

COMPOSITIONS AND METHODS FOR ENGRAFTMENT OF BASE EDITED CELLS**CROSS REFERENCE TO RELATED APPLICATION**

This application claims priority to and benefit of provisional application number
5 62/976,239, filed on February 13, 2020, the entire contents of which are incorporated by
reference herein in their entirety.

SEQUENCE LISTING

10 This application contains a Sequence Listing which has been submitted electronically in
ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created
on February 12, 2021, is named 180802-043701PCT_SL.txt and is 2,097,152 bytes in size.

BACKGROUND

15 Targeted editing of nucleic acid sequences, for example, the targeted cleavage or the
targeted modification of genomic DNA is a highly promising approach for the study of gene
function and also has the potential to provide new therapies for human genetic diseases.
Currently available base editors include cytidine base editors (e.g., BE4) that convert target C•G
base pairs to T•A and adenine base editors (e.g., ABE7.10) that convert A•T to G•C. There is a
20 need in the art for improved targeted editing of nucleic acids for use in treatment of specific
diseases, such as for engraftment to treat a genetic disorder, for example, a genetic disorder
leading to a hematopoietic diseases or disorders, such as Sickle Cell Disease (SCD). Current
methods of treatment are focused on managing the symptoms of the disease. Methods for
editing the genetic mutations that cause sickle cell disease (SCD) are urgently required.

25

SUMMARY OF THE INVENTION

As described below, the present invention features compositions and methods involving
the use of adenine base editors (ABEs), e.g. ABE8.8, that have increased efficiency and methods
of using base editors comprising adenosine deaminase variants for editing a target sequence. As
30 further described herein, such base editors, when introduced (e.g., by electroporation) into
hematopoietic stem cells, hematopoietic progenitor cells and descendants thereof, provide viable
and robust base-edited donor cells, which exhibit stem cell phenotype and activity, and which
demonstrate successful engraftment into the bone marrow of animals in an *in vivo* mouse model.
The base-edited ("edited") cells described and used in the methods herein maintain a high level

of base editing and function over a long term time period, e.g., for at least 8 weeks or for at least 16 weeks, post engraftment.

In an aspect, the invention features a method of engrafting nucleobase-edited hematopoietic stem cells or progenitors thereof in a subject having a hemoglobinopathy. The method involves: (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and a base editor containing a polynucleotide programmable DNA binding domain and a deaminase domain, or a polynucleotide encoding the base editor, where the guide RNA targets the polynucleotide programmable DNA binding domain to induce a nucleobase change in a target hemoglobin (HBB) gene or in the promoter region of *HBG1/2*, thereby obtaining nucleobase-edited hematopoietic stem cells or progenitors thereof; and where the nucleobase-edited hematopoietic stem cells or progenitors thereof are contacted with the gRNA and the base editor within 48 hours following collection from a donor; and (b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after administration. In embodiments, the nucleobase-edited hematopoietic stem cells or progenitors thereof include CD34⁺ cells enriched from polymorphonuclear blood cells (PBMCs) collected from the donor.

In an aspect, the invention features a method of engrafting nucleobase-edited hematopoietic stem cells or progenitors thereof in a subject having a hemoglobinopathy. The method involves: (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and a base editor containing a polynucleotide programmable DNA binding domain and a deaminase domain, or a polynucleotide encoding the base editor, where the guide RNA targets the polynucleotide programmable DNA binding domain to induce a nucleobase change in a target hemoglobin (HBB) gene or in the promoter region of *HBG1/2*, thereby obtaining nucleobase-edited hematopoietic stem cells or progenitors thereof; and (b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after administration.

In an aspect, the invention features a method of treating a hemoglobinopathy in a subject. The method involves: (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and a base editor containing a polynucleotide programmable DNA binding domain and a deaminase domain, or a polynucleotide encoding the base editor, where the guide RNA targets the polynucleotide programmable DNA binding domain to induce a nucleobase change in a target hemoglobin (HBB) gene or in a target hemoglobin (HBB) gene in the promoter region of

HBG1/2, thereby obtaining nucleobase-edited hematopoietic stem cells or progenitors thereof; and (b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after administration.

5 In an aspect, the invention features a method of engrafting nucleobase-edited hematopoietic stem cells or progenitors thereof in a subject having a hemoglobinopathy. The method involves: (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and an adenosine base editor containing a polynucleotide programmable DNA binding domain and an adenosine deaminase domain containing an amino acid sequence with at
10 least 85% sequence identity to the sequence
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID
NO: 3) and containing the alterations Y123H, Y147R, and Q154R, or a polynucleotide encoding
15 the base editor, where the adenosine deaminase domain catalyzes the hydrolytic deamination of adenine or adenosine, and where the guide RNA targets the polynucleotide programmable DNA binding domain to induce an A to G nucleobase change in a target hemoglobin (*HBB*) gene or in the promoter region of *HBG1/2*, thereby obtaining nucleobase-edited hematopoietic stem cells or progenitors thereof; and (b) administering the nucleobase-edited hematopoietic stem cells or
20 progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after administration.

In an aspect, the invention features a method of treating a hemoglobinopathy in a subject. The method involves: (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with
25 a guide RNA and an adenosine base editor containing a polynucleotide programmable DNA binding domain and an adenosine deaminase domain containing an amino acid sequence with at least 85% sequence identity to
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
30 SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 3) and containing the alterations Y123H, Y147R, and Q154R, or a polynucleotide encoding the base editor, where the adenosine deaminase domain catalyzes the hydrolytic deamination of adenine or adenosine, and where the guide RNA targets the polynucleotide programmable DNA binding domain to induce an A to G nucleobase change in a target hemoglobin (*HBB*) gene or in

the promoter region of *HBG1/2*, thereby obtaining nucleobase-edited hematopoietic stem cells or progenitors thereof; and (b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after

5 administration.

In an aspect, the invention features a method of engrafting edited hematopoietic stem cells or progenitors thereof in a subject having a hemoglobinopathy. The method involves: (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and a base editor containing an amino acid sequence with at least 80% sequence identity to one of the

10 following two amino acid sequences:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
 AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
 SLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSSGSSG
 GSSGSETPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVL
 15 GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
 DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLLI
 YLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
 RLKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDITYDDD
 LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL
 20 LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
 REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLA
 RGNRSFAWMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLY
 EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
 CFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
 25 LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
 MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
 RHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
 LYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
 PSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
 30 HVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYK VREINNYHHAHDA
 YLNAVVG TALIKKYPKLESEFVYGDYKVYDVVRKMIKSEQEIGKATAKYFFYSNIMNFF
 KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFS
 KESILPKRNSDKLIARKKDWDPKKGFFSPTVAYSVLVVAKVEKGKSKKLKSVKELL
 GITIMERSSEKPNIDFLEAKGYKEVKKDLIKLPKYSLELENGRKRMLASAGELQKGN

ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLAD
ANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 258),

and

5 MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAA
GSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSSGGSSG
GSSGSETPGTSESATPESSGGSSGSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVL
VLNNRVIGEGWNR AIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAG
10 AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFR
MPRRVFNAQKKAQSSTDSSGGSSGSSGSETPGTSESATPESSGGSSGSSGSDKKYSIGLAIG
TNSVGWAVITDEYKVPSKKFKVLGNDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLV DSTDKADLRLLIYALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQT
15 YNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNF
KSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEE
FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGASAQSEFI
20 ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNE
ENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
25 KELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVS
DFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
30 RKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKEK GKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEF

ESPKKKRKV (SEQ ID NO: 259), or a polynucleotide encoding the base editor, where the guide RNA targets the polynucleotide programmable DNA binding domain to induce an A to G nucleobase change in the promoter region of HBG1/2, thereby obtaining edited hematopoietic stem cells or progenitors thereof; and (b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after administration.

In an aspect, the invention features a method of treating a hemoglobinopathy in a subject. The method involves: (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and a base editor containing an amino acid sequence with at least 80% sequence identity to one of the following two amino acid sequences

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSGSSG
GSSGSETPGTSESATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVL
GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLLI
YLALAHMIKFRGHFLIEGDLNPDNSDVKLFQQLVQTYNQLFEENPINASGVDAKAILS
RLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSAILLSDILRVNTEITKAPLSASMIKRYDEHHQDRTL
LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEIFYKFIKPILEKMDGTEELLVKLN
REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSHLLY
EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMG
RHKPENIVIAMARENQTTQKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDA
YLNNAVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEIEGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL

GITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSTGLYETRIDLSQLGGDEGADKRTADGSEFESPCKKRKV (SEQ ID NO: 258),
5 and
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAA
GSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSG
GSSGSETPGTSESATPESSGGSSGSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVL
10 VLNNRVIGEGWNRRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFR
MPRRVFNAQKKAQSSTDSGGSSGSSGSETPGTSESATPESSGGSSGSSGSKKYSIGLAIG
TNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
15 YPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALLSLGLTPNF
KSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE
FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
20 LKDNREKIEKILTFRIPYYVGPLARGNSRFAMTRKSEETITPWNFEEVVDKGASQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKKIHKDKDFLDNE
ENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
25 SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYG DYKVYDVRKMI
30 AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKEVGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRRDKPIREQAENIIHLFTLTNLGAP

AAFKYFDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDLSQLGGDEGADKRTADGSEF
 ESPKKKRKV (SEQ ID NO: 259), or a polynucleotide encoding the base editor, where the
 guide RNA targets the polynucleotide programmable DNA binding domain to induce an A to G
 nucleobase change in a target hemoglobin (HBB) gene or in the promoter region of HBG1/2,
 5 thereby obtaining edited hematopoietic stem cells or progenitors thereof; and
 (b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a
 subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem
 cells or progenitors thereof in tissues of the subject after administration.

10 In an aspect, the invention features a kit for use in the method of any one of the above
 aspects, where the kit contains the guide RNA and a polynucleotide encoding the base editor.

In any of the above aspects and/or embodiments thereof, the nucleobase change is an A
 to G nucleobase change.

In any of the above aspects and/or embodiments thereof, the deaminase domain is an
 adenosine deaminase domain and shares at least 85% sequence identity with the sequence
 15 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
 AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
 SLMDVLHYPGMNRHVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID
 NO: 3), and the adenosine deaminase domain is capable of catalyzing the hydrolytic deamination
 of adenine or adenosine. In embodiments, the adenosine deaminase domain contains one or
 20 more of the following alterations: Y123H, Q154S, and Q154R. In embodiments, the adenosine
 deaminase domain contains one or more of the following alterations: Y147T, Y147R, Q154S,
 Y123H, and Q154R. In embodiments, the adenosine deaminase domain contains a combination
 of alterations selected from one or more of the following: Y147R, Q154R, and Y123H; Y147R,
 Q154R, and I76Y; Y147R, Q154R, and T166R; Y147T and Q154R; Y147T and Q154S; and
 25 Y123H, Y147R, Q154R, and I76Y. In embodiments, the adenosine deaminase domain contains
 the alterations Y147R, Q154R, and Y123H. In embodiments, the adenosine deaminase domain
 contains an alteration at position 82 or 166. In embodiments, the alteration at position 82 is
 V82S. In embodiments, the alteration at position 166 is T166R. In embodiments, the adenosine
 deaminase domain contains an alteration at positions 166 and 82. In embodiments, the
 30 adenosine deaminase domain has at least 90% sequence identity to the sequence.

In any of the above aspects and/or embodiments thereof, the deaminase domain is a
 TadA*8 variant. In any of the above aspects and/or embodiments thereof, the TadA*8 variant is
 selected from one or more of the following: TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4,
 TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12,

and TadA*8.13. In any of the above aspects and/or embodiments thereof, the base editor is an ABE8 base editor selected from one or more of the following: ABE8.1, ABE8.2, ABE8.3, ABE8.4, ABE8.5, ABE8.6, ABE8.7, ABE8.8, ABE8.9, ABE8.10, ABE8.11, ABE8.12, and ABE8.13.

5 In any of the above aspects and/or embodiments thereof, the base editor further contains a wild-type adenosine deaminase domain.

In any of the above aspects and/or embodiments thereof, the polynucleotide programmable DNA binding domain is a Cas9. In embodiments, the Cas9 is a SpCas9, a SaCas9, or a variant thereof.

10 In any of the above aspects and/or embodiments thereof, the polynucleotide programmable DNA binding domain contains a modified Cas9 having an altered protospacer-adjacent motif (PAM) specificity. In embodiments, the Cas9 has specificity for a PAM sequence selected from one or more of the following NGG, NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, and NGC, where N is A, G, C, or T and where R is A or G.

15 In any of the above aspects and/or embodiments thereof, the polynucleotide programmable DNA binding domain is nuclease inactive. In any of the above aspects and/or embodiments thereof, the polynucleotide programmable DNA binding domain is a nickase. In any of the above aspects and/or embodiments thereof, the polynucleotide programmable DNA binding domain contains the alterations D10A and/or H840A. In any of the above aspects and/or
20 embodiments thereof, the polynucleotide programmable DNA binding domain contains the alteration D10A.

In any of the above aspects and/or embodiments thereof, the deaminase domain contains an adenosine deaminase monomer. In any of the above aspects and/or embodiments thereof, the deaminase domain contains an adenosine deaminase dimer.

25 In any of the above aspects and/or embodiments thereof, the engraftment efficiency of the nucleobase-edited hematopoietic stem cells or progenitors thereof is measured in the subject at about 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, or 8 or more weeks after administering the cells to the subject. In any of the above aspects and/or embodiments thereof, the engraftment efficiency of the nucleobase-edited hematopoietic stem cells or progenitors
30 thereof is measured in the subject at least 8 weeks after administering the cells to the subject. In any of the above aspects and/or embodiments thereof, the engraftment efficiency of the nucleobase-edited hematopoietic stem cells or progenitors thereof is measured in the subject at least 16 weeks after administering the cells to the subject. In embodiments, the measured engraftment efficiency is at least about 20%. In embodiments, the measured engraftment

efficiency is at least about 30%. In embodiments, the measured engraftment efficiency is at least about 40%. In embodiments, the measured engraftment efficiency is at least about 50%.

In any of the above aspects and/or embodiments thereof, at least about 50% of the hematopoietic cells or progenitors thereof in (b) are viable. In any of the above aspects and/or
5 embodiments thereof, at least 30% of the hematopoietic cells or progenitors thereof in (b) contain the nucleobase change. In any of the above aspects and/or embodiments thereof, at least 50% of the hematopoietic cells or progenitors thereof in (b) contain the nucleobase change. In any of the above aspects and/or
10 embodiments thereof, at least 60% of the hematopoietic cells or progenitors thereof in (b) contain the nucleobase change. In any of the above aspects and/or
10 embodiments thereof, at least 70% of the hematopoietic cells or progenitors thereof in (b) contain the nucleobase change.

In any of the above aspects and/or embodiments thereof, the hematopoietic cells or progenitors thereof are isolated or derived from the subject.

In any of the above aspects and/or embodiments thereof, the hematopoietic stem cells or
15 progenitors thereof contain a single-nucleotide polymorphism (SNP) associated with sickle cell disease (SCD). In embodiments, the SNP associated with SCD results in a E6V substitution in a hemoglobin beta unit encoded by the HBB gene. In any of the above aspects and/or
embodiments thereof, the nucleobase change results in a E6A substitution in the hemoglobin beta unit encoded by the HBB gene.

20 In any of the above aspects and/or embodiments thereof, at least 30% of the hematopoietic stem cells or progenitors thereof retain base editing activity following engraftment. In any of the above aspects and/or embodiments thereof, at least 50% of the hematopoietic stem cells or progenitors thereof retain base editing activity following engraftment. In any of the above aspects and/or embodiments thereof, at least 60% of the
25 hematopoietic stem cells or progenitors thereof retain base editing activity following engraftment. In any of the above aspects and/or embodiments thereof, at least 70% of the hematopoietic stem cells or progenitors thereof retain base editing activity following engraftment. In any of the above aspects and/or embodiments thereof, at least 80% of the hematopoietic stem cells or progenitors thereof retain base editing activity following
30 engraftment. In any of the above aspects and/or embodiments thereof, at least 90% of the hematopoietic stem cells or progenitors thereof retain base editing activity following engraftment.

In any of the above aspects and/or embodiments thereof, the hematopoietic cells or progenitors thereof retain the ability to differentiate following administration. In any of the

above aspects and/or embodiments thereof, the hematopoietic cells or progenitors thereof are capable of generating erythrocytes. In any of the above aspects and/or embodiments thereof, the polynucleotide encoding the base editor contains mRNA or is mRNA.

In any of the above aspects and/or embodiments thereof, the hematopoietic stem cells or progenitors thereof are contacted with at least about 1 nM of mRNA encoding the base editor. In any of the above aspects and/or embodiments thereof, the hematopoietic stem cells or progenitors thereof are contacted with at least about 3 nM RNA encoding the base editor. In any of the above aspects and/or embodiments thereof, the hematopoietic stem cells or progenitors thereof are contacted with at least about 10 nM RNA encoding the base editor. In any of the above aspects and/or embodiments thereof, the hematopoietic stem cells or progenitors thereof are contacted with at least about 30 nM RNA encoding the base editor. In any of the above aspects and/or embodiments thereof, the hematopoietic stem cells or progenitors thereof are contacted with at least about 50 nM RNA encoding the base editor. In any of the above aspects and/or embodiments thereof, the hematopoietic stem cells or progenitors thereof are contacted with at least about 3000 nM of the gRNA.

In any of the above aspects and/or embodiments thereof, levels of fetal hemoglobin (HbF) are increased in the subject following engraftment relative to the levels in a control subject that received unedited hematopoietic stem cells or progenitors thereof. In any of the above aspects and/or embodiments thereof, levels of fetal hemoglobin (HbF) are increased in the subject by at least about 20% relative to the levels in a control subject that received unedited hematopoietic stem cells or progenitors thereof. In any of the above aspects and/or embodiments thereof, HbS expression is reduced in the subject following engraftment relative to HbS expression in a control subject that received unedited hematopoietic stem cells or progenitors thereof. In any of the above aspects and/or embodiments thereof, HbS expression is reduced in the subject by at least about 20% relative to HbS expression in a control subject that received unedited hematopoietic stem cells or progenitors thereof.

In any of the above aspects and/or embodiments thereof, the nucleobase-edited hematopoietic stem cells or progenitors thereof express CD34 (e.g., are CD34⁺). In any of the above aspects and/or embodiments thereof, the nucleobase-edited hematopoietic stem cells or progenitors thereof express one or more of CD34, CD45, CD19, and GlyA. In any of the above aspects and/or embodiments thereof, the nucleobase-edited hematopoietic stem cells or progenitors thereof are GlyA⁺.

In any of the above aspects and/or embodiments thereof, the nucleobase-edited hematopoietic stem cells or progenitors thereof express fetal hemoglobin (HbF).

In any of the above aspects and/or embodiments thereof, the hematopoietic stem cells or progenitors thereof are human hematopoietic stem cells or progenitors thereof. In any of the above aspects and/or embodiments thereof, the subject is a mammal. In any of the above aspects and/or embodiments thereof, the subject is a human.

5 In any of the above aspects and/or embodiments thereof, the subject has sickle cell disease (SCD), thalassemia, and/or anemia. In any of the above aspects and/or embodiments thereof, the subject has SCD.

In any of the above aspects and/or embodiments thereof, the nucleobase-edited hematopoietic stem cells or progenitors thereof are autologous to the subject.

10 In any of the above aspects and/or embodiments thereof, the nucleobase-edited hematopoietic stem cells or progenitors thereof are not enriched prior to administration. In any of the above aspects and/or embodiments thereof, the nucleobase-edited hematopoietic stem cells or progenitors thereof are enriched prior to administration.

In any of the above aspects and/or embodiments thereof, the nucleobase change
15 abolishes, disrupts, or reduces BCL11A binding in the promoter region of *HBG1/2*. In any of the above aspects and/or embodiments thereof, the nucleobase change is at a position selected from -114, -117, -175, and -198 in the promoter region of *HBG1/2*. In any of the above aspects and/or embodiments thereof, the nucleobase change is associated with an increase in expression of *HBG1/2*.

20 In any of the above aspects and/or embodiments thereof, the nucleobase change is associated with an increase in levels of hemoglobin gamma subunit in the hematopoietic stem cells or progenitors thereof. In any of the above aspects and/or embodiments thereof, an increased level of HbF protein is expressed in the subject after administration. In any of the above aspects and/or embodiments thereof, the administration results in expression of HbF in the
25 subject for at least 8 weeks. In any of the above aspects and/or embodiments thereof, the administration results in expression of HbF in the subject for at least 16 weeks.

In any of the above aspects and/or embodiments thereof, the administration reduces or ameliorates a symptom associated with sickle cell disease in the subject. In any of the above aspects and/or embodiments thereof, erythrocytes generated from the hematopoietic cells or
30 progenitors thereof exhibit reduced sickling.

In any of the above aspects and/or embodiments thereof, at least 50% editing is retained at least 16 weeks after the administration in a tissue of the subject. In any of the above aspects and/or embodiments thereof, at least 80% editing is retained 16 weeks after the administration in a tissue of the subject.

In any of the above aspects and/or embodiments thereof, administration is performed multiple times. In any of the above aspects and/or embodiments thereof, administration is performed multiple times at an interval of at least about one month.

In any of the above aspects and/or embodiments thereof, the guide RNA contains a nucleotide sequence selected from SEQ ID NOs: 130-155 listed in Table 1. In any of the above aspects and/or embodiments thereof, the gRNA contains or is the sequence, from 5'-3':
 GACCAAUAGCCUUGACAGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUA
 GUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU, corresponding to bases 4-97 of SEQ ID NO: 129. In any of the above aspects and/or embodiments thereof, the guide
 RNA contains or is the nucleotide sequence, from 5'-3':
 csususGACCAAUAGCCUUGACAGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAG
 GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU_{sususu} (SEQ ID
 NO: 129), where lowercase characters indicate 2'-O-methylated nucleobases, and "s" indicates
 phosphorothioates (SEQ ID NO: 129). In any of the above aspects and/or embodiments thereof,
 the guide RNA contains or is the nucleotide sequence of any one of 5'-
 gsascsUUCUCCACAGGAGUCAGGGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAA
 GGCUAGUCCGUUAUUAACUUGAAAAAGUGGCACCGAGUCGGUGCU_{sususu}-3' (SEQ
 ID NO: 126), 5'-
 ascsusUCUCCACAGGAGUCAGGGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAG
 GCUAGUCCGUUAUUAACUUGAAAAAGUGGCACCGAGUCGGUGCU_{sususu}-3' (SEQ
 ID NO: 127), and 5'-
 csususCUCCACAGGAGUCAGGGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGG
 CUAGUCCGUUAUUAACUUGAAAAAGUGGCACCGAGUCGGUGCU_{sususu}-3' (SEQ ID
 NO: 128), where lowercase characters indicate 2'-O-methylated nucleobases, and "s" indicates
 phosphorothioates.

In any of the above aspects and/or embodiments thereof, the administration is associated with hemoglobin subunit gamma being expressed in at least 50% of cells in the bone marrow of the subject. In any of the above aspects and/or embodiments thereof, the administration is associated with hemoglobin subunit gamma being expressed in at least 60% of cells in the bone marrow of the subject.

In any of the above aspects and/or embodiments thereof, the method further involves depletion of one or more lymphocytic lineage cells in the subject prior to administering the hematopoietic stem cells or progenitors thereof.

In any of the above aspects and/or embodiments thereof, the hematopoietic stem cells or progenitors thereof are enriched CD34⁺ cells, and the CD34⁺ cells are enriched from donor peripheral blood mononuclear cells (PBMCs) less than 24 hours after the PBMCs are collected or isolated from a donor. In any of the above aspects and/or embodiments thereof, the

5 hematopoietic stem cells or progenitors thereof are enriched CD34⁺ cells, and the CD34⁺ cells are enriched from donor peripheral blood mononuclear cells (PBMCs) less than 48 hours after the PBMCs are collected or isolated from a donor. In any of the above aspects and/or

embodiments thereof, the hematopoietic stem cells or progenitors thereof are cryopreserved following collection or isolation from a donor.

10 In any of the above aspects and/or embodiments thereof, the gRNA and/or the polynucleotide encoding the base editor contains a 2'-O-Methyl nucleotide modification. In any of the above aspects and/or embodiments thereof, the 2'-O-Methyl nucleotide modification is disposed at a 3' or 5' end of the gRNA and/or the polynucleotide encoding the base editor. In any of the above aspects and/or embodiments thereof, the gRNA and/or the polynucleotide encoding

15 the base editor contains a phosphorothioate internucleotide linkage.

In any of the above aspects and/or embodiments thereof, the hematopoietic stem cells or progenitors thereof are contacted with the polynucleotide encoding the base editor. In any of the above aspects and/or embodiments thereof, the base editor is delivered as a polynucleotide that is expressed in the hematopoietic stem cells or progenitors thereof.

20 In any of the above aspects and/or embodiments thereof, engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof is maintained in the subject for at least 8 weeks. In any of the above aspects and/or embodiments thereof, engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof is maintained in the subject for at least 16 weeks. In any of the above aspects and/or embodiments thereof, the nucleobase-edited

25 hematopoietic stem cells or progenitors thereof are contacted with the gRNA and the base editor within 24 hours following collection from a donor.

In any of the above aspects and/or embodiments thereof, the base editor shares at least 90% sequence identity to one of the following two sequences:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAH

30 AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG

SLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSG

GSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVL

GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD

DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LI

YLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDRTL
LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
5 REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLY
EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
10 MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIAMARENQTTQKGQKNSRERMKRIEIGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVHDHVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVBREINNYHHAHDA
15 YLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
GITIMERSSEFKNPIDFLEAKGYKEVKDLIJKLPKYSLELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLAD
20 ANLDKVL SAYNKHDKPIREQAENIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSTGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 258),
and
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAA
25 GSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSG
GSSGSETPGTSESATPESSGGSSGSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVL
VLNNRVIGEGWNRAGLHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFR
MPRRVFNAQKKAQSSTDSGGSSGSSGSETPGTSESATPESSGGSSGSSGSDKKYSIGLAIG
30 TNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLVDSADKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTPNF
KSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI

TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE
 FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFP
 LKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF
 ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 5 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNE
 ENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
 SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
 KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHDHIVPQSFL
 10 KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAE
 RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVS
 DFRKDFQFYKVRINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMI
 AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKPLIETNGETGEIVWDKGRDFATV
 RKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
 15 VLVVAKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYS
 LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVE
 QHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAP
 AAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEF
 ESPKKKRKV (SEQ ID NO: 259).

20 In any of the above aspects and/or embodiments thereof, the base editor shares at least
 95% sequence identity to one of the following two sequences:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGA VLVLNNRVIGEGWNRAIGLHDPTAH
 AEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
 SLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSG
 25 GSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVL
 GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVD
 DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LI
 YLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSA
 RLKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
 30 LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL
 LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
 REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKDNREKIEKILTRIPYYVGPLA
 RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY
 EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE

CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMG
RHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLY
5 LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYK VREINNYHHAHDA
YLNAAVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFS
10 KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEK GKSKKLKSVKELL
GITIMERSSEKPNIDFLEAKGYKEVKKDLIILPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 258),
15 and
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAA
GSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAKKAQSSTDSGGSSG
GSSGSETPGTSESATPESSGGSSGSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVL
20 VLNNRVIGEGWNR AIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFR
MPRRVFNAQKKAQSSTDSGGSSGSSGSETPGTSESATPESSGGSSGSSGSDKKYSIGLAIG
TNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
25 YPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIASLGLTPNF
KSNFDLAEDAQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE
FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
30 LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV DKGASAQSFI
ERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNE
ENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
SPAIIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGI

KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVS
DFRKDFQFYKVRINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI
5 AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIILPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAP
10 AAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEF
ESPKKKRKV (SEQ ID NO: 259).

The description and examples herein illustrate embodiments of the present disclosure in detail. It is to be understood that this disclosure is not limited to the particular embodiments described herein and as such can vary. Those of skill in the art will recognize that there are
15 numerous variations and modifications of this disclosure, which are encompassed within its scope.

Although various features of the present disclosure can be described in the context of a single embodiment, the features can also be provided separately or in any suitable combination. Conversely, although the present disclosure can be described herein in the context of separate
20 embodiments for clarity, the present disclosure can also be implemented in a single embodiment. The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

The features of the present disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present will be obtained by
25 reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and in view of the accompanying drawings as described hereinbelow.

Definitions

30 Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton *et al.*, Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary

of Genetics, 5th Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them below, unless specified otherwise.

The term “engraftment” or “engrafting” as used herein refers to the process by which
 5 cells administered to a subject, e.g. a recipient, as well as precursors and descendants of the cells, are incorporated into a tissue or organ of the subject. In an embodiment, the tissue is bone marrow. In embodiments, the cell is a hematopoietic stem cell (HSC), a progenitor of a hematopoietic stem cell, or a bone marrow cell. In embodiments, cells administered, introduced, or transplanted into a recipient for engraftment, travel through the bloodstream and home to free
 10 bone marrow (BM) niches which provide optimal conditions for their survival, proliferation, and generation of new blood cells, including red blood cells (erythrocytes), white blood cells (leukocytes, such as monocytes, macrophages and neutrophils) and platelets.

“Engraftment efficiency” refers to the fraction or percentage of cells (e.g., donor cells) incorporated in a tissue (e.g., bone marrow) or organ following administration (e.g.,
 15 transplantation) into a recipient subject. In embodiments, engraftment efficiency is measured 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, 8 weeks, 9 weeks, 10 weeks, 11 weeks, 12 weeks, 13 weeks, 14 weeks, 15 weeks, 16 weeks, 17 weeks, 18 weeks, 19 weeks, or 20 weeks following administration of the cells to a subject. Such incorporated cells constitute those that were administered to the subject (and/or descendants thereof following administration
 20 of the cell(s) to the subject). For example, engraftment efficiency of a donor hematopoietic stem cell (HSC) administered to a subject and comprising a nucleobase change (i.e., an “edited” or “nucleobase-edited” cell) may be expressed as the percentage of donor cells in a tissue (e.g., bone marrow) of the subject comprising the nucleobase change and/or cells descended from the HSC administered. Engraftment efficiency may be monitored by measuring complete blood cell
 25 count (and assessing blood cell lineages and phenotypes) over repeated time periods. An increase in counts of cells administered to a subject and descendants thereof over time indicates that engraftment is occurring or has occurred. In an embodiment, the HSC, progenitors of hematopoietic stem cells, or bone marrow cells that are engrafted are nucleobase-edited. In an embodiment, the nucleobase editing induces an A to G nucleobase change in the promoter region
 30 of the *HBG1/2* polynucleotide. In general, the cells or nucleobase-edited cells that are engrafted, e.g., HSC, progenitors of hematopoietic stem cells, or bone marrow cells, into tissues or organs of a recipient subject are also termed “donor” cells. In an embodiment, the cells are obtained from a donor subject.

As used herein, sickle cell disease (SCD) refers to a group of disorders that affects hemoglobin, the molecule in red blood cells that delivers oxygen to cells throughout the body. Individuals with this disorder have atypical hemoglobin molecules, which can distort red blood cells into a sickle, or crescent, shape. SCD affects beta globin function and can lead to severe anemia and progressive multiple organ failure. The clinical manifestations of sickle cell disease (SCD) result from intermittent episodes of microvascular occlusion leading to tissue ischemia/reperfusion injury and chronic hemolysis. Vaso-occlusive events are associated with ischemia/reperfusion damage to tissues resulting in pain and acute or chronic injury affecting any organ system. The bones/marrow, spleen, liver, brain, lungs, kidneys, and joints are often affected. SCD is a genetic disorder characterized by the presence of at least one hemoglobin S allele (HbS; p.Glu6Val in *HbB*) and a second *HbB* pathogenic variant resulting in abnormal hemoglobin polymerization. HbS/S (homozygous p.Glu6Val in *HbB*) accounts for 60%-70% of sickle cell disease (SCD) in the United States. The life expectancy for men and women suffering from sickle cell disease (SCD) is only 42 and 48 years, respectively.

By “ β -globin (HbB) protein” is meant a polypeptide or fragment thereof having at least about 95% amino acid sequence identity to NCBI Accession No. NP_000509. In particular embodiments, a β -globin protein comprises one or more alterations relative to the following reference sequence. In one particular embodiment, a β -globin protein associated with sickle cell disease comprises an E6V (also termed E7V) mutation.

By “HbB polynucleotide” is meant a nucleic acid molecule encoding β -globin protein or a fragment thereof. The sequence of an exemplary HbB polynucleotide, which is available at NCBI Accession No. NM_000518, is provided below:

```

1 acatttgctt ctgacacaac tgtgttcact agcaacctca aacagacacc atggtgcac
61 tgactcctga ggagaagtct gccgttactg cctgtggtgg caaggtgaac gtggatgaag
25 121 ttggtggtga ggccttgggc aggtctgtgg tggctaccc ttggaccag aggttctttg
181 agtccttttg ggatctgtcc actcctgatg ctgttatggg caaccttaag gtgaaggctc
241 atggcaagaa agtgcctggg gcccttagtg atggcctggc tcacctggac aacctcaagg
301 gcacctttgc cactctgagt gagctgcact gtgacaagct gcacgtggat cctgagaact
361 tcaggctcct gggcaacgtg ctggtctgtg tgcctggcca tcaactttgg aaagaattca
30 421 cccaccagtg gcaggtctgc tatcagaaag tgggtggtgg tgtggctaata gccctggccc
481 acaagtatca ctaagctgcg tttcttgctg tccaatttct attaaagggt cctttgttcc
541 ctaagtccaa ctactaaact ggggatatt atgaagggcc ttgagcatct ggattctgcc
601 taataaaaaa catttatttt cattgcaa (SEQ ID NO: 1)

```

An exemplary hemoglobin subunit beta polypeptide sequence is provided below:

VHLTPEEKSAVTALWGKVVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHG
KKVLGAFSDGLAHLDDLKGTFTATLSELHCDKLHVDPENFRLLGNVLCVLAHHFGKEFTTPVQA
AYQKVVAGVANALAHKYH (SEQ ID NO: 2).

By “adenosine deaminase” is meant a polypeptide or fragment thereof capable of
5 catalyzing the hydrolytic deamination of adenine or adenosine. In some embodiments, the
deaminase or deaminase domain is an adenosine deaminase catalyzing the hydrolytic
deamination of adenosine to inosine or deoxy adenosine to deoxyinosine. In some embodiments,
the adenosine deaminase catalyzes the hydrolytic deamination of adenine or adenosine in
deoxyribonucleic acid (DNA). The adenosine deaminases (e.g. engineered adenosine
10 deaminases, evolved adenosine deaminases) provided herein may be from any organism, such as
a bacterium.

By “Adenosine Deaminase Base Editor 8 (ABE8) polypeptide” or “ABE8” is meant a
base editor as defined herein comprising an adenosine deaminase variant comprising an
alteration at amino acid position 82 and/or 166 of the following reference sequence:
15 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHYPGMNRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID
NO: 3).

In some embodiments, ABE8 comprises further alterations, as described herein, relative
20 to the reference sequence.

By “Adenosine Deaminase Base Editor 8 (ABE8) polynucleotide” is meant a
polynucleotide encoding an ABE8.

“Administering” is referred to herein as providing one or more compositions described
herein to a patient or a subject.

25 By “agent” is meant any small molecule chemical compound, antibody, nucleic acid
molecule, or polypeptide, or fragments thereof.

By “alteration” is meant a change (increase or decrease) in the level, structure, or activity
of an analyte, gene or polypeptide as detected by standard art known methods such as those
described herein. As used herein, an alteration includes a 10% change in expression levels, a
30 25% change, a 40% change, and a 50% or greater change in expression levels. In some
embodiments, an alteration includes an insertion, deletion, or substitution of a nucleobase or
amino acid.

By “ameliorate” is meant decrease, suppress, attenuate, diminish, arrest, or stabilize the
development or progression of a disease, such as a hemoglobinopathy, sickle cell disease, or

thalassemia, which is an inherited blood disorder in which red blood cells contain less hemoglobin than normal, thus resulting in less oxygen being carried by the blood. Thalassemia can cause anemia.

By "analog" is meant a molecule that is not identical, but has analogous functional or structural features. For example, a polypeptide analog retains the biological activity of a corresponding naturally-occurring polypeptide, while having certain biochemical modifications that enhance the analog's function relative to a naturally occurring polypeptide. Such biochemical modifications could increase the analog's protease resistance, membrane permeability, or half-life, without altering, for example, ligand binding. An analog may include an unnatural amino acid.

By "base editor (BE)," or "nucleobase editor polypeptide (NBE)" is meant an agent that binds a polynucleotide and has nucleobase modifying activity. In various embodiments, the base editor comprises a nucleobase modifying polypeptide (*e.g.*, a deaminase) and a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9 or Cpf1) in conjunction with a guide polynucleotide (*e.g.*, guide RNA (gRNA)). Representative nucleic acid and protein sequences of base editors are provided in the Sequence Listing as SEQ ID NOs: 4-13.

By "base editing activity" is meant acting to chemically alter a base within a polynucleotide. In one embodiment, a first base is converted to a second base. In one embodiment, the base editing activity is cytidine deaminase activity, *e.g.*, converting target C•G to T•A. In another embodiment, the base editing activity is adenosine or adenine deaminase activity, *e.g.*, converting A•T to G•C.

The term "base editor system" refers to an intermolecular complex for editing a nucleobase of a target nucleotide sequence. In various embodiments, the base editor (BE) system comprises (1) a polynucleotide programmable nucleotide binding domain, a deaminase domain (*e.g.*, cytidine deaminase or adenosine deaminase) for deaminating nucleobases in the target nucleotide sequence; and (2) one or more guide polynucleotides (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. In various embodiments, the base editor (BE) system comprises a nucleobase editor domain selected from an adenosine deaminase or a cytidine deaminase, and a domain having nucleic acid sequence specific binding activity. In some embodiments, the base editor system comprises (1) a base editor (BE) comprising a polynucleotide programmable DNA binding domain and a deaminase domain for deaminating one or more nucleobases in a target nucleotide sequence; and (2) one or more guide RNAs in conjunction with the polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a

polynucleotide programmable DNA binding domain. In some embodiments, the base editor is a cytidine base editor (CBE). In some embodiments, the base editor is an adenine or adenosine base editor (ABE). In some embodiments, the base editor is an adenine or adenosine base editor (ABE) or a cytidine base editor (CBE).

5 By “base editing activity” is meant acting to chemically alter a base within a polynucleotide. In one embodiment, a first base is converted to a second base. In one embodiment, the base editing activity is cytidine deaminase activity, e.g., converting target C•G to T•A. In another embodiment, the base editing activity is adenosine deaminase activity, e.g., converting A•T to G•C.

10 The term “Cas9” or “Cas9 domain” refers to an RNA guided nuclease comprising a Cas9 protein, or a fragment thereof (e.g., a protein comprising an active, inactive, or partially active DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casnI nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat) associated nuclease.

15 The term “conservative amino acid substitution” or “conservative mutation” refers to the replacement of one amino acid by another amino acid with a common property. A functional way to define common properties between individual amino acids is to analyze the normalized frequencies of amino acid changes between corresponding proteins of homologous organisms (Schulz, G. E. and Schirmer, R. H., Principles of Protein Structure, Springer-Verlag, New York
20 (1979)). According to such analyses, groups of amino acids can be defined where amino acids within a group exchange preferentially with each other, and therefore resemble each other most in their impact on the overall protein structure (Schulz, G. E. and Schirmer, R. H., supra). Non-limiting examples of conservative mutations include amino acid substitutions of amino acids, for example, lysine for arginine and vice versa such that a positive charge can be maintained;
25 glutamic acid for aspartic acid and vice versa such that a negative charge can be maintained; serine for threonine such that a free –OH can be maintained; and glutamine for asparagine such that a free –NH₂ can be maintained.

The term “coding sequence” or “protein coding sequence” as used interchangeably herein refers to a segment of a polynucleotide that codes for a protein. Coding sequences can also be
30 referred to as open reading frames. The region or sequence is bounded nearer the 5' end by a start codon and nearer the 3' end with a stop codon. Stop codons useful with the base editors described herein include the following:

Glutamine CAG → TAG Stop codon
CAA → TAA

Arginine CGA → TGA
 Tryptophan TGG → TGA
 TGG → TAG
 TGG → TAA

5 As used herein, the terms "condition" and "conditioning" refer to processes by which a patient is prepared for receipt of a transplant containing hematopoietic stem cells. Such procedures promote the engraftment of a hematopoietic stem cell transplant (for instance, as inferred from a sustained increase in the quantity of viable hematopoietic stem cells within a blood sample isolated from a patient following a conditioning procedure and subsequent
 10 hematopoietic stem cell transplantation). According to the methods described herein, a patient may be conditioned for hematopoietic stem cell transplant therapy by administration to the patient of an antibody or antigen-binding fragment thereof capable of binding an antigen expressed by hematopoietic stem cells, such as CD117, CXCR4, CD135, CD90, CD45, and/or CD34. Such antibodies are expected to act via complement-mediated cytotoxicity and antibody-
 15 dependent cell-mediated cytotoxicity. As described herein, the transplanted cells have been edited so that the antibody no longer binds the antigen (e.g., CD117, CXCR4, CD135, CD90, CD45, and/or CD34). Administration of an antibody, antigen-binding fragment thereof, drug-antibody conjugate, or chimeric antigen receptor expressing T-cell (CAR-T) capable of binding one or more antigens (e.g., CD117, CXCR4, CD135, CD90, CD45, CD34) to a patient in need of
 20 hematopoietic stem cell transplant therapy can promote the engraftment of a hematopoietic stem cell graft, for example, by selectively depleting endogenous hematopoietic stem cells, thereby creating a vacancy filled by an exogenous hematopoietic stem cell transplant.

By "cytidine deaminase" is meant a polypeptide or fragment thereof capable of catalyzing a deamination reaction that converts an amino group to a carbonyl group. In one
 25 embodiment, the cytidine deaminase converts cytosine to uracil or 5-methylcytosine to thymine. PmCDA1 (SEQ ID NO: 14 and 15), which is derived from *Petromyzon marinus* (*Petromyzon marinus* cytosine deaminase 1, "PmCDA1"), AID (Activation-induced cytidine deaminase; AICDA). Exemplary AID polypeptide sequences are provided in the Sequence Listing as SEQ ID NOs: 16-28 and 20-23, which are derived from a mammal (e.g., human, swine, bovine, horse,
 30 monkey etc.). Exemplary APOBEC cytidine deaminase polypeptide sequences are provided in the Sequence Listing as SEQ ID NOs: 24-64. Additional exemplary cytidine deaminase (CDA) sequences are provided in the Sequence Listing as SEQ ID NOs: 19 and 65-68. Other exemplary cytidine deaminase sequences, including APOBEC polypeptide sequences, are provided in the Sequence Listing as SEQ ID NOs: 291-413.

The term “deaminase” or “deaminase domain,” as used herein, refers to a protein or enzyme that catalyzes a deamination reaction.

“Detect” refers to identifying the presence, absence or amount of the analyte to be detected. In one embodiment, a sequence alteration in a polynucleotide or polypeptide is detected. In another embodiment, the presence of indels is detected.

By “detectable label” is meant a composition that when linked to a molecule of interest renders the latter detectable, via spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes, magnetic beads, metallic beads, colloidal particles, fluorescent dyes, electron-dense reagents, enzymes (for example, as commonly used in an enzyme linked immunosorbent assay (ELISA)), biotin, digoxigenin, or haptens.

By “disease” is meant any condition or disorder that damages or interferes with the normal function of a cell, tissue, or organ. Exemplary diseases include hemoglobinopathies (e.g., sickle cell disease).

By “effective amount” is meant the amount of an agent or active compound, e.g., a base editor as described herein, that is required to ameliorate the symptoms of a disease relative to an untreated patient or an individual without disease, i.e., a healthy individual, or is the amount of the agent or active compound sufficient to elicit a desired biological response. The effective amount of active compound(s) used to practice the present invention for therapeutic treatment of a disease varies depending upon the manner of administration, the age, body weight, and general health of the subject. Ultimately, the attending physician or veterinarian will decide the appropriate amount and dosage regimen. Such amount is referred to as an “effective” amount. In one embodiment, an effective amount is the amount of a base editor of the invention sufficient to introduce an alteration in a gene of interest in a cell (e.g., a cell *in vitro* or *in vivo*). In one embodiment, an effective amount is the amount of a base editor required to achieve a therapeutic effect. Such therapeutic effect need not be sufficient to alter a pathogenic gene in all cells of a subject, tissue or organ, but only to alter the pathogenic gene in about 1%, 5%, 10%, 25%, 50%, 75% or more of the cells present in a subject, tissue or organ. In one embodiment, an effective amount is sufficient to ameliorate one or more symptoms of a disease.

The term “exonuclease” refers to a protein or polypeptide capable of digesting a nucleic acid (e.g., RNA or DNA) from free ends.

The term “endonuclease” refers to a protein or polypeptide capable of catalyzing (e.g., cleaving) internal regions in a nucleic acid (e.g., DNA or RNA).

By "fragment" is meant a portion of a polypeptide or nucleic acid molecule. This portion contains, at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the entire length of the reference nucleic acid molecule or polypeptide. A fragment may contain 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or amino acids.

By "guide RNA" or "gRNA" is meant a polynucleotide or polynucleotide complex which is specific for a target sequence and can form a complex with a polynucleotide programmable nucleotide binding domain protein (*e.g.*, Cas9 or Cpf1). In an embodiment, the guide polynucleotide is a guide RNA (gRNA). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule.

As used herein, the term "hematopoietic stem cells" ("HSCs") refers to immature blood cells having the multipotential capacity to self-renew and to differentiate into mature blood cells containing diverse lineages, including, but not limited to, granulocytes (*e.g.*, promyelocytes, neutrophils, eosinophils, basophils), erythrocytes (*e.g.*, reticulocytes, erythrocytes), thrombocytes (*e.g.*, megakaryoblasts, platelet producing megakaryocytes, platelets), monocytes (*e.g.*, monocytes, macrophages), dendritic cells, microglia, osteoclasts, and lymphocytes (*e.g.*, NK cells, B-cells and T-cells). Such cells may include CD34⁺ cells, which are immature cells (or HSCs) that express the CD34 cell surface marker. CD34 is a marker of human HSCs, and the colony-forming activity of human bone marrow (BM) cells is found in the CD34⁺ fraction. In humans, CD34⁺ cells are believed to include a subpopulation of cells with the stem cell properties defined above, whereas in mice, HSCs are CD34⁺. In an embodiment, transplantation studies using enriched CD34⁺ BM cells indicated the presence of HSCs with long-term BM reconstitutive ability within this fraction. In addition, HSCs also refer to long term repopulating HSCs (LT-HSC) and short term repopulating HSCs (ST-HSC). LT-HSCs and ST-HSCs are differentiated, based on functional potential and on cell surface marker expression. For example, human HSCs are CD34⁺, CD38⁻, CD45RA⁻, CD90⁺, CD49F⁺, and lin⁻ (negative for mature lineage markers including CD2, CD3, CD4, CD7, CD8, CD10, CD11b, CD19, CD20, CD56, CD235A). In mice, bone marrow LT-HSCs are CD34⁺, SCA-1⁺, C-kit⁺, CD135⁻, Slamf1/CD150⁺, CD48⁻, and lin⁻ (negative for mature lineage markers including Ter119, CD11b, Gr1, CD3, CD4, CD8, B220, IL7ra), whereas ST-HSCs are CD34⁺, SCA-1⁺, C-kit⁺, CD135⁻, Slamf1/CD150⁺, and lin⁻ (negative for mature lineage markers including Ter119, CD11b, Gr1, CD3, CD4, CD8, B220, IL7ra). In addition, ST-HSCs are less quiescent and more proliferative than LT-HSCs under homeostatic conditions. However, LT-HSC have greater self-renewal potential (*i.e.*, they survive throughout adulthood, and can be serially transplanted through

successive recipients), whereas ST-HSCs have limited self-renewal (i.e., they survive for only a limited period of time, and do not possess serial transplantation potential). Any of these HSCs can be used in the methods described herein. ST-HSCs are particularly useful because they are highly proliferative and thus, can more quickly give rise to differentiated progeny.

5 As used herein, the term "hematopoietic stem cell functional potential" refers to the functional properties of hematopoietic stem cells which include 1) multi-potency (which refers to the ability to differentiate into multiple different blood lineages including, but not limited to, granulocytes (*e.g.*, promyelocytes, neutrophils, eosinophils, basophils), erythrocytes (*e.g.*, reticulocytes, erythrocytes), thrombocytes (*e.g.*, megakaryoblasts, platelet producing
10 megakaryocytes, platelets), monocytes (*e.g.*, monocytes, macrophages), dendritic cells, microglia, osteoclasts, and lymphocytes (*e.g.*, NK cells, B-cells and T-cells); 2) self-renewal (which refers to the ability of hematopoietic stem cells to give rise to daughter cells that have equivalent potential as the mother cell, and further that this ability can repeatedly occur throughout the lifetime of an individual without exhaustion); and 3) the ability of hematopoietic
15 stem cells or progeny thereof to be reintroduced into a transplant recipient whereupon they home to the hematopoietic stem cell niche and re-establish productive and sustained hematopoiesis."

"Hybridization" means hydrogen bonding, which may be Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary nucleobases. For example, adenine and thymine are complementary nucleobases that pair through the formation of
20 hydrogen bonds.

By "increases" is meant a positive alteration of at least 5%, 10%, 25%, 50%, 75%, or 100%. Percentages between these values are encompassed by the term.

The terms "inhibitor of base repair", "base repair inhibitor", "IBR" or their grammatical equivalents refer to a protein that is capable in inhibiting the activity of a nucleic acid repair
25 enzyme, for example a base excision repair enzyme.

An "intein" is a fragment of a protein that is able to excise itself and join the remaining fragments (the exteins) with a peptide bond in a process known as protein splicing.

The terms "isolated," "purified," or "biologically pure" refer to material that is free to varying degrees from components which normally accompany it as found in its native state.

30 "Isolate" denotes a degree of separation from original source or surroundings. "Purify" denotes a degree of separation that is higher than isolation. A "purified" or "biologically pure" protein is sufficiently free of other materials such that any impurities do not materially affect the biological properties of the protein or cause other adverse consequences. That is, a nucleic acid or peptide of this invention is purified if it is substantially free of cellular material, viral material, or culture

medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Purity and homogeneity are typically determined using analytical chemistry techniques, for example, polyacrylamide gel electrophoresis or high performance liquid chromatography. The term "purified" can denote that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. For a protein that can be subjected to modifications, for example, phosphorylation or glycosylation, different modifications may give rise to different isolated proteins, which can be separately purified.

By "isolated polynucleotide" is meant a nucleic acid (*e.g.*, a DNA) that is free of the genes which, in the naturally-occurring genome of the organism from which the nucleic acid molecule of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. In addition, the term includes an RNA molecule that is transcribed from a DNA molecule, as well as a recombinant DNA that is part of a hybrid gene encoding additional polypeptide sequence.

By an "isolated polypeptide" is meant a polypeptide of the invention that has been separated from components that naturally accompany it. Typically, the polypeptide is isolated when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, a polypeptide of the invention. An isolated polypeptide of the invention may be obtained, for example, by extraction from a natural source, by expression of a recombinant nucleic acid encoding such a polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

By "CD117 (C-kit; SCFR) polypeptide" is meant a polypeptide or fragment thereof having at least about 95% amino acid sequence identity to an amino acid sequence provided at GenBank Accession No. NP_000213 that binds an anti-CD117 antibody. In some embodiments, an CD117 polypeptide or fragment thereof has SCF signaling activity. An exemplary CD117 polypeptide sequence follows:

>NP_000213.1 mast/stem cell growth factor receptor Kit isoform 1 precursor [Homo sapiens]
MRGARGAWDFLCVLLLLLRVQTGSSQPSVSPGEPSPPSIHPGKSDLIVRVGDEIRLLCTDP
GFVKWTFEILDETENKQNEWITEKAEATNTGKYTCTNKHGLSNSIYVFVRDPAKLFLV

DRSLYGKEDNDTLVRCPLTDPEVTNYSKGCQGKPLPKDLRFIPDPKAGIMIKSVKRAY
 HRLCLHCSVDQEGKSVLSEKFIKVRPAFKAVPVVSVSKASYLLREGEEFTVTCTIKDVS
 SSVYSTWKRENSQTKLQEKYNSWHHGDFNYERQATLTISSARVNDSGVFMCIYANNTFG
 SANVTTLLEVVDKGFNIFPMINTTVFVNDGENVDLIVEYEAFPKPEHQQWIYMNRTFTD
 5 KWEDYPKSENESENIRYVSELHLTRLKGTEGGTYTFLVSNSDVNAIAFNVYVNTKPEILT
 YDRLVNGMLQCVAAGFPEPTIDWYFCPGTEQRCSASVLPVDVQTLNSSGPPFGKLVVQS
 SIDSSAFKHNGTVECKAYNDVGKTSAYFNFAFKGNNKEQIHPHTLFTPLLIGFVIVAGM
 MCIIVMILTYKYQLQKPMYEVQWKVVEEINGNNYVYIDPTQLPYDHWKWEFPRNRLSFGK
 TLGAGAFGKVVEATAYGLIKSDAAMTVAVKMLKPSAHLTEREALMSELKVL SYLGNH
 10 MNIVNLLGACTIGGPTLVITEYCCYGDLLNFLRRKRDSFICSKQEDHAEAAKYKNLLHSK
 ESSCSDSTNEYMDMKPGVSYVVPKADKRRSVRIGSYIERDVTPAIMEDELALDLEDL
 LSFSYQVAKGMAFLASKNCIHRDLAARNILLTHGRITKICDFGLARDIKNDSNYVVKGN
 ARLPVKWMAPESIFNCVYTFESDVWSYGIFLWELFSLGSSPYGMPVDSKFYKMIKEGF
 RMLSPEHAPAEMYDIMKTCWDADPLKRPTFKQIVQLIEKQISESTNHIYSNLANCSPNRQ
 15 KPVVDHSVRINSVGSTASSSQPLL VHDDV (SEQ ID NO: 69).

By "CD117 polynucleotide" is meant a nucleic acid molecule that encodes a CD117 polypeptide. An exemplary CD117 polynucleotide sequence follows:

>NM_000222.2 Homo sapiens KIT proto-oncogene, receptor tyrosine kinase (KIT), transcript variant 1, mRNA

20 TCTGGGGGCTCGGCTTTGCCGCGCTCGCTGCACTTGGGCGAGAGCTGGAACGTGGAC
 CAGAGCTCGGATCCCATCGCAGCTACCGCGATGAGAGGCGCTCGCGGCGCCTGGGA
 TTTTCTCTGCGTTCTGCTCCTACTGCTTCGCGTCCAGACAGGCTCTTCTCAACCATCT
 GTGAGTCCAGGGGAACCGTCTCCACCATCCATCCATCCAGGAAAATCAGACTTAAT
 AGTCCGCGTGGGCGACGAGATTAGGCTGTTATGCACTGATCCGGGCTTTGTCAAATG
 25 GACTTTTGAGATCCTGGATGAAACGAATGAGAATAAGCAGAATGAATGGATCACGG
 AAAAGGCAGAAGCCACCAACACCGGCAAATACACGTGCACCAACAAACACGGCTT
 AAGCAATTCCATTTATGTGTTTGTAGAGATCCTGCCAAGCTTTTCCTTGTTGACCGC
 TCCTTGATGGGAAAGAAGACAACGACACGCTGGTCCGCTGTCCTCTCACAGACCCA
 GAAGTGACCAATTATCCCTCAAGGGGTGCCAGGGGAAGCCTCTTCCCAAGGACTT
 30 GAGGTTTATTCCTGACCCCAAGGCGGGCATCATGATCAAAAGTGTGAAACGCGCCT
 ACCATCGGCTCTGTCTGCATTGTTCTGTGGACCAGGAGGGCAAGTCAGTGCTGTGCG
 AAAAATTCATCCTGAAAGTGAGGCCAGCCTTCAAAGCTGTGCCTGTTGTGTCTGTGT
 CCAAAGCAAGCTATCTTCTTAGGGGAAGGGGAAGAATTCACAGTGACGTGCACAATA
 AAAGATGTGTCTAGTTCTGTGTACTCAACGTGGAAAAGAGAAAACAGTCAGACTAA

ACTACAGGAGAAATATAATAGCTGGCATCACGGTGACTTCAATTATGAACGTCAGG
CAACGTTGACTATCAGTTCAGCGAGAGTTAATGATTCTGGAGTGTTTCATGTGTTATG
CCAATAATACTTTTGGATCAGCAAATGTCACAACAACCTTGGAAGTAGTAGATAAA
GGATTCATTAATATCTTCCCCATGATAAAACACTACAGTATTTGTAAACGATGGAGAA
5 AATGTAGATTTGATTGTTGAATATGAAGCATTCCCCAAACCTGAACACCAGCAGTGG
ATCTATATGAACAGAACCTTCACTGATAAATGGGAAGATTATCCCAAGTCTGAGAAT
GAAAGTAATATCAGATACGTAAGTGAACCTTCATCTAACGAGATTAAAAGGCACCGA
AGGAGGCACTTACACATTCCTAGTGTCCAATTCTGACGTCAATGCTGCCATAGCATT
TAATGTTTATGTGAATACAAAACCAGAAATCCTGACTTACGACAGGCTCGTGAATGG
10 CATGCTCCAATGTGTGGCAGCAGGATTCCCAGAGCCCACAATAGATTGGTATTTTTG
TCCAGGAACTGAGCAGAGATGCTCTGCTTCTGTACTGCCAGTGGATGTGCAGACACT
AAACTCATCTGGGCCACCGTTTGGAAAGCTAGTGGTTCAGAGTTCTATAGATTCTAG
TGCATTCAAGCACAATGGCACGGTTGAATGTAAGGCTTACAACGATGTGGGCAAGA
CTTCTGCCTATTTTAACTTTGCATTTAAAGGTAACAACAAGAGCAAATCCATCCCC
15 ACACCCTGTTCACTCCTTTGCTGATTGGTTTCGTAATCGTAGCTGGCATGATGTGCAT
TATTGTGATGATTCTGACCTACAAATATTTACAGAAACCCATGTATGAAGTACAGTG
GAAGGTTGTTGAGGAGATAAATGGAAACAATTATGTTTACATAGACCCAACACAAC
TTCCTTATGATCACAAATGGGAGTTTCCCAGAAACAGGCTGAGTTTTGGGAAAACCC
TGGGTGCTGGAGCTTTCGGGAAGGTTGTTGAGGCAACTGCTTATGGCTTAATTAAGT
20 CAGATGCGGCCATGACTGTCGCTGTAAAGATGCTCAAGCCGAGTGCCCATTTGACA
GAACGGGAAGCCCTCATGTCTGAACTCAAAGTCCTGAGTTACCTTGGTAATCACATG
AATATTGTGAATCTACTTGGAGCCTGCACCATTGGAGGGCCCCACCCTGGTCATTACA
GAATATTGTTGCTATGGTGATCTTTTGAATTTTTTGAGAAGAAAACGTGATTCATTTA
TTTGTTCAAAGCAGGAAGATCATGCAGAAGCTGCACTTTATAAGAATCTTCTGCATT
25 CAAAGGAGTCTTCCTGCAGCGATAGTACTAATGAGTACATGGACATGAAACCTGGA
GTTTCTTATGTTGTCCCAACCAAGGCCGACAAAAGGAGATCTGTGAGAATAGGCTCA
TACATAGAAAGAGATGTGACTCCCGCCATCATGGAGGATGACGAGTTGGCCCTAGA
CTTAGAAGACTTGCTGAGCTTTTCTTACCAGGTGGCAAAGGGCATGGCTTTCCTCGC
CTCCAAGAATTGTATTCACAGAGACTTGGCAGCCAGAAATATCCTCCTTACTCATGG
30 TCGGATCACAAAGATTTGTGATTTTGGTCTAGCCAGAGACATCAAGAATGATTCTAA
TTATGTGGTTAAAGGAAACGCTCGACTACCTGTGAAGTGGATGGCACCTGAAAGCA
TTTTCAACTGTGTATACACGTTTGAAAGTGACGTCTGGTCCTATGGGATTTTTCTTTG
GGAGCTGTTCTCTTTAGGAAGCAGCCCCTATCCTGGAATGCCGGTCGATTCTAAGTT
CTACAAGATGATCAAGGAAGGCTTCCGGATGCTCAGCCCTGAACACGCACCTGCTG

AAATGTATGACATAATGAAGACTTGCTGGGATGCAGATCCCCATAAAAAGACCAACA
TTCAAGCAAATTGTTTCAGCTAATTGAGAAGCAGATTTTCAGAGAGCACCAATCATATT
TACTCCAACTTAGCAAACTGCAGCCCCAACCGACAGAAGCCCGTGGTAGACCATTCT
GTGCGGATCAATTCTGTCGGCAGCACCGCTTCCTCCTCCCAGCCTCTGCTTGTGCAC
5 GACGATGTCTGAGCAGAATCAGTGTGTTGGGTCACCCCTCCAGGAATGATCTCTTCTT
TTGGCTTCCATGATGGTTATTTTCTTTTCTTTCAACTTGCATCCAACTCCAGGATAGT
GGGCACCCCACTGCAATCCTGTCTTTCTGAGCACACTTTAGTGGCCGATGATTTTTGT
CATCAGCCACCATCCTATTGCAAAGGTTCCAAGTGTATATATTTCCCAATAGCAACGT
AGCTTCTACCATGAACAGAAAACATTCTGATTTGGAAAAAGAGAGGGAGGTATGGA
10 CTGGGGGGCCAGAGTCCTTTCCAAGGCTTCTCCAATTCTGCCCAAAAATATGGTTGAT
AGTTTACCTGAATAAATGGTAGTAATCACAGTTGGCCTTCAGAACCATCCATAGTAG
TATGATGATACAAGATTAGAAGCTGAAAAACCTAAGTCCTTTATGTGGAAAAACAGAA
CATCATTAGAACAAAGGACAGAGTATGAACACCTGGGCTTAAGAAATCTAGTATTT
CATGCTGGGAATGAGACATAGGCCATGAAAAAAATGATCCCCAAGTGTGAACAAAA
15 GATGCTCTTCTGTGGACCACTGCATGAGCTTTTATACTACCGACCTGGTTTTTTAAATA
GAGTTTGCTATTAGAGCATTGAATTGGAGAGAAGGCCTCCCTAGCCAGCACTTGTAT
ATACGCATCTATAAATTGTCCGTGTTTCATACATTTGAGGGGAAAAACACCATAAGGTT
TCGTTTCTGTATACAACCCTGGCATTATGTCCACTGTGTATAGAAGTAGATTAAGAG
CCATATAAGTTTGAAGGAAACAGTTAATACCATTTTTTAAGGAAACAATATAACCAC
20 AAAGCACAGTTTGAACAAAATCTCCTCTTTTAGCTGATGAACCTTATTCTGTAGATTCT
GTGGAACAAGCCTATCAGCTTCAGAATGGCATTGTACTCAATGGATTTGATGCTGTT
TGACAAAGTTACTGATTCACTGCATGGCTCCACAGGAGTGGGAAAAACACTGCCAT
CTTAGTTTGGATTCTTATGTAGCAGGAAATAAAGTATAGGTTTAGCCTCCTTCGCAG
GCATGTCCTGGACACCGGGCCAGTATCTATATATGTGTATGTACGTTTGTATGTGTG
25 TAGACAAATATTTGGAGGGGTATTTTTGCCCTGAGTCCAAGAGGGTCCTTTAGTACC
TGAAAAGTAACTTGGCTTTCATTATTAGTACTGCTCTTGTTTCTTTTCACATAGCTGT
CTAGAGTAGCTTACCAGAAGCTTCCATAGTGGTGCAGAGGAAGTGAAGGCATCAG
TCCCTATGTATTTGCAGTTCACCTGCACTTAAGGCACTCTGTTATTTAGACTCATCTT
ACTGTACCTGTTTCCTTAGACCTTCCATAATGCTACTGTCTCACTGAAACATTTAAATT
30 TTACCCCTTTAGACTGTAGCCTGGATATTATTCTTGTTAGTTTACCTCTTTAAAAACAAA
ACAAAACAAAACAAAAAACTCCCTTCCTCACTGCCCAATATAAAAGGCAAATGTG
TACATGGCAGAGTTTGTGTGTTGTCTTGAAAGATTCAGGTATGTTGCCTTTATGGTTT
CCCCCTTCTACATTTCTTAGACTACATTTAGAGAACTGTGGCCGTTATCTGGAAGTA
ACCATTTGCACTGGAGTTCTATGCTCTCGCACCTTTCCAAAGTTAACAGATTTTGGG

GTTGTGTTGTCACCCAAGAGATTGTTGTTTGGCATACTTTGTCTGAAAAATTCCTTTG
 TGTTTCTATTGACTTCAATGATAGTAAGAAAAGTGGTTGTTAGTTATAGATGTCTAG
 GTA CTTCAGGGGCACTTCATTGAGAGTTTTGTCTTGGATATTCTTGAAAGTTTATATT
 TTTATAATTTTTTCTTACATCAGATGTTTCTTTGCAGTGGCTTAATGTTTGAAATTATT
 5 TTGTGGCTTTTTTTGTAAATATTGAAATGTAGCAATAATGTCTTTTGAATATTCCCAA
 GCCCATGAGTCCTTGAAAATATTTTTTATATATACAGTAACTTTATGTGTAAATACAT
 AAGCGGCGTAAGTTTAAAGGATGTTGGTGTTCACGTGTTTTATTCTGTATGTTGTC
 CAATTGTTGACAGTTCTGAAGAATTCTAATAAAATGTACATATATAAATCAAAAAA
 AAAAAAAA (SEQ ID NO: 70).

10 By “C-X-C chemokine receptor type 4 (CXCR4) polypeptide” is meant a polypeptide or
 fragment thereof having at least about 95% amino acid sequence identity to an amino acid
 sequence provided at GenBank Accession NP_001008540 that binds an anti- CXCR4 antibody.
 An exemplary CXCR4 polypeptide sequence follows:

>NP_001008540.1 C-X-C chemokine receptor type 4 isoform a [Homo sapiens]
 15 MSIPLPLLQIYTS DNYTEEMGSGDYDSMKEPCFREENANFNKIFLPTIYSIIFLTGIVGNGL
 VILVMGYQKKLRSM TDKYRLHLSVADLLFVITLPFWAVDAVANWYFGNFLCKAVHVI
 YTVNLYSSVLILAFISLDRYLAIVHATNSQRPRKLLAEKV VYVGVWIPALLLTIPDFIFAN
 VSEADDRYICDRFYPNDLWVVVFQFQHIMVGLILPGIVILSCYCHISKLSHSGKHQKRKA
 LKTTVILILAFFACWLPYYIGISIDSFILLEIKQGCEFENTVHKWISITEALAFFHCCLNPIL
 20 YAFLGAKFKTSAQHALTSVSRGSSLKILSKGKRGGHSSVSTESESSSFHSS (SEQ ID NO:
 71).

By “CXCR4 polynucleotide” is meant a nucleic acid molecule that encodes a CXCR4
 polypeptide. An exemplary CXCR4 polynucleotide sequence follows:

>NM_003467.2 Homo sapiens C-X-C motif chemokine receptor 4 (CXCR4), transcript
 25 variant 2, mRNA
 AACTTCAGTTTGTG GCTGCGGCAGCAGGTAGCAAAGTGACGCCGAGGGCCTGAGT
 GCTCCAGTAGCCACCGCATCTGGAGAACCAGCGGTTACCATGGAGGGGATCAGTAT
 ATACACTTCAGATAACTACACCGAGGAAATGGGCTCAGGGGACTATGACTCCATGA
 AGGAACCCTGTTTCCGTGAAGAAAATGCTAATTTCAATAAAATCTTCCTGCCACCA
 30 TCTACTCCATCATCTTCTTA ACTGGCATTGTGGGCAATGGATTGGTCATCCTGGTCAT
 GGGTTACCAGAAGAACTGAGAAGCATGACGGACAAGTACAGGCTGCACCTGTCAG
 TGGCCGACCTCCTCTTTGTCATCACGCTTCCCTTCTGGGCAGTTGATGCCGTGGCAA
 CTGGTACTTTGGGA ACTTCCTATGCAAGGCAGTCCATGTCATCTACACAGTCAACCT
 CTACAGCAGTGTCTCATCTGGCCTTCATCAGTCTGGACCGCTACCTGGCCATCGT

CCACGCCACCAACAGTCAGAGGCCAAGGAAGCTGTTGGCTGAAAAGGTGGTCTATG
 TTGGCGTCTGGATCCCTGCCCTCCTGCTGACTATCCCCGACTTCATCTTTGCCAACGT
 CAGTGAGGCAGATGACAGATATATCTGTGACCGCTTCTACCCCAATGACTTGTGGGT
 GGTTGTGTTCCAGTTTCAGCACATCATGGTTGGCCTTATCCTGCCTGGTATTGTCATC
 5 CTGTCCTGCTATTGCATTATCATCTCCAAGCTGTCACACTCCAAGGGCCACCAGAAG
 CGCAAGGCCCTCAAGACCACAGTCATCCTCATCCTGGCTTTCTTCGCCTGTTGGCTG
 CCTTACTACATTGGGATCAGCATCGACTCCTTCATCCTCCTGGAAATCATCAAGCAA
 GGGTGTGAGTTTGAGAACACTGTGCACAAGTGGATTTCCATCACCGAGGCCCTAGCT
 TTCTTCCACTGTTGTCTGAACCCCATCCTCTATGCTTTCCTTGGAGCCAAATTTAAAA
 10 CCTCTGCCCAGCACGCACTCACCTCTGTGAGCAGAGGGTCCAGCCTCAAGATCCTCT
 CCAAAGGAAAGCGAGGTGGACATTCATCTGTTTCCACTGAGTCTGAGTCTTCAAGTT
 TTCCTCCAGCTAACACAGATGTAAAAGACTTTTTTTTATACGATAAAATAACTTTTTT
 TTAAGTTACACATTTTTTCAGATATAAAAAGACTGACCAATATTGTACAGTTTTTATTGC
 TTGTTGGATTTTTGTCTTGTGTTTCTTTAGTTTTTGTGAAGTTTAATTGACTTATTTAT
 15 ATAAATTTTTTTTGTTCATATTGATGTGTGTCTAGGCAGGACCTGTGGCCAAGTTCT
 TAGTTGCTGTATGTCTCGTGGTAGGACTGTAGAAAAGGGAAGTGAACATTCCAGAG
 CGTGTAGTGAATCACGTAAAGCTAGAAATGATCCCCAGCTGTTTATGCATAGATAAT
 CTCTCCATTCCCGTGGAAACGTTTTTCTGTTCTTAAGACGTGATTTTGCTGTAGAAGA
 TGGCACTTATAACCAAAGCCCAAAGTGGTATAGAAATGCTGGTTTTTTCAGTTTTTCAG
 20 GAGTGGGTTGATTTTCAGCACCTACAGTGTACAGTCTTGTATTAAGTTGTAAATAAAA
 GTACATGTAAACTTAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 72).

By "CD135 polypeptide" is meant a polypeptide or fragment thereof having at least about
 95% amino acid sequence identity to an amino acid sequence provided at GenBank Accession
 No. NP__004110 that binds an anti-CD135 antibody. An exemplary CD135 polypeptide sequence
 25 follows:

>NP__004110.2 receptor-type tyrosine-protein kinase FLT3 precursor [Homo sapiens]
 MPALARDGGQLPLLVFVSAMIFGTITNQDLPIKCVLINHKNNDSVVGKSSSYPMVSESP
 EDLGCALRPQSSGTVYEAAAVEVDVSASITLQVLVDAPGNISCLWVFKHSSLNCQPHFD
 LQNRGVVSMVILKMTETQAGEYLLFIQSEATNYTILFTVSIRNTLLYTLRRPYFRKMENQ
 30 DALVCISESVPEPIVEWVLCDSQGESCKEESPAVVKKEEKVLHELFGTDIRCCARNELGR
 ECTRLFTIDLNQTPQTTLPLQLFLKVGEPWIRCKAVHVNHGFGTLWELENKALEEGNYF
 EMSTYSTNRTMIRILFAFVSSVARNDTGYTCCSSKHPSQSALVTIVEKGFINATNSSDY
 EIDQYEEFCFSVRFKAYPQIRCTWTFSRKSFPCEQKGLDNGYSISKFCNKHQPGYIFHA
 ENDDAQFTKMFTLNIRRKPVLAEEASASQASCFSDGYPLPSWTWKKCSDKSPNCTEEIT

EGVWNRKANRKVFGQWVSSSTLNMSEAIKGFLVKCCAYNSLGTSCETILLNSPGPFPIQ
 DNISFYATIGVCLLFIVVLTLICHKYKKQFRYESQLQMVQVTGSSDNEYFYVDFREY
 DLKWEFPRENLEFGKVLGSGAFGKVMNATAYGISKTGVSIQVAVKMLKEKADSSEREA
 LMSELKMMTQLGSHENIVNLLGACTLSGPIYLIFEYCCYGDLLNYLRSKREKFHRTWTEI
 5 FKEHNFSFYPTFQSHPNSSMPGSREVQIHPDSDQISGLHGNSFHSEDEIEYENQKRLEEEE
 DLNVLTFEDLLCFAYQVAKGMEFLEFKSCVHRDLAARNVLVTHGKVVKICDFGLARDI
 MSDSNYVVRGNARLPVKWMAPESLFEGIYTIKSDVWSYGILLWEIFSLGVNPYPGIPVD
 ANFYKLIQNGFKMDQPFYATEEIIYIMQSCWAFDSRKRPSPNLTSLGCGQLADAEEM
 YQNVDGRVSECPHTYQNRPFPSREMDLGLLSPQAQVEDS (SEQ ID NO: 73).

10 By “CD135 polynucleotide” is meant a nucleic acid molecule that encodes a CD135
 polypeptide. An exemplary CD135 polynucleotide sequence follows:

>NM_004119.2 Homo sapiens fms related tyrosine kinase 3 (FLT3), transcript variant 1,
 mRNA
 ACCTGCAGCGCGAGGCGCGCCGCTCCAGGCGGCATCGCAGGGCTGGGCCGGCGCGG
 15 CCTGGGGACCCCGGGCTCCGGAGGCCATGCCGGCGTTGGCGCGCGACGGCGGCCAG
 CTGCCGCTGCTCGTTGTTTTTCTGCAATGATATTTGGGACTATTACAAATCAAGATC
 TGCCTGTGATCAAGTGTGTTTTAATCAATCATAAGAACAATGATTCATCAGTGGGGA
 AGTCATCATCATATCCCATGGTATCAGAATCCCCGGAAGACCTCGGGTGTGCGTTGA
 GACCCAGAGCTCAGGGACAGTGTACGAAGCTGCCGCTGTGGAAGTGGATGTATCT
 20 GCTTCCATCACACTGCAAGTGCTGGTCGACGCCCCAGGGAACATTTCTGTCTCTGG
 GTCTTTAAGCACAGCTCCCTGAATTGCCAGCCACATTTTGATTTACAAAACAGAGGA
 GTTGTTCATGCTCATTTTGAAGTACAGAAACCCCAAGCTGGAGAATACCTACTT
 TTTATTCAGAGTGAAGCTACCAATTACACAATATTGTTTACAGTGAGTATAAGAAAT
 ACCCTGCTTTACACATTAAGAAGACCTTACTTTAGAAAAATGGAAAACAGGACGC
 25 CCTGGTCTGCATATCTGAGAGCGTTCCAGAGCCGATCGTGGAATGGGTGCTTTGCGA
 TTCACAGGGGGAAAGCTGTAAAGAAGAAAGTCCAGCTGTTGTTAAAAAGGAGGAA
 AAAGTGCTTCATGAATTATTTGGGACGGACATAAGGTGCTGTGCCAGAAATGAACT
 GGGCAGGGAATGCACCAGGCTGTTACAATAGATCTAAATCAAACCTCCTCAGACCA
 CATTGCCACAATTATTCTTAAAGTAGGGGAACCCCTATGGATAAGGTGCAAAGCTG
 30 TTCATGTGAACCATGGATTCTGGGCTCACCTGGGAATTAGAAAACAAAGCACTCGAG
 GAGGGCAACTACTTTGAGATGAGTACCTATTCAACAAACAGAACTATGATACGGAT
 TCTGTTTGCTTTTGTATCATCAGTGGCAAGAAACGACACCGGATACTACACTTGTTT
 CTCTTCAAAGCATCCCAGTCAATCAGCTTTGGTTACCATCGTAGAAAAGGGATTTAT
 AAATGCTACCAATTCAAGTGAAGATTATGAAATTGACCAATATGAAGAGTTTGT

TTCTGTCAGGTTTAAAGCCTACCCACAAATCAGATGTACGTGGACCTTCTCTCGAAA
ATCATTTTCCTTGTGAGCAAAAGGGTCTTGATAACGGATACAGCATATCCAAGTTTTCG
CAATCATAAGCACCAGCCAGGAGAATATATATTCCATGCAGAAAATGATGATGCCC
AATTTACCAAAAATGTTACGCTGAATATAAGAAGGAAACCTCAAGTGCTCGCAGAA
5 GCATCGGCAAGTCAGGCGTCCTGTTTCTCGGATGGATACCCATTACCATCTTGGACC
TGGAAGAAGTGTTTACAGACAAGTCTCCCAACTGCACAGAAGAGATCACAGAAGGAGT
CTGGAATAGAAAGGCTAACAGAAAAGTGTTTGGACAGTGGGTGTGAGCAGTACTC
TAAACATGAGTGAAGCCATAAAAGGGTTCCTGGTCAAGTGCTGTGCATACAATTCCC
TTGGCACATCTTGTGAGACGATCCTTTTAAACTCTCCAGGCCCTTCCCTTTCATCCA
10 AGACAACATCTCATTTCTATGCAACAATTGGTGTTTGTCTCCTCTTCATTGTGCTTTTA
ACCCTGCTAATTTGTGACAAAGTACAAAAAGCAATTTAGGTATGAAAGCCAGCTACA
GATGGTACAGGTGACCGGCTCCTCAGATAATGAGTACTTCTACGTTGATTTTCAGAGA
ATATGAATATGATCTCAAATGGGAGTTTCCAAGAGAAAAATTTAGAGTTTGGGAAGG
TACTAGGATCAGGTGCTTTTGGAAAAGTGATGAACGCAACAGCTTATGGAATTAGC
15 AAAACAGGAGTCTCAATCCAGGTTGCCGTCAAAATGCTGAAAGAAAAAGCAGACAG
CTCTGAAAGAGAGGCACTCATGTCAGAACTCAAGATGATGACCCAGCTGGGAAGCC
ACGAGAATATTGTGAACCTGCTGGGGGCGTGCACACTGTCAGGACCAATTTACTTGA
TTTTTGAATACTGTTGCTATGGTGATCTTCTCAACTATCTAAGAAGTAAAAGAGAAA
AATTTACAGGACTTGGACAGAGATTTTCAAGGAACACAATTTACAGTTTTTACCCCA
20 CTTTCCAATCACATCCAAATTCAGCATGCCTGGTTCAAGAGAAGTTCAGATACACC
CGGACTCGGATCAAATCTCAGGGCTTCATGGGAATTCATTTCACTCTGAAGATGAAA
TTGAATATGAAAACCAAAAAAGGCTGGAAGAAGAGGAGGACTTGAATGTGCTTACA
TTTGAAGATCTTCTTTGCTTTGCATATCAAGTTGCCAAAGGAATGGAATTTCTGGAA
TTTAAGTCGTGTGTTTACAGAGACCTGGCCGCCAGGAACGTGCTTGTACCCACGGG
25 AAAGTGGTGAAGATATGTGACTTTGGATTGGCTCGAGATATCATGAGTGATTCCAAC
TATGTTGTACAGGGGCAATGCCCCGTCTGCCTGTAAAATGGATGGCCCCCGAAAGCCTG
TTTGAAGGCATCTACACCATTAAGAGTGATGTCTGGTCATATGGAATATTACTGTGG
GAAATCTTCTCACTTGGTGTGAATCCTTACCCTGGCATTCCGGTTGATGCTAACTTCT
ACAAACTGATTCAAAATGGATTTAAAATGGATCAGCCATTTTATGCTACAGAAGAA
30 ATATACATTATAATGCAATCCTGCTGGGCTTTTGAAGTCAAGGAAACGGCCATCCTTC
CCTAATTTGACTTCGTTTTTAGGATGTCAGCTGGCAGATGCAGAAGAAGCGATGTAT
CAGAATGTGGATGGCCGTGTTTCGGAATGTCCTCACACCTACCAAAACAGGCGACCT
TTCAGCAGAGAGATGGATTTGGGGCTACTCTCTCCGCAGGCTCAGGTGGAAGATTCC
TAGAGGAACAATTTAGTTTTAAGGACTTCATCCCTCCACCTATCCCTAACAGGCTGT

AGATTACCAAAACAAGATTAATTTTCATCACTAAAAGAAAATCTATTATCAACTGCTG
 CTTCCACCAGACTTTTTCTCTAGAAAGCTGTCGTGCTTTACTCTTGTTTTCAAAGGGACTT
 TTGTA AAAATCAAATCATCCTGTCACAAGGCAGGAGGAGCTGATAATGAACTTTATTG
 GAGCATTGATCTGCATCCAAGGCCTTCTCAGGCTGGCTTGAGTGAATTGTGTACCTG
 5 AAGTACAGTATATTCTTGTAATAACATAAAACAAAAGCATTTTGCTAAGGAGAAGC
 TAATATGATTTTTTAAGTCTATGTTTTAAATAATATGTAAATTTTTCAGCTATTTAG
 TGATATATTTTATGGGTGGGAATAAAATTTCTACTACAGAATTGCCATTATTGAAT
 TATTTACATGGTATAATTAGGGCAAGTCTTAAGTGGAGTTCACGAACCCCTGAAAT
 TGTGCACCCATAGCCACCTACACATTCCTTCCAGAGCACGTGTGCTTTTACCCCAAG
 10 ATACAAGGAATGTGTAGGCAGCTATGGTTGTCACAGCCTAAGATTTCTGCAACAAC
 AGGGGTTGTATTGGGGGAAGTTTATAATGAATAGGTGTTCTACCATAAAGAGTAAT
 ACATCACCTAGACACTTTGGCGGCCTTCCCAGACTCAGGGCCAGTCAGAAGTAACAT
 GGAGGATTAGTATTTTCAATAAAGTTACTCTTGTCCTCCACAAAAAAA (SEQ ID NO:
 74).

15 By “CD90 polypeptide” is meant a polypeptide or fragment thereof having at least about 95% amino acid sequence identity to an amino acid sequence provided at GenBank Accession No. NP_001298089 that binds an anti-CD90 antibody. An exemplary CD90 polypeptide sequence follows:

>NP_001298089.1 thy-1 membrane glycoprotein isoform 1 preproprotein [Homo
 20 sapiens]
 MNLAIISIALLLTVLQVSRGQKVTSLTACLVDQSLRLDCRHENTSSSPIQYEFSLTRETKKH
 VLFGTGVGPEHTYRSRTNFTSKYNMKVLVLSAFTSKDEGTYTCALHHS GHSPPISSQNV
 TVLRDKLVKCEGISLLAQNTSWLLLLLLSLSLQATDFMSL (SEQ ID NO: 75).

By “CD90 polynucleotide” is meant a nucleic acid molecule that encodes a CD90
 25 polypeptide. An exemplary CD90 polynucleotide sequence follows:

>NM_006288.5 Homo sapiens Thy-1 cell surface antigen (THY1), transcript variant 1,
 mRNA
 AGCAACCGGAGGCGGCGGCGCTGAGGAGGCTGCAGCAGCGGAAGACCCAG
 TCCAGATCCAGGACTGAGATCCCAGAACCATGAACCTGGCCATCAGCATCGCTCTCC
 30 TGCTAACAGTCTTGACAGGTCTCCCGAGGGCAGAAGGTGACCAGCCTAACGGCCTGC
 CTAGTGGACCAGAGCCTTCGTCTGGACTGCCGCCATGAGAATACCAGCAGTTCACCC
 ATCCAGTACGAGTTCAGCCTGACCCGTGAGACAAAGAAGCACGTGCTCTTTGGCACT
 GTGGGGGTGCCTGAGCACACATACCGCTCCCGAACCAACTTCACCAGCAAATACAA
 CATGAAGGTCCTCTACTTATCCGCCTTCACTAGCAAGGACGAGGGCACCTACACGTG

TGCACTCCACCACTCTGGCCATTCCCCACCCATCTCCTCCCAGAACGTACACAGTGCT
CAGAGACAAACTGGTCAAGTGTGAGGGGCATCAGCCTGCTGGCTCAGAACACCTCGT
GGCTGCTGCTGCTCCTGCTCTCCCTCTCCCTCCTCCAGGCCACGGATTTTCATGTCCCT
GTGACTGGTGGGGCCCATGGAGGAGACAGGAAGCCTCAAGTTCCAGTGCAGAGATC
5 CTACTTCTCTGAGTCAGCTGACCCCCCTCCCCCAATCCCTCAAACCTTGAGGAGAAG
TGGGGACCCCCACCCCTCATCAGGAGTTCCAGTGCTGCATGCGATTATCTACCCACGT
CCACGCGGCCACCTCACCTCTCCGCACACCTCTGGCTGTCTTTTTGTACTTTTTGTT
CCAGAGCTGCTTCTGTCTGGTTTATTTAGGTTTTATCCTTCCTTTTCTTTGAGAGTTTCG
TGAAGAGGGAAGCCAGGATTGGGGACCTGATGGAGAGTGAGAGCATGTGAGGGGT
10 AGTGGGATGGTGGGGTACCAGCCACTGGAGGGGTCATCCTTGCCCATCGGGACCAG
AAACCTGGGAGAGACTTGGATGAGGAGTGGTTGGGCTGTGCCTGGGCCTAGCACGG
ACATGGTCTGTCTGACAGCACTCCTCGGCAGGCATGGCTGGTGCCTGAAGACCCCA
GATGTGAGGGCACCAACAAGAATTTGTGGCCTACCTTGTGAGGGAGAGAACTGAGC
ATCTCCAGCATTTCTCAGCCACAACCAAAAAAAAAAATAAAAAGGGCAGCCCTCCTTAC
15 CACTGTGGAAGTCCCTCAGAGGCCTTGGGGCATGACCCAGTGAAGATGCAGGTTTG
ACCAGGAAAGCAGCGCTAGTGGAGGGTTGGAGAAGGAGGTAAAGGATGAGGGTTC
ATCATCCCTCCCTGCCTAAGGAAGCTAAAAGCATGGCCCTGCTGCCCCCTCCCTGCCT
CCACCCACAGTGGAGAGGGCTACAAAGGAGGACAAGACCCTCTCAGGCTGTCCCAA
GCTCCCAAGAGCTTCCAGAGCTCTGACCCACAGCCTCCAAGTCAGGTGGGGTGGAG
20 TCCAGAGCTGCACAGGGTTTGGCCCAAGTTTCTAAGGGAGGCACTTCCCTCCCCTCG
CCCATCAGTGCCAGCCCCCTGCTGGCTGGTGCCTGAGCCCCCTCAGACAGCCCCCTGCC
CCGCAGGCCTGCCTTCTCAGGGACTTCTGCGGGGCCTGAGGCAAGCCATGGAGTGA
GACCCAGGAGCCGGACACTTCTCAGGAAATGGCTTTTCCCAACCCCCAGCCCCACC
CGGTGGTTCTTCCCTGTTCTGTGACTGTGTATAGTGCCACCACAGCTTATGGCATCTCA
25 TTGAGGACAAAGAAAACCTGCACAATAAAACCAAGCCTCTGGAATCTGTCTCGTGT
CCACCTGGCCTTCGCTCCTCCAGCAGTGCCTGCCTGCCCCGCTTCGCTGGGGTCTCC
ACGGGTGAGGCTGGGGAACGCCACCTCTTCCCTCTTCCCTGACTTCTCCCCAACCCT
TAGTAGCAACGCTACCCCAGGGGCTAATGACTGCACACTGGGCTTCTTTTCAGAATG
ACCCTAACGAGACACATTTGCCCAAATAAACGAACATCCCATGTCTGCTGACTCACC
30 TGGCTGGAACAACATGCTTACTGCCAACATGTGGGCCGAACCACATGGCCCTGGCTC
TGGAATGCACAAGTGGCTTTGCGTGAATCTGCGCTAAGCTATGCAGTCTGCTTTTTC
TTCTCAGCTCTGGTAGTTCTTCAGAAATGTACCCTCCAGGCACATCCACTATTGCGA
GGGTGAGCACGAAGGGTGGGAGATGCCCATGTCTCAAGGCATCACTTCCTAAATC
CAAAAGCATCGGCAGGAGAAAGGACTGGGGACAAATACTGTCCCTTCGGGAGTAGG

GAGGGAACACTGAGGCCCCATCCCTGGCTCCTTCCCTAAAAGTAGAGTAAAATGGAA
GCGAGCATCCTGGGATTGGGGGCAAGAGGGGGGACCGCAGGGTAGCTGTGGGTTCCA
ACTGCTGTCAGAGTCAGAGAGGCAGCCCCAAGCCAGCCTCCCTGCTTTGCCAGGGA
ATTTGGGGGAGGAAGGTGACAGCTGCCCAGAGGCTGACTCATCTGATATTTAGCAC
5 TGGGTAGGATGATTGTTTCTGAGCATTTTTCTTAAAGGCCTCAGATCTAAATTATGCC
ACCGGCTCCCCTCTTGCTACCTCCCGTCAACTTCTCTGCCTTGCTTCCACCCCTGT
AGTTACCATACACAGAGGAGGAGGAGCTGTCCTTGTCAGGTTGGGAGGCTGACA
ACCCCTTAGCAAGATGCTGCCAGCCCAGAGCTCTCCAAGGGGAGGAACACCCCTGA
GACTCAGGCCCCCTCTCCTTCAGCCCTGCTTGGGCTGCAAGCGCCGTGCCAAGGAAAG
10 GCATCTTGGTGAGAAGAGCTGCTGTGGGGGAAGGGAGATCAAATGCCAGAGAAATG
TGGGGTGCCCCACCCTCAGGATAGTAAAAGAGTATGGAGGTATTTCTGGAAGGAAA
TGAGCGGCACTGTGTGAAGCCTCGCACCTGTGTGACACTTCCTATGGGGTCTTTGTC
ACACTCTAGTACTATGTCCCTGAAGAGTTTAGCAGCCACACTCTTAGAAGGGTGCTG
GGAGATGGTGTTGCCCTCTGCAGCCATGTTTAGGGGAGCGGAACCTGAGGCCACA
15 GTGGGTGAGATTAGCTCAAGAAGCCACAGAGGCCACCAGAGGGCCACGGACTTCGG
AAAGGAGAAGAGAAGAACAGGGCATCAGGCCTCACAACGCAAACCTACCCAGAGA
TGGGCACAGTGGCTCATGCCTGTAATCCCACCACTTTGGGAAGAGGCGGATCGCTTG
AGGTCAGGAGTTCGAGACTAGCCTCGAAACCCTATCTCTACTAAAAATACAAAAAT
TAGCCAGGCATGGTGGCCTGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAG
20 GAGAATCACTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCAAGATTGCACCACTG
CACTCCAGCCTGGGAGATAGAGTGAGACTCCATCTCAAAAAAATAAAAAATAAAAT
AAACCTACCCGGAATGACCATGCTGAGGACTGGGAGCCCGCAGACTTTCAGCCACA
GGCCGCGACAGCCGTGGGTCCCTCCCTGGTCAAGTCAGCAGGCCTTGTGGAGGCTGT
GGGGTATCTGTGGTGACTCAGGTAATTATAGAGGGCTGGCCCCCAGCCCTGGTTCCT
25 GTACACATGCCCCAACCCCATCCCCATCCACTCCCTCGCCAGTCCTAACCTCTTTCCT
GGGTCCCCCCCCCTTCAGCACCTAAGTCCATACCTAGGGCCGTGGAATTCCCGCTCAA
GAGCAACAGAAGCCCCCTCTCTGCACCCCCATTTCTGGACTGGATTGTCCACTGAGAC
GCGCAATGTCTGCATCTCTGACATCTAGAGGCTTCCTCGGGAAGGGCATGGGGATCT
CCGTGAGATGTGGGGACTTTCCTGAGGCAACCAAGAAATCTACACAGCGTCCGGGG
30 ACCTGTGACACACATCCCTCCCGCCTCCTCAACCTGATGTCCCTCTCTGAATCTGCAG
CTTTCGTGCTGTGAAGGTGTCTTTACATGTGAAACAAACAAACCAAGTCAAGAGTA
AATCATCTCATTTACTAGTGAGAAAATGTTGGAGCTGGAGTCCTTCAGAGAGTCCTG
GCCAGGCAAGAGGGGCCATCAGCTCTCTTCTGCTCAACAGGGGCTCTCAGCCTCAGG
ACACTCTCAGGCCTGGAATGTCCCCAACACACTCAAGGAGAAACATGTCTGTGCA

GACCCACAGGAGGCATCTTTGCCCCGGCACAAGGAAGAGCTGGGGTCAGTGGGACCT
 GTAGATGTAGACACATCATATGGAGGGTGGGTAGGACCAATGTGGCAGCTTCATGG
 AGGCCAAGTGTGGCTCTGCACCAGGAAGGGGCTGTGATGGCTGGAGGTGCCCAGCA
 GTGCAGGCGGGGAGTGCCTGGCAGTGGCGTGGCCAGGTGGAGGCCACCTGTCAAGT
 5 TTGCAATAAAGCAGTTTCCTGAATTTGGTGAGAA (SEQ ID NO: 76).

By “CD45 polypeptide” is meant a polypeptide or fragment thereof having at least about 95% amino acid sequence identity to an amino acid sequence provided at GenBank Accession No. NP_001254727 that binds an anti-CD45 antibody. An exemplary CD45 polypeptide sequence follows:

10 >NP_001254727.1 receptor-type tyrosine-protein phosphatase C isoform 5 precursor
 [Homo sapiens]
 MTMYLWLKLLAFGFAFLDTEVFVTGQSPTPSPTGHLQAEEQGSQSKSPNLKSREADSSA
 FSWWPKAREPLTNHWSKSKSPKAEELGV (SEQ ID NO: 77).

By “CD45 polynucleotide” is meant a nucleic acid molecule that encodes a CD45
 15 polypeptide. An exemplary CD45 polynucleotide sequence follows:

>NM_001267798.2 Homo sapiens protein tyrosine phosphatase receptor type C
 (PTPRC), transcript variant 5, mRNA
 GACATCATCACCTAGCAGTTCATGCAGCTAGCAAGTGGTTTGTCTTAGGGTAACAG
 AGGAGGAAATTGTTCTCGTCTGATAAGACAACAGTGGAGAAAGGACGCATGCTGT
 20 TTCTTAGGGACACGGCTGACTTCCAGATATGACCATGTATTTGTGGCTTAAACTCTT
 GGCATTTGGCTTTGCCTTTCTGGACACAGAAGTATTTGTGACAGGGCAAAGCCCAAC
 ACCTTCCCCCACTGGCCATCTGCAAGCTGAGGAGCAAGGAAGCCAATCCAAGTCAC
 CAAACCTCAAAAGTAGGGAAGCTGACAGTTCAGCCTTCAGTTGGTGGCCAAAGGCC
 CGAGAGCCCCCTCACAAACCACTGGAGTAAGTCCAAGAGTCCAAAAGCTGAGGAACT
 25 TGGAGTCTGATGTTCAAGAGCAGGAAGCAGCCAGCACGAGAGAAAGATGAAGACC
 AGAAGACTCAGCAAGCTCACTTCTCCTACCTTCTTGTGCCTGCTTTTTCTAGCCGTGC
 TGGCAGTTGCTTGGATGATGCCCCACTCATATTGGGTGGGGGTGGGGGGGTGGGGA
 GGGTCTGCCTCCCCCAGTCCACTGACTCAAATGTTAATCTCCCTTGGCAATACGCTC
 ACAGGCACACCCAGGAACAATACTTTGCATCCTTCAATCCAATCAAGTTGACACTCA
 30 ATATTAACCATCAAATACTATTATAAGGAGAATGTTGCATGATTTTCCTTCTAGTCTG
 TTTGTAATTCACATCTAATGAAAGAGTGAGAGTGGACGATAAAGGGAAGTTGTTGA
 AACATTTCTCTCAAAGCAAAAGGGATCATTGGAAGCAGGCAGACACCAGAATTGGT
 TTAACCTAAAAATAACAAATTAATAATTATCAAGTCTATAATGATGACAGTGACTTA
 ATGTGAATAGAAAGAATTCTAAACTCTCTCCTTCCTTCCTCCCTCCCTTCTTTCTAC

TTTCTTTCCACTCCCTTTCTCCCACCCCCTTTTCTTTTCCTTTCTTTTCTCCCACCCCTCT
 CTCCCTCCCTTTTCTTTTATTCAATGCATAGTAGTTGAAAAAATCTAAAGTTAGACCTG
 ATTTTACACTGAAGACTAGAGGTAGTTACTATCCTATTACTGTACTTAGTTGGCTATG
 CTGGCATGTCATTATGGGTAAAAGTTTGATGGATTTATTTGTGAGTTATTTGGTTATG
 5 AAAATCTAGAGATTGAAGTTTTTCATTAGAAAATAACACACATAACAAGTCTATGAT
 CATTTTGCATTTCTGTAATCACAGAATAGTTCTGCAATATTTTCATGTATATTGGAATT
 GAAGTTCAATTGAATTTTATCTGTATTTAGTAAAAATTAACCTTTAGCTTTGATACTAA
 TGAATAAAGCTGGGTTTTTTATTTA (SEQ ID NO: 78).

By “CD34 polypeptide” is meant a polypeptide or fragment thereof having at least about
 10 95% amino acid sequence identity to an amino acid sequence provided at GenBank Accession
 No. NP_001020280 that binds an anti-CD34 antibody. An exemplary CD34 polypeptide
 sequence follows:

>NP_001020280.1 hematopoietic progenitor cell antigen CD34 isoform a precursor
 [Homo sapiens]
 15 MLVRRGARAGPRMPRGWTALCLLSLLPSGFMSLDNNGTATPELPTQGTFNSVSTNVSY
 QETTTTSTLGSLSLHPVSQHGNEATTNITETTIVKFTSTSVITSVYGNNTSSVQSQTSVISTV
 FTTPANVSTPETTLKPSLSPGNVSDLSTTSTSLATSPTKPYTSSSPILSDIKAEIKCSGIREVK
 LTQGICLEQNKTSSEAEFKKDRGEGLARVLCGEEQADADAGAQCVCALLAQSEVRPQC
 LLLVLANRTEISSKLQLMKKHQSDLKKGILDFTEQDVASHQSYSQKTLIALVTSGALLA
 20 VLGITGYFLMNRWSPTGERLGEDPYTENGSGGQGYSSGPGTSPEAQGKASVNRGAQ
 ENGTGQATSRNGHSARQHVVADTEL (SEQ ID NO: 79).

By “CD34 polynucleotide” is meant a nucleic acid molecule that encodes a CD34
 polypeptide. An exemplary CD34 polynucleotide sequence follows:

>NM_001025109.2 Homo sapiens CD34 molecule (CD34), transcript variant 1, mRNA
 25 AGTGTCTTCCACTCGGTGCGTCTCTCTAGGAGCCGCGCGGGAAGGATGCTGGTCCGC
 AGGGGCGCGCGCGCAGGGCCCAGGATGCCGCGGGGCTGGACCGCGCTTTGCTTGCT
 GAGTTTGCTGCCTTCTGGGTTTCATGAGTCTTGACAACAACGGTACTGCTACCCAGAG
 GTTACCTACCCAGGGAACATTTTCAAATGTTTCTACAAATGTATCCTACCAAGAAAC
 TACAACACCTAGTACCCTTGGAAGTACCAGCCTGCACCCTGTGTCTCAACATGGCAA
 30 TGAGGCCACAACAACATCACAGAAACGACAGTCAAATTCACATCTACCTCTGTGA
 TAACCTCAGTTTATGGAAACACAACTCTTCTGTCCAGTCACAGACCTCTGTAATCA
 GCACAGTGTTACCCACCCAGCCAACGTTTCAACTCCAGAGACAACCTTGAAGCCTA
 GCCTGTACCTGGAAATGTTTCAGACCTTTCAACCACTAGCACTAGCCTTGCAACAT
 CTCCCCTAAACCCTATACATCATCTTCTCCTATCCTAAGTGACATCAAGGCAGAAA

TCAAATGTTTCAGGCATCAGAGAAGTGAAATTGACTCAGGGCATCTGCCTGGAGCAA
AATAAGACCTCCAGCTGTGCGGAGTTTAAGAAGGACAGGGGAGAGGGCCTGGCCCCG
AGTGCTGTGTGGGGAGGAGCAGGCTGATGCTGATGCTGGGGCCCAGGTATGCTCCC
TGCTCCTTGCCCAGTCTGAGGTGAGGCCTCAGTGTCTACTGCTGGTCTTGGCCAACA
5 GAACAGAAATTTCCAGCAAACCTCCAACCTTATGAAAAAGCACCAATCTGACCTGAAA
AAGCTGGGGATCCTAGATTTCACTGAGCAAGATGTTGCAAGCCACCAGAGCTATTCC
CAAAAGACCCTGATTGCACTGGTCACCTCGGGAGCCCTGCTGGCTGTCTTGGGCATC
ACTGGCTATTTCTGATGAATCGCCGAGCTGGAGCCCCACAGGAGAAAGGCTGGG
CGAAGACCCTTATTACACGGAAAACGGTGGAGGCCAGGGCTATAGCTCAGGACCTG
10 GGACCTCCCCTGAGGCTCAGGGAAAGGCCAGTGTGAACCGAGGGGCTCAGGAAAAC
GGGACCGGCCAGGCCACCTCCAGAAACGGCCATTTCAGCAAGACAACACGTGGTGGC
TGATACCGAATTGTGACTCGGCTAGGTGGGGCAAGGCTGGGCAGTGTCCGAGAGAG
CACCCCTCTCTGCATCTGACCACGTGCTACCCCCATGCTGGAGGTGACATCTCTTAC
GCCCAACCCTTCCCCACTGCACACACCTCAGAGGCTGTTCTTGGGGCCCTACACCTT
15 GAGGAGGGGCAGGTAAACTCCTGTCTTTACACATTCGGCTCCCTGGAGCCAGACTC
TGGTCTTCTTTGGGTAAACGTGTGACGGGGGAAAGCCAAGGTCTGGAGAAGCTCCC
AGGAACAATCGATGGCCTTGCAGCACTCACACAGGACCCCCCTTCCCCCTACCCCCCTCC
TCTCTGCCGCAATACAGGAACCCCCAGGGGAAAGATGAGCTTTTCTAGGCTACAATT
TTCTCCCAGGAAGCTTTGATTTTTACCGTTTCTTCCCTGTATTTTCTTTCTACTTTG
20 AGGAAACCAAAGTAACCTTTTGCACCTGCTCTCTTGTAATGATATAGCCAGAAAAAC
GTGTTGCCTTGAACCACTTCCCTCATCTCTCCTCCAAGACACTGTGGACTTGGTCACC
AGCTCCTCCCTTGTTCTCTAAGTTCCACTGAGCTCCATGTGCCCCCTCTACCATTTGC
AGAGTCCTGCACAGTTTTTCTGGCTGGAGCCTAGAACAGGCCTCCCAAGTTTTAGGAC
AAACAGCTCAGTTCTAGTCTCTCTGGGGCCACACAGAACTCTTTTTGGGCTCCTTTT
25 TCTCCCTCTGGATCAAAGTAGGCAGGACCATGGGACCAGGTCTTGGAGCTGAGCCTC
TCACCTGTACTCTTCCGAAAAATCCTCTTCCCTCTGAGGCTGGATCCTAGCCTTATCCT
CTGATCTCCATGGCTTCCCTCCTCCCTCCTGCCGACTCCTGGGTGAGCTGTTGCCTCA
GTCCCCCAACAGATGCTTTTTCTGTCTCTGCCTCCCTCACCTGAGCCCCCTCCTTGCT
CTGCACCCCCATATGGTCATAGCCCAGATCAGCTCCTAACCCCTTATCACCAGCTGCC
30 TCTTCTGTGGGTGACCCAGGTCCTTGTGTGCTGTTGATTTCTTTCCAGAGGGGGTTGAG
CAGGGATCCTGGTTTCAATGACGGTTGGAAATAGAAATTTCCAGAGAAGAGAGTAT
TGGGTAGATATTTTTCTGAATACAAAGTGATGTGTTTAAATACTGCAATTAAAGTG
ATACTGAAACACATCTGTTATGTGACTCTGTCTTAGCTGGGTGTGTCTGCATGCAAG
AGTGACACCCTCCATTAGACCTAGCTAGACTGTGCAGTGATGTGGTGGGGAGGACC

AGCCAGGGAAGAGGGAGCACCTCAGCAGACACAGGCACCAGCCAGGATGCTAAGG
ACCTTTAGCCAAGTCTGCCAACTATTCTCCTCCATGGGGAGAGGAAACATCCATTTC
CAGTGGTAGAAAGGCAGACCCGAATGTACCAGGGAGCTTCCAAATGGAGGGTGGTA
TGTTGGGTTCTTAGGAGCTGTACCCTTCATGAACACCCTTCTGAGAAGAGGAGCATG
5 CTGATCACTGCTGCAAAATATGCAAAACAAAGGGAAGGGGCAATGTCCTGTGCACC
CTTTATTATCAGGCCACCCCCCTCCCCAGCCCCCAGGTCAGAGTAGACACAGTGAA
GGACTATGTGGGGACTGTTGTTCTAGAGACCTGGCAGCCAACTCAGGGAGGGGGCT
GGTTTCCACCCTCAAGATTAAGACAGCAGCCTAATTAATAAAAAAAAAATCTGTAAGCA
TGTACCTCCCCCAGCTTCCAAAACAACCCCCACCCACCCCTACCAGGCCATAGGA
10 AGTTGGGGAGGGAGTGCTGAGGAGCTCCAGGAAACACTCCCAAGTGTGTGACAGT
GGCAGAGGCAGTTGGGGCCAAACAAAGGTTGATTCTTCCATTCTTATCTCCATAAAG
CCAGACCTTTCCCTTCAGCACTCCTCCACCCCCATCTCCTTCTTGCTTTTCTCCAACCTC
CTCTAATCATAGGTTCTTCCCTAGGACAGAGGGGAGGCGAAATGATGAGGTTTCAGA
GTCTTCCCTCAAAGGCGATGGCTGCCTTGAGGGTTGGAGCAAAGGATGATGAGCAA
15 AAGACGATGGTAATCAGTAGGGAAGTCCAGCCCACTTGCATCTAGTTGCACATCTTG
CCTTGAGAGTAATCCAGTGAGGGTCTGTCCCAGCTAGGACATCAAGTAGGAGGGGT
GGGTTTCAGGGTTTCAGATTCTTAGGAAATATGGGAGGAGAGGAAAAGGCAACTTGGA
TGCACCTCCAGCTTCAGGCCTAGCAACCTGCAATGCATCTCACCTGAGTTTGCTGG
AATGTGTATGTATGCTTTGGGAGGAAGGGCTGTGTGTGTATTGCGGGGTGGGGTGG
20 GGCAGCTGGTTCCCTCTGACAGCTGGACAGCTTGCCCTGAAGAATTTGCCTGCTTTC
TGGAATAATCCAACCTTTCCACCGTGGGCCTGAGCGTCCTGGTACAGCAATGGCGCC
ACCTGCTGGCCTTATTGAGGTCCTACTGCTCAGCCTCAGCTCAATCGCCTCCATGTTG
GGCTTCTCTCCCTGGCTGCCCCACCCTCTAGTCCAATTTCTCTTGACACAAAGCTCA
TATAACTATAGAACGTCACCTGTTGAAGAGAACTTTAAAGATACATTTAATTAACTC
25 CCTTATGGTATAGTTAAAGACAAACTAAGGCTCAGAGAAGGGAGGTGGCTTGCCCA
ATCACCCAGAATTCCAAAGTCCTGAATCTGTAGTTTTCCCTTCCATCATATCATCCTA
CTCTTCTGCCGAGTCTCCGTGTTACTCCAGTTGGATGTCATGAAGCCAGTGTGGCA
GTGTGAAGATAGGTTTGGGACTTCACTTCTGGAGCATTTTCATCAACATAAGCTATCC
TAGGCCTGGCCAGCCAAGCAGGTCCTGGAGGAGCCCCAGGACAAAGATCACAGGA
30 GGCCATGAGGTTCCGGCTTCTTCGGCGCCACAGTGAGCCCAGGAAAATTAGCTGTA
GGGTATTACACTGTTGACTATGGAGAGCATATCTGGAATTATCTTCAGCCAGATTTT
CATCTGAATGGATAAATGGGAATACCATCTAAGTCCAGATAAATAGATCACTTCCAT
CTCATCCCTTCTAGGTAGATTAATCCCACACTTCCTCTTCACACAAAACCAGTAATA
GGTCATCGATTTTGTGCAACAGGATGCTGCTTCTCTTCTTAAAGCCCCCATCGAAGA

GGCTTCCAGCCACCATTCAATCATTTCATCAAGTCTTATGATGTGCCAGACACTGCGC
GAAATGTGCCAGAACATCTGTTATGTGCCAGACACTGTTCTTGAGACTGGGGATACA
GCAAACACTCATGAAGCTTATAATTCTAGCAGAAGAGGACAGTAAACAATGTCATC
TCAGTAAGTATATACATGTGTTTTTCAGGATTGAGAGCTATGAAAAACATAAAATATA
5 TTGAGAATAATGGTTGGTATTTTACATATGGTGGTTACTTTTAGAAAAATAACAGTG
GAGAGCACAGCTTCACTTGAATGAAGTGGAGAAGCAGGTTGTATGCCAAGCTGGGA
GAGATTATCCCACACAGGGGAAAGGACAAGTGCAAAGCCCTATGATGAAAAGCTGC
CAAGTGCGAGAAAGCCTCAGATGGCAGGGGGCAAGATGGCCATGAGGTTGTGTCAGT
GAGTGGGGGTGGGGAGAGGCAGGAGGTCAGACTACATGGGGCCTTTTTAGTTGTAG
10 ATTGGGAAGCCACTGGAGGGTTTTGAGCAGAGAAGTCATATCATCTGCTTTATGTTT
TAAAAGGATCATGCTGGCTGCTGAGTAGAGAATAGAGGTTGAGGGATAAGAAAGTA
GAAGGAGACCGTAGCAAGAAGAACGATCATGGCTGGGAGCAGGTGATCATATTGGC
AGTGATGAGATCAAGCAGAATTCAAAAAGTGGTTTCAAAGTAGAGGTAACAGGACT
TGCTCAGTCTATTTATTTCTTCAAATAATAATCATATTTACAATGATAGTAGCTAACA
15 GTTTTTGAGTGCTTACTGTATGAAAATTGAGATATGGTGCCAATATTTAAATAGCAT
ATTTTACTTAACATTCACAGAAACCCTGTGAAGTAGGTTCTATTATCTCAGAAAAAG
AAACTGAAACTCAGAGAATAACAAGGGACTGTGTTACGTGCACAGTGGCAGAGGCA
AAGATGAATAGGATGTGAGTTTATTTGAACCCCAAATGTTTAAATCTTGGGGATAAT
ACAACACACATTTAAACAAAGAAGCAAGAAAAAAAAATGCACAACAGAAAGTGAGA
20 AATAACACGAGGAAAGACTAAATGAAGTGCTTTGTATCTAGATGTGGGCAGGACCC
TTTCCAGCTGAGAAGATCTGAGACTGGGTCATGAACAGGTGGTTTTCTGAGTGGGTCC
TGTA AAAATGAATACGATTTTGTATGATAGTAATGAGTAAGGACATTTGAGACTGAT
AGAAGAGTACATACAATATGTAGTGATGGGGAAAGATAAGGTACTGTCAAAGGACA
ATGTGTTTTCTGGTATGACAGAGAAGTAGAATGTGTTAAGGGAAGCCGAGTACCAG
25 AAAGATCCGGGTGTCACAGTTTGTGTAGGGTGTTTAAAGCTAAACCACAGAGTTTAA
TTTTATCCAATAGAAGAGGAGCCACAGAAGAGTTTCCATTTATTCATTAATTTATTC
ATTTATTCAAAAAATATTTGAGTGCTTATTATAAGCCAGGTACTATGCCAGGCACCT
GGGATAAGACATAGTCCCTTCTGTCAAGTCTTTACATTGGGTGGATGTGGGAGGGAC
AGATGACAGAACAATATGCATTGAGTGTAAGTGCTATGGTATAGGAAGCTCTGAGT
30 GGGAGGGGCATGGAAGCCGTGGAAGACCATGGAAGGCTTCCAGGAGAAGTGACG
TCTGGACTGATCCTTTGGTCAAGCAGGAGTTAAAGAGGAGAAAAGGAGAGATATGG
GTGTTCCCGAGAGAGGAAGAAGCCTTGTCCCAGGAGCAAAGTGAGGGTGATTGTTC
CAGAAATGTGAGTGATTCTTTTAAGGCTCAAGCAAAGCATGTGATTCTTCTTTATAC
CTTCTATTTCTTTGCTGAGTGTTTCTGTTCTTTTGTTC AAGCATGCTGCAATTGCTCA

TTAAAGCATGTTTATGATGGCTGTCTGTTTTAAAATTCTTGTCAGATGGTTTCAACAT
 CTTTATCATCTCAATGTTGGCATCTGTTAATGGTTTTTCTCAATCAAATTGAGATTT
 TCCTGGTTCCTTGGTATTACCAGTGATTTTAATTGCATCTGGAAATTTGGGATTTATGT
 TGAAAGACTGGATCTTATTGAAAGATTCTGTTTAGCACCCCTCCTTTGATAACCACAC
 5 TGGTGGGTCCAGGTTCCTTCAGCTGTTGACACCTTCAGGGCAGAGAGGTGGGAT
 GGGGTGAAGGGGGTACCTCATTATTGCTGGCCCAGGTTAGAAGTTCAGGCTTCCCAG
 TAGATCTCTGCTGATAACCACCTGGTGCCATGTCATTCTTGAGTCCAAAAGTCCCTC
 CCAATTCTGCCTTCTTCTCTCTACATATCGGAGTCTCCCTATGTTTGACTTATATATA
 ATGTCCAGGGTTTTTAGAGTTAGTTAACAGGAGGCATAAGAAAAAGTGTGTCCACTC
 10 CATCTTGTCTGGAAGTTCAGGAGTTCGAATATAAGAGAGAGGAGAGGAAATTA
 CAAGCCATGAGACTGGAGAGTTAGGCAGGTTCTACACCAGCTATTCTCAAAGCCCTC
 TTACACTCTTAAAAATTTAGAACTTCAAAGAGCTTTTGATTTTGAAAGTTACATCTAT
 CAATTATTACTGTTTCAAAAATTTAAATTTGAGAAAAATTTTATTTATTAATTTGTTTAA
 AAATAACAATAATTATTCAATTACATGATAATGTAAGTAATGCTTTTCTTAATGAAA
 15 AATAATTATATTTTCCAAAACAAAAACAATTAGGAAAAAGAGTGTCATTGTTTTAGA
 CTTTGGTAAATCTCTCTAATATCTGGCTGAAGAGAAGAATGCTGATTCTTTTTTTTT
 TTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCACCCAGGCTGGAGTGTAGTGGTGTGA
 TCTCGGCTCACTGCAAGCTCTGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTC
 CCAAGTAGCTGGGACTACAGGCACCCGCCACCACGCCCGGCTAATTTTTTTGTATTT
 20 TTAGTAGAGATGGGGTTTTACCGTGTTAGCCAGGCTGGTCTCGATCTCCTGACCTCA
 TGATCCACCCACCTCAGCCTCCCAAAGCGCTGGGATTACAGGTGTGAGACACCGCG
 CCCAGCCCCCGAATGCTGATTCTTTTATCTGCTTCTGTATTCAATCTGTTGTGATATG
 ATGGGTAGCCTCTGAAACACTCCACTGTATACTTGTGAAAGAATGAATGTGAAAAA
 GGAAAATAGATTTGTAGTATTATTATTCAAATTGTTTTGACCTCAGAGACCACTTGG
 25 AAATGTTTTAGGGAACCCCCAGAGGACCTTGGATCATGCTTTGAGAACCGCGGCTCT
 AGATATGTTACTATTTTCAGTAGCATCTAAGTACATGTGGCTGCTGAGCACTTGTAAT
 GTGGCTAGTGCAAATGAGAGACAGGACTTCCAGCTATATGTAATTTAATAAACTCA
 AATTTAAAACTGGAACCTCATAAAATGTTTTGTTGTTGTTGTTAAACATGACCTTAT
 AGTTTTGGTAGGAA (SEQ ID NO: 80).

30 By “Stem Cell Factor (SCF) polypeptide” is meant a polypeptide or fragment thereof
 having at least about 95% amino acid sequence identity to an amino acid sequence provided at
 GenBank Accession No. NP_000890 that functions in hematopoiesis. In some embodiments, a
 SCF polypeptide or fragment thereof binds CD117. An exemplary SCF polypeptide sequence
 follows:

>NP_000890.1 kit ligand isoform b precursor [Homo sapiens]

MKKTQTWILTCIYLQLLLFNPLVKTEGICRNRVTNNVKDVTCLVANLPKDYMITLKYP
 GMDVLPSHCWISSEMVVQLSDSLDILLDKFSNISEGLSNYSIIDKLVNIVDDLVECVKENS
 SKDLKKSFKSPEPRLFTPEEFFRIFNRSIDAFKDFVVASETSDCVVSSTLSPEKDSRVSVTK
 5 PFMLPPVAASSLRNDSSSSNRKAKNPPGDSSLHWAAMALPALFSLIIGFAFGALYWKKR
 QPSLTRAVENIQINEEDNEISMLQEKEREFQEV (SEQ ID NO: 81).

By “SCF polynucleotide” is meant a nucleic acid molecule that encodes a SCF polypeptide. An exemplary SCF polynucleotide sequence follows:

>NM_003994.5 Homo sapiens KIT ligand (KITLG), transcript variant a, mRNA

10 GGGCTTCGCTCGCCGCCTCGCGCCGAGACTAGAAGCGCTGCGGGAAGCAGGGACAG
 TGGAGAGGGCGCTGCGCTCGGGCTACCCAATGCGTGGACTATCTGCCGCCGCTGTTC
 GTGCAATATGCTGGAGCTCCAGAACAGCTAAACGGAGTCGCCACACCACTGTTTGT
 GCTGGATCGCAGCGCTGCCTTTCCTTATGAAGAAGACACAACTTGGATTCTCACTT
 GCATTTATCTTCAGCTGCTCCTATTTAATCCTCTCGTCAAACTGAAGGGATCTGCAG
 15 GAATCGTGTGACTAATAATGTAAAAGACGTCCTAAATTGGTGGCAAATCTTCCAA
 AAGACTACATGATAACCCCTCAAATATGTCCCCGGGATGGATGTTTTGCCAAGTCATT
 GTTGGATAAGCGAGATGGTAGTACAATTGTCAGACAGCTTGACTGATCTTCTGGACA
 AGTTTTCAAATATTTCTGAAGGCTTGAGTAATTATTCCATCATAGACAACTTGTGA
 ATATAGTGGATGACCTTGTGGAGTGCGTGAAAGAAAACTCATCTAAGGATCTAAAA
 20 AAATCATTCAAGAGCCCGAGAACCAGGCTCTTTACTCCTGAAGAATTCTTTAGAATT
 TTTAATAGATCCATTGATGCCTTCAAGGACTTTGTAGTGGCATCTGAACTAGTGAT
 TGTGTGGTTTTCTTCAACATTAAGTCCTGAGAAAGGGAAGGCCAAAAATCCCCCTGGA
 GACTCCAGCCTACACTGGGCAGCCATGGCATTGCCAGCATTGTTTTCTCTTATAATT
 GGCTTTTGCTTTTGGAGCCTTATACTGGAAGAAGAGACAGCCAAGTCTTACAAGGGC
 25 AGTTGAAAATATACAAATTAATGAAGAGGATAATGAGATAAGTATGTTGCAAGAGA
 AAGAGAGAGAGTTTCAAGAAGTGTAATTGTGGCTTGATCAACACTGTTACTTTTCGT
 ACATTGGCTGGTAACAGTTCATGTTTGCTTCATAAATGAAGCAGCTTTAAACAAATT
 CATATTCTGTCTGGAGTGACAGACCACATCTTTATCTGTTCTTGCTACCCATGACTTT
 ATATGGATGATTAGAAATTGGAACAGAATGTTTTACTGTGAACTGGCACTGAATT
 30 AATCATCTATAAAGAAGAACTTGCATGGAGCAGGACTCTATTTTAAGGACTGCGGG
 ACTTGGGTCTCATTTAGAACTTGCAGCTGATGTTGGAAGAGAAAGCACGTGTCTCAG
 ACTGCATGTACCATTTGCATGGCTCCAGAAATGTCTAAATGCTGAAAAAACACCTAG
 CTTTATTCTTCAGATACAACTGCAGCCTGTAGTTATCCTGGTCTCTGCAAGTAGATT
 TCAGCTTGGATAGTGAGGGTAACAATTTTCTCAAAGGGATCTGGAAAAAATGTTTA

AAACTCAGTAGTGTCTAGCCACTGTACAGTGTAGAAAGCAGTGGGAACTGTGATTGG
ATTTGGCAACATGTCAGCTTTATAGTTGCCGATTAGTGATATGGGTCTGATTTTCGATC
TCTTCCTGATGTAAACCATGCTCACCCATATCCCCTATACAAATGCAAATGGTTGC
CTGGTTCCATTTATGCAAGGGAGCCAGTACTGAATTATGCCTTGGCAGAGGGGAGA
5 CTCCAAAAGAGTCATCGCAGGAAGAAGTTAAGAACACTGAACATCAGAACAGTCTG
CCAAGAAGGACATTGGCATCCTGGGAAAAGTCCGCCTTTTCCCTTGACCACTATAGGG
TGTATAAATCGTGTTTGCAAAATGTGTTATGATGTGTTTATATTCTAAAACTATTACA
GAGCTATGTAAAGGGACTTAGGAGAAAAATGCTGAATGTAAGATGGTCCCATTTCAA
TTTCCACCATGGGAGAGCCTAAAAATAAATTATGACATTTAGTATCTAAGGTTAGAA
10 AACCACGCCACATGCTAATATGGGTGTTGAAAAGTACTAGGTTACTTATAATGCAAGG
AATCAGGAACTTTAGTTATTTATAGTATAATCACCATTATCTGTTTAAAGGATCCA
TTTAGTTAAAATCGGGCACTCTATATTCATTAAGGTTTATGAATTAAGAAAGAAAGCT
TTATGTAGTTATGCATGTCAGTTTGCTATTTAAAATGTGTGACAGTGTGTCATATT
AAGAGTGAATTTGGCAGGAATTCCCAAGATGGACATTGTGCTTTTAACTAGAACTT
15 GTAAGACATTATGTGAATATCCCTTGCCAATTTTTTTTATAATAAGAAAACATCTGA
CTAAAGTCAAAGAATGATTTCTTATGGTTTATTTTGATGAAAGTTCTTTTAACATGTC
TTGAATGTACACATAAAGGAATCCAAAGCTTTCCATTCTAACTTAATCTTTGTGATA
ACATTATTGCCATGTTCTACAACCGTAAGATGACAGTTTTCAATGTAGTGACACAAA
AGGGCATGAAAACTAACTGCTAGCTTTCCCTTTCAATTTCAAAAAGTCCAAGAATTTCT
20 AGTATAATTTGGATTTTAGCTTCTGTTCAAAGCAAATCCAGATGCAACTCCAGTAAGT
GGCCTTTGCTCTTTTTTTGTACCAAAGAGCCCAGATGATTCCTACAGTCCCTTTCTTCT
CTAACATGCTGTGGTTCCTTAAATATGAGTAATTTCTCTAAGATATAACCCAGGTGC
TTTGAGAAGCTGCATTAAGGTGTTTCAAGGCCCTCAGATATCACATGGTACACTTGATT
AGTAATAAAACCAGAGATCAATTTAAATTGCTGATAGGTCCTGTCTCAGTGTGTGGC
25 ATTGACTGTTTTTCAGGAAAATAGATACAGATTAATATGAGTTATGCGTGTAGGTTGT
GTATAGATTGAGAAGATAGATACTTCTCAATCTAGTAGTTTGATTTATTTAACCAAT
GGTTTCAGTTTGCTTGAGCATATGAAAATCCTGCTTAATGTGCTTAAGAGTATAATA
AATGTGTACTTTTGTCTCAAACCTAGTAGCTGGGTTTTAACACTCATGGACATGGT
CTTAATCAATGGAGTTAAATAAACAATTCAGCAAGTTATTAATCTGACATGGTAG
30 GAGAGGGGAGATGTGTCCTGCTTATTAAATGTGTTGGTCCATTGAAAGTTACATGGA
TTGCCAATTTTTTAAACACTAAAGTTGAATAAAATGCATGAACAATAGAAAAATGC
TGAACATTATTTGGATGCTAGCTGCTTGGACATTAACACTGTGTTATTTCTGCTTTGAG
ATGAAAATATATATTTATCTTTGCTTATTTTATCCCAGATGTGTTCTGAATATCCTTC
TTCATAAATCATGGAAAACACTGCTGAGATAGTAAACCATGAAATCGCCTTTTCA

GTTGGTGCCATGTATCTGACAGTTCCATCTTGGAAGGTTTCAAAATTACCTTTTAAA
ATGATCTCAGAAAGTCTGTAGATTCTCAATGATACTGAAAGCTTTGCACCTCTTTGGT
AGAAACCAGGTCTATTTAGAAAATGGCTTTATGATAAATGTTGCCTCCTGAGTGATA
ATGAAGTGTCCTGGATATTGTATTGTAATTTAATGTGCTTACCACACTGCCACATTT
5 TAATGAGTCAGAGAAAAATTAATTTTTCTTCAATACAATAATAGAACAAGTAGCCTA
TTCTCTTAAAAAGTATGTGAAAAGAAAATTATGAAAAAATATGCATACCTAATGAA
GTATTGGTTTTAGTAAGAATTAAATACATTTTCATTGAGCTTTAAAGTACTTTGGAGA
AACTTTGGGGCACGTTTTCTACTCTAATTCAACTAAAGTTATAAATAAAGAGAAAA
ACTCATTGAGAAATCATGGATTTTAAAAATATTTTACTGCAGCCAAGTTTTCATTTCA
10 AAATGTAATTTTCAGTTTGGAGCTTTTAGGCATTATGTATATTTAAAAAATATATTCTT
CAAAAATGCATTTTGGCATGGTGGGATGGATGTTGCAAAAGATATCCGGAGCCTCC
AGTCTGTCATTAAGTATGATATGGTAAATCACCTCTCTTCTTTGGGTCTCAATTTTTTAT
TTATCTATATGGTAACTCAGAGATCACTCCTTAGGGGTGAGTCCTATTGCAATATG
ACCGACAAAGAAGACAAAATAGCATTGAACTAACCACATACAAAATATCCAACCTCT
15 GGATTCTGTGAATAAGTATCTTGACCATAAAAAGTCATTGCTGTTCTTGTTTCTAATG
TAAATAGTGTCCATTAGTAAAAGTGAAATTCAGTCTTAAGTAGGGTGAATTGGATCA
CCATTTACACAAGAGATGGCTTTTTCTTTGCTTGAATAAACATTTTGGATCACCTCC
AAAGAATGAAAACCAGTAGTACGTTTTAGTCATATTAGTCAGGATGAGAACTATA
AGATGTGTGTAACATTTGGAAATGCACCAAAGTGAGCGTTTAAATCTTCTCATTTTA
20 TTGAAAACATAAGAGCAGAAAATGTAAAATGCTCATGAAGGTTTTGAATGCCAAAAG
ATATTTTAGAATCAATTTATAAAGGGGTAATTCATTAATTACACTTTAAAATTGGAA
AGTGGGATAAGAAATCTAAAGTAAACCAGCTTATCTTTGAAACAATATTATTTTGAA
ATTGGCTTTAAAATAAAACCATTTCAGATTGAAATTCATAATTAGCTCATTTGTGGAGT
TTGATCACACAATTCATAATGTTGCTGCTTTCCATTAAGTCTTGAAATGCCTTTG
25 TTIGTAAAAATAAAATAATGGTACTTTTCATTTTATAACAAGGTGTTTTTTTCAAGAA
ATAATCCATGCTAAAATGGATATTTGTGATCCTGAAATGTTTACTAAGCATTGTAAA
TTATTTTATAACTGCCATCTCCAACCTACATCCTTATGATGTTTTTAACAATAAAATTA
AAACAACCTGTTAACTAAAAACCACACCGTTTTCCAGTACTTGATCTCTGAGCTACA
ATACTCACTAAATATAATTTTCCAATCAAAATATTCTATTCTATATTCTAAGGGTTAA
30 TATGTGATTATAGTGTCCACTTGCCACCATTTTTTTTAAATCAATGGACTTGAAAAGTA
TTAATTTAGATGGATGCGCAGATATACCCTCAGTTCAGTCATAGATTGGAGTTTGCA
TATAATAATGTAAATGTATGTCGACACTATTCTAAATAGTTCTATTATGACTGAAAT
TTAATTAAATAAAAAAGGTTGTAAAATGTGATGTGTATGTGTATATACTGTATGTGT
ACTTTTTAAATAGGTGTATGTCCCAACCCTTTTTTATACAGGTTTGAATTTAAATTT

ACATGATATATACATATACTTTATTGTTCTAAATAAAGAATTTTATGCACTCTCAAA
AAAAAAAAAAAAAAAA (SEQ ID NO: 82).

The term “linker”, as used herein, refers to a molecule that links two moieties. In one embodiment, the term “linker” refers to a covalent linker (*e.g.*, covalent bond) or a non-covalent

5 “Makassar” or “Hb G-Makassar” refers to a human β -hemoglobin variant, the human Hemoglobin (Hb) of G-Makassar variant or mutation (HB Makassar variant), which is an asymptomatic, naturally-occurring variant (E6A) hemoglobin. Hb G-Makassar was first identified in Indonesia. (Mohamad, A.S. et al., 2018, *Hematol. Rep.*, 10(3):7210 (doi:10.4081/hr.2018.7210). The Hb G-Makassar mobility is slower when subjected to
10 electrophoresis. The Makassar β -hemoglobin variant has its anatomical abnormality at the β -6 or A3 location where the glutamyl residue typically is replaced by an alanyl residue. The substitution of single amino acid in the gene encoding the β -globin subunit β -6 glutamyl to valine will result as sickle cell disease. Routine procedures, such as isoelectric focusing, hemoglobin electrophoresis separation by cation-exchange High Performance Liquid
15 Chromatography (HPLC) and cellulose acetate electrophoresis, have been unable to separate the Hb G-Makassar and HbS globin forms, as they were found to have identical properties when analyzed by these methods. Consequently, Hb G-Makassar and HbS have been incorrectly identified and mistaken for each other by those skilled in the art, thus leading to misdiagnosis of Sickle Cell Disease (SCD). In one embodiment, the valine at amino acid position 6, which
20 causes sickle cell disease, is replaced with an alanine, to thereby generate an Hb variant (Hb Makassar) that does not generate a sickle cell phenotype. In some embodiments, a Val to Ala (GTG to GCG) replacement (*i.e.*, the Hb Makassar variant) can be generated using an A•T to G•C base editor (ABE).

Thus, the present invention includes compositions and methods for base editing a
25 thymidine (T) to a cytidine (C) in the codon of the sixth amino acid of a sickle cell disease variant of the β -globin protein (Sickle HbS; E6V), thereby substituting an alanine for a valine (V6A) at this amino acid position. Substitution of alanine for valine at position 6 of HbS generates a β -globin protein variant that does not have a sickle cell phenotype (*e.g.*, does not have the potential to polymerize as in the case of the pathogenic variant HbS). Accordingly, the
30 compositions and methods of the invention are useful for the treatment of sickle cell disease (SCD).

By “marker” is meant any protein or polynucleotide having an alteration in expression level or activity that is associated with a disease or disorder, such as, for example, sickle cell

disease (SCD), thalassemia, anemia, hemoglobin C disease, hemoglobin S-C disease, or other hemaglobinopathies involving the abnormal or aberrant production or structure of hemoglobin.

The term “mutation,” as used herein, refers to a substitution of a residue within a sequence, *e.g.*, a nucleic acid or amino acid sequence, with another residue, or a deletion or
5 insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Various methods for making the amino acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring
10 Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

The term “single nucleotide polymorphism (SNP)” is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population (*e.g.*, > 1%). For example, at a specific base position in the human genome, the C nucleotide can appear in most individuals, but in a minority of
15 individuals, the position is occupied by an A. This means that there is a SNP at this specific position, and the two possible nucleotide variations, C or A, are said to be alleles for this position. SNPs underlie differences in susceptibility to disease. The severity of illness and the way our body responds to treatments are also manifestations of genetic variations. SNPs can fall within coding regions of genes, non-coding regions of genes, or in the intergenic regions
20 (regions between genes). In some embodiments, SNPs within a coding sequence do not necessarily change the amino acid sequence of the protein that is produced, due to degeneracy of the genetic code. SNPs in the coding region are of two types: synonymous and nonsynonymous SNPs. Synonymous SNPs do not affect the protein sequence, while nonsynonymous SNPs change the amino acid sequence of protein. The nonsynonymous SNPs are of two types:
25 missense and nonsense. SNPs that are not in protein-coding regions can still affect gene splicing, transcription factor binding, messenger RNA degradation, or the sequence of noncoding RNA. Gene expression affected by this type of SNP is referred to as an eSNP (expression SNP) and can be upstream or downstream from the gene. A single nucleotide variant (SNV) is a variation in a single nucleotide without any limitations of frequency and can arise in somatic
30 cells. A somatic single nucleotide variation can also be called a single-nucleotide alteration.

The terms “nucleic acid” and “nucleic acid molecule,” as used herein, refer to a compound comprising a nucleobase and an acidic moiety, *e.g.*, a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, *e.g.*, nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are

linked to each other via a phosphodiester linkage. In some embodiments, “nucleic acid” refers to individual nucleic acid residues (*e.g.* nucleotides and/or nucleosides). In some embodiments, “nucleic acid” refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms “oligonucleotide” and “polynucleotide” can be used interchangeably to refer to a polymer of nucleotides (*e.g.*, a string of at least three nucleotides). In some embodiments, “nucleic acid” encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g.*, a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms “nucleic acid,” “DNA,” “RNA,” and/or similar terms include nucleic acid analogs, *e.g.*, analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, *e.g.*, in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5' to 3' direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.* adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (*e.g.*, methylated bases); intercalated bases; modified sugars (*e.g.*, fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (*e.g.*, phosphorothioates and 5'-*N*-phosphoramidite linkages).

The term “nuclear localization sequence,” “nuclear localization signal,” or “NLS” refers to an amino acid sequence that promotes import of a protein into the cell nucleus. Nuclear localization sequences are known in the art and described, for example, in Plank *et al.*, International PCT application, PCT/EP2000/011690, filed November 23, 2000, published as WO/2001/038547 on May 31, 2001, the contents of which are incorporated herein by reference

for their disclosure of exemplary nuclear localization sequences. In other embodiments, the NLS is an optimized NLS described, for example, by Koblan *et al.*, Nature Biotech. 2018 doi:10.1038/nbt.4172. In some embodiments, an NLS comprises the amino acid sequence KRTADGSEFESPKKKRKV (SEQ ID NO: 83), KRPAATKKAGQAKKKK (SEQ ID NO: 84),
 5 KKTELQTTNAENKTKKL (SEQ ID NO: 85), KRGINDRNFWRGENGGRKTR (SEQ ID NO: 86), RKSGKIAAIVVKRPRK (SEQ ID NO: 87), PKKKRKV (SEQ ID NO: 88), or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 89).

The term “nucleobase,” “nitrogenous base,” or “base,” used interchangeably herein, refers to a nitrogen-containing biological compound that forms a nucleoside, which in turn is a
 10 component of a nucleotide. The ability of nucleobases to form base pairs and to stack one upon another leads directly to long-chain helical structures such as ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). Five nucleobases – adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U) – are called primary or canonical. Adenine and guanine are derived from purine, and cytosine, uracil, and thymine are derived from pyrimidine. DNA and RNA can
 15 also contain other (non-primary) bases that are modified. Non-limiting exemplary modified nucleobases can include hypoxanthine, xanthine, 7-methylguanine, 5,6-dihydrouracil, 5-methylcytosine (m5C), and 5-hydrmethylcytosine. Hypoxanthine and xanthine can be created through mutagen presence, both of them through deamination (replacement of the amine group with a carbonyl group). Hypoxanthine can be modified from adenine. Xanthine can be modified
 20 from guanine. Uracil can result from deamination of cytosine. A “nucleoside” consists of a nucleobase and a five carbon sugar (either ribose or deoxyribose). Examples of a nucleoside include adenosine, guanosine, uridine, cytidine, 5-methyluridine (m5U), deoxyadenosine, deoxyguanosine, thymidine, deoxyuridine, and deoxycytidine. Examples of a nucleoside with a modified nucleobase includes inosine (I), xanthosine (X), 7-methylguanosine (m7G),
 25 dihydrouridine (D), 5-methylcytidine (m5C), and pseudouridine (Ψ). A “nucleotide” consists of a nucleobase, a five carbon sugar (either ribose or deoxyribose), and at least one phosphate group.

The terms “nucleic acid” and “nucleic acid molecule,” as used herein, refer to a compound comprising a nucleobase and an acidic moiety, *e.g.*, a nucleoside, a nucleotide, or a
 30 polymer of nucleotides.

As used herein, the terms “oligonucleotide” and “polynucleotide” can be used interchangeably to refer to a polymer of nucleotides.

The term “nucleic acid programmable DNA binding protein” or “napDNAbp” may be used interchangeably with “polynucleotide programmable nucleotide binding domain” to refer to

a protein that associates with a nucleic acid (*e.g.*, DNA or RNA), such as a guide nucleic acid or guide polynucleotide (*e.g.*, gRNA), that guides the napDNAbp to a specific nucleic acid sequence. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable RNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a Cas9 protein. A Cas9 protein can associate with a guide RNA that guides the Cas9 protein to a specific DNA sequence that is complementary to the guide RNA. In some embodiments, the napDNAbp is a Cas9 domain, for example a nuclease active Cas9, a Cas9 nickase (nCas9), or a nuclease inactive Cas9 (dCas9). Non-limiting examples of nucleic acid programmable DNA binding proteins include, Cas9 (*e.g.*, dCas9 and nCas9), Cas12a/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ (Cas12j/Casphi). Non-limiting examples of Cas enzymes include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (also known as Csn1 or Csx12), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/CasΦ, Cpf1, Csy1, Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Type II Cas effector proteins, Type V Cas effector proteins, Type VI Cas effector proteins, CARE, DinG, homologues thereof, or modified or engineered versions thereof. Other nucleic acid programmable DNA binding proteins are also within the scope of this disclosure, although they may not be specifically listed in this disclosure. See, *e.g.*, Makarova *et al.* "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?" *CRISPR J.* 2018 Oct;1:325-336. doi: 10.1089/crispr.2018.0033; Yan *et al.*, "Functionally diverse type V CRISPR-Cas systems" *Science*. 2019 Jan 4;363(6422):88-91. doi: 10.1126/science.aav7271, the entire contents of each are hereby incorporated by reference. Exemplary nucleic acid programmable DNA binding proteins and nucleic acid sequences encoding nucleic acid programmable DNA binding proteins are provided in the Sequence Listing as SEQ ID NOs: 90-123 and 158.

The terms "nucleobase editing domain" or "nucleobase editing protein," as used herein, refers to a protein or enzyme that can catalyze a nucleobase modification in RNA or DNA, such as cytosine (or cytidine) to uracil (or uridine) or thymine (or thymidine), and adenine (or adenosine) to hypoxanthine (or inosine) deaminations, as well as non-templated nucleotide

additions and insertions. In some embodiments, the nucleobase editing domain is a deaminase domain (*e.g.*, an adenine deaminase, an adenosine deaminase; a cytidine deaminase or a cytosine deaminase).

As used herein, “obtaining” as in “obtaining an agent” includes synthesizing, generating,
5 producing, isolating, purchasing, or otherwise acquiring the agent.

A “patient” or “subject” as used herein refers to a mammalian subject or individual diagnosed with, at risk of having or developing, susceptible to having or developing, or suspected of having or developing a disease or a disorder. In some embodiments, the term “patient” refers to a mammalian subject with a higher than average likelihood of developing a
10 disease or a disorder. Exemplary patients can be humans, non-human primates, cats, dogs, pigs, cattle, horses, camels, llamas, goats, sheep, rodents (*e.g.*, mice, rabbits, rats, or guinea pigs) and other mammals that can benefit from the therapies disclosed herein. Exemplary human patients can be male and/or female.

“Patient in need thereof” or “subject in need thereof” is referred to herein as a patient
15 diagnosed with, at risk or having, predetermined to have, or suspected of having a disease or disorder.

The terms “pathogenic mutation”, “pathogenic variant”, “disease causing mutation”, “disease causing variant”, “deleterious mutation”, or “predisposing mutation” refers to a genetic alteration or mutation that increases an individual’s susceptibility or predisposition to a certain
20 disease or disorder. In some embodiments, the pathogenic mutation comprises at least one wild-type amino acid substituted by at least one pathogenic amino acid in a protein encoded by a gene.

The terms “protein”, “peptide”, “polypeptide”, and their grammatical equivalents are used interchangeably herein, and refer to a polymer of amino acid residues linked together by
25 peptide (amide) bonds. A protein, peptide, or polypeptide can be naturally occurring, recombinant, or synthetic, or any combination thereof.

The term “fusion protein” as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins.

The term “recombinant” as used herein in the context of proteins or nucleic acids refers to
30 proteins or nucleic acids that do not occur in nature, but are the product of human engineering. For example, in some embodiments, a recombinant protein or nucleic acid molecule comprises an amino acid or nucleotide sequence that comprises at least one, at least two, at least three, at least four, at least five, at least six, or at least seven mutations as compared to any naturally occurring sequence.

By “reduces” is meant a negative alteration of at least 10%, 25%, 50%, 75%, or 100%.

By “reference” is meant a standard or control condition. In one embodiment, the reference is a wild-type or healthy cell. In other embodiments and without limitation, a reference is an untreated cell that is not subjected to a test condition, or is subjected to placebo or normal saline, medium, buffer, and/or a control vector that does not harbor a polynucleotide of interest. In some embodiments, a reference is a subject that has not been administered a treatment. In some embodiments, a reference is a subject that has not been administered a composition of the invention. In some embodiments, a reference is a subject that has not been administered a cell of the invention.

A “reference sequence” is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset of or the entirety of a specified sequence; for example, a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence. For polypeptides, the length of the reference polypeptide sequence will generally be at least about 16 amino acids, at least about 20 amino acids, at least about 25 amino acids, about 35 amino acids, about 50 amino acids, or about 100 amino acids. For nucleic acids, the length of the reference nucleic acid sequence will generally be at least about 50 nucleotides, at least about 60 nucleotides, at least about 75 nucleotides, about 100 nucleotides or about 300 nucleotides or any integer thereabout or therebetween. In some embodiments, a reference sequence is a wild-type or naturally occurring sequence of a protein or polypeptide of interest. In other embodiments, a reference sequence is a polynucleotide sequence encoding a wild-type or naturally occurring protein or polynucleotide. In some embodiments, a reference sequence may be a nonmutated or normal sequence.

The term “RNA-programmable nuclease,” and “RNA-guided nuclease” are used with (*e.g.*, binds or associates with) one or more RNA(s) that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). In some embodiments, the RNA-programmable nuclease is the (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (CsnI) from *Streptococcus pyogenes*.

The term “single nucleotide polymorphism (SNP)” is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population (*e.g.*, > 1%).

By “specifically binds” is meant a nucleic acid molecule, polypeptide, polypeptide/polynucleotide complex, compound, or molecule that recognizes and binds a

polypeptide and/or nucleic acid molecule of the invention, but which does not substantially recognize and bind other molecules in a sample, for example, a biological sample.

By "substantially identical" is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence. In one embodiment, a reference sequence is a wild-type amino acid or nucleic acid sequence. In another embodiment, a reference sequence is any one of the amino acid or nucleic acid sequences described herein. In one embodiment, such a sequence is at least 60%, 80%, 85%, 90%, 95% or even 99% identical at the amino acid level or nucleic acid level to the sequence used for comparison.

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence.

COBALT is used, for example, with the following parameters:

- a) alignment parameters: Gap penalties-11,-1 and End-Gap penalties-5,-1,
- b) CDD Parameters: Use RPS BLAST on; Blast E-value 0.003; Find Conserved columns and Recompute on, and
- c) Query Clustering Parameters: Use query clusters on; Word Size 4; Max cluster distance 0.8; Alphabet Regular.

EMBOSS Needle is used, for example, with the following parameters:

- a) Matrix: BLOSUM62;
- b) GAP OPEN: 10;
- c) GAP EXTEND: 0.5;
- d) OUTPUT FORMAT: pair;
- e) END GAP PENALTY: false;
- f) END GAP OPEN: 10; and
- g) END GAP EXTEND: 0.5.

Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will

typically exhibit substantial identity. Polynucleotides having "substantial identity" to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having "substantial identity" to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. By "hybridize" is meant pair to form a double-stranded molecule between complementary polynucleotide sequences (*e.g.*, a gene described herein), or portions thereof, under various conditions of stringency. (See, *e.g.*, Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507).

For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and more preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, *e.g.*, formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and more preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, *e.g.*, sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

For most applications, washing steps that follow hybridization will also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM

NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include a temperature of at least about 25° C, more preferably of at least about 42° C, and even more preferably of at least about 68° C. In an embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In another embodiment, wash steps will occur at 42 C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art. Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel *et al.* (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

By "split" is meant divided into two or more fragments.

A "split Cas9 protein" or "split Cas9" refers to a Cas9 protein that is provided as an N-terminal fragment and a C-terminal fragment encoded by two separate nucleotide sequences. The polypeptides corresponding to the N-terminal portion and the C-terminal portion of the Cas9 protein may be spliced to form a "reconstituted" Cas9 protein.

The term "target site" refers to a sequence within a nucleic acid molecule that is deaminated by a deaminase (*e.g.*, cytidine or adenine deaminase) or a fusion protein comprising a deaminase (*e.g.*, a dCas9-adenosine deaminase fusion protein or a base editor disclosed herein). In embodiments, the fusion protein comprises ABE8. In an embodiment, the fusion protein comprises ABE8.8.

As used herein, the terms "treat," "treating," "treatment," and the like refer to reducing or ameliorating a disorder and/or symptoms associated therewith or obtaining a desired pharmacologic and/or physiologic effect. It will be appreciated that, although not precluded, treating a disorder or condition does not require that the disorder, condition or symptoms associated therewith be completely eliminated. In some embodiments, the effect is therapeutic, *i.e.*, without limitation, the effect partially or completely reduces, diminishes, abrogates, abates, alleviates, decreases the intensity of, or cures a disease and/or adverse symptom attributable to the disease. In some embodiments, the effect is preventative, *i.e.*, the effect protects or prevents an occurrence or reoccurrence of a disease or condition. To this end, the presently disclosed methods comprise administering a therapeutically effective amount of a compositions as described herein.

By “uracil glycosylase inhibitor” or “UGI” is meant an agent that inhibits the uracil-excision repair system. Base editors comprising a cytidine deaminase convert cytosine to uracil, which is then converted to thymine through DNA replication or repair. Including an inhibitor of uracil DNA glycosylase (UGI) in the base editor prevents base excision repair which changes the U back to a C. An exemplary UGI comprises an amino acid sequence as follows:

5 >sp|P14739|IUNGI__BPPB2 Uracil-DNA glycosylase inhibitor
MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
APEYKPWALVIQDSNGENKIKML (SEQ ID NO: 124).

Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting 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, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50.

The recitation of a listing of chemical groups in any definition of a variable herein includes definitions of that variable as any single group or combination of listed groups. The recitation of an embodiment for a variable or aspect herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

All terms are intended to be understood as they would be understood by a person skilled in the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the disclosure pertains.

In this application, the use of the singular includes the plural unless specifically stated otherwise. It must be noted that, as used in the specification, the singular forms “a,” “an” and “the” include plural referents unless the context clearly dictates otherwise. In this application, the use of “or” means “and/or” unless stated otherwise. Furthermore, use of the term “including” as well as other forms, such as “include,” “includes,” and “included,” is not limiting.

As used in this specification and claim(s), the words “comprising” (and any form of comprising, such as “comprise” and “comprises”), “having” (and any form of having, such as “have” and “has”), “including” (and any form of including, such as “includes” and “include”) or “containing” (and any form of containing, such as “contains” and “contain”) are inclusive or open-ended and do not exclude additional, unrecited elements or method steps. It is contemplated that any embodiment discussed in this specification can be implemented with respect to any method or composition of the present disclosure, and vice versa. Furthermore, compositions of the present disclosure can be used to achieve methods of the present disclosure.

The term “about” or “approximately” means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, *i.e.*, the limitations of the measurement system. For example, “about” can mean within 1 or more than 1 standard deviation, per the practice in the art. Alternatively, “about” can mean a range of up to 20%, up to 10%, up to 5%, or up to 1% of a given value. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, *e.g.*, within 5-fold, within 2-fold of a value. Where particular values are described in the application and claims, unless otherwise stated, the term “about” means within an acceptable error range for the particular value should be assumed.

Reference in the specification to “some embodiments,” “an embodiment,” “one embodiment” or “other embodiments” means that a particular feature, structure, or characteristic described in connection with the embodiments is included in at least some embodiments, but not necessarily all embodiments, of the present disclosures.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts plasmids containing polynucleotides encoding an adenosine deaminase, *e.g.*, TadA tRNA deaminase, and a Cas9 protein, *e.g.*, dCas9, and gRNA as used in ABE nucleobase editing in mammalian cells.

FIG. 2 is a schematic depicting an exemplary study design workflow involving electroporation (EP) of human CD34⁺ cells for engraftment in a mouse for proof of concept experiments. As shown in the schematic, the study design includes the following procedure: thaw the cells; culture the cells for 2 days in cell culture flasks (or plates or conical tubes); EP buffer exchange and wash the cells; electroporate the cells, *e.g.*, with mRNA encoding ABE base editor and gRNA; 20 minute, 37°C EP incubation; 2 day culture of the cells in cell culture flasks (or plates or conical tubes); cryopreservation of the cells; and engraftment of the cells into a mouse model.

FIGs. 3A and 3B are bar graphs. FIG. 3A shows the percent of A → G (A → G%) edited CD34⁺ cells from two donors (donor 1, donor 2) using ABE 8.8 (50nM), ABE 8.8 (20nM) and ABE 7.10 (50nM) adenosine nucleobase editing systems. FIG. 3B shows viability of edited cells as a percentage (%) of total edited cells at 48 hours post-electroporation (EP). In the sets of bar graphs shown, the leftmost bar (at baseline) represents unedited cells; the second bar from the left represents cells treated with 50 nM ABE8.8; the third bar from the left represents cells treated with 20 nM ABE8.8; and the fourth bar from the left represents cells treated with 50 nM ABE7.10.

FIGs. 4A and 4B are bar graphs. FIG. 4A (cells from donor 1) and FIG. 4B (cells from donor 2) present data showing hCD45⁺ cells as a percentage (%) of total CD45⁺ cells in mouse bone marrow (BM) at the indicated periods after injection (engraftment). Indicated mice groups received unedited cells, and hCD34⁺ cells edited using ABE 8.8 (50nM), ABE 8.8 (20nM) or ABE 7.10 (50nM) ABE nucleobase editing systems as indicated. In the sets of bar graphs shown, the leftmost set of bars represents unedited cells; the set of bars second from the left represents cells treated with 50 nM ABE8.8; the set of bars third from the left represents cells treated with 20 nM ABE8.8; and the set of bars fourth from the left represents cells treated with 50 nM ABE7.10.

FIGs. 5Ai, 5Aii, and 5B-5E are bar graphs. FIGs. 5Ai and 5Aii present data showing A→G% edited cells in mouse bone marrow, at the time of injection (In), at 8 weeks and at 16 weeks post-injection in the indicated groups of engrafted mice. Treated (edited) hCD34⁺ cells were edited using ABE 8.8 (50nM), ABE 8.8 (20nM) and ABE 7.10 (50nM) ABE nucleobase editing systems. FIGs. 5B (cells from donor 1) and 5C (cells from donor 1) present results from sorted cell populations of the indicated mice groups at 16 weeks post-injection (dose). Sorting was performed using flow cytometry. CD34⁺ cells were further sorted for Lin 34 and Gly A markers. FIGs. 5D and 5E show results of expression levels of gamma globin protein in engrafted mice at 16 weeks after injection with edited donor cells. (FIG. 5D- results from recipients of donor 1 cells; FIG. 5E- results from recipients of donor 2 cells). n=3-6 mice per group. In FIGs. 5Ai – 5C, the leftmost set of bars represents unedited cells (baseline); the set of bars second from the left represents cells treated with 50 nM ABE8.8; the set of bars third from the left represents cells treated with 20 nM ABE8.8; and the set of bars fourth from the left represents cells treated with 50 nM ABE7.10. In FIGs. 5D and 5E, the leftmost bar represents unedited cells; the bar second from the left represents cells treated with 50 nM ABE8.8; the bar third from the left represents cells treated with 20 nM ABE8.8; and the bar fourth from the left represents cells treated with 50 nM ABE7.10.

FIGs. 6A-6C are bar graphs presenting data collected at 16-weeks following engraftment of human CD34⁺ cells from a single healthy donor into NOD.Cg-Kit^{W-41J} Tyr⁺ Prkdc^{scid} Il2rg^{tm1Wjl}/ThomJ (NBSGW) mouse bone marrow (N=6 for chimerism and editing, N=5 for induction). FIG. 6A is a bar graph comparing the percentage of edited or unedited CD34⁺ cells that were engrafted. FIG. 6B is a bar graph showing the efficiency of base editing. FIG 6C is a bar graph showing expression levels of gamma globin in edited and unedited cells.

FIGs. 7A-7B present bar graphs and stacked bar graphs relating to CD34⁺ cells from a sickle cell disease (SCD) patient that were transfected with ABE8.8 mRNA and sgRNA using

electroporation. FIG. 7A is a bar graph showing percent cells edited at 48 hours and at 14 days post-electroporation. FIG. 7B is a stacked bar graph showing the different edits (including bystander editing) contained within each indicated cell population at the indicated time points.

FIGs. 8A-8D are plots and bar graphs relating to globin levels analyzed on day 18 post-differentiation for edited sickle cell disease (SCD)-CD34⁺ cells that were differentiated to erythroid cells. FIGs. 8A and 8B are plots showing peaks corresponding to identified globin polypeptides. FIG. 8C is a bar graph presenting percent change in expression of gamma globin (corresponding to HbF levels) in the edited cells, and FIG. 8D is a bar graph presenting a concurrent percent reduction in S globin in the edited cells. In the bar graphs, the leftmost bar represents unedited cells, and the rightmost bar represents base-edited cells. The y-axis of FIG. 8C reflects $\gamma/(\gamma+S+\Delta)*100$ and the y-axis of FIG. 8D reflects $S/(\gamma+S+\Delta)*100$.

FIGs. 9A-9C present a schematic depiction and bar graphs. FIG. 9A depicts the experimental design and treatment conditions used in the study described in Example 5 herein. FIGs. 9B and 9C shows bar graphs and results demonstrating long term (16 weeks) engraftment and *HBG1/2* gene promoter base editing retention in NBSGW mice (NBSGW mouse model). FIG. 9B shows % hCD45⁺/(hCD45⁺+mCD45⁺) human cell chimerism in bone marrow (BM). FIG. 9C shows % *HBG1/2* promoter base editing in bulk BM cells. For the sets of bars in the graphs, the leftmost bar represents unedited cells; the second bar from the left represents cells treated with 1 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the third bar from the left represents cells treated with 3 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the fourth bar from the left represents cells treated with 10 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the fifth bar from the left represents cells treated with 30 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the sixth bar from the left represents cells treated with 10 nM ABE mRNA (Lot R34) + 3000 nM gRNA; and the seventh bar from the left represents cells treated with 10 nM ABE mRNA (Lot R34) + 3000 nM gRNA. In the experiments, the ABE mRNA is ABE8.8 mRNA and the gRNA is *HBG1/2* gRNA. ABE8.8 encoding mRNA, MRNA288 (produced by CRO); ABE8.8 encoding mRNA, Lot R34 (research grade); and pilot *HBG1/2* gRNA (GMP-like gRNA) are as described in Example 5. The legend provided in FIG. 9C applies to FIGs. 9B and 9C.

FIGs. 10A-10D present bar graphs demonstrating that *HBG1/2* gene promoter-edited human stem cells (HSCs) display long term, multi-lineage (e.g., erythroid, myeloid, lymphoid) hematopoietic reconstitution in NBSGW mice (the NBSGW mouse model). In the bar graphs, the leftmost bar represents unedited cells; the second bar from the left represents cells treated with 1 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the third bar from the left represents

cells treated with 3 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the fourth bar from the left represents cells treated with 10 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the fifth bar from the left represents cells treated with 30 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the sixth bar from the left represents cells treated with 10 nM ABE mRNA (Lot R34) + 3000 nM gRNA; and the seventh bar from the left represents cells treated with 10 nM ABE mRNA (Lot R34) + 3000 nM gRNA. In the experiments, the ABE mRNA is ABE8.8 mRNA and the gRNA is *HBG1/2* gRNA. The legend to the right of FIG. 10B applies to FIGs. 10A-10D.

FIG. 11 presents a bar graph showing results that demonstrate long term human hematopoietic, multi-lineage reconstitution in NBSGW mice at 16 weeks post electroporation of the cells with the base editor (ABE mRNA) and gRNA. Percent (%) *HBG1/2* promoter base editing in human hematopoietic cell subpopulations was assessed. In the figure, the leftmost series of 5 bars (i.e., Bulk BM, CD15+, CD19+, Lin-CD34+, BlyA+) represent unedited cells; the second series of 5 bars from the left represents cells treated with 1 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the third series of 5 bars from the left represents cells treated with 3 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the fourth series of 5 bars from the left represents cells treated with 10 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the fifth series of 5 bars from the left represents cells treated with 30 nM ABE mRNA (MRNA288) + 3000 nM gRNA; the sixth series of 5 bars from the left represents cells treated with 10 nM ABE mRNA (Lot R34) + 3000 nM gRNA; and the seventh series of 5 bars from the left represents cells treated with 10 nM ABE mRNA (Lot R34) + 3000 nM gRNA. In the experiments, the ABE mRNA is ABE8.8 mRNA and the gRNA is *HBG1/2* gRNA.

FIGs. 12A and 12B present bar graphs demonstrating that *HBG1/2* gene promoter base editing is maintained long term (16 weeks) post-engraftment with elevated gamma globin (γ -globin) levels in NBSGW mice. In FIG. 12A, the % *HBG1/2* promoter base editing in bulk BM cells at 16 weeks was assessed. In FIG. 12B, the % γ -globin protein levels in flow cytometry-sorted BM-derived human erythroid cells were assessed. The cell treatments represented in the bar graphs shown are the same as those described above for FIGs. 10A-D. In the experiments, the ABE mRNA is ABE8.8 mRNA and the gRNA is *HBG1/2* gRNA. The legend to the right of FIG. 12B applies to FIGs. 12A and 12B.

FIGs. 13A and 13B present bar graphs demonstrating that long term engraftment and *HBG1/2* gene promoter base editing are retained in irradiated NSG (irrNSG) mice. FIG. 13A shows % hCD45/(hCD45+mCD45+) human cell chimerism in bone marrow (BM). FIG. 13B shows % *HBG1/2* promoter base editing in bulk BM cells. The bar graphs and sets of bar graphs shown in the figure represent the cell treatments as described above for FIGs. 9B and 9C. In the

experiments, the ABE mRNA is ABE8.8 mRNA and the gRNA is *HBG1/2* gRNA. The legend to the right of FIG. 13B applies to FIGs. 13A and 13B.

FIGs. 14A-14C present bar graphs demonstrating that *HBG1/2* gene promoter-edited HPSCs displayed long term, multi-lineage (e.g., erythroid, myeloid, lymphoid) hematopoietic reconstitution in irrNSG mice. Shown are human progenitor stem cells (HSPCs), (FIG. 14A); human myeloid cells (FIG. 14B) and human lymphoid cells (FIG. 14C). The bar graphs in the figures represent the cell treatments as described above, e.g., in FIGs. 10A-10D. In the experiments, the ABE mRNA is ABE8.8 mRNA and the gRNA is *HBG1/2* gRNA. The legend to the right of FIG. 14C applies to FIGs. 14A-14C.

FIG. 15 presents a bar graph showing % *HBG1/2* promoter base editing in bulk BM cells assessed from NBSGW mice and from irrNSG mice at 16 weeks. FIG. 15 shows that comparable *HBG1/2* gene promoter base editing was retained long term (16 weeks) in NBSGW mice and in irrNSG mice as determined by analysis of bulk bone marrow (BM) cells obtained from the mice. The bar graphs in the figure represent the cell treatments as described above, e.g. for FIGs. 13A and 13B. In the experiment, the ABE mRNA is ABE8.8 mRNA and the gRNA is *HBG1/2* gRNA.

FIG. 16 presents a schematic and bar graphs showing the results of a long term engraftment study using the NBSGW mouse model and including a secondary engraftment component (16 weeks + 8 weeks) of donor cells (HPSCs). The leftmost graph shows the percent human cell chimerism (hCD45+/(hCD45+mCD45+)) in engrafted mice at 16 weeks + 8 weeks post dose; the middle graph shows the % LIN-hCD34+ cells in engrafted mice at 16 weeks + 8 weeks post dose; and the rightmost graph shows the % base editing (A→G) of bone marrow cells assessed from engrafted mice at 16 + 8 weeks post dose. In each of the bar graphs, the leftmost bar represents unedited HPSCs used for engraftment in NBSGW mice; the middle bar represents base-edited HPSCs electroporated using small scale electroporation (OC-400) used for engraftment into NBSGW mice; and the rightmost bar represents base-edited HPSCs electroporated using large scale electroporation (CL1.1) used for engraftment into NBSGW mice. In the experiments, the ABE mRNA is ABE8.8 mRNA and the gRNA is *HBG1/2* gRNA.

FIGs. 17A and 17B present bar graphs showing assessments of human bone marrow (BM) cell chimerism (hCD45+/(hCD45+mCD45+)), (FIG. 17A) and percent base editing in BM cells (FIG. 17B) at 13 weeks post-dose of ABE8.8 mRNA conducted during a dose titration study in the NBSGW mouse model (Example 5). Mean +/- SEM: N=1 at 13 weeks; N=3 at 8 weeks; N=1 at 0 weeks. For the bar graph or sets of bar graphs shown in the figure, the leftmost bar represents cells treated with 10 nM ABE8.8 mRNA (R34) + 3000 nM gRNA; the rightmost

bar represents cells treated with 30 nM mRNA (288) + 3000 nM gRNA. In the experiments, the ABE mRNA is ABE8.8 mRNA and the gRNA is *HBG1/2* gRNA.

FIGs. 18A and 18B present graphs showing apoptosis / cell viability as determined by flow cytometry analyses performed on freshly thawed donor cells after cryopreservation. The results show the percentage of live, dead and apoptotic CD34+ cells at 24hr isolation and 48+hr isolation versus control PBMCs, as described in Example 5. Also shown are the locations of live, dead and apoptotic cells on the graphs generated using an apoptosis detection kit and assessed by flow cytometry and antibody reagents directed against 7-AAD and Annexin V.

FIGs. 19A-19C present flow cytometry graphs showing the results of the assessment of apoptosis / cell viability; measurement of apoptosis; and lineage analysis of donor CD34+ cells. FIG. 19A shows apoptosis / cell viability as determined by flow cytometry analyses performed on "Pre-EP" CD34+ cell samples at 24hr isolation and 48+hr isolation, as described in Example 5. Cells were in culture for 48+ hours post thawing after cryopreservation (FIG. 19A). FIG. 19B shows the measurement of apoptosis determined by flow cytometry analyses performed on different groups of "Post-EP" CD34+ cell samples at 24hr isolation and 48+hr isolation (unedited versus base-edited CD34+ cells), as described in Example 5. FIG. 19C shows flow cytometry results of a lineage analysis performed on freshly thawed donor cells through 24hr post electroporation using antibody reagents specific for the lineage markers analyzed. The SSC-A ordinate values are in 50k increments ranging from 0 to 250k; the CD15 ordinate values and the CD34 and CD19 abscissa values range logarithmically from 0 to 10^5 .

FIGs. 20A and 20B show bar graphs which present the results of assessing percent cell viability and percent base editing (A to G) in unedited and base-edited cells. The bar graphs in FIG. 20A shows viability of the cells at pre-electroporation (Pre-EP), and at 24, 48 and 72 hours post electroporation, as described in Example 6. The bar graphs in FIG. 20B shows the percent of base editing achieved in the base-edited, transplanted cells at the indicated time periods. For the bar graph or sets of bar graphs shown in the FIGs. 20A and 20B, the leftmost bar represents unedited cells that were collected after 48+ hr pre-enrichment (48+ hr Pre-Enrich); the second bar from the left represents base-edited cells electroporated using the small scale OC-400 cell electroporation cartridge and treated as shown (48+ hr Pre-Enrich); the third bar from the left represents unedited cells that were collected after 24 hours pre-enrichment (24 hr Pre-Enrich); the fourth bar from the left represents base-edited cells electroporated using the small scale OC-400 cell electroporation cartridge (24 hr Pre-Enrich); and the rightmost bar represents edited cells electroporated using the large scale CL1.1 cell electroporation cartridge (24 hr Pre-Enrich). The "24 hr or 48+ hr pre-enrichment" of unedited and base-edited cells refer to the time period

between isolating a blood sample (PBMCs) from a donor and enriching for CD34+ cells in the sample, as described in Example 6 herein. The legend to the right in FIG. 20B applies to FIGs. 20A and 20B.

FIGs. 21A and 21B present a bar graph and a graph depicting a cell growth curve. The bar graph shown in FIG. 21A presents the percentage of enucleated cells (%DAPI-/NucRed-) after thawing. The treatment conditions of cells represented by the bar graphs are shown along the abscissa of FIG. 21A. The graph in FIG. 21B presents the 'theoretical total cells' assessed on the indicated day after thawing. For FIGs. 21A and 21B, mean +/- SEM; N=3.

FIGs. 22A and 22B show bar graphs which present the results of assessing the amount of gamma globin induction (gamma/beta-like) and the number of colonies (CFUs) detected in unedited or base edited cells. FIG. 22A shows the amount of gamma globin induction (gamma/beta-like) produced or expressed by unedited cells under the pre-enrichment conditions shown and by base-edited cells subjected to small or large scale electroporation and the pre-enrichment conditions shown (mean +/- SEM; N=3). FIG. 22B shows the number of colonies of the types shown (BFU-E, CFU-GM, and CFU-GEMM) produced by either unedited cells under the pre-enrichment conditions shown or by base-edited cells subjected to small scale (OC400) or large scale (CL1.1) electroporation and the pre-enrichment conditions shown (mean +/- SEM; N=2).

FIGs. 23A and 23B show bar graphs which present the results of assessing human donor cell chimerism in mouse bone marrow (BM) and percent base editing (A to G) in animals at 8 weeks post dosing with unedited or base-edited donor CD34+ cells. FIG. 23A shows the percentage of human donor cell chimerism (hCD45+/(hCD45+ + mCD45+)) in mouse bone marrow (BM) assessed at 8 weeks after mice were dosed (transplanted) with unedited or base-edited CD34+ cells that had been electroporated under small scale (OC400) or large scale (CL1.1) electroporation conditions and subjected to either 24 or 48+ hour pre-enrichment conditions. FIG. 23B shows the percentage of base editing (A to G) in the cell materials shown on the x-axis (input; bulk BM; CD34+/LIN-; and whole blood) at 8 weeks after administration/transplant into animals (n=3 at 8 weeks). The bars and sets of bars in the graphs represent the cells and conditions as described for FIGs. 20A and 20B.

FIGs. 24A-24D show bar graphs which present the results of assessing human donor cell chimerism in mouse bone marrow (BM), percent hCD15+ cells, percent GlyA+ cells and percent human CD34+ cells in animals at 8 and 16 weeks post dosing with unedited or base-edited donor CD34+ cells. FIG. 24A shows the percentage of human donor cell chimerism (hCD45+/(hCD45+ + mCD45+)) in mouse bone marrow (BM) detected at 16 weeks after mice

were dosed (transplanted) with unedited or based-edited CD34+ cells that had been electroporated under small scale (OC400) or large scale (CL1.1) electroporation and subjected to either 24 or 48+ hour pre-enrichment conditions. FIG. 24B shows the percentage of hCD15+ cells detected in mice at 8 weeks after mice were dosed (transplanted) with unedited or based-edited CD34+ cells that had been electroporated under small scale (OC400) or large scale (CL1.1) electroporation conditions and subjected to either 24 or 48+ hour pre-enrichment conditions. FIG. 24C shows the percentage of GlyA+ cells detected in mice at 16 weeks after mice were dosed (transplanted) with unedited or based-edited CD34+ cells that had been electroporated under small scale (OC400) or large scale (CL1.1) electroporation conditions and subjected to either 24 or 48+ hour pre-enrichment conditions. FIG. 24D shows the percentage of hCD34+ cells (hCD34+/hCD45+ cells) detected in mice at 16 weeks after mice were dosed (transplanted) with unedited or based-edited CD34+ cells that had been electroporated under small scale (OC400) or large scale (CL1.1) electroporation conditions and subjected to either 24 or 48+ hour pre-enrichment conditions. The bars and sets of bars in the graphs of FIGs. 24A-24D represent the cells and conditions as described above (mean +/- SEM, n=4-5).

FIGs. 25A-25C show bar graphs which present the results of assessing at 8 and 16 weeks post dosing chimerism, base editing and globin reactivation in unedited and base-edited cells administered to animals. FIG. 25A shows the percentage of human donor cell chimerism (hCD45+/(hCD45+ + mCD45+)) in mouse bone marrow (BM) assessed at 8 weeks and 16 weeks after mice were dosed (transplanted) with unedited or based-edited CD34+ cells that had been electroporated under small scale (OC400) or large scale (CL1.1) electroporation conditions and subjected to either 24 or 48+ hour pre-enrichment conditions. FIG. 25B shows the percent of base editing at 8 and 16 weeks as assessed in unedited cells and base-edited cells as described. FIG. 25C shows the percent of gamma/beta globin-like fetal globin reactivation in the described unedited cells and base-edited cells at 16 weeks after dosing in animals. In FIGs. 25A-C, the bar graphs or sets of bar graphs represent the cells and conditions that are as described for the above drawings (e.g., the leftmost bar or the leftmost bar in the set of bars represents unedited cells, 48+ hours; the second to the left bar or the second bar to the left in the set of bars represents edited cells, 48+ hour, OC-400; the third to the left bar or the third bar to the left in the set of bars represents unedited cells, 24 hours; the fourth to the left bar or the fourth bar to the left in the set of bars represents edited cells, 24 hours, OC-400; and the fifth to the left bar or the fifth bar to the left in the set of bars represents edited cells, 24 hours, CL1.1).

FIG. 26 presents sets of bar graphs demonstrating the percent base editing in the cell subpopulations having the phenotypes and lineages shown (i.e., GlyA+, CD15+, CD19+, LIN-

CD34+, BM) as assessed at 16 weeks post dosing of unedited or base edited cells into animals. The leftmost set of bars represents the % base editing in subpopulations of cells detected at 16 weeks after transplanting animals with unedited CD34+ cells (CD34+ cells isolated 24 hours from the time of collecting the human donor blood sample, "24 hr"). The middle set of bars represents the % base editing in subpopulations of cells detected at 16 weeks after transplanting animals with base-edited CD34+ cells subjected to small scale electroporation (OC-400) (CD34+ cells isolated 24 hours from the time of collecting the human donor blood sample, "24 hr"). The rightmost set of bars represents the % base editing in subpopulations of cells detected at 16 weeks after transplanting animals with base-edited CD34+ cells subjected to large scale electroporation (CL1.1) (CD34+ cells isolated 24 hours from the time of collecting the human donor blood sample, "24 hr").

FIG. 27 presents a schematic illustration of a target site for editing the HBG1/2 locus. In the figure, the sequences from top to bottom are SEQ ID NOs: 289 and 290.

DETAILED DESCRIPTION OF THE INVENTION

The invention features compositions containing novel adenine base editors (e.g., ABE8) that have increased efficiency and methods of using the compositions to generate modifications at target sites within nucleic acid molecules, in particular, with the treatment of hemoglobinopathies, such as sickle cell disease (SCD), anemia, thalassemia, etc..

Sickle cell disease (SCD) is a monogenic disorder affecting beta globin function, which leads to severe anemia and progressive multiple organ failure. A promising treatment for sickle cell disease (SCD) is the re-expression of fetal hemoglobin (HbF), which occurs naturally in individuals with hereditary persistence of fetal hemoglobin (HPFH). High levels of HbF are sometimes a result of β -globin gene deletions or point mutations in the promoters of the HbF genes. Sickle cell disease (SCD) patients harboring natural genetic variations in the human gamma globin gene promoters, *HBG1* and *HBG2* (*HBG1/2*), display elevated HbF levels and are typically afflicted with significantly fewer complications from sickle cell disease (SCD).

Featured herein are compositions and methods for long-term engraftment treatment with modified cells, e.g. hematopoietic cells modified, e.g., base-edited, with base editor systems as described herein. For example, a single nucleobase polymorphism (SNP) may be edited in hematopoietic stem cells or progenitor cells (HSPCs), e.g. human CD34⁺ cells, for engraftment in order to generate a desired treatment and/or phenotype. In some embodiments, base edited human CD34⁺ cells (donor cells) are engrafted into a recipient having sickle cell disease for the treatment of SCD. Base editing modifications may correct a mutation associated with sickle cell

disease (SCD), or may create one or more nucleobase modifications that ameliorate sickle cell disease (SCD) symptoms. In some embodiments, modified human CD34⁺ hematopoietic stem/progenitor cells (HSPCs) are introduced into (e.g., engrafted) a subject in need thereof to generate increased and/or persistent expression of HbF. In some embodiments, base edited
5 human CD34⁺ cells are introduced into (e.g., engrafted) a subject in need thereof for the treatment of sickle cell disease (SCD). In some embodiments, modified human CD34⁺ hematopoietic stem/progenitor cells (HSPCs) are introduced into (e.g., engrafted) a subject in need thereof to recreate an HPFH phenotype as a treatment for sickle cell disease (SCD).

In one aspect, the present disclosure provides nucleobase editor and base editor systems
10 with improved base editing functions that generate a high percentage of nucleobase-edited cells, which become engrafted in a subject following delivery or administration to the subject. Following introduction into a subject, such base-edited cells become grafted and perform their functions as transplanted bone marrow cells. In certain embodiments, a base editor system provided herein effects editing at a single target nucleobase with increased editing efficiency,
15 reduced off target effects, reduced indel formation, reduced bystander modifications, reduced spurious modifications, or a combination thereof.

Base Editing at the HBG1/2 Locus

In some embodiments, the adenosine base editing system targeted for editing a
20 hemoglobin gene or a regulatory element thereof provides base-edited cells that are advantageous for transplantation and engraftment in a subject in need thereof, e.g., a subject afflicted with a hemoglobinopathy, such as sickle cell disease or thalassemia. In some embodiments, the methods provide for editing human *HBG1/2* gene promoters in HSPCs. In some embodiments, the method for editing a hemoglobin gene or a regulatory subunit thereof is
25 an improved method over currently available methods for gene editing and for generating base edited cells that are suitable and beneficial for transplantation and engraftment. In some embodiments, the adenosine base editing system provided herein for editing a hemoglobin gene or a regulatory subunit thereof comprise one or more, or a combination of two or more, of the following advantages: higher editing efficiency; higher fidelity and significantly lower off-target
30 editing events; higher survival of edited cells; higher persistence of edited cells *in vitro*; higher survival and persistence of edited cells *in vivo*; higher engraftment potential; higher ability to differentiate into erythropoietic lineage; higher proliferation capability *in vitro*, higher proliferation capability *in vivo*, higher expression of HbF; and higher reduction in a defective globin gene expression such as HbS, when compared to previously reported or existing base

editing systems. In embodiments, the higher expression of HbF compensates for a hemoglobin deficiency in a subject. In embodiments, the hemoglobin deficiency is alpha thalassemia or beta thalassemia. Thalassemias are blood disorders characterized by decreased hemoglobin production. Thalassemia typically is associated with a deficiency in alpha and/or beta globin production in a subject.

In one aspect, the present disclosure provides a method for editing human *HBG1/2* gene promoters in HSPCs for long-term engraftment potential in a subject having sickle cell disease (SCD). FIG. 27 illustrates a target sequence for editing human *HBG1/2* gene promoters. In embodiments, editing the human *HBG1/2* gene promoters disrupts and/or eliminates binding of BCL11A in the promoter region. In embodiments, editing the *HBG1/2* gene promoter is associated with de-repression of the *HBG1/2* gene. In embodiments, editing the *HBG1/2* gene promoter abolishes, disrupts, or reduces BCL11A binding in the promoter region of the *HBG1/2* gene. In embodiments, editing of human *HBG1/2* gene results in a nucleobase change at position -144 relative to the canonical transcription start site of the *HBG1/2* gene. In one embodiment, the present disclosure provides a method for editing human *HBG1/2* gene promoters in HSPCs using an improved adenosine base editing system (ABE) for long-term engraftment potential in a subject having sickle cell disease (SCD). In some embodiments, several improvements are incorporated in the present disclosure for an improved adenosine base editing system that is targeted for editing a hemoglobin gene or a regulatory subunit thereof, such as, for example, editing human *HBG1/2* gene promoters in HSPCs.

HbB Gene Editing

In one aspect, the methods described herein are useful in HbB gene editing. In particular, the compositions and methods of the invention are useful for the treatment of sickle cell disease (SCD), which is caused by a Glu to Val mutation at the sixth amino acid of the β -globin protein encoded by the *HbB* gene. Despite many developments to date in the field of gene editing, precise correction of the diseased *HbB* gene to revert Val to Glu remains elusive, and is presently not achievable using either CRISPR/Cas nuclease or CRISPR/Cas base editing approaches.

Genome editing of the *HbB* gene to replace the affected nucleotide using a CRISPR/Cas nuclease approach requires cleavage of genomic DNA. However, cleavage of genomic DNA carries an increased risk of generating base insertions/deletions (indels), which have the potential to cause unintended and undesirable consequences, including generating premature stop codons, altering the codon reading frame, etc. Furthermore, generating double-stranded breaks at the beta globin (β -globin) locus has the potential to radically alter the locus through recombination

events. The beta-globin locus contains a cluster of globin genes having sequence identity to one another. Because of the structure of the beta-globin locus, recombination repair of a double-stranded break within the locus has the potential to result in gene loss of intervening sequences between globin genes, for example between gamma- and beta-globin genes. Unintended alterations to the locus also carry a risk of causing thalassemia. CRISPR/Cas base editing approaches hold promise in that they have the ability to generate precise alterations at the nucleobase level. However, precise correction of Val - Glu (GTG - GAG) requires a T•A to A•T transversion editor, which is not presently known to exist.

Additionally, the specificity of CRISPR/Cas base editing is due in part to a limited window of editable nucleotides created by R-loop formation upon CRISPR/Cas binding to DNA. Thus, CRISPR/Cas targeting must occur at or near the sickle cell site to allow base editing to be possible, and there may be additional sequence requirements for optimal editing within the window. One requirement for CRISPR/Cas targeting is the presence of a protospacer-adjacent motif (PAM) flanking the site to be targeted. For example, many base editors are based on SpCas9 which requires an NGG PAM. Even assuming hypothetically that a T•A to A•T transversion were possible, no NGG PAM exists that would place the target “A” at a desirable position for such an SpCas9 base editor. Although many new CRISPR/Cas proteins have been discovered or generated that expand the collection of available PAMs, PAM requirements remain a limiting factor in the ability to direct CRISPR/Cas base editors to specific nucleotides at any location in the genome.

The present invention is based, at least in part, on several discoveries described herein that address the foregoing challenges for providing a genome editing approach for treatment of sickle cell anemia. In one aspect, the invention is based in part on the ability to replace the valine at amino acid position 6, which causes sickle cell disease, with an alanine, to thereby generate an Hb variant (Hb Makassar) that does not generate a sickle cell phenotype. While precise correction (GTG - GAG) is not possible without a T•A to A•T transversion base editor, the studies performed herein have found that a Val - Ala (GTG - GCG) replacement (*i.e.*, the Hb Makassar variant) can be generated using an A•T to G•C base editor (adenine base editor or ABE). This was achieved in part by the development of novel base editors and novel base editing strategies, as provided herein. For example, novel ABE base editors (*i.e.*, having an adenosine deaminase domain) that utilize flanking sequences (*e.g.*, PAM sequences; zinc finger binding sequences) for optimal base editing at the sickle cell target site.

Thus, the present invention includes compositions and methods for base editing a thymidine (T) to a cytidine (C) in the codon of the sixth amino acid of a sickle cell disease

variant of the β -globin protein (Sickle HbS; E6V), thereby substituting an amino acid at position 6 of the β -globin protein for a valine (V6A or E6A) at this amino acid position. Substitution of alanine for valine at position 6 of HbS generates a β -globin protein variant that does not have a sickle cell phenotype (e.g., does not have the potential to polymerize as in the case of the pathogenic variant HbS). Accordingly, the compositions and methods of the invention are useful for the treatment of sickle cell disease (SCD).

In some embodiments, several improvements are incorporated in the present disclosure for an improved adenosine base editing system that is targeted for editing a hemoglobin gene or a regulatory subunit thereof. In some embodiments, the method provides for editing human HbB gene for generating a the Hb Makassar (E6A) variant in place of the Sickle HbS; E6V. In some embodiments, the improvements are useful in engraftment of the HbB edited hematopoietic stem cells.

In some embodiments, the target polynucleotide (DNA) sequence encodes a protein (e.g., HbB), and the gene edit is in a codon of the polynucleotide (DNA) sequence and results in a change in the amino acid encoded by the mutant codon as compared to the wild-type codon. In some embodiments, the deamination of a mutant A results in a change of the amino acid encoded by the mutant codon. In some embodiments, the deamination of the mutant C results in a change of the amino acid encoded by the mutant codon.

Guide RNA (gRNA) Sequences

To produce the gene edits described above, hematopoietic stem/progenitor cells (HSPCs) are collected from a subject and contacted with a guide RNA and a nucleobase editor polypeptide comprising a nucleic acid programmable DNA binding protein (napDNAbp) and a cytidine deaminase or adenosine deaminase. In some embodiments, multiple target sites are edited simultaneously. In some embodiments, editing the multiple target sites simultaneously comprises contacting the HSPCs with two or more gRNAs. In embodiments, the HSPCs are contacted with multiple distinct gRNAs, each targeting a different sequence. The guide RNA can be a single guide or a dual guide. In some embodiments, cells to be edited are contacted with at least one nucleic acid, wherein at least one nucleic acid encodes a guide RNA, or two or more guide RNAs, and a nucleobase editor polypeptide comprising a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase, e.g., an adenosine or a cytidine deaminase. In some embodiments, the gRNA comprises nucleotide analogs. These nucleotide analogs can inhibit degradation of the gRNA by cellular processes. An exemplary target sequence for base

editing of the *HBG1/2* promoter is CTTGACCAATAGCCTTGACAAGG-3' (SEQ ID NO: 125), wherein AGG is the PAM sequence (see FIG. 27).

In some embodiments, a guide RNA provided herein directs the base editor to effect a nucleobase substitution in an HbB gene, thereby replacing a E6V mutation with a E6A

5 substitution in a hemoglobin beta subunit encoded by the HbB gene. In some embodiments, the HbB gene comprises one or more mutations or SNPs associated with sickle cell disease, for example, a GAG-GTG substitution that results in the E6V amino acid mutation. Exemplary guide RNA sequences targeting the HbB gene include the nucleic acid sequences 5'-
gsascsUUCUCCACAGGAGUCAGGGUUUUAGAGCUAGAAAUAGCAAGUUAUUUUAA
10 GGCUAGUCCGUUAUUAACUUGAAAAAGUGGCACCGAGUCGGUGCUssususu-3' (SEQ ID NO: 126), 5'-
ascsusUCUCCACAGGAGUCAGGGUUUUAGAGCUAGAAAUAGCAAGUUAUUUUAAAG
GCUAGUCCGUUAUUAACUUGAAAAAGUGGCACCGAGUCGGUGCUssususu-3' (SEQ ID NO: 127), or 5'-
15 csususCUCCACAGGAGUCAGGGUUUUAGAGCUAGAAAUAGCAAGUUAUUUUAAAGG
CUAGUCCGUUAUUAACUUGAAAAAGUGGCACCGAGUCGGUGCUssususu-3' (SEQ ID NO: 128), wherein lowercase characters indicate 2'-O-methylated nucleobases, and "s" indicates phosphorothioates.

In some embodiments, a guide RNA provided herein directs the base editor to effect a
20 nucleobase substitution in a promoter region of a *HBG1/2* gene, thereby generating enhanced or elongated expression of hemoglobin gamma subunit and increased level of HbF. An exemplary guide RNA targeting the promoter region of an *HBG1/2* gene is the nucleic acid sequence 5'-
csususGACCAAUAGCCUUGACAGUUUUAGAGCUAGAAAUAGCAAGUUAUUUUAAAG
GCUAGUCCGUUAUUAACUUGAAAAAGUGGCACCGAGUCGGUGCUssususu-3' (SEQ ID
25 NO: 129), wherein lowercase characters indicate 2'-O-methylated nucleobases, and "s" indicates phosphorothioates.

Exemplary guide RNA spacer sequences and nucleobase changes are provided in **Table 1** below.

Table 1. Introduction of Gene Regulator Edits					
gRNA Spacer SEQ ID NO	Gene	Nucleotide change	Base Editor	gRNA Spacer Sequence	PAM
130	<i>HBG</i> 1/2	c. -198 T>C	ABE	GUGGGGAAGGGGCCCCCAAG	AGG
131	<i>HBG</i> 1/2	c. -198 T>C	ABE	AUUGAGAUAGUGUGGGGAAG	GGG
132	<i>HBG</i> 1/2	c. -198 T>C	ABE	CAUUGAGAUAGUGUGGGGAA	GGG
133	<i>HBG</i> 1/2	c. -198 T>C	ABE	GCAUUGAGAUAGUGUGGGGA	AGG
134	<i>HBG</i> 1/2	c. -198 T>C	ABE	GUGGGGAAGGGGCCCCCAAG	AGG
135	<i>HBG</i> 1/2	c. -114 ~ -102 deletion	CBE and/or ABE	GCUAUUGGUCAAGGCAAGGC	TGG
136	<i>HBG</i> 1/2	c. -114 ~ -102 deletion	CBE and/or ABE	CAAGGCUAUUGGUCAAGGCA	AGG
137	<i>HBG</i> 1/2	c. -114 ~ -102 deletion	CBE and/or ABE	CUUGUCAAGGCUAUUGGUCA	AGG
138	<i>HBG</i> 1/2	c. -114 ~ -102 deletion	CBE and/or ABE	CUUGACCAAUAGCCUUGACA	AGG
139	<i>HBG</i> 1/2	c. -114 ~ -102 deletion	CBE and/or ABE	GUUUGCCUUGUCAAGGCUAU	TGG
140	<i>HBG</i> 1/2	c. -114 ~ -102 deletion	CBE and/or ABE	UGGUCAAGUUUGCCUUGUCA	AGG
141	<i>HBG</i> 1/2	c. -198 T>C	ABE	UGGGGAAGGGGCCCCCAAGA	GGA

142	HBG 1/2	c. -198 T>C	ABE	GUGUGGGGAAGGGGCCCCCA	AGA
143	HBG 1/2	c. -175 T>C	ABE	UCAGACAGAUUUUGCAUUG	AGA
144	HBG 1/2	c. -175 T>C	ABE	UUUCAGACAGAUUUUGCAU	TGA
145	HBG 1/2	c. -114 ~ -102 deletion	CBE and/or ABE	CUUGCCUUGACCAAUAGCCU	TGA
146	HBG 1/2	c. -114 ~ -102 deletion	CBE and/or ABE	UAGCCUUGACAAGGCAAACU	TGA
147	HBG 1/2	c. -90 BCL11A binding	CBE and/or ABE	CAAACUUGACCAAUAGUCUU	AGA
148	HBG 1/2	c. -198 T>C	ABE	UGUGGGGAAGGGGCCCCCAA	GAGGAT
149	HBG 1/2	c. -202 C>T, -201 C>T, -198 T>C, - 197 C>T, -196 C>T, -195 C>G	CBE and/or ABE	GGGCCCCUCCCCACACUAU	CTCAAT
150	HBG 1/2	c. -175 T>C	ABE	CAGACAGAUUUUGCAUUGA	GATAGT
151	HBG 1/2	c. -175 T>C	ABE	UUUCAGACAGAUUUUGCAU	TGAGAT
152	HBG 1/2	c. -114 ~ -102 deletion	CBE and/or ABE	GCCUUGACAAGGCAAACUUG	ACCAAT
153	HBG 1/2	c. -114 ~ -102 deletion	CBE and/or ABE	UUGACAAGGCAAACUUGACC	AATAGT
154	HBG 1/2	c. -90 BCL11A binding	CBE and/or ABE	UGACCAAUAGUCUUAGAGUA	TCCAGT
155	HBG 1/2	c. -175 T>C	ABE	AGACAGAUUUUGCAUUGAGUA	TTT

In some embodiments, any of the fusion proteins provided herein may have a Cas9 domain that does not have nuclease activity (dCas9), or a Cas9 domain that cuts one strand of a duplexed DNA molecule, referred to as a Cas9 nickase (nCas9). Without wishing to be bound by any particular theory, the presence of the catalytic residue (e.g., H840) maintains the activity of the Cas9 to cleave the non-edited (e.g., non-methylated) strand opposite the targeted nucleobase. Mutation of the catalytic residue (e.g., D10 to A10) prevents cleavage of the edited strand containing the targeted A residue. Such Cas9 variants can generate a single-strand DNA break (nick) at a specific location based on the gRNA-defined target sequence, leading to repair of the non-edited strand, ultimately resulting in a nucleobase change on the non-edited strand.

Base editors of the invention can be used for targeted editing of DNA *in vitro* or *in vivo*. In non-limiting examples, base editors of the invention are used for the generation of mutant cells or animals, for the correction of genetic defects in cells *ex vivo* (such as in cells obtained from a subject that are subsequently re-introduced into the same or another subject), or for the introduction of targeted mutations *in vivo* (e.g., the correction of genetic defects or the introduction of deactivating mutations in disease-associated genes in a G to A, or a T to C to mutation).

Nucleobase Editors

Useful in the methods and compositions described herein are nucleobase editors that edit, modify or alter a target nucleotide sequence of a polynucleotide. Nucleobase editors described herein typically include a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (e.g., adenosine deaminase or cytidine deaminase). A polynucleotide programmable nucleotide binding domain, when in conjunction with a bound guide polynucleotide (e.g., gRNA), can specifically bind to a target polynucleotide sequence and thereby localize the base editor to the target nucleic acid sequence desired to be edited.

Polynucleotide Programmable Nucleotide Binding Domain

Polynucleotide programmable nucleotide binding domains bind polynucleotides (e.g., RNA, DNA). A polynucleotide programmable nucleotide binding domain of a base editor can itself comprise one or more domains (e.g., one or more nuclease domains). In some embodiments, the nuclease domain of a polynucleotide programmable nucleotide binding domain can comprise an endonuclease or an exonuclease. An endonuclease can cleave a single strand of a double-stranded nucleic acid or both strands of a double-stranded nucleic acid

molecule. In some embodiments, a nuclease domain of a polynucleotide programmable nucleotide binding domain can cut zero, one, or two strands of a target polynucleotide.

Non-limiting examples of a polynucleotide programmable nucleotide binding domain which can be incorporated into a base editor include a CRISPR protein-derived domain, a restriction nuclease, a meganuclease, TAL nuclease (TALEN), and a zinc finger nuclease (ZFN). In some embodiments, a base editor comprises a polynucleotide programmable nucleotide binding domain comprising a natural or modified protein or portion thereof which via a bound guide nucleic acid is capable of binding to a nucleic acid sequence during CRISPR (i.e., Clustered Regularly Interspaced Short Palindromic Repeats)-mediated modification of a nucleic acid. Such a protein is referred to herein as a “CRISPR protein.” Accordingly, disclosed herein is a base editor comprising a polynucleotide programmable nucleotide binding domain comprising all or a portion of a CRISPR protein (i.e. a base editor comprising as a domain all or a portion of a CRISPR protein, also referred to as a “CRISPR protein-derived domain” of the base editor). A CRISPR protein-derived domain incorporated into a base editor can be modified compared to a wild-type or natural version of the CRISPR protein. For example, as described below a CRISPR protein-derived domain can comprise one or more mutations, insertions, deletions, rearrangements and/or recombinations relative to a wild-type or natural version of the CRISPR protein.

Cas proteins that can be used herein include class 1 and class 2. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 or Csx12), Cas10, Csy1, Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Cas12a/Cpf1, Cas12b/C2c1 (e.g., SEQ ID NO: 156), Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ, CARF, DinG, homologues thereof, or modified versions thereof. A CRISPR enzyme can direct cleavage of one or both strands at a target sequence, such as within a target sequence and/or within a complement of a target sequence. For example, a CRISPR enzyme can direct cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence.

A vector that encodes a CRISPR enzyme that is mutated to with respect, to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence can be used. A

Cas protein (e.g., Cas9, Cas12) or a Cas domain (e.g., Cas9, Cas12) can refer to a polypeptide or domain with at least or at least about 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity and/or sequence homology to a wild-type exemplary Cas polypeptide or Cas domain. Cas (e.g., Cas9, Cas12) can refer to the wild-type or a modified form of the Cas protein that can comprise an amino acid change such as a deletion, insertion, substitution, variant, mutation, fusion, chimera, or any combination thereof.

In some embodiments, a CRISPR protein-derived domain of a base editor can include all or a portion of Cas9 from *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrophidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquis* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); *Neisseria meningitidis* (NCBI Ref: YP_002342100.1), *Streptococcus pyogenes*, or *Staphylococcus aureus*.

Cas9 nuclease sequences and structures are well known to those of skill in the art (See, e.g., “Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.” Ferretti et al., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); “CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.” Deltcheva E., et al., *Nature* 471:602-607(2011); and “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” Jinek M., et al., *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, “The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems” (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by reference.

High Fidelity Cas9 Domains

Some aspects of the disclosure provide high fidelity Cas9 domains. High fidelity Cas9 domains are known in the art and described, for example, in Kleinstiver, B.P., et al. “High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.” *Nature* 529, 490-495 (2016); and Slaymaker, I.M., et al. “Rationally engineered Cas9 nucleases with

improved specificity.” *Science* 351, 84-88 (2015); the entire contents of each of which are incorporated herein by reference. An Exemplary high fidelity Cas9 domain is provided in the Sequence Listing as SEQ ID NO: 157. In some embodiments, high fidelity Cas9 domains are engineered Cas9 domains comprising one or more mutations that decrease electrostatic

5 interactions between the Cas9 domain and the sugar-phosphate backbone of a DNA, relative to a corresponding wild-type Cas9 domain. High fidelity Cas9 domains that have decreased electrostatic interactions with the sugar-phosphate backbone of DNA have less off-target effects. In some embodiments, the Cas9 domain (*e.g.*, a wild type Cas9 domain (SEQ ID NOs: 93 and 158) comprises one or more mutations that decrease the association between the Cas9 domain

10 and the sugar-phosphate backbone of a DNA. In some embodiments, a Cas9 domain comprises one or more mutations that decreases the association between the Cas9 domain and the sugar-phosphate backbone of DNA by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, or at least 70%.

15 In some embodiments, any of the Cas9 fusion proteins provided herein comprise one or more of a D10A, N497X, a R661X, a Q695X, and/or a Q926X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the high fidelity Cas9 enzyme is SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, or hyper accurate Cas9 variant (HypaCas9). In some embodiments, the modified Cas9

20 eSpCas9(1.1) contains alanine substitutions that weaken the interactions between the HNH/RuvC groove and the non-target DNA strand, preventing strand separation and cutting at off-target sites. Similarly, SpCas9-HF1 lowers off-target editing through alanine substitutions that disrupt Cas9's interactions with the DNA phosphate backbone. HypaCas9 contains mutations (SpCas9 N692A/M694A/Q695A/H698A) in the REC3 domain that increase Cas9 proofreading and target

25 discrimination. All three high fidelity enzymes generate less off-target editing than wildtype Cas9.

Cas9 Domains with Reduced Exclusivity

Typically, Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a “protospacer adjacent motif (PAM)” or PAM-like motif, which is a 2-6 base pair DNA sequence immediately

30 following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system. The presence of an NGG PAM sequence is required to bind a particular nucleic acid region, where the “N” in “NGG” is adenosine (A), thymidine (T), or cytosine (C), and the G is guanosine. This may limit the ability to edit desired bases within a genome. In some

embodiments, the base editing fusion proteins provided herein may need to be placed at a precise location, for example a region comprising a target base that is upstream of the PAM. See *e.g.*, Komor, A.C., *et al.*, “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage” *Nature* 533, 420-424 (2016), the entire contents of which are hereby
5 incorporated by reference. Exemplary polypeptide sequences for spCas9 proteins capable of binding a PAM sequence are provided in the Sequenc Listing as SEQ ID NOs: 158-161. Accordingly, in some embodiments, any of the fusion proteins provided herein may contain a Cas9 domain that is capable of binding a nucleotide sequence that does not contain a canonical (*e.g.*, NGG) PAM sequence. Cas9 domains that bind to non-canonical PAM sequences have
10 been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.*, “Engineered CRISPR-Cas9 nucleases with altered PAM specificities” *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, “Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition” *Nature Biotechnology* 33, 1293-1298 (2015); the
15 entire contents of each are hereby incorporated by reference.

Nickases

In some embodiments, the polynucleotide programmable nucleotide binding domain can comprise a nickase domain. Herein the term “nickase” refers to a polynucleotide programmable
20 nucleotide binding domain comprising a nuclease domain that is capable of cleaving only one strand of the two strands in a duplexed nucleic acid molecule (*e.g.*, DNA). In some embodiments, a nickase can be derived from a fully catalytically active (*e.g.*, natural) form of a polynucleotide programmable nucleotide binding domain by introducing one or more mutations into the active polynucleotide programmable nucleotide binding domain. For example, where a
25 polynucleotide programmable nucleotide binding domain comprises a nickase domain derived from Cas9, the Cas9-derived nickase domain can include a D10A mutation and a histidine at position 840. In such embodiments, the residue H840 retains catalytic activity and can thereby cleave a single strand of the nucleic acid duplex. In another example, a Cas9-derived nickase domain can comprise an H840A mutation, while the amino acid residue at position 10 remains a
30 D. In some embodiments, a nickase can be derived from a fully catalytically active (*e.g.*, natural) form of a polynucleotide programmable nucleotide binding domain by removing all or a portion of a nuclease domain that is not required for the nickase activity. For example, where a polynucleotide programmable nucleotide binding domain comprises a nickase domain derived

from Cas9, the Cas9-derived nickase domain can comprise a deletion of all or a portion of the RuvC domain or the HNH domain.

In some embodiments, wild-type Cas9 corresponds to, or comprises the following amino acid sequence:

5 MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLV DSTDKADLRLLIYLA LAHMIKFRGHFLIEGDLNPDNSDVKLFQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
10 EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHSLLEYEFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF
NASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLK
15 RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILOTVKVVDDELVKVMGRHKPENIVIAMARENQTTQKGOKNSRERM
KRIEEGIKELGSQILKEHPVENTQLONEKLYLYYLONGRDMYVDQELDINRLSDYDVDHIVPOSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKROLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFOFYK
20 VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDEFATVRKVL SMPQVNIVKKTEVOT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGI
TIMERSSSFENPIDFLEAKGYKEVKKDLIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH
25 RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQL
GGD (SEQ ID NO:158) (single underline: HNH domain; double underline: RuvC domain).

Throughout the disclosure, the initial methionine of a polypeptide sequence (e.g., the wild-type-Cas9 sequence provided immediately above) is omitted in some embodiments where the polypeptide sequence is incorporated into a base editor and/or a fusion protein.

30 In some embodiments, the strand of a nucleic acid duplex target polynucleotide sequence that is cleaved by a base editor comprising a nickase domain (e.g., Cas9-derived nickase domain, Cas12-derived nickase domain) is the strand that is not edited by the base editor (i.e., the strand that is cleaved by the base editor is opposite to a strand comprising a base to be edited). In other
embodiments, a base editor comprising a nickase domain (e.g., Cas9-derived nickase domain,
35 Cas12-derived nickase domain) can cleave the strand of a DNA molecule which is being targeted for editing. In such embodiments, the non-targeted strand is not cleaved.

In some embodiments, a Cas9 nuclease has an inactive (e.g., an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase, referred to as an “nCas9” protein (for “nickase” Cas9). The Cas9 nickase may be a Cas9 protein that is capable of cleaving only one strand of a duplexed nucleic acid molecule (e.g., a duplexed DNA molecule). In some embodiments the Cas9 nickase cleaves the target strand of a duplexed nucleic acid molecule, meaning that the Cas9 nickase cleaves the strand that is base paired to (complementary to) a gRNA (e.g., an sgRNA) that is bound to the Cas9. In some embodiments, a Cas9 nickase comprises a D10A mutation and has a histidine at position 840. In some embodiments the Cas9 nickase cleaves the non-target, non-base-edited strand of a duplexed nucleic acid molecule, meaning that the Cas9 nickase cleaves the strand that is not base paired to a gRNA (e.g., an sgRNA) that is bound to the Cas9. In some embodiments, a Cas9 nickase comprises an H840A mutation and has an aspartic acid residue at position 10, or a corresponding mutation. In some embodiments the Cas9 nickase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the Cas9 nickases provided herein. Additional suitable Cas9 nickases will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure.

The amino acid sequence of an exemplary catalytically active Cas9 nickase (nCas9) is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAE
 ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLVESFLVEEDKKHERHPIFGNIVDEV
 AYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVDFLIQLVQT
 YNQLFEENPINASGVDAKILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
 LAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
 KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
 ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGP
 LARGNSRFAWMTRKSEETITPWNFEVVDKGGASQSFIERMTNFDKNLPNEKVLPHSLLYEYF
 TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
 VEDRFNASLGTYHDLKKIHKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVM
 KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQV
 SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
 RERMKRIIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
 VPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
 RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKD
 FQFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKA
 TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKK

EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKE
 LLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLEFLENGRKRMLASAGELQKGNELAL
 PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL
 YNKHARDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI
 TGLYETRI
 5 DLSQLGGD (SEQ ID NO: 94)

The Cas9 nuclease has two functional endonuclease domains: RuvC and HNH. Cas9 undergoes a conformational change upon target binding that positions the nuclease domains to cleave opposite strands of the target DNA. The end result of Cas9-mediated DNA cleavage is a double-strand break (DSB) within the target DNA (~3-4 nucleotides upstream of the PAM sequence). The resulting DSB is then repaired by one of two general repair pathways: (1) the efficient but error-prone non-homologous end joining (NHEJ) pathway; or (2) the less efficient but high-fidelity homology directed repair (HDR) pathway.

The “efficiency” of non-homologous end joining (NHEJ) and/or homology directed repair (HDR) can be calculated by any convenient method. For example, in some embodiments, efficiency can be expressed in terms of percentage of successful HDR. For example, a surveyor nuclease assay can be used to generate cleavage products and the ratio of products to substrate can be used to calculate the percentage. For example, a surveyor nuclease enzyme can be used that directly cleaves DNA containing a newly integrated restriction sequence as the result of successful HDR. More cleaved substrate indicates a greater percent HDR (a greater efficiency of HDR). As an illustrative example, a fraction (percentage) of HDR can be calculated using the following equation $[(\text{cleavage products})/(\text{substrate plus cleavage products})]$ (e.g., $(b+c)/(a+b+c)$, where “a” is the band intensity of DNA substrate and “b” and “c” are the cleavage products). In some embodiments, efficiency can be expressed in terms of percentage of successful NHEJ. For example, a T7 endonuclease I assay can be used to generate cleavage products and the ratio of products to substrate can be used to calculate the percentage NHEJ. T7 endonuclease I cleaves mismatched heteroduplex DNA which arises from hybridization of wild-type and mutant DNA strands (NHEJ generates small random insertions or deletions (indels) at the site of the original break). More cleavage indicates a greater percent NHEJ (a greater efficiency of NHEJ). As an illustrative example, a fraction (percentage) of NHEJ can be calculated using the following equation: $(1-(1-(b+c)/(a+b+c))^{1/2}) \times 100$, where “a” is the band intensity of DNA substrate and “b” and “c” are the cleavage products (Ran *et al.*, *Cell*. 2013 Sep. 12; 154(6):1380-9; and Ran *et al.*, *Nat Protoc.* 2013 Nov.; 8(11): 2281–2308).

The NHEJ repair pathway is the most active repair mechanism, and it frequently causes small nucleotide insertions or deletions (indels) at the DSB site. The randomness of NHEJ-mediated DSB repair has important practical implications, because a population of cells

expressing Cas9 and a gRNA or a guide polynucleotide can result in a diverse array of mutations. In most embodiments, NHEJ gives rise to small indels in the target DNA that result in amino acid deletions, insertions, or frameshift mutations leading to premature stop codons within the open reading frame (ORF) of the targeted gene. The ideal end result is a loss-of-
5 function mutation within the targeted gene.

While NHEJ-mediated DSB repair often disrupts the open reading frame of the gene, homology directed repair (HDR) can be used to generate specific nucleotide changes ranging from a single nucleotide change to large insertions like the addition of a fluorophore or tag. In order to utilize HDR for gene editing, a DNA repair template containing the desired sequence
10 can be delivered into the cell type of interest with the gRNA(s) and Cas9 or Cas9 nickase. The repair template can contain the desired edit as well as additional homologous sequence immediately upstream and downstream of the target (termed left & right homology arms). The length of each homology arm can be dependent on the size of the change being introduced, with larger insertions requiring longer homology arms. The repair template can be a single-stranded
15 oligonucleotide, double-stranded oligonucleotide, or a double-stranded DNA plasmid. The efficiency of HDR is generally low (<10% of modified alleles) even in cells that express Cas9, gRNA and an exogenous repair template. The efficiency of HDR can be enhanced by synchronizing the cells, since HDR takes place during the S and G2 phases of the cell cycle. Chemically or genetically inhibiting genes involved in NHEJ can also increase HDR frequency.
20 In some embodiments, Cas9 is a modified Cas9. A given gRNA targeting sequence can have additional sites throughout the genome where partial homology exists. These sites are called off-targets and need to be considered when designing a gRNA. In addition to optimizing gRNA design, CRISPR specificity can also be increased through modifications to Cas9. Cas9 generates double-strand breaks (DSBs) through the combined activity of two nuclease domains, RuvC and
25 HNH. Cas9 nickase, a D10A mutant of SpCas9, retains one nuclease domain and generates a DNA nick rather than a DSB. The nickase system can also be combined with HDR-mediated gene editing for specific gene edits.

30 *Catalytically Dead Nucleases*

Also provided herein are base editors comprising a polynucleotide programmable nucleotide binding domain which is catalytically dead (*i.e.*, incapable of cleaving a target polynucleotide sequence). Herein the terms “catalytically dead” and “nuclease dead” are used interchangeably to refer to a polynucleotide programmable nucleotide binding domain which has

one or more mutations and/or deletions resulting in its inability to cleave a strand of a nucleic acid. In some embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain base editor can lack nuclease activity as a result of specific point mutations in one or more nuclease domains. For example, in the case of a base editor comprising a Cas9 domain, the Cas9 can comprise both a D10A mutation and an H840A mutation. Such mutations inactivate both nuclease domains, thereby resulting in the loss of nuclease activity. In other embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain can comprise one or more deletions of all or a portion of a catalytic domain (e.g., RuvC1 and/or HNH domains). In further embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain comprises a point mutation (e.g., D10A or H840A) as well as a deletion of all or a portion of a nuclease domain. dCas9 domains are known in the art and described, for example, in Qi *et al.*, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell*. 2013; 152(5):1173-83, the entire contents of which are incorporated herein by reference.

Additional suitable nuclease-inactive dCas9 domains will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A/H840A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (See, e.g., Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838, the entire contents of which are incorporated herein by reference).

In some embodiments, dCas9 corresponds to, or comprises in part or in whole, a Cas9 amino acid sequence having one or more mutations that inactivate the Cas9 nuclease activity. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10X mutation and a H840X mutation of the amino acid sequence set forth herein, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid change. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10A mutation and a H840A mutation of the amino acid sequence set forth herein, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, a nuclease-inactive Cas9 domain comprises the amino acid sequence set forth in Cloning vector pPlatTET-gRNA2 (Accession No. BAV54124).

In some embodiments, a variant Cas9 protein can cleave the complementary strand of a guide target sequence but has reduced ability to cleave the non-complementary strand of a double stranded guide target sequence. For example, the variant Cas9 protein can have a

mutation (amino acid substitution) that reduces the function of the RuvC domain. As a non-limiting example, in some embodiments, a variant Cas9 protein has a D10A (aspartate to alanine at amino acid position 10) and can therefore cleave the complementary strand of a double stranded guide target sequence but has reduced ability to cleave the non-complementary strand of a double stranded guide target sequence (thus resulting in a single strand break (SSB) instead of a double strand break (DSB) when the variant Cas9 protein cleaves a double stranded target nucleic acid) (see, for example, Jinek *et al.*, *Science*. 2012 Aug. 17; 337(6096):816-21).

In some embodiments, a variant Cas9 protein can cleave the non-complementary strand of a double stranded guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence. For example, the variant Cas9 protein can have a mutation (amino acid substitution) that reduces the function of the HNH domain (RuvC/HNH/RuvC domain motifs). As a non-limiting example, in some embodiments, the variant Cas9 protein has an H840A (histidine to alanine at amino acid position 840) mutation and can therefore cleave the non-complementary strand of the guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence (thus resulting in a SSB instead of a DSB when the variant Cas9 protein cleaves a double stranded guide target sequence). Such a Cas9 protein has a reduced ability to cleave a guide target sequence (*e.g.*, a single stranded guide target sequence) but retains the ability to bind a guide target sequence (*e.g.*, a single stranded guide target sequence).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors W476A and W1126A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors H840A, W476A, and W1126A, mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). As another non-limiting example, in some embodiments, the variant Cas9 protein harbors H840A, D10A, W476A, and W1126A, mutations such that the polypeptide has a

reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). In some embodiments, the variant Cas9 has restored catalytic His residue at position 840 in the Cas9 HNH domain (A840H).

5 As another non-limiting example, in some embodiments, the variant Cas9 protein harbors, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). As another non-limiting example, in some
10 embodiments, the variant Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). In some embodiments, when a variant Cas9 protein harbors W476A and W1126A
15 mutations or when the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations, the variant Cas9 protein does not bind efficiently to a PAM sequence. Thus, in some such embodiments, when such a variant Cas9 protein is used in a method of binding, the method does not require a PAM sequence. In other words, in some embodiments, when such a variant Cas9 protein is used in a method of binding, the method can
20 include a guide RNA, but the method can be performed in the absence of a PAM sequence (and the specificity of binding is therefore provided by the targeting segment of the guide RNA). Other residues can be mutated to achieve the above effects (*i.e.*, inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 can be altered (*i.e.*, substituted). Also, mutations
25 other than alanine substitutions are suitable.

In some embodiments, a variant Cas9 protein that has reduced catalytic activity (*e.g.*, when a Cas9 protein has a D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or a A987 mutation, *e.g.*, D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A), the variant Cas9 protein can still bind to target DNA in a site-
30 specific manner (because it is still guided to a target DNA sequence by a guide RNA) as long as it retains the ability to interact with the guide RNA.

In some embodiments, the variant Cas protein can be spCas9, spCas9-VRQR, spCas9-VRER, xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER, spCas9-LRKIQK, or spCas9-LRVSQL.

In some embodiments, the Cas9 domain is a Cas9 domain from *Staphylococcus aureus* (SaCas9). In some embodiments, the SaCas9 domain is a nuclease active SaCas9, a nuclease inactive SaCas9 (SaCas9d), or a SaCas9 nickase (SaCas9n). In some embodiments, the SaCas9 comprises a N579A mutation, or a corresponding mutation in any of the Cas9 or SaCas9 amino acid sequences provided in the Sequence Listing submitted herewith.

In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a NNGRRT or a NNGRRV PAM sequence. In some embodiments, the SaCas9 domain comprises one or more of a E781X, a N967X, and a R1014X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SaCas9 domain comprises one or more of a E781K, a N967K, and a R1014H mutation, or one or more corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SaCas9 domain comprises a E781K, a N967K, or a R1014H mutation, or corresponding mutations in any of the amino acid sequences provided herein.

In some embodiments, one of the Cas9 domains present in the fusion protein may be replaced with a guide nucleotide sequence-programmable DNA-binding protein domain that has no requirements for a PAM sequence. In some embodiments, the Cas9 is an SaCas9. Residue A579 of SaCas9 can be mutated from N579 to yield a SaCas9 nickase. Residues K781, K967, and H1014 can be mutated from E781, N967, and R1014 to yield a SaKKH Cas9.

In some embodiments, a modified SpCas9 including amino acid substitutions D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R (SpCas9-MQKFRAER) and having specificity for the altered PAM 5'-NGC-3' was used.

Alternatives to *S. pyogenes* Cas9 can include RNA-guided endonucleases from the Cpf1 family that display cleavage activity in mammalian cells. CRISPR from *Prevotella* and *Francisella I* (CRISPR/Cpf1) is a DNA-editing technology analogous to the CRISPR/Cas9 system. Cpf1 is an RNA-guided endonuclease of a class II CRISPR/Cas system. This acquired immune mechanism is found in *Prevotella* and *Francisella* bacteria. Cpf1 genes are associated with the CRISPR locus, coding for an endonuclease that use a guide RNA to find and cleave viral DNA. Cpf1 is a smaller and simpler endonuclease than Cas9, overcoming some of the CRISPR/Cas9 system limitations. Unlike Cas9 nucleases, the result of Cpf1-mediated DNA cleavage is a double-strand break with a short 3' overhang. Cpf1's staggered cleavage pattern can open up the possibility of directional gene transfer, analogous to traditional

restriction enzyme cloning, which can increase the efficiency of gene editing. Like the Cas9 variants and orthologues described above, Cpf1 can also expand the number of sites that can be targeted by CRISPR to AT-rich regions or AT-rich genomes that lack the NGG PAM sites favored by SpCas9. The Cpf1 locus contains a mixed alpha/beta domain, a RuvC-I followed by
 5 a helical region, a RuvC-II and a zinc finger-like domain. The Cpf1 protein has a RuvC-like endonuclease domain that is similar to the RuvC domain of Cas9.

Furthermore, Cpf1, unlike Cas9, does not have a HNH endonuclease domain, and the N-terminal of Cpf1 does not have the alpha-helical recognition lobe of Cas9. Cpf1 CRISPR-Cas domain architecture shows that Cpf1 is functionally unique, being classified as Class 2, type V
 10 CRISPR system. The Cpf1 loci encode Cas1, Cas2 and Cas4 proteins that are more similar to types I and III than type II systems. Functional Cpf1 does not require the trans-activating CRISPR RNA (tracrRNA), therefore, only CRISPR (crRNA) is required. This benefits genome editing because Cpf1 is not only smaller than Cas9, but also it has a smaller sgRNA molecule (approximately half as many nucleotides as Cas9). The Cpf1-crRNA complex cleaves target
 15 DNA or RNA by identification of a protospacer adjacent motif 5'-YTN-3' or 5'-TTN-3' in contrast to the G-rich PAM targeted by Cas9. After identification of PAM, Cpf1 introduces a sticky-end-like DNA double-stranded break having an overhang of 4 or 5 nucleotides.

In some embodiments, the Cas9 is a Cas9 variant having specificity for an altered PAM sequence. In some embodiments, the Additional Cas9 variants and PAM sequences are described
 20 in Miller, S.M., *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs, Nat. Biotechnol. (2020), the entirety of which is incorporated herein by reference. In some embodiments, a Cas9 variant have no specific PAM requirements. In some embodiments, a Cas9 variant, *e.g.* a SpCas9 variant has specificity for a NRNH PAM, wherein R is A or G and H is A, C, or T. In some embodiments, the SpCas9 variant has specificity for a PAM sequence AAA,
 25 TAA, CAA, GAA, TAT, GAT, or CAC. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1218, 1219, 1221, 1249, 1256, 1264, 1290, 1318, 1317, 1320, 1321, 1323, 1332, 1333, 1335, 1337, or 1339 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1135, 1218, 1219, 1221, 1249, 1320, 1321, 1323,
 30 1332, 1333, 1335, or 1337 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1219, 1221, 1256, 1264, 1290, 1318, 1317, 1320, 1323, 1333 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1131, 1135, 1150, 1156, 1180, 1191, 1218, 1219, 1221, 1227,

- 1249, 1253, 1286, 1293, 1320, 1321, 1332, 1335, 1339 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1127, 1135, 1180, 1207, 1219, 1234, 1286, 1301, 1332, 1335, 1337, 1338, 1349 or a corresponding position thereof. Exemplary amino acid substitutions and PAM specificity of
- 5 SpCas9 variants are shown in **Tables 2A-2B and 3**.

Table 2A SpCas9 Variants

SpCas9	SpCas9 amino acid position												
	1114	1135	1218	1219	1221	1249	1320	1321	1323	1332	1333	1335	1337
	R	D	G	E	Q	P	A	P	A	D	R	R	T
AAA		N		V	H						G		
AAA		N		V	H						G		
AAA				V							G		
TAA	G	N		V							I		
TAA		N		V							I		A
TAA	G	N		V							I		A
CAA				V							K		
CAA		N		V							K		
CAA		N		V							K		
GAA				V	H		V				K		
GAA		N		V			V				K		
GAA				V	H		V				K		
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
GAT				V							I		
GAT				V					D			Q	
GAT				V					D			Q	
CAC				V						N		Q	N
CAC		N		V								Q	N
CAC				V						N		Q	N

Table 2B

SpCas9	SpCas9 amino acid position															
	1114	1134	1135	1137	1139	1151	1180	1188	1211	1219	1221	1256	1264	1290	1318	1317
	R	F	D	P	V	K	D	K	K	E	Q	Q	H	V	L	N
GAA										V	H					
GAA			N	S						V						
GAA			N							V	H					
CAA			N							V	H					
CAA	G		N	S						V	H					
CAA			N					R		V	H					
CAA			N				G		R	V	H					
CAA			N							V	H					
AAA			N				G			V	H					
CAA	G		N				G			V	H					
CAA		L	N				G			V	H					
TAA	G		N				G			V	H					
TAA	G		N			E	G			V	H					
TAA	G		N				G			V	H					
TAA	G		N				G		R	V	H					
TAA			N				G		R	V	H					
TAA	G		N		A		G			V	H					
TAA	G		N							V	H					

Table 2C

SpCas9	SpCas9 amino acid position																			
	1114	1131	1135	1150	1156	1180	1191	1218	1219	1221	1227	1249	1253	1286	1293	1320	1321	1332	1335	1339
R	Y	D	E	K		D	K	G	E	Q	A	P	E	N	A	A	P	D	R	T
SacB.TAT		N					N		V	H						V	S		L	
SacB.TAT		N						S	V	H	V	S			T		S	G	L	
AAT		N						S	V	H		S		K			S	G	L	I
TAT	G	N				G		S	V	H		S	K				S	G	L	
TAT	G	N				G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N	E		G		S	V	H		S					S	G	L	
TAT	G	C	N	V		G		S	V	H		S					S	G	L	
TAT		C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	

Table 3

SpCas9	SpCas9 amino acid position															
	1114	1127	1135	1180	1207	1219	1234	1286	1301	1332	1335	1337	1338	1349		
R	D	D	D	E	E	E	N	N	P	D	R	T	S	H		
SacB.CAC			N			V				N	Q	N				
AAC	G		N			V				N	Q	N				
AAC	G		N			V				N	Q	N				
TAC	G		N			V				N	Q	N				
TAC	G		N			V		H		N	Q	N				
TAC	G		N		G	V	D	H		N	Q	N				
TAC	G		N			V				N	Q	N				
TAC	G	G	N	E		V		H		N	Q	N				

	SpCas9 amino acid position													
	1114	1127	1135	1180	1207	1219	1234	1286	1301	1332	1335	1337	1338	1349
SpCas9	R	D	D	D	E	E	N	N	P	D	R	T	S	H
TAC	G		N			V		H		N	Q	N		
TAC	G		N			V				N	Q	N	T	R

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a single effector of a microbial CRISPR-Cas system. Single effectors of microbial CRISPR-Cas systems include, without limitation, Cas9, Cpf1, Cas12b/C2c1, and Cas12c/C2c3. Typically, microbial CRISPR-Cas systems are divided into Class 1 and Class 2 systems. Class 1 systems have multisubunit effector complexes, while Class 2 systems have a single protein effector. For example, Cas9 and Cpf1 are Class 2 effectors. In addition to Cas9 and Cpf1, three distinct Class 2 CRISPR-Cas systems (Cas12b/C2c1, and Cas12c/C2c3) have been described by Shmakov *et al.*, "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", *Mol. Cell*, 2015 Nov. 5; 60(3): 385-397, the entire contents of which is hereby incorporated by reference. Effectors of two of the systems, Cas12b/C2c1, and Cas12c/C2c3, contain RuvC-like endonuclease domains related to Cpf1. A third system contains an effector with two predicated HEPN RNase domains. Production of mature CRISPR RNA is tracrRNA-independent, unlike production of CRISPR RNA by Cas12b/C2c1. Cas12b/C2c1 depends on both CRISPR RNA and tracrRNA for DNA cleavage.

In some embodiments, the napDNAbp is a circular permutant (e.g., SEQ ID NO: 163).

The crystal structure of *Alicyclobacillus acidoterrestris* Cas12b/C2c1 (AacC2c1) has been reported in complex with a chimeric single-molecule guide RNA (sgRNA). See *e.g.*, Liu *et al.*, "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", *Mol. Cell*, 2017 Jan. 19; 65(2):310-322, the entire contents of which are hereby incorporated by reference. The crystal structure has also been reported in *Alicyclobacillus acidoterrestris* C2c1 bound to target DNAs as ternary complexes. See *e.g.*, Yang *et al.*, "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell*, 2016 Dec. 15; 167(7):1814-1828, the entire contents of which are hereby incorporated by reference. Catalytically competent conformations of AacC2c1, both with target and non-target DNA strands, have been captured independently positioned within a single RuvC catalytic pocket, with Cas12b/C2c1-mediated cleavage resulting in a staggered seven-nucleotide break of target DNA. Structural comparisons between Cas12b/C2c1 ternary complexes and previously identified Cas9 and Cpf1 counterparts demonstrate the diversity of mechanisms used by CRISPR-Cas9 systems.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein may be a Cas12b/C2c1, or a Cas12c/C2c3 protein. In some embodiments, the napDNAbp is a Cas12b/C2c1 protein. In some embodiments, the napDNAbp is a Cas12c/C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Cas12b/C2c1 or Cas12c/C2c3 protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12b/C2c1 or Cas12c/C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the napDNAbp sequences provided herein. It should be appreciated that Cas12b/C2c1 or Cas12c/C2c3 from other bacterial species may also be used in accordance with the present disclosure.

In some embodiments, a napDNAbp refers to Cas12c. In some embodiments, the Cas12c protein is a Cas12c1 (SEQ ID NO: 164) or a variant of Cas12c1. In some embodiments, the Cas12 protein is a Cas12c2 (SEQ ID NO: 165) or a variant of Cas12c2. In some embodiments, the Cas12 protein is a Cas12c protein from *Oleiphilus* sp. HI0009 (*i.e.*, OspCas12c; SEQ ID NO: 166) or a variant of OspCas12c. These Cas12c molecules have been described in Yan *et al.*, "Functionally Diverse Type V CRISPR-Cas Systems," Science, 2019 Jan. 4; 363: 88-91; the entire contents of which is hereby incorporated by reference. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Cas12c1, Cas12c2, or OspCas12c protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12c1, Cas12c2, or OspCas12c protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any Cas12c1, Cas12c2, or OspCas12c protein described herein. It should be appreciated that Cas12c1, Cas12c2, or OspCas12c from other bacterial species may also be used in accordance with the present disclosure.

In some embodiments, a napDNAbp refers to Cas12g, Cas12h, or Cas12i, which have been described in, for example, Yan *et al.*, "Functionally Diverse Type V CRISPR-Cas Systems," *Science*, 2019 Jan. 4; 363: 88-91; the entire contents of each is hereby incorporated by reference. Exemplary Cas12g, Cas12h, and Cas12i polypeptide sequences are provided in the Sequence Listing as SEQ ID NOs: 167-170. By aggregating more than 10 terabytes of sequence data, new classifications of Type V Cas proteins were identified that showed weak similarity to previously characterized Class V protein, including Cas12g, Cas12h, and Cas12i. In some embodiments, the Cas12 protein is a Cas12g or a variant of Cas12g. In some embodiments, the Cas12 protein is a Cas12h or a variant of Cas12h. In some embodiments, the Cas12 protein is a Cas12i or a variant of Cas12i. It should be appreciated that other RNA-guided DNA binding proteins may be used as a napDNAbp, and are within the scope of this disclosure. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Cas12g, Cas12h, or Cas12i protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12g, Cas12h, or Cas12i protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any Cas12g, Cas12h, or Cas12i protein described herein. It should be appreciated that Cas12g, Cas12h, or Cas12i from other bacterial species may also be used in accordance with the present disclosure. In some embodiments, the Cas12i is a Cas12i1 or a Cas12i2.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein may be a Cas12j/CasΦ protein. Cas12j/CasΦ is described in Pausch *et al.*, "CRISPR-CasΦ from huge phages is a hypercompact genome editor," *Science*, 17 July 2020, Vol. 369, Issue 6501, pp. 333-337, which is incorporated herein by reference in its entirety. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Cas12j/CasΦ protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12j/CasΦ protein. In some embodiments, the napDNAbp is a nuclease inactive ("dead")

Cas12j/CasΦ protein. It should be appreciated that Cas12j/CasΦ from other species may also be used in accordance with the present disclosure.

Fusion Proteins with Internal Insertion

5 Provided herein are fusion proteins comprising a heterologous polypeptide fused to a nucleic acid programmable nucleic acid binding protein, for example, a napDNAbp. A heterologous polypeptide can be a polypeptide that is not found in the native or wild-type napDNAbp polypeptide sequence. The heterologous polypeptide can be fused to the napDNAbp at a C-terminal end of the napDNAbp, an N-terminal end of the napDNAbp,
10 or inserted at an internal location of the napDNAbp. In some embodiments, the heterologous polypeptide is a deaminase (e.g., cytidine or adenosine deaminase) or a functional fragment thereof. For example, a fusion protein can comprise a deaminase flanked by an N-terminal fragment and a C-terminal fragment of a Cas9 or Cas12 (e.g., Cas12b/C2c1), polypeptide. In some embodiments, the cytidine deaminase is an
15 APOBEC deaminase (e.g., APOBEC1). In some embodiments, the adenosine deaminase is a TadA (e.g., TadA*7.10 or TadA*8). In some embodiments, the TadA is a TadA*8 or a TadA*9. TadA sequences (e.g., TadA7.10 or TadA*8) as described herein are suitable deaminases for the above-described fusion proteins.

In some embodiments, the fusion protein comprises the structure:

20 NH₂-[N-terminal fragment of a napDNAbp]-[deaminase]-[C-terminal fragment of a napDNAbp]-COOH;
 NH₂-[N-terminal fragment of a Cas9]-[adenosine deaminase]-[C-terminal fragment of a Cas9]-COOH;
 NH₂-[N-terminal fragment of a Cas12]-[adenosine deaminase]-[C-terminal fragment of a
25 Cas12]-COOH;
 NH₂-[N-terminal fragment of a Cas9]-[cytidine deaminase]-[C-terminal fragment of a Cas9]-COOH;
 NH₂-[N-terminal fragment of a Cas12]-[cytidine deaminase]-[C-terminal fragment of a Cas12]-COOH;
30 where each instance of “[”-]” is an optional linker.

The deaminase can be a circular permutant deaminase. For example, the deaminase can be a circular permutant adenosine deaminase. In some embodiments, the

deaminase is a circular permutant TadA, circularly permuted at amino acid residue 116, 136, or 65 as numbered in the TadA reference sequence.

The fusion protein can comprise more than one deaminase. The fusion protein can comprise, for example, 1, 2, 3, 4, 5 or more deaminases. In some embodiments, the fusion protein comprises one or two deaminase. The two or more deaminases in a fusion protein can be an adenosine deaminase, a cytidine deaminase, or a combination thereof. The two or more deaminases can be homodimers or heterodimers. The two or more deaminases can be inserted in tandem in the napDNAbp. In some embodiments, the two or more deaminases may not be in tandem in the napDNAbp.

In some embodiments, the napDNAbp in the fusion protein is a Cas9 polypeptide or a fragment thereof. The Cas9 polypeptide can be a variant Cas9 polypeptide. In some embodiments, the Cas9 polypeptide is a Cas9 nickase (nCas9) polypeptide or a fragment thereof. In some embodiments, the Cas9 polypeptide is a nuclease dead Cas9 (dCas9) polypeptide or a fragment thereof. The Cas9 polypeptide in a fusion protein can be a full-length Cas9 polypeptide. In some cases, the Cas9 polypeptide in a fusion protein may not be a full length Cas9 polypeptide. The Cas9 polypeptide can be truncated, for example, at a N-terminal or C-terminal end relative to a naturally-occurring Cas9 protein. The Cas9 polypeptide can be a circularly permuted Cas9 protein. The Cas9 polypeptide can be a fragment, a portion, or a domain of a Cas9 polypeptide, that is still capable of binding the target polynucleotide and a guide nucleic acid sequence.

In some embodiments, the Cas9 polypeptide is a *Streptococcus pyogenes* Cas9 (SpCas9), *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus* 1 Cas9 (St1Cas9), or fragments or variants of any of the Cas9 polypeptides described herein.

In some embodiments, the fusion protein comprises an adenosine deaminase domain and a cytidine deaminase domain inserted within a Cas9. In some embodiments, an adenosine deaminase is fused within a Cas9 and a cytidine deaminase is fused to the C-terminus. In some embodiments, an adenosine deaminase is fused within Cas9 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and an adenosine deaminase is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and an adenosine deaminase fused to the N-terminus.

Exemplary structures of a fusion protein with an adenosine deaminase and a cytidine deaminase and a Cas9 are provided as follows:

NH₂-[Cas9(adenosine deaminase)]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[Cas9(adenosine deaminase)]-COOH;
 NH₂-[Cas9(cytidine deaminase)]-[adenosine deaminase]-COOH; or
 NH₂-[adenosine deaminase]-[Cas9(cytidine deaminase)]-COOH.

5 In some embodiments, the “-” used in the general architecture above indicates the presence of an optional linker.

In various embodiments, the catalytic domain has DNA modifying activity (e.g., deaminase activity), such as adenosine deaminase activity. In some embodiments, the adenosine deaminase is a TadA (e.g., TadA*7.10). In some embodiments, the TadA is a
 10 TadA*8. In some embodiments, a TadA*8 is fused within Cas9 and a cytidine deaminase is fused to the C-terminus. In some embodiments, a TadA*8 is fused within Cas9 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and a TadA*8 is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and a TadA*8 fused to the N-terminus.

15 Exemplary structures of a fusion protein with a TadA*8 and a cytidine deaminase and a Cas9 are provided as follows:

NH₂-[Cas9(TadA*8)]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[Cas9(TadA*8)]-COOH;
 NH₂-[Cas9(cytidine deaminase)]-[TadA*8]-COOH; or
 20 NH₂-[TadA*8]-[Cas9(cytidine deaminase)]-COOH.

In some embodiments, the “-” used in the general architecture above indicates the presence of an optional linker.

The heterologous polypeptide (e.g., deaminase) can be inserted in the napDNAbp (e.g., Cas9 or Cas12 (e.g., Cas12b/C2c1)) at a suitable location, for example, such that the
 25 napDNAbp retains its ability to bind the target polynucleotide and a guide nucleic acid. A deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted into a napDNAbp without compromising function of the deaminase (e.g., base editing activity) or the napDNAbp (e.g., ability to bind to target nucleic acid and guide nucleic acid). A deaminase (e.g., adenosine
 30 deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted in the napDNAbp at, for example, a disordered region or a region comprising a high temperature factor or B-factor as shown by crystallographic studies. Regions of a protein that are less ordered, disordered, or unstructured, for example solvent exposed

regions and loops, can be used for insertion without compromising structure or function. A deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted in the napDNAbp in a flexible loop region or a solvent-exposed region. In some embodiments, the deaminase (e.g., adenosine
5 deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted in a flexible loop of the Cas9 or the Cas12b/C2c1 polypeptide.

In some embodiments, the insertion location of a deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is determined by B-factor analysis of the crystal structure of Cas9 polypeptide. In some
10 embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted in regions of the Cas9 polypeptide comprising higher than average B-factors (e.g., higher B factors compared to the total protein or the protein domain comprising the disordered region). B-factor or temperature factor can indicate the fluctuation of atoms from their average position (for
15 example, as a result of temperature-dependent atomic vibrations or static disorder in a crystal lattice). A high B-factor (e.g., higher than average B-factor) for backbone atoms can be indicative of a region with relatively high local mobility. Such a region can be used for inserting a deaminase without compromising structure or function. A deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine
20 deaminase) can be inserted at a location with a residue having a C α atom with a B-factor that is 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200%, or greater than 200% more than the average B-factor for the total protein. A deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at a location with a residue having a
25 C α atom with a B-factor that is 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200% or greater than 200% more than the average B-factor for a Cas9 protein domain comprising the residue. Cas9 polypeptide positions comprising a higher than average B-factor can include, for example, residues 768, 792, 1052, 1015, 1022, 1026, 1029, 1067, 1040, 1054, 1068, 1246, 1247, and 1248
30 as numbered in the Cas9 reference sequence. Cas9 polypeptide regions comprising a higher than average B-factor can include, for example, residues 792-872, 792-906, and 2-791 as numbered in the Cas9 reference sequence (SEQ ID NO: 158).

A heterologous polypeptide (e.g., deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, and 1248 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, or 1248-1249 as numbered in the Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 769-770, 792-793, 793-794, 1016-1017, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1053-1054, 1055-1056, 1068-1069, 1069-1070, 1248-1249, or 1249-1250 as numbered in the Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide replaces an amino acid residue selected from the group consisting of: 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, and 1248 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. It should be understood that the reference to the Cas9 reference sequence with respect to insertion positions is for illustrative purposes. The insertions as discussed herein are not limited to the Cas9 polypeptide sequence of the Cas9 reference sequence, but include insertion at corresponding locations in variant Cas9 polypeptides, for example a Cas9 nickase (nCas9), nuclease dead Cas9 (dCas9), a Cas9 variant lacking a nuclease domain, a truncated Cas9, or a Cas9 domain lacking partial or complete HNH domain.

A heterologous polypeptide (e.g., deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 768, 792, 1022, 1026, 1040, 1068, and 1247 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 768-769, 792-793, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1068-1069, or 1247-1248 as numbered in the Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 769-770, 793-794, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1069-1070, or 1248-1249 as numbered in the Cas9 reference sequence or corresponding amino acid positions

thereof. In some embodiments, the heterologous polypeptide replaces an amino acid residue selected from the group consisting of: 768, 792, 1022, 1026, 1040, 1068, and 1247 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

5 A heterologous polypeptide (e.g., deaminase) can be inserted in the napDNAbp at an amino acid residue as described herein, or a corresponding amino acid residue in another Cas9 polypeptide. In an embodiment, a heterologous polypeptide (e.g., deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 1002, 1003, 1025, 1052-1056, 1242-1247, 1061-1077, 943-947, 686-
10 691, 569-578, 530-539, and 1060-1077 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at the N-terminus or the C-terminus of the residue or replace the residue. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine
15 deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of the residue.

 In some embodiments, an adenosine deaminase (e.g., TadA) is inserted at an amino acid residue selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the Cas9 reference
20 sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, an adenosine deaminase (e.g., TadA) is inserted in place of residues 792-872, 792-906, or 2-791 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted at the N-terminus of an amino acid selected from the group
25 consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted at the C-terminus of an amino acid selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered
30 in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted to replace an amino acid selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247,

1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, a cytidine deaminase (e.g., APOBEC1) is inserted at an amino acid residue selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the cytidine deaminase is inserted at the N-terminus of an amino acid selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the cytidine deaminase is inserted at the C-terminus of an amino acid selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the cytidine deaminase is inserted to replace an amino acid selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 768 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 768 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 768 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 768 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 791 or is inserted at amino acid residue 792, as numbered in the Cas9 reference

sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some
embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or
adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid
residue 791 or is inserted at the N-terminus of amino acid 792, as numbered in the Cas9
5 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.
In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or
adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid
791 or is inserted at the N-terminus of amino acid 792, as numbered in the Cas9 reference
sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some
10 embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or
adenosine deaminase and cytidine deaminase) is inserted to replace amino acid 791, or is
inserted to replace amino acid 792, as numbered in the Cas9 reference sequence, or a
corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine
15 deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid
residue 1016 as numbered in the Cas9 reference sequence, or a corresponding amino acid
residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g.,
adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine
deaminase) is inserted at the N-terminus of amino acid residue 1016 as numbered in the
20 Cas9 reference sequence, or a corresponding amino acid residue in another Cas9
polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine
deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus
of amino acid residue 1016 as numbered in the Cas9 reference sequence, or a
corresponding amino acid residue in another Cas9 polypeptide. In some embodiments,
25 the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase
and cytidine deaminase) is inserted to replace amino acid residue 1016 as numbered in the
Cas9 reference sequence, or a corresponding amino acid residue in another Cas9
polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine
30 deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid
residue 1022, or is inserted at amino acid residue 1023, as numbered in the Cas9
reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.
In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or

adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1022 or is inserted at the N-terminus of amino acid residue 1023, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1022 or is inserted at the C-terminus of amino acid residue 1023, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1022, or is inserted to replace amino acid residue 1023, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1026, or is inserted at amino acid residue 1029, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1026 or is inserted at the N-terminus of amino acid residue 1029, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1026 or is inserted at the C-terminus of amino acid residue 1029, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1026, or is inserted to replace amino acid residue 1029, as numbered in the Cas9 reference sequence, or corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1040 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g.,

adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1040 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1040 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1040 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1052, or is inserted at amino acid residue 1054, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1052 or is inserted at the N-terminus of amino acid residue 1054, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1052 or is inserted at the C-terminus of amino acid residue 1054, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1052, or is inserted to replace amino acid residue 1054, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1067, or is inserted at amino acid residue 1068, or is inserted at amino acid residue 1069, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g.,

adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1067 or is inserted at the N-terminus of amino acid residue 1068 or is inserted at the N-terminus of amino acid residue 1069, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1067 or is inserted at the C-terminus of amino acid residue 1068 or is inserted at the C-terminus of amino acid residue 1069, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1067, or is inserted to replace amino acid residue 1068, or is inserted to replace amino acid residue 1069, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1246, or is inserted at amino acid residue 1247, or is inserted at amino acid residue 1248, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1246 or is inserted at the N-terminus of amino acid residue 1247 or is inserted at the N-terminus of amino acid residue 1248, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1246 or is inserted at the C-terminus of amino acid residue 1247 or is inserted at the C-terminus of amino acid residue 1248, as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1246, or is inserted to replace amino acid residue 1247, or is inserted to replace amino acid residue 1248, as numbered in the

Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, a heterologous polypeptide (e.g., deaminase) is inserted in a flexible loop of a Cas9 polypeptide. The flexible loop portions can be selected from the group consisting of 530-537, 569-570, 686-691, 943-947, 1002-1025, 1052-1077, 1232-1247, or 1298-1300 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The flexible loop portions can be selected from the group consisting of: 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, or 1248-1297 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (e.g., adenine deaminase) can be inserted into a Cas9 polypeptide region corresponding to amino acid residues: 1017-1069, 1242-1247, 1052-1056, 1060-1077, 1002 – 1003, 943-947, 530-537, 568-579, 686-691, 1242-1247, 1298 – 1300, 1066-1077, 1052-1056, or 1060-1077 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (e.g., adenine deaminase) can be inserted in place of a deleted region of a Cas9 polypeptide. The deleted region can correspond to an N-terminal or C-terminal portion of the Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 792-872 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 792-906 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 2-791 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 1017-1069 as numbered in the Cas9 reference sequence, or corresponding amino acid residues thereof.

Exemplary internal fusions base editors are provided in **Table 4** below:

Table 4: Insertion loci in Cas9 proteins

BE ID	Modification	Other ID
IBE001	Cas9 TadA ins 1015	ISLAY01
IBE002	Cas9 TadA ins 1022	ISLAY02
IBE003	Cas9 TadA ins 1029	ISLAY03

BE ID	Modification	Other ID
IBE004	Cas9 TadA ins 1040	ISLAY04
IBE005	Cas9 TadA ins 1068	ISLAY05
IBE006	Cas9 TadA ins 1247	ISLAY06
IBE007	Cas9 TadA ins 1054	ISLAY07
IBE008	Cas9 TadA ins 1026	ISLAY08
IBE009	Cas9 TadA ins 768	ISLAY09
IBE020	delta HNH TadA 792	ISLAY20
IBE021	N-term fusion single TadA helix truncated 165-end	ISLAY21
IBE029	TadA-Circular Permutant 116 ins 1067	ISLAY29
IBE031	TadA- Circular Permutant 136 ins 1248	ISLAY31
IBE032	TadA- Circular Permutant 136 ins 1052	ISLAY32
IBE035	delta 792-872 TadA ins	ISLAY35
IBE036	delta 792-906 TadA ins	ISLAY36
IBE043	TadA-Circular Permutant 65 ins 1246	ISLAY43
IBE044	TadA ins C-term truncate 2791	ISLAY44

A heterologous polypeptide (e.g., deaminase) can be inserted within a structural or functional domain of a Cas9 polypeptide. A heterologous polypeptide (e.g., deaminase) can be inserted between two structural or functional domains of a Cas9 polypeptide. A heterologous polypeptide (e.g., deaminase) can be inserted in place of a structural or functional domain of a Cas9 polypeptide, for example, after deleting the domain from the Cas9 polypeptide. The structural or functional domains of a Cas9 polypeptide can include, for example, RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, or HNH.

In some embodiments, the Cas9 polypeptide lacks one or more domains selected from the group consisting of: RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, or HNH domain. In some embodiments, the Cas9 polypeptide lacks a nuclease domain. In some embodiments, the Cas9 polypeptide lacks an HNH domain. In some embodiments, the Cas9 polypeptide lacks a portion of the HNH domain such that the Cas9 polypeptide has reduced or abolished HNH activity. In some embodiments, the Cas9 polypeptide comprises a deletion of the nuclease domain, and the deaminase is inserted to replace the nuclease domain. In some embodiments, the HNH domain is deleted and the deaminase

is inserted in its place. In some embodiments, one or more of the RuvC domains is deleted and the deaminase is inserted in its place.

A fusion protein comprising a heterologous polypeptide can be flanked by a N-terminal and a C-terminal fragment of a napDNAbp. In some embodiments, the fusion protein comprises a deaminase flanked by a N-terminal fragment and a C-terminal fragment of a Cas9 polypeptide. The N terminal fragment or the C terminal fragment can bind the target polynucleotide sequence. The C-terminus of the N terminal fragment or the N-terminus of the C terminal fragment can comprise a part of a flexible loop of a Cas9 polypeptide. The C-terminus of the N terminal fragment or the N-terminus of the C terminal fragment can comprise a part of an alpha-helix structure of the Cas9 polypeptide. The N-terminal fragment or the C-terminal fragment can comprise a DNA binding domain. The N-terminal fragment or the C-terminal fragment can comprise a RuvC domain. The N-terminal fragment or the C-terminal fragment can comprise an HNH domain. In some embodiments, neither of the N-terminal fragment and the C-terminal fragment comprises an HNH domain.

In some embodiments, the C-terminus of the N terminal Cas9 fragment comprises an amino acid that is in proximity to a target nucleobase when the fusion protein deaminates the target nucleobase. In some embodiments, the N-terminus of the C terminal Cas9 fragment comprises an amino acid that is in proximity to a target nucleobase when the fusion protein deaminates the target nucleobase. The insertion location of different deaminases can be different in order to have proximity between the target nucleobase and an amino acid in the C-terminus of the N terminal Cas9 fragment or the N-terminus of the C terminal Cas9 fragment. For example, the insertion position of an deaminase can be at an amino acid residue selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The N-terminal Cas9 fragment of a fusion protein (i.e. the N-terminal Cas9 fragment flanking the deaminase in a fusion protein) can comprise the N-terminus of a Cas9 polypeptide. The N-terminal Cas9 fragment of a fusion protein can comprise a length of at least about: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino acids. The N-terminal Cas9 fragment of a fusion protein can comprise a sequence corresponding to amino acid residues: 1-56, 1-95, 1-200, 1-300, 1-400, 1-500,

1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, or 1-1100 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The N-terminal Cas9 fragment can comprise a sequence comprising at least: 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to amino acid residues: 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, or 1-1100 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The C-terminal Cas9 fragment of a fusion protein (i.e. the C-terminal Cas9 fragment flanking the deaminase in a fusion protein) can comprise the C-terminus of a Cas9 polypeptide. The C-terminal Cas9 fragment of a fusion protein can comprise a length of at least about: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino acids. The C-terminal Cas9 fragment of a fusion protein can comprise a sequence corresponding to amino acid residues: 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, or 56-1368 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The N-terminal Cas9 fragment can comprise a sequence comprising at least: 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to amino acid residues: 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, or 56-1368 as numbered in the Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The N-terminal Cas9 fragment and C-terminal Cas9 fragment of a fusion protein taken together may not correspond to a full-length naturally occurring Cas9 polypeptide sequence, for example, as set forth in the Cas9 reference sequence.

The fusion protein described herein can effect targeted deamination with reduced deamination at non-target sites (e.g., off-target sites), such as reduced genome wide spurious deamination. The fusion protein described herein can effect targeted deamination with reduced bystander deamination at non-target sites. The undesired deamination or off-target deamination can be reduced by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% compared with, for example, an end terminus fusion protein comprising the deaminase fused to a N terminus or a C terminus of a Cas9 polypeptide. The undesired

deamination or off-target deamination can be reduced by at least one-fold, at least two-fold, at least three-fold, at least four-fold, at least five-fold, at least tenfold, at least fifteen fold, at least twenty fold, at least thirty fold, at least forty fold, at least fifty fold, at least 60 fold, at least 70 fold, at least 80 fold, at least 90 fold, or at least hundred fold,

- 5 compared with, for example, an end terminus fusion protein comprising the deaminase fused to a N terminus or a C terminus of a Cas9 polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) of the fusion protein deaminates no more than two nucleobases within the range of an R-loop. In some
10 embodiments, the deaminase of the fusion protein deaminates no more than three nucleobases within the range of the R-loop. In some embodiments, the deaminase of the fusion protein deaminates no more than 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleobases within the range of the R-loop. An R-loop is a three-stranded nucleic acid structure including a DNA:RNA hybrid, a DNA:DNA or an RNA: RNA complementary structure and the
15 associated with single-stranded DNA. As used herein, an R-loop may be formed when a target polynucleotide is contacted with a CRISPR complex or a base editing complex, wherein a portion of a guide polynucleotide, e.g. a guide RNA, hybridizes with and displaces with a portion of a target polynucleotide, e.g. a target DNA. In some embodiments, an R-loop comprises a hybridized region of a spacer sequence and a target
20 DNA complementary sequence. An R-loop region may be of about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleobase pairs in length. In some embodiments, the R-loop region is about 20 nucleobase pairs in length. It should be understood that, as used herein, an R-loop region is not limited to the target DNA strand
25 that hybridizes with the guide polynucleotide. For example, editing of a target nucleobase within an R-loop region may be to a DNA strand that comprises the complementary strand to a guide RNA, or may be to a DNA strand that is the opposing strand of the strand complementary to the guide RNA. In some embodiments, editing in the region of the R-loop comprises editing a nucleobase on non-complementary strand
30 (protospacer strand) to a guide RNA in a target DNA sequence.

The fusion protein described herein can effect target deamination in an editing window different from canonical base editing. In some embodiments, a target nucleobase is from about 1 to about 20 bases upstream of a PAM sequence in the target

polynucleotide sequence. In some embodiments, a target nucleobase is from about 2 to about 12 bases upstream of a PAM sequence in the target polynucleotide sequence. In some embodiments, a target nucleobase is from about 1 to 9 base pairs, about 2 to 10 base pairs, about 3 to 11 base pairs, about 4 to 12 base pairs, about 5 to 13 base pairs, about 6 to 14 base pairs, about 7 to 15 base pairs, about 8 to 16 base pairs, about 9 to 17 base pairs, about 10 to 18 base pairs, about 11 to 19 base pairs, about 12 to 20 base pairs, about 1 to 7 base pairs, about 2 to 8 base pairs, about 3 to 9 base pairs, about 4 to 10 base pairs, about 5 to 11 base pairs, about 6 to 12 base pairs, about 7 to 13 base pairs, about 8 to 14 base pairs, about 9 to 15 base pairs, about 10 to 16 base pairs, about 11 to 17 base pairs, about 12 to 18 base pairs, about 13 to 19 base pairs, about 14 to 20 base pairs, about 1 to 5 base pairs, about 2 to 6 base pairs, about 3 to 7 base pairs, about 4 to 8 base pairs, about 5 to 9 base pairs, about 6 to 10 base pairs, about 7 to 11 base pairs, about 8 to 12 base pairs, about 9 to 13 base pairs, about 10 to 14 base pairs, about 11 to 15 base pairs, about 12 to 16 base pairs, about 13 to 17 base pairs, about 14 to 18 base pairs, about 15 to 19 base pairs, about 16 to 20 base pairs, about 1 to 3 base pairs, about 2 to 4 base pairs, about 3 to 5 base pairs, about 4 to 6 base pairs, about 5 to 7 base pairs, about 6 to 8 base pairs, about 7 to 9 base pairs, about 8 to 10 base pairs, about 9 to 11 base pairs, about 10 to 12 base pairs, about 11 to 13 base pairs, about 12 to 14 base pairs, about 13 to 15 base pairs, about 14 to 16 base pairs, about 15 to 17 base pairs, about 16 to 18 base pairs, about 17 to 19 base pairs, about 18 to 20 base pairs away or upstream of the PAM sequence. In some embodiments, a target nucleobase is about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more base pairs away from or upstream of the PAM sequence. In some embodiments, a target nucleobase is about 1, 2, 3, 4, 5, 6, 7, 8, or 9 base pairs upstream of the PAM sequence. In some embodiments, a target nucleobase is about 2, 3, 4, or 6 base pairs upstream of the PAM sequence.

The fusion protein can comprise more than one heterologous polypeptide. For example, the fusion protein can additionally comprise one or more UGI domains and/or one or more nuclear localization signals. The two or more heterologous domains can be inserted in tandem. The two or more heterologous domains can be inserted at locations such that they are not in tandem in the NapDNAbp.

A fusion protein can comprise a linker between the deaminase and the napDNAbp polypeptide. The linker can be a peptide or a non-peptide linker. For example, the linker can be an XTEN, (GGGS)_n (SEQ ID NO: 171), (GGGGS)_n (SEQ ID NO: 172), (G)_n,

(EAAAK)_n (SEQ ID NO: 173), (GGG)_n, SGSETPGTSESATPES (SEQ ID NO: 174). In some embodiments, the fusion protein comprises a linker between the N-terminal Cas9 fragment and the deaminase. In some embodiments, the fusion protein comprises a linker between the C-terminal Cas9 fragment and the deaminase. In some embodiments, the N-terminal and C-terminal fragments of napDNAbp are connected to the deaminase with a linker. In some embodiments, the N-terminal and C-terminal fragments are joined to the deaminase domain without a linker. In some embodiments, the fusion protein comprises a linker between the N-terminal Cas9 fragment and the deaminase, but does not comprise a linker between the C-terminal Cas9 fragment and the deaminase. In some

5

10

embodiments, the fusion protein comprises a linker between the C-terminal Cas9 fragment and the deaminase, but does not comprise a linker between the N-terminal Cas9 fragment and the deaminase.

In some embodiments, the napDNAbp in the fusion protein is a Cas12 polypeptide, e.g., Cas12b/C2c1, or a fragment thereof. The Cas12 polypeptide can be a variant Cas12 polypeptide. In other embodiments, the N- or C-terminal fragments of the Cas12 polypeptide comprise a nucleic acid programmable DNA binding domain or a RuvC domain. In other embodiments, the fusion protein contains a linker between the Cas12 polypeptide and the catalytic domain. In other embodiments, the amino acid sequence of the linker is GGSGGS (SEQ ID NO: 175) or GSSGSETPGTSESATPESG (SEQ ID NO: 176). In other embodiments, the linker is a rigid linker. In other

15

20

embodiments of the above aspects, the linker is encoded by GGAGGCTCTGGAGGAAGC (SEQ ID NO: 177) or GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGC (SEQ ID NO: 178).

25

Fusion proteins comprising a heterologous catalytic domain flanked by N- and C-terminal fragments of a Cas12 polypeptide are also useful for base editing in the methods as described herein. Fusion proteins comprising Cas12 and one or more deaminase domains, e.g., adenosine deaminase, or comprising an adenosine deaminase domain flanked by Cas12 sequences are also useful for highly specific and efficient base editing

30

of target sequences. In an embodiment, a chimeric Cas12 fusion protein contains a heterologous catalytic domain (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) inserted within a Cas12 polypeptide. In some embodiments, the fusion protein comprises an adenosine deaminase domain and a

cytidine deaminase domain inserted within a Cas12. In some embodiments, an adenosine deaminase is fused within Cas12 and a cytidine deaminase is fused to the C-terminus. In some embodiments, an adenosine deaminase is fused within Cas12 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and an adenosine deaminase is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and an adenosine deaminase fused to the N-terminus. Exemplary structures of a fusion protein with an adenosine deaminase and a cytidine deaminase and a Cas12 are provided as follows:

NH₂-[Cas12(adenosine deaminase)]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[Cas12(adenosine deaminase)]-COOH;
 NH₂-[Cas12(cytidine deaminase)]-[adenosine deaminase]-COOH; or
 NH₂-[adenosine deaminase]-[Cas12(cytidine deaminase)]-COOH;

In some embodiments, the “-” used in the general architecture above indicates the presence of an optional linker.

In various embodiments, the catalytic domain has DNA modifying activity (e.g., deaminase activity), such as adenosine deaminase activity. In some embodiments, the adenosine deaminase is a TadA (e.g., TadA*7.10). In some embodiments, the TadA is a TadA*8. In some embodiments, a TadA*8 is fused within Cas12 and a cytidine deaminase is fused to the C-terminus. In some embodiments, a TadA*8 is fused within Cas12 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and a TadA*8 is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and a TadA*8 fused to the N-terminus. Exemplary structures of a fusion protein with a TadA*8 and a cytidine deaminase and a Cas12 are provided as follows:

N-[Cas12(TadA*8)]-[cytidine deaminase]-C;
 N-[cytidine deaminase]-[Cas12(TadA*8)]-C;
 N-[Cas12(cytidine deaminase)]-[TadA*8]-C; or
 N-[TadA*8]-[Cas12(cytidine deaminase)]-C.

In some embodiments, the “-” used in the general architecture above indicates the presence of an optional linker.

In other embodiments, the fusion protein contains one or more catalytic domains. In other embodiments, at least one of the one or more catalytic domains is inserted within the Cas12 polypeptide or is fused at the Cas12 N-terminus or C-terminus. In other

embodiments, at least one of the one or more catalytic domains is inserted within a loop, an alpha helix region, an unstructured portion, or a solvent accessible portion of the Cas12 polypeptide. In other embodiments, the Cas12 polypeptide is Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other

5 embodiments, the Cas12 polypeptide has at least about 85% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus* sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b (SEQ ID NO: 179). In other

embodiments, the Cas12 polypeptide has at least about 90% amino acid sequence identity to *Bacillus hisashii* Cas12b (SEQ ID NO: 180), *Bacillus thermoamylovorans* Cas12b,

10 *Bacillus* sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b. In other

embodiments, the Cas12 polypeptide has at least about 95% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b (SEQ ID NO: 181), *Bacillus* sp. V3-13 Cas12b (SEQ ID NO: 182), or *Alicyclobacillus acidiphilus* Cas12b. In

other embodiments, the Cas12 polypeptide contains or consists essentially of a fragment

15 of *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus* sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b. In embodiments, the Cas12 polypeptide contains BvCas12b (V4), which in some embodiments is expressed as 5' mRNA Cap---5' UTR---bhCas12b---STOP sequence --- 3' UTR --- 120polyA tail (SEQ ID NOs: 183-185).

20 In other embodiments, the catalytic domain is inserted between amino acid positions 153-154, 255-256, 306-307, 980-981, 1019-1020, 534-535, 604-605, or 344-345 of BhCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic

domain is inserted between amino acids P153 and S154 of BhCas12b. In other

25 embodiments, the catalytic domain is inserted between amino acids K255 and E256 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids D980 and G981 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K1019 and L1020 of BhCas12b. In other embodiments, the

catalytic domain is inserted between amino acids F534 and P535 of BhCas12b. In other

30 embodiments, the catalytic domain is inserted between amino acids K604 and G605 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids H344 and F345 of BhCas12b. In other embodiments, catalytic domain is inserted between amino acid positions 147 and 148, 248 and 249, 299 and 300, 991 and 992, or

1031 and 1032 of BvCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P147 and D148 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids G248 and G249 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids P299 and E300 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids G991 and E992 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids K1031 and M1032 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acid positions 157 and 158, 258 and 259, 310 and 311, 1008 and 1009, or 1044 and 1045 of AaCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P157 and G158 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids V258 and G259 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids D310 and P311 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids G1008 and E1009 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids G1044 and K1045 at of AaCas12b.

In other embodiments, the fusion protein contains a nuclear localization signal (e.g., a bipartite nuclear localization signal). In other embodiments, the amino acid sequence of the nuclear localization signal is MAPKKKRKVGIHGVPAA (SEQ ID NO: 186). In other embodiments of the above aspects, the nuclear localization signal is encoded by the following sequence:

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC (SEQ ID NO: 187). In other embodiments, the Cas12b polypeptide contains a mutation that silences the catalytic activity of a RuvC domain. In other embodiments, the Cas12b polypeptide contains D574A, D829A and/or D952A mutations. In other embodiments, the fusion protein further contains a tag (e.g., an influenza hemagglutinin tag).

In some embodiments, the fusion protein comprises a napDNAbp domain (e.g., Cas12-derived domain) with an internally fused nucleobase editing domain (e.g., all or a portion of a deaminase domain, e.g., an adenosine deaminase domain). In some embodiments, the napDNAbp is a Cas12b. In some embodiments, the base editor

comprises a BhCas12b domain with an internally fused TadA*8 domain inserted at the loci provided in **Table 5** below.

Table 5: Insertion loci in Cas12b proteins

BhCas12b	Insertion site	Inserted between aa
position 1	153	PS
position 2	255	KE
position 3	306	DE
position 4	980	DG
position 5	1019	KL
position 6	534	FP
position 7	604	KG
position 8	344	HF
BvCas12b	Insertion site	Inserted between aa
position 1	147	PD
position 2	248	GG
position 3	299	PE
position 4	991	GE
position 5	1031	KM
AaCas12b	Insertion site	Inserted between aa
position 1	157	PG
position 2	258	VG
position 3	310	DP
position 4	1008	GE
position 5	1044	GK

By way of nonlimiting example, an adenosine deaminase (e.g., TadA*8.13) may be inserted into a BhCas12b to produce a fusion protein (e.g., TadA*8.13-BhCas12b) that effectively edits a nucleic acid sequence.

In some embodiments, the base editing system described herein is an ABE with
5 TadA inserted into a Cas9. Polypeptide sequences of relevant ABEs with TadA inserted into a Cas9 are provided in the attached Sequence Listing as SEQ ID NOs: 188-233.

In some embodiments, adenosine deaminase base editors were generated to insert TadA or variants thereof into the Cas9 polypeptide at the identified positions.

Exemplary, yet nonlimiting, fusion proteins are described in International PCT
10 Application Nos. PCT/US2020/016285 and U.S. Provisional Application Nos. 62/852,228 and 62/852,224, the contents of which are incorporated by reference herein in their entireties.

A to G Editing

15 In some embodiments, a base editor described herein comprises an adenosine deaminase domain. Such an adenosine deaminase domain of a base editor can facilitate the editing of an adenine (A) nucleobase to a guanine (G) nucleobase by deaminating the A to form inosine (I), which exhibits base pairing properties of G. Adenosine deaminase is capable of deaminating (i.e., removing an amine group) adenine of a deoxyadenosine
20 residue in deoxyribonucleic acid (DNA). In some embodiments, an A-to-G base editor further comprises an inhibitor of inosine base excision repair, for example, a uracil glycosylase inhibitor (UGI) domain or a catalytically inactive inosine specific nuclease. Without wishing to be bound by any particular theory, the UGI domain or catalytically inactive inosine specific nuclease can inhibit or prevent base excision repair of a
25 deaminated adenosine residue (e.g., inosine), which can improve the activity or efficiency of the base editor.

A base editor comprising an adenosine deaminase can act on any polynucleotide, including DNA, RNA and DNA-RNA hybrids. In certain embodiments, a base editor comprising an adenosine deaminase can deaminate a target A of a polynucleotide
30 comprising RNA. For example, the base editor can comprise an adenosine deaminase domain capable of deaminating a target A of an RNA polynucleotide and/or a DNA-RNA hybrid polynucleotide. In an embodiment, an adenosine deaminase incorporated into a base editor comprises all or a portion of adenosine deaminase acting on RNA (ADAR,

e.g., ADAR1 or ADAR2) or tRNA (ADAT). A base editor comprising an adenosine deaminase domain can also be capable of deaminating an A nucleobase of a DNA polynucleotide. In an embodiment an adenosine deaminase domain of a base editor comprises all or a portion of an ADAT comprising one or more mutations which permit the ADAT to deaminate a target A in DNA. For example, the base editor can comprise all or a portion of an ADAT from *Escherichia coli* (EcTadA) comprising one or more of the following mutations: D108N, A106V, D147Y, E155V, L84F, H123Y, I156F, or a corresponding mutation in another adenosine deaminase. Exemplary ADAT homolog polypeptide sequences are provided in the Sequence Listing as SEQ ID NOs: 234-241.

The adenosine deaminase can be derived from any suitable organism (e.g., *E. coli*). In some embodiments, the adenosine deaminase is from a prokaryote. In some embodiments, the adenosine deaminase is from a bacterium. In some embodiments, the adenosine deaminase is from *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Shewanella putrefaciens*, *Haemophilus influenzae*, *Caulobacter crescentus*, or *Bacillus subtilis*. In some embodiments, the adenosine deaminase is from *E. coli*. In some embodiments, the adenine deaminase is a naturally-occurring adenosine deaminase that includes one or more mutations corresponding to any of the mutations provided herein (e.g., mutations in ecTadA). The corresponding residue in any homologous protein can be identified by e.g., sequence alignment and determination of homologous residues. The mutations in any naturally-occurring adenosine deaminase (e.g., having homology to ecTadA) that correspond to any of the mutations described herein (e.g., any of the mutations identified in ecTadA) can be generated accordingly.

In some embodiments, the adenosine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the amino acid sequences set forth in any of the adenosine deaminases provided herein. It should be appreciated that adenosine deaminases provided herein may include one or more mutations (e.g., any of the mutations provided herein). The disclosure provides any deaminase domains with a certain percent identity plus any of the mutations or combinations thereof described herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has 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, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46,

47, 48, 49, 50, or more mutations compared to a reference sequence, or any of the adenosine deaminases provided herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, or at least 170 identical contiguous amino acid residues as compared to any one of the amino acid sequences known in the art or described herein.

It should be appreciated that any of the mutations provided herein (*e.g.*, based on the TadA reference sequence) can be introduced into other adenosine deaminases, such as *E. coli* TadA (ecTadA), *S. aureus* TadA (saTadA), or other adenosine deaminases (*e.g.*, bacterial adenosine deaminases). It would be apparent to the skilled artisan that additional deaminases may similarly be aligned to identify homologous amino acid residues that can be mutated as provided herein. Thus, any of the mutations identified in the TadA reference sequence can be made in other adenosine deaminases (*e.g.*, ecTadA) that have homologous amino acid residues. It should also be appreciated that any of the mutations provided herein can be made individually or in any combination in the TadA reference sequence or another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises a D108X mutation in the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108G, D108N, D108V, D108A, or D108Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase. It should be appreciated, however, that additional deaminases may similarly be aligned to identify homologous amino acid residues that can be mutated as provided herein.

In some embodiments, the adenosine deaminase comprises an A106X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A106V mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises a E155X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase,

where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a E155D, E155G, or E155V mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

5 In some embodiments, the adenosine deaminase comprises a D147X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D147Y, mutation in TadA reference sequence, or a corresponding
10 mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A106X, E155X, or D147X, mutation in the TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments,
15 the adenosine deaminase comprises an E155D, E155G, or E155V mutation. In some embodiments, the adenosine deaminase comprises a D147Y.

It should also be appreciated that any of the mutations provided herein may be made individually or in any combination in ecTadA or another adenosine deaminase. For example, an adenosine deaminase may contain a D108N, a A106V, a E155V, and/or a
20 D147Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, an adenosine deaminase comprises the following group of mutations (groups of mutations are separated by a ";") in TadA reference sequence, or corresponding mutations in another adenosine deaminase:
D108N and A106V; D108N and E155V; D108N and D147Y; A106V and E155V;
25 A106V and D147Y; E155V and D147Y; D108N, A106V, and E155V; D108N, A106V, and D147Y; D108N, E155V, and D147Y; A106V, E155V, and D147Y; and D108N, A106V, E155V, and D147Y. It should be appreciated, however, that any combination of corresponding mutations provided herein may be made in an adenosine deaminase (*e.g.*, ecTadA).

30 In some embodiments, the adenosine deaminase comprises one or more of a H8X, T17X, L18X, W23X, L34X, W45X, R51X, A56X, E59X, E85X, M94X, I95X, V102X, F104X, A106X, R107X, D108X, K110X, M118X, N127X, A138X, F149X, M151X, R153X, Q154X, I156X, and/or K157X mutation in TadA reference sequence, or one or

more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y, T17S, L18E, W23L, L34S, W45L, R51H, A56E, or A56S, E59G, E85K, or
 5 E85G, M94L, I95L, V102A, F104L, A106V, R107C, or R107H, or R107P, D108G, or D108N, or D108V, or D108A, or D108Y, K110I, M118K, N127S, A138V, F149Y, M151V, R153C, Q154L, I156D, and/or K157R mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one or more of a H8X,
 10 D108X, and/or N127X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where X indicates the presence of any amino acid. In some embodiments, the adenosine deaminase comprises one or more of a H8Y, D108N, and/or N127S mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

15 In some embodiments, the adenosine deaminase comprises one or more of H8X, R26X, M61X, L68X, M70X, A106X, D108X, A109X, N127X, D147X, R152X, Q154X, E155X, K161X, Q163X, and/or T166X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type
 20 adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y, R26W, M61I, L68Q, M70V, A106T, D108N, A109T, N127S, D147Y, R152C, Q154H or Q154R, E155G or E155V or E155D, K161Q, Q163H, and/or T166P mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

25 In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, D108X, N127X, D147X, R152X, and Q154X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA), where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine
 30 deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8X, M61X, M70X, D108X, N127X, Q154X, E155X, and Q163X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*,

ecTadA), where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8X, D108X, N127X, E155X, and T166X in TadA reference sequence, or a
5 corresponding mutation or mutations in another adenosine deaminase (e.g., ecTadA), where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, A106X, and D108X, or
10 a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8X, R26X, L68X, D108X, N127X, D147X, and E155X, or a
15 corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of H8X, R126X, L68X,
20 D108X, N127X, D147X, and E155X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five mutations selected from the group consisting of H8X, D108X, A109X,
25 N127X, and E155X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8Y, D108N, N127S,
30 D147Y, R152C, and Q154H in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (e.g., ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, M61I, M70V, D108N, N127S, Q154R,

E155G and Q163H in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, N127S, E155V, and T166P in TadA reference
5 sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8Y, A106T, D108N, N127S, E155D, and K161Q in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the
10 adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, R26W, L68Q, D108N, N127S, D147Y, and E155V in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of
15 H8Y, D108N, A109T, N127S, and E155G in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of the or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108N, D108G, or D108V mutation
20 in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A106V and D108N mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises R107C and D108N mutations in TadA reference sequence, or corresponding mutations in another
25 adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and Q154H mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase.
30 In some embodiments, the adenosine deaminase comprises a D108N, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, and N127S mutation in TadA reference sequence, or corresponding mutations in another

adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A106V, D108N, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of S2X, H8X, I49X, L84X, H123X, N127X, I156X, and/or K160X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of S2A, H8Y, I49F, L84F, H123Y, N127S, I156F, and/or K160S mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an L84X mutation adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an L84F mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an H123X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an H123Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an I156X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an I156F mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of L84X, A106X, D108X, H123X, D147X, E155X, and I156X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine

deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of S2X, I49X, A106X, D108X, D147X, and E155X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five mutations selected from the group consisting of H8X, A106X, D108X, N127X, and K160X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of L84F, A106V, D108N, H123Y, D147Y, E155V, and I156F in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of S2A, I49F, A106V, D108N, D147Y, and E155V in TadA reference sequence.

In some embodiments, the adenosine deaminase comprises one, two, three, four, or five mutations selected from the group consisting of H8Y, A106T, D108N, N127S, and K160S in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one or more of a E25X, R26X, R107X, A142X, and/or A143X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of E25M, E25D, E25A, E25R, E25V, E25S, E25Y, R26G, R26N, R26Q, R26C, R26L, R26K, R107P, R107K, R107A, R107N, R107W, R107H, R107S, A142N, A142D, A142G, A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q, and/or A143R mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of the mutations described herein

corresponding to TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an E25X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an E25M, E25D, E25A, E25R, E25V, E25S, or E25Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R26X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises R26G, R26N, R26Q, R26C, R26L, or R26K mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R107X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an R107P, R107K, R107A, R107N, R107W, R107H, or R107S mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A142X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A142N, A142D, A142G, mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A143X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q, and/or A143R mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of a H36X, N37X, P48X, I49X, R51X, M70X, N72X, D77X, E134X, S146X, Q154X, K157X, and/or K161X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H36L, N37T, N37S, P48T, P48L, I49V, R51H, R51L, M70L, N72S, D77G, E134G, S146R, S146C, Q154H, K157N, and/or K161T mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an H36X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an H36L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an N37X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an N37T or N37S mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an P48X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an P48T or P48L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an R51X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an

R51H or R51L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an S146X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an S146R or S146C mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an K157X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a K157N mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an P48X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a P48S, P48T, or P48A mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an A142X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A142N mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an W23X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a W23R or W23L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an R152X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a
 5 R152P or R52H mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In one embodiment, the adenosine deaminase may comprise the mutations H36L, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F, and K157N. In some embodiments, the adenosine deaminase comprises the following combination of
 10 mutations relative to TadA reference sequence, where each mutation of a combination is separated by a "_" and each combination of mutations is between parentheses:

(A106V_D108N),
 (R107C_D108N),
 (H8Y_D108N_N127S_D147Y_Q154H),
 15 (H8Y_D108N_N127S_D147Y_E155V),
 (D108N_D147Y_E155V),
 (H8Y_D108N_N127S),
 (H8Y_D108N_N127S_D147Y_Q154H),
 (A106V_D108N_D147Y_E155V),
 20 (D108Q_D147Y_E155V),
 (D108M_D147Y_E155V),
 (D108L_D147Y_E155V),
 (D108K_D147Y_E155V),
 (D108I_D147Y_E155V),
 25 (D108F_D147Y_E155V),
 (A106V_D108N_D147Y),
 (A106V_D108M_D147Y_E155V),
 (E59A_A106V_D108N_D147Y_E155V),
 (E59A cat dead_A106V_D108N_D147Y_E155V),
 30 (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y),
 (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (D103A_D104N),
 (G22P_D103A_D104N),
 (D103A_D104N_S138A),
 35 (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

- (E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
 (E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F),
 (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
 (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
 5 (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F),
 (L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F),
 (R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
 (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F),
 (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
 10 (A106V_D108N_A142N_D147Y_E155V),
 (R26G_A106V_D108N_A142N_D147Y_E155V),
 (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V),
 (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V),
 (E25D_R26G_A106V_D108N_A142N_D147Y_E155V),
 15 (A106V_R107K_D108N_A142N_D147Y_E155V),
 (A106V_D108N_A142N_A143G_D147Y_E155V),
 (A106V_D108N_A142N_A143L_D147Y_E155V),
 (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
 (N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),
 20 (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T),
 (H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),
 (N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F),
 (H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),
 (H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)
 25 (H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F),
 (L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),
 (N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
 (D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),
 30 (H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F),
 (Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F),
 (E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
 (L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),
 35 (P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

- (D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
 (L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
 (H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N),
 (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),
 5 (L84F_A106V_D108N_D147Y_E155V_I156F),
 (R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),
 (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T),
 (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),
 (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E),
 10 (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F),
 15 (L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F),
 (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
 (P48S_A142N),
 (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),
 (P48T_I49V_A142N),
 20 (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
 (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F
 (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
 _K157N),
 (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_
 25 I156F_K157N),
 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F
 _K157N),
 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F
 30 _K157N),
 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
 _K157N),
 (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
 _K157N),
 35 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F
 _K161T),

- (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N),
 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N),
 5 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N),
 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F_K157N),
 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N),
 10 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),
 (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N),
 15 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155V_I156F_K157N).

In some embodiments, the TadA deaminase is TadA variant. In some embodiments, the TadA variant is TadA*7.10. In particular embodiments, the fusion proteins comprise a single TadA*7.10 domain (*e.g.*, provided as a monomer). In other
 20 embodiments, the fusion protein comprises TadA*7.10 and TadA(wt), which are capable of forming heterodimers. In one embodiment, a fusion protein of the invention comprises a wild-type TadA linked to TadA*7.10, which is linked to Cas9 nickase.

In some embodiments, TadA*7.10 comprises at least one alteration. In some embodiments, the adenosine deaminase comprises an alteration in the following
 25 sequence:

TadA*7.10
 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHD
 PTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
 NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQK
 30 KAQSSTD (SEQ ID NO: 3)

In some embodiments, TadA*7.10 comprises an alteration at amino acid 82 and/or 166. In particular embodiments, TadA*7.10 comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R. In other embodiments, a variant of TadA*7.10 comprises a combination of alterations selected

from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; 5 V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R.

In some embodiments, an adenosine deaminase variant (*e.g.*, TadA*8) comprises a deletion. In some embodiments, an adenosine deaminase variant comprises a deletion of the C terminus. In particular embodiments, an adenosine deaminase variant comprises a deletion of the C terminus beginning at residue 149, 150, 151, 152, 153, 154, 155, 156, 10 and 157, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, an adenosine deaminase variant (*e.g.*, TadA*8) is a monomer comprising one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference 15 sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) is a monomer comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + 20 Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a homodimer 25 comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having a combination 30 of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R

+ Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*,
 5 TadA*8) comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain
 10 and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R
 15 + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8)
 20 comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the
 25 group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to
 30 TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In particular embodiments, an adenosine deaminase heterodimer comprises a TadA*8 domain and an adenosine deaminase domain selected from *Staphylococcus aureus* (*S. aureus*) TadA, *Bacillus subtilis* (*B. subtilis*) TadA, *Salmonella typhimurium* (*S.*

typhimurium) TadA, *Shewanella putrefaciens* (*S. putrefaciens*) TadA, *Haemophilus influenzae* F3031 (*H. influenzae*) TadA, *Caulobacter crescentus* (*C. crescentus*) TadA, *Geobacter sulfurreducens* (*G. sulfurreducens*) TadA, or TadA*7.10.

In some embodiments, an adenosine deaminase is a TadA*8. In one embodiment, an adenosine deaminase is a TadA*8 that comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHD
PTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQK
KAQSSTD (SEQ ID NO: 242)

In some embodiments, the TadA*8 is truncated. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length TadA*8. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length TadA*8. In some embodiments the adenosine deaminase variant is a full-length TadA*8.

In some embodiments the TadA*8 is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24.

In other embodiments, a base editor of the disclosure comprising an adenosine deaminase variant (*e.g.*, TadA*8) monomer comprising one or more of the following alterations: R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) monomer comprises a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor comprises a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the base editor comprises a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor comprises a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the base editor comprises a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, the TadA*8 is a variant as shown in **Table 6**. **Table 6** shows certain amino acid position numbers in the TadA amino acid sequence and the amino acids present in those positions in the TadA-7.10 adenosine deaminase. **Table 6** also shows amino acid changes in TadA variants relative to TadA-7.10 following phage-assisted non-continuous evolution (PANCE) and phage-assisted continuous evolution (PACE), as described in M. Richter *et al.*, 2020, Nature Biotechnology, doi.org/10.1038/s41587-020-0453-z, the entire contents of which are incorporated by

reference herein. In some embodiments, the TadA*8 is TadA*8a, TadA*8b, TadA*8c, TadA*8d, or TadA*8e. In some embodiments, the TadA*8 is TadA*8e.

Table 6. Select TadA*8 Variants

	TadA	TadA amino acid number									
		26	88	109	111	119	122	147	149	166	167
	TadA-7.10	R	V	A	T	D	H	Y	F	T	D
PANCE 1					R						
PANCE 2				S/T	R						
	TadA-8a	C		S	R	N	N	D	Y	I	N
	TadA-8b		A	S	R	N	N		Y	I	N
PACE	TadA-8c	C		S	R	N	N		Y	I	N
	TadA-8d		A		R	N			Y		
	TadA-8e			S	R	N	N	D	Y	I	N

5

In one embodiment, a fusion protein of the invention comprises a wild-type TadA is linked to an adenosine deaminase variant described herein (*e.g.*, TadA*8), which is linked to Cas9 nickase. In particular embodiments, the fusion proteins comprise a single TadA*8 domain (*e.g.*, provided as a monomer). In other embodiments, the fusion protein comprises TadA*8 and TadA(wt), which are capable of forming heterodimers.

In some embodiments, the adenosine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the amino acid sequences set forth in any of the adenosine deaminases provided herein. It should be appreciated that adenosine deaminases provided herein may include one or more mutations (*e.g.*, any of the mutations provided herein). The disclosure provides any deaminase domains with a certain percent identity plus any of the mutations or combinations thereof described herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more mutations compared to a reference sequence, or any of the adenosine deaminases provided herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has at least 5, at least 10, at least 15, at least 20, at

15

20

least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, or at least 170 identical contiguous amino acid residues as compared to any one of the amino acid sequences known in the art or described herein.

5 In particular embodiments, a TadA*8 comprises one or more mutations at any of the following positions shown in bold. In other embodiments, a TadA*8 comprises one or more mutations at any of the positions shown with underlining:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVN⁵⁰NRV IGEGWNRAIG
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG¹⁰⁰
10 RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR¹⁵⁰
MPRQVFNAQK KAQSSTD (SEQ ID NO: 3)

For example, the TadA*8 comprises alterations at amino acid position 82 and/or 166 (*e.g.*, V82S, T166R) alone or in combination with any one or more of the following Y147T, Y147R, Q154S, Y123H, and/or Q154R, relative to TadA*7.10, the TadA
15 reference sequence, or a corresponding mutation in another TadA. In particular embodiments, a combination of alterations is selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y;
20 Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, the TadA*8 is truncated. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19,
25 or 20 N-terminal amino acid residues relative to the full length TadA*8. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length TadA*8. In some embodiments the adenosine deaminase variant is a full-length TadA*8.

In one embodiment, a fusion protein of the invention comprises a wild-type TadA
30 is linked to an adenosine deaminase variant described herein (*e.g.*, TadA*8), which is linked to Cas9 nickase. In particular embodiments, the fusion proteins comprise a single TadA*8 domain (*e.g.*, provided as a monomer). In other embodiments, the base editor comprises TadA*8 and TadA(wt), which are capable of forming heterodimers.

In particular embodiments, the fusion proteins comprise a single (*e.g.*, provided as a monomer) TadA*8. In some embodiments, the TadA*8 is linked to a Cas9 nickase. In some embodiments, the fusion proteins of the invention comprise as a heterodimer of a wild-type TadA (TadA(wt)) linked to a TadA*8. In other embodiments, the fusion proteins of the invention comprise as a heterodimer of a TadA*7.10 linked to a TadA*8. In some embodiments, the base editor is ABE8 comprising a TadA*8 variant monomer. In some embodiments, the base editor is ABE8 comprising a heterodimer of a TadA*8 and a TadA(wt). In some embodiments, the base editor is ABE8 comprising a heterodimer of a TadA*8 and TadA*7.10. In some embodiments, the base editor is ABE8 comprising a heterodimer of a TadA*8. In some embodiments, the TadA*8 is selected from **Table 6, 12, or 13**. In some embodiments, the ABE8 is selected from **Table 12, 13, or 15**.

In some embodiments, the adenosine deaminase is a TadA*9 variant. In some embodiments, the adenosine deaminase is a TadA*9 variant selected from the variants described below and with reference to the following sequence (termed TadA*7.10):

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
LHDPTHAHEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITTEGILADE CAALLCYFFR
MPRQVFNAQK KAQSSTD (SEQ ID NO: 3)

In some embodiments, an adenosine deaminase comprises one or more of the following alterations: R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K, D139L, D139M, C146R, and A158K. The one or more alterations are shown in the sequence above in underlining and bold font.

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: V82S + Q154R + Y147R; V82S + Q154R + Y123H; V82S + Q154R + Y147R + Y123H; Q154R + Y147R + Y123H + I76Y + V82S; V82S + I76Y; V82S + Y147R; V82S + Y147R + Y123H; V82S + Q154R + Y123H; Q154R + Y147R + Y123H + I76Y; V82S + Y147R; V82S + Y147R + Y123H; V82S + Q154R + Y123H; V82S + Q154R + Y147R; V82S + Q154R + Y147R; Q154R + Y147R + Y123H + I76Y; Q154R + Y147R + Y123H + I76Y + V82S; I76Y_V82S_Y123H_Y147R_Q154R; Y147R + Q154R + H123H; and V82S + Q154R.

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: E25F + V82S + Y123H, T133K + Y147R +

Q154R; E25F + V82S + Y123H + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R + Q154R; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R;
 5 Q71M + V82S + Y123H + Y147R + Q154R; E25F + V82S + Y123H + T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; V82S + Y123H + P124W + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R + Q154R; R23H + V82S + Y123H + Y147R + Q154R; R21N + V82S +
 10 Y123H + Y147R + Q154R; V82S + Y123H + Y147R + Q154R + A158K; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; and M70V + V82S + M94V + Y123H + Y147R + Q154R

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: Q71M + V82S + Y123H + Y147R + Q154R;
 15 E25F + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D139M + Y147R + Q154R; Y73S + I76Y + V82S + Y123H + Y147R + Q154R; E25F + I76Y + V82S +
 20 Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D139M + Y147R + Q154R; Y73S + I76Y + V82S + Y123H + Y147R + Q154R; and V82S + Q154R; N72K_V82S + Y123H +
 25 Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R; N72K_V82S + Y123H + Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; M70V + V82S + M94V + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H +
 30 T133K + Y147R + Q154R + A158K; and M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R. In some embodiments, the adenosine deaminase is expressed as a monomer. In other embodiments, the adenosine deaminase is expressed as a heterodimer. In some embodiments, the deaminase or other polypeptide sequence lacks a methionine,

for example when included as a component of a fusion protein. This can alter the numbering of positions. However, the skilled person will understand that such corresponding mutations refer to the same mutation, *e.g.*, Y73S and Y72S and D139M and D138M.

5 In some embodiments, the TadA*9 variant comprises the alterations described in **Table 16** as described herein. In some embodiments, the TadA*9 variant is a monomer. In some embodiments, the TadA*9 variant is a heterodimer with a wild-type TadA adenosine deaminase. In some embodiments, the TadA*9 variant is a heterodimer with another TadA variant (*e.g.*, TadA*8, TadA*9). Additional details of TadA*9 adenosine
10 deaminases are described in International PCT Application No. PCT/2020/049975, which is incorporated herein by reference for its entirety.

Any of the mutations provided herein and any additional mutations (*e.g.*, based on the ecTadA amino acid sequence) can be introduced into any other adenosine deaminases. Any of the mutations provided herein can be made individually or in any combination in
15 TadA reference sequence or another adenosine deaminase (*e.g.*, ecTadA).

Details of A to G nucleobase editing proteins are described in International PCT Application No. PCT/2017/045381 (WO2018/027078) and Gaudelli, N.M., *et al.*, “Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage” *Nature*, 551, 464-471 (2017), the entire contents of which are hereby incorporated by
20 reference.

Guide Polynucleotides

A polynucleotide programmable nucleotide binding domain, when in conjunction with a bound guide polynucleotide (*e.g.*, gRNA), can specifically bind to a target
25 polynucleotide sequence (*i.e.*, via complementary base pairing between bases of the bound guide nucleic acid and bases of the target polynucleotide sequence) and thereby localize the base editor to the target nucleic acid sequence desired to be edited. In some embodiments, the target polynucleotide sequence comprises single-stranded DNA or double-stranded DNA. In some embodiments, the target polynucleotide sequence
30 comprises RNA. In some embodiments, the target polynucleotide sequence comprises a DNA-RNA hybrid.

CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR

clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems, correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, and then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See, e.g., Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. See e.g., "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti, J.J. et al., Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E. et al., Nature 471:602-607(2011); and "Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M. et al., Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference).

The PAM sequence can be any PAM sequence known in the art. Suitable PAM sequences include, but are not limited to, NGG, NGA, NGC, NGN, NGT, NGCG, NGAG, NGAN, NGNG, NGCN, NGCG, NGTN, NNGRRT, NNNRRT, NNGRR(N), TTTV, TYCV, TYCV, TATV, NNNNGATT, NNAGAAW, or NAAAAC. Y is a pyrimidine; N is any nucleotide base; W is A or T.

In an embodiment, a guide polynucleotide described herein can be RNA or DNA. In one embodiment, the guide polynucleotide is a gRNA. An RNA/Cas complex can assist in "guiding" a Cas protein to a target DNA. Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gRNA")

can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. *See, e.g.,* Jinek M. *et al.*, Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference.

In some embodiments, the guide polynucleotide is at least one single guide RNA ("sgRNA" or "gNRA"). In some embodiments, a guide polynucleotide comprises two or more individual polynucleotides, which can interact with one another via, for example, complementary base pairing (*e.g.*, a dual guide polynucleotide, dual gRNA). For example, a guide polynucleotide can comprise a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA) or can comprise one or more trans-activating CRISPR RNA (tracrRNA).

In some embodiments, the guide polynucleotide is at least one tracrRNA. In some embodiments, the guide polynucleotide does not require PAM sequence to guide the polynucleotide-programmable DNA-binding domain (*e.g.*, Cas9 or Cpf1) to the target nucleotide sequence.

A guide polynucleotide may include natural or non-natural (or unnatural) nucleotides (*e.g.*, peptide nucleic acid or nucleotide analogs). In some cases, the targeting region of a guide nucleic acid sequence can be at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. A targeting region of a guide nucleic acid can be between 10-30 nucleotides in length, or between 15-25 nucleotides in length, or between 15-20 nucleotides in length.

In some embodiments, the base editor provided herein utilizes one or more guide polynucleotide (*e.g.*, multiple gRNA). In some embodiments, a single guide polynucleotide is utilized for different base editors described herein. For example, a single guide polynucleotide can be utilized for a cytidine base editor and an adenosine base editor.

In some embodiments, the methods described herein can utilize an engineered Cas protein. A guide RNA (gRNA) is a short synthetic RNA composed of a scaffold sequence necessary for Cas-binding and a user-defined ~20 nucleotide spacer that defines the genomic target to be modified. Exemplary gRNA scaffold sequences are provided in the sequence listing as SEQ ID NOs: 90 and 243-252. Thus, a skilled artisan can change the genomic target of the Cas protein specificity is partially determined by how specific the gRNA targeting sequence is for the genomic target compared to the rest of the genome.

In other embodiments, a guide polynucleotide can comprise both the polynucleotide targeting portion of the nucleic acid and the scaffold portion of the nucleic acid in a single molecule (*i.e.*, a single-molecule guide nucleic acid). For example, a single-molecule guide polynucleotide can be a single guide RNA (sgRNA or gRNA).

5 Herein the term guide polynucleotide sequence contemplates any single, dual or multi-molecule nucleic acid capable of interacting with and directing a base editor to a target polynucleotide sequence.

Typically, a guide polynucleotide (*e.g.*, crRNA/trRNA complex or a gRNA) comprises a “polynucleotide-targeting segment” that includes a sequence capable of
10 recognizing and binding to a target polynucleotide sequence, and a “protein-binding segment” that stabilizes the guide polynucleotide within a polynucleotide programmable nucleotide binding domain component of a base editor. In some embodiments, the polynucleotide targeting segment of the guide polynucleotide recognizes and binds to a DNA polynucleotide, thereby facilitating the editing of a base in DNA. In other cases,
15 the polynucleotide targeting segment of the guide polynucleotide recognizes and binds to an RNA polynucleotide, thereby facilitating the editing of a base in RNA. Herein a “segment” refers to a section or region of a molecule, *e.g.*, a contiguous stretch of nucleotides in the guide polynucleotide. A segment can also refer to a region/section of a complex such that a segment can comprise regions of more than one molecule. For
20 example, where a guide polynucleotide comprises multiple nucleic acid molecules, the protein-binding segment of can include all or a portion of multiple separate molecules that are for instance hybridized along a region of complementarity. In some embodiments, a protein-binding segment of a DNA-targeting RNA that comprises two separate molecules can comprise (i) base pairs 40-75 of a first RNA molecule that is 100
25 base pairs in length; and (ii) base pairs 10-25 of a second RNA molecule that is 50 base pairs in length. The definition of “segment,” unless otherwise specifically defined in a particular context, is not limited to a specific number of total base pairs, is not limited to any particular number of base pairs from a given RNA molecule, is not limited to a particular number of separate molecules within a complex, and can include regions of
30 RNA molecules that are of any total length and can include regions with complementarity to other molecules.

The guide polynucleotides can be synthesized chemically, synthesized enzymatically, or a combination thereof. For example, the gRNA can be synthesized

using standard phosphoramidite-based solid-phase synthesis methods. Alternatively, the gRNA can be synthesized *in vitro* by operably linking DNA encoding the gRNA to a promoter control sequence that is recognized by a phage RNA polymerase. Examples of suitable phage promoter sequences include T7, T3, SP6 promoter sequences, or variations thereof. In embodiments in which the gRNA comprises two separate molecules (e.g., crRNA and tracrRNA), the crRNA can be chemically synthesized and the tracrRNA can be enzymatically synthesized.

A gRNA molecule can be transcribed *in vitro*.

A guide polynucleotide may be expressed, for example, by a DNA that encodes the gRNA, e.g., a DNA vector comprising a sequence encoding the gRNA. The gRNA may be encoded alone or together with an encoded base editor. Such DNA sequences may be introduced into an expression system, e.g., a cell, together or separately. For example, DNA sequences encoding a polynucleotide programmable nucleotide binding domain and a gRNA may be introduced into a cell, each DNA sequence can be part of a separate molecule (e.g., one vector containing the polynucleotide programmable nucleotide binding domain coding sequence and a second vector containing the gRNA coding sequence) or both can be part of a same molecule (e.g., one vector containing coding (and regulatory) sequence for both the polynucleotide programmable nucleotide binding domain and the gRNA). An RNA can be transcribed from a synthetic DNA molecule, e.g., a gBlocks® gene fragment.

A gRNA or a guide polynucleotide can comprise three regions: a first region at the 5' end that can be complementary to a target site in a chromosomal sequence, a second internal region that can form a stem loop structure, and a third 3' region that can be single-stranded. A first region of each gRNA can also be different such that each gRNA guides a fusion protein to a specific target site. Further, second and third regions of each gRNA can be identical in all gRNAs.

A first region of a gRNA or a guide polynucleotide can be complementary to sequence at a target site in a chromosomal sequence such that the first region of the gRNA can base pair with the target site. In some cases, a first region of a gRNA can comprise from or from about 10 nucleotides to 25 nucleotides (*i.e.*, from 10 nucleotides to nucleotides; or from about 10 nucleotides to about 25 nucleotides; or from 10 nucleotides to about 25 nucleotides; or from about 10 nucleotides to 25 nucleotides) or more. For example, a region of base pairing between a first region of a gRNA and a target site in a

chromosomal sequence can be or can be about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, or more nucleotides in length. Sometimes, a first region of a gRNA can be or can be about 19, 20, or 21 nucleotides in length.

5 A gRNA or a guide polynucleotide can also comprise a second region that forms a secondary structure. For example, a secondary structure formed by a gRNA can comprise a stem (or hairpin) and a loop. A length of a loop and a stem can vary. For example, a loop can range from or from about 3 to 10 nucleotides in length, and a stem can range from or from about 6 to 20 base pairs in length. A stem can comprise one or more bulges of 1 to 10 or about 10 nucleotides. The overall length of a second region can range from 10 or from about 16 to 60 nucleotides in length. For example, a loop can be or can be about 4 nucleotides in length and a stem can be or can be about 12 base pairs.

A gRNA or a guide polynucleotide can also comprise a third region at the 3' end that can be essentially single-stranded. For example, a third region is sometimes not complementarity to any chromosomal sequence in a cell of interest and is sometimes not 15 complementarity to the rest of a gRNA. Further, the length of a third region can vary. A third region can be more than or more than about 4 nucleotides in length. For example, the length of a third region can range from or from about 5 to 60 nucleotides in length.

A gRNA or a guide polynucleotide can target any exon or intron of a gene target. In some cases, a guide can target exon 1 or 2 of a gene, in other cases; a guide can target 20 exon 3 or 4 of a gene. In some embodiments, a composition comprises multiple gRNAs that all target the same exon or multiple gRNAs that target different exons. An exon and/or an intron of a gene can be targeted.

A gRNA or a guide polynucleotide can target a nucleic acid sequence of about 20 nucleotides or less than about 20 nucleotides (e.g., at least about 5, 10, 15, 16, 17, 18, 19, 25 20, 21, 22, 23, 24, 25, 30 nucleotides), or anywhere between about 1-100 nucleotides (e.g., 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 40, 50, 60, 70, 80, 90, 100). A target nucleic acid sequence can be or can be about 20 bases immediately 5' of the first nucleotide of the PAM. A gRNA can target a nucleic acid sequence. A target nucleic acid can be at least or at least about 1-10, 1-20, 1-30, 1-40, 1-50, 1-60, 1-70, 1-80, 1-90, 30 or 1-100 nucleotides.

Methods for selecting, designing, and validating guide polynucleotides, *e.g.*, gRNAs and targeting sequences are described herein and known to those skilled in the art. For example, to minimize the impact of potential substrate promiscuity of a

deaminase domain in the nucleobase editor system (*e.g.*, an AID domain), the number of residues that could unintentionally be targeted for deamination (*e.g.*, off-target C residues that could potentially reside on single strand DNA within the target nucleic acid locus) may be minimized. In addition, software tools can be used to optimize the gRNAs

5 corresponding to a target nucleic acid sequence, *e.g.*, to minimize total off-target activity across the genome. For example, for each possible targeting domain choice using *S. pyogenes* Cas9, all off-target sequences (preceding selected PAMs, *e.g.*, NAG or NGG) may be identified across the genome that contain up to certain number (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of mismatched base-pairs. First regions of gRNAs complementary to a

10 target site can be identified, and all first regions (*e.g.*, crRNAs) can be ranked according to its total predicted off-target score; the top-ranked targeting domains represent those that are likely to have the greatest on-target and the least off-target activity. Candidate targeting gRNAs can be functionally evaluated by using methods known in the art and/or as set forth herein.

15 As a non-limiting example, target DNA hybridizing sequences in crRNAs of a gRNA for use with Cas9s may be identified using a DNA sequence searching algorithm. gRNA design is carried out using custom gRNA design software based on the public tool Cas-OFFinder as described in Bae S., Park J., & Kim J.-S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided

20 endonucleases. *Bioinformatics* 30, 1473-1475 (2014). This software scores guides after calculating their genome-wide off-target propensity. Typically matches ranging from perfect matches to 7 mismatches are considered for guides ranging in length from 17 to 24. Once the off-target sites are computationally-determined, an aggregate score is calculated for each guide and summarized in a tabular output using a web-interface. In

25 addition to identifying potential target sites adjacent to PAM sequences, the software also identifies all PAM adjacent sequences that differ by 1, 2, 3 or more than 3 nucleotides from the selected target sites. Genomic DNA sequences for a target nucleic acid sequence, *e.g.*, a target gene may be obtained and repeat elements may be screened using publicly available tools, for example, the RepeatMasker program. RepeatMasker searches

30 input DNA sequences for repeated elements and regions of low complexity. The output is a detailed annotation of the repeats present in a given query sequence.

Following identification, first regions of gRNAs, *e.g.*, crRNAs, are ranked into tiers based on their distance to the target site, their orthogonality and presence of 5'

nucleotides for close matches with relevant PAM sequences (for example, a 5' G based on identification of close matches in the human genome containing a relevant PAM *e.g.*, NGG PAM for *S. pyogenes*, NNGRRRT or NNGRRV PAM for *S. aureus*). As used herein, orthogonality refers to the number of sequences in the human genome that contain a minimum number of mismatches to the target sequence. A “high level of orthogonality” or “good orthogonality” may, for example, refer to 20-mer targeting domains that have no identical sequences in the human genome besides the intended target, nor any sequences that contain one or two mismatches in the target sequence. Targeting domains with good orthogonality may be selected to minimize off-target DNA cleavage.

10 A gRNA can then be introduced into a cell or embryo as an RNA molecule or a non-RNA nucleic acid molecule, *e.g.*, DNA molecule. In one embodiment, a DNA encoding a gRNA is operably linked to promoter control sequence for expression of the gRNA in a cell or embryo of interest. A RNA coding sequence can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). Plasmid vectors that can be used to express gRNA include, but are not limited to, px330 vectors and px333 vectors. In some cases, a plasmid vector (*e.g.*, px333 vector) can comprise at least two gRNA-encoding DNA sequences. Further, a vector can comprise additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, etc.), selectable marker sequences (*e.g.*, GFP or antibiotic resistance genes such as puromycin), origins of replication, and the like. A DNA molecule encoding a gRNA can also be linear. A DNA molecule encoding a gRNA or a guide polynucleotide can also be circular.

 In some embodiments, a reporter system is used for detecting base-editing activity and testing candidate guide polynucleotides. In some embodiments, a reporter system comprises a reporter gene based assay where base editing activity leads to expression of the reporter gene. For example, a reporter system may include a reporter gene comprising a deactivated start codon, *e.g.*, a mutation on the template strand from 3'-TAC-5' to 3'-CAC-5'. Upon successful deamination of the target C, the corresponding mRNA will be transcribed as 5'-AUG-3' instead of 5'-GUG-3', enabling the translation of the reporter gene. Suitable reporter genes will be apparent to those of skill in the art. Non-limiting examples of reporter genes include gene encoding green fluorescence protein (GFP), red fluorescence protein (RFP), luciferase, secreted alkaline phosphatase (SEAP), or any other gene whose expression are detectable and apparent to those skilled in the art. The

reporter system can be used to test many different gRNAs, *e.g.*, in order to determine which residue(s) with respect to the target DNA sequence the respective deaminase will target. sgRNAs that target non-template strand can also be tested in order to assess off-target effects of a specific base editing protein, *e.g.*, a Cas9 deaminase fusion protein. In some embodiments, such gRNAs can be designed such that the mutated start codon will not be base-paired with the gRNA. The guide polynucleotides can comprise standard ribonucleotides, modified ribonucleotides (*e.g.*, pseudouridine), ribonucleotide isomers, and/or ribonucleotide analogs. In some embodiments, the guide polynucleotide can comprise at least one detectable label. The detectable label can be a fluorophore (*e.g.*, FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluors, Halo tags, or suitable fluorescent dye), a detection tag (*e.g.*, biotin, digoxigenin, and the like), quantum dots, or gold particles.

In some embodiments, a base editor system may comprise multiple guide polynucleotides, *e.g.*, gRNAs. For example, the gRNAs may target to one or more target loci (*e.g.*, at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 g RNA, at least 50 gRNA) comprised in a base editor system. The multiple gRNA sequences can be tandemly arranged and are preferably separated by a direct repeat.

A guide polynucleotide can comprise one or more modifications to provide a nucleic acid with a new or enhanced feature. A guide polynucleotide can comprise a nucleic acid affinity tag. A guide polynucleotide can comprise synthetic nucleotide, synthetic nucleotide analog, nucleotide derivatives, and/or modified nucleotides.

In some cases, a gRNA or a guide polynucleotide can comprise modifications. A modification can be made at any location of a gRNA or a guide polynucleotide. More than one modification can be made to a single gRNA or a guide polynucleotide. A gRNA or a guide polynucleotide can undergo quality control after a modification. In some cases, quality control can include PAGE, HPLC, MS, or any combination thereof.

A modification of a gRNA or a guide polynucleotide can be a substitution, insertion, deletion, chemical modification, physical modification, stabilization, purification, or any combination thereof.

A gRNA or a guide polynucleotide can also be modified by 5'adenylate, 5' guanosine-triphosphate cap, 5'N7-Methylguanosine-triphosphate cap, 5'triphosphate cap, 3' phosphate, 3' thiophosphate, 5' phosphate, 5' thiophosphate, Cis-Syn thymidine dimer,

trimers, C12 spacer, C3 spacer, C6 spacer, dSpacer, PC spacer, rSpacer, Spacer 18, Spacer 9, 3'-3' modifications, 5'-5' modifications, abasic, acridine, azobenzene, biotin, biotin BB, biotin TEG, cholesteryl TEG, desthiobiotin TEG, DNP TEG, DNP-X, DOTA, dT-Biotin, dual biotin, PC biotin, psoralen C2, psoralen C6, TINA, 3' DABCYL, black
 5 hole quencher 1, black hole quencher 2, DABCYL SE, dT-DABCYL, IRDye QC-1, QSY-21, QSY-35, QSY-7, QSY-9, carboxyl linker, thiol linkers, 2'-deoxyribonucleoside analog purine, 2'-deoxyribonucleoside analog pyrimidine, ribonucleoside analog, 2'-O-methyl ribonucleoside analog, sugar modified analogs, wobble/universal bases, fluorescent dye label, 2'-fluoro RNA, 2'-O-methyl RNA, methylphosphonate,
 10 phosphodiester DNA, phosphodiester RNA, phosphothioate DNA, phosphorothioate RNA, UNA, pseudouridine-5'-triphosphate, 5'-methylcytidine-5'-triphosphate, or any combination thereof.

In some cases, a modification is permanent. In other cases, a modification is transient. In some cases, multiple modifications are made to a gRNA or a guide
 15 polynucleotide. A gRNA or a guide polynucleotide modification can alter physiochemical properties of a nucleotide, such as their conformation, polarity, hydrophobicity, chemical reactivity, base-pairing interactions, or any combination thereof.

A guide polynucleotide can be transferred into a cell by transfecting the cell with
 20 an isolated gRNA or a plasmid DNA comprising a sequence coding for the guide RNA and a promoter. A gRNA or a guide polynucleotide can also be transferred into a cell in other way, such as using virus-mediated gene delivery. A gRNA or a guide polynucleotide can be isolated. For example, a gRNA can be transfected in the form of an isolated RNA into a cell or organism. A gRNA can be prepared by *in vitro*
 25 transcription using any *in vitro* transcription system known in the art. A gRNA can be transferred to a cell in the form of isolated RNA rather than in the form of plasmid comprising encoding sequence for a gRNA.

A modification can also be a phosphorothioate substitute. In some cases, a natural phosphodiester bond can be susceptible to rapid degradation by cellular nucleases and; a
 30 modification of internucleotide linkage using phosphorothioate (PS) bond substitutes can be more stable towards hydrolysis by cellular degradation. A modification can increase stability in a gRNA or a guide polynucleotide. A modification can also enhance biological activity. In some cases, a phosphorothioate enhanced RNA gRNA can inhibit

5 RNase A, RNase T1, calf serum nucleases, or any combinations thereof. These properties can allow the use of PS-RNA gRNAs to be used in applications where exposure to nucleases is of high probability *in vivo* or *in vitro*. For example, phosphorothioate (PS) bonds can be introduced between the last 3-5 nucleotides at the 5'- or 3'-end of a gRNA which can inhibit exonuclease degradation. In some cases, phosphorothioate bonds can be added throughout an entire gRNA to reduce attack by endonucleases.

In some embodiments, the guide RNA is designed to disrupt a splice site (*i.e.*, a splice acceptor (SA) or a splice donor (SD)). In some embodiments, the guide RNA is designed such that the base editing results in a premature STOP codon.

10

Protospacer Adjacent Motif

The term "protospacer adjacent motif (PAM)" or PAM-like motif refers to a 2-6 base pair DNA sequence immediately following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system. In some embodiments, the PAM can be a 5' PAM (*i.e.*, located upstream of the 5' end of the protospacer). In other 15 embodiments, the PAM can be a 3' PAM (*i.e.*, located downstream of the 5' end of the protospacer). The PAM sequence is essential for target binding, but the exact sequence depends on a type of Cas protein. The PAM sequence can be any PAM sequence known in the art. Suitable PAM sequences include, but are not limited to, NGG, NGA, NGC, NGN, NGT, NGTT, NGCG, NGAG, NGAN, NGNG, NGCN, NGCG, NGTN, 20 NNGRRT, NNNRRT, NNGRR(N), TTTV, TYCV, TYCV, TATV, NNNNGATT, NNAGAAW, or NAAAAC. Y is a pyrimidine; N is any nucleotide base; W is A or T.

A base editor provided herein can comprise a CRISPR protein-derived domain that is capable of binding a nucleotide sequence that contains a canonical or non- 25 canonical protospacer adjacent motif (PAM) sequence. A PAM site is a nucleotide sequence in proximity to a target polynucleotide sequence. Some aspects of the disclosure provide for base editors comprising all or a portion of CRISPR proteins that have different PAM specificities.

For example, typically Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), 30 require a canonical NGG PAM sequence to bind a particular nucleic acid region, where the "N" in "NGG" is adenine (A), thymine (T), guanine (G), or cytosine (C), and the G is guanine. A PAM can be CRISPR protein-specific and can be different between different base editors comprising different CRISPR protein-derived domains. A PAM can be 5' or

3' of a target sequence. A PAM can be upstream or downstream of a target sequence. A PAM can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides in length. Often, a PAM is between 2-6 nucleotides in length.

In some embodiments, the PAM is an “NRN” PAM where the “N” in “NRN” is adenine (A), thymine (T), guanine (G), or cytosine (C), and the R is adenine (A) or guanine (G); or the PAM is an “NYN” PAM, wherein the “N” in NYN is adenine (A), thymine (T), guanine (G), or cytosine (C), and the Y is cytosine (C) or thymine (T), for example, as described in R.T. Walton *et al.*, 2020, *Science*, 10.1126/science.aba8853 (2020), the entire contents of which are incorporated herein by reference.

Several PAM variants are described in **Table 7** below.

Table 7. Cas9 proteins and corresponding PAM sequences

Variant	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
spCas9-MQKFRAER	NGC
Cpf1	5' (TTTV)
SpyMac	5'-NAA-3'

In some embodiments, the PAM is NGC. In some embodiments, the NGC PAM is recognized by a Cas9 variant. In some embodiments, the NGC PAM variant includes one or more amino acid substitutions selected from D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R (collectively termed “MQKFRAER”).

In some embodiments, the PAM is NGT. In some embodiments, the NGT PAM is recognized by a Cas9 variant. In some embodiments, the NGT PAM variant is generated through targeted mutations at one or more residues 1335, 1337, 1135, 1136, 1218, and/or 1219. In some embodiments, the NGT PAM variant is created through targeted mutations at one or more residues 1219, 1335, 1337, 1218. In some embodiments, the NGT PAM variant is created through targeted mutations at one or more residues 1135, 1136, 1218, 1219, and 1335. In some embodiments, the NGT PAM variant is selected from the set of targeted mutations provided in **Tables 8A** and **8B** below.

10 **Table 8A: NGT PAM Variant Mutations at residues 1219, 1335, 1337, 1218**

Variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R
9	L	L	T	
10	L	L	R	
11	L	L	Q	
12	L	L	L	
13	F	I	T	
14	F	I	R	
15	F	I	Q	
16	F	I	L	
17	F	G	C	
18	H	L	N	
19	F	G	C	A
20	H	L	N	V
21	L	A	W	
22	L	A	F	
23	L	A	Y	
24	I	A	W	
25	I	A	F	
26	I	A	Y	

Table 8B: NGT PAM Variant Mutations at residues 1135, 1136, 1218, 1219, and 1335

Variant	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				
28	V				
29	I				
30		A			
31		W			
32		H			
33		K			
34			K		
35			R		
36			Q		
37			T		
38			N		
39				I	
40				A	
41				N	
42				Q	
43				G	
44				L	
45				S	
46				T	
47					L
48					I
49					V
50					N
51					S
52					T
53					F
54					Y
55	N1286Q	I1331F			

In some embodiments, the NGT PAM variant is selected from variant 5, 7, 28, 31, or 36 in **Table 8A** and **Table 8B**. In some embodiments, the variants have improved NGT PAM recognition.

In some embodiments, the NGT PAM variants have mutations at residues 1219, 1335, 1337, and/or 1218. In some embodiments, the NGT PAM variant is selected with mutations for improved recognition from the variants provided in **Table 9** below.

Table 9: NGT PAM Variant Mutations at residues 1219, 1335, 1337, and 1218

Variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R

In some embodiments, the NGT PAM is selected from the variants provided in **Table 10** below.

5

Table 10. NGT PAM variants

	NGTN variant	D1135	S1136	G1218	E1219	A1322R	R1335	T1337
Variant 1	LRKIQK	L	R	K	I	-	Q	K
Variant 2	LRSVQK	L	R	S	V	-	Q	K
Variant 3	LRSVQL	L	R	S	V	-	Q	L
Variant 4	LRKIRQK	L	R	K	I	R	Q	K
Variant 5	LRSVRQK	L	R	S	V	R	Q	K
Variant 6	LRSVRQL	L	R	S	V	R	Q	L

In some embodiments the NGTN variant is variant 1. In some embodiments, the NGTN variant is variant 2. In some embodiments, the NGTN variant is variant 3. In some embodiments, the NGTN variant is variant 4. In some embodiments, the NGTN variant is variant 5. In some embodiments, the NGTN variant is variant 6.

In some embodiments, the Cas9 domain is a Cas9 domain from *Streptococcus pyogenes* (SpCas9). In some embodiments, the SpCas9 domain is a nuclease active SpCas9, a nuclease inactive SpCas9 (SpCas9d), or a SpCas9 nickase (SpCas9n). In some embodiments, the SpCas9 comprises a D9X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid except for D. In some embodiments, the SpCas9 comprises a D9A mutation, or a corresponding

mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having an
5 NGG, a NGA, or a NGCG PAM sequence.

In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135E, R1335Q, and T1337R mutation, or
10 a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135E, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid
15 sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135V, a R1335Q, and a T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135V, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein.
20 In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a G1218X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135V, a G1218R, a R1335Q, and a T1337R mutation, or a corresponding mutation in any of the amino acid
25 sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135V, a G1218R, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein.

In some examples, a PAM recognized by a CRISPR protein-derived domain of a base editor disclosed herein can be provided to a cell on a separate oligonucleotide to an
30 insert (*e.g.*, an AAV insert) encoding the base editor. In such embodiments, providing PAM on a separate oligonucleotide can allow cleavage of a target sequence that otherwise would not be able to be cleaved, because no adjacent PAM is present on the same polynucleotide as the target sequence.

In an embodiment, *S. pyogenes* Cas9 (SpCas9) can be used as a CRISPR endonuclease for genome engineering. However, others can be used. In some embodiments, a different endonuclease can be used to target certain genomic targets. In some embodiments, synthetic SpCas9-derived variants with non-NGG PAM sequences can be used. Additionally, other Cas9 orthologues from various species have been identified and these “non-SpCas9s” can bind a variety of PAM sequences that can also be useful for the present disclosure. For example, the relatively large size of SpCas9 (approximately 4kb coding sequence) can lead to plasmids carrying the SpCas9 cDNA that cannot be efficiently expressed in a cell. Conversely, the coding sequence for *Staphylococcus aureus* Cas9 (SaCas9) is approximately 1 kilobase shorter than SpCas9, possibly allowing it to be efficiently expressed in a cell. Similar to SpCas9, the SaCas9 endonuclease is capable of modifying target genes in mammalian cells *in vitro* and in mice *in vivo*. In some embodiments, a Cas protein can target a different PAM sequence. In some embodiments, a target gene can be adjacent to a Cas9 PAM, 5'-NGG, for example. In other embodiments, other Cas9 orthologs can have different PAM requirements. For example, other PAMs such as those of *S. thermophilus* (5'-NNAGAA for CRISPR1 and 5'-NGGNG for CRISPR3) and *Neisseria meningitidis* (5'-NNNNGATT) can also be found adjacent to a target gene.

In some embodiments, for a *S. pyogenes* system, a target gene sequence can precede (*i.e.*, be 5' to) a 5'-NGG PAM, and a 20-nt guide RNA sequence can base pair with an opposite strand to mediate a Cas9 cleavage adjacent to a PAM. In some embodiments, an adjacent cut can be or can be about 3 base pairs upstream of a PAM. In some embodiments, an adjacent cut can be or can be about 10 base pairs upstream of a PAM. In some embodiments, an adjacent cut can be or can be about 0-20 base pairs upstream of a PAM. For example, an adjacent cut can be next to, 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, 26, 27, 28, 29, or 30 base pairs upstream of a PAM. An adjacent cut can also be downstream of a PAM by 1 to 30 base pairs. The sequences of exemplary SpCas9 proteins capable of binding a PAM sequence follow:

In some embodiments, engineered SpCas9 variants are capable of recognizing protospacer adjacent motif (PAM) sequences flanked by a 3' H (non-G PAM) (*see Tables 2A-2B and 3*). In some embodiments, the SpCas9 variants recognize NRNH PAMs (where R is A or G and H is A, C or T). In some embodiments, the non-G PAM is

NRRH, NRTH, or NRCH (*see e.g.*, Miller, S.M., *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs, Nat. Biotechnol. (2020), the contents of which is incorporated herein by reference in its entirety).

In some embodiments, the Cas9 domain is a recombinant Cas9 domain. In some
5 embodiments, the recombinant Cas9 domain is a SpyMacCas9 domain. In some
embodiments, the SpyMacCas9 domain is a nuclease active SpyMacCas9, a nuclease
inactive SpyMacCas9 (SpyMacCas9d), or a SpyMacCas9 nickase (SpyMacCas9n). In
some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain
can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments,
10 the SpyMacCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a
nucleic acid sequence having a NAA PAM sequence.

The sequence of an exemplary Cas9 A homolog of Spy Cas9 in *Streptococcus*
macacae with native 5'-NAAN-3' PAM specificity is known in the art and described, for
example, by Jakimo *et al.*,
15 (www.biorxiv.org/content/biorxiv/early/2018/09/27/429654.full.pdf), and is provided as
SEQ ID NO: 162.

In some embodiments, a variant Cas9 protein harbors, H840A, P475A, W476A,
N477A, D1125A, W1126A, and D1218A mutations such that the polypeptide has a
reduced ability to cleave a target DNA or RNA. Such a Cas9 protein has a reduced
20 ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability
to bind a target DNA (*e.g.*, a single stranded target DNA). As another non-limiting
example, in some embodiments, the variant Cas9 protein harbors D10A, H840A, P475A,
W476A, N477A, D1125A, W1126A, and D1218A mutations such that the polypeptide
has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to
25 cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a
target DNA (*e.g.*, a single stranded target DNA). In some embodiments, when a variant
Cas9 protein harbors W476A and W1126A mutations or when the variant Cas9 protein
harbors P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations, the variant
Cas9 protein does not bind efficiently to a PAM sequence. Thus, in some such cases,
30 when such a variant Cas9 protein is used in a method of binding, the method does not
require a PAM sequence. In other words, in some embodiments, when such a variant
Cas9 protein is used in a method of binding, the method can include a guide RNA, but the
method can be performed in the absence of a PAM sequence (and the specificity of

binding is therefore provided by the targeting segment of the guide RNA). Other residues can be mutated to achieve the above effects (*i.e.*, inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 can be altered (*i.e.*, substituted). Also, mutations
 5 other than alanine substitutions are suitable.

In some embodiments, a CRISPR protein-derived domain of a base editor can comprise all or a portion of a Cas9 protein with a canonical PAM sequence (NGG). In other embodiments, a Cas9-derived domain of a base editor can employ a non-canonical PAM sequence. Such sequences have been described in the art and would be apparent to
 10 the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.*, “Engineered CRISPR-Cas9 nucleases with altered PAM specificities” *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, “Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition” *Nature Biotechnology* 33, 1293-1298 (2015); R.T. Walton *et al.*
 15 “Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants” *Science* 10.1126/science.aba8853 (2020); Hu *et al.* “Evolved Cas9 variants with broad PAM compatibility and high DNA specificity,” *Nature*, 2018 Apr. 5, 556(7699), 57-63; Miller *et al.*, “Continuous evolution of SpCas9 variants compatible with non-G PAMs” *Nat. Biotechnol.*, 2020 Apr;38(4):471-481; the entire contents of each are hereby
 20 incorporated by reference.

Fusion Proteins Comprising a NapDNAbp and a Cytidine Deaminase and/or Adenosine Deaminase

Some aspects of the disclosure provide fusion proteins comprising a Cas9 domain
 25 or other nucleic acid programmable DNA binding protein (*e.g.*, Cas12) and one or more cytidine deaminase or adenosine deaminase domains. It should be appreciated that the Cas9 domain may be any of the Cas9 domains or Cas9 proteins (*e.g.*, dCas9 or nCas9) provided herein. In some embodiments, any of the Cas9 domains or Cas9 proteins (*e.g.*, dCas9 or nCas9) provided herein may be fused with any of the cytidine deaminases
 30 and/or adenosine deaminases provided herein. The domains of the base editors disclosed herein can be arranged in any order.

In some embodiments, the fusion protein comprises the following domains A-C, A-D, or A-E:

NH₂-[A-B-C]-COOH;

NH₂-[A-B-C-D]-COOH; or

NH₂-[A-B-C-D-E]-COOH;

wherein A and C or A, C, and E, each comprises one or more of the following:

- 5 an adenosine deaminase domain or an active fragment thereof,
 a cytidine deaminase domain or an active fragment thereof, and
 wherein B or B and D, each comprises one or more domains having nucleic acid
 sequence specific binding activity.

In some embodiments, the fusion protein comprises the following structure:

10 NH₂-[A_n-B_o-C_n]-COOH;

NH₂-[A_n-B_o-C_n-D_o]-COOH; or

NH₂-[A_n-B_o-C_p-D_o-E_q]-COOH;

wherein A and C or A, C, and E, each comprises one or more of the following:

- an adenosine deaminase domain or an active fragment thereof,
 15 a cytidine deaminase domain or an active fragment thereof, and
 wherein n is an integer: 1, 2, 3, 4, or 5, wherein p is an integer: 0, 1, 2, 3, 4, or 5; wherein
 q is an integer 0, 1, 2, 3, 4, or 5; and wherein B or B and D each comprises a domain
 having nucleic acid sequence specific binding activity; and wherein o is an integer: 1, 2,
 3, 4, or 5.

- 20 For example, and without limitation, in some embodiments, the fusion protein
 comprises the structure:

NH₂-[adenosine deaminase]-[Cas9 domain]-COOH;

NH₂-[Cas9 domain]-[adenosine deaminase]-COOH;

NH₂-[cytidine deaminase]-[Cas9 domain]-COOH;

25 NH₂-[Cas9 domain]-[cytidine deaminase]-COOH;

NH₂-[cytidine deaminase]-[Cas9 domain]-[adenosine deaminase]-COOH;

NH₂-[adenosine deaminase]-[Cas9 domain]-[cytidine deaminase]-COOH;

NH₂-[adenosine deaminase]-[cytidine deaminase]-[Cas9 domain]-COOH;

NH₂-[cytidine deaminase]-[adenosine deaminase]-[Cas9 domain]-COOH;

30 NH₂-[Cas9 domain]-[adenosine deaminase]-[cytidine deaminase]-COOH; or

NH₂-[Cas9 domain]-[cytidine deaminase]-[adenosine deaminase]-COOH.

In some embodiments, any of the Cas12 domains or Cas12 proteins provided
 herein may be fused with any of the cytidine or adenosine deaminases provided herein.

For example, and without limitation, in some embodiments, the fusion protein comprises the structure:

- NH₂-[adenosine deaminase]-[Cas12 domain]-COOH;
 NH₂-[Cas12 domain]-[adenosine deaminase]-COOH;
 5 NH₂-[cytidine deaminase]-[Cas12 domain]-COOH;
 NH₂-[Cas12 domain]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[Cas12 domain]-[adenosine deaminase]-COOH;
 NH₂-[adenosine deaminase]-[Cas12 domain]-[cytidine deaminase]-COOH;
 NH₂-[adenosine deaminase]-[cytidine deaminase]-[Cas12 domain]-COOH;
 10 NH₂-[cytidine deaminase]-[adenosine deaminase]-[Cas12 domain]-COOH;
 NH₂-[Cas12 domain]-[adenosine deaminase]-[cytidine deaminase]-COOH; or
 NH₂-[Cas12 domain]-[cytidine deaminase]-[adenosine deaminase]-COOH.

In some embodiments, the adenosine deaminase is a TadA*8. Exemplary fusion protein structures include the following:

- 15 NH₂-[TadA*8]-[Cas9 domain]-COOH;
 NH₂-[Cas9 domain]-[TadA*8]-COOH;
 NH₂-[TadA*8]-[Cas12 domain]-COOH; or
 NH₂-[Cas12 domain]-[TadA*8]-COOH.

- In some embodiments, the adenosine deaminase of the fusion protein comprises a
 20 TadA*8 and a cytidine deaminase and/or an adenosine deaminase. In some
 embodiments, the TadA*8 is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5,
 TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12,
 TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19,
 TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24.

- 25 Exemplary fusion protein structures include the following:

NH₂-[TadA*8]-[Cas9/Cas12]-[adenosine deaminase]-COOH;
 NH₂-[adenosine deaminase]-[Cas9/Cas12]-[TadA*8]-COOH;
 NH₂-[TadA*8]-[Cas9/Cas12]-[cytidine deaminase]-COOH; or
 NH₂-[cytidine deaminase]-[Cas9/Cas12]-[TadA*8]-COOH.

- 30 In some embodiments, the adenosine deaminase of the fusion protein comprises a
 TadA*9 and a cytidine deaminase and/or an adenosine deaminase. Exemplary fusion
 protein structures include the following:

NH₂-[TadA*9]-[Cas9/Cas12]-[adenosine deaminase]-COOH;

NH₂-[adenosine deaminase]-[Cas9/Cas12]-[TadA*9]-COOH;
NH₂-[TadA*9]-[Cas9/Cas12]-[cytidine deaminase]-COOH; or
NH₂-[cytidine deaminase]-[Cas9/Cas12]-[TadA*9]-COOH.

5 In some embodiments, the fusion protein can comprise a deaminase flanked by an N- terminal fragment and a C-terminal fragment of a Cas9 or Cas12 polypeptide. In some embodiments, the fusion protein comprises a cytidine deaminase flanked by an N-terminal fragment and a C-terminal fragment of a Cas9 or Cas12 polypeptide. In some embodiments, the fusion protein comprises an adenosine deaminase flanked by an N-terminal fragment and a C-terminal fragment of a Cas9 or Cas 12 polypeptide.

10 In some embodiments, the fusion proteins comprising a cytidine deaminase or adenosine deaminase and a napDNAbp (*e.g.*, Cas9 or Cas12 domain) do not include a linker sequence. In some embodiments, a linker is present between the cytidine or adenosine deaminase and the napDNAbp. In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker. In some
15 embodiments, cytidine or adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein. For example, in some embodiments the cytidine or adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein.

It should be appreciated that the fusion proteins of the present disclosure may comprise one or more additional features. For example, in some embodiments, the fusion
20 protein may comprise inhibitors, cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin
25 (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags, biotin ligase tags, FAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. In some embodiments, the fusion
30 protein comprises one or more His tags.

Exemplary, yet nonlimiting, fusion proteins are described in International PCT Application Nos. PCT/2017/044935, PCT/US2019/044935, and PCT/US2020/016288, each of which is incorporated herein by reference for its entirety.

Fusion Proteins Comprising a Nuclear Localization Sequence (NLS)

In some embodiments, the fusion proteins provided herein further comprise one or more (*e.g.*, 2, 3, 4, 5) nuclear targeting sequences, for example a nuclear localization sequence (NLS). In one embodiment, a bipartite NLS is used. In some embodiments, a NLS comprises an amino acid sequence that facilitates the importation of a protein, that comprises an NLS, into the cell nucleus (*e.g.*, by nuclear transport). In some embodiments, the NLS is fused to the N-terminus or the C-terminus of the fusion protein. In some embodiments, the NLS is fused to the C-terminus or N-terminus of an nCas9 domain or a dCas9 domain. In some embodiments, the NLS is fused to the N-terminus or C-terminus of the Cas12 domain. In some embodiments, the NLS is fused to the N-terminus or C-terminus of the cytidine or adenosine deaminase. In some embodiments, the NLS is fused to the fusion protein via one or more linkers. In some embodiments, the NLS is fused to the fusion protein without a linker. In some embodiments, the NLS comprises an amino acid sequence of any one of the NLS sequences provided or referenced herein. Additional nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.*, PCT/EP2000/011690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, an NLS comprises the amino acid sequence

PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 253),
 KRTADGSEFESPKKKRKV (SEQ ID NO: 83), KRPAATKKAGQAKKKK (SEQ ID NO: 84), KKTELQTTNAENKTKKL (SEQ ID NO: 85), KRGINDRNFWRGENGRKTR (SEQ ID NO: 86), RKSGKIAAIVVKRPRKPKKKRKV (SEQ ID NO: 254), or
 MDSLLMNRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 89).

In some embodiments, the fusion proteins comprising a cytidine or adenosine deaminase, a Cas9 domain, and an NLS do not comprise a linker sequence. In some embodiments, linker sequences between one or more of the domains or proteins (*e.g.*, cytidine or adenosine deaminase, Cas9 domain or NLS) are present. In some embodiments, a linker is present between the cytidine deaminase and adenosine deaminase domains and the napDNAbp. In some embodiments, the “-” used in the general architecture below indicates the presence of an optional linker. In some embodiments, the cytidine deaminase and adenosine deaminase and the napDNAbp are

fused via any of the linkers provided herein. For example, in some embodiments the cytidine deaminase and adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein.

- In some embodiments, the general architecture of exemplary napDNAbp (*e.g.*,
 5 Cas9 or Cas12) fusion proteins with a cytidine or adenosine deaminase and a napDNAbp (*e.g.*, Cas9 or Cas12) domain comprises any one of the following structures, where NLS is a nuclear localization sequence (*e.g.*, any NLS provided herein), NH₂ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein:
- NH₂-NLS-[cytidine deaminase]-[napDNAbp domain]-COOH;
 10 NH₂-NLS [napDNAbp domain]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[napDNAbp domain]-NLS-COOH;
 NH₂-[napDNAbp domain]-[cytidine deaminase]-NLS-COOH;
 NH₂-NLS-[adenosine deaminase]-[napDNAbp domain]-COOH;
 NH₂-NLS [napDNAbp domain]-[adenosine deaminase]-COOH;
 15 NH₂-[adenosine deaminase]-[napDNAbp domain]-NLS-COOH;
 NH₂-[napDNAbp domain]-[adenosine deaminase]-NLS-COOH;
 NH₂-NLS-[cytidine deaminase]-[napDNAbp domain]-[adenosine deaminase]-COOH;
 NH₂-NLS-[adenosine deaminase]-[napDNAbp domain]-[cytidine deaminase]-COOH;
 NH₂-NLS-[adenosine deaminase] [cytidine deaminase]-[napDNAbp domain]-COOH;
 20 NH₂-NLS-[cytidine deaminase]-[adenosine deaminase]-[napDNAbp domain]-COOH;
 NH₂-NLS-[napDNAbp domain]-[adenosine deaminase]-[cytidine deaminase]-COOH;
 NH₂-NLS-[napDNAbp domain]-[cytidine deaminase]-[adenosine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[napDNAbp domain]-[adenosine deaminase]-NLS-COOH;
 NH₂-[adenosine deaminase]-[napDNAbp domain]-[cytidine deaminase]-NLS-COOH;
 25 NH₂-[adenosine deaminase] [cytidine deaminase]-[napDNAbp domain]-NLS-COOH;
 NH₂-[cytidine deaminase]-[adenosine deaminase]-[napDNAbp domain]-NLS-COOH;
 NH₂-[napDNAbp domain]-[adenosine deaminase]-[cytidine deaminase]-NLS-COOH; or
 NH₂-[napDNAbp domain]-[cytidine deaminase]-[adenosine deaminase]-NLS-COOH. In some embodiments, the NLS is present in a linker or the NLS is flanked by
 30 linkers, for example described herein. A bipartite NLS comprises two basic amino acid clusters, which are separated by a relatively short spacer sequence (hence bipartite - 2 parts, while monopartite NLSs are not). The NLS of nucleoplasmin, KR[PAATKKAGQA]KKKK (SEQ ID NO: 84), is the prototype of the ubiquitous

bipartite signal: two clusters of basic amino acids, separated by a spacer of about 10 amino acids. The sequence of an exemplary bipartite NLS follows:

PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 253)

A vector that encodes a CRISPR enzyme comprising one or more nuclear
5 localization sequences (NLSs) can be used. For example, there can be or be about 1, 2, 3,
4, 5, 6, 7, 8, 9, 10 NLSs used. A CRISPR enzyme can comprise the NLSs at or near the
amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 NLSs at or near the
carboxy-terminus, or any combination thereof (*e.g.*, one or more NLS at the amino-
terminus and one or more NLS at the carboxy terminus). When more than one NLS is
10 present, each can be selected independently of others, such that a single NLS can be
present in more than one copy and/or in combination with one or more other NLSs
present in one or more copies.

CRISPR enzymes used in the methods can comprise about 6 NLSs. An NLS is
considered near the N- or C-terminus when the nearest amino acid to the NLS is within
15 about 50 amino acids along a polypeptide chain from the N- or C-terminus, *e.g.*, within 1,
2, 3, 4, 5, 10, 15, 20, 25, 30, 40, or 50 amino acids.

Additional Domains

A base editor described herein can include any domain which helps to facilitate
20 the nucleobase editing, modification or altering of a nucleobase of a polynucleotide. In
some embodiments, a base editor comprises a polynucleotide programmable nucleotide
binding domain (*e.g.*, Cas9), a nucleobase editing domain (*e.g.*, deaminase domain), and
one or more additional domains. In some embodiments, the additional domain can
facilitate enzymatic or catalytic functions of the base editor, binding functions of the base
25 editor, or be inhibitors of cellular machinery (*e.g.*, enzymes) that could interfere with the
desired base editing result. In some embodiments, a base editor can comprise a nuclease,
a nickase, a recombinase, a deaminase, a methyltransferase, a methylase, an acetylase, an
acetyltransferase, a transcriptional activator, or a transcriptional repressor domain.

In some embodiments, a base editor can comprise an uracil glycosylase inhibitor
30 (UGI) domain. In some embodiments, cellular DNA repair response to the presence of U:
G heteroduplex DNA can be responsible for a decrease in nucleobase editing efficiency in
cells. In such embodiments, uracil DNA glycosylase (UDG) can catalyze removal of U
from DNA in cells, which can initiate base excision repair (BER), mostly resulting in

reversion of the U:G pair to a C:G pair. In such embodiments, BER can be inhibited in base editors comprising one or more domains that bind the single strand, block the edited base, inhibit UGI, inhibit BER, protect the edited base, and /or promote repairing of the non-edited strand. Thus, this disclosure contemplates a base editor fusion protein

5 comprising a UGI domain.

In some embodiments, a base editor comprises as a domain all or a portion of a double-strand break (DSB) binding protein. For example, a DSB binding protein can include a Gam protein of bacteriophage Mu that can bind to the ends of DSBs and can protect them from degradation. See Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaa04774 (2017), the entire content of which is hereby incorporated by reference.

Additionally, in some embodiments, a Gam protein can be fused to an N terminus of a base editor. In some embodiments, a Gam protein can be fused to a C terminus of a base editor. The Gam protein of bacteriophage Mu can bind to the ends of double strand breaks (DSBs) and protect them from degradation. In some embodiments, using Gam to bind the free ends of DSB can reduce indel formation during the process of base editing. In some embodiments, 174-residue Gam protein is fused to the N terminus of the base editors. See Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaa04774 (2017). In some embodiments, a mutation or mutations can change the length of a base editor domain relative to a wild type domain. For example, a deletion of at least one amino acid in at least one domain can reduce the length of the base editor. In another case, a mutation or mutations do not change the length of a domain relative to a wild type domain. For example, substitutions in any domain does not change the length of the base editor.

Non-limiting examples of such base editors, where the length of all the domains is the same as the wild type domains, can include:

NH₂-[nucleobase editing domain]-Linker1-[APOBEC1]-Linker2-[nucleobase editing domain]-COOH;

NH₂-[nucleobase editing domain]-Linker1-[APOBEC1]-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-[APOBEC1]-Linker2-[nucleobase editing domain]-
 COOH;
 NH2-[nucleobase editing domain]-[APOBEC1]-[nucleobase editing domain]-COOH;
 NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-Linker2-[nucleobase editing
 5 domain]-[UGI]-COOH;
 NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-[nucleobase editing domain]-
 [UGI]-COOH;
 NH2-[nucleobase editing domain]-[APOBEC1]-Linker2-[nucleobase editing domain]-
 [UGI]-COOH;
 10 NH2-[nucleobase editing domain]-[APOBEC1]-[nucleobase editing domain]-[UGI]-
 COOH;
 NH2-[UGI]-[nucleobase editing domain]-Linker1-[APOBEC1]-Linker2-[nucleobase
 editing domain]-COOH;
 NH2-[UGI]-[nucleobase editing domain]-Linker1-[APOBEC1]-[nucleobase editing
 15 domain]-COOH;
 NH2-[UGI]-[nucleobase editing domain]-[APOBEC1]-Linker2-[nucleobase editing
 domain]-COOH; or
 NH2-[UGI]-[nucleobase editing domain]-[APOBEC1]-[nucleobase editing domain]-
 COOH.
 20

BASE EDITOR SYSTEM

Provided herein are systems, compositions, and methods for editing a nucleobase
 using a base editor system. In some embodiments, the base editor system comprises (1) a
 base editor (BE) comprising a polynucleotide programmable nucleotide binding domain
 25 and a nucleobase editing domain (e.g., a deaminase domain) for editing the nucleobase;
 and (2) a guide polynucleotide (e.g., guide RNA) in conjunction with the polynucleotide
 programmable nucleotide binding domain. In some embodiments, the base editor system
 is a cytidine base editor (CBE) or an adenosine base editor (ABE). In some
 embodiments, the polynucleotide programmable nucleotide binding domain is a
 30 polynucleotide programmable DNA or RNA binding domain. In some embodiments, the
 nucleobase editing domain is a deaminase domain. In some embodiments, a deaminase
 domain can be a cytidine deaminase or a cytosine deaminase. In some embodiments, a
 deaminase domain can be an adenine deaminase or an adenosine deaminase. In some

embodiments, the adenosine base editor can deaminate adenine in DNA. In some embodiments, the base editor is capable of deaminating a cytidine in DNA.

In some embodiments, a base editing system as provided herein provides a new approach to genome editing that uses a fusion protein containing a catalytically defective
5 Streptococcus pyogenes Cas9, a deaminase (e.g., cytidine or adenosine deaminase), and an inhibitor of base excision repair to induce programmable, single nucleotide (C→T or A→G) changes in DNA without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions.

10 Details of nucleobase editing proteins are described in International PCT Application Nos. PCT/2017/045381 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., *et al.*, “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage” *Nature* 533, 420-424 (2016); Gaudelli, N.M., *et al.*,
15 “Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage” *Nature* 551, 464-471 (2017); and Komor, A.C., *et al.*, “Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity” *Science Advances* 3:eaa04774 (2017), the entire contents of which are hereby incorporated by reference.

20 Use of the base editor system provided herein comprises the steps of: (a) contacting a target nucleotide sequence of a polynucleotide (e.g., double- or single stranded DNA or RNA) of a subject with a base editor system comprising a nucleobase editor (e.g., an adenosine base editor or a cytidine base editor) and a guide polynucleic acid (e.g., gRNA), wherein the target nucleotide sequence comprises a targeted
25 nucleobase pair; (b) inducing strand separation of said target region; (c) converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase; and (d) cutting no more than one strand of said target region, where a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase. It should be appreciated that in some
30 embodiments, step (b) is omitted. In some embodiments, said targeted nucleobase pair is a plurality of nucleobase pairs in one or more genes. In some embodiments, the base editor system provided herein is capable of multiplex editing of a plurality of nucleobase pairs in one or more genes. In some embodiments, the plurality of nucleobase pairs is

located in the same gene. In some embodiments, the plurality of nucleobase pairs is located in one or more genes, wherein at least one gene is located in a different locus.

In some embodiments, the cut single strand (nicked strand) is hybridized to the guide nucleic acid. In some embodiments, the cut single strand is opposite to the strand comprising the first nucleobase. In some embodiments, the base editor comprises a Cas9 domain. In some embodiments, the first base is adenine, and the second base is not a G, C, A, or T. In some embodiments, the second base is inosine.

In some embodiments, a single guide polynucleotide may be utilized to target a deaminase to a target nucleic acid sequence. In some embodiments, a single pair of guide polynucleotides may be utilized to target different deaminases to a target nucleic acid sequence.

The nucleobase components and the polynucleotide programmable nucleotide binding component of a base editor system may be associated with each other covalently or non-covalently. For example, in some embodiments, the deaminase domain can be targeted to a target nucleotide sequence by a polynucleotide programmable nucleotide binding domain. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to a deaminase domain. In some embodiments, a polynucleotide programmable nucleotide binding domain can target a deaminase domain to a target nucleotide sequence by non-covalently interacting with or associating with the deaminase domain. For example, in some embodiments, the nucleobase editing component, e.g., the deaminase component can comprise an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with an additional heterologous portion or domain that is part of a polynucleotide programmable nucleotide binding domain. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polypeptide. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a

K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or an RNA recognition motif.

5 A base editor system may further comprise a guide polynucleotide component. It should be appreciated that components of the base editor system may be associated with each other via covalent bonds, noncovalent interactions, or any combination of associations and interactions thereof. In some embodiments, a deaminase domain can be targeted to a target nucleotide sequence by a guide polynucleotide. For example, in some
10 embodiments, the nucleobase editing component of the base editor system, e.g., the deaminase component, can comprise an additional heterologous portion or domain (e.g., polynucleotide binding domain such as an RNA or DNA binding protein) that is capable of interacting with, associating with, or capable of forming a complex with a portion or segment (e.g., a polynucleotide motif) of a guide polynucleotide. In some embodiments,
15 the additional heterologous portion or domain (e.g., polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or linked to the deaminase domain. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polypeptide. In some
20 embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional
25 heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or an RNA recognition motif.

30 In some embodiments, a base editor system can further comprise an inhibitor of base excision repair (BER) component. It should be appreciated that components of the base editor system may be associated with each other via covalent bonds, noncovalent interactions, or any combination of associations and interactions thereof. The inhibitor of

BER component may comprise a base excision repair inhibitor. In some embodiments, the inhibitor of base excision repair can be a uracil DNA glycosylase inhibitor (UGI). In some embodiments, the inhibitor of base excision repair can be an inosine base excision repair inhibitor. In some embodiments, the inhibitor of base excision repair can be
5 targeted to the target nucleotide sequence by the polynucleotide programmable nucleotide binding domain. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to an inhibitor of base excision repair. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to a deaminase domain and an inhibitor of base excision repair. In some
10 embodiments, a polynucleotide programmable nucleotide binding domain can target an inhibitor of base excision repair to a target nucleotide sequence by non-covalently interacting with or associating with the inhibitor of base excision repair. For example, in some embodiments, the inhibitor of base excision repair component can comprise an additional heterologous portion or domain that is capable of interacting with, associating
15 with, or capable of forming a complex with an additional heterologous portion or domain that is part of a polynucleotide programmable nucleotide binding domain. In some embodiments, the inhibitor of base excision repair can be targeted to the target nucleotide sequence by the guide polynucleotide. For example, in some embodiments, the inhibitor of base excision repair can comprise an additional heterologous portion or domain (*e.g.*,
20 polynucleotide binding domain such as an RNA or DNA binding protein) that is capable of interacting with, associating with, or capable of forming a complex with a portion or segment (*e.g.*, a polynucleotide motif) of a guide polynucleotide. In some embodiments, the additional heterologous portion or domain of the guide polynucleotide (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or
25 linked to the inhibitor of base excision repair. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some
30 embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu

Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or an RNA recognition motif.

In some embodiments, the base editor inhibits base excision repair (BER) of the edited strand. In some embodiments, the base editor protects or binds the non-edited
5 strand. In some embodiments, the base editor comprises UGI activity. In some
embodiments, the base editor comprises a catalytically inactive inosine-specific nuclease. In some embodiments, the base editor comprises nickase activity. In some embodiments, the intended edit of base pair is upstream of a PAM site. In some embodiments, the
10 intended edit of base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or
20 nucleotides upstream of the PAM site. In some embodiments, the intended edit of
base-pair is downstream of a PAM site. In some embodiments, the intended edited base
pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides
downstream stream of the PAM site.

In some embodiments, the method does not require a canonical (*e.g.*, NGG) PAM
15 site. In some embodiments, the nucleobase editor comprises a linker or a spacer. In some
embodiments, the linker or spacer is 1-25 amino acids in length. In some embodiments,
the linker or spacer is 5-20 amino acids in length. In some embodiments, the linker or
spacer is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length.

In some embodiments, the base editing fusion proteins provided herein need to be
20 positioned at a precise location, for example, where a target base is placed within a
defined region (*e.g.*, a “deamination window”). In some embodiments, a target can be
within a 4 base region. In some embodiments, such a defined target region can be
approximately 15 bases upstream of the PAM. See Komor, A.C., *et al.*, “Programmable
editing of a target base in genomic DNA without double-stranded DNA cleavage” *Nature*
25 533, 420-424 (2016); Gaudelli, N.M., *et al.*, “Programmable base editing of A•T to G•C
in genomic DNA without DNA cleavage” *Nature* 551, 464-471 (2017); and Komor,
A.C., *et al.*, “Improved base excision repair inhibition and bacteriophage Mu Gam protein
yields C:G-to-T:A base editors with higher efficiency and product purity” *Science*
Advances 3:eaa04774 (2017), the entire contents of which are hereby incorporated by
30 reference.

In some embodiments, the target region comprises a target window, wherein the
target window comprises the target nucleobase pair. In some embodiments, the target
window comprises 1- 10 nucleotides. In some embodiments, the target window is 1, 2, 3,

4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some embodiments, the intended edit of base pair is within the target window. In some embodiments, the target window comprises the intended edit of base pair. In some embodiments, the method is performed using any of the base editors provided herein. In some embodiments, a target window is a deamination window. A deamination window can be the defined region in which a base editor acts upon and deaminates a target nucleotide. In some embodiments, the deamination window is within a 2, 3, 4, 5, 6, 7, 8, 9, or 10 base regions. In some embodiments, the deamination window is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 bases upstream of the PAM.

10 The base editors of the present disclosure can comprise any domain, feature or amino acid sequence which facilitates the editing of a target polynucleotide sequence. For example, in some embodiments, the base editor comprises a nuclear localization sequence (NLS). In some embodiments, an NLS of the base editor is localized between a deaminase domain and a polynucleotide programmable nucleotide binding domain. In some embodiments, an NLS of the base editor is localized C-terminal to a polynucleotide programmable nucleotide binding domain.

Other exemplary features that can be present in a base editor as disclosed herein are localization sequences, such as cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags, biotin ligase tags, FAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. In some embodiments, the fusion protein comprises one or more His tags.

30 In some embodiments, non-limiting exemplary cytidine base editors (CBE) include BE1 (APOBEC1-XTEN-dCas9), BE2 (APOBEC1-XTEN-dCas9-UGI), BE3 (APOBEC1-XTEN-dCas9(A840H)-UGI), BE3-Gam, saBE3, saBE4-Gam, BE4, BE4-Gam, saBE4, or saB4E-Gam. BE4 extends the APOBEC1-Cas9n(D10A) linker to 32 amino acids and the Cas9n-UGI linker to 9 amino acids, and appends a second copy of

UGI to the C-terminus of the construct with another 9-amino acid linker into a single base editor construct. The base editors saBE3 and saBE4 have the *S. pyogenes* Cas9n(D10A) replaced with the smaller *S. aureus* Cas9n(D10A). BE3-Gam, saBE3-Gam, BE4-Gam, and saBE4-Gam have 174 residues of Gam protein fused to the N-terminus of BE3,
 5 saBE3, BE4, and saBE4 via the 16 amino acid XTEN linker.

In some embodiments, the adenosine base editor (ABE) can deaminate adenine in DNA. In some embodiments, ABE is generated by replacing APOBEC1 component of BE3 with natural or engineered *E. coli* TadA, human ADAR2, mouse ADA, or human ADAT2. In some embodiments, ABE comprises evolved TadA variant. In some
 10 embodiments, the ABE is ABE 1.2 (TadA*-XTEN-nCas9-NLS). In some embodiments, TadA* comprises A106V and D108N mutations.

In some embodiments, the ABE is a second-generation ABE. In some embodiments, the ABE is ABE2.1, which comprises additional mutations D147Y and E155V in TadA* (TadA*2.1). In some embodiments, the ABE is ABE2.2, ABE2.1 fused
 15 to catalytically inactivated version of human alkyl adenine DNA glycosylase (AAG with E125Q mutation). In some embodiments, the ABE is ABE2.3, ABE2.1 fused to catalytically inactivated version of *E. coli* Endo V (inactivated with D35A mutation). In some embodiments, the ABE is ABE2.6 which has a linker twice as long (32 amino acids, (SGGS)₂ (SEQ ID NO: 255)-XTEN-(SGGS)₂ (SEQ ID NO: 255)) as the linker in
 20 ABE2.1. In some embodiments, the ABE is ABE2.7, which is ABE2.1 tethered with an additional wild-type TadA monomer. In some embodiments, the ABE is ABE2.8, which is ABE2.1 tethered with an additional TadA*2.1 monomer. In some embodiments, the ABE is ABE2.9, which is a direct fusion of evolved TadA (TadA*2.1) to the N-terminus of ABE2.1. In some embodiments, the ABE is ABE2.10, which is a direct fusion of wild-
 25 type TadA to the N-terminus of ABE2.1. In some embodiments, the ABE is ABE2.11, which is ABE2.9 with an inactivating E59A mutation at the N-terminus of TadA* monomer. In some embodiments, the ABE is ABE2.12, which is ABE2.9 with an inactivating E59A mutation in the internal TadA* monomer.

In some embodiments, the ABE is a third generation ABE. In some embodiments,
 30 the ABE is ABE3.1, which is ABE2.3 with three additional TadA mutations (L84F, H123Y, and I156F).

In some embodiments, the ABE is a fourth generation ABE. In some embodiments, the ABE is ABE4.3, which is ABE3.1 with an additional TadA mutation A142N (TadA*4.3).

In some embodiments, the ABE is a fifth generation ABE. In some embodiments, the ABE is ABE5.1, which is generated by importing a consensus set of mutations from surviving clones (H36L, R51L, S146C, and K157N) into ABE3.1. In some embodiments, the ABE is ABE5.3, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to an internal evolved TadA*. In some embodiments, the ABE is ABE5.2, ABE5.4, ABE5.5, ABE5.6, ABE5.7, ABE5.8, ABE5.9, ABE5.10, ABE5.11, ABE5.12, ABE5.13, or ABE5.14, as shown in **Table 11** below. In some embodiments, the ABE is a sixth generation ABE. In some embodiments, the ABE is ABE6.1, ABE6.2, ABE6.3, ABE6.4, ABE6.5, or ABE6.6, as shown in **Table 11** below. In some embodiments, the ABE is a seventh generation ABE. In some embodiments, the ABE is ABE7.1, ABE7.2, ABE7.3, ABE7.4, ABE7.5, ABE7.6, ABE7.7, ABE7.8, ABE 7.9, or ABE7.10, as shown in **Table 11** below.

Table 11. Genotypes of ABEs

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
ABE1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
ABE2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
ABE5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
ABE6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE6.5	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.3	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K
ABE7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.8	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

In some embodiments, the base editor is an eighth generation ABE (ABE8). In some embodiments, the ABE8 contains a TadA*8 variant. In some embodiments, the ABE8 has a monomeric construct containing a TadA*8 variant ("ABE8.x-m"). In some

5 embodiments, the ABE8 is ABE8.1-m, which has a monomeric construct containing TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is ABE8.2-m, which has a monomeric construct containing TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-m, which has a monomeric construct containing TadA*7.10 with a Q154S mutation (TadA*8.3). In

some embodiments, the ABE8 is ABE8.4-m, which has a monomeric construct containing TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-m, which has a monomeric construct containing TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-m, which has a monomeric construct containing TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-m, which has a monomeric construct containing TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-m, which has a monomeric construct containing TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-m, which has a monomeric construct containing TadA*7.10 with Y147T and Q154S mutations (TadA*8.12).

In some embodiments, the ABE8 is ABE8.13-m, which has a monomeric construct containing TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-m, which has a monomeric construct containing TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-m, which has a monomeric construct containing TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the ABE8 is ABE8.16-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-m, which has a monomeric construct containing TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-m, which has a monomeric construct containing TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations

(TadA*8.20). In some embodiments, the ABE8 is ABE8.21-m, which has a monomeric construct containing TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-m, which has a monomeric construct containing TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some

5 embodiments, the ABE8 is ABE8.23-m, which has a monomeric construct containing TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24).

10 In some embodiments, the ABE8 has a heterodimeric construct containing wild-type *E. coli* TadA fused to a TadA*8 variant ("ABE8.x-d"). In some embodiments, the ABE8 is ABE8.1-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is ABE8.2-d, which has a heterodimeric construct containing wild-type *E. coli*

15 TadA fused to TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a Q154S mutation (TadA*8.3). In some embodiments, the ABE8 is ABE8.4-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments,

20 the ABE8 is ABE8.5-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-d, which has a heterodimeric construct containing wild-type *E. coli*

25 TadA fused to TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147R, Q154R and I76Y

30 mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused

to TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147T and Q154S mutations (TadA*8.12). In some embodiments, the ABE8 is ABE8.13-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the ABE8 is ABE8.16-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24).

In some embodiments, the ABE8 has a heterodimeric construct containing TadA*7.10 fused to a TadA*8 variant ("ABE8.x-7"). In some embodiments, the ABE8 is ABE8.1-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is

5 ABE8.2-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Q154S mutation (TadA*8.3). In some embodiments, the ABE8 is ABE8.4-7, which has a heterodimeric construct containing TadA*7.10 fused to

10 TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is

15 ABE8.7-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some

20 embodiments, the ABE8 is ABE8.9-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147T and Q154R mutations

25 (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147T and Q154S mutations (TadA*8.12). In some embodiments, the ABE8 is ABE8.13-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the

30 ABE8 is ABE8.14-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the

ABE8 is ABE8.16-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24).

In some embodiments, the ABE is ABE8.1-m, ABE8.2-m, ABE8.3-m, ABE8.4-m, ABE8.5-m, ABE8.6-m, ABE8.7-m, ABE8.8-m, ABE8.9-m, ABE8.10-m, ABE8.11-m, ABE8.12-m, ABE8.13-m, ABE8.14-m, ABE8.15-m, ABE8.16-m, ABE8.17-m, ABE8.18-m, ABE8.19-m, ABE8.20-m, ABE8.21-m, ABE8.22-m, ABE8.23-m, ABE8.24-m, ABE8.1-d, ABE8.2-d, ABE8.3-d, ABE8.4-d, ABE8.5-d, ABE8.6-d, ABE8.7-d, ABE8.8-d, ABE8.9-d, ABE8.10-d, ABE8.11-d, ABE8.12-d, ABE8.13-d, ABE8.14-d, ABE8.15-d, ABE8.16-d, ABE8.17-d, ABE8.18-d, ABE8.19-d, ABE8.20-d, ABE8.21-d, ABE8.22-d, ABE8.23-d, or ABE8.24-d as shown in **Table 12** below. In **Table 12**, “monomer” indicates an ABE comprising a single TadA*7.10 comprising the indicated alterations, and “heterodimer” indicates an ABE comprising a TadA*7.10 comprising the indicated alterations fused to an *E. coli* TadA adenosine deaminase.

Table 12: Adenosine Deaminase Base Editor 8 (ABE8) Variants.

ABE8	Adenosine Deaminase	Adenosine Deaminase Description
ABE8.1-m	TadA*8.1	Monomer_TadA*7.10 + Y147T
ABE8.2-m	TadA*8.2	Monomer_TadA*7.10 + Y147R
ABE8.3-m	TadA*8.3	Monomer_TadA*7.10 + Q154S
ABE8.4-m	TadA*8.4	Monomer_TadA*7.10 + Y123H
ABE8.5-m	TadA*8.5	Monomer_TadA*7.10 + V82S
ABE8.6-m	TadA*8.6	Monomer_TadA*7.10 + T166R
ABE8.7-m	TadA*8.7	Monomer_TadA*7.10 + Q154R
ABE8.8-m	TadA*8.8	Monomer_TadA*7.10 + Y147R_Q154R_Y123H
ABE8.9-m	TadA*8.9	Monomer_TadA*7.10 + Y147R_Q154R_I76Y
ABE8.10-m	TadA*8.10	Monomer_TadA*7.10 + Y147R_Q154R_T166R
ABE8.11-m	TadA*8.11	Monomer_TadA*7.10 + Y147T_Q154R
ABE8.12-m	TadA*8.12	Monomer_TadA*7.10 + Y147T_Q154S
ABE8.13-m	TadA*8.13	Monomer_TadA*7.10 + Y123H_Y147R_Q154R_I76Y
ABE8.14-m	TadA*8.14	Monomer_TadA*7.10 + I76Y_V82S
ABE8.15-m	TadA*8.15	Monomer_TadA*7.10 + V82S_Y147R
ABE8.16-m	TadA*8.16	Monomer_TadA*7.10 + V82S_Y123H_Y147R
ABE8.17-m	TadA*8.17	Monomer_TadA*7.10 + V82S_Q154R
ABE8.18-m	TadA*8.18	Monomer_TadA*7.10 + V82S_Y123H_Q154R
ABE8.19-m	TadA*8.19	Monomer_TadA*7.10 + V82S_Y123H_Y147R_Q154R
ABE8.20-m	TadA*8.20	Monomer_TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R
ABE8.21-m	TadA*8.21	Monomer_TadA*7.10 + Y147R_Q154S
ABE8.22-m	TadA*8.22	Monomer_TadA*7.10 + V82S_Q154S
ABE8.23-m	TadA*8.23	Monomer_TadA*7.10 + V82S_Y123H
ABE8.24-m	TadA*8.24	Monomer_TadA*7.10 + V82S_Y123H_Y147T
ABE8.1-d	TadA*8.1	Heterodimer_(WT) + (TadA*7.10 + Y147T)
ABE8.2-d	TadA*8.2	Heterodimer_(WT) + (TadA*7.10 + Y147R)
ABE8.3-d	TadA*8.3	Heterodimer_(WT) + (TadA*7.10 + Q154S)
ABE8.4-d	TadA*8.4	Heterodimer_(WT) + (TadA*7.10 + Y123H)
ABE8.5-d	TadA*8.5	Heterodimer_(WT) + (TadA*7.10 + V82S)
ABE8.6-d	TadA*8.6	Heterodimer_(WT) + (TadA*7.10 + T166R)
ABE8.7-d	TadA*8.7	Heterodimer_(WT) + (TadA*7.10 + Q154R)
ABE8.8-d	TadA*8.8	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_Y123H)
ABE8.9-d	TadA*8.9	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_I76Y)
ABE8.10-d	TadA*8.10	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_T166R)
ABE8.11-d	TadA*8.11	Heterodimer_(WT) + (TadA*7.10 + Y147T_Q154R)

ABE8	Adenosine Deaminase	Adenosine Deaminase Description
ABE8.12-d	TadA*8.12	Heterodimer (WT) + (TadA*7.10 + Y147T_Q154S)
ABE8.13-d	TadA*8.13	Heterodimer (WT) + (TadA*7.10 + Y123H Y147T Q154R I76Y)
ABE8.14-d	TadA*8.14	Heterodimer (WT) + (TadA*7.10 + I76Y V82S)
ABE8.15-d	TadA*8.15	Heterodimer (WT) + (TadA*7.10 + V82S Y147R)
ABE8.16-d	TadA*8.16	Heterodimer (WT) + (TadA*7.10 + V82S Y123H Y147R)
ABE8.17-d	TadA*8.17	Heterodimer (WT) + (TadA*7.10 + V82S_Q154R)
ABE8.18-d	TadA*8.18	Heterodimer (WT) + (TadA*7.10 + V82S Y123H Q154R)
ABE8.19-d	TadA*8.19	Heterodimer (WT) + (TadA*7.10 + V82S Y123H Y147R Q154R)
ABE8.20-d	TadA*8.20	Heterodimer (WT) + (TadA*7.10 + I76Y V82S Y123H Y147R Q154R)
ABE8.21-d	TadA*8.21	Heterodimer (WT) + (TadA*7.10 + Y147R_Q154S)
ABE8.22-d	TadA*8.22	Heterodimer (WT) + (TadA*7.10 + V82S_Q154S)
ABE8.23-d	TadA*8.23	Heterodimer (WT) + (TadA*7.10 + V82S Y123H)
ABE8.24-d	TadA*8.24	Heterodimer (WT) + (TadA*7.10 + V82S Y123H Y147T)

In some embodiments, the ABE8 is ABE8a-m, which has a monomeric construct containing TadA*7.10 with R26C, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-m, which has a monomeric construct containing TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-m, which has a monomeric construct containing TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some embodiments, the ABE8 is ABE8d-m, which has a monomeric construct containing TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some embodiments, the ABE8 is ABE8e-m, which has a monomeric construct containing TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).

In some embodiments, the ABE8 is ABE8a-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with R26C, A109S, T111R, D119, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some

embodiments, the ABE8 is ABE8c-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some embodiments, the ABE8 is ABE8d-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some
 5 embodiments, the ABE8 is ABE8e-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).

In some embodiments, the ABE8 is ABE8a-7, which has a heterodimeric
 10 construct containing TadA*7.10 fused to TadA*7.10 with R26C, A109S, T111R, D119, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-7,
 15 which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some embodiments, the ABE8 is ABE8d-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some embodiments, the ABE8 is ABE8e-7, which has a
 20 heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e). In some embodiments, the ABE is ABE8a-m, ABE8b-m, ABE8c-m, ABE8d-m, ABE8e-m, ABE8a-d, ABE8b-d, ABE8c-d, ABE8d-d, or ABE8e-d, as shown in **Table 13** below. In some embodiments, the ABE is ABE8e-m or ABE8e-d. ABE8e shows efficient
 25 adenine base editing activity and low indel formation when used with Cas homologues other than SpCas9, for example, SaCas9, SaCas9-KKH, Cas12a homologues, *e.g.*, LbCas12a, enAs-Cas12a, SpCas9-NG and circularly permuted CP1028-SpCas9 and CP1041-SpCas9. In addition to the mutations shown for ABE8e in **Table 13**, off-target RNA and DNA editing were reduced by introducing a V106W substitution into the TadA
 30 domain (as described in M. Richter *et al.*, 2020, Nature Biotechnology, doi.org/10.1038/s41587-020-0453-z, the entire contents of which are incorporated by reference herein).

Table 13: Additional Adenosine Deaminase Base Editor 8 Variants

ABE8 Base Editor	Adenosine Deaminase	Adenosine Deaminase Description
ABE8a-m	TadA*8a	Monomer_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8b-m	TadA*8b	Monomer_TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8c-m	TadA*8c	Monomer_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8d-m	TadA*8d	Monomer_TadA*7.10 + V88A + T111R + D119N + F149Y
ABE8e-m	TadA*8e	Monomer_TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8a-d	TadA*8a	Heterodimer_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)
ABE8b-d	TadA*8b	Heterodimer_(WT) + (TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8c-d	TadA*8c	Heterodimer_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8d-d	TadA*8d	Heterodimer_(WT) + (TadA*7.10 + V88A + T111R + D119N + F149Y)
ABE8e-d	TadA*8e	Heterodimer_(WT) + (TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)

In some embodiments, base editors (e.g., ABE8) are generated by cloning an adenosine deaminase variant (e.g., TadA*8) into a scaffold that includes a circular permutant Cas9 (e.g., CP5 or CP6) and a bipartite nuclear localization sequence. In some embodiments, the base editor (e.g., ABE7.9, ABE7.10, or ABE8) is an NGC PAM CP5 variant (*S. pyogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (e.g., ABE7.9, ABE7.10, or ABE8) is an AGA PAM CP5 variant (*S. pyogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (e.g., ABE7.9, ABE7.10, or ABE8) is an NGC PAM CP6 variant (*S. pyogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (e.g., ABE7.9, ABE7.10, or ABE8) is an AGA PAM CP6 variant (*S. pyogenes* Cas9 or spVRQR Cas9).

In some embodiments, the ABE has a genotype as shown in **Table 14** below.

Table 14. Genotypes of ABEs

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

As shown in **Table 15** below, genotypes of 40 ABE8s are described. Residue positions in the evolved *E. coli* TadA portion of ABE are indicated. Mutational changes in ABE8 are

shown when distinct from ABE7.10 mutations. In some embodiments, the ABE has a genotype of one of the ABEs as shown in Table 15 below.

Table 15. Residue Identity in Evolved TadA

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	R	L	A	L	I	V	F	V	N	Y	C	Y	P	Q	V	F	N	T
ABE8.1-m												T						
ABE8.2-m												R						
ABE8.3-m														S				
ABE8.4-m										H								
ABE8.5-m						S												
ABE8.6-m																		R
ABE8.7-m														R				
ABE8.8-m										H		R		R				
ABE8.9-m					Y							R		R				
ABE8.10-m												R		R				R
ABE8.11-m												T		R				
ABE8.12-m												T		S				
ABE8.13-m					Y					H		R		R				
ABE8.14-m					Y	S												
ABE8.15-m						S						R						
ABE8.16-m						S				H		R						
ABE8.17-m						S								R				
ABE8.18-m						S				H				R				
ABE8.19-m						S				H		R		R				
ABE8.20-m					Y	S				H		R		R				
ABE8.21-m												R		S				
ABE8.22-m						S								S				
ABE8.23-m						S				H								
ABE8.24-m						S				H		T						
ABE8.1-d												T						
ABE8.2-d												R						
ABE8.3-d														S				
ABE8.4-d										H								
ABE8.5-d						S												
ABE8.6-d																		R
ABE8.7-d														R				
ABE8.8-d										H		R		R				
ABE8.9-d					Y							R		R				
ABE8.10-d												R		R				R
ABE8.11-d												T		R				
ABE8.12-d												T		S				

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE8.13-d					Y					H		R		R				
ABE8.14-d					Y	S												
ABE8.15-d						S						R						
ABE8.16-d						S				H		R						
ABE8.17-d						S								R				
ABE8.18-d						S				H				R				
ABE8.19-d						S				H		R		R				
ABE8.20-d					Y	S				H		R		R				
ABE8.21-d												R		S				
ABE8.22-d						S								S				
ABE8.23-d						S				H								
ABE8.24-d						S				H		T						

In some embodiments, the base editor is ABE8.1, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

5 ABE8.1_Y147T_CP5_NGC PAM_monomer
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHD
PTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQK
KAQSSTDSSGSSGGSSGSETPGTSESATPESSGGSSGGSEIGKATAKYFFYSNIM
10 NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIV
KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSSF EKNPIDFLEAKGYKEVKKDLIHKLP
KYSLFELENGRRKMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPE
DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIR
15 EQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVL DATLIHQ SITGLYE
TRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVIT
DEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR
RKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNS
20 DVDKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKS NFDLAEDAKLQLSKD TYDDDLNLLAQIGD
QYADFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA
LVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL

LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKDNREKIE
 KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEFVVDKGASAQSFIER
 MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
 KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
 5 LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTIAHLFDDKVMK
 QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDS
 LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
 RHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
 QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNK
 10 VLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAER
 GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
 LKSKLVSDFRKDFQFYKVINNYHHAHDAYLNNAVVGTAIIKKYPKLESEFV
 YGDYKVYDVRKMIKSEQEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 256).

In the above sequence, the plain text denotes an adenosine deaminase sequence,
 15 bold sequence indicates sequence derived from Cas9, the italicized sequence denotes a
 linker sequence, and the underlined sequence denotes a bipartite nuclear localization
 sequence. Other ABE8 sequences are provided in the attached sequence listing (SEQ ID
 NOs: 257-279).

In some embodiments, the base editor is a ninth generation ABE (ABE9). In
 20 some embodiments, the ABE9 contains a Tada*9 variant. ABE9 base editors include an
 adenosine deaminase variant comprising an amino acid sequence, which contains
 alterations relative to an ABE 7*10 reference sequence, as described herein. Exemplary
 ABE9 variants are listed in **Table 16**. Details of ABE9 base editors are described in
 International PCT Application No. PCT/2020/049975, which is incorporated herein by
 25 reference for its entirety. In **Table 16**, “monomer” indicates an ABE comprising a single
 Tada*7.10 comprising the indicated alterations and “heterodimer” indicates an ABE
 comprising a Tada*7.10 comprising the indicated alterations fused to an *E. coli* Tada
 adenosine deaminase.

Table 16. Adenosine Deaminase Base Editor 9 (ABE9) Variants.

ABE9 Description	Alterations
ABE9.1 monomer	E25F, V82S, Y123H, T133K, Y147R, Q154R
ABE9.2 monomer	E25F, V82S, Y123H, Y147R, Q154R
ABE9.3 monomer	V82S, Y123H, P124W, Y147R, Q154R
ABE9.4 monomer	L51W, V82S, Y123H, C146R, Y147R, Q154R

ABE9 Description	Alterations
ABE9.5 monomer	P54C, V82S, Y123H, Y147R, Q154R
ABE9.6 monomer	Y73S, V82S, Y123H, Y147R, Q154R
ABE9.7 monomer	N38G, V82T, Y123H, Y147R, Q154R
ABE9.8 monomer	R23H, V82S, Y123H, Y147R, Q154R
ABE9.9 monomer	R21N, V82S, Y123H, Y147R, Q154R
ABE9.10 monomer	V82S, Y123H, Y147R, Q154R, A158K
ABE9.11 monomer	N72K, V82S, Y123H, D139L, Y147R, Q154R,
ABE9.12 monomer	E25F, V82S, Y123H, D139M, Y147R, Q154R
ABE9.13 monomer	M70V, V82S, M94V, Y123H, Y147R, Q154R
ABE9.14 monomer	Q71M, V82S, Y123H, Y147R, Q154R
ABE9.15 heterodimer	E25F, V82S, Y123H, T133K, Y147R, Q154R
ABE9.16 heterodimer	E25F, V82S, Y123H, Y147R, Q154R
ABE9.17 heterodimer	V82S, Y123H, P124W, Y147R, Q154R
ABE9.18 heterodimer	L51W, V82S, Y123H, C146R, Y147R, Q154R
ABE9.19 heterodimer	P54C, V82S, Y123H, Y147R, Q154R
ABE9.2 heterodimer	Y73S, V82S, Y123H, Y147R, Q154R
ABE9.21 heterodimer	N38G, V82T, Y123H, Y147R, Q154R
ABE9.22 heterodimer	R23H, V82S, Y123H, Y147R, Q154R
ABE9.23 heterodimer	R21N, V82S, Y123H, Y147R, Q154R
ABE9.24 heterodimer	V82S, Y123H, Y147R, Q154R, A158K
ABE9.25 heterodimer	N72K, V82S, Y123H, D139L, Y147R, Q154R,
ABE9.26 heterodimer	E25F, V82S, Y123H, D139M, Y147R, Q154R
ABE9.27 heterodimer	M70V, V82S, M94V, Y123H, Y147R, Q154R
ABE9.28 heterodimer	Q71M, V82S, Y123H, Y147R, Q154R
ABE9.29 monomer	E25F I76Y V82S Y123H Y147R Q154R
ABE9.30 monomer	I76Y V82T Y123H Y147R Q154R
ABE9.31 monomer	N38G I76Y V82S Y123H Y147R Q154R
ABE9.32 monomer	N38G I76Y V82T Y123H Y147R Q154R
ABE9.33 monomer	R23H I76Y V82S Y123H Y147R Q154R
ABE9.34 monomer	P54C I76Y V82S Y123H Y147R Q154R
ABE9.35 monomer	R21N I76Y V82S Y123H Y147R Q154R
ABE9.36 monomer	I76Y V82S Y123H D138M Y147R Q154R
ABE9.37 monomer	Y72S I76Y V82S Y123H Y147R Q154R
ABE9.38 heterodimer	E25F I76Y V82S Y123H Y147R Q154R
ABE9.39 heterodimer	I76Y V82T Y123H Y147R Q154R
ABE9.40 heterodimer	N38G I76Y V82S Y123H Y147R Q154R
ABE9.41 heterodimer	N38G I76Y V82T Y123H Y147R Q154R
ABE9.42 heterodimer	R23H I76Y V82S Y123H Y147R Q154R
ABE9.43 heterodimer	P54C I76Y V82S Y123H Y147R Q154R
ABE9.44 heterodimer	R21N I76Y V82S Y123H Y147R Q154R
ABE9.45 heterodimer	I76Y V82S Y123H D138M Y147R Q154R
ABE9.46 heterodimer	Y72S I76Y V82S Y123H Y147R Q154R
ABE9.47 monomer	N72K V82S, Y123H, Y147R, Q154R
ABE9.48 monomer	Q71M V82S, Y123H, Y147R, Q154R

ABE9 Description	Alterations
ABE9.49 monomer	M70V, V82S, M94V, Y123H, Y147R, Q154R
ABE9.50 monomer	V82S, Y123H, T133K, Y147R, Q154R
ABE9.51 monomer	V82S, Y123H, T133K, Y147R, Q154R, A158K
ABE9.52 monomer	M70V, Q71M, N72K, V82S, Y123H, Y147R, Q154R
ABE9.53 heterodimer	N72K V82S, Y123H, Y147R, Q154R
ABE9.54 heterodimer	Q71M V82S, Y123H, Y147R, Q154R
ABE9.55 heterodimer	M70V, V82S, M94V, Y123H, Y147R, Q154R
ABE9.56 heterodimer	V82S, Y123H, T133K, Y147R, Q154R
ABE9.57 heterodimer	V82S, Y123H, T133K, Y147R, Q154R, A158K
ABE9.58 heterodimer	M70V, Q71M, N72K, V82S, Y123H, Y147R, Q154R

In some embodiments, the base editor comprises a domain comprising all or a portion of a uracil glycosylase inhibitor (UGI). In some embodiments, the base editor comprises a domain comprising all or a portion of a nucleic acid polymerase. In some

5 embodiments, a base editor can comprise as a domain all or a portion of a nucleic acid polymerase (NAP). For example, a base editor can comprise all or a portion of a eukaryotic NAP. In some embodiments, a NAP or portion thereof incorporated into a base editor is a DNA polymerase. In some embodiments, a NAP or portion thereof incorporated into a base editor has translesion polymerase activity. In some

10 embodiments, a NAP or portion thereof incorporated into a base editor is a translesion DNA polymerase. In some embodiments, a NAP or portion thereof incorporated into a base editor is a Rev7, Rev1 complex, polymerase iota, polymerase kappa, or polymerase eta. In some embodiments, a NAP or portion thereof incorporated into a base editor is a eukaryotic polymerase alpha, beta, gamma, delta, epsilon, gamma, eta, iota, kappa,

15 lambda, mu, or nu component. In some embodiments, a NAP or portion thereof incorporated into a base editor comprises an amino acid sequence that is at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to a nucleic acid polymerase (*e.g.*, a translesion DNA polymerase). In some embodiments, a nucleic acid polymerase or portion thereof incorporated into a base editor is a translesion DNA

20 polymerase.

In some embodiments, a domain of the base editor can comprise multiple domains. For example, the base editor comprising a polynucleotide programmable nucleotide binding domain derived from Cas9 can comprise a REC lobe and an NUC lobe

corresponding to the REC lobe and NUC lobe of a wild-type or natural Cas9.

Structurally, Cas9 includes two lobes, an α -helical recognition lobe (REC) and a nuclease lobe (NUC). REC is comprised of three α -helical domains (REC1 and REC2) and has no structural similarity to any other known protein. REC1 forms an elongated, alpha-helical structure comprising 25 alpha helices and two beta-sheets; REC2 inserted within REC1 adopts a six-helix bundle structure. The NUC lobe includes the nuclease domains RuvCs, HNH, and the C-terminal domain (CTD). The REC lobe and the NUC lobe of Cas9 fold to present a positively charged groove at their interface which accommodates the negatively charged sgRNA:target DNA heteroduplex (Nishimasu H. et al., 2014, *Cell* 156:935-49; Jiang, F. et al., 2017, *Ann. Rev. Biophysics*, 46(1):505-529). An sgRNA-DNA complex is bound at the interface between the two lobes.

In another example, the base editor can comprise one or more of a RuvCI domain, BH domain (bridge helix domain which connects the REC and NUC lobes), REC1 domain, REC2 domain, RuvCII domain, L1 domain, HNH domain, L2 domain, RuvCIII domain, WED domain, TOPO domain or CTD domain. In some embodiments, one or more domains of the base editor comprise a mutation (*e.g.*, substitution, insertion, deletion) relative to a wild-type version of a polypeptide comprising the domain. For example, an HNH domain of a polynucleotide programmable DNA binding domain can comprise an H840A substitution. In another example, a RuvCI domain of a polynucleotide programmable DNA binding domain can comprise a D10A substitution.

Different domains (*e.g.*, adjacent domains) of the base editor disclosed herein can be connected to each other with or without the use of one or more linker domains (*e.g.*, an XTEN linker domain). In some embodiments, a linker domain can be a bond (*e.g.*, covalent bond), chemical group, or a molecule linking two molecules or moieties, *e.g.*, two domains of a fusion protein, such as, for example, a first domain (*e.g.*, Cas9-derived domain) and a second domain (*e.g.*, an adenosine deaminase domain or a cytidine deaminase domain). In some embodiments, a linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-hetero atom bond, etc.). In certain embodiments, a linker is a carbon nitrogen bond of an amide linkage. In certain embodiments, a linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In certain embodiments, a linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, etc.). In certain embodiments, a linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In some embodiments,

a linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, *etc.*). In some embodiments, a linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain embodiments, a linker is based on a carbocyclic moiety (*e.g.*,
5 cyclopentane, cyclohexane). In other embodiments, a linker comprises a polyethylene glycol moiety (PEG). In certain embodiments, a linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. A linker can include functionalized moieties to facilitate attachment of a nucleophile (*e.g.*, thiol, amino) from the peptide to the linker. Any electrophile can be used as part of the linker.
10 Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates. In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of a nucleic acid editing protein. In some embodiments, a linker joins a dCas9 and a second domain (*e.g.*, UGI,
15 *etc.*).

Linkers

In certain embodiments, linkers may be used to link any of the peptides or peptide domains of the invention. The linker may be as simple as a covalent bond, or it may be a
20 polymeric linker many atoms in length. In certain embodiments, the linker is a polypeptide or based on amino acids. In other embodiments, the linker is not peptide-like. In certain embodiments, the linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-heteroatom bond, *etc.*). In certain embodiments, the linker is a carbon-nitrogen bond of an amide linkage. In certain embodiments, the linker is a cyclic
25 or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In certain embodiments, the linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, *etc.*). In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In certain embodiments, the linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid,
30 alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, *etc.*). In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain embodiments, the linker is based on a carbocyclic moiety (*e.g.*, cyclopentane, cyclohexane). In other embodiments, the linker comprises a

polyethylene glycol moiety (PEG). In other embodiments, the linker comprises amino acids. In certain embodiments, the linker comprises a peptide. In certain embodiments, the linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. The linker may include functionalized moieties to facilitate
5 attachment of a nucleophile (*e.g.*, thiol, amino) from the peptide to the linker. Any electrophile may be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates.

Typically, a linker is positioned between, or flanked by, two groups, molecules, or
10 other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, a linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, a linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, a linker is 2-100 amino acids in length, for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
15 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. In some embodiments, the linker is about 3 to about 104 (*e.g.*, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100) amino acids in length. Longer or
20 shorter linkers are also contemplated.

In some embodiments, any of the fusion proteins provided herein, comprise a cytidine or adenosine deaminase and a Cas9 domain that are fused to each other via a linker. Various linker lengths and flexibilities between the cytidine or adenosine deaminase and the Cas9 domain can be employed (*e.g.*, ranging from very flexible linkers
25 of the form (GGGS)_n (SEQ ID NO: 171), (GGGGS)_n (SEQ ID NO: 172), and (G)_n to more rigid linkers of the form (EAAAK)_n (SEQ ID NO: 173), (SGGS)_n (SEQ ID NO: 280), SGSETPGTSESATPES (SEQ ID NO: 174) (see, *e.g.*, Guilinger JP, et al. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome
modification. Nat. Biotechnol. 2014; 32(6): 577-82; the entire contents are incorporated
30 herein by reference) and (XP)_n in order to achieve the optimal length for activity for the cytidine or adenosine deaminase nucleobase editor. In some embodiments, *n* is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. In some embodiments, the linker comprises a (GGS)_n motif, wherein *n* is 1, 3, or 7. In some embodiments, cytidine deaminase or

adenosine deaminase and the Cas9 domain of any of the fusion proteins provided herein are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 174), which can also be referred to as the XTEN linker. In some embodiments, a linker comprises a plurality of proline residues and is 5-21, 5-14, 5-9, 5-7 amino acids in length, e.g., PAPAP (SEQ ID NO: 281), PAPAPA (SEQ ID NO: 282), PAPAPAP (SEQ ID NO: 283), PAPAPAPA (SEQ ID NO: 284), P(AP)4 (SEQ ID NO: 285), P(AP)7 (SEQ ID NO: 286), P(AP)10 (SEQ ID NO: 287) (see, e.g., Tan J, Zhang F, Karcher D, Bock R. Engineering of high-precision base editors for site-specific single nucleotide replacement. Nat Commun. 2019 Jan 25;10(1):439; the entire contents are incorporated herein by reference). Such proline-rich linkers are also termed “rigid” linkers.

In another embodiment, the base editor system comprises a component (protein) that interacts non-covalently with a deaminase (DNA deaminase), e.g., an adenosine or a cytidine deaminase, and transiently attracts the adenosine or cytidine deaminase to the target nucleobase in a target polynucleotide sequence for specific editing, with minimal or reduced bystander or target-adjacent effects. Such a non-covalent system and method involving deaminase-interacting proteins serves to attract a DNA deaminase to a particular genomic target nucleobase and decouples the events of on-target and target-adjacent editing, thus enhancing the achievement of more precise single base substitution mutations. In an embodiment, the deaminase-interacting protein binds to the deaminase (e.g., adenosine deaminase or cytidine deaminase) without blocking or interfering with the active (catalytic) site of the deaminase from engaging the target nucleobase (e.g., adenosine or cytidine, respectively). Such a system, termed “MagnEdit,” involves interacting proteins tethered to a Cas9 and gRNA complex and can attract a co-expressed adenosine or cytidine deaminase (either exogenous or endogenous) to edit a specific genomic target site, and is described in McCann, J. et al., 2020, “MagnEdit – interacting factors that recruit DNA-editing enzymes to single base targets,” Life-Science-Alliance, Vol. 3, No. 4 (e201900606), (doi 10.26508/lsa.201900606), the contents of which are incorporated by reference herein in their entirety. In an embodiment, the DNA deaminase is an adenosine deaminase variant (e.g., TadA*8) as described herein.

In another embodiment, a system called “Suntag,” involves non-covalently interacting components used for recruiting protein (e.g., adenosine deaminase or cytidine deaminase) components, or multiple copies thereof, of base editors to polynucleotide target sites to achieve base editing at the site with reduced adjacent target editing, for

example, as described in Tanenbaum, M.E. et al., “A protein tagging system for signal amplification in gene expression and fluorescence imaging,” *Cell*. 2014 October 23; 159(3): 635–646. doi:10.1016/j.cell.2014.09.039; and in Huang, Y.-H. et al., 2017, “DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A,” *Genome Biol* 18: 176. doi:10.1186/s13059-017-1306-z, the contents of each of which are incorporated by reference herein in their entirety. In an embodiment, the DNA deaminase is an adenosine deaminase variant (e.g., TadA*8) as described herein.

Nucleic Acid Programmable DNA Binding Proteins with Guide RNAs

10 Provided herein are compositions and methods for base editing in cells. Further provided herein are compositions comprising a guide polynucleic acid sequence, e.g. a guide RNA sequence, or a combination of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more guide RNAs as provided herein. In some embodiments, a composition for base editing as provided herein further comprises a polynucleotide that encodes a base editor, e.g. a C-base editor or an A-base editor. For example, a composition for base editing may comprise a mRNA sequence encoding a BE, a BE4, an ABE, and a combination of one or more guide RNAs as provided. A composition for base editing may comprise a base editor polypeptide and a combination of one or more of any guide RNAs provided herein. Such a composition may be used to effect base editing in a cell through different delivery approaches, for example, electroporation, nucleofection, viral transduction or transfection. In some embodiments, the composition for base editing comprises an mRNA sequence that encodes a base editor and a combination of one or more guide RNA sequences provided herein for electroporation (EP).

25 Some aspects of this disclosure provide complexes comprising any of the fusion proteins provided herein, and a guide RNA bound to a nucleic acid programmable DNA binding protein (napDNAbp) domain (e.g., a Cas9 (e.g., a dCas9, a nuclease active Cas9, or a Cas9 nickase) or Cas12) of the fusion protein. These complexes are also termed ribonucleoproteins (RNPs). In some embodiments, the guide nucleic acid (e.g., guide RNA) is from 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the guide RNA is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides long. In some embodiments, the guide RNA comprises a sequence of 15, 16, 17, 18, 19, 20, 21, 22, 23,

24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the target sequence is a DNA sequence. In some embodiments, the target sequence is an RNA sequence. In some embodiments, the target sequence is a sequence in the genome of a bacteria, yeast, fungi, insect, plant, or animal. In some embodiments, the target sequence is a sequence in the genome of a human. In some embodiments, the 3' end of the target sequence is immediately adjacent to a canonical PAM sequence (NGG). In some embodiments, the 3' end of the target sequence is immediately adjacent to a non-canonical PAM sequence (*e.g.*, a sequence listed in **Table 7** or 5'-NAA-3'). In some embodiments, the guide nucleic acid (*e.g.*, guide RNA) is complementary to a sequence in a gene of interest (*e.g.*, a gene associated with a disease or disorder).

Some aspects of this disclosure provide methods of using the fusion proteins, or complexes provided herein. For example, some aspects of this disclosure provide methods comprising contacting a DNA molecule with any of the fusion proteins provided herein, and with at least one guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an AGC, GAG, TTT, GTG, or CAA sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, or 5' (TTTV) sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an *e.g.*, TTN, DTTN, GTTN, ATTN, ATTC, DTTNT, WTTN, HATY, TTTN, TTTV, TTTC, TG, RTR, or YTN PAM site.

It will be understood that the numbering of the specific positions or residues in the respective sequences depends on the particular protein and numbering scheme used. Numbering might differ, *e.g.*, in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species may affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, *e.g.*, by sequence alignment and determination of homologous residues.

It will be apparent to those of skill in the art that in order to target any of the fusion proteins disclosed herein, to a target site, *e.g.*, a site comprising a mutation to be edited, it is typically necessary to co-express the fusion protein together with a guide

RNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for napDNAbp (e.g., Cas9 or Cas12) binding, and a guide sequence, which confers sequence specificity to the napDNAbp:nucleic acid editing enzyme/domain fusion protein. Alternatively, the guide RNA and tracrRNA may be provided separately, as two nucleic acid molecules. In some embodiments, the guide RNA comprises a structure, wherein the guide sequence comprises a sequence that is complementary to the target sequence. The guide sequence is typically 20 nucleotides long. The sequences of suitable guide RNAs for targeting napDNAbp:nucleic acid editing enzyme/domain fusion proteins to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins to specific target sequences are provided herein.

Distinct portions of sgRNA are predicted to form various features that interact with Cas9 (e.g., SpyCas9) and/or the DNA target. Six conserved modules have been identified within native crRNA:tracrRNA duplexes and single guide RNAs (sgRNAs) that direct Cas9 endonuclease activity (see Briner et al., Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality *Mol Cell*. 2014 Oct 23;56(2):333-339). The six modules include the spacer responsible for DNA targeting, the upper stem, bulge, lower stem formed by the CRISPR repeat:tracrRNA duplex, the nexus, and hairpins from the 3' end of the tracrRNA. The upper and lower stems interact with Cas9 mainly through sequence-independent interactions with the phosphate backbone. In some embodiments, the upper stem is dispensable. In some embodiments, the conserved uracil nucleotide sequence at the base of the lower stem is dispensable. The bulge participates in specific side-chain interactions with the Rec1 domain of Cas9. The nucleobase of U44 interacts with the side chains of Tyr 325 and His 328, while G43 interacts with Tyr 329. The nexus forms the core of the sgRNA:Cas9 interactions and lies at the intersection between the sgRNA and both Cas9 and the target DNA. The nucleobases of A51 and A52 interact with the side chain of Phe 1105; U56 interacts with Arg 457 and Asn 459; the nucleobase of U59 inserts into a hydrophobic pocket defined by side chains of Arg 74, Asn 77, Pro 475, Leu 455, Phe 446, and Ile 448; C60 interacts with Leu 455, Ala 456, and Asn 459, and C61 interacts with the side chain of Arg 70, which in turn interacts with C15. In

some embodiments, one or more of these mutations are made in the bulge and/or the nexus of a sgRNA for a Cas9 (e.g., spyCas9) to optimize sgRNA:Cas9 interactions.

Moreover, the tracrRNA nexus and hairpins are critical for Cas9 pairing and can be swapped to cross orthogonality barriers separating disparate Cas9 proteins, which is instrumental for further harnessing of orthogonal Cas9 proteins. In some embodiments, the nexus and hairpins are swapped to target orthogonal Cas9 proteins. In some embodiments, a sgRNA is dispensed of the upper stem, hairpin 1, and/or the sequence flexibility of the lower stem to design a guide RNA that is more compact and conformationally stable. In some embodiments, the modules are modified to optimize multiplex editing using a single Cas9 with various chimeric guides or by concurrently using orthogonal systems with different combinations of chimeric sgRNAs. Details regarding guide functional modules and methods thereof are described, for example, in Briner et al., Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell. 2014 Oct 23;56(2):333-339, the contents of which is incorporated by reference herein in its entirety.

The domains of the base editor disclosed herein can be arranged in any order. Non-limiting examples of a base editor comprising a fusion protein comprising e.g., a polynucleotide-programmable nucleotide-binding domain (e.g., Cas9 or Cas12) and a deaminase domain (e.g., cytidine or adenosine deaminase) can be arranged as follows:

NH2-[nucleobase editing domain]-Linker1-[nucleobase editing domain]-COOH;
 NH2-[deaminase]-Linker1-[nucleobase editing domain]-COOH;
 NH2-[deaminase]-Linker1-[nucleobase editing domain]-Linker2-[UGI]-COOH;
 NH2-[deaminase]-Linker1-[nucleobase editing domain]-COOH;
 NH2-[adenosine deaminase]-Linker1-[nucleobase editing domain]-COOH;
 NH2-[nucleobase editing domain]-[deaminase]-COOH;
 NH2-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
 NH2-[deaminase]-[inosine BER inhibitor]-[nucleobase editing domain]-COOH;
 NH2-[inosine BER inhibitor]-[deaminase]-[nucleobase editing domain]-COOH;
 NH2-[nucleobase editing domain]-[deaminase]-[inosine BER inhibitor]-COOH;
 NH2-[nucleobase editing domain]-[inosine BER inhibitor]-[deaminase]-COOH;
 NH2-[inosine BER inhibitor]-[nucleobase editing domain]-[deaminase]-COOH;
 NH2-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-COOH;

5 NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;

10 NH2-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;

NH2-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;

NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;

15 NH2-[inosine BER inhibitor]-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-COOH;

NH2-[inosine BER inhibitor]-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-COOH;

20 NH2-[inosine BER inhibitor]-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-COOH; or

NH2-[inosine BER inhibitor]NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-COOH.

In some embodiments, the base editing fusion proteins provided herein need to be positioned at a precise location, for example, where a target base is placed within a defined region (e.g., a “deamination window”). In some embodiments, a target can be within a 4-base region. In some embodiments, such a defined target region can be approximately 15 bases upstream of the PAM. See Komor, A.C., et al., “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage” Nature 533, 420-424 (2016); Gaudelli, N.M., et al., “Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage” Nature 551, 464-471 (2017); and Komor, A.C., et al., “Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity” Science

Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

A defined target region can be a deamination window. A deamination window can be the defined region in which a base editor acts upon and deaminates a target
5 nucleotide. In some embodiments, the deamination window is within a 2, 3, 4, 5, 6, 7, 8, 9, or 10 base regions. In some embodiments, the deamination window is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 bases upstream of the PAM.

The base editors of the present disclosure can comprise any domain, feature or amino acid sequence which facilitates the editing of a target polynucleotide sequence.
10 For example, in some embodiments, the base editor comprises a nuclear localization sequence (NLS). In some embodiments, an NLS of the base editor is localized between a deaminase domain and a napDNAbp domain. In some embodiments, an NLS of the base editor is localized C-terminal to a napDNAbp domain.

Non-limiting examples of protein domains which can be included in the fusion
15 protein include a deaminase domain (e.g., adenosine deaminase or cytidine deaminase), a uracil glycosylase inhibitor (UGI) domain, epitope tags, reporter gene sequences, and/or protein domains having one or more of the activities described herein.

A domain may be detected or labeled with an epitope tag, a reporter protein, other binding domains. Non-limiting examples of epitope tags include histidine (His) tags, V5
20 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow
25 fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). Additional protein sequences can include amino acid sequences that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions.

30

Methods of Using Fusion Proteins Comprising a Cytidine or Adenosine Deaminase and a Cas9 Domain

Some aspects of this disclosure provide methods of using the fusion proteins, or complexes provided herein. For example, some aspects of this disclosure provide
5 methods comprising contacting a DNA molecule with any of the fusion proteins provided herein, and with at least one guide RNA described herein.

In some embodiments, a fusion protein of the invention is used for editing a target gene of interest. In particular, a cytidine deaminase or adenosine deaminase nucleobase editor described herein is capable of making multiple mutations within a target sequence.
10 These mutations may affect the function of the target. For example, when a cytidine deaminase or adenosine deaminase nucleobase editor is used to target a regulatory region the function of the regulatory region is altered and the expression of the downstream protein is reduced or eliminated.

It will be understood that the numbering of the specific positions or residues in the
15 respective sequences depends on the particular protein and numbering scheme used. Numbering might be different, e.g., in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species may affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art,
20 e.g., by sequence alignment and determination of homologous residues.

It will be apparent to those of skill in the art that in order to target any of the fusion proteins comprising a Cas9 domain and a cytidine or adenosine deaminase, as disclosed herein, to a target site, e.g., a site comprising a mutation to be edited, it is typically necessary to co-express the fusion protein together with a guide RNA, e.g., an
25 sgRNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for Cas9 binding, and a guide sequence, which confers sequence specificity to the Cas9:nucleic acid editing enzyme/domain fusion protein. Alternatively, the guide RNA and tracrRNA may be provided separately, as two nucleic acid molecules. In some embodiments, the guide RNA comprises a structure, wherein the
30 guide sequence comprises a sequence that is complementary to the target sequence. The guide sequence is typically 20 nucleotides long. The sequences of suitable guide RNAs for targeting Cas9:nucleic acid editing enzyme/domain fusion proteins to specific genomic target sites will be apparent to those of skill in the art based on the instant

disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins to specific target sequences are provided
5 herein.

Base Editor Efficiency

In some embodiments, the purpose of the methods provided herein is to alter a gene and/or gene product via gene editing. The nucleobase editing proteins provided
10 herein can be used for gene editing-based human therapeutics in vitro or in vivo. It will be understood by the skilled artisan that the nucleobase editing proteins provided herein, e.g., the fusion proteins comprising a polynucleotide programmable nucleotide binding domain (e.g., Cas9) and a nucleobase editing domain (e.g., an adenosine deaminase domain or a cytidine deaminase domain) can be used to edit a nucleotide from A to G or
15 C to T.

Advantageously, base editing systems as provided herein provide genome editing without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions as CRISPR may do. In some embodiments, the present disclosure provides base editors that efficiently
20 generate an intended mutation, such as a STOP codon, in a nucleic acid (e.g., a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations, such as unintended point mutations. In some embodiments, an intended mutation is a mutation that is generated by a specific base editor (e.g., adenosine base editor or cytidine base editor) bound to a guide polynucleotide (e.g., gRNA), specifically
25 designed to generate the intended mutation. In some embodiments, the intended mutation is in a gene associated with a target antigen associated with a disease or disorder, e.g., a hemoglobinopathy (e.g., sickle cell disease). In some embodiments, the intended mutation is an adenine (A) to guanine (G) point mutation (e.g., SNP) in a gene associated with a target antigen associated with a disease or disorder, e.g., a hemoglobinopathy (e.g.,
30 sickle cell disease). In some embodiments, the intended mutation is an adenine (A) to guanine (G) point mutation within the coding region or non-coding region of a gene (e.g., regulatory region or element). In some embodiments, the intended mutation is a cytosine (C) to thymine (T) point mutation (e.g., SNP) in a gene associated with a target antigen

associated with a disease or disorder, *e.g.*, a hemoglobinopathy (*e.g.*, sickle cell disease). In some embodiments, the intended mutation is a cytosine (C) to thymine (T) point mutation within the coding region or non-coding region of a gene (*e.g.*, regulatory region or element). In some embodiments, the intended mutation is a point mutation that
 5 generates a STOP codon, for example, a premature STOP codon within the coding region of a gene. In some embodiments, the intended mutation is a mutation that eliminates a stop codon.

The base editors of the invention advantageously modify a specific nucleotide base encoding a protein without generating a significant proportion of indels. An "indel",
 10 as used herein, refers to the insertion or deletion of a nucleotide base within a nucleic acid. Such insertions or deletions can lead to frame shift mutations within a coding region of a gene. In some embodiments, it is desirable to generate base editors that efficiently modify (*e.g.* mutate) a specific nucleotide within a nucleic acid, without generating a large number of insertions or deletions (*i.e.*, indels) in the nucleic acid. In
 15 some embodiments, it is desirable to generate base editors that efficiently modify (*e.g.* mutate or methylate) a specific nucleotide within a nucleic acid, without generating a large number of insertions or deletions (*i.e.*, indels) in the nucleic acid. In certain embodiments, any of the base editors provided herein can generate a greater proportion of intended modifications (*e.g.*, methylations) versus indels. In certain embodiments, any of
 20 the base editors provided herein can generate a greater proportion of intended modifications (*e.g.*, mutations) versus indels.

In some embodiments, the base editors provided herein are capable of generating a ratio of intended mutations to indels (*i.e.*, intended point mutations:unintended point mutations) that is greater than 1:1. In some embodiments, the base editors provided
 25 herein are capable of generating a ratio of intended mutations to indels that is at least 1.5:1, at least 2:1, at least 2.5:1, at least 3:1, at least 3.5:1, at least 4:1, at least 4.5:1, at least 5:1, at least 5.5:1, at least 6:1, at least 6.5:1, at least 7:1, at least 7.5:1, at least 8:1, at least 10:1, at least 12:1, at least 15:1, at least 20:1, at least 25:1, at least 30:1, at least 40:1, at least 50:1, at least 100:1, at least 200:1, at least 300:1, at least 400:1, at least 500:1, at
 30 least 600:1, at least 700:1, at least 800:1, at least 900:1, or at least 1000:1, or more. The number of intended mutations and indels may be determined using any suitable method.

In some embodiments, the base editors provided herein can limit formation of indels in a region of a nucleic acid. In some embodiments, the region is at a nucleotide

targeted by a base editor or a region within 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nucleotide targeted by a base editor. In some embodiments, any of the base editors provided herein can limit the formation of indels at a region of a nucleic acid to less than 1%, less than 1.5%, less than 2%, less than 2.5%, less than 3%, less than 3.5%, less than 4%, less than 4.5%, less than 5%, less than 6%, less than 7%, less than 8%, less than 9%, less than 10%, less than 12%, less than 15%, or less than 20%. The number of indels formed at a nucleic acid region may depend on the amount of time a nucleic acid (e.g., a nucleic acid within the genome of a cell) is exposed to a base editor. In some embodiments, a number or proportion of indels is determined after at least 1 hour, at least 2 hours, at least 6 hours, at least 12 hours, at least 24 hours, at least 36 hours, at least 48 hours, at least 3 days, at least 4 days, at least 5 days, at least 7 days, at least 10 days, or at least 14 days of exposing a nucleic acid (e.g., a nucleic acid within the genome of a cell) to a base editor.

Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of efficiently generating an intended mutation in a nucleic acid (e.g. a nucleic acid within a genome of a subject) without generating a considerable number of unintended mutations (e.g., spurious off-target editing or bystander editing). In some embodiments, an intended mutation is a mutation that is generated by a specific base editor bound to a gRNA, specifically designed to generate the intended mutation. In some embodiments, the intended mutation is a mutation that generates a stop codon, for example, a premature stop codon within the coding region of a gene. In some embodiments, the intended mutation is a mutation that eliminates a stop codon. In some embodiments, the intended mutation is a mutation that alters the splicing of a gene. In some embodiments, the intended mutation is a mutation that alters the regulatory sequence of a gene (e.g., a gene promotor or gene repressor). In some embodiments, any of the base editors provided herein are capable of generating a ratio of intended mutations to unintended mutations (e.g., intended mutations:unintended mutations) that is greater than 1:1. In some embodiments, any of the base editors provided herein are capable of generating a ratio of intended mutations to unintended mutations that is at least 1.5:1, at least 2:1, at least 2.5:1, at least 3:1, at least 3.5:1, at least 4:1, at least 4.5:1, at least 5:1, at least 5.5:1, at least 6:1, at least 6.5:1, at least 7:1, at least 7.5:1, at least 8:1, at least 10:1, at least 12:1, at least 15:1, at least 20:1, at least 25:1, at least 30:1, at least 40:1, at least 50:1, at least 100:1, at least 150:1, at least 200:1, at least

250:1, at least 500:1, or at least 1000:1, or more. It should be appreciated that the characteristics of the base editors described herein may be applied to any of the fusion proteins, or methods of using the fusion proteins provided herein.

Base editing is often referred to as a “modification”, such as, a genetic
5 modification, a gene modification and modification of the nucleic acid sequence and is clearly understandable based on the context that the modification is a base editing modification. A base editing modification is therefore a modification at the nucleotide
10 base level, for example as a result of the deaminase activity discussed throughout the disclosure, which then results in a change in the gene sequence, and may affect the gene product. In essence therefore, the gene editing modification described herein may result in
15 a modification of the gene, structurally and/or functionally, wherein the expression of the gene product may be modified, for example, the expression of the gene is knocked out; or conversely, enhanced, or, in some circumstances, the gene function or activity may be modified. Using the methods disclosed herein, a base editing efficiency may be
20 determined as the knockdown efficiency of the gene in which the base editing is performed, wherein the base editing is intended to knockdown the expression of the gene. A knockdown level may be validated quantitatively by determining the expression level by any detection assay, such as assay for protein expression level, for example, by flow cytometry; assay for detecting RNA expression such as quantitative RT-PCR, northern
25 blot analysis, or any other suitable assay such as pyrosequencing; and may be validated qualitatively by nucleotide sequencing reactions.

In some embodiments, the modification, e.g., single base edit results in at least 10% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 10% reduction of the gene targeted expression. In some
25 embodiments, the base editing efficiency may result in at least 20% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 30% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 40% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 50% reduction of the
30 gene targeted expression. In some embodiments, the base editing efficiency may result in at least 60% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 70% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 80% reduction of the

targeted gene expression. In some embodiments, the base editing efficiency may result in at least 90% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 91% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 92% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 93% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 94% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 95% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 96% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 97% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 98% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 99% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in knockout (100% knockdown of the gene expression) of the gene that is targeted.

In some embodiments, any of base editor systems provided herein result in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than 0.02%, or less than 0.01% indel formation in the target polynucleotide sequence.

In some embodiments, targeted modifications, e.g., single base editing, are used simultaneously to target at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 different endogenous sequences for base editing with different guide RNAs. In some embodiments, targeted modifications, e.g. single base editing, are used to sequentially target at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44,

45, 46, 47, 48, 49 50, or more different endogenous gene sequences for base editing with different guide RNAs.

Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of efficiently generating an intended mutation, such as
5 a point mutation, in a nucleic acid (e.g., a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations, such as unintended point mutations (i.e., mutation of bystanders). In some embodiments, any of the base editors provided herein are capable of generating at least 0.01% of intended mutations (i.e., at least 0.01% base editing efficiency). In some embodiments, any of the base
10 editors provided herein are capable of generating at least 0.01%, 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% of intended mutations.

In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result in less than 50%, less than 40%, less than
15 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than
20 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than 0.02%, or less than 0.01% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result in less than 0.8% indel formation in the target polynucleotide sequence. In some embodiments, any of base
25 editor systems comprising one of the ABE8 base editor variants described herein result in at most 0.8% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result in less than 0.3% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising
30 one of the ABE8 base editor variants described results in lower indel formation in the target polynucleotide sequence compared to a base editor system comprising one of ABE7 base editors. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein results in lower indel formation in the

target polynucleotide sequence compared to a base editor system comprising an ABE7.10.

In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein has reduction in indel frequency compared to a base editor system comprising one of the ABE7 base editors. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein has at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% reduction in indel frequency compared to a base editor system comprising one of the ABE7 base editors. In some embodiments, a base editor system comprising one of the ABE8 base editor variants described herein has at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% reduction in indel frequency compared to a base editor system comprising an ABE7.10.

The invention provides adenosine deaminase variants (e.g., ABE8 variants) that have increased efficiency and specificity. In particular, the adenosine deaminase variants described herein are more likely to edit a desired base within a polynucleotide, and are less likely to edit bases that are not intended to be altered (e.g., “bystanders”).

In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations. In some embodiments, an unintended editing or mutation is a bystander mutation or bystander editing, for example, base editing of a target base (e.g., A or C) in an unintended or non-target position in a target window of a target nucleotide sequence. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations compared to a base editor system comprising an ABE7 base editor, e.g., ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least

55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% compared to a base editor system comprising an ABE7 base editor, e.g., ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations by at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold compared to a base editor system comprising an ABE7 base editor, e.g., ABE7.10.

10 In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing. In some embodiments, an unintended editing or mutation is a spurious mutation or spurious editing, for example, non-specific editing or guide independent editing of a target base (e.g., A or C) in an unintended or non-target region of the genome. In some
15 embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing compared to a base editor system comprising an ABE7 base editor, e.g., ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%,
20 at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% compared to a base editor system comprising an ABE7 base editor, e.g., ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing by at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold compared to a base editor system
25 comprising an ABE7 base editor, e.g., ABE7.10.
30

In some embodiments, any of the ABE8 base editor variants described herein have at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least

45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% base editing efficiency.

In some embodiments, the base editing efficiency may be measured by calculating the percentage of edited nucleobases in a population of cells. In some embodiments, any of the ABE8 base editor variants described herein have base editing efficiency of at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited nucleobases in a population of cells.

In some embodiments, any of the ABE8 base editor variants described herein has higher base editing efficiency compared to the ABE7 base editors. In some embodiments, any of the ABE8 base editor variants described herein have at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300%, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher base editing efficiency compared to an ABE7 base editor, e.g., ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at

least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher base editing efficiency compared to an ABE7 base editor, e.g., ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein have at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%,
5 at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% on-target base editing efficiency. In some embodiments, any of the ABE8 base editor variants described herein have on-target base editing efficiency of at least 0.01%, at least 1%, at least 2%, at least
10 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited target nucleobases in a population of cells.

In some embodiments, any of the ABE8 base editor variants described herein has
15 higher on-target base editing efficiency compared to the ABE7 base editors. In some embodiments, any of the ABE8 base editor variants described herein have at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at
20 least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least
25 270%, at least 280%, at least 290%, at least 300%, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher on-target base editing efficiency compared to an ABE7 base editor, e.g., ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein has
30 at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold,

at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher on-target base editing efficiency compared to an ABE7 base editor, e.g., ABE7.10.

The ABE8 base editor variants described herein may be delivered to a host cell via a plasmid, a vector, a LNP complex, or an mRNA. In some embodiments, any of the ABE8 base editor variants described herein is delivered to a host cell as an mRNA. In some embodiments, an ABE8 base editor delivered via a nucleic acid based delivery system, e.g., an mRNA, has on-target editing efficiency of at least at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited nucleobases. In some embodiments, an ABE8 base editor delivered by an mRNA system has higher base editing efficiency compared to an ABE8 base editor delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300% higher, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% on-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at

least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher on-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system.

In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than 0.02%, or less than 0.01% off-target editing in the target polynucleotide sequence.

In some embodiments, any of the ABE8 base editor variants described herein has lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described

herein has at least about 2.2 fold decrease in guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system.

In some embodiments, any of the ABE8 base editor variants described herein has
5 lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least
10 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5
15 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 5.0 fold, at least 10.0 fold, at least 20.0 fold, at least 50.0 fold, at least 70.0 fold, at least 100.0 fold, at least 120.0 fold, at least 130.0 fold, or at least 150.0 fold lower guide-
20 independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, ABE8 base editor variants described herein has 134.0 fold decrease in guide-independent off-target editing efficiency (e.g., spurious RNA deamination) when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments,
25 ABE8 base editor variants described herein does not increase guide-independent mutation rates across the genome.

In some embodiments, a single gene delivery event (e.g., by transduction, transfection, electroporation or any other method) can be used to target base editing of 5 sequences within a cell's genome. In some embodiments, a single gene delivery event can
30 be used to target base editing of 6 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 7 sequences within a cell's genome. In some embodiments, a single electroporation event can be used to target base editing of 8 sequences within a cell's genome. In some

embodiments, a single gene delivery event can be used to target base editing of 9 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 10 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 20 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 30 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 40 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 50 sequences within a cell's genome.

10 In some embodiments, the method described herein, for example, the base editing methods has minimum to no off-target effects.

In some embodiments, the base editing method described herein results in at least 50% of a cell population that have been successfully edited (i.e., cells that have been successfully engineered). In some embodiments, the base editing method described herein results in at least 55% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 60% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 65% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 70% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 75% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 80% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 85% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 90% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 95% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% of a cell population that have been successfully edited.

In some embodiments, the live cell recovery following a base editing intervention is greater than at least 60%, 70%, 80%, 90% of the starting cell population at the time of

the base editing event. In some embodiments, the live cell recovery as described above is about 70%. In some embodiments, the live cell recovery as described above is about 75%. In some embodiments, the live cell recovery as described above is about 80%. In some embodiments, the live cell recovery as described above is about 85%. In some
5 embodiments, the live cell recovery as described above is about 90%, or about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, or 100% of the cells in the population at the time of the base editing event.

In some embodiments the engineered cell population can be further expanded in vitro by about 2 fold, about 3-fold, about 4-fold, about 5-fold, about 6-fold, about 7-fold,
10 about 8-fold, about 9-fold, about 10-fold, about 15-fold, about 20-fold, about 25-fold, about 30-fold, about 35-fold, about 40-fold, about 45-fold, about 50-fold, or about 100-fold.

The number of intended mutations and indels can be determined using any suitable method, for example, as described in International PCT Application Nos.
15 PCT/2017/045381 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632); Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" *Nature* 551, 464-471 (2017); and Komor, A.C., et al., "Improved base excision repair
20 inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" *Science Advances* 3:eaao4774 (2017); the entire contents of which are hereby incorporated by reference.

In some embodiments, to calculate indel frequencies, sequencing reads are scanned for exact matches to two 10-bp sequences that flank both sides of a window in
25 which indels can occur. If no exact matches are located, the read is excluded from analysis. If the length of this indel window exactly matches the reference sequence the read is classified as not containing an indel. If the indel window is two or more bases longer or shorter than the reference sequence, then the sequencing read is classified as an insertion or deletion, respectively. In some embodiments, the base editors provided
30 herein can limit formation of indels in a region of a nucleic acid. In some embodiments, the region is at a nucleotide targeted by a base editor or a region within 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nucleotide targeted by a base editor.

The number of indels formed at a target nucleotide region can depend on the amount of time a nucleic acid (e.g., a nucleic acid within the genome of a cell) is exposed to a base editor. In some embodiments, the number or proportion of indels is determined after at least 1 hour, at least 2 hours, at least 6 hours, at least 12 hours, at least 24 hours, at least 36 hours, at least 48 hours, at least 3 days, at least 4 days, at least 5 days, at least 7 days, at least 10 days, or at least 14 days of exposing the target nucleotide sequence (e.g., a nucleic acid within the genome of a cell) to a base editor. It should be appreciated that the characteristics of the base editors as described herein can be applied to any of the fusion proteins, or methods of using the fusion proteins provided herein.

Details of base editor efficiency are described in International PCT Application Nos. PCT/2017/045381 (WO 2018/027078) and PCT/US2016/058344 (WO 2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" *Nature* 551, 464-471 (2017); and Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" *Science Advances* 3:eaa04774 (2017), the entire contents of which are hereby incorporated by reference. In some embodiments, editing of a plurality of nucleobase pairs in one or more genes using the methods provided herein results in formation of at least one intended mutation. In some embodiments, said formation of said at least one intended mutation results in the disruption the normal function of a gene. In some embodiments, said formation of said at least one intended mutation results decreases or eliminates the expression of a protein encoded by a gene. It should be appreciated that multiplex editing can be accomplished using any method or combination of methods provided herein.

Multiplex Editing

In some embodiments, the base editor system provided herein is capable of multiplex editing of a plurality of nucleobase pairs in one or more genes. In some embodiments, the plurality of nucleobase pairs is located in the same gene or in one or more genes, wherein at least one gene is located in a different locus. In some embodiments, the multiplex editing can comprise one or more guide polynucleotides. In

some embodiments, the multiplex editing can comprise one or more base editor systems. In some embodiments, the multiplex editing can comprise one or more base editor systems with a single guide polynucleotide or a plurality of guide polynucleotides. In some embodiments, the multiplex editing can comprise one or more guide polynucleotides with a single base editor system. In some embodiments, the multiplex editing can comprise at least one guide polynucleotide that does or does not require a PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, the multiplex editing can comprise a mix of at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence and at least one guide polynucleotide that require a PAM sequence to target binding to a target polynucleotide sequence. It should be appreciated that the characteristics of the multiplex editing using any of the base editors as described herein can be applied to any combination of methods using any base editor provided herein. It should also be appreciated that the multiplex editing using any of the base editors as described herein can comprise a sequential editing of a plurality of nucleobase pairs.

In some embodiments, the plurality of nucleobase pairs are in one more genes. In some embodiments, the plurality of nucleobase pairs is in the same gene. In some embodiments, at least one gene in the one more genes is located in a different locus.

In some embodiments, the editing is editing of the plurality of nucleobase pairs in at least one protein coding region, in at least one protein non-coding region, or in at least one protein coding region and at least one protein non-coding region.

In some embodiments, the editing is in conjunction with one or more guide polynucleotides. In some embodiments, the base editor system can comprise one or more base editor systems. In some embodiments, the base editor system can comprise one or more base editor systems in conjunction with a single guide polynucleotide or a plurality of guide polynucleotides. In some embodiments, the editing is in conjunction with one or more guide polynucleotide with a single base editor system. In some embodiments, the editing is in conjunction with at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence or with at least one guide polynucleotide that requires a PAM sequence to target binding to a target polynucleotide sequence, or with a mix of at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence and at least

one guide polynucleotide that does require a PAM sequence to target binding to a target polynucleotide sequence. It should be appreciated that the characteristics of the multiplex editing using any of the base editors as described herein can be applied to any of combination of the methods of using any of the base editors provided herein. It should
5 also be appreciated that the editing can comprise a sequential editing of a plurality of nucleobase pairs.

In some embodiments, the base editor system capable of multiplex editing of a plurality of nucleobase pairs in one or more genes comprises one of ABE7, ABE8, and/or ABE9 base editors. In some embodiments, the base editor system capable of multiplex
10 editing comprising one of the ABE8 base editor variants described herein has higher multiplex editing efficiency compared to the base editor system capable of multiplex editing comprising one of ABE7 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least
15 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at
20 least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300% higher, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at
25 least 500% higher multiplex editing efficiency compared the base editor system capable of multiplex editing comprising one of ABE7 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9
30 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 4.0 fold, at least 4.5 fold, at least 5.0 fold, at least 5.5 fold, or at least 6.0

fold higher multiplex editing efficiency compared the base editor system capable of multiplex editing comprising one of ABE7 base editors.

Expression of Fusion Proteins in a Host Cell

5 Fusion proteins of the invention comprising an adenosine deaminase variant may be expressed in virtually any host cell of interest, including but not limited to bacteria, yeast, fungi, insects, plants, and animal cells using routine methods known to the skilled artisan. For example, a DNA encoding an adenosine deaminase of the invention can be cloned by designing suitable primers for the upstream and downstream of CDS based on
10 the cDNA sequence. The cloned DNA may be directly, or after digestion with a restriction enzyme when desired, or after addition of a suitable linker and/or a nuclear localization signal, ligated with a DNA encoding one or more additional components of a base editing system. The base editing system is translated in a host cell to form a complex.

15 A DNA encoding a protein domain described herein can be obtained by chemically synthesizing the DNA, or by connecting synthesized partly overlapping oligoDNA short chains by utilizing the PCR method and the Gibson Assembly method to construct a DNA encoding the full length thereof. The advantage of constructing a full-length DNA by chemical synthesis or a combination of PCR method or Gibson Assembly
20 method is that the codon to be used can be designed in CDS full-length according to the host into which the DNA is introduced. In the expression of a heterologous DNA, the protein expression level is expected to increase by converting the DNA sequence thereof to a codon highly frequently used in the host organism. As the data of codon use frequency in host to be used, for example, the genetic code use frequency database
25 (<http://www.kazusa.or.jp/codon/index.html>) disclosed in the home page of Kazusa DNA Research Institute can be used, or documents showing the codon use frequency in each host may be referred to. By reference to the obtained data and the DNA sequence to be introduced, codons showing low use frequency in the host from among those used for the DNA sequence may be converted to a codon coding the same amino acid and showing
30 high use frequency.

 An expression vector containing a DNA encoding a nucleic acid sequence-recognizing module and/or a nucleic acid base converting enzyme can be produced, for

example, by linking the DNA to the downstream of a promoter in a suitable expression vector.

As the expression vector, *Escherichia coli*-derived plasmids (e.g., pBR322, pBR325, pUC12, pUC13); *Bacillus subtilis*-derived plasmids (e.g., pUB110, pTP5, pC194); yeast-derived plasmids (e.g., pSH19, pSH15); insect cell expression plasmids (e.g., pFast-Bac); animal cell expression plasmids (e.g., pA1-11, pXT1, pRc/CMV, pRc/RSV, pcDNA1/Neo); bacteriophages such as lambda phage and the like; insect virus vectors such as baculovirus and the like (e.g., BmNPV, AcNPV); animal virus vectors such as retrovirus, vaccinia virus, adenovirus and the like, and the like are used.

Regarding the promoter to be used, any promoter appropriate for a host to be used for gene expression can be used. In a conventional method using double-stranded breaks, since the survival rate of the host cell sometimes decreases markedly due to the toxicity, it is desirable to increase the number of cells by the start of the induction by using an inductive promoter. However, since sufficient cell proliferation can also be afforded by expressing the nucleic acid-modifying enzyme complex of the present invention, a constitutive promoter can be used without limitation.

For example, when the host is an animal cell, an SR.alpha. promoter, SV40 promoter, LTR promoter, cytomegalovirus (CMV) promoter, Rous sarcoma virus (RSV) promoter, Moloney mouse leukemia virus (MoMuLV), LTR, herpes simplex virus thymidine kinase (HSV-TK) promoter, and the like can be used. Of these, CMV promoter, SR.alpha. promoter and the like are preferable.

When the host is *Escherichia coli*, a trp promoter, lac promoter, recA promoter, lambda.P.sub.L promoter, lpp promoter, T7 promoter, and the like can be used.

When the host is in the genus *Bacillus*, the SPO1 promoter, SPO2 promoter, penP promoter, and the like can be used.

When the host is a yeast, the Gal1/10 promoter, PHO5 promoter, PGK promoter, GAP promoter, ADH promoter, and the like can be used.

When the host is an insect cell, the polyhedrin promoter, P10 promoter, and the like can be used.

When the host is a plant cell, the CaMV35S promoter, CaMV19S promoter, NOS promoter, and the like can be used.

Expression vectors for use in the present invention, besides those mentioned above, can comprise an enhancer, a splicing signal, a terminator, a polyA addition signal,

a selection marker such as drug resistance gene, an auxotrophic complementary gene and the like, a replication origin, and the like can be used.

An RNA encoding a protein domain described herein can be prepared by, for example, *in vitro* transcription of a nucleic acid sequence encoding any of the fusion
5 proteins disclosed herein.

A fusion protein of the invention can be intracellularly expressed by introducing into the cell an expression vector comprising a nucleic acid sequence encoding the fusion protein.

Host cells of interest include, but are not limited to, bacteria, yeast, fungi, insects,
10 plants, and animal cells. For example, a host cell may comprise bacteria from the genus *Escherichia*, such as *Escherichia coli* K12.cndot.DH1 [Proc. Natl. Acad. Sci. USA, 60, 160 (1968)], *Escherichia coli* JM103 [Nucleic Acids Research, 9, 309 (1981)], *Escherichia coli* JA221 [Journal of Molecular Biology, 120, 517 (1978)], *Escherichia coli* HB101 [Journal of Molecular Biology, 41, 459 (1969)], *Escherichia coli* C600 [Genetics,
15 39, 440 (1954)] and the like.

A host cell may comprise bacteria from the genus *Bacillus*, for example *Bacillus subtilis* M1114 [Gene, 24, 255 (1983)], *Bacillus subtilis* 207-21 [Journal of Biochemistry, 95, 87 (1984)] and the like.

A host cell may be a yeast cell. Examples of yeast cells include *Saccharomyces*
20 *cerevisiae* AH22, AH22R.sup.-, NA87-11A, DKD-5D, 20B-12, *Schizosaccharomyces pombe* NCYC1913, NCYC2036, *Pichia pastoris* KM71 and the like.

When the viral delivery methods utilize the virus AcNPV, cells from a cabbage armyworm larva-derived established line (*Spodoptera frugiperda* cell, Sf cell), MG1 cells derived from the mid-intestine of *Trichoplusia ni*, High Five™ cells derived from an
25 ovary of *Trichoplusia ni*, *Mamestra brassicae*-derived cells, *Estigmene acrea*-derived cells and the like can be used. When the virus is BmNPV, cells of *Bombyx mori*-derived established line (*Bombyx mori* N cell; BmN cell) and the like are used. As the Sf cell, for example, Sf9 cell (ATCC CRL1711), Sf21 cell [all above, In Vivo, 13, 213-217 (1977)] and the like are used.

30 An insect can be any insect, for example, larva of *Bombyx mori*, *Drosophila*, cricket, and the like [Nature, 315, 592 (1985)].

Animal cells contemplated in the present invention include, but are not limited to, cell lines such as monkey COS-7 cells, monkey Vero cells, Chinese hamster ovary (CHO)

cells, dhfr gene-deficient CHO cells, mouse L cells, mouse AtT-20 cells, mouse myeloma cells, rat GH3 cells, human FL cells and the like, pluripotent stem cells such as iPS cells, ES cells derived humans and other mammals, and primary cultured cells prepared from various tissues. Furthermore, zebrafish embryo, *Xenopus* oocyte, and the like can also be
5 used.

Plant cells are also contemplated in the present invention. Plant cells include, but are not limited to, suspended cultured cells, callus, protoplast, leaf segment, root segment and the like prepared from various plants (e.g., grain such as rice, wheat, corn, and the like; product crops such as tomato, cucumber, eggplant and the like; garden plants such as
10 carnations, *Eustoma russellianum*, and the like; and other plants such as tobacco, *Arabidopsis thaliana* and the like) are used.

All the above-mentioned host cells may be haploid (monoploid), or polyploid (e.g., diploid, triploid, tetraploid, etc.). Using conventional methods, mutations, in principle, introduced into only one homologous chromosome produce a heterogenous
15 cell. Therefore, the desired phenotype is not expressed unless the mutation is dominant. For recessive mutations, acquiring a homozygous cell can be inconvenient due to labor and time requirements. In contrast, according to the present invention, since a mutation can be introduced into any allele on the homologous chromosome in the genome, the desired phenotype can be expressed in a single generation even in the case of recessive
20 mutation, thereby solving the problem associated with conventional mutagenesis methods.

An expression vector can be introduced by a known method (e.g., the lysozyme method, the competent method, the PEG method, the CaCl₂ coprecipitation method, electroporation, microinjection, particle gun method, lipofection, *Agrobacterium*-
25 mediated delivery, etc.) according to the kind of the host.

Escherichia coli can be transformed according to the methods described in, for example, Proc. Natl. Acad. Sci. USA, 69, 2110 (1972), Gene, 17, 107 (1982).

The genus *Bacillus* can be introduced into a vector according to the methods described in, for example, Molecular & General Genetics, 168, 111 (1979).

30 A yeast can be introduced into a vector according to the methods described in, for example, Methods in Enzymology, 194, 182-187 (1991), Proc. Natl. Acad. Sci. USA, 75, 1929 (1978).

An insect cell and an insect can be introduced into a vector according to the methods described in, for example, Bio/Technology, 6, 47-55 (1988).

A vector can be introduced into an animal cell according to the methods described in, for example, Cell Engineering additional volume 8, New Cell Engineering Experiment
5 Protocol, 263-267 (1995) (published by Shujunsha), and Virology, 52, 456 (1973).

A cell comprising a vector can be cultured according to a known method according to the kind of the host. For example, when *Escherichia coli* or genus *Bacillus* is cultured, a liquid medium is preferable as a medium to be used for the culture. The medium preferably contains a carbon source, nitrogen source, inorganic substance and the
10 like necessary for the growth of the transformant. Examples of the carbon source include glucose, dextrin, soluble starch, sucrose and the like; examples of the nitrogen source include inorganic or organic substances such as ammonium salts, nitrate salts, corn steep liquor, peptone, casein, meat extract, soybean cake, potato extract and the like; and examples of the inorganic substance include calcium chloride, sodium dihydrogen
15 phosphate, magnesium chloride and the like. The medium may contain yeast extract, vitamins, growth promoting factor and the like. The pH of the medium is preferably about 5 about 8.

As a medium for culturing *Escherichia coli*, for example, M9 medium containing glucose, casamino acid [Journal of Experiments in Molecular Genetics, 431-433, Cold
20 Spring Harbor Laboratory, New York 1972] is preferable. Where necessary, for example, agents such as 3 β -indolylacrylic acid may be added to the medium to ensure an efficient function of a promoter. *Escherichia coli* is cultured at generally about 15 to about 43°C. Where necessary, aeration and stirring may be performed.

The genus *Bacillus* is cultured at generally about 30 to about 40°C. Where
25 necessary, aeration and stirring may be performed.

Examples of the medium for culturing yeast include Burkholder minimum medium [Proc. Natl. Acad. Sci. USA, 77, 4505 (1980)], SD medium containing 0.5% casamino acid [Proc. Natl. Acad. Sci. USA, 81, 5330 (1984)] and the like. The pH of the medium is preferably about 5 to about 8. The culture is performed at generally about
30 20°C to about 35°C. Where necessary, aeration and stirring may be performed.

As a medium for culturing an insect cell or insect, for example, Grace's Insect Medium [Nature, 195, 788 (1962)] containing an additive such as inactivated 10% bovine serum and the like as appropriate and the like are used. The pH of the medium is

preferably about 6.2 to about 6.4. The culture is performed at generally about 27°C. Where necessary, aeration and stirring may be performed.

As a medium for culturing an animal cell, for example, minimum essential medium (MEM) containing about 5 to about 20% of fetal bovine serum [Science, 122, 501 (1952)], Dulbecco's modified Eagle medium (DMEM) [Virology, 8, 396 (1959)], RPMI 1640 medium [The Journal of the American Medical Association, 199, 519 (1967)], 199 medium [Proceeding of the Society for the Biological Medicine, 73, 1 (1950)] and the like are used. The pH of the medium is preferably about 6 to about 8. The culture is performed at generally about 30°C. to about 40°C. Where necessary, aeration and stirring may be performed.

As a medium for culturing a plant cell, for example, MS medium, LS medium, B5 medium and the like are used. The pH of the medium is preferably about 5- about 8. The culture is performed at generally about 20°C to about 30°C. Where necessary, aeration and stirring may be performed.

When a higher eukaryotic cell, such as animal cell, insect cell, plant cell and the like is used as a host cell, a DNA encoding a base editing system of the present invention (e.g., comprising an adenosine deaminase variant) is introduced into a host cell under the regulation of an inducible promoter (e.g., metallothionein promoter (induced by heavy metal ion), heat shock protein promoter (induced by heat shock), Tet-ON/Tet-OFF system promoter (induced by addition or removal of tetracycline or a derivative thereof), steroid-responsive promoter (induced by steroid hormone or a derivative thereof) etc.), the induction substance is added to the medium (or removed from the medium) at an appropriate stage to induce expression of the nucleic acid-modifying enzyme complex, culture is performed for a given period to carry out a base editing and, introduction of a mutation into a target gene, transient expression of the base editing system can be realized.

Prokaryotic cells such as *Escherichia coli* and the like can utilize an inducible promoter. Examples of the inducible promoter include, but are not limited to, lac promoter (induced by IPTG), cspA promoter (induced by cold shock), araBAD promoter (induced by arabinose) and the like.

Alternatively, the above-mentioned inductive promoter can also be utilized as a vector removal mechanism when higher eukaryotic cells, such as animal cell, insect cell, plant cell and the like are used as a host cell. That is, a vector is mounted with a

replication origin that functions in a host cell, and a nucleic acid encoding a protein necessary for replication (e.g., SV40 on and large T antigen, oriP and EBNA-1 etc. for animal cells), of the expression of the nucleic acid encoding the protein is regulated by the above-mentioned inducible promoter. As a result, while the vector is autonomously replicable in the presence of an induction substance, when the induction substance is removed, autonomous replication is not available, and the vector naturally falls off along with cell division (autonomous replication is not possible by the addition of tetracycline and doxycycline in Tet-OFF system vector).

10

DELIVERY

The suitability of nucleobase editors to target one or more nucleotides in a target sequence (e.g., the hemoglobin beta subunit (HbB) gene or a promoter region of an HbG1/2 gene) is evaluated as described herein. In one embodiment, a single cell of interest is transfected, transduced, or otherwise modified with a nucleic acid molecule or molecules encoding a base editing system described herein together with a small amount of a vector encoding a reporter (e.g., GFP). These cells can be any cell line known in the art, including . Alternatively, primary cells (e.g., human) may be used. Cells may also be hematopoietic stem/ and progenitor cells (HSPCs) obtained from a subject or individual, such as from tissue biopsy, surgery, blood, plasma, serum, or other biological fluid. Such cells may be relevant to the eventual cell target. In embodiments, the HSPC cells are human CD34⁺ hematopoietic stem/progenitor cells (HSPCs).

Delivery may be performed using a viral vector. In one embodiment, transfection may be performed using lipid transfection (such as Lipofectamine or Fugene) or by electroporation. Following transfection, expression of a reporter (e.g., GFP) can be determined either by fluorescence microscopy or by flow cytometry to confirm consistent and high levels of transfection. These preliminary transfections can comprise different nucleobase editors to determine which combinations of editors give the greatest activity. The system can comprise one or more different vectors. In one embodiment, the base editor is codon optimized for expression of the desired cell type, preferentially a eukaryotic cell, preferably a mammalian cell or a human cell.

The activity of the nucleobase editor is assessed as described herein, i.e., by sequencing the genome of the cells to detect alterations in a target sequence. For Sanger sequencing, purified PCR amplicons are cloned into a plasmid backbone, transformed,

miniprep and sequenced with a single primer. Sequencing may also be performed using next generation sequencing (NGS) techniques. When using next generation sequencing, amplicons may be 300-500 bp with the intended cut site placed asymmetrically. Following PCR, next generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of the amplicon, e.g., for use in high throughput sequencing (for example on an Illumina MiSeq). The fusion proteins that induce the greatest levels of target specific alterations in initial tests can be selected for further evaluation.

In particular embodiments, the nucleobase editors are used to target polynucleotides of interest. In one embodiment, a nucleobase editor of the invention is delivered to cells (e.g., hematopoietic stem/ and progenitor cells (HSPCs)) in conjunction with one or more guide RNAs that are used to target one or more nucleic acid sequences of interest within the genome of a cell, thereby altering the target gene(s) (e.g., hematopoietic stem/ and progenitor cells (HSPCs)). In some embodiments, a base editor is targeted by one or more guide RNAs to introduce one or more edits to the sequence of one or more target sequences of interest (e.g., the hemoglobin beta subunit (HbB) gene or a promoter region of an HBG1/2 gene). In some embodiments, the one or more edits to the sequence of one or more genes of interest decrease or eliminate expression of the protein encoded by the gene in the host cell (e.g., hematopoietic stem/ and progenitor cells (HSPCs)). In some embodiments, expression of one or more proteins encoded by one or more target sequences of interest (e.g., the hemoglobin beta subunit (HbB) gene or a promoter region of an HBG1/2 gene) is completely knocked out or eliminated in the host cell (e.g., hematopoietic stem/ and progenitor cells (HSPCs)).

In some embodiments, the host cell is a mammalian cell. In some embodiments, the host cell is a human cell.

Nucleic Acid-Based Delivery of Base Editor Systems

Nucleic acid molecules encoding a base editor system according to the present disclosure can be administered to subjects or delivered into cells *in vitro* or *in vivo* by art-known methods or as described herein. For example, a base editor system comprising a deaminase (e.g., cytidine or adenine deaminase) can be delivered by vectors (e.g., viral or non-viral vectors), or by naked DNA, DNA complexes, lipid nanoparticles, or a combination of the aforementioned compositions.

Any RNA of the systems, for example a guide RNA or a base editor-encoding mRNA, can be delivered in the form of RNA. Base editor-encoding mRNA can be generated using in vitro transcription. For example, nuclease mRNA can be synthesized using a PCR cassette containing the following elements: T7 promoter, optional kozak sequence (GCCACC), nuclease sequence, and 3' UTR such as a 3' UTR from beta globin-polyA tail. The cassette can be used for transcription by T7 polymerase. Guide polynucleotides (e.g., gRNA) can also be transcribed using in vitro transcription from a cassette containing a T7 promoter, followed by the sequence "GG", and guide polynucleotide sequence.

To enhance expression and reduce possible toxicity (e.g., immunogenicity), the base editor-coding sequence and/or the guide nucleic acid can be modified to include one or more modified nucleosides e.g. using pseudo-U or 5-Methyl-C. In some embodiments, the base editor-coding sequence and/or the guide nucleic acid contains one or more chemically modified nucleobases, such as 2' -O-methyl (2' -OMe), 2' -deoxy (2' -H), 2' -O-C1-3alkyl-O-C1-3alkyl such as 2' -methoxyethyl ("2' -MOE"), 2' -fluoro ("2' -F"), 2' -amino ("2' -NH₂"), 2' -arabinosyl ("2' -arabino") nucleotide, 2' -F-arabinosyl ("2' -F-arabino") nucleotide, 2' -locked nucleic acid ("LNA") nucleotide, 2' -unlocked nucleic acid ("ULNA") nucleotide, a sugar in L form ("L-sugar"), 4' -thioribosyl nucleotide, or any chemical modification as described herein. In some embodiments, the base editor-coding sequence and/or the guide nucleic acid contains an internucleotide linkage modification such as phosphorothioate "P(S)" (P(S)), phosphonocarboxylate (P(CH₂)_nCOOR) such as phosphonoacetate "PACE" (P(CH₂COO-)), thiophosphonocarboxylate ((S)P(CH₂)_nCOOR) such as thiophosphonoacetate "thioPACE" ((S)P(CH₂)_nCOO-)), alkylphosphonate (P(C1-3alkyl)) such as methylphosphonate —P(CH₃), boranophosphonate (P(BH₃)), and phosphorodithioate (P(S)₂). In some embodiments, the base editor-coding sequence and/or the guide nucleic acid contains a nucleobase chemical modification such as 2-thiouracil ("2-thioU"), 2-thiocytosine ("2-thioC"), 4-thiouracil ("4-thioU"), 6-thioguanine ("6-thioG"), 2-aminoadenine ("2-aminoA"), 2-aminopurine, pseudouracil, hypoxanthine, 7-deazaguanine, 7-deaza-8-azaguanine, 7-deazaadenine, 7-deaza-8-azaadenine, 5-methylcytosine ("5-methylC"), 5-methyluracil ("5-methylU"), 5-hydroxymethylcytosine, 5-hydroxymethyluracil, 5,6-dehydrouracil, 5-propynylcytosine, 5-propynyluracil, 5-

ethynylcytosine, 5-ethynyluracil, 5-allyluracil ("5-allylU"), 5-allylcytosine ("5-allylC"), 5-aminoallyluracil ("5-aminoallylU"), 5-aminoallyl-cytosine ("5-aminoallylC"), an abasic nucleotide, Z base, P base, Unstructured Nucleic Acid ("UNA"), isoguanine ("isoG"), isocytosine ("isoC"). In some embodiments, the base editor-coding sequence and/or the

5 guide nucleic acid contains one or more isotopic modifications on the nucleotide sugar, the nucleobase, the phosphodiester linkage and/or the nucleotide phosphates. Such modifications include nucleotides comprising one or more ¹⁵N, ¹³C, ¹⁴C, Deuterium, ³H, ³²P, ¹²⁵I, ¹³¹I atoms or other atoms or elements thereof. In various embodiments, the modified nucleobase(s) reduce toxicity (e.g., immunogenicity) and/or stability (e.g.,

10 increase serum half-life) of the base editor-coding sequence (e.g., mRNA) and/or the guide nucleic acid.

Nanoparticles, which can be organic or inorganic, are useful for delivering a base editor system or component thereof. Nanoparticles are well known in the art and any suitable nanoparticle can be used to deliver a base editor system or component thereof, or

15 a nucleic acid molecule encoding such components. In one example, organic (e.g. lipid and/or polymer) nanoparticles are suitable for use as delivery vehicles in certain embodiments of this disclosure. Exemplary lipids for use in nanoparticle formulations, and/or gene transfer are shown in **Table 17** (below).

20 **Table 17**

Lipids Used for Gene Transfer		
Lipid	Abbreviation	Feature
1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine	DOPC	Helper
1,2-Dioleoyl-sn-glycero-3-phosphatidylethanolamine	DOPE	Helper
Cholesterol		Helper
N-[1-(2,3-Dioleoyloxy)propyl]N,N,N-trimethylammonium chloride	DOTMA	Cationic
1,2-Dioleoyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Diocetadecylamidoglycylspermine	DOGS	Cationic
N-(3-Aminopropyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide	GAP-DLRIE	Cationic
Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleoyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleoyloxy-N-[2(sperminecarboxamido-ethyl)-N,N-dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleoyl-3-trimethylammonium-propane	DOPA	Cationic
N-(2-Hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic

Lipids Used for Gene Transfer		
Lipid	Abbreviation	Feature
Dimyristoxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3 β -[N-(N',N'-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Dioleoyl-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Diocetadecylamidoglycylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxymethyloxy)ethyl]trimethylammonium bromide	CLIP-6	Cationic
Ethylidimyristoylphosphatidylcholine	EDMPC	Cationic
1,2-Distearoyl-N,N-dimethyl-3-aminopropane	DSDMA	Cationic
1,2-Dimyristoyl-trimethylammonium propane	DMTAP	Cationic
O,O'-Dimyristyl-N-lysyl aspartate	DMKE	Cationic
1,2-Distearoyl-sn-glycero-3-ethylphosphocholine	DSEPC	Cationic
N-Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
N-t-Butyl-N0-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl]imidazolium chloride	DOTIM	Cationic
N1 -Cholesterylloxycarbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)-N-ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoleoyl-3-dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane	DLin-KC2-DMA	Cationic
dilinoleyl-methyl-4-dimethylaminobutyrate	DLin-MC3-DMA	Cationic

Table 18 lists exemplary polymers for use in gene transfer and/or nanoparticle formulations.

5 Table 18

Polymers Used for Gene Transfer	
Polymer	Abbreviation
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis (succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimidate	DTBP
Poly(ethylene imine)biscarbamate	PEIC
Poly(L-lysine)	PLL
Histidine modified PLL	
Poly(N-vinylpyrrolidone)	PVP
Poly(propylenimine)	PPI
Poly(amidoamine)	PAMAM

Polymers Used for Gene Transfer	
Polymer	Abbreviation
Poly(amidoethylenimine)	SS-PAEI
Triethylenetetramine	TETA
Poly(β -aminoester)	
Poly(4-hydroxy-L-proline ester)	PHP
Poly(allylamine)	
Poly(α -[4-aminobutyl]-L-glycolic acid)	PAGA
Poly(D,L-lactic-co-glycolic acid)	PLGA
Poly(N-ethyl-4-vinylpyridinium bromide)	
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(N-2-hydroxypropylmethacrylamide)	pHPMA
Poly(2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	
Galactosylated chitosan	
N-Dodacylated chitosan	
Histone	
Collagen	
Dextran-spermine	D-SPM

Table 19 summarizes delivery methods for a polynucleotide encoding a fusion protein described herein.

5 **Table 19**

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Physical	(<i>e.g.</i> , electroporation, particle gun, Calcium Phosphate transfection)	YES	Transient	NO	Nucleic Acids and Proteins
Viral	Retrovirus	NO	Stable	YES	RNA
	Lentivirus	YES	Stable	YES/NO with modification	RNA
	Adenovirus	YES	Transient	NO	DNA
	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA
	Vaccinia Virus	YES	Very Transient	NO	DNA

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
	Herpes Simplex Virus	YES	Stable	NO	DNA
Non-Viral	Cationic Liposomes	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
	Polymeric Nanoparticles	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
Biological Non-Viral	Attenuated Bacteria	YES	Transient	NO	Nucleic Acids
Delivery Vehicles	Engineered Bacteriophages	YES	Transient	NO	Nucleic Acids
	Mammalian Virus-like Particles	YES	Transient	NO	Nucleic Acids
	Biological liposomes: Erythrocyte Ghosts and Exosomes	YES	Transient	NO	Nucleic Acids

In another aspect, the delivery of base editor system components or nucleic acids encoding such components, for example, a polynucleotide programmable nucleotide binding domain (e.g., Cas9) such as, for example, Cas9 or variants thereof, and a gRNA targeting a nucleic acid sequence of interest, may be accomplished by delivering a ribonucleoprotein (RNP) to cells. In general, a ribonucleoprotein (RNP) is a complex of ribonucleic acid and RNA-binding protein. The RNP comprises a polynucleotide programmable nucleotide binding domain (e.g., Cas9), in complex with the targeting gRNA. RNPs or polynucleotides described herein may be delivered to cells using known methods, such as electroporation, nucleofection, or cationic lipid-mediated methods, for example, as reported by Zuris, J.A. et al., 2015, Nat. Biotechnology, 33(1):73-80, which is incorporated by reference in its entirety. RNPs are advantageous for use in CRISPR base editing systems, particularly for cells that are difficult to transfect, such as primary cells. In addition, RNPs can also alleviate difficulties that may occur with protein expression in cells, especially when eukaryotic promoters, e.g., CMV or EF1A, which may be used in CRISPR plasmids, are not well-expressed. Advantageously, the use of RNPs does not require the delivery of foreign DNA into cells. Moreover, because an

RNP comprising a nucleic acid binding protein and gRNA complex is degraded over time, the use of RNPs has the potential to limit off-target effects. In a manner similar to that for plasmid based techniques, RNPs can be used to deliver binding protein (e.g., Cas9 variants) and to direct homology directed repair (HDR).

5 Nucleic acid molecules encoding a base editor system can be delivered directly to cells (e.g., hematopoietic stem/progenitor cells) as naked DNA or RNA by means of transfection or electroporation, for example, or can be conjugated to molecules (e.g., N-acetylgalactosamine) promoting uptake by the target cells. Vectors encoding base editor systems and/or their components can also be used. In particular embodiments, a
10 polynucleotide, e.g. a mRNA encoding a base editor system or a functional component thereof, may be co-electroporated with one or more guide RNAs as described herein.

Nucleic acid vectors can comprise one or more sequences encoding a domain of a fusion protein described herein. A vector can also encode a protein component of a base editor system operably linked to a nuclear localization signal, nucleolar localization
15 signal, or mitochondrial localization signal. As one example, a vector can include a Cas9 coding sequence that includes one or more nuclear localization sequences (e.g., a nuclear localization sequence from SV40), and one or more deaminases.

The vector can also include any suitable number of regulatory/control elements, e.g., promoters, enhancers, introns, polyadenylation signals, Kozak consensus sequences,
20 or internal ribosome entry sites (IRES). These elements are well known in the art.

Vectors according to this disclosure include recombinant viral vectors. Exemplary viral vectors are set forth herein above. Other viral vectors known in the art can also be used. In addition, viral particles can be used to deliver base editor system components in nucleic acid and/or protein form. For example, "empty" viral particles can be assembled
25 to contain a base editor system or component as cargo. Viral vectors and viral particles can also be engineered to incorporate targeting ligands to alter target tissue specificity.

Vectors described herein may comprise regulatory elements to drive expression of a base editor system or component thereof. Such vectors include adeno-associated viruses with inverted long terminal repeats (AAV ITR). The use of AAV-ITR can be
30 advantageous for eliminating the need for an additional promoter element, which can take up space in the vector. The additional space freed up can be used to drive the expression of additional elements, such as a guide nucleic acid or a selectable marker. ITR activity can be used to reduce potential toxicity due to over expression.

Any suitable promoter can be used to drive expression of a base editor system or component thereof and, where appropriate, the guide nucleic acid. For ubiquitous expression, promoters include CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains. For brain or other CNS cell expression, suitable promoters include: SynapsinI for
5 all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons. For liver cell expression, suitable promoters include the Albumin promoter. For lung cell expression, suitable promoters include SP-B. For endothelial cells, suitable promoters include ICAM. For hematopoietic cell expression suitable promoters include IFNbeta or CD45. For osteoblast expression suitable promoters can
10 include OG-2.

In some embodiments, a base editor system of the present disclosure is of small enough size to allow separate promoters to drive expression of the base editor and a compatible guide nucleic acid within the same nucleic acid molecule. For instance, a vector or viral vector can comprise a first promoter operably linked to a nucleic acid
15 encoding the base editor and a second promoter operably linked to the guide nucleic acid.

The promoter used to drive expression of a guide nucleic acid can include: Pol III promoters, such as U6 or H1 Use of Pol II promoter and intronic cassettes to express gRNA Adeno Associated Virus (AAV).

In particular embodiments, a fusion protein of the invention is encoded by a
20 polynucleotide present in a viral vector (e.g., adeno-associated virus (AAV), AAV3, AAV3b, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh8, AAV10, and variants thereof), or a suitable capsid protein of any viral vector. Thus, in some aspects, the disclosure relates to the viral delivery of a fusion protein. Examples of viral vectors include retroviral vectors (e.g. Maloney murine leukemia virus, MML-V), adenoviral
25 vectors (e.g. AD100), lentiviral vectors (HIV and FIV-based vectors), herpesvirus vectors (e.g. HSV-2).

In some aspects, the methods described herein for editing specific genes in a cell can be used to genetically modify the cell (e.g., a hematopoietic stem/progenitor cell (HSPCs)).
30

Viral Vectors

A base editor described herein can therefore be delivered with viral vectors. In some embodiments, a base editor disclosed herein can be encoded on a nucleic acid that is

contained in a viral vector. In some embodiments, one or more components of the base editor system can be encoded on one or more viral vectors. For example, a base editor and guide nucleic acid can be encoded on a single viral vector. In other embodiments, the base editor and guide nucleic acid are encoded on different viral vectors. In either case, the base editor and guide nucleic acid can each be operably linked to a promoter and terminator. The combination of components encoded on a viral vector can be determined by the cargo size constraints of the chosen viral vector.

The use of RNA or DNA viral based systems for the delivery of a base editor takes advantage of highly evolved processes for targeting a virus to specific cells in culture or in the host and trafficking the viral payload to the nucleus or host cell genome. Viral vectors can be administered directly to cells in culture, patients (in vivo), or they can be used to treat cells in vitro, and the modified cells can optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

Viral vectors can include lentivirus (e.g., HIV and FIV-based vectors), Adenovirus (e.g., AD100), Retrovirus (e.g., Maloney murine leukemia virus, MML-V), herpesvirus vectors (e.g., HSV-2), and Adeno-associated viruses (AAVs), or other plasmid or viral vector types, in particular, using formulations and doses from, for example, U.S. Patent No. 8,454,972 (formulations, doses for adenovirus), U.S. Patent No. 8,404,658 (formulations, doses for AAV) and U.S. Patent No. 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For example, for AAV, the route of administration, formulation and dose can be as in U.S. Patent No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in U.S. Patent No. 8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as in U.S. Patent No. 5,846,946 and as in clinical studies involving plasmids. Doses can be based on or extrapolated to an average 70 kg individual (e.g. a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species.

Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. The viral vectors can be injected into the tissue of interest. For cell-type specific base editing, the expression of the base editor and optional guide nucleic acid can be driven by a cell-type specific promoter.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (See, e.g., Buchscher et al., *J. Virol.* 66:2731-2739 (1992); Johann et al., *J. Virol.* 66:1635-1640 (1992); Sommerfelt et al., *Virol.* 176:58-59 (1990); Wilson et al., *J. Virol.* 63:2374-2378 (1989); Müller et al., *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

Retroviral vectors, especially lentiviral vectors, can require polynucleotide sequences smaller than a given length for efficient integration into a target cell. For example, retroviral vectors of length greater than 9 kb can result in low viral titers compared with those of smaller size. In some aspects, a base editor of the present disclosure is of sufficient size so as to enable efficient packaging and delivery into a target cell via a retroviral vector. In some embodiments, a base editor is of a size so as to allow efficient packing and delivery even when expressed together with a guide nucleic acid and/or other components of a targetable nuclease system.

Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and psi.2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for

packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, Adeno-associated virus ("AAV") vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA can be packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line can also be infected with adenovirus as a helper. The helper virus can promote replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid in some cases is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

In applications where transient expression is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors can also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (See, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). The construction of recombinant AAV vectors is described in a number of publications, including U.S. Patent No. 5,173,414; Tratschin et al., *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski et al., *J. Virol.* 63:03822-3828 (1989).

AAV is a small, single-stranded DNA dependent virus belonging to the parvovirus family. The 4.7 kb wild-type (wt) AAV genome is made up of two genes that encode four replication proteins and three capsid proteins, respectively, and is flanked on either side by 145-bp inverted terminal repeats (ITRs). The virion is composed of three capsid proteins, Vp1, Vp2, and Vp3, produced in a 1:1:10 ratio from the same open reading frame but from differential splicing (Vp1) and alternative translational start sites (Vp2 and Vp3, respectively). Vp3 is the most abundant subunit in the virion and participates in receptor recognition at the cell surface defining the tropism of the virus. A phospholipase

domain, which functions in viral infectivity, has been identified in the unique N terminus of Vp1.

Similar to wt AAV, recombinant AAV (rAAV) utilizes the cis-acting 145-bp ITRs to flank vector transgene cassettes, providing up to 4.5 kb for packaging of foreign DNA.

5 Subsequent to infection, rAAV can express a fusion protein of the invention and persist without integration into the host genome by existing episomally in circular head-to-tail concatemers. Although there are numerous examples of rAAV success using this system, in vitro and in vivo, the limited packaging capacity has limited the use of AAV-mediated gene delivery when the length of the coding sequence of the gene is equal or greater in
10 size than the wt AAV genome.

Viral vectors can be selected based on the application. For example, for in vivo gene delivery, AAV can be advantageous over other viral vectors. In some embodiments, AAV allows low toxicity, which can be due to the purification method not requiring ultra-centrifugation of cell particles that can activate the immune response. In some
15 embodiments, AAV allows low probability of causing insertional mutagenesis because it doesn't integrate into the host genome. Adenoviruses are commonly used as vaccines because of the strong immunogenic response they induce. Packaging capacity of the viral vectors can limit the size of the base editor that can be packaged into the vector.

AAV has a packaging capacity of about 4.5 Kb or 4.75 Kb including two 145 base
20 inverted terminal repeats (ITRs). This means disclosed base editor as well as a promoter and transcription terminator can fit into a single viral vector. Constructs larger than 4.5 or 4.75 Kb can lead to significantly reduced virus production. For example, SpCas9 is quite large, the gene itself is over 4.1 Kb, which makes it difficult for packing into AAV. Therefore, embodiments of the present disclosure include utilizing a disclosed base editor
25 which is shorter in length than conventional base editors. In some examples, the base editors are less than 4 kb. Disclosed base editors can be less than 4.5 kb, 4.4 kb, 4.3 kb, 4.2 kb, 4.1 kb, 4 kb, 3.9 kb, 3.8 kb, 3.7 kb, 3.6 kb, 3.5 kb, 3.4 kb, 3.3 kb, 3.2 kb, 3.1 kb, 3 kb, 2.9 kb, 2.8 kb, 2.7 kb, 2.6 kb, 2.5 kb, 2 kb, or 1.5 kb. In some embodiments, the disclosed base editors are 4.5 kb or less in length.

30 An AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the type of AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue.

AAV8 is useful for delivery to the liver. A tabulation of certain AAV serotypes as to these cells can be found in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)).

In some embodiments, lentiviral vectors are used to transduce a cell of interest with a polynucleotide encoding a base editor system. Lentiviruses are complex
5 retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. The most commonly known lentivirus is the human immunodeficiency virus (HIV), which uses the envelope glycoproteins of other viruses to target a broad range of cell types.

Lentiviruses can be prepared as follows. After cloning pCasES10 (which contains
10 a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) were seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum and without antibiotics. After 20 hours, media is changed to OptiMEM (serum-free) media and transfection was done 4 hours later. Cells are transfected with 10 µg of lentiviral transfer plasmid (pCasES10) and the following packaging plasmids: 5 µg
15 of pMD2.G (VSV-g pseudotype), and 7.5 µg of psPAX2 (gag/pol/rev/tat). Transfection can be done in 4 mL OptiMEM with a cationic lipid delivery agent (50 µl Lipofectamine 2000 and 100 µl Plus reagent). After 6 hours, the media is changed to antibiotic-free DMEM with 10% fetal bovine serum. These methods use serum during cell culture, but serum-free methods are preferred.

20 Lentivirus can be purified as follows. Viral supernatants are harvested after 48 hours. Supernatants are first cleared of debris and filtered through a 0.45 µm low protein binding (PVDF) filter. They are then spun in an ultracentrifuge for 2 hours at 24,000 rpm. Viral pellets are resuspended in 50 µl of DMEM overnight at 4° C. They are then aliquoted and immediately frozen at -80°C.

25 In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated. In another embodiment, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostatin and angiostatin that is contemplated to be delivered via a subretinal injection. In another embodiment, use of self-inactivating
30 lentiviral vectors are contemplated.

The small packaging capacity of AAV vectors makes the delivery of a number of genes that exceed this size and/or the use of large physiological regulatory elements challenging. These challenges can be addressed, for example, by dividing the protein(s)

to be delivered into two or more fragments, wherein the N-terminal fragment is fused to a split intein-N and the C-terminal fragment is fused to a split intein-C. These fragments are then packaged into two or more AAV vectors. As used herein, "intein" refers to a self-splicing protein intron (e.g., peptide) that ligates flanking N-terminal and C-terminal
5 exteins (e.g., fragments to be joined). The use of certain inteins for joining heterologous protein fragments is described, for example, in Wood et al., J. Biol. Chem. 289(21): 14512-9 (2014). For example, when fused to separate protein fragments, the inteins IntN and IntC recognize each other, splice themselves out and simultaneously ligate the flanking N- and C-terminal exteins of the protein fragments to which they were fused,
10 thereby reconstituting a full-length protein from the two protein fragments. Other suitable inteins will be apparent to a person of skill in the art.

A fragment of a fusion protein of the invention can vary in length. In some embodiments, a protein fragment ranges from 2 amino acids to about 1000 amino acids in length. In some embodiments, a protein fragment ranges from about 5 amino acids to
15 about 500 amino acids in length. In some embodiments, a protein fragment ranges from about 20 amino acids to about 200 amino acids in length. In some embodiments, a protein fragment ranges from about 10 amino acids to about 100 amino acids in length. Suitable protein fragments of other lengths will be apparent to a person of skill in the art.

In one embodiment, dual AAV vectors are generated by splitting a large transgene
20 expression cassette in two separate halves (5' and 3' ends, or head and tail), where each half of the cassette is packaged in a single AAV vector (of <5 kb). The re-assembly of the full-length transgene expression cassette is then achieved upon co-infection of the same cell by both dual AAV vectors followed by: (1) homologous recombination (HR) between 5' and 3' genomes (dual AAV overlapping vectors); (2) ITR-mediated tail-to-
25 head concatemerization of 5' and 3' genomes (dual AAV trans-splicing vectors); or (3) a combination of these two mechanisms (dual AAV hybrid vectors). The use of dual AAV vectors in vivo results in the expression of full-length proteins. The use of the dual AAV vector platform represents an efficient and viable gene transfer strategy for transgenes of >4.7 kb in size.

30

Inteins

Inteins (intervening protein) are auto-processing domains found in a variety of diverse organisms, which carry out a process known as protein splicing. Protein splicing

is a multi-step biochemical reaction comprised of both the cleavage and formation of peptide bonds. While the endogenous substrates of protein splicing are proteins found in intein-containing organisms, inteins can also be used to chemically manipulate virtually any polypeptide backbone.

5 In protein splicing, the intein excises itself out of a precursor polypeptide by cleaving two peptide bonds, thereby ligating the flanking extein (external protein) sequences via the formation of a new peptide bond. This rearrangement occurs post-translationally (or possibly co-translationally). Intein-mediated protein splicing occurs spontaneously, requiring only the folding of the intein domain.

10 About 5% of inteins are split inteins, which are transcribed and translated as two separate polypeptides, the N-intein and C-intein, each fused to one extein. Upon translation, the intein fragments spontaneously and non-covalently assemble into the canonical intein structure to carry out protein splicing in trans. The mechanism of protein splicing entails a series of acyl-transfer reactions that result in the cleavage of two peptide
15 bonds at the intein-extein junctions and the formation of a new peptide bond between the N- and C-exteins. This process is initiated by activation of the peptide bond joining the N-extein and the N-terminus of the intein. Virtually all inteins have a cysteine or serine at their N-terminus that attacks the carbonyl carbon of the C-terminal N-extein residue. This N to O/S acyl-shift is facilitated by a conserved threonine and histidine (referred to
20 as the TXXH motif (SEQ ID NO: 288)), along with a commonly found aspartate, which results in the formation of a linear (thio)ester intermediate. Next, this intermediate is subject to trans-(thio)esterification by nucleophilic attack of the first C-extein residue (+1), which is a cysteine, serine, or threonine. The resulting branched (thio)ester intermediate is resolved through a unique transformation: cyclization of the highly
25 conserved C-terminal asparagine of the intein. This process is facilitated by the histidine (found in a highly conserved HNF motif) and the penultimate histidine and may also involve the aspartate. This succinimide formation reaction excises the intein from the reactive complex and leaves behind the exteins attached through a non-peptidic linkage. This structure rapidly rearranges into a stable peptide bond in an intein-independent
30 fashion.

 In some embodiments, a portion or fragment of a nuclease (e.g., Cas9) is fused to an intein. The nuclease can be fused to the N-terminus or the C-terminus of the intein. In some embodiments, a portion or fragment of a fusion protein is fused to an intein and

fused to an AAV capsid protein. The intein, nuclease and capsid protein can be fused together in any arrangement (e.g., nuclease-intein-capsid, intein-nuclease-capsid, capsid-intein-nuclease, etc.). In some embodiments, an N-terminal fragment of a base editor (e.g., ABE, CBE) is fused to a split intein-N and a C-terminal fragment is fused to a split
 5 intein-C. These fragments are then packaged into two or more AAV vectors. In some embodiments, the N-terminus of an intein is fused to the C-terminus of a fusion protein and the C-terminus of the intein is fused to the N-terminus of an AAV capsid protein.

In one embodiment, inteins are utilized to join fragments or portions of a cytidine or adenosine deaminase base editor protein that is grafted onto an AAV capsid protein.
 10 The use of certain inteins for joining heterologous protein fragments is described, for example, in Wood et al., J. Biol. Chem. 289(21); 14512-9 (2014). For example, when fused to separate protein fragments, the inteins IntN and IntC recognize each other, splice themselves out and simultaneously ligate the flanking N- and C-terminal exteins of the protein fragments to which they were fused, thereby reconstituting a full-length protein
 15 from the two protein fragments. Other suitable inteins will be apparent to a person of skill in the art.

In some embodiments, an ABE was split into N- and C- terminal fragments at Ala, Ser, Thr, or Cys residues within selected regions of SpCas9. These regions correspond to loop regions identified by Cas9 crystal structure analysis.

20 The N-terminus of each fragment is fused to an intein-N and the C- terminus of each fragment is fused to an intein C at amino acid positions S303, T310, T313, S355, A456, S460, A463, T466, S469, T472, T474, C574, S577, A589, and S590, which are indicated in capital letters in the sequence below (called the “Cas9 reference sequence”).

```

  1 mdkkysigld igtntsvgwav itdeykvpsk kfkvgntdr hsikknliga llfdsgetae
  25 61 atrlkrtrarr rytrrrknric ylqeifsnem akvddsffhr leesflveed kkherhpifg
  121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlnpdnsd
  181 vdklfiglvq tynqlfeenp inasgvdaka ilsarlkskr rlenliaqlp gekknglfgn
  241 liaislgltp nfksnfdlae daklqlskdt ydddldnlla qigdqyadlf laaknlsdai
  301 llSdilrvnT eiTKaplsas mikrydehhq ditllkalvr qqlpekykei ffdqSkngya
  361 gyidggasqe efykfikpil ekmdgteell vkinredllr kqrtfdngsi phqihlgeih
  30 421 ailrrqedfy pflkdnreki ekiltfripy yvgplArgnS rfAwmTrkSe eTiTpwnfee
  481 vvdkgasags fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkyy tegmrkpafl
  541 sgeqkkaivd llfktnrkvt vkqlkedyfk kieCfdSvei sgvedrfnAS lgtyhdlkvi
  601 ikdkdfldne enedilediv ltltlfedre mieerlktya hlfdckvmkq lkrrrytgwg
  35 661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediqk aqvsgggdsl
  721 behianlags paikkgilqt vkvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
  781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrldsydvdh
  841 ivpqsflkdd sidnkvltrs dknrgksdnv pseevvkkmk nywrqlinak litqrkfdnl
  901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrnm tkydendkli revkvitlks
  40 961 klvsdfrkdf qfykvrein yhhahdayln avvgtalikk ypklesefvy gdykvydvrk
  1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
  1081 atvrkvlsmq qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfidsptva
  1141 ysvlvvakve kgkskkklsv kellgitime rssfeknpid fleakgykev kkdliiklpk

```

1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
 1261 qhkhyldelii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hiftltnlga
 1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqllggd

(SEQ ID NO: 158).

5

Pharmaceutical Compositions

In some aspects, the present invention provides a pharmaceutical composition comprising any of the genetically modified cells, base editors, fusion proteins, or the fusion protein-guide polynucleotide complexes described herein

- 10 The pharmaceutical compositions of the present invention can be prepared in accordance with known techniques. See, e.g., Remington, The Science And Practice of Pharmacy (21st ed. 2005). In general, the cell, or population thereof is admixed with a suitable carrier prior to administration or storage, and in some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable carrier.
- 15 Suitable pharmaceutically acceptable carriers generally comprise inert substances that aid in administering the pharmaceutical composition to a subject, aid in processing the pharmaceutical compositions into deliverable preparations, or aid in storing the pharmaceutical composition prior to administration. Pharmaceutically acceptable carriers can include agents that can stabilize, optimize or otherwise alter the form, consistency,
- 20 viscosity, pH, pharmacokinetics, solubility of the formulation. Such agents include buffering agents, wetting agents, emulsifying agents, diluents, encapsulating agents, and skin penetration enhancers. For example, carriers can include, but are not limited to, saline, buffered saline, dextrose, arginine, sucrose, water, glycerol, ethanol, sorbitol, dextran, sodium carboxymethyl cellulose, and combinations thereof.
- 25 Some nonlimiting examples of materials which can serve as pharmaceutically-acceptable carriers include: (1) sugars, such as lactose, glucose and sucrose; (2) starches, such as corn starch and potato starch; (3) cellulose, and its derivatives, such as sodium carboxymethyl cellulose, methylcellulose, ethyl cellulose, microcrystalline cellulose and cellulose acetate; (4) powdered tragacanth; (5) malt; (6) gelatin; (7) lubricating agents,
- 30 such as magnesium stearate, sodium lauryl sulfate and talc; (8) excipients, such as cocoa butter and suppository waxes; (9) oils, such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; (10) glycols, such as propylene glycol; (11) polyols, such as glycerin, sorbitol, mannitol and polyethylene glycol (PEG); (12) esters, such as ethyl oleate and ethyl laurate; (13) agar; (14) buffering agents, such as
- 35 magnesium hydroxide and aluminum hydroxide; (15) alginic acid; (16) pyrogen-free

water; (17) isotonic saline; (18) Ringer's solution; (19) ethyl alcohol; (20) pH buffered solutions; (21) polyesters, polycarbonates and/or polyanhydrides; (22) bulking agents, such as polypeptides and amino acids (23) serum alcohols, such as ethanol; and (23) other non-toxic compatible substances employed in pharmaceutical formulations. Wetting
5 agents, coloring agents, release agents, coating agents, sweetening agents, flavoring agents, perfuming agents, preservative and antioxidants can also be present in the formulation.

Pharmaceutical compositions can comprise one or more pH buffering compounds to maintain the pH of the formulation at a predetermined level that reflects physiological
10 pH, such as in the range of about 5.0 to about 8.0. The pH buffering compound used in the aqueous liquid formulation can be an amino acid or mixture of amino acids, such as histidine or a mixture of amino acids such as histidine and glycine. Alternatively, the pH buffering compound is preferably an agent which maintains the pH of the formulation at a predetermined level, such as in the range of about 5.0 to about 8.0, and which does not
15 chelate calcium ions. Illustrative examples of such pH buffering compounds include, but are not limited to, imidazole and acetate ions. The pH buffering compound may be present in any amount suitable to maintain the pH of the formulation at a predetermined level.

Pharmaceutical compositions can also contain one or more osmotic modulating
20 agents, i.e., a compound that modulates the osmotic properties (e.g., tonicity, osmolality, and/or osmotic pressure) of the formulation to a level that is acceptable to the blood stream and blood cells of recipient individuals. The osmotic modulating agent can be an agent that does not chelate calcium ions. The osmotic modulating agent can be any compound known or available to those skilled in the art that modulates the osmotic
25 properties of the formulation. One skilled in the art may empirically determine the suitability of a given osmotic modulating agent for use in the inventive formulation. Illustrative examples of suitable types of osmotic modulating agents include, but are not limited to: salts, such as sodium chloride and sodium acetate; sugars, such as sucrose, dextrose, and mannitol; amino acids, such as glycine; and mixtures of one or more of
30 these agents and/or types of agents. The osmotic modulating agent(s) may be present in any concentration sufficient to modulate the osmotic properties of the formulation.

In addition to a modified cell, or population thereof, and a carrier, the pharmaceutical compositions of the present invention can include at least one additional

therapeutic agent useful in the treatment of disease. For example, some embodiments of the pharmaceutical composition described herein further comprises a chemotherapeutic agent. In some embodiments, the pharmaceutical composition further comprises a cytokine peptide or a nucleic acid sequence encoding a cytokine peptide. In some
5 embodiments, the pharmaceutical compositions comprising the cell or population thereof can be administered separately from an additional therapeutic agent.

One consideration concerning the therapeutic use of genetically modified cells of the invention is the quantity of cells necessary to achieve an optimal or satisfactory effect. The quantity of cells to be administered may vary for the subject being treated. In one
10 embodiment, between 10^4 to 10^{10} , between 10^5 to 10^9 , or between 10^6 and 10^8 genetically modified cells of the invention are administered to a human subject. In some embodiments, at least about 1×10^8 , 2×10^8 , 3×10^8 , 4×10^8 , and 5×10^8 genetically modified cells of the invention are administered to a human subject. Determining the precise effective dose may be based on factors for each individual
15 subject, including their size, age, sex, weight, and condition. Dosages can be readily ascertained by those skilled in the art from this disclosure and the knowledge in the art.

The skilled artisan can readily determine the number of cells and amount of optional additives, vehicles, and/or carriers in compositions and to be administered in methods of the invention. Typically, additives (in addition to the cell(s)) are present in an
20 amount of 0.001 to 50 % (weight) solution in phosphate buffered saline, and the active ingredient is present in the order of micrograms to milligrams, such as about 0.0001 to about 5 wt%, preferably about 0.0001 to about 1 wt%, still more preferably about 0.0001 to about 0.05 wt% or about 0.001 to about 20 wt%, preferably about 0.01 to about 10 wt%, and still more preferably about 0.05 to about 5 wt %. Of course, for any
25 composition to be administered to an animal or human, and for any particular method of administration, it is preferred to determine therefore: toxicity, such as by determining the lethal dose (LD) and LD₅₀ in a suitable animal model (e.g., a rodent such as a mouse); and, the dosage of the composition(s), concentration of components therein, and the timing of administering the composition(s), which elicit a suitable response. Such
30 determinations do not require undue experimentation from the knowledge of the skilled artisan, this disclosure and the documents cited herein. And, the time for sequential administrations can be ascertained without undue experimentation.

In some embodiments, the pharmaceutical composition is formulated for delivery to a subject. Suitable routes of administering the pharmaceutical composition described herein include, without limitation: topical, subcutaneous, transdermal, intradermal, intralesional, intraarticular, intraperitoneal, intravesical, transmucosal, gingival, intradental, intracochlear, transtympanic, intraorgan, epidural, intrathecal, intramuscular, intravenous, intravascular, intraosseous, periocular, intratumoral, intracerebral, and intracerebroventricular administration.

In some embodiments, the pharmaceutical composition described herein is administered locally to a diseased site. In some embodiments, the pharmaceutical composition described herein is administered to a subject by injection, by means of a catheter, by means of a suppository, or by means of an implant, the implant being of a porous, non-porous, or gelatinous material, including a membrane, such as a sialastic membrane, or a fiber.

In other embodiments, the pharmaceutical composition described herein is delivered in a controlled release system. In one embodiment, a pump can be used (see, e.g., Langer, 1990, *Science* 249: 1527-1533; Sefton, 1989, *CRC Crit. Ref. Biomed. Eng.* 14:201; Buchwald et al., 1980, *Surgery* 88:507; Saudek et al., 1989, *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used. (See, e.g., *Medical Applications of Controlled Release* (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); *Controlled Drug Bioavailability, Drug Product Design and Performance* (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, *Macromol. Sci. Rev. Macromol. Chem.* 23:61. See also Levy et al., 1985, *Science* 228: 190; During et al., 1989, *Ann. Neurol.* 25:351; Howard et al., 1989, *J. Neurosurg.* 71: 105.) Other controlled release systems are discussed, for example, in Langer, *supra*.

In some embodiments, the pharmaceutical composition is formulated in accordance with routine procedures as a composition adapted for intravenous or subcutaneous administration to a subject, e.g., a human. In some embodiments, pharmaceutical composition for administration by injection are solutions in sterile isotonic use as solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the pharmaceutical is to be administered by infusion,

it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the pharmaceutical composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients can be mixed prior to administration.

5 A pharmaceutical composition for systemic administration can be a liquid, e.g., sterile saline, lactated Ringer's or Hank's solution. In addition, the pharmaceutical composition can be in solid forms and re-dissolved or suspended immediately prior to use. Lyophilized forms are also contemplated. The pharmaceutical composition can be contained within a lipid particle or vesicle, such as a liposome or microcrystal, which is
10 also suitable for parenteral administration. The particles can be of any suitable structure, such as unilamellar or plurilamellar, so long as compositions are contained therein. Compounds can be entrapped in "stabilized plasmid-lipid particles" (SPLP) containing the fusogenic lipid dioleoylphosphatidylethanolamine (DOPE), low levels (5-10 mol%) of cationic lipid, and stabilized by a polyethyleneglycol (PEG) coating (Zhang Y. P. et al.,
15 Gene Ther. 1999, 6: 1438-47). Positively charged lipids such as N-[1-(2,3-dioleoyloxi)propyl]-N,N,N-trimethyl-amoniummethylsulfate, or "DOTAP," are particularly preferred for such particles and vesicles. The preparation of such lipid particles is well known. See, e.g., U.S. Patent Nos. 4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016; and 4,921,757; each of which is incorporated herein by reference.

20 The pharmaceutical composition described herein can be administered or packaged as a unit dose, for example. The term "unit dose" when used in reference to a pharmaceutical composition of the present disclosure refers to physically discrete units suitable as unitary dosage for the subject, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with
25 the required diluent; i.e., carrier, or vehicle.

 Further, the pharmaceutical composition can be provided as a pharmaceutical kit comprising (a) a container containing a compound of the invention in lyophilized form and (b) a second container containing a pharmaceutically acceptable diluent (e.g., sterile used for reconstitution or dilution of the lyophilized compound of the invention.

30 Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In another aspect, an article of manufacture containing materials useful for the treatment of the diseases described above is included. In some embodiments, the article of manufacture comprises a container and a label. Suitable containers include, for example, bottles, vials, syringes, and test tubes. The containers can be formed from a variety of materials such as glass or plastic. In some embodiments, the container holds a composition that is effective for treating a disease described herein and can have a sterile access port. For example, the container can be an intravenous solution bag or a vial having a stopper pierceable by a hypodermic injection needle. The active agent in the composition is a compound of the invention. In some embodiments, the label on or associated with the container indicates that the composition is used for treating the disease of choice. The article of manufacture can further comprise a second container comprising a pharmaceutically-acceptable buffer, such as phosphate-buffered saline, Ringer's solution, or dextrose solution. It can further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for use.

In some embodiments, any of the fusion proteins, gRNAs, and/or complexes described herein are provided as part of a pharmaceutical composition. In some embodiments, the pharmaceutical composition comprises any of the fusion proteins provided herein. In some embodiments, the pharmaceutical composition comprises any of the complexes provided herein. In some embodiments, the pharmaceutical composition comprises a ribonucleoprotein complex comprising an RNA-guided nuclease (e.g., Cas9) that forms a complex with a gRNA and a cationic lipid. In some embodiments pharmaceutical composition comprises a gRNA, a nucleic acid programmable DNA binding protein, a cationic lipid, and a pharmaceutically acceptable excipient. Pharmaceutical compositions can optionally comprise one or more additional therapeutically active substances.

In some embodiments, compositions provided herein are administered to a subject, for example, to a human subject, in order to effect a targeted genomic modification within the subject. In some embodiments, cells are obtained from the subject and contacted with any of the pharmaceutical compositions provided herein. In some embodiments, cells removed from a subject and contacted ex vivo with a pharmaceutical composition are re-introduced into the subject, optionally after the desired genomic modification has been effected or detected in the cells. Methods of delivering

pharmaceutical compositions comprising nucleases are known, and are described, for example, in U.S. Patent Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties. Although the
5 descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals or organisms of all sorts, for example, for veterinary use.

Modification of pharmaceutical compositions suitable for administration to
10 humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, domesticated animals, pets, and
15 commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as chickens, ducks, geese, and/or turkeys.

Formulations of the pharmaceutical compositions described herein can be prepared by any method known or hereafter developed in the art of pharmacology. In
20 general, such preparatory methods include the step of bringing the active ingredient(s) into association with an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit. Pharmaceutical formulations can additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all
25 solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage form desired. Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated in its
30 entirety herein by reference) discloses various excipients used in formulating pharmaceutical compositions and known techniques for the preparation thereof. See also PCT application PCT/US2010/055131 (Publication number WO2011/053982 A8, filed Nov. 2, 2010), incorporated in its entirety herein by reference, for additional suitable

methods, reagents, excipients and solvents for producing pharmaceutical compositions comprising a nuclease.

Except insofar as any conventional excipient medium is incompatible with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition, its use is contemplated to be within the scope of this disclosure.

The compositions, as described above, can be administered in effective amounts. The effective amount will depend upon the mode of administration, the particular condition being treated, and the desired outcome. It may also depend upon the stage of the condition, the age and physical condition of the subject, the nature of concurrent therapy, if any, and like factors well-known to the medical practitioner. For therapeutic applications, it is that amount sufficient to achieve a medically desirable result. In some embodiments, compositions in accordance with the present disclosure can be used for treatment of any of a variety of diseases, disorders, and/or conditions.

Methods of Treatment

Some aspects of the present invention provide methods of treating a subject in need, the method comprising administering to a subject in need an effective therapeutic amount of a pharmaceutical composition as described herein. More specifically, the methods of treatment include administering to a subject in need thereof one or more pharmaceutical compositions comprising one or more cells having at least one edited gene. In other embodiments, the methods of the invention comprise expressing or introducing into a cell a base editor polypeptide and one or more guide RNAs capable of targeting a nucleic acid molecule encoding at least one polypeptide.

Provided herein are methods and compositions involving or containing modified cells, e.g. base edited progenitor or stem cells for engraftment purposes. Base editor systems provided herein may be used to generate modifications in target polynucleotides, e.g., a target gene or a regulatory element thereof, in a progenitor cell or a stem cell. In some embodiments, base editing is performed in a population of progenitor cells or stem cells. In some embodiments, a target gene or a regulatory element thereof may comprise a mutation or a SNP that is associated with a genetic condition, disorder, or disease. In some embodiments, a base editor system provided herein is capable of effecting a single

nucleobase modification that corrects the mutation associated with the genetic condition, disorder, or disease. In some embodiments, a base editor system provided herein is capable of effecting a single nucleobase modification that does not revert the mutation or SNP associated with the genetic condition, disorder, or disease to a wild type nucleobase, but replaces the mutation or SNP with another nucleobase that ameliorates at least one symptom of the genetic condition, disorder, or disease. In some embodiments, a base editor system provided herein is capable of effecting a single nucleobase modification that does not revert the mutation or SNP associated with the genetic condition, disorder, or disease to a wild type, but introduces a nucleobase modification elsewhere in the genome that ameliorates at least one symptom of the genetic condition, disorder, or disease. In one aspect, provided herein are base editor systems and methods of using the same to generate modified cells that allow for long-term engraftment in subjects in need thereof. Genetic manipulation of target cells, e.g. stem cells, may be performed when the stem cells are resting cells (non-cycling), i.e., cells that are not dividing, or when the stem cells are cycling, i.e., cells that are dividing. Additionally, modified stem cells may be introduced into a subject for engrafting into a desired tissue or tissues. When cells of hematopoietic progenitor lineages with a gene modification are introduced into a subject, the cells are required to home into the desired tissue, be stabilized, be able to proliferate, be able to differentiate into the cell lineages and retain the gene modification and remain functionally active for a prolonged period in order for the gene-manipulated hematopoietic progenitor cells to be useful in a gene therapy.

In some embodiments, the purpose of the methods provided herein is to restore the function of a dysfunctional gene via gene editing. In some embodiments, the function of a dysfunctional gene is restored by introducing an intended mutation. In some embodiments, the methods provided herein can be used to disrupt the normal function of a gene product. The nucleobase editing proteins provided herein can be validated for gene editing-based human therapeutics *in vitro*, e.g., by correcting a disease-associated mutation in human cell culture. It will be understood by the skilled artisan that the nucleobase editing proteins provided herein, e.g., the fusion proteins comprising a napDNAbp domain (e.g., Cas12) and a nucleobase editing domain (e.g., an adenosine deaminase domain or a cytidine deaminase domain) can be used to correct any single point A to G or C to T mutation. In the first case, deamination of the mutant A to I

corrects the mutation, and in the latter case, deamination of the A that is base-paired with the mutant T, followed by a round of replication, corrects the mutation.

5 Provided herein is a method of gene editing in hematopoietic stem cells, for example, a adenosine base edited hemoglobin gene or a regulatory region thereof, such as a promoter region, wherein the gene edited hematopoietic stem cells exhibit one or more of: higher editing efficiency; higher fidelity and significantly lower off-target editing events; higher survival of edited cells; higher persistence of edited cells in vitro; higher survival and persistence of edited cells in vivo; higher engraftment potential; higher ability to differentiate into erythropoietic lineage; higher proliferation capability in vitro, 10 higher proliferation capability in vivo, higher expression of HbF; and higher reduction in a defective globin gene expression such as HbS, when compared to previously reported or existing base editing systems. In some embodiments, the improvements in the adenosine base editing system provided herein is associated with at least one of the following advantages: higher editing efficiency; higher fidelity and significantly reduced or lower off-target editing events; higher survival of edited cells; higher persistence of edited cells 15 in vitro; higher survival and persistence of edited cells in vivo; higher engraftment potential; higher ability to differentiate into erythropoietic lineage; higher proliferation capability in vitro, higher proliferation capability in vivo, higher expression of HbF; and higher reduction in a defective globin gene expression such as HbS when compared to previously reported or existing base editing systems. 20

In an aspect, provided herein are methods for engrafting a population of edited cells in a subject in need thereof. The engrafted cells may be autologous or allogeneic cells. In some embodiments, the cells for engraftment are allogeneic cells. In some embodiments, the cells for engraftment are obtained from a donor. In some embodiments, 25 the donor is a healthy donor, or a donor that is histocompatibility matched with the subject. In some embodiments, the cells are isolated from the subject. In some embodiments, a target cell or population of cells for editing are isolated from a subject. In some embodiments, a target cell or population of cells for editing are derived from a donor other than the subject. In some embodiments, the target cell or population of cells 30 are contacted with a base editor system provided herein to generate a desired nucleobase modification. For example, a population of hematopoietic stem cells may be isolated from a patient with sickle cell disease (SCD) and base edited for engraftment. The cell isolated from the patient may comprise a mutation or SNP associated with a genetic disease,

disorder, or condition that may be corrected or ameliorated by base editing. In some embodiments, at least about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the population of cells contacted with the base editor system comprise the desired nucleobase modification. In

5 some embodiments, the percentage of edited cells in the population is at least about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, 110%, 120%, 130%, 140%, 150%, 200%, 250%, 300% or 350% higher than in a population edited with a base editor system comprising a wild type deaminase. In some embodiments, the percentage of edited cells in the population edited

10 with a ABE8 base editor is at least about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, 100%, 110%, 120%, 130%, 140%, 150%, 200%, 250%, 300% or 350% higher than in a population edited with ABE7.10. In some embodiments, the editing efficiency with a ABE8 in a population of cells is at least about 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 9, 10,

15 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, or 60 folds higher than the editing efficiency achieved with a base editor comprising a wild type deaminase. In some embodiments, the editing efficiency with a ABE8 in a population of cells is at least about 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, or 60 folds higher than the editing efficiency achieved with

20 a base editor comprising an ABE7.10. In some embodiments, at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% of the edited population of cells retain viability. In some embodiments, the editing efficiency with a ABE8 in a population of cells is at least about 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, or 60 folds higher than the editing

25 efficiency achieved with a base editor comprising an ABE7.10. In some embodiments, at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% of the edited population of cells retain ability to differentiate. In some embodiments, a hematopoietic cell or a population of hematopoietic cells are contacted with a base editor system, e.g. an ABE8 system, provided herein to effect a nucleobase modification that corrects a

30 mutation or ameliorates at least one detrimental effect of the mutation. For example, a nucleobase modification in the promoter region of HBG1/2 may increase the expression of hemoglobin gamma subunit and/or the expression of HbF protein, thereby compensating for at least one sickle cell disease (SCD) detrimental effect. In some

embodiments, expression of a hemoglobin gamma subunit is increased by at least about 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, or 60 folds in a hematopoietic cell as compared to a control cell without the base edit. In some embodiments, expression of a hemoglobin gamma subunit is increased by at least about 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, or 60 folds in a population of hematopoietic cells as compared to a population of control cells without the base edit. In some embodiments, expression of a hemoglobin gamma subunit is increased by at least about 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, or 60 folds in a subject engrafted with a hematopoietic cell or a population of hematopoietic cells with the base edit as compared to pre-engraftment level of hemoglobin gamma subunit. In some embodiments, sickling is reduced by at least about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95% in a population of base edited hematopoietic cells as compared to a control population of cells, wherein sickling is measured by the percentage of cells exhibiting sickling phenotype. In some embodiments, expression of a HbF protein is increased by at least about 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, or 60 folds in a subject engrafted with a hematopoietic cell or a population of hematopoietic cells with the base edit as compared to pre-engraftment level of HbF in the subject. In some embodiment, the subject is a mammal. In some embodiments, the subject is a human, a non-human primate, a cat, a dog, a pig, a cattle, a house, a camel, a llama, a goat, a sheep, a rodent, a mouse, a rat, a rabbit, a guinea pig, or any other suitable mammal.

Base edited cells for engraftment may be any type of suitable cells. In some embodiments, the base edited cell or cells for engraftment are hematopoietic stem cells, common myeloid progenitors, proerythroblasts, erythroblasts, reticulocytes, or erythrocytes.

In some embodiments, a subject receives a base edited cell or a population of base edited cells, e.g., a population of hematopoietic stem cells, for long term engraftment. In some embodiments, one or more lymphocytic lineage cells are depleted prior to engraftment. The engraftment may be directed to a particular tissue or particular tissues or organs, for example, blood or bone marrow.

In some embodiments, the subject (e.g., human subject) is prepared with a conditioning regimen before transplantation and engraftment of the base edited cells. In some embodiments, the transplant is autologous (e.g., obtained or derived from the subject). In some embodiments, the transplant is allogeneic (e.g., obtained or derived from a donor). Depending on the type of transplant, a myeloablative or non-myeloablative conditioning may be used. Myeloablative conditioning results in ablation of bone marrow and may comprise chemical agents, radiation, or combinations thereof (e.g., cyclophosphamide with total body irradiation). Chemical agents useful for conditioning include without limitation, busulfan, treosulfan, cyclophosphamide, fludarabine, and the like. Non-myeloablative conditioning uses lower doses of chemical agents and radiation. Non-myeloablative conditioning may comprise the use of antibody and antibody drug conjugates that target and eliminate hematopoietic stem cells.

An edited population of cells with high editing efficiency as provided herein may not need to be enriched for engraftment to a subject. In some embodiments, the base edited cell or population of cells retain the ability to differentiate after engraftment. In some embodiments, the base edited cell or population of cells retain the ability to differentiate at least about 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, 8 weeks, 9 weeks, 10 weeks, 11 weeks, 12 weeks, 13 weeks, 14 weeks, 15 weeks, 16 weeks, 17 weeks, 18 weeks, 19 weeks, 20 weeks, 6 months, 7 months, 8 months, 9 months, 10 months, 11 months, 12 months, 1.5 year, 2 years, 2.5 years, 3 years, 3.5 years, or 4 years after engraftment. In some embodiments, the tissue or organ that is engrafted with base edited cells retain the nucleobase modification effected by the base editor system post engraftment. In some embodiments, the tissue or organ that is engrafted with base edited cells retain the nucleobase modification effected by the base editor system after differentiation of the engrafted cell or population of cells. In some embodiments, at least about 1%, 2%, 5%, 7%, 10%, 15%, 20%, 30% or 40% or 50%, 60%, 70%, 80%, 90% or more of cells in the tissue or organ engrafted with the base edited cell or population of cells retain the nucleobase modification after engraftment. In some embodiments, at least about 1%, 2%, 5%, 7%, 10%, 15%, 20%, 30% or 40% or 50%, 60%, 70%, 80%, 90% or more of cells in the tissue or organ engrafted with the base edited cell or population of cells retain the nucleobase modification after differentiation of the engrafted cell or population of cells. In some embodiments, at least about 1%, 2%, 5%, 7%, 10%, 15%, 20%, 30% or 40% or 50%, 60%, 70%, 80%, 90% or more of cells in

the tissue or organ engrafted with the base edited cell or population of cells retain the nucleobase modification at least about 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, 8 weeks, 9 weeks, 10 weeks, 11 weeks, 12 weeks, 13 weeks, 14 weeks, 15 weeks, 16 weeks, 17 weeks, 18 weeks, 19 weeks, 20 weeks, 6 months, 7 months, 8 months, 9 months, 10 months, 11 months, 12 months, 1.5 year, 2 years, 2.5 years, 3 years, 3.5 years, or 4 years after engraftment. In some embodiments, the methods and compositions provided herein lead to at least 10%, at least 20%, at least 30% or 40% or 50%, 60%, 70%, 80%, 90% or more improvement in engraftment efficiency of the base edited hematopoietic cells compared to a previously reported method or compositions using an existing base editing system. In some embodiments, the ABE edited hematopoietic cells generated by the methods described herein are at least 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10 fold, 15-fold, 20-fold, 25-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold or at least 100-fold or more efficient in engrafting than ABE edited hematopoietic cells generated by an ABE7.10.

The compositions, as described above, can be administered in effective amounts. The effective amount will depend upon the mode of administration, the particular condition being treated, and the desired outcome. It may also depend upon the stage of the condition, the age and physical condition of the subject, the nature of concurrent therapy, if any, and like factors well-known to the medical practitioner. For therapeutic applications, it is that amount sufficient to achieve a medically desirable result.

In one embodiment, a subject is administered at least 0.1×10^5 cells, at least 0.5×10^5 cells, at least 1×10^5 cells, at least 5×10^5 cells, at least 1×10^6 cells, at least 0.5×10^7 cells, at least 1×10^7 cells, at least 0.5×10^8 cells, at least 1×10^8 cells, at least 0.5×10^9 cells, at least 1×10^9 cells, at least 2×10^9 cells, at least 3×10^9 cells, at least 4×10^9 cells, at least 5×10^9 cells, or at least 1×10^{10} cells. In particular embodiments, about 1×10^7 cells to about 1×10^9 cells, about 2×10^7 cells to about 0.9×10^9 cells, about 3×10^7 cells to about 0.8×10^9 cells, about 4×10^7 cells to about 0.7×10^9 cells, about 5×10^7 cells to about 0.6×10^9 cells, or about 5×10^7 cells to about 0.5×10^9 cells are administered to the subject.

In one embodiment, a subject is administered at least 0.1×10^4 cells/kg of bodyweight, at least 0.5×10^4 cells/kg of bodyweight, at least 1×10^4 cells/kg of bodyweight, at least 5×10^4 cells/kg of bodyweight, at least 1×10^5 cells/kg of bodyweight, at least 0.5×10^6 cells/kg of bodyweight, at least 1×10^6 cells/kg of bodyweight, at least 0.5×10^7 cells/kg of bodyweight, at least 1×10^7 cells/kg of bodyweight, at least 0.5×10^8

cells/kg of bodyweight, at least 1×10^8 cells/kg of bodyweight, at least 2×10^8 cells/kg of bodyweight, at least 3×10^8 cells/kg of bodyweight, at least 4×10^8 cells/kg of bodyweight, at least 5×10^8 cells/kg of bodyweight, or at least 1×10^9 cells/kg of bodyweight. In particular embodiments, about 1×10^6 cells/kg of bodyweight to about 1×10^8 cells/kg of bodyweight, about 2×10^6 cells/kg of bodyweight to about 0.9×10^8 cells/kg of bodyweight, about 3×10^6 cells/kg of bodyweight to about 0.8×10^8 cells/kg of bodyweight, about 4×10^6 cells/kg of bodyweight to about 0.7×10^8 cells/kg of bodyweight, about 5×10^6 cells/kg of bodyweight to about 0.6×10^8 cells/kg of bodyweight, or about 5×10^6 cells/kg of bodyweight to about 0.5×10^8 cells/kg of bodyweight are administered to the subject.

One of ordinary skill in the art would recognize that multiple administrations of the pharmaceutical compositions contemplated in particular embodiments may be required to affect the desired therapy. For example, a composition may be administered to the subject 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more times over a span of 1 week, 2 weeks, 3 weeks, 1 month, 2 months, 3 months, 4 months, 5 months, 6 months, 1 year, 2 years, 5 years, 10 years, or more. In any of such methods, the methods may comprise administering to the subject an effective amount of an edited cell or a base editor system or polynucleotide encoding such system. In any of such methods, the methods may comprise administering one or more doses of an effective amount of the edited cells per day. In any of such methods, the methods may comprise administering two or more doses of an effective amount of the modified cells per day. In any of such methods, the methods may comprise administering three or more doses of an effective amount of the edited cells per day. In any of such methods, the methods may comprise administering one or more doses of an effective amount of the edited cells per week. In any of such methods, the methods may comprise administering two or more doses of an effective amount of the edited cells per week. In any of such methods, the methods may comprise administering three or more doses of an effective amount of the edited cells per week. In any of such methods, the methods may comprise administering one or more doses of an effective amount of the edited cells per month. In any of such methods, the methods may comprise administering two or more doses of an effective amount of the edited cells per month. In any of such methods, the methods may comprise administering three or more doses of an effective amount of the edited cells per month.

Administration of the pharmaceutical compositions contemplated herein may be carried out using conventional techniques including, but not limited to, infusion, transfusion, or parenterally. In some embodiments, parenteral administration includes infusing or injecting intravascularly, intravenously, intramuscularly, intraarterially, 5 intrathecally, intratumorally, intradermally, intraperitoneally, transtracheally, subcutaneously, subcuticularly, intraarticularly, subcapsularly, subarachnoidly and intrasternally.

In some embodiments, a composition described herein (e.g., edited cell, base editor system) is administered in a dosage that is about 0.5-30 mg per kilogram body 10 weight of the human subject. In another embodiment, the amount of the composition administered is about 0.5-20 mg per kilogram body weight of the human subject. In another embodiment, the amount of the composition administered is about 0.5-10 mg per kilogram body weight of the human subject. In another embodiment, the amount of the composition administered is about 0.04 mg, about 0.08 mg, about 0.16 mg, about 0.32 15 mg, about 0.64 mg, about 1.25 mg, about 1.28 mg, about 1.92 mg, about 2.5 mg, about 3.56 mg, about 3.75 mg, about 5.0 mg, about 7.12 mg, about 7.5 mg, about 10 mg, about 14.24 mg, about 15 mg, about 20 mg, or about 30 mg per kilogram body weight of the human subject. In another embodiment, the amount of the composition administered is about 1.92 mg, about 3.75 mg, about 7.5 mg, about 15.0 mg, or about 20 30.0 mg per kilogram body weight of the human subject and the composition is administered two times a week. In another embodiment, the amount of the composition administered is about 1.28 mg, about 2.56 mg, about 5.0 mg, about 10 mg, or about 20 mg per kilogram body weight of the human subject and the composition is administered two times a week. In another embodiment, the amount of the composition administered is 25 about 1.92 mg, about 3.75 mg, about 7.5 mg, about 15.0 mg, or about 30.0 mg per kilogram body weight of the human subject and the composition is administered once a week. In another embodiment, the amount of the composition administered is about 1.28 mg, about 2.56 mg, about 5.0 mg, about 10 mg, or about 20 mg per kilogram body weight of the human subject and the composition is administered once a week. In another 30 embodiment, the amount of the composition administered is about 1.92 mg, about 3.75 mg, about 7.5 mg, about 15.0 mg, or about 30.0 mg per kilogram body weight of the human subject and the composition is administered once a day three, five or seven times in a seven day period. In another embodiment, the composition is administered

intravenously once a day, seven times in a seven day period. In another embodiment, the amount of the composition administered is about 1.28 mg, about 2.56 mg, about 5.0 mg, about 10 mg, or about 20 mg per kilogram body weight of the human subject and the composition is administered once a day three, five or seven times in a seven day period.

- 5 In another embodiment, the composition is administered intravenously once a day, seven times in a seven day period.

In some embodiments, the composition is administered over a period of 0.25 h (hours), 0.5 h, 1 h, 2 h, 3 h, 4 h, 5 h, 6 h, 7 h, 8 h, 9 h, 10 h, 11 h, or 12 h. In another embodiment, the composition is administered over a period of 0.25-2 h. In another
10 embodiment, the composition is gradually administered over a period of 1 h. In another embodiment, the composition is gradually administered over a period of 2 h.

In one embodiment, the invention provides a method of monitoring treatment progress is provided. The method includes the step of determining a level of diagnostic marker (Marker) (e.g., SNP associated with the disease or condition) or diagnostic
15 measurement (e.g., screen, assay) in a subject suffering from or susceptible to a disease, disorder, or symptoms thereof in which the subject has been administered a therapeutic amount of a composition herein sufficient to treat the disease or symptoms thereof. The level of Marker determined in the method can be compared to known levels of Marker in either healthy normal controls or in other afflicted patients to establish the subject's
20 disease status. In preferred embodiments, a second level of Marker in the subject is determined at a time point later than the determination of the first level, and the two levels are compared to monitor the course of disease or the efficacy of the therapy. In certain preferred embodiments, a pre-treatment level of Marker in the subject is determined prior to beginning treatment according to this invention; this pre-treatment
25 level of Marker can then be compared to the level of Marker in the subject after the treatment commences, to determine the efficacy of the treatment.

In some embodiments, cells are obtained from the subject and contacted with a pharmaceutical composition as provided herein. In some embodiments, cells removed from a subject and contacted ex vivo with a pharmaceutical composition are re-introduced
30 into the subject, optionally after the desired genomic modification has been effected or detected in the cells. Methods of delivering pharmaceutical compositions comprising nucleases are described, for example, in U.S. Patent Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219;

and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties. Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions that are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally
5 suitable for administration to animals or organisms of all sorts, for example, for veterinary use.

The response in individual subjects can be characterized as a complete response, a partial response, or stable disease. In some embodiments, the response is a partial response (PR). In some embodiments, the response is a complete response (CR). In some
10 embodiments, the response results in progression-free survival of the subject (e.g., stable disease). In some embodiments, the treatment results in an increased survival time of the human subject as compared to the expected survival time of the human subject if the human subject was not treated with the compound. In some embodiments, the human subject to be treated with the described methods is a child (e.g., 0-18 years of age). In
15 other embodiments, the human subject to be treated with the described methods is an adult (e.g., 18+ years of age).

Such fusion proteins can be used for targeted editing of DNA *in vitro*, e.g., for the generation of mutant cells or animals; for the introduction of targeted mutations, e.g., for the correction of genetic defects in cells *ex vivo*, e.g., in cells obtained from a subject that
20 are subsequently re-introduced into the same or another subject; and for the introduction of targeted mutations *in vivo*, e.g., the correction of genetic defects or the introduction of deactivating mutations in disease-associated genes in a G to A, or a T to C to mutation can be treated using the nucleobase editors provided herein.

Kits

25 The invention provides kits for the treatment of a hemoglobinopathy (e.g. sickle cell disease) in a subject. In some embodiments, the kit further includes a base editor system or a polynucleotide encoding a base editor system, wherein the base editor polypeptide system a nucleic acid programmable DNA binding protein (napDNAbp), a deaminase, and a guide RNA. In some embodiments, the napDNAbp is Cas9 or Cas12.
30 In some embodiments, the polynucleotide encoding the base editor is a mRNA sequence. In some embodiments, the deaminase is a cytidine deaminase or an adenosine deaminase. In some embodiments, the kit comprises an edited cell and instructions regarding the use of such cell.

The kits may further comprise written instructions for using the base editor system and/or edited cell. In other embodiments, the instructions include at least one of the following: precautions; warnings; clinical studies; and/or references. The instructions may be printed directly on the container (when present), or as a label applied to the container, or as a separate sheet, pamphlet, card, or folder supplied in or with the container. In a further embodiment, a kit can comprise instructions in the form of a label or separate insert (package insert) for suitable operational parameters. In yet another embodiment, the kit can comprise one or more containers with appropriate positive and negative controls or control samples, to be used as standard(s) for detection, calibration, or normalization. The kit can further comprise a second container comprising a pharmaceutically-acceptable buffer, such as (sterile) phosphate-buffered saline, Ringer's solution, or dextrose solution. It can further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for use.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are well within the purview of the skilled artisan. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook, 1989); "Oligonucleotide Synthesis" (Gait, 1984); "Animal Cell Culture" (Freshney, 1987); "Methods in Enzymology" "Handbook of Experimental Immunology" (Weir, 1996); "Gene Transfer Vectors for Mammalian Cells" (Miller and Calos, 1987); "Current Protocols in Molecular Biology" (Ausubel, 1987); "PCR: The Polymerase Chain Reaction", (Mullis, 1994); "Current Protocols in Immunology" (Coligan, 1991). These techniques are applicable to the production of the polynucleotides and polypeptides of the invention, and, as such, may be considered in making and practicing the invention. Particularly useful techniques for particular embodiments will be discussed in the sections that follow.

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the assay, screening, and therapeutic methods of the invention, and are not intended to limit the scope of what the inventors regard as their invention.

EXAMPLES

Example 1: NOD.Cg-Kit^{W-41J} Tyr⁺ Prkdc^{scid} Il2rg^{tm1Wjl}/ThomJ (NBSGW) Mouse Engraftment Protocol

This example describes an engraftment protocol using human cells engrafted into a NBSGW mouse model. In particular, female NOD.Cg-Kit^{W-41J} Tyr⁺ Prkdc^{scid} Il2rg^{tm1Wjl}/ThomJ (NBSGW) mice (Stock 026622, Jackson Laboratories) at the age of 6–7 weeks were transplanted with human CD34⁺ hematopoietic stem cells or progenitor cells (HSPCs) via the tail vein at approximately 1 million (1x10⁶) viable cells per mouse. NBSGW mice support engraftment of human hematopoietic stem cells without irradiation and are suitable for xenotransplantation use. Xenograft NBSGW mice exhibit enhanced human hematopoietic chimerism in peripheral blood, bone marrow, and spleen when compared to non-irradiated engrafted NSG mice, with a level of chimerism similar to that seen in irradiated NSG (irrNSG) mice.

Human chimerism in bone marrow was evaluated by flow cytometry at weeks 8 and 16 or at 18 weeks post-transplantation. A fraction of cells obtained from bone marrow were centrifuged and stored at –20°C for DNA extraction and deep sequencing. For flow cytometry, cells were first incubated with anti-Fc blocking antibodies (catalog no. 422302 and 101320; BioLegend) for 15 minutes, followed by incubation with anti-human CD45 (hCD45) and anti-mouse CD45 (mCD45) antibodies. The percentage human chimerism in bone marrow was defined as $[\text{hCD45}^+ / (\text{hCD45}^+ + \text{mCD45}^+)] * 100$.

Example 2: Base editor gene editing in human CD34⁺ cells

In this example, ABEs (delivered as mRNA) were transfected into human CD34⁺ cells, followed by measurement of cell viability and editing efficiency. Transfected cells were cryopreserved for use in a mouse engraftment study, as described in Example 1.

FIG. 1 represents an exemplary procedure to generate a mammalian vector containing a polynucleotide encoding an adenosine deaminase base editor (ABE), e.g., ABE8.8, for electroporation. **FIG. 2** provides an overview of a workflow for electroporating CD34⁺ cells and preparing them for mouse engraftment. In this example, cells were edited to induce a A>G nucleobase change in the promoter of HBG1/2 at position -198, which led to increased expression of HbF, gamma globin gene product.

Electroporation Cell Preparation and Procedure

Cell Thaw

To prepare cells for electroporation, human CD34⁺ cells were isolated from peripheral blood (PB) of healthy G-CSF/Plerixafor mobilized human donors and cryopreserved. On the day of electroporation, for each donor, a vial containing 11 x 10⁶ cells/mL was removed from liquid nitrogen and placed on dry ice. The vial was then placed on a ThawStar Cell Thawing Unit and was removed from the unit immediately following cell thawing. A suspension of the cells (1 mL) was admixed with X-Vivo 10 cell medium w/o Gentamicin or Phenol Red, 4°C, (Lonza Pharma and Research, Basel, Switzerland), and the cells were then transferred into a conical tube (15 mL), (8 mL cell medium per 1 mL cell suspension). The cell suspension was centrifuged gently in the 15 mL conical tube at 300 x g for 10 minutes at ambient temperature. The medium was aspirated from the centrifuged cells, and the centrifugation process was repeated for an additional one minute. The above process was repeated for each frozen cell aliquot. The cell pellet was next resuspended in prewarmed (37°C) culture medium (Lonza X-Vivo 10 medium containing 1X glutamax, and 100ng/mL TPO, 100ng/mL SCF and 100ng/mL Flt-3L at an approximate cell concentration of 1 x 10⁶ cells/mL. If multiple frozen aliquots were processed as above, the centrifuged cell pellets were pooled in the culture medium. The resuspended cells were transferred to a non-tissue culture treated T-75 flask at an approximate cell concentration of 1 x 10⁶ cells/mL. This process was repeated for multiple vials from each donor and the cells corresponding to each donor were respectively pooled and resuspended in 37°C culture medium.

Electroporation procedure

The cells (or pooled cells) resuspended in cell medium and the mRNA for electroporation were maintained on ice. Electroporation was carried out using a MaxCyte Flow Electroporation instrument (Gaithersburg, MD). The cell cassettes for use in the MaxCyte instrument were also prechilled to -20°C. Immediately prior the electroporation, the cells were counted; each culture flask (or cell culture plate) was rinsed with the 37°C medium described above. By way of example, the results of the cell counts from different donors were as follows:

DONOR #1, D328644-1

Count: 1.3e6 viable cells/mL (1.3 x 10⁶ viable cells/mL)

95.8% viability

Volume = 160mL

Total Cells = 208e6 (2.08×10^8 viable cells/mL)

5 **DONOR #2, D327579-1**

Count: 1.34e6 viable cells/mL (1.34×10^6 viable cells/mL)

95.2% viability

Volume = 160mL

Total Cells = 214.4e6 (2.144×10^8 viable cells/mL)

10

For the MaxCyte OC 400 electroporation cartridge, which is used for a smaller volume of cells (400 μ L), cells were electroporated at a cell concentration of 100×10^6 cells/mL. To prepare the cells for electroporation, the cells were centrifuged at 300 x g for 10 minutes at 4°C, the supernatant was aspirated, the cells were suspended in cell medium and again centrifuged at 300 x g for an additional minute at 4°C. Following this centrifugation, the remaining supernatant was removed, and the cells were resuspended in 4°C MaxCyte EP-Buffer containing 0.1% HSA (10 mL) and counted (e.g., on an NC-200 cell counter device (ChemoMetec A/S, Allerød, Denmark)). The results of the cell counts were as follows:

20

DONOR #1, D328644-1

Count #1: 1.52e6 viable cells/mL (1.52×10^6 viable cells/mL)

98.7% viability

Volume = 150mL

25 Count #2: 1.49e6 viable cells/mL (1.49×10^6 viable cells/mL)

97.5% viability

Volume = 150mL

Total Cells (Ave) = 225.75e6 ($\sim 2.26 \times 10^8$ viable cells/mL)

30 **DONOR #2, D327579-1**

Count #1: 1.42e6 viable cells/mL (1.42×10^6 viable cells/mL)

96.6% viability

Volume = 150mL

Count #2: 1.45e6 viable cells/mL (1.45×10^6 viable cells/mL)

98.2% viability

Volume = 150mL

Total Cells (Ave) = 215.25e6 (~2.15 x 10e8 viable cells/mL)

5 For each electroporation, the cell suspension was centrifuged at 300 x g for 10 minutes at 4°C. The supernatant was aspirated; the cells were resuspended and again centrifuged at 300 x g for 1 minute at 4°C. The supernatant was removed, and the cells were resuspended in 4°C MaxCyte EP-Buffer based on the below **Table 20**, which provides electroporation compositions (containing 4°C MaxCyte EP-Buffer) used to
10 transfect cells. In **Table 20**, the *HBG1/2* guide RNA (gRNA), i.e., “g1,” nucleotide sequence is as follows:

5’-

csususGACCAAUAGCCUUGACAGUUUUAGAGCUAGAAAUAGCAAGUUA
10 UAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU

15 su-3’ (SEQ ID NO: 129). In the g1 gRNA sequence, A, G, U, C represent RNA nucleotides; a, g, u, c represent 2’-O-Methyl nucleotides; and s represents Phosphorothioate.

Aliquots of the resuspended cells were used to prepare the electroporation
20 compositions (groups) presented in **Table 20**. Fifteen minutes prior to electroporation, the base editor-encoding mRNA (e.g., ABE8.8 or ABE7.10) and the gRNA were combined together and kept on ice. The cell suspension that was kept on ice was added to the gRNA/mRNA combination. The mixture was pipetted (e.g., 3 times), and the cell suspension was transferred into the MaxCyte OC-400 cartridge, which was prechilled in
25 beads that were stored at -20°C. Immediately thereafter, the electroporation process was started using the MaxCyte GT system. Following electroporation, the cartridge was placed in a sterile BSC and the cell suspension was removed without touching the sides of the cartridge or rinsing the cartridge with medium or buffer. A suspension of the electroporated cells was placed in the middle of a 6-well, untreated tissue culture plate,
30 and the plate was placed in a 37°C / 5% CO₂ incubator for 20 minutes. Such an incubation is advantageous for achieving a higher level of base editing in the electroporated cells without adversely impacting cell viability. Ten (10) mL of prewarmed 37°C culture medium (X-vivo medium containing 1X glutamax, and

100ng/mL TPO, 100ng/mL SCF, and 100ng/mL Flt-3L) was added to the cell suspension, and the cells were cultured for 72 hours. Thereafter, 100,000 cells were collected at 24, 48, and 72 hours post electroporation for Next Genome Sequencing (NGS) analysis. The cell suspension was counted at 24, 48, and 72 hours using the NC-200 cell counter device
5 to determine the number of live cells and cell viability. A majority of the cells were cryopreserved at 48 hours post electroporation.

Table 20

Gp	Donor ID #	MaxCyte Program	Editor	Guide	Cells	mRNA $\mu\text{g}/\mu\text{L}$	sgRNA $\mu\text{g}/\mu\text{L}$	mRNA μg	sgRNA μg	mRNA μL	sgRNA μL	Additional Maxcyte Buffer μL
1	D328644-1	N/A	N/A	N/A	40	N/A	N/A	N/A	N/A	N/A	N/A	80
2		HSC-3	ABE8.8 (R34)	g1	40	2.5	12	35.2	51.84	14.08	4.32	61.6
3			ABE8.8 (R34)	g1	40	2.5	12	12	17.6	4.8	1.47	73.73
4			ABE7.10 (R35)	g1	40	2	12	35.2	51.84	17.6	4.32	58.08
5	D327579-1	N/A	N/A	N/A	40	N/A	N/A	N/A	N/A	N/A	N/A	80
6		HSC-3	ABE8.8 (R34)	g1	40	2.5	12	35.2	51.84	14.08	4.32	61.6
7			ABE8.8 (R34)	g1	40	2.5	12	12	17.6	4.8	1.47	73.73
8			ABE7.10 (R35)	g1	40	2	12	35.2	51.84	17.6	4.32	58.08

Exemplary cell counts are presented in **Table 21**:

Table 21: Exemplary cell counts.

Sample Number	Timepoint Post Dose	Live Cells (e5/mL)	Viability (%)	Volume (mL)	Total Cells (e6/mL)
1	24 Hours	9.54	84.9	40	38.16
2	24 Hours	7.9	85.7	40	31.6
3	24 Hours	7.68	82.8	40	30.72
4	24 Hours	9.44	90.8	40	37.76
5	24 Hours	9.48	93.5	40	37.92
6	24 Hours	8.38	84.6	40	33.52
7	24 Hours	8.46	90.2	40	33.84
8	24 Hours	8.78	90.2	40	35.12
1	48 Hours	9.91	95.2	39	38.6
2	48 Hours	8.2	94.7	39	32.0
3	48 Hours	8.14	93.6	39	31.7
4	48 Hours	9.45	92.8	39	36.9
5	48 Hours	9.47	95.4	39	36.9
6	48 Hours	9.12	95.5	39	35.6
7	48 Hours	8.97	91.8	39	35.0
8	48 Hours	9.23	96.3	39	36.0
1	Pre-Dose (30 min Post Cryo)	32.5	98.5	10	32.5
2	Pre-Dose (30 min Post Cryo)	31.6	96.1	10	31.6
3	Pre-Dose (30 min Post Cryo)	31.8	97.2	10	31.8
4	Pre-Dose (30 min Post Cryo)	32.4	95.8	10	32.4
5	Pre-Dose (30 min Post Cryo)	30.9	95.4	10	30.9
6	Pre-Dose (30 min Post Cryo)	29.8	93.2	10	29.8
7	Pre-Dose (30 min Post Cryo)	27.6	92.7	10	27.6
8	Pre-Dose (30 min Post Cryo)	32.2	92.6	10	32.2

FIG. 3A presents editing efficiency measured in donor cells edited using ABE7.10 and ABE 8.8 base editor systems, at the indicated RNA concentrations in the electroporation reactions. As demonstrated in **FIG. 3A**, ABE 8.8 at 20 nM and 50 nM showed high efficiency of editing in both donors. At 48 hours post electroporation, all cells showed viabilities of greater than 90% (**FIG. 3B**).

10 Example 3. Engraftment efficiency of human ABE edited cells in mice

This example provides representative results of engrafting ABE edited human CD34⁺ cells from two donors into Female NOD.Cg-Kit^{W-4J} Tyr⁺ Prkdc^{scid}

Il2rg^{tm1Wjl}/ThomJ mice (NBSGW) mice via tail vein injection. The same engraftment protocol was used as that described in Example 1. The percentage of human chimerism (i.e., percentage of donor human BM cells engrafted in bone marrow ("engraftment")) was measured at 8 weeks and 16 weeks post-injection for the donor 1 cells (**FIG. 4A**) and the donor 2 cells (**FIG. 4B**). High percentages of engraftment of the human cells (edited and unedited) were observed and edited cells displayed similar levels of chimerism compared to the unedited controls. This indicated that the edited donor cells were highly viable *in vivo* over prolonged periods. **FIG. 5Ai and FIG. 5Aii** demonstrate proliferation of ABE edited cells *in vivo* during a time course study by demonstrating an increase over time of the prevalence of A → G edited cells in the bone marrow. At the time of injection (In), both the doses of ABE 8.8 resulted in a higher percentage of edited cells in mouse bone marrow (BM) than ABE 7.10. When examined at 8 and 16 weeks post-injection, all of the cells proliferated with high efficiency. Cells electroporated with ABE 8.8 using 50 nM RNA had slightly higher engraftment efficiencies than cells electroporated with ABE 8.8 using 20 nM RNA, and both of the ABE 8.8 transfected cell populations showed higher engraftment efficiencies at all time points evaluated relative to ABE 7.10 transduced cells.

At 16 and 18 weeks, ABE 8.8 (50 and 20 nM doses) edited cells showed higher editing efficiencies relative to ABE 7.10 edited cells (**FIG. 5B and FIG. 5C**). Cells were sorted via flow cytometry for expression of LIN-hCD34⁺ and GlyA⁺.

GlyA⁺ cells were sorted and globin levels were measured by Ultra-High Performance Liquid Chromatography (UHPLC) after 16- or 18-weeks post-transplantation. **FIGs. 5D and 5E** show expression levels of gamma globin at 16 and 18 weeks, respectively. In **FIGs. 5D and 5E**, the expression levels are indicated as percent total beta-like proteins. High expression levels of gamma globin was noted at 16 or 18 weeks post-dose in both donors. Cells transcribed with ABE 8.8 RNA used in an amount of 50 nM showed higher levels of expression than cells transcribed with ABE 7.10 RNA used in an amount of 50 nM, indicating that the ABE 8.8 editing system was a significant improvement over the ABE 7.10 editing system.

To further assess long term engraftment potential and HbF generation capacity of cells containing an *HBG1/2* gene promoter edited using the ABE8.8 editing system, hCD34⁺ cells from a healthy individual were edited using the ABE8.8 editing system. The cells were then introduced into NOD.Cg-Kit^{W-11J} Tyr⁺ Prkdc^{scid} Il2rg^{tm1Wjl}/ThomJ

(NBSGW) mouse bone marrow for engraftment and measured after 16 weeks. As shown in FIG. 6A, engraftment of unedited and base-edited hCD34⁺ cells was assessed. The similar engraftment efficiencies of the edited and unedited cells indicated that there was minimal to no toxicity as a result of editing of the *HBG1/2* gene promoter. This observation was consistent with the observations described for the experiments discussed above. The ABE8.8 editing system showed high base editing efficiency (FIG. 6B). Over 80% of CD34⁺ cells had A>G edits. Cells containing an edited *HBG1/2* promoter region showed high expression levels of gamma globin protein (FIG. 6C), expressed as a percentage of gamma globin protein relative to total of gamma and beta hemoglobin proteins.

Example 4. Differentiation and persistence efficiency of human ABE edited cells obtained from a sickle cell disease (SCD) patient tested *in vitro*:

In this example, CD34⁺ cells obtained from a sickle cell disease patient were transfected with ABE8.8 mRNA and sgRNA by electroporation. Time course studies of the cells grown *in vitro* demonstrated viability and proliferation of the cells, as well as high base editing efficiencies (FIGs. 7A, and 7B). Of note, 16.5% of editing was observed at 48 hr post differentiation, and at 14 days post differentiation, an impressive level of 89.2% edited cells was observed (shown in FIG. 7A). The breakdown of by-stander editing was measured, as shown in FIG. 7B, and revealed a high degree of editing specificity.

Edited sickle cell disease (SCD) CD34⁺ cells were differentiated to erythroid cells, and the levels of globin in the cells were analyzed on day 18 post differentiation by UHPLC (FIGs. 8A and 8B). A 63.2% γ globin level was detected and the S globin level was reduced from 86% to 32.9% (FIGs. 8C and 8D).

Example 5. Transplantation study using different doses of ABE 8.8-encoding mRNA

A dose titration study was performed using mRNA encoding ABE 8.8 and *HBG1/2* gRNA, which were introduced into human hematopoietic cells, e.g., CD34⁺ cells, (“donor cells”), by electroporation (EP). In particular, base editing of the *HBG1/2* gene promoter in the cells was evaluated following electroporation of Granulocyte Colony Stimulating Factor (GCSF) + Plerixafor-mobilized (“seed mobilized”) peripheral blood (PB) CD34⁺ human stem progenitor cells (HSPCs) with different concentrations

(doses) of ABE8.8 mRNA and *HBG1/2* gRNA. Increasing concentrations of ABE8.8 mRNA (1 nM to 30 nM) were used with a fixed concentration (3000 nM) of target guide RNA (*HBG1/2a* gRNA). In the study, two ABE8.8 materials, research scale (large scale) ABE8.8-encoding mRNA (Lot R34) and ABE8.8-encoding mRNA prepared by a contract research organization (CRO), (MRNA288; TriLink) were compared. **FIG. 9A** provides the experimental design of the study. Briefly, the seed mobilized CD34⁺ HSPCs (10⁸ cells/mL) were electroporated (EP) to introduce the ABE8.8 and the *HBG1/2* gRNA into the cells, and cell viability and NGS analyses were performed at 48 hours post EP. Base editing of the *HBG1/2* gene promoter using ABE8.8 and *HBG1/2a* gRNA was assessed in the cells. The treatment conditions of the cells involved either the use of 1, 3, 10, or 30 nM of mRNA (MRNA288) + 3000 nM gRNA (pilot grade) or the use of 10 nM mRNA (MRNA288 versus Lot R34) + 3000 nM gRNA (pilot versus R&D grade). Pilot grade gRNA refers to material that may be considered to closely approximate “Good Manufacturing Practice” (GMP) grade material. The gRNA (*HBG1/2*), or “g1,” is as described in Example 2 *supra*. “MRNA407” is an alternative name for “Lot R34” material.

The study also involved the introduction, e.g., via intravenous tail vein injection, (also referred to as “transplantation”) of the donor, base edited human CD34⁺ HSPCs into two different mouse models, i.e., a non-irradiated NBSGW mouse model and an irradiated NSG (irrNSG) mouse model. The NBSGW mouse model provides an *in vivo* animal system that allows for a high percentage of cell engraftment following transplant and does not result in anemic animals. The NBSGW and irrNSG mouse models were used to evaluate the engraftment capability of the *HBG1/2* gene promoter-base edited human stem cells (HSCs) as described compared to that of unedited (control) cells.

The non-irradiated NBSGW mouse model was used to further evaluate the base editing levels achieved in CD34⁺ HSPCs (and progenitors of these cells) following a dose titration of ABE8.8 mRNA and to determine whether the mRNA (MRNA288) material performed similarly to the mRNA material (Lot R34) over time post-engraftment. The irradiated NSG (irrNSG) mouse model was used to evaluate base the fitness of base-edited HSCs. In addition, the two mouse models were used to evaluate whether human HSCs had the ability to differentiate into various hematopoietic lineages and retain long-term base editing *in vivo*. In particular, the NBSGW mouse model was used to evaluate gamma globin (γ globin) protein induction in NBSGW bone marrow-derived human

erythroid cells via UHPLC analysis. The irrNSG mouse model was used to evaluate multi-lineage hematopoietic reconstitution *in vivo*. As illustrated in **FIG. 9A**, bone marrow (BM) analysis was conducted at 8 weeks following IV injection of cells, and BM and erythroid phenotyping and gamma globin (γ globin) analyses were performed at 16 weeks (long term engraftment).

The experimental regimens and materials used in the studies are summarized in **Table 22** below.

Table 22

Group	Treatment	Materials	Dose	Mouse Strain	Animal Number	Dosing Route	Number of Cells Per Transplant	Scheduled Takedown (Week)
1	Unedited	-	-	NBSGW	8	IV	1 x 10 ⁶	8 weeks (n=3 mice) 16 Weeks (n=5 mice)
				irrNSG	8			
2	Edited	MRNA 288; Pilot HBG1/2a gRNA	1 nM ABE8.8 mRNA + 3000 nM gRNA	NBSGW	8			
				irrNSG	8			
3	Edited	MRNA 288; Pilot HBG1/2a gRNA	3 nM ABE8.8 mRNA + 3000 nM gRNA	NBSGW	8			
				irrNSG	8			
4	Edited	MRNA 288; Pilot HBG1/2a gRNA	10 nM ABE8.8 mRNA + 3000 nM gRNA	NBSGW	8			
				irrNSG	8			
5	Edited	MRNA 288; Pilot HBG1/2a gRNA	30 nM ABE8.8 mRNA + 3000 nM gRNA	NBSGW	8			
				irrNSG	8			
6	Edited	Lot R34 (MRNA 407); Pilot HBG1/2a gRNA	10 nM ABE8.8 mRNA + 3000 nM gRNA	NBSGW	8			
				irrNSG	8			
7	Edited			NBSGW	8			

Group	Treatment	Materials	Dose	Mouse Strain	Animal Number	Dosing Route	Number of Cells Per Transplant	Scheduled Takedown (Week)
		Lot R34; R&D grade HBG1/2a gRNA	10 nM ABE8.8 mRNA + 3000 nM gRNA	irrNSG	8			

In brief, for the study and as reflected in Table 22, G+P mobilized CD34+ cells (HemaCare) were thawed and cultured in 2+2 day culture in X-Vivo10 culture medium plus cytokines (Lonza Pharma and Biotech, Basel, Switzerland). Cells were base edited (Treatment – Edited) using a MaxCyte ATx Flow Electroporation instrument (HSC-3 Program; OC-400 processing component). For the NBSGW mouse model, a total of 56 NBSGW mice were studied. At 8 weeks (8 week takedown) and 16 weeks (16 week takedown), the bone marrow (BM) of 21 mice and 35 mice, respectively, was analyzed by flow cytometry and NGS analysis (n = 3-4 per group for cell sorting). For the irrNSG mouse model, a total of 56 irrNSG mice were studied. At 8 weeks (8 week takedown) and 16 weeks (16 week takedown), the bone marrow (BM) cells of 21 mice and 35 mice, respectively, were analyzed by flow cytometry and NGS analysis (n = 3-4 per group for cell sorting).

FIGS. 9B and 9C show long term (16 weeks) engraftment and *HBG1/2* gene promoter base editing retention in NBSGW mice. The results demonstrate that similar human cell chimerism (>90%) was achieved in bone marrow (BM) samples obtained from *HBG1/2* promoter-base edited and unedited mouse treatment groups at 16 weeks post-transplantation. In addition, *HBG1/2* gene promoter base editing reached >88% in bulk BM cells with increasing ABE8.8 mRNA dose. **FIGS. 10A-10D** demonstrate that *HBG1/2* gene promoter-base edited human stem cells (HSCs) displayed long term (16 weeks), multi-lineage hematopoietic reconstitution in NBSGW mice (the NBSGW mouse model). Flow cytometry analyses were performed to identify phenotypic markers on human progenitor stem cells (HSPCs), human erythroid cells, human myeloid cells and human lymphoid cells (% cells of the different types) as shown. **FIG. 11** demonstrates long term human hematopoietic, multi-lineage reconstitution in NBSGW mice at 16 weeks. The percent base editing of the various human hematopoietic cell subpopulations analyzed and the amounts of electroporated ABE mRNA and gRNA are shown. The

amount of *HBG1/2* gene promoter base editing achieved was shown to be similar in sorted human HSPCs (Lin-CD34+), human myeloid (CD15+), human lymphoid (CD19+) and human erythroid (GlyA+) cells at 16 weeks post-transplantation. The results shown in **FIGS. 12A and 12B** demonstrate that *HBG1/2* gene promoter base editing was maintained long term (16 weeks) post-engraftment with elevated γ -globin levels in NBSGW mice. As shown in **FIGS. 12A and B**, *HBG1/2* gene promoter A-to-G base editing reached >88% in bulk bone marrow cell samples at 16 weeks post-transplantation. In addition, >53% of gamma globin (γ -globin) protein levels were expressed in sorted, base edited, bone marrow-derived human erythroid cells compared to unedited BM-derived erythroid cells (<0.3%). The standard-error-of-the-mean (SEM) ranged from 0.001 to 0.022 in these experiments.

The use of unedited and base-edited donor cells for transplant and engraftment was compared in irradiated NSG (irrNSG) mice. Compared with unedited cells, edited cells transfected with ABE8.8 mRNA and *HBG1/2* gRNA by electroporation retained base editing function over time, e.g., for at least 16 weeks. The base-edited donor cells were robust and functional and successfully engrafted *in vivo* in two mouse models, namely, the NBSGW and irrNSG mouse models as described herein. As demonstrated in **FIGS. 13A and 13B**, long term engraftment and *HBG1/2* gene promoter base editing were retained in irradiated NSG (irrNSG) mice. Human cell chimerism was comparable in BM samples of *HBG1/2* promoter base edited and unedited treatment groups at 16 weeks post-transplantation, i.e., long term engraftment (**FIG. 13A**). *HBG1/2* gene promoter base editing reached >85% in bulk BM cells with increasing doses of ABE8.8 mRNA (**FIG. 13B**). **FIGS. 14A-14C** present results demonstrating that *HBG1/2* gene promoter-edited HSCs also displayed long term, multi-lineage hematopoietic reconstitution in irrNSG mice in the irrNSG mouse model. **FIG. 15** shows that comparable *HBG1/2* gene promoter base editing was retained long term (16 weeks) in NBSGW and irrNSG mice. As observed in **FIG. 15**, similar *HBG1/2* gene promoter base edited levels were detected in bulk bone marrow (BM) cells obtained from NBSGW (30 nM: $88.3 \pm 0.6\%$) mice and irrNSG (30 nM: $85.1 \pm 3.7\%$) mice at 16 weeks post-transplantation.

The results of the studies of dose titration of mRNA-encoded ABE8.8 (ABE8.8 mRNA) for *HBG1/2* gene promoter base editing as described in this Example demonstrate that >85-88% of on-target base editing was attained following titration of

mRNA (MRNA288) at doses from 1 nM to 30 nM. Similar base editing levels, human chimerism and multi-lineage human hematopoietic cell reconstitution were achieved in bulk bone marrow cells from NBSGW and irrNSG mice using mRNA encoding the ABE base editors that were assessed. In addition, on-target base editing using a 30 nM ABE8.8 mRNA dose resulted in >53% gamma globin (γ -globulin) protein induction *in vivo* in NBSGW bone marrow-derived human erythroid cells compared to the percent of γ -globulin protein induction (<0.3%) assessed in unedited cells. Moreover, similar on-target base editing levels were achieved *in vivo* in the various sorted bone marrow-derived hematopoietic cell sub-populations analyzed compared to the levels achieved in bulk bone marrow cells (16 weeks) for any of the ABE8.8 mRNA doses tested, mRNA material (GMP-like versus R&D grade) used, or mouse model NBSGW and irrNSG) tested, thus indicating that *HBG1/2* gene promoter base editing in HSCs is maintained long-term *in vivo*.

FIG. 16 presents a schematic showing a further long term engraftment study that includes a secondary engraftment component. The study used the NBSGW mouse model as described above. Mice were transplanted with PB-CD34⁺ cells (1×10^6 cells) electroporated with ABE mRNA and gRNA. Cell marker phenotype and % base editing analyses were performed on bone marrow cells obtained from the engrafted mice at 8 and 16 weeks post transplantation. After 16 weeks, mice were again injected with bone marrow cells (5×10^6 cells) and bone marrow phenotype and % base editing were analyzed 8 weeks after the second injection. **FIG 16** shows graphs of chimerism, phenotype analysis and % base editing analyses of BM cells obtained from mice at 16 + 8 weeks post dose and demonstrates that the % of LIN-hCD34⁺ cells is consistent between unedited and the base-edited groups and that base editing persisted in secondary engraftment. Of note, for secondary transplantation of cells into mice as described herein, the animals did not require an ablation procedure, e.g., treatment with busulfan, to achieve engraftment of the base-edited donor cells. **FIGS. 17A and 17B** show the results of human BM chimerism (hCD45⁺/(hCD45⁺ + mCD45⁺)) and percent base editing in cells assessed from transplanted mice at week 13 of the dose titration study.

Example 6. Characterization of base-edited CD34+ cells prior to in vivo transplantation, electroporation and transplant studies

Prior to introduction and engraftment in recipient animals in the mouse models of the above-described Examples, donor human CD34+ cells were assessed for the expression of apoptosis, lineage and phenotype protein antigen markers by flow cytometry methods using specific, labeled anti-protein marker antibodies. The assessed lineage and phenotyping markers included the apoptosis markers Annexin V and 7-ADD. The assessed antigen markers included hCD45, mCD45, CD3, CD235a, CD19, CD34, CD15, CD33, and DAPI, and the assessed phenotyping antigen markers included hCD45, mCD45, CD45RA, CD90, CD34, CD15, CD38 and DAPI.

For the analyses, PBMCs were collected from a 31 year old male Caucasian donor and cryopreserved. Apoptosis was measured (using 7-AAD and Annexin V markers) in samples of freshly thawed, donor cells at 24 hours from the time of collecting cells (PBMCs) from the donor to the time of enriching CD34+ cells ("24hr isolation"), e.g., using apheresis, from the collected PBMCs, and at 48+ hours from the time of collecting cells from the donor to the time of enriching CD34+ cells ("48+hr isolation"). The assessment of apoptosis of donor cells was compared with that of peripheral blood mononuclear cells (PBMCs) as control (**FIGs. 18A and 18B**). The number of donor CD34+ cells assessed for apoptosis at 24hr isolation was 6.75×10^8 CD34+ cells; the number of CD34+ cells assessed for apoptosis at 48+hr isolation was 1.60×10^9 CD34+ cells. The 24hr isolation CD34+ cell population analyzed was found to contain 96.3% live cells; the 48+hr isolation CD34+ cell population analyzed was found to contain 96.7% live cells. Without wishing to be bound by theory, donor CD34+ cells are considered to be stem cells or stem-like cells, with stem cell properties and which have the potential to differentiate into other hemapoietic cell lineages. An apoptosis detection kit (BioLegend, Catalog # 640926) was used for the Annexin V marker analysis. **FIG. 18C** graphically shows the locations of live cells (lower left quadrant of graph), dead cells (upper right quadrant of graph) and apoptotic cells (lower right quadrant of graph) following flow cytometry analysis using the BioLegend kit.

As used in the examples herein and as reflected in the drawings, 24hr isolation or pre-enrichment refers to a 24 hour time period between collecting donor cells (PBMCs) from a donor to the time of isolating or enriching CD34+ cells, e.g., using apheresis, from the collected PBMCs ("24hr isolation" or "24hr pre-enrichment"), and 48+hr isolation or

pre-enrichment refers to a 48+ hour time period between collecting donor cells (PBMCs) from a donor to the time of isolating or enriching CD34+ cells from the collected PMBCs (“48+hr isolation” or “48+hr pre-enrichment”). These time periods generally reflect the amount of time that donor cells (peripheral blood cells or PBMCs) are outside of the body prior to isolating or enriching for CD34+ cells (i.e., human stem cells or stem-like cells) for electroporation with mRNA encoding a base editor (e.g., ABE such as ABE8.8) and *HBG1/2* gRNA. Cryopreservation of the donor cells and thawing of the cells for use prior to electroporation and base editing did not adversely affect the ability of the edited cells to engraft in *in vivo* mouse models following transplantation.

Apoptosis was also measured in CD34+ cells prior to electrophoresis (“Pre EP”) of the cells and following electrophoresis (“Post EP”). **FIG. 19A** shows the results of an apoptosis analysis performed on “Pre-EP” samples in which cells were cultured for 48+ hour post thawing after cryopreservation. **FIG. 19B** shows the measurement of apoptosis determined by flow cytometry analyses performed on different groups of “Post-EP” CD34+ cell samples enriched at 24hr or 48+hr (unedited versus base-edited CD34+ cells). **FIG. 19C** shows flow cytometry results of a lineage analysis performed on freshly thawed donor cells through 24hr post electroporation using antibody reagents specific for the lineage markers analyzed.

Both small and large scale electroporation systems were used for the base editing studies. For small scale electroporation, an OC-400 (total volume 400 μ L) cell electroporation cartridge is used; for large scale electroporation, a CL1.1 cell electroporation cartridge (total volume 3 mL) is used. Flow electroporation is performed using the MaxCyte instruments in conjunction with the small and large cartridges. For optimum base editing efficiency, cell viability, and base editing retention over time, CD34+ cells should be electroporated 24 to 48+ hours after collection from the donor and cryopreservation.

By way of example, a large scale process that resulted in an at least 16 week base editing retention in CD34+ cells electroporated with ABE8.8 mRNA and *HBG1/2* gRNA involved the use of cryopreserved cells as starting material, Lonza X-vivo 10 containing glutamax supplement as cell culture medium; cells cultured for 48+ hours post electroporation as described herein; an OC-400 electroporation cell cartridge (prechilled) or a CL1.1 electroporation cell cartridge (ambient temperature) for flow electroporation using a MaxCyte instrument; a cell culture / cell transfer vessel or container which is a

flask, a culture plate, or a conical cell culture tube (50 mL); a post electroporation incubation temperature for the electroporated cells of 37°C; and a pre-electroporation cell process involving multiple centrifugations and washes at 4°C such as described hereinabove.

- 5 A cell electroporation, transplant and engraftment study was performed based on the study design of **FIG. 2**. **Table 23** below presents the study parameters.

Table 23

Group	Treatment	Timepoint	mRNA and gRNA	Cell Source	Viability % at transplant
1	Unedited	N/A	N/A	48+-Hr Pre- Enrichment Collection	90
2	OC-400	Research	Research		91.7
3	Unedited	N/A	N/A	24-Hr Pre- Enrichment Collection	93
4	OC-400	Research	Research		93.9
5	CL-1.1	Research	Research		90.4

- The NBSGW mouse model was used for injection/transplant of base-edited
- 10 CD34⁺ human cells and subsequent engraftment in the study. Transplanted cells were collected and evaluated at 8-weeks and 16-weeks post injection. At the 8-week evaluation, BM FACS (lineage and hematopoietic stem cell (HSC) phenotyping), bulk BM and Blood NGS, sorting of Lin⁻CD34⁺ cells, NGS were performed and cells were cultured. At the 16-week collection, BM FACS (lineage and hematopoietic stem cell
- 15 (HSC) phenotyping), bulk BM and blood cell NGS, cell sorting based on (hCD15⁺ hCD19⁺, hCD34⁺ and GlyA⁺) cell markers, secondary engraftment from Groups 3,4 and 5, cell culture and CFU analyses and GlyA⁺ for UPLC were carried out. **FIG. 20A** shows viability of the cells at pre-electroporation (Pre-EP), and at 24, 48 and 72 hours post electroporation. **FIG. 20B** shows the percent of base editing in the transplanted cells
- 20 at the indicated time periods. **FIG. 21A** shows the percentage of enucleated cells (%DAPI-/NucRed-) after thawing the cells. **FIG. 21B** shows a growth curve of cells after thawing on days 0-14. The results demonstrate that both unedited cells and edited cells that were electroporated using small or large scale electroporation, and both the 24 and

48+ hour pre-enrichment of unedited or base-edited cells after thawing, showed nearly identical enucleation and cell growth.

The induction of gamma globin expression was demonstrated to be similar (about 60%) in base-edited cells following small scale electroporation and 48+ or 24 hour pre-enrichment and in base-edited cells following large scale electroporation and 24 hour pre-enrichment (**FIG. 22A**). The numbers of cell colonies (colony forming unit (CFU)) formed by thawed unedited CD34+ cells and base edited CD34+ cells that had undergone either small or large scale electroporation and 24 or 48+ hour pre-enrichment treatment were found to be similar (**FIG. 22B**). In addition, the numbers of different, specific colony forming unit cell types were also very similar among the aforementioned groups, namely, for CFUs such as burst-forming unit-erythroid (BFU-E) cells, which are the earliest precursor cells specific to the erythroid lineage; the CFU-GM, i.e., granulocyte-macrophage progenitor cells, which are precursors of monoblasts and myeloblasts; and CFU-GEMM ("Colony Forming Units-Granulocyte, Erythrocyte, Monocyte, Megakaryocyte"), which is a colony forming unit that generates myeloid cells. CFU-GEMM cells are the oligopotential progenitor cells for myeloid cells; thus, they are also known as common myeloid progenitor cells or myeloid stem cells. Red blood cells, white blood cells and platelets derive from CFU-GEMM.

The results from the study showed that at 8 weeks post dosing of animals with unedited or base-edited CD34+ cells, the percent human cell chimerism in mouse bone marrow (BM), (hCD45+/(hCD45+ + mCD45+), was high among the different cell treatment groups (**FIG. 23A**). As observed, based-edited CD34+ cells that had been electroporated under small scale (OC400) or large scale (CL1.1) electroporation conditions and subjected to either 24 or 48+ hour pre-enrichment conditions showed similar percent chimerism. The percentage of base editing (A to G) in input, bulk BM, CD34+/LIN-, and whole blood assessed at 8 weeks post-transplant was similar between base-edited CD34+ cells (24hr pre-enrichment) electroporated under small and large scale conditions (**FIG. 23B**). **FIGS. 24A-24D** demonstrate that similar percentages of human donor cell chimerism in mouse bone marrow (BM), hCD15+ cells, GlyA+ cells and CD34+ human cells were detected in animals at 8 and 16 weeks post dosing with unedited or base-edited donor CD34+ cells, independent of the electroporation type or time of isolation/enrichment of CD34+ cells. Similar results were determined for assessments of chimerism, base editing percent and fetal globin reactivation after 16

weeks of engraftment (**FIGs. 25A-25C**). The engrafted, base edited cells showed fetal hemoglobin (HbF) upregulation.

The percent base editing in different cell phenotype and lineage subpopulations (i.e., GlyA+, CD15+, CD19+, LIN-CD34+, BM) as assessed at 16 weeks post dosing of unedited or base edited cells into animals was determined to be high, (approximately 80% or greater), for animals transplanted with base-edited CD34+ cells subjected to small scale electroporation (OC-400) (CD34+ cells isolated 24 hours from the time of collecting the human donor blood sample, "24 hr") and for animals transplanted with base-edited CD34+ cells subjected to large scale electroporation (CL1.1) (CD34+ cells isolated 24 hours from the time of collecting the human donor blood sample, "24 hr"). (**FIG. 26**). Similar amounts of base editing were detected in each of the cell subtypes at 16 weeks following transplant of base-edited CD34+ cells that had been electroporated using either a small or large scale electroporation process, where the CD34+ cells were isolated 24 hours from the time of collecting the human donor blood sample, "24 hr".

The results of the studies described in this Example demonstrate that cells, namely, CD34+ cells isolated within 24 hours post apheresis, engrafted in animals and retained base editing levels out to at least 16 weeks post transplantation. CD34+ cells isolated 48+ hours post apheresis engrafted but did not retain editing levels out to 16 weeks post transplantation (e.g., **FIGS. 23B and 25B**). The percent (%) Gamma/beta-like levels in 24-hour post apheresis groups were determined to be within the therapeutic range. Base editing levels of CD34+ cells electroporated using small and large scale electroporation methods, OC-400 and CL1.1, respectively, and retention are similar at 16 weeks post transplantation in NBSGW mice (the NBSGW mouse model).

As described in the above Examples, base-edited, transplanted donor cells successfully engrafted in recipient animals (e.g., the NBSGW mouse model) and were sustained in the animals for a long term period of at least 16 weeks or longer. Engraftment in the animals was achieved using edited donor cells (e.g., CD34+ cells) that were electroporated using both small and large scale electroporation procedures and using mRNA encoding the base editor, e.g., an adenosine deaminase-containing base editor, ABE8.8, and a guide RNA, e.g., *HBG1/2* gRNA (g1). Long term engraftment and base editing (at least 16 weeks or longer) was demonstrated. No known off-target effects were detected using engrafted, edited donor cells transfected with mRNA encoding the adenosine base editor (e.g., ABE8/8) and a gRNA (g1) in the long term engraftment

studies described herein. In addition, the base edited cells used for transplant and engraftment offer greater levels of efficacy and a better safety profile in animals, as the base editing function within the cells does not involve double stranded DNA breaks. Without wishing to be bound by theory, the CD34⁺ cells (donor cells) used for base-
5 editing and transplantation as described herein provide a higher quality and more robust starting cell population for transplantation and engraftment, as the CD34⁺ cells were enriched from donor PBMCs and electroporated with ABE (e.g., ABE8.8) mRNA and gRNA within a relatively short period of time before electroporation, base editing and *in vivo* transplantation, e.g., a 24 or 48+ hour period of time after collection and apheresis of
10 the cells from the donor. In addition, a higher yield of viable, base-edited donor cells, which maintain stem cell like properties and functional activity, is provided by the cells and methods herein. The cells that were base-edited using the base editors and editing techniques described herein were demonstrated to be highly suitable for *in vivo* transplantation and long term engraftment versus other types of gene editing technologies,
15 which frequently involve the use of different nucleases, induce DNA strand breaks and result in more off-target editing effects.

Example 7. Engraftment method

In vivo animal studies were conducted using healthy 5- to 7-week-old female
20 NBSGW mice (Stock Number 026622, Jackson Laboratory). Procedures involving animals were reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at the Charles River CRADL Facility (IACUC Protocol CR-0071) and complied with all applicable sections of the Animal Welfare Act regulations (9 CFR), the Public Health Service Policy on Human Care and Use of Laboratory Animals, the
25 Guide for the Care and Use of Laboratory Animals, and the standards set forth in the USDA Animal Welfare Act.

Female NBSGW mice were randomly assigned to groups as per the study design. On Day 0, mice were weighed prior to receiving 500 μ L of a single intravenous (IV) injection of the test article (1×10^6 edited hCD34⁺ HSPCs) or control cells (1×10^6
30 unedited hCD34⁺ HSPCs) via the tail vein. Post-transplantation, the mice were observed daily for general health and weighed periodically throughout the study.

Mice were sacrificed at terminal timepoints per study design. To harvest bone marrow samples, the femur, tibia, and pelvic bones of the sacrificed mice were collected

in IMDM/10% FBS. The total bone marrow was flushed and filtered using a 70 μ m nylon strainer. The resulting bone marrow cell pellet was stored on ice until use.

Other Embodiments

5 From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

 The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or
10 subcombination) of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

 All patents and publications mentioned in this specification are herein
incorporated by reference to the same extent as if each independent patent and publication
15 was specifically and individually indicated to be incorporated by reference.

CLAIMS

What is claimed is:

1. A method of engrafting nucleobase-edited hematopoietic stem cells or progenitors
5 thereof in a subject having a hemoglobinopathy, the method comprising:
 - (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and a base editor comprising a polynucleotide programmable DNA binding domain and a deaminase domain, or a polynucleotide encoding the base editor, wherein the guide RNA targets the polynucleotide programmable DNA binding domain to induce a
10 nucleobase change in a target hemoglobin (HBB) gene or in the promoter region of *HBG1/2*, thereby obtaining nucleobase-edited hematopoietic stem cells or progenitors thereof; and wherein the nucleobase-edited hematopoietic stem cells or progenitors thereof are contacted with the gRNA and the base editor within 48 hours following collection from a donor; and
15 (b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after administration.
- 20 2. The method of claim 1, wherein the nucleobase-edited hematopoietic stem cells or progenitors thereof comprise CD34⁺ cells enriched from polymorphonuclear blood cells (PBMCs) collected from the donor.
3. A method of engrafting nucleobase-edited hematopoietic stem cells or progenitors
25 thereof in a subject having a hemoglobinopathy, the method comprising:
 - (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and a base editor comprising a polynucleotide programmable DNA binding domain and a deaminase domain, or a polynucleotide encoding the base editor, wherein the guide RNA targets the polynucleotide programmable DNA binding domain to induce a
30 nucleobase change in a target hemoglobin (HBB) gene or in the promoter region of *HBG1/2*, thereby obtaining nucleobase-edited hematopoietic stem cells or progenitors thereof; and

(b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after administration.

5

4. A method of treating a hemoglobinopathy in a subject, the method comprising:

(a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and a base editor comprising a polynucleotide programmable DNA binding domain and a deaminase domain, or a polynucleotide encoding the base editor, wherein the guide
10 RNA targets the polynucleotide programmable DNA binding domain to induce a nucleobase change in a target hemoglobin (HBB) gene or in a target hemoglobin (HBB) gene in the promoter region of *HBG1/2*, thereby obtaining nucleobase-edited hematopoietic stem cells or progenitors thereof; and

(b) administering the nucleobase-edited hematopoietic stem cells or progenitors
15 thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after administration.

5. The method of any one of claims 1-4, wherein the nucleobase change is an A to G
20 nucleobase change.

6. The method of any one of claims 1-5, wherein the deaminase domain is an adenosine deaminase domain and shares at least 85% sequence identity with the sequence
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD
25 PTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQK
KAQSSTD (SEQ ID NO: 3), and wherein the adenosine deaminase domain is capable of catalyzing the hydrolytic deamination of adenine or adenosine.

30 7. The method of claim 6, wherein the adenosine deaminase domain comprises one or more of the following alterations: Y123H, Q154S, and Q154R.

8. The method of claim 6 or claim 7, wherein the adenosine deaminase domain comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, and Q154R.
- 5 9. The method of any one of claims 6-8, wherein the adenosine deaminase domain comprises a combination of alterations selected from the group consisting of:
Y147R, Q154R, and Y123H;
Y147R, Q154R, and I76Y;
Y147R, Q154R, and T166R;
10 Y147T and Q154R; Y147T and Q154S; and
Y123H, Y147R, Q154R, and I76Y.
10. The method of any one of claims 6-9, wherein the adenosine deaminase domain comprises the alterations Y147R, Q154R, and Y123H.
- 15 11. The method of any one of claims 1-10, wherein the deaminase domain is a TadA*8 variant.
12. The method of claim 11, wherein the TadA*8 variant is selected from the group
20 consisting of: TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6,
TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, and TadA*8.13.
13. The method of any one of claims 1-12, wherein the base editor is an ABE8 base editor selected from the group consisting of: ABE8.1, ABE8.2, ABE8.3, ABE8.4,
25 ABE8.5, ABE8.6, ABE8.7, ABE8.8, ABE8.9, ABE8.10, ABE8.11, ABE8.12, and
ABE8.13.
14. A method of engrafting nucleobase-edited hematopoietic stem cells or progenitors thereof in a subject having a hemoglobinopathy, the method comprising:
30 (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and an adenosine base editor comprising a polynucleotide programmable DNA binding domain and an adenosine deaminase domain comprising an amino acid sequence with at least 85% sequence identity to the sequence

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD
PTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQK
KAQSSTD (SEQ ID NO: 3) and comprising the alterations Y123H, Y147R, and Q154R,

5 or a polynucleotide encoding the base editor, wherein the adenosine deaminase domain catalyzes the hydrolytic deamination of adenine or adenosine, and wherein said guide RNA targets said polynucleotide programmable DNA binding domain to induce an A to G nucleobase change in a target hemoglobin (HBB) gene or in the promoter region of *HBG1/2*, thereby obtaining nucleobase-edited hematopoietic stem cells or progenitors
10 thereof; and

(b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after administration.

15

15. A method of treating a hemoglobinopathy in a subject, the method comprising:

(a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and an adenosine base editor comprising a polynucleotide programmable DNA binding domain and an adenosine deaminase domain comprising an amino acid sequence
20 with at least 85% sequence identity to

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD
PTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQK
KAQSSTD (SEQ ID NO: 3) and comprising the alterations Y123H, Y147R, and Q154R,

25 or a polynucleotide encoding the base editor, wherein the adenosine deaminase domain catalyzes the hydrolytic deamination of adenine or adenosine, and wherein said guide RNA targets said polynucleotide programmable DNA binding domain to induce an A to G nucleobase change in a target hemoglobin (HBB) gene or in the promoter region of *HBG1/2*, thereby obtaining nucleobase-edited hematopoietic stem cells or progenitors
30 thereof; and

(b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited

hematopoietic stem cells or progenitors thereof in tissues of the subject after administration.

16. The method of any one of claims 6-15, wherein the adenosine deaminase domain
5 comprises an alteration at position 82 or 166.
17. The method of claim 16, wherein the alteration at position 82 is V82S.
18. The method of claim 16 or claim 17, wherein the alteration at position 166 is
10 T166R.
19. The method of any one of claims 6-18, wherein the adenosine deaminase domain comprises an alteration at positions 166 and 82.
20. The method of any one of claims 6-19, wherein the deaminase domain has at least
15 90% sequence identity to the sequence.
21. The method any one of claims 7-20, wherein the base editor further comprises a wild-type adenosine deaminase domain.
20
22. The method of any one of claims 1-21, wherein the polynucleotide programmable DNA binding domain is a Cas9.
23. The method of claim 22, wherein the Cas9 is a SpCas9, a SaCas9, or a variant
25 thereof.
24. The method of any one of claims 1-23, wherein the polynucleotide programmable DNA binding domain comprises a modified Cas9 having an altered protospacer-adjacent motif (PAM) specificity.
30
25. The method of claim 24, wherein the Cas9 has specificity for a PAM sequence selected from the group consisting of NGG, NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, and NGC, wherein N is A, G, C, or T and wherein R is A or G.

26. The method of any one of claims 1-25, wherein the polynucleotide programmable DNA binding domain is nuclease inactive.
- 5 27. The method of any one of claims 1-25, wherein the polynucleotide programmable DNA binding domain is a nickase.
28. The method of claim 26 or claim 27, wherein the polynucleotide programmable DNA binding domain comprises the alterations D10A and/or H840A.
- 10 29. The method of claim 28, wherein the polynucleotide programmable DNA binding domain comprises the alteration D10A.
30. The method of any one of claims 1-29, wherein the deaminase domain comprises an adenosine deaminase monomer.
- 15 31. The method of any one of claims 1-30, wherein the deaminase domain comprises an adenosine deaminase dimer.
- 20 32. A method of engrafting edited hematopoietic stem cells or progenitors thereof in a subject having a hemoglobinopathy, the method comprising:
- (a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide RNA and a base editor comprising an amino acid sequence with at least 80% sequence identity to one of the following two amino acid sequences:
- 25 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD
PTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGW
AVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
30 RRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGFLFGN
LIA LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK

NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE
5 YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICK
GILQTVKVVDDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
10 ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLES
EFVYGDYK VYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
15 LIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFLENKRKMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGGSPEDNEQKQLFVEQHKHYLDEHEQISEFSKRVI
LADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
20 TSTKEVL DATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV
(SEQ ID NO: 258), and
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGA
RDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
25 KAQSSTDSSGGSSGGSSGSETPGTSESATPESSGGSSGGSSSEVEFSHEYWMRHALTL
AKRARDEREVPVGAVLVHNNRVIGEGWNR AIGLHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLCRFFRMRRVFNAQKKAQSSTDSSGGSSGGSSGSE
TPGTSESATPESSGGSSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGN
30 TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK
VDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTD
KADLR LIYLA LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINA
SGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA

EDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
 APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
 QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
 RQEDFYFPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
 5 VVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGM
 RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
 SLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDD
 KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHD
 DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
 10 HKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
 EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
 NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGSELDDKAG
 FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQ
 FYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIK
 15 SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
 ATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
 DSPTVAYSVLVVAKEGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKE
 VKKDLIHLKPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYASHYEK
 LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD
 20 KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLY
 ETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 259), or a

polynucleotide encoding the base editor, wherein said guide RNA targets said
 polynucleotide programmable DNA binding domain to induce an A to G nucleobase
 change in the promoter region of HBG1/2, thereby obtaining edited hematopoietic stem
 25 cells or progenitors thereof;

(b) administering the nucleobase-edited hematopoietic stem cells or progenitors
 thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited
 hematopoietic stem cells or progenitors thereof in tissues of the subject after
 administration.

30

33. A method of treating a hemoglobinopathy in a subject, the method comprising:

(a) contacting hematopoietic stem cells or progenitors thereof *in vitro* with a guide
 RNA and a base editor comprising an amino acid sequence with at least 80% sequence

identity to one of the following two amino acid sequences

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD
 PTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
 NAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQK
 5 KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGW
 AVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
 RRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
 EKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKL
 FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGN
 10 LIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAK
 NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
 DNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRF
 AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE
 15 YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGTYHDLKKIHKDKDFLDNEENEDILEDIVLTLTLF
 EDREMIEERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
 DFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICK
 GILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
 20 ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
 QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
 DNLTKAERGGSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
 VKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLES
 EFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
 25 LIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIME
 RSSFEKNPIDFLEAKGYKEVKKDLIHKLPKYSLEFLENKRKMLASAGELQKGNE
 LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
 LADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
 30 TSTKEVLDTLIHQITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV
 (SEQ ID NO: 258), and
 MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
 DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGA

RDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
 KAQSSTDSSGGSSGGSSGSETPGTSESATPESSGGSSGGSSSEVEFSHEYWMRHALTL
 AKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQ
 NYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 5 GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSSGGSSGGSSGSE
 TPGTSESATPESSGGSSGGSSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGN
 TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK
 VDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVSTD
 KADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINA
 10 SGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
 EDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
 APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
 QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
 RQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
 15 VVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGM
 RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
 SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD
 KVMKQLKRRRYTGWGRLSRKLINGIRDQKQSGKTILDFLKSDGFANRNFQMQLIHD
 DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
 20 HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
 EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
 NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
 FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQ
 FYKVVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMIK
 25 SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
 ATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
 DSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
 VKKDLIHLKPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEK
 LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHARD
 30 KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLY
 ETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 259) , or a
 polynucleotide encoding the base editor , wherein said guide RNA targets said
 polynucleotide programmable DNA binding domain to induce an A to G nucleobase

change in a target hemoglobin (HBB) gene or in the promoter region of HBG1/2, thereby obtaining edited hematopoietic stem cells or progenitors thereof;

- (b) administering the nucleobase-edited hematopoietic stem cells or progenitors thereof to a subject in an effective amount to obtain engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof in tissues of the subject after administration.

34. The method of any one of claims 1-33, wherein engraftment efficiency of the nucleobase-edited hematopoietic stem cells or progenitors thereof is measured in the subject at about 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, or 8 or more weeks after administering the cells to the subject.

35. The method of any one of claims 1-34, wherein engraftment efficiency of the nucleobase-edited hematopoietic stem cells or progenitors thereof is measured in the subject at least 8 weeks after administering the cells to the subject.

36. The method of any one of claims 1-35, wherein engraftment efficiency of the nucleobase-edited hematopoietic stem cells or progenitors thereof is measured in the subject at least 16 weeks after administering the cells to the subject.

37. The method of any one of claims 34-36, wherein the measured engraftment efficiency is at least about 20%.

38. The method of any one of claims 34-37, wherein the measured engraftment efficiency is at least about 30%.

39. The method of any one of claims 34-38, wherein the measured engraftment efficiency is at least about 40%.

40. The method of any one of claims 34-39, wherein the measured engraftment efficiency is at least about 50%.

41. The method of any one of claims 1-40, wherein at least about 50% of the hematopoietic cells or progenitors thereof in (b) are viable.
42. The method of any one of claims 1-41, wherein at least 30% of the hematopoietic
5 cells or progenitors thereof in (b) comprise the nucleobase change.
43. The method of any one of claims 1-42, wherein at least 50% of the hematopoietic cells or progenitors thereof in (b) comprise the nucleobase change.
- 10 44. The method of any one of claims 1-43, wherein at least 60% of the hematopoietic cells or progenitors thereof in (b) comprise the nucleobase change.
45. The method of any one of claims 1-44, wherein at least 70% of the hematopoietic cells or progenitors thereof in (b) comprise the nucleobase change.
- 15 46. The method of any one of claims 1-45, wherein the hematopoietic cells or progenitors thereof are isolated or derived from the subject.
47. The method of any one of claims 1-46, wherein the hematopoietic stem cells or
20 progenitors thereof comprise a single-nucleotide polymorphism (SNP) associated with sickle cell disease (SCD).
48. The method of claim 47, wherein the SNP associated with SCD results in a E6V substitution in a hemoglobin beta unit encoded by the HBB gene.
- 25 49. The method of any one of claims 1-48, wherein at least 30% of the hematopoietic stem cells or progenitors thereof retain base editing activity following engraftment.
50. The method of any one of claims 1-49, wherein at least 50% of the hematopoietic
30 stem cells or progenitors thereof retain base editing activity following engraftment.
51. The method of any one of claims 1-50, wherein at least 60% of the hematopoietic stem cells or progenitors thereof retain base editing activity following engraftment.

52. The method of any one of claims 1-51, wherein at least 70% of the hematopoietic stem cells or progenitors thereof retain base editing activity following engraftment.
- 5 53. The method of any one of claims 1-52, wherein at least 80% of the hematopoietic stem cells or progenitors thereof retain base editing activity following engraftment.
54. The method of any one of claims 1-53, wherein at least 90% of the hematopoietic stem cells or progenitors thereof retain base editing activity following engraftment.
- 10 55. The method of any one of claim 1-54, wherein the nucleobase change results in a E6A substitution in the hemoglobin beta unit encoded by the HBB gene.
56. The method of any one of claims 1-55 wherein the hematopoietic cells or
15 progenitors thereof retain the ability to differentiate following administration.
57. The method of any one of claims 1-56, wherein the hematopoietic cells or progenitors thereof are capable of generating erythrocytes.
- 20 58. The method of any one of claims 1-57, wherein the polynucleotide encoding the base editor comprises mRNA or is mRNA.
59. The method any one of claims 1-58, wherein the hematopoietic stem cells or progenitors thereof are contacted with at least about 1 nM of mRNA encoding the base
25 editor.
60. The method any one of claims 1-59, wherein the hematopoietic stem cells or progenitors thereof are contacted with at least about 3 nM RNA encoding the base editor.
- 30 61. The method of any one of claims 1-60, wherein the hematopoietic stem cells or progenitors thereof are contacted with at least about 10 nM RNA encoding the base editor.

62. The method of any one of claims 1-61, wherein the hematopoietic stem cells or progenitors thereof are contacted with at least about 30 nM RNA encoding the base editor.
- 5 63. The method of any one of claims 1-62, wherein the hematopoietic stem cells or progenitors thereof are contacted with at least about 50 nM RNA encoding the base editor.
64. The method of any one of claims 1-63, wherein the hematopoietic stem cells or
10 progenitors thereof are contacted with at least about 3000 nM of the gRNA.
65. The method of any one of claims 1-64, wherein levels of fetal hemoglobin (HbF) are increased in the subject following engraftment relative to the levels in a control subject that received unedited hematopoietic stem cells or progenitors thereof.
15
66. The method of any one of claims 1-65, wherein levels of fetal hemoglobin (HbF) are increased in the subject by at least about 20% relative to the levels in a control subject that received unedited hematopoietic stem cells or progenitors thereof.
- 20 67. The method of any one of claims 1-66, wherein HbS expression is reduced in the subject in the subject following engraftment relative to HbS expression in a control subject that received unedited hematopoietic stem cells or progenitors thereof.
68. The method of any one of claims 1-67, wherein HbS expression is reduced in the
25 subject by at least about 20% relative to HbS expression in a control subject that received unedited hematopoietic stem cells or progenitors thereof.
69. The method of any one of claims 1-68, wherein the nucleobase-edited hematopoietic stem cells or progenitors thereof express CD34.
30
70. The method of any one of claims 1-69, wherein the nucleobase-edited hematopoietic stem cells or progenitors thereof express one or more of CD34, CD45, CD19, and GlyA.

71. The method of any one of claims 1-70, wherein the nucleobase-edited hematopoietic stem cells or progenitors thereof express HbF.
- 5 72. The method of any one of claims 1-71, wherein the hematopoietic stem cells or progenitors thereof are human hematopoietic stem cells or progenitors thereof.
73. The method of any one of claims 1-72, wherein the subject is a mammal.
- 10 74. The method of any one of claims 1-73, wherein the subject is a human.
75. The method of any one of claims 1-74, wherein the nucleobase-edited hematopoietic stem cells or progenitors thereof are GlyA⁺.
- 15 76. The method of any one of claims 1-75, wherein the subject has sickle cell disease (SCD), thalassemia, and/or anemia.
77. The method of claim 76, wherein the subject has SCD.
- 20 78. The method of any one of claims 1-77, wherein the nucleobase-edited hematopoietic stem cells or progenitors thereof are autologous to the subject.
79. The method of any one of claims 3-68 and 70-78, wherein the nucleobase-edited hematopoietic stem cells or progenitors thereof are not enriched prior to administration.
- 25 80. The method of any one of claims 1-79, wherein the nucleobase-edited hematopoietic stem cells or progenitors thereof are enriched prior to administration.
81. The method of any one of claims 1-80, wherein the nucleobase change abolishes, 30 disrupts, or reduces BCL11A binding in the promoter region of *HBG1/2*.
82. The method of any one of claims 1-81, wherein the nucleobase change is at a position selected from -114, -117, -175, and -198 in the promoter region of *HBG1/2*.

83. The method of any one of claims 1-82, wherein the nucleobase change is associated with an increase in expression of HBG1/2.
84. The method of any one of claims 1-83, wherein the nucleobase change is associated with an increase in levels of hemoglobin gamma subunit in the hematopoietic stem cells or progenitors thereof.
85. The method of any one of claims 1-84, wherein an increased level of HbF protein is expressed in the subject after administration.
86. The method of any one of claims 1-85, wherein the administration results in expression of HbF in the subject for at least 8 weeks.
87. The method of any one of claims 1-86, wherein the administration results in expression of HbF in the subject for at least 16 weeks.
88. The method of any one of claims 1-87, wherein the administration reduces or ameliorates a symptom associated with sickle cell disease in the subject.
89. The method of any one of claims 1-88, wherein erythrocytes generated from the hematopoietic cells or progenitors thereof exhibit reduced sickling.
90. The method of any one of claims 1-89, wherein at least 50% editing is retained at least 16 weeks after the administration in a tissue of the subject.
91. The method of any one of claims 1-90, wherein at least 80% editing is retained 16 weeks after the administration in a tissue of the subject.
92. The method of any one of claims 1-91, wherein administration is performed multiple times.
93. The method of any one of claims 1-92, wherein administration is performed multiple times at an interval of at least about one month.

94. The method of any one of claims 1-93, wherein the guide RNA comprises a nucleotide sequence selected from SEQ ID NOs: 130-155 listed in Table 1.

- 5 95. The method of any one of claims 1-93, wherein the gRNA comprises or consists of the sequence, from 5'-3':
 GACCAAUAGCCUUGACAGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAG
 GCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU,
 corresponding to bases 4-97 of SEQ ID NO: 129.

10

96. The method of any one of claims 1-93, wherein the guide RNA comprises or consists of the nucleotide sequence, from 5'-
 3':csususGACCAAUAGCCUUGACAGUUUUAGAGCUAGAAAUAGCAAGUUAAA
 AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU_{su}
 15 susu (SEQ ID NO: 129), wherein lowercase characters indicate 2'-O-methylated nucleobases, and "s" indicates phosphorothioates (SEQ ID NO: 129).

97. The method of any one of claims 1-93, wherein the guide RNA comprises or consists of the nucleotide sequence of any one of 5'-

- 20 gsascsUUCUCCACAGGAGUCAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAA
 AUAAGGCUAGUCCGUUAUUAACUUGAAAAAGUGGCACCGAGUCGGUGCU_s
 ususu-3' (SEQ ID NO: 126), 5'-
 ascsusUCUCCACAGGAGUCAGGGUUUUAGAGCUAGAAAUAGCAAGUAAAA
 UAAGGCUAGUCCGUUAUUAACUUGAAAAAGUGGCACCGAGUCGGUGCU_{su}
 25 susu-3' (SEQ ID NO: 127), and 5'-
 csususCUCCACAGGAGUCAGGGUUUUAGAGCUAGAAAUAGCAAGUAAAAU
 AAGGCUAGUCCGUUAUUAACUUGAAAAAGUGGCACCGAGUCGGUGCU_{susu}
 su-3' (SEQ ID NO: 128), wherein lowercase characters indicate 2'-O-methylated nucleobases, and "s" indicates phosphorothioates.

30

98. The method of any one of claims 1-97, wherein the administration is associated with hemoglobin subunit gamma being expressed in at least 50% of cells in the bone marrow of the subject.

99. The method of any one of claims 1-98, wherein the administration is associated with hemoglobin subunit gamma being expressed in at least 60% of cells in the bone marrow of the subject.
- 5 100. The method of any one of claims 1-99, further comprising depletion of one or more lymphocytic lineage cells in the subject prior to administering the hematopoietic stem cells or progenitors thereof.
- 10 101. The method of any one of claims 3-68, and 70-100, wherein the hematopoietic stem cells or progenitors thereof are enriched CD34⁺ cells, and wherein the CD34⁺ cells are enriched from donor peripheral blood mononuclear cells (PBMCs) less than 48 hours after the PBMCs are collected or isolated from a donor.
- 15 102. The method of any one of claims 1-101, wherein the hematopoietic stem cells or progenitors thereof are cryopreserved following collection or isolation from a donor.
- 20 103. The method of any one of claims 1-102, wherein the gRNA and/or the polynucleotide encoding the base editor comprises a 2'-O-Methyl nucleotide modification.
104. The method of claim 103, wherein the 2'-O-Methyl nucleotide modification is disposed at a 3' or 5' end of the gRNA and/or the polynucleotide encoding the base editor.
- 25 105. The method of any one of claims 1-104, wherein the gRNA and/or the polynucleotide encoding the base editor comprises a phosphorothioate internucleotide linkage.
- 30 106. The method of any one of claims 1-105, wherein the hematopoietic stem cells or progenitors thereof are contacted with the polynucleotide encoding the base editor.
107. The method of any one of claims 1-106, wherein the base editor is delivered as a polynucleotide that is expressed in the hematopoietic stem cells or progenitors thereof.

108. The method of any one of claims 1-107, wherein engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof is maintained in the subject for at least 8 weeks.

5

109. The method of any one of claims 1-108, wherein engraftment of the nucleobase-edited hematopoietic stem cells or progenitors thereof is maintained in the subject for at least 16 weeks.

10 110. The method of any one of claims 1-109, wherein the nucleobase-edited hematopoietic stem cells or progenitors thereof are contacted with the gRNA and the base editor within 24 hours following collection from a donor.

111. The method of any one of claims 32-110, wherein the base editor shares at least 90% sequence identity to one of the following two sequences:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHD
PTAHAEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQK
KAQSSTDSSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGW
20 AVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLVDSTDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLLAAK
25 NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE
YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
30 KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICK
GILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK

ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLES
5 EFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
LIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
10 LADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVL DATLIHQ SITGLYETRIDL SQLGGDEGADKRTADGSEFESPKKKRKV
(SEQ ID NO: 258), and
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGA
15 RDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSSEVEFSHEYWMRHALTL
AKRARDEREVPVGAVLVHNNRVIGEGWNR AIGLHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSE
20 TPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGN
TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAK
VDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTD
KADLR LIYLA LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINA
SGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
25 EDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
RQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGAS AQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGM
30 RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD
KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFQMQLIHD
DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR

HKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
 EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
 NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGELSEDKAG
 FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQ
 5 FYKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIK
 SEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
 ATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
 DSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKE
 VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
 10 LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD
 KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLY
 ETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 259).

112. The method of any one of claims 32-111, wherein the base editor shares at least
 15 95% sequence identity to one of the following two sequences:
 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN RVIGEGWNRAIGLHD
 PTAHAEIMALRQGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
 NAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQK
 KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGW
 20 AVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
 RRKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
 EKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKL
 FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGN
 LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK
 25 NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
 DNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRF
 AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE
 YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 30 KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLF
 EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
 DFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICK
 GILQTVKVVDLVKVMGRHHPENIVIAMARENQTTQKGQKNSRERMKRIEEGIK

ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLES
5 EFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
LIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
10 LADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVL DATLIHQ SITGLYETRIDL SQLGGDEGADKRTADGSEFESPKKKRKV
(SEQ ID NO: 258), and
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGA
15 RDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGSSGSEVEFSHEYWMRHALTL
AKRARDEREVPVGAVLVHNNRVIGEGWNR AIGLHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSE
20 TPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGN
TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAK
VDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTD
KADLR LIYLA LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINA
SGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
25 EDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
RQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGAS AQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGM
30 RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDD
KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFQMQLIHD
DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR

HKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQ
5 FYKVVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSPQVNVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
DSPTVAYSVLVVAKEGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKE
VKKDLIIKLPKYSLENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
10 LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 259).

113. The method of any one of claims 3-68, and 70-112, wherein the hematopoietic
15 stem cells or progenitors thereof are enriched CD34⁺ cells, and wherein the CD34⁺ cells
are enriched from donor peripheral blood mononuclear cells (PBMCs) less than 24 hours
after the PBMCs are collected or isolated from a donor.

114. A kit for use in the method of any one of claims 1-113, wherein the kit comprises
20 the guide RNA and a polynucleotide encoding the base editor.