[0007] Recently, Liu *et al.* developed base editing as a technology that edits target nucleotides without creating DSBs or relying on HDR[4–6,24–27]. Direct modification of DNA bases by Cas-fused deaminases allows for C•G to T•A, or A•T to G•C, base pair conversions in a short target window (~5-7 bases) with high efficiency. As a result, base editors have been rapidly adopted by the scientific community. However, several factors may limit their generality for precision genome editing.

[0008] Therefore, the development of programmable editors that are capable of introducing any desired single or multiple nucleotide change, which could install nucleotide insertions or deletions (e.g., at least 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more base pair insertions or deletions), and/or which could alter or modify the nucleotide sequence at a target site with high specificity and efficiency would substantially expand the scope and therapeutic potential of genome editing technologies based on CRISPR.

SUMMARY OF THE INVENTION

[0009] The present disclosure disclosed new compositions (e.g., new PEgRNA and PE complexes comprising same) and methods for using prime editing (PE) to repair therapeutic targets, e.g., those targets identified in the ClinVar database, using PEgRNA designed using a specialized algorithm that is described herein. Thus, in one example, the present application discloses an algorithm for predicting on a large-scale the sequences for PEgRNA that may be used to repair therapeutic targets (e.g., those included in the ClinVar database). In addition, the present application discloses predicted sequences for therapeutic PEgRNAs designed and which can be designed using the disclosed algorithm and which may be used with prime editing to repair therapeutic targets.

[0010] The herein disclosed algorithm and the predicted PEgRNA sequences relate in general to prime editing. Thus, this disclosure also provides a description for the various components and examples of prime editing, including suitable napDNAbp (e.g., Cas9 nickase) and a polymerase (e.g., a reverse transcriptase), as well as other suitable components (e.g., linkers, NLS) and PE fusion proteins, that may be used with the therapeutic PEgRNA disclosed herein.

[0011] As disclosed herein, prime editing is a versatile and precise genome editing method that directly writes new genetic information into a specified DNA site using a nucleic acid programmable DNA binding protein ("napDNAbp") working in association with a polymerase

(i.e., in the form of a fusion protein or otherwise provided *in trans* with the napDNAbp), wherein the prime editing system is programmed with a prime editing (PE) guide RNA ("PEgRNA") that both specifies the target site and templates the synthesis of the desired edit in the form of a replacement DNA strand by way of an extension (either DNA or RNA) engineered onto a guide RNA (e.g., at the 5′ or 3′ end, or at an internal portion of a guide RNA). The replacement strand containing the desired edit (e.g., a single nucleobase substitution) shares the same sequence as the endogenous strand of the target site to be edited (with the exception that it includes the desired edit). Through DNA repair and/or replication machinery, the endogenous strand of the target site is replaced by the newly synthesized replacement strand containing the desired edit. In some cases, prime editing may be thought of as a "search-and-replace" genome editing technology since the prime editors, as described herein, not only search and locate the desired target site to be edited, but at the same time, encode a replacement strand containing a desired edit which is installed in place of the corresponding target site endogenous DNA strand. The prime editors of the present disclosure relate, in part, to the discovery that the mechanism of target-primed reverse transcription (TPRT) or "prime editing" can be leveraged or adapted for conducting precision CRISPR/Cas-based genome editing with high efficiency and genetic flexibility (e.g., as depicted in various embodiments of FIGs. 1A-1F). TPRT is naturally used by mobile DNA elements, such as mammalian non-LTR retrotransposons and bacterial Group II introns[28,29]. The inventors have herein used Cas protein-reverse transcriptase fusions or related systems to target a specific DNA sequence with a guide RNA, generate a single strand nick at the target site, and use the nicked DNA as a primer for reverse transcription of an engineered reverse transcriptase template that is integrated with the guide RNA. However, while the concept begins with prime editors that use reverse trancriptases as the DNA polymerase component, the prime editors described herein are not limited to reverse transcriptases but may include the use of virtually and DNA polymerase. Indeed, while the application throughout may refer to prime editors with "reverse transcriptases," it is set forth here that reverse transcriptases are only one type of DNA polymerase that may work with prime editing. Thus, where ever the specification mentions "reverse transcriptases," the person having ordinary skill in the art should appreciate that any suitable DNA polymerase may be used in place of the reverse transcriptase. Thus, in one example, the prime editors may comprise Cas9 (or an equivalent napDNAbp) which is programmed to target a DNA sequence by associating it with a specialized guide RNA (i.e.,

assimilation of the desired nucleotide change as a result of cellular DNA repair and/or replication processes. Because templated DNA synthesis offers single nucleotide precision for the modification of any nucleotide, including insertions and deletions, the scope of this approach is very broad and could foreseeably be used for myriad applications in basic science and therapeutics.

[0013a] In one aspect, the present disclosure provides a guide RNA comprising a spacer sequence, a gRNA core, and an extension arm, wherein the extension arm comprises (i) a primer binding site, (ii) a reverse transcription (RT) edit template, and (iii) a homology arm;

wherein the spacer sequence comprises from 10 to 40 nucleotides which contain a nucleotide sequence that is complementary to a protospacer sequence in a target strand of a target DNA sequence;

wherein the gRNA core is capable of binding a nucleic acid programmable DNA binding protein (napDNAbp) that is a Cas9 nickase;

wherein the extension arm is capable of serving as a template for reverse transcription synthesis of a single strand of DNA and comprises, in a 5′ to 3′ direction, the homology arm, the RT edit template, and the primer binding site;

wherein the primer binding site is 8-20 nucleotides in length and is capable of annealing to a 3′-ended ssDNA flap that is formed after napDNAbp nicking of a non-target strand of the target DNA sequence, and which serves as a primer sequence to prime reverse transcription of the single strand of DNA;

wherein the RT edit template encodes a desired edit in the single strand of DNA; and

wherein the homology arm encodes a portion of the single strand of DNA that is complementary to the target strand of the target DNA sequence.

[0013b] In another aspect, the present disclosure provides a polynucleotide encoding the guide RNA as defined herein.

[0013c] In another aspect, the present disclosure provides a vector comprising the polynucleotide as defined herein.

[0013d] In another aspect, the present disclosure provides a pharmaceutical composition comprising: (a) the guide RNA as defined herein, the polynucleotide as defined herein, or the vector as defined herein, and (b) a pharmaceutically acceptable excipient, wherein the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the desired edit

in the single strand of DNA corrects one or more mutations associated with the disease or disorder.

[0013e] In another aspect, the present disclosure provides a prime editing complex comprising a napDNAbp, a reverse transcriptase, and the guide RNA as defined herein, wherein the napDNAbp is a Cas9 nickase.

[0013f] In another aspect, the present disclosure provides one or more polynucleotides encoding the prime editing complex as defined herein.

[0013g] In another aspect, the present disclosure provides one or more vectors comprising the one or more polynucleotides as defined herein.

[0013h] In another aspect, the present disclosure provides a pharmaceutical composition comprising the prime editing complex as defined herein, the one or more polynucleotides as defined herein, or the one or more vectors as defined herein, and a pharmaceutically acceptable excipient, wherein the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the desired edit in the single strand of DNA corrects one or more mutations associated with the disease or disorder.

[0013i] In another aspect, the present disclosure provides a composition comprising the guide RNA as defined herein and one or more polynucleotides encoding a prime editor comprising a napDNAbp and a reverse transcriptase, wherein the napDNAbp is a Cas9 nickase.

[0013j] In another aspect, the present disclosure provides a composition comprising the polynucleotide as defined herein or the vector as defined herein, and one or more polynucleotides or vectors encoding a prime editor comprising a napDNAbp and a reverse transcriptase, wherein the napDNAbp is a Cas9 nickase.

[0013k] In another aspect, the present disclosure provides the gRNA as defined herein, the polynucleotide as defined herein, the vector as defined herein, the pharmaceutical composition as defined herein, the prime editing complex as defined herein, the one or more polynucleotides as defined herein, the one or more vectors as defined herein, the pharmaceutical composition as defined herein, or the composition as defined herein for use as a medicament, wherein the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the desired edit in the single strand of DNA corrects one or more mutations associated with the disease or disorder.

Algorithm and methods of designing therapeutic PEgRNA

[0014] In one aspect, the present disclosure relates to a novel algorithm for designing therapeutic PEgRNA, in particular, on a large-scale as opposed to a one-off PEgRNA design exercise.

[0015] Accordingly, some aspects relate to a computerized method for determining a sequence of a prime editor guide RNA (PEgRNA). The method includes using at least one computer hardware processor to access data indicative of an input allele, an output allele, and a fusion protein comprising a nucleic acid programmable DNA binding protein and a polymerase (e.g., a reverse transcriptase). The method includes determining the PEgRNA sequence based on the input allele, the output allele, and the fusion protein, wherein the PEgRNA sequence is designed to be associated with the fusion protein to change the input allele to the output allele, including determining for the PEgRNA sequence one or more of the following features: a spacer complementary to a target nucleotide sequence in the input allele (i.e., the spacer, as defined in FIG. 27); a gRNA backbone for interacting with the fusion protein (i.e., the gRNA core as defined in FIG. 27); and an extension (i.e., the extension arm as shown in FIG. 27) comprising one or more of: a DNA synthesis template(as shown in FIG. 27) comprising a desired nucleotide change to change the input allele to the output allele; primer binding site (i.e., the primer binding site as shown in FIG. 27). The PEgRNA may also comprise a 3' termination signal that terminates transcription from a promoter. In addition, the PEgRNA may include a first modifier at the 5' end of the extension arm and a second modifier at the 3' end of the extension arm. Such sequences (shown as "e1" and "e2" in FIG. 27) may include stem-loop sequences, which may increase the stability of the PEgRNA.

[0016] In some examples, the method includes determining the spacer and the extension, and determining the spacer is at the 5′ end of the PEgRNA , and the extension is at a 3′ end of the PEgRNA structure.

[0017] In some examples, the method includes determining the spacer and the extension, and determining the spacer is at the 5′ end of the PEgRNA , and the extension is 3′ to the spacer.

CLAIMS

1. A guide RNA comprising a spacer sequence, a gRNA core, and an extension arm, wherein the extension arm comprises (i) a primer binding site, (ii) a reverse transcription (RT) edit template, and (iii) a homology arm;

wherein the spacer sequence comprises from 10 to 40 nucleotides which contain a nucleotide sequence that is complementary to a protospacer sequence in a target strand of a target DNA sequence;

wherein the gRNA core is capable of binding a nucleic acid programmable DNA binding protein (napDNAbp) that is a Cas9 nickase;

wherein the extension arm is capable of serving as a template for reverse transcription synthesis of a single strand of DNA and comprises, in a 5′ to 3′ direction, the homology arm, the RT edit template, and the primer binding site;

wherein the primer binding site is 8-20 nucleotides in length and is capable of annealing to a 3′-ended ssDNA flap that is formed after napDNAbp nicking of a non-target strand of the target DNA sequence, and which serves as a primer sequence to prime reverse transcription of the single strand of DNA;

wherein the RT edit template encodes a desired edit in the single strand of DNA; and

wherein the homology arm encodes a portion of the single strand of DNA that is complementary to the target strand of the target DNA sequence.

2. The guide RNA of claim 1, wherein the protospacer sequence comprises about 20 nucleotides.

3. The guide RNA of claim 1 or 2, wherein the extension arm is between 10 and 500 nucleotides in length.

4. The guide RNA of claim 1 or 2, wherein the extension arm is at least 10 nucleotides, at least 11 nucleotides, at least 12 nucleotides, at least 13 nucleotides, at least 14 nucleotides, at least 15 nucleotides, at least 16 nucleotides, at least 17 nucleotides, at least 18 nucleotides, at least 19 nucleotides, at least 20 nucleotides, at least 21 nucleotides, at least 22 nucleotides, at least 23 nucleotides, at least 24 nucleotides, at least 25

nucleotides, at least 26 nucleotides, at least 27 nucleotides, at least 28 nucleotides, at least 29 nucleotides, at least 30 nucleotides, at least 31 nucleotides, at least 32 nucleotides, at least 33 nucleotides, at least 34 nucleotides, at least 35 nucleotides, at least 36 nucleotides, at least 37 nucleotides, at least 38 nucleotides, at least 39 nucleotides, at least 40 nucleotides, or at least 100 nucleotides in length.

5. The guide RNA of any one of claims 1-4, wherein the extension arm is a 5′ extension arm.

6. The guide RNA of any one of claims 1-4, wherein the extension arm is a 3′ extension arm.

7. The guide RNA of any one of claims 1-6, wherein the RT edit template is at least 1 nucleotide, at least 2 nucleotides, at least 3 nucleotides, at least 4 nucleotides, at least 5 nucleotides, at least 6 nucleotides, at least 7 nucleotides, at least 8 nucleotides, at least 9 nucleotides, at least 10 nucleotides, at least 11 nucleotides, at least 12 nucleotides, at least 13 nucleotides, at least 14 nucleotides, at least 15 nucleotides, at least 16 nucleotides, at least 17 nucleotides, at least 18 nucleotides, at least 19 nucleotides, at least 20 nucleotides, at least 21 nucleotides, at least 22 nucleotides, at least 23 nucleotides, at least 24 nucleotides, at least 25 nucleotides, at least 26 nucleotides, at least 27 nucleotides, at least 28 nucleotides, at least 29 nucleotides, at least 30 nucleotides, at least 31 nucleotides, at least 32 nucleotides, at least 33 nucleotides, at least 34 nucleotides, at least 35 nucleotides, at least 36 nucleotides, at least 37 nucleotides, at least 38 nucleotides, at least 39 nucleotides, at least 40 nucleotides, or at least 100 nucleotides in length.

8. The guide RNA of any one of claims 1-7, wherein the desired edit comprises one or more single-base nucleotide changes, one or more deletions, one or more insertions, or a combination thereof.

9. The guide RNA of any one of claims 1-8, wherein the homology arm is at least 1 nucleotide, at least 2 nucleotides, at least 3 nucleotides, at least 4 nucleotides, at least 5 nucleotides, at least 6 nucleotides, at least 7 nucleotides, at least 8 nucleotides, at least 9 nucleotides, at least 10 nucleotides, at least 11 nucleotides, at least 12 nucleotides, at least 13 nucleotides, at least 14 nucleotides, at least 15 nucleotides, at least 16 nucleotides, at

least 17 nucleotides, at least 18 nucleotides, at least 19 nucleotides, at least 20 nucleotides, at least 21 nucleotides, at least 22 nucleotides, at least 23 nucleotides, at least 24 nucleotides, at least 25 nucleotides, at least 26 nucleotides, at least 27 nucleotides, at least 28 nucleotides, at least 29 nucleotides, or at least 30 nucleotides in length.

10. The guide RNA of any one of claims 1-9 further comprising a 5′ end modifier region comprising a hairpin sequence, a stem/loop sequence, or a toeloop sequence.

11. The guide RNA of any one of claims 1-10 further comprising a 3′ end modifier region comprising a hairpin sequence, a stem/loop sequence, or a toeloop sequence.

12. The guide RNA of any one of claims 1-11, wherein the spacer sequence comprises the nucleotide sequence of any one of SEQ ID NOs: 135515-271028 or 880463-947840.

13. The guide RNA of any one of claims 1-12, wherein the gRNA core comprises the nucleotide sequence of SEQ ID NO: 1361579 or 1361580.

14. The guide RNA of any one of claims 1-13, wherein the primer binding site comprises the nucleotide sequence of any one of SEQ ID NOs: 406543-542056 or 1015219-1082596.

15. The guide RNA of any one of claims 1-14, wherein the RT edit template comprises the nucleotide sequence of any one of SEQ ID NOs: 542057-677570 or 1082597-1149974.

16. The guide RNA of any one of claims 1-15, wherein the homology arm comprises the nucleotide sequence of any one of SEQ ID NOs: 677571-813084 or 1149975-1217352.

17. The guide RNA of any one of claims 1-16, wherein the extension arm comprises the nucleotide sequence of any one of SEQ ID NOs: 271029-406542 or 947841-1015218.

18. The guide RNA of any one of claims 1-17 comprising the sequence of any one of

SEQ ID NOs: 1-135514 or 813085-880462.

19. A polynucleotide encoding the guide RNA of any one of claims 1-18.

20. A vector comprising the polynucleotide of claim 19.

21. A pharmaceutical composition comprising: (a) the guide RNA of any one of claims 1-18, the polynucleotide of claim 19, or the vector of claim 20, and (b) a pharmaceutically acceptable excipient, wherein the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the desired edit in the single strand of DNA corrects one or more mutations associated with the disease or disorder.

22. A prime editing complex comprising a napDNAbp, a reverse transcriptase, and the guide RNA of any one of claims 1-18, wherein the napDNAbp is a Cas9 nickase.

23. The prime editing complex of claim 22, wherein the napDNAbp and the reverse transcriptase are formed as a fusion protein.

24. One or more polynucleotides encoding the prime editing complex of claim 22 or 23.

25. One or more vectors comprising the one or more polynucleotides of claim 24.

26. A pharmaceutical composition comprising the prime editing complex of claim 22 or 23, the one or more polynucleotides of claim 24, or the one or more vectors of claim 25 and a pharmaceutically acceptable excipient, wherein the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the desired edit in the single strand of DNA corrects one or more mutations associated with the disease or disorder.

27. A composition comprising the guide RNA of any one of claims 1-18 and one or more polynucleotides encoding a prime editor comprising a napDNAbp and a reverse transcriptase, wherein the napDNAbp is a Cas9 nickase.

28. A composition comprising the polynucleotide of claim 19 or the vector of claim

20, and one or more polynucleotides or vectors encoding a prime editor comprising a napDNAbp and a reverse transcriptase, wherein the napDNAbp is a Cas9 nickase.

29. The composition of claim 27 or 28, wherein the prime editor is a fusion protein.

30. The composition of any one of claims 27-29, wherein the reverse transcriptase is a Moloney Murine Leukemia Virus reverse transcriptase (MMLV-RT) or a variant of MMLV-RT.

31. The composition of claim 30, wherein the variant of MMLV-RT comprises one or more of the following mutations: P51X, S67X, E69X, L139X, T197X, D200X, H204X, F209X, E302X, T306X, F309X, W313X, T330X, L345X, L435X, N454X, D524X, E562X, D583X, H594X, L603X, E607X, and D653X relative to the wild type MMLV-RT of SEQ ID NO: 1361485 or at a corresponding amino acid position in another wild type MMLV-RT polypeptide sequence, and wherein X is any amino acid.

32. The composition of claim 30, wherein the variant of MMLV-RT comprises one or more of the following mutations: P51L, S67K, E69K, L139P, T197A, D200N, H204R, F209N, E302K, E302R, T306K, F309N, W313F, T330P, L345G, L435G, N454K, D524G, E562Q, D583N, H594Q, L603W, E607K, and D653N relative to the wild type MMLV-RT of SEQ ID NO: 1361485 or at a corresponding amino acid position in another wild type MMLV-RT polypeptide sequence.

33. The composition of claim 30, wherein the variant of MMLV-RT comprises amino acid substitutions D200N, T330P, and/or L603W relative to the wild type MMLV-RT of SEQ ID NO: 1361485 or at a corresponding amino acid position in another wild type MMLV-RT polypeptide sequence.

34. The composition of claim 30, wherein the variant of MMLV-RT comprises amino acid substitutions D200N, T306K, W313F, T330P, and L603W relative to the wild type MMLV-RT of SEQ ID NO: 1361485 or at a corresponding amino acid position in another wild type MMLV-RT polypeptide sequence.

35. The composition of claim 30, wherein the variant of MMLV-RT comprises the sequence of SEQ ID NO: 1361514.

36. The composition of claim 29, wherein the fusion protein comprises the sequence of SEQ ID NO: 1361516.

37. The gRNA of any one of claims 1-18, the polynucleotide of claim 19, the vector of claim 20, the pharmaceutical composition of claim 21, the prime editing complex of claim 22 or 23, the one or more polynucleotides of claim 24, the one or more vectors of claim 25, the pharmaceutical composition of claim 26, or the composition of any one of claims 27-36 for use as a medicament, wherein the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the desired edit in the single strand of DNA corrects one or more mutations associated with the disease or disorder.