

[0036] In some embodiments, the amount of genetic variation, such as polymorphisms or causal variants is analyzed. In some embodiments, the presence or absence of genetic alterations is detected.

[0037] The disclosure also provides for a method for detecting a rare mutation in a cell-free or a substantially cell free sample obtained from a subject comprising: a) sequencing extracellular polynucleotides from a bodily sample from a subject, wherein each of the extracellular polynucleotides generate a plurality of sequencing reads; b) filtering out reads that fail to meet a set threshold; c) mapping sequence reads derived from the sequencing onto a reference sequence; d) identifying a subset of mapped sequence reads that align with a variant of the reference sequence at each mappable base position; e) for each mappable base position, calculating a ratio of (a) a number of mapped sequence reads that include a variant as compared to the reference sequence, to (b) a number of total sequence reads for each mappable base position; f) normalizing the ratios or frequency of variance for each mappable base position and determining potential rare variant(s) or other genetic alteration(s); and g) comparing the resulting number for each of the regions

[0038] This disclosure also provides for a method comprising: a. providing at least one set of tagged parent polynucleotides, and for each set of tagged parent polynucleotides; b. amplifying the tagged parent polynucleotides in the set to produce a corresponding set of amplified progeny polynucleotides; c. sequencing a subset (including a proper subset) of the set of amplified progeny polynucleotides, to produce a set of sequencing reads; and d. collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides. In certain embodiments the method further comprises: e. analyzing the set of consensus sequences for each set of tagged parent molecules.

[0038a] This disclosure also provides for a method comprising: (a) providing at least one set of tagged parent polynucleotides by converting initial starting genetic material into the tagged parent polynucleotides, wherein the initial starting genetic material comprises between 100 and 100,000 haploid human genome equivalents of cell-free DNA polynucleotides, and the tagging is by enzymatic ligation with between 2 and 1,000,000 unique identifiers; (b) amplifying the tagged parent polynucleotides in the set to produce a corresponding set of amplified progeny polynucleotides; (c) sequencing a subset of the set of amplified progeny polynucleotides, to produce a set of sequencing reads; and (d) collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides,

wherein the generation of consensus sequences is based on information from the tag and at least one of (i) sequence information at the beginning region of the sequence read (ii) sequence information at the end region of the sequence read, and (iii) the length of the sequence read.

[0039] In some embodiments each polynucleotide in a set is mappable to a reference sequence.

[0040] In some embodiments the method comprises providing a plurality of sets of tagged parent polynucleotides, wherein each set is mappable to a different reference sequence.

[0041] In some embodiments the method further comprises converting initial starting genetic material into the tagged parent polynucleotides.

[0042] In some embodiments the initial starting genetic material comprises no more than 100 ng of polynucleotides.

CLAIMS

WHAT IS CLAIMED IS:

1. A method comprising:
 - (a) providing at least one set of tagged parent polynucleotides by converting initial starting genetic material into the tagged parent polynucleotides, wherein the initial starting genetic material comprises between 100 and 100,000 haploid human genome equivalents of cell-free DNA polynucleotides, and the tagging is by enzymatic ligation with between 2 and 1,000,000 unique identifiers;
 - (b) amplifying the tagged parent polynucleotides in the set to produce a corresponding set of amplified progeny polynucleotides;
 - (c) sequencing a subset of the set of amplified progeny polynucleotides, to produce a set of sequencing reads; and
 - (d) collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides, wherein the generation of consensus sequences is based on information from the tag and at least one of (i) sequence information at the beginning region of the sequence read (ii) sequence information at the end region of the sequence read, and (iii) the length of the sequence read.
2. The method of claim 1 wherein each polynucleotide in a set is mappable to a reference sequence.
3. The method of claim 1, comprising providing a plurality of sets of tagged parent polynucleotides, wherein each set is mappable to a different mappable position in the reference sequence.
4. The method of claim 1 further comprising: (e) analyzing the set of consensus sequences for each set of tagged parent polynucleotides separately or in combination.
5. The method of any preceding claim, wherein the initial starting genetic material comprises no more than 100 ng of polynucleotides.
6. The method of any preceding claim, comprising converting the initial starting genetic material into tagged parent polynucleotides with a conversion efficiency of at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 80% or at least 90%.

7. The method of any preceding claim, wherein converting comprises any of blunt-end ligation, sticky end ligation and single strand ligation.
8. The method of claim 2, wherein a plurality of the sets map to different mappable positions in a reference sequence from the same genome.
9. The method of claim 4, comprising providing a plurality of sets of tagged parent polynucleotides, wherein each set is mappable to a different mappable position in a reference sequence and wherein the mappable position in the reference sequence is the locus of a tumor marker and analyzing comprises detecting the tumor marker in the set of consensus sequences.
10. The method of claim 4, comprising providing a plurality of sets of tagged parent polynucleotides, wherein each set is mappable to a different mappable position in a reference sequence and wherein the mappable position of the reference sequence comprise a plurality of mappable positions in the reference sequence, each of which mappable position is the locus of a tumor marker.
11. The method of claim 4, comprising providing a plurality of sets of tagged parent polynucleotides, wherein each set is mappable to a different mappable position in a reference sequence and wherein analyzing comprises (i) detecting copy number variation of consensus sequences between at least two sets of parent polynucleotides; (ii) detecting the presence of sequence variations compared with the reference sequences; or (iii) detecting the presence of sequence variations compared with the reference sequences and detecting copy number variation of consensus sequences between at least two sets of parent polynucleotides.
12. The method of any preceding claim, wherein the subset of the set of amplified progeny polynucleotides sequenced is of sufficient size so that any nucleotide sequence represented in the set of tagged parent polynucleotides at a percentage that is the same as the percentage per-base sequencing error rate of the sequencing platform used, has at least a 50%, at least a 60%, at least a 70%, at least a 80%, at least a 90% at least a 95%, at least a 98%, at least a 99%, at least a 99.9% or at least a 99.99% chance of being represented among the set of consensus sequences.
13. The method of any preceding claim, comprising enriching the set of amplified progeny polynucleotides for polynucleotides mapping to one or more selected mappable positions in a reference sequence by: (i) selective amplification of sequences from initial starting genetic material converted to tagged parent polynucleotides; (ii) selective amplification of tagged parent polynucleotides; (iii) selective sequence capture of amplified progeny polynucleotides; or (iv) selective sequence capture of initial starting genetic material.

14. The method of claim 4, wherein analyzing comprises detecting mutations, rare mutations, indels, copy number variations, transversions, translocations, inversion, deletions, aneuploidy, partial aneuploidy, polyploidy, chromosomal instability, chromosomal structure alterations, gene fusions, chromosome fusions, gene truncations, gene amplification, gene duplications, chromosomal lesions, DNA lesions, abnormal changes in nucleic acid chemical modifications, abnormal changes in epigenetic patterns, abnormal changes in nucleic acid methylation infection or cancer.

15. The method of claim 4, wherein analysis further comprises detection and monitoring of an abnormality or disease within an individual, such as, infection and/or cancer; optionally wherein the method is performed in combination with immune repertoire profiling.

16. The method of any preceding claim, wherein the polynucleotides are extracted from a sample selected from the group consisting of blood, plasma, serum, urine, saliva, mucosal excretions, sputum, stool, and tears.

17. The method of any preceding claim, comprising filtering out reads with an accuracy or quality score of less than a threshold.

18. The method of any preceding claim, wherein the number of unique identifiers is:

- i) at least 3 and at most 100;
- ii) at least 5 and at most 100;
- iii) at least 10 and at most 100;
- iv) at least 15 and at most 100;
- v) at least 25 and at most 100;
- vi) at least 3 and at most 1000;
- vii) at least 5 and at most 1000;
- viii) at least 10 and at most 1000;
- ix) at least 15 and at most 1000;
- x) at least 25 and at most 1000;
- xi) at least 3 and at most 10,000;
- xii) at least 5 and at most 10,000;
- xiii) at least 10 and at most 10,000;
- xiv) at least 15 and at most 10,000; or
- xv) at least 25 and at most 10,000.