può comprendere due filamenti di RNA. Il primo filamento comprende nella direzione da 5' a 3' una sequenza di estensione spaziatrice opzionale, una sequenza spaziatrice e una sequenza di ripetizione CRISPR minima. Il secondo filamento può comprendere una sequenza di tracrRNA minima (complementare alla sequenza di ripetizione CRISPR minima), una sequenza di 3' tracrRNA e una sequenza di estensione di tracrRNA opzionale.

Un RNA guida monomolecolare (sgRNA) in un sistema di Tipo II può comprendere, nella direzione da 5' a 3', una sequenza di estensione spaziatrice opzionale, una sequenza spaziatrice, una sequenza di ripetizione CRISPR minima, un linker guida a singola molecola, una sequenza di tracrRNA minima, una sequenza di 3' tracrRNA e una sequenza di estensione di tracrRNA opzionale. L'estensione di tracrRNA opzionale può comprendere elementi che contribuiscono a una funzionalità supplementare (per esempio stabilità) al RNA

U001.D1.SM.61E

guida. Il linker guida a singola molecola può collegare la ripetizione CRISPR minima e la sequenza di tracrRNA minima per formare una struttura a forcina. L'estensione di tracrRNA opzionale può comprendere una o più forcine.

Un RNA guida monomolecolare (sgRNA) in un sistema di Tipo V può comprendere, nella direzione da 5' a 3', una sequenza di ripetizione CRISPR minima e una sequenza spaziatrice.

Il sgRNA può comprendere una sequenza spaziatrice di 20 nucleotidi all'estremità 5' della sequenza di sgRNA. Il sgRNA può comprendere una sequenza spaziatrice di meno di 20 nucleotidi all'estremità 5' della sequenza di sgRNA. Il sgRNA può comprendere una sequenza spaziatrice di più di 20 nucleotidi all'estremità 5' della sequenza di sgRNA. Il sgRNA può comprendere una sequenza spaziatrice di lunghezza variabile con 17-30 nucleotidi all'estremità 5' della sequenza di sgRNA (si veda la Tabella 1).

Il sgRNA può comprendere nessun uracile all'estremità 3' della sequenza di sgRNA, come in SEQ ID NO: 71,961 della Tabella 1. Il sgRNA può comprendere uno o più uracili all'estremità 3' della sequenza di sgRNA, come in SEQ ID NO: 71.962 nella Tabella 1. Per esempio, il sgRNA può comprendere 1 uracile (U) all'estremità 3' della sequenza di sgRNA. Il sgRNA può comprendere 2 uracili (UU) all'estremità 3' della sequenza di sgRNA. Il sgRNA può comprendere 3 uracili (UUU) all'estremità 3' della sequenza di sgRNA. Il sgRNA può comprendere 4 uracili (UUUU) all'estremità 3' della sequenza di sgRNA. Il sgRNA può comprendere 5 uracili (UUUUU) all'estremità 3' della sequenza di sgRNA. Il sgRNA può comprendere 6 uracili (UUUUUU)

U001.D1.SM.61E

più facilmente. Vari tipi di modificazioni di RNA possono essere introdotti durante o dopo sintesi chimica e/o generazione enzimatica di RNA, per esempio modificazioni che potenziano la stabilità, riducono la probabilità o il grado di risposta immunitaria innata, e/o potenziano altri attributi, come descritto nell'arte.

Sequenza di estensione spaziatrice

In alcuni esempi di acidi nucleici di bersagliamento o targeting, una sequenza di estensione spaziatrice può modificare l'attività, fornire stabilità e/o fornire una locazione per modificazioni di un acido nucleico di bersagliamento o targeting genomico. Una sequenza di estensione spaziatrice può modificare l'attività o specificità on-target o off-target. In alcuni esempi una sequenza di estensione spaziatrice può venire fornita. La sequenza di estensione spaziatrice può avere una lunghezza superiore a 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 1000, 2000, 3000, 4000, 5000, 6000, o 7000 o più nucleotidi. La sequenza di estensione spaziatrice può avere una lunghezza inferiore a 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 1000, 2000, 3000, 4000, 5000, 6000, 7000 o più nucleotidi. La sequenza di estensione spaziatrice può avere una lunghezza inferiore a 10 nucleotidi. La sequenza di estensione spaziatrice può avere una lunghezza compresa tra 10 e 30 nucleotidi. La sequenza di estensione spaziatrice può avere una lunghezza compresa tra 30 e 70 nucleotidi.

La sequenza di estensione spaziatrice può comprendere un'altra frazione (per

U001.D1.SM.61E

esempio una sequenza di controllo di stabilità, una sequenza di legame a endoribonucleasi, un ribozima). La frazione può diminuire o aumentare la stabilità di un acido nucleico di bersagliamento o targeting di acido nucleico. La frazione può essere un segmento terminatore di trascrizione (vale a dire una sequenza di terminazione di trascrizione). La frazione può agire in una cellula eucariotica. La frazione può agire in una cellula procariotica. La frazione può agire sia in cellule eucariotiche che procariotiche. Esempi non limitativi di frazioni adatte includono: un 5' cap (per esempio un cap di 7-metilguanilato (m7 G)), una sequenza riboswitch (per esempio per permettere una stabilità regolata e/o accessibilità regolata da parte di proteine e complessi proteici), una sequenza che forma un duplex dsRNA (vale a dire una forcina), una sequenza che bersaglia il RNA in una locazione subcellulare (per esempio nucleo, mitocondri, cloroplasti e simili), una modificazione o sequenza che fornisce il tracciamento (per esempio coniugazione diretta a una molecola fluorescente, coniugazione a una frazione che facilita il rilevamento fluorescente, una sequenza che consente il rilevamento fluorescente, etc.) e/o una modificazione o sequenza che fornisce un sito di legame per proteine (per esempio proteine che agiscono su DNA, inclusi attivatori di trascrizione, controlli di trascrizione, DNA metiltransferasi, DNA demetilasi, istone acetiltransferasi, istone deacetilasi, e simili).

Sequenza spaziatrice

La sequenza spaziatrice si ibrida a una sequenza in un acido nucleico bersaglio d'interesse. Lo spaziatore di un acido nucleico di bersagliamento o targeting genomico può interagire con un acido nucleico bersaglio in una maniera

U001.D1.SM.61E

sequenza-specifica tramite ibridazione (vale a dire appaiamento di basi). La sequenza nucleotidica dello spaziatore può variare a seconda della sequenza dell'acido nucleico bersaglio di interesse.

In un sistema CRISPR/Cas nel presente documento la sequenza spaziatrice può essere progettata per ibridarsi a un acido nucleico bersaglio che è situato in 5' di un PAM dell'enzima Cas9 utilizzato nel sistema. Lo spaziatore può perfettamente corrispondere alla sequenza bersaglio o può avere appaiamenti errati o mismatch. Ciascun enzima Cas9 ha una particolare sequenza di PAM che esso riconosce in un DNA bersaglio. Per esempio, *S. pyogenes* riconosce in un acido nucleico bersaglio un PAM che comprende la sequenza 5'-NRG-3', dove R comprende A o G, dove N è qualsiasi nucleotide e N è immediatamente 3' della sequenza bersaglio di acido nucleico bersagliata dalla sequenza spaziatrice.

La sequenza di acido nucleico bersaglio può comprendere 20 nucleotidi. L'acido nucleico bersaglio può comprendere meno di 20 nucleotidi. L'acido nucleico bersaglio può comprendere più di 20 nucleotidi. L'acido nucleico bersaglio può comprendere almeno: 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 o più nucleotidi. L'acido nucleico bersaglio può comprendere al massimo: 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 o più nucleotidi. La sequenza di acido nucleico bersaglio può comprendere 20 basi immediatamente 5' del primo nucleotide del PAM. Per esempio, in una sequenza comprendente 5'-NNNNNNNNNNNNNNNNNNNNNNNNNNNNNRG-3' (SEQ ID NO: 71,948), l'acido nucleico bersaglio può comprendere la sequenza che corrisponde ai N, dove N è

U001.D1.SM.61E

qualsiasi nucleotide, e la sequenza NRG sottolineata è PAM di *S. pyogenes*.

La sequenza spaziatrice che si ibrida all'acido nucleico bersaglio può avere una lunghezza di almeno circa 6 nucleotidi (nt). La sequenza spaziatrice può essere almeno circa 6 nt, almeno circa 10 nt, almeno circa 15 nt, almeno circa 18 nt, almeno circa 19 nt, almeno circa 20 nt, almeno circa 25 nt, almeno circa 30 nt, almeno circa 35 nt o almeno circa 40 nt, da circa 6 nt a circa 80 nt, da circa 6 nt a circa 50 nt, da circa 6 nt a circa 45 nt, da circa 6 nt a circa 40 nt, da circa 6 nt a circa 35 nt, da circa 6 nt a circa 30 nt, da circa 6 nt a circa 25 nt, da circa 6 nt a circa 20 nt, da circa 6 nt a circa 19 nt, da circa 10 nt a circa 50 nt, da circa 10 nt a circa 45 nt, da circa 10 nt a circa 40 nt, da circa 10 nt a circa 35 nt, da circa 10 nt a circa 30 nt, da circa 10 nt a circa 25 nt, da circa 10 nt a circa 20 nt, da circa 10 nt a circa 19 nt, da circa 19 nt a circa 25 nt, da circa 19 nt a circa 30 nt, da circa 19 nt a circa 35 nt, da circa 19 nt a circa 40 nt, da circa 19 nt a circa 45 nt, da circa 19 nt a circa 50 nt, da circa 19 nt a circa 60 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, o da circa 20 nt a circa 60 nt. In alcuni esempi, la sequenza spaziatrice può comprendere 20 nucleotidi.

In alcuni esempi, lo spaziatore può comprendere 19 nucleotidi.

In alcuni esempi la complementarietà in percentuale tra la sequenza spaziatrice e l'acido nucleico bersaglio è almeno circa del 30%, almeno circa del 40%, almeno circa del 50%, almeno circa del 60%, almeno circa del 65%, almeno circa del 70%, almeno circa del 75%, almeno circa dell'80%, almeno circa

U001.D1.SM.61E

dell'85%, almeno circa del 90%, almeno circa del 95%, almeno circa del 97%, almeno circa del 98%, almeno circa del 99%, o del 100%. In alcuni esempi, la complementarità in percentuale tra la sequenza spaziatrice e l'acido nucleico bersaglio è al massimo circa del 30%, al massimo circa del 40%, al massimo circa del 50%, al massimo circa del 60%, al massimo circa del 65%, al massimo circa del 70%, al massimo circa del 75%, al massimo circa dell'80%, al massimo circa dell'85%, al massimo circa del 90%, al massimo circa del 95%, al massimo circa del 97%, al massimo circa del 98%, al massimo circa del 99%, o del 100%. In alcuni esempi, la complementarità in percentuale tra la sequenza spaziatrice e l'acido nucleico bersaglio è del 100% sui sei nucleotidi contigui a 5' della sequenza bersaglio del filamento complementare dell'acido nucleico bersaglio. La complementarità in percentuale tra la sequenza spaziatrice e l'acido nucleico bersaglio può essere almeno del 60% su circa 20 nucleotidi contigui. La lunghezza della sequenza spaziatrice e la lunghezza dell'acido nucleico bersaglio possono differire da 1 a 6 nucleotidi, il che può essere pensato come un rigonfiamento o più rigonfiamenti (bulge).

La sequenza spaziatrice può essere progettata o scelta utilizzando un programma informatico. Il programma informatico può utilizzare variabili, come temperatura di fusione prevista, formazione di strutture secondarie, temperatura di ricottura o annealing prevista, identità di sequenza, contesto genomico, accessibilità di cromatina, % GC, frequenza di occorrenza genomica (per esempio di sequenze che sono identiche o sono simili ma variano in uno o più punti come risultato di un appaiamento errato o mismatch, inserimento o

U001.D1.SM.61E

delezione), stato di metilazione, presenza di SNP, e simili.

Sequenza di ripetizione CRISPR minima

Una sequenza di ripetizione CRISPR minima può essere una sequenza con una identità di sequenza almeno circa del 30%, circa del 40%, circa del 50%, circa del 60%, circa del 65%, circa del 70%, circa del 75%, circa dell'80%, circa dell'85%, circa del 90%, circa del 95%, o del 100% a una sequenza di ripetizione CRISPR di riferimento (per esempio crRNA da *S. pyogenes*).

Una sequenza di ripetizione CRISPR minima può comprendere nucleotidi che possono ibridarsi a una sequenza di tracrRNA minima in una cellula. La sequenza di ripetizione CRISPR minima e una sequenza di tracrRNA minima possono formare un duplex, vale a dire una struttura a doppio filamento a basi appaiate. Insieme, la sequenza di ripetizione CRISPR minima e la sequenza di tracrRNA minima possono legarsi al polipeptide sito-diretto. Almeno una parte della sequenza di ripetizione CRISPR minima può ibridarsi alla sequenza di tracrRNA minima. Almeno una parte della sequenza di ripetizione CRISPR minima può essere almeno circa al 30%, circa al 40%, circa al 50%, circa al 60%, circa al 65%, circa al 70%, circa al 75%, circa all'80%, circa all'85%, circa al 90%, circa al 95%, o al 100% complementare alla sequenza di tracrRNA minima. Almeno una parte della sequenza di ripetizione CRISPR minima può essere al massimo circa al 30%, circa al 40%, circa al 50%, circa al 60%, circa al 65%, circa al 70%, circa al 75%, circa all'80%, circa all'85%, circa al 90%, circa al 95%, o al 100% complementare alla sequenza di tracrRNA minima.

La sequenza di ripetizione CRISPR minima può avere una lunghezza da circa 7

U001.D1.SM.61E

nucleotidi a circa 100 nucleotidi. Per esempio, la lunghezza della sequenza di ripetizione CRISPR minima è da circa 7 nucleotidi (nt) a circa 50 nt, da circa 7 nt a circa 40 nt, da circa 7 nt a circa 30 nt, da circa 7 nt a circa 25 nt, da circa 7 nt a circa 20 nt, da circa 7 nt a circa 15 nt, da circa 8 nt a circa 40 nt, da circa 8 nt a circa 30 nt, da circa 8 nt a circa 25 nt, da circa 8 nt a circa 20 nt, da circa 8 nt a circa 15 nt, da circa 15 nt a circa 100 nt, da circa 15 nt a circa 80 nt, da circa 15 nt a circa 50 nt, da circa 15 nt a circa 40 nt, da circa 15 nt a circa 30 nt, o da circa 15 nt a circa 25 nt. In alcuni esempi, la sequenza di ripetizione CRISPR minima può avere una lunghezza di approssimativamente 9 nucleotidi. La sequenza di ripetizione CRISPR minima può avere una lunghezza di approssimativamente 12 nucleotidi.

La sequenza di ripetizione CRISPR minima può essere almeno circa al 60% identica a sequenza di ripetizione CRISPR minima di riferimento (per esempio crRNA di tipo selvatico da *S. pyogenes*) su una striscia di almeno 6, 7, o 8 nucleotidi contigui.

Per esempio, la sequenza di ripetizione CRISPR minima può essere identica almeno circa al 65%, identica almeno circa al 70%, identica almeno circa al 75%, identica almeno circa all'80%, identica almeno circa all'85%, identica almeno circa al 90%, identica almeno circa al 95%, identica almeno circa al 98%, identica almeno circa al 99%, o identica al 100% a una sequenza di ripetizione CRISPR minima di riferimento su una striscia di almeno 6, 7, o 8 nucleotidi contigui.

Sequenza di tracrRNA minima

U001.D1.SM.61E

Una sequenza di tracrRNA minima può essere una sequenza con una identità di sequenza almeno circa del 30%, circa del 40%, circa del 50%, circa del 60%, circa del 65%, circa del 70%, circa del 75%, circa dell'80%, circa dell'85%, circa del 90%, circa del 95%, o del 100% a una sequenza di tracrRNA di riferimento (per esempio tracrRNA di tipo selvatico da *S. pyogenes*).

Una sequenza di tracrRNA minima può comprendere nucleotidi che si ibridano a una sequenza di ripetizione CRISPR minima in una cellula. Una sequenza di tracrRNA minima e una sequenza di ripetizione CRISPR minima formano un duplex, vale a dire una struttura a doppio filamento a basi appaiate. Insieme, la sequenza di tracrRNA minima e la ripetizione CRISPR minima possono legarsi a un polipeptide sito-diretto. Almeno una parte della sequenza di tracrRNA minima può ibridarsi alla sequenza di ripetizione CRISPR minima. La sequenza di tracrRNA minima può essere almeno circa al 30%, circa al 40%, circa al 50%, circa al 60%, circa al 65%, circa al 70%, circa al 75%, circa all'80%, circa all'85%, circa al 90%, circa al 95%, o al 100% complementare alla sequenza di ripetizione CRISPR minima.

La sequenza di tracrRNA minima può avere una lunghezza da circa 7 nucleotidi a circa 100 nucleotidi. Per esempio, la sequenza di tracrRNA minima può avere una lunghezza da circa 7 nucleotidi (nt) a circa 50 nt, da circa 7 nt a circa 40 nt, da circa 7 nt a circa 30 nt, da circa 7 nt a circa 25 nt, da circa 7 nt a circa 20 nt, da circa 7 nt a circa 15 nt, da circa 8 nt a circa 40 nt, da circa 8 nt a circa 30 nt, da circa 8 nt a circa 25 nt, da circa 8 nt a circa 20 nt, da circa 8 nt a circa 15 nt, da circa 15 nt a circa 100 nt, da circa 15 nt a circa 80 nt, da circa 15

U001.D1.SM.61E

nt a circa 50 nt, da circa 15 nt a circa 40 nt, da circa 15 nt a circa 30 nt o da circa 15 nt a circa 25 nt. La sequenza di tracrRNA minima può avere una lunghezza di approssimativamente 9 nucleotidi. La sequenza di tracrRNA minima può avere una lunghezza di approssimativamente 12 nucleotidi. Il tracrRNA minimo è costituito da tracrRNA nt 23-48 descritto in Jinek *et al.*, *supra*.

La sequenza di tracrRNA minima può essere identica almeno circa al 60% a una sequenza di tracrRNA minima di riferimento (per esempio tracrRNA di tipo selvatico da *S. pyogenes*) su una striscia di almeno 6, 7, o 8 nucleotidi contigui. Per esempio, la sequenza di tracrRNA minima può essere identica almeno circa al 65%, identica circa al 70%, identica circa al 75%, identica circa all'80%, identica circa all'85%, identica circa al 90% identica circa al 95%, identica circa al 98%, identica circa al 99%, o identica al 100% a una sequenza di tracrRNA minima di riferimento su una striscia di almeno 6, 7, o 8 nucleotidi contigui.

Il duplex tra il RNA di CRISPR minimo e il tracrRNA minimo può comprendere una doppia elica. Il duplex tra il RNA di CRISPR minimo e il tracrRNA minimo può comprendere almeno circa 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o più nucleotidi. Il duplex tra il RNA di CRISPR minimo e il tracrRNA minimo può comprendere al massimo circa 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o più nucleotidi.

Il duplex può comprendere un appaiamento errato o mismatch (vale a dire i due filamenti del duplex non sono complementari al 100%). Il duplex può comprendere almeno circa 1, 2, 3, 4, o 5 appaiamenti errati o mismatch. Il duplex può comprendere al massimo circa 1, 2, 3, 4, o 5 appaiamenti errati o mismatch. Il duplex può comprendere non più di 2 appaiamenti errati o

U001.D1.SM.61E

mismatch.

Rigonfiamenti o bulge

In alcuni casi ci può essere un "rigonfiamento o bulge" nel duplex tra il RNA di CRISPR minimo e il tracrRNA minimo. Un rigonfiamento o bulge è una regione spaiata di nucleotidi all'interno del duplex. Un rigonfiamento o bulge può contribuire al legame del duplex al polipeptide sito-diretto. Il rigonfiamento o bulge può comprendere, su un lato del duplex, un 5'-XXX-3' spaiato dove X è qualsiasi purina e Y comprende un nucleotide che può formare una coppia tentennante con un nucleotide sul filamento opposto, e una regione nucleotidica spaiata sull'altro lato del duplex. Il numero di nucleotidi spaiati sui due lati del duplex può essere differente.

In un esempio, il rigonfiamento o bulge può comprendere una purina spaiata (per esempio adenina) sul filamento di ripetizione CRISPR minima del rigonfiamento o bulge. In alcuni esempi, il rigonfiamento o bulge può comprendere un 5'-AAGY-3' spaiato del filamento di sequenza di tracrRNA minima del rigonfiamento o bulge, dove Y comprende un nucleotide che può formare un appaiamento tentennante con un nucleotide sul filamento di ripetizione CRISPR minima.

Un rigonfiamento o bulge sul lato di ripetizione CRISPR minima del duplex può comprendere almeno 1, 2, 3, 4, o 5 o più nucleotidi spaiati. Un rigonfiamento o bulge sul lato di ripetizione CRISPR minima del duplex può comprendere al massimo 1, 2, 3, 4, o 5 o più nucleotidi spaiati. Un rigonfiamento o bulge sul lato di ripetizione CRISPR minima del duplex può

U001.D1.SM.61E

comprendere un nucleotide spaiato.

Un rigonfiamento o bulge sul lato di sequenza di tracrRNA minima del duplex può comprendere almeno 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o più nucleotidi spaiati. Un rigonfiamento o bulge sul lato di sequenza di tracrRNA minima del duplex può comprendere al massimo 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o più nucleotidi spaiati. Un rigonfiamento o bulge su un secondo lato del duplex (per esempio il lato di sequenza di tracrRNA minima del duplex) può comprendere 4 nucleotidi spaiati.

Un rigonfiamento o bulge può comprendere almeno un appaiamento tentennante. In alcuni esempi, a rigonfiamento o bulge può comprendere al massimo un appaiamento tentennante. Un rigonfiamento o bulge può comprendere almeno un nucleotide purinico. Un rigonfiamento o bulge può comprendere almeno 3 nucleotidi purinici. Una sequenza di rigonfiamento o bulge può comprendere almeno 5 nucleotidi purinici. Una sequenza di rigonfiamento o bulge può comprendere almeno un nucleotide guaninico. In alcuni esempi, una sequenza di rigonfiamento o bulge può comprendere almeno un nucleotide adeninico.

Forcine o hairpin

In vari esempi, una o più forcine può essere situata 3' al tracrRNA minimo nella sequenza 3' tracrRNA.

La forcina può iniziare almeno circa 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, o 20 o più nucleotidi 3' dall'ultimo nucleotide appaiato nel duplex di sequenza di ripetizione CRISPR minima e sequenza di tracrRNA minima. La forcina può iniziare al massimo circa 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o più nucleotidi 3'

U001.D1.SM.61E

dell'ultimo nucleotide appaiato nel duplex di sequenza di ripetizione CRISPR minima e sequenza di tracrRNA minima.

La forcina può comprendere almeno circa 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, o 20 o più nucleotidi consecutivi. La forcina può comprendere al massimo circa 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, o più nucleotidi consecutivi.

La forcina può comprendere un dinucleotide CC (vale a dire due nucleotidi citosinici consecutivi).

La forcina può comprendere nucleotidi in duplex (per esempio nucleotidi in una forcina, ibridati insieme). Per esempio, una forcina può comprendere un dinucleotide CC che è ibridato a un dinucleotide GG in un duplex a forcina della sequenza 3' tracrRNA.

Una o più delle forcine possono interagire con regioni interagenti con RNA di un polipeptide sito-diretto.

In alcuni esempi ci sono due o più forcine, e in altri esempi ci sono tre o più forcine.

Sequenza 3' tracrRNA

Una sequenza 3' tracrRNA può comprendere una sequenza con una identità di sequenza almeno circa del 30%, circa del 40%, circa del 50%, circa del 60%, circa del 65%, circa del 70%, circa del 75%, circa dell'80%, circa dell'85%, circa del 90%, circa del 95%, o del 100% a una sequenza di tracrRNA di riferimento (per esempio un tracrRNA da *S. pyogenes*).

La sequenza 3' tracrRNA può avere una lunghezza da circa 6 nucleotidi a circa 100 nucleotidi. Per esempio,

U001.D1.SM.61E

la sequenza 3' tracrRNA può avere una lunghezza da circa 6 nucleotidi (nt) a circa 50 nt, da circa 6 nt a circa 40 nt, da circa 6 nt a circa 30 nt, da circa 6 nt a circa 25 nt, da circa 6 nt a circa 20 nt, da circa 6 nt a circa 15 nt, da circa 8 nt a circa 40 nt, da circa 8 nt a circa 30 nt, da circa 8 nt a circa 25 nt, da circa 8 nt a circa 20 nt, da circa 8 nt a circa 15 nt, da circa 15 nt a circa 100 nt, da circa 15 nt a circa 80 nt, da circa 15 nt a circa 50 nt, da circa 15 nt a circa 40 nt, da circa 15 nt a circa 30 nt, o da circa 15 nt a circa 25 nt. La sequenza 3' tracrRNA può avere una lunghezza di approssimativamente 14 nucleotidi.

La sequenza 3' tracrRNA può essere identica almeno circa al 60% a una sequenza 3' tracrRNA di riferimento (per esempio sequenza 3' tracrRNA di tipo selvatico da *S. pyogenes*) su una striscia di almeno 6, 7, o 8 nucleotidi contigui. Per esempio, la sequenza 3' tracrRNA può essere identica almeno circa al 60%, identica circa al 65%, identica circa al 70%, identica circa al 75%, identica circa all'80%, identica circa all'85%, identica circa al 90%, identica circa al 95%, identica circa al 98%, identica circa al 99%, o identica al 100% a una sequenza 3' tracrRNA di riferimento (per esempio sequenza 3' tracrRNA di tipo selvatico da *S. pyogenes*) su una striscia di almeno 6, 7, o 8 nucleotidi contigui.

La sequenza 3' tracrRNA può comprendere più di una regione in duplex (per esempio forcina, regione ibridata). La sequenza 3' tracrRNA può comprendere due regioni in duplex.

La sequenza 3' tracrRNA può comprendere una struttura stem-loop. La struttura stem-loop nel 3' tracrRNA può comprendere almeno 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o 20 o più nucleotidi. La struttura stem-loop nel 3' tracrRNA può

U001.D1.SM.61E

comprendere al massimo 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o più nucleotidi. La struttura stem-loop può comprendere una frazione funzionale. Per esempio, la struttura stem-loop può comprendere un aptamero, un ribozima, una forcina di interazione proteica, un array CRISPR, un introne, o un esone. La struttura stem-loop può comprendere almeno circa 1, 2, 3, 4, or 5 o più frazioni funzionali. La struttura stem-loop può comprendere al massimo circa 1, 2, 3, 4, o 5 o più frazioni funzionali.

La forcina nella sequenza 3' tracrRNA può comprendere un dominio P. In alcuni esempi, il dominio P può comprendere una regione a doppio filamento nella forcina.

Sequenza di estensione di tracrRNA

Una sequenza di estensione di tracrRNA può venire fornita sia che il tracrRNA sia nel contesto di guide a singola molecola o guide a doppia molecola. La sequenza di estensione di tracrRNA può avere una lunghezza da circa 1 nucleotide a circa 400 nucleotidi. La sequenza di estensione di tracrRNA può avere una lunghezza superiore a 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, o 400 nucleotidi. La sequenza di estensione di tracrRNA può avere una lunghezza da circa 20 a circa 5000 o più nucleotidi. La sequenza di estensione di tracrRNA può avere una lunghezza superiore a 1000 nucleotidi. La sequenza di estensione di tracrRNA può avere una lunghezza inferiore a 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400 o più nucleotidi. La sequenza di estensione di

U001.D1.SM.61E

tracrRNA può avere una lunghezza inferiore a 1000 nucleotidi. La sequenza di estensione di tracrRNA può avere una lunghezza inferiore a 10 nucleotidi. La sequenza di estensione di tracrRNA può avere una lunghezza da 10 a 30 nucleotidi. La sequenza di estensione di tracrRNA può avere una lunghezza da 30 a 70 nucleotidi.

La sequenza di estensione di tracrRNA può comprendere una frazione funzionale (per esempio una sequenza di controllo di stabilità, un ribozima, una sequenza di legame a endoribonucleasi). La frazione funzionale può comprendere un segmento terminatore di trascrizione (vale a dire una sequenza di terminazione di trascrizione). La frazione funzionale può avere una lunghezza totale da circa 10 nucleotidi (nt) a circa 100 nucleotidi, da circa 10 nt a circa 20 nt, da circa 20 nt a circa 30 nt, da circa 30 nt a circa 40 nt, da circa 40 nt a circa 50 nt, da circa 50 nt a circa 60 nt, da circa 60 nt a circa 70 nt, da circa 70 nt a circa 80 nt, da circa 80 nt a circa 90 nt, o da circa 90 nt a circa 100 nt, da circa 15 nt a circa 80 nt, da circa 15 nt a circa 50 nt, da circa 15 nt a circa 40 nt, da circa 15 nt a circa 30 nt, o da circa 15 nt a circa 25 nt. La frazione funzionale può funzionare in una cellula eucariotica. La frazione funzionale può funzionare in una cellula procariotica. La frazione funzionale può funzionare sia in cellule eucariotiche che procariotiche.

Esempi non limitativi di frazioni funzionali di estensione di tracrRNA adatte includono una coda 3' poliadenilata, una sequenza riboswitch (per esempio per permettere una stabilità regolata e/o accessibilità regolata da parte di proteine e complessi proteici), una sequenza che forma un duplex dsRNA (vale a dire una

U001.D1.SM.61E

forcina), una sequenza che bersaglia il RNA in una locazione subcellulare (per esempio nucleo, mitocondri, cloroplasti e simili), una modificazione o sequenza che fornisce il tracciamento (per esempio coniugazione diretta a una molecola fluorescente, coniugazione a una frazione che facilita il rilevamento fluorescente, una sequenza che consente il rilevamento fluorescente, etc.) e/o una modificazione o sequenza che fornisce un sito di legame per proteine (per esempio proteine che agiscono su DNA, inclusi attivatori di trascrizione, controlli di trascrizione, DNA metiltransferasi, DNA demetilasi, istone acetiltransferasi, istone deacetilasi, e simili). La sequenza di estensione di tracrRNA può comprendere un sito di legame a primer o un indice molecolare (per esempio sequenza a codice a barre). La sequenza di estensione di tracrRNA può comprendere uno o più tag di affinità.

Sequenza Linker di guida a singola molecola

La sequenza linker di un acido nucleico guida a singola molecola può avere una lunghezza da circa 3 nucleotidi a circa 100 nucleotidi. In Jinek *et al.*, *supra*, per esempio, è stato utilizzato un semplice “tetraloop” a 4 nucleotidi (-GAAA-), *Science*, 337(6096):816-821 (2012). Un linker illustrativo ha una lunghezza da circa 3 nucleotidi (nt) a circa 90 nt, da circa 3 nt a circa 80 nt, da circa 3 nt a circa 70 nt, da circa 3 nt a circa 60 nt, da circa 3 nt a circa 50 nt, da circa 3 nt a circa 40 nt, da circa 3 nt a circa 30 nt, da circa 3 nt a circa 20 nt, da circa 3 nt a circa 10 nt. Per esempio, il linker può avere una lunghezza da circa 3 nt a circa 5 nt, da circa 5 nt a circa 10 nt, da circa 10 nt a circa 15 nt, da circa 15 nt a circa 20 nt, da circa 20 nt a circa 25 nt, da circa 25 nt a circa 30 nt, da circa 30 nt a

U001.D1.SM.61E

circa 35 nt, da circa 35 nt a circa 40 nt, da circa 40 nt a circa 50 nt, da circa 50 nt a circa 60 nt, da circa 60 nt a circa 70 nt, da circa 70 nt a circa 80 nt, da circa 80 nt a circa 90 nt, o da circa 90 nt a circa 100 nt. Il linker di un acido nucleico guida a singola molecola può essere tra 4 e 40 nucleotidi. Il linker può essere almeno circa 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, o 7000 o più nucleotidi. Il linker può essere al massimo circa 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, o 7000 o più nucleotidi.

Linker possono comprendere qualsiasi di una varietà di sequenze, sebbene in alcuni esempi il linker non comprenderà sequenze che hanno regioni estensive di omologia con altre porzioni del RNA guida, che potrebbero causare legame intramolecolare che potrebbe interferire con altre regioni funzionali della guida. In Jinek *et al.*, *supra* è stata utilizzata una semplice sequenza a 4 nucleotidi - GAAA-, Science, 337(6096):816-821 (2012), ma numerose altre sequenze, incluse sequenze più lunghe, possono analogamente essere utilizzate.

La sequenza linker può comprendere una frazione funzionale. Per esempio, la sequenza linker può comprendere una o più caratteristiche, incluso un aptamero, un ribozima, una forcina interagente con proteina, un sito di legame a proteina, un array CRISPR, un introne, o un esone. La sequenza linker può comprendere almeno circa 1, 2, 3, 4, o 5 o più frazioni funzionali. In alcuni esempi la sequenza linker può comprendere al massimo circa 1, 2, 3, 4, o 5 o più frazioni funzionali.

Una fase dei metodi *ex vivo* della presente descrizione può comprendere

U001.D1.SM.61E

l'editing delle cellule iPSC paziente-specifiche utilizzando ingegneria genomica. In alternativa, una fase dei metodi *ex vivo* della presente descrizione può comprendere l'editing di cellule staminali mesenchimali, o cellule progenitrici ematopoietiche. Analogamente, una fase dei metodi *in vivo* della presente descrizione può comprendere l'editing delle cellule in un paziente affetto da una emoglobinopatia utilizzando ingegneria genomica. Similmente, una fase nei metodi cellulari della presente descrizione può comprendere l'editing all'interno o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A in una cellula umana mediante ingegneria genomica.

Differenti pazienti affetti da emoglobinopatia generalmente richiederanno strategie differenti di delezione, modulazione o inattivazione. Qualsiasi endonucleasi CRISPR può venire utilizzata nei metodi della presente descrizione, ciascuna endonucleasi CRISPR avendo il suo proprio PAM associato, che può essere o meno specifico per malattia. Per esempio, sequenze spaziatrici di gRNA per bersagliamento o targeting all'interno di o vicino a un gene BCL11A o altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi CRISPR/Cas9 da *S. pyogenes* sono state identificate in SEQ ID NOs: 1- 29.482 della Lista di sequenze. Sequenze spaziatrici di gRNA per bersagliamento o targeting all'interno o vicino a un gene BCL1 1A o altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi CRISPR/Cas9 da *S. aureus* sono state identificate in SEQ ID NOs: 29.483-32.387 della Lista di sequenze. Sequenze

U001.D1.SM.61E

spaziatrici di gRNA per bersagliamento o targeting all'interno di o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi CRISPR/Cas9 da *S. thermophilus* sono state identificate in SEQ ID NOs. 32.388-33.420 della Lista di sequenze. Sequenze spaziatrici di gRNA per bersagliamento o targeting all'interno di o vicino a un gene BCL11A o altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi CRISPR/Cas9 da *T. denticola* sono state identificate in SEQ ID NOs. 33.421-33.851 della Lista di sequenze. Sequenze spaziatrici di gRNA per bersagliamento o targeting all'interno di o vicino a un gene BCL11A o altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi CRISPR/Cas9 da *N. meningitides* sono state identificate in SEQ ID NOs. 33.852-36.731. Sequenze spaziatrici di gRNA per bersagliamento o targeting all'interno di o vicino a un gene BCL11A o altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A con una endonucleasi CRISPR/Cpf1 da *Acidominococcus*, *Lachnospiraceae*, e *Franciscella Novicida* sono state identificate in SEQ ID NOs. 36.732-71.947.

Per esempio, la sequenza di controllo di trascrizione del gene BCL11A può essere modulata o inattivata mediante delezioni che insorgono a causa del pathway di NHEJ. NHEJ può essere utilizzata per cancellare segmenti della sequenza di controllo di trascrizione del gene BCL11A, direttamente o alterando siti donatori o accettori di splice attraverso il clivaggio da un gRNA che bersaglia più posizioni, o numerosi gRNA.

U001.D1.SM.61E

La sequenza di controllo di trascrizione del gene BCL11A può anche essere modulata o inattivata inserendo un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata. Per esempio, il donatore per modulazione o attivazione mediante HDR contiene la sequenza di controllo di trascrizione modificata del gene BCL11A con piccoli o grandi bracci di omologia fiancheggianti per consentire l'annealing. HDR è essenzialmente un meccanismo privo di errore che utilizza una sequenza di DNA omologa fornita come uno stampo o template durante la riparazione di DSB. Il tasso di riparazione diretta da omologia (HDR) è in funzione della distanza tra la sequenza di controllo di trascrizione e il sito tagliato, così la scelta di siti bersaglio sovrapposti o più vicini è importante. Template o stampi possono includere sequenze extra affiancate dalle regioni omologhe o possono contenere una sequenza che differisce dalla sequenza genomica, consentendo quindi l'editing di sequenza.

In aggiunta alla modulazione o inattivazione della sequenza di controllo di trascrizione del gene BCL11A mediante NHEJ o HDR, è possibile una gamma di altre opzioni. Se vi sono delezioni piccole o grandi, può essere introdotto con knock in un cDNA che contiene una sequenza di controllo di trascrizione modificata. Un cDNA a piena lunghezza può essere introdotto in qualsiasi "porto sicuro" o "safe harbour", ma deve utilizzare un promotore fornito o altro promotore. Se questo costrutto è inserito con knock in nella posizione corretta, esso avrà un controllo fisiologico simile al gene normale. Coppie di nucleasi possono venire utilizzate per cancellare regioni geniche, sebbene dovrebbe

U001.D1.SM.61E

venire solitamente fornito un donatore per modulare o inattivare la funzione. In questo caso sarebbero forniti il gRNA e altra sequenza donatrice.

Alcune strategie di ingegneria genomica implicano la modulazione o inattivazione di una sequenza di controllo di trascrizione del gene BCL11A mediante la delezione di almeno una porzione della sequenza di controllo di trascrizione del gene BCL11A e/o knocking-in di un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata nel locus del gene corrispondente o in un locus “safe harbour” mediante riparazione diretta da omologia (HDR), che è anche nota come ricombinazione omologa (HR). Questa strategia può modulare o inattivare la sequenza di controllo di trascrizione del gene BCL11A e annullare, trattare e/o mitigare lo stato di malattia. Nucleotidi donatori per modulazione/inattivazione di sequenze di controllo di trascrizione sono spesso piccoli (< 300 bp). Questo è vantaggioso poiché efficienze di HDR possono essere inversamente correlate alla dimensione della molecola donatrice. Inoltre, è previsto che i template donatori possano adattarsi in molecole di virus-adoassociati (AAV) con dimensioni vincolate, che sono state mostrate essere un mezzo efficace di veicolazione di template donatore.

Riparazione diretta da omologia è un meccanismo cellulare per riparare rotture a doppio filamento (DSB). La forma più comune è ricombinazione omologa. Esistono altre vie o pathway per HDR, incluso annealing a singolo filamento e HDR alternativa. Strumenti di ingegneria genomica consentono a ricercatori di manipolare vie o pathway di ricombinazione omologa cellulare per creare

U001.D1.SM.61E

modificazioni sito-specifiche al genoma. È stato trovato che cellule possono riparare una rottura a doppio filamento utilizzando una molecola donatrice sintetica fornita in trans. Pertanto, introducendo una rottura a doppio filamento vicino a una mutazione specifica e fornendo un donatore adatto, possono venire prodotti nel genoma cambiamenti mirati. Clivaggio specifico aumenta il tasso di HDR più di 1000 volte al di sopra del tasso di 1 in 10^6 cellule riceventi un donatore omologo da solo. Il tasso di riparazione diretta da omologia (HDR) in un particolare nucleotide è in funzione della distanza al sito di taglio, così è importante scegliere siti bersagli sovrapposti o più vicini. Editing genomico offre il vantaggio rispetto all'addizione genica dato che la correzione in situ lascia inalterato il resto del genoma.

Donatori forniti per editing mediante HDR variano notevolmente ma possono contenere la sequenza desiderata con piccoli o grandi bracci di omologia affiancati per permettere l'annealing al DNA genomico. Le regioni di omologia fiancheggianti i cambiamenti genetici introdotti possono essere di 30 bp o più piccole, oppure grandi quando una cassetta multi-kilobasica che può contenere promotori, cDNA, etc. Sono stati utilizzati sia donatori oligonucleotidici a singolo filamento che a doppio filamento. Questi oligonucleotidi variano nelle dimensioni da meno di 100 nt a più di molte kb, sebbene ssDNA può peggio possa anche essere generato e utilizzato. Donatori a doppio filamento possono venire utilizzati, inclusi ampliconi di PCR, plasmidi e mini-cerchi. In generale, è stato trovato che un vettore AAV può essere un mezzo molto efficace di veicolazione di un template donatore, sebbene il limite di impacchettamento per

U001.D1.SM.61E

donatori individuali sia <5kb. La trascrizione attiva del donatore aumentava HDR di tre volte, il che indica che l'inclusione di promotore può aumentare la conversione. Di contro, la metilazione di CpG del donatore diminuiva l'espressione genica e HDR.

In aggiunta a endonucleasi di tipo selvatico, come Cas9, esistono varianti di nickasi che hanno uno o l'altro dominio nucleasico inattivato, dando luogo al taglio di soltanto un filamento di DNA. HDR può essere diretta da Cas nickasi individuali o utilizzando coppie di nickasi che fiancheggiano l'area bersaglio. Donatori possono essere a singolo filamento, nicked o dsDNA.

Il DNA donatore può essere fornito con la nucleasi o indipendentemente da una varietà di differenti metodi, per esempio mediante trasfezione, nanoparticelle, microiniezione, o trasduzione virale. Una gamma di opzioni di tethering sono state proposte per aumentare la disponibilità dei donatori per HDR. Esempi includono il fissaggio del donatore alla nucleasi, il fissaggio a proteine di legame al DNA che si legano vicino, o il fissaggio a proteine che sono implicate nel legame o nella riparazione di estremità di DNA.

La scelta della via o pathway di riparazione può essere guidata da un numero di condizioni di coltura, come quelle che influenzano il ciclo cellulare, o dal bersagliamento o targeting di riparazione di DNA e proteine associate. Per esempio, per aumentare HDR, molecole chiave di NHEJ possono venire soppresse, come KU70, KU80 o DNA ligasi IV.

Senza un donatore presente, le estremità da una rottura di DNA o estremità da rotture differenti possono essere unite utilizzando le numerose vie o pathway di

U001.D1.SM.61E

riparazione non omologa in cui le estremità di DNA sono unite con un piccolo appaiamento di basi o senza alcun appaiamento di basi alla giunzione. In aggiunta a NHEJ canonica esistono meccanismi simili di riparazione, come alt-NHEJ. Se ci sono due rotture, il segmento intermedio può essere cancellato o invertito. Vie o pathway di riparazione di NHEJ possono portare a inserzioni, delezioni o mutazioni a livello delle giunzioni.

NHEJ è stata utilizzata per inserire una cassetta di espressione genica inducibile di 15-kb in un locus definito in linee cellulari umane dopo clivaggio nucleasico. Maresca, M., Lin, V.G., Guo, N. & Yang, Y., *Genome Res* 23, 539-546 (2013).

Oltre all'editing genomico mediante NHEJ o HDR, sono state condotte inserzioni geniche sito-specifiche che utilizzano sia la via o pathway di NHEJ che HR. Un approccio di combinazione può essere applicabile in certi contesti, possibilmente compresi confini di introni/esoni. NHEJ può risultare efficace per ligazione nell'introne, mentre la HDR priva di errori può essere meglio adatta nella regione codificante.

Come ulteriore alternativa, il gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata può essere introdotta tramite knock in nel locus del gene corrispondente oppure in un sito "safe harbor", come AAVS1. In alcuni esempi i metodi possono fornire un gRNA o una coppia di gRNA che possono venire utilizzati per facilitare l'incorporazione di una nuova sequenza da un template donatore polinucleotidico per l'introduzione con knock-in di una parte del o l'intero gene

U001.D1.SM.61E

BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata.

I metodi possono fornire coppie di gRNA che producono una delezione tagliando il gene due volte, un gRNA tagliando all'estremità 5' di una o più mutazioni e l'altro gRNA tagliando all'estremità 3' di una o più mutazioni che facilita l'inserzione di una nuova sequenza da un template donatore polinucleotidico per sostituire la sequenza di controllo di trascrizione del gene BCL11A. Il taglio può essere effettuato mediante una coppia di DNA endonucleasi che producono ciascuna una DSB nel genoma, o mediante molteplici nickasi che producono insieme una DSB nel genoma.

In alternativa, i metodi possono fornire un gRNA per produrre un taglio a doppio filamento attorno a una sequenza di controllo di trascrizione del gene BCL11A che facilita l'inserzione di una nuova sequenza da un template donatore polinucleotidico per sostituire la sequenza di controllo di trascrizione del gene BCL11A con un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata. Il taglio a doppio filamento può essere prodotto da una singola DNA endonucleasi o da molteplici nickasi che insieme producono una DSB nel genoma.

Modificazioni illustrative all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A includono sostituzioni all'interno o vicino (prossimali) alla sequenza di controllo di trascrizione del gene BCL11A a cui si fa riferimento sopra, come all'interno della regione di meno di 3 kb, meno di 2kb, meno di 1 kb, meno di 0,5 kb a

U001.D1.SM.61E

monte o a valle della sequenza di controllo di trascrizione.

Tali varianti possono includere sostituzioni che sono maggiori nella direzione 5' e/o 3' rispetto alla sostituzione specifica in questione, o più piccole in una delle due direzioni. Pertanto, per “vicino” o “prossimale” in merito a tali sostituzioni specifiche, si intende che il locus di SSB o DSB associato con un confine di sostituzione desiderato (indicato qui anche come un punto terminale) può essere all'interno di una regione che è meno di circa 3 kb dal locus di riferimento indicato. Il locus di SSB o DSB può essere più prossimale e entro 2 kb, entro 1 kb, entro 0,5 kb, o entro 0,1 kb. Nel caso di una piccola sostituzione, il punto terminale desiderato può essere in corrispondenza del o “adiacente al” locus di riferimento, con cui si intende che il punto terminale può essere entro 100 bp, entro 50 bp, entro 25 bp o da meno di circa 10 bp a 5 bp dal locus di riferimento.

Esempi comprendenti sostituzioni maggiori o minori possono essere attesi fornire lo stesso beneficio, purché l'attività di controllo di trascrizione sia modulata o inattivata. È quindi previsto che molte variazioni delle sostituzioni descritte e illustrate nel presente documento possano essere efficaci per migliorare emoglobinopatie.

Un'altra strategia di ingegneria genomica implica la delezione di esoni o introni. La delezione mirata di specifici esoni o introni può essere una strategia attraente per trattare un grande sottoinsieme di pazienti con un singolo cocktail terapeutico. Delezioni possono essere delezioni di singoli esoni o introni o delezioni di molteplici esoni o introni. Mentre delezioni di molteplici esoni

U001.D1.SM.61E

possono raggiungere un numero maggiore di pazienti, per delezioni più grandi l'efficienza di delezione diminuisce notevolmente con una dimensione aumentata. Pertanto, l'intervallo di delezioni può avere una dimensione da 40 a 10.000 paia di basi (bp). Per esempio, dimensioni di delezioni possono variare da 40 a 100; da 100 a 300; da 300 a 500; da 500 a 1.000; da 1.000 a 2.000; da 2.000 a 3.000; da 3.000 a 5.000; o da 5.000 a 10.000 paia di basi. Può essere desiderabile cancellare un introne se l'introne contiene un elemento regolatorio, come un elemento di controllo di trascrizione (per esempio un sito di legame a fattore di trascrizione).

Al fine di assicurare che il pre-mRNA sia correttamente processato dopo la delezione, i segnali di splicing circostanti possono essere cancellati. Donatori e accettori di splicing sono generalmente entro 100 paia di basi dell'introne adiacente. Pertanto, in alcuni esempi, metodi possono fornire tutti i gRNA che tagliano approssimativamente +/- 100-3100 bp rispetto a ciascuna giunzione di esone/introne d'interesse.

Per qualsiasi delle strategie di editing genomico, editing genico può essere confermato mediante analisi di sequenziamento o di PCR.

Selezione di sequenza bersaglio

Spostamenti nella posizione del confine 5' e/o 3' rispetto a loci di riferimento particolari possono essere utilizzati per facilitare o potenziare particolari applicazioni di editing genico, che dipendono in parte dal sistema endonucleasico selezionato per l'editing, come descritto e illustrato più sotto nel presente documento.

U001.D1.SM.61E

In un primo esempio non limitativo di tale selezione di sequenza bersaglio, molti sistemi endonucleasici hanno regole o criteri che possono guidare la selezione iniziale di potenziali siti bersaglio per clivaggio, come il requisito di un motivo di sequenza di PAM in una particolare posizione adiacente ai siti di clivaggio di DNA nel caso di endonucleasi di Tipo II o di Tipo V di CRISPR.

In un altro esempio non limitativo di selezione o ottimizzazione di sequenza bersaglio, la frequenza di attività off-target per una particolare combinazione di sequenza bersaglio e endonucleasi di editing genico (vale a dire la frequenza di DSB che si verificano in siti differenti dalla sequenza bersaglio selezionata) può venire valutata rispetto alla frequenza di attività on-target. In alcuni casi, cellule che sono state correttamente editate a livello del locus desiderato possono avere un vantaggio selettivo rispetto a altre cellule. Esempi illustrativi, ma non limitative, di un vantaggio selettivo includono l'acquisizione di attributi come potenziato tasso di replicazione, persistenza, resistenza a alcune condizioni, aumentati tassi di attecchimento o engraftment successivo o persistenza *in vivo* dopo l'introduzione in un paziente, e altri attributi associati al mantenimento o aumentati numeri o aumentata vitalità di tali cellule. In altri casi, cellule che sono state correttamente editate al locus desiderato possono essere positivamente selezionate mediante uno o più metodi di screening utilizzati per identificare, smistare o altrimenti selezionare cellule che sono state correttamente editate. Sia vantaggio selettivo che metodi di selezione diretta possono trarre vantaggio del fenotipo associato con la correzione. In alcuni casi cellule possono essere editate due o più volte al fine di creare una seconda

U001.D1.SM.61E

modificazione che crea un nuovo fenotipo che è utilizzato per selezionare o purificare la popolazione di cellule desiderata. Una tale seconda modificazione potrebbe essere creata aggiungendo un secondo gRNA per un marcatore selezionabile o vagliabile. In alcuni casi cellule possono essere correttamente editate a livello del locus desiderato utilizzando un frammento di DNA che contiene il cDNA e anche un marcatore selezionabile.

Sia che qualsiasi vantaggio selettivo sia applicabile o qualsiasi selezione diretta sia da applicare in un particolare caso, la selezione di sequenza bersaglio può anche essere guidata dalla considerazione di frequenze off-target al fine di aumentare l'efficacia dell'applicazione e/o ridurre il potenziale per alterazioni indesiderate in siti differenti dal bersaglio desiderato. Come descritto ulteriormente e illustrato nel presente documento e nell'arte, l'occorrenza di attività off-target può essere influenzata da un numero di fattori comprese similarità e dissimilarità tra il sito bersaglio e vari siti off-target, nonché la particolare endonucleasi utilizzata. Sono disponibili strumenti di bioinformatica che aiutano nella predizione di attività off-target, e frequentemente tali strumenti possono anche venire utilizzati per identificare i siti con maggiore probabilità di attività off-target, che possono poi venire valutati in contesti sperimentali per valutare frequenze relative di attività da off-target a on-target, consentendo quindi la selezione di sequenze che hanno relative attività on-target superiori. Esempi illustrativi di tali tecniche sono qui forniti, e altri sono noti nell'arte.

Un altro aspetto di selezione di sequenza bersaglio si riferisce a eventi di ricombinazione omologa. Sequenze che condividono regioni di omologia

U001.D1.SM.61E

possono servire come punti focali per eventi di ricombinazione omologa che portano alla delezione di sequenze intermedie. Tali eventi di ricombinazione si verificano durante il normale corso di replicazione di cromosomi e altre sequenze di DNA, e anche in altri momenti quando sequenze di DNA sono sintetizzate, come nel caso di riparazioni di rotture a doppio filamento (DSB), che hanno luogo su base regolare durante il normale ciclo di replicazione cellulare ma possono anche essere potenziate dall'occorrenza di vari eventi (come luce UV e altri induttori di rottura di DNA) o dalla presenza di alcuni agenti (come vari induttori chimici). Molti di tali induttori fanno sì che DSB avvengano indiscriminatamente nel genoma, e DSB possono essere regolarmente indotte e riparate in cellule normali. Durante la riparazione la sequenza originale può venire ricostruita con completa fedeltà, tuttavia, in alcuni casi, piccole inserzioni o delezioni (indicate come “indel”) sono introdotte nel sito di DSB.

DSB possono anche essere specificamente indotte in particolari posizioni, come nel caso dei sistemi di endonucleasi descritti nel presente documento, che possono venire utilizzati per provocare eventi di modificazione genica diretti o preferenziali in posizioni cromosomiche selezionate. La tendenza per sequenze omologhe di essere soggette a ricombinazione nel contesto di riparazione di DNA (nonché replicazione) può essere sfruttata in un numero di circostanze, ed è la base per una applicazione di sistemi di editing genico, come CRISPR, in cui riparazione diretta da omologia è utilizzata per inserire una sequenza d'interesse, fornita attraverso l'uso di un polinucleotide “donatore”, in una posizione

U001.D1.SM.61E

cromosomica desiderata.

Regioni di omologia tra sequenze particolari, che possono essere piccole regioni di “micro-omologia” che possono comprendere appena dieci paia di basi o meno, possono anche essere utilizzate per provocare delezioni desiderate. Per esempio, una singola DSB può venire introdotta a livello di un sito che presenta micro-omologia con una sequenza adiacente. Durante il normale corso di riparazione di tale DSB, un risultato che si verifica con elevata frequenza è la delezione della sequenza intermedia come risultato di ricombinazione facilitata dalla DSB e dal processo di riparazione cellulare concomitante.

In alcune circostanze, tuttavia, la selezione di sequenze bersaglio all'interno di regioni di omologia può dare origine a delezioni molto più grandi, incluse fusioni geniche (quando le delezioni sono in regioni codificanti), che possono o meno essere desiderate date le particolari circostanze.

Gli esempi forniti nel presente documento illustrano inoltre la selezione di varie regioni bersaglio per la creazione di DSB progettate per indurre sostituzioni che hanno come risultato la modulazione o inattivazione di attività di proteine di controllo di trascrizione, nonché la selezione di specifiche sequenze bersaglio all'interno di tali regioni che sono progettate per minimizzare eventi off-target rispetto a eventi on-target.

Modificazioni di acido nucleico

In alcuni casi, polinucleotidi introdotti in cellule possono comprendere una o più modificazioni che possono essere utilizzate individualmente o in combinazione, per esempio per potenziare attività, stabilità o specificità, dopo

U001.D1.SM.61E

veicolazione, ridurre risposte immunitarie innate in cellule ospiti, o per altri potenziamenti, come descritto ulteriormente nel presente documento e come noto nell'arte.

In alcuni esempi polinucleotidi modificati possono venire utilizzati nel sistema CRISPR/Cas9/Cpf1, nel qual caso gli RNA guida (guide monomolecolari o guide a doppia molecola) e/o un DNA o un RNA codificante una endonucleasi Cas o Cpf1 introdotta in una cellula possono essere modificati, come descritto e illustrato sotto. Tali polinucleotidi modificati possono essere utilizzati nel sistema CRISPR/Cas9/Cpf1 per editare uno qualsiasi o più loci genomici.

Utilizzando il sistema CRISPR/Cas9/Cpf1 ai fini di illustrazioni non limitative di tali usi, modificazioni di RNA guida possono venire utilizzate per potenziare la formazione o stabilità del complesso di editing genomico CRISPR/Cas9/Cpf1 comprendente RNA guida, che possono essere guide monomolecolari o a doppia molecola, e una endonucleasi Cas o Cpf1. Modificazioni di RNA guida possono anche o in alternativa essere utilizzate per potenziare l'inizio, la stabilità o la cinetica di interazioni tra il complesso di editing genomico con la sequenza bersaglio nel genoma, che può essere utilizzata, per esempio, per potenziare attività on-target. Modificazioni di RNA guida possono anche o in alternativa essere utilizzati per potenziare specificità, per esempio i tassi relativi di editing genomico a livello del sito on-target rispetto a effetti a livello di altri siti (off-target).

Modificazioni possono anche o in alternativa essere utilizzate per aumentare la stabilità di un RNA guida, per esempio aumentando la sua resistenza a

U001.D1.SM.61E

degradazione da ribonucleasi (RNasi) presenti in una cellula, causando quindi un aumento della sua emivita nella cellula. Modificazioni che potenziano l'emivita di RNA guida possono essere particolarmente utili in aspetti in cui una endonucleasi Cas o Cpf1 è introdotta nella cellula per essere editata tramite un RNA che necessita di venire tradotto al fine di generare endonucleasi, poiché un aumento dell'emivita di RNA guida introdotti contemporaneamente all'RNA codificante l'endonucleasi può essere utilizzato per aumentare il tempo che gli RNA guida e l'endonucleasi Cas o Cpf1 codificata coesistono nella cellula.

Modificazioni possono anche o in alternativa essere utilizzate per diminuire la probabilità o il grado a cui RNA introdotti in cellule elicitano risposte immunitarie innate. Tali risposte, che sono state ben caratterizzate nel contesto di interferenza di RNA (RNAi), inclusi piccoli RNA interferenti (siRNA), come descritto sotto e nell'arte, tendono a essere associate con una emivita ridotta del RNA e/o la sollecitazione di citochine o altri fattori associati con risposte immunitarie.

Uno o più tipi di modificazioni possono anche essere apportati a RNA codificanti una endonucleasi che sono introdotti in una cellula, incluse, senza limitazione, modificazioni che potenziano la stabilità del RNA (come aumentando la sua degradazione da RNasi presenti nella cellula), modificazioni che potenziano la traduzione del prodotto risultante (vale a dire l'endonucleasi), e/o modificazioni che diminuiscono la probabilità o il grado a cui gli RNA introdotti in cellule elicitano risposte immunitarie innate.

Combinazioni di modificazioni, come quelle menzionate e altre, possono

U001.D1.SM.61E

similmente venire utilizzate. Nel caso di CRISPR/Cas9/Cpf1, per esempio, uno o più tipi di modificazioni possono venire apportati a RNA guida (inclusi quelle esemplificate sopra), e/o uno o più tipi di modificazioni possono venire apportati a RNA codificanti endonucleasi Cas (incluse quelle esemplificate sopra).

A titolo illustrativo, RNA guida utilizzati nel sistema CRISPR/Cas9/Cpf1, o altri RNA più piccoli, possono venire prontamente sintetizzati con mezzi chimici, consentendo a un numero di modificazioni di venire prontamente incorporate, come illustrato sotto e descritto nell'arte. Molte procedure sintetiche chimiche sono in continua espansione, purificazioni di tali RNA mediante procedure come cromatografia liquida ad elevate prestazioni (HPLC, che evita l'uso di gel come PAGE) tendono a diventare più impegnative quando lunghezze di polinucleotidi aumentano significativamente oltre i cento nucleotidi o poco più. Un approccio che può venire utilizzato per generare RNA chimicamente modificati di lunghezza maggiore consiste nel produrre due o più molecole che sono ligate insieme. RNA molto più lunghi, come quelli codificanti una endonucleasi Cas9, sono più prontamente generati enzimaticamente. Mentre meno tipi di modificazioni sono disponibili per l'uso in RNA prodotti enzimaticamente, esistono ancora modificazioni che possono venire utilizzate per, ad esempio, potenziare la stabilità, ridurre la probabilità o il grado di riposte immunitarie innate, e/o potenziare altri attributi, come descritto ulteriormente sotto e nell'arte; e nuovi tipi di modificazioni sono regolarmente in fase di sviluppo.

A titolo illustrativo di vari tipi di modificazioni, specialmente quelle utilizzate

U001.D1.SM.61E

frequentemente con RNA più piccoli sintetizzati chimicamente, modificazioni possono comprendere uno o più nucleotidi modificati alla posizione 2' dello zucchero, in alcuni aspetti un nucleotide 2'-O-alchile-, 2'-O-alchil-O-alchile-, o 2'-fluoro-modificato. In alcuni esempi, modificazioni di RNA possono comprendere modificazioni 2'-fluoro, 2'-ammino o 2' O-metile sul ribosio di pirimidine, residui abasici, o una base invertita all'estremità 3' del RNA. Tali modificazioni possono essere ordinariamente incorporate in oligonucleotidi e questi oligonucleotidi sono stati mostrati avere una T_m maggiore (vale a dire una maggiore affinità di legame a bersaglio) rispetto a 2'-deossioligo-nucleotidi contro un dato bersaglio.

Sono state mostrate numerose modificazioni nucleotidiche e nucleosidiche per rendere l'oligonucleotide in cui esse sono incorporate più resistente a digestione nucleasica rispetto all'oligonucleotide nativo; questi oligonucleotidi modificati sopravvivono intatti per un tempo più lungo rispetto a oligonucleotidi non modificati. Esempi specifici di oligonucleotidi modificati includono quelli comprendenti scheletri o backbone modificati, per esempio fosforotioati, fosfotriesteri, metil fosfonati, legami interzuccheri alchilici a catena corta o cicloalchilici o legami interzuccheri eteroatomici o eterociclici a catena corta. Alcuni oligonucleotide sono oligonucleotide con scheletri di fosforotioato e quelli con scheletri di eteroatomi, in particolare $\text{CH}_2\text{-NH-O-CH}_2$, $\text{CH}_2\text{-N(CH}_3\text{)-O-CH}_2$ (noto come uno scheletro di metilene(metilimmino) o MMI), scheletri di $\text{CH}_2\text{-ON(CH}_3\text{)-CH}_2$, $\text{CH}_2\text{-N(CH}_3\text{)-N(CH}_3\text{)-CH}_2$ e $\text{O-N(CH}_3\text{)-CH}_2\text{-CH}_2$, in cui lo scheletro di fosfodiesteri nativo è rappresentato

U001.D1.SM.61E

come O- P-- O- CH,); scheletri ammidici [si veda De Mesmaecker et al., *Ace. Chem. Res.*, 28:366-374 (1995)]; strutture di scheletro di morfolino (si veda Summerton e Weller, brevetto statunitense n° 5.034.506); scheletro di acido peptidonucleico (PNA) (in cui lo scheletro di fosfodiesteri dell'oligo-nucleotide è sostituito con uno scheletro di poliammide, i nucleotidi essendo legati direttamente o indirettamente agli atomi di azoto aza dello scheletro di poliammide, si veda Nielsen et al., *Science* 1991, 254, 1497). Legami contenenti fosforo includono, ma non sono limitati a, fosforotioati, fosforotioati chirali, fosforoditioati, fosfotriesteri, amminoalchilfosfotriesteri, metil- e altri alchil fosfonati comprendenti 3'-alchilene fosfonati e fosfonati chirali, fosfinati, fosforamidati comprendenti 3'-ammino fosforamidato e amminoalchilfosforamidati, tionofosforamidati, tionoalchilfosfonati, tionoalchilfosfotriesteri, e boranofosfati aventi legami 3'-5' normali, analoghi legati a 2'-5' di questi e quelli aventi polarità invertita in cui le coppie adiacenti di unità nucleosidiche sono legate 3'-5' a 5'-3' o 2'-5' a 5'-2'; si vedano i brevetti statunitensi n° 3.687.808; 4.469.863; 4.476.301; 5.023.243; 5.177.196; 5.188.897; 5.264.423; 5.276.019; 5.278.302; 5.286.717; 5.321.131; 5.399.676; 5.405.939; 5.453.496; 5.455.233; 5.466.677; 5.476.925; 5.519.126; 5.536.821; 5.541.306; 5.550.111; 5.563.253; 5.571.799; 5.587.361; e 5.625.050.

Composti oligomerici a base di morfolino sono descritti in Braasch e David Corey, *Biochemistry*, 41(14): 4503-4510 (2002); Genesis, Volume 30, Issue 3, (2001); Heasman, *Dev. Biol.*, 243: 209-214 (2002); Nasevicius et al., *Nat. Genet.*, 26:216-220 (2000); Lacerra et al., *Proc. Natl. Acad. Sci.*, 97: 9591-9596

U001.D1.SM.61E

(2000); e brevetto statunitense n° 5.034.506, concesso in data 23 luglio 1991.

Mimetici oligonucleotidici di cicloesenile e acido nucleico sono descritti in Wang et al., J. Am. Chem. Soc., 122: 8595-8602 (2000).

Scheletri oligonucleotidici modificati che non includono un atomo di fosforo hanno qui scheletri che sono formati da legami internucleosidici alchilici o cicloalchilici a catena corta, legami internucleosidici eteroatomici e alchilici o cicloalchilici misti, o uno o più legami internucleosidici eteroatomici o eterociclici a catena corta. Questi comprendono quelli aventi legami morfolino (formati in parte dalla porzione di zucchero di un nucleoside); scheletri silossanici; scheletri di solfuro, solfossido e solfone; scheletri di formacetile e tioformacetile; scheletri di metilene formacetile e tioformacetile; scheletri contenenti alchene; scheletri di solfammato; scheletri di metileneimmino e metileneidrazino; scheletri di solfonato e solfonammide; scheletri di ammido; e altri aventi parti di componenti di N, O, S e CH₂ misti; si vedano i brevetti U.S. n° 5.034.506; 5.166.315; 5.185.444; 5.214.134; 5.216.141; 5.235.033; 5.264.562; 5.264.564; 5.405.938; 5.434.257; 5.466.677; 5.470.967; 5.489.677; 5.541.307; 5.561.225; 5.596.086; 5.602.240; 5.610.289; 5.602.240; 5.608.046; 5.610.289; 5.618.704; 5.623.070; 5.663.312; 5.633.360; 5.677.437; e 5.677.439.

Una o più porzioni di zucchero sostituite possono anche venire incluse, per esempio una delle seguenti alla posizione 2': OH, SH, SCH₃, F, OCN, OCH₃, OCH₃ O(CH₂)_n CH₃, O(CH₂)_n NH₂, o O(CH₂)_n CH₃, dove n è da 1 a circa 10; alchile inferiore a C₁ a C₁₀, alcossialcossi, alchile inferiore sostituito, alcarile o aralchile; Cl; Br; CN; CF₃; OCF₃; O-, S-, o N-alchile; O-, S-, o N-alchenile;

U001.D1.SM.61E

SOCH₃; SO₂; CH₃; ONO₂; NO₂; N₃; NH₂; eterocicloalchile; eterocicloalcarile; amminoalchilammino; polialchilammino; silile sostituito; un gruppo di clivaggio di RNA; un gruppo reporter; un intercalante; un gruppo per migliorare le proprietà farmacocinetiche di un oligonucleotide; o un gruppo per migliorare le proprietà farmacocinetiche di un oligonucleotide e altri sostituenti aventi proprietà simili. In alcuni aspetti una modificazione include 2'-metossietossi (2'-O-CH₂CH₂OCH₃, noto anche come 2'-O-(2-metossietile)) (Martin et al, *Helv. Chim. Acta*, 1995, 78, 486). Altre modificazioni includono 2'-metossi (2'-O-CH₃), 2'-propossi (2'-OCH₂ CH₂CH₃) e 2'-fluoro (2'-F). Modificazioni simili possono anche venire prodotte in altre posizioni nell'oligo-nucleotide, in particolare nella posizione 3' dello zucchero sul nucleotide 3' terminale e sulla posizione 5' del nucleotide 5' terminale. Oligonucleotidi possono anche avere mimetici di zucchero, come ciclobutili al posto del gruppo pentofuranosile.

In alcuni esempi, sia uno zucchero che un legame internucleosidico, vale a dire lo scheletro, delle unità nucleotidiche possono essere sostituiti con nuovi gruppi. Le unità di base possono essere mantenute per ibridazione con un composto bersaglio di acido nucleico appropriato. Un tale composto oligomero, un mimetico oligonucleotidico che è stato mostrato avere eccellenti proprietà di ibridazione, è indicato come acido peptidonucleico (PNA). In composti PNA lo scheletro di zucchero di un oligonucleotide può essere sostituito con uno scheletro contenente ammidi, per esempio uno scheletro di amminoetilglicina. Le nucleobasi possono essere mantenute e legate direttamente o indirettamente a atomi di azoto aza della porzione ammidica dello scheletro. Brevetti statunitensi

U001.D1.SM.61E

rappresentativi che insegnano la preparazione di composti PNA comprendono, ma non sono limitati a, i brevetti statunitensi n° 5.539.082; 5.714.331; e 5.719.262. Un ulteriore insegnamento di composti PNA può essere trovato in Nielsen et al, Science, 254: 1497-1500 (1991).

RNA guida possono anche includere, in aggiunta o in alternativa, modificazioni o sostituzioni di nucleobasi (spesso indicate nell'arte semplicemente come "basi"). Come qui utilizzate, nucleobasi "non modificate" o "naturali" includono (A), guanina (G), timina (T), citosina (C), e uracile (U). Nucleobasi modificate includono nucleobasi trovate soltanto non frequentemente o transitoriamente in acidi nucleici naturali, per esempio ipoxantina, 6-metiladenina, 5-Me pirimidine, in particolare 5-metilcitosina (indicata anche come 5-metil-2' deossicitosina e spesso indicata nell'arte come 5-Me-C), 5-idrossimetilcitosina (HMC), glicosil HMC e gentobiosil HMC, nonché nucleobasi sintetiche, per esempio 2-amminoadenina, 2-(metilammino)adenina, 2-(imidazolilalchil)adenina, 2-(amminoachilammino)adenina o altre alchiladenine eterosostituite, 2-tiouracile, 2-tiotimina, 5-bromo-uracile, 5-idrossimetiluracile, 8-azaguanina, 7-deazaguanina, N6 (6-amminoetil)adenina, e 2,6-diamminopurina. Kornberg, A., DNA Replication, W. H. Freeman & Co., San Francisco, pp75-77 (1980); Gebeyehu et al., Nucl. Acids Res. 15:4513 (1997). Una base "universale" nota nell'arte, per esempio inosina, può anche venire inclusa. Sostituzioni 5-Me-C sono state mostrate aumentare la stabilità di duplex di acido nucleico di 0,6-1,2°C.

U001.D1.SM.61E

(Sanghvi, Y. S., in Crooke, S.

T. e Lebleu, B., eds., *Antisense Research and Applications*, CRC Press, Boca Raton, 1993, pp. 276-278) e sono aspetti di sostituzioni di basi.

Nucleobasi modificate possono comprendere altre nucleobasi sintetiche e naturali, come 5-metilcitosina (5-me-C), 5-idrossimetil citosina, xantina, ipoxantina, 2-amminoadenina, derivati 6-metilici e altri derivati alchilici di adenina e guanina, derivati 2-propilici e altri derivati alchilici di adenina e guanina, 2-tiouracile, 2-tiotimine e 2-tiocitosina, 5-alouracile e citosina, 5-propinil uracile e citosina, 6-azo uracile, citosina e timina, 5-uracile (pseudouracile), 4-tiouracile, 8-alo, 8-ammino, 8-tiolo, 8-tioalchile, 8-idrossile e altre adenine e guanine 8-sostituite, 5-alo, in particolare 5-bromo, 5-trifluorometile e altri uracili e citosine 5-sostituiti, 7-metilguanina e 7-metiladenina, 8-azaguanina e 8-azaadenina, 7-deazaguanina e 7-deazaadenina, e 3-deazaguanina e 3-deazaadenina.

Inoltre, nucleobasi possono comprendere quelle descritte nel brevetto statunitense n° 3.687.808, quelle descritte in 'The Concise Encyclopedia of Polymer Science And Engineering', pp. 858-859, Kroschwitz, J.I., ed. John Wiley & Sons, 1990, quelle descritte da Englisch et al., *Angewandte Chemie, International Edition*', 1991, 30, p. 613, e quelle descritte da Sanghvi, Y. S., Capitolo 15, *Antisense Research and Applications*', pp. 289- 302, Crooke, S.T. e Lebleu, B. ea., CRC Press, 1993. Alcune di queste nucleobasi sono particolarmente utili per aumentare l'affinità di legame dei composti oligomerici dell'invenzione. Queste includono pirimidine 5-sostituite, 6-azapirimidine e

U001.D1.SM.61E

purine N-2, N-6 e O-6 sostituite, comprendenti 2-amminopropiladenina, 5-propiniluracile e 5-propinilcitosina. Sostituzioni di 5-metilcitosina sono state mostrate aumentare la stabilità di duplex di acido nucleico di 0,6-1,2°C (Sanghvi, Y.S., Crooke, S.T. and Lebleu, B., eds, 'Antisense Research and Applications', CRC Press, Boca Raton, 1993, pp. 276-278) e sono aspetti di sostituzioni di basi, ancora più in particolare quando combinate con modificazioni di zucchero di 2'-O-metossietile. Nucleobasi modificate sono descritte nei brevetti statunitensi n° 3.687.808, nonché 4.845.205; 5.130.302; 5.134.066; 5.175.273; 5.367.066; 5.432.272; 5.457.187; 5.459.255; 5.484.908; 5.502.177; 5.525.711; 5.552.540; 5.587.469; 5.596.091; 5.614.617; 5.681.941; 5.750.692; 5.763.588; 5.830.653; 6.005.096; e nella Pubblicazione di domanda di brevetto U.S. 2003/0158403.

Quindi, il termine "modificato" si riferisce a uno zucchero, fosfato o una base non naturale che è incorporato(a) in un RNA guida, una endonucleasi, o una sequenza di controllo di trascrizione di BCL11A o sia un RNA guida che una endonucleasi. Non è necessario che tutte le posizioni in un dato oligonucleotide siano modificate uniformemente, e di fatto più di una delle modificazioni menzionate sopra può venire incorporata in un singolo oligonucleotide, o persino in un singolo nucleoside all'interno di un oligonucleotide.

Gli RNA guida e/o mRNA (o DNA) codificanti una endonucleasi possono essere legati chimicamente a una o più frazioni o coniugati che potenziano l'attività, la distribuzione cellulare, o l'assorbimento cellulare dell'oligonucleotide. Tali frazioni comprendono, ma non sono limitate a,

U001.D1.SM.61E

frazioni lipidiche come una frazione di colesterolo [Letsinger et al., Proc. Natl. Acad. Sci. USA, 86: 6553-6556 (1989)]; acido colico [Manoharan et al., Bioorg. Med. Chem. Lett., 4: 1053-1060 (1994)]; un tioetere, per esempio esil-S-tritiltiolo [Manoharan et al., Ann. N. Y. Acad. Sci., 660: 306-309 (1992) e Manoharan et al., Bioorg. Med. Chem. Lett., 3: 2765-2770 (1993)]; un tiocolesterolo [Oberhauser et al., Nucl. Acids Res., 20: 533-538 (1992)]; una catena alifatica, per esempio residui di dodecandiolo o undecile [Kabanov et al., FEBS Lett., 259: 327-330 (1990) e Svinarchuk et al., Biochimie, 75: 49- 54 (1993)]; un fosfolipide, per esempio di-esadecil-rac-glicerolo o trietilammonio 1,2-di-O-esadecil-rac-glicero-3-H-fosfonato [Manoharan et al., Tetrahedron Lett., 36: 3651-3654 (1995) e Shea et al., Nucl. Acids Res., 18: 3777-3783 (1990)]; una poliammina o una catena di polietilenglicole [Mancharan et al., Nucleosides & Nucleotides, 14: 969-973 (1995)]; acido adamantan-acetico [Manoharan et al., Tetrahedron Lett., 36: 3651-3654 (1995)]; una frazione di palmitile [(Mishra et al., Biochim. Biophys. Acta, 1264: 229-237 (1995)]; o una frazione di ottadecilammina o esilammino-carbonil-t-ossicolesterolo [Crooke et al., J. Pharmacol. Exp. Ther., 277: 923-937 (1996)]. Si vedano anche i brevetti statunitensi n° 4.828.979; 4.948.882; 5.218.105; 5.525.465; 5.541.313; 5.545.730; 5.552.538; 5.578.717. 5.580.731; 5.580.731; 5.591.584; 5.109.124; 5.118.802; 5.138.045; 5.414.077; 5.486.603; 5.512.439; 5.578.718; 5.608.046; 4.587.044; 4.605.735; 4.667.025; 4.762.779; 4.789.737; 4.824.941; 4.835.263; 4.876.335; 4.904.582; 4.958.013; 5.082.830; 5.112.963; 5.214.136; 5.082.830; 5.112.963; 5.214.136; 5.245.022; 5.254.469; 5.258.506; 5.262.536; 5.272.250;

U001.D1.SM.61E

5.292.873; 5.317.098; 5.371.241. 5.391.723; 5.416.203. 5.451.463; 5.510.475;
5.512.667; 5.514.785; 5.565.552; 5.567.810; 5.574.142; 5.585.481; 5.587.371;
5.595.726; 5.597.696; 5.599.923; 5.599. 928 e 5.688.941.

Zuccheri e altre frazioni possono venire utilizzati per bersagliare proteine e complessi comprendenti nucleotidi, come polisomi e liposomi cationici, in siti particolari. Per esempio il trasferimento diretto da cellule epatiche può essere mediato tramite recettori di asialoglicoproteina (ASGPR); si veda, per esempio, Hu, et al., *Protein Pept Lett.* 21(10):1025-30 (2014). Altri sistemi noti nell'arte e sviluppati regolarmente possono venire utilizzati per bersagliare biomolecole d'uso nel presente caso e/o complessi di esse a particolari cellule bersaglio d'interesse.

Queste frazioni o coniugati di bersagliamento o targeting possono includere gruppi coniugati legati covalentemente a gruppi funzionali, come gruppi idrossilici primari o secondari. Gruppi coniugati della descrizione includono intercalanti, molecole reporter, poliammine, poliammidi, polietilenglicoli, polieteri, gruppi che potenziano le proprietà farmacodinamiche di oligomeri, e gruppi che potenziano le proprietà farmacocinetiche di oligomeri. Gruppi coniugati tipici includono colesteroli, lipidi, fosfolipidi, biotina, fenazina, folato, fenantridina, antrachinone, acridina, fluoresceine, rodamine, cumarine, e coloranti. Gruppi che potenziano le proprietà farmacodinamiche, nel contesto di questa descrizione, includono gruppi che migliorano l'assorbimento, migliorano la resistenza a degradazione, e/o rafforzano l'ibridazione sequenza-specifica con l'acido nucleico bersaglio. Gruppi che potenziano le proprietà farmacocinetiche,

U001.D1.SM.61E

nel contesto di questa descrizione, includono gruppi che migliorano l'assorbimento, la distribuzione, il metabolismo o l'escrezione dei composti della presente descrizione. Gruppi coniugati rappresentativi sono descritti nel brevetto statunitense n°6.287.860. Frazioni coniugate includono, ma non sono limitate a, frazioni lipidiche come una frazione di colesterolo, acido colico, un tioetere, per esempio esil-5-tritiltiolo, un tiocolesterolo, una catena alifatica, per esempio residui di dodecandiolo o undecile, un fosfolipide, per esempio di esadecil-rac-glicerolo o trietilammonio 1,2-di-O-esadecil-rac-glicero-3-H-fosfonato, una poliammina o una catena di polietilenglicole, o acido adamantanacetico, una frazione di palmitile, o una frazione di ottadecilammina o esilammino-carbonil-ossi-colesterolo. Si vedano, per esempio, i brevetti statunitensi n° 4.828.979; 4.948.882; 5.218.105; 5.525.465; 5.541.313; 5.545.730; 5.552.538; 5.578.717. 5.580.731; 5.580.731; 5.591.584; 5.109.124; 5.118.802; 5.138.045; 5.414.077; 5.486.603; 5.512.439; 5.578.718; 5.608.046; 4.587.044; 4.605.735; 4.667.025; 4.762.779; 4.789.737; 4.824.941; 4.835.263; 4.876.335; 4.904.582; 4.958.013; 5.082.830; 5.112.963; 5.214.136; 5.082.830; 5.112.963; 5.214.136; 5.245.022; 5.254.469; 5.258.506; 5.262.536; 5.272.250; 5.292.873; 5.317.098; 5.371.241. 5.391.723; 5.416.203. 5.451.463; 5.510.475; 5.512.667; 5.514.785; 5.565.552; 5.567.810; 5.574.142; 5.585.481; 5.587.371; 5.595.726; 5.597.696; 5.599.923; 5.599.928 e 5.688.941.

Polinucleotidi più lunghi che sono meno suscettibili a sintesi chimica e sono tipicamente prodotti mediante sintesi enzimatica possono anch'essi venire modificati in vari modi. Tali modificazioni possono includere, per esempio,

U001.D1.SM.61E

l'introduzione di alcuni analoghi nucleotidici, l'incorporazione di sequenze particolari o altre frazioni alle estremità 5' o 3' di molecole, e altre modificazioni. A titolo illustrativo, il mRNA codificante Cas9 ha una lunghezza di circa 4 kb, e può venire sintetizzato mediante trascrizione *in vitro*. Modificazioni al mRNA possono venire applicate, per esempio, per aumentare la sua traduzione o stabilità (come aumentando la sua resistenza a degradazione con una cellula) o ridurre la tendenza del RNA a elicitare una risposta immunitaria innata che è spesso osservata in cellule dopo l'introduzione di RNA esogeni, in particolare RNA più lunghi come quello codificante Cas9.

Numerose modificazioni di questo tipo sono state descritte nell'arte, come code polyA, analoghi di 5' cap (per esempio l'analogo "Anti Reverse Cap Analog" (ARCA) o m7G(5')ppp(5')G (mCAP)), regioni 5' o 3' non tradotte (UTR), uso di basi modificate (come Pseudo-UTP, 2-Tio-UTP, 5-Metilcitidina-5'-Trifosfato (5-Metil-CTP) o N6-Metil-ATP), o trattamento con fosfatasi per rimuovere fosfatasi 5' terminali. Queste e altre modificazioni sono note nell'arte, e nuove modificazioni di RNA sono regolarmente in fase di sviluppo.

Esistono numerosi fornitori commerciali di RNA modificati, inclusi, per esempio, TriLink Biotech, AxoLabs, Bio-Synthesis Inc., Dharmacon e molti altri. Come descritto da TriLink, per esempio, 5-Metil-CTP può venire utilizzato per impartire caratteristiche desiderabili, come una aumentata stabilità nucleasica, aumentata traduzione o ridotta interazione di recettori immunitari innati con RNA trascritto *in vitro*. 5-Metilcitidina-5'-Trifosfato (5-Metil-CTP), N6-Metil-ATP, nonché Pseudo-UTP e 2-Tio-UTP, sono anche stati mostrati

U001.D1.SM.61E

ridurre la stimolazione immunitaria innata in coltura e *in vivo* potenziando nel contempo la traduzione, come illustrato in pubblicazioni da Kormann *et al.* e Warren *et al.* a cui si fa riferimento sotto.

È stato mostrato che RNA modificato chimicamente veicolato *in vivo* può venire utilizzato per ottenere effetti terapeutici migliorati; si veda, per esempio, Kormann *et al.*, *Nature Biotechnology* 29, 154-157 (2011). Tali modificazioni possono venire utilizzate, per esempio, per aumentare la stabilità della molecola di RNA e/o ridurre la sua immunogenicità. Utilizzando modificazioni chimiche come Pseudo-U, N6-Metil-A, 2-Tio-U e 5-Metil-C, è stato trovato che la sostituzione di appena un quarto dei residui di uridina e citidina con 2-Tio-U e 5-Metil-C rispettivamente dava come risultato una riduzione significativa in riconoscimento mediato da recettore toll-simile (TLR) del mRNA in topi. Riducendo l'attivazione del sistema immunitario innato, queste modificazioni possono venire utilizzate per aumentare efficacemente la stabilità e la longevità del mRNA *in vivo*; si veda, per esempio, Kormann *et al.*, *supra*.

È stato anche mostrato che la somministrazione ripetuta di RNA messaggeri sintetici incorporanti modificazioni progettate per bypassare risposte antivirali innate può riprogrammare cellule umane differenziate a pluripotenza. Si veda, per esempio, Warren, *et al.*, *Cell Stem Cell*, 7(5):618-30 (2010). Tali mRNA modificati che agiscono come proteine di riprogrammazione primarie possono essere un mezzo efficiente di riprogrammazione di molteplici tipi di cellule umane. Tali cellule sono indicate come cellule staminali pluripotenti indotte (iPSC), ed è stato trovato che RNA sintetizzato enzimaticamente incorporante 5-

U001.D1.SM.61E

Metil-CTP, Pseudo-UTP e un analogo “Anti Reverse Cap Analog” (ARCA) potrebbe venire utilizzato per evadere efficacemente la risposta antivirale delle cellule; si veda, per esempio, Warren *et al.*, *supra*.

Altre modificazioni di polinucleotidi descritte nell’arte includono, per esempio, l’uso di code polyA, l’addizione di analoghi 5’ cap (come m7G(5’)ppp(5’)G (mCAP)), modificazioni di regioni 5’ o 3’ non tradotte (UTR) o trattamento con fosfatasi per rimuovere fosfatasi 5’ terminali – e nuovi approcci sono regolarmente in fase di sviluppo.

Numerose composizioni e tecniche applicabili alla generazione di RNA modificati per l’uso nel presente documento sono state sviluppate congiuntamente alla modificazione di RNA interferenza (RNAi), inclusi piccoli RNA interferenti (siRNA). siRNA presentano particolari sfide *in vivo* poiché i loro effetti di silenziamento genico tramite interferenza di mRNA sono generalmente transitori, il che può richiedere una somministrazione ripetuta. In aggiunta, siRNA sono RNA a doppio filamento (dsRNA) e cellule di mammifero hanno risposte immunitarie che si sono evolute per rilevare e neutralizzare dsRNA, che è spesso un sottoprodotto di infezione virale. Quindi, esistono enzimi di mammiferi come PKR (chinasi responsiva a dsRNA), e gene potenzialmente inducibile da acido retinoico - I (RIG-I), che possono mediare risposte cellulari a dsRNA, nonché recettori Toll-simili (come TLR3, TLR7 e TLR8) che possono innescare l’induzione di citochine in risposta a tali molecole; si vedano, per esempio, le disamine da Angart *et al.*, *Pharmaceuticals* (Basel) 6(4): 440-468 (2013); Kanasty *et al.*, *Molecular Therapy* 20(3): 513-524

U001.D1.SM.61E

(2012); Burnett et al., *Biotechnol J.* 6(9):1130-46 (2011); Judge e MacLachlan, *Hum Gene Ther* 19(2):111-24 (2008); e riferimenti citati all'interno di essi.

Un'ampia varietà di modificazioni sono state sviluppate e applicate per potenziare la stabilità a RNA, ridurre risposte immunitarie innate, e/o conseguire altri benefici che possono essere utili congiuntamente all'introduzione di polinucleotidi in cellule umane, come descritto nel presente documento; si vedano, per esempio, le disamine da Whitehead KA et al., *Annual Review of Chemical and Biomolecular Engineering*, 2: 77-96 (2011); Gaglione e Messere, *Mini Rev Med Chem*, 10(7):578-95 (2010); Chernolovskaya et al, *Curr Opin Mol Ther.*, 12(2):158-67 (2010); Deleavey et al., *Curr Protoc Nucleic Acid Chem* Capitolo 16: Unit 16.3 (2009); Behlke, *Oligonucleotides* 18(4):305-19 (2008); Fucini et al., *Nucleic Acid Ther* 22(3): 205-210 (2012); Bremsen et al., *Front Genet* 3:154 (2012).

Come notato sopra, esistono numerosi fornitori commerciali di RNA modificati, molti dei quali si sono specializzati in modificazioni progettate per migliorare l'efficacia di siRNA. Vari approcci sono offerti sulla base di varie scoperte riportate nella letteratura. Per esempio, Dharmacon nota che la sostituzione di un ossigeno non potante con zolfo (fosforotioato, PS) è stata estensivamente utilizzata per migliorare la resistenza a nucleasi di siRNA, come riportato da Kole, *Nature Reviews Drug Discovery* 11:125-140 (2012). Modificazioni della posizione 2' del ribosio sono state riportate migliorare la resistenza nucleasica del legame di fosfato internucleotidico aumentando la stabilità a duplex (T_m), che è stata anche mostrata fornire protezione da

U001.D1.SM.61E

attivazione immunitaria. Una combinazione di modificazioni moderate di scheletro di PS con piccole sostituzioni 2' ben tollerate (2'-O-Metile, 2'-Fluoro, 2'-Idro) sono state associate con siRNA altamente stabili per applicazioni *in vivo*, come riportato da Soutschek et al. Nature 432:173-178 (2004); e modificazioni di 2'-O-Metile sono state riportate essere efficaci nel miglioramento della stabilità come riportato da Volkov, Oligonucleotides 19:191-202 (2009). In merito alla diminuzione dell'induzione di risposte immunitarie innate, la modificazione di sequenze specifiche con 2'-O-Metile, 2'-Fluoro, 2'-Idro è stata riportata ridurre l'interazione TLR7/TLR8, al tempo stesso conservando generalmente attività di silenziamento; si vedano, per esempio, Judge et al., Mol. Ther. 13:494-505 (2006); e Cekaite et al., J. Mol. Biol. 365:90-108 (2007). Modificazioni supplementari, come 2-tiouracile, pseudouracile, 5-metilcitosina, 5-metiluracile, e N6-metiladenosina, sono state anche mostrate minimizzare gli effetti immunitari mediati da TLR3, TLR7, e TLR8; si veda, per esempio, Kariko, K. et al., Immunity 23:165-175 (2005).

Come noto nell'arte, e disponibile in commercio, numerosi coniugati possono venire applicati a polinucleotidi, come RNA, per l'uso nel presente documento, che possono potenziare la loro veicolazione e/o il loro assorbimento da parte di cellule, inclusi, per esempio, colesterolo, tocoferolo e acido folico, lipidi, peptidi, polimeri, linker e aptameri; si veda, per esempio, la disamina da Winkler, Ther. Deliv. 4:791-809 (2013), e riferimenti citati all'interno.

Ottimizzazione di codoni

Un polinucleotide codificante un polipeptide sito-diretto può venire ottimizzato

U001.D1.SM.61E

nel codone secondo metodi standard nell'arte per l'espressione nella cellula contenente il DNA bersaglio d'interesse. Per esempio, se l'acido nucleico bersaglio desiderato è in una cellula umana, un polinucleotide con ottimizzazione di codone codificante Cas9 è contemplato per l'uso per produrre il polipeptide Cas9.

Complessi di un acido nucleico di bersagliamento o targeting genomico e un polipeptide sito-diretto

Un acido nucleico di bersagliamento o targeting genomico interagisce con un polipeptide sito-diretto (per esempio una nucleasi guidata da acido nucleico come Cas9), formando quindi un complesso. L'acido nucleico di bersagliamento o targeting genomico guida il polipeptide sito-diretto a un acido nucleico bersaglio

RNP

Il polipeptide sito-diretto e l'acido nucleico di bersagliamento o targeting genomico possono venire somministrati ciascuno separatamente a una cellula o a un paziente. D'altro canto, il polipeptide sito-diretto può venire pre-complessato con uno o più acidi nucleici di bersagliamento o targeting di genoma (RNA guida, sgRNA, o crRNA insieme con un tracrRNA). Il materiale pre-complessato può poi venire somministrato a una cellula o a un paziente. Tale materiale pre-complessato è noto come una particella ribonucleoproteica (RNP). Il polipeptide sito-diretto nella RNP può essere, per esempio, una endonucleasi Cas9 o una endonucleasi Cpf1. Il polipeptide sito-diretto può essere fiancheggiato all'estremità N, all'estremità C, o sia all'estremità N che

U001.D1.SM.61E

all'estremità C da uno o più segnali di localizzazione nucleare (NLS). Per esempio, una endonucleasi Cas9 può venire fiancheggiata da due NLS, un NLS situato all'estremità N e il secondo NLS situato all'estremità C. Il NLS può essere qualsiasi NLS noto nell'arte, come NLS SV40. Il rapporto in peso tra acido nucleico di bersagliamento o targeting genomico e polipeptide sito-diretto RNP può essere 1:1. Per esempio, il rapporto in peso tra sgRNA e endonucleasi Cas9 nella RNP può essere 1:1 per esempio, il sgRNA può comprendere la sequenza di acido nucleico di SEQ ID NO: 71.959, l'endonucleasi Cas9 può essere una Cas9 di *S. pyogenes* comprendente un NLS SV40 all'estremità N e un NLS SV40 all'estremità C, e il rapporto in peso tra sgRNA e endonucleasi Cas9 può essere di 1:1.

Acidi nucleici codificanti componenti di sistema

La presente descrizione divulga un acido nucleico comprendente una sequenza nucleotidica codificante un acido nucleico di bersagliamento o targeting genomico della descrizione, e/o qualsiasi acido nucleico o molecola proteica necessaria per realizzare gli aspetti dei metodi della descrizione.

L'acido nucleico codificante un acido nucleico di bersagliamento o targeting genomico della descrizione, un polipeptide sito-diretto della descrizione, e/o qualsiasi acido nucleico o molecola proteica necessaria per realizzare gli aspetti dei metodi della descrizione possono comprendere un vettore (per esempio un vettore di espressione ricombinante).

Il termine "vettore" si riferisce a una molecola di acido nucleico in grado di trasportare un altro acido nucleico a cui essa è stata legata. Un tipo di vettore è

U001.D1.SM.61E

un “plasmide”, che si riferisce a un loop di DNA a doppio filamento circolare in cui segmenti di acido nucleico possono essere legati. Un altro tipo di vettore è un vettore virale, in cui segmenti di acido nucleico supplementari possono essere legati nel genoma virale. Alcuni vettori sono capaci di replicazione autonoma in una cellula ospite in cui essi sono introdotti (per esempio vettori batterici aventi una origine batterica di replicazione e vettori di mammiferi episomiali). Altri vettori (per esempio vettori di mammiferi non episomiali) sono integrati nel genoma di una cellula ospite con introduzione nella cellula ospite, e quindi sono replicati insieme al genoma ospite.

In alcuni esempi vettori possono essere in grado di dirigere l’espressione di acidi nucleici a cui essi sono legati operativamente. Tali vettori sono indicati nel presente documento come “vettori di espressione ricombinante” o più semplicemente “vettori di espressione”, che adempiono funzioni equivalenti.

L’espressione “legato operativamente” significa che la sequenza nucleotidica d’interesse è legata a una o più sequenze regolatorie in una maniera che consente l’espressione della sequenza nucleotidica. L’espressione “sequenza regolatoria” è intesa includere, per esempio, promotori, potenziatori e altri elementi di controllo di espressione (per esempio segnali di poliadenilazione). Tali sequenze regolatorie sono ben note nell’arte e sono descritte, per esempio, in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Sequenze regolatorie includono quelle che dirigono l’espressione costitutiva di una sequenza nucleotidica in molti tipi di cellule ospiti, e quelle che dirigono l’espressione della sequenza nucleotidica

U001.D1.SM.61E

soltanto in alcune cellule ospiti (per esempio sequenze regolatorie tessuto-specifiche). Gli individui esperti nel ramo comprenderanno che la progettazione del vettore di espressione può dipendere da fattori quali la scelta della cellula bersaglio, il livello di espressione desiderato, e simili.

Vettori di espressione contemplati includono, ma non sono limitati a, vettori virali basati su virus vaccinico, poliovirus, adenovirus, virus adeno-associato, SV40, virus dell'herpes simplex, virus dell'immunodeficienza umana, retrovirus (per esempio virus della leucemia murina, virus della necrosi della milza, e vettori derivati da retrovirus come virus del sarcoma di Rous, virus del sarcoma di Harvey, virus della leucosi aviaria, un lentivirus, virus dell'immunodeficienza umana, virus del sarcoma mieloproliferativo, e virus del tumore mammario) e altri vettori ricombinanti. Altri vettori contemplati per cellule bersaglio eucariotiche includono, ma non sono limitati a, i vettori pXT1, pSG5, pSVK3, pBPV, pMSG, e pSVLSV40 (Pharmacia). Vettori supplementari contemplati per cellule bersaglio eucariotiche includono, ma non sono limitati a, i vettori pCTx-1, pCTx-2, e pCTx-3, che sono descritti nelle Figure da 1A a 1C. Altri vettori possono venire utilizzati purché essi siano compatibili con la cellula ospite.

In alcuni esempi un vettore può comprendere uno o più elementi di controllo di trascrizione e/o traduzione. A seconda del sistema ospite/vettore utilizzato, qualsiasi di un numero di elementi di controllo di trascrizione e traduzione adatti, inclusi promotori costitutivi e inducibili, elementi potenziatori di trascrizione, terminatori di trascrizione, etc, possono venire utilizzati nel vettore di espressione. Il vettore può essere un vettore autoinattivante che inattiva le

U001.D1.SM.61E

sequenze virali o i componenti del macchinario CRISPR o altri elementi.

Esempi non limitativi di promotori eucariotici adatti (vale a dire promotori funzionali in una cellula eucariotica) includono quelli da citomegalovirus (CMV) immediato precoce, timidina chinasi di virus dell'herpes simplex (HSV), SV40 precoce e avanzato, lunghe ripetizioni terminali (LTR) da retrovirus, promotore del fattore di allungamento umano-1 (EF1), un costrutto ibrido comprendente il potenziatore di citomegalovirus (CMV) fuso al promotore di beta-actina di pollo (CAG), promotore di virus di cellule staminali murine (MSCV), promotore di locus di fosfoglicerato chinasi-1 (PGK), e metallothioneina-I di topo.

Per l'espressione di piccoli RNA, inclusi RNA guida utilizzati congiuntamente a endonucleasi Cas, vari promotori come promotori di RNA polimerasi III, inclusi per esempio U6 e H1, possono essere vantaggiosi. Descrizioni di e parametri per potenziare l'uso di tali promotori sono noti nell'arte, e informazioni e approcci supplementari sono regolarmente descritti; si veda, per esempio, Ma, H. et al., *Molecular Therapy - Nucleic Acids* 3, e161 (2014) doi:10.1038/mtna.2014.12.

Il vettore di espressione può anche contenere un sito di legame a ribosoma per l'inizio di traduzione e un terminatore di trascrizione. Il vettore di espressione può anche comprendere sequenze appropriate per amplificare espressione. Il vettore di espressione può anche includere sequenze nucleotidiche codificanti tag non nativi (per esempio tag di istidina, tag di emagglutinina, proteina verde fluorescente, etc.) che sono fuse al polipeptide sito-diretto, dando così luogo a

U001.D1.SM.61E

una proteina di fusione.

Un promotore può essere un promotore inducibile (per esempio un promotore di shock termico, promotore regolato da tetraciclina, promotore regolato da steroide, promotore regolato da metallo, promotore regolato da recettore per estrogeni, etc.). Il promotore può essere un promotore costitutivo, (per esempio promotore di CMV, promotore di UBC). In alcuni casi il promotore può essere un promotore spazialmente ristretto e/o temporalmente ristretto (per esempio un promotore tessuto-specifico, un promotore specifico per tipo di cellula, etc.).

L'acido nucleico codificante un acido nucleico di bersagliamento o targeting genomico della descrizione e/o un polipeptide sito-diretto può essere impacchettato in o sulla superficie di veicoli di veicolazione per la veicolazione a cellule. Veicoli di veicolazione contemplati includono, ma non sono limitati a, nanosfere, liposomi, dot quantici, nanoparticelle, particelle di polietilenglicole, idrogel e micelle. Come descritto nell'arte, una varietà di frazioni di targeting o bersagliamento possono venire utilizzate per potenziare l'interazione preferenziale di tali veicoli con tipi di cellule desiderati o posizioni.

L'introduzione dei complessi, polipeptidi e acidi nucleici della descrizione in cellule può avvenire mediante infezione virale o batteriofagica, trasfezione, coniugazione, fusione protoplastica, lipofezione, elettroporazione, nucleofezione, precipitazione con fosfato di calcio, trasfezione mediata da polietilenimina (PEI), trasfezione mediata da DEAE-destrano, trasfezione mediata da liposoma, tecnologia di cannone particellare o "particle gun", precipitazione con fosfato di calcio, microiniezione diretta, veicolazione di acido

U001.D1.SM.61E

nucleico mediata da nanoparticelle, e simili.

Veicolazione

Polinucleotidi di RNA guida (RNA o DNA) e/o polinucleotide(i) di endonucleasi (RNA o DNA) possono venire veicolati mediante veicoli di veicolazione virali o non virali noti nell'arte, come elettroporazione, forza meccanica, deformazione cellulare (SQZ Biotech), e peptidi di penetrazione cellulare. In alternativa, polipeptide(i) di endonucleasi possono venire veicolati mediante veicoli di veicolazione virali o non virali noti nell'arte, come elettroporazione o nanoparticelle lipidiche. In altri aspetti alternativi, la DNA endonucleasi può essere veicolata come uno o più polipeptidi, da sola o pre-complettata con uno o più RNA guida, o uno o più crRNA insieme con un tracrRNA.

Elettroporazione è una tecnica di veicolazione in cui un campo elettrico è applicato a una o più cellule al fine di aumentare la permeabilità della membrana cellulare, che consente a sostanze come farmaci, acidi nucleici (acidi nucleici di bersagliamento o targeting genomico), proteine (polipeptidi sito-diretti) o RNP di venire introdotti nella cellula. In generale, elettroporazione opera passando migliaia di volt attraverso una distanza di uno o due millimetri di cellule sospese in una cuvetta di elettroporazione (1,0 – 1,5 kV, 250 - 750V/cm).

Polinucleotidi possono venire veicolati da veicoli di veicolazione non virali, inclusi, ma non limitati a, nanoparticelle, liposomi, ribonucleoproteine, peptidi caricati positivamente, coniugati di RNA a piccola molecola, chimere aptamero-RNA, e complessi di proteine di fusione-RNA. Alcuni veicoli di veicolazione

U001.D1.SM.61E

non virali sono descritti in Peer e Lieberman, *Gene Therapy*, 18: 1127-1133 (2011) (che si focalizza su veicoli di veicolazione non virali per siRNA che sono anche utili per la veicolazione di altri polinucleotidi).

Polinucleotidi, come RNA guida, sgRNA e mRNA codificanti una endonucleasi, possono venire veicolati a una cellula o a un paziente da una nanoparticella lipidica (LNP).

Una LNP si riferisce a qualsiasi particella avente un diametro inferiore a 1000 nm, 500 nm, 250 nm, 200 nm, 150 nm, 100 nm, 75 nm, 50 nm, o 25 nm. In alternativa, una nanoparticella può avere dimensioni variabili da 1 a 1000 nm, da 1 a 500 nm, da 1 a 250 nm, da 25 a 200 nm, da 25 a 100 nm, da 35 a 75 nm, o da 25 a 60 nm.

LNP possono essere costituite da lipidi cationici, anionici o neutri. Lipidi neutri, come il fosfolipide fusogenico DOPE o il componente di membrana colesterolo, possono venire inclusi in LNP come “lipidi aiutanti o helper” per potenziare l’attività di trasfezione e la stabilità nanoparticellare. Limitazioni di lipidi cationici includono bassa efficacia dovuta a una scarsa stabilità e una rapida clearance, nonché la generazione di risposte infiammatorie o antiinfiammatorie.

LNP possono anche essere costituite da lipidi idrofobi, lipidi idrofili o sia lipidi idrofobi che idrofili.

Qualsiasi lipide o combinazione di lipidi che sono noti nell’arte può venire utilizzato(a) per produrre una LNP. Esempi di lipidi utilizzati per produrre LNP sono: DOTMA, DOSPA, DOTAP, DMRIE, DC-colesterolo, DOTAP-

U001.D1.SM.61E

colesterolo, GAP-DMORIE-DPyPE, e GL67A-DOPE-DMPE-polietilenglicole (PEG). Esempi di lipidi cationici sono: 98N12-5, C12-200, DLin-KC2-DMA (KC2), DLin-MC3-DMA (MC3), XTC, MD1, e 7C1. Esempi di lipidi neutri sono: DPSC, DPPC, POPC, DOPE, e SM. Esempi di lipidi modificati da PEG sono: PEG-DMG, PEG-CerC14, e PEG-CerC20.

I lipidi possono essere combinati in qualsiasi numero di rapporti molari per produrre una LNP. In aggiunta, il polinucleotide o i polinucleotidi possono venire combinati con lipide(i) in un'ampia gamma di rapporti molari per produrre una LNP.

Come dichiarato precedentemente, il polipeptide sito-diretto e l'acido nucleico di bersagliamento o targeting genomico possono venire somministrati ciascuno separatamente a una cellula o a un paziente. D'altro canto, il polipeptide sito-diretto può venire pre-complessato con uno o più RNA guida, o uno o più crRNA insieme con un tracrRNA. Il materiale pre-complessato può poi venire somministrato a una cellula o a un paziente. Tale materiale pre-complessato è noto come una particella ribonucleoproteica (RNP).

RNA è in grado di formare interazioni specifiche con RNA o DNA. Mentre la proprietà è sfruttata in molti processi biologici, essa si accompagna anche al rischio di interazioni promiscue in un ambiente cellulare ricco di acidi nucleici. Una soluzione a questo problema è la formazione di particelle ribonucleoproteiche (RNP), in cui il RNA è pre-complessato con una endonucleasi. Un altro beneficio della RNP è la protezione del RNA da degradazione.

U001.D1.SM.61E

L'endonucleasi nella RNP può essere modificata o non modificata. Analogamente, il gRNA, crRNA, tracrRNA, o sgRNA può essere modificato o non modificato. Numerose modificazioni sono note nell'arte e possono essere utilizzate.

L'endonucleasi e il sgRNA possono venire generalmente combinati in un rapporto molare 1:1. In alternativa, l'endonucleasi, crRNA e tracrRNA possono venire generalmente combinati in un rapporto molare 1:1:1. Tuttavia, un'ampia gamma di rapporti molari può venire utilizzata per produrre una RNP.

Un vettore di virus adeno-associato (AAV) ricombinante può venire utilizzato per la veicolazione. Tecniche per produrre particelle di rAAV, in cui il genoma di AAV da impacchettare che include il polinucleotide da veicolare, geni rep e cap, e funzioni di virus helper sono fornite a una cellula sono standard nell'arte. La produzione di rAAV tipicamente richiede che i seguenti componenti siano presenti all'interno di una singola cellula (qui denotata come una cellula di impacchettamento): un genoma di rAAV, geni rep e cap di AAV separati dal (vale a dire non nel) genoma di rAAV e funzioni di virus helper. I geni rep e cap di AAV possono essere da qualsiasi sierotipo di AAV per cui virus ricombinante può essere derivato, e possono essere da un differente sierotipo di AAV rispetto a ITR di genoma di rAAV, inclusi, ma non limitati a, sierotipi di AAV AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-7, AAV-8, AAV-9, AAV-10, AAV-11, AAV-12, AAV-13 e AAV rh.74. La produzione di rAAV pseudotipato è descritta, per esempio, nella pubblicazione di domanda di brevetto internazionale WO 01/83692. Si veda la Tabella 2.

Tabella 2

Sierotipo di AAV	N° di accessione Genbank
AAV-1	NC 002077.1
AAV-2	NC 001401.2
AAV-3	NC 001729.1
AAV-3B	AF028705.1
AAV-4	NC 001829.1
AAV-5	NC 006152.1
AAV-6	AF028704.1
AAV-7	NC 006260.1
AAV-8	NC 006261.1
AAV-9	AX753250.1
AAV-10	AY631965.1
AAV-11	AY631966.1
AAV-12	DQ813647.1
AAV-13	EU285562.1

Un metodo di generazione di una cellula di impacchettamento implica la creazione di una linea cellulare che esprime stabilmente tutti i componenti necessari per la produzione di particelle di AAV. Per esempio, un plasmide (o molteplici plasmidi) comprendente(i) un genoma di rAAV privo di geni rep e cap di AAV, geni rep e cap di AAV separati dal genoma di rAAV, e un marcatore selezionabile come un gene di resistenza a neomicina, sono integrati nel genoma di una cellula. Genomi di AAV sono stati introdotti in plasmidi batterici mediante procedure come GC tailing (Samulski et al., 1982, Proc. Natl. Acad. S6. USA, 79:2077-2081), addizione di linker sintetici contenenti siti di clivaggio di endonucleasi di restrizione (Laughlin et al., 1983, Gene, 23:65-73) o mediante ligazione diretta di estremità smussate (Senapathy & Carter, 1984, J. Biol. Chem., 259:4661-4666). La linea di cellule di impacchettamento può poi venire infettata con un virus helper, come adenovirus. I vantaggi di questo metodo sono che le cellule sono selezionabili e sono adatte per la produzione su larga scala di rAAV. Altri esempi di metodi adatti impiegano adenovirus o

U001.D1.SM.61E

baculovirus, anziché plasmidi, per introdurre genomi di rAAV e/o geni rep e cap in cellule di impacchettamento.

Principi generali di produzione di rAAV sono esaminati in, per esempio, Carter, 1992, *Current Opinions in Biotechnology*, 1533-539; e Muzyczka, 1992, *Curr. Topics in Microbial. and Immunol.*, 158:97-129). Vari approcci sono descritti in Ratschin et al., *Mol. Cell. Biol.* 4:2072 (1984); Hermonat et al., *Proc. Natl. Acad. Sci. USA*, 81:6466 (1984); Tratschin et al., *Mol. Cell. Biol.* 5:3251 (1985); McLaughlin et al., *J. Virol.*, 62:1963 (1988); e Lebkowski et al., 1988 *Mol. Cell. Biol.*, 7:349 (1988). Samulski et al. (1989, *J. Virol.*, 63:3822-3828); Brevetto statunitense n° 5,173,414; WO 95/13365 e Brevetto statunitense corrispondente n° 5,658.776; WO 95/13392; WO 96/17947; PCT/US98/18600; WO 97/09441 (PCT/US96/ 14423); WO 97/08298 (PCT/US96/13872); WO 97/21825 (PCT/US96/20777); WO 97/06243 (PCT/FR96/01064); WO 99/11764; Perrin et al. (1995) *Vaccine* 13:1244-1250; Paul et al. (1993) *Human Gene Therapy* 4:609-615; Clark et al. (1996) *Gene Therapy* 3:1124-1132; Brevetto statunitense n° 5.786.211; Brevetto statunitense n° 5.871.982; e Brevetto Statunitense n° 6.258.595.

Sierotipi di vettore AAV possono venire abbinati a tipi di cellule bersaglio. Per esempio, i seguenti tipi di cellule esemplari possono venire trasdotti dai sierotipi AAV indicati, tra gli altri. Si veda la Tabella 3.

Tabella 3

Tipo di tessuto/cellula	Sierotipo
Fegato	AAV8, AAV3, AA5, AAV9
Muscolo scheletrico	AAV1, AAV7, AAV6, AAV8, AAV9
Sistema nervoso centrale	AAV5, AAV1, AAV4
RPE	AAV5, AAV4

U001.D1.SM.61E

Cellule fotorecetttrici	AAV5
Polmone	AAV9
Cuore	AAV8
Pancreas	AAV8
Rene	AAV2, AA8

Oltre a vettori virali adeno-associati, possono venire utilizzati altri vettori virali. Tali vettori virali includono, ma non sono limitati a, lentivirus, alfavirus, enterovirus, pestivirus, baculovirus, herpesvirus, virus Epstein Barr, papovavirus, poxvirus, virus vaccinico, e virus dell'herpes simplex.

In alcuni casi, Cas9 mRNA, sgRNA avente come bersaglio uno o due loci all'interno del o vicino al gene BCL11A o altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A, e DNA donatore possono venire formulati separatamente in nanoparticelle lipidiche, o sono tutti formulati in una nanoparticella lipidica.

In alcuni casi, Cas9 mRNA può venire formulato in una nanoparticella lipidica, mentre sgRNA e DNA donatore possono venire veicolati in un vettore AAV.

Opzioni sono disponibili per veicolare la nucleasi Cas9 come un plasmide a DNA, come mRNA o come una proteina. Il RNA guida può essere espresso dallo stesso DNA, oppure può anche venire veicolato come un RNA. Il RNA può essere chimicamente modificato per alterare o migliorare la sua emivita, o diminuire la probabilità o il grado di risposta immunitaria. La proteina endonucleasica può venire complessata con il gRNA prima della veicolazione. Vettori virali consentono una veicolazione efficiente; versioni separate di Cas9 e ortologhi più piccoli di Cas9 possono venire impacchettate in AAV, così come lo possono donatori per HDR. Esistono anche svariati metodi di veicolazione

U001.D1.SM.61E

non virali che possono veicolare ciascuno di questi componenti, oppure metodi non virali e virali possono venire impiegati in tandem. Per esempio, nanoparticelle possono venire utilizzate per veicolare la proteina e RNA guida, mentre AAV può venire utilizzato per veicolare un DNA donatore.

Cellule geneticamente modificate

L'espressione "cellula geneticamente modificata" si riferisce a una cellula che comprende almeno una modificazione genetica introdotta mediante editing genomico (per esempio utilizzando il sistema CRISPR/Cas9/Cpf1). In alcuni esempi *ex vivo* del presente documento, la cellula geneticamente modificata può essere una cellula progenitrice geneticamente modificata. In alcuni esempi *in vivo* del presente documento, la cellula geneticamente modificata può essere una cellula progenitrice ematopoietica geneticamente modificata. Nel presente documento è contemplata una cellula geneticamente modificata comprendente un acido nucleico di targeting genomico esogeno e/o un acido nucleico esogeno codificante un acido nucleico di bersagliamento o targeting genomico.

L'espressione "popolazione trattata con controllo" descrive una popolazione di cellule che è stata trattata con mezzi, induzione virale, sequenze di acidi nucleici, temperatura, confluenza, dimensione di beute, pH, etc., identici con l'eccezione dell'aggiunta dei componenti di editing genomico. Qualsiasi metodo noto nell'arte può venire utilizzato per misurare la modulazione o inattivazione della sequenza di controllo di trascrizione del gene BCL11A o l'espressione o attività proteica, per esempio analisi Western Blot della sequenza di controllo di trascrizione del gene BCL11A o la quantificazione della sequenza di controllo di

U001.D1.SM.61E

trascrizione del mRNA di gene BCL11A.

L'espressione "cellula isolata" si riferisce a una cellula che è stata rimossa da un organismo in cui essa era originariamente presente, o un discendente di una tale cellula. Opzionalmente, la cellula può venire coltivata *in vitro*, per esempio sotto condizioni definite in presenza di altre cellule. Opzionalmente la cellula può venire successivamente introdotta in un secondo organismo o reintrodotta nell'organismo da cui essa (o la cellula da cui essa è discesa) è stata isolata.

L'espressione "popolazione isolata" in merito a una popolazione isolata di cellule si riferisce a una popolazione di cellule che è stata rimossa e separata da una popolazione mista o eterogenea di cellule. In alcuni casi la popolazione isolate può essere una popolazione sostanzialmente pura di cellule, rispetto alla popolazione eterogenea da cui le cellule sono state isolate o arricchite. In alcuni casi la popolazione isolate può essere una popolazione isolata di cellule progenitrici umane, per esempio una popolazione sostanzialmente pura di cellule progenitrici umane, rispetto a una popolazione eterogenea di cellule comprendente cellule progenitrici umane e cellule da cui le cellule progenitrici umane sono state derivate.

L'espressione "sostanzialmente potenziata" con riferimento a una particolare popolazione di cellule, si riferisce a una popolazione di cellule in cui l'occorrenza di un particolare tipo di cellula è aumentata rispetto a livelli pre-esistenti o di riferimento, di almeno 2 volte, almeno 3, almeno 4, almeno 5, almeno 6, almeno 7, almeno 8, almeno 9, almeno 10, almeno 20, almeno 50, almeno 100, almeno 400, almeno 1000, almeno 5000, almeno 20000, almeno

U001.D1.SM.61E

100000 o più volte a seconda, per esempio, dei livelli desiderati di tali cellule per migliorare una emoglobinopatia.

L'espressione "sostanzialmente arricchita" con riferimento a una particolare popolazione di cellule si riferisce a una popolazione di cellule che è almeno circa 10%, circa 20%, circa 30%, circa 40%, circa 50%, circa 60%, circa 70% o più con riferimento alle cellule che costituiscono una popolazione totale di cellule.

L'espressione "sostanzialmente pura" con riferimento a una particolare popolazione di cellule si riferisce a una popolazione di cellule che è pura almeno circa al 75%, almeno circa all'85%, almeno circa al 90%, o almeno circa al 95%, con riferimento alle cellule che costituiscono una popolazione totale di cellule. Vale a dire, le espressioni "sostanzialmente pura" o "essenzialmente purificata" con riferimento a una popolazione di cellule progenitrici si riferisce a una popolazione di cellule che contengono meno di circa il 20%, circa il 15%, circa il 10%, circa il 9%, circa l'8%, circa il 7%, circa il 6%, circa il 5%, circa il 4%, circa il 3%, circa il 2%, circa l'1%, o meno dell'1%, di cellule che non sono cellule progenitrici come definite dai termini del presente documento.

Differenziazione di iPSC sottoposte a editing genomico in cellule progenitrici ematopoietiche

Un'altra fase dei metodi *ex vivo* della presente descrizione può comprendere la differenziazione delle iPSC sottoposte a editing genomico in cellule progenitrici ematopoietiche. La fase di differenziazione può venire effettuata secondo qualsiasi metodo noto nell'arte.

U001.D1.SM.61E

Differenziazione di cellule staminali mesenchimali sottoposte a editing genomico in cellule progenitrici ematopoietiche

Un'altra fase dei metodi *ex vivo* della presente descrizione può comprendere la differenziazione delle cellule staminali mesenchimali sottoposte a editing genomico in cellule progenitrici ematopoietiche. La fase di differenziazione può venire effettuata secondo qualsiasi metodo noto nell'arte.

Impiantazione di cellule in pazienti

Un'altra fase dei metodi *ex vivo* della presente descrizione può comprendere l'impiantazione delle cellule in pazienti. La fase di impiantazione può venire effettuata utilizzando qualsiasi metodo di impiantazione noto nell'arte. Per esempio, le cellule geneticamente modificate possono venire iniettate direttamente nel sangue del paziente o altrimenti somministrate al paziente. Le cellule geneticamente modificate possono venire purificate *ex vivo* utilizzando un marcatore selezionato.

Carrier farmaceuticamente accettabili

I metodi *ex vivo* di somministrazione di cellule progenitrici a un soggetto contemplati nel presente documento possono implicare l'uso di composizioni terapeutiche comprendenti cellule progenitrici.

Composizioni terapeutiche possono contenere un carrier fisiologicamente tollerabile insieme alla composizione cellulare, e opzionalmente almeno un agente bioattivo supplementare come descritto nel presente documento, disciolto o disperso all'interno come un principio attivo. In alcuni casi la composizione terapeutica non è sostanzialmente immunogena quando somministrata a un

U001.D1.SM.61E

mammifero o paziente umano per scopi terapeutici, a meno che non lo si desideri.

In generale, le cellule progenitrici descritte nel presente documento possono venire somministrate come una sospensione con un carrier farmaceuticamente accettabile. Un individuo esperto nel ramo riconoscerà che un carrier farmaceuticamente accettabile da utilizzare in una composizione cellulare non includerà tamponi, composti, agenti di crioconservazione, conservanti o altri agenti in quantità che interferiscono sostanzialmente con la vitalità delle cellule da veicolare al soggetto. Una formulazione comprendente cellule può includere, per esempio, tamponi osmotici che permettono di mantenere l'integrità della membrana cellulare e, opzionalmente, nutrienti per mantenere la vitalità cellulare o potenziare l'attecchimento o engraftment durante la somministrazione. Tali formulazioni e sospensioni sono note agli individui esperti nel ramo e/o possono venire adattate per l'uso con le cellule progenitrici, come descritto nel presente documento, utilizzando sperimentazione ordinaria.

Una composizione cellulare può anche venire emulsionata o presentata come una composizione liposomica, purché la procedura di emulsificazione non influisca negativamente sulla vitalità cellulare. Le cellule e qualsiasi altro principio attivo possono venire miscelati con eccipienti che sono farmaceuticamente accettabili e compatibili con il principio attivo, e in quantità adatte per l'uso nei metodi terapeutici descritti nel presente documento.

Agenti aggiuntivi inclusi in una composizione cellulare possono includere sali farmaceuticamente accettabili dei componenti presenti. Sali farmaceuticamente

U001.D1.SM.61E

accettabili includono i sali di addizione a acidi (formati con i gruppi amminici liberi del polipeptide) che sono formati con acidi inorganici, come, per esempio, acido cloridrico o acido fosforico, o acidi organici come acido acetico, tartarico, mandelico e simili. Sali formati con i gruppi carbossilici liberi possono anche essere derivati da basi inorganiche come, per esempio, idrossidi di sodio, potassio, ammonio, calcio o ferro, e basi organiche come isopropilammina, trietilammina, 2-etilammino etanolo, istidina, procaina e simili.

Carrier fisiologicamente tollerabili sono ben noti nell'arte. Carrier liquidi esemplari sono soluzioni acquose sterili che non contengono alcun materiale in aggiunta ai principi attivi e acqua, o contengono un tampone come fosfato di sodio a valore di pH fisiologico, soluzione salina fisiologica o entrambi, come una soluzione salina tamponata con fosfato. Inoltre, carrier acquosi possono contenere più di un sale tampone, nonché sali come cloruri di sodio e potassio, destrosio, glicole polietilenico e altri soluti. Composizioni liquide possono anche contenere fasi liquide in aggiunta e all'esclusione di acqua. Esempi di tali fasi liquide addizionali sono glicerina, oli vegetali come olio di semi di cotone e emulsioni acqua-olio. La quantità di un composto attivo utilizzato nelle composizioni cellulari che è efficace nel trattamento di un particolare disturbo o condizione può dipendere dalla natura del disturbo o della condizione, e può venire determinata mediante tecniche cliniche standard.

Somministrazione e efficacia

I termini "somministrazione", "introduzione" e "trapianto" sono utilizzati in modo intercambiabile nel contesto del posizionamento di cellule, per esempio

U001.D1.SM.61E

cellule progenitrici, in un soggetto, mediante un metodo o una via che porta ad una localizzazione almeno parziale delle cellule introdotte in un sito desiderato, come un sito di lesione o riparazione, in modo tale da produrre uno o più effetti desiderati. Le cellule, per esempio cellule progenitrici, o loro progenie differenziata, possono venire somministrate mediante qualsiasi via appropriata che porta a una veicolazione a una posizione desiderata nel soggetto dove almeno una porzione delle cellule impiantate o componenti delle cellule rimangono vitali. Il periodo di vitalità delle cellule dopo somministrazione a un soggetto può essere da poche ore, per esempio ventiquattro ore, a qualche giorno, fino ad alcuni anni o persino per la durata di vita del paziente, vale a dire attecchimento o engraftment a lungo termine. Per esempio in alcuni aspetti descritti nel presente documento, una quantità efficace di cellule progenitrici miogeniche è somministrata tramite una via sistemica di somministrazione, come una via intraperitoneale o endovenosa.

I termini “individuo”, “soggetto”, “ospite” e “paziente” sono utilizzati in modo intercambiabile nel presente documento e si riferiscono a qualsiasi soggetto per cui una diagnosi, un trattamento o una terapia sono desiderati. In alcuni aspetti il soggetto è un mammifero. In alcuni aspetti il soggetto è un essere umano.

Quando fornite profilatticamente, cellule progenitrici descritte nel presente documento possono venire somministrate anticipatamente a un soggetto prima di qualsiasi sintomo di una emoglobinopatia, per esempio prima dello sviluppo di stanchezza, respiro affannoso, ittero, pubertà ritardata a crescita lenta, dolore alle articolazioni, alle ossa e al petto, milza e fegato ingrossati. Pertanto, la

U001.D1.SM.61E

somministrazione profilattica di una popolazione di cellule progenitrici ematopoietiche serve a prevenire una emoglobinopatia, come β -talassemia o anemia falciforme.

Quando fornite terapeuticamente, cellule progenitrici ematopoietiche sono fornite all'insorgenza (o dopo) di un sintomo o indicazione di emoglobinopatia, per esempio all'insorgenza della malattia.

La popolazione di cellule progenitrici ematopoietiche da somministrare secondo i metodi descritti nel presente documento può comprendere cellule progenitrici ematopoietiche allogeniche ottenute da uno o più donatori. "Allogenica" si riferisce a una cellula progenitrice ematopoietica o a campioni biologici comprendenti cellule progenitrici ematopoietiche ottenute da uno o più donatori differenti della stessa specie, dove i geni a livello di uno o più loci non sono identici. Per esempio, una popolazione di cellule progenitrici ematopoietiche da somministrare a un soggetto può essere derivata da uno o più soggetti donatori non parenti, o da uno o più fratelli non identici. In alcuni casi possono venire utilizzate popolazioni di cellule progenitrici ematopoietiche singeniche, come quelle ottenute da animali geneticamente identici o da gemelli identici. Le cellule progenitrici ematopoietiche possono essere cellule autologhe; vale a dire, le cellule progenitrici ematopoietiche sono ottenute o isolate da un soggetto e somministrate allo stesso soggetto, vale a dire il donatore e il ricevente sono la stessa persona.

L'espressione "quantità efficace" si riferisce alla quantità di una popolazione di cellule progenitrici o loro progenie necessaria per impedire o alleviare almeno

U001.D1.SM.61E

uno o più segni o sintomi di una emoglobinopatia, e si riferisce a una quantità sufficiente di una composizione per fornire l'effetto desiderato, per esempio per trattare un soggetto affetto da una emoglobinopatia. L'espressione "quantità terapeuticamente efficace" si riferisce pertanto a una quantità di cellule progenitrici o di una composizione comprendente cellule progenitrici che è sufficiente per promuovere un particolare effetto quando somministrata a un comune soggetto, come uno che è affetto da una emoglobinopatia o è a rischio per una emoglobinopatia. Una quantità efficace includerebbe anche una quantità sufficiente a impedire o ritardare lo sviluppo di un sintomo della malattia, dopo il decorso di un sintomo della malattia (per esempio, ma senza che ciò costituisca una limitazione, rallentare la progressione di un sintomo della malattia), o annullare un sintomo della malattia. Resta sottinteso che, per qualsiasi caso dato, una "quantità efficace" appropriata può venire determinata da un individuo di comune esperienza nel ramo utilizzando una sperimentazione ordinaria.

Per l'uso nei vari aspetti descritti nel presente documento, una quantità efficace di cellule progenitrici comprende almeno 10^2 cellule progenitrici, almeno 5×10^2 cellule progenitrici, almeno 10^3 cellule progenitrici, almeno 5×10^3 cellule progenitrici, almeno 10^4 cellule progenitrici, almeno 5×10^4 cellule progenitrici, almeno 10^5 cellule progenitrici, almeno 2×10^5 cellule progenitrici, almeno 3×10^5 cellule progenitrici, almeno 4×10^5 cellule progenitrici, almeno 5×10^5 cellule progenitrici, almeno 6×10^5 cellule progenitrici, almeno 7×10^5 cellule progenitrici, almeno 8×10^5 cellule progenitrici, almeno 9×10^5 cellule

U001.D1.SM.61E

progenitrici, almeno 1×10^6 cellule progenitrici, almeno 2×10^6 cellule progenitrici, almeno 3×10^6 cellule progenitrici, almeno 4×10^6 cellule progenitrici, almeno 5×10^6 cellule progenitrici, almeno 6×10^6 cellule progenitrici, almeno 7×10^6 cellule progenitrici, almeno 8×10^6 cellule progenitrici, almeno 9×10^6 cellule progenitrici, o multipli di esse. Le cellule progenitrici possono essere derivate da uno o più donatori, oppure possono essere ottenute da una fonte autologa. In alcuni esempi descritti nel presente documento le cellule progenitrici possono venire espanse in coltura prima della somministrazione a un soggetto che ne ha necessità.

“Somministrare” si riferisce alla veicolazione di una composizione di cellule progenitrici in un soggetto mediante un metodo o una via che porta a una localizzazione almeno parziale della composizione cellulare a livello di un sito desiderato. Una composizione cellulare può venire somministrata attraverso qualsiasi via appropriata che porta a un trattamento efficace nel soggetto, vale a dire la somministrazione dà luogo a una veicolazione a una posizione desiderata nel soggetto dove almeno una porzione della composizione veicolata, vale almeno 1×10^4 cellule, è veicolata al sito desiderato per un periodo di tempo. Modi di somministrazione includono iniezione, infusione, instillazione o ingestione. “Iniezione” include, senza limitazione, iniezione e infusione endovenosa, intramuscolare, intraarteriosa, intratecale, intraventricolare, intracapsulare, intraorbitale, intracardiaca, intradermica, intraperitoneale, transtracheale, sottocutanea, subcuticolare, intraarticolare, subcapsulare, subaracnoide, intraspinale, intracerebrospinale e intrasternale. In alcuni esempi

U001.D1.SM.61E

la via è endovenosa. Per la veicolazione di cellule può venire prodotta una somministrazione mediante iniezione o infusione.

Le cellule possono venire somministrate sistemicamente. Le espressioni “somministrazione sistemica, “sommministrata sistemicamente”, “somministrazione periferica” e “sommministrata perifericamente” si riferiscono alla somministrazione di una popolazione di cellule progenitrici differente da direttamente in un sito, tessuto o organo bersaglio, dal momento che essa entra, invece, nel sistema circolatorio del soggetto e, quindi, è soggetta a metabolismo e altri processi simili.

L'efficacia di un trattamento comprendente una composizione per il trattamento di emoglobinopatie può venire determinata dal clinico esperto. Tuttavia, un trattamento è considerato “trattamento efficace” se uno qualsiasi dei o tutti i segni o sintomi di, a titolo di esempio, livelli di BCL11A funzionale e HbF funzionale sono alterati in una maniera benefica (per esempio diminuiti almeno del 10% per BCL11A e/o aumentati almeno del 10% per HbF), o altri sintomi clinicamente accettati o marcatori di malattia sono migliorati o alleviati. L'efficacia può anche venire misurata dall'incapacità di un individuo a peggiorare quando valutato da ospedalizzazione o la necessità di interventi medici (per esempio la progressione della malattia è arrestata o almeno rallentata). Metodi di misurazione di questi indicatori sono noti agli esperti nel ramo e/o descritti nel presente documento. Un trattamento include qualsiasi trattamento di una malattia in un individuo o un animale (alcuni esempi non limitativi includono un essere umano, o un mammifero) e include: (1)

U001.D1.SM.61E

l'inibizione della malattia, per esempio l'arresto o il rallentamento della progressione di sintomi; oppure (2) l'alleviamento della malattia, per esempio provocando una regressione di sintomi; e (3) la prevenzione o la riduzione della probabilità dello sviluppo di sintomi.

Il trattamento secondo la presente invenzione può migliorare uno o più sintomi associati con emoglobinopatie diminuendo la quantità di BCL11A funzionale e/o aumentando la quantità di HbF funzionale nell'individuo. Segni precoci tipicamente associati con emoglobinopatie includono, per esempio, stanchezza, respiro affannoso, ittero, pubertà ritardata a crescita lenta, dolore alle articolazioni, alle ossa e al petto, milza e fegato ingrossati.

Kit

La presente descrizione divulga kit per eseguire i metodi descritti nel presente documento. Un kit può includere uno o più di un acido nucleico di bersagliamento o targeting genomico, un polinucleotide codificante un acido nucleico di bersagliamento o targeting genomico, un polipeptide sito-diretto, un polinucleotide codificante un polipeptide sito-diretto, e/o qualsiasi acido nucleico o molecola proteica necessaria per effettuare gli aspetti dei metodi descritti nel presente documento, o qualsiasi combinazione di essi.

Un kit può comprendere: (1) un vettore comprendente una sequenza nucleotidica codificante un acido nucleico di bersagliamento o targeting genomico, (2) il polipeptide sito-diretto o un vettore comprendente una sequenza nucleotidica codificante il polipeptide sito-diretto, e (3) un reagente per ricostituzione e/o diluizione del o dei vettori e/o polipeptide.

U001.D1.SM.61E

Un kit può comprendere: (1) un vettore comprendente (i) una sequenza nucleotidica codificante un acido nucleico di targeting o bersagliamento genomico, e (ii) una sequenza nucleotidica codificante il polipeptide sito-diretto; e (2) un reagente per ricostituzione e/o diluizione del vettore.

In qualsiasi dei kit menzionati sopra il kit può comprendere un acido nucleico guida monomolecolare di bersagliamento o targeting genomico. In qualsiasi dei kit menzionati sopra, il kit può comprendere un acido nucleico a doppia molecola di bersagliamento o targeting genomico. In qualsiasi dei kit menzionati sopra, il kit può comprendere due o più guide a doppia molecola o guide monomolecolari. Il kit può comprendere un vettore che codifica l'acido nucleico di bersagliamento o targeting di acido nucleico.

In qualsiasi dei kit menzionati sopra il kit può inoltre comprendere un polinucleotide da inserire per effettuare la modificazione genetica desiderata.

Componenti di un kit possono essere in contenitori separati, oppure combinati in un singolo contenitore.

Qualsiasi kit descritto sopra può inoltre comprendere uno o più reagenti supplementari, dove tali reagenti supplementari sono scelti tra un tampone, un tampone per introdurre un polipeptide o polinucleotide in una cellula, un tampone di lavaggio, un reagente di controllo, un vettore di controllo, un polinucleotide di RNA di controllo, un reagente per la produzione *in vitro* del polipeptide da DNA, adattori per sequenziamento e simili. Un tampone può essere un tampone di stabilizzazione, un tampone di ricostituzione, un tampone di diluizione, o simile. Un kit può anche comprendere uno o più componenti che

U001.D1.SM.61E

possono venire utilizzati per facilitare o potenziare il legame on-target o il clivaggio di DNA mediante l'endonucleasi, o migliorare la specificità di bersagliamento o targeting.

In aggiunta ai componenti menzionati sopra, un kit può inoltre comprendere istruzioni per utilizzare i componenti del kit per effettuare i metodi. Le istruzioni per eseguire i metodi possono venire registrate su un supporto di registrazione adatto. Per esempio, le istruzioni possono essere stampate su un substrato, come carta o plastica, etc. Le istruzioni possono essere presenti nei kit come un inserto di confezione, nell'etichettatura del contenitore del kit o componenti di esso (vale a dire associate all'imballaggio o al sottoimballaggio), etc. Le istruzioni possono essere presenti come un file di dati di archiviazione elettronica presente su un supporto di archiviazione leggibile da computer, per esempio CD-ROM, floppy disk, flash drive, etc. In alcuni casi le reali istruzioni non sono presenti nel kit, ma possono essere forniti mezzi per ottenere le istruzioni da una fonte remota (per esempio tramite Internet). Un esempio di questo caso è un kit che comprende un indirizzo internet dove le istruzioni possono venire visualizzate e/o da cui le istruzioni possono venire scaricate. Come con le istruzioni, questo mezzo per ottenere le istruzioni può venire registrato su un substrato adatto.

Formulazione di RNA guida

RNA guida della presente descrizione possono venire formulati con eccipienti farmaceuticamente accettabili come carrier, solventi, stabilizzanti, adiuvanti, diluenti, etc., a seconda del particolare modo di somministrazione e della forma di dosaggio. Composizioni di RNA guida possono venire formulate per ottenere

U001.D1.SM.61E

un pH fisiologicamente compatibile, e variano da un pH di circa 3 a un pH di circa 11, da circa pH 3 a circa pH 7, a seconda della formulazione e della via di somministrazione. In alcuni casi il pH può venire regolato a un intervallo da circa pH 5,0 a circa pH 8. In alcuni casi le composizioni possono comprendere una quantità terapeuticamente efficace di almeno un composto come descritto nel presente documento, insieme con uno o più eccipienti farmaceuticamente accettabili. Opzionalmente, le composizioni possono comprendere una combinazione dei composti descritti nel presente documento, oppure possono includere un secondo principio attivo utile nel trattamento o nella prevenzione di crescita batterica (per esempio, e senza limitazione, agenti antibatterici o antimicrobici), oppure possono includere una combinazione di reagenti della presente descrizione.

Eccipienti adatti includono, per esempio, molecole di trasporto o carrier che includono grandi macromolecole metabolizzate lentamente come proteine, polisaccaridi, acidi polilattici, acidi poliglicolici, amminoacidi polimerici, copolimeri di amminoacidi, e particelle virali inattive. Altri eccipienti esemplari possono includere antiossidanti (per esempio, e senza limitazione, acido ascorbico), agenti chelanti (per esempio, e senza limitazione, EDTA), carboidrati (per esempio, e senza limitazione, destrina, idrossialchilcellulosa e idrossialchilmetilcellulosa), acido stearico, liquidi (per esempio, e senza limitazione, oli, acqua, soluzione salina, glicerolo e etanolo), agenti umettanti o emulsionanti, sostanze di tamponamento del pH, e simili

Altri approcci terapeutici possibili

U001.D1.SM.61E

Editing genico può venire condotto utilizzando nucleasi ingegnerizzate per bersagliare sequenze specifiche. Ad oggi esistono quattro tipi maggiori di nucleasi: meganucleasi e loro derivati, nucleasi “zinc finger” o “a dito di zinco” (ZFN), nucleasi effettrici simili a fattori di trascrizione (TALEN), e sistemi di nucleasi CRISPR-Cas9. Le piattaforme nucleasiche variano nella difficoltà di design, nella densità di bersagliamento o targeting e nel modo di azione, in particolare quando la specificità di ZFN e TALEN è attraverso interazioni proteina-DNA, mentre interazioni RNA-DNA primariamente guidano Cas9. Il clivaggio di Cas9 richiede anche un motivo adiacente, il PAM, che differisce tra differenti sistemi di CRISPR. Cas9 da *Streptococcus pyogenes* cliva utilizzando un NGG PAM, CRISPR da *Neisseria meningitidis* può clivare a livello di siti con PAM, inclusi NNNNGATT, NNNNNGTTT e NNNNGCTT. Un numero di altri ortologi di Cas bersagliano protospaziatori adiacenti a PAM alternativi.

Endonucleasi CRISPR, come Cas9, possono venire utilizzate nei metodi della presente descrizione. Tuttavia, gli insegnamenti descritti nel presente documento, come siti bersaglio terapeutici, potrebbero venire applicati a altre forme di endonucleasi, come ZFN, TALEN, HE, o MegaTAL, oppure utilizzando combinazioni di nucleasi. Tuttavia, al fine di applicare gli insegnamenti della presente descrizione a tali endonucleasi, si necessiterebbe, tra le altre cose, di proteine ingegnerizzate dirette ai siti bersaglio specifici.

Domini di legame addizionali possono venire fusi alla proteina Cas9 per aumentare la specificità. I siti bersaglio di questi costrutti mapperebbero il sito specificato di gRNA identificato, ma richiederebbero motivi di legame

U001.D1.SM.61E

supplementari, come per un dominio “zinc finger” o “a dito di zinco”. Nel caso di Mega-TAL, una meganucleasi può essere fusa a un dominio di legame a DNA di TALE. Il dominio di meganucleasi può aumentare la specificità e fornire il clivaggio. Similmente, Cas9 inattivata o morta (dCas9) può essere fusa a un dominio di clivaggio e richiedere il sito bersaglio di sgRNA/Cas9 e il sito di legame adiacente per il dominio di legame a DNA fuso. Questo probabilmente richiederebbe una qualche ingegnerizzazione proteica della dCas9, in aggiunta all’inattivazione catalitica, per diminuire il legame senza il sito di legame addizionale.

Nucleasi “zinc finger” o “a dito di zinco”

Nucleasi zinc finger o “a dito di zinco” (ZFN) sono proteine modulari costituite da un dominio di legame a DNA zinc finger ingegnerizzato legato al dominio catalitico dell’endonucleasi FokI di tipo II. Poiché FokI funziona soltanto come un dimero, una coppia di ZFN deve essere ingegnerizzata per legarsi a sequenze di “emisisito” bersaglio cognate su filamenti di DNA opposti e con un preciso distanziamento tra esse per consentire al dimero FokI cataliticamente attivo di formarsi. Alla dimerizzazione del dominio FokI, che non ha di per sé alcuna specificità di sequenza, una rottura a doppio filamento di DNA è generata tra gli emisiti di ZFN come fase di inizio nell’editing genomico.

Il dominio di legame a DNA di ciascun ZFN è tipicamente costituito da 3-6 dita di zinco o zinc finger dell’architettura abbondante di Cys2-His2, con ciascun dito che riconosce primariamente una tripletta di nucleotidi su un

U001.D1.SM.61E

filamento della sequenza di DNA bersaglio, sebbene possa anche essere importante una interazione tra filamenti con un quarto nucleotide. L'alterazione degli amminoacidi di un dito in posizioni che producono contatti chiave con il DNA altera la specificità di sequenza di un dato dito. Quindi, una proteina a dito di zinco a quattro dita riconoscerà selettivamente una sequenza bersaglio di 12 bp, dove la sequenza bersaglio è un composto delle preferenze di tripletta contribute da ciascun dito, sebbene la preferenza di tripletta può essere influenzata a vari gradi da dita adiacenti. Un importante aspetto di ZFN è che esse possono essere prontamente ri-bersagliate pressoché a qualsiasi indirizzo genomico semplicemente modificando dita individuali, sebbene competenze notevoli siano richieste per fare bene questo. Nella maggior parte delle applicazioni di ZFN sono utilizzate proteine di 4-6 dita, che riconoscono 12-18 bp rispettivamente. Quindi, una coppia di ZFN riconoscerà tipicamente una sequenza bersaglio combinata di 24-36 bp, non includendo lo spaziatore tipico di 5-7 bp tra emisiti. I siti di legame possono essere separati ulteriormente con spaziatori più grandi, inclusi 15-17 bp. Una sequenza bersaglio di questa lunghezza è probabile che sia unica nel genoma umano, ipotizzando che sequenze ripetitive o omologhi genici sono esclusi durante il processo di design. Ciononostante, le interazioni proteina-DNA di ZFN non sono assolute nella loro specificità così si verificano eventi di legame e clivaggio off-target, come un eterodimero tra le due ZFN, o come un omodimero di una o l'altra delle ZFN. Quest'ultima possibilità è stata effettivamente eliminata ingegnerizzando l'interfaccia di dimerizzazione del dominio FokI per creare varianti "più" e

U001.D1.SM.61E

“meno”, note anche come varianti eterodimeriche obbligate, che possono soltanto dimerizzare l’una con l’altra e non con se stesse. Forzare l’eterodimero obbligato previene la formazione dell’omodimero. Ciò ha notevolmente potenziato la specificità di ZFN, nonché qualsiasi altra nucleasi che adotta queste varianti di FokI.

Una varietà di sistemi basati su ZFN sono stati descritti nell’arte, loro modificazioni sono regolarmente riportate, e numerosi riferimenti descrivono regole e parametri che sono utilizzati per guidare la progettazione di ZFN; si vedano, per esempio, Segal et al., Proc Natl Acad Sci USA 96(6):2758-63 (1999); Dreier B et al., J Mol Biol. 303(4):489-502 (2000); Liu Q et al., J Biol Chem. 277(6):3850-6 (2002); Dreier et al., J Biol Chem 280(42):35588-97 (2005); e Dreier et al., J Biol Chem. 276(31):29466-78 (2001).

Nucleasi effettrici simili a attivatori di trascrizione (TALEN)

TALEN rappresentano un altro formato di nucleasi modulari tramite cui, come con ZFN, un dominio di legame a DNA ingegnerizzato è legato al dominio nucleasico FokI, e una coppia di TALEN operano in tandem per ottenere un clivaggio di DNA mirato. La maggiore differenza da ZFN è la natura del dominio di legame a DNA e le proprietà di riconoscimento di sequenze di DNA bersaglio associate. Il dominio di legame a DNA di TALEN deriva da proteine TALE, che sono state originariamente descritte nel patogeno batterico vegetale *Xanthomonas sp.* TALE sono costituite da array tandem di ripetizioni di 33-35 amminoacidi, ciascuna ripetizione riconoscendo una singola coppia di basi nella sequenza di DNA bersaglio che ha tipicamente una lunghezza fino a 20 bp,

U001.D1.SM.61E

dando una lunghezza di sequenza bersaglio di fino a 40 bp. La specificità nucleotidica di ciascuna ripetizione è determinata dal diresiduo variabile da ripetere (RVD), che include soltanto due amminoacidi alle posizioni 12 e 13. Le basi guanina, adenina, citosina e timina sono riconosciute in modo predominante dai quattro RVD: Asn-Asn, Asn-Ile, His-Asp e Asn-Gly, rispettivamente. Questo costituisce un codice di riconoscimento molto più semplice rispetto a per dita di zinco, e quindi rappresenta un vantaggio rispetto alle ultime per design nucleasico. Tuttavia, come con ZFN, le interazioni proteina-DNA di TALEN non sono assolute nella loro specificità, e TALEN hanno anch'esse beneficiato dell'uso di varianti eterodimeriche obbligate del dominio FokI per ridurre attività off-target.

Sono state create varianti addizionali del dominio FokI che sono deattivate nella loro funzione catalitica. Se metà di una TALEN o una coppia di ZFN contiene un dominio FokI inattivo, allora soltanto clivaggio (nicking) di DNA a singolo filamento avrà luogo a livello del sito bersaglio, anziché una DSB. L'esito è paragonabile all'uso di mutanti di "nickasi" di CRISPR/Cas9/CpfI in cui uno dei domini di clivaggio di Cas9 è stato deattivato. Nick di DNA possono essere utilizzati per guidare editing genomico mediante HDR, ma a una efficienza inferiore rispetto a con una DSB. Il principale beneficio è che nick off-target sono rapidamente e accuratamente riparati, a differenza della DSB, che è incline a riparazione errata mediata da NHEJ.

Vari sistemi basati su TALEN sono stati descritti nell'arte, e loro modificazioni sono regolarmente riportate; si vedano, per esempio, Boch, Science

U001.D1.SM.61E

326(5959):1509-12 (2009); Mak et al., *Science* 335(6069):716-9 (2012); e Moscou et al., *Science* 326(5959):1501 (2009). L'uso di TALEN basati sulla piattaforma "Golden Gate", o schema di clonaggio, è stato descritto da molteplici gruppi; si vedano, per esempio, Cermak et al., *Nucleic Acids Res.* 39(12):e82 (2011); Li et al., *Nucleic Acids Res.* 39(14):6315-25(2011); Weber et al., *PLoS One.* 6(2):e16765 (2011); Wang et al., *J Genet Genomics* 41(6):339-47, Epub 2014 May 17 (2014); e Cermak T et al., *Methods MolBiol.* 1239:133-59 (2015).

Endonucleasi di inserimento o homing

Endonucleasi di inserimento o homing (HE) sono endonucleasi sequenza-specifiche che hanno lunghe sequenze di riconoscimento (14-44 paia di basi) e clivano DNA con elevata specificità – spesso in siti unici nel genoma. Esistono almeno sei famiglie note di HE come classificate dalla loro struttura, incluse LAGLIDADG (SEQ ID NO. 71,949), GIY-YIG, His-Cis box, H-N-H, PD-(D/E)xK, e Vsr-simili che sono derivate da un'ampia gamma di ospiti, inclusi eucarioti, protisti, batteri, archei, cianobatteri e fagi. Come con ZFN e TALEN, HE possono essere utilizzate per creare una DSB a livello di un locus bersaglio come la fase iniziale in editing genomico. In aggiunta, alcune HE naturali e ingegnerizzate tagliano soltanto un singolo filamento di DNA funzionando quindi come nickasi sito-specifiche. La grande sequenza bersaglio di HE e la specificità che esse offrono le ha rese candidati attraenti per creare DSB sito-specifiche.

Una varietà di sistemi basati su HE sono stati descritti nell'arte, e loro

U001.D1.SM.61E

modificazioni sono regolarmente riportate; si veda, per esempio, le disamine di Steentoft et al., *Glycobiology* 24(8):663-80 (2014); Belfort e Bonocora, *Methods Mol Biol.* 1123:1-26 (2014); Hafez e Hausner, *Genome* 55(8):553-69 (2012); e riferimenti citati all'interno.

MegaTAL / Tev-mTALEN / MegaTev

Come esempi ulteriori di nucleasi ibride, la piattaforma MegaTAL e la piattaforma Tev-mTALEN utilizzano una fusione di domini di legame a DNA di TALE e HE cataliticamente attivi, sfruttando sia il legame a DNA regolabile che la specificità della TALE, nonché la specificità di sequenza di clivaggio di HE; si veda, per esempio Boissel et al., *NAR* 42: 2591-2601 (2014); Kleinstiver et al., *G3* 4:1155-65 (2014); e Boissel e Scharenberg, *Methods Mol. Biol.* 1239: 171-96 (2015).

In una ulteriore variazione l'architettura di MegaTev è la fusione di meganucleasi (Mega) con il dominio nucleasico derivato dall'endonucleasi di inserimento o homing di GIY-YIG I-Tev1 (Tev). I due siti attivi sono posizionati a circa 30 bp di distanza su un substrato di DNA e generano due DSB con estremità coesive non compatibili; si veda, per esempio, Wolfs et al., *NAR* 42, 8816-29 (2014). È previsto che altre combinazioni di approcci basati su nucleasi esistenti evolveranno e siano utili nel raggiungimento delle modificazioni genomiche bersaglio descritte nel presente documento.

dCas9-FokI o dCpf1-FokI e altre nucleasi

La combinazione delle proprietà strutturali e funzionali delle piattaforme nucleasiche descritte sopra offre un ulteriore approccio a editing genomico che

U001.D1.SM.61E

può potenzialmente superare alcune delle carenze intrinseche. A titolo di esempio, il sistema di editing genomico di CRISPR tipicamente utilizza una singola endonucleasi Cas9 per creare una DSB. La specificità di targeting o bersagliamento è guidata da una sequenza di 20 o 24 nucleotidi nel RNA guida che subisce un appaiamento di basi di Watson-Crick con il DNA bersaglio (più altre 2 basi nella sequenza NAG o NGG PAM adiacente nel caso di Cas9 da *S. pyogenes*). Tale sequenza è sufficientemente lunga per essere unica nel genoma umano, tuttavia la specificità dell'interazione RNA/DNA non è assoluta, con una promiscuità significativa talvolta tollerata, in particolare nella metà 5' della sequenza bersaglio, riducendo efficacemente il numero di basi che guidano specificità. Una soluzione a ciò è stato disattivare completamente la funzione catalitica di Cas9 o Cpf1 – mantenendo soltanto la funzione di legame a DNA guidata da RNA – e fondere invece un dominio FokI alla Cas9 deattivata; si vedano, per esempio, Tsai et al., Nature Biotech 32: 569-76 (2014); e Guilinger et al., Nature Biotech. 32: 577-82 (2014). Poiché FokI deve dimerizzare per diventare cataliticamente attivo, due RNA guida sono richiesti per legare due fusioni di FokI in stretta prossimità per formare il dimero e clivare DNA. Ciò essenzialmente raddoppia il numero di basi nei siti bersaglio combinati, aumentando quindi la stringenza di targeting o bersagliamento da sistemi basati su CRISPR.

Come ulteriore esempio, la fusione del dominio di legame a DNA di TALE a una HE cataliticamente attiva, come I-TevI, trae vantaggio sia dal legame a DNA regolabile che dalla specificità della TALE, nonché dalla specificità di

U001.D1.SM.61E

sequenza di clivaggio di I-TevI, con l'aspettativa che clivaggio off-target può essere ulteriormente ridotto.

Rilevamento di mutazioni on- e off-target mediante sequenziamento

Per sequenziare siti on-target e siti off-target putativi, i primer di amplificazione appropriati sono stati identificati e reazioni sono state impostate con questi primer utilizzando il DNA genomico raccolto utilizzando soluzione di estrazione di DNA QuickExtract (Epicentre) da cellule trattate tre giorni post-trasfezione. I primer di amplificazione contengono la porzione gene-specifica fiancheggiata da adattatori. L'estremità 5' del primer forward include un sito di legame a primer forward (read1) modificato. L'estremità 5' di primer reverse contiene un sito di legame a primer a codice a barre e reverse (read2) modificato combinato, in orientamento opposto. Le reazioni di PCR individuali sono state validate mediante separazione su gel di agarosio, poi purificate e ri-amplificate. I primer forward del secondo turno contengono la sequenza Illumina P5, seguita da una porzione del sito di legame a primer (read1) forward modificato. I primer reverse del secondo turno contengono la sequenza Illumina P7 (all'estremità 5'), seguita dal codice a barre a 6 basi e dal sito di legame a primer a codice a barre e reverse (read2) modificato combinato. Le amplificazioni del secondo turno sono stata anch'esse controllate su gel di agarosio, poi purificate e quantificate utilizzando uno spettrofotometro NanoDrop. I prodotti di amplificazione sono stati riuniti per abbinare la concentrazione e poi sottoposti all'Emory Integrated Genomic core per prepping e sequenziamento di libreria su una macchina Illumina Miseq.

U001.D1.SM.61E

Le letture di sequenziamento sono state smistate tramite codice a barre e poi allineate alle sequenze di riferimento fornite da bioinformatica per ciascun prodotto. Tassi di inserzione e delezione nelle letture di sequenziamento allineate sono state rilevate nella regione dei siti tagliati putativi utilizzando il software precedentemente descritto; si veda, per esempio, Lin et al., *Nucleic Acids Res.*, 42: 7473-7485 (2014). I livelli di inserzioni e delezioni rilevati in questa finestra sono stati poi comparati al livello osservato nella stessa posizione in DNA genomico isolato da cellule trasfettate in mock per minimizzare gli effetti di artefatti di sequenziamento.

Saggi di rilevamento di mutazione

Le attività di clivaggio on-target e off-target di combinazioni di Cas9 e RNA guida sono state misurate utilizzando i tassi di mutazione risultanti dalla riparazione imperfetta di rotture a doppio filamento mediante NHEJ.

Loci on-target sono stati amplificati utilizzando AccuPrime Taq DNA Polimerasi ad alta fedeltà (Life Technologies, Carlsbad, CA) seguendo le istruzioni del produttore per 40 cicli (94°C, 30 s; 52-60°C, 30 s; 68°C, 60 s) in 50 µl di reazioni contenenti 1 µl del lisato cellulare, e 1 µl di ciascun primer di amplificazione 10 µM. Saggi di rilevamento di mutazione di T7EI sono stati effettuati secondo il protocollo del produttore [Reyon et al., *Nat. Biotechnol.*, 30: 460-465 (2012)], con le digestioni separate su gel di agarosio al 2% e quantificati utilizzando ImageJ [Guschin et al., *Methods Mol. Biol.*, 649: 247-256 (2010)]. I saggi determinano la percentuale di inserzioni/delezioni (“indel”) nella popolazione in massa di cellule.

U001.D1.SM.61E

Metodi e composizioni

Pertanto, la presente descrizione si riferisce in particolare ai seguenti aspetti non limitativi: in un primo metodo, Metodo 1, la presente descrizione fornisce un metodo per editare un gene BCL11A in una cellula umana mediante editing genomico, il metodo comprendendo la fase di: introduzione nella cellula umana di una o più endonucleasi di acido desossiribonucleico (DNS) per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB) all'interno del o vicino al gene BCL11A o altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che porta a una delezione, modulazione o inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A.

In un altro metodo, Metodo 2, la presente descrizione fornisce un metodo per editare un gene BCL11A in una cellula umana mediante editing genomico, come fornito nel Metodo 1, in cui la sequenza di controllo di trascrizione è situata all'interno di un secondo introne del gene BCL11A.

In un altro metodo, Metodo 3, la presente descrizione fornisce un metodo per editare un gene BCL11A in una cellula umana mediante editing genomico, come fornito in Metodi 1 o 2, in cui la sequenza di controllo di trascrizione è situata all'interno di un sito ipersensibile a DNA +58 (DHS) del gene BCL11A.

In un altro metodo, Metodo 4, la presente descrizione descrive un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia, il metodo comprendendo le fasi di: creazione di una cellula staminale pluripotente indotta (iPSC) specifica per paziente; editing all'interno o vicino a un gene BCL11A o

U001.D1.SM.61E

di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A della iPSC; differenziazione della iPSC sottoposta a editing genomico in una cellula progenitrice ematopoietica; e impiantazione della cellula progenitrice ematopoietica nel paziente.

In un altro metodo, Metodo 5, la presente descrizione descrive un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come descritto nel Metodo 4, in cui la fase di creazione comprende: l'isolamento di una cellula somatica dal paziente; e l'introduzione di un set di geni associati a pluripotenza nella cellula somatica per indurre la cellula somatica a diventare una cellula staminale pluripotente.

In un altro metodo, Metodo 6, la presente descrizione descrive un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come descritto nel Metodo 5, in cui la cellula somatica è un fibroblasto.

In un altro metodo, Metodo 7, la presente descrizione descrive un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come descritto nei Metodi 5 o 6, in cui il set di geni associati a pluripotenza è uno o più dei geni scelti nel gruppo costituito da OCT4, SOX2, KLF4, Lin28, NANOG e cMYC.

In un altro metodo, Metodo 8, la presente descrizione descrive un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come descritto in qualsiasi dei Metodi 4-7, dove la fase di editing comprende l'introduzione nella iPSC di una o più acido desossiribonucleico (DNS) endonucleasi per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB) all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che

U001.D1.SM.61E

codifica un elemento regolatorio del gene BCL11A che ha come risultato una delezione, modulazione o inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A.

In un altro metodo, Metodo 9, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato in qualsiasi dei Metodi 4-8, dove la fase di differenziazione comprende uno o più dei seguenti per differenziare la iPSC con editing genomico in una cellula progenitrice ematopoietica: il trattamento con una combinazione di piccole molecole, la veicolazione di fattori di trascrizione master, la veicolazione di mRNA codificante fattori di trascrizione master, o la veicolazione di mRNA codificante fattori di trascrizione.

In un altro metodo, Metodo 10, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato in qualsiasi dei Metodi 4-9, dove la fase di impiantazione comprende l'impiantazione della cellula progenitrice ematopoietica nel paziente mediante trapianto, iniezione locale, infusione sistemica, o combinazioni di esse.

In un altro metodo, Metodo 11, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia, il metodo comprendendo le fasi di: isolamento di una cellula staminale mesenchimale dal paziente; editing all'interno o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A della cellula staminale mesenchimale; differenziazione della cellula staminale mesenchimale sottoposta a editing genomico in una cellula progenitrice ematopoietica; e

U001.D1.SM.61E

impiantazione della cellula progenitrice ematopoietica nel paziente.

In un altro metodo, Metodo 12, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato nel Metodo 11, dove la cellula staminale mesenchimale è isolata dal midollo osseo o sangue periferico del paziente.

In un altro metodo, Metodo 13, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato nei Metodi 11 o 12, dove la fase di isolamento comprende: l'aspirazione di midollo osseo e l'isolamento di cellule mesenchimali utilizzando mezzi di centrifugazione in gradiente di densità.

In un altro metodo, Metodo 14, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato in qualsiasi dei Metodi 11-13, dove la fase di editing comprende l'introduzione nella cellula staminale mesenchimale di una o più acido desossiribonucleico (DNS) endonucleasi per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB) all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che dà luogo a una delezione, modulazione o inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A.

In un altro metodo, Metodo 15, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato in qualsiasi dei Metodi 11-14, dove la fase di differenziazione comprende uno o più dei seguenti per differenziare la cellula staminale mesenchimale sottoposta a

U001.D1.SM.61E

editing genomico in una cellula progenitrice ematopoietica: il trattamento con una combinazione di piccole molecole, la veicolazione di fattori di trascrizione master, la veicolazione di mRNA codificante fattori di trascrizione master, o la veicolazione di mRNA codificante fattori di trascrizione.

In un altro metodo, Metodo 16, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato in qualsiasi dei Metodi 11-15, dove la fase di impiantazione comprende l'impiantazione della cellula progenitrice ematopoietica nel paziente mediante trapianto, iniezione locale, infusione sistemica, o combinazioni di esse.

In un altro metodo, Metodo 17, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia, il metodo comprendendo le fasi di: isolamento di una cellula progenitrice ematopoietica dal paziente; editing all'interno o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A della cellula progenitrice ematopoietica; e impiantazione della cellula progenitrice ematopoietica sottoposta a editing genomico nel paziente.

In un altro metodo, Metodo 18, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come fornito nel Metodo 17, dove il metodo comprende inoltre il trattamento del paziente con fattore di stimolazione di colonie di granulociti (GCSF) prima della fase di isolamento.

In un altro metodo, Metodo 19, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato nel

U001.D1.SM.61E

Metodo 18, dove la fase di trattamento è effettuata in combinazione con Plerixaflor.

In un altro metodo, Metodo 20, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato in qualsiasi dei Metodi 17-19, dove la fase di isolamento comprende l'isolamento di cellule CD34+.

In un altro metodo, Metodo 21, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato in qualsiasi dei Metodi 17-20, dove la fase di editing comprende l'introduzione nella cellula progenitrice ematopoietica di una o più acido desossiribonucleico (DNS) endonucleasi per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSBs) all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che porta a una delezione, mutazione o inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A.

In un altro metodo, Metodo 22, la presente descrizione divulga un metodo *ex vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato in qualsiasi dei Metodi 17-21, dove la fase di impiantazione comprende l'impiantazione della cellula progenitrice ematopoietica sottoposta a editing genomico nel paziente mediante trapianto, iniezione locale, infusione sistemica, o combinazioni di esse.

In un altro metodo, Metodo 23, la presente descrizione divulga un metodo *in vivo* per trattare un paziente affetto da una emoglobinopatia, il metodo

U001.D1.SM.61E

comprendendo la fase di editing di un gene BCL11A in una cellula del paziente.

In un altro metodo, Metodo 24, la presente descrizione divulga un metodo *in vivo* per trattare un paziente affetto da una emoglobinopatia come divulgato nel Metodo 23, dove la fase di editing comprende l'introduzione nella cellula di una o più acido desossiribonucleico (DNS) endonucleasi per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB) all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che dà luogo a una delezione, modulazione, o inattivazione permanente di un controllo di trascrizione del gene del gene BCL11A.

In un altro metodo, Metodo 25, la presente descrizione divulga un metodo *in vivo* per trattare un paziente affetto da una emoglobinopatia come fornito nei Metodi 23 o 24, in cui la cellula è una cellula di midollo spinale, una cellula progenitrice ematopoietica, o una cellula CD34+.

In un altro metodo, Metodo 26, la presente descrizione divulga un metodo secondo uno qualsiasi dei Metodi 1, 8, 14, 21 e 24, in cui le una o più DNA endonucleasi sono una endonucleasi Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (note anche come Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, o Cpf1; un omologo di esse, una ricombinazione della molecola di esse presente naturalmente, e combinazioni di esse.

U001.D1.SM.61E

In un altro metodo, Metodo 27, la presente descrizione fornisce un metodo come divulgato nel Metodo 26, dove il metodo comprende l'introduzione nella cellula di uno o più polinucleotidi codificanti le una o più DNA endonucleasi.

In un altro metodo, Metodo 28, la presente descrizione fornisce un metodo come fornito nei Metodi 26 o 27, dove il metodo comprende l'introduzione nella cellula di uno o più acidi ribonucleici (RNA) codificanti le una o più DNA endonucleasi.

In un altro metodo, Metodo 29, la presente descrizione divulga un metodo come divulgato nei Metodi 27 o 28, in cui gli uno o più polinucleotidi o uno o più RNA sono uno o più polinucleotidi modificati o uno o più RNA modificati.

In un altro metodo, Metodo 30, la presente descrizione divulga un metodo come divulgato nel Metodo 26, dove le una o più DNA endonucleasi sono una o più proteine o polipeptidi.

In un altro metodo, Metodo 31, la presente descrizione fornisce un metodo come divulgato nel Metodo 30, dove le una o più proteine o gli uno o più polipeptidi sono fiancheggiati all'estremità N, all'estremità C o sia all'estremità N che all'estremità C da uno o più segnali di localizzazione nucleare (NLS).

In un altro metodo, Metodo 32, la presente descrizione fornisce un metodo come fornito nel Metodo 31, dove le una o più proteine o gli uno o più polipeptidi sono fiancheggiati da due NLS, un NLS situato all'estremità N e il secondo NLS situato all'estremità C.

In un altro metodo, Metodo 33, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 31-32, dove gli uno o più NLS sono un

U001.D1.SM.61E

NLS SV40.

In un altro metodo, Metodo 34, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1-33, dove il metodo comprende inoltre l'introduzione nella cellula di uno o più acidi ribonucleici guida (gRNA).

In un altro metodo, Metodo 35, la presente descrizione fornisce un metodo come fornito nel Metodo 34, dove gli uno o più gRNA sono RNA guida monomolecolari (sgRNA).

In un altro metodo, Metodo 36, la presente descrizione fornisce un metodo come fornito nei Metodi 34 o 35, dove gli uno o più gRNA o uno o più sgRNA sono una o più gRNA modificati o uno o più sgRNA modificati.

In un altro metodo, Metodo 37, la presente descrizione fornisce un metodo come fornito nel Metodo 36, dove gli uno o più sgRNA modificati comprendono tre residui di 2'-O-metil-fosforotioato a livello di o vicino a ciascuna delle loro estremità 5' e 3'.

In un altro metodo, Metodo 38, la presente descrizione fornisce un metodo come fornito nel Metodo 37, dove il sgRNA modificato è la sequenza di acido nucleico di SEQ ID NO: 71.959.

In un altro metodo, Metodo 39, la presente descrizione fornisce un metodo come fornito nei Metodi 34-38, dove le una o più DNA endonucleasi sono pre-complesate con uno o più gRNA o uno o più sgRNA per formare una o più ribonucleoproteine (RNP).

In un altro metodo, Metodo 40, la presente descrizione fornisce un metodo come fornito nel Metodo 39, dove il rapporto in peso tra sgRNA e DNA

U001.D1.SM.61E

endonucleasi nella RNP è 1:1.

In un altro metodo, Metodo 41, la presente descrizione fornisce un metodo come fornito nel Metodo 40, dove il sgRNA comprende la sequenza di acido nucleico di SEQ ID NO: 71.959, la DNA endonucleasi è una Cas9 di *S. pyogenes* comprendente un NLS SV40 all'estremità N e un NLS SV40 all'estremità C, dove il rapporto in peso tra sgRNA e DNA endonucleasi è 1:1.

In un altro metodo, Metodo 42, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1-41, dove il metodo comprende inoltre l'introduzione nella cellula di un template donatore polinucleotidico comprendente un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata.

In un altro metodo, Metodo 43, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1, 8, 14, 21, o 24, dove il metodo comprende inoltre l'introduzione nella cellula di un acido ribonucleico guida (gRNA) e di un template donatore polinucleotidico comprendente un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata, e in cui le una o più DNA endonucleasi sono una o più endonucleasi Cas9 o Cpf1 che effettuano una rottura a singolo filamento (SSB) o una rottura a doppio filamento (DSB) a livello di un locus all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che facilita l'inserzione di una nuova sequenza dal template donatore polinucleotidico nel DNA cromosomico a livello del locus che dà luogo a una inserzione, modulazione o inattivazione permanente della

U001.D1.SM.61E

sequenza di controllo di trascrizione del DNA cromosomico prossimale al locus, e in cui il gRNA comprende una sequenza spaziatrice che è complementare a un segmento del locus.

In un altro metodo, Metodo 44, la presente descrizione fornisce un metodo come fornito nel Metodo 43, in cui prossimale significa nucleotidi sia a monte che a valle del locus.

In un altro metodo, Metodo 45, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1, 8, 14, 21, o 24, dove il metodo comprende inoltre l'introduzione nella cellula di uno o più acidi ribonucleici guida (gRNA) e di un template donatore polinucleotidico comprendente un gene BCL11A di tipo selvatico o cDNA comprendente una sequenza di controllo di trascrizione modificata, e in cui le una o più DNA endonucleasi sono una o più endonucleasi Cas9 o Cpf1 che effettuano o creano una coppia di rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB), la prima rottura a un locus 5' e la seconda rottura a un locus 3', all'interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che facilita l'inserzione di una nuova sequenza dal template donatore polinucleotidico nel DNA cromosomico tra il locus 5' e il locus 3' che dà luogo a una inserzione, modulazione o inattivazione permanente della sequenza di controllo di trascrizione del DNA cromosomico tra il locus 5' e il locus 3'.

In un altro metodo, Metodo 46, la presente descrizione fornisce un metodo come fornito nel Metodo 45, in cui un gRNA crea una coppia di SSB o DSB.

U001.D1.SM.61E

In un altro metodo, Metodo 47, la presente descrizione fornisce un metodo come fornito nel Metodo 45, in cui un gRNA comprende una sequenza spaziatrice che è complementare al locus 5' o al locus 3'.

In un altro metodo, Metodo 48, la presente descrizione fornisce un metodo come fornito nel Metodo 45, dove il metodo comprende un primo RNA guida e un secondo RNA guida, dove il primo RNA guida comprende una sequenza spaziatrice che è complementare a un segmento del locus 5' e il secondo RNA guida comprende una sequenza spaziatrice che è complementare a un segmento del locus 3'.

In un altro metodo, Metodo 49, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 43-48, dove gli uno o due gRNA sono uno o due RNA guida monomolecolari (sgRNA).

In un altro metodo, Metodo 50, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 43-49, dove gli uno o due gRNA o uno o due sgRNA sono uno o due gRNA modificati o uno o due sgRNA modificati.

In un altro metodo, Metodo 51, la presente descrizione fornisce un metodo come fornito nel Metodo 50, dove il sgRNA modificato comprende tre residui di 2'-O-metil-fosforotioato a livello di o vicino a ciascuna delle sue estremità 5' e 3'.

In un altro metodo, Metodo 52, la presente descrizione fornisce un metodo come fornito nel Metodo 51, dove il sgRNA modificato è la sequenza di acido nucleico di SEQ ID NO: 71.959.

In un altro metodo, Metodo 53, la presente descrizione fornisce un metodo

U001.D1.SM.61E

come fornito in uno qualsiasi dei Metodi 43-52, dove le una o più endonucleasi Cas9 sono pre-complesate con uno o due gRNA o uno o due sgRNA per formare una o più ribonucleoproteine (RNP).

In un altro metodo, Metodo 54, la presente descrizione fornisce un metodo come fornito nel Metodo 53, dove le una o più endonucleasi Cas9 sono fiancheggiate all'estremità N, all'estremità C o sia all'estremità N che all'estremità C da uno o più segnali di localizzazione nucleare (NLS).

In un altro metodo, Metodo 55, la presente descrizione fornisce un metodo come fornito nel Metodo 54, dove le una o più endonucleasi Cas9 sono fiancheggiate da due NLS, un NLS situato all'estremità N e il secondo NLS situato all'estremità C.

In un altro metodo, Metodo 56, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 54-55, dove gli uno o più NLS sono un NLS SV40.

In un altro metodo, Metodo 57, la presente descrizione fornisce un metodo come fornito nel Metodo 53, dove il rapporto in peso tra sgRNA e endonucleasi Cas9 nella RNP è 1:1.

In un altro metodo, Metodo 58, la presente descrizione fornisce un metodo come fornito nel Metodo 53, dove il sgRNA comprende la sequenza di acido nucleico di SEQ ID NO: 71.959, l'endonucleasi Cas9 è una endonucleasi Cas9 di *S. pyogenes* comprendente un NLS SV40 all'estremità N e un NLS SV40 all'estremità C, dove il rapporto in peso tra sgRNA e endonucleasi Cas9 è 1:1.

In un altro metodo, Metodo 59, la presente descrizione fornisce un metodo

U001.D1.SM.61E

come fornito in uno qualsiasi dei Metodi 43-58, dove il template donatore è a singolo o a doppio filamento.

In un altro metodo, Metodo 60, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 42-59, dove la sequenza di controllo di trascrizione modificata è situata all'interno di un secondo introne del gene BCL11A.

In un altro metodo, Metodo 61, come fornito in qualsiasi dei Metodi 42-59, dove la sequenza di controllo di trascrizione modificata è situata all'interno di un sito ipersensibile a DNA +58 (DHS) del gene BCL11A.

In un altro metodo, Metodo 62, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 42-61, dove l'inserzione avviene mediante riparazione diretta da omologia (HDR).

In un altro metodo, Metodo 63, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 8, 14, 21, 24, 43, e 45, dove la SSB, DSB, o il locus 5' e il locus 3' sono situati all'interno di un secondo introne del gene BCL11A.

In un altro metodo, Metodo 64, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 8, 14, 21, 24, 43, e 45, in cui la SSB, DSB, o 5' DSB e 3' DSB sono situate all'interno di un sito ipersensibile a DNA +58 DNA (DHS) del gene BCL11A.

In un altro metodo, Metodo 65, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1, 8, 14, 21, o 24, dove il metodo comprende inoltre l'introduzione nella cellula di uno o più acidi ribonucleici

U001.D1.SM.61E

guida (gRNA), e dove le una o più DNA endonucleasi sono una o più endonucleasi Cas9 o Cpf1 che effettuano o creano una coppia di rotture a singolo filamento (SSB) o rotture a doppio filamento (DSBs), una prima SSB o DSB a livello di un locus 5' e una seconda SSB o DSB a livello di un locus 3', all'interno del o vicino al gene BCL11A o altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che causa una delezione del DNA cromosomico tra il locus 5' e il locus 3' che porta a una delezione, modulazione o inattivazione permanente della sequenza di controllo di trascrizione del DNA cromosomico tra il locus 5' e il locus 3'.

In un altro metodo, Metodo 66, la presente descrizione fornisce un metodo come fornito nel Metodo 65, in cui un gRNA crea una coppia di SSB o DSB.

In un altro metodo, Metodo 67, la presente descrizione fornisce un metodo come fornito nel Metodo 65, in cui un gRNA comprende una sequenza spaziatrice che è complementare al locus 5' o al locus 3'.

In un altro metodo, Metodo 68, la presente descrizione fornisce un metodo come fornito nel Metodo 65, dove il metodo comprende un primo RNA guida e un secondo RNA guida, in cui il primo RNA guida comprende una sequenza spaziatrice che è complementare a un segmento del locus 5' e il secondo RNA guida comprende una sequenza spaziatrice che è complementare a un segmento del locus 3'.

In un altro metodo, Metodo 69, la presente descrizione fornisce un metodo come fornito nei Metodi 65-68, in cui gli uno o più gRNA sono uno o più RNA guida monomolecolari (sgRNA).

U001.D1.SM.61E

In un altro metodo, Metodo 70, la presente descrizione fornisce un metodo come fornito nei Metodi 65-69 in cui gli uno o più gRNA o uno o più sgRNA sono uno o più gRNA modificati o uno o più sgRNA modificati.

In un altro metodo, Metodo 71, la presente descrizione fornisce un metodo come fornito nel Metodo 70, in cui il sgRNA modificato comprende tre residui di 2'-O-metil-fosforotioato a livello di o vicino a ciascuna delle sue estremità 5' e 3'.

In un altro metodo, Metodo 72, la presente descrizione fornisce un metodo come fornito nel Metodo 71, in cui il sgRNA modificato è la sequenza di acido nucleico di SEQ ID NO: 71.959.

In un altro metodo, Metodo 73, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 65-72, in cui le una o più endonucleasi Cas9 sono pre-complesstate con uno o più gRNA o uno o più sgRNA per formare una o più ribonucleoproteine (RNP).

In un altro metodo, Metodo 74, la presente descrizione fornisce un metodo come fornito nel Metodo 73, in cui le una o più endonucleasi Cas9 sono fiancheggiate all'estremità N, all'estremità C, o sia all'estremità N che all'estremità C da uno o più segnali di localizzazione nucleare (NLS).

In un altro metodo, Metodo 75, la presente descrizione fornisce un metodo come fornito nel Metodo 74, in cui le una o più endonucleasi Cas9 sono fiancheggiate da due NLS, un NLS situato all'estremità N e il secondo NLS situato all'estremità C.

In un altro metodo, Metodo 76, la presente descrizione fornisce un metodo

U001.D1.SM.61E

come fornito in uno qualsiasi dei Metodi 74-75, in cui gli uno o più NLS sono un NLS SV40.

In un altro metodo, Metodo 77, la presente descrizione fornisce un metodo come fornito nel Metodo 73, in cui il rapporto in peso tra sgRNA e endonucleasi Cas9 nella RNP è 1:1.

In un altro metodo, Metodo 78, la presente descrizione fornisce un metodo come fornito nel Metodo 73, in cui il sgRNA comprende la sequenza di acido nucleico di SEQ ID NO: 71.959, l'endonucleasi Cas9 è una Cas9 di *S. pyogenes* comprendente un NLS SV40 all'estremità N e un NLS SV40 all'estremità C, in cui il rapporto in peso tra sgRNA e endonucleasi Cas9 è 1:1.

In un altro metodo, Metodo 79, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 65-78, in cui sia il locus 5' che il locus 3' sono situati all'interno di un secondo introne del gene BCL11A.

In un altro metodo, Metodo 80, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 65-78, in cui sia il locus 5' che il locus 3' sono situati all'interno di un sito ipersensibile a DNA +58 (DHS) del gene BCL11A.

In un altro metodo, Metodo 81, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1, 8, 14, 21, o 24-80, in cui il Cas9 o Cpf1 mRNA, gRNA, e template donatore sono formulati in nanoparticelle lipidiche separate o tutti co-formulati in una nanoparticella lipidica.

In un altro metodo, Metodo 82, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1, 8, 14, 21, o 24-80, in cui il Cas9 o

U001.D1.SM.61E

Cpf1 mRNA è formulato in una nanoparticella lipidica, e sia gRNA che template donatore sono veicolati alla cellula da un vettore virale adeno-associato (AAV).

In un altro metodo, Metodo 83, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1, 8, 14, 21, o 24-80, in cui il Cas9 o Cpf1 mRNA è formulato in una nanoparticella lipidica, e il gRNA è veicolato alla cellula mediante elettroporazione e template donatore è veicolato alla cellula da un vettore virale adeno-associato (AAV).

In un altro metodo, Metodo 84, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1, 8, 14, 21, o 24-80, in cui le una o più RNP sono veicolate alla cellula mediante elettroporazione.

In un altro metodo, Metodo 85, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1-84, in cui il gene BCL11A è situato sul Cromosoma 2: 60.451.167 -60.553.567 (Consorzio di Riferimento sul Genoma - Genome Reference Consortium - GRCh38).

In un altro metodo, Metodo 86, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1-85, in cui l'emoglobinopatia è scelta in un gruppo costituito da anemia falciforme e talassemia (α , β , δ , γ , e combinazione di esse).

In un altro metodo, Metodo 87, la presente descrizione fornisce un metodo come fornito in uno qualsiasi dei Metodi 1-86, in cui l'editing all'interno o vicino a un gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A può ridurre l'espressione del gene BCL11A.

U001.D1.SM.61E

In una prima composizione, Composizione 1, la presente descrizione fornisce uno o più acidi ribonucleici guida (gRNA) per editare un gene BCL11A in una cellula da un paziente affetto da una emoglobinopatia, gli uno o più gRNA comprendendo una sequenza spaziatrice scelta nel gruppo costituito da sequenze di acidi nucleici in SEQ ID NOs: 1 – 71.947 della Lista di sequenze.

In un'altra composizione, Composizione 2, la presente descrizione fornisce gli uno o più gRNA della Composizione 1, dove gli uno o più gRNA sono uno o più RNA guida monomolecolari (sgRNA).

In un'altra composizione, Composizione 3, la presente descrizione fornisce gli uno o più gRNA o sgRNA di Composizioni 1 o 2, dove gli uno o più gRNA o uno o più sgRNA sono uno o più gRNA modificati o uno o più sgRNA modificati.

In un'altra composizione, Composizione 4, la presente descrizione fornisce gli uno o più sgRNA della Composizione 3, dove gli uno o più sgRNA modificati comprendono tre residui di 2'-O-metil-fosforotioato a livello di o vicino a ciascuna delle loro estremità 5' e 3'.

In un'altra composizione, Composizione 5, la presente descrizione fornisce gli uno o più sgRNA della Composizione 3, dove gli uno o più sgRNA modificati comprendono la sequenza di acido nucleico di SEQ ID NO: 71.959.

In un'altra composizione, Composizione 6, la presente descrizione fornisce un RNA guida monomolecolare (sgRNA) comprendente la sequenza di acido nucleico di SEQ ID NO: 71.959.

Definizioni

U001.D1.SM.61E

Il termine "comprendente" o "comprende" è utilizzato in riferimento a composizioni, metodi e uno o più rispettivi componenti, che sono essenziali all'invenzione, tuttavia aperto all'inclusione di elementi non specificati, sia che siano essenziali o meno.

L'espressione "costituito essenzialmente da" si riferisce a quegli elementi richiesti per un dato aspetto. L'espressione consente la presenza di elementi supplementari che non influiscono materialmente sulla o sulle caratteristiche basilari e innovative o funzionali di quell'aspetto dell'invenzione.

L'espressione "costituito essenzialmente da" si riferisce a aspetti dell'invenzione.

L'espressione "costituito da" si riferisce a composizioni, metodi e rispettivi componenti di essi come descritti nel presente documento, che sono esclusivi di qualsiasi elemento non indicati in quella descrizione dell'aspetto.

Le forme singolari "un" "uno(a)" e "il (la)" includono riferimenti plurali, salvo che il contesto non indichi chiaramente altrimenti.

Qualsiasi intervallo numerico indicato in questa specificazione descrive tutti i sotto-intervalli della stessa precisione numerica (vale a dire aventi lo stesso numero di cifre specificato) sussunti all'interno dell'intervallo indicato. Per esempio, un intervallo indicato da "1,0 a 10,0" descrive tutti i sottointervalli tra (e includenti) il valore minimo indicato di 1,0 e il valore massimo indicato di 10,0, come, per esempio, "da 2,4 a 7,6," anche se l'intervallo "da 2,4 a 7,6" non è espressamente indicato nel testo della specificazione. Pertanto, la Richiedente si riserva il diritto di modificare questa specificazione, incluse le rivendicazioni,

U001.D1.SM.61E

per indicare espressamente qualsiasi sottointervallo della stessa precisione numerica sussunta all'interno degli intervalli espressamente indicati in questa specificazione. Tutti questi intervalli sono intrinsecamente descritti in questa specificazione in modo tale che una modifica per indicare espressamente qualsiasi sottointervallo si conformerà con la descrizione scritta, sufficienza di descrizione, e requisiti di materia aggiunti, inclusi i requisiti sotto 35 U.S.C. § 112(a) e Articolo 123(2) EPC. Inoltre, se non espressamente specificato o altrimenti richiesto dal contesto, tutti i parametri numerici descritti in questa specificazione (come quelli esprimenti valori, intervalli, quantità, percentuali, e simili) possono essere letti come se preceduti dalla parola “circa”, anche se la parola “circa” non appare espressamente prima di un numero. Inoltre, parametri numerici descritti in questa specificazione dovrebbero essere intesi alla luce del numero di cifre significative riportate, della precisione numerica, e applicando tecniche ordinarie di arrotondamento. È anche sottinteso che parametri numerici descritti in questa specificazione possiederanno necessariamente la variabilità intrinseca caratteristica delle tecniche di misurazione sottostanti utilizzate per determinare il valore numerico del parametro.

Esempi

L'invenzione sarà compresa in modo più completo facendo riferimento agli esempi seguenti, che forniscono aspetti illustrativi non limitativi dell'invenzione.

Gli esempi descrivono l'uso del sistema CRISPR come una tecnica di editing genomico illustrativa per creare delezioni, inserzioni o sostituzioni genomiche

U001.D1.SM.61E

definite, chiamate “modificazioni genomiche” nel presente documento, all’interno del o vicino al gene BCL11A o di altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A che portano a una delezione, modulazione o inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A. L’introduzione delle modificazioni terapeutiche definite rappresenta una strategia terapeutica innovativa per il potenziale miglioramento di una emoglobinopatia, come descritto e illustrato nel presente documento.

Esempio 1 – Siti bersaglio CRISPR/SpCas9 per la sequenza di controllo di trascrizione del gene BCL11A

Regioni della sequenza di 12,4 kb di controllo di trascrizione del gene BCL11A sono state scansionate per siti bersaglio. Ciascuna area è stata scansionata per un motivo adiacente protospaziatore (PAM) avente la sequenza NRG. Sequenze spaziatrici di 20 bp di gRNA corrispondenti al PAM sono state identificate, come mostrato in SEQ ID NOs: 1 – 29.482 della Lista di sequenze.

Esempio 2 – Siti bersaglio CRISPR/SaCas9 per la sequenza di controllo di trascrizione del gene BCL11A

Regioni della sequenza di 12,4 kb di controllo di trascrizione del gene BCL11A sono state scansionate per siti bersaglio. Ciascuna area è stata scansionata per un motivo adiacente protospaziatore (PAM) avente la sequenza NNGRRT. Sequenze spaziatrici di 20 bp di gRNA corrispondenti al PAM sono state identificate, come mostrato in SEQ ID NOs: 29.483 – 32.387 della Lista di sequenze.

U001.D1.SM.61E

Esempio 3 – Siti bersaglio RISPR/StCas9 per la sequenza di controllo di trascrizione del gene BCL11A

Regioni della sequenza di 12,4 kb di controllo di trascrizione del gene BCL11A sono state scansionate per siti bersaglio. Ciascuna area è stata scansionata per un motivo adiacente protospaziatore (PAM) avente la sequenza NNAGAAW. Sequenze spaziatrici di 20 bp di gRNA corrispondenti al PAM sono state identificate, come mostrato in SEQ ID NOs: 32.388 – 33.420 della Lista di sequenze.

Esempio 4 – Siti bersaglio CRISPR/TdCas9 per la sequenza di controllo di trascrizione del gene BCL11A

Regioni della sequenza di 12,4 kb di controllo di trascrizione del gene BCL11A sono state scansionate per siti bersaglio. Ciascuna area è stata scansionata per un motivo adiacente protospaziatore (PAM) avente la sequenza NAAAAC. Sequenze spaziatrici di 20 bp di gRNA corrispondenti al PAM sono state identificate, come mostrato in SEQ ID NOs: 33.421 – 33.851 della Lista di sequenze.

Esempio 5 -Siti bersaglio CRISPR/NmCas9 per la sequenza di controllo di trascrizione del gene BCL11A

Regioni della sequenza di 12,4 kb di controllo di trascrizione del gene BCL11A sono state scansionate per siti bersaglio. Ciascuna area è stata scansionata per un motivo adiacente protospaziatore (PAM) avente la sequenza NNNNGHTT. Sequenze spaziatrici di 20 bp di gRNA corrispondenti al PAM sono state identificate, come mostrato in SEQ ID NOs: 33.852 – 36.731 della

U001.D1.SM.61E

Lista di sequenze.

Esempio 6 – Siti bersaglio CRISPR/Cpf1 per la sequenza di controllo di trascrizione del gene BCL11A

Regioni della sequenza di 12,4 kb di controllo di trascrizione del gene BCL11A sono state scansionate per siti bersaglio. Ciascuna area è stata scansionata per un motivo adiacente protospaziatore (PAM) avente la sequenza YTN. Sequenze spaziatrici di 22 bp di gRNA corrispondenti al PAM sono state identificate, come mostrato in SEQ ID NOs: 36.732 – 71.947 della Lista di sequenze.

Esempio 7 – Analisi di bioinformatica dei filamenti guida

Guide candidate saranno vagliate e selezionate in un processo a più fasi che implica sia legame teorico che attività valutata sperimentalmente. A titolo illustrativo, guide candidate aventi sequenze che si abbinano a un particolare sito on-target, come un sito all'interno della sequenza di controllo di trascrizione del gene BCL11A, con PAM adiacente possono venire valutate per il loro potenziale per clivare a livello di siti off-target aventi sequenze simili, utilizzando uno o più di una varietà di strumenti di bioinformatica disponibili per valutare legame off-target, come descritto e illustrato più dettagliatamente sotto, al fine di valutare la probabilità di effetti a livello di posizioni cromosomiche differenti da quelle desiderate. Candidati che sono predetti avere un potenziale inferiore per attività off-target possono poi venire valutati sperimentalmente per misurare la loro attività on-target, e poi attività off-target a livello di vari siti. Guide preferite hanno una attività on-target sufficientemente

U001.D1.SM.61E

elevata per raggiungere livelli desiderati di editing genico a livello del locus selezionato, e attività off-target relativamente inferiore per ridurre la probabilità di alterazioni a livello di altri loci cromosomici. Il rapporto tra attività on-target e off-target è spesso indicato come “specificità” di una guida.

Per lo screening iniziale di attività off-target predette esistono numerosi strumenti di bioinformatica noti e disponibili pubblicamente che possono venire utilizzati per predire i siti off-target più probabili; e dal momento che il legame a siti bersaglio nel sistema di nucleasi CRISPR/Cas9/Cpf1 è guidato da appaiamento di basi di Watson-Crick tra sequenze complementari, il grado di dissimilarità (e quindi il potenziale ridotto per legame off-target) è essenzialmente correlato a differenze di sequenze primarie: appaiamenti errati o mismatch e rigonfiamenti o bulge, vale a dire basi che sono cambiate in una base non complementare, e inserzioni o delezioni di basi nel potenziale sito off-target rispetto al sito bersaglio. Uno strumento di bioinformatica esemplare chiamato COSMID (Siti off-target di CRISPR con appaiamenti errati, inserzioni e delezioni) (disponibile su internet al sito crispr.bme.gatech.edu) compila tali similarità. Altri strumenti di bioinformatica includono, ma non sono limitati a, GUIDO, autoCOSMID, e CCTop.

Bioinformatica è stata utilizzata per minimizzare clivaggio off-target al fine di ridurre gli effetti dannosi di mutazioni e riarrangiamenti cromosomici. Studi su sistemi CRISPR/Cas9 hanno suggerito la possibilità di elevata attività off-target dovuta a una ibridazione non specifica del filamento guida a sequenze di DNA con appaiamenti di basi errati e/o rigonfiamenti o bulge, in particolare a

U001.D1.SM.61E

posizioni distali dalla regione PAM. Pertanto, è importante avere uno strumento di bioinformatica che sia in grado di identificare potenziali siti off-target che hanno inserzioni e/o delezioni tra il filamento guida di RNA e sequenze genomiche, in aggiunta a appaiamenti errati di basi. Lo strumento basato su bioinformatica, COSMID (Siti off-target di CRISPR con appaiamenti errati di basi, inserzioni e delezioni) è stato pertanto utilizzato per ricercare genomi per potenziali siti off-target di CRISPR (disponibile su internet al sito crispr.bme.gatech.edu). COSMID ha prodotto liste ordinate dei potenziali siti off-target sul numero e sulla posizione di appaiamenti errati, consentendo una scelta più informata di siti bersaglio, e evitando l'uso di siti con clivaggio off-target più probabile.

Sono stati impiegati pipeline di bioinformatica supplementari che ponderano l'attività su e/o off-target stimata di siti di targeting di gRNA in una regione. Altre caratteristiche che possono venire utilizzate per predire attività includono informazioni riguardo il tipo di cellula in questione, accessibilità a DNA, stato di cromatina, siti di legame a fattori di trascrizione, dati di legame a fattori di trascrizione, e altri dati di CHIP-seq. Sono ponderati fattori addizionali che predicano l'efficienza di editing, come relative posizioni e direzioni di coppie di gRNA, caratteristiche di sequenze locali e micro-omologie.

Esempio 8 – Test di guide preferite in cellule per attività on-target

I gRNA predetti avere la più bassa attività off-target saranno testati per attività on-target in cellule K562, e valutati per la frequenza di indel utilizzando TIDE.

TIDE è uno strumento internet per valutare rapidamente l'editing genomico

U001.D1.SM.61E

mediante CRISPR-Cas9 di un locus bersaglio determinato da un RNA guida (gRNA o sgRNA). Sulla base di dati di traccia di sequenze quantitative da due reazioni di sequenziamento capillare standard, il software TIDE quantifica l'efficacia di editing e identifica i tipi predominanti di inserzioni e delezioni (indel) nel DNA di un pool cellulare mirato. Si veda Brinkman *et al*, Nucl. Acids Res. (2014) per una spiegazione dettagliata e esempi. Un metodo alternativo è il Sequenziamento di Nuova Generazione (NGS), noto anche come sequenziamento ad elevata produttività, che è il termine onnicomprensivo per descrivere un numero di differenti tecnologie moderne di sequenziamento, incluse: sequenziamento Illumina (Solexa), sequenziamento Roche 454, sequenziamento Ion torrent: Proton/PGM, e sequenziamento SOLiD. Queste tecnologie recenti consentono di sequenziare DNA e RNA in modo molto più rapido e economico rispetto al sequenziamento Sanger precedentemente utilizzato, e quindi hanno rivoluzionato lo studio della genomica e biologia molecolare.

La trasfezione di cellule di coltura tissutale consente lo screening di differenti costrutti e un mezzo robusto per testare l'attività e la specificità. Linee di cellule di coltura tissutale, come K562 o HEK293T sono facilmente trasfettate e portano a una elevata attività. Queste o altre linee cellulari saranno valutate per determinare le linee cellulari che si abbinano con CD34+ e forniscono il surrogato migliore. Queste cellule saranno poi utilizzate per molti test di stadio iniziale. Per esempio, gRNA individuali per Cas9 di *S. pyogenes* possono venire trasfettati nelle cellule utilizzando plasmidi come, per esempio, CTx-1, CTx-2, o

U001.D1.SM.61E

CTx-3 descritti nella Figura 1A-1C, che sono adatti per l'espressione in cellule umane. In alternativa possono anche venire utilizzati vettori disponibili in commercio. Per la valutazione di Indel Freq dei gRNA di BCL11A descritti nel presente documento è stato impiegato un plasmide di espressione di Cas9 disponibile in commercio (GeneArt, Thermo Fisher). Alcuni giorni dopo (48 ore per questo esperimento), il DNA genomico è stato raccolto e il sito bersaglio è stato amplificato mediante PCR. L'attività di taglio è stata misurata dal tasso di inserzioni, delezioni e mutazioni introdotte mediante riparazione di NHEJ delle estremità di DNA libere. Sebbene questo metodo non possa differenziare sequenze riparate correttamente da DNA non clivato, il livello di taglio può essere calibrato dalla quantità di riparazione errata. L'attività off-target può venire osservata amplificando siti off-target putativi identificati e utilizzando metodi simili per rilevare clivaggio. Traslocazione può anche venire valutata utilizzando primer che fiancheggiano siti tagliati, per determinare se un taglio specifico e traslocazioni sono avvenuti. Sono stati sviluppati saggi non guidati che consentono il test complementare di clivaggio off-target, incluso guideseq. I gRNA o coppie di gRNA con attività significative possono poi essere seguiti in cellule coltivate per misurare la modulazione o inattivazione del sito ipersensibile a DNA +58 (DHS) all'interno della sequenza di controllo di trascrizione del gene BCL11A. Eventi off-target possono essere seguiti nuovamente. Similmente, cellule CD34+ possono essere trasfettate e il livello di modulazione o inattivazione del sito ipersensibile a DNA +58 (DHS) all'interno della sequenza di controllo di trascrizione del gene BCL11A e possibili eventi

U001.D1.SM.61E

off-target possono venire misurati. Questi esperimenti consentono una ottimizzazione di progettazione e veicolazione di nucleasi e donatori.

Esempio 9 – Test di guide preferite in cellule per attività off-target

I gRNAs aventi la migliore attività on-target da studi di sequenziamento di nuova generazione e TIDE nell'esempio riportato sopra saranno poi testati per attività off-target utilizzando sequenziamento di genoma intero. gRNA candidati saranno valutati in modo più completo in cellule CD34+ o iPSC.

Esempio 10 – Test di combinazioni preferite di gRNA in cellule

I gRNA aventi la migliore attività on-target dagli studi di sequenziamento di nuova generazione e TIDE e l'attività off-target più bassa saranno testati in combinazioni per valutare la dimensione della delezione risultante dall'uso di ciascuna combinazione di gRNA. Combinazioni potenziali di gRNA saranno valutate in cellule CD34+ umane primarie.

Per esempio, combinazioni di gRNA saranno testate per efficienza di delezione di tutta o una porzione della sequenza di controllo di trascrizione del gene BCL11A. Le combinazioni di gRNA saranno anche testate per l'efficienza di delezione di tutto o una porzione del sito ipersensibile a DNA +58 (DHS) del gene BCL11A.

Esempio 11 – Test di differenti approcci per editing genico di HDR

Dopo aver testato i gRNA sia per l'attività on-target che per l'attività off-target, strategie di modulazione/inattivazione e knock-in saranno testate per editing genico di HDR.

Per l'approccio di modulazione/inattivazione, DNA template donatore sarà

U001.D1.SM.61E

fornito come un oligonucleotide a singolo filamento, un oligonucleotide a doppio filamento (con sequenza PAM intatta/sequenza PAM mutata), una molecola lunga di DNA a singolo filamento (sequenza PAM intatta/sequenza PAM mutata) o una molecola lunga di DNA a doppio filamento (sequenza PAM intatta/sequenza PAM mutata). Il DNA template donatore comprenderà un gene BCL11A di tipo selvatico o cDNA comprendente un sito ipersensibile a DNA +58 (DHS) modificato (per esempio mutato). In aggiunta, il DNA template donatore sarà veicolato da AAV.

Per l'approccio di knock-in di cDNA, un DNA a singolo filamento o a doppio filamento può includere più di 40 nt della sequenza di controllo di trascrizione modificata del gene BCL11A. Il DNA a singolo filamento o a doppio filamento può includere più di 80 nt della sequenza di controllo di trascrizione modificata del gene BCL11A. il DNA a singolo filamento o a doppio filamento può includere più di 100 nt della sequenza di controllo di trascrizione modificata del gene BCL11A. Il DNA a singolo filamento o a doppio filamento può includere più di 150 nt della sequenza di controllo di trascrizione modificata del gene BCL11A. Il DNA a singolo filamento o a doppio filamento può includere più di 300 nt della sequenza di controllo di trascrizione modificata del gene BCL11A. Il DNA a singolo filamento o a doppio filamento può includere più di 400 nt della sequenza di controllo di trascrizione modificata del gene BCL11A. In alternativa, il DNA template sarà veicolato da AAV.

Per l'approccio di knock-in di cDNA, un DNA a singolo filamento o a doppio filamento può includere più di 40 nt del sito ipersensibile a DNA +58 (DHS)

U001.D1.SM.61E

modificato del gene BCL11A. Il DNA a singolo filamento o a doppio filamento può includere più di 80 nt del sito ipersensibile a DNA +58 (DHS) modificato del gene BCL11A. Il DNA a singolo filamento o a doppio filamento può includere più di 100 nt del sito ipersensibile a DNA +58 (DHS) modificato del gene BCL11A. Il DNA a singolo filamento o a doppio filamento può includere più di 150 nt del sito ipersensibile a DNA +58 (DHS) modificato del gene BCL11A. Il DNA a singolo filamento o a doppio filamento può includere più di 300 nt del sito ipersensibile a DNA +58 (DHS) modificato del gene BCL11A. Il DNA a singolo filamento o a doppio filamento può includere più di 400 nt del sito ipersensibile a DNA +58 (DHS) modificato del gene BCL11A. In alternativa, il DNA template sarà veicolato da AAV.

Esempio 12 – Ri-valutazione di combinazioni principali di CRISPR-Cas9/DNA donatore

Dopo aver testato le differenti strategie per editing genico di HDR, le combinazioni principali di CRISPR-Cas9/DNA donatore saranno valutate nuovamente in cellule umane primarie per efficienza di delezione, ricombinazione e specificità off-target. mRNA di Cas9 o RNP saranno formulati in nanoparticelle lipidiche per veicolazione, sgRNA saranno formulati in nanoparticelle o veicolati come AAV, e DNA donatore sarà formulato in nanoparticelle o veicolato come AAV.

Esempio 13 – Test *in vivo* in modello animale attinente

Dopo che le combinazioni CRISPR-Cas9/DNA donatore sono state nuovamente valutate, le formulazioni principali saranno testate *in vivo* in un

U001.D1.SM.61E

modello animale.

La coltura in cellule umane consente di effettuare test diretti sul bersaglio umano e sul genoma umano di fondo, come descritto sopra.

Valutazioni di efficacia e sicurezza preclinica possono venire osservate attraverso l'attecchimento o engraftment di cellule CD34+ di topo o umane in NSG o topi simili. Le cellule modificate possono essere osservate nei mesi dopo l'attecchimento o engraftment.

Esempio 14- Editing di cellule con vari gRNA

Cellule CD34+ di sangue periferico umano mobilizzate da donatori umani 1-3 sono state coltivate in Mezzo StemSpan senza siero con supplemento di espansione di CD34+ per due giorni. 100.000 cellule sono state lavate e elettroporate utilizzando mRNA di Cas9 con gRNA di Corfu Large (CLO), gRNA di Corfu Small (CSO), gRNA di HPFH5, gRNA di Kenya, sRNA di SD2 o sgRNA di SPY101. Le cellule sono state lasciate recuperare per due giorni prima di venire passate a un mezzo di differenziazione eritroide (IMDM+Glutamax supplementato con siero umano al 5%, insulina 10ug/ml, SCF 20ng/ml, IL-3 5ng/ml, EPO 3U/ml, dexamethasone 1uM, β -estradiolo 1uM, Olo-transferrina 330ug/ml e eparina 2U/ml). La percentuale di inserzioni/delezioni ("indels") è stata determinata per ciascuna delle cellule elettroporate con gRNA di Corfu Large (CLO), cellule elettroporate con gRNA di Corfu Small (CSO), cellule elettroporate con gRNA di HPFH5, cellule elettroporate con gRNA di Kenya, cellule elettroporate con sgRNA di SD2, e le cellule elettroporate con sgRNA di SPY101 (Figura 3), come descritto nelle

U001.D1.SM.61E

sezioni "Rilevamento di mutazioni on-target e off-target mediante sequenza" e "Saggi di rilevamento di misurazioni" descritte nel presente documento. Dopo la differenziazione di queste cellule per 12 giorni in mezzo di differenziazione eritroide, RNA è stato raccolto per valutare livelli di emoglobina mediante PCR quantitativa in tempo reale (Figure 4A-4C).

Singoli progenitori eritroidi sono stati generati utilizzando citometria a flusso un giorno dopo e coltivati nel mezzo di differenziazione eritroide per espandersi e crescere come colonie. Ciascuna colonia è stata divisa e raccolta 12 giorni post-separazione per analisi di DNA e RNA. Le colonie sorelle sono state raccolte 15 giorni post-separazione per l'analisi di proteine emoglobiniche. L'espressione di globine (rapporto di $\gamma/18sRNA$ o rapporto γ/α) è stata determinata mediante PCR quantitativa e comparata per ciascuna delle colonie eritroidi editate (Figure 5A-5B).

Esempio 15 – Test di sgRNA di SPY101

Tre esiti possibili di editing genico possono verificarsi all'interno dell'introne 2 del gene BCL11A quando viene utilizzato sgRNA di SPY101. Il primo esito di editing genico che può avvenire quando si utilizza sgRNA di SPY101 dà luogo a soltanto indel in entrambi gli alleli (Indel/Indel, Figura 6). Il secondo esito di editing genico che può verificarsi quando si utilizza sgRNA di SPY101 dà luogo a un clone con sia indel che sequenze di tipo selvatico nei due alleli (Indel/WT, Figura 6). Il terzo esito di editing genico che può avvenire quando si utilizza sgRNA di SPY101 porta a una colonia con sequenze di tipo selvatico in entrambi gli alleli (WT/WT, Figura 6).

U001.D1.SM.61E

Quando si utilizza sgRNA di SPY101, il 92% delle colonie eritroidi sono state editate. Per esempio, il 92% delle colonie eritroidi avevano alleli con indel (Figura 6).

L'espressione di γ -globina (rapporto di mRNA di γ/α globina o rapporto di mRNA di $\gamma/(\gamma+\beta)$ globina) è stata misurata in singole colonie eritroidi editate con SPY101 (Figure 7A-B). Le singole colonie eritroidi includevano colonie con indel biallelico o omozigote (indel/indel), colonie con un indel monoallelico o eterozigote (indel/WT), e colonie con sequenze di tipo selvatico in entrambi gli alleli (WT/WT). Le colonie eritroidi aventi indel erano in grado di esprimere livelli superiori di gamma globina rispetto ai cloni con sequenze di tipo selvatico in entrambi gli alleli (Figure 7A-B).

Esempio 16 – Strategia terapeutica per anemia falciforme (SCD) e β -talassemia

La tabella seguente (Tabella 4) fornisce informazioni in merito ai gRNA utilizzati negli Esempi 16-17

Tabella 4

gRNA	Nome	Sequenza	SEQ ID NO:
gRNA A	CL01	5'usgsusGUGCUGGCCCGCAACUUGUU UUAGAGCUAGAAAUAGCAAGUAAAA	71950

U001.D1.SM.61E

		UAAGGCUAGUCCGUUAUCAACUUGAA AAAGUGGCACCGAGUCGGUGCususus U3'	
gRNA B	CL08	5'cscscsACUCAAGAGAUUUGGUGGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71951
gRNA C	CS02	5'gsusasGACCACCAGUAAUCUGAGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71952
gRNA D	CS06	5'asgsusAUACCUCCAUACCAUGGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71953
gRNA E	HPFH5-15	5'csusgsUCUUAUUACCCUGUCAUGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71954
gRNA F	HPFH5-4	5'ascsusGAGUUCUAAAAUCAUCGGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71955

U001.D1.SM.61E

gRNA G	Kenya 02	5'gsuscsUUCAGCCUACAACAUACGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71956
gRNA H	Kenya 17	5'gsususAAGUUCAUGUCAUAGGAGUU UUAGAGCUAGAAAUAGCAAGUUAAAA UAAGGCUAGUCCGUUAUCAACUUGAA AAAGUGGCACCGAGUCGGUGCususus U3'	71957
gRNA I	SD2	5'csususGUCAAGGCUAUUGGUCAGUU UUAGAGCUAGAAAUAGCAAGUUAAAA UAAGGCUAGUCCGUUAUCAACUUGAA AAAGUGGCACCGAGUCGGUGCususus U3'	71958
gRNA J	SPY	5'csusasACAGUUGCUUUUAUCACGUUU UAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCusususU 3'	71959

a, g, u: residui di 2'-O-metile

s: fosforotioato

A, C, G, U: residui di RNA

La Tabella seguente (Tabella 5) fornisce informazioni legate ai bersagli a cui si fa riferimento negli Esempi 16-17.

Tabella 5

Bersaglio 1	Corfu Large
Bersaglio 2	Corfu Small
Bersaglio 3	HPFH5
Bersaglio 4	KENYA
Bersaglio 5	SD2
Bersaglio 6	SPY101

Una strategia terapeutica per SCD e β -talassemia utilizzava CRISPR/Cas9 per ricreare le stesse mutazioni genetiche che avvengono naturalmente in pazienti con HPFH. Cellule staminali ematopoietiche di pazienti sono state isolate,

U001.D1.SM.61E

queste cellule sono state trattate *ex vivo* con CRISPR/Cas9 per creare edit genetici di HPFH, e poi le cellule editate sono state reintrodotte nel paziente. Le cellule staminali geneticamente modificate davano origine a eritrociti contenenti livelli sufficienti di HBF per ridurre significativamente la severità di sintomi di malattia. Un numero di edit genetici sono stati prioritizzati sulla base del grado di sovraregolazione di HBF osservata in naturale, la capacità di ricreare questi edit a una elevata efficienza utilizzando CRISPR/Cas9, e l'assenza di editing off-target.

Sequenze di RNA guida (gRNA) candidate sono state selezionate computazionalmente e poi vagliate per efficacia di editing on-target in cellule CD34+. Nella Figura 8 sono mostrati i risultati di un tale screening. gRNA sono stati identificati con elevato (>70%) e costante editing di bersaglio attraverso molteplici campioni di donatori. Ciascun donatore di cellule CD34+ è rappresentato da un unico simbolo (▲, ★, ●) e l'efficienza di editing on-target per ciascun donatore è misurata due volte.

gRNA candidati sono stati vagliati in cellule CD34+ per attività off-target esaminando centinaia di siti computazionalmente identificati per essere più simili nella sequenza al sito on-target desiderato, e quindi avere il potenziale più alto per attività off-target. Le Figure 9A-B mostrano l'approccio sperimentale (Figura 9A) e risultati (Figura 9B) per ciascuno dei gRNA testati nella Figura 8. La maggior parte dei gRNA non presentava alcuna attività off-target rilevabile, anche a livello di siti predetti. Soltanto gRNA C e gRNA G mostrano attività off-target. Molteplici sonde sono state utilizzate per ciascun sito predetto per

U001.D1.SM.61E

aumentare la sensibilità del saggio.

gRNA candidati sono stati utilizzati per ricreare HPFH specifica o altre modificazioni in cellule eritroidi ottenute da pazienti affetti da SCD e β -talassemia, nonché da donatori sani. Dopo differenziazione eritroide, livelli di trascritti di globine sono stati misurati per valutare l'aumento in γ -globina rispetto a α - o β -globina. Come mostrato nelle Figure 10A-B, livelli di mRNA di γ -globina superiori al 30% sono stati osservati in cellule di paziente editate con gRNA per ricreare HPFH Bersaglio 5 e 6. Campioni di pazienti affetti da SCD e β -talassemia esibivano un aumento assoluto maggiore in γ -globina rispetto a quelli da donatori sani, coerente con l'osservazione di HbF superiore in pazienti rispetto a carrier eterozigoti con HPFH. Il livello di fondo per cellule trattate con mock da ciascun donatore era sottratto dai valori mostrati. I dati rappresentano un singolo esperimento, eccetto per dati di pazienti con SCD che rappresentano la media di 3 campioni di donatori differenti. L'efficienza di editing era simile per tutti gli esperimenti.

Per assicurare che efficienze di editing nelle popolazioni CD34+ in bulk erano rappresentative di quelle HSC ripopolanti a lungo termine (LT-HSC), cellule CD34+ in bulk sono state smistate in sottopopolazioni specifiche e saggate per efficienza di editing on-target come mostrato nelle Figure 11A-C. E' stata osservata una elevata efficienza di editing nella popolazione LT-HSC. Esperimenti sono stati effettuati utilizzando SPY101 e proteina Cas9 attraverso 4 donatori. Le barre rappresentano media \pm SEM. LT-HSC, cellula staminale ematopoietica a lungo termine; MPP, Progenitore multipotente; MLP,

U001.D1.SM.61E

Progenitore multilinfoide; *CMP*, Progenitore mieloide comune; *MEP*, Progenitore di megacariociti eritrociti; *GMP*, Progenitore di granulociti macrofagi.

Studi di attecchimento o engraftment *in vivo* sono stati effettuati in topi immunocompromessi per confermare che le HSPC sottoposte a editing genico mantengono il potenziale per ripopolazione a lungo termine sul sistema ematopoietico. Cellule CD34+ umane da donatori sani erano non trattate, non editate, o sottoposte a editing genico utilizzando SPY101 gRNA e introdotte in topi NSG. Come mostrato nella Figura 12, la presenza di livelli simili di cellule hCD45RA+ (a 8 settimane post-attecchimento o engraftment) in topi iniettati con HSPC non trattate/non editate e topi iniettati con HSPC sottoposte a editing genico di SPY101 ha confermato che le cellule editate con SPY101 mantenevano il potenziale di attecchimento o engraftment. Punti dati rappresentano animali individuali e illustrano la percentuale di cellule vive che erano CD45RA+ umane. Media \pm SD. "Non trattate" rappresenta HSPC che non erano elettroporate e iniettate in topi immunocompromessi. "Non editate" rappresenta HSPC che erano elettroporate, ma non sottoposte a editing genico e iniettate in topi immunocompromessi. "SPY101" rappresenta HSPC che erano elettroporate con Cas9 e SPY101 gRNA e iniettate in topi immunocompromessi.

Lo sviluppo del processo è stato iniziato in una struttura capace di GMP in preparazione per studi clinici. Come mostrato nella Figura 13, nessuna perdita significativa di efficacia di editing genico è stata osservata su scala clinica in un processo compatibile con GMP. I dati erano la media attraverso 4 o più

U001.D1.SM.61E

esperimenti, \pm SD.

Studi di GLP/tossicologia sono stati iniziati per i nostri candidati guida, come mostrato nella Figura 14. Due studi separati in topi NSG consentiranno una caratterizzazione comprensiva di biodistribuzione e tossicologia di cellule CD34+ editate.

Esempio 17 – Strategia terapeutica per anemia falciforme (SCD) e β -talassemia

Risultati da ricreazione di sei differenti varianti di HPFH, o “bersagli o target” di editing, in cellule CD34+ umane di mPB sono mostrati nelle Figure 16A-B e Figure 17. Le cellule CD34+ sono state trattate con CRISPR/Cas9, differenziate in eritrociti, e poi saggiate per l’espressione di HBF mRNA e proteine in bulk (Figure 16A-B) e colonie (Figura 17), utilizzando un processo sperimentale mostrato nella Figura 15.

I risultati presentati nelle Figure 16A-B erano da 3 donatori differenti per bersagli 1-3, e 7 donatori differenti per bersagli 4-6. Il livello di fondo per cellule trattate con mock è stato sottratto. I dati sono media \pm SEM. L’analisi in bulk ha confermato la sovraregolazione di HBF e ha consentito la prioritizzazione di bersagli che dimostravano i livelli più alti di HBF.

L’analisi clonale presentata nelle Figure 17 ha consentito la conferma che edit genetici causati da CRISPR/Cas9 erano realmente la causa dell’aumento di HBF a livello delle cellule individuali. Livelli di trascritti di mRNA erano misurati mediante qRT-PCR. I dati sono media \pm SEM.

I bersagli 5 e 6 presentavano i livelli più alti di HBF e sono stati ulteriormente

U001.D1.SM.61E

analizzati nelle Figure 18A-B. I dati sono media \pm SEM. WT denota colonie che non mostrano evidenza di editing genico, Eterozigote o Het denota colonie con un allele editato, e Omozigote o Omo denota colonie con entrambi gli alleli editati. Evidenze nelle Figure 16A-B, 17, e 18A-B supportano la relazione causale tra gli edit genetici prodotti, e la sovraregolazione desiderata di HBF, fornendo ulteriore validazione per la strategia terapeutica proposta.

Esempio 18 – Test di RNA guida preferiti in cellule per attività on-target

Cellule CD34+ di sangue periferico umano (mPB) mobilizzate da quattro donatori indipendenti sono state coltivate in mezzo CellGro® privo di siero comprendente fattore di cellule staminali (SCF) umane ricombinanti 100 ng/ml, ligando di Fit 3 (FLT3L) umano ricombinante 100 ng/ml, e Trombopoietina (TPO) 100 ng/ml. 200.000 cellule per donatore sono state lavate e elettroporate utilizzando un elettroporatore Lonza senza alcun componente di editing di CRISPR/Cas9 (campione di elettroporazione mock), con GFP gRNA e proteina Cas9 come controllo negativo (GFP), con SPY101 gRNA e proteina Cas9 (SPY), con SD2 gRNA e proteina Cas9 (SD2), o duplici BCL11A Esone 2 gRNA e proteina Cas9 (Ex2). La proteina Cas9 ricombinante codifica per Cas9 di *S. pyogenes* fiancheggiata da due sequenze di localizzazione nucleare di SV40 (NLS). Questi esperimenti sono stati effettuati utilizzando un rapporto in peso di ribonucleoproteina (RNP) 1:1 di gRNA:Cas9. Il gRNA di SPY101 crea una distruzione di InDel di sito di legame a DHS+58 Gata1 nell'Introne 2 del locus di BCL11a. Il SD2 gRNA crea InDel e una delezione di 4,9 Kb nel locus di beta-globina umana. La delezione di 4,9 Kb è situata a monte di HBG1 e

U001.D1.SM.61E

include l'intera sequenza di HBG2. La delezione di 4,9 Kb inizia 168 bp 5' alla sequenza codificante di HBG2 e termina 168 bp 5' alla sequenza codificante di HBG1. gRNA di Esone 2 creano una delezione di 196 bp sull'Esone 2 del locus di BCL11A e servono da controllo positivo. Cellule CD34+ umane di mPB che non sono state elettroporate servono da controllo negativo (senza EP).

Dopo elettroporazione, cellule CD34+ di mPB sottoposte a editing genico sono state lasciate recuperare per due giorni prima di essere passate a un mezzo di differenziazione eritroide (IMDM+L-glutamina supplementato con siero umano al 5%, insulina 10 ug/mL, SCF 20 ng/mL, IL-3 5 ng/mL, EPO 3 U/mL, dexametasone 1uM, olo-transferrina 330ug/ml e eparina 2 U/mL). Le cellule CD34+ di mPB sottoposte a editing genico erano differenziate in eritrociti e ulteriormente testate tramite analisi di TIDE, analisi di ddPCR, analisi di PCR quantitativa in tempo reale, FACS, e LC-MS (Figure 20A-B, 21A-D, 22A-B, e 23A-D). Il processo sperimentale globale è mostrato nella Figura 19.

Analisi di TIDE / Analisi di ddPCR

DNA genomico è stato isolato e testato per ciascuno dei campioni di cellule CD34+ umane di mPB sottoposte a editing genico cresciuti in mezzo di differenziazione. DNA genomico è stato isolato dalle cellule i giorni 1, 11, 13 e 15 post-differenziazione. Il DNA genomico è stato analizzato tramite analisi di TIDE, che è uno strumento internet per valutare rapidamente l'editing genomico mediante CRISPR-Cas9 di un locus bersaglio determinato da un RNA guida (gRNA o sgRNA). I risultati presentati nelle Figure 20A-B erano da 4 donatori differenti e dimostravano che la percentuale di editing genico era mantenuta per

U001.D1.SM.61E

tutta la differenziazione eritroide *ex vivo* di cellule CD34+ di mPB editate con SD2 gRNA (Figura 20B) e Cellule CD34+ di mPB editate con SPY101 gRNA (Figura 20A). I dati sono media \pm SD.

Il DNA genomico è stato anche analizzato tramite analisi di ddPCR per rilevare la frequenza di delezione di 4,9 kb con trattamento con SD2. I risultati presentati nella Figura 20B erano da 4 donatori differenti e mostravano che la percentuale di editing genico era mantenuta per tutta la differenziazione eritroide *ex vivo* di cellule CD34+ di mPB editate con SD2 gRNA (Figura 20B). I dati sono media \pm SD.

Analisi di PCR quantitativa in tempo reale

mRNA è stato isolato e testato per ciascuno dei campioni di cellule CD34+ umane di mPB sottoposte a editing genico cresciuti in mezzo di differenziazione. L'isolamento di mRNA è stato effettuato nei giorni 11 e 15 post-differenziazione. L'espressione globinica (rapporto di γ/α e rapporto di $\gamma/(\gamma+\beta)$) è stata determinata mediante PCR quantitativa in tempo reale e comparata per ciascuna delle cellule CD34+ umane di mPB editate con SD2 gRNA e cellule CD34+ umane di mPB editate con SPY101 gRNA (Figure 21A-D). I risultati presentati nelle Figure 21A-D erano da quattro donatori differenti e dimostravano un aumento nel trascritto di γ -globina in cellule CD34+ umane di mPB editate con SD2 gRNA e cellule CD34+ umane di mPB editate con SPY101 gRNA a paragone con controllo negativo. I dati sono media \pm SD.

FACS / LC-MS

Cellule CD34+ umane di mPB editate con SD2 gRNA e cellule CD34+ umane

U001.D1.SM.61E

di mPB editate con SPY101 gRNA sono state cresciute in mezzo di differenziazione per 15 giorni. Cellule CD34+ umane di mPB sono state anche editate con duplici gRNA di esone 2 di BCL11A (Ex2) o GFP gRNA e cresciute in mezzo di differenziazione per 15 giorni. Alcune cellule CD34+ umane di mPB non erano editate con alcun componente di editing di CRISPR/Cas9 (campione di elettroporazione mock) e alcune cellule CD34+ umane di mPB non erano elettroporate (senza EP). Le cellule vive sono state colorate con Glicoforina A, un marcatore di maturazione eritroide. Le cellule sono state poi fissate e permeabilizzate. Le cellule fissate sono state colorate con anticorpo coniugato con fluoroforo per ciascuna subunità di globina. Le cellule colorate sono state poi analizzate tramite FACS, un esempio di γ -globina rappresentato nella Figura 22A. Le intensità fluorescenti mediane medie per γ -globina da 4 donatori differenti sono riportate nella Figura 22B (media \pm SEM) e dimostravano una sovraregolazione in γ -globina in cellule CD34+ umane di mPB editate con SD2 gRNA e cellule CD34+ umane di mPB editate con SPY101 gRNA.

Cellule CD34+ umane di mPB editate con SD2 gRNA e cellule CD34+ di mPB editate con SPY101 gRNA sono state cresciute in mezzo di differenziazione per 15 giorni. Cellule CD34+ umane di mPB sono state anche editate con duplici gRNA di Esone 2 di BCL11A (Ex2) o GFP gRNA e cresciute in mezzo di differenziazione per 15 giorni. Alcune cellule CD34+ umane di mPB non erano editate con alcun componente di editing di CRISPR/Cas9 (campione di elettroporazione mock) e alcune cellule CD34+

U001.D1.SM.61E

umane di mPB non erano elettroporate (senza EP). Cromatografia liquida – spettrometria di massa (LC-MS) è stata utilizzata per rilevare monomeri globinici denaturati (Figure 23A-D). I risultati presentati nelle Figure 23A-D erano da 4 donatori differenti e hanno anche dimostrato una sovraregolazione di γ -globina in cellule CD34+ di mPB editate con SD2 gRNA e cellule CD34+ di mPB editate con SPY101 gRNA. I dati sono media \pm SD.

Esempio 19 – Test di RNA guida preferiti in cellule per attività fuori-bersagliamento

Mentre editing off-target del genoma è fondamentale per una terapia soddisfacente, il rilevamento di qualsiasi evento di editing off-target è una componente importante per garantire la sicurezza del prodotto. Un metodo per rilevare modificazioni di siti off-target implica l'arricchimento per regioni del genoma che sono più simili al sito on-target tramite sequenziamento di cattura di ibridi e quantificazione di qualsiasi indel che è rilevato.

Sequenziamento di cattura di ibridi è un metodo che quantifica edit off-target in cellule editate con CRISPR-Cas9 e DNA. Dettagli relativi al metodo di sequenziamento di cattura di ibridi sono i seguenti:

MATERIALI E METODI

Materiali e fonti

1.1.1. DNA genomico

Dato che il fine di questo metodo è determinare se editing mediante CRISPR-Cas9 è avvenuto a livello di siti off-target nel genoma, vengono tipicamente utilizzati almeno due campioni – campioni trattati e controllo (non trattati,

U001.D1.SM.61E

elettroporati con mock, etc.). Ciascun campione ha DNA genomico (gDNA) estratto mediante un metodo appropriato e quel gDNA è ibridato con le librerie di cattura di ibridi (1.1.2) seguite dalla parte restante del protocollo come descritto sotto.

1.1.2. Librerie di cattura di ibridi

Librerie di cattura di ibridi come descritte in (1.2.2) sono generate fornendo una lista di fino a 57.000 sequenze esca o bait di oligonucleotidi 120-mer che sono poi sintetizzate come un kit di cattura di ibridi SureSelect XT personalizzato.

1.2 Metodi

1.2.1. Algoritmi di rilevamento di siti off-target

Per determinare i siti che hanno maggiore probabilità di avere editing off-target sono stati utilizzati algoritmi con differenti caratteristiche per garantire la copertura di un'ampia gamma di siti off-target.

1.1.1.1. CCTop

Per una data sequenza guida CCTop utilizza l'algoritmo di mappatura di sequenze Bowtie 1 per cercare il genoma per siti off-target con fino a 5 mismatch tra il sito e la guida. Si fa riferimento a questo sito come "siti off-target omologhi" (anziché "siti off-target predetti") dal momento che soltanto omologia di sequenza è utilizzata per determinare i potenziali siti off-target nel genoma. Questi 5 mismatch sono limitati a più di 2 mismatch nella regione seed di allineamento di 5 basi più vicina all'estremità di PAM delle sequenze. L'algoritmo CRISPOR (1.2.1.2) non ha la limitazione nella regione seed e

U001.D1.SM.61E

quindi complementa CCTop.

1.2.1.1. COSMID

Dal momento che alcuni siti di clivaggio di Cas9 off-target possono avere brevi indel (indicati anche come rigonfiamenti o bulge) tra loro stessi e la guida, la Richiedente effettua una ricerca anche con l'algoritmo COSMID che può rilevare siti off-target con indel (tipicamente limitati fino a 2 indel) e quindi complementa la ricerca effettuata con CCTop.

1.2.1.2. CRISPOR

CRISPOR è uno strumento che implementa molte funzioni differenti di punteggio on- e off-target di CRISPR pubblicate al fine di comparare vari metodi. Esso utilizza l'algoritmo BWA per cercare sequenze guida contro il genoma per trovare i loro siti off-target. Questo differisce dall'algoritmo Bowtie 1 utilizzato in CCTop e consente una ricerca che è leggermente più permissiva nel senso che mismatch vicino alla regione non sono limitati a 2 di 5 basi come in CCTop.

1.2.1.3. PAM

Per impostazione predefinita, screening vengono effettuati con una ricerca per guide con NGG o NAG PAM dal momento che essi hanno alcune delle attività maggiori. Screening in stadio avanzato possono includere più PAM per garantire che non venga mancato alcun sito off-target, neppure quelli con attività molto bassa.

1.2.1.4. Combinazione di algoritmi

Le guide emesse da ciascun algoritmo sono unite insieme per eliminare siti off-

U001.D1.SM.61E

target identici e introdotte nel componente di progettazione di esche o bait di cattura di ibridi.

1.2.2. Esche di cattura di ibridi

1.2.2.1 Progettazione

La lista di siti prodotta dagli algoritmi di rilevamento di siti off-target (1.2.1.) è poi utilizzata per generare sonde di cattura di ibridi che arricchiranno ciascuno dei siti off-target nei campioni di gDNA immessi. Sebbene un'esca o bait possa essere sufficiente per arricchire in modo soddisfacente una sequenza di DNA bersaglio, numerose esche o bait sono generalmente progettate e inserite lungo il sito bersaglio (Figura 24) al fine di rendere più probabile che un'esca tiri giù una regione bersaglio anche se essa è fiancheggiata su un lato da una sequenza ripetitiva che può essere difficile da legare specificamente. Esche di cattura di ibridi (120-meri, porzioni di colore scuro) sono introdotte attraverso un'esca (20-meri, porzione leggera denotata dall'asterisco *) (Figura 24).

1.2.3. Sequenziamento

Dopo l'arricchimento di cattura di ibridi, un sequenziamento è effettuato sul sequenziatore Illumina HiSeq con letture di 125 bp a estremità appaiate e una dimensione di inserto di 175 bp. Il sequenziamento è tipicamente effettuato per bersagliare una profondità di copertura con bersagli aventi 5 letture rilevate da un evento di frequenza minimo. Per rilevare, per esempio, eventi di indel allo 0,5%, sequenziamento a copertura 1000x è effettuato in modo tale che un evento allo 0,5 potrebbe avere 5 letture.

1.2.4. Efficacia di esche o bait

U001.D1.SM.61E

In un esperimento tipico si trova che esche coprono la grande maggioranza dei siti bersaglio con livelli elevati di copertura di sequenziamento. Esistono alcune limitazioni alla copertura di sequenziamento che può essere raggiunta da metodi di sequenziamento di nuova generazione (NGS) dovute a: elevata o bassa percentuale di GC, sequenze a bassa complessità, bassa affinità per esca, non specificità di esca e altre ragioni. Il reale potere di rilevare indel in un esperimento è stimato calcolando il potere di campionamento di differente copertura di sequenziamento per siti con differenti frequenze di indel reali. Generalmente, una aumentata copertura di sequenza fornisce un aumentato potere di rilevare siti con indel a bassa frequenza. Per esempio, se un sito ha una copertura di sequenziamento di 2500x, cattura di ibridi avrà un potere del 99% di vedere siti con frequenza di indel allo 0,4%, e un potere del 94% di vedere siti con frequenza di indel allo 0,3% (Figura 25).

1.2.5. Quantificazione

Dati di sequenziamento sono allineati con l'algoritmo BWA utilizzando parametri predefiniti al build hg38 di genoma umano. Per ciascun sito off-target potenziale, tutti gli indel all'interno di 3 bp del potenziale sito di clivaggio di Cas9 sono contati e divisi per la copertura a livello del sito tagliato, e ciò fornisce una quantità di indel a livello di un particolare sito tagliato.

1.2.6. Valutazione statistica di siti tagliati significativi

Vari eventi possono portare a indel che non sono un risultato del rilevamento di CRISPR-Cas9 a livello di siti attraverso tutto il genoma: varianti di indel di linea germinale o polimorfismi, regioni suscettibili a rotture genomiche, regioni con

U001.D1.SM.61E

run omopolimerici, e regioni che sono altrimenti difficili da sequenziare.

1.2.6.1. Siti esclusi da analisi

La Richiedente ha escluso dall'analisi: qualsiasi sito con un indel di "linea germinale" su una base donatore a donatore (il donatore ha una frequenza di indel > 30% in ogni campione), qualsiasi sito di cromosoma Y in campioni femminili, e qualsiasi sito con 0 copertura.

1.2.6.2. Test statistico

Per valutare se un indel osservato a livello di un potenziale sito off-target è realmente un evento indotto da CRISPR-Cas9, la Richiedente ha testato se i campioni trattati con Cas9 e guida hanno una frequenza significativamente superiore di indel rispetto ai campioni non trattati utilizzando sia il test di Mann-Whitney Wilcoxon che il test t di Student. Se uno di questi test è significativo ($p < 0,05$) si considera il sito contrassegnato per follow-up con PCR per determinare se vi è editing significativo. Per garantire di aver contrassegnato siti per follow-up nel modo più aggressivo possibile, la Richiedente non effettua la correzione di test a ipotesi multiple, che diminuirebbe il numero di siti che risultano essere significativi.

La Richiedente stabilisce anche una analisi di controllo negativo, dove essa ripete l'analisi, eccetto che essa cerca siti con una maggiore frequenza di indel nel campione non trattato rispetto al campione trattato. Biologicamente non vi è alcuna ragione per cui ci si aspetterebbe di trovare "corrispondenze reali" in questa analisi, che fornisce informazioni empiriche riguardo il numero di falsi positivi che ci si aspetta di trovare in questo set di dati che possono essere

U001.D1.SM.61E

attribuibili a rumore di fondo. Inoltre, si può espandere ciò in una distribuzione nulla empirica sfruttando due campioni di controllo negativo addizionali, incluse cellule elettroporate senza Cas9 o guida, e cellule elettroporate con Cas9 e una guida di GFP. Testando corrispondenti in campioni che sono “meno trattati” rispetto a campioni che sono “più trattati” si determina una distribuzione nulla empirica conservativa di corrispondenze false positive, che possono essere utilizzate per informare la credibilità delle corrispondenze nella nostra analisi originale per campioni trattati vs non trattati.

Cellule CD34+ umane di mPB editate con SD2 gRNA e cellule CD34+ umane di mPB editate con SPY101 gRNA sono state analizzate tramite un metodo di sequenziamento di cattura di ibridi descritto nel presente documento. I risultati presentati nelle Figure 26-27 erano da 3-4 donatori differenti e dimostravano 0 siti off-target con evidenza di taglio in cellule CD34+ di mPB sottoposte a editing genico, che erano editate con SD2 gRNA (Figura 27) e cellule CD34+ di mPB sottoposte a editing genico editate con SPY101 gRNA (Figura 26). La frequenza di indel per il turno 1 è superiore a ($>$) 0,5%. La frequenza di indel per il turno 2 è superiore a ($>$) 0,2%.

Esempio 20 – Esperimenti di attecchimento o engraftment

Cellule CD34+ di sangue umano periferico (mPB) umano mobilizzate sono state isolate da donatori sani utilizzando microsfere CliniMACS CD34 con CliniMACS Prodigy (Miltenyi Biotec) e coltivate in mezzo CellGro® senza siero comprendente fattore di cellule staminali umane ricombinanti 100 ng/ml (SCF), ligando di Fit 3 umano ricombinante (FLT3L) 100 ng/ml, e

U001.D1.SM.61E

Trombopoietina (TPO) 100 ng/ml. Le cellule sono state poi elettroporate utilizzando un dispositivo Maxcyte® seguendo le istruzioni del produttore con uno dei seguenti: un vettore vuoto che non contiene alcun componente di editing di CRISPR/Cas9 (campione di elettroporazione mock), con GFP gRNA e proteina Cas9 come un controllo negativo (GFP), con SPY101 gRNA e proteina Cas9 (SPY101), o con SD2 gRNA e proteina Cas9 (SD2). La proteina Cas9 ricombinante codifica per Cas9 di *S. pyogenes* fiancheggiata da due sequenze di localizzazione nucleare di SV40 (NLS). Questi esperimenti sono stati effettuati utilizzando un rapporto in peso di ribonucleoproteina (RNP) tra gRNS e Cas9 di 1:1.

Ciascuna delle cellule umane CD34+ di mPB sottoposte a editing genico è stata iniettata attraverso la vena caudale in 16 topi immunodeficienti ("NSG" o NOD scid gamma -NOD) per dimostrare capacità di homing e di attecchimento o engraftment. NSG è un ceppo di topi da laboratorio consanguinei e tra i più immunodeficienti descritti fino ad oggi; si veda, per esempio, Shultz *et al.*, *Nat. Rev. Immunol.* 7(2): 118-130 (2007). Dettagli relativi all'esperimento di attecchimento o engraftment sono presentati nella Figura 28. 8 settimane dopo l'iniezione i topi NSG sono stati sottoposti a prelievo di sangue e il sangue periferico è stato analizzato tramite FACS per cellule vive CD45RA+ umane e CD45+ di topo. 16 settimane dopo l'iniezione, i topi NSG sono stati sacrificati, e il midollo osseo, la milza e il sangue periferico sono stati analizzati tramite FACS per cellule vive CD45RA+ umane e CD45+ di topo. L'attecchimento o engraftment di cellule umane CD34+ di mPB in topi NSG irradiati in tutti i

U001.D1.SM.61E

gruppi di trattamento è stato osservato da tutti e tre i donatori sani. Cellule CD45RA+ umane sono state rilevate utilizzando FACS in tutti i 3 organi ematopoietici da tutti i donatori. Cellule di controllo CD34+ non trasfettate esibivano percentuali migliori di attecchimento o engraftment. Tutti i gruppi di cellule trasfettate avevano percentuali simili di attecchimento o engraftment, incluso il gruppo trasfettato con mock attraverso tutti i tre donatori sani. In generale, l'aggiunta di Cas9-gRNA RNP non influiva sull'attecchimento o engraftment rispetto al controllo di trasfezione mock (Figure 29A-E e Figure 30). Punti dati nelle Figure 29A-E rappresentano topi individuali e illustrano la percentuale di cellule vive che erano cellule CD45RA+ umane. I dati sono in media + SEM.

Esempio 21 – Valutazione di efficienza e efficacia di editing di SPY101 utilizzando Cas9 RNP

Al fine di conseguire la massima efficacia utilizzando CRISPR-Cas9 per trattare SCD e β -talassemia utilizzando SPY101, la Richiedente ha valutato due formati differenti di Cas9, Cas9 mRNA o proteina Cas9, per la loro efficienza, efficacia e tossicità di editing in cellule CD34+ umane da sangue periferico mobilizzato (mPB). La Richiedente ha comparato fonti per Cas9 mRNA a proteina Cas9 mediante elettroporazione di Cas9 mRNA e SPY101 gRNA o proteina Cas9 complessata con SPY101 gRNA (come complesso ribonucleoproteico (RNP) in cellule CD34+ umane di mPB e valutate per la loro efficienza di editing e vitalità cellulare a 48 ore post-elettroporazione. La Richiedente ha comparato varie fonti per Cas9 mRNA a proteina Cas9 e ha

U001.D1.SM.61E

trovato che, mentre si possono raggiungere livelli simili di efficienza di editing tra alcuni Cas9 mRNA a proteina Cas9 (Figura 31), la maggior parte avevano una vitalità cellulare significativamente inferiore rispetto a campioni di controllo (campioni con nessuna elettroporazione (No EP) o nessuna elettroporazione su substrato (EP mock)) mostrati nelle Figure 32A-B. Ciò indica che Cas9 RNP è il miglior formato da utilizzare per una efficiente veicolazione di Cas9 e gRNA in cellule CD34+ umane di mPB.

La Richiedente ha successivamente comparato differenti fonti di proteina Cas9 nonché proteina Cas9 con un numero variabile di segnali di localizzazione nucleare (NLS) all'estremità N o C dato che questo può influire su una efficiente localizzazione di Cas9 nel nucleo per fornire editing. Come mostrato nelle Figure 33A-C, la Richiedente ha trovato che proteina Cas9 Aldevron con un NLS sia all'estremità N che all'estremità C dava la migliore efficienza di editing senza alcuna variazione nella vitalità cellulare.

Successivamente, la Richiedente ha compilato efficienza di editing di SPY101 esaminata attraverso vari donatori di CD34+ di mPB umano utilizzando Cas9 mRNA o proteina Cas9 (Feldan o Aldevron scelta dall'esempio precedente) e osservato che proteina Cas9 Aldevron portava alla più elevata efficacia di editing (Figura 34A). Inoltre, la Richiedente è stata in grado di conseguire tassi simili di efficienza di editing nella produzione compatibile con GMP su scala clinica utilizzando proteina Cas9 (Figura 34B).

La Richiedente ha successivamente esaminato l'efficacia di SPY101 attraverso numerosi donatori di CD34+ derivati da mPB (Figure 35A-B) o midollo osseo

U001.D1.SM.61E

(BM, Figure 36A-B). La Richiedente ha osservato che, per tutto il suo processo di ottimizzazione, ha conseguito una migliore efficacia, misurata come espressione di γ -globina a α -globina o a globine simili a β -globina (β -globina + γ -globina) mediante PCR quantitativa in tempo reale in cellule CD34+ differenziate eritroidi.

La Richiedente ha poi indagato se SPY101 fosse efficace in cellule ottenute da pazienti affetti da SCD o β -talassemia. Cellule mononucleari di sangue periferico da donatori sani o pazienti sono state elettroporate con SPY101 Cas9 RNP e differenziate eritroidi simili agli esempi riportati sopra prima dell'estrazione di RNA per misurare l'espressione di γ -globina. La Richiedente ha osservato che SPY101 era realmente efficace nell'aumento di γ -globina in campioni di pazienti (Figure 37A-B).

Al fine di comprendere meglio una relazione tra genotipo e fenotipo in cellule eritroidi editate con SPY101, la Richiedente ha effettuato una analisi su singole colonie simile all'Esempio 15 con Cas9 RNP e ha trovato che questo aumenta una frazione maggiore di colonie editate bi-alleleliche utilizzando Cas9 RNP rispetto a Cas9 mRNA (Figure 38A-B). Inoltre, la scomposizione dettagliata di colonie non editate, distruzione mono-allelelica di sito di legame a GATA1 bersagliata da SPY101 e distruzione bi-allelelica di sito di legame a GATA1 rivelavano un'efficacia dipendente da dose di SPY101, misurata come aumento di γ -globina a paragone con cellule trattate con GFP gRNA di controllo (Figure 39A-B).

Per esaminare la percentuale di cellule esprimenti γ -globina, la Richiedente ha

U001.D1.SM.61E

effettuato analisi di FACS in cellule CD34+ da mPB umano editate con SPY101 Cas9 RNP. A paragone con cellule trattate con GFP gRNA di controllo, si osserva una maggiore percentuale di cellule differenziate eritroidi esprimenti γ -globina (Figure 40A-D), nonché un aumento di espressione di γ -globina per cellula (Figure 40E) in cellule trattate con SPY101.

RIVENDICAZIONI

1. RNA guida monomolecolare (sgRNA) comprendente la sequenza di acido nucleico di SEQ ID NO: 71.959, in cui il sgRNA comprende tre residui di 2'-O-metil-fosforotioato a ciascuna delle sue estremità 5' e 3'.
2. RNA guida monomolecolare (sgRNA) comprendente la sequenza di acido nucleico di SEQ ID NO: 71.959, in cui le basi 1-3 e 97-99 sono nucleotidi 2'-O-metil fosforotioato modificati.
3. Acido nucleico comprendente una sequenza nucleotidica codificante il sgRNA della rivendicazione 1 o 2.
4. Vettore comprendente l'acido nucleico della rivendicazione 3.
5. SgRNA secondo la rivendicazione 1 o 2, acido nucleico secondo la rivendicazione 3, o vettore secondo la rivendicazione 4, per l'uso nel trattamento di un paziente affetto da una emoglobinopatia.
6. SgRNA per uso secondo la rivendicazione 5, acido nucleico per uso secondo la rivendicazione 5, o vettore per uso secondo la rivendicazione 5, in cui l'emoglobinopatia è anemia falciforme o drepanocitosi o talassemia.
7. SgRNA per uso secondo la rivendicazione 6, acido nucleico per uso secondo la rivendicazione 6 o vettore per uso secondo la rivendicazione 6, in cui l'emoglobinopatia è una talassemia e la talassemia è scelta nel gruppo costituito da α , β , δ , γ , e loro combinazioni, opzionalmente inoltre in cui la talassemia è β -talassemia.
8. Metodo *ex vivo* di editing di un gene di linfoma a cellule B 11A (BCL11A) in una cellula umana mediante editing genomico, il metodo comprendendo:

U001.D1.SM.61E

L'introduzione nella cellula umana di una o più acido desossiribonucleico (DNA) endonucleasi o di un acido nucleico comprendente una sequenza nucleotidica codificante le una o più DNA endonucleasi o un vettore comprendente detto acido nucleico per effettuare una o più rotture a singolo filamento (SSB) o rotture a doppio filamento (DSB) all'interno del o vicino al gene BCL11A o di un'altra sequenza di DNA che codifica un elemento regolatorio del gene BCL11A, che ha come risultato una delezione, una modulazione o una inattivazione permanente di una sequenza di controllo di trascrizione del gene BCL11A, in cui il metodo comprende inoltre l'introduzione nella cellula del sgRNA della rivendicazione 1 o 2, di un acido nucleico comprendente una sequenza nucleotidica codificante la sequenza di acido nucleico di detto sgRNA, o di un vettore comprendente detto acido nucleico, e in cui le una o più DNA endonucleasi comprendono una endonucleasi Cas9.

9. Metodo secondo la rivendicazione 8, in cui l'acido nucleico comprendente una sequenza nucleotidica codificante le una o più DNA endonucleasi è uno o più acidi ribonucleici (RNA).
10. Metodo secondo la rivendicazione 8 o 9, in cui le una o più DNA endonucleasi comprendono ciascuna, all'estremità N, all'estremità C o sia all'estremità N che all'estremità C, uno o più segnali di localizzazione nucleare (NLS), opzionalmente in cui gli uno o più NLS comprendono un NLS SV40.
11. Metodo secondo la rivendicazione 8 o 10, in cui le una o più DNA

U001.D1.SM.61E

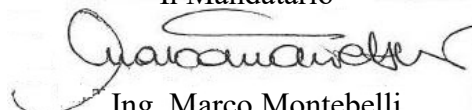
endonucleasi sono pre-complesstate con il sgRNA per formare una o più ribonucleoproteine (RNP), opzionalmente in cui le una o più DNA endonucleasi sono combinate con il sgRNA in un rapporto in peso di 1:1 di DNA endonucleasi/sgRNA per formare le una o più RNP.

12. Metodo secondo la rivendicazione 11, in cui le una o più RNP sono veicolate alla cellula umana mediante elettroporazione, opzionalmente in cui le una o più DNA endonucleasi sono una Cas9 di *S. pyogenes* comprendente un NLS SV40 all'estremità N e un NLS SV40 all'estremità C.
13. Metodo secondo una qualsiasi delle rivendicazioni da 8 a 12 in cui la cellula umana è una cellula progenitrice ematopoietica, opzionalmente in cui la cellula progenitrice ematopoietica è una cellula CD34+.

Dogana, 22/05/2024

In fede

Il Mandatario



Ing. Marco Montebelli
USBM CPI 001

U001.D1.SM.61E

TRADUZIONE DELLE FIGURE

Figura 1A

Chicken β -actin promoter – Promotore di β -actina di pollo

Dasher GFP – Proteina verde fluorescente Dasher

gRNA scaffold – Scaffold di gRNA

Kanamycin – Kanamicina

Figura 1B

Chicken β -actin promoter – promotore di β -actina di pollo

Dasher GFP – Proteina verde fluorescente Dasher

gRNA scaffold – Scaffold di gRNA

Kanamycin – Kanamicina

Figura 1C

Chicken β -actin promoter – promotore di β -actina di pollo

Dasher GFP – Proteina verde fluorescente Dasher

gRNA scaffold – Scaffold di gRNA

Kanamycin – Kanamicina

Figura 2A

Complement of PAM – Complemento di PAM

Gene of interest – Gene d'interesse

Protospacer – Protospaziatore

Spacer sequence – Sequenza spaziatrice

Target DNA – DNA bersaglio

Figura 2B

Complement of PAM – Complemento di PAM

U001.D1.SM.61E

Gene of interest – Gene d’interesse

Linker loop – Ansa linker

Protospacer – Protospaziatore

Spacer sequence – Sequenza spaziatrice

Target DNA – DNA bersaglio

Figura 3

Donor – Donatore

Indels – Indel

Figure 4A, 4B, 4C

Deletion/modification – Delezione/modificazione

Donor – donatore

Erythrocytes - Eritrociti

Expansion - espansione

Hematopoiesis from CD34+ stem cell to erythrocyte (red blood cell) –

Ematopoiesi da cellula staminale CD34+ a eritrocito (globulo rosso)

Mixed population of edited and unedited CD34+ HSC – Popolazione mista di

cellule HSC CD34+ editate e non editate

Normalized to unedited control – Normalizzato a controllo non editato

Polychromatic erythroblast – Eritroblasto policromatico

Proerythroblast - Proeritroblasto

Ratio of – Rapporto di

Reticulocytes - Reticolociti

Terminal differentiation – Differenziazione terminale

Figure 5A, 5B

U001.D1.SM.61E

Both edited and unedited CD34+ clones differentiated into RBC – Cloni CD34+
sia editati che non editati differenziati in RBC

γ/α globin mRNA ratio – rapporto di mRNA di γ/α globina

Small deletion – Piccola delezione

Figura 6

Bcl11a Intron (SPY101) Rate of Editing – Tasso di editing di introne Bcl11a
(SPY101)

Data from 60 colonies – Dati da 60 colonie

Edited colonies – Colonie editate

Indel – Indel

Total indel frequency – Frequenza di indel totali

Figure 7A, 7B

Bc11a intron (SPY101) edit – Edit di introne (SPY101) di Bc11a

Indel – Indel

Percentage β -like globins HBG/(HBB+HBG) – Percentuale di globine β -simili
HBG/(HBB+HBG)

Percentage of HBG to HBA – Percentuale di HBG/HBA

Figura 8

Donor – Donatore

Editing efficiency – Efficienza di editing

Figura 9A

Sequencing - Sequenziamento

SureSelect target enrichment system capture process – Processo di cattura di
sistema di arricchimento di bersaglio SureSelect

U001.D1.SM.61E

Figura 9B

Max off target cleavage efficiency – Efficienza massima di clivaggio off-target

of off target cut sites detected – numero di siti tagliati off-target rilevati

of off target cut sites examined – numero di siti tagliati off-target esaminati

Figura 10A

$\gamma/\gamma+\beta$ globin mRNA ratio (%) – Rapporto (%) di mRNA di $\gamma/\gamma+\beta$ globina

Healthy donor – Donatore sano

Target – bersaglio

Figura 10B

γ/α globin mRNA ratio (%) – Rapporto (%) di mRNA di γ/α globina

Healthy donor – Donatore sano

Target – bersaglio

Figura 11B

Bulk - Bulk

Edited – editato

Frequency – Frequenza

Mock - Mock

Population - Popolazione

Figura 11C

Bulk – Bulk

Editing efficiency – Efficienza di editing

High editing efficiencies in all cell types – Elevate efficienze di editing in tutti i

tipi di cellule

Figura 12

U001.D1.SM.61E

8 week interim bleeds – Sanguinamenti ad interim di 8 settimane

Donor - Donatore

Mean – Media

Unedited – Non editato

Untreated – Non trattato

Figura 13

Clinical scale – Scala clinica

Editing efficiency – Efficienza di editing

Lab scale – Scala di laboratorio

Figura 14

Biodistribution and persistence – Biodistribuzione e persistenza

Biodistribution/persistence study – Studi di biodistribuzione/persistenza

Clinical signs – Segni clinici

Detailed analysis to exclude malignancies – Analisi dettagliata per escludere malignità

Engraftment - Attecchimento

Fresh CD34+ hHSPCs – hHSPC CD34+ fresche

Full necropsy – Necropsia completa

General toxicology assessments – Valutazioni di tossicologia generali

Mice – topi

Figura 15

Days - giorni

Electroporation – Elettroporazione

Erythroid differentiation – Differenziazione eritroide

U001.D1.SM.61E

Experimental setup – Setup sperimentale

Genomic DNA isolation, on target editing (sequencing) – Isolamento di DNA genomico, editing su bersaglio (sequenziamento)

Human mPB CD34+ cells – Cellule CD34+ di mpB umane

Protein – proteina

Rna isolation – Isolamento di RNA

Single cell sorting for erythroid progenitors – Separazione di singole cellule per progenitori eritroidi

Sister pellets – Pellet sorelle

Transcript – Trascritto

Figura 16A

$\gamma/(\gamma+\beta)$ globin mRNA ratio (% over mock) – Rapporto (%) di mRNA di $\gamma/(\gamma+\beta)$ globina (% rispetto a mock)

Target – bersaglio

Figura 16B

$\gamma/(\gamma+\beta)$ globin protein ratio (% over mock) – Rapporto (%) di proteine $\gamma/(\gamma+\beta)$ globina (% rispetto a mock)

Target – bersaglio

Figura 17

$\gamma/(\gamma+\beta)$ globin mRNA ratio (%) – Rapporto (%) di mRNA di $\gamma/(\gamma+\beta)$ globina (%)

Target – bersaglio

Figura 18A

Eterozygous - Eterozigote

U001.D1.SM.61E

Genotype – Genotipo

Homozygous - Omozigote

Target - Bersaglio

WT – Tipo selvatico

Figura 18B

$\gamma/(\gamma+\beta)$ globin mRNA ratio (%) – Rapporto (%) di mRNA di $\gamma/(\gamma+\beta)$ globina (%)

Target – bersaglio

Figura 19

Bulk workflow and the List of Bulk Data – Flusso di lavoro di bulk e lista di dati bulk

Day – giorno

Erythroid differentiation for 15 days – Differenziazione eritroide per 15 giorni

Off target analysis (hybrid capture) – Analisi fuori bersaglio (cattura di ibridi)

On target editing – Editing su bersaglio

Protein - Proteina

Thaw and culture CD34+ from 4 donors – Scongellamento e coltura di CD34+ da 4 donatori

Transcript – Trascritto

Figure 20A, 20B

Day – Giorno

Deletion - Delezione

Donor - Donatore

Editing % maintained throughout *ex vivo* erythroid differentiation of mPB

U001.D1.SM.61E

CD34+ cells – Percentuale (%) di editing mantenuta attraverso tutta la differenziazione eritroide *ex vivo* di cellule CD34+ di mPB

Figure 21A, 21B, 21C e 21D

Day – Giorno

Donor - Donatore

γ/α globin transcript – trascritto di γ/α globina

$\gamma/(\gamma+\beta)$ globin transcript – trascritto di $\gamma/(\gamma+\beta)$ globina

Increase in γ -globin transcript post editing – Aumento in trascritto di γ -globina dopo l'editing

Mock EP – EP mock

No EP – Nessuna EP

Figure 22A, 22B

Exon – Esone

HbG (Median Fluor Intensity) relative to CD34 bulk GFP – HbG (Intensità di fluorescenza mediana) rispetto a CD34 bulk GFP

Mock EP – EP mock

No EP – Nessuna EP

Figure 23A, 23B, 23C e 23D

Bulk LC-MS data further validates upregulation of γ -globin post editing – Dati in bulk di LC-MS validano ulteriormente la sovraregolazione di γ -globina dopo editing

Donor – Donatore

Mock EP – EP mock

Normalized to – Normalizzato a

U001.D1.SM.61E

No EP – Nessuna EP

γ/α globin protein – Proteina γ/α globina

$\gamma/(\gamma+\beta)$ globin protein – Proteina $\gamma/(\gamma+\beta)$ globina

Figura 24

Hybrid capture bait design – Progettazione di esche (bait) per cattura di ibridi

Figura 25

Coverage (#reads) – Copertura (# di letture)

Coverage increases power – Copertura aumenta potenza

True editing – Editing reale

Figura 26

And – E

Cas9 format – Formato di Cas9

Hybrid capture summary – Sommario di cattura di ibridi

Indel freq – Frequenza di indel

Median sequencing coverage – Copertura di sequenziamento mediana

Number of donors – Numero di donatori

Number of mismatches and types of PAMs – Numero di mismatch e tipi di PAM

Number of sites – Numero di siti

Round – Turno

of off-target sites with evidence of cutting - # di siti off-target con evidenza di taglio

Figura 27

And – E

Cas9 format – Formato di Cas9

U001.D1.SM.61E

Hybrid capture summary – Sommario di cattura di ibridi

Indel freq – Frequenza di indel

Median sequencing coverage – Copertura di sequenziamento mediana

Number of donors – Numero di donatori

Number of mismatches and types of PAMs – Numero di mismatch e tipi di PAM

Number of sites – Numero di siti

Round – Turno

of off-target sites with evidence of cutting - # di siti off-target con evidenza di taglio

Figura 28

Cell origin – Origine cellulare

Cohort – Coorte

Control untreated – Controllo non trattato

Donor – Donatore

Electroporation of cells with GFP Guide RNA and Cas9 protein –
Elettroporazione di cellule con RNA guida di GFP e proteina Cas9

Electroporation of cells with SD2 gRNA and Cas9 protein – Elettroporazione di
cellule con SD2 gRNA e proteina Cas9

Electroporation of cells with SPY101 gRNA and Cas9 protein –
Elettroporazione di cellule con SPY101 gRNA e proteina Cas9

Group – Gruppo

“Mock electroporation” of cells without any CRISPR/Cas9 editing components
– “Elettroporazione mock” di cellule senza alcun componente di editing di
CRISPR/Cas9

U001.D1.SM.61E

Number of female animals – Numero di animali femmina

Pilot engraftment study plan – Piano di studio di attecchimento o engraftment pilota

Radiation dose – Dose di radiazioni

Radiation dose rate – Rateo di dose di radiazioni

Treatment – Trattamento

Untreated cells – Cellule non trattate

Figure 29A, 29B e 29C

Week – Settimana

% of total hCD45RA+ live cells - % di cellule vive hCD45RA+

Figure 29D, 29E

Week – Settimana

% of total hCD45RA+ live cells - % di cellule vive hCD45RA+

Figura 30

8 week interim blood analysis – Analisi del sangue ad interim a 8 settimane

% of total hCD45RA+ live cells - % di cellule vive hCD45RA+

Donor – Donatore

Figura 31

Cas9 RNP is as efficient in editing as Cas9 mRNA but superior in post-electroporation viability – Cas9 RNP è efficiente in editing quanto Cas9 mRNA ma superiore nella vitalità post-elettroporazione

Compared to RNP by One Way ANOVA – Comparato a RNP mediante ANOVA a una via

Figura 32A

U001.D1.SM.61E

Cas9 RNP is superior to Cas9 mRNA in post-electroporation viability – Cas9

RNP è superiore a Cas9 mRNA nella vitalità post-elettroporazione

Compared to No EP by One Way ANOVA – Comparato a Nessuna EP mediante

ANOVA a una via

Fold increase in cell count at 48 h post-electroporation – Numero di volte di aumento nel conteggio cellulare 48 ore dopo l'elettroporazione

No EP – Nessuna EP

Figura 32B

Cas9 RNP is superior to Cas9 mRNA in post-electroporation viability – Cas9

RNP è superiore a Cas9 mRNA nella vitalità post-elettroporazione

Cell viability at 48h post-electroporation – Vitalità cellulare 48 ore dopo l'elettroporazione

Compared to No EP by One Way ANOVA – Comparato a Nessuna EP mediante

ANOVA a una via

No EP – Nessuna EP

Figura 33A

Cas9 RNP optimization – Ottimizzazione di Cas9 RNP

Figure 34A, 34B

Bulk editing efficiency for SPY101 – Efficienza di editing in bulk per SPY101

Editing efficiency – Efficienza di editing

Figura 35A, 35B

All unlabeled data points had > 70% editing efficiency – Tutti i punti dati non etichettati avevano una efficienza di editing superiore al 70%

Better optimized – Meglio ottimizzato

U001.D1.SM.61E

Editing efficiency – Efficienza di editing

Efficacy of SPY101 in mPB CD34+ cells – Efficacia di SPY101 in cellule CD34+ di mPB

Non optimized differentiation – Differenziazione non ottimizzata

Optimized differentiation – Differenziazione ottimizzata

γ/α globin mRNA ratio in % (Ctrl subtracted) – rapporto di mRNA di γ/α globina in % (sottratto il controllo)

$\gamma/(\gamma+\beta)$ globin mRNA ratio in % (Ctrl subtracted) – rapporto di mRNA di $\gamma/(\gamma+\beta)$ globina in % (sottratto il controllo)

Figure 36A, 36B

All unlabeled data points had > 80% editing efficiency – Tutti i punti dati non etichettati avevano una efficienza di editing superiore all'80%

Better optimized – Meglio ottimizzato

Editing efficiency – Efficienza di editing

Efficacy of SPY101 in BM CD34+ cells – Efficacia di SPY101 in cellule CD34+ di midollo spinale

Non optimized – Non ottimizzata

Optimized – Ottimizzata

γ/α globin mRNA ratio in % (Ctrl subtracted) – rapporto di mRNA di γ/α globina in % (sottratto il controllo)

$\gamma/(\gamma+\beta)$ globin mRNA ratio in % (Ctrl subtracted) – rapporto di mRNA di $\gamma/(\gamma+\beta)$ globina in % (sottratto il controllo)

Figure 37A e 37B

Efficacy of SPY101 in patient samples – Efficacia di SPY101 in campioni di

U001.D1.SM.61E

pazienti

From β -Thalassemic patient – Da paziente affetto da β -talassemia

From SCD patients – Da pazienti affetti da anemia falciforme

Healthy - Sano

γ/α globin mRNA ratio in % (Ctrl subtracted) – rapporto di mRNA di γ/α globina in % (sottratto il controllo)

$\gamma/(\gamma+\beta)$ globin mRNA ratio in % (Ctrl subtracted) – rapporto di mRNA di $\gamma/(\gamma+\beta)$ globina in % (sottratto il controllo)

Figure 38A, 38B

Data from 60 colonies – Dati da 60 colonie

Data from 164 colonies – Dati da 164 colonie

Edited colonies – Colonie editate

Indel frequency – Frequenza di indel

SPY edited genotype breakdown – Scomposizione di genotipo editato con SPY

Figure 39A, 39B

Bi-allelic disrupt – Distrutto bi-allelico

Distribution of SPY-edited colonies by GBS disruption – Distribuzione di colonie editate con SPY mediante distruzione di GBS

Gata1 binding site disruption by SPY101 linked to increased γ -globin expression – Distruzione del sito di legame a Gata1 mediante SPY101 legata a espressione aumentata di γ -globina

Median with interquartile range – Mediana con intervallo interquartile

Mono-allelic disrupt – Distrutto mono-allelico

No disrupt – Nessun distrutto

U001.D1.SM.61E

One Way ANOVA comp to GFP or No Disrupt – ANOVA a una via comparato a
GFP o Nessun Distrutto

γ/α -globin mRNA ratio – Rapporto di mRNA di γ/α -globina

Figure 40A, 40B, 40C, 40D, 40E


% γ Hb positive cells – Percentuale (%) di cellule positive a γ Hb

Median fluorescence intensity – Intensità di fluorescenza mediana

SPY101 increased γ -globin expression – SPY101 aumentava l'espressione di γ -
globina

In fede

Il Mandatario


Ing. Marco Montebelli
USBM CPI 001

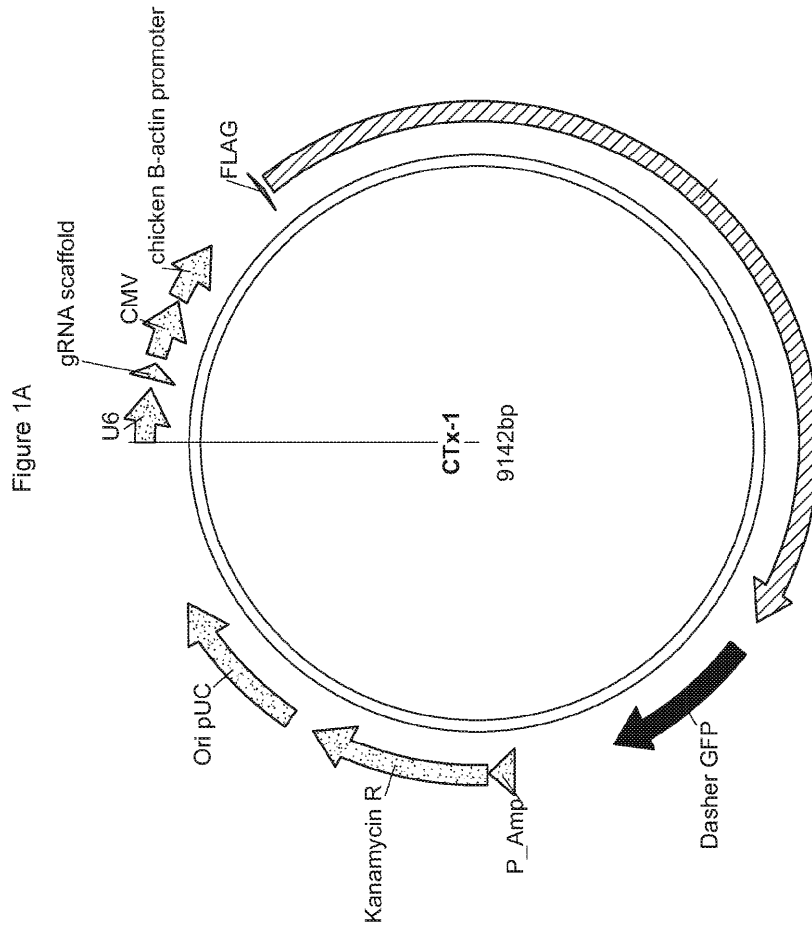


Figure 1B

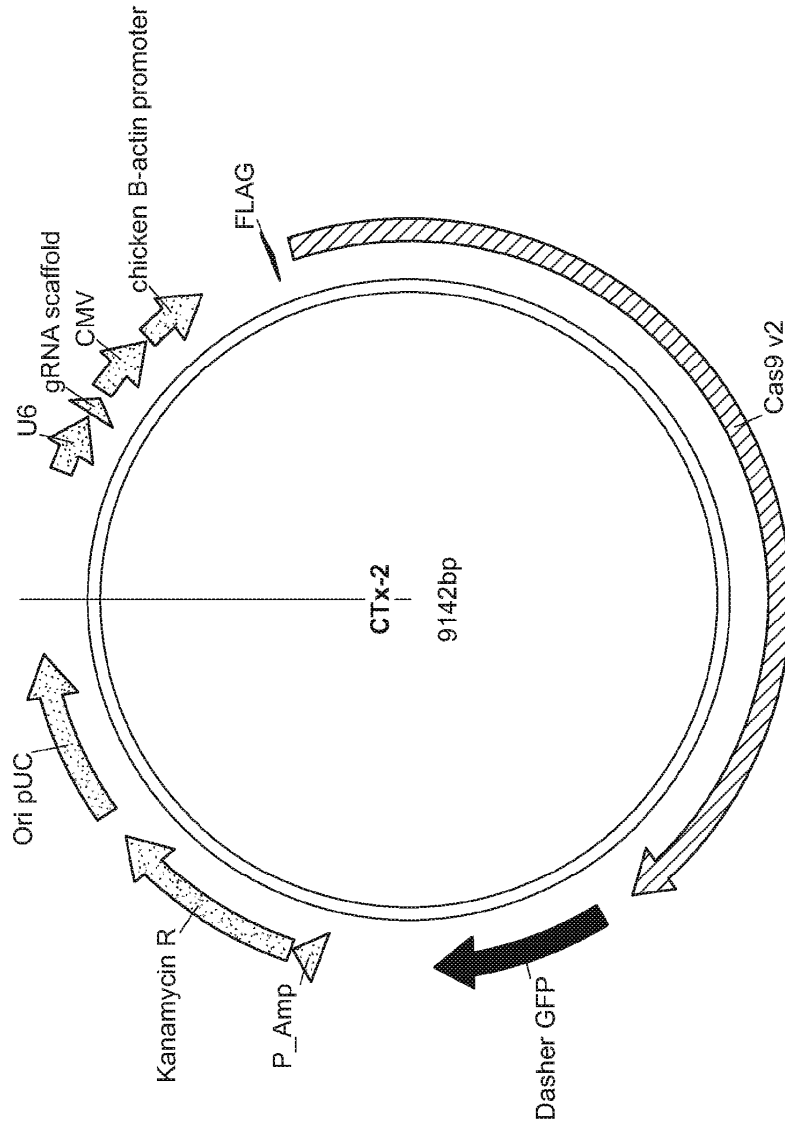


Figure 1C

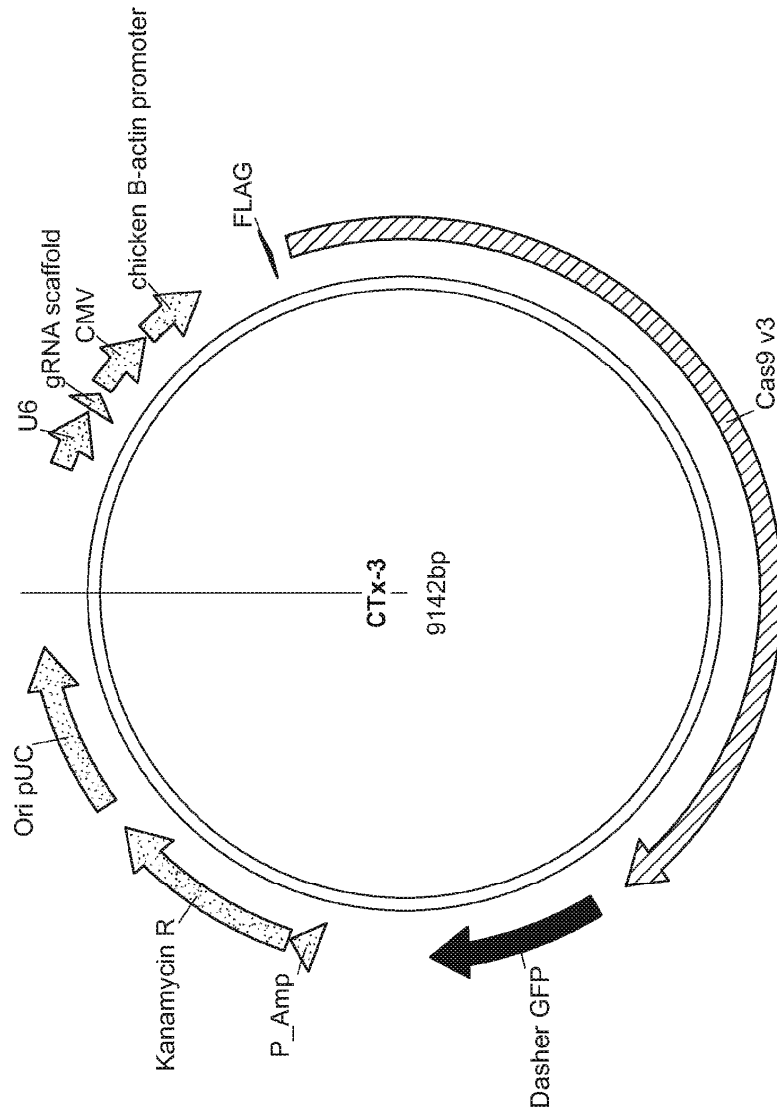


Figure 2A

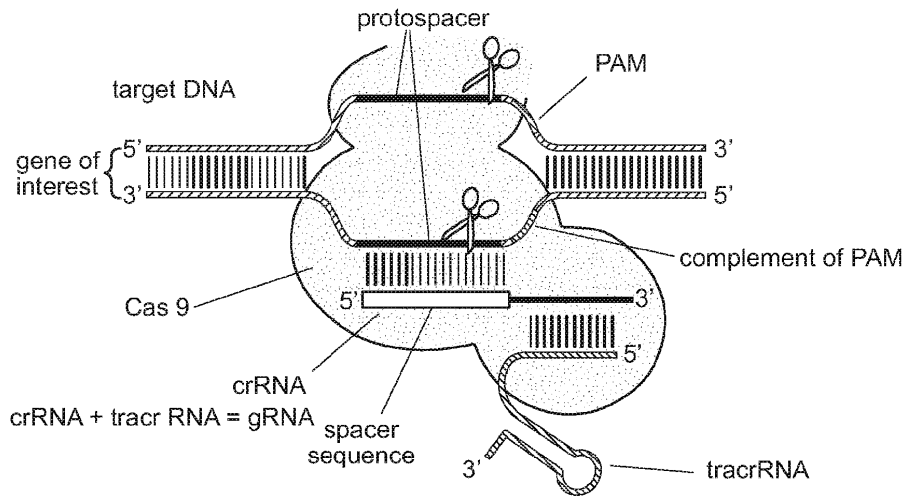


Figure 2B

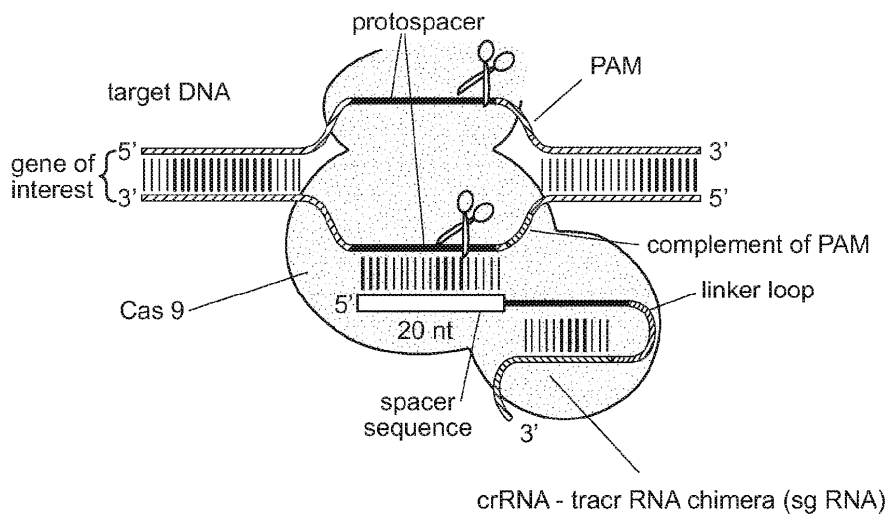


FIGURE 3

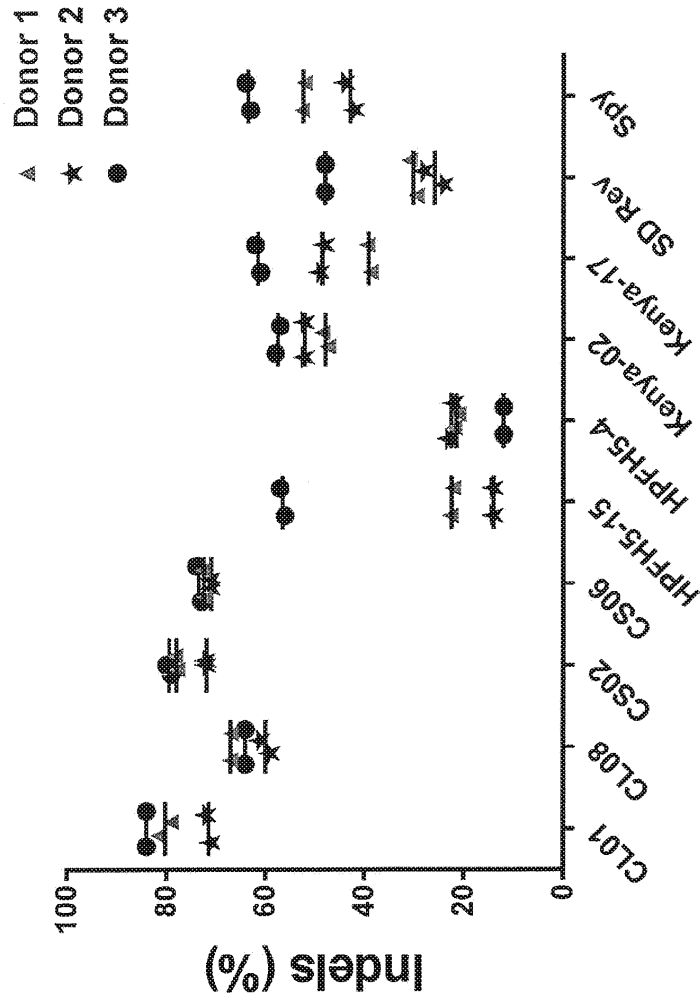


Figure 4A

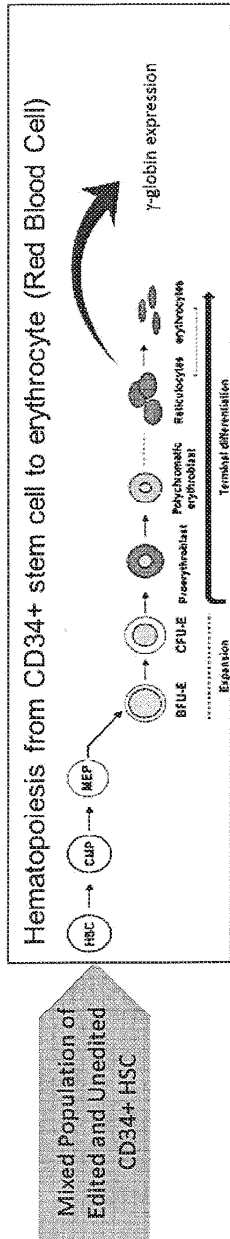


Figure 4B

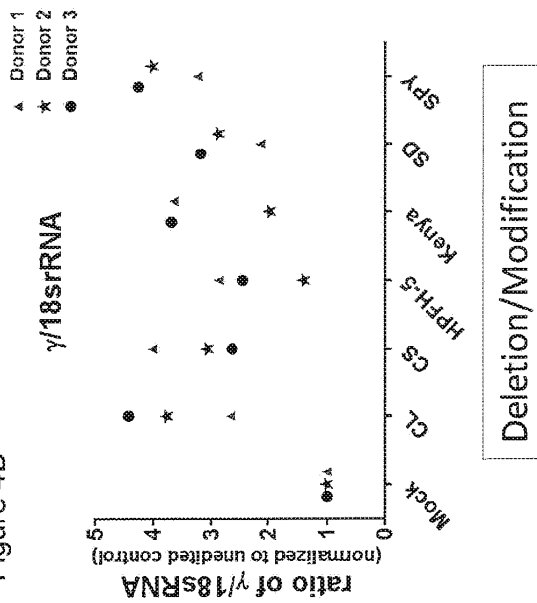
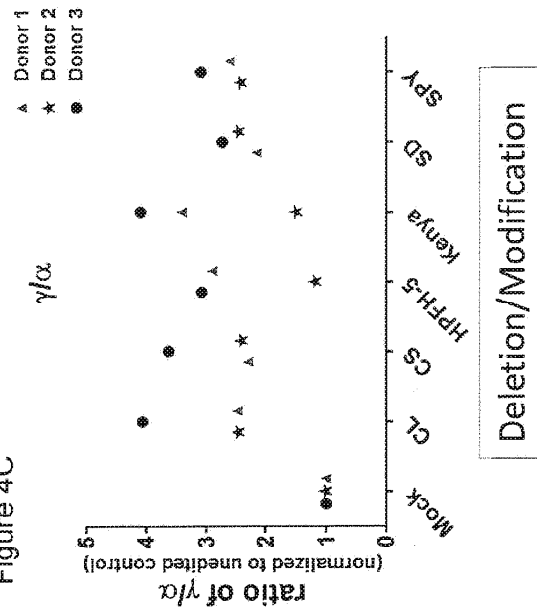


Figure 4C



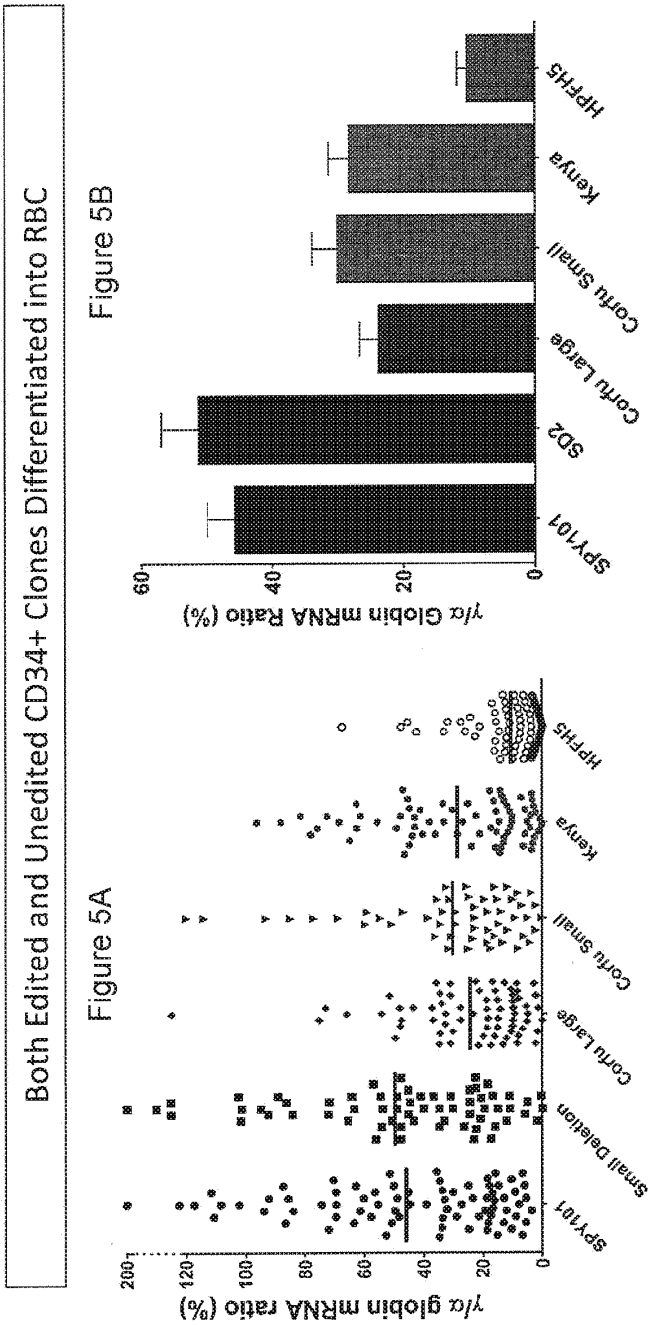
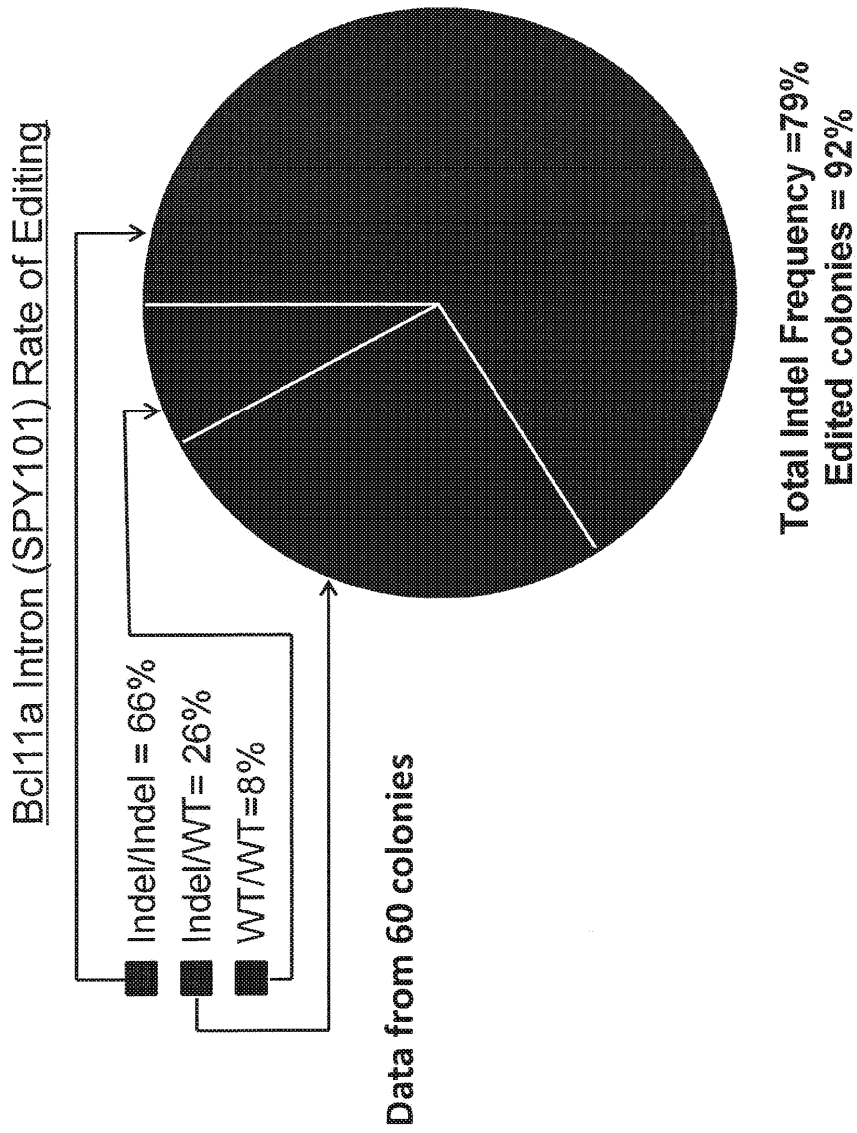
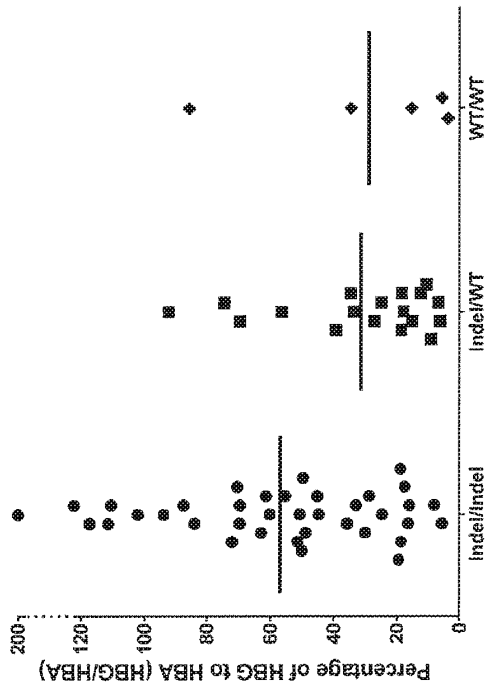


FIGURE 6



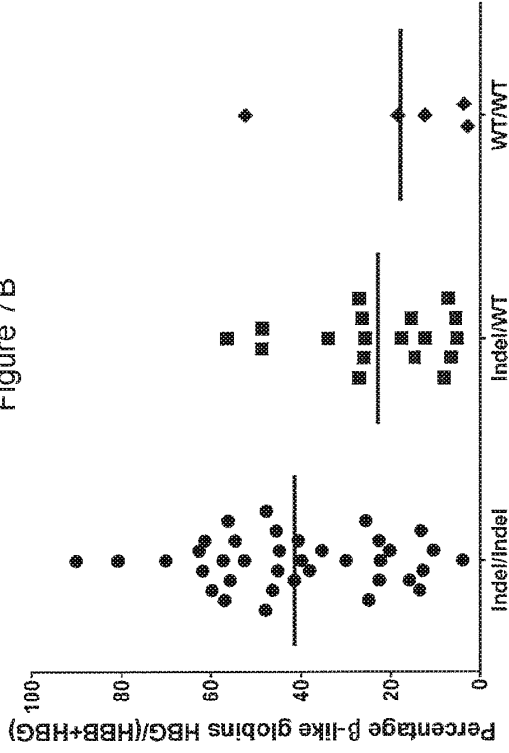
Bcl11a Intron (SPY101) Edit

Figure 7A



HBG/HBA

Figure 7B



HBG/(HBB+HBG)

FIGURE 8

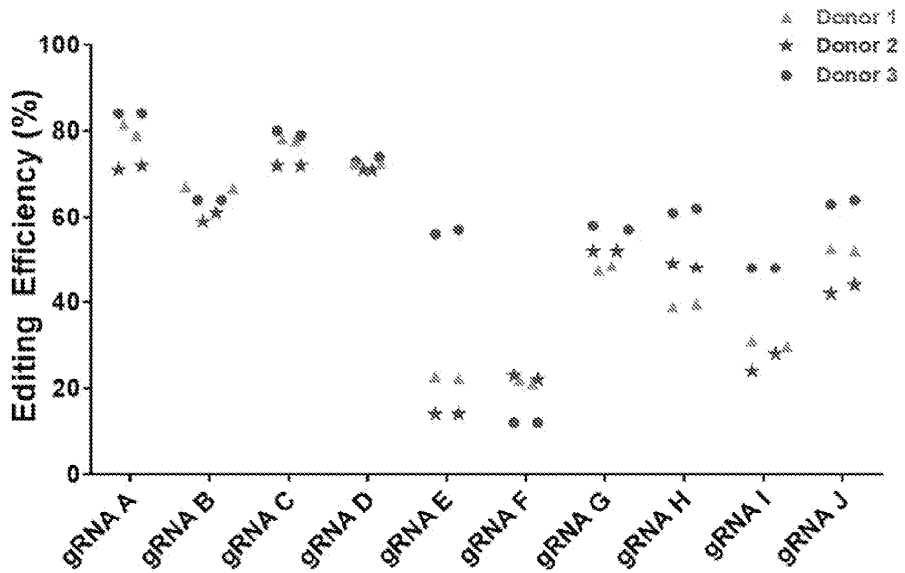


Figure 10A

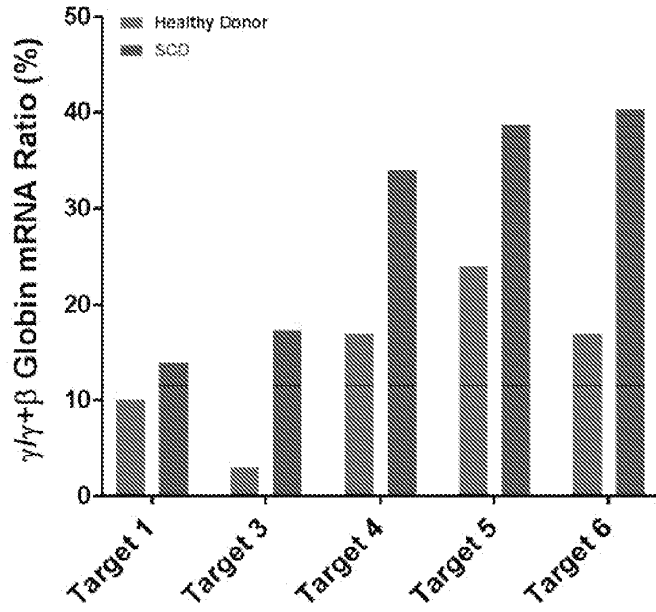


Figure 10B

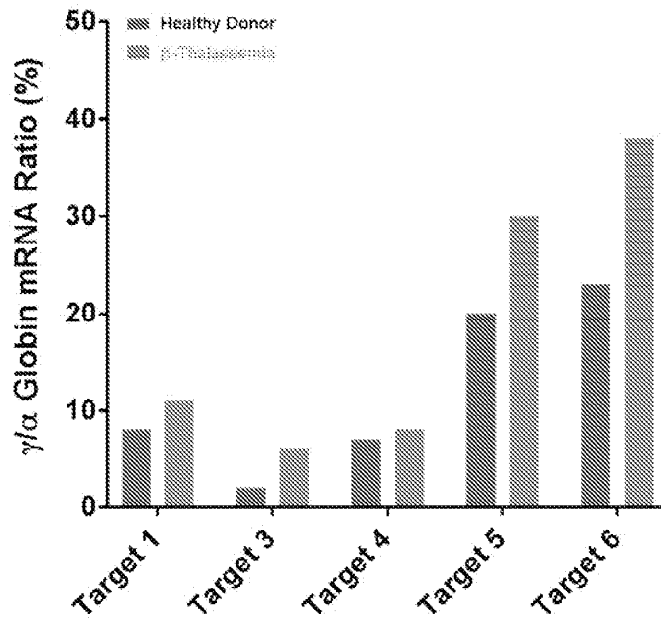


Figure 11A

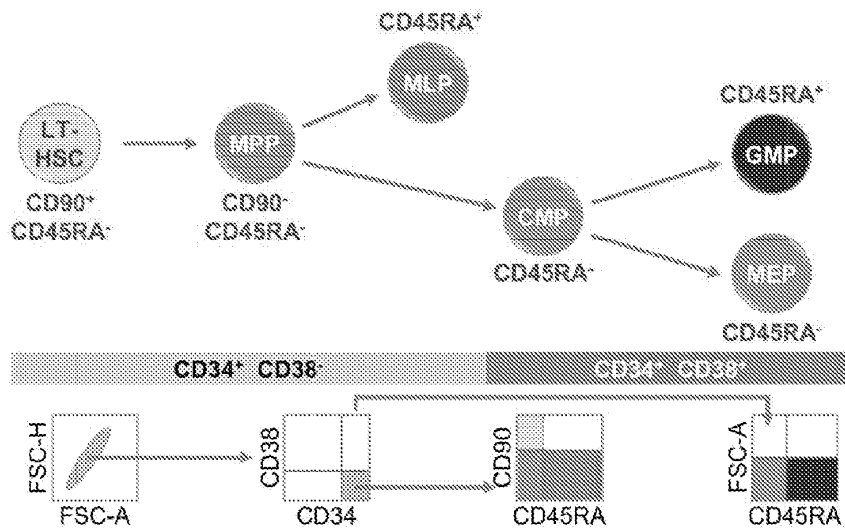


Figure 11B

Distribution of cell types remains unchanged after editing

Population	Frequency	
	Mock	Edited
Bulk	94.83	94.30
LT-HSC	7.82	7.98
MPP	18.97	16.38
MLP	13.23	8.47
CMP/MEP	11.23	11.32
GMP	8.39	13.26

Figure 11C

High editing efficiencies in all cell types

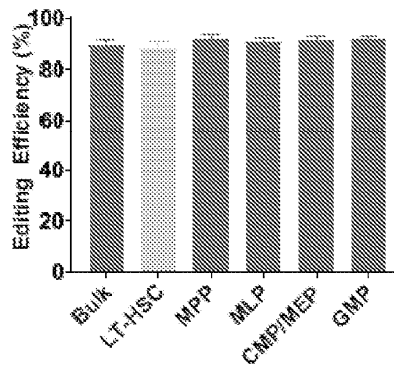


Figure 12

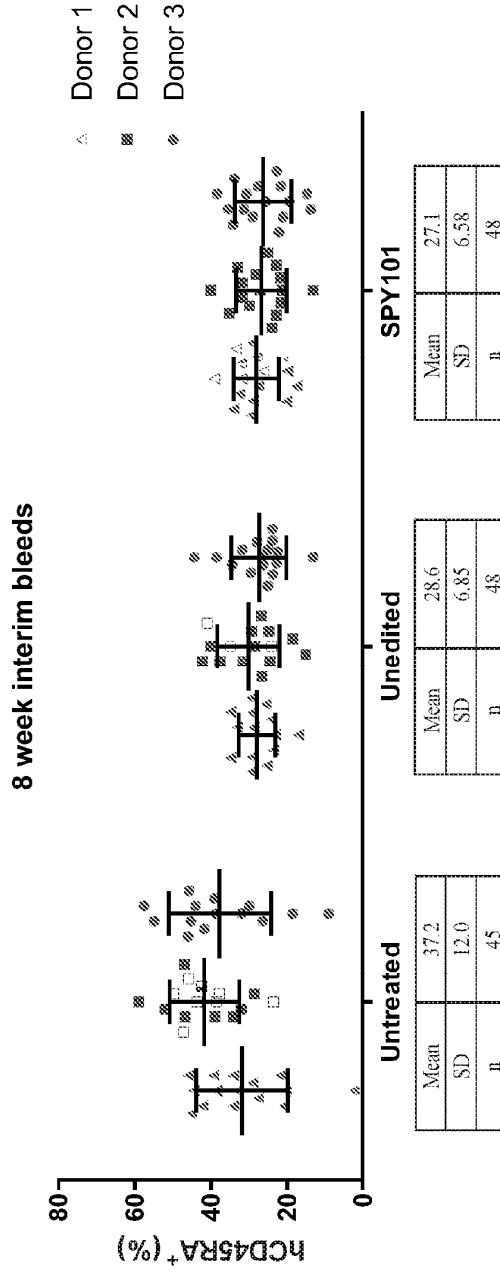


Figure 13

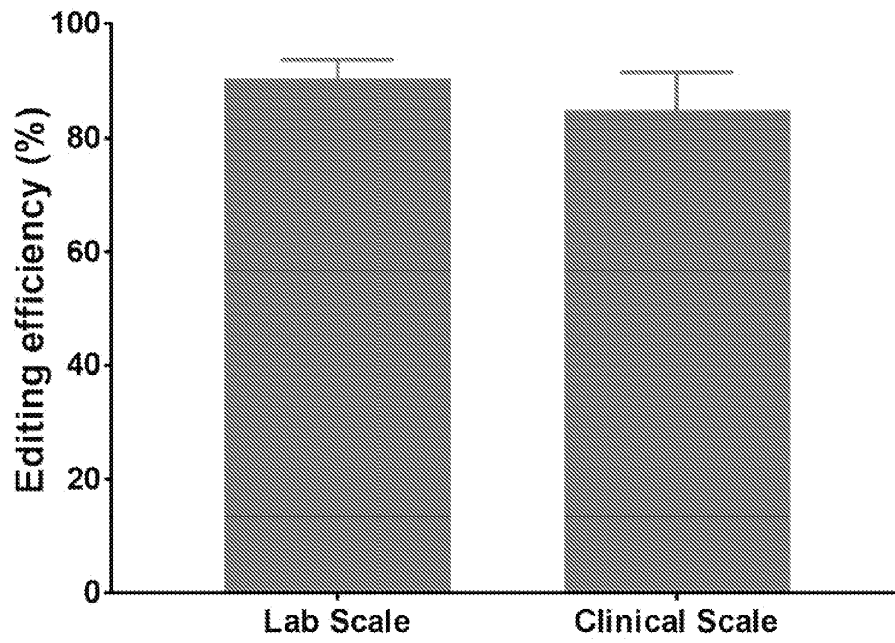
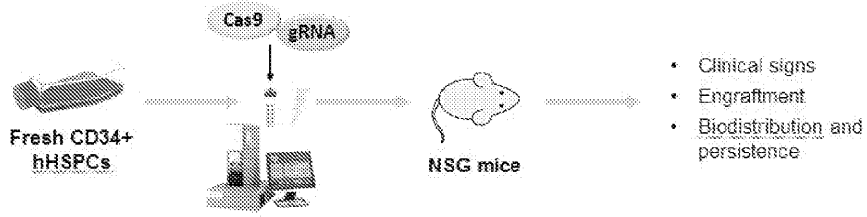


Figure 14

Biodistribution / Persistence Study



Tumorigenicity / General Toxicology Study

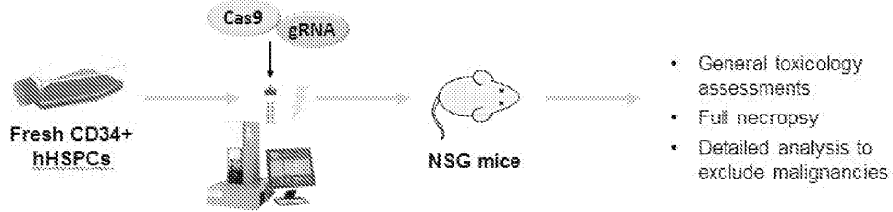


Figure 15

Experimental setup

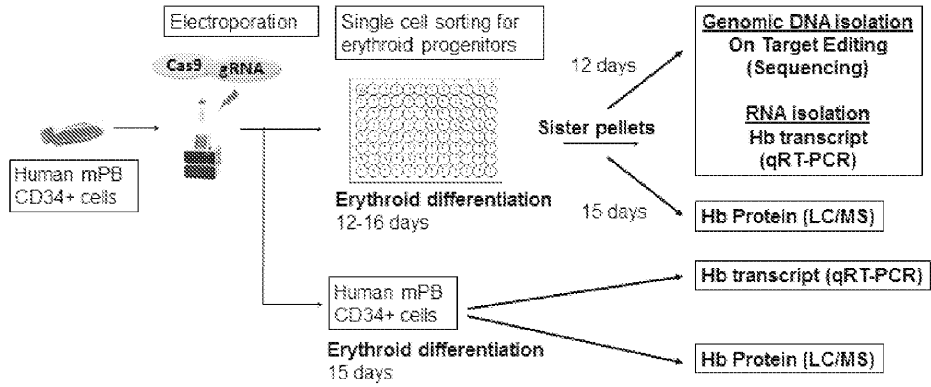


Figure 16A

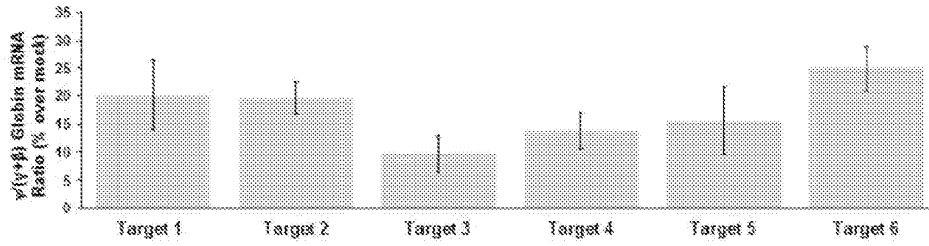


Figure 16B

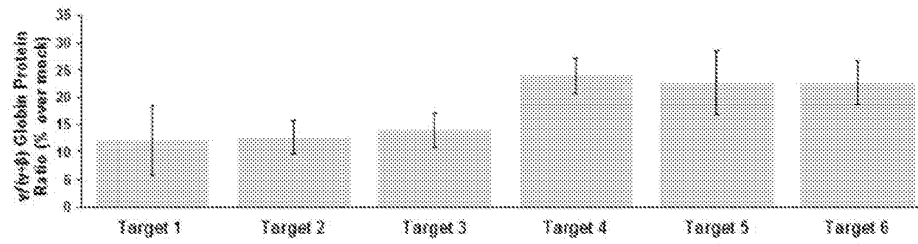


Figure 17

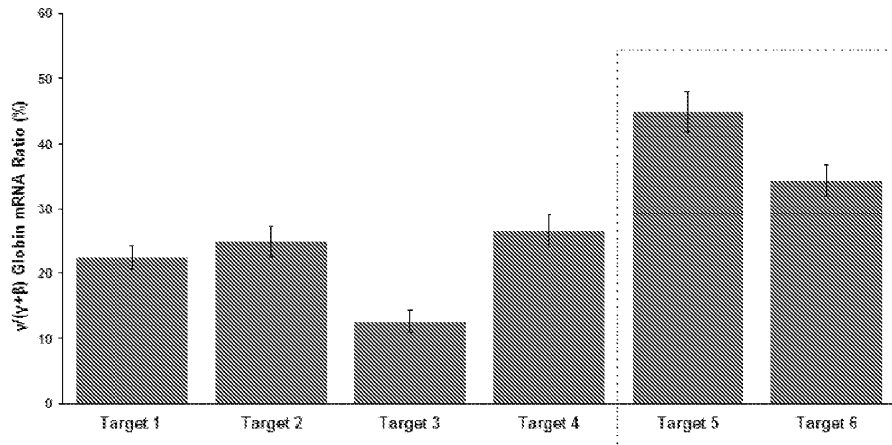


Figure 18B

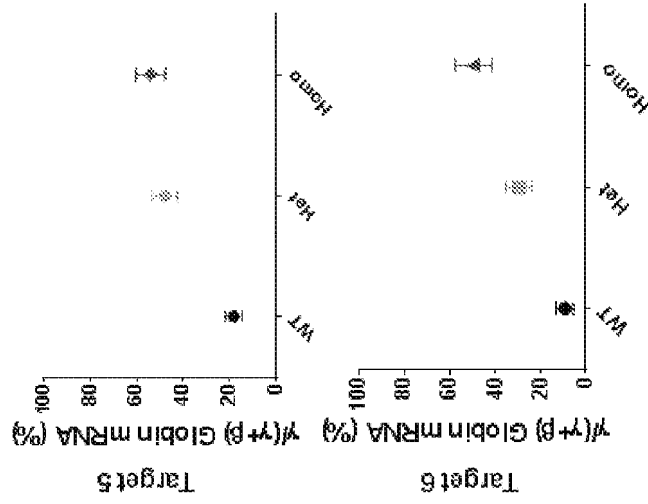


Figure 18A

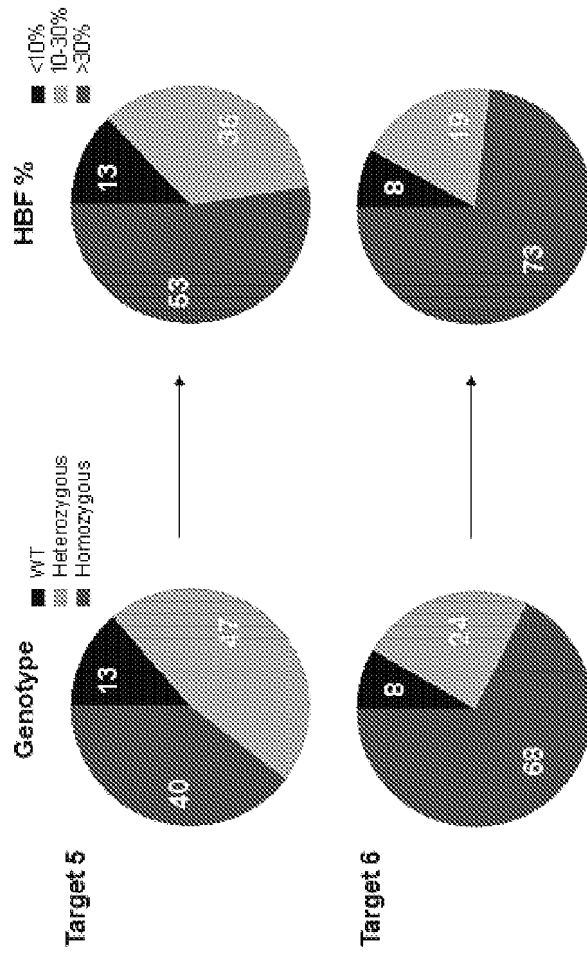
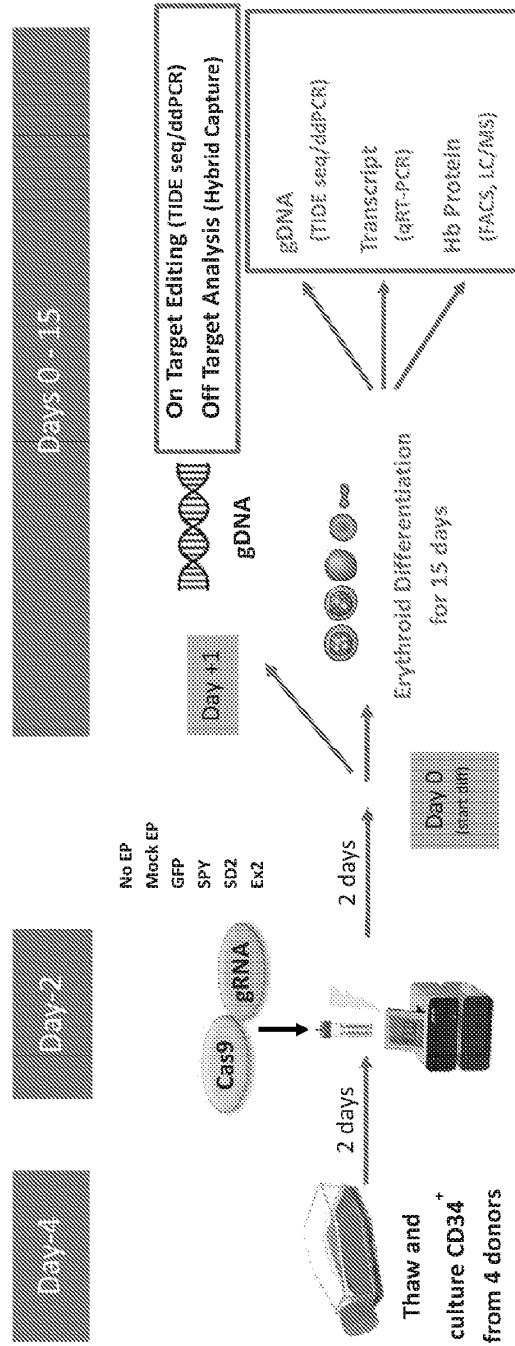


Figure 19
Bulk Workflow and the List of Bulk Data



Editing % maintained throughout ex vivo erythroid differentiation of mPB CD34+ cells

Figure 20A

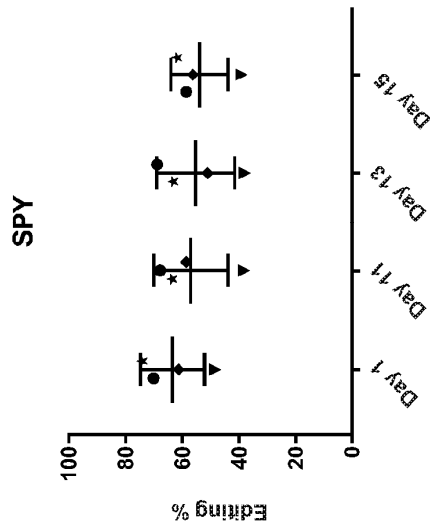
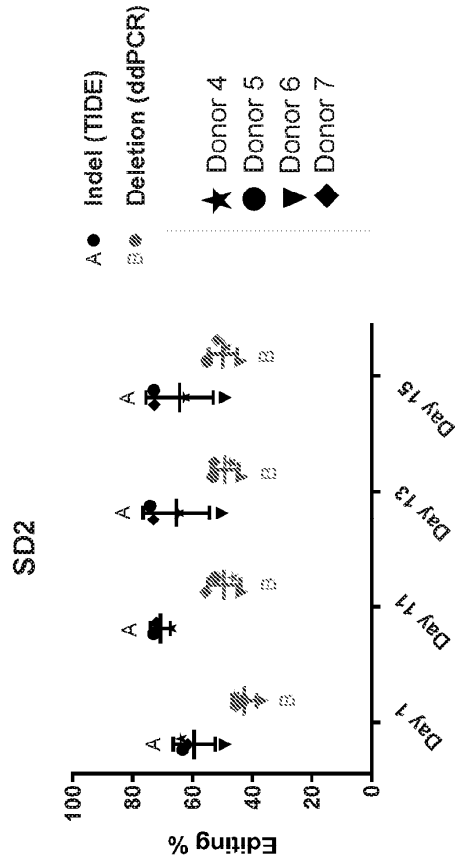


Figure 20B



Increase in γ -globin transcript post editing

Figure 21A

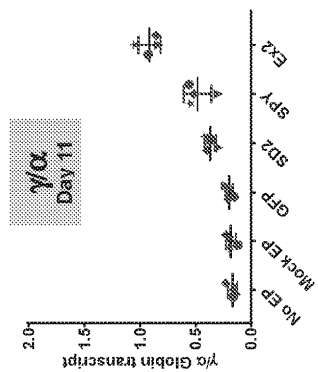


Figure 21B

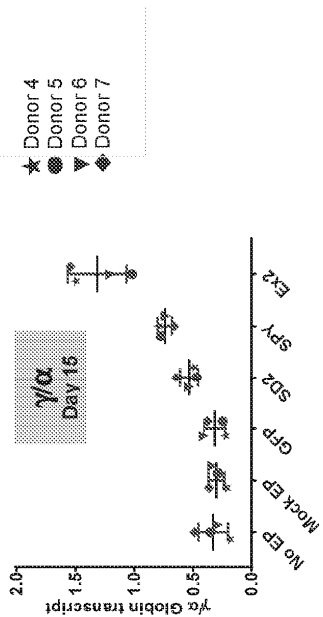


Figure 21C

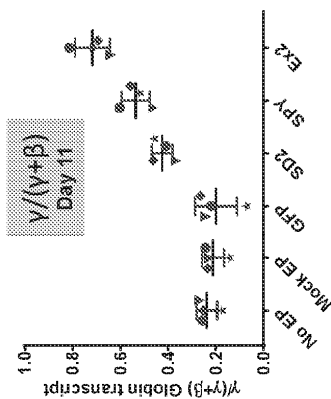
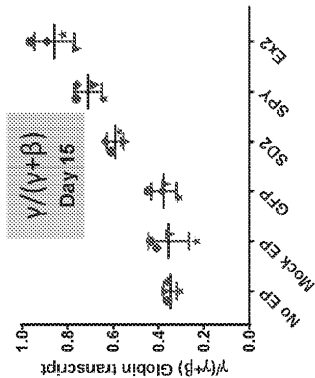


Figure 21D



FACS analysis shows upregulation of γ -globin in edited cells

Figure 22A

FACS

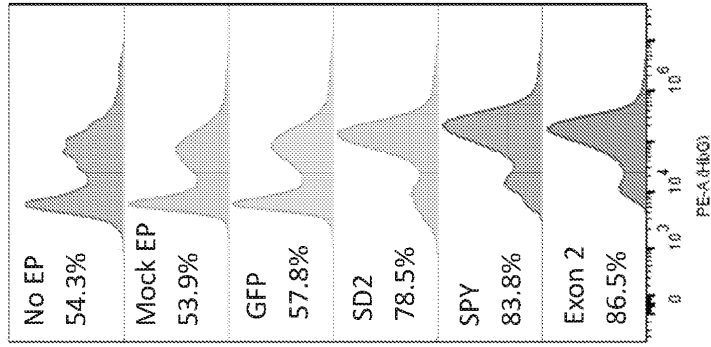
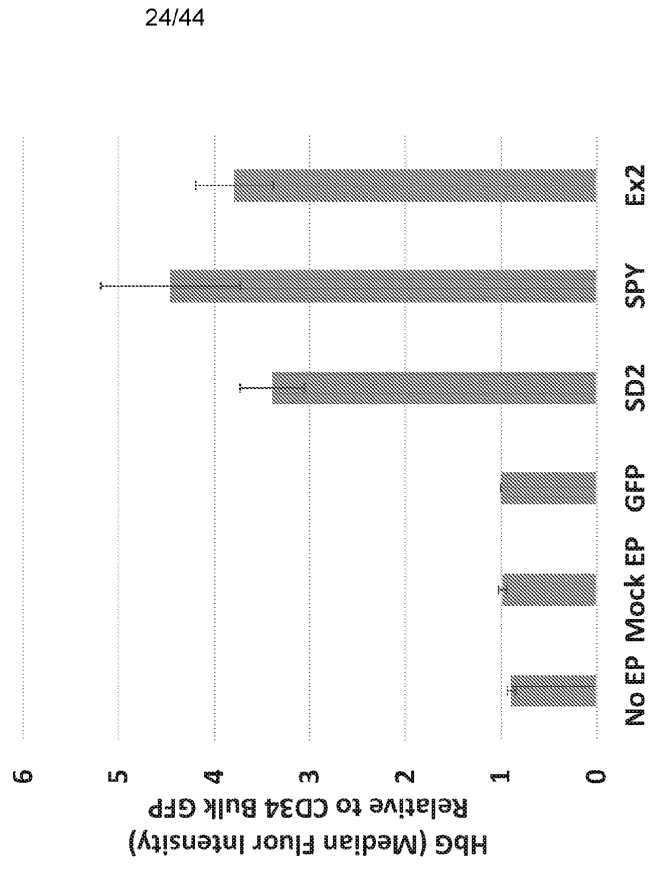


Figure 22B

MFI



Bulk LC-MS data further validates upregulation of γ -globin post editing

Figure 23A

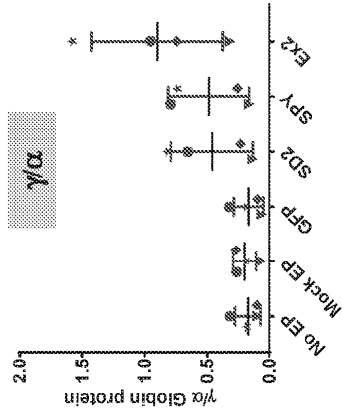


Figure 23B

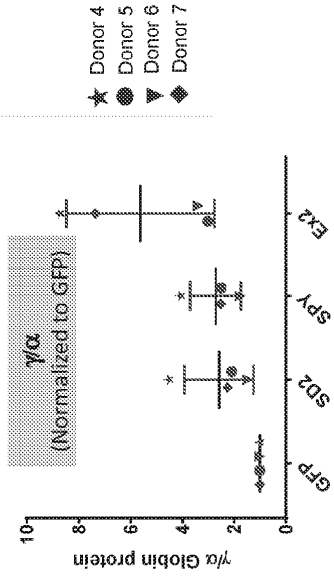


Figure 23C

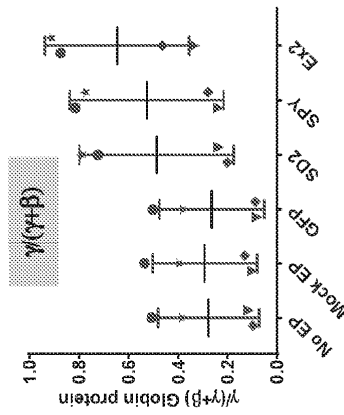


Figure 23D

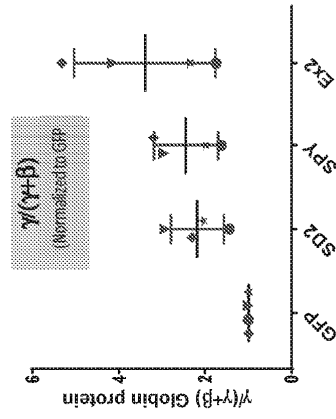
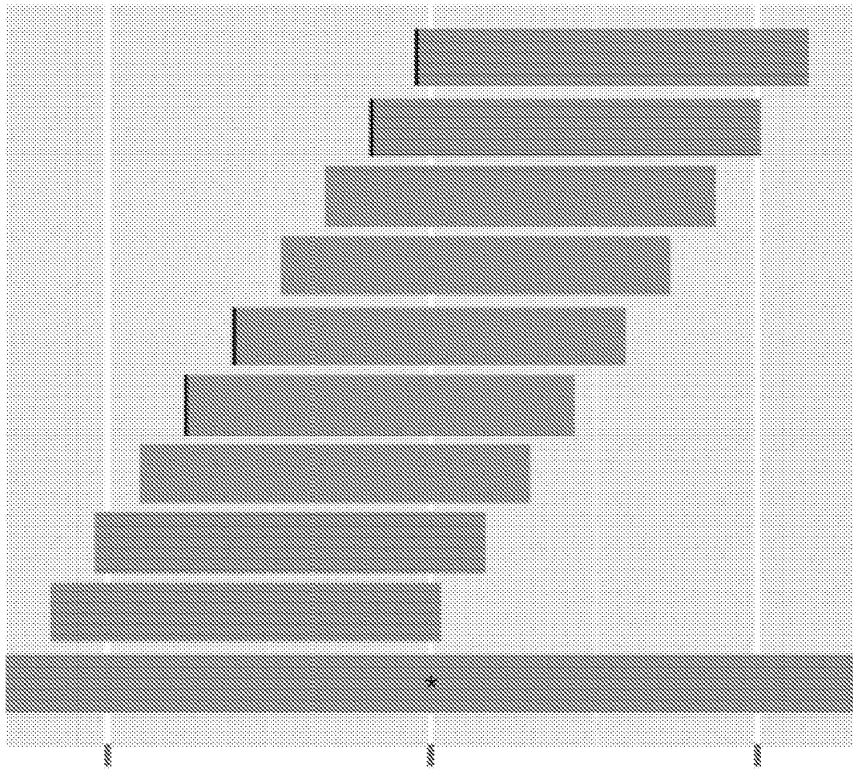


Figure 24



Hybrid Capture Bait Design

Figure 25

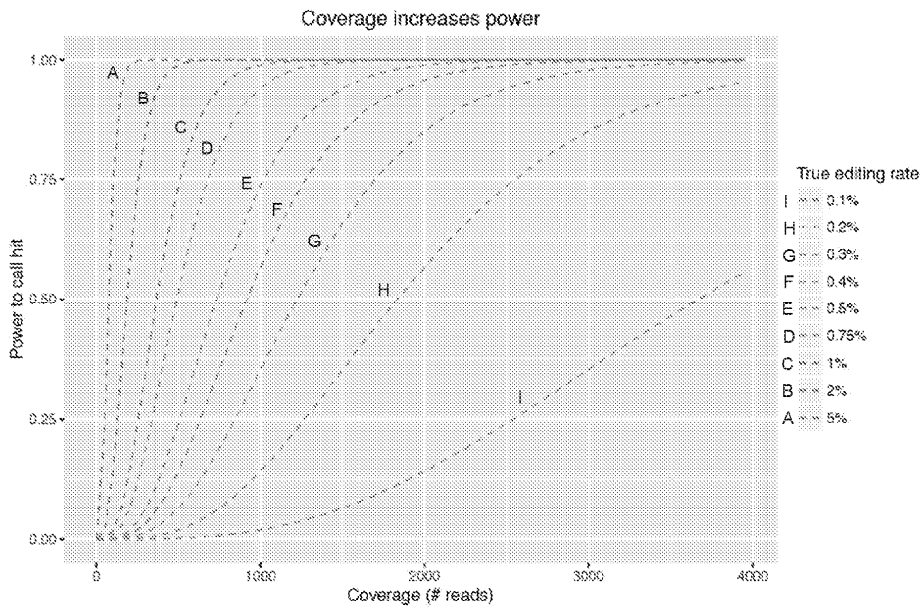


Figure 26
Hybrid capture summary (SPY101)

SPY101	Round 1	Round 2
Number of donors	3	4
Cas9 format	mRNA	RNP
Median sequencing coverage	~800	2492
Number of sites	840	5003
Number of mismatches and types of PAMs	4: NGG and NAG	5: NGG, NAG 4: NGA, NAA, NCG, NGC, NTG, NGT
# of off-target sites with evidence of cutting	0* *indel freq > 0.5%	0 ^Δ ^Δ indel freq > 0.2%

Figure 27
Hybrid capture summary (SD2)

SD2	Round 1	Round 2
Number of donors	3	4
Cas9 format	mRNA	RNP
Median sequencing coverage	~800	2775
Number of sites	677	3272
Number of mismatches and types of PAMs	4: NGG and NAG	5: NGG, NAG 4: NGA, NAA, NCG, NGC, NTG, NGT
# of off-target sites with evidence of cutting*	0* *Indel freq >0.5%	0 ^A ^A Indel freq >0.2%

Figure 28
Pilot Engraftment Study Plan

Group	Treatment	Radiation dose (cGy)	Radiation Dose Rate (cGy/min)	Number of Female Animals		
				Cohort 1 cell origin: Donor #1	Cohort 2 cell origin: Donor #2	Cohort 3 cell origin: Donor #3
1	Untreated cells	200		16	16	16
2	"mock electroporation" of cells without any CRISPR/Cas9 editing components	200		16	16	16
3	Electroporation of cells with GFP Guide RNA and Cas9 protein	200	160	16	16	16
4	Electroporation of cells with SPY101 gRNA and Cas9 protein	200		16	16	16
5	Electroporation of cells with SD2 gRNA and Cas9 protein	200		16	16	16
6	Control Untreated	200		10	10	10

Figure 29B

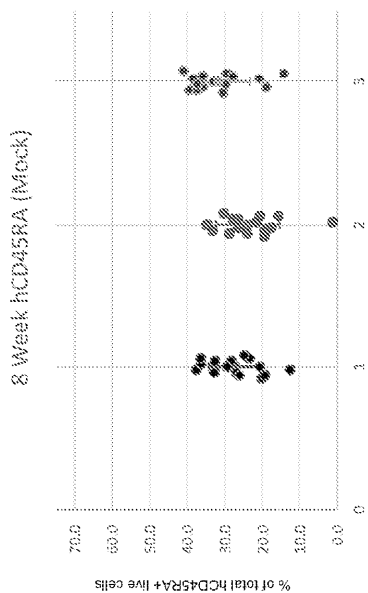


Figure 29A

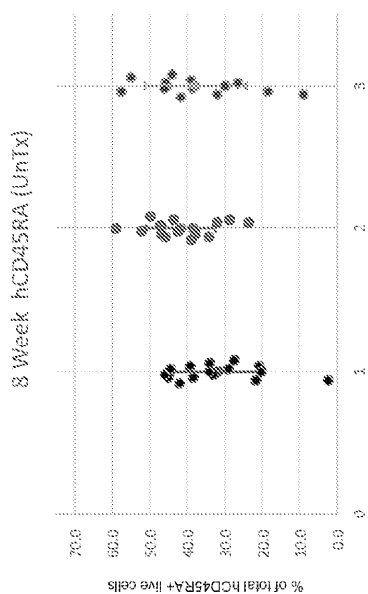


Figure 29C

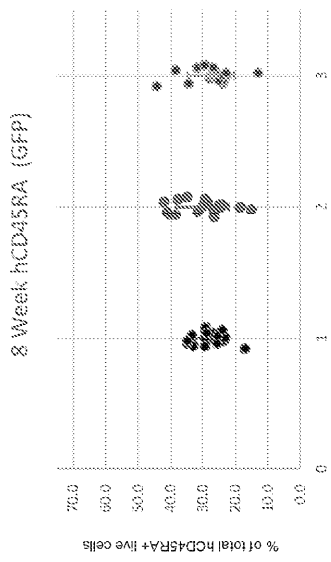


Figure 29D

8 Week hCD45RA (Spy101)

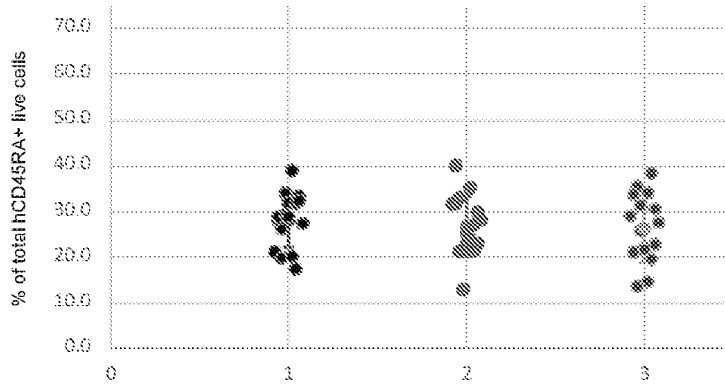


Figure 29E

8 Week hCD45RA (SD2)

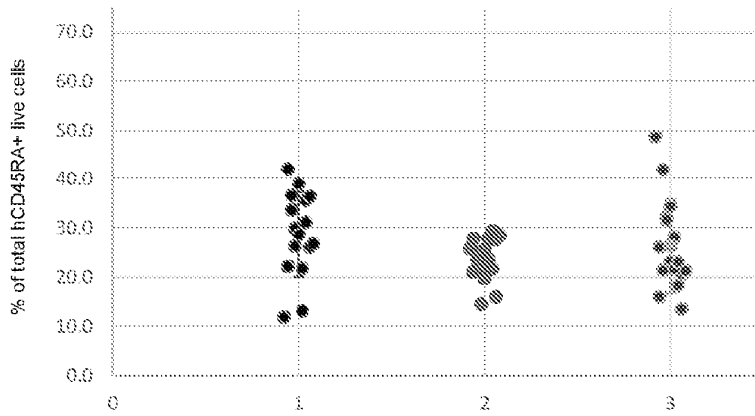


Figure 30

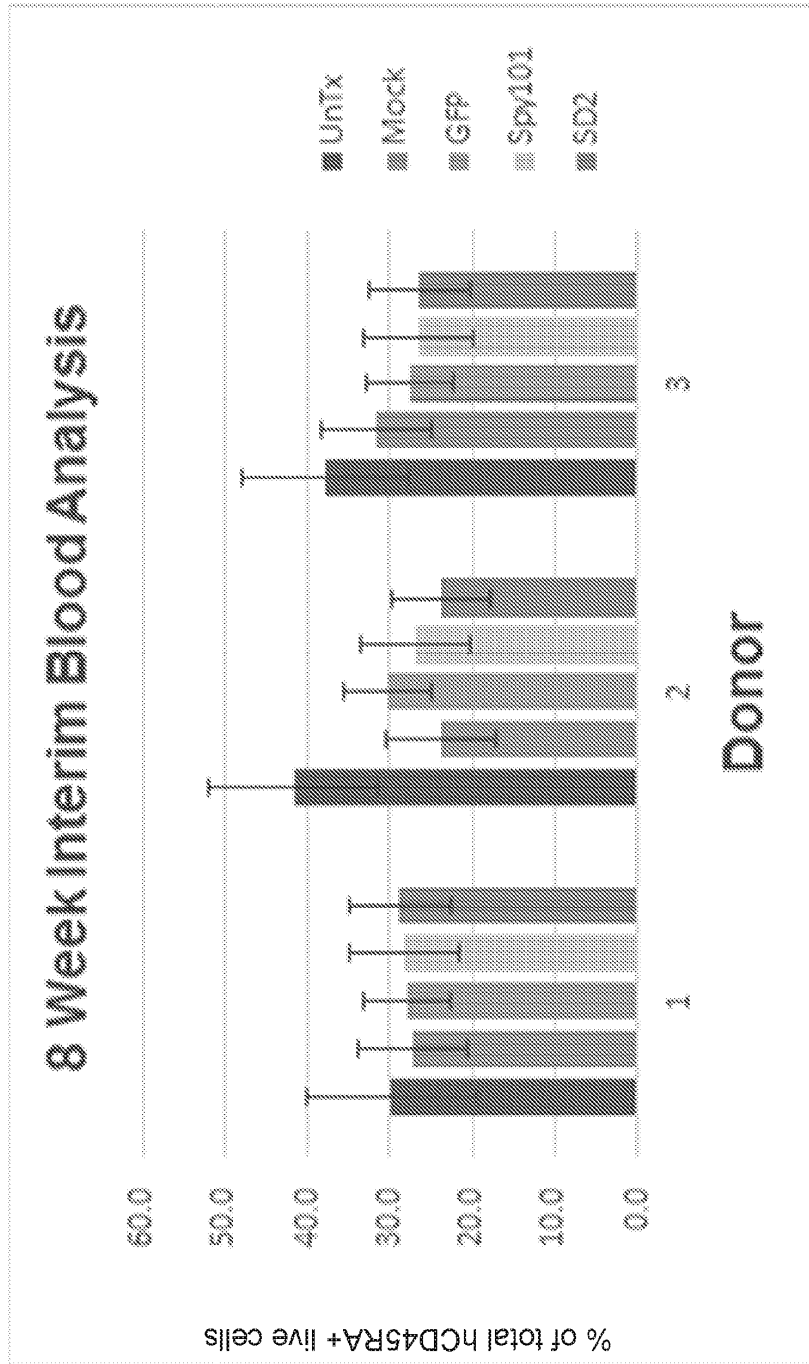
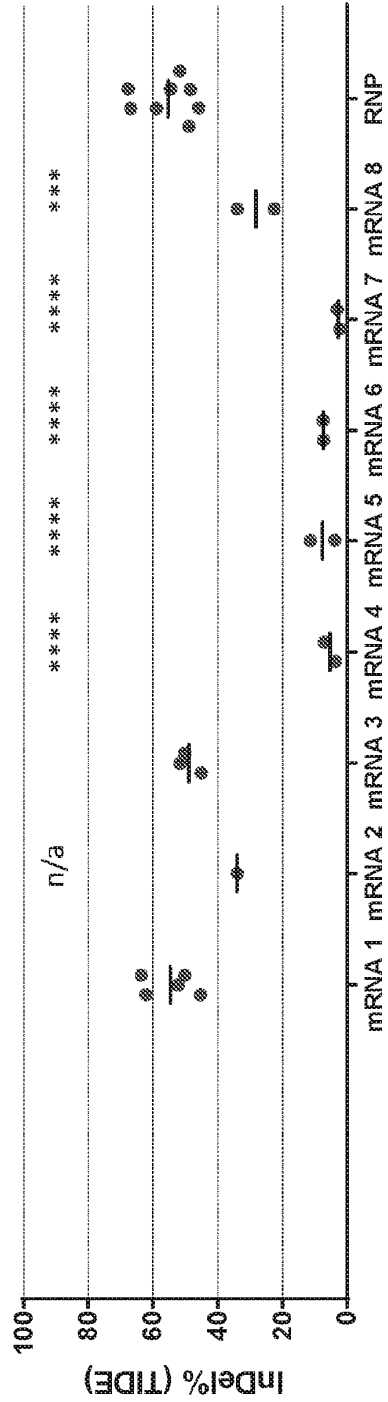


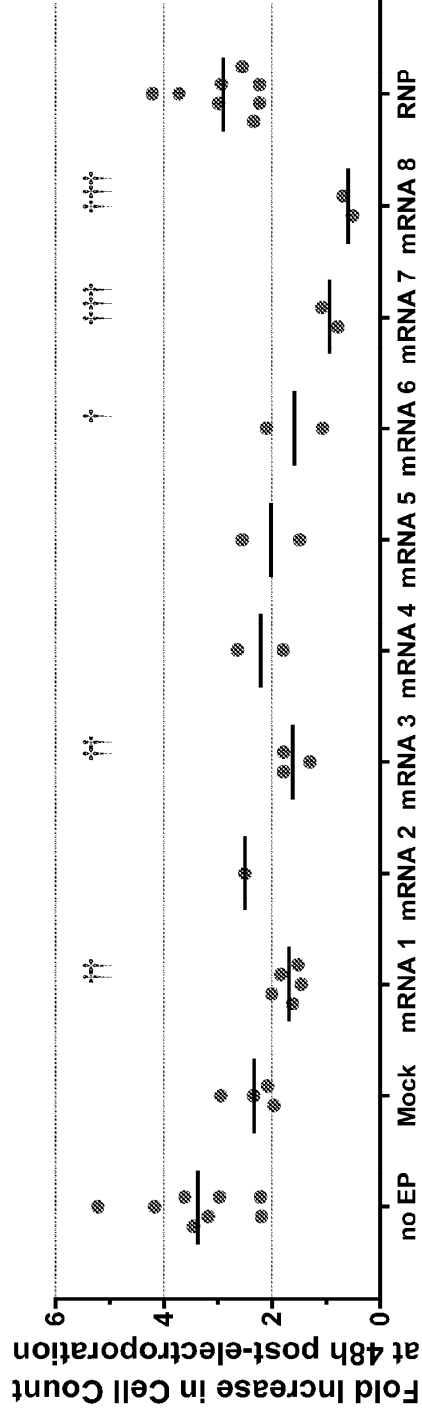
Figure 31
Cas9 RNP is as efficient in editing as Cas9 mRNA but superior in post-electroporation viability



*** p<0.001 compared to RNP by One Way ANOVA
**** p<0.0001 compared to RNP by One Way ANOVA

Cas9 RNP is superior to Cas9 mRNA in post-electroporation cell viability

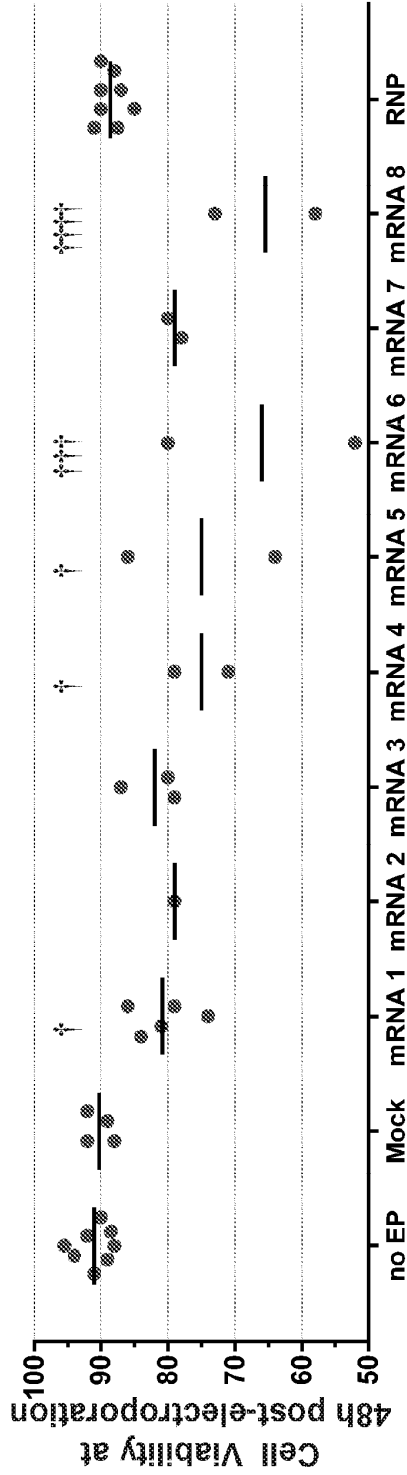
Figure 32A



† p<0.05, †† p<0.01, ††† p<0.001, †††† p<0.0001 compared to No EP by One Way ANOVA

Cas9 RNP is superior to Cas9 mRNA in post-electroporation cell viability

Figure 32B



† p<0.05, †† p<0.01, ††† p<0.001, †††† p<0.0001 compared to No EP by One Way ANOVA

Cas9 RNP optimization

Figure 33A

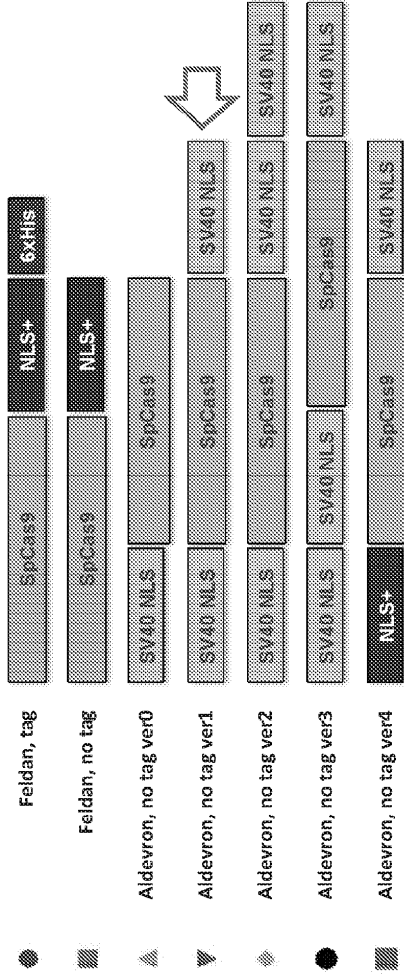


Figure 33B

SPY101 (Cas9 1ug: gRNA 1ug)

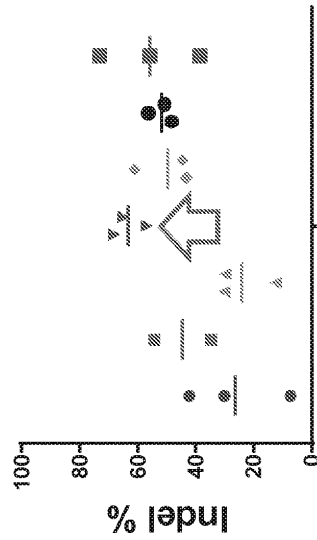
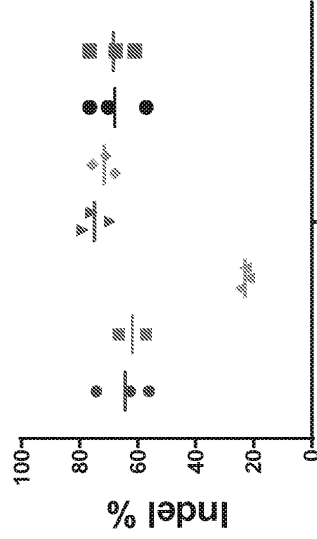


Figure 33C

SPY101 Cas9 3ug: gRNA 3ug)



Bulk Editing Efficiency for SPY101

Figure 34A

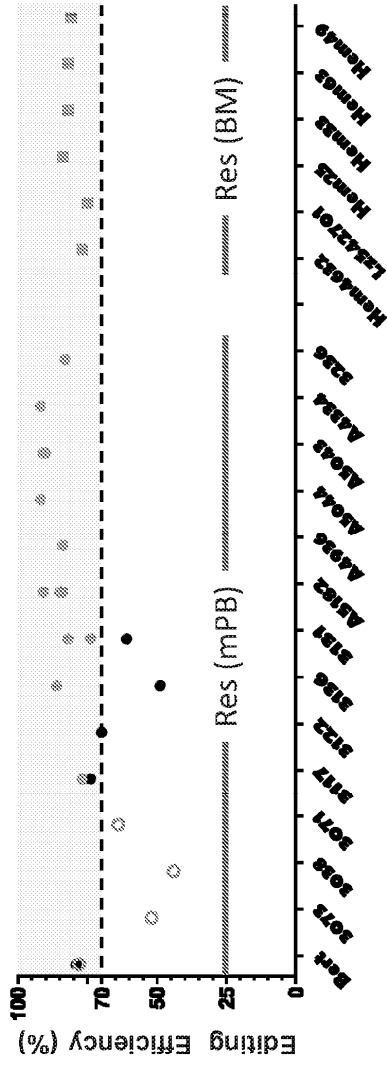
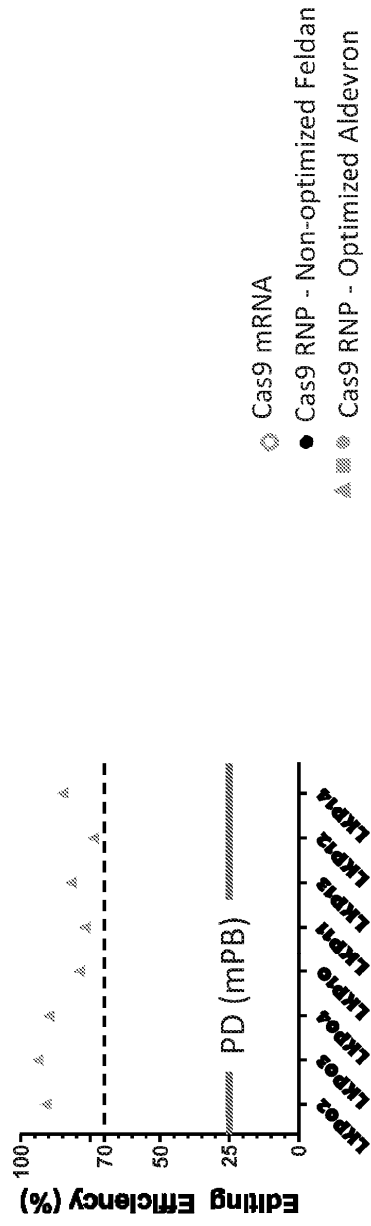


Figure 34B



Efficacy of SPY101 in mPB CD34+ cells

Figure 35A

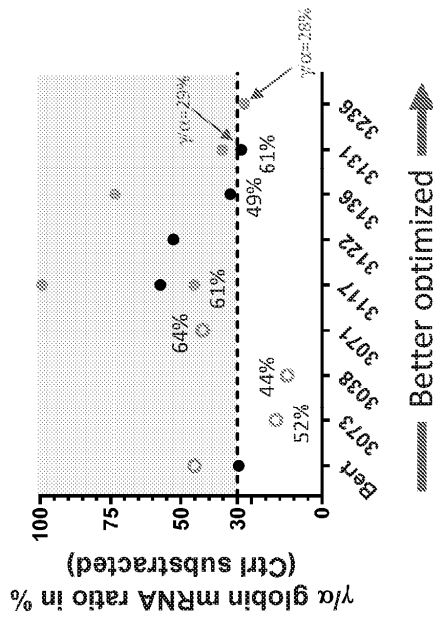
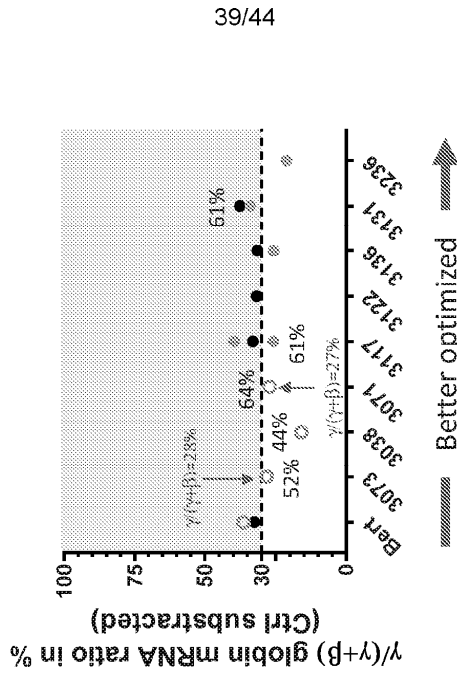


Figure 35B



% = Editing efficiency

All unlabeled data points had
> 70% editing efficiency

- Cas9 mRNA - Non optimized Differentiation
- Cas9 RNP - Non-optimized Differentiation
- ⊗ Cas9 RNP - Optimized Differentiation

➤ Better optimized

Efficacy of SPY101 in BM CD34+ cells

Figure 36A

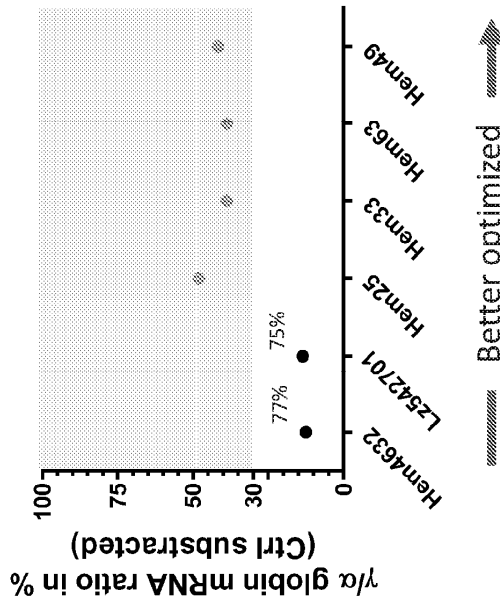
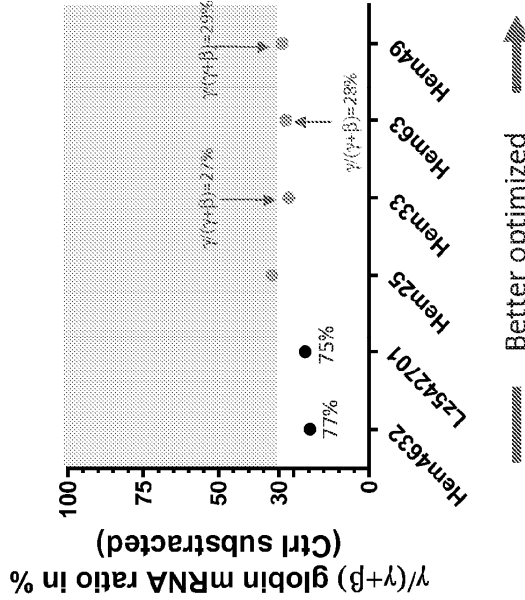


Figure 36B



% = Editing efficiency
 All unlabeled data points had
 > 80% editing efficiency

● Cas9 RNP - Non-optimized
 ◐ Cas9 RNP - Optimized

▨ Better optimized

▨ Better optimized

Efficacy of SPY101 in patient samples

Figure 37A

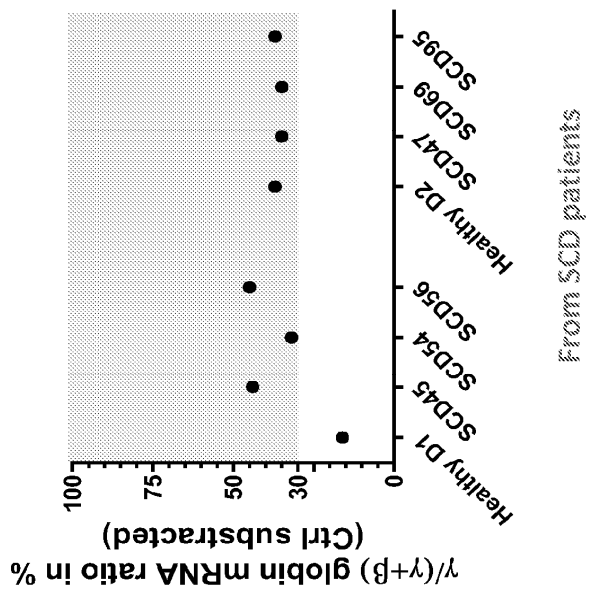
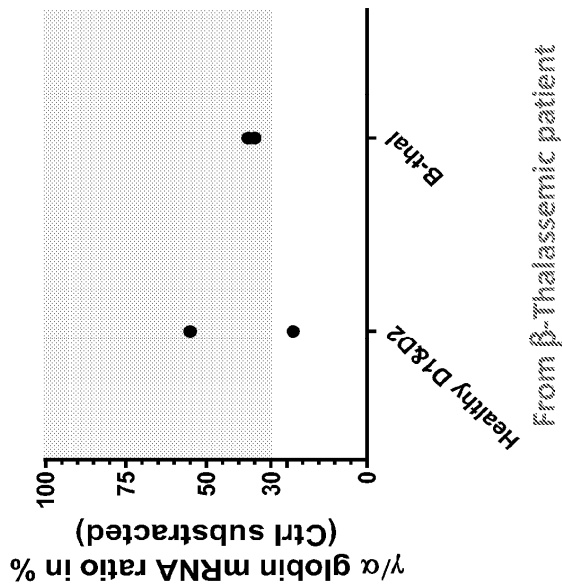


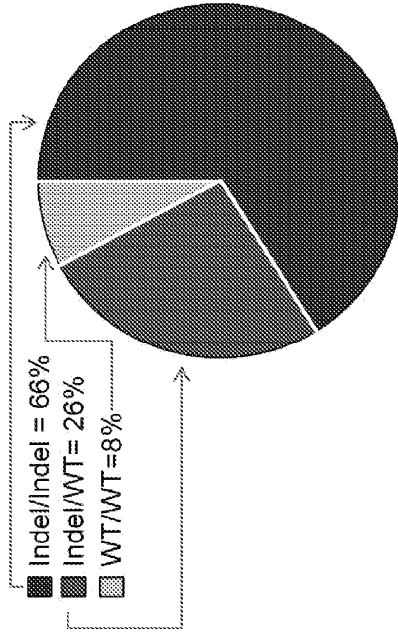
Figure 37B



SPY101 edited genotype breakdown

Figure 38A

Cas9 mRNA

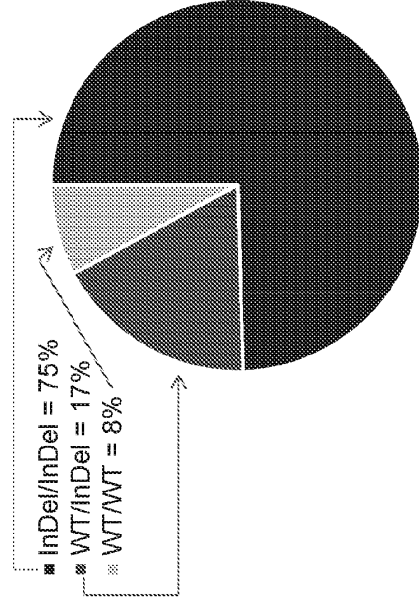


Indel Frequency = 79%
Edited colonies = 92%

Data from 60 colonies

Figure 38B

Cas9 RNP



Indel Frequency = 83%
Edited colonies = 92%

Data from 164 colonies

Gata1 Binding Site disruption by SPY101 linked to increased γ -globin expression

Figure 39A

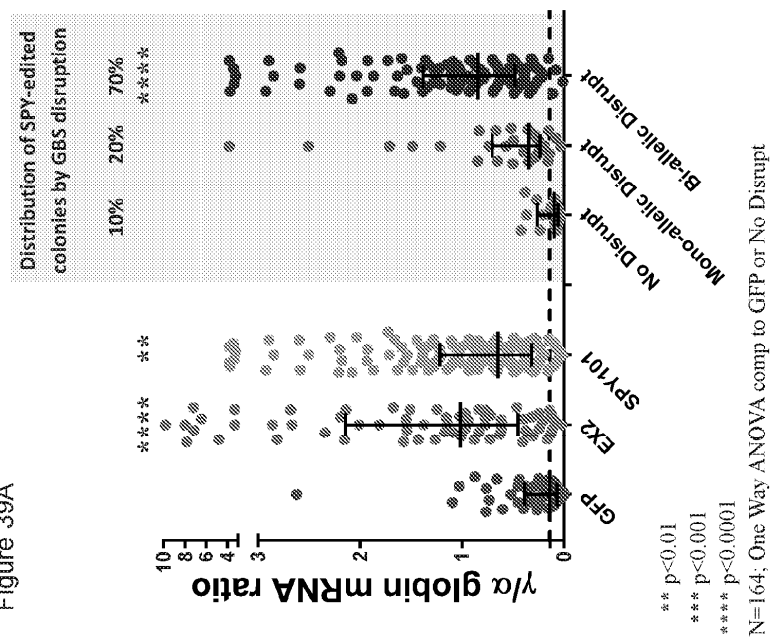
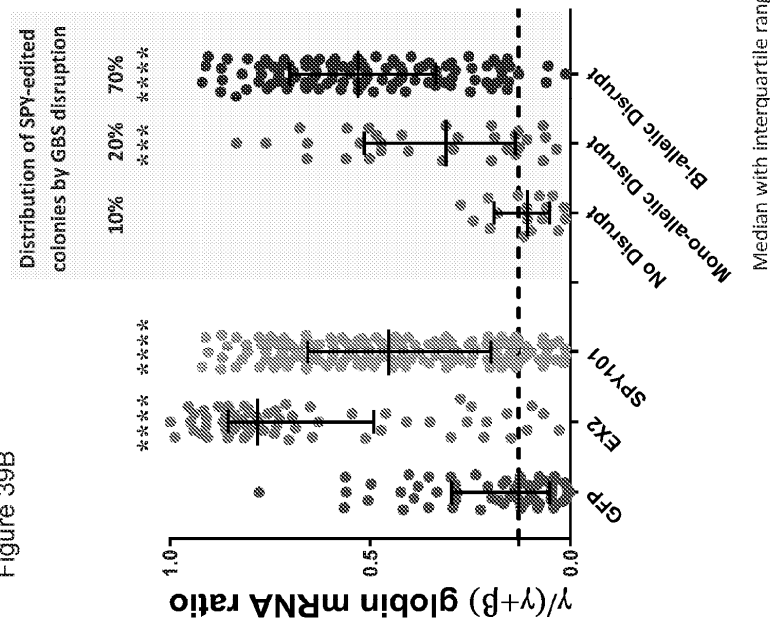


Figure 39B



SPY101 increased γ -globin expression

Figure 40A

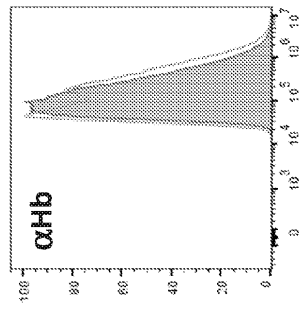


Figure 40B

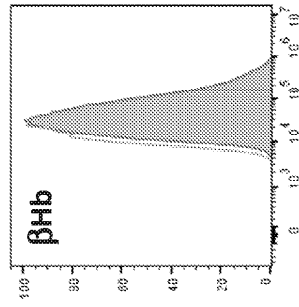


Figure 40C

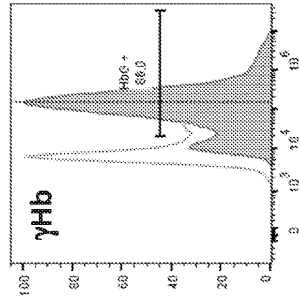


Figure 40D

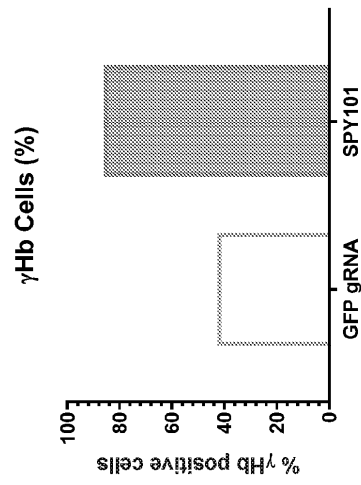


Figure 40E

