

BRIEF SUMMARY

[0005a] In one aspect, the present invention provides a method for identifying somatic mutations in a human subject by analyzing a biological sample of the human subject, the biological sample including DNA fragments originating from normal cells and potentially from tumor cells or cells associated with cancer, the biological sample including cell-free DNA fragments, the method comprising:

obtaining template DNA fragments from the biological sample to be analyzed, the template DNA fragments including cell-free DNA fragments;

preparing a sequencing library of analyzable DNA molecules using the template DNA fragments, the preparation of the sequencing library of analyzable DNA molecules not including a step of DNA amplification of the template DNA fragments;

sequencing the sequencing library of analyzable DNA molecules to obtain a plurality of sequence reads;

receiving, at a computer system, the plurality of sequence reads;

aligning, by the computer system, the plurality of sequence reads to a reference human genome to determine genomic positions for the plurality of sequence reads;

obtaining, by the computer system, information about a consensus sequence corresponding to the human subject; and

comparing, by the computer system, the sequence reads to the consensus sequence to identify a filtered set of loci as having somatic mutations in some tissue of the human subject, wherein:

at each locus of the filtered set, a number of the sequence reads having a sequence variant relative to the consensus sequence is above a cutoff value, the cutoff value being greater than one.

[0005b] In another aspect, the present invention provides a method for identifying somatic mutations in a human subject by analyzing a biological sample of the human subject, the biological sample including DNA fragments originating from normal cells and potentially from tumor cells or cells associated with cancer, the biological sample including cell-free DNA fragments, the method comprising:

obtaining template DNA fragments from the biological sample to be analyzed, the template DNA fragments including cell-free DNA fragments;

preparing a sequencing library of analyzable DNA molecules using the template DNA fragments, wherein a duplication rate of the sequencing library from the template DNA fragments is less than 5%;

sequencing the sequencing library of analyzable DNA molecules to obtain a plurality of sequence reads;

receiving, at a computer system, the plurality of sequence reads;

aligning, by the computer system, the plurality of sequence reads to a reference human genome to determine genomic positions for the plurality of sequence reads;

obtaining, by the computer system, information about a consensus sequence corresponding to the human subject; and

comparing, by the computer system, the sequence reads to the consensus sequence to identify a filtered set of loci as having somatic mutations in some tissue of the human subject, wherein:

at each locus of the filtered set, a number of the sequence reads having a sequence variant relative to the consensus sequence is above a cutoff value, the cutoff value being greater than one.

[0005c] In yet another aspect, the present invention provides a method for identifying somatic mutations in a human subject by analyzing a biological sample of the human subject, the biological sample including DNA fragments originating from normal cells and potentially from tumor cells or cells associated with cancer, the biological sample including cell-free DNA fragments, the method comprising, performing, by a computer system:

obtaining information about a consensus sequence corresponding to the human subject;

receiving one or more sequence reads for each of a plurality of DNA fragments in the biological sample;

aligning the plurality of sequence reads to a reference human genome using a first alignment procedure to determine genomic positions for the plurality of sequence reads;

comparing the sequence reads to the consensus sequence to identify a filtered set of loci as having somatic mutations in some tissue of the human subject, wherein:

at each locus of the filtered set, a number of the sequence reads having a sequence variant relative to the consensus sequence is above a cutoff value, the cutoff value being greater than one;

for each of a first set of candidate loci identified as potentially having a somatic mutation:

for each of the sequence reads aligning to the candidate locus using the first alignment procedure and having the sequence variant:

5 determining whether the sequence read aligns to the candidate locus using a second alignment procedure that uses a different matching algorithm than that used for the first alignment procedure;

 comparing the mapping quality to a quality threshold; and

 determining whether to discard the sequence read based on the

10 comparing of the mapping quality to the quality threshold, wherein the mapping quality being less than the quality threshold provides a higher likelihood of discarding the sequence read than the mapping quality being greater than the quality threshold, thereby obtaining a number of remaining sequence reads;

 comparing the number of remaining sequence reads to a candidate

15 threshold; and

 determining whether to discard the candidate locus based on the comparing of the number of remaining sequence reads to the candidate threshold, wherein the number of remaining sequence reads being less than the candidate threshold provides a higher likelihood of discarding the candidate locus than the number of remaining

20 sequence reads being greater than the candidate threshold; and

identifying the filtered set of loci as having somatic mutations using the remaining candidate loci.

[0005d] In yet another aspect, the present invention provides a method for identifying somatic mutations in a human subject by analyzing a biological sample of the human subject,

25 the biological sample including DNA fragments originating from normal cells and potentially from tumor cells or cells associated with cancer, the biological sample including cell-free DNA fragments, the method comprising, performing, by a computer system:

 obtaining information about a consensus sequence corresponding to the human subject; and

30 receiving one or more sequence reads for each of a plurality of DNA fragments in the biological sample;

 aligning the plurality of sequence reads to a reference human genome using a first alignment procedure to determine genomic positions for the plurality of sequence reads;

comparing the sequence reads to the consensus sequence to identify a filtered set of loci as having somatic mutations in some tissue of the human subject, wherein:

at each locus of the filtered set, a number of the sequence reads having a sequence variant relative to the consensus sequence is above a cutoff value, the cutoff

5 value being greater than one;

for each of a first set of candidate loci identified as potentially having a somatic mutation:

determining a size difference between a first group of DNA fragments having the sequence variant and a second group of DNA fragments having a wildtype

10 allele;

comparing the size difference to a size threshold;

when the size difference is less than the size threshold, discarding the candidate locus as a potential mutation; and

identifying the filtered set of loci as having somatic mutations in the human subject using the

15 remaining candidate loci.

[0005e] In yet another aspect, the present invention provides a method for identifying somatic mutations in a human subject by analyzing a biological sample of the human subject, the biological sample including DNA fragments originating from normal cells and potentially from tumor cells or cells associated with cancer, the biological sample including cell-free

20 DNA fragments, the method comprising, performing, by a computer system:

obtaining information about a consensus sequence corresponding to the human subject; and

receiving one or more sequence reads for each of a plurality of DNA fragments in the biological sample;

25 aligning the plurality of sequence reads to a reference human genome using a first alignment procedure to determine genomic positions for the plurality of sequence reads;

comparing the sequence reads to the consensus sequence to identify a filtered set of loci as having somatic mutations in some tissue of the human subject, wherein:

at each locus of the filtered set, a number of the sequence reads having a

30

sequence variant relative to the consensus sequence is above a cutoff value, the cutoff value being greater than one;

identifying a group of regions known to be associated with histone modifications that are associated with cancer;

for each of a first set of candidate loci identified as potentially having a somatic mutation:

determining whether the candidate locus is in one of the group of regions;

determining whether to discard the candidate locus based on whether the

5 candidate locus is in one of the group of regions, wherein the candidate locus not being in one of the group of regions provides a higher likelihood of discarding the candidate locus than when the candidate locus is in one of the group of regions;

identifying the filtered set of loci as having somatic mutations using the remaining candidate loci.

10 **[0005f]** In yet another aspect, the present invention provides a method for identifying *de novo* mutations of a fetus by analyzing a biological sample of a female subject pregnant with the fetus, the biological sample including cell-free DNA fragments from the fetus and the female subject, the method comprising:

obtaining template DNA fragments from the biological sample to be analyzed,

15 the template DNA fragments including cell-free DNA fragments;

preparing a sequencing library of analyzable DNA molecules using the template DNA fragments, the preparation of the sequencing library of analyzable DNA molecules not including a step of DNA amplification of the template DNA fragments;

20 sequencing the sequencing library of analyzable DNA molecules to obtain a plurality of sequence reads;

receiving, at a computer system, the plurality of sequence reads;

aligning, by the computer system, the plurality of sequence reads to a reference human genome to determine genomic positions for the plurality of sequence reads;

25 obtaining, by the computer system, information about a maternal genome of the female subject and a paternal genome of a father of the fetus; and

comparing, by the computer system, the sequence reads to the maternal genome and the paternal genome to identify a filtered set of loci as having *de novo* mutations in the fetus, wherein:

30 at each locus of the filtered set, a number of the sequence reads having a sequence variant not in the maternal genome and not in the paternal genome is above a cutoff value, the cutoff value being greater than one.

[0005g] In yet another aspect, the present invention provides a method for identifying *de novo* mutations of a fetus by analyzing a biological sample of a female subject pregnant with the fetus, the biological sample including cell-free DNA fragments from the fetus and the female subject, the method comprising:

- 5 obtaining template DNA fragments from the biological sample to be analyzed, the template DNA fragments including cell-free DNA fragments;
 - preparing a sequencing library of analyzable DNA molecules using the template DNA fragments, wherein a duplication rate of the sequencing library from the template DNA fragments is less than 5%;
- 10 sequencing the sequencing library of analyzable DNA molecules to obtain a plurality of sequence reads;
 - receiving, at a computer system, the plurality of sequence reads;
 - aligning, by the computer system, the plurality of sequence reads to a reference human genome to determine genomic positions for the plurality of sequence reads;
- 15 obtaining, by the computer system, information about a maternal genome of the female subject and a paternal genome of a father of the fetus; and
 - comparing, by the computer system, the sequence reads to the maternal genome and the paternal genome to identify a filtered set of loci as having *de novo* mutations in the fetus, wherein:
- 20 at each locus of the filtered set, a number of the sequence reads having a sequence variant not in the maternal genome and not in the paternal genome is above a cutoff value, the cutoff value being greater than one.

[0005h] In yet another aspect, the present invention provides a method for identifying *de novo* mutations of a fetus by analyzing a biological sample of a female subject pregnant with the fetus, the biological sample including cell-free DNA fragments from the fetus and the female subject, the method comprising, performing, by a computer system:

- 25 obtaining information about a maternal genome of the female subject and a paternal genome of a father of the fetus;
 - receiving one or more sequence reads for each of a plurality of DNA
- 30 fragments in the biological sample;
 - aligning the plurality of sequence reads to a reference human genome using a first alignment procedure to determine genomic positions for the plurality of sequence reads;

comparing the sequence reads to the maternal genome and the paternal genome to identify a filtered set of loci as having *de novo* mutations in the fetus, wherein:

at each locus of the filtered set, a number of the sequence reads having a sequence variant not in the maternal genome and not in the paternal genome is above a

cutoff value, the cutoff value being greater than one;

for each of a first set of candidate loci identified as potentially having a *de novo* mutation:

for each of the sequence reads aligning to the candidate locus using the first alignment procedure and having the sequence variant:

determining whether the sequence read aligns to the candidate locus using a second alignment procedure that uses a different matching algorithm than used for the first alignment procedure;

comparing the mapping quality to a quality threshold; and

determining whether to discard the sequence read based on the

comparing of the mapping quality to the quality threshold, wherein the mapping quality being less than the quality threshold provides a higher likelihood of discarding the sequence read than the mapping quality being greater than the quality threshold, thereby obtaining a number of remaining sequence reads;

comparing the number of remaining sequence reads to a candidate

threshold; and

determining whether to discard the candidate locus based on the

comparing of the number of remaining sequence reads to the candidate threshold, wherein the number of remaining sequence reads being less than the candidate threshold provides a higher likelihood of discarding the candidate locus than the number of remaining

sequence reads being greater than the candidate threshold; and

identifying the filtered set of loci as having *de novo* mutations using the remaining candidate loci.

[0005i] In yet another aspect, the present invention provides a method for identifying *de novo* mutations of a fetus by analyzing a biological sample of a female subject pregnant with the fetus, the biological sample including cell-free DNA fragments from the fetus and the female subject, the method comprising, performing, by a computer system:

obtaining information about a maternal genome of the female subject and a paternal genome of a father of the fetus;

receiving one or more sequence reads for each of a plurality of DNA fragments in the biological sample;

aligning the plurality of sequence reads to a reference human genome using a first alignment procedure to determine genomic positions for the plurality of sequence reads;

5 comparing the sequence reads to the maternal genome and the paternal genome to identify a filtered set of loci as having *de novo* mutations in the fetus, wherein:

at each locus of the filtered set, a number of the sequence reads having a sequence variant not in the maternal genome and not in the paternal genome is above a cutoff value, the cutoff value being greater than one;

10 for each of a first set of candidate loci identified as potentially having a *de novo* mutation:

determining a size difference between a first group of DNA fragments having the sequence variant and a second group of DNA fragments having a wildtype allele;

15 comparing the size difference to a size threshold;
when the size difference is less than the size threshold, discarding the candidate locus as a potential mutation; and
identifying the filtered set of loci as having *de novo* mutations in the fetus using the remaining candidate loci.

20 **[0005j]** In yet another aspect, the present invention provides a method for identifying *de novo* mutations of a fetus by analyzing a biological sample of a female subject pregnant with the fetus, the biological sample including cell-free DNA fragments from the fetus and the female subject, the method comprising, performing, by a computer system:

25 obtaining information about a maternal genome of the female subject and a paternal genome of a father of the fetus;

receiving one or more sequence reads for each of a plurality of DNA fragments in the biological sample;

aligning the plurality of sequence reads to a reference human genome using a first alignment procedure to determine genomic positions for the plurality of sequence reads;

30 comparing the sequence reads to the maternal genome and the paternal genome to identify a filtered set of loci as having *de novo* mutations in the fetus, wherein:

at each locus of the filtered set, a number of the sequence reads having a sequence variant not in the maternal genome and not in the paternal genome is above a cutoff value, the cutoff value being greater than one;

identifying a group of regions known to be associated with histone
5 modifications that are associated with fetal tissue;

for each of a first set of candidate loci identified as potentially having a *de novo* mutation:

determining whether the candidate locus is in one of the group of regions;

determining whether to discard the candidate locus based on whether the

10 candidate locus is in one of the group of regions, wherein the candidate locus not being in one of the group of regions provides a higher likelihood of discarding the candidate locus than when the candidate locus is in one of the group of regions;

identifying the filtered set of loci as having *de novo* mutations using the remaining candidate loci.

15 **[0005k]** In yet another aspect, the present invention provides a computer product comprising a computer readable medium storing a plurality of instructions for controlling a computer system to perform an operation as disclosed herein.

[0005l] In yet another aspect, the present invention provides a system
comprising:

20 the computer product as disclosed herein; and
one or more processors for executing instructions stored on the computer readable medium.

[0005m] In yet another aspect, the present invention provides a system comprising means for performing the method as disclosed herein.

[0005n] In yet another aspect, the present invention provides a system configured to perform
25 the method as disclosed herein.

[0005o] In yet another aspect, the present invention provides a system comprising modules that respectively perform the steps of any of the method as disclosed herein.

[0006] Examples are related to the accurate detection of somatic mutations in the plasma
(or other samples containing cell-free DNA) of cancer patients and for subjects being
30 screened for cancer. The detection of these molecular markers would be useful for the screening, detection, monitoring, management, and prognostication of cancer patients. For

example, a mutational load can be determined from the identified somatic mutations, and the mutational load can be used to screen for any or various types of cancers, where no prior knowledge about a tumor or possible cancer of the subject may be required. Examples can be useful for guiding the use of therapies (e.g. targeted therapy, immunotherapy, genome editing, surgery, chemotherapy, embolization therapy, anti-angiogenesis therapy) for cancers. Examples are also directed to identifying de novo mutations in a fetus by analyzing a maternal sample having cell-free DNA from the fetus.

[0007] Other examples are directed to systems and computer readable media associated with methods described herein.

[0008] A better understanding of the nature and advantages of embodiments of the present invention may be gained with reference to the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. 1 shows a table 100 of the top 28 most commonly identified mutations among cancers.

[0010] FIG. 2 is a table 200 showing an expected number of mutations to be detected for different tumor DNA fractions, sequencing depths, number of mutation per genome and the fraction of genome searched.

[0011] FIG. 3 is a plot 300 showing the relationship between the percentage of sequence reads from PCR replicates and sequencing depth.

[0012] FIGS. 4A and 4B show a comparison between the sequencing depth required for PCR and PCR-free protocols to detect cancer-associated mutations in the plasma of a cancer subject at various tumor DNA fractions according to embodiments of the present invention.

WHAT IS CLAIMED IS:

1 1. A method for identifying somatic mutations in a human subject by
2 analyzing a biological sample of the human subject, the biological sample including DNA
3 fragments originating from normal cells and potentially from tumor cells or cells associated
4 with cancer, the biological sample including cell-free DNA fragments, the method
5 comprising:
6 obtaining template DNA fragments from the biological sample to be analyzed,
7 the template DNA fragments including cell-free DNA fragments;
8 preparing a sequencing library of analyzable DNA molecules using the
9 template DNA fragments, the preparation of the sequencing library of analyzable DNA
10 molecules not including a step of DNA amplification of the template DNA fragments;
11 sequencing the sequencing library of analyzable DNA molecules to obtain a
12 plurality of sequence reads;
13 receiving, at a computer system, the plurality of sequence reads;
14 aligning, by the computer system, the plurality of sequence reads to a
15 reference human genome to determine genomic positions for the plurality of sequence reads;
16 obtaining, by the computer system, information about a consensus sequence
17 corresponding to the human subject; and
18 comparing, by the computer system, the sequence reads to the consensus
19 sequence to identify a filtered set of loci as having somatic mutations in some tissue of the
20 human subject, wherein:
21 at each locus of the filtered set, a number of the sequence reads having a
22 sequence variant relative to the consensus sequence is above a cutoff value, the cutoff
23 value being greater than one.

1 2. A method for identifying somatic mutations in a human subject by
2 analyzing a biological sample of the human subject, the biological sample including DNA
3 fragments originating from normal cells and potentially from tumor cells or cells associated
4 with cancer, the biological sample including cell-free DNA fragments, the method
5 comprising:

obtaining template DNA fragments from the biological sample to be analyzed,
the template DNA fragments including cell-free DNA fragments;
preparing a sequencing library of analyzable DNA molecules using the
template DNA fragments, wherein a duplication rate of the sequencing library from the
template DNA fragments is less than 5%;
sequencing the sequencing library of analyzable DNA molecules to obtain a
plurality of sequence reads;
receiving, at a computer system, the plurality of sequence reads;
aligning, by the computer system, the plurality of sequence reads to a
reference human genome to determine genomic positions for the plurality of sequence reads;
obtaining, by the computer system, information about a consensus sequence
corresponding to the human subject; and
comparing, by the computer system, the sequence reads to the consensus
sequence to identify a filtered set of loci as having somatic mutations in some tissue of the
human subject, wherein:
at each locus of the filtered set, a number of the sequence reads having a
sequence variant relative to the consensus sequence is above a cutoff value, the cutoff
value being greater than one.

3. The method of claim 1 or claim 2, wherein identifying the filtered set
of loci as having somatic mutations in some tissue of the human subject further includes:
for each of a first set of candidate loci identified as potentially having a
somatic mutation:
for each of the sequence reads aligning to the candidate locus using a first
alignment procedure and having the sequence variant:
determining whether the sequence read aligns to the candidate locus
using a second alignment procedure that uses a different matching algorithm than that
used for the first alignment procedure;
when the sequence read realigns to the candidate locus using the
second alignment procedure, determining a mapping quality of the realignment for the
second alignment procedure;
comparing the mapping quality to a quality threshold; and

14 determining whether to discard the sequence read based on the
15 comparing of the mapping quality to the quality threshold, wherein the mapping
16 quality being less than the quality threshold provides a higher likelihood of discarding
17 the sequence read than the mapping quality being greater than the quality threshold,
18 thereby obtaining a number of remaining sequence reads;
19 comparing the number of remaining sequence reads to a candidate
20 threshold; and
21 determining whether to discard the candidate locus based on the
22 comparing of the number of remaining sequence reads to the candidate threshold, wherein
23 the number of remaining sequence reads being less than the candidate threshold provides
24 a higher likelihood of discarding the candidate locus than the number of remaining
25 sequence reads being greater than the candidate threshold; and
26 identifying the filtered set of loci as having somatic mutations using the
27 remaining candidate loci.

1 4. The method of claim 2, wherein the duplication rate is less than 2%.

1 5. The method of claim 4, wherein the number of analyzable DNA
2 molecules in the sequencing library is less than the number of template DNA fragments
3 originally present in the biological sample before library preparation.

1 6. A method for identifying somatic mutations in a human subject by
2 analyzing a biological sample of the human subject, the biological sample including DNA
3 fragments originating from normal cells and potentially from tumor cells or cells associated
4 with cancer, the biological sample including cell-free DNA fragments, the method
5 comprising, performing, by a computer system:
6 obtaining information about a consensus sequence corresponding to the human
7 subject;
8 receiving one or more sequence reads for each of a plurality of DNA
9 fragments in the biological sample;
10 aligning the plurality of sequence reads to a reference human genome using a
11 first alignment procedure to determine genomic positions for the plurality of sequence reads;
12 comparing the sequence reads to the consensus sequence to identify a filtered
13 set of loci as having somatic mutations in some tissue of the human subject, wherein:

14 at each locus of the filtered set, a number of the sequence reads having a
15 sequence variant relative to the consensus sequence is above a cutoff value, the cutoff
16 value being greater than one;
17 for each of a first set of candidate loci identified as potentially having a
18 somatic mutation:
19 for each of the sequence reads aligning to the candidate locus using the
20 first alignment procedure and having the sequence variant:
21 determining whether the sequence read aligns to the candidate locus
22 using a second alignment procedure that uses a different matching algorithm than that
23 used for the first alignment procedure;
24 comparing the mapping quality to a quality threshold; and
25 determining whether to discard the sequence read based on the
26 comparing of the mapping quality to the quality threshold, wherein the mapping
27 quality being less than the quality threshold provides a higher likelihood of discarding
28 the sequence read than the mapping quality being greater than the quality threshold,
29 thereby obtaining a number of remaining sequence reads;
30 comparing the number of remaining sequence reads to a candidate
31 threshold; and
32 determining whether to discard the candidate locus based on the
33 comparing of the number of remaining sequence reads to the candidate threshold, wherein
34 the number of remaining sequence reads being less than the candidate threshold provides
35 a higher likelihood of discarding the candidate locus than the number of remaining
36 sequence reads being greater than the candidate threshold; and
37 identifying the filtered set of loci as having somatic mutations using the
38 remaining candidate loci.

1 7. The method of any one of claims 1, 2, or 6, wherein identifying the
2 filtered set of loci as having somatic mutations in some tissue of the human subject further
3 includes:

4 for each of a second set of candidate loci identified as potentially having a
5 somatic mutation:

6 determining a size difference between a first group of DNA fragments
7 having the sequence variant and a second group of DNA fragments having a wildtype
8 allele;

9 comparing the size difference to a size threshold;
10 determining whether to discard the candidate locus as a potential mutation
11 based on the comparison, wherein the size difference being less than the size threshold
12 provides a higher likelihood of discarding the candidate locus than the size difference
13 being greater than the size threshold; and
14 identifying the filtered set of loci as having somatic mutations in the human
15 subject using the remaining candidate loci.

1 8. The method of claim 7, wherein the size difference is a difference in a
2 median size of the first group of DNA fragments and the second group of DNA fragments.

1 9. The method of claim 7, wherein the size difference is a maximum in a
2 cumulative frequency by size between the first group and the second group.

1 10. A method for identifying somatic mutations in a human subject by
2 analyzing a biological sample of the human subject, the biological sample including DNA
3 fragments originating from normal cells and potentially from tumor cells or cells associated
4 with cancer, the biological sample including cell-free DNA fragments, the method
5 comprising, performing, by a computer system:
6 obtaining information about a consensus sequence corresponding to the human
7 subject; and
8 receiving one or more sequence reads for each of a plurality of DNA
9 fragments in the biological sample;
10 aligning the plurality of sequence reads to a reference human genome using a
11 first alignment procedure to determine genomic positions for the plurality of sequence reads;
12 comparing the sequence reads to the consensus sequence to identify a filtered
13 set of loci as having somatic mutations in some tissue of the human subject, wherein:
14 at each locus of the filtered set, a number of the sequence reads having a
15 sequence variant relative to the consensus sequence is above a cutoff value, the cutoff
16 value being greater than one;
17 for each of a first set of candidate loci identified as potentially having a
18 somatic mutation:
19 determining a size difference between a first group of DNA fragments
20 having the sequence variant and a second group of DNA fragments having a wildtype
21 allele;

comparing the size difference to a size threshold;
when the size difference is less than the size threshold, discarding the
candidate locus as a potential mutation; and
identifying the filtered set of loci as having somatic mutations in the human
subject using the remaining candidate loci.

11. The method of any one of claims 1, 2, 6, or 10, wherein identifying the
filtered set of loci as having somatic mutations in some tissue of the human subject further
includes:

identifying a group of regions known to be associated with histone
modifications that are associated with cancer;
for each of a second first set of candidate loci identified as potentially having a
somatic mutation:
determining whether the candidate locus is in one of the group of regions;
determining whether to discard the candidate locus based on whether the
candidate locus is in one of the group of regions, wherein the candidate locus not being in
one of the group of regions provides a higher likelihood of discarding the candidate locus
than when the candidate locus is in one of the group of regions;
identifying the filtered set of loci as having somatic mutations using the
remaining candidate loci.

12. A method for identifying somatic mutations in a human subject by
analyzing a biological sample of the human subject, the biological sample including DNA
fragments originating from normal cells and potentially from tumor cells or cells associated
with cancer, the biological sample including cell-free DNA fragments, the method
comprising, performing, by a computer system:
obtaining information about a consensus sequence corresponding to the human
subject; and
receiving one or more sequence reads for each of a plurality of DNA
fragments in the biological sample;
aligning the plurality of sequence reads to a reference human genome using a
first alignment procedure to determine genomic positions for the plurality of sequence reads;
comparing the sequence reads to the consensus sequence to identify a filtered
set of loci as having somatic mutations in some tissue of the human subject, wherein:

14 at each locus of the filtered set, a number of the sequence reads having a
15 sequence variant relative to the consensus sequence is above a cutoff value, the cutoff
16 value being greater than one;
17 identifying a group of regions known to be associated with histone
18 modifications that are associated with cancer;
19 for each of a first set of candidate loci identified as potentially having a
20 somatic mutation:
21 determining whether the candidate locus is in one of the group of regions;
22 determining whether to discard the candidate locus based on whether the
23 candidate locus is in one of the group of regions, wherein the candidate locus not being in
24 one of the group of regions provides a higher likelihood of discarding the candidate locus
25 than when the candidate locus is in one of the group of regions;
26 identifying the filtered set of loci as having somatic mutations using the
27 remaining candidate loci.

1 13. The method of any one of claims 1, 2, 6, 10, or 12, further comprising:
2 determining a mutational load for the human subject using an amount of loci
3 in the filtered set of loci.

1 14. The method of claim 13, wherein the mutational load is determined as
2 a raw number of somatic mutations, a density of somatic mutations per number of bases, a
3 percentage of loci of a genomic region that are identified as having somatic mutations, a
4 number of somatic mutations observed in a particular amount of sample, or an increase
5 compared with a reference load.

1 15. The method of claim 13, further comprising:
2 comparing the mutational load to a cancer threshold to determine a level of
3 cancer.

1 16. The method of claim 15, wherein the level of cancer indicates a tumor,
2 further comprising:
3 determining a first amount of histone modifications for each of a first plurality
4 of segments of the reference human genome;
5 determining a second amount of the filtered set of loci for each of a second
6 plurality of segments of the reference human genome;

7 determining a first set of segments having the first amount of histone
8 modifications above a first threshold and having the second amount of the filtered set of loci
9 above a second threshold; and
10 identifying a tissue of origin of the tumor based on the first set of segments.

1 17. The method of any one of claims 1, 2, 6, 10, or 12, wherein identifying
2 the filtered set of loci as having somatic mutations in some tissue of the human subject
3 further includes:

4 for each of a second set of candidate loci identified as potentially having a
5 somatic mutation:

6 determining a fraction of sequence reads having the sequence variant;
7 comparing the fraction to a fraction threshold;

8 determining whether to discard the candidate locus as a potential mutation
9 based on the comparison, wherein the fraction being less than the fraction threshold
10 provides a higher likelihood of discarding the candidate locus than the fraction being
11 greater than the fraction threshold; and

12 identifying the filtered set of loci as having somatic mutations in the human
13 subject using the remaining candidate loci.

1 18. The method of claim 17, wherein the fraction threshold is 20%.

1 19. The method of claim 17, wherein the fraction threshold is 30%.

1 20. The method of claim 17, further comprising:
2 measuring a fractional concentration of tumor DNA in the biological sample,
3 wherein the fraction threshold is determined based on the fractional concentration.

1 21. The method of claim 20, wherein the fractional concentration of tumor
2 DNA in the biological sample is measured for each of a plurality of regions, and wherein the
3 fraction threshold used for a candidate locus is dependent on the fractional concentration
4 measured for the region that the candidate locus resides.

1 22. The method of claim 17, further comprising:
2 identifying one or more aberrant regions that have a copy number aberration,
3 wherein the fraction threshold used for a candidate locus in an aberrant region is dependent
4 on whether the aberrant region exhibits a copy number gain or a copy number loss.

1 23. The method of claim 17, further comprising:
2 identifying one or more aberrant regions that have a copy number aberration;
3 and
4 identifying a first sequence read from a first aberrant region exhibiting a copy
5 number gain to be more likely to have a somatic mutation than a second sequence read from a
6 second aberrant region exhibiting a copy number loss as part of determining whether to
7 discard sequence reads for determining the number of the sequence reads having a sequence
8 variant relative to the consensus sequence for each of the filtered set of loci.

1 24. The method of claim 23, wherein the one or more aberrant regions are
2 identified by:
3 for each of the second set of candidate loci identified as potentially having a
4 somatic mutation:
5 calculating an apparent mutant fraction of a sequence variant relative to
6 the consensus sequence;
7 for each of a plurality of regions:
8 determining a variance in the apparent mutant fractions of the candidate
9 loci in the aberrant region;
10 comparing the variance to a variance threshold, where an aberrant region
11 exhibiting a copy number gain has a variance greater than the threshold.

1 25. The method of any one of claims 1, 2, 6, 10, or 12, wherein the
2 sequencing is methylation-aware sequencing, and wherein identifying the filtered set of loci
3 as having somatic mutations in some tissue of the human subject further includes:
4 for each of a second set of candidate loci identified as potentially having a
5 somatic mutation:
6 for each of the sequence reads aligning to the candidate locus and having
7 the sequence variant:
8 determining a methylation status of the corresponding analyzable DNA
9 molecule at one or more sites;
10 determining whether to discard the sequence read based on the
11 methylation status, wherein the methylation status not being methylated provides a
12 higher likelihood of discarding the sequence read than the methylation status being
13 methylated, thereby obtaining a number of remaining sequence reads;

14 comparing the number of remaining sequence reads to a candidate
15 threshold; and
16 determining whether to discard the candidate locus based on the
17 comparing of the number of remaining sequence reads to the candidate threshold, wherein
18 the number of remaining sequence reads being less than the candidate threshold provides
19 a higher likelihood of discarding the candidate locus than the number of remaining
20 sequence reads being greater than the candidate threshold; and
21 identifying the filtered set of loci as having somatic mutations using the
22 remaining candidate loci.

1 26. The method of any one of claims 1, 2, 6, 10, or 12, wherein identifying
2 the filtered set of loci as having somatic mutations in some tissue of the human subject
3 further includes:
4 for each of a second set of candidate loci identified as potentially having a
5 somatic mutation:
6 for each of the sequence reads aligning to the candidate locus and having
7 the sequence variant:
8 determining an end location corresponding to where an end of the
9 sequence read aligns;
10 comparing the end location to a plurality of cancer-specific or cancer-
11 associated terminal locations;
12 determining whether to discard the sequence read based on the
13 comparison, wherein the end location not being a cancer-specific or cancer-associated
14 terminal location provides a higher likelihood of discarding the sequence read than the
15 end location being a cancer-specific or cancer-associated terminal location, thereby
16 obtaining a number of remaining sequence reads;
17 comparing the number of remaining sequence reads to a candidate
18 threshold; and
19 determining whether to discard the candidate locus based on the
20 comparing of the number of remaining sequence reads to the candidate threshold, wherein
21 the number of remaining sequence reads being less than the candidate threshold provides
22 a higher likelihood of discarding the candidate locus than the number of remaining
23 sequence reads being greater than the candidate threshold; and

identifying the filtered set of loci as having somatic mutations using the remaining candidate loci.

27. The method of any one of claims 1, 2, 6, 10, or 12, wherein the sequencing is performed using a single-stranded sequencing library preparation process that provides a subsequent sequencing step to yield two strand reads for each template DNA molecule, wherein identifying the filtered set of loci as having somatic mutations in some tissue of the human subject further includes:

for each of a second set of candidate loci identified as potentially having a somatic mutation:

for each pair of strand reads aligning to the candidate locus:

determining whether both strands have the sequence variant;

determining whether to discard the sequence read based on whether both strands have the sequence variant, wherein both strands not having the sequence variant provides a higher likelihood of discarding the strand reads than the only one strand read having the sequence variant, thereby obtaining a number of remaining sequence reads;

comparing the number of remaining sequence reads to a candidate threshold; and

determining whether to discard the candidate locus based on the comparing of the number of remaining sequence reads to the candidate threshold, wherein the number of remaining sequence reads being less than the candidate threshold provides a higher likelihood of discarding the candidate locus than the number of remaining sequence reads being greater than the candidate threshold; and

identifying the filtered set of loci as having somatic mutations using the remaining candidate loci.

28. The method of any one of claims 1, 2, 6, 10, or 12, wherein the consensus sequence corresponding to the human subject is a reference genome for a specified population of human subjects.

29. The method of any one of claims 1, 2, 6, 10, or 12, wherein cell-free DNA fragments from tumor cells or cells associated with cancer comprise less than 50% of the cell-free DNA fragments in the biological sample.

1 30. The method of any one of claims 1, 2, 6, 10, or 12, wherein the
2 biological sample includes plasma or serum.

1 31. The method of any one of claims 1, 2, 6, 10, or 12, wherein the aligned
2 sequence reads comprise at least 5% of the reference human genome.

1 32. The method of claim 31, wherein the aligned sequence reads comprise
2 at least 10% of the reference human genome.

1 33. The method of any one of claims 1, 2, 6, 10, or 12, wherein a
2 sequencing depth of at least 25x is used.

1 34. The method of claim 33, wherein the sequencing depth is at least 50x.

1 35. The method of claim 34, wherein the sequencing depth is at least 100x.

1 36. A method for identifying *de novo* mutations of a fetus by analyzing a
2 biological sample of a female subject pregnant with the fetus, the biological sample including
3 cell-free DNA fragments from the fetus and the female subject, the method comprising:
4 obtaining template DNA fragments from the biological sample to be analyzed,
5 the template DNA fragments including cell-free DNA fragments;
6 preparing a sequencing library of analyzable DNA molecules using the
7 template DNA fragments, the preparation of the sequencing library of analyzable DNA
8 molecules not including a step of DNA amplification of the template DNA fragments;
9 sequencing the sequencing library of analyzable DNA molecules to obtain a
10 plurality of sequence reads;
11 receiving, at a computer system, the plurality of sequence reads;
12 aligning, by the computer system, the plurality of sequence reads to a
13 reference human genome to determine genomic positions for the plurality of sequence reads;
14 obtaining, by the computer system, information about a maternal genome of
15 the female subject and a paternal genome of a father of the fetus; and
16 comparing, by the computer system, the sequence reads to the maternal
17 genome and the paternal genome to identify a filtered set of loci as having *de novo* mutations
18 in the fetus, wherein:

19 at each locus of the filtered set, a number of the sequence reads having a
20 sequence variant not in the maternal genome and not in the paternal genome is above a
21 cutoff value, the cutoff value being greater than one.

1 37. A method for identifying *de novo* mutations of a fetus by analyzing a
2 biological sample of a female subject pregnant with the fetus, the biological sample including
3 cell-free DNA fragments from the fetus and the female subject, the method comprising:

4 obtaining template DNA fragments from the biological sample to be analyzed,
5 the template DNA fragments including cell-free DNA fragments;

6 preparing a sequencing library of analyzable DNA molecules using the
7 template DNA fragments, wherein a duplication rate of the sequencing library from the
8 template DNA fragments is less than 5%;

9 sequencing the sequencing library of analyzable DNA molecules to obtain a
10 plurality of sequence reads;

11 receiving, at a computer system, the plurality of sequence reads;

12 aligning, by the computer system, the plurality of sequence reads to a
13 reference human genome to determine genomic positions for the plurality of sequence reads;

14 obtaining, by the computer system, information about a maternal genome of
15 the female subject and a paternal genome of a father of the fetus; and

16 comparing, by the computer system, the sequence reads to the maternal
17 genome and the paternal genome to identify a filtered set of loci as having *de novo* mutations
18 in the fetus, wherein:

19 at each locus of the filtered set, a number of the sequence reads having a
20 sequence variant not in the maternal genome and not in the paternal genome is above a
21 cutoff value, the cutoff value being greater than one.

1 38. The method of claim 36 or claim 37, wherein identifying the filtered
2 set of loci as having *de novo* mutations in the fetus further includes:

3 for each of a first set of candidate loci identified as potentially having a *de*
4 *novo* mutation:

5 for each of the sequence reads aligning to the candidate locus using a first
6 alignment procedure and having the sequence variant:

determining whether the sequence read aligns to the candidate locus using a second alignment procedure that uses a different matching algorithm than used for the first alignment procedure;

when the sequence read realigns to the candidate locus using the second alignment procedure, determining a mapping quality of the realignment for the second alignment procedure;

comparing the mapping quality to a quality threshold; and

determining whether to discard the sequence read based on the comparing of the mapping quality to the quality threshold, wherein the mapping quality being less than the quality threshold provides a higher likelihood of discarding the sequence read than the mapping quality being greater than the quality threshold, thereby obtaining a number of remaining sequence reads;

comparing the number of remaining sequence reads to a candidate threshold; and

determining whether to discard the candidate locus based on the comparing of the number of remaining sequence reads to the candidate threshold, wherein the number of remaining sequence reads being less than the candidate threshold provides a higher likelihood of discarding the candidate locus than the number of remaining sequence reads being greater than the candidate threshold; and

identifying the filtered set of loci as having *de novo* mutations using the remaining candidate loci.

39. The method of claim 37, wherein the duplication rate is less than 2%.

40. The method of claim 39, wherein the number of analyzable DNA molecules in the sequencing library is less than the number of template DNA fragments.

41. A method for identifying *de novo* mutations of a fetus by analyzing a biological sample of a female subject pregnant with the fetus, the biological sample including cell-free DNA fragments from the fetus and the female subject, the method comprising, performing, by a computer system:

obtaining information about a maternal genome of the female subject and a paternal genome of a father of the fetus;

receiving one or more sequence reads for each of a plurality of DNA fragments in the biological sample;

aligning the plurality of sequence reads to a reference human genome using a first alignment procedure to determine genomic positions for the plurality of sequence reads; comparing the sequence reads to the maternal genome and the paternal genome to identify a filtered set of loci as having *de novo* mutations in the fetus, wherein: at each locus of the filtered set, a number of the sequence reads having a sequence variant not in the maternal genome and not in the paternal genome is above a cutoff value, the cutoff value being greater than one; for each of a first set of candidate loci identified as potentially having a *de novo* mutation: for each of the sequence reads aligning to the candidate locus using the first alignment procedure and having the sequence variant: determining whether the sequence read aligns to the candidate locus using a second alignment procedure that uses a different matching algorithm than used for the first alignment procedure; comparing the mapping quality to a quality threshold; and determining whether to discard the sequence read based on the comparing of the mapping quality to the quality threshold, wherein the mapping quality being less than the quality threshold provides a higher likelihood of discarding the sequence read than the mapping quality being greater than the quality threshold, thereby obtaining a number of remaining sequence reads; comparing the number of remaining sequence reads to a candidate threshold; and determining whether to discard the candidate locus based on the comparing of the number of remaining sequence reads to the candidate threshold, wherein the number of remaining sequence reads being less than the candidate threshold provides a higher likelihood of discarding the candidate locus than the number of remaining sequence reads being greater than the candidate threshold; and identifying the filtered set of loci as having *de novo* mutations using the remaining candidate loci.

42. The method of any one of claims 36, 37, or 41, wherein identifying the filtered set of loci as having *de novo* mutations in the fetus further includes:

for each of a second set of candidate loci identified as potentially having a *de novo* mutation:

- determining a size difference between a first group of DNA fragments having the sequence variant and a second group of DNA fragments having a wildtype allele;
- comparing the size difference to a size threshold;
- determining whether to discard the candidate locus as a potential mutation based on the comparison, wherein the size difference being less than the size threshold provides a higher likelihood of discarding the candidate locus than the size difference being greater than the size threshold; and
- identifying the filtered set of loci as having *de novo* mutations in the fetus using the remaining candidate loci.

43. The method of claim 42, wherein the size difference is a difference in a median size of the first group of DNA fragments and the second group of DNA fragments.

44. The method of claim 42, wherein the size difference is a maximum in a cumulative frequency by size between the first group and the second group.

45. A method for identifying *de novo* mutations of a fetus by analyzing a biological sample of a female subject pregnant with the fetus, the biological sample including cell-free DNA fragments from the fetus and the female subject, the method comprising, performing, by a computer system:

- obtaining information about a maternal genome of the female subject and a paternal genome of a father of the fetus;
- receiving one or more sequence reads for each of a plurality of DNA fragments in the biological sample;
- aligning the plurality of sequence reads to a reference human genome using a first alignment procedure to determine genomic positions for the plurality of sequence reads;
- comparing the sequence reads to the maternal genome and the paternal genome to identify a filtered set of loci as having *de novo* mutations in the fetus, wherein:
 - at each locus of the filtered set, a number of the sequence reads having a sequence variant not in the maternal genome and not in the paternal genome is above a cutoff value, the cutoff value being greater than one;

16 for each of a first set of candidate loci identified as potentially having a *de*
17 *nov*o mutation:
18 determining a size difference between a first group of DNA fragments
19 having the sequence variant and a second group of DNA fragments having a wildtype
20 allele;
21 comparing the size difference to a size threshold;
22 when the size difference is less than the size threshold, discarding the
23 candidate locus as a potential mutation; and
24 identifying the filtered set of loci as having *de novo* mutations in the fetus
25 using the remaining candidate loci.

1 46. The method of any one of claims 36, 37, 41, or 45, wherein identifying
2 the filtered set of loci as having *de novo* mutations in the fetus further includes:
3 identifying a group of regions known to be associated with histone
4 modifications that are associated with cancer;
5 for each of a second first set of candidate loci identified as potentially having a
6 *de novo* mutation:
7 determining whether the candidate locus is in one of the group of regions;
8 determining whether to discard the candidate locus based on whether the
9 candidate locus is in one of the group of regions, wherein the candidate locus not being in
10 one of the group of regions provides a higher likelihood of discarding the candidate locus
11 than when the candidate locus is in one of the group of regions;
12 identifying the filtered set of loci as having *de novo* mutations using the
13 remaining candidate loci.

1 47. A method for identifying *de novo* mutations of a fetus by analyzing a
2 biological sample of a female subject pregnant with the fetus, the biological sample including
3 cell-free DNA fragments from the fetus and the female subject, the method comprising,
4 performing, by a computer system:
5 obtaining information about a maternal genome of the female subject and a
6 paternal genome of a father of the fetus;
7 receiving one or more sequence reads for each of a plurality of DNA
8 fragments in the biological sample;

aligning the plurality of sequence reads to a reference human genome using a first alignment procedure to determine genomic positions for the plurality of sequence reads; comparing the sequence reads to the maternal genome and the paternal genome to identify a filtered set of loci as having *de novo* mutations in the fetus, wherein: at each locus of the filtered set, a number of the sequence reads having a sequence variant not in the maternal genome and not in the paternal genome is above a cutoff value, the cutoff value being greater than one; identifying a group of regions known to be associated with histone modifications that are associated with fetal tissue; for each of a first set of candidate loci identified as potentially having a *de novo* mutation: determining whether the candidate locus is in one of the group of regions; determining whether to discard the candidate locus based on whether the candidate locus is in one of the group of regions, wherein the candidate locus not being in one of the group of regions provides a higher likelihood of discarding the candidate locus than when the candidate locus is in one of the group of regions; identifying the filtered set of loci as having *de novo* mutations using the remaining candidate loci.

48. The method of any one of claims 36, 37, 41, 45, or 47, wherein identifying the filtered set of loci as having *de novo* mutations in the fetus further includes: for each of a second set of candidate loci identified as potentially having a *de novo* mutation: determining a fraction of sequence reads having the sequence variant; comparing the fraction to a fraction threshold; determining whether to discard the candidate locus as a potential mutation based on the comparison, wherein the fraction being less than the fraction threshold provides a higher likelihood of discarding the candidate locus than the fraction being greater than the fraction threshold; and identifying the filtered set of loci as having *de novo* mutations in the fetus using the remaining candidate loci.

49. The method of claim 48, wherein the fraction threshold is 20%.

50. The method of claim 48, wherein the fraction threshold is 30%.

1 51. The method of claim 48, further comprising:
2 measuring a fractional concentration of fetal DNA in the biological sample,
3 wherein the fraction threshold is determined based on the fractional concentration.

1 52. The method of claim 51, wherein the fractional concentration of fetal
2 DNA in the biological sample is measured for each of a plurality of regions, and wherein the
3 fraction threshold used for a candidate locus is dependent on the fractional concentration
4 measured for the region that the candidate locus resides.

1 53. The method of claim 48, further comprising:
2 identifying one or more aberrant regions that have a copy number aberration,
3 wherein the fraction threshold used for a candidate locus in an aberrant region is dependent
4 on whether the aberrant region exhibits a copy number gain or a copy number loss.

1 54. The method of claim 48, further comprising:
2 identifying one or more aberrant regions that have a copy number aberration in
3 the fetus; and
4 identifying a first sequence read from a first aberrant region exhibiting a copy
5 number gain to be more likely to have a *de novo* mutation than a second sequence read from a
6 second aberrant region exhibiting a copy number loss as part of determining whether to
7 discard sequence reads for determining the number of the sequence reads having a sequence
8 variant relative to the maternal genome and the paternal genome for each of the filtered set of
9 loci.

1 55. The method of claim 54, wherein the one or more aberrant regions are
2 identified by:
3 for each of the second set of candidate loci identified as potentially having a
4 *de novo* mutation:
5 calculating an apparent mutant fraction of a sequence variant not in the
6 maternal genome and not in the paternal genome;
7 for each of a plurality of regions:
8 determining a variance in the apparent mutant fractions of the candidate
9 loci in the aberrant region;
10 comparing the variance to a variance threshold, where an aberrant region
11 exhibiting a copy number gain has a variance greater than the threshold.

1 56. The method of any one of claims 36, 37, 41, 45, or 47, wherein the
2 sequencing is methylation-aware sequencing, and wherein identifying the filtered set of loci
3 as having *de novo* mutations in the fetus further includes:
4 for each of a second set of candidate loci identified as potentially having a *de*
5 *novo* mutation:
6 for each of the sequence reads aligning to the candidate locus and having
7 the sequence variant:
8 determining a methylation status of the corresponding analyzable DNA
9 molecule at one or more sites;
10 determining whether to discard the sequence read based on the
11 methylation status, wherein the methylation status not being methylated provides a
12 higher likelihood of discarding the sequence read than the methylation status being
13 methylated, thereby obtaining a number of remaining sequence reads;
14 comparing the number of remaining sequence reads to a candidate
15 threshold; and
16 determining whether to discard the candidate locus based on the
17 comparing of the number of remaining sequence reads to the candidate threshold, wherein
18 the number of remaining sequence reads being less than the candidate threshold provides
19 a higher likelihood of discarding the candidate locus than the number of remaining
20 sequence reads being greater than the candidate threshold; and
21 identifying the filtered set of loci as having *de novo* mutations using the
22 remaining candidate loci.

1 57. The method of any one of claims 36, 37, 41, 45, or 47, wherein
2 identifying the filtered set of loci as having *de novo* mutations in the fetus further includes:
3 for each of a second set of candidate loci identified as potentially having a *de*
4 *novo* mutation:
5 for each of the sequence reads aligning to the candidate locus and having
6 the sequence variant:
7 determining an end location corresponding to where an end of the
8 sequence read aligns;

9 comparing the end location to a plurality of cancer-specific or cancer-
10 associated terminal locations;
11 determining whether to discard the sequence read based on the
12 comparison, wherein the end location not being a cancer-specific or cancer-associated
13 terminal location provides a higher likelihood of discarding the sequence read than the
14 end location being a cancer-specific or cancer-associated terminal location, thereby
15 obtaining a number of remaining sequence reads;
16 comparing the number of remaining sequence reads to a candidate
17 threshold; and
18 determining whether to discard the candidate locus based on the
19 comparing of the number of remaining sequence reads to the candidate threshold, wherein
20 the number of remaining sequence reads being less than the candidate threshold provides
21 a higher likelihood of discarding the candidate locus than the number of remaining
22 sequence reads being greater than the candidate threshold; and
23 identifying the filtered set of loci as having *de novo* mutations using the
24 remaining candidate loci.

1 58. The method of any one of claims 36, 37, 41, 45, or 47, wherein the
2 sequencing is performed using a single-stranded sequencing library preparation process that
3 provides a subsequent sequencing step to yield two strand reads for each template DNA
4 molecule, wherein identifying the filtered set of loci as having *de novo* mutations in the fetus
5 further includes:

6 for each of a second set of candidate loci identified as potentially having a *de*
7 *novo* mutation:

8 for each pair of strand reads aligning to the candidate locus:

9 determining whether both strands have the sequence variant;

10 determining whether to discard the sequence read based on whether
11 both strands have the sequence variant, wherein both strands not having the sequence
12 variant provides a higher likelihood of discarding the strand reads than the only one
13 strand read having the sequence variant, thereby obtaining a number of remaining
14 sequence reads;

15 comparing the number of remaining sequence reads to a candidate
16 threshold; and

17 determining whether to discard the candidate locus based on the
18 comparing of the number of remaining sequence reads to the candidate threshold, wherein
19 the number of remaining sequence reads being less than the candidate threshold provides
20 a higher likelihood of discarding the candidate locus than the number of remaining
21 sequence reads being greater than the candidate threshold; and
22 identifying the filtered set of loci as having *de novo* mutations using the
23 remaining candidate loci.

1 59. The method of any one of claims 36, 37, 41, 45, or 47, wherein cell-
2 free DNA fragments from the fetus comprise less than 50% of the cell-free DNA fragments
3 in the biological sample.

1 60. The method of any one of claims 36, 37, 41, 45, or 47, wherein the
2 biological sample includes plasma or serum.

1 61. The method of any one of claims 36, 37, 41, 45, or 47, wherein the
2 aligned sequence reads comprise at least 5% of the reference genome.

1 62. The method of claim 61, wherein the aligned sequence reads comprise
2 at least 10% of the reference genome.

1 63. The method of any one of claims 36, 37, 41, 45, or 47, wherein a
2 sequencing depth of at least 25x is used.

1 64. The method of claim 63, wherein the sequencing depth is at least 50x.

1 65. The method of claim 64, wherein the sequencing depth is at least 100x.

1 66. A computer product comprising a computer readable medium storing a
2 plurality of instructions for controlling a computer system to perform an operation of any of
3 claims 1, 2, 6, 10, 12, 36, 37, 41, 45, or 47.

1 67. A system comprising:
2 the computer product of claim 66; and
3 one or more processors for executing instructions stored on the computer
4 readable medium.

1 68. A system comprising means for performing the method of any of
2 claims 1, 2, 6, 10, 12, 36, 37, 41, 45, or 47.

1 69. A system configured to perform the method of any of claims 1, 2, 6,
2 10, 12, 36, 37, 41, 45, or 47.

1 70. A system comprising modules that respectively perform the steps of
2 any of the method of any of claims 1, 2, 6, 10, 12, 36, 37, 41, 45, or 47.