

CLAIMS

1. A method of analyzing a gene, the method comprising:

acquiring data for split reads and discordantly aligned paired-end (PE) reads from
5 next-generation sequencing (NGS) data of a sample of a subject acquired by
deep-sequencing;

extracting first candidate gene pairs having a possibility of translocation in a
chromosome of the sample by using break points that are indicated by the split reads and the
PE reads; and

10 extracting, from among the extracted first candidate gene pairs, second candidate
gene pairs comprising genes in which a plurality of split reads having breakpoints within
the same coverage are aligned;

extracting, from among the extracted second candidate gene pairs, third candidate
gene pairs in which the direction of fusion between different genes goes from 5' end to 3'
15 end or from 3' end to 5' end; and

identifying translocation genes from among the extracted third candidate gene pairs,
wherein, in the genes included in the extracted second candidate gene pairs, the
number of the split reads having the breakpoints within the same coverage is a
predetermined threshold value or greater.

20 2. The method of claim 1, wherein the NGS data comprises data in a binary
version of sequence alignment/map (SAM) (BAM) format or in an SAM format.

3. The method of claim 2, wherein the acquiring comprises acquiring, from the
25 data in a BAM format or in an SAM format, data of an FLAG and a compact idiosyncratic
gapped alignment report (CIGAR) string for each of the split reads and the PE reads.

4. The method of claim 1, wherein the NGS data is generated by targeted
sequencing for identifying base sequences of target genes from the sample.

30 5. The method of claim 1, wherein the sample is a biopsy sample or a
formalin-fixed, paraffin-embedded (FFPE) sample.

6. A computer-readable recording medium recording a program for executing the method according to any one of claims 1 to 5 in a computer.

7. An apparatus for analyzing a gene, the apparatus comprising:

5 a read analyzing unit to acquire, from next-generation sequencing (NGS) data of a sample of a subject acquired by deep-sequencing, data for split reads and discordantly aligned paired-end (PE) reads; and

a translocation identifying unit to extract first candidate gene pairs having a possibility of translocation in a chromosome of the sample by using break points that are indicated by the split reads and the PE reads, extract, from among the extracted first candidate gene pairs, second candidate gene pairs comprising genes in which a plurality of split reads having breakpoints within the same coverage are aligned, extract, from among the extracted second candidate gene pairs, third candidate gene pairs in which the direction of fusion between different genes goes from 5' end to 3' end or from 3' end to 5' end, and
10 identify translocation genes from among the extracted third candidate gene pairs,

wherein, in the genes included in the extracted second candidate gene pairs, the number of the split reads having the breakpoints within the same coverage is a predetermined threshold value or greater.

20 8. The apparatus of claim 7, wherein the NGS data comprises data in a binary version of sequence alignment/map (SAM) (BAM) or in an SAM format.

9. The apparatus of claim 8, wherein the read analyzing unit acquires, from the data in a BAM format or in an SAM format, data of an FLAG and a compact idiosyncratic
25 gapped alignment report (CIGAR) string for each of the split reads and the PE reads.

10. The apparatus of claim 7, wherein the NGS data is generated by targeted sequencing for identifying base sequences of target genes from the sample.

30 11. The apparatus of claim 7, wherein the sample is a biopsy sample or a formalin-fixed, paraffin-embedded (FFPE) sample.