

WHAT IS CALIMED IS:

1. A method of determining whether copy number variation presents in a genome sample, comprising following steps:

sequencing the genome sample, to obtain a sequencing result consisting of a plurality of reads;

aligning the sequencing result to a reference genome sequence, to determine a distribution of the reads in the reference genome sequence;

determining a plurality of breakpoints in the reference genome sequence based on the distribution of the reads in the reference genome sequence, wherein the number of reads has significance at both sides of the breakpoints;

determining a detection window in the reference genome based on the plurality of the breakpoints;

determining a first parameter based on reads falling in the detection window; and

determining whether the copy number variation presents in the genome sample against the detection window based on difference between the first parameter and a preset threshold.

2. The method of claim 1, further comprising a step of extracting the genome sample from a biological sample.

3. The method of claim 2, wherein the biological sample is pregnant sample or fetal sample,

optionally, the biological sample is at least one selected from a group consisting of pregnant plasma, chorionic villi, amniotic fluid, umbilical cord blood, placenta and fetal heel blood.

4. The method of claim 2, wherein the biological sample is at least one selected from a group consisting of blood, urine, saliva, tissue, germ cells, oosperm, blastomere and embryo,

optionally, the biological sample is a single cell.

5. The method of claim 1, wherein the step of sequencing the genome sample further comprises following sub-steps of:

amplifying the genome sample;
constructing a sequencing-library with amplified genome sample; and
sequencing the sequencing-library.

6. The method of claim 4, further comprising a step of:

lysing the single cell, to release whole genome of the single cell.

7. The method of claim 6, wherein the single cell is lysed using an alkaline lysate, to release whole genome of the single cell.

8. The method of claim 7, wherein the whole genome is amplified by means of PCR-based whole genome amplification method.

9. The method of claim 8, wherein the PCR-based whole genome amplification method is OmniPlex WGA.

10. The method of claim 5, wherein the sub-step of sequencing the sequencing-library is performed by at least one selected from a group consisting of Hiseq system, Miseq system, Genome Analyzer system, 454 FLX, SOLiD system, Ion Torrent system and single molecule sequencing apparatus.

11. The method of claim 1, wherein the copy number variation is at least one selected from aneuploidy of chromosome, deletion of chromosome, and addition micro-deletion and micro-repetition of chromosome fragments.

12. The method of claim 1, wherein the step of determining a plurality of breakpoints in the reference genome sequence further comprises following sub-steps:

dividing the reference genome sequence into a plurality of primary windows having a predetermined length, and determining reads falling in each of the plurality of the primary windows;

for at least one site in the reference genome sequence, determining the number of reads falling in the same number of the plurality of primary windows at both sides of the site;

determining p value of the site, wherein the p value represents that the number of reads falling in either side of the site has significance; and

determining that the site is the breakpoints, if the p value of the site is smaller than a final p value.

13. The method of claim 12, wherein the reads falling in each of the plurality of primary windows are uniquely-aligned reads.

14. The method of claim 12, wherein 100 of primary windows are selected from either side of the site.

15. The method of claim 12, wherein the plurality of primary windows have a length of 100 Kbp to 200 Kbp, preferably 150 Kbp.

16. The method of claim 12, wherein the final p value is 1.1×10^{-50} or less.

17. The method of claim 12, wherein the sub-step of determining p value of the site further comprises:

for the site, selecting the plurality of primary windows having the same number at either side of the site respectively, and calculating the relative number of reads falling in each of the plurality of primary windows R_i , wherein i represents No. of the plurality of primary windows,

subjecting all of the relative numbers of reads falling in the plurality of primary windows R_i to Run-Test, to determine the p value of the site,

wherein the relative number of reads is determined by following formula:

$$R_i = \log_2 \left(\frac{r_i}{\bar{r}} \right)$$

wherein r_i represents the number of reads falling in the i-th primary window,

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n r_i,$$

n represents the total number of the plurality of primary windows.

18. The method of claim 17, wherein subjecting all of the relative numbers of the reads falling in each of the plurality of primary windows to Run-Test further comprises:

subjecting the relative number of reads falling in each of the plurality of primary windows R_i to a correction of GC content, to obtain corrected relative number of reads $\tilde{R}_i$;

determining the normalized number of reads falling in each of the plurality of primary windows Z_i based on the corrected relative number of reads; and

subjecting all of the normalized numbers of reads falling in each of the plurality of primary windows Z_i to Run-Test.

19. The method of claim 18, wherein the corrected relative number of reads $\tilde{R}_i$ is obtained by following steps:

calculating GC content of each of the plurality of primary windows;

dividing the GC content into a plurality of regions in a unit of 0.001, and calculating a mean value M_s among all of the relative numbers of reads falling in each of the plurality of regions, wherein s is No. of the plurality of regions;

determining the corrected relative number of reads $\tilde{R}_i$ based on the following formula:

$$\tilde{R}_i = R_i - M_s,$$

determining the normalized number of reads Z_i based on the following formula:

wherein

$$Z_i = (R_i - \tilde{R}_i - mean) / SD,$$

$$mean = \frac{1}{n} \sum_{i=1}^n (R_i - \tilde{R}_i),$$

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (R_i - \tilde{R}_i - mean)^2}$$

20. The method of claim 19, wherein the step of determining a detection window in the reference genome based on the plurality of the breakpoints further comprises:

1) determining a plurality of candidate breakpoints, wherein other breakpoints present both before and after the candidate breakpoints;

2) determining p value of each candidate breakpoint, and removing a candidate breakpoint having the maximal p value;

3) performing the step 2) with rest of the candidate breakpoints until all p values of the rest of the candidate breakpoints are smaller than the final p value, wherein the rest of the candidate breakpoints are taken as screened candidate breakpoints; and

4) determining a region between two successive screened candidate breakpoints as the detection window,

wherein the p value of the candidate breakpoint is obtained by following steps:

selecting a region between the candidate breakpoint and previous candidate breakpoint as a first candidate region and selecting a region between the candidate breakpoint and next candidate breakpoint as a second candidate region;

subjecting the normalized number of reads falling in the primary windows Z_i which are included both in the first candidate region and the second candidate region to Run-Test, to determine the p value of the candidate breakpoints,

optionally, the final p value is obtained by following steps:

based on a sequencing result of a control sample, repeating the step of determining a detection window in the reference genome, and recording p values of the breakpoints which are removed each time until the number of the breakpoints is zero; and

determining the final p value, based on p values of removed breakpoints,

optionally, the final p value is 1.1×10^{-50} .

21. The method of claim 20, wherein the step of determining a first parameter based on the reads falling in the detection window further comprises:

determining a mean value among all of the normalized numbers of reads falling in each of the plurality of primary windows $\bar{Z}$ which are included in the detection windows, wherein the mean value of the normalized numbers of reads $\bar{Z}$ is taken as the first

parameter.

22. The method of claim 1, wherein the preset threshold comprises:
a first threshold being -1.645 and a second threshold being 1.645.

23. The method of claim 1, wherein the reference genome sequence is at least one selected from human chromosome 21, chromosome 18, chromosome 13, chromosome X and chromosome Y.

24. A system for determining whether copy number variation presents in a genome sample, comprising:

a sequencing apparatus, configured to sequence the genome sample, to obtain a sequencing result consisting of a plurality of reads;

an analysis apparatus, connected to the sequencing apparatus, configured to determine whether copy number variation presents in the genome sample based on the sequencing result, wherein the analysis apparatus further comprises:

an aligning unit, configured to align the sequencing result to a reference genome sequence, to determine a distribution of the reads in the reference genome sequence;

a breakpoint determining unit, connected to the aligning unit, configured to determine a plurality of breakpoints in the reference genome sequence based on the distribution of the reads in the reference genome sequence, wherein the number of reads has significance between two sides of the breakpoints;

a detection window determining unit, connected to the breakpoint determining unit, configured to determine a detection window in the reference genome based on the plurality of the breakpoints;

a parameter determining unit, connected to the detection window determining unit, configured to determine a first parameter based on reads falling in the detection window; and

a determining unit, connected to the parameter determining unit, configured to determine whether the copy number variation presents in the genome sample against the detection window based on difference between the first parameter and a preset

threshold.

25. The system of claim 24, further comprising a genome extracting apparatus, configured to extract the genome sample from a biological sample.

26. The system of claim 24, wherein the sequencing apparatus further comprises:
a genome amplifying unit, configured to amplify the genome sample;
a sequencing-library constructing unit, connected to the genome amplifying unit, configured to construct a sequencing-library with amplified genome sample; and
a sequencing unit, connected to the sequencing-library constructing unit, configured to sequence the sequencing-library.

27. The system of claim 26, wherein the sequencing unit is at least one selected from a group consisting of Hiseq system, Miseq system, Genome Analyzer system, 454 FLX, SOLiD system, Ion Torrent system and single molecule sequencing apparatus.

28. The system of claim 24, wherein the breakpoint determining unit further comprises a module for performing following sub-steps:

dividing the reference genome sequence into a plurality of primary windows having a predetermined length, and determining reads falling in each of the plurality of the primary windows;

for at least one site in the reference genome sequence, determining the number of reads falling in the same number of the plurality of primary windows at both sides of the site;

determining p value of the site, wherein the p value represents that the number of reads falling in either side of the site has significance; and

determining the site is the breakpoints, if the p value of the site is smaller than a final p value.

29. The system of claim 28, wherein the breakpoint determining unit further comprises a module for performing followings to determine the p value:

for the site, selecting the plurality of primary windows having the same number at

either side of the site respectively, and calculating the relative number of reads falling in each of the plurality of primary windows R_i , wherein i represents No. of the plurality of primary windows,

subjecting all of the relative numbers of reads falling in the plurality of primary windows R_i to Run-Test, to determine the p value of the site,

wherein the relative number of reads is determined by following formula:

$$R_i = \log_2 \left(\frac{r_i}{\bar{r}} \right)$$

wherein r_i represents the number of reads falling in the i -th primary window,

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n r_i,$$

n represents the total number of the plurality of primary windows.

30. The system of claim 29, wherein the breakpoint determining unit further comprises a module for performing followings to subject all of the relative numbers of the reads falling in the plurality of primary windows to Run-Test:

subjecting the relative number of reads falling in each of the plurality of primary windows R_i to a correction of GC content, to obtain corrected relative number of reads $\tilde{R}_i$;

determining the normalized number of reads falling in each of the plurality of primary windows Z_i based on the corrected relative number of reads; and

subjecting all of the normalized numbers of reads falling in each of the plurality of primary windows Z_i to Run-Test.

31. The system of claim 30, wherein the corrected relative number of reads $\tilde{R}_i$ is obtained by a module for performing following steps:

calculating GC content of each of the plurality of primary windows;

dividing the GC content into a plurality of regions in a unit of 0.001, and calculating a mean value M_s among all of the relative numbers of reads falling in each of the plurality of regions, wherein s is No. of the plurality of regions;

determining the corrected relative number of reads $\tilde{R}_i$ based on the following

formula:

$$\tilde{R}_i = R_i - M_s;$$

determining the normalized number of reads Z_i based on the following formula:

wherein

$$Z_i = (R_i - \tilde{R}_i - mean) / SD, \text{ wherein}$$

$$mean = \frac{1}{n} \sum_{i=1}^n (R_i - \tilde{R}_i),$$

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (R_i - \tilde{R}_i - mean)^2}$$

32. The system of claim 31, wherein based on the plurality of the breakpoints, the detection window determining unit further comprises a module for performing followings:

1) determining a plurality of candidate breakpoints, wherein other breakpoints present both before and after the candidate breakpoints;

2) determining p value of each candidate breakpoint, and removing a candidate breakpoint having the maximal p value;

3) performing the step 2) with rest of the candidate breakpoints until all p values of the rest of the candidate breakpoints are smaller than the final p value, wherein the rest of the candidate breakpoints are taken as screened candidate breakpoints; and

4) determining a region between two successive screened candidate breakpoints as the detection window,

wherein the p value of the candidate breakpoint is obtained by following steps:

selecting a region between the candidate breakpoint and previous candidate breakpoint as a first candidate region, and selecting a region between the candidate breakpoint and next candidate breakpoint as a second candidate region;

subjecting the normalized number of reads falling in the primary windows Z_i which are included both in the first candidate region and the second candidate region to Run-Test, to determine the p value of the candidate breakpoints,

optionally, the final p value is obtained by following steps:

based on a sequencing result of a control sample, repeating the step of determining a detection window in the reference genome, and recording p values of the breakpoints

which are removed each time until the number of the breakpoints is zero; and
determining the final p value, based on a distribution of the p values of removed breakpoints,
optionally, the final p value is 1.1×10^{-50} .

33. The system of claim 32, wherein the parameter determining unit further comprises a module for performing followings:

determining a mean value among all of the normalized numbers of reads falling in each of the plurality of primary windows $\bar{Z}$ which are included in the detection windows, wherein the mean value of the normalized numbers of reads $\bar{Z}$ is taken as the first parameter.

34. The system of the claim 24, wherein a preset threshold is preserved in the determining unit, wherein the preset threshold comprises: a first threshold being -1.645 and a second threshold being 1.645.

35. The system of claim 24, wherein the reference genome sequence is preserved in the aligning unit, wherein the reference genome sequence is known human genome sequence, optionally, the reference genome sequence is at least one selected from human chromosome 21, chromosome 18, chromosome 13, chromosome X and chromosome Y.

36. A computer readable medium, comprising an order, configured to perform by a processor to determine whether copy number variation presents in a genome sample through following steps:

aligning the sequencing result to a reference genome sequence, to determine a distribution of the reads in the reference genome sequence;

determining a plurality of breakpoints in the reference genome sequence based on the distribution of the reads in the reference genome sequence, wherein the number of reads has significance at both sides of the breakpoints;

determining a detection window in the reference genome based on the plurality of

the breakpoints;

determining a first parameter based on reads falling in the detection window; and

determining whether the copy number variation presents in the genome sample against the detection window based on difference between the first parameter and a preset threshold.

37. The computer readable medium of claim 36, wherein the step of determining a plurality of breakpoints in the reference genome sequence further comprises following sub-steps:

dividing the reference genome sequence into a plurality of primary windows having a predetermined length, and determining reads falling in each of the plurality of the primary windows;

for at least one site in the reference genome sequence, determining the number of reads falling in the same number of the plurality of primary windows at both sides of the site;

determining p value of the site, wherein the p value represents that the number of reads falling in either side of the site has significance; and

determining the site is the breakpoints, if the p value of the site is smaller than a final p value.

38. The computer readable medium of claim 37, wherein the reads falling in each of plurality of the primary windows are uniquely-aligned reads.

39. The computer readable medium of claim 37, wherein 100 of primary windows are selected from either side of the site.

40. The computer readable medium of claim 37, wherein the plurality of primary windows have a length of 100 Kbp to 200 Kbp, preferably 150 Kbp.

41. The computer readable medium of claim 37, wherein the final p value is 1.1×10^{-50} or less.

42. The computer readable medium of claim 37, wherein the sub-step of determining p value of the site further comprises:

for the site, selecting the plurality of primary windows having the same number at both sides of the site respectively, and calculating the relative number of reads falling in each of the plurality of primary windows R_i , wherein i represents No. of the plurality of primary windows,

subjecting all of the relative numbers of reads falling in the plurality of primary windows R_i to Run-Test, to determine the p value of the site,

wherein the relative number of reads is determined by following formula:

$$R_i = \log_2 \left(\frac{r_i}{\bar{r}} \right)$$

wherein r_i represents the number of reads falling in the i -th primary window,

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n r_i$$

n represents the total number of the plurality of primary windows.

43. The computer readable medium of claim 42, wherein subjecting the all of the relative numbers of the reads falling in the plurality of primary windows to Run-Test further comprises:

subjecting the relative number of reads falling in each of the plurality of primary windows R_i to a correction of GC content, to obtain corrected relative number of reads $\tilde{R}_i$;

determining the normalized number of reads falling in each of the plurality of primary windows Z_i based on the corrected relative number of reads; and

subjecting all of the normalized number of reads falling in each of the plurality of primary windows Z_i to Run-Test.

44. The computer readable medium of claim 43, wherein the corrected relative number of reads $\tilde{R}_i$ is obtained by following steps:

calculating GC content of each of the plurality of primary windows;

dividing the GC content into a plurality of regions in a unit of 0.001, and calculating a

mean value M_s among all of the relative numbers of reads falling in each of the plurality of regions, wherein s is No. of the plurality of regions;

determining the corrected relative number of reads $\tilde{R}_i$ based on the following formula:

$$\tilde{R}_i = R_i - M_s;$$

determining the normalized number of reads Z_i based on the following formula:

wherein

$$Z_i = (R_i - \tilde{R}_i - mean) / SD, \text{ wherein}$$

$$mean = \frac{1}{n} \sum_{i=1}^n (R_i - \tilde{R}_i),$$

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (R_i - \tilde{R}_i - mean)^2}$$

45. The computer readable medium of claim 43, wherein based on the plurality of the breakpoints, the step of determining a detection window in the reference genome further comprises:

1) determining a plurality of candidate breakpoints, wherein other breakpoints present both before and after the candidate breakpoints;

2) determining p value of each candidate breakpoint, and removing a candidate breakpoint having the maximal p value;

3) performing the step 2) with rest of the candidate breakpoints until all p values of the rest of the candidate breakpoints are smaller than the final p value, wherein the rest of the candidate breakpoints are taken as screened candidate breakpoints; and

4) determining a region between two successive screened candidate breakpoints as the detection window,

wherein the p value of the candidate breakpoint is obtained by following steps:

selecting a region between the candidate breakpoint and previous candidate breakpoint as a first candidate region, and selecting a region between the candidate breakpoint and next candidate breakpoint as a second candidate region;

subjecting the normalized number of reads falling in the primary windows Z_i which are included both in the first candidate region and the second candidate region to

Run-Test, to determine the p value of the candidate breakpoints,

optionally, the final p value is obtained by following steps:

based on a sequencing result of a control sample, repeating the step of determining a detection window in the reference genome, and recording p values of the breakpoints which are removed each time until the number of the breakpoints is zero; and

determining the final p value, based on p values of removed breakpoints,

optionally, the final p value is 1.1×10^{-50} .

46. The computer readable medium of claim 45, wherein based on the reads falling in the detection window, the step of determining a first parameter further comprises:

determining a mean value among all of the normalized numbers of reads falling in all each of the plurality of primary windows $\bar{Z}$ which are included in the detection windows, wherein the mean value of the normalized numbers of reads $\bar{Z}$ is taken as the first parameter.

47. The computer readable medium of claim 46, wherein the preset threshold comprises: a first threshold being -1.645 and a second threshold being 1.645.

48. The computer readable medium of claim 36, wherein the reference genome sequence is at least one selected from human chromosome 21, chromosome 18, chromosome 13, chromosome X and chromosome Y.

WHAT IS CALIMED IS:

1. A method of determining whether copy number variation presents in a genome sample, comprising following steps:

sequencing the genome sample, to obtain a sequencing result consisting of a plurality of reads;

aligning the sequencing result to a reference genome sequence, to determine a distribution of the reads in the reference genome sequence;

determining a plurality of breakpoints in the reference genome sequence based on the distribution of the reads in the reference genome sequence, wherein the number of reads has significance at both sides of the breakpoints;

determining a detection window in the reference genome based on the plurality of the breakpoints;

determining a first parameter based on reads falling in the detection window; and

determining whether the copy number variation presents in the genome sample against the detection window based on difference between the first parameter and a preset threshold.

2. The method of claim 1, further comprising a step of extracting the genome sample from a biological sample.

3. The method of claim 2, wherein the biological sample is pregnant sample or fetal sample.

4. The method of claim 2, wherein the biological sample is at least one selected from a group consisting of blood, urine, saliva, tissue, germ cells, oosperm, blastomere and embryo.

5. The method of claim 1, wherein the step of sequencing the genome

sample further comprises following sub-steps of:

amplifying the genome sample;
constructing a sequencing-library with amplified genome sample; and
sequencing the sequencing-library.

6. The method of claim 4, further comprising a step of:

lysing the single cell, to release whole genome of the single cell.

7. The method of claim 6, wherein the single cell is lysed using an alkaline lysate, to release whole genome of the single cell.

8. The method of claim 7, wherein the whole genome is amplified by means of PCR-based whole genome amplification method.

9. The method of claim 8, wherein the PCR-based whole genome amplification method is OmniPlex WGA.

10. The method of claim 5, wherein the sub-step of sequencing the sequencing-library is performed by at least one selected from a group consisting of Hiseq system, Miseq system, Genome Analyzer system, 454 FLX, SOLiD system, Ion Torrent system and single molecule sequencing apparatus.

11. The method of claim 1, wherein the copy number variation is at least one selected from aneuploidy of chromosome, deletion of chromosome, and addition micro-deletion and micro-repetition of chromosome fragments.

12. The method of claim 1, wherein the step of determining a plurality of breakpoints in the reference genome sequence further comprises following sub-steps:

dividing the reference genome sequence into a plurality of primary windows having a predetermined length, and determining reads falling in each

of the plurality of the primary windows;

for at least one site in the reference genome sequence, determining the number of reads falling in the same number of the plurality of primary windows at both sides of the site;

determining p value of the site, wherein the p value represents that the number of reads falling in either side of the site has significance; and

determining that the site is the breakpoints, if the p value of the site is smaller than a final p value.

13. The method of claim 12, wherein the reads falling in each of the plurality of primary windows are uniquely-aligned reads.

14. The method of claim 12, wherein 100 of primary windows are selected from either side of the site.

15. The method of claim 12, wherein the plurality of primary windows have a length of 100 Kbp to 200 Kbp.

16. The method of claim 12, wherein the final p value is 1.1×10^{-50} or less.

17. The method of claim 12, wherein the sub-step of determining p value of the site further comprises:

for the site, selecting the plurality of primary windows having the same number at either side of the site respectively, and calculating the relative number of reads falling in each of the plurality of primary windows R_i , wherein i represents No. of the plurality of primary windows,

subjecting all of the relative numbers of reads falling in the plurality of primary windows R_i to Run-Test, to determine the p value of the site,

wherein the relative number of reads is determined by following formula:

$$R_i = \log_2 \left(\frac{r_i}{\bar{r}} \right)$$

wherein n represents the number of reads falling in the i -th primary window,

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n r_i,$$

n represents the total number of the plurality of primary windows.

18. The method of claim 17, wherein subjecting all of the relative numbers of the reads falling in each of the plurality of primary windows to Run-Test further comprises:

subjecting the relative number of reads falling in each of the plurality of primary windows R_i to a correction of GC content, to obtain corrected relative number of reads $\tilde{R}_i$;

determining the normalized number of reads falling in each of the plurality of primary windows Z_i based on the corrected relative number of reads; and

subjecting all of the normalized numbers of reads falling in each of the plurality of primary windows Z_i to Run-Test.

19. The method of claim 18, wherein the corrected relative number of reads $\tilde{R}_i$ is obtained by following steps:

calculating GC content of each of the plurality of primary windows;

dividing the GC content into a plurality of regions in a unit of 0.001, and calculating a mean value M_s among all of the relative numbers of reads falling in each of the plurality of regions, wherein s is No. of the plurality of regions;

determining the corrected relative number of reads $\tilde{R}_i$ based on the following formula:

$$\tilde{R}_i = R_i - M_s;$$

determining the normalized number of reads Z_i based on the following formula:

wherein

$$Z_i = (R_i - \tilde{R}_i - mean) / SD,$$

$$mean = \frac{1}{n} \sum_{i=1}^n (R_i - \tilde{R}_i)$$

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (R_i - \tilde{R}_i - mean)^2}$$

20. The method of claim 19, wherein the step of determining a detection window in the reference genome based on the plurality of the breakpoints further comprises:

1) determining a plurality of candidate breakpoints, wherein other breakpoints present both before and after the candidate breakpoints;

2) determining p value of each candidate breakpoint, and removing a candidate breakpoint having the maximal p value;

3) performing the step 2) with rest of the candidate breakpoints until all p values of the rest of the candidate breakpoints are smaller than the final p value, wherein the rest of the candidate breakpoints are taken as screened candidate breakpoints; and

4) determining a region between two successive screened candidate breakpoints as the detection window,

wherein the p value of the candidate breakpoint is obtained by following steps:

selecting a region between the candidate breakpoint and previous candidate breakpoint as a first candidate region, and selecting a region between the candidate breakpoint and next candidate breakpoint as a second candidate region;

subjecting the normalized number of reads falling in the primary windows Z_i which are included both in the first candidate region and the second candidate region to Run-Test, to determine the p value of the candidate breakpoints.

21. The method of claim 20, wherein the step of determining a first parameter based on the reads falling in the detection window further comprises:

determining a mean value among all of the normalized numbers of reads falling in each of the plurality of primary windows $\bar{Z}$ which are included in the detection windows, wherein the mean value of the normalized numbers of reads $\bar{Z}$ is taken as the first parameter.

22. The method of claim 1, wherein the preset threshold comprises:
a first threshold being -1.645 and a second threshold being 1.645.

23. The method of claim 1, wherein the reference genome sequence is at least one selected from human chromosome 21, chromosome 18, chromosome 13, chromosome X and chromosome Y.

24. A system for determining whether copy number variation presents in a genome sample, comprising:

a sequencing apparatus, configured to sequence the genome sample, to obtain a sequencing result consisting of a plurality of reads;

an analysis apparatus, connected to the sequencing apparatus, configured to determine whether copy number variation presents in the genome sample based on the sequencing result, wherein the analysis apparatus further comprises:

an aligning unit, configured to align the sequencing result to a reference genome sequence, to determine a distribution of the reads in the reference genome sequence;

a breakpoint determining unit, connected to the aligning unit, configured to determine a plurality of breakpoints in the reference genome sequence based on the distribution of the reads in the reference genome sequence, wherein the number of reads has significance between two sides of the breakpoints;

a detection window determining unit, connected to the breakpoint

determining unit, configured to determine a detection window in the reference genome based on the plurality of the breakpoints;

a parameter determining unit, connected to the detection window determining unit, configured to determine a first parameter based on reads falling in the detection window; and

a determining unit, connected to the parameter determining unit, configured to determine whether the copy number variation presents in the genome sample against the detection window based on difference between the first parameter and a preset threshold.

25. The system of claim 24, further comprising a genome extracting apparatus, configured to extract the genome sample from a biological sample.

26. The system of claim 24, wherein the sequencing apparatus further comprises:

a genome amplifying unit, configured to amplify the genome sample;

a sequencing-library constructing unit, connected to the genome amplifying unit, configured to construct a sequencing-library with amplified genome sample; and

a sequencing unit, connected to the sequencing-library constructing unit, configured to sequence the sequencing-library.

27. The system of claim 26, wherein the sequencing unit is at least one selected from a group consisting of Hiseq system, Miseq system, Genome Analyzer system, 454 FLX, SOLiD system, Ion Torrent system and single molecule sequencing apparatus.

28. The system of claim 24, wherein the breakpoint determining unit further comprises a module for performing following sub-steps:

dividing the reference genome sequence into a plurality of primary windows having a predetermined length, and determining reads falling in each

of the plurality of the primary windows;

for at least one site in the reference genome sequence, determining the number of reads falling in the same number of the plurality of primary windows at both sides of the site;

determining p value of the site, wherein the p value represents that the number of reads falling in either side of the site has significance; and

determining the site is the breakpoints, if the p value of the site is smaller than a final p value.

29. The system of claim 28, wherein the breakpoint determining unit further comprises a module for performing followings to determine the p value:

for the site, selecting the plurality of primary windows having the same number at either side of the site respectively, and calculating the relative number of reads falling in each of the plurality of primary windows R_i , wherein i represents No. of the plurality of primary windows,

subjecting all of the relative numbers of reads falling in the plurality of primary windows R_i to Run-Test, to determine the p value of the site,

wherein the relative number of reads is determined by following formula:

$$R_i = \log_2 \left(\frac{r_i}{\bar{r}} \right)$$

wherein r_i represents the number of reads falling in the i -th primary window,

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n r_i,$$

n represents the total number of the plurality of primary windows.

30. The system of claim 29, wherein the breakpoint determining unit further comprises a module for performing followings to subject all of the relative numbers of the reads falling in the plurality of primary windows to Run-Test:

subjecting the relative number of reads falling in each of the plurality of primary windows R_i to a correction of GC content, to obtain corrected relative number of reads $\tilde{R}_i$;

determining the normalized number of reads falling in each of the plurality of primary windows Z_i based on the corrected relative number of reads; and

subjecting all of the normalized numbers of reads falling in each of the plurality of primary windows Z_i to Run-Test.

31. The system of claim 30, wherein the corrected relative number of reads $\tilde{R}_i$ is obtained by a module for performing following steps:

calculating GC content of each of the plurality of primary windows;

dividing the GC content into a plurality of regions in a unit of 0.001, and calculating a mean value M_s among all of the relative numbers of reads falling in each of the plurality of regions, wherein s is No. of the plurality of regions;

determining the corrected relative number of reads $\tilde{R}_i$ based on the following formula:

$$\tilde{R}_i = R_i - M_s;$$

determining the normalized number of reads Z_i based on the following formula:

wherein

$$Z_i = (R_i - \tilde{R}_i - mean) / SD, \text{ wherein}$$

$$mean = \frac{1}{n} \sum_{i=1}^n (R_i - \tilde{R}_i)$$

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (R_i - \tilde{R}_i - mean)^2}$$

32. The system of claim 31, wherein based on the plurality of the breakpoints, the detection window determining unit further comprises a module for performing followings:

1) determining a plurality of candidate breakpoints, wherein other breakpoints present both before and after the candidate breakpoints;

2) determining p value of each candidate breakpoint, and removing a

candidate breakpoint having the maximal p value;

3) performing the step 2) with rest of the candidate breakpoints until all p values of the rest of the candidate breakpoints are smaller than the final p value, wherein the rest of the candidate breakpoints are taken as screened candidate breakpoints; and

4) determining a region between two successive screened candidate breakpoints as the detection window,

wherein the p value of the candidate breakpoint is obtained by following steps:

selecting a region between the candidate breakpoint and previous candidate breakpoint as a first candidate region, and selecting a region between the candidate breakpoint and next candidate breakpoint as a second candidate region;

subjecting the normalized number of reads falling in the primary windows Z_i which are included both in the first candidate region and the second candidate region to Run-Test, to determine the p value of the candidate breakpoints.

33. The system of claim 32, wherein the parameter determining unit further comprises a module for performing followings:

determining a mean value among all of the normalized numbers of reads falling in each of the plurality of primary windows $\bar{Z}$ which are included in the detection windows, wherein the mean value of the normalized numbers of reads $\bar{Z}$ is taken as the first parameter.

34. The system of the claim 24, wherein a preset threshold is preserved in the determining unit, wherein the preset threshold comprises: a first threshold being -1.645 and a second threshold being 1.645.

35. The system of claim 24, wherein the reference genome sequence is

preserved in the aligning unit, wherein the reference genome sequence is known human genome sequence.