CLAIMS

1. A method for determining a sequence of at least a portion of at least one target template nucleic acid molecule comprising:

(a) providing a pair of samples, each sample comprising at least one target template nucleic acid molecule;

(b) sequencing regions of the at least one target template nucleic acid molecule in a first of the pair of samples to provide non-mutated sequence reads;

(c) introducing mutations into the at least one target template nucleic acid molecule in a second of the pair of samples to provide at least one mutated target template nucleic acid molecule;

(d) sequencing regions of the at least one mutated target template nucleic acid molecule to provide mutated sequence reads;

(e) analysing the mutated sequence reads, and using information obtained from analysing the mutated sequence reads to determine the sequence for at least a portion of at least one target template nucleic acid molecule from the non-mutated sequence reads.

2. A method for generating a sequence of at least a portion of at least one target template nucleic acid molecule comprising:

(a) obtaining data comprising:

(i) non-mutated sequence reads; and

(ii) mutated sequence reads; and

(b) analysing the mutated sequence reads, and using information obtained from analysing the mutated sequence reads to generate a sequence for at least a portion of at least one target template nucleic acid molecule from the non-mutated sequence reads.

3. A method according to Claim 1 or 2, in which the step of analysing the mutated sequence reads, and using information obtained from analysing the mutated sequence reads to determine or generate a sequence for at least a portion of at least one target template nucleic acid molecule from the non-mutated sequence reads comprises preparing an assembly graph.

4. A method according to Claim 3, in which the assembly graph comprises nodes computed from non-mutated sequence reads, and each valid route through the assembly graph

comprising the nodes represents the sequence of at least a portion of the at least one target template nucleic acid molecule.

5. A method according to Claim 4, in which a sequence is determined or generated for at least a portion of at least one target template nucleic acid molecule from nodes that form part of a valid route through the assembly graph.

6. A method according to Claim 1, in which analysing the mutated sequence reads comprises identifying mutated sequence reads that are likely to have originated from the same at least one mutated target template nucleic acid molecule.

7. A method according to Claim 5, in which identifying nodes that form part of a valid route through the assembly graph using information obtained by analysing the mutated sequence reads comprises:

(i) computing nodes from non-mutated sequence reads;

(ii) mapping the mutated sequence reads to the assembly graph;

(iii) identifying mutated sequence reads that are likely to have originated from the same at least one mutated target template nucleic acid molecule; and

(iv) identifying nodes that are linked by mutated sequence reads that are likely to have originated from the same at least one mutated target template nucleic acid molecule,

in which nodes that are linked by mutated sequence reads are likely to have originated from the same at least one mutated target template nucleic acid molecule and form part of a valid route through the assembly graph.

8. A method according to Claim 6, in which mutated sequence reads are likely to have originated from the same mutated target template nucleic acid molecule if the odds ratio probability that the mutated sequence reads originated from the same mutated target template nucleic acid molecule: probability that the mutated sequence reads did not originate from the same mutated target template nucleic acid molecule exceeds a threshold.

9. A method according to Claim 6, in which identifying mutated sequence reads that are likely to have originated from the same mutated target template nucleic acid molecule comprises using a probability function based on the following parameters:

e. a matrix (N) of nucleotides in each position of the mutated sequence reads and the assembly graph;

f. a probability (M) that a given nucleotide (i) was mutated to read nucleotide (j);

g. a probability (E) that a given nucleotide (i) was read erroneously to read nucleotide (j) conditioned on the nucleotide having been read erroneously; and

h. a probability (Q) that a nucleotide in position Y was read erroneously.

10. A method according to Claim 6, in which identifying mutated sequence reads that are likely to have originated from the same mutated target template nucleic acid molecule comprises using machine learning or neural networks to identify the mutated sequence reads.

11. A method according to Claim 6, in which the method comprises a pre-clustering step.

12. A method according to Claim 1 or 2, in which the method comprises sequencing the ends of the at least one target template nucleic acid molecule using paired-end sequencing.

13. A method according to Claim 1 or 2, in which the method comprises mapping the sequences of the ends of the at least one target template nucleic acid molecule to an assembly graph.

14. A method according to Claim 1 or 2, in which the at least one target template nucleic acid molecule comprises a barcode at each end.

15. A method according to Claim 1 or 2, in which the at least one mutated target template nucleic acid molecule comprises between 1% and 50%, between 3% and 25%, between 5% and 20%, or around 8% mutations.

16. A method according to Claim 1 or 2, in which the at least one mutated target template nucleic acid molecule comprises unevenly distributed mutations.

17. A method according to Claim 1 or 2, in which the mutated sequence reads and/or the non-mutated sequence reads comprise sequencing errors that are unevenly distributed.

18. A method according to Claim 1 or 2, in which the step of introducing mutations into the at least one mutated target template nucleic acid molecule introduces mutations that are unevenly distributed.

19. A computer program adapted to perform a method according to Claim 1 or 2.

20. A computer readable medium comprising a computer program according to Claim 19.

21. A method for determining a sequence of at least one target template nucleic acid molecule comprising:

(a) providing at least one sample comprising the at least one target template nucleic acid molecule;

(b) sequencing regions of the at least one target template nucleic acid molecule; and

(c) assembling a sequence of the at least one target template nucleic acid molecule from the sequences of the regions of the at least one target template nucleic acid molecule,

in which:

(i) the step of providing at least one sample comprising the at least one target template nucleic acid molecule comprises controlling the number of target template nucleic acid molecules in the at least one sample.

22. A method according to Claim 21, in which controlling the number of target template nucleic acid molecules comprises measuring the number of target template nucleic acid molecules in the first of the pair of samples, the second of the pair of samples, or the at least one sample.

23. A method according to Claim 22, in which measuring the number of target template nucleic acid molecules comprises preparing a dilution series of the first of the pair of samples, the second of the pair of samples, or the at least one sample to provide a dilution series comprising diluted samples.

CLAIMS

1. A method for determining a sequence of at least a portion of at least one target template nucleic acid molecule comprising:

(a) providing a pair of samples, each sample comprising at least one target template nucleic acid molecule;

(b) sequencing regions of the at least one target template nucleic acid molecule in a first of the pair of samples to provide non-mutated sequence reads;

(c) introducing mutations into the at least one target template nucleic acid molecule in a second of the pair of samples to provide at least one mutated target template nucleic acid molecule;

(d) sequencing regions of the at least one mutated target template nucleic acid molecule to provide mutated sequence reads;

(e) analysing the mutated sequence reads, and using information obtained from analysing the mutated sequence reads to determine the sequence for at least a portion of at least one target template nucleic acid molecule from the non-mutated sequence reads.

2. A method for generating a sequence of at least a portion of at least one target template nucleic acid molecule comprising:

(a) obtaining data comprising:

(i) non-mutated sequence reads; and

(ii) mutated sequence reads; and

(b) analysing the mutated sequence reads, and using information obtained from analysing the mutated sequence reads to generate a sequence for at least a portion of at least one target template nucleic acid molecule from the non-mutated sequence reads.

3. A method according to Claim 1 or 2, in which the step of analysing the mutated sequence reads, and using information obtained from analysing the mutated sequence reads to determine or generate a sequence for at least a portion of at least one target template nucleic acid molecule from the non-mutated sequence reads comprises preparing an assembly graph.

4. A method according to Claim 3, in which the assembly graph comprises nodes computed from non-mutated sequence reads, and each valid route through the assembly graph

comprising the nodes represents the sequence of at least a portion of the at least one target template nucleic acid molecule.

5. A method according to Claim 4, in which a sequence is determined or generated for at least a portion of at least one target template nucleic acid molecule from nodes that form part of a valid route through the assembly graph.

6. A method according to Claim 1, in which analysing the mutated sequence reads comprises identifying mutated sequence reads that are likely to have originated from the same at least one mutated target template nucleic acid molecule.

7. A method according to Claim 5, in which identifying nodes that form part of a valid route through the assembly graph using information obtained by analysing the mutated sequence reads comprises:

(i) computing nodes from non-mutated sequence reads;

(ii) mapping the mutated sequence reads to the assembly graph;

(iii) identifying mutated sequence reads that are likely to have originated from the same at least one mutated target template nucleic acid molecule; and

(iv) identifying nodes that are linked by mutated sequence reads that are likely to have originated from the same at least one mutated target template nucleic acid molecule,

in which nodes that are linked by mutated sequence reads are likely to have originated from the same at least one mutated target template nucleic acid molecule and form part of a valid route through the assembly graph.

8. A method according to Claim 6, in which mutated sequence reads are likely to have originated from the same mutated target template nucleic acid molecule if the odds ratio probability that the mutated sequence reads originated from the same mutated target template nucleic acid molecule: probability that the mutated sequence reads did not originate from the same mutated target template nucleic acid molecule exceeds a threshold.

9. A method according to Claim 6, in which identifying mutated sequence reads that are likely to have originated from the same mutated target template nucleic acid molecule comprises using a probability function based on the following parameters:

e. a matrix (N) of nucleotides in each position of the mutated sequence reads and the assembly graph;

f. a probability (M) that a given nucleotide (i) was mutated to read nucleotide (j);

g. a probability (E) that a given nucleotide (i) was read erroneously to read nucleotide (j) conditioned on the nucleotide having been read erroneously; and

h. a probability (Q) that a nucleotide in position Y was read erroneously.

10. A method according to Claim 6, in which identifying mutated sequence reads that are likely to have originated from the same mutated target template nucleic acid molecule comprises using machine learning or neural networks to identify the mutated sequence reads.

11. A method according to Claim 6, in which the method comprises a pre-clustering step.

12. A method according to Claim 1 or 2, in which the method comprises sequencing the ends of the at least one target template nucleic acid molecule using paired-end sequencing.

13. A method according to Claim 1 or 2, in which the method comprises mapping the sequences of the ends of the at least one target template nucleic acid molecule to an assembly graph.

14. A method according to Claim 1 or 2, in which the at least one target template nucleic acid molecule comprises a barcode at each end.

15. A method according to Claim 1 or 2, in which the at least one mutated target template nucleic acid molecule comprises between 1% and 50%, between 3% and 25%, between 5% and 20%, or around 8% mutations.

16. A method according to Claim 1 or 2, in which the at least one mutated target template nucleic acid molecule comprises unevenly distributed mutations.

17. A method according to Claim 1 or 2, in which the mutated sequence reads and/or the non-mutated sequence reads comprise sequencing errors that are unevenly distributed.

18. A method according to Claim 1 or 2, in which the step of introducing mutations into the at least one mutated target template nucleic acid molecule introduces mutations that are unevenly distributed.

19. A computer program adapted to perform a method according to Claim 1 or 2.

20. A computer readable medium comprising a computer program according to Claim 19.