

resequencing data.

For example, Figure 3, to which is referred, is a schematic diagram 1 showing a FPGA-based resequencing analysis system according to examples of the present disclosure, where the resequencing analysis system includes a task scheduling unit 1, an analyzing unit 2 and
5 a result generating unit 3. The task scheduling unit 1 is configured to receive the basic genome resequencing data and the parameter information describing the basic genome resequencing data.

At Step 202: a first task for outputting the alignment result and a second task for sorting and deduplicating the alignment result are determined based on the basic genome resequencing data and the parameter information describing the basic genome resequencing data.

10 In examples of the present disclosure, the processes of aligning, sorting and deduplicating are executed in parallel, thus reducing program running time and calculation cost, thereby improving resequencing efficiency. Therefore, once the genome resequencing data is received, the first task for outputting the alignment result and the second task for sorting and deduplicating the alignment result both may be determined in parallel based on the basic genome resequencing
15 data and the parameter information describing the basic genome resequencing data.

For example, referring to Figure 3, the task scheduling unit 1 may be configured to determine both the first task for outputting the alignment result and the second task for sorting and deduplicating the alignment result in parallel based on the basic genome resequencing data and the parameter information describing the basic genome resequencing data. The first task and
20 the second task may be then sent to the analyzing unit 2, which is process-scheduled. For example, the analyzing unit 2 may be process-scheduled by a bioinformatics software and a system scheduling script.

At Step 203: a first thread corresponding to the first task and a second thread corresponding to the second task are allocated from the FPGA and a CPU, respectively.

25 In examples of the present disclosure, after the first task and the second task are determined, the first thread corresponding to the first task and the second thread corresponding to the second task may be allocated from the FPGA and the CPU, respectively.

At Step 204: the first thread corresponding to the first task and the second thread corresponding to the second task are started and executed in parallel, such that the genome
30 resequencing data is called via respective threads.

method of the present disclosure are basically as same as that in the related technology; meanwhile although the accuracy in INDEL detection by the claimed method of the present disclosure is slightly worse than that in the related technology, the sensitivity basically maintains. Therefore, the overall result obtained by the claimed method of the present disclosure is
5 equivalent to the related technology.

In order to implement the above examples, the present disclosure further provides in an example a FPGA-based resequencing analysis device.

Figure 9 is a schematic diagram showing a FPGA-based resequencing analysis device in the Example 4 of the present disclosure.

10 As shown in Figure 9, the FPGA-based resequencing analysis device 100 includes a receiving module 101, a processing module 120, a correcting module 130 and a detecting module 140.

The receiving module 101 is configured to receive genome resequencing data.

The processing module 120 is configured to take the resequencing data as an input of an
15 FPGA, to determine an alignment result in a process of resequencing based on an output of the FPGA, and at the same time, to sort and deduplicate the alignment result.

The correcting module 130 is configured to subject a sorted and deduplicated alignment result to correction in accordance with a base quality value.

The detecting module 140 is configured to detect a variation result based on a corrected
20 alignment result.

Further, as a possible implementation of the present disclosure, referring to Figure 10 on the basis of Figure 9, the FPGA-based resequencing analysis device 100 may further include: a calculating module 150, configured to subject the genome resequencing data to hash-calculation based on a sequence of a preset length after the genome resequencing data is received.

25 Accordingly, the processing module 120 is specifically configured to take hash-calculated genome resequencing data as the input of the FPGA.

As a possible implementation, the genome resequencing data include basic genome resequencing data and parameter information describing the basic genome resequencing data.

The FPGA-based resequencing analysis device 100 may further include:

30 a determining module 160, configured to determine a first task for outputting the alignment result and a second task for sorting and deduplicating the alignment result based on the basic genome resequencing data and the parameter information describing the basic genome

What is claim is:

1. A Field-Programmable Gate Array (FPGA)-based resequencing analysis method, comprising:

receiving genome resequencing data;

taking the genome resequencing data as an input of an FPGA, determining an alignment result in a process of resequencing based on an output of the FPGA, and at the same time, sorting and deduplicating the alignment result;

subjecting a sorted and deduplicated alignment result to correction in accordance with a base quality value; and

detecting a variation result based on a corrected alignment result.

2. The FPGA-based resequencing analysis method of claim 1, wherein subsequent to receiving genome resequencing data, the method further comprises:

subjecting the genome resequencing data to hash-calculation based on a sequence of a preset length,

taking the genome resequencing data as an input of an FPGA further comprising

taking hash-calculated genome resequencing data as the input of the FPGA.

3. The FPGA-based resequencing analysis method of claim 1, wherein the genome resequencing data comprises basic genome resequencing data and parameter information describing the basic genome resequencing data,

wherein subsequent to receiving genome resequencing data, the method further comprises:

determining a first task for outputting the alignment result and a second task for sorting and deduplicating the alignment result based on the basic genome resequencing data and the parameter information describing the basic genome resequencing data; and

allocating a first thread corresponding to the first task from the FPGA and a second thread corresponding to the second task from a CPU.

4. The FPGA-based resequencing analysis method of claim 3, wherein taking the genome resequencing data as an input of an FPGA, determining an alignment result in a process of

resequencing based on an output of the FPGA, and at the same time, sorting and deduplicating the alignment result further comprises:

starting and executing the first thread corresponding to the first task and the second thread corresponding to the second task in parallel, such that the genome resequencing data is called via
5 respective threads; and

determining the alignment result in the process of resequencing via an output result of the first thread corresponding to the first task, and at the same time, sorting and deduplicating the alignment result via an output result of the second thread corresponding to the second task.

10 5. The FPGA-based resequencing analysis method of claim 4, wherein determining the alignment result in the process of resequencing via an output result of the first thread corresponding to the first task further comprises:

executing the first thread corresponding to the first task to determine a targeted alignment position;

15 triggering alignment on duplicate base pairs in the basic genome resequencing data in accordance with the targeted alignment position; and

taking an alignment result of the duplicate base pairs as the alignment result in the process of resequencing.

20 6. The FPGA-based resequencing analysis method of claim 4, wherein subsequent to determining a targeted alignment position, the method further comprises:

 caching the targeted alignment position by using a ping-pong buffering technology.

7. A FPGA-based resequencing analysis device, comprising:

25 a receiving module, configured to receive genome resequencing data;

 a processing module, configured to take the resequencing data as an input of an FPGA, to determine an alignment result in a process of resequencing based on an output of the FPGA, and at the same time, to sort and deduplicate the alignment result;

 a correcting module, configured to subject a sorted and deduplicated alignment result to
30 correction in accordance with a base quality value; and

 a detecting module, configured to detect a variation result based on a corrected alignment

result.

8. The FPGA-based resequencing analysis device of claim 7, wherein the FPGA-based resequencing analysis device further comprising: a calculating module, configured to subject the genome resequencing data to hash-calculation based on a sequence of a preset length after the genome resequencing data is received; and

wherein the processing module is specifically configured to take hash-calculated genome resequencing data as the input of the FPGA.

9. The FPGA-based resequencing analysis device of claim 7, wherein the genome resequencing data comprises basic genome resequencing data and parameter information describing the basic genome resequencing data, wherein the FPGA-based resequencing analysis device further comprises:

a determining module, configured to determine a first task for outputting the alignment result and a second task for sorting and deduplicating the alignment result based on the basic genome resequencing data and the parameter information describing the basic genome resequencing data after the genome resequencing data is received; and

an allocating module, configured to allocate a first thread corresponding to the first task from the FPGA and a second thread corresponding to the second task from a CPU.

10. The FPGA-based resequencing analysis device of claim 9, wherein the processing module further comprises:

a starting and executing unit, configured to start and execute the first thread corresponding to the first task and the second thread corresponding to the second task in parallel, so as to call the genome resequencing data via respective threads; and

a determining and processing unit, configured to determine the alignment result in the process of resequencing via an output result of the first thread corresponding to the first task, and at the same time, to sort and deduplicate the alignment result via an output result of the second thread corresponding to the second task.

11. The FPGA-based resequencing analysis device of claim 10, wherein the processing

module is specifically configured to:

execute the first thread corresponding to the first task to determine a targeted alignment position;

trigger alignment on duplicated base pairs in the basic genome resequencing data in accordance with the targeted alignment position; and

take an alignment result of the duplicated base pairs as the alignment result in the process of resequencing.

12. The FPGA-based resequencing analysis device of claim 10, wherein the determining unit is further configured to cache the targeted alignment position by using a ping-pong buffering technology.

13. A computer device, comprising:

a memory;

a processor; and

a computer program stored in the memory and executable by the processor,

wherein the computer program, when executed by the processor, achieves a Field-Programmable Gate Array (FPGA)-based resequencing analysis method according to any one of claims 1 to 6.

14. A computer-readable storage medium having stored therein computer-readable instructions that causes the computer to perform a Field-Programmable Gate Array (FPGA)-based resequencing analysis method according to any one of claims 1 to 6.