

WHAT IS CLAIMED IS

1. A method of detecting absence of heterozygosity (AOH), e.g. copy-number neutral loss of heterozygosity (CN-LOH), in a biological sample from a subject, comprising

(i) receiving sequence reads from low-pass genome sequencing of genomic DNA of the biological sample;

(ii) aligning the sequence reads to a human genome reference, and selecting and sorting sequence reads aligned to the human genome reference based on the aligned chromosome and genomic coordinates;

(iii) identifying single-nucleotide variants (SNVs) in the aligned sequence reads, wherein a single-nucleotide variant at each site has a mutant base type different from the base type at the corresponding site from the human genome reference;

(iv) identifying homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs from the SNVs identified in step (iii), wherein

a homozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being 100%,

a diploid heterozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being no less than 25% and no large than 75%,

a non-diploid heterozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being less than 25% and larger than 0% or larger than 75% and less than 100%;

(v) determining a rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs identified in step (iv) for a window, wherein the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs represents the ratio of the number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window to the average number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs among all windows in the biological sample; and

(vi) comparing the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for individual windows determined from step (v) with an average rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for corresponding individual windows established from a control population.

2. The method of claim 1, wherein the biological sample is selected from the group consisting of peripheral blood, chorionic villus, amniotic fluid, cord blood, placental tissue, and tissue samples from organs.

3. The method of claim 1 or 2, wherein the subject is a pregnant female, an infant, a subject suffering from a cancer, or a subject suspected of suffering from a cancer.

4. The method of any one of preceding claims, wherein the sequence reads are single-end sequence reads or paired-end sequence reads.
5. The method of any one of preceding claims, wherein the low-pass genome sequencing has a read depth of 3~5 folds, e.g. 3-fold.
6. The method of any one of preceding claims, wherein the human genome reference is GRCh37/hg19 or GRCh38/hg38.
7. The method of any one of preceding claims, wherein the aligning step is done using Short Oligonucleotide Alignment Program 2 (SOAP2) or Burrows-Wheeler Aligner (BWA) and Bowtie2.
8. The method of any one of preceding claims, wherein step (ii) further includes removing sequence reads due to polymerase chain reaction (PCR) duplication.
9. The method of any one of preceding claims, wherein step (iii) further includes discarding a site as described below:
 - (a) a minimal read-depth of the site is determined by the minimal read-depth of the biological sample;
 - (b) a maximum read-depth of the site is determined by the maximal read-depth of the biological sample; or
 - (c) a site where no sequence read supports a mutant base type.
10. The method of any one of preceding claims, wherein the window in step (v) has a fixed length, e.g. 100-kb.
11. The method of any one of preceding claims, wherein step (v) comprises
 - determining the number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window,
 - determining the average number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs among all windows in the biological sample, and
 - calculating the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window by dividing the number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs identified for the window by the average number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs among all windows in the biological sample.

12. The method of any one of preceding claims, wherein the control population has the same gender as the subject, and optionally, has at least 30 control subjects.

13. The method of any one of preceding claims, wherein step (vi) comprises

normalizing the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for a window by an average rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the corresponding window established from the control population, thereby providing a corresponding rate ratio of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window.

14. The method of claim 13, wherein in step (vi), decreased rate of diploid heterozygous SNVs and increased rate of homozygous SNVs indicate AOH, and preferably, step (vi) further comprises

where the copy-number neutral expressing as copy-ratios equal to 1, for all windows with diploid heterozygous SNVs rate ratios less than 1, defining a region if there are a plurality of windows with consecutive diploid heterozygous SNVs rate ratios less than 0.5 and the percentage of windows with homozygous rate ratios larger than 1.25 is at least 30%, and optionally combining two regions into one if there are no more than one windows with diploid heterozygous SNVs rate ratios larger than 0.5 but less than 1; and

reporting the region as presence of AOH.

15. The method of claim 13, wherein in step (vi), increased rate of non-diploid heterozygous SNVs indicates mosaic AOH, and preferably, step (vi) further comprises

where the copy-number mosaic duplication represented as copy-ratio larger than 1, for all windows with non-diploid heterozygous SNVs rate ratios larger than 1, defining a region if there are a plurality of windows with consecutive non-diploid heterozygous SNVs rate ratios larger than 1.15; and

reporting the region as presence of mosaic AOH.

16. The method of any one of preceding claims, wherein an average rate of heterozygous SNVs for corresponding individual windows established from a control population is determined by

(ci) receiving sequence reads from low-pass genome sequencing of genomic DNA of a biological sample from a control subject from the control population;

(cii) aligning the sequence reads to a human genome reference, and selecting and sorting sequence reads aligned to the human genome reference based on the aligned chromosome and genomic coordinates;

(ciii) identifying single-nucleotide variants (SNVs) in the aligned sequence reads, wherein a single-nucleotide variant at each site has a mutant base type different from the base type at the corresponding site from the human genome reference;

(civ) identifying homozygous SNVs, diploid heterozygous SNVs, or non-diploid

heterozygous SNVs from the SNVs identified in step (ciii), wherein

a homozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being 100%,

a diploid heterozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being no less than 25% and no large than 75%,

a non-diploid heterozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being less than 25% and larger than 0% or larger than 75% and less than 100%;

(cv) determining a rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs identified in step (civ) for a window, wherein the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs represents the ratio of the number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window to the average number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs among all windows in the biological sample from the control subject; and

(cvi) averaging rates of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for a window from all control subjects to provide an average rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the corresponding window in the control population.

17. The method of claim 16, further comprising, between step (cii) and (ciii), a step of sex determination, wherein the aligned ratios of chromosome X, chromosome Y and the whole genome are calculated as the numbers of sequence reads aligned to the chromosome/genome dividing by the length defined by the humane reference genome, respectively, the chromosome Y percentage is calculated as the aligned ratio of chromosome Y dividing by the aligned ratio of the whole genome, and a control subject is considered as male if the chromosome Y percentage is larger than 0.1.

18. The method of claim 17, wherein steps (ciii) to (cvi) are carried out on male and female control subjects respectively, based on the result of the step of sex determination.

19. The method of claim 16, wherein, in step (cvi), if rates of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for a window among control subjects have substantial deviation, the average rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window is calculated as an average of the rates of the window and its flanking windows (e.g., two upstream and two downstream windows).

20. A computer system for detecting absence of heterozygosity (AOH), e.g. copy-number neutral loss of heterozygosity (CN-LOH), in a biological sample from a subject, comprising a processor and a memory storing a plurality of instructions, wherein the

processor, upon processing the instructions, is configured to:

(i) receive sequence reads from low-pass genome sequencing of genomic DNA of the biological sample;

(ii) align the sequence reads to a human genome reference, and selecting and sorting sequence reads aligned to the human genome reference based on the aligned chromosome and genomic coordinates;

(iii) identify single-nucleotide variants (SNVs) in the aligned sequence reads, wherein a single-nucleotide variant at each site has a mutant base type different from the base type at the corresponding site from the human genome reference;

(iv) identify homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs from the SNVs identified in (iii), wherein

a homozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being 100%,

a diploid heterozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being no less than 25% and no large than 75%,

a non-diploid heterozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being less than 25% and larger than 0% or larger than 75% and less than 100%;

(v) determine a rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs identified in (iv) for a window, wherein the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs represents the ratio of the number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window to the average number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs among all windows in the biological sample; and

(vi) compare the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for individual windows determined from (v) with an average rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for corresponding individual windows established from a control population.

21. The computer system of claim 20, wherein the biological sample is selected from the group consisting of peripheral blood, chorionic villus, amniotic fluid, cord blood, placental tissue, and tissue samples from organs.

22. The computer system of claim 20 or 21, wherein the subject is a pregnant female, an infant, a subject suffering from a cancer, or a subject suspected of suffering from a cancer.

23. The computer system of any one of claims 20 to 22, wherein the sequence reads are single-end sequence reads or paired-end sequence reads.

24. The computer system of any one of claims 20 to 23, wherein the low-pass genome sequencing has a read depth of 3~5 folds, e.g. 3 folds.

25. The computer system of any one of claims 20 to 24, wherein the human genome reference is GRCh37/hg19 or GRCh38/hg38.

26. The computer system of any one of claims 20 to 25, wherein the aligning operation is done using Short Oligonucleotide Alignment Program 2 (SOAP2) or Burrows-Wheeler Aligner (BWA) and Bowtie2.

27. The computer system of any one of claims 20 to 26, wherein the processor, upon processing the instructions, is further configured to remove sequence reads due to polymerase chain reaction (PCR) duplication.

28. The computer system of any one of claims 20 to 27, wherein the processor, upon processing the instructions, is further configured to discard a site as described below:

(a) a minimal read-depth of the site is determined by the minimal read-depth of the biological sample;

(b) a maximum read-depth of the site is determined by the maximal read-depth of the biological sample; or

(c) a site where no sequence read supports a mutant base type.

29. The computer system of any one of claims 20 to 28, wherein the window has a fixed length, e.g. 100-kb.

30. The computer system of any one of claims 20 to 29, wherein the processor, upon processing the instructions, is further configured to:

determine the number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window,

determine the average number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs among all windows in the biological sample, and

calculate the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window by dividing the number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs identified for the window by the average number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs among all windows in the biological sample.

31. The computer system of any one of claims 20 to 30, wherein the control population has the same gender as the subject, and optionally, has at least 30 control subjects.

32. The computer system of any one of claims 20 to 31, wherein the processor, upon processing the instructions, is further configured to:

normalizing the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for a window by an average rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the corresponding window established from the control population, thereby providing a corresponding rate ratio of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window.

33. The computer system of claim 32, wherein the processor, upon processing the instructions, is further configured to:

where the copy-number neutral expressing as copy-ratios equal to 1, for all windows with diploid heterozygous SNVs rate ratios less than 1, define a region if there are a plurality of windows with consecutive diploid heterozygous SNVs rate ratios less than 0.5 and the percentage of windows with homozygous rate ratios larger than 1.25 is at least 30%, and optionally combining two regions into one if there are no more than one windows with diploid heterozygous SNVs rate ratios larger than 0.5 but less than 1; and

report the region as presence of AOH.

34. The computer system of claim 32, wherein the processor, upon processing the instructions, is further configured to:

where the copy-number mosaic duplication represented as copy-ratio larger than 1 or the copy-number neutral expressing as copy-ratios equal to 1, for all windows with non-diploid heterozygous SNVs rate ratios larger than 1, define a region if there are a plurality of windows with consecutive non-diploid heterozygous SNVs rate ratios larger than 1.15; and

report the region as presence of mosaic AOH.

35. A computer readable medium storing a plurality of instructions, wherein the plurality of instructions, upon executed by one or more processors, perform an operation including

(i) receiving sequence reads from low-pass genome sequencing of genomic DNA of a biological sample from a subject;

(ii) aligning the sequence reads to a human genome reference, and selecting and sorting sequence reads aligned to the human genome reference based on the aligned chromosome and genomic coordinates;

(iii) identifying single-nucleotide variants (SNVs) in the aligned sequence reads, wherein a single-nucleotide variant at each site has a mutant base type different from the base type at the corresponding site from the human genome reference;

(iv) identifying homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs from the SNVs identified in (iii), wherein

a homozygous SNV is define based on the percentage of sequence reads

supporting the mutant base type different from the base type at the corresponding site from the human genome reference being 100%,

a diploid heterozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being no less than 25% and no large than 75%,

a non-diploid heterozygous SNV is define based on the percentage of sequence reads supporting the mutant base type different from the base type at the corresponding site from the human genome reference being less than 25% and larger than 0% or larger than 75% and less than 100%;

(v) determining a rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs identified in (iv) for a window, wherein the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs represents the ratio of the number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window to the average number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs among all windows in the biological sample; and

(vi) comparing the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for individual windows determined from (v) with an average rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for corresponding individual windows established from a control population.

36. The computer readable medium of claim 35, wherein the biological sample is selected from the group consisting of peripheral blood, chorionic villus, amniotic fluid, cord blood, placental, and tissue samples from organs.

37. The computer readable medium of claim 35 or 36, wherein the subject is a pregnant female, an infant, a subject suffering from a cancer, or a subject suspected of suffering from a cancer.

38. The computer readable medium of any one of claims 35 to 37, wherein the sequence reads are single-end sequence reads or paired-end sequence reads.

39. The computer readable medium of any one of claims 35 to 38, wherein the low-pass genome sequencing has a read depth of 3~5 folds, e.g. 3 folds.

40. The computer readable medium of any one of claims 35 to 39, wherein the human genome reference is GRCh37/hg19 or GRCh38/hg38.

41. The computer readable medium of any one of claims 35 to 40, wherein the aligning is done using Short Oligonucleotide Alignment Program 2 (SOAP2) or Burrows-Wheeler Aligner (BWA) and Bowtie2.

42. The computer readable medium of any one of claims 35 to 41, wherein (ii) further includes removing sequence reads due to polymerase chain reaction (PCR) duplication.

43. The computer readable medium of any one of claims 35 to 42, wherein (iii) further includes discarding a site as described below:

(a) a minimal read-depth of the site is determined by the minimal read-depth of the biological sample;

(b) a maximum read-depth of the site is determined by the maximal read-depth of the biological sample; or

(c) a site where no sequence read supports a mutant base type.

44. The computer readable medium of any one of claims 35 to 43, wherein the window in (iv) has a fixed length, e.g. 100 kb.

45. The computer readable medium of any one of claims 35 to 44, wherein (v) comprises

determining the number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window,

determining the average number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs among all windows in the biological sample, and

calculating the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window by dividing the number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs identified for the window by the average number of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs among all windows in the biological sample.

46. The computer readable medium of any one of claims 35 to 45, wherein the control population has the same gender as the subject, and optionally, has at least 30 control subjects.

47. The computer readable medium of any one of claims 35 to 46, wherein step (vi) comprises

normalizing the rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for a window by an average rate of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the corresponding window established from the control population, thereby providing a corresponding rate ratio of homozygous SNVs, diploid heterozygous SNVs, or non-diploid heterozygous SNVs for the window.

48. The computer readable medium of claim 47, wherein in (vi), decreased rate of diploid heterozygous SNVs and increased rate of homozygous SNVs indicate AOH,

and preferably, (vi) further comprises

where the copy-number neutral expressing as copy-ratios equal to 1, for all windows with diploid heterozygous SNVs rate ratios less than 1, defining a region if there are a plurality of windows with consecutive diploid heterozygous SNVs rate ratios less than 0.5 and the percentage of windows with homozygous rate ratios larger than 1.25 is at least 30%, and optionally combining two regions into one if there are no more than one windows with diploid heterozygous SNVs rate ratios larger than 0.5 but less than 1; and

reporting the region as presence of AOH.

49. The computer readable medium of claim 47, wherein in (vi), increased rate of non-diploid heterozygous SNVs indicates mosaic AOH, and preferably, (vi) further comprises

where the copy-number mosaic duplication represented as copy-ratio larger than 1 or the copy-number neutral expressing as copy-ratios equal to 1, for all windows with non-diploid heterozygous SNVs rate ratios larger than 1, defining a region if there are a plurality of windows with consecutive non-diploid heterozygous SNVs rate ratios larger than 1.15; and

reporting the region as presence of mosaic AOH.

50. A device comprising one or more processors and a computer readable medium of any one of claims 35-49.