

【CLAIMS】

【Claim 1】 A method of detecting a gene rearrangement in a sample comprising:

(a) acquiring a library including a plurality of nucleic acid molecules from a sample of a subject;

(b) bringing the library into contact with a plurality of bait sets to enrich the library with respect to a preselected sequence to provide a selected nucleic acid molecule and thereby provide a library catch;

(c) acquiring reads from the nucleic acid molecules of the library catch through next-generation sequencing;

(d) aligning the reads to a reference genome;

(e) filtering the reads with information of the region of interest and dividing the filtered reads to concordant read pair having secondary alignment tag (SA tag) information, discordant read pair having SA tag information, and concordant read pair not having SA tag information, and discordant read pair not having SA tag information by matching filtered reads with reference genome,

wherein the information of the region of interest comprises information on the location of a known fusion gene and information on a target gene region;

wherein information on the location of a known fusion gene and information on a target gene region comprises chromosome information and information of start and end

positions on the chromosome;

(f) for concordant read pair having SA tag information, finding a matching region of a virtual second read that forms a pair using a soft-clip segment portion of a first read as a query, or finding a matching region of ~~the~~ a virtual first read that forms a pair using a soft-clip segment portion of the second read as a query by using blast, while assuming, a secondary mapping region, obtained by determining whether or not there is mapping to other genes with reference to secondary alignment tag information of the first read, as ~~a~~ the virtual second read, ~~a secondary mapping region obtained by determining whether or not there is mapping to other genes with reference to secondary alignment tag information of the first read;~~

(g) for discordant read pair having SA tag information, finding a matching region of a second read that forms a pair using a soft-clip segment portion of a first read as a query, or finding a matching region of the first read that forms a pair using a soft-clip segment portion of the second read as a query by using blast;

(h) deriving read pair, the matching region of which is found in each step (f) and (g) as a rearrangement gene candidate group to determine a supporting pair count, wherein the supporting pair count is increased when the rearrangement gene candidate group is determined in both steps (f) and (g)

~~than supporting pair count of the rearrangement gene candidate group is determined in each step; and~~

(i) determining the rearrangement gene candidate group having a supporting pair count equal to or greater than a cut-off value as a rearrangement gene.

【Claim 2】 The method according to claim 1, wherein the information of the region of interest comprises Table 1.

【Claim 3】 The method according to claim 1, wherein the supporting pair counts is not increased when a position of the gene rearrangement is different, although a type of the gene rearrangement is identical.

【Claim 4】 The method according to claim 1, wherein the cut-off value of supporting pair count is 2.

【Claim 5】 The method according to claim 1, further comprising,

(j) producing a gene rearrangement template based on the gene rearrangement candidate group derived in step (h);

(k) determining a supporting read count as a number of reads comprising a breakpoint of the gene rearrangement in the gene rearrangement template after performing blast using

the concordant read pair not having SA tag information, discordant read pair not having SA tag information obtained from step (e) as blastdb and using the gene rearrangement template produced in step (j) as a query; and

(l) determining the rearrangement gene candidate group of step (h) having a supporting read counts equal to or greater than a cut-off value as a rearrangement gene,

wherein the gene rearrangement template is a base sequence on the reference genome including 300 bp to 500 bp in the 5' direction and 300 bp to 500 bp in the 3' direction from the gene rearrangement position.

【Claim 6】 The method according to claim 5, wherein cut-off value of supporting read count is 5.

【Claim 7】 The method according to claim 5, wherein the concordant read pair not having SA tag information and the discordant read pair not having SA tag information comprise a soft-clip segment.

【Claim 8】 A computer system comprising a computer-readable medium encoded with a plurality of instructions for controlling a computing system to perform a method of

detecting a gene rearrangement using next-generation sequencing (NGS),

wherein the method comprises:

(a) acquiring a library including a plurality of nucleic acid molecules from a sample of a subject;

(b) bringing the library into contact with a plurality of bait sets to enrich the library with respect to a preselected sequence to provide a selected nucleic acid molecule and thereby provide a library catch;

(c) acquiring reads from the nucleic acid molecules of the library catch through next-generation sequencing;

(d) aligning the reads to a reference genome;

(e) filtering the reads with information of the region of interest and dividing the filtered reads to concordant read pair having secondary alignment tag(SA tag) information, discordant read pair having SA tag information, and concordant read pair not having SA tag information, and discordant read pair not having SA tag information by matching filtered reads with reference genome,

wherein the information of the region of interest comprises information on the location of a known fusion gene and information on a target gene region;

wherein information on the location of a known fusion gene and information on a target gene region comprises

chromosome information and information of start and end positions on the chromosome;

(f) for concordant read pair having SA tag information, finding a matching region of a virtual second read that forms a pair using a soft-clip segment portion of a first read as a query, or finding a matching region of ~~the~~ a virtual first read that forms a pair using a soft-clip segment portion of the second read as a query by using blast, while assuming, a secondary mapping region, obtained by determining whether or not there is mapping to other genes with reference to secondary alignment tag information of the first read, as ~~a~~ the virtual second read, ~~a secondary mapping region obtained by determining whether or not there is mapping to other genes with reference to secondary alignment tag information of the first read;~~

(g) for discordant read pair having SA tag information, finding a matching region of a second read that forms a pair using a soft-clip segment portion of a first read as a query, or finding a matching region of the first read that forms a pair using a soft-clip segment portion of the second read as a query by using blast;

(h) deriving read pair, the matching region of which is found in each step (f) and (g) as a rearrangement gene candidate group to determine a supporting pair count,

wherein the supporting pair count is increased when the rearrangement gene candidate group is determined in both steps (f) and (g) ~~than supporting pair count of the rearrangement gene candidate group is determined in each step;~~ and

(i) determining the rearrangement gene candidate group having a supporting pair count equal to or greater than a cut-off value as a rearrangement gene.

【Claim 9】 The method according to any one of claims 1 to 8, wherein the sample is selected from the group consisting of: one or more premalignant or malignant cells; cells selected from solid tumors, soft tissue tumors or metastatic lesions; tissue or cells from surgical resections; histologically normal tissue; at least one blood tumor cell (CTC); and blood samples from the same subject having or at risk of developing a normal adjacent tumor (NAT) and a tumor.