

Claims

1. A method for determining the nucleotide sequence of a nucleic acid of interest in a sample, comprising:

1) providing n samples, wherein n is an integer of ≥ 1 ; optionally, the n samples to be analyzed are divided into m groups, m is an integer and $n \geq m \geq 1$;

2) amplifying: a pair or multiple pairs of index primers are used for each sample, when there are templates from the sample, PCR amplification is performed under conditions suitable for amplifying the nucleic acid of interest, wherein each pair of index primers consist of a forward index primer and a reverse index primer comprising primer indexes, wherein the primer indexes comprised in the forward index primer and reverse index primer may be identical or different: the primer indexes in the pairs of index primers used for different samples are different;

wherein each pair of index primers are formed by a pair of primer indexes and a pair of PCR primers, wherein forward and reverse PCR primers have a forward primer index and a reverse primer index optionally linked by a linker sequence at 5' end, respectively; and

wherein said pair of primer indexes are selected from the 95 pairs of primer indexes as shown in Table 6; and said PCR primers are selected from SEQ ID NOs: 1-40 and 231-238;

3) pooling: when $n > 1$, pooling PCR products from each of the samples together:

4) shearing: subjecting the amplified products to incomplete shearing, and purifying and recovering; wherein said incomplete shearing produces intact un-sheared PCR products and partially sheared PCR products;

5) sequencing: subjecting the recovered DNA mixture to sequencing by using the second generation sequencing technique, to obtain sequences of the sheared DNA; and

6) assembling: corresponding the obtained sequencing data to samples one by one based on the unique primer index for each sample, aligning each sequence read to the DNA reference sequence corresponding to the PCR products by using alignment program, assembling a complete sequence of the nucleic acid of interest from the sequences of the sheared DNA by virtue of sequence overlapping and linkage relationship, whereby a length of the complete sequence of the nucleic acid of interest exceeds a maximum read length of a sequencer used for said sequencing.

2. A method of claim 1, wherein said PCR primers are selected from PCR primers for amplification of Exons 2, 3 and 4 of HLA-A/B as shown in Table 1 or Table 2, or PCR primers for amplification of Exon 2 of HLA-DRB1 as shown in Table 7; PCR primers for amplification of Exons 2, 3 and/or 4 of HLA-C as shown in Table 3 or Table 4; or PCR primers for amplification of Exon 2 and/or 3 of HLA-DQB1 gene as shown in Table 5.

3. A method of claim 1 or 2, wherein said primer indexes comprise at least 10, or at least 20, or at least 30, or at least 40, or at least 50, or at least 60, or at least 70, or at least 80, or at least 90, or 95 pairs of the 95 pairs of primer indexes as shown in Table 6; and/or

wherein said primer indexes consisting of 10-95 pairs 20-95 pairs, 30-95 pairs, 40-95 pairs, 50-95 pairs, 60-95 pairs, 70-95 pairs, 80-95 pairs, 90-95 pairs, or 95 pairs of the 95 pairs of primer indexes as shown in Table 6; and/or

wherein said primer indexes comprises at least PI-1 to PI-10, or PI-11 to PI-20, or PI-21 to PI-30, or PI-31 to PI-40, or PI-41 to PI-50, or PI-51 to PI-60, or PI-61 to PI-70, or PI-71 to PI-80, or PI-81 to PI-90, or PI-91 to PI-95 of the 95 pairs of primer indexes as shown in Table 6, or combinations of any two or more of them.

4. A method of anyone of claims 1-3, wherein said DNA shearing includes chemical shearing methods and physical shearing methods, wherein the chemical shearing methods include enzymatic digestion, and the physical shearing methods include ultrasonic shearing methods or mechanical shearing methods.

5. A method of anyone of claims 1-4, wherein after said DNA shearing, all DNA fragments between the maximum read length of the sequencer and the applicable maximum DNA length of the sequencer are purified and recovered, wherein said purification and recovery methods include recovery by electrophoresis and gel slicing, and recovery by magnetic beads.

6. A method of claim 1, comprising, in addition to steps 1) to 4) of claim 1, the following steps:

5) constructing a library: constructing a PCR-free sequencing library by using the

library of the sheared PCR products, wherein different library adapters may be added to distinguish different PCR-Free sequencing libraries, all DNA fragments between the maximum read length of the sequencer and the applicable maximum DNA length of the sequencer are purified and recovered;

6) sequencing: subjecting the recovered DNA mixture to sequencing by using the second generation sequencing technique, obtaining the sequences of the sheared DNAs;

7) assembling: corresponding the obtained sequencing data to the samples one by one based on different library adapter sequences of the libraries and the unique primer index for each sample, aligning each sequence read to the DNA reference sequence corresponding to the PCR products by using alignment program, assembling a complete sequence of the nucleic acid of interest from the sequences of the sheared DNA based on sequence overlapping and linkage relationship.

7. A method of claim 6, wherein in step 5, the DNA fragments of 450 to 750 bp are purified and recovered.

8. A method of claim 1 or 6, wherein the forward index primer and/or the reverse index primer are degenerate primers; and/or wherein the second generation sequencing technique is Paired-End technique.

9. A method of claim 1 or 6, wherein the second generation sequencing technique is Illumina GA or Illumina Hiseq 2000 sequencing technique.

10. A method of claim 1 or 6, wherein the alignment program is Blast or BWA program.

11. A method of anyone of claims 1-10, wherein the samples are from mammalian.

12. A method of claim 11, wherein the samples are from human.

13. A method of claim 11, wherein the samples are human blood sample.

14. Use of a method of any one of claims 1 to 13 in HLA typing, characterized by comprising: sequencing a sample from a patient by the method of any one of claims 1 to 13, and aligning the sequencing results with sequence data of Exons 2, 3, 4 of HLA-A/B, Exons 2, 3 and/or 4 of HLA-C, Exon 2 and/or 3 of HLA-DQB1 gene and/or Exon 2 of HLA-DRB1 in HLA database; wherein if the result of sequence alignment shows 100% match, the HLA genotype of the corresponding sample is determined.

15. Use of claim 14, wherein the sample is blood sample; and/or wherein the HLA database is IMGT HLA professional database.

16. A method of anyone of claims 1-13, wherein the method is used for HLA typing and further comprises the following steps:

typing: aligning the sequencing results with sequence data of Exons 2, 3, 4 of HLA-A/B, Exons 2, 3 and/or 4 of HLA-C, Exon 2 and/or 3 of HLA-DQB1 gene and/or Exon 2 of HLA-DRB1 in HLA database, wherein if the result of sequence alignment shows 100% match, the HLA genotype of the corresponding sample is determined.

17. A method of claim 16, wherein the HLA database is IMGT HLA professional database.