(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0015713
(43) 공개일자 2023년01월31일

(51) 국제특허분류(Int. Cl.)
G16B 50/40 (2019.01) *G16B 20/20* (2019.01)
G16B 30/10 (2019.01) *G16B 50/20* (2019.01)
(52) CPC특허분류
G16B 50/40 (2019.02)
G16B 20/20 (2019.02)
(21) 출원번호 10-2021-0097179
(22) 출원일자 2021년07월23일
심사청구일자 2021년07월23일

(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
주식회사 클리노믹스
울산광역시 울주군 언양읍 유니스트길 50,110동 3층 307호
(72) 발명자
박종화
울산광역시 울주군 언양읍 반천강변길 51, 101동 1701호(현대아파트)
조윤성
경기도 용인시 기흥구 흥덕1로79번길 37, 503동 1702호(영덕동, 흥덕마을5단지 호반베르디움)
이황열
경기도 용인시 기흥구 흥덕1로79번길 37, 504동 1201호(영덕동, 흥덕마을5단지 호반베르디움)
(74) 대리인
특허법인메이저

전체 청구항 수 : 총 17 항

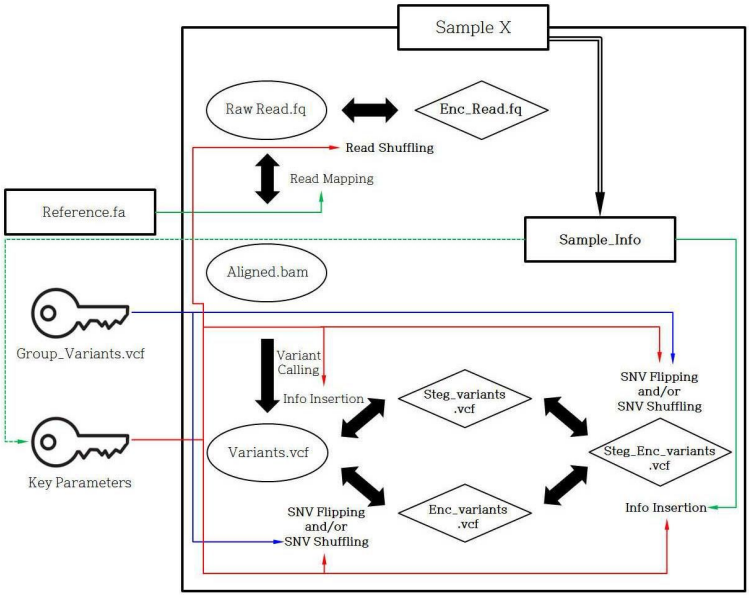
(54) 발명의 명칭 **유전체 정보의 암호화 및 복호화 방법**

(57) 요 약

본 발명은 유전체 정보의 암호화 및 복호화 방법에 관한 것으로, 해결하고자 하는 과제는 유전정보를 이진 형식으로 인코딩하거나, 암호화하기에 앞서, 암호화되지 않은 유전체 정보처럼 보이지만 실제 유전정보를 숨김으로써 유출된 유전체 정보의 진위여부를 판단할 수 없도록 함으로써, 개인의 변이정보나 원시서열정보를 암호화하여 개

(뒷면에 계속)

대 표 도



인의 유전체정보를 숨기고, 정해진 키 값에 의해서만 실제 정보를 복호화 가능하도록 하여 컴퓨터에 저장되거나 전송되는 과정에서 개인의 유전체정보가 유출되는 것을 방지하는데 있다.

일례로, (a) 게놈지도 매핑이 완료된 유전체 원천서열정보로부터 피검자의 유전형 및 변이정보를 확인하는 단계; (b) 상기 유전형 변이정보 중 키 샘플 군에서 단일염기변이가 존재하는 영역을 염색체 순서대로 정렬하는 단계; (c) 정렬된 단일염기변이영역 중 미리 저장된 변조/복조 규칙정보에 포함된 제1 키 파라미터 값에 해당하는 위치의 변이마다 키 샘플 군에서의 유전형 빈도정보를 기준으로 피검자의 유전형을 각각 변조하여 유전체 정보를 암호화하는 단계; 및 (d) 암호화된 유전체 정보를 상기 변조/복조 규칙정보와 상기 유전형 빈도정보를 기초로 복조하여 유전체 정보를 복호화하는 단계를 포함하는 유전체 정보의 암호화 및 복호화 방법을 개시한다.

(52) CPC특허분류
G16B 30/10 (2019.02)
G16B 50/20 (2019.02)

명 세 서

청구범위

청구항 1

(a) 게놈지도 매핑이 완료된 유전체 원천서열정보로부터 피검자의 유전형 및 변이정보를 확인하는 단계;

(b) 상기 유전형 변이정보 중 키 샘플군에서 단일염기변이가 존재하는 영역을 염색체 순서대로 정렬하는 단계;

(c) 정렬된 단일염기변이영역 중 미리 저장된 변조/복조 규칙정보에 포함된 제1 키 파라미터 값에 해당하는 위치의 변이마다 키 샘플군에서의 유전형 빈도정보를 기준으로 피검자의 유전형을 각각 변조하여 유전체 정보를 암호화하는 단계; 및

(d) 암호화된 유전체 정보를 상기 변조/복조 규칙정보와 상기 유전형 빈도정보를 기초로 복조하여 유전체 정보를 복호화하는 단계를 포함하는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 2

제1 항에 있어서,

(c-1) 상기 (c) 단계 이후에, 정렬된 단일염기변이영역 중 n번째 변이를 기준으로 제2 키 파라미터 값에 해당하는 bp 위치의 좌우변이영역의 유전형을 서로 바꾸어 유전체 정보를 암호화하는 단계; 및

(d-1) 상기 (d) 단계 이전에, 상기 (c-1) 단계를 통해 암호화된 유전체 정보의 단일염기변이 중 n번째 변이를 기준으로 상기 (c-1) 단계에서 선택된 좌우 변이영역에 대한 유전형을 서로 바꾸어 유전체 정보를 사전 복호화하는 단계를 더 포함하는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 3

제2 항에 있어서,

상기 변조/복조 규칙정보는,

암호화/복호화 종류, 순서, 상기 제1 키 파라미터 값, 제2 키 파라미터 값, 유전형 변조순서정보를 포함하는 파일 형태로 제공되는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 4

청구항 제1항 내지 제3항 중 어느 한 항의 방법을 실행하기 위한 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체.

청구항 5

(a) 유전체 원천서열정보의 단편서열 수와 단편서열 길이, 또는 키 샘플군의 단일염기변이영역을 기준으로 정렬된 피검자의 유전형 변이정보를 확인하는 단계;

(b) 상기 유전체 원천서열정보 및 상기 유전형 변이정보 중 적어도 하나에 포함된 염기서열을 미리 저장된 변조/복조 규칙정보에 포함되어 있는 제1 키 파라미터 값을 기본 단위길이로 나누어 다수의 블록 그룹을 생성하는 단계;

(c) 상기 블록 그룹의 순서를 변조/복조 규칙정보에 포함되어 있는 제2 키 파라미터 값에 따라 재배열하여 유전체 정보를 암호화하는 단계; 및

(d) 암호화된 블록 그룹의 순서를 상기 제2 키 파라미터 값에 따라 역순으로 재배열하여 유전체 정보를 복호화하는 단계를 포함하는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 6

제5 항에 있어서,

(c-1) 상기 (c) 단계 이후에, 블록 그룹에 따라 재배열된 염기서열에 대해 제3 키 파라미터 값(양 또는 음의 정수)에 따라 전체 염기서열을 특정 bp 만큼 좌측이나 우측에서 반대방향으로 옮겨 유전체 유전체 정보를 암호화하는 단계; 및

(d-1) 상기 (d) 단계 이전에, 상기 (c-1) 단계를 통해 암호화된 유전체 정보의 염기서열 중 n번째 변이를 기준으로 상기 (c-1) 단계에서 반대방향으로 옮겨진 염기서열을 유전체 정보를 사전 복호화하는 단계를 더 포함하는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 7

제6 항에 있어서,

상기 변조/복조 규칙정보는,

암호화/복호화 종류, 순서, 상기 제1 키 파라미터 값, 상기 제2 키 파라미터 값, 상기 제3 키 파라미터 값을 포함하는 파일 형태로 제공되는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 8

청구항 제5항 내지 제7항 중 어느 한 항의 방법을 실행하기 위한 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체.

청구항 9

(a) 문자열 및 수치 형식으로 표현되는 피검자의 개인정보 및 임상정보를 미리 저장된 인코딩/디코딩 변환테이블을 기초로 4진수 형태로 인코딩하여 염기서열형태로 변환하는 단계;

(b) 피검자의 개인정보 및 임상정보에 포함된 각 항목에 대하여 항목번호와 미리 설정된 제1 키 파라미터 값을 곱한 후 염색체 수로 나눈 나머지 제1 값과, 항목번호와 상기 제1 키 파라미터 값의 곱을 미리 설정된 제2 키 파라미터 값만큼 제곱 연산한 제2 값을 산출하여 피검자의 개인정보 및 임상정보를 은닉할 염색체의 위치를 설정하는 단계;

(c) 상기 (a) 단계를 통해 변환된 가상염기서열을 상기 (b) 단계를 통해 변환된 염색체 위치에 변이정보로서 삽입하여 피검자 정보를 은닉하는 단계; 및

(d) 상기 (a) 단계를 통해 암호화된 가상염기서열을 상기 제1 키 파라미터 값, 상기 제2 키 파라미터 값, 상기 제1 값, 상기 제2 값 및 상기 인코딩/디코딩 변환테이블을 기초로 역으로 연산을 진행하여 피검자의 개인정보 및 임상정보를 복호화하는 단계를 포함하는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 10

제9 항에 있어서,

상기 변조/복조 규칙정보는,

암호화/복호화 종류, 순서, 상기 인코딩/디코딩 변환테이블, 상기 제1 키 파라미터 값, 상기 제2 키 파라미터

값을 포함하는 파일 형태로 제공되는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 11

청구항 제9항 및 제10항 중 어느 한 항의 방법을 실행하기 위한 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체.

청구항 12

(a) 게놈지도 매핑이 완료된 유전체 원천서열정보로부터 피검자의 유전형 및 변이정보를 확인하는 단계; (b) 상기 유전형 변이정보 중 키 샘플군에서 단일염기변이가 존재하는 영역을 염색체 순서대로 정렬하는 단계; (c) 정렬된 단일염기변이영역 중 미리 저장된 변조/복조 규칙정보에 포함된 제1 키 파라미터 값에 해당하는 위치의 변이마다 키 샘플군에서의 유전형 빈도정보를 기준으로 피검자의 유전형을 각각 변조하여 유전체 정보를 암호화하는 단계를 포함하는 단일염기변이영역에서의 유전형 변조를 통한 암호화 단계;

(d) 문자열 및 수치 형식으로 표현되는 피검자의 개인정보 및 임상정보를 미리 저장된 인코딩/디코딩 변환테이블을 기초로 4진수 형태로 인코딩하여 염기서열형태로 변환하는 단계;

(e) 피검자의 개인정보 및 임상정보에 포함된 각 항목에 대하여 항목번호와 미리 설정된 제2 키 파라미터 값을 곱한 후 염색체 수로 나눈 나머지 제1 값과, 항목번호와 상기 제2 키 파라미터 값의 곱을 미리 설정된 제3 키 파라미터 값만큼 제곱 연산한 제2 값을 산출하여 피검자의 개인정보 및 임상정보를 은닉할 염색체의 위치를 설정하는 단계;

(f) 상기 (d) 단계를 통해 변환된 가상염기서열을 상기 (e) 단계를 통해 변환된 염색체 위치에 변이정보로서 삽입하여 피검자 정보를 은닉하는 단계를 포함하는 염기서열화 및 삽입변이 은닉을 통한 암호화 단계;

(g) 상기 염기서열화 및 삽입변이 은닉을 통한 암호화 단계를 통해 암호화된 가상염기서열을 상기 제2 키 파라미터 값, 상기 제3 키 파라미터 값, 상기 제1 값, 상기 제2 값 및 상기 인코딩/디코딩 변환테이블을 기초로 역으로 연산을 진행하여 피검자의 개인정보 및 임상정보를 복호화하는 염기서열화 은닉 삽입변이에 대한 복호화 단계; 및

(h) 상기 단일염기변이영역에서의 유전형 변조를 통한 암호화 단계를 통해 암호화된 유전체 정보를 상기 변조/복조 규칙정보와 상기 유전형 빈도정보를 기초로 복조하여 유전체 정보를 복호화하는 단일염기변이영역 변조 유전형에 대한 복호화 단계를 포함하는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 13

청구항 제12항의 방법을 실행하기 위한 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체.

청구항 14

(a) 키 샘플군의 단일염기변이영역을 기준으로 정렬된 피검자의 유전형 변이정보를 확인하는 단계;

(b) 상기 단일염기변이영역의 염기서열을 미리 저장된 변조/복조 규칙정보에 포함되어 있는 제1 키 파라미터 값을 기본 단위길이로 나누어 다수의 블록 그룹을 생성하는 단계;

(c) 상기 블록 그룹의 순서를 변조/복조 규칙정보에 포함되어 있는 제2 키 파라미터 값에 따라 재배열하여 유전체 정보를 암호화하는 단계를 포함하는 염기서열의 블록화 및 재배열을 통한 암호화 단계;

(d) 문자열 및 수치 형식으로 표현되는 피검자의 개인정보 및 임상정보를 미리 저장된 인코딩/디코딩 변환테이블을 기초로 4진수 형태로 인코딩하여 염기서열형태로 변환하는 단계;

(e) 피검자의 개인정보 및 임상정보에 포함된 각 항목에 대하여 항목번호와 미리 설정된 제3 키 파라미터 값을 곱한 후 염색체 수로 나눈 나머지 제1 값과, 항목번호와 상기 제3 키 파라미터 값의 곱을 미리 설정된 제4 키

파라미터 값만큼 제곱 연산한 제2 값을 산출하여 피검자의 개인정보 및 임상정보를 은닉할 염색체의 위치를 설정하는 단계;

(f) 상기 (d) 단계를 통해 변환된 가상염기서열을 상기 (e) 단계를 통해 변환된 염색체 위치에 변이정보로서 삽입하여 피검자 정보를 은닉하는 단계를 포함하는 염기서열화 및 삽입변이 은닉을 통한 암호화 단계;

(g) 상기 염기서열화 및 삽입변이 은닉을 통한 암호화 단계를 통해 암호화된 가상염기서열을 상기 제3 키 파라미터 값, 상기 제4 키 파라미터 값, 상기 제1 값, 상기 제2 값 및 상기 인코딩/디코딩 변환테이블을 기초로 역으로 연산을 진행하여 피검자의 개인정보 및 임상정보를 복호화하는 염기서열화 은닉 삽입변이에 대한 복호화 단계; 및

(h) 상기 염기서열의 블록화 및 재배열을 통한 암호화 단계를 통해 암호화된 블록 그룹의 순서를 상기 제2 키 파라미터 값에 따라 역순으로 재배열하여 복호화하는 블록화 및 재배열 염기서열에 통한 복호화 단계를 포함하는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 15

청구항 제14항의 방법을 실행하기 위한 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체.

청구항 16

(a1) 게놈지도 매핑이 완료된 유전체 원천서열정보로부터 피검자의 유전형 및 변이정보를 확인하는 단계; (a2) 상기 유전형 변이정보 중 키 샘플군에서 단일염기변이가 존재하는 영역을 염색체 순서대로 정렬하는 단계; (a3) 정렬된 단일염기변이영역 중 미리 저장된 변조/복조 규칙정보에 포함된 제1 키 파라미터 값에 해당하는 위치의 변이마다 키 샘플군에서의 유전형 빈도정보를 기준으로 피검자의 유전형을 각각 변조하여 유전체 정보를 암호화하는 단계를 포함하는 단일염기변이영역에서의 유전형 변조를 통한 암호화 단계;

(b1) 상기 유전체 정보를 암호화하는 단계를 통해 변조된 단일염기변이영역의 염기서열을 미리 저장된 변조/복조 규칙정보에 포함되어 있는 제2 키 파라미터 값을 기본 단위길이로 나누어 다수의 블록 그룹을 생성하는 단계; (b2) 상기 블록 그룹의 순서를 변조/복조 규칙정보에 포함되어 있는 제3 키 파라미터 값에 따라 재배열하여 유전체 정보를 암호화하는 단계를 포함하는 염기서열의 블록화 및 재배열을 통한 암호화 단계;

(c1) 문자열 및 수치 형식으로 표현되는 피검자의 개인정보 및 임상정보를 미리 저장된 인코딩/디코딩 변환테이블을 기초로 4진수 형태로 인코딩하여 염기서열형태로 변환하는 단계; (c2) 피검자의 개인정보 및 임상정보에 포함된 각 항목에 대하여 항목번호와 미리 설정된 제4 키 파라미터 값을 곱한 후 염색체 수로 나눈 나머지 제1 값과, 항목번호와 상기 제4 키 파라미터 값의 곱을 미리 설정된 제5 키 파라미터 값만큼 제곱 연산한 제2 값을 산출하여 피검자의 개인정보 및 임상정보를 은닉할 염색체의 위치를 설정하는 단계; (c3) 상기 (c1) 단계를 통해 변환된 가상염기서열을 상기 (c2) 단계를 통해 변환된 염색체 위치에 변이정보로서 삽입하여 피검자 정보를 은닉하는 단계를 포함하는 염기서열화 및 삽입변이 은닉을 통한 암호화 단계;

(d) 상기 염기서열화 및 삽입변이 은닉을 통한 암호화 단계를 통해 암호화된 가상염기서열을 상기 제4 키 파라미터 값, 상기 제4 키 파라미터 값, 상기 제1 값, 상기 제2 값 및 상기 인코딩/디코딩 변환테이블을 기초로 역으로 연산을 진행하여 피검자의 개인정보 및 임상정보를 복호화하는 염기서열화 은닉 삽입변이에 대한 복호화 단계;

(e) 상기 염기서열의 블록화 및 재배열을 통한 암호화 단계를 통해 암호화된 블록 그룹의 순서를 상기 제3 키 파라미터 값에 따라 역순으로 재배열하여 복호화하는 블록화 및 재배열 염기서열에 대한 복호화 단계; 및

(f) 상기 단일염기변이영역에서의 유전형 변조를 통한 암호화 단계를 통해 암호화된 유전체 정보를 상기 변조/복조 규칙정보와 상기 유전형 빈도정보를 기초로 복조하여 유전체 정보를 복호화하는 단일염기변이영역 변조 유전형에 대한 복호화 단계를 포함하는 것을 특징으로 하는 유전체 정보의 암호화 및 복호화 방법.

청구항 17

청구항 제16항의 방법을 실행하기 위한 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체.

발명의 설명

기 술 분 야

[0001] 본 발명의 실시예는 유전체 정보의 암호화 및 복호화 방법에 관한 것으로, 좀 더 구체적으로는 단일염기변이영역에서의 유전형 변조를 통한 암호화 복호화 방법, 염기서열의 블록화 및 재배열을 통한 암호화 복호화 방법, 피검자 정보의 염기서열화 및 삽입변이 유전체 은닉을 통한 암호화 및 복호화 방법에 관한 것이다.

배 경 기 술

[0002] 유전체는 생명의 진화와 발달 과정에서 축적된 개별 생명체들의 설계도로, 인간은 23개의 염색체로 구성된 30억 개의 염기서열을 가진다.

[0003] 이 중 1,000만여 개의 위치에 대한 염기서열변화로 인해 개인의 특성이 결정되며, 개인마다 고유한 염기서열의 특성은 실종자, 범죄자의 특정이나 질병위험도의 예측, 환자맞춤치료제 등에 활용되기도 한다. 검사기술의 발달로 유전체 검사비용은 초기보다 많이 낮아졌지만, 아직 기술의 도입기로 전체 인구 중 일부에 대해서만 특정 목적으로 사용되고 있어, 개개인의 유전체 정보가 축적될수록 앞으로의 활용 범위는 현재 예측 가능한 수준보다 넓어질 것이다.

[0004] 개인의 유전체 정보는 의도치 않게 노출될 경우 악용될 가능성이 있으며, 지문이나 홍채와 같이 출생 시 결정되어 일생 동안 거의 변하지 않고, 인위적으로 바꾸기도 어렵기 때문에, 이러한 피해는 돌이킬 수 없다. 따라서 유전체 정보의 보안은 필수적이다.

[0005] 개인의 유전체정보를 암호화하고 유출되는 것으로부터 보호하기 위해, 변이위치와 형질을 사전에 정의된 이진 형식으로 변환하는 방법, 유전체정보에 매칭되는 피검자 정보를 암호화하는 방법, 서버에서 공개/개인 키를 확인하여 허가된 영역의 변이정보만을 제공하는 방법, 실험적으로 PCR 기반의 프라이머 또는 칩 기반의 프로브 정보를 암호 키로 하여 삽입된 워터마크를 통해 유출 시 출처를 확인할 수 있는 방법 등이 개발되어 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 등록특허공보 제10-2030800호(등록일자: 2019년10월02일)

(특허문헌 0002) 공개특허공보 제10-2009-0068686호(공개일자: 2009년06월29일)

(특허문헌 0003) 공개특허공보 제10-2019-0118853호(공개일자: 2019년10월21일)

(특허문헌 0004) 공개특허공보 제10-2011-0033880호(공개일자: 2011년04월01일)

발명의 내용

해결하려는 과제

[0007] 본 발명의 실시예는, 유전정보를 이진 형식으로 인코딩하거나, 암호화하기에 앞서, 암호화되지 않은 유전체 정보처럼 보이지만 실제 유전정보를 숨김으로써 유출된 유전체 정보의 진위여부를 판단할 수 없도록 함으로써, 개인의 변이정보나 원시서열정보를 암호화하여 개인의 유전체정보를 숨기고, 정해진 키 값에 의해서만 실제 정보를 복호화 가능하도록 하여 컴퓨터에 저장되거나 전송되는 과정에서 개인의 유전체정보가 유출되는 것을 방지할 수 있는 유전체 정보의 암호화 및 복호화 방법을 제공한다.

과제의 해결 수단

[0008] 본 발명의 일 실시예에 따른 유전체 정보의 암호화 및 복호화 방법은, (a) 게놈지도 매핑이 완료된 유전체 원천 서열정보로부터 피검자의 유전형 및 변이정보를 확인하는 단계; (b) 상기 유전형 변이정보 중 키 샘플군에서 단일염기변이가 존재하는 영역을 염색체 순서대로 정렬하는 단계; (c) 정렬된 단일염기변이영역 중 미리 저장된

변조/복조 규칙정보에 포함된 제1 키 파라미터 값에 해당하는 위치의 변이마다 키 샘플군에서의 유전형 빈도정보를 기준으로 피검자의 유전형을 각각 변조하여 유전체 정보를 암호화하는 단계; 및 (d) 암호화된 유전체 정보를 상기 변조/복조 규칙정보와 상기 유전형 빈도정보를 기초로 복조하여 유전체 정보를 복호화하는 단계를 포함한다.

[0009] 또한, (c-1) 상기 (c) 단계 이후에, 정렬된 단일염기변이영역 중 n번째 변이를 기준으로 제2 키 파라미터 값에 해당하는 bp 위치의 좌우변이영역의 유전형을 서로 바꾸어 유전체 정보를 암호화하는 단계; 및 (d-1) 상기 (d) 단계 이전에, 상기 (c-1) 단계를 통해 암호화된 유전체 정보의 단일염기변이 중 n번째 변이를 기준으로 상기 (c-1) 단계에서 선택된 좌우 변이영역에 대한 유전형을 서로 바꾸어 유전체 정보를 사전 복호화하는 단계를 더 포함할 수 있다.

[0010] 또한, 상기 변조/복조 규칙정보는, 암호화/복호화 종류, 순서, 상기 제1 키 파라미터 값, 제2 키 파라미터 값, 유전형 변조순서정보를 포함하는 파일 형태로 제공될 수 있다.

[0011] 본 발명의 다른 실시예에 따른 유전체 정보의 암호화 및 복호화 방법은, (a) 유전체 원천서열정보의 단편서열 수와 단편서열 길이, 또는 키 샘플군의 단일염기변이영역을 기준으로 정렬된 피검자의 유전형 변이정보를 확인하는 단계; (b) 상기 유전체 원천서열정보 및 상기 유전형 변이정보 중 적어도 하나에 포함된 염기서열을 미리 저장된 변조/복조 규칙정보에 포함되어 있는 제1 키 파라미터 값을 기본 단위길이로 나누어 다수의 블록 그룹을 생성하는 단계; (c) 상기 블록 그룹의 순서를 변조/복조 규칙정보에 포함되어 있는 제2 키 파라미터 값에 따라 재배열하여 유전체 정보를 암호화하는 단계; 및 (d) 암호화된 블록 그룹의 순서를 상기 제2 키 파라미터 값에 따라 역순으로 재배열하여 유전체 정보를 복호화하는 단계를 포함할 수 있다.

[0012] 또한, (c-1) 상기 (c) 단계 이후에, 블록 그룹에 따라 재배열된 염기서열에 대해 제3 키 파라미터 값(양 또는 음의 정수)에 따라 전체 염기서열을 특정 bp 만큼 좌측이나 우측에서 반대방향으로 옮겨 유전체 유전체 정보를 암호화하는 단계; 및 (d-1) 상기 (d) 단계 이전에, 상기 (c-1) 단계를 통해 암호화된 유전체 정보의 염기서열 중 n번째 변이를 기준으로 상기 (c-1) 단계에서 반대방향으로 옮겨진 염기서열을 유전체 정보를 사전 복호화하는 단계를 더 포함할 수 있다.

[0013] 또한, 상기 변조/복조 규칙정보는, 암호화/복호화 종류, 순서, 상기 제1 키 파라미터 값, 상기 제2 키 파라미터 값, 상기 제3 키 파라미터 값을 포함할 수 있다.

[0014] 본 발명의 또 다른 실시예에 따른 유전체 정보의 암호화 및 복호화 방법은, (a) 문자열 및 수치 형식으로 표현되는 피검자의 개인정보 및 임상정보를 미리 저장된 인코딩/디코딩 변환테이블을 기초로 4진수 형태로 인코딩하여 염기서열형태로 변환하는 단계; (b) 피검자의 개인정보 및 임상정보에 포함된 각 항목에 대하여 항목번호와 미리 설정된 제1 키 파라미터 값을 곱한 후 염색체 수로 나눈 나머지 제1 값과, 항목번호와 상기 제1 키 파라미터 값의 곱을 미리 설정된 제2 키 파라미터 값만큼 제곱 연산한 제2 값을 산출하여 피검자의 개인정보 및 임상정보를 은닉할 염색체의 위치를 설정하는 단계; (c) 상기 (a) 단계를 통해 변환된 가상염기서열을 상기 (b) 단계를 통해 변환된 염색체 위치에 변이정보로서 삽입하여 피검자 정보를 은닉하는 단계; 및 (d) 상기 (a) 단계를 통해 암호화된 가상염기서열을 상기 제1 키 파라미터 값, 상기 제2 키 파라미터 값, 상기 제1 값, 상기 제2 값 및 상기 인코딩/디코딩 변환테이블을 기초로 역으로 연산을 진행하여 피검자의 개인정보 및 임상정보를 복호화하는 단계를 포함할 수 있다.

[0015] 또한, 상기 변조/복조 규칙정보는, 암호화/복호화 종류, 순서, 상기 인코딩/디코딩 변환테이블, 상기 제1 키 파라미터 값, 상기 제2 키 파라미터 값을 포함하는 파일 형태로 제공될 수 있다.

[0016] 본 발명의 또 다른 실시예에 따른 유전체 정보의 암호화 및 복호화 방법은, (a) 게놈지도 매핑이 완료된 유전체 원천서열정보로부터 피검자의 유전형 및 변이정보를 확인하는 단계; (b) 상기 유전형 변이정보 중 키 샘플군에서 단일염기변이가 존재하는 영역을 염색체 순서대로 정렬하는 단계; (c) 정렬된 단일염기변이영역 중 미리 저장된 변조/복조 규칙정보에 포함된 제1 키 파라미터 값에 해당하는 위치의 변이마다 키 샘플군에서의 유전형 빈도정보를 기준으로 피검자의 유전형을 각각 변조하여 유전체 정보를 암호화하는 단계를 포함하는 단일염기변이영역에서의 유전형 변조를 통한 암호화 단계; (d) 문자열 및 수치 형식으로 표현되는 피검자의 개인정보 및 임상정보를 미리 저장된 인코딩/디코딩 변환테이블을 기초로 4진수 형태로 인코딩하여 염기서열형태로 변환하는 단계; (e) 피검자의 개인정보 및 임상정보에 포함된 각 항목에 대하여 항목번호와 미리 설정된 제2 키 파라미터 값을 곱한 후 염색체 수로 나눈 나머지 제1 값과, 항목번호와 상기 제2 키 파라미터 값의 곱을 미리 설정된 제3 키 파라미터 값만큼 제곱 연산한 제2 값을 산출하여 피검자의 개인정보 및 임상정보를 은닉할 염색체의 위치를

설정하는 단계; (f) 상기 (d) 단계를 통해 변환된 가상염기서열을 상기 (e) 단계를 통해 변환된 염색체 위치에 변이정보로서 삽입하여 피검자 정보를 은닉하는 단계를 포함하는 염기서열화 및 삽입변이 은닉을 통한 암호화 단계; (g) 상기 염기서열화 및 삽입변이 은닉을 통한 암호화 단계를 통해 암호화된 가상염기서열을 상기 제2 키 파라미터 값, 상기 제3 키 파라미터 값, 상기 제1 값, 상기 제2 값 및 상기 인코딩/디코딩 변환테이블을 기초로 역으로 연산을 진행하여 피검자의 개인정보 및 임상정보를 복호화하는 염기서열화 은닉 삽입변이에 대한 복호화 단계; 및 (h) 상기 단일염기변이영역에서의 유전형 변조를 통한 암호화 단계를 통해 암호화된 유전체 정보를 상기 변조/복조 규칙정보와 상기 유전형 빈도정보를 기초로 복조하여 유전체 정보를 복호화하는 단일염기변이영역 변조 유전형에 대한 복호화 단계를 포함한다.

[0017] 본 발명의 또 다른 실시예에 따른 유전체 정보의 암호화 및 복호화 방법은, (a) 키 샘플군의 단일염기변이영역을 기준으로 정렬된 피검자의 유전형 변이정보를 확인하는 단계; (b) 상기 단일염기변이영역의 염기서열을 미리 저장된 변조/복조 규칙정보에 포함되어 있는 제1 키 파라미터 값을 기본 단위길이로 나누어 다수의 블록 그룹을 생성하는 단계; (c) 상기 블록 그룹의 순서를 변조/복조 규칙정보에 포함되어 있는 제2 키 파라미터 값에 따라 재배열하여 유전체 정보를 암호화하는 단계를 포함하는 염기서열의 블록화 및 재배열을 통한 암호화 단계; (d) 문자열 및 수치 형식으로 표현되는 피검자의 개인정보 및 임상정보를 미리 저장된 인코딩/디코딩 변환테이블을 기초로 4진수 형태로 인코딩하여 염기서열형태로 변환하는 단계; (e) 피검자의 개인정보 및 임상정보에 포함된 각 항목에 대하여 항목번호와 미리 설정된 제3 키 파라미터 값을 곱한 후 염색체 수로 나눈 나머지 제1 값과, 항목번호와 상기 제3 키 파라미터 값의 곱을 미리 설정된 제4 키 파라미터 값만큼 제곱 연산한 제2 값을 산출하여 피검자의 개인정보 및 임상정보를 은닉할 염색체의 위치를 설정하는 단계; (f) 상기 (d) 단계를 통해 변환된 가상염기서열을 상기 (e) 단계를 통해 변환된 염색체 위치에 변이정보로서 삽입하여 피검자 정보를 은닉하는 단계를 포함하는 염기서열화 및 삽입변이 은닉을 통한 암호화 단계; (g) 상기 염기서열화 및 삽입변이 은닉을 통한 암호화 단계를 통해 암호화된 가상염기서열을 상기 제3 키 파라미터 값, 상기 제4 키 파라미터 값, 상기 제1 값, 상기 제2 값 및 상기 인코딩/디코딩 변환테이블을 기초로 역으로 연산을 진행하여 피검자의 개인정보 및 임상정보를 복호화하는 염기서열화 은닉 삽입변이에 대한 복호화 단계; 및 (h) 상기 염기서열의 블록화 및 재배열을 통한 암호화 단계를 통해 암호화된 블록 그룹의 순서를 상기 제2 키 파라미터 값에 따라 역순으로 재배열하여 복호화하는 블록화 및 재배열 염기서열에 통한 복호화 단계를 포함한다.

[0018] 본 발명의 또 다른 실시예에 따른 유전체 정보의 암호화 및 복호화 방법은, (a1) 게놈지도 매핑이 완료된 유전체 원천서열정보로부터 피검자의 유전형 및 변이정보를 확인하는 단계; (a2) 상기 유전형 변이정보 중 키 샘플군에서 단일염기변이가 존재하는 영역을 염색체 순서대로 정렬하는 단계; (a3) 정렬된 단일염기변이영역 중 미리 저장된 변조/복조 규칙정보에 포함된 제1 키 파라미터 값에 해당하는 위치의 변이마다 키 샘플군에서의 유전형 빈도정보를 기준으로 피검자의 유전형을 각각 변조하여 유전체 정보를 암호화하는 단계를 포함하는 단일염기변이영역에서의 유전형 변조를 통한 암호화 단계; (b1) 상기 유전체 정보를 암호화하는 단계를 통해 변조된 단일염기변이영역의 염기서열을 미리 저장된 변조/복조 규칙정보에 포함되어 있는 제2 키 파라미터 값을 기본 단위길이로 나누어 다수의 블록 그룹을 생성하는 단계; (b2) 상기 블록 그룹의 순서를 변조/복조 규칙정보에 포함되어 있는 제3 키 파라미터 값에 따라 재배열하여 유전체 정보를 암호화하는 단계를 포함하는 염기서열의 블록화 및 재배열을 통한 암호화 단계; (c1) 문자열 및 수치 형식으로 표현되는 피검자의 개인정보 및 임상정보를 미리 저장된 인코딩/디코딩 변환테이블을 기초로 4진수 형태로 인코딩하여 염기서열형태로 변환하는 단계; (c2) 피검자의 개인정보 및 임상정보에 포함된 각 항목에 대하여 항목번호와 미리 설정된 제4 키 파라미터 값을 곱한 후 염색체 수로 나눈 나머지 제1 값과, 항목번호와 상기 제4 키 파라미터 값의 곱을 미리 설정된 제5 키 파라미터 값만큼 제곱 연산한 제2 값을 산출하여 피검자의 개인정보 및 임상정보를 은닉할 염색체의 위치를 설정하는 단계; (c3) 상기 (c1) 단계를 통해 변환된 가상염기서열을 상기 (c2) 단계를 통해 변환된 염색체 위치에 변이정보로서 삽입하여 피검자 정보를 은닉하는 단계를 포함하는 염기서열화 및 삽입변이 은닉을 통한 암호화 단계; (d) 상기 염기서열화 및 삽입변이 은닉을 통한 암호화 단계를 통해 암호화된 가상염기서열을 상기 제4 키 파라미터 값, 상기 제4 키 파라미터 값, 상기 제1 값, 상기 제2 값 및 상기 인코딩/디코딩 변환테이블을 기초로 역으로 연산을 진행하여 피검자의 개인정보 및 임상정보를 복호화하는 염기서열화 은닉 삽입변이에 대한 복호화 단계; (e) 상기 염기서열의 블록화 및 재배열을 통한 암호화 단계를 통해 암호화된 블록 그룹의 순서를 상기 제3 키 파라미터 값에 따라 역순으로 재배열하여 복호화하는 블록화 및 재배열 염기서열에 대한 복호화 단계; 및 (f) 상기 단일염기변이영역에서의 유전형 변조를 통한 암호화 단계를 통해 암호화된 유전체 정보를 상기 변조/복조 규칙정보와 상기 유전형 빈도정보를 기초로 복조하여 유전체 정보를 복호화하는 단일염기변이영역 변조 유전형에 대한 복호화 단계를 포함한다.

발명의 효과

[0019] 본 발명에 따르면, 유전정보를 이진 형식으로 인코딩하거나, 암호화하기에 앞서, 암호화되지 않은 유전체 정보처럼 보이지만 실제 유전정보를 숨김으로써 유출된 유전체 정보의 진위여부를 판단할 수 없도록 함으로써, 개인의 변이정보나 원시서열정보를 암호화하여 개인의 유전체정보를 숨기고, 정해진 키 값에 의해서만 실제 정보를 복호화 가능하도록 하여 컴퓨터에 저장되거나 전송되는 과정에서 개인의 유전체정보가 유출되는 것을 방지할 수 있는 유전체 정보의 암호화 및 복호화 방법을 제공할 수 있다.

도면의 간단한 설명

[0020] 도 1은 본 발명의 전반적인 실시예에 따른 NGS(Next Generation Sequencing)를 통한 유전체 데이터의 생산 및 분석 과정을 나타낸 도면이다.

도 2는 본 발명의 제1 실시예에 따른 단일염기변이영역에서의 유전형 변조를 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 흐름도이다.

도 3은 본 발명의 제1 실시예에 따른 단일염기변이영역에서의 유전형 변조를 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 도면이다.

도 4는 본 발명의 제1 실시예에 따른 키(key) 샘플 군의 단일염기변이영역 및 유전형 빈도 정보를 나타낸 도면이다.

도 5는 본 발명의 제2 실시예에 따른 염기서열의 블록화 및 재배열을 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 흐름도이다.

도 6은 본 발명의 제2 실시예에 따른 염기서열의 블록화 및 재배열을 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 도면이다.

도 7은 본 발명의 제3 실시예에 따른 피검자 정보의 염기서열화와 삽입변이 유전체 은닉을 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 흐름도이다.

도 8은 본 발명의 제3 실시예에 따른 피검자 정보의 염기서열화와 삽입변이 유전체 은닉을 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 도면이다.

도 9는 본 발명의 각 실시예에 따른 변조/복조 규칙정보에 포함된 키 파라미터에 대한 예시를 나타낸 도면이다.

도 10은 본 발명의 제1 및 제3 실시예를 결합한 제4 실시예에 따른 유전체 정보의 암호화 및 복호화 방법의 각 단계를 나타낸 흐름도이다.

도 11은 본 발명의 제2 및 제3 실시예를 결합한 제5 실시예에 따른 유전체 정보의 암호화 및 복호화 방법의 각 단계를 나타낸 흐름도이다.

도 12는 본 발명의 제1 내지 제3 실시예를 결합한 제6 실시예에 따른 유전체 정보의 암호화 및 복호화 방법의 각 단계를 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0021] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 본 발명에 대해 구체적으로 설명하기로 한다.

[0022] 본 발명에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 발명에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 발명의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0023] 명세서 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나 이상의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

[0024] 아래에서는 첨부한 도면을 참고하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을

가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0025] [제1 실시예]

[0026] 도 1은 본 발명의 전반적인 실시예에 따른 NGS(Next Generation Sequencing)를 통한 유전체 데이터의 생산 및 분석 과정을 나타낸 도면이고, 도 2는 본 발명의 제1 실시예에 따른 단일염기변이영역에서의 유전형 변조를 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 흐름도이고, 도 3은 본 발명의 제1 실시예에 따른 단일염기변이영역에서의 유전형 변조를 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 도면이며, 도 4는 본 발명의 제1 실시예에 따른 키(key) 샘플 군의 단일염기변이영역 및 유전형 빈도 정보를 나타낸 도면이다.

[0027] 도 2를 참조하면, 본 발명의 제1 실시예에 따른 단일염기변이영역에서의 유전형 변조를 통한 유전체 정보의 암호화 및 복호화 방법(S100)은, 유전체 변이정보 확인 단계(S110), 단일염기변이영역 정렬 단계(S120), 제1 유전체 정보 암호화 단계(S130-1), 제2 유전체 정보 암호화 단계(S130-2), 제2 유전체 정보 복호화 단계(S140-1) 및 제1 유전체 정보 복호화 단계(S140-2) 중 적어도 하나를 포함할 수 있다.

[0028] 상기 유전체 변이정보 확인 단계(S110)에서는, 도 1에 도시된 바와 같이 차세대 염기서열 분석(NGS: Next Generation Sequencing)을 통한 유전체 데이터의 생산과 분석 과정을 거치면서 원천서열정보(raw read data)를 생산하고, 게놈지도에 대한 매핑을 통해 원천서열의 염색체와 염색체 위치를 확인하고, 확인된 염색체와 염색체 위치 중 게놈지도와 피검자의 유전형(allele)이 다른 서열위치 및 유전형을 확인하는 과정을 거칠 수 있다. 이러한 과정에서 유전체 원천서열정보는 fq 확장자의 파일로 생성되며, 서열 아이디, 실험기기 관련정보, 서열정보, 서열 및 유전형에 대한 실험 정확도 정보를 포함할 수 있다.

[0029] 본 실시예에서의 게놈지도 매핑은 유전체 원천서열정보에 포함된 수백 만개의 짧은 서열이 생체 세포 내에 존재하는 게놈지도 중 어떠한 염색체에 어떠한 위치의 서열인지를 확인하는 작업을 말하며, 이러한 게놈지도 매핑 정보는 sam 확장자 또는 binary code로 압축된 bam 확장자 파일로 저장될 수 있다.

[0030] 이러한 유전체 변이정보 확인 단계(S110)에서는 매핑된 원천서열로부터 게놈지도와 다른 유전형이 피검자에서 얼마나 발견되었는지를 비교함으로써 유전형 변이정보를 확인할 수 있으며, 이러한 정보들은 vcf 확장자 파일로 저장될 수 있다.

[0031] 상기 단일염기변이영역 정렬 단계(S120)에서는, 유전체 변이정보 중 단일염기변이가 존재하는 영역을 염색체 순서대로 정렬할 수 있다. 예를 들어, 한국인 변이체(혹은 보유한 유전체 정보 중 특정 그룹에 해당하는 키 샘플군)를 기준으로 단일염기변이(SNP)가 존재하는 영역만 'chr1-chrX/Y/M' 순으로 정렬할 수 있다(ex. 30억 bp -> 300만 bp).

[0032] 상기 제1 유전체 정보 암호화 단계(S130-1)에서는, 정렬된 단일염기변이 중 미리 저장된 변조/복조 규칙정보에 포함된 제1 키 파라미터 값 k_{SF}에 해당하는 위치의 변이마다 유전형 빈도정보를 기준으로 해당 변이의 유전형을 각각 변조하여 유전체 정보를 암호화할 수 있다. 여기서, 변조/복조 규칙정보는, 암호화/복호화 종류, 순서, 제1 키 파라미터 값과 키 샘플군의 유전형빈도 대신 미리 약속된 유전형 순서에 따라 변이영역을 변조하고자 하는 경우 유전형 변조 순서정보를 포함하는 파일 형태로 제공될 수 있다. 여기서, 제1 키 파라미터 값 k_{SF}는 단일염기변이 변조(SNV Flip)를 통한 암호화 방법에 사용되는 키 파라미터 값으로서, 본 실시예에서는 키 파라미터에 해당하는 특정 길이마다 염기를 변조하기 위한 위치지정 값으로 사용될 수 있다.

[0033] 예를 들어, 정렬된 단일염기변이영역 중 특정 키 파라미터 값(k_{SF} = 7)에 해당하는 위치 또는 영역마다 유전형(allele)을 하나씩 변조할 수 있는데, 유전형의 변조는 각 위치 또는 영역에 대한 키 샘플군에서의 유전형 빈도(allele frequency)정보를 기준으로 수행될 수 있으며, 가령 T:50%, C:30%, A:20% 영역의 경우 유전형 T는 C로 변조하고, A는 T로 변조하여 암호화 작업을 수행할 수 있다. 이러한 유전형 변조 과정을 통해 유전체 샘플이 보유하지 않은 변이가 생성되거나, 보유한 변이가 숨겨질 수 있다. 또한, 도 3에 나타난 바와 같이, 가령 G:50%, T:30%, A:20% 영역의 경우 유전형 G는 T로 변조하여 암호화 작업을 수행할 수 있다.

[0034] 상기 제2 유전체 정보 암호화 단계(S130-2)에서는, 제1 유전체 정보 암호화 단계(S130-1) 이후에, 정렬된 단일염기변이 중 k_{SF}번째 변이를 기준으로 제2 파라미터값 k_{BP}에 해당하는 bp만큼 좌 또는 우측에 위치한 변이영역과

유전형을 서로 바꾸어 유전체 정보를 암호화할 수 있다. 즉, 키 샘플군의 단일염기변이영역에 대하여 해당 샘플군내 유전형 빈도에 따라 변조하는 제1 유전체 정보 암호화 단계(S130-1) 외에도, 단일염기변이영역 중 k_{SF}번째 변이로부터 k_{BP}(양의 정수) bp(염기서열 길이)좌우에 위치한 변이영역을 대상으로 유전형을 서로 바꾸어 미리 약속된 형태로 변이정보를 변조할 수 있다. 여기서, 제2 파라미터값 k_{BP}는 단일염기변이 변조를 통한 암호화 방법에서 추가적인 암호화를 위해 사용되는 키 파라미터로서, 본 실시예에서는 k_{SF} 값에 따라 변조하는 영역으로부터 특정길이(base pair)에 위치한 좌우 유전형을 서로 바꾸기 위한 위치지정 값으로 사용될 수 있다. 이에 따라, 제1 키 파라미터는 단일염기변이영역 유전형변조 암호화 방법에서 변이영역 중 특정 bp 마다 변조할 영역 단위를 지정하는 키 파라미터 값을 의미하며, 제2 키 파라미터는 추가적인 암호화를 위하여, 제1 키 파라미터에 의해 정의된 위치를 기준으로 특정 bp 좌우 유전형을 바꾸도록 지정할 수 있는 파라미터로서 구분될 수 있다.

[0035] 상기 제2 유전체 정보 복호화 단계(S140-1)에서는, 제2 유전체 정보 암호화 단계(S130-2)를 통해 암호화된 유전체 정보의 단일염기변이 중 k_{SF}번째 변이를 기준으로 제2 유전체 정보 암호화 단계(S130-2)에서 선택된 좌우 변이에 대한 유전형을 역으로 서로 바꾸어 유전체 정보에 대한 사전 복호화 작업을 수행할 수 있다.

[0036] 상기 제1 유전체 정보 복호화 단계(S140-2)에서는, 제1 유전체 정보 암호화 단계(S130-1)를 통해 암호화된 유전체 정보를 미리 저장된 변조/복조 규칙정보와 유전형 빈도정보를 기초로 역순으로 복조하여 유전체 정보를 복호화할 수 있다.

[0037] 예를 들어, 변이정보의 복호화를 위해서는, 암호화에 사용된 키 샘플 군의 단일염기변이영역 및 유전형 빈도정보와 제1 키 파라미터 값(ex. k_{SF} = 7)이 필요하며, 해당 변이영역 중 매 k_{SF} 번째 위치 또는 영역을 해당 변이영역의 유전형 빈도를 기준 역으로 변조하여 복호화할 수 있으며, 가령, T:50%, C:30%, A:20% 영역의 경우 유전형 C는 T로 변조하고, T는 A로 변조하여 복호화 작업을 수행할 수 있다. 또한, 도 3과 관련하여, 예를 들면, 가령 G:50%, T:30%, A:20% 영역의 경우 유전형 T는 G로 복조하여 복호화 작업을 수행할 수 있다.

[0038] 이러한 변이정보의 변조를 통한 암호화 및 복호화 방법은 변이정보 데이터 형식을 유지하면서 변이정보를 변조함으로써, 변이의 진위여부 파악이 불가능하며 복호화를 위한 규칙과 키 파라미터에 대한 정보 없이는 원본의 형태로 변환할 수 없으며, 본 실시예에 따른 변조/복조 규칙이나 키 파라미터의 적용 방법을 한정하지 않는다.

[0039] 본 실시예의 단일염기변이영역에서의 유전형 변조는, 사전에 정의된 염색체 위치에 한하여 일정한 규칙(특정 위치를 특정 유전형으로)으로 적용 가능하며, 어떠한 염색체 위치를 변조영역으로 정의할지는 실제 유전체 샘플집단 중 변이가 존재하는 영역 및 유전형 빈도를 기준으로 함으로써, 일반적인 유전체 데이터의 형식으로 보이지만 원본과 다른 정보로 변조하고자 하는 목적을 달성할 수 있다.

[0040] 예를 들면, 도 4에 도시된 바와 같이, 샘플집단의 유전형 정보는 vcf 파일과 유사한 형태로 저장 가능하며, 염색체(CHR: Chromosome), 위치(POS: Position), 게놈지도 유전형(REF: Reference allele), 샘플집단 보유유전형(ALT: Altered allele), 샘플집단 유전형 빈도(AF: Allele Frequency) 정보를 포함할 수 있다.

[0041] 샘플집단(이하 key 샘플군)은 한국인 표준 유전체 샘플, 또는 서버에 저장된 유전체 샘플 중 일부 등에 대해 적용할 수 있으며, 어떠한 샘플집단을 기준으로 하는지에 따라 변조가 적용되는 유전체 위치, 유전형을 결정하므로, 암호화 및 복호화 결과가 달라질 수 있다.

[0042] 변이정보의 변조에 있어서 기준이 되는 키 샘플 군과 함께 해당 영역을 어떤 규칙으로 변조할 것인지(SNV_Flip)에 대한 정보는 별도의 키 파라미터 파일(변조/복조 규칙정보)로 저장할 수 있으며, 변조의 방식이나 기준이 되는 키 파라미터 값의 개수는 정의하는 방식에 따라 달라질 수 있다.

[0043] 본 발명의 제1 실시예에 따른 단일염기변이영역에서의 유전형 변조를 통한 유전체 정보의 암호화 및 복호화 방법에서 키 파라미터 값과 함께 키 정보로 사용되는 키 샘플 군 단일염기변이영역 및 유전형 빈도 정보는 파일로 저장되거나 서버를 통해 인증된 사용자에게 전달될 수 있다.

[0044] 키 샘플 군 단일염기변이영역 및 유전형 빈도를 기준으로 암호화하고자 하는 샘플의 단일염기변이영역의 유전형을 변조함으로써, 해당 피검자가 보유한 변이를 다른 유전형의 변이로 변환하거나 표준유전형으로 변환하여 변이를 숨길 수 있으며, 피검자가 표준유전형을 보유한 영역을 다른 유전형의 변이로 변환하여 가상변이를 생성할 수 있다.

[0045] 본 발명의 제1 실시예에 따른 단일염기변이영역에서의 유전형 변조를 통한 암호화 및 복호화 방법은, 변이정보 데이터 형식을 유지하면서 변이정보를 변조함으로써, 복호화를 위한 규칙과 키 파라미터에 대한 정보가 없이는

변이의 진위여부 파악이 불가능한 형태로 변환할 수 있으며, 변조 규칙이나 키 파라미터의 적용 방법에 대해서는 구체적으로 한정하지 않는다.

[0046] [제2 실시예]

[0047] 도 5는 본 발명의 제2 실시예에 따른 염기서열의 블록화 및 재배열을 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 흐름도이며, 도 6은 본 발명의 제2 실시예에 따른 염기서열의 블록화 및 재배열을 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 도면이다.

[0048] 도 5를 참조하면, 본 발명의 제2 실시예에 따른 염기서열의 블록화 및 재배열을 통한 암호화 및 복호화 방법(S200)은, 원천서열정보 및 변이정보 확인 단계(S210), 염기서열 블록 그룹 생성 단계(S220), 제1 유전체 정보 암호화 단계(S230-1), 제2 유전체 정보 암호화 단계(S230-2), 제2 유전체 정보 복호화 단계(S240-1), 제1 유전체 정보 복호화 단계(S240-2) 중 적어도 하나를 포함할 수 있다.

[0049] 상기 원천서열정보 및 변이정보 확인 단계(S210)는, 도 1에 도시된 바와 같이 차세대 염기서열 분석(NGS: Next Generation Sequencing)을 통한 유전체 데이터의 생산과 분석 과정을 거치면서 원천서열정보(raw read data)를 생산하고, 게놈지도에 대한 매핑을 통해 원천서열의 염색체와 염색체 위치를 확인하고, 확인된 염색체와 염색체 위치 중 게놈지도와 피검자의 유전형(allele)이 다른 서열위치 및 유전형을 확인하는 과정을 거칠 수 있다. 이러한 과정에서 유전체 원천서열정보는 fq 확장자의 파일로 생성되며, 서열 아이디, 실험기기 관련정보, 서열정보, 서열 및 유전형에 대한 실험 정확도 정보를 포함할 수 있다.

[0050] 본 실시예에서의 염기서열의 블록화 및 재배열을 통한 암호화 및 복호화 방법은 게놈지도에 매핑되기 전 단계의 원천서열정보 또는 상기 실시예에서의 유전체 변이정보 확인 단계(S110) 및 단일염기서열변이 정렬 단계(S120)와 동일한 단계를 통해 키 샘플군의 단일염기변이영역을 기준으로 정렬된 변이정보 대해 적용 가능하다.

[0051] 이러한 원천서열정보 및 변이정보 확인 단계(S210)에서는, 암호화를 적용할 대상이 원천서열정보일 경우, fq 확장자 파일로 저장된 원천서열들의 단편서열 길이(read length)와 단편서열 수를 확인할 수 있다. 암호화를 적용할 대상이 변이정보일 경우, 매핑된 원천서열로부터 게놈지도와 다른 유전형이 피검자에서 얼마나 발견되었는지를 비교함으로써 유전형 변이정보를 확인할 수 있으며, 확인된 변이정보는 특정 변이체 또는 키 샘플군의 변이영역을 기준으로 정렬하여 vcf 확장자 파일로 저장할 수 있다.

[0052] 상기 염기서열 블록 그룹 생성 단계(S220)에서는, 유전체 원천서열정보 및 유전형 변이정보 중 적어도 하나에 포함된 염기서열을 미리 저장된 변조/복조 규칙정보에 포함되어 있는 제1 키 파라미터 값 k_{BS}을 기본 단위길이로 나누어 다수의 블록 그룹을 생성할 수 있다.

[0053] 즉, 원천서열정보 또는 키 샘플 군의 변이영역으로 정렬된 모든 염기서열을 제1 키 파라미터 값(ex. k_{BS}=7)의 길이로 나누어 블록을 만들 수 있으며, 예를 들어 40bp 염기서열의 경우에는 '7a, 7b, 7c, 7d, 7e, 5f'와 같은 블록으로 각각 형성할 수 있다.

[0054] 상기 제1 유전체 정보 암호화 단계(S230-1)에서는, 블록 그룹의 순서를 변조/복조 규칙정보에 포함되어 있는 제2 키 파라미터 값 j_{BS}에 따라 재배열하여 유전체 정보를 암호화할 수 있다.

[0055] 예를 들어, 도 6에 도시된 바와 같이 제2 키 파라미터 값이 'j_{BS}=2인 경우, 해당 수의 그룹으로 나눈 후 재배열하여 암호화할 수 있는데, '7a 7b 7c 7d 7e 5f' 블록 그룹의 경우, '7a 7c 7e 7b 7d 5f' 등과 같은 형태로 재배열할 수 있으며, 제2 키 파라미터 값이 j_{BS} = 3인 경우, '7a 7b 7c 7d 7e 5f'의 블록 그룹을 '7a 7d 7b 7e 7c 5f' 등과 같은 형태로 재배열할 수 있다. 여기서, 변조/복조 규칙정보는, 암호화/복호화 종류, 순서, 상기 제1 키 파라미터 값, 상기 제2 키 파라미터 값을 포함하는 파일 형태로 제공될 수 있다.

[0056] 상기 제2 유전체 정보 암호화 단계(S230-2)에서는, 블록 그룹에 따라 재배열된 염기서열에 대해 제3 키 파라미터 값 k_R에 따라 전체 염기서열 중 n(양 혹은 음의 정수) bp를 좌측이나 우측에서 반대방향으로 이동시키는 등 미리 약속된 형태의 방식으로 염기서열을 추가적 또는 대체적으로 변형할 수 있다. 여기서, 제3 키 파라미터 값 k_R은 블록화 및 재배열을 통한 암호화 방법에서 추가적인 암호화를 위해 사용되는 키 파라미터 값으로서, 본 실시예에서는 재배열된 염기서열 중 좌측서열 또는 우측서열 일부를 잘라서 반대방향으로 옮길(Rotate) 영역을 지정하기 위한 값으로 사용될 수 있다.

[0057] 상기 제2 유전체 정보 복호화 단계(S240-1)에서는, 제2 유전체 정보 암호화 단계(S230-2)를 통해 암호화된 유전

체 정보의 염기서열 중 제2 유전체 정보 암호화 단계(S230-2) 단계에서 반대방향으로 옮겨진 염기서열을 제3 키 파라미터 값에 따라 원래 위치로 돌려놓음으로써 유전체 정보를 사전 복호화할 수 있다.

[0058] 상기 제1 유전체 정보 복호화 단계(S240-2)에서는, 제1 유전체 정보 암호화 단계(S230-1)를 통해 암호화된 블록 그룹의 순서를 제2 키 파라미터 값에 따라 역순으로 재배열하여 유전체 정보를 복호화할 수 있다.

[0059] 예를 들어, 블록을 제2 키 파라미터 값(ex. j_{BS}=2)의 그룹으로 나누고 역으로 재배열하여'7a, 7c, 7e, 7b, 7d, 5f'의 블록을 '7a, 7b, 7c, 7d, 7e, 5f' 블록의 원본 염기서열정보로 복호화할 수 있다.

[0060] 본 발명의 제2 실시예에 따른 염기서열 블록화 및 재배열을 통한 암호화 및 복호화 방법은 원천서열 또는 변이정보 형식을 유지하면서 서열정보를 변조함으로써, 복호화를 위한 규칙과 키 파라미터(key parameter)에 대한 정보 없이는 원본 서열의 유전체지도에 대한 매핑 또는 변이정보 확인이 불가능한 형태로 변환하며, 블록화 또는 재배열 규칙이나 키 파라미터(key parameter)의 적용 방법에 대해서는 구체적으로 한정하지 않는다.

[0061] [제3 실시예]

[0062] 도 7은 본 발명의 제3 실시예에 따른 피검자 정보의 염기서열화와 삽입변이 유전체 은닉을 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 흐름도이고, 도 8은 본 발명의 제3 실시예에 따른 피검자 정보의 염기서열화와 삽입변이 유전체 은닉을 통한 유전체 정보의 암호화 및 복호화 방법을 설명하기 위해 나타낸 도면이며, 도 9는 본 발명의 각 실시예에 따른 변조/복조 규칙정보에 포함된 키 파라미터에 대한 예시를 나타낸 도면이다.

[0063] 도 7을 참조하면, 본 발명의 제2 실시예에 따른 피검자 정보의 염기서열화와 삽입변이 유전체 은닉을 통한 변이정보의 암호화 및 복호화 방법(S300)은, 피검자 정보 염기서열 변환 단계(S310), 가상변이 삽입위치 변환 단계(S320), 피검자 정보 암호화 단계(S330) 및 피검자 정보 복호화 단계(S340) 중 적어도 하나를 포함할 수 있다.

[0064] 상기 피검자 정보 염기서열 변환 단계(S310)는, 문자열 및 수치 형식으로 표현되는 피검자의 개인정보 및 임상정보를 미리 저장된 인코딩/디코딩 변환테이블을 기초로 4진수 형태로 인코딩하여 염기서열형태로 변환할 수 있다.

[0065] 예를 들어, 도 8에 도시된 바와 같이 피검자의 개인정보 및 임상정보가 "1.이름: 홍길동, 2.생년월일: 2020.01.01, 3.혈당: 72, 4.키:173.2"를 포함하는 경우, 'ATCGCATC, GGTCACAC, TCGT, TACGAA'와 같은 염기서열 형태로 변환할 수 있다.

[0066] 이러한 피검자 정보의 염기서열 변환 단계(S310)에서, 문자열 또는 수치 형식으로 이루어진 피검자 정보의 4진수 염기서열화 변형은 UTF-8 등 알려진 인코딩/디코딩 변환테이블 또는 데이터 소유자와 인증된 사용자간 정의한 변환테이블에 따라 달라질 수 있으며, 각각의 정보에 대한 가상변이를 삽입할 염색체나 변이위치 또한 정의한 수식에 따라 달라질 수도 있다.

[0067] 상기 가상변이 삽입위치 변환 단계(S320)에서는, 피검자의 개인정보 및 임상정보에 포함된 각 항목에 대하여 항목번호와 미리 설정된 제1 키 파라미터 값 k_{VI}을 곱한 후 염색체 수로 나눈 나머지 제1 값과, 항목번호와 제1 키 파라미터 값의 곱을 미리 설정된 제2 키 파라미터 값 j_{VI}만큼 제곱 연산한 제2 값을 산출할 수 있다. 여기서, 제1 키 파라미터 값 k_{VI}은 피검자 정보의 염기서열화 및 삽입변이 은닉을 위한 암호화 방법에서 변이삽입(Variant Insertion) 위치를 결정하기 위해 사용되는 파라미터 값으로서, 본 실시예에서는 0~22 범위의 임의의 수를 구하기 위해 염색체 수로 나눈 나머지 값을 연산하기 전 항목번호를 1차 암호화하기 위한 배수 값으로 사용될 수 있다. 또한, 제2 키 파라미터 값 j_{VI}은 피검자 정보의 염기서열화 및 삽입변이 은닉을 위한 암호화 방법에서 변이삽입(Variant Insertion) 위치를 결정하기 위해 사용되는 파라미터 값으로서, 본 실시예에서는 k_{SF} 값에 따라 구한 0~22 범위의 나머지 제 1값으로부터, 변이를 은닉할 500만~2500만bp에 해당하는 임의의 염색체 위치를 구하기 위한 거듭제곱 값으로 사용될 수 있다.

[0068] 상기 피검자 정보 암호화 단계(S330)에서는, 가상변이 삽입위치 변환 단계(S320)를 통해 변환된 피검자의 개인정보 및 임상정보가 포함된 각 항목에 해당하는 염색체 위치마다 피검자 정보 염기서열 변환 단계(S310)를 통해 변환된 염기서열을 변이정보로서 삽입하여 피검자 정보를 유전체 데이터에 은닉할 수 있다.

[0069] 예를 들어, 항목번호(키의 경우, N = 4)에 특정 제1 키 파라미터의 값(ex. k_{VI}=7)을 곱하여 염색체 수(23)로 나눈 나머지에 해당하는 염색체에 삽입변이(Insertion)를 은닉할 수 있다. 삽입위치는 항목번호와 제1 키 파라미

터 값(ex. k_{VI}=7)의 곱을 제2 키 파라미터 값(ex. j_{VI}=4)만큼 제곱하는 연산으로 지정할 수 있으며, 가령 '4.키:173.2' 항목의 경우 '(4*7) mod 23 = 5 AND $(4*7)^4$ = 614,656'와 같은 계산 식을 통하여 'chr5:614,656G' 위치에 'GTACGAA'의 유전형으로 삽입하여 은닉할 수 있다.

[0070] 여기서, 변조/복조 규칙정보는, 암호화/복호화 종류, 순서, 상기 인코딩/디코딩 변환테이블, 제1 키 파라미터 값, 제2 키 파라미터 값을 포함하는 파일 형태로 제공될 수 있다.

[0071] 상기 피검자 정보 복호화 단계(S340)에서는, 피검자 정보 염기서열 변환 단계(S310)를 통해 암호화된 가상염기서열을 제1 키 파라미터 값, 제2 키 파라미터 값, 제1 값, 상기 제2 값, 및 인코딩/디코딩 변환테이블을 기초로 역으로 연산을 진행하여 피검자의 개인정보 및 임상정보를 복호화할 수 있다.

[0072] 예를 들어, 항목(키의 경우, N = 4)마다 특정 제1 키 파라미터 값(ex. k_{VI}=7)으로부터 염색체와 각 영역 위치를 역으로 계산하고, 해당 위치의 삽입서열을 디코딩하여 문자열 혹은 수치 형식의 개인정보 및 임상정보를 복원할 수 있으며, 피검자 정보 은닉을 위해 삽입된 가상변이는 원본의 변이정보 복호화 과정에서 선택적으로, 제외할 수 있다. 예를 들어, '4.키'에 대한 항목에 대하여 'chr5:614,656GTACGAA' 로 암호화된 삽입변이 'TACGAA'를 디코딩하여 '173.2'라는 키에 해당하는 값으로 변환하고, 암호화된 피검자의 유전형 데이터에서 해당 가상변이를 제외함으로써 이 영역의 원본 유전형을 복호화할 수 있다.

[0073] 본 발명의 제3 실시예에 따른 피검자 정보의 염기서열화와 삽입변이 유전체 은닉을 통한 변이정보의 암호화 및 복호화 방법은 변이정보 데이터 형식을 유지하면서 샘플의 개인정보 또는 임상정보를 염기서열 변이정보로 변환하여 유전체 데이터에 은닉함으로써, 도 8의 (a)의 샘플정보 및 임상정보 항목이 비어있거나, 가상으로 기록되어 있더라도, 적용 방법과 키 파라미터(key parameter) 정보를 활용하여 변이 데이터에 은닉되어 있는 실제 정보를 다시 복원할 수 있으며, 문자열 또는 수치형 정보의 염기서열 변환 방식이나 정보은닉 위치를 정의하는 특정 수식에 대해서는 구체적으로 한정하지 않는다.

[0074] 도 10은 본 발명의 제1 및 제3 실시예를 결합한 제4 실시예에 따른 유전체 정보의 암호화 및 복호화 방법의 각 단계를 나타낸 흐름도이고, 도 11은 본 발명의 제2 및 제3 실시예를 결합한 제5 실시예에 따른 유전체 정보의 암호화 및 복호화 방법의 각 단계를 나타낸 흐름도이며, 도 12는 본 발명의 제1 내지 제3 실시예를 결합한 제6 실시예에 따른 유전체 정보의 암호화 및 복호화 방법의 각 단계를 나타낸 흐름도이다.

[0075] 상술한 제1 실시예에 따른 단일염기변이영역에서의 유전형 변조를 통한 암호화 및 복호화 방법, 제2 실시예에 따른 염기서열의 블록화 및 재배열을 통한 암호화 및 복호화 방법, 및 제3 실시예에 따른 피검자 정보의 염기서열화 및 삽입변이 유전체 은닉을 통한 암호화 및 복호화 방법을 두 개 이상을 조합하여 암호화할 수 있으며, 적용순서와 각 키 파라미터를 역으로 적용하여 복호화할 수 있다.

[0076] 예를 들어, '제1 실시예에 따른 단일염기변이영역에서의 유전형 변조를 통한 암호화 및 복호화 방법, 및 제3 실시예에 따른 피검자 정보의 염기서열화 및 삽입변이 은닉을 통한 암호화 및 복호화 방법이 순차적으로 조합된 제4 실시예', '제2 실시예에 따른 염기서열의 블록화 및 재배열을 통한 암호화 및 복호화 방법, 및 제3 실시예에 따른 피검자 정보의 염기서열화 및 삽입변이 은닉을 통한 암호화 및 복호화 방법이 순차적으로 조합된 제5 실시예'와, '제1 실시예에 따른 단일염기변이영역에서의 유전형 변조를 통한 암호화 및 복호화 방법, 제2 실시예에 따른 염기서열의 블록화 및 재배열을 통한 암호화 및 복호화 방법, 및 제3 실시예에 따른 피검자 정보의 염기서열화 및 삽입변이 은닉을 통한 암호화 및 복호화 방법이 순차적으로 조합된 제6 실시예'가 추가적으로 실시 가능하다.

[0077] 이와 같이 조합된 실시예의 경우, 키 샘플 군의 단일변이영역을 기준으로 정렬된 변이정보 형식(.vcf)의 유전체 데이터에 개인정보 또는 임상정보의 염기변환과 유전체 은닉을 통한 암호화 및 복호화 방법을 적용하여 가상의 삽입변이를 생성할 수 있다.

[0078] 또한, 키 샘플 군의 단일변이영역 및 가상으로 생성된 삽입변이 영역에 대해 단일염기 유전형 변조를 통한 암호화 및 복호화 방법을 적용하여 일정한 변이영역의 유전형을 변조할 수 있으며, 가상으로 생성된 삽입변이는 변조할 유전형의 기준이 되는 키 샘플 군의 유전형빈도 정보가 없으므로, 도 4에 도시된 'nt_order'에 대한 키 파라미터의 예시처럼 미리 정의된 유전형 순서에 따라 변조하여 암호화 및 복호화를 진행할 수 있다.

[0079] 또한, 염기서열의 블록화 및 재배열을 통한 암호화 및 복호화 방법은 정렬된 키 샘플군의 단일염기변이영역에 대해서도 동일하게 적용 가능하며, 이때 서열 블록화 및 재배열 암호화 방법은 샘플이 보유한 삽입변이, 결실변이 및 개인정보 또는 임상정보의 유전체 은닉을 위해 가상으로 생성된 삽입변이를 모두 제외하고 키 샘플 군의

단일염기변이영역에만 적용하는 등 각 암호화 방법의 적용 대상을 구분하여 적용함으로써, 암호화의 영향이 중첩되어 복호화 오류가 발생하는 것을 피할 수 있으며, 특정 유전체 데이터에 복합적으로 적용된 암호화 방법과 키 파라미터를 역순으로 적용하여 원본 변이 데이터와 피검자의 개인정보 또는 임상정보를 복호화할 수 있다.

[0080] 원천서열정보, 게놈지도 매핑 정보, 변이정보의 형태로 서버에 저장된 유전체 데이터를 통상적인 NGS 분석 방법에 따라 원천서열정보(.fq) 또는 변이정보(.vcf) 형태로 각각 변환하여 암호화 및 복호화 방법을 위한 프로세스에 적용할 수 있다.

[0081] 데이터 소유자(또는 피검자)는 특정 유전체 데이터에 적용한 암호화 또는 데이터 은닉 방법의 종류 및 순서와 파라미터 값들을 키 파라미터 정보로 기록될 수 있으며, 인증된 사용자는 전달받은 키 파라미터 정보에 따라 적용된 암호화 방법의 종류 및 순서와 해당 파라미터 값들을 역으로 계산하여 원본 유전체 정보 또는 해당 유전체 데이터의 피검자 정보를 복호화할 수 있다.

[0082] 암호화 또는 데이터 은닉 방법에 사용되는 키 파라미터 값은 도 9에 도시된 바와 같은 항목을 포함하며, 유효한 범위(ex. 염기서열 재배열 시 블록을 나누는 기준이 되는 k_{BS}의 경우, 1보다 큰 정수)에서 난수를 생성하여 지정하거나, 피검자의 개인정보 또는 임상정보 중 특정 항목 값의 연산을 통해 확인할 수 있는 형태로 기록되고 전달될 수 있다.

[0083] 각각의 키 파라미터 값은 암호화 또는 데이터 은닉 방법에서의 유전체 데이터의 변이정보 변조의 빈도에 영향을 줄 수 있으며, 데이터마다 요구되는 보안의 수준에 따라 각 파라미터 값을 크게 혹은 작게 지정함으로써, 원본 변이정보와 유사하거나 상이한 패턴으로 유전체 변이정보를 암호화할 수 있다.

[0084] 이와 같이 본 실시예들은, 유전체 변이 분석 과정에서 서버에 생산되고 저장되는 파일 중 원천서열정보 또는 변이정보에 포함된 데이터를 일정한 규칙으로 변조하거나, 피검자의 정보를 변이정보 형태로 은닉함으로써, 의도하지 않게 서버 내 유전체 정보가 유출되더라도 일반적인 유전체 데이터의 형식으로 보이지만 원본과 다른 정보를 포함하는 가상 데이터를 생산할 수 있다.

[0085] 본 실시예는, 암호화 및 복호화 방법을 각각 실행할 수 있는 프로그램이 기록된 컴퓨터 판독 가능 기록매체로서 구현될 수 있으며, 해당 기록매체를 이용하여 상기 방법에 따른 프로세스를 동작시키기 위한 암호화 및 복호화 실행 장치로도 구현될 수도 있다.

[0086] 이상에서 설명한 것은 본 발명에 의한 유전체 정보의 암호화 및 복호화 방법을 실시하기 위한 하나의 실시예에 불과한 것으로서, 본 발명은 상기 실시예에 한정되지 않고, 이하의 특허청구범위에서 청구하는 바와 같이 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변경 실시가 가능한 범위까지 본 발명의 기술적 정신이 있다고 할 것이다.

부호의 설명

[0087] S100: 유전체 정보의 암호화 및 복호화 방법(단일염기변이영역에서의 유전체 변조를 통한 암호화 및 복호화 방법)

S110: 유전체 변이정보 확인 단계

S120: 단일염기서열변이 정렬 단계

S130-1: 제1 유전체 정보 암호화 단계

S130-2: 제2 유전체 정보 암호화 단계

S140-1: 제2 유전체 정보 복호화 단계

S140-2: 제1 유전체 정보 복호화 단계

S200: 유전체 정보의 암호화 및 복호화 방법(염기서열의 블록화 및 재배열을 통한 암호화 및 복호화 방법)

S210: 원천서열정보 및 변이정보 확인 단계

S220: 염기서열 블록 그룹 생성 단계

S230-1: 제1 유전체 정보 암호화 단계

S230-2: 제2 유전체 정보 암호화 단계

S240-1: 제2 유전체 정보 복호화 단계

S240-2: 제1 유전체 정보 복호화 단계

S300: 유전체 정보의 암호화 및 복호화 방법(피검자의 염기서열화 및 삽입변이 유전체 은닉을 통한 암호화 및 복호화 방법)

S310: 피검자 정보 염기서열 변환 단계

S320: 가상변이 삽입위치 변환 단계

S330: 피검자 정보 암호화 단계

S340: 피검자 정보 복호화 단계

S400: 유전체 정보의 암호화 및 복호화 방법(단일염기변이영역에서의 유전체 변조와, 피검자의 염기서열화 및 삽입변이 유전체 은닉을 통한 암호화 및 복호화 방법)

S410: 단일염기변이영역에서의 유전형 변조를 통한 암호화 단계

S420: 염기서열화 및 삽입변이 은닉을 통한 암호화 단계

S430: 염기서열화 및 은닉 삽입변이에 대한 복호화 단계

S440: 단일염기변이영역 변조 유전형에 대한 복호화 단계

S500: 유전체 정보의 암호화 및 복호화 방법(염기서열의 블록화 및 재배열과, 피검자의 염기서열화 및 삽입변이 유전체 은닉을 통한 암호화 및 복호화 방법)

S510: 염기서열의 블록화 및 재배열을 통한 암호화 단계

S520: 염기서열화 및 삽입변이 은닉을 통한 암호화 단계

S530: 염기서열화 은닉 삽입변이에 대한 복호화 단계

S540: 블록화 및 재배열 염기서열에 대한 복호화 단계

S600: 유전체 정보의 암호화 및 복호화 방법(단일염기변이영역에서의 유전형 변조, 염기서열의 블록화 및 재배열, 및 피검자 정보의 염기서열화 및 삽입변이 은닉을 통한 유전체 정보의 암호화 및 복호화 방법)

S610: 단일염기변이영역에서의 유전형 변조를 통한 암호화 단계

S620: 염기서열의 블록화 및 재배열을 통한 암호화 단계

S630: 염기서열화 및 삽입변이 은닉을 통한 암호화 단계

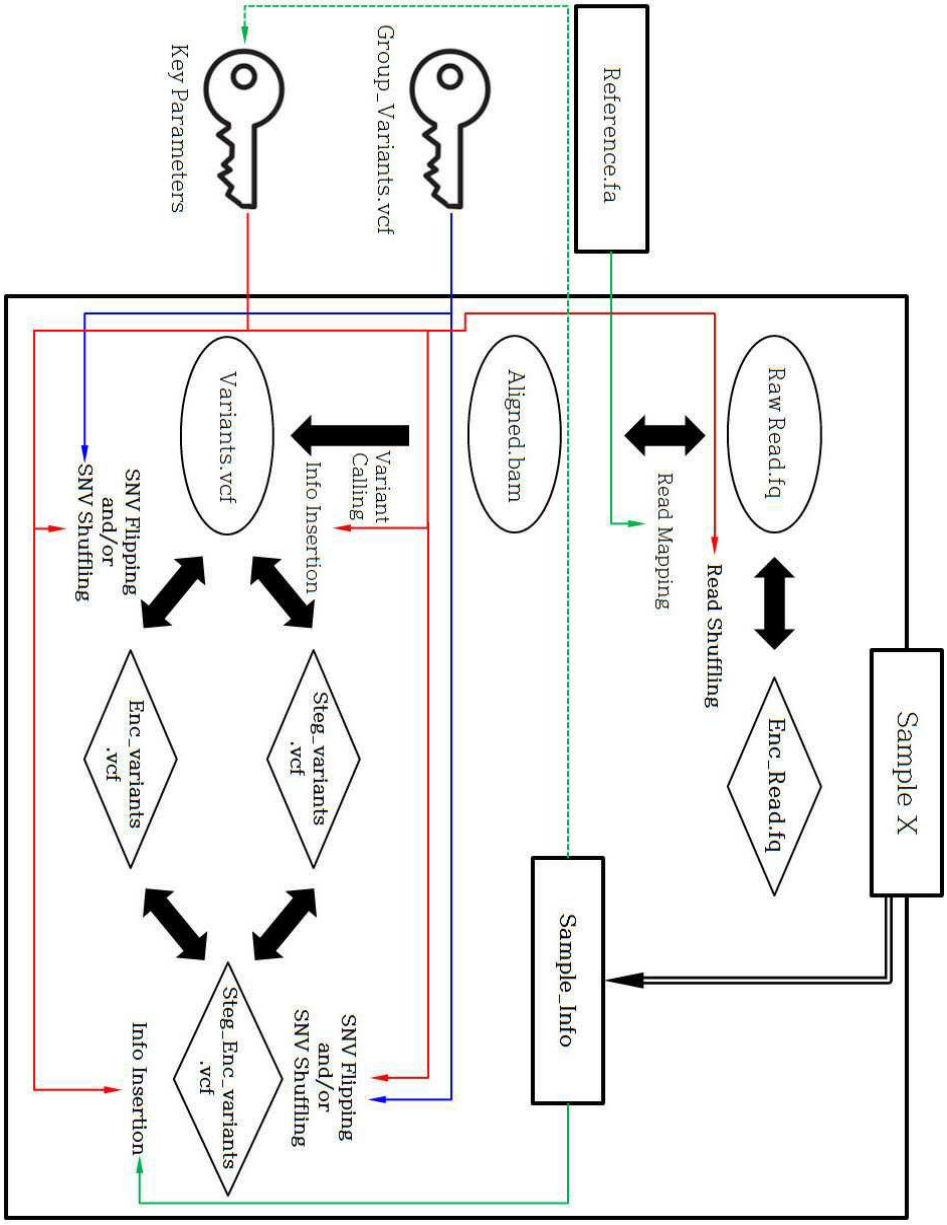
S640: 염기서열의 은닉 삽입변이에 대한 복호화 단계

S650: 블록화 및 재배열 염기서열에 대한 복호화 단계

S660: 단일염기변이영역 변조 유전형에 대한 복호화 단계

도면

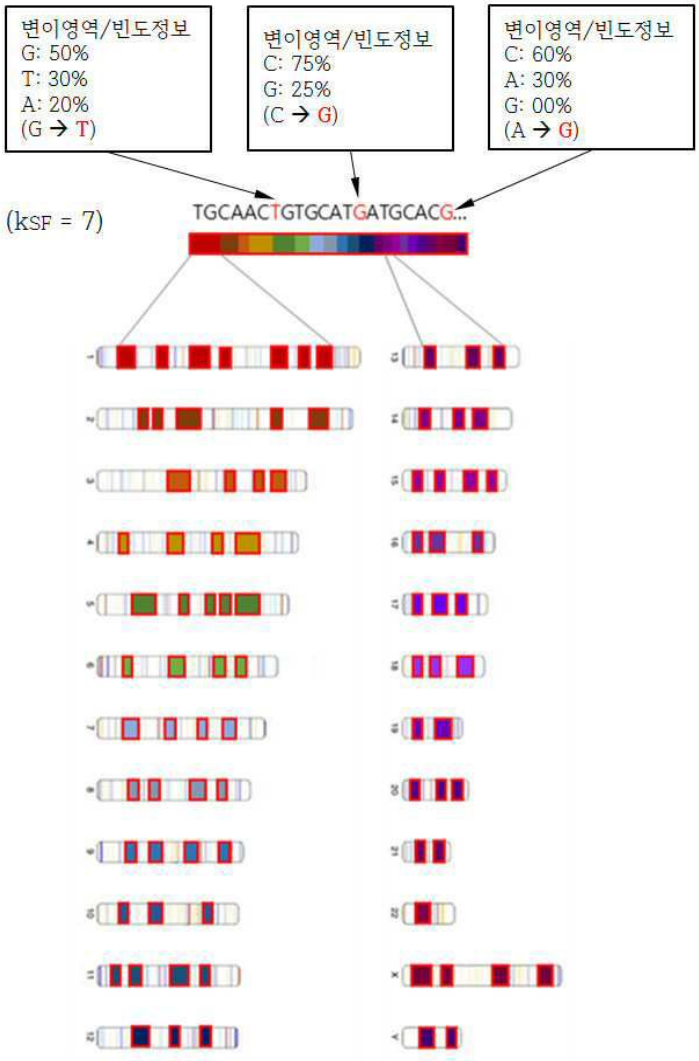
도면1


Reference.fa
Group_Variants.vcf
Key Parameters
Raw Read.fq
Read Shuffling
Enc_Read.fq
Sample X
Read Mapping
Aligned.bam
Variant Calling
Variants.vcf
Info Insertion
SNV Flipping and/or SNV Shuffling
Steg_variants .vcf
Enc_variants .vcf
Sample_Info
Steg_Enc_variants .vcf
SNV Flipping and/or SNV Shuffling
Info Insertion

도면2


S100
유전체 변이정보 확인 단계
S110
단일염기변이영역 정렬 단계
S120
제1 유전체 정보 암호화 단계
S130-1
제2 유전체 정보 암호화 단계
S130-2
제2 유전체 정보 복호화 단계
S140-1
제1 유전체 정보 복호화 단계
S140-2

도면3


변이영역/빈도정보
G: 50%
T: 30%
A: 20%
(G → T)
변이영역/빈도정보
C: 75%
G: 25%
(C → G)
변이영역/빈도정보
C: 60%
A: 30%
G: 00%
(A → G)
(kSF = 7)
TGCAACTGTGCATGATGCACG...

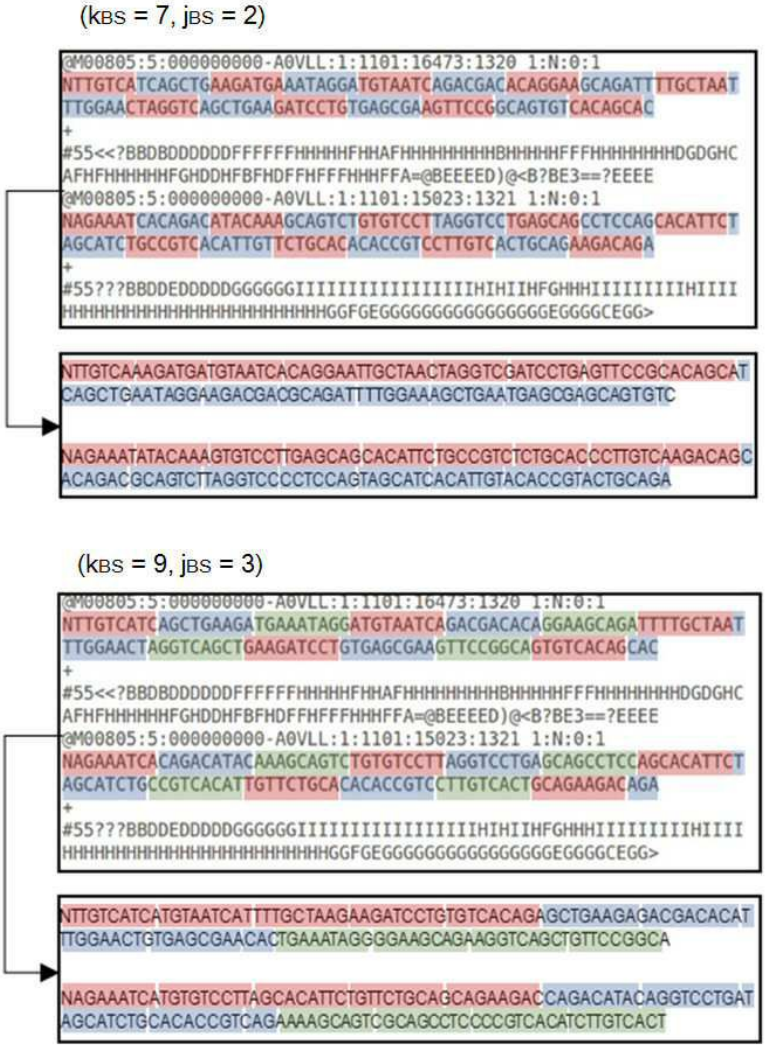
도면4

CHR	POS	REF	ALT	AF
1	611317	A	G	0.7, 0.3
1	817341	G	C,A	0.6, 0.2, 0.2
1	1277925	C	A	0.9, 0.1
1	2356177	T	C	0.5, 0.5
1	3341978	T	G	0.6, 0.4
2	279705	C	T,A	0.6, 0.3, 0.1
2	865143	G	C	0.8, 0.2
2	1315829	T	A,G,C	0.4, 0.3, 0.2, 0.1
2	1371538	T	C,A	0.5, 0.4, 0.1
3	211996	C	A	0.7, 0.3
3	846227	T	A	0.6, 0.4
…				
…				
X	304103	A	T	0.6, 0.4
X	356577	C	G,A	0.5, 0.3, 0.2
Y	278913	G	C,T	0.6, 0.2, 0.2
M	4923	C	A	0.5, 0.5

도면5



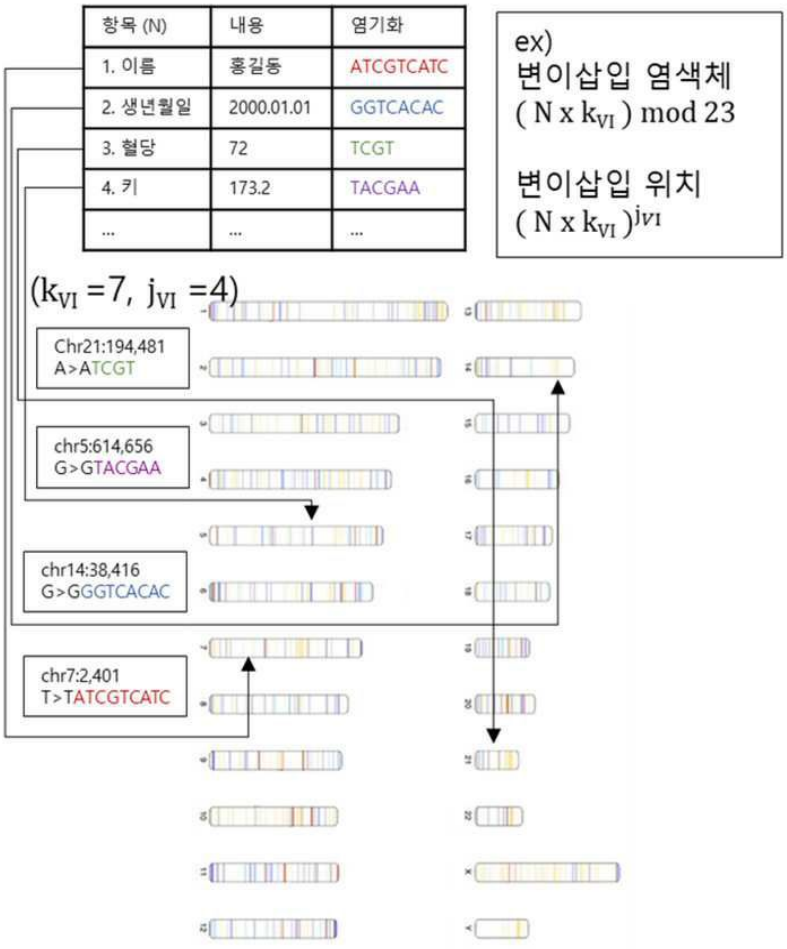
도면6


(kBS = 7, jBS = 2)
@M00805:5:000000000-A0VLL:1:1101:16473:1320 1:N:0:1
NTTGTCATCAGCTGAAGATGAAATAGGATGTAATCAGACGACACAGGAAGCAGATTTTGCTAAT
TTGGAACTAGGTCAGCTGAAGATCCTGTGAGCGAAGTTCCGGCAGTGTCACAGCAC
+
#55<<?BBDBDDDDDDFFFFFFHHHHHFHHAFHHHHHHHHHHBHHHHHFFFHHHHHHHHDGDGHC
AFHFHHHHHHFGHDDHFBFHDFFHFFFHHHFFA=@BEEEED)@<B?BE3==?EEEE
@M00805:5:000000000-A0VLL:1:1101:15023:1321 1:N:0:1
NAGAAATCACAGACATACAAAGCAGTCTGTGTCCTTAGGTCCTGAGCAGCCTCCAGCACATTCT
AGCATCTGCCGTCACATTGTTCTGCACACACCGTCCTTGTCACTGCAGAAGACAGA
+
#55???BBDDEDDDDDGGGGGGIIIIIIIIIIIIIIIIIIHIHIIHFGHHHIIIIIIIIIHIIII
HHHHHHHHHHHHHHHHHHHHHHHHHGGFGEGGGGGGGGGGGGGGGGEGGGGCEGG>
NTTGTCAAAGATGATGTAATCACAGGAATTGCTAACTAGGTCGATCCTGAGTTCCGCACAGCAT
CAGCTGAATAGGAAGACGACGCAGATTTTGGAAAGCTGAATGAGCGAGCAGTGTC
NAGAAATATACAAAGTGTCCTTGAGCAGCACATTCTGCCGTCTCTGCACCCTTGTCAAGACAGC
ACAGACGCAGTCTTAGGTCCCCTCCAGTAGCATCACATTGTACACCGTACTGCAGA
(kBS = 9, jBS = 3)
@M00805:5:000000000-A0VLL:1:1101:16473:1320 1:N:0:1
NTTGTCATCAGCTGAAGATGAAATAGGATGTAATCAGACGACACAGGAAGCAGATTTTGCTAAT
TTGGAACTAGGTCAGCTGAAGATCCTGTGAGCGAAGTTCCGGCAGTGTCACAGCAC
+
#55<<?BBDBDDDDDDFFFFFFHHHHHFHHAFHHHHHHHHHHBHHHHHFFFHHHHHHHHDGDGHC
AFHFHHHHHHFGHDDHFBFHDFFHFFFHHHFFA=@BEEEED)@<B?BE3==?EEEE
@M00805:5:000000000-A0VLL:1:1101:15023:1321 1:N:0:1
NAGAAATCACAGACATACAAAGCAGTCTGTGTCCTTAGGTCCTGAGCAGCCTCCAGCACATTCT
AGCATCTGCCGTCACATTGTTCTGCACACACCGTCCTTGTCACTGCAGAAGACAGA
+
#55???BBDDEDDDDDGGGGGGIIIIIIIIIIIIIIIIIIHIHIIHFGHHHIIIIIIIIIHIIII
HHHHHHHHHHHHHHHHHHHHHHHHHGGFGEGGGGGGGGGGGGGGGGEGGGGCEGG>
NTTGTCATCATGTAATCATTTTGCTAAGAAGATCCTGTGTCACAGAGCTGAAGAGACGACACAT
TTGGAACTGTGAGCGAACACTGAAATAGGGGAAGCAGAAGGTCAGCTGTTCCGGCA
NAGAAATCATGTGTCCTTAGCACATTCTGTTCTGCAGCAGAAGACCAGACATACAGGTCCTGAT
AGCATCTGCACACCGTCAGAAAAGCAGTCGCAGCCTCCCCGTCACATCTTGTCACT

도면7


S300
피검자 정보 염기서열 변환 단계
S310
가상변이 삽입위치 변환 단계
S320
피검자 정보 암호화 단계
S330
피검자 정보 복호화 단계
S340

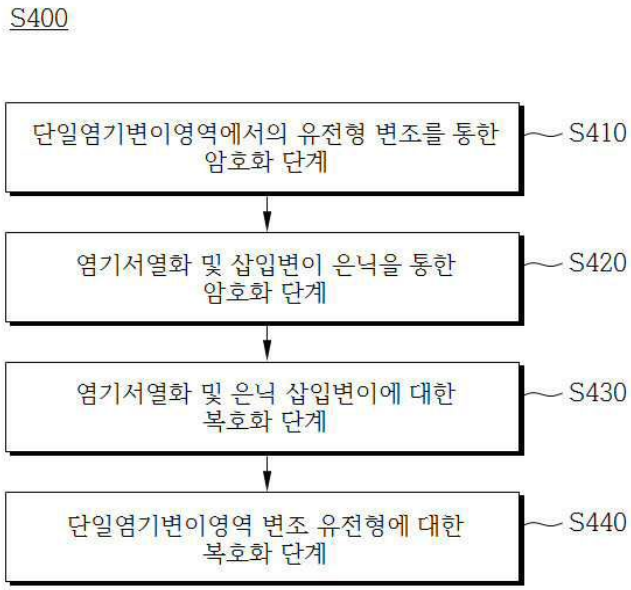
도면8


항목 (N)
내용
염기화
1. 이름
홍길동
ATCGTCATC
2. 생년월일
2000.01.01
GGTCACAC
3. 혈당
72
TCGT
4. 키
173.2
TACGAA
...
...
...
ex)
변이삽입 염색체
(N x k_{VI}) mod 23
변이삽입 위치
(N x k_{VI})^{jVI}
(k_{VI} =7, j_{VI} =4)
Chr21:194,481
A>ATCGT
chr5:614,656
G>GTACGAA
chr14:38,416
G>GGGTCACAC
chr7:2,401
T>TATCGTCATC

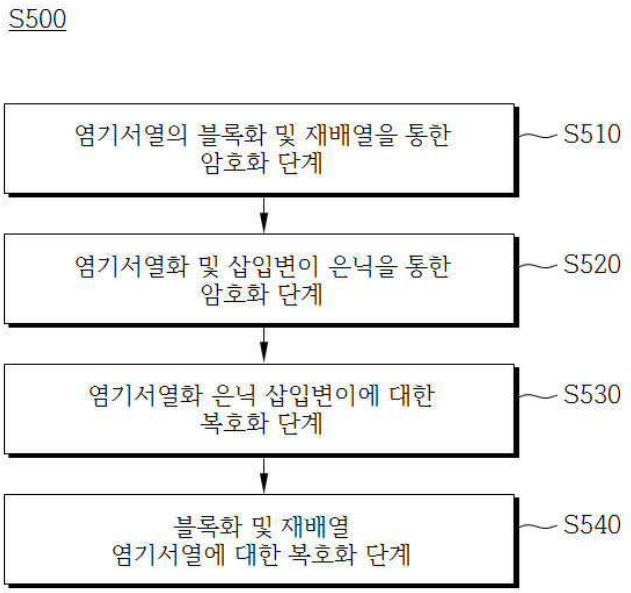
도면9


Reference_Info = [GRCJ38.p13]
Key_Group_File = [group_var.vcf]
Sample_Info =
[
1. 이름: 홍길동,
2. 생년월일: 2000.01.01,
3. 키: 173.2,
4. 혈당: 72,
…
]
Method_Parameter =
[
1-0. Block_Shuffle = False,
1-1. kBS = None,
1-2. jBS = None,
2-1. SNV_Flip = 2,
2-2. kSF = 7,
3-1. Variant_Insertion = 1,
3-2. kVI = 7,
3-3. jVI = 4,
3-4. encoding = UTF-8,
3-5. nt_order = 'ACGT'
]
…
(a)
Reference_Info = [GRCJ38.p13]
Key_Group_File = [group_var.vcf]
Sample_Info =
[
1. 이름: 홍길순,
2. 생년월일: 2000.01.01,
3. 키: 160.5,
4. 혈당: 65,
…
]
Method_Parameter =
[
1-0. Block_Shuffle = 1,
1-1. kBS = Sample_Info_1 mod 5 + 5,
1-2. jBS = Sample_Info_2 mod 3 + 2,
2-1. SNV_Flip = False,
2-2. kSF = None,
3-1. Variant_Insertion = False,
3-2. kVI = None,
3-3. jVI = None,
3-4. encoding = None,
3-5. nt_order = None
]
…
(b)

도면10


S400
단일염기변이영역에서의 유전형 변조를 통한 암호화 단계
S410
염기서열화 및 삽입변이 은닉을 통한 암호화 단계
S420
염기서열화 및 은닉 삽입변이에 대한 복호화 단계
S430
단일염기변이영역 변조 유전형에 대한 복호화 단계
S440

도면11


S500
염기서열의 블록화 및 재배열을 통한 암호화 단계
S510
염기서열화 및 삽입변이 은닉을 통한 암호화 단계
S520
염기서열화 은닉 삽입변이에 대한 복호화 단계
S530
블록화 및 재배열 염기서열에 대한 복호화 단계
S540

도면12


S600
단일염기변이영역에서의 유전형 변조를 통한 암호화 단계
S610
염기서열의 블록화 및 재배열을 통한 암호화 단계
S620
염기서열화 및 삽입변이 은닉을 통한 암호화 단계
S630
염기서열의 은닉 삽입변이에 대한 복호화 단계
S640
블록화 및 재배열 염기서열에 대한 복호화 단계
S650
단일염기변이영역 변조 유전형에 대한 복호화 단계
S660