Claims

1. An altered cytidine deaminase comprising amino acid substitution mutations in a cytidine deaminase at positions functionally equivalent to Tyr130Ala and Tyr132 in a wild-type APOBEC3A protein of SEQ ID NO:3, wherein the altered cytidine deaminase converts 5-methyl cytosine (5mC) to thymidine (T) by deamination at a greater rate than conversion of cytosine (C) to uracil (U) by deamination.

2. The altered cytidine deaminase of claim 1, wherein the substitution mutation at the position functionally equivalent to Tyr132 comprises a mutation to His, Arg, Gln, or Lys.

3. The altered cytidine deaminase of claim 1, wherein the rate is at least 10-fold greater for converting 5mC to T by deamination than conversion of C to U.

4. The altered cytidine deaminase of claim 1, wherein the altered cytidine deaminase is member of the AID subfamily, the APOBEC1 subfamily, the APOBEC2 subfamily, the APOBEC3A subfamily, the APOBEC3B subfamily, the APOBEC3C subfamily, the APOBEC3D subfamily, the APOBEC3F subfamily, the APOBEC3G subfamily, the APOBEC3H subfamily, or the APOBEC4 subfamily.

5. The altered cytidine deaminase of claim 1, wherein the altered cytidine deaminase comprises a ZDD motif H- [P/A/V]-E-$X_{[23-28]}$-P-C-$X_{[2-4]}$-C (SEQ ID NO:12).

6. The altered cytidine deaminase of claim 1, wherein the altered cytidine deaminase is a member of the APOBEC3A subfamily and comprises a ZDD motif HXEX_{24}SW(S/T)PC$X_{[2-4]}$CX_6FX_8LX_5R(L/I)Y$X_{[8-11]}$LX_2L$X_{[10]}$M (SEQ ID NO:13), wherein the amino acid substitution mutation at the position functionally equivalent to (Tyr/Phe)130 of the wild-type APOBEC3A protein is the Tyr (Y) amino acid of the ZDD motif.

7. The altered cytidine deaminase of claim 1, wherein the altered cytidine deaminase is a member of the APOBEC3A subfamily and comprises $X_{[16-26]}$-GRXXTXLCYXV-X_{15}-GXXXN-

X_{12}-HAEXXF-X_{14}-YXXTWXXSWSPC- $X_{[2-4]}$-CA-X_5-FL-X_7-LXIXXXR(L/I)Y-X_8-GLXXLXXXG-X_5-M-X_4-FXXCWXXFV-X_6-FXPW-X_{13}-LXXI- $X_{[2-6]}$ (SEQ ID NO:14).

8. The altered cytidine deaminase of claim 1, wherein the altered cytidine deaminase is a member of the APOBEC3A family and comprises SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:59.

9. A polynucleotide encoding the altered cytidine deaminase of any one of the preceding claims.

10. A composition comprising the altered cytidine deaminase of any one of claims 1-8 and a buffer.

11. The composition of claim 10, wherein the composition further comprises:

at least one of

(i) a sample comprising DNA comprising at least one modified cytosine, wherein the modified cytosine is 5-methyl cytosine (5mC); or

(ii) a buffer having a pH that is lower than 7; or

(iii) combinations thereof.

12. The composition of claim 11, wherein the DNA comprises single-stranded DNA.

13. The composition of claim 11, wherein the sample comprises genomic DNA or cell free DNA.

14. The composition of claim 13, wherein the genomic DNA is from a single cell or is a mixture from a plurality of cells.

15. A method comprising:

providing a sample of DNA suspected of comprising single-stranded DNA comprising at least one 5-methyl cytosine (5mC);

contacting the single-stranded DNA with an altered cytidine deaminase of any one of claims 1-8 under conditions suitable for conversion of 5-methylcytosine (5mC) to thymidine (T) by deamination at a greater rate than conversion of cytosine (C) to uracil (U) by deamination, to result in converted single-stranded DNA,

; and

processing the converted single-stranded DNA to produce a sequencing library.

16. The method of claim 15, wherein the method further comprises, before the providing, denaturing double-stranded DNA present in the sample to result in single-stranded DNA.

17. A method comprising:

providing a sample of DNA suspected of comprising double-stranded DNA comprising at least one 5-methyl cytosine (5mC);

processing the double-stranded DNA to produce a sequencing library;

denaturing the sequencing library to result in a single-stranded DNA;

contacting the single-stranded DNA with an altered cytidine deaminase of any one of claims 1-8 under conditions suitable for conversion of 5-methylcytosine (5mC) to thymidine (T) by deamination at a greater rate than conversion of cytosine (C) to uracil (U) by deamination, ; and

converting the converted single-stranded DNA to a converted double-stranded DNA sequencing library.

18. The method of claim 15, wherein the processing comprises converting the converted single-stranded DNA to double-stranded DNA before producing the sequencing library.

19. The method of claim 17 or 18, wherein the processing comprises fragmentation or tagmentation of the double-stranded DNA and addition of a universal sequence to the double-stranded DNA fragments.

20. The method of claim 15 or 17, wherein the converting comprises amplifying the converted single-stranded DNA to be the converted double-stranded DNA.

21. The method of claim 15 or 17, wherein the sample is a biological sample.

22. The method of claim 21, wherein the biological sample comprises cell-free DNA.

23. The method of claim 21, wherein the biological sample comprises a fluid selected from blood or serum.

24. The method of claim 21, wherein the sample comprises single cells or isolated nuclei.

25. The method of claim 21, wherein the biological sample comprises a tissue.

26. The method of claim 25, wherein the tissue comprises tumor tissue.

27. The method of claim 15, further comprising:

providing a surface comprising a plurality of amplification sites,

wherein the amplification sites comprise at least two populations of attached single-stranded capture oligonucleotides having a free 3' end, and

contacting the surface comprising amplification sites with the sequencing library under conditions suitable to produce a plurality of amplification sites that each comprise a clonal population of amplicons from an individual member of the sequencing library.

28. A method of detecting the location of a 5-methyl cytosine (5mC) in a target nucleic acid, the method comprising:

(a) contacting target nucleic acids suspected of comprising at least one 5mC with the altered cytidine deaminase of any one of claims 1-8 to produce converted nucleic acids comprising at least one converted 5mC;

(b) detecting the at least one converted cytosine in the converted nucleic acids of (a).

29. The method of claim 28, wherein the altered cytidine deaminase has cytosine-defective deaminase activity, and wherein the detecting comprises identifying thymidine nucleotides in the converted nucleic acid to determine the location of 5mC nucleotides in the target nucleic acid.

30. The method of claim 28, wherein the detecting comprises sequencing the converted nucleic acids or hybridizing nucleic acid probes to the converted nucleic acids.

31. The method of claim 30, wherein the detecting comprises sequencing the converted nucleic acids, the method further comprising:

(c) comparing the sequence of the converted nucleic acids with an untreated reference sequence to determine which cytosines in the target nucleic acids are modified.

32. The method of claim 31, wherein a predetermined sequence of the untreated reference sequence and a predetermined sequence of the converted nucleic acid are compared.

33. The method of claim 32, wherein the predetermined sequence comprises a CpG island or a promoter.

34. The method of claim 30, wherein the detecting comprises hybridizing the converted nucleic acids to the nucleic acid probes, optionally wherein the nucleic acid probes are present on an analyte array, the method further comprising sequencing the hybridized converted nucleic acids.

35. The method of claim 30, wherein the detecting comprises hybridizing the converted nucleic acids to the nucleic acid probes, the method further comprising (i) amplifying the converted nucleic acid, wherein the nucleic acid probes comprise two primers for amplification of a predetermined sequence, wherein the primers anneal to regions of converted nucleic acids comprising at least one converted cytosine with a greater affinity than to the regions of converted nucleic acids wherein at least one cytosine is not a converted cytosine, wherein the presence of an amplified product is indicative of a modified cytosine in the target nucleic acid, or (ii) cleaving a single stranded DNA (ssDNA) reporter substrate by a CRISPR-based system, wherein the ssDNA reporter substrate comprises a fluorophore and a quencher, wherein the presence of fluorescence is indicative of a modified cytosine in the target nucleic acid.

36. The method of claim 35, wherein the CRISPR-based system comprises a guide RNA sequence that anneals to a predetermined sequence of a nucleic acid comprising at least one converted cytosine and anneals at lower affinity to the predetermined sequence of the nucleic acid when at least one cytosine is not a converted cytosine.

37. The method of claim 30, wherein the detecting comprises hybridizing the converted nucleic acids to the nucleic acid probes, wherein the converted nucleic acids are present in a fixed cell, wherein the nucleic acid probes comprise a fluorescent labeled probe, and wherein the nucleic acid probes anneal to a predetermined sequence of converted nucleic acids comprising at least one converted cytosine with a greater affinity than to the regions of converted nucleic acids wherein at least one cytosine is not a converted cytosine, wherein the presence of cell-associated fluorescence is indicative of a modified cytosine in the target nucleic acid.

38. The method of claim 37, wherein the fixed cell is present in formalin-fixed paraffin-embedded (FFPE) sample.

39. The method of claim 31, wherein the untreated reference sequence is a predetermined sequence.

40. The method of claim 28, further comprising providing the target nucleic acids, wherein the providing comprises preparing single stranded (ss) DNA from a sample comprising the target nucleic acids.

41. The method of claim 28, wherein the target nucleic acids are genomic DNA or cell free DNA.

42. The method of claim 41, wherein the genomic DNA is from a single cell or is a mixture from a plurality of cells.

43. The method of claim 30, wherein the sequencing comprises processing the converted nucleic acids to produce a sequencing library.

44. The method of claim 28 or 43, further comprising:

providing a surface comprising a plurality of amplification sites,

wherein the amplification sites comprise at least two populations of attached single-stranded capture oligonucleotides having a free 3' end, and

contacting the surface comprising amplification sites with the sequencing library under conditions suitable to produce a plurality of amplification sites that each comprise a clonal population of amplicons from an individual member of the sequencing library.

45. The method of claim 28, wherein the target nucleic acids are obtained from a subject, wherein the detecting comprises obtaining a pattern of cytosine modification in the converted nucleic acids, the method further comprising comparing the pattern of cytosine modification in the converted nucleic acids with the pattern of cytosine modification in a reference nucleic acid.

46. The method of claim 45, wherein the subject has or is at risk of having a disease or condition, wherein the reference nucleic acid is from a normal subject.

47. The method of claim 46, wherein the pattern of cytosine modification is linked in-*cis* to a coding region that is correlated with a disease or condition.

48. The method of claim 45, wherein the pattern of cytosine modification is linked in-*cis* to a coding region, wherein the coding region in the reference nucleic acid is transcriptionally active or transcriptionally inactive, wherein the comparing further comprises determining if the pattern of cytosine modification of the converted nucleic acid indicates the coding region is transcriptionally active or transcriptionally inactive in the subject.

49. The method of claim 48, wherein transcription of the coding region is correlated with a disease or condition.

50. The method of claim 46, wherein the subject has the disease or condition and is undergoing treatment for the disease or condition, the method further comprising determining if the treatment is correlated with a change in the pattern of cytosine modification in the subject.

51. The method of claim 50, wherein the subject previously had the disease or condition, the method further comprising comparing the pattern of cytosine modification in the subject with the pattern of cytosine modification in the subject when the subject had the disease or condition.