

WHAT IS CLAIMED IS:

1. A method of determining base information of a predetermined region in a fetal genome, comprising following steps:

constructing a sequencing library based on a genomic DNA sample of a fetus;

subjecting the sequencing library to sequencing, to obtain a sequencing result of the fetus consisting of a plurality of sequencing data; and

determining the base information of the predetermined region based on the sequencing result of the fetus combining with genetic information of a related individual using a hidden Markov Model.

2. The method of claim 1, wherein the genomic DNA sample of the fetus is extracted from pregnant peripheral blood.

3. The method of claim 1, wherein the sequencing library is subjected to sequencing by at least one selected from Illumina-Solexa, ABI-Solid, Roche-454 and a single molecule sequencing apparatus.

4. The method of claim 1, further comprising a step of aligning the sequencing result of the fetus to a reference sequence, to determine sequencing result deriving from the predetermined region.

5. The method of claim 4, wherein the reference sequence is a human reference genome.

6. The method of claim 1, wherein the related individual is parents of the fetus.

7. The method of claim 1, wherein the step of determining the base information of the predetermined region using the hidden Markov Model is performed based on Viterbi algorithm.

8. The method of claim 7, wherein in the Viterbi algorithm, 0.25 is used as a

probability distribution of an initial status, re/N is used as a recombination probability, with re being 25~30, preferably re being 25, and N being a length of the predetermined region,

$$a_{jk} = \Pr(q_i = k | q_{i-1} = j) = \begin{cases} (1-p_r)^2 & x_i = x_{i-1}, y_i = y_{i-1} \\ (1-p_r) \cdot p_r & x_i = x_{i-1}, y_i \neq y_{i-1} \text{ or } x_i \neq x_{i-1}, y_i = y_{i-1} \\ p_r^2 & x_i \neq x_{i-1}, y_i \neq y_{i-1} \end{cases}$$

is

used as a recombination transition matrix with p_r being re/N .

9. The method of claim 4, wherein the step of aligning the sequencing result of the fetal genome to the reference sequence to determine sequencing result deriving from the predetermined region further comprises:

determining a base having the highest probability based on a formula of

$$P_{i,base} = \sum_{k \in \{0,1\}} \frac{1}{2} (1-\varepsilon) \Delta(base, m_k) + \frac{1}{2} \varepsilon \cdot \Delta(base, m_i) + \frac{1}{2} \varepsilon \cdot \Delta(base, f_{y_i})$$

$$\Delta(x, y) = \begin{cases} 1-e & x=y \\ e/3 & x \neq y \end{cases}$$

wherein

10. The method of claim 1, wherein the predetermined region is a site previously determined as having a genetic polymorphism.

11. The method of claim 10, wherein the genetic polymorphism is at least one selected from single nucleotide polymorphism and STR.

12. A system for determining base information of a predetermined region in a fetal genome, comprising:

a library constructing apparatus, adapted for constructing sequencing library based on a genomic DNA sample of a fetus;

a sequencing apparatus, connected to the library constructing apparatus, and adapted for subjecting the sequencing library to sequencing, to obtain a sequencing result of the fetus consisting of a plurality of sequencing data; and

an analyzing apparatus, connected to the sequencing apparatus, and adapted for determining the base information of the predetermined region based on the sequencing

result of the fetus combining with genetic information of a related individual using a hidden Markov Model.

13. The system of claim 12, further comprising a DNA sample extracting apparatus, adapted for extracting the genomic DNA sample of the fetus from pregnant peripheral blood.

14. The system of claim 12, wherein the sequencing apparatus is at least one selected from Illumina-Solexa, ABI-Solid, Roche-454 and a single molecule sequencing apparatus.

15. The system of claim 12, further comprising an aligning apparatus, connected to the sequencing apparatus, and adapted for aligning the sequencing result of the fetus to a reference sequence, to determine sequencing result deriving from the predetermined region.

16. The system of claim 12, wherein the analyzing apparatus is adapted for determining the base information of the predetermined region using a hidden Markov Model based on Viterbi algorithm.

17. The system of claim 16, wherein in the Viterbi algorithm, 0.25 is used as a probability distribution of an initial status, re/N is used as a recombination probability, with re being 25~30, preferably re being 25, and N being a length of the predetermined region,

$$a_{jk} = \Pr(q_i = k | q_{i-1} = j) = \begin{cases} (1-p_r)^2 & x_i = x_{i-1}, y_i = y_{i-1} \\ (1-p_r) \cdot p_r & x_i = x_{i-1}, y_i \neq y_{i-1} \text{ or } x_i \neq x_{i-1}, y_i = y_{i-1} \\ p_r^2 & x_i \neq x_{i-1}, y_i \neq y_{i-1} \end{cases}$$

is

used as a recombination transition matrix with p_r being re/N .

18. The system of claim 15, wherein the aligning apparatus is adapted for determining a base having the highest probability based on a formula of

$$P_{i,base} = \sum_{k \in \{0,1\}} \frac{1}{2} (1-\varepsilon) \Delta(base, m_k) + \frac{1}{2} \varepsilon \cdot \Delta(base, m_{x_i}) + \frac{1}{2} \varepsilon \cdot \Delta(base, f_{y_i})$$

$$\Delta(x, y) = \begin{cases} 1-e & x=y \\ e/3 & x \neq y \end{cases}$$

wherein

19. A computer readable medium comprising a plurality of instructions, adapted for determining base information of a predetermined region based on a sequencing result of a fetus combining with genetic information of a related individual using a hidden Markov Model.

20. The computer readable medium of claim 19, wherein the plurality of instructions are adapted for determining the base information of the predetermined region using the hidden Markov model based on Viterbi algorithm

21. The computer readable medium of claim 20, wherein in the Viterbi algorithm, 0.25 is used as a probability distribution of an initial status, re/N is used as a recombination probability, with re being 25~30, preferably re being 25, and N being a length of the predetermined region,

$$a_{jk} = \Pr(q_i = k | q_{i-1} = j) = \begin{cases} (1-p_r)^2 & x_i = x_{i-1}, y_i = y_{i-1} \\ (1-p_r) \cdot p_r & x_i = x_{i-1}, y_i \neq y_{i-1} \text{ or } x_i \neq x_{i-1}, y_i = y_{i-1} \\ p_r^2 & x_i \neq x_{i-1}, y_i \neq y_{i-1} \end{cases}$$

is

used as a recombination transition matrix with p_r being re/N .

22. The computer readable medium of claim 19, wherein the plurality of instructions are adapted for aligning the sequencing result of the fetus to a reference sequence, to determine sequencing result deriving from the predetermined region.

23. The computer readable medium of claim 22, wherein the plurality of instructions are further adapted for determining a base having the highest probability based on based on a formula of

$$P_{i,base} = \sum_{k \in \{0,1\}} \frac{1}{2} (1 - \varepsilon) \Delta(base, m_k) + \frac{1}{2} \varepsilon \cdot \Delta(base, m_{x_i}) + \frac{1}{2} \varepsilon \cdot \Delta(base, f_{y_i})$$

$$\Delta(x, y) = \begin{cases} 1 - e & x = y \\ e/3 & x \neq y \end{cases}$$

wherein

24. The computer readable medium of claim 19, wherein the predetermined region is a site previously determined as having a genetic polymorphism.

25. The computer readable medium of claim 24, wherein the genetic polymorphism is at least one selected from single nucleotide polymorphism and STR.

WHAT IS CLAIMED IS:

1. A method of determining base information of a predetermined region in a fetal genome, comprising following steps:

constructing a sequencing library based on a genomic DNA sample of a fetus;
subjecting the sequencing library to sequencing, to obtain a sequencing result of the fetus consisting of a plurality of sequencing data; and

determining the base information of the predetermined region based on the sequencing result of the fetus combining with genetic information of a related individual using a hidden Markov Model,

wherein the step of determining the base information of the predetermined region using the hidden Markov Model is performed based on Viterbi algorithm.

2. The method of claim 1, wherein the genomic DNA sample of the fetus is extracted from pregnant peripheral blood.

3. The method of claim 1, wherein the sequencing library is subjected to sequencing by at least one selected from Illumina-Solexa, ABI-Solid, Roche-454 and a single molecule sequencing apparatus.

4. The method of claim 1, further comprising a step of aligning the sequencing result of the fetus to a reference sequence, to determine sequencing result deriving from the predetermined region.

5. The method of claim 4, wherein the reference sequence is a human reference genome.

6. The method of claim 1, wherein the related individual is parents of the fetus.

7. The method of claim 1, wherein in the Viterbi algorithm, 0.25 is used as a probability distribution of an initial status, re/N is used as a recombination probability,

with re being 25~30, preferably re being 25, and N being a length of the predetermined region,

$$a_{jk} = \Pr(q_i = k | q_{i-1} = j) = \begin{cases} (1-p_r)^2 & x_i = x_{i-1}, y_i = y_{i-1} \\ (1-p_r) \cdot p_r & x_i = x_{i-1}, y_i \neq y_{i-1} \text{ or } x_i \neq x_{i-1}, y_i = y_{i-1} \\ p_r^2 & x_i \neq x_{i-1}, y_i \neq y_{i-1} \end{cases} \quad \text{is}$$

used as a recombination transition matrix with p_r being re/N .

8. The method of claim 4, wherein the step of aligning the sequencing result of the fetal genome to the reference sequence to determine sequencing result deriving from the predetermined region further comprises:

determining a base having the highest probability based on a formula of

$$P_{i,base} = \sum_{k \in \{0,1\}} \frac{1}{2} (1-\varepsilon) \Delta(base, m_k) + \frac{1}{2} \varepsilon \cdot \Delta(base, m_{x_i}) + \frac{1}{2} \varepsilon \cdot \Delta(base, f_{y_i})$$

wherein

$$\Delta(x, y) = \begin{cases} 1-e & x = y \\ e/3 & x \neq y \end{cases}.$$

9. The method of claim 1, wherein the predetermined region is a site previously determined as having a genetic polymorphism.

10. The method of claim 9, wherein the genetic polymorphism is at least one selected from single nucleotide polymorphism and STR.

11. A system for determining base information of a predetermined region in a fetal genome, comprising:

a library constructing apparatus, adapted for constructing sequencing library based on a genomic DNA sample of a fetus;

a sequencing apparatus, connected to the library constructing apparatus, and adapted for subjecting the sequencing library to sequencing, to obtain a sequencing result of the fetus consisting of a plurality of sequencing data; and

an analyzing apparatus, connected to the sequencing apparatus, and adapted for

determining the base information of the predetermined region based on the sequencing result of the fetus combining with genetic information of a related individual using a hidden Markov Model,

wherein the analyzing apparatus is adapted for determining the base information of the predetermined region using a hidden Markov Model based on Viterbi algorithm.

12. The system of claim 11, further comprising a DNA sample extracting apparatus, adapted for extracting the genomic DNA sample of the fetus from pregnant peripheral blood.

13. The system of claim 11, wherein the sequencing apparatus is at least one selected from Illumina-Solexa, ABI-Solid, Roche-454 and a single molecule sequencing apparatus.

14. The system of claim 11, further comprising an aligning apparatus, connected to the sequencing apparatus, and adapted for aligning the sequencing result of the fetus to a reference sequence, to determine sequencing result deriving from the predetermined region.

15. The system of claim 11, wherein in the Viterbi algorithm, 0.25 is used as a probability distribution of an initial status, re/N is used as a recombination probability, with re being 25~30, preferably re being 25, and N being a length of the predetermined region,

$$a_{jk} = \Pr(q_i = k | q_{i-1} = j) = \begin{cases} (1-p_r)^2 & x_i = x_{i-1}, y_i = y_{i-1} \\ (1-p_r) \cdot p_r & x_i = x_{i-1}, y_i \neq y_{i-1} \text{ or } x_i \neq x_{i-1}, y_i = y_{i-1} \\ p_r^2 & x_i \neq x_{i-1}, y_i \neq y_{i-1} \end{cases} \quad \text{is}$$

used as a recombination transition matrix with p_r being re/N .

16. The system of claim 14, wherein the aligning apparatus is adapted for determining a base having the highest probability based on a formula of

$$P_{i,base} = \sum_{k \in \{0,1\}} \frac{1}{2} (1-\varepsilon) \Delta(base, m_k) + \frac{1}{2} \varepsilon \cdot \Delta(base, m_{x_i}) + \frac{1}{2} \varepsilon \cdot \Delta(base, f_{y_i})$$

$$\Delta(x, y) = \begin{cases} 1-e & x = y \\ e/3 & x \neq y \end{cases}$$

wherein

17. A computer readable medium comprising a plurality of instructions, adapted for determining base information of a predetermined region based on a sequencing result of a fetus combining with genetic information of a related individual using a hidden Markov Model,

wherein the plurality of instructions are adapted for determining the base information of the predetermined region using the hidden Markov model based on Viterbi algorithm.

18. The computer readable medium of claim 17, wherein in the Viterbi algorithm, 0.25 is used as a probability distribution of an initial status, re/N is used as a recombination probability, with re being 25~30, preferably re being 25, and N being a length of the predetermined region,

$$a_{jk} = \Pr(q_i = k | q_{i-1} = j) = \begin{cases} (1-p_r)^2 & x_i = x_{i-1}, y_i = y_{i-1} \\ (1-p_r) \cdot p_r & x_i = x_{i-1}, y_i \neq y_{i-1} \text{ or } x_i \neq x_{i-1}, y_i = y_{i-1} \\ p_r^2 & x_i \neq x_{i-1}, y_i \neq y_{i-1} \end{cases}$$

is

used as a recombination transition matrix with p_r being re/N .

19. The computer readable medium of claim 17, wherein the plurality of instructions are adapted for aligning the sequencing result of the fetus to a reference sequence, to determine sequencing result deriving from the predetermined region.

20. The computer readable medium of claim 19, wherein the plurality of instructions are further adapted for determining a base having the highest probability based on based on a formula of

$$P_{i,base} = \sum_{k \in \{0,1\}} \frac{1}{2} (1 - \varepsilon) \Delta(base, m_k) + \frac{1}{2} \varepsilon \cdot \Delta(base, m_{x_i}) + \frac{1}{2} \varepsilon \cdot \Delta(base, f_{y_i})$$

wherein

$$\Delta(x, y) = \begin{cases} 1 - e & x = y \\ e/3 & x \neq y \end{cases} .$$

21. The computer readable medium of claim 17, wherein the predetermined region is a site previously determined as having a genetic polymorphism.

22. The computer readable medium of claim 21, wherein the genetic polymorphism is at least one selected from single nucleotide polymorphism and STR.