

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2015년 12월 30일 (30.12.2015)



(10) 국제공개번호
WO 2015/199440 A1

- (51) 국제특허분류:
G06F 21/60 (2013.01) *G06F 19/10* (2011.01)
- (21) 국제출원번호: PCT/KR2015/006427
- (22) 국제출원일: 2015년 6월 24일 (24.06.2015)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2014-0077019 2014년 6월 24일 (24.06.2014) KR
10-2015-0088901 2015년 6월 23일 (23.06.2015) KR
- (71) 출원인: 서울대학교산학협력단 (SEOUL NATIONAL UNIVERSITY R&DB FOUNDATION) [KR/KR]; 151-015 서울시 관악구 관악로 1, Seoul (KR).
- (72) 발명자: 김주한 (KIM, Ju Han); 138-911 서울시 송파구 올림픽로 135, 203 동 1501 호, Seoul (KR).
- (74) 대리인: 정부연 (JUNG, Bu Yon); 137-873 서울시 서초구 반포대로 26 길, 29, 원플러스빌딩, 2층, Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO,

AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

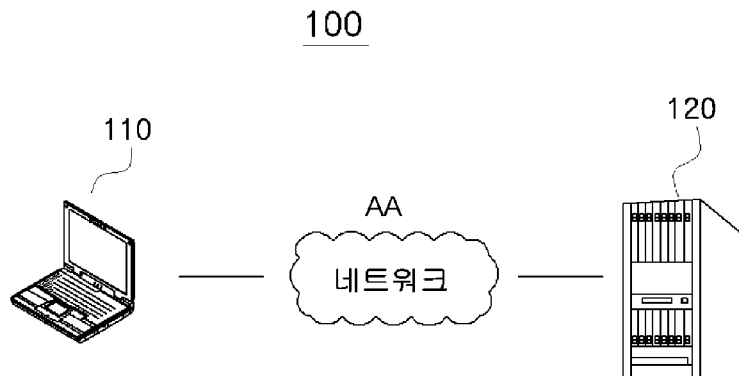
(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제 21 조(3))

(54) Title: NUCLEIC ACID SEQUENCE SECURITY METHOD, DEVICE, AND RECORDING MEDIUM HAVING SAME SAVED THEREIN

(54) 발명의 명칭 : 핵산염기서열 보안 방법, 장치 및 이를 저장한 기록매체



AA ... Network

(57) Abstract: A nucleic acid sequence security method comprises the steps of: (a) forming a plurality of composites each of which comprises a target element derived from a nucleic acid sequence of an analysis requestor and/or a disguising element which is the same as or different from the target element; and (b) providing an analyst with the formed plurality of composites.

(57) 요약서: 핵산염기서열 보안 방법은 (a) 각각이 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체 (target element) 또는 상기 표적 단위체와 동일하거나 또는 다른 위장 단위체 (disguising element) 중 적어도 하나를 포함하는 복수의 복합체들 (composites)을 생성하는 단계 및 (b) 상기 생성된 복수의 복합체들을 분석자에게 제공하는 단계를 포함한다.



WO 2015/199440 A1

명세서

발명의 명칭: 핵산염기서열 보안 방법, 장치 및 이를 저장한 기록매체

기술분야

- [1] 본 발명은 핵산염기서열 보안 기술에 관한 것으로, 보다 상세하게는 분석 요청자의 핵산염기서열을 분석자에게 노출하지 않고 분석 요청자의 핵산염기서열을 분석할 수 있는 핵산염기서열 보안 방법, 장치 및 이를 저장한 기록매체에 관한 것이다.

배경기술

- [2] 게놈 (Genome) 또는 유전체는 한 개체가 갖는 염색체의 총 염기서열을 의미한다. 유전체는 한 생물종의 거의 완전한 유전정보의 총합이고 핵산염기서열 정보를 저장하고 있다. 인간의 유전체는 한 인간 개체를 만들기 위해 필요한 모든 유전자들과 유전자 이외의 부분을 포함하는 약 32억쌍 정도의 모든 핵산염기서열로, 22쌍 44개의 상염색체와 1쌍의 성염색체(X,Y) 및 미토콘드리아에 나뉘어져 있다. 유전체의 핵산은 인산, 당과 함께 아데닌(A), 티아민(T), 구아닌(G) 및 시토신(C) 염기로 이루어진 이중나선형의 물질이다. 유전정보는 핵산의 위 4 가지 염기의 서열의 배열에 의해 부호화된다.
- [3] 사람의 유전자는 각각 수백개 또는 수천개의 아미노산배열을 가지며 몸에서 생산하는 모든 단백질의 설계도를 제공한다. 사람의 유전자는 약 3만~5만개로 추정되며 전체 유전체의 약 3~5%를 차지한다. 약 3만~5만개에 달하는 사람의 유전자는 염기서열로 부호화되어 있고 이중 극히 일 부분만이 그 의미가 밝혀져 있다. 유전체의 핵산염기서열 중 유전자 영역에 속하지 않는 95~97% 부위에 존재하는 핵산염기서열의 유전형에도 개체의 표현형에 관련되는 수많은 부위가 있고 그에 대한 의학적 의미도 지속적으로 밝혀지고 있다. 염기서열로 부호화되어있는 유전체 정보의 의학적, 생물학적 의미의 해석은 지속적으로 발전하고 있다. 따라서, 염기서열을 일차분석한 개인은 새롭게 밝혀진 유전체 염기서열 해석정보에 대한 본인의 해당여부를 확인하기 위해 본인의 염기서열정보에 대한 반복적인 조회 및 해석을 필요로 할 수 있다.
- [4] 염기서열분석기술 발달은 분석 시에 요구되는 비용을 감소시켰고 각 개인의 인간 유전체 분석과 유전체 정보의 활용을 가능하게 하였다. 이에 따라, 많은 사람들은 자신의 유전적 특성인 유전체 염기서열정보를 보유하고 있고 질병 발생률을 포함한 관련된 표현형 위험률은 개인이 보유한 유전체 염기서열 정보를 통해 계산될 수 있다. 따라서, 개인의 유전체 염기서열 정보는 노출이 부담스러운 민감정보로 분류되며 가급적 타인에게 노출되지 않는 것이 좋다. 그러나, 개인이 소유한 자신의 유전형 정보의 해석에 필요한 계산방법 및 그에 필요한 정보자원은 정보해독용 분석장비와 분석방법 등을 보유한 분석자 또는

분석기관이 보유하고 있어서, 개인은 자신의 유전형 정보로부터 자신의 표현형 발생확률을 계산하기 위해 분석자 또는 분석기관에게 자신의 유전형 정보를 전송해야 한다. 이러한 유전형 정보 전송은 개인의 유전형 정보 노출위험을 증가시킨다는 문제가 있다.

- [5] 대한민국공개특허 2013-0075559는 유전자 정보 관리 방법에 관한것으로, 보다 구체적으로, 개인의 유전자 정보를 복수개의 부분서열로 분할하고 저장하여 관리함으로써, 저장 정보에 대한 해킹 혹은 유출 등이 일어난 경우에 그 분할저장된 정보를 분할전의 원 상태로 복원하는 것을 어렵게 하는 방법을 개시한다. 상기의 특허는 민감한 유전자 정보의 저장보관시의 보안에 관한 것으로, 분석대상 유전정보를 보유한 개인 또는 기관이 핵산염기서열 정보를 분석자 또는 분석기관으로 전송하고 그 분석결과를 전송받는 과정에 관여되는 정보 보안 관리 기술에 관한 본 발명과 구별된다. 상기 공개특허는 구분자 정보가 노출되면 개인의 유전자 정보가 노출될 수 있고, 구분자 정보 자체의 보안도 취약할 수 있다는 문제점을 갖는다.

발명의 상세한 설명

기술적 과제

- [6] 본 발명의 일 실시예는 분석 요청자의 핵산염기서열을 분석자에게 노출하지 않고 분석 요청자의 핵산염기서열을 분석할 수 있는 핵산염기서열 보안 방법을 제공하고자 한다.
- [7] 본 발명의 일 실시예는 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체와 위장 단위체를 각각 포함하는 복합체를 표현하고 복수의 표적 단위체들의 관점에서 정의된 희소 행렬을 통해 분석 요청자의 핵산염기서열을 분석할 수 있는 핵산염기서열 보안 방법을 제공하고자 한다.
- [8] 본 발명의 일 실시예는 표적 단위체와 위장 단위체를 각각 포함하는 복수의 복합체들의 개수 또는 복수의 복합체들 각각의 크기를 결정하여 핵산염기서열의 보안강도를 설정할 수 있는 핵산염기서열 보안 방법을 제공하고자 한다.
- [9] 본 발명의 일 실시예는 분석자로부터 복수의 복합체들에 대한 해석결과를 수신하여 분석 요청자의 핵산염기서열의 노출 없이 핵산염기서열에 대한 분석 결과를 획득할 수 있는 핵산염기서열 보안 방법을 제공하고자 한다.
- [10] 본 발명의 일 실시예는 분석자의 관점에서 핵산염기서열에 대한 단위체 단위로 해석결과를 송신함으로써 분석자가 보유한 개별 염기에 대한 우도비(odds ratio)와 같은 고유의 정보나 노하우를 지킬 수 있는 핵산염기서열 보안 방법을 제공하고자 한다.
- [11] 본 발명의 일 실시예는 핵산염기서열 보안 장치의 계산량을 최소화하면서도 보안의 수준을 향상시킬 수 있는 핵산염기서열 보안 방법을 제공하고자 한다.
- [12]

과제 해결 수단

- [13] 실시예들 중에서, 핵산염기서열 보안 방법은 (a) 각각이 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체(target element) 또는 상기 표적 단위체와 동일하거나 또는 다른 위장 단위체(disguising element) 중 적어도 하나를 포함하는 복수의 복합체들(composites)을 생성하는 단계 및(b) 상기 생성된 복수의 복합체들을 분석자에게 제공하는 단계를 포함한다.
- [14] 상기 (a) 단계는 상기 복수의 복합체들을 복수의 표적 단위체들에 대한 최소 행렬로 표현하는 단계를 포함할 수 있다.상기 (a) 단계는 해당 표적 단위체가 해당 복합체에 포함되면 상기 해당 표적 단위체의 위치를 결정하는 단계를 더 포함할 수 있다.
- [15] 일 실시예에서, 상기 (a) 단계는 상기 최소 행렬에 있는 표적 단위체 셀을 상기 표적 단위체에 대한 적어도 하나의 염기와 유전좌위를 포함하는 염기-좌위 집합으로 정의하는 단계를 더 포함할 수 있다. 다른 일 실시예에서, 상기 (a) 단계는 상기 최소 행렬에 있는 표적 단위체 셀을 유전좌위와 연관된 적어도 하나의 염기 집합으로 정의하는 단계를 더 포함할 수 있다. 상기 (a) 단계는 상기 최소 행렬에 있는 표적 단위체 셀의 위치를 동적으로 결정하여 디코딩 과정에서 필요한 표적 단위체 맵을 생성하는 단계를 더 포함할 수 있다.
- [16] 일 실시예에서, 상기 (a) 단계는 상기 핵산염기서열로부터 적어도 하나의 염기와 유전좌위를 추출하여 상기 표적 단위체를 생성하는 단계를 포함할 수 있다. 다른 일 실시예에서, 상기 (a) 단계는 상기 핵산염기서열을 부분 염기서열로 분절하여 상기 표적 단위체를 생성하는 단계를 포함할 수 있다.
- [17] 상기 (a) 단계는 상기 표적 단위체와의 유사도를 기초로 해당 복합체에 있는 적어도 하나의 위장 단위체를 생성하는 단계를 포함할 수 있다. 상기 (a) 단계는상기 표적 단위체와의 유전자 거리 또는 진화적 거리가 특정 거리 이하인 적어도 하나의 위장 단위체를 생성하는 단계를 더 포함할 수 있다. 상기 (a) 단계는상기 분석 요청자에 의하여 설정된 보안 강도에 따라 상기 복수의 복합체들의 개수 또는 상기 복수의 복합체들 각각의 크기를 결정하는 단계를 포함할 수 있다.
- [18] 상기 (b) 단계는 상기 생성된 복수의 복합체들을 분할하여 복수의 직접 또는 간접 분석자들에게 제공하는 단계를 포함할 수 있다.
- [19] 상기 방법은 (c) 상기 분석자로부터 상기 복수의 복합체들에 대한 해석결과를 나타내는 복수의 분석 복합체(analysis composites)들을 수신하여 상기 핵산염기서열에 대한 분석 결과를 획득하는 단계를 더 포함할 수 있다.
- [20] 상기 (c) 단계는 표적 단위체 맵을 기초로 복수의 표적 단위체들에 대한 해석결과를 나타내는 복수의 표적 분석 단위체들(target analysis elements)을 결정하는 단계를 포함할 수 있다. 상기 (c) 단계는 상기 결정된 복수의 표적 분석 단위체들을 조합하여 상기 분석 요청자의 사후오즈(Posterior Odds)를 계산하는

단계를 더 포함할 수 있다.

- [21] 실시예들 중에서, 핵산염기서열 보안 장치는 각각이 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체(target element) 또는 상기 표적 단위체와 동일하거나 또는 다른 위장 단위체(disguising element) 중 적어도 하나를 포함하는 복수의 복합체들(composites)을 생성하는 복합체 생성부 및 상기 생성된 복수의 복합체들을 분석자에게 제공하는 복합체 제공부를 포함한다.
- [22] 상기 장치는 상기 분석자로부터 상기 복수의 복합체들에 대한 해석결과를 나타내는 복수의 분석 복합체들을 수신하여 상기 핵산염기서열에 대한 분석 결과를 획득하는 분석 복합체 해석부를 더 포함할 수 있다.
- [23] 상기 분석 복합체 해석부는 표적 단위체 맵을 기초로 복수의 표적 단위체들에 대한 해석결과를 나타내는 복수의 표적 분석 단위체들(target analysis elements)을 결정할 수 있다. 상기 분석 복합체 해석부는 상기 결정된 복수의 표적 분석 단위체들을 조합하여 상기 분석 요청자의 사후오즈(Posterior Odds)를 계산할 수 있다.
- [24] 실시예들 중에서, 핵산염기서열 보안 방법에 관한 컴퓨터 프로그램을 기록한 기록매체는 각각이 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체(target element) 또는 상기 표적 단위체와 동일하거나 또는 다른 위장 단위체(disguising element) 중 적어도 하나를 포함하는 복수의 복합체들(composites)을 생성하는 기능 및 상기 생성된 복수의 복합체들을 분석자에게 제공하는 기능을 포함한다.
- [25] 상기 기록매체는 상기 분석자로부터 상기 복수의 복합체들에 대한 해석결과를 나타내는 복수의 분석 복합체들을 수신하여 상기 핵산염기서열에 대한 분석 결과를 획득하는 기능을 더 포함할 수 있다.
- [26]

발명의 효과

- [27] 개시된 기술은 다음의 효과를 가질 수 있다. 다만, 특정 실시예가 다음의 효과를 전부 포함하여야 한다거나 다음의 효과만을 포함하여야 한다는 의미는 아니므로, 개시된 기술의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.
- [28] 본 발명의 일 실시예에 따른 핵산염기서열 보안 방법은 분석 요청자의 핵산염기서열을 분석자에게 노출하지 않고 분석 요청자의 핵산염기서열을 분석할 수 있다.
- [29] 본 발명의 일 실시예에 따른 핵산염기서열 보안 방법은 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체와 위장 단위체를 각각 포함하는 복합체를 표현하고 복수의 표적 단위체들의 관점에서 정의된 회소 행렬을 통해 분석 요청자의 핵산염기서열을 분석할 수 있다.

- [30] 본 발명의 일 실시예에 따른 핵산염기서열 보안 방법은 표적 단위체와 위장 단위체를 각각 포함하는 복수의 복합체들의 개수 또는 복수의 복합체들 각각의 크기를 결정하여 핵산염기서열의 보안강도를 설정할 수 있다.
- [31] 본 발명의 일 실시예에 따른 핵산염기서열 보안 방법은 복수의 복합체들에 대한 해석결과를 수신하여 분석 요청자의 핵산염기서열의 노출 없이 핵산염기서열에 대한 분석 결과를 획득할 수 있다.
- [32] 본 발명의 일 실시예에 따른 핵산염기서열 보안 방법은 분석자의 관점에서 핵산염기서열에 대한 단위체 단위로 해석결과를 송신함으로써 분석자가 보유한 개별 염기에 대한 우도비(odds ratio)와 같은 고유의 정보나 노하우를 지킬 수 있다.
- [33] 본 발명의 일 실시예에 따른 핵산염기서열 보안 방법은 핵산염기서열 보안 장치의 계산량을 최소화하면서도 보안의 수준을 향상시킬 수 있다.

[34]

도면의 간단한 설명

- [35] 도 1은 본 발명의 일 실시예에 따른 핵산염기서열 보안 시스템을 설명하는 도면이다.
- [36] 도 2는 도 1에 있는 핵산염기서열 보안장치를 설명하는 블록도이다.
- [37] 도 3은 본 발명의 일 실시예에 따른 복수의 복합체들 생성 및 복수의 분석 복합체들 해석 프로시저를 설명하는 도면이다.
- [38] 도 4은 본 발명의 다른 일 실시예에 따른 복수의 복합체들 생성 및 복수의 분석 복합체들 해석프로시저를 설명하는 도면이다.
- [39] 도 5a는 본 발명의 일 실시예에 따른 표적 단위체에 대한 회소 행렬을 시각화한 도면이고, 도 5b는 본 발명의 다른 일 실시예에 따른 표적 단위체에 대한 회소 행렬을 시각화한 도면이다.
- [40] 도 6은 도 2에 있는 핵산염기서열 보안 장치에 의하여 수행되는 핵산염기서열 보안 방법을 설명하는 흐름도이다.

[41]

발명의 실시를 위한 형태

- [42] 본 발명에 관한 설명은 구조적 내지 기능적 설명을 위한 실시예에 불과하므로, 본 발명의 권리범위는 본문에 설명된 실시예에 의하여 제한되는 것으로 해석되어서는 아니 된다. 즉, 실시예는 다양한 변경이 가능하고 여러 가지 형태를 가질 수 있으므로 본 발명의 권리범위는 기술적 사상을 실현할 수 있는 균등물들을 포함하는 것으로 이해되어야 한다. 또한, 본 발명에서 제시된 목적 또는 효과는 특정 실시예가 이를 전부 포함하여야 한다거나 그러한 효과만을 포함하여야 한다는 의미는 아니므로, 본 발명의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.
- [43] 한편, 본 출원에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.

- [44] "제1", "제2" 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하기 위한 것으로, 이들 용어들에 의해 권리범위가 한정되어서는 아니 된다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다.
- [45] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결될 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다고 언급된 때에는 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 한편, 구성요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.
- [46] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함하다" 또는 "가지다" 등의 용어는 실시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [47] 각 단계들에 있어 식별부호(예를 들어, a, b, c 등)는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 설명하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 단계들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.
- [48] 본 발명은 컴퓨터가 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현될 수 있고, 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있으며, 또한, 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한, 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.
- [49] 염기는 핵염기(nucleobase) 혹은 줄여서 염기(base)로 불리기도 하며, DNA(deoxyribonucleic acid), RNA(ribonucleic acid), 뉴클레오타이드 및 뉴클레오사이드에서 발견되는 질소염기(nitrogenous bases)인 시토신(Cytosine), 구아닌(Guanine), 아데닌(Adenine), 티민(Thymine), 우라실(Uracil)을 포함할 수 있다. 뉴클레오타이드(nucleotide)는 DNA, RNA와 같은 핵산을 구성하는 유기 분자로, 염기-당-인산의 결합으로 이루어진다. 핵산은 뉴클레오타이드가 긴 사슬 모양으로 중합된 고분자 유기물의 한 종류이다.
- [50] 핵산염기서열은 배열을 통해 구현될 수 있고, 배열의 주소(즉, 인덱스)는

유전체 좌표값(이하, 유전좌위)(예, rsID 번호), 배열에 저장된 자료 값은 염기(예, A, G, T, C)로 구성된다. 한편, 여기에서, 염기는 유전자형을 포함할 수 있다. 유전좌위는 연속된 값일 수도 있고, 유전체의 서로 다른부위에서 필요한 부분들만을 골라서 추출한 불연속 좌표값의 목록일 수도 있다. 핵산염기서열은 그 유전좌위가 연속된 좌표값인 경우 배열의 주소 중 한 개만 명시해도 같은 배열내의 나머지 주소는 명확히 알 수 있으므로 배열의 주소 중 한 개만 명시하고 나머지는 생략할 수 있다.

- [51] 개인 유전체 핵산염기서열은 한 개인의 유전체(genome)의 전체 핵산 염기서열을 포함하고, 분석 요청자의 핵산염기서열에 해당할 수 있다.
- [52] 표적 핵산염기서열은 분석 요청자의 핵산염기서열과 연관된 복수의 표적단위체들에 해당하고, 개인 유전체 핵산염기서열에서 분석 대상 부위의 유전좌위-염기 쌍을 추출하고 일정한 순서로 배열하여 생성하거나 또는 개인 유전체 핵산염기서열을 분절하여 생성할 수 있다. 즉, 표적 핵산염기서열은 복수의 표적 단위체들을 일정한 순서로 연결하여 생성될 수 있다. 표적 핵산염기서열은 원래의 개인 유전체 핵산염기서열의 전체이거나, 개인 유전체 핵산염기서열에서 추출해낸 부분 염기서열이거나 또는 개인 유전체 핵산염기서열의 여러 부위에서 추출한 부분 염기서열을 조합한 염기서열일 수 있다.
- [53] 본 발명의 일 실시예에서, 표적 핵산염기서열은 개인별 유전형 정보일 수 있다. 특히, 인간 유전체에 대한 염기서열 분석이 이루어진 뒤, 단순히 유전체 전체 염기서열 해독에 대한 분석뿐만 아니라 인종별, 개인별 다양성을 기반으로 한 단일염기다형성 (single nucleotide polymorphism, SNP) 유전체 염기서열 변이에 대한 분석이 활발하게 진행되는 현실이 고려되었다. 다형성은 유전체상에 존재하는 염기서열의 개인간 차이를 말하는 것으로, 단일염기다형성은 A, T, C 및 G로 이루어진 염기서열 중 하나의 염기에 개인간 차이가 있는 것으로 유전자 다형성 중 그 수가 가장 많다. 인간의 유전자는 약 99.9% 일치하지만 약 0.1%의 단일 염기다형성 차이로 인해 체질, 외모, 질병 등 개인과 인종의 유전적 특성을 나타내게 되며 이로 인해 예컨대 사람마다 동일한 약을 사용해도 약의 효능 및 반응이 다르게 된다. 염기서열의 개인간 차이는 다양한 유전체 변이를 포함하고 이는 염기의 치환, 부가 또는 결실에 의한 것으로, 단일 염기다형성을 포함하는 SNV(Single Nucleotide Variation), STRP (short tandem repeatpolymorphism) 또는 VNTR (various number of tandem repeat) 및 CNV (Copy numbervariation)를 포함하는 다수체 (polyallelic) 변이의 형태로 나타날 수 있다.
- [54] 표적단위체(target element)는 분석 요청자의 핵산염기서열에서 파생된 부분 염기서열에 해당한다. 표적 단위체는 분석 요청자의 핵산염기서열로부터 적어도 하나의 염기와 유전좌위를 추출하여 생성되거나 또는 분석 요청자의 핵산염기서열을 부분 염기서열로 분절하여 생성될 수 있다.
- [55] 위장단위체(disguising element)는 표적 단위체의 식별을 어렵게 하기 위해 표적

단위체와 동일하거나 또는 다른 염기서열을 포함할 수 있다. 위장 단위체는 실제 자연계에 존재하는 핵산염기서열을 활용하거나 또는 실제 자연계에 존재하는 핵산염기서열을 참고하여 생성될 수 있다. 일 실시예에서, 위장 단위체는 표적 단위체와 일치하지 않는 부분 염기서열로 구성되거나, 표적 단위체와 부분적으로 일치하고 부분적으로 불일치하는 부분 염기서열로 구성되거나 또는 표적 단위체와 동일한 부분 염기서열로 구성될 수 있다. 다른 일 실시예에서, 위장 단위체는 분석 요청자의 핵산염기서열을 참고하거나, 무작위적으로 생성된 적어도 하나의 핵산염기서열을 적어도 두 개의 부분 염기서열들로 분절하여 생성되거나 또는 분석 요청자의 핵산염기서열을 분절하여 생성된 표적 단위체의 염기서열을 참고하거나 또는 무작위적으로 생성된 적어도 하나의 핵산염기서열로 구성될 수 있다.

[56] 복합체는 표적 단위체 또는 표적 단위체와 동일하거나 또는 다른 위장 단위체 중 적어도 하나를 포함한다. 즉, 복합체는 표적 단위체와 위장 단위체를 함께 포함하거나, 위장 단위체만을 포함하거나 또는 표적 단위체를 포함할 수 있다.

[57] 분석 복합체는 복합체에 대한 해석결과를 나타낼 수 있다. 예를 들어, 분석 복합체는 복합체에 포함된 단위체(들)에 각각 해당하는 핵산염기서열에 대한 해석결과의 집합을 포함할 수 있다. 다른 예를 들어, 분석 복합체는 복합체에 포함된 단위체(들)에 각각 해당하는 핵산염기서열 및 해당 해석결과를 함께 포함한 염기서열-해석결과 쌍의 집합을 포함할 수 있다. 분석 복합체는 복합체에 포함된 단위체(들)의 핵산염기서열 정보를 표기하지 않거나 생략하고 해당 핵산 염기서열의 식별을 위해 부여된 식별자로 대체한 양태의 해석결과도 포함할 수 있다. 본 발명은 이러한 복합체 및 분석 복합체의 크기(복합체에 포함되는 단위체의 개수) 또는 개수 변화를 통한 높은 수준의 보안을 가능하게 할 수 있다.

[58] 여기서 사용되는 모든 용어들은 다르게 정의되지 않는 한, 본 발명이 속하는 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한 이상적이거나 과도하게 형식적인 의미를 지니는 것으로 해석될 수 없다.

[59]

[60] 도 1은 본 발명의 일 실시예에 따른 핵산염기서열 보안 시스템을 설명하는 도면이다.

[61] 도 1을 참조하면, 핵산염기서열 보안 시스템(100)은 핵산염기서열 보안장치(110) 및 분석 서버(이하, 분석자)(120)를 포함하고, 이들은 네트워크를 통해 연결될 수 있다.

[62] 핵산염기서열 보안 장치(110)는 분석 요청자의 핵산염기서열을 기초로 생성된 복합체(composite)에 관한 분석을 요청할 수 있고, 예를 들어, 데스크탑, 노트북,

태블릿 PC 또는 스마트폰으로 구현될 수 있다. 분석 요청자의 핵산염기서열은 복수의 메모리 영역들(후술될 희소 행렬 메모리 영역 및 표적 단위체 맵 영역)을 통해 관리될 수 있다.

- [63] 분석자(120)는 핵산염기서열 보안 장치(110)와 네트워크를 통해 연결되고 핵산염기서열 보안 장치(110)로부터 복합체(composite)를 수신하고, 수신한 복합체를 분석하여 분석 복합체(analysis composite)를 핵산염기서열 보안 장치(110)에 제공할 수 있다.

[64]

- [65] 도 2는 도 1에 있는 핵산염기서열 보안 장치를 설명하는 블록도이다.

- [66] 도 2를 참조하면, 핵산염기서열 보안 장치(110)는 프로세서(210), 메모리(220), 네트워크 인터페이스(230), 사용자 입력 장치(240) 및 사용자 출력 장치(250) 및 저장장치(260)를 포함한다.

- [67] 프로세서(210)는 복합체 생성부(212), 복합체 제공부(214), 분석 복합체 수신부(216), 분석 복합체 해석부(218) 및 핵산염기서열 보안 제어부(219)를 포함한다.

- [68] 복합체 생성부(212)는 각각이 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체 또는 표적 단위체와 동일하거나 또는 다른 위장 단위체 중 적어도 하나를 포함하는 복수의 복합체들을 생성한다. 복합체 생성부(212)는 생성된 복합체에 표적 단위체가 포함되면 해당 복합체에서의 해당 표적 단위체의 위치를 결정할 수 있다. 예를 들어, 복합체 생성부(212)는 한 개의 표적 단위체를 포함하는 제1 복합체, 네 개의 위장 단위체를 포함하는 제2 복합체 및 한 개의 표적 단위체와 한 개의 위장 단위체를 포함하는 제3 복합체를 생성하고 제1 복합체 및 제3 복합체에 포함된 표적 단위체의 해당 복합체에서의 위치를 결정할 수 있다.

- [69] 복합체 생성부(212)는 복수의 복합체들을 복수의 표적 단위체들에 대한 희소 행렬로 표현할 수 있다. 복합체 생성부(212)는 희소 행렬에 있는 표적 단위체 셀의 위치를 동적으로 결정하여 디코딩 과정(복수의 분석 복합체로부터 표적 단위체에 대한 해석결과를 획득하는 과정)에서 필요한 표적 단위체 맵(224)을 생성할 수 있다.

- [70] 복합체 생성부(212)는 설정된 보안 강도에 따라 복수의 복합체들의 개수 또는 복수의 복합체들 각각의 크기(복합체에 포함된 표적 단위체 또는 위장 단위체의 개수)를 결정할 수 있다. 복수의 복합체들의 개수는 복수의 복합체들로 구성된 행렬의 열의 개수, 복수의 복합체들 각각의 크기는 복수의 복합체들로 구성된 행렬의 행의 개수와 관련된다. 복합체 생성부(212)는 분석 요청자로부터 보안 강도 설정 요청을 수신하여 요청된 보안 강도를 만족시키는 복수의 복합체들을 생성할 수 있다.

- [71] 복합체 생성부(212)는 각각이 표적 단위체 또는 위장 단위체 중 적어도 하나를 포함하는 복수의 복합체들을 생성하여 분석자에게 전송함으로써, 핵산염기서열에 대한 보안 등급 증가율에 비해서 분석자의 핵산염기서열

계산량 증가율은 현저히 낮은 보안 방법을 제공할 수 있다.

- [72] 일반적인 암호화 기술은 정보 K 를 암호화 함수(key) E 로 변환한 $E(K)$ 형태로 전송하고, 수신자는 복호화 함수(key) D 로 복호화 $K = D(E(K))$ 하는 과정이며, 이때 보안키에 해당하는 암호화, 복호화 함수는 각자의 책임으로 보관하는 것이며, 노출되면 암호화 기술은 무력화된다. 보안키는 일종의 정보이므로 그 크기가 보안의 핵심이다. 암호 해독을 위한 가장 단순한 공정은 랜덤 키 발생기(random key generator, 모든 조합을 다 테스트 해보는 기법)이다. 이를 통해, 이론적으로 모든 보안키는 해독 가능하다. 예를 들어 4자리수 보안키는 1만번의 공격으로 해독할 수 있다. 따라서 현대보안 알고리즘은 32비트 보안키와 같은 매우 큰 보안키를 사용함으로써 현존하는 컴퓨터 기술로는 적당한 시간내(예를 들어, 1억년)에 모든 경우의 수를 도저히 테스트해볼 수 없도록 한다. 결국, 해결과제의 크기(Problem Size)를 현실적으로 계산 불가능한 수준으로 크게 만드는 것이다.
- [73] 본 발명의 실시예는 개인의 강력한 식별정보인 유전자 정보를 대상으로 하므로, 데이터를 암호화해서 전송하는 문제에 관한 것이라기 보다는 보안키 자체를 전송하는 문제와 더 유사하다. 본 발명은 보안키를 위장키가 포함될 수 있는 적어도 2개 이상의 단위체로 분산(보안키를 구성하는 요소를 추출하거나, 분절하거나 또는 추출하여 조합하는 방식이 모두 가능)시킴으로써 보안키를 크게하는 효과를 갖는다. 예를 들어, 복합체 생성부(212)는 길이가 100인 핵산염기서열을 기초로 10개의 복합체를 생성하고 각 복합체는 1개의 표적 단위체(길이가 100인 핵산염기서열에서 유래)와 4개의 위장단위체를 포함함으로써 핵산염기서열 보안키를 크게 할 수 있다.
- [74] 예를 들어, 생성된 복합체의 개수가 i , 각 복합체가 포함하는 단위체의 수가 j 인 경우 뒤섞여 버린 것을 모두 조합 추출하는 경우의 수는 j^i 개 만큼 존재하게 되므로, 조합 가능한 j^i 개의 키 중 원본 키 한 개를 찾아내는 문제로 변환되어 해결과제의 크기가 기하급수적으로 증가하게 된다. 예를 들어, $i=10, j=10$ 일 때 문제크기는 10^{10} 즉 100억 가지로 증가하는 것이다. 비교를 위해 현존하는 보안기술인 16 비트보안 (65536), 32비트(43억), 64비트(1.8×10^{19})를 고려하면, 64비트 보안과 유사수준을 얻기 위해서 $i=19, j=10$ 으로 충분하다. 즉 각 복합체가 포함하는 단위체의 개수가 2인 경우, 복합체수는 보안 비트수와 동일해지며, 이때 보안수준은 $1/j^i$ 이 된다.
- [75] 이와 같이 해결과제의 크기가 커지는 경우 분석자 또는 분석기관인 수신측의 계산부하의 증가문제도 함께 해결하는 것이 중요하다. 분석자 또는 분석기관이 원래의 표적염기서열 한 개를 직접 받은경우에 수행하는 해석 또는 분석 연산량을 의미하는 계산복잡도는 해당 알고리즘에 따라 달라지며, 입력한 분석 요청자의 핵산염기서열의 길이에 무관한 상수(K), 해당 길이에 비례하는 선형알고리즘, 해당 길이의 제곱에 비례하는 이차함수(quadratic) 또는 해당 길이의 기하급수증 등에 해당할 수 있다.

- [76] 즉, 본 발명은 알고리즘에 입력되는 분석 요청자의 핵산염기서열의 길이는 변하지 않는 반면, 분석자(120)의 계산부하는 각 복합체에 포함된 단위체 부가 배수만큼 증가하게 되므로 j^i 인 보안등급 증가율에 비해서 분석자(120)의 계산량 증가율($i*j$)은 현저히 낮은 매우 유리한 방식을 제공한다. 예를 들어, 핵산염기서열 보안 장치(110)는 분석 요청자의 핵산염기서열을 기초로 10개의 표적 단위체를 생성하고 표적 단위체 마다 4개의 위장 단위체를 부가하여 10개의 복합체를 생성한 경우, 5^{10} 배 증가된 보안 등급을 가지는데 이때, 분석자(120)의 계산량은 $5*10$ 배 증가하고 이는 보안 등급 증가량에 비해 현저히 낮은 수치이다.
- [77] 복합체 제공부(214)는 복합체 생성부(212)가 생성한 복수의 복합체들을 분석자(120)에게 제공한다. 복합체 제공부(214)는 생성된 복수의 복합체들 전부를 한 분석자에게 제공하거나 생성된 복수의 복합체들을 분할하여 복수의 직접 또는 간접 분석자들에게 제공할 수 있다. 예를 들어, 복합체 제공부(214)는 6개의 복합체를 생성하여 3개의 복합체는 분석자 A에게, 복합체는 분석자 B에게 제공할 수 있다.
- [78] 분석 복합체 수신부(216)는 분석자(120)로부터 복수의 복합체들에 대한 해석결과를 나타내는 복수의 분석 복합체들(analysis composites)을 수신한다. 예를 들어, 복합체 수신부(216)는 제1 복합체에 대한 해석결과를 나타내는 제1 분석 복합체, 제2 복합체에 대한 해석결과를 나타내는 제2 분석 복합체 및 제3 복합체에 대한 해석결과를 나타내는 제3 분석 복합체를 수신할 수 있다.
- [79] 분석 복합체 해석부(218)는 수신된 복수의 분석 복합체들을 통해 핵산염기서열에 대한 분석 결과를 획득한다. 분석 복합체 해석부(218)는 복수의 분석 복합체들에서 복수의 표적 단위체들의 해석결과에 해당하는 복수의 표적 해석 단위체들(target analysis elements)를 결정하고 이를 이용하여 분석 요청자의 핵산염기서열에 대한 최종 분석결과를 도출할 수 있다.
- [80] 핵산염기서열 보안 제어부(219)는 핵산염기서열 보안 장치(110)의 전체적인 동작을 제어하고, 복합체 생성부(212), 복합체 제공부(214), 분석 복합체 수신부(216) 및 분석 복합체 해석부(218) 간의 데이터 흐름을 제어할 수 있다.
- [81] 메모리(220)는 희소 행렬 메모리 영역(Sparse Matrix Memory Area, S.M.M.A) 및 표적 단위체 맵 영역(Target Element Map, T.E.M.A)을 포함한다.
- [82] 희소 행렬 메모리 영역(S.M.M.A)는 복수의 복합체들(composites, 223) 및 복합체(223)에 포함된 복수의 표적 단위체들에 대한 희소 행렬(222)을 저장하는 공간에 해당한다.
- [83] 복수의 복합체들(223) 각각은 표적 단위체(T.E) 또는 위장 단위체(D.E) 중 적어도 하나를 포함할 수 있다. 예를 들어, 복합체 1, 2, 3 및 5(223a, 223b, 223c 및 223e)는 하나의 표적 단위체와 네 개의 위장 단위체를, 복합체 4(223d)는 다섯 개의 위장 단위체를 포함한다.
- [84] 도 2에서, 복수의 복합체들(223)은 4개의 표적 단위체와 21개의 위장 단위체를

포함하고, 복수의 복합체들(223)에 포함된 복수의 표적 단위체들과 위장 단위체들은 5행 5열 행렬을 구성할 수 있다(복수의 복합체들(223) 각각은 구성된 행렬의 열에 위치). 따라서, 복수의 복합체들(223)은 복수의 표적 단위체들에 대한 희소 행렬(222)로 표현될 수 있다.

- [85] 도 2에서, 복수의 표적 단위체들에 대한 희소 행렬(222)은 2행 1열에 복합체 1(223a)의 표적 단위체, 4행 2열에 복합체 2(223b)의 표적 단위체, 1행 3열에 복합체 3(223c)의 표적 단위체 및 5행 5열에 복합체 5(223e)의 표적 단위체를 행렬의 셀로 포함한다.
- [86] 표적 단위체 맵 영역(T.E.M.A)은 희소 행렬 메모리 영역(S.M.M.A)에 저장된 희소 행렬(222)에 대한 표적 단위체 맵(224)을 저장하는 공간에 해당한다.
- [87] 표적 단위체 맵(224)은 희소 행렬(222)에 포함된 복수의 표적 단위체들의 위치에 대한 맵에 해당한다. 예를 들어, 표적 단위체 맵(224)은 희소 행렬 메모리 영역(S.M.M.A)에 저장된 희소 행렬(222)의 열의 개수가 5개에 해당하면 제1 열에 있는 표적 단위체부터 제5 열에 있는 표적 단위체까지의 위치를 포함한다. 표적 단위체 맵(224)은 희소 행렬(222)의 해당 열이 표적 단위체를 포함하고 있지 않는 경우 해당 열에 대한 표적 단위체의 위치를 마이너스 값(예를 들어 -1)으로 저장하여 표적 단위체의 유무를 구분할 수 있다.
- [88] 네트워크 인터페이스(230)는 네트워크를 통해 분석자(120)와 연결하기 위한 환경을 포함하고, 예를 들어, LAN(Local Area Network) 통신을 위한 어댑터를 포함할 수 있다.
- [89] 사용자 입력 장치(240)는 사용자 입력을 수신하기 위한 환경을 포함하고, 예를 들어, 마우스, 트랙볼, 터치 패드, 그래픽 태블릿, 스캐너, 터치 스크린, 키보드 또는 포인팅 장치와 같은 어댑터를 포함할 수 있다.
- [90] 사용자 출력 장치(250)는 사용자에게 특정 정보(예를 들어, 분석 요청자의 핵산염기서열 분석결과)를 출력하기 위한 환경을 포함하고, 예를 들어, 모니터 또는 터치스크린과 같은 어댑터를 포함할 수 있다. 일 실시예에서, 사용자 입력 장치(240)와 사용자 출력 장치(250)는 원격 접속을 통해 접속될 수 있다.
- [91] 저장장치(260)는 SSD(Solid State Disk) 또는 HDD(Hard Disk Drive)와 같은 비휘발성 메모리로 구현될 수 있고, 핵산염기서열 보안 장치(110)에 필요한 데이터를 저장하는데 사용된다.
- [92]
- [93] 도 3은 본 발명의 일 실시예에 따른 복수의 복합체들 생성 및 복수의 분석 복합체들 해석프로시저를 설명하는 도면이다.
- [94] 도 3에서, 각 염기는 유전좌위(유전좌위는 염기의 고유한 주소/좌표로 염색체상의 위치, 일종의 좌표값, 12번 염색체의 1234501번째 위치 또는 rsID 등의 값으로 표기될 수 있다)를 가지고 있으며, 이처럼 유전좌위 + 염기 + 기타 정보를 담는 파일 포맷은 표준화되어 사용되는 많은 파일이 있다. 예를 들면 VCF (Variant Call Format)은 전체 서열이 아니라, 추출된 개인별 변이 부위만 좌표와

검출된 유전형 + 기타의 정보로 표현하는 파일 포맷이다. 일 실시예에서, 분석 요청자의 핵산염기서열 및 상기 표적 단위체, 위장 단위체, 복합체 또는 분석 복합체에 포함된 핵산염기서열의 유전좌위 + 염기 + 기타정보를 담는 표준화된 포맷은 VCF(Variant CallFormat), BCF(Binary version of VCF), GFF(Gene-Finding Format, GenericFeature Format, 현재버전은 4.1), GTF(Gene Transfer Format), GVF(GenomeVariation Format), SAM/BAM(Sequence Alignment Map/ Binary version of SAM), QUAL, SCARF, QSEQ, IG, Maq, SOAP, bcf, pileup, mpileup, CASAVA, MaCH, GLFv2, GPFv2, axt, BED, BED15, BED detail, BEDPE, bedGraph, bigBed, bigWig, Chain, GenePred table, HAL, HDF5, MAF, Net, Personal Genome SNP, PSL, WIG (Wiggleformat), .2bit, .nib, CSFASTQ, CSFASTA, FASTA, FASTQ 형식 또는 그 확장형식들을 포함할 수 있다. 이때 표적 핵산염기서열은 SNP(single nucleotide polymorphism) 등 선택된 특정 염기(마커)만 선별해서 보낼 수도 있고 연속된 전체서열을 보낼 수도 있다. 유전자서열은 염기가 일정순서로 나열된 서열 자체이고, 또한 본 발명은 전송된 표적 핵산염기서열의 보안에 관한 것이기 때문에, 그 전송서열이 원 서열이든, 일부만 선택된 서열이든 모두 가능한 것이다. 일 실시예에서, 표적 핵산염기서열은 특정 유전자 또는 비유전자 좌위의 SNP(single nucleotide polymorphism), 염기의 치환, 부가 또는 결실을 포함하는 이수체 돌연변이 또는 STRP (short tandem repeat polymorphism) 또는 VNTR (various number of tandem repeat)를 포함하는 다수체 (polyallelic) 변이에서 결정된 염기일 수 있다.

- [95] 복합체 생성부(212)는 분석 요청자의 핵산염기서열로부터 적어도 하나의 염기와 유전좌위를 추출하여 표적 단위체를 생성한다. 복합체 생성부(212)는 복수의 복합체들을 생성하기 위하여 분석 요청자의 핵산염기서열에서 적어도 하나의 염기와 유전좌위를 추출하는 방식으로 GAAT의 핵산염기서열을 가지는 표적 단위체 T.E1, ACGCA의 핵산염기서열을 가지는 표적 단위체 T.E2, TCCTGAT의 핵산염기서열을 가지는 표적 단위체 T.E3, GACAC의 핵산염기서열을 가지는 표적 단위체 T.E4, CCAGCG의 핵산염기서열을 가지는 표적 단위체 T.E5를 생성할 수 있다.
- [96] 다음으로, 복합체 생성부(212)는 복수의 복합체들을 구성하는 위장 단위체를 생성한다. 일 실시예에서, 복합체 생성부(212)는 생성된 표적 단위체와의 유사도를 기초로 해당 복합체에 있는 적어도 하나의 위장 단위체를 생성할 수 있다. 다른 일 실시예에서, 복합체 생성부(212)는 생성된 표적 단위체와의 유전자 거리 또는 진화적 거리가 특정 거리 이하인 적어도 하나의 위장 단위체를 생성할 수 있다.
- [97] 최종적으로, 복합체 생성부(212)는 표적 단위체 T.E1을 포함하는 복합체 C1, 표적 단위체 T.E2를 포함하는 복합체 C2, 표적 단위체 T.E3를 포함하는 복합체 C3, 표적 단위체 T.E4를 포함하는 복합체 C4, 표적 단위체, 표적 단위체 T.E5를 포함하는 복합체 C5 및 표적 단위체를 포함하지 않고 복수의 위장 단위체들만으로

구성된 복합체 C6를 생성할 수 있다. 복합체 생성부(212)는 복수의 복합체들(C1~C6)을 생성하면서 복합체에 포함된 표적 단위체의 위치를 결정할 수 있다. 도 3에서, 복합체 생성부(212)는 표적 단위체 T.E1을 복합체 C1의 3번 위치에, 표적 단위체 T.E2을 복합체 C2의 1번 위치에, 표적 단위체 T.E3를 복합체 C3의 4번 위치에, 표적 단위체 T.E4를 복합체 C4의 1번 위치에 마지막으로 표적 단위체 T.E5를 복합체 C5의 3번 위치에 배치할 수 있다.

- [98] 생성된 복수의 복합체들(C1~C6)은 분석자에게 제공되고 분석자(120)는 제공받은 복수의 복합체들(C1~C6)을 분석하여 복수의 복합체들(C1~C6)에 대한 해석결과를 나타내는 복수의 분석 복합체들(analysis composite 1~analysis composite 6, A.C1~A.C6)을 분석 복합체 수신부(216)에 제공한다. 즉, 분석자(120)는 분절된 핵산염기서열들에 대한 제공 가능한 분석결과값을 생성한다. 예를 들어, 복합체 생성부(212)가 10개의 복합체를 생성하고, 각 복합체는 1개의 표적 단위체와 4개의 위장단위체를 포함하고 있는 경우, 분석자(120)는 50개의 분석결과값을 생성하고 생성된 분석결과값을 포함하는 복수의 분석 복합체들을 핵산염기서열 보안 장치(110)로 전송하게 된다. 핵산염기서열 보안 장치(110)는 50개의 분석결과값을 수신하여 표적 단위체 염기서열 또는 표적 단위체 생성규칙(예를 들어, 표적 단위체 맵(224)과 같은 표적 단위체 위치 정보)을 참조하여 각 분석 복합체에서 표적 단위체의 해석 또는 분석의 결과값 10개(10개의 표적 분석 단위체)를 추출하고, 이들을 조합하여 분석 요청자의 핵산염기서열에 대한 해석 또는 분석의 최종결과물과 동일한 결과값을 획득할 수 있다. 이때, 분석 복합체에서 추출한 표적 단위체의 결과값 10개를 조합하기 위한 함수는, 함수를 미리 약속해두거나, 분석기관이 생성/전송하거나 클라이언트(예를 들어, 분석 요청자)가 요청할 수 있다. 여기에서, 함수는 10개의 수를 곱하거나, 합하거나 또는 평균치를 계산하는 등의 다양한 계산 방법을 포함한다

- [99] 도 3에서, 복수의 분석 복합체들(A.C1~A.C6)은 복수의 복합체들(C1~C6)에 포함된 복수의 표적 단위체들(T.E1~T.E5)에 각각 대응하는 복수의 표적 해석 단위체(R13, R21, R34, R41 및 R53)을 포함한다. 핵산염기서열 보안 장치(110)는 분석을 요청한 표적 단위체와 위장 단위체의 핵산염기서열 정보를 보유하고 있으므로, 각 복합체별 해당 표적단위체의 핵산염기서열 정보(도3에서, T.E1은 GAAT, T.E2는 ACGCA, T.E3는 TCCTGAT, T.E4 GACAC, T.E5는 CCAGCG에 해당)와 일치하는 서열과 논리적으로 연결된 분석결과 (여기서는 R13, R21, R34, R41, R53)를 용이하게 추출할 수 있다. 이때 분석자(120)는 복수의 분석 복합체들에 포함된 각 단위체에 표적 단위체 및 위장 단위체의 핵산염기서열정보를 반드시 다 포함시켜 전송할 필요는 없이 부호화 등을 통해 정보량을 줄여(논리적으로는 여전히 식별가능하게 줄이거나 부호화) 전송하는 것이 가능하다.

- [100] 일 실시예에서, 복수의 표적 해석 단위체(R13, R21, R34, R41 및 R53)는 복수의

표적 단위체들을 구성하는 핵산염기서열과 해당 핵산염기서열에 대한 분석결과값을 연결하는 염기서열-결과값 자료쌍의 집합을 의미할 수 있다. 다른 일 실시예에서, 복수의 표적 해석 단위체(R13, R21, R34, R41 및 R53)는 복수의 표적 단위체들을 구성하는 핵산염기서열 정보가 표기되지 않거나 생략되고 해당 핵산염기서열의 식별을 위해 부여된 식별자로 대체된 식별자-결과값 자료쌍의 집합을 의미할 수 있다.

- [101] 분석 복합체 해석부(218)는 표적 단위체 맵(224)을 기초로 분석자로부터 수신한 복수의 분석 복합체들(A.C1~A.C6)에서 복수의 표적 단위체들(T.E1~T.E5)에 대응하는 복수의 표적 해석 단위체(R13, R21, R34, R41 및 R53)를 결정할 수 있다.
- [102] 분석 복합체 해석부(218)는 결정된 복수의 표적 분석 단위체들((R13, R21, R34, R41 및 R53))을 특정 함수 $f(R_k)$ 를 이용하여 연산하고 분석 요청자의 핵산염기서열에 대한 최종 분석 결과를 도출할 수 있다. 예를 들어 분석 복합체 해석부(218)는 결정된 복수의 표적 분석 단위체들((R13, R21, R34, R41 및 R53))을 조합하여 분석 요청자의 사후오즈(Posterior Odds)를 계산할 수 있고 계산된 분석 요청자의 사후오즈는 분석 요청자의 핵산염기서열에 대한 최종 분석 결과를 제공할 수 있다. 이하, 분석 복합체 해석부(218)에서 수행되는 분석 요청자의 사후오즈 계산과정을 상세히 설명한다.
- [103] 분석 복합체 해석부(218)는 베이즈 이론(Bayes Theorem)에 따라 분석 요청자의 핵산염기서열의 사후오즈(Posterior Odds, 해당 핵산염기서열을 가진 개인의 질병위험률)를 계산할 수 있다. 사후오즈(Posterior Odds)는 베이즈 이론(Bayes Theorem)에 따르면 하기의 수학적 식 1과 같이 사전오즈(Prior Odds)와 우도비(Likelihood Ratio)의 곱으로 표현될 수 있다.
- [104] [수학적 식 1]
- [105] $\text{Posterior Odds} = \text{Prior Odds} \times \text{Likelihood Ratio}$
- [106]
$$\frac{p(D | S)}{p(\bar{D} | S)} = \frac{p(D)}{p(\bar{D})} \cdot \frac{p(S | D)}{p(S | \bar{D})}$$
- [107] 이때, 사전오즈는 사전지식이 없는 경우 중립 값인 1로 치환될 수 있다. 따라서, 사후오즈는 사전오즈 값 1에 우도비를 계속 곱해 나가는 연쇄계산(Chain calculation)을 통해 얻어질 수 있다. 여기에서, 유전형에 따른 우도비는 하기의 수학적 식 2와 같이 계산될 수 있다.
- [108] [수학적 식 2]
- [109] $\text{우도비} = (\text{증례군의 유전형 확률}) / (\text{대조군 유전형 확률})$

[110]

$$LR = \frac{\text{probability of the genotype in the case population}}{\text{probability of the genotype in the control population}}$$

	AA	Aa	aa
Case	a	b	c
Control	d	e	f
LR	$\frac{a}{(a+b+c)} \cdot \frac{d}{(d+e+f)}$	$\frac{b}{(a+b+c)} \cdot \frac{e}{(d+e+f)}$	$\frac{c}{(a+b+c)} \cdot \frac{f}{(d+e+f)}$

[111] 예를들어, n 개의 유전좌위 $a_1, a_2, \dots, a_n$ 에서 획득한 염기 $G(a_1), G(a_2), \dots, G(a_n)$ 으로 구성된 핵산염기서열 $[a_1, a_2, \dots, a_n]$ 의 각 위치에 대한 우도비 $LR(G(a_1)), LR(G(a_2)), \dots, LR(G(a_n))$ 가 주어진 경우, 해당 핵산염기서열을 가진 개인의 질병위험률에 해당하는 사후오즈는 하기의 수학식 3과 같이 계산될 수 있다.

[112] [수학식 3]

[113]

$$Posterior Odds = \prod_{i=1}^n G(a_i)$$

[114] 분석 복합체 해석부(218)는 수학식 3을 이용하여 분석 요청자의 핵산염기서열의 사후오즈를 계산할 수 있다. 예를 들어, 핵산염기서열 보안 장치(110)가 분석 요청자의 핵산염기서열 $[a_1, a_2, \dots, a_n]$ 로부터 k개의 표적단위체를 결정한 경우, 곱셈의 교환법칙($a \times b = b \times a$)과, 곱셈의 결합법칙($(a \times b) \times c = a \times (b \times c)$)에 의해, 분석 복합체 해석부(218)는 하기의 수학식 4와 같이 k개의 표적 단위체의 사후오즈(R_k)를 곱함(프로덕트 연산수행)으로써 분석 요청자의 핵산염기서열에 대한 최종 사후오즈를 계산할 수 있다.

[115] [수학식 4]

[116]

$$Posterior Odds = \prod_{k=1}^k (R_k)$$

[117] 본 발명에서, 수학식 4의 R_k 는 복수의 분석 복합체들에 포함된 표적 단위체에 대한 분석결과 염기서열의 해석결과 값에 해당할 수 있다.

[118] 수학식 4는 변수가 k 및 R_k 뿐이므로, 분석 요청자와 분석자(120) 사이에서 사전에 미리 정해질 수 있고, 분석자(120)에 의해 복수의 분석 복합체들과 함께 전송될 수도 있으며, 분석 요청자에 의해 적어도 하나의 표적 단위체 각각의 해석결과값($R_1 \sim R_k$)과 함께 연산에 필요한 해당 수식이 요청될 수도 있다.

[119] 또한, 수학식 4는 표적 단위체의 개수나, 각 표적 단위체의 핵산염기서열의

길이에 무관하게 성립한다. 그러므로 핵산염기서열 보안 장치(110)는 표적 단위체의 개수를 n 보다 작거나 같은 모든 수로 결정하는 것이 가능하다. 그러나 분석 요청자의 핵산염기서열의 위치를 하나씩 분절하는 것은 분석을 수행하는 분석자(120)의 관점에서 핵심적인 지식재산인 각 위치에 대한 우도 $LR(G(a_1))$, $LR(G(a_2)), \dots, LR(G(a_n))$ 를 모두 노출하는 것이 되므로 적절하지 않다.

핵산염기서열 보안 장치(110)는 분석자(120)의 데이터 처리에 사용되는 알고리즘등과 같은 지식재산보호를 위해서 최소 2 이상의 길이를 갖는 표적 단위체를 생성할 필요가 있다. 표적 단위체의 길이는 보안강도 설정의 목표에 따라 변동될 수 있다. 보안강도는 표적 단위체 길이에 대해 지수적으로 증가하므로 손쉽게 고도의 보안등급을 설정할 수 있다. 반면 표적 단위체의 길이가 너무 길어져 분석 요청자의 핵산염기서열의 길이에 근접하게 되면 상대적으로 송신자의 표적 핵산염기서열 노출확률이 증가하게 되어, 이를 방어하기 위해서는 복합체에 포함된 위장 단위체의 수를 늘려야 한다. 그러나 위장 단위체 수와 보안강도는 선형적으로 밖에는 증가하지 않으므로, 그 수를 매우 크게 증가시켜야 하고, 이는 분석자(120)의 분석부하를 기하급수적으로 증가시켜 바람직한 방법으로 보기 어렵다. 그러므로 본 발명은 표적 핵산염기서열을 보호하기 위한 복합체의 개수, 복수의 복합체들에 포함된 표적 단위체의 개수 또는 각 복합체에 포함된 위장 단위체의 개수를 조절하여 분석 요청자와 분석자(120) 각각의 프라이버시와 지식자산 보안등급 요구를 충족할 수 있다. 즉, 본 발명은 표적 단위체의 개수, 위장 단위체의 개수 및 복수의 복합체들의 개수를 조절하여 분석 요청자의 핵산염기서열 보안 및 분석자(120)의 지식자산 보안 요구를 동시에 충족시킬 수 있다.

[120] 예를 들어, 도 3은 복합체 생성부(212)가 6개의 복합체를 생성한 경우로, 복합체 C1부터 복합체 C5까지는 하나의 표적 단위체와 약 4개의 위장 단위체를 포함하고, C6는 4개의 위장 단위체만을 포함하고 있다. 이때 표적 단위체를 포함하지 않는 복합체 C6를 포함하는 6개의 복합체의 표적 핵산염기서열의 정보보호 등급은 $4^6 = 2^{12} = 1/4096$ 으로 12비트 보안수준에 해당한다. 보안수준 목표를 매우 강력한 32비트로 정한 경우, $2^{32} = 4^{16} = 16^8 = 256^4 = 65546^2$ 이므로 각 2개의 단위체를 포함하는 32개의 복합체, 각 4개의 단위체를 포함하는 16개의 복합체, 각 16개의 단위체를 포함하는 8개의 복합체, 각 256개의 단위체를 포함하는 4개의 복합체 또는 각 65546개의 단위체를 포함하는 2개의 복합체를 생성하여 32비트의 보안수준이 만족될 수 있다.

[121] 그러므로 본 발명은 먼저 정보보호등급을 정한 후, 분석자(120)의 계산부하를 고려한 복합체의 개수 및 복합체에 포함되는 단위체의 개수를 결정할 수 있다.

[122]

[123] 도 4는 본 발명의 다른 일 실시예에 따른 복수의 복합체들 생성 및 복수의 분석 복합체들 해석프로시저를 설명하는 도면이다.

[124] 도 4에서, 복합체 생성부(212)는 분석 요청자의 핵산염기서열을 부분

염기서열로 분절하여 상기 표적 단위체를 생성한다. 복합체 생성부(212)는 복수의 복합체들을 생성하기 위하여 분석 요청자의 핵산염기서열을 부분 염기서열로 분절하는 방식으로 GGAA의 핵산염기서열을 가지는 표적 단위체 T.E1, TCAAC의 핵산염기서열을 가지는 표적 단위체 T.E2, CGGCGGA의 핵산염기서열을 가지는 표적 단위체 T.E3, CTGAT의 핵산염기서열을 가지는 표적 단위체 T.E4, TACACCC의 핵산염기서열을 가지는 표적 단위체 T.E5를 생성할 수 있다. 이후에 핵산염기서열 보안 장치(110)에서 수행되는 프로시저는 도 3과 같다.

[125]

[126] 도 5a는 본 발명의 일 실시예에 따른 표적 단위체에 대한 회소 행렬을 시각화한 도면이고, 도 5b는 본 발명의 다른 일 실시예에 따른 표적 단위체에 대한 회소 행렬을 시각화한 도면이다.

[127] 도 5a에서, 복합체 생성부(212)는 회소 행렬에 있는 표적 단위체 셀을 표적 단위체에 대한 적어도 하나의 염기와 유전좌위를 포함하는 염기-좌위 집합으로 정의할 수 있다. 예를 들어, 회소 행렬에 있는 표적 단위체 셀은 해당 표적 단위체가 유전좌위가 12인 염기 A, 유전좌위가 15인 염기 G 및 유전좌위가 332인 염기 G로 구성된 핵산염기서열에 해당하면 {(A,12),(G,15),(G,332)}의 집합으로 정의될 수 있다.

[128] 도 5b에서, 복합체 생성부(212)는 회소 행렬에 있는 표적 단위체 셀을 유전좌위와 연관된 적어도 하나의 염기 집합으로 정의할 수 있다. 예를 들어, 회소 행렬에 있는 표적 단위체 셀은 해당 표적 단위체가 해당 표적 단위체가 유전좌위가 12인 염기 A, 유전좌위가 15인 염기 G 및 유전좌위가 332인 염기 G로 구성된 핵산염기서열에 해당하면 {A,G,G}의 집합으로 정의될 수 있다. 이때, 해당 염기들의 유전좌위 12,15 및 332는 행렬에서 공통으로 표현될 수 있다.

[129]

[130] 도 6은 도 2에 있는 핵산염기서열 보안 장치에 의하여 수행되는 핵산염기서열 보안 방법을 설명하는 흐름도이다.

[131] 복합체 생성부(212)는 각각이 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체 또는 표적 단위체와 동일하거나 또는 다른 위장 단위체 중 적어도 하나를 포함하는 복수의 복합체들을 생성한다(S601). 이때, 복합체 생성부(212)는 복수의 복합체들을 복수의 표적 단위체들에 대한 회소 행렬로 표현할 수 있다(S602).

[132] 복합체 생성부(212)는 표적 단위체가 해당 복합체에 포함되면 해당 표적 단위체의 위치를 결정할 수 있다(S603). 복합체 생성부(212)는 복수의 복합체들을 복수의 표적 단위체들에 대한 회소 행렬로 표현하면서, 회소 행렬에 있는 표적 단위체 셀의 위치를 동적으로 결정하여 디코딩 과정에서 필요한 표적 단위체 맵을 생성할 수 있다(S604)

[133] 복합체 제공부(214)는 생성된 복수의 복합체들을 분석자에게 제공하고 해석

복합체 수신부(216)은 분석자로부터 복수의 복합체들에 대한 해석결과를 나타내는 복수의 분석 복합체들을 수신할 수 있다(S605 및 S606).

- [134] 해석 복합체 해석부(218)는 표적 단위체 맵에 저장된 표적 단위체의 위치 정보를 기초로 복수의 표적 단위체들에 대한 해석결과를 나타내는 복수의 표적 분석 단위체들을 결정할 수 있다(S607). 해석 복합체 해석부(218)는 결정된 복수의 표적 분석 단위체들을 조합하여 분석 요청자의 사후오즈를 계산하고, 계산된 사후오즈를 통해 분석 요청자의 핵산염기서열의 최종 분석결과를 획득할 수 있다(S608 및 S609).

[135]

- [136] 상기에서는 본 출원의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 출원을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

[137]

산업상 이용가능성

- [138] 본 발명은 핵산염기서열 보안 기술에 관한 것으로, 보다 상세하게는 분석 요청자의 핵산염기서열을 분석자에게 노출하지 않고 분석 요청자의 핵산염기서열을 분석할 수 있는 핵산염기서열 보안 방법, 장치 및 이를 저장한 기록매체에 관한 것이다.

청구범위

- [청구항 1] (a) 각각이 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체(target element) 또는 상기 표적 단위체와 동일하거나 또는 다른 위장 단위체(disguising element) 중 적어도 하나를 포함하는 복수의 복합체들(composites)을 생성하는 단계; 및
(b) 상기 생성된 복수의 복합체들을 분석자에게 제공하는 단계를 포함하는 핵산염기서열 보안 방법.
- [청구항 2] 제1항에 있어서, 상기 (a) 단계는
상기 복수의 복합체들을 복수의 표적 단위체들에 대한 최소 행렬로 표현하는 단계를 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 3] 제2항에 있어서, 상기 (a) 단계는
해당 표적 단위체가 해당 복합체에 포함되면 상기 해당 표적 단위체의 위치를 결정하는 단계를 더 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 4] 제2항에 있어서, 상기 (a) 단계는
상기 최소 행렬에 있는 표적 단위체 셀을 상기 표적 단위체에 대한 적어도 하나의 염기와 유전좌위를 포함하는 염기-좌위 집합으로 정의하는 단계를 더 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 5] 제2항에 있어서, 상기 (a) 단계는
상기 최소 행렬에 있는 표적 단위체 셀을 유전좌위와 연관된 적어도 하나의 염기 집합으로 정의하는 단계를 더 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 6] 제2항에 있어서, 상기 (a) 단계는
상기 최소 행렬에 있는 표적 단위체 셀의 위치를 동적으로 결정하여 디코딩 과정에서 필요한 표적 단위체 맵을 생성하는 단계를 더 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 7] 제1항에 있어서, 상기 (a) 단계는
상기 핵산염기서열로부터 적어도 하나의 염기와 유전좌위를 추출하여 상기 표적 단위체를 생성하는 단계를 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 8] 제1항에 있어서, 상기 (a) 단계는
상기 핵산염기서열을 부분 염기서열로 분절하여 상기 표적 단위체를 생성하는 단계를 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 9] 제1항에 있어서, 상기 (a) 단계는
상기 표적 단위체와의 유사도를 기초로 해당 복합체에 있는 적어도 하나의 위장 단위체를 생성하는 단계를 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.

- [청구항 10] 제9항에 있어서, 상기 (a) 단계는
상기 표적 단위체와의 유전자 거리 또는 진화적 거리가 특정 거리 이하인 적어도 하나의 위장 단위체를 생성하는 단계를 더 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 11] 제1항에 있어서, 상기 (a) 단계는
상기 분석 요청자에 의하여 설정된 보안 강도에 따라 상기 복수의 복합체들의 개수 또는 상기 복수의 복합체들 각각의 크기를 결정하는 단계를 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 12] 제1항에 있어서,
(c) 상기 분석자로부터 상기 복수의 복합체들에 대한 해석결과를 나타내는 복수의 분석 복합체들(analysis composites)을 수신하여 상기 핵산염기서열에 대한 분석 결과를 획득하는 단계를 더 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 13] 제12항에 있어서, 상기 (c) 단계는
표적 단위체 맵을 기초로 복수의 표적 단위체들에 대한 해석결과를 나타내는 복수의 표적 분석 단위체들(target analysis elements)을 결정하는 단계를 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 14] 제13항에 있어서, 상기 (c) 단계는
상기 결정된 복수의 표적 분석 단위체들을 조합하여 상기 분석 요청자의 사후오즈(Posterior Odds)를 계산하는 단계를 더 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 15] 제1항에 있어서, 상기 (b) 단계는
상기 생성된 복수의 복합체들을 분할하여 복수의 직접 또는 간접 분석자들에게 제공하는 단계를 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법.
- [청구항 16] 각각이 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체(target element) 또는 상기 표적 단위체와 동일하거나 또는 다른 위장 단위체(disguising element) 중 적어도 하나를 포함하는 복수의 복합체들(composites)을 생성하는 복합체 생성부; 및
상기 생성된 복수의 복합체들을 분석자에게 제공하는 복합체 제공부를 포함하는 핵산염기서열 보안 장치.
- [청구항 17] 제16항에 있어서,
상기 분석자로부터 상기 복수의 복합체들에 대한 해석결과를 나타내는 복수의 분석 복합체들을 수신하여 상기 핵산염기서열에 대한 분석 결과를 획득하는 분석 복합체 해석부를 더 포함하는 것을 특징으로 하는 핵산염기서열 보안 장치.
- [청구항 18] 제17항에 있어서, 상기 분석 복합체 해석부는
표적 단위체 맵을 기초로 복수의 표적 단위체들에 대한 해석결과를

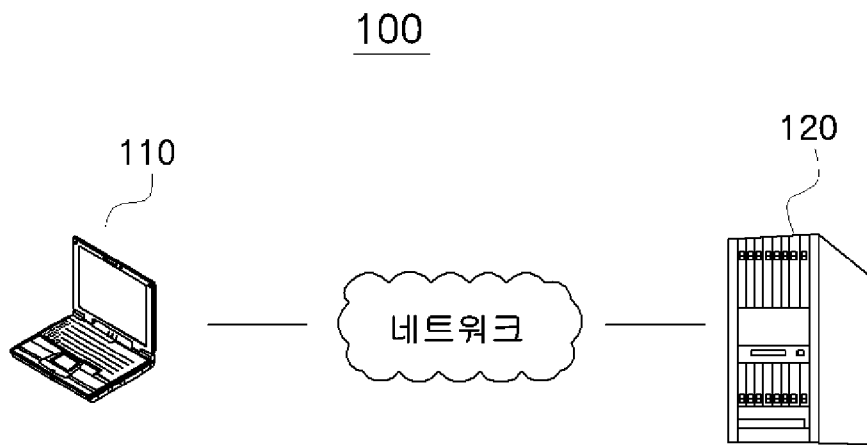
나타내는 복수의 표적 분석 단위체들(target analysis elements)을 결정하는 것을 특징으로 하는 핵산염기서열 보안 장치.

[청구항 19] 제18항에 있어서, 상기 분석 복합체 해석부는
상기 결정된 복수의 표적 분석 단위체들을 조합하여 상기 분석 요청자의 사후오즈(Posterior Odds)를 계산하는 것을 특징으로 하는 핵산염기서열 보안 장치.

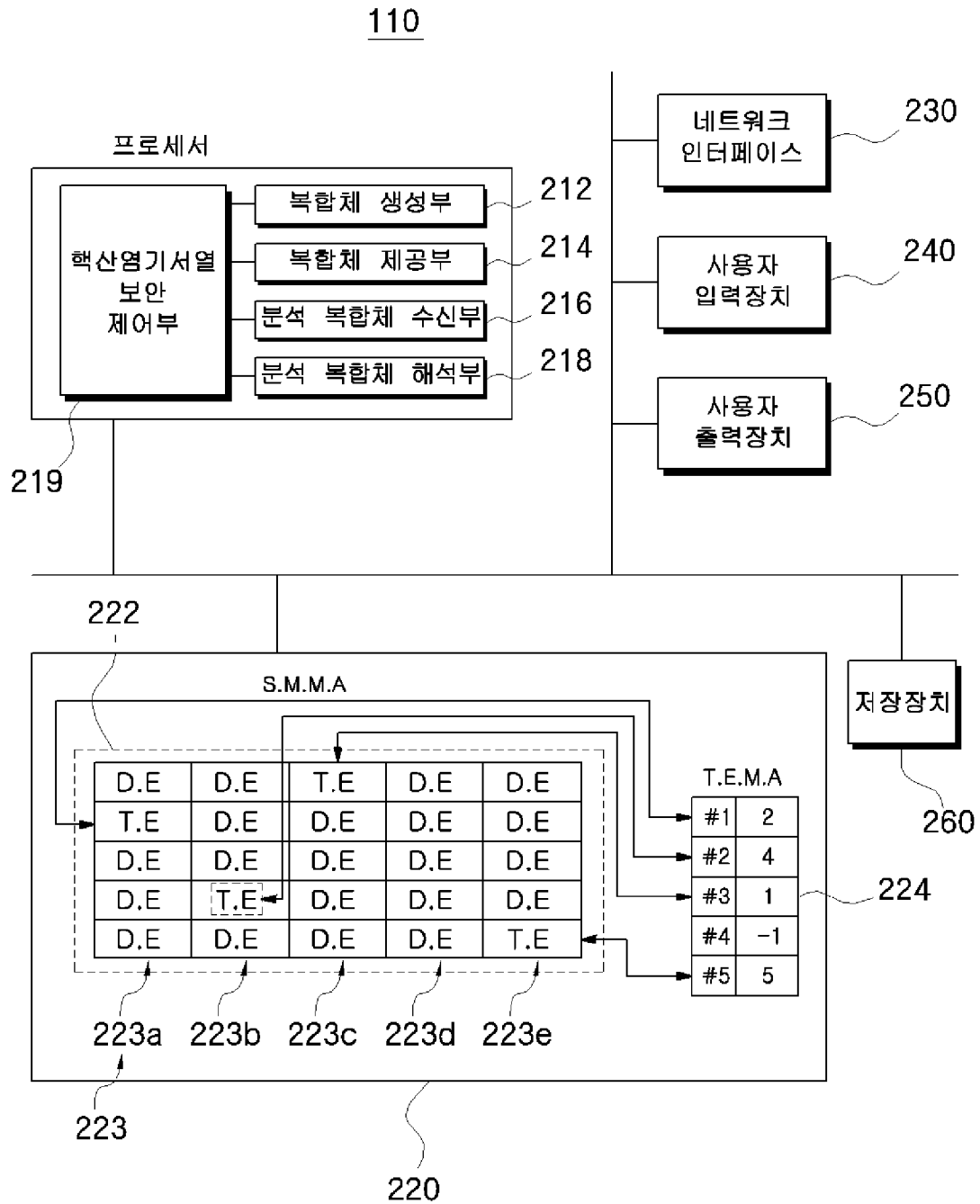
[청구항 20] 각각이 분석 요청자의 핵산염기서열로부터 파생된 표적 단위체(target element) 또는 상기 표적 단위체와 동일하거나 또는 다른 위장 단위체(disguising element) 중 적어도 하나를 포함하는 복수의 복합체들(composites)을 생성하는 기능; 및
상기 생성된 복수의 복합체들을 분석자에게 제공하는 기능을 포함하는 핵산염기서열 보안 방법에 관한 컴퓨터 프로그램을 기록한 기록매체.

[청구항 21] 제20항에 있어서,
상기 분석자로부터 상기 복수의 복합체들에 대한 해석결과를 나타내는 복수의 분석 복합체들을 수신하여 상기 핵산염기서열에 대한 분석 결과를 획득하는 기능을 더 포함하는 것을 특징으로 하는 핵산염기서열 보안 방법에 관한 컴퓨터 프로그램을 기록한 기록매체.

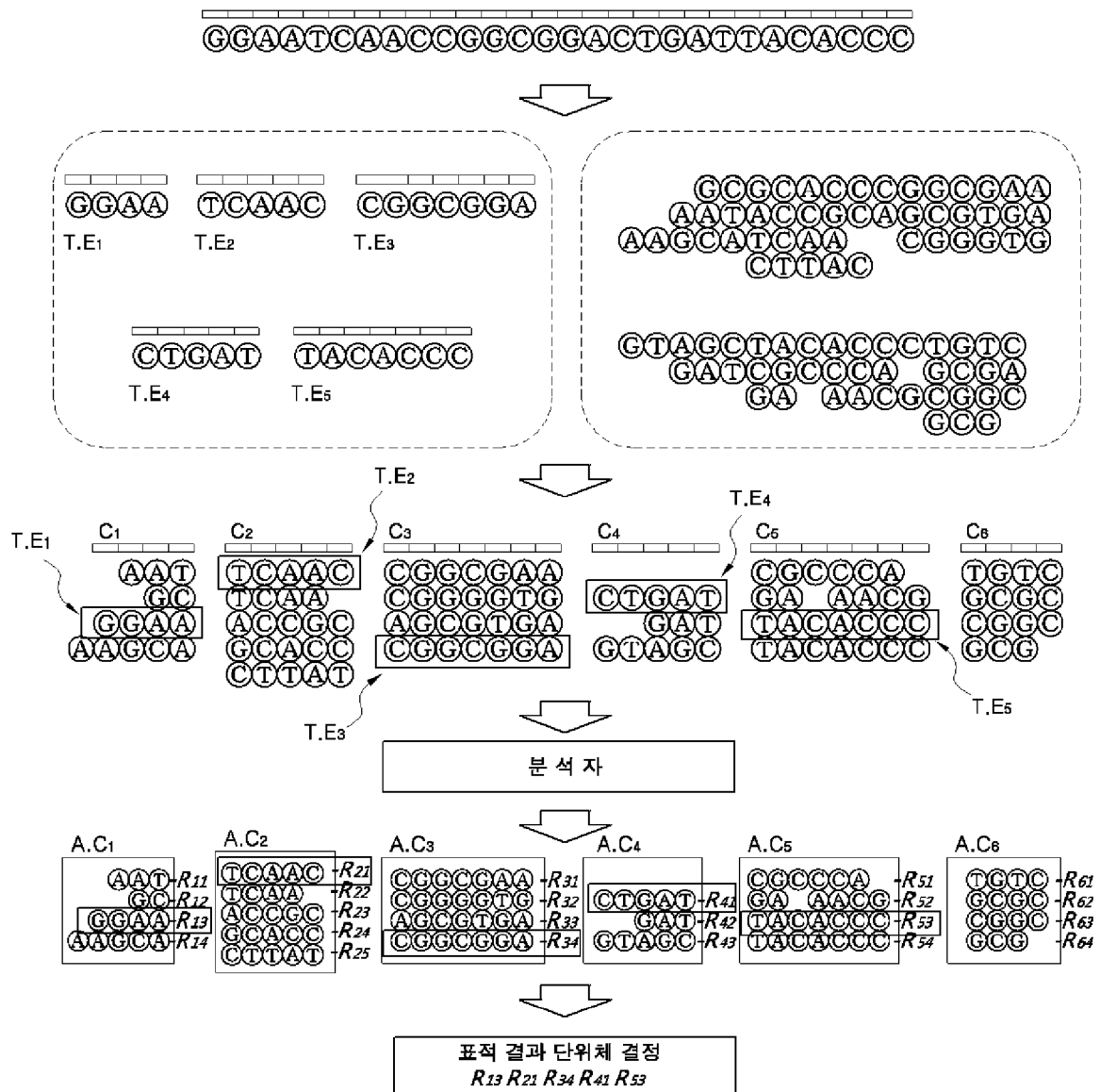
[도1]



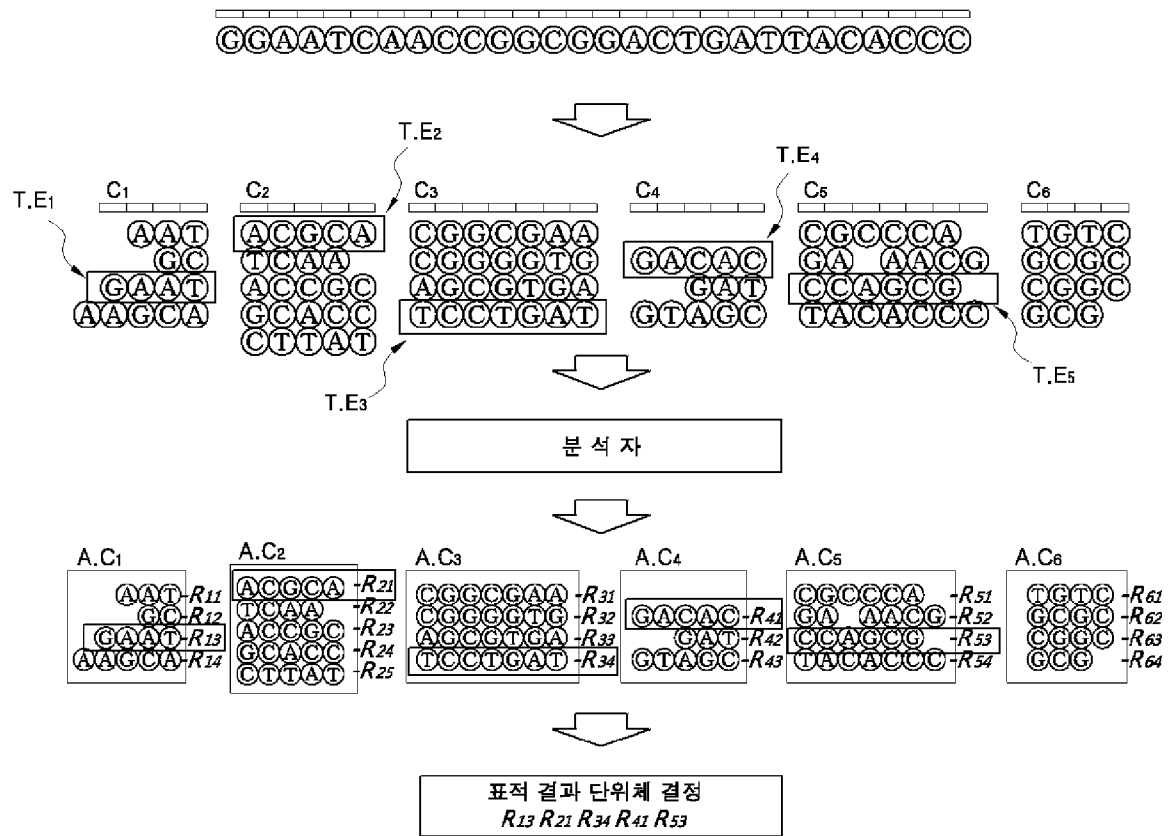
[도2]



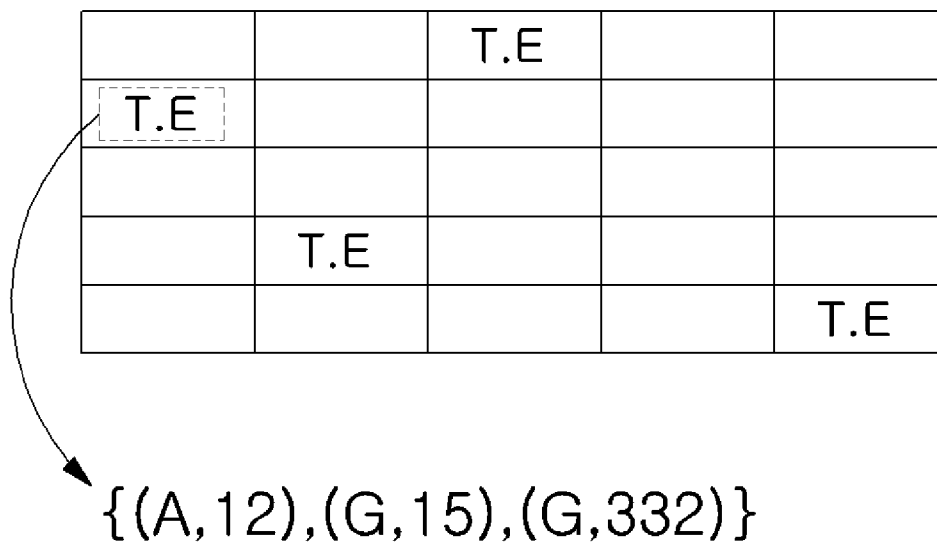
[도3]



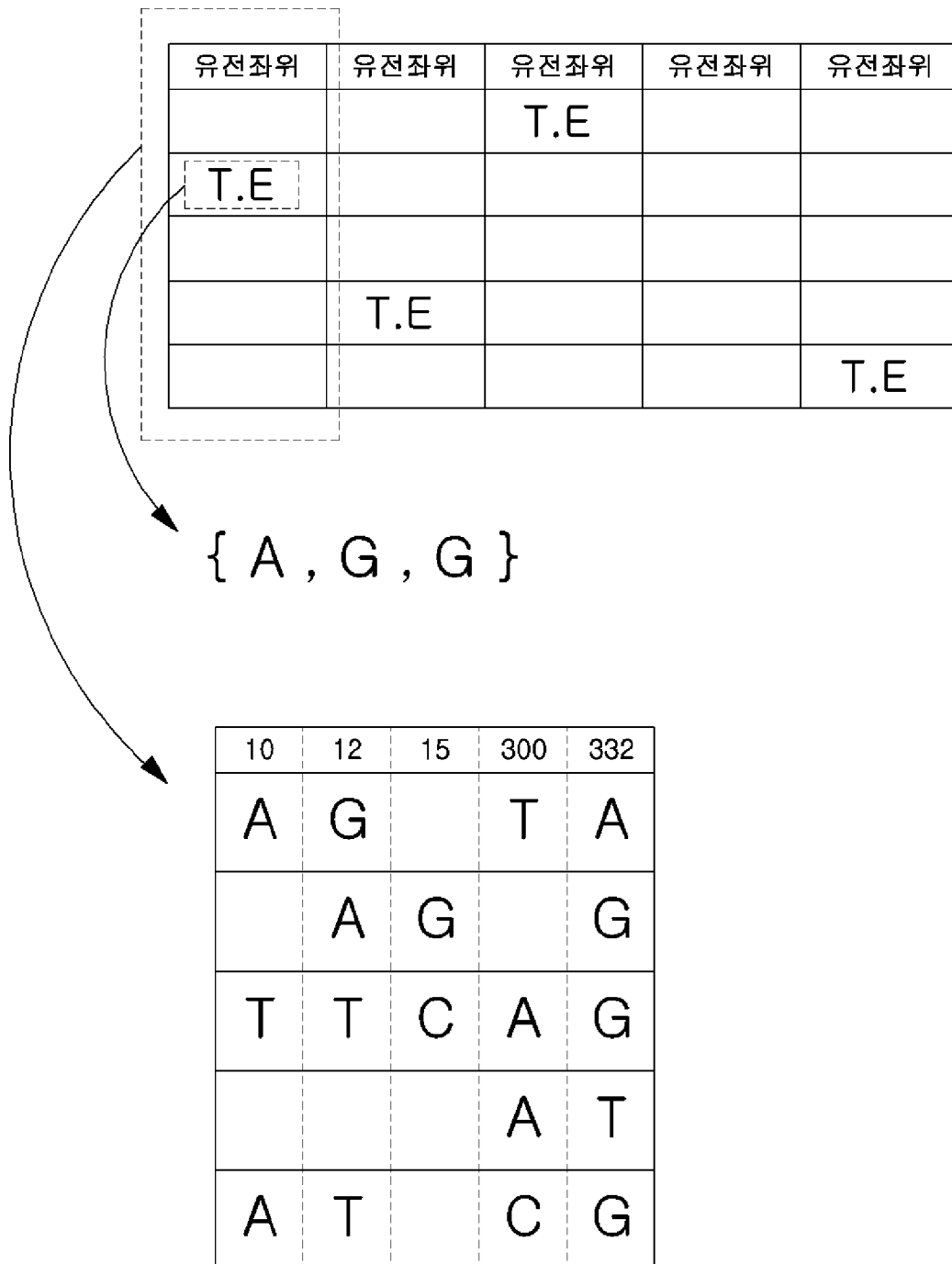
[도4]



[도5a]



[도5b]



[도6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2015/006427**A. CLASSIFICATION OF SUBJECT MATTER****G06F 21/60(2013.01)i, G06F 19/10(2011.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 21/60; C12M 1/00; G06F 19/28; G06F 19/00; G06F 17/30; G06F 19/10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean Utility models and applications for Utility models: IPC as above

Japanese Utility models and applications for Utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: "nucleotide sequence, security, element, disguising"

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2013-178801 A2 (EUROPEAN MOLECULAR BIOLOGY LABORATORY) 05 December 2013 See abstract, figures 2, 7 and paragraphs [0036]-[0049]	1,8,11,12,15,16,17 ,20,21
A		2-7,9,10,13,14,18 ,19
A	KR 10-2013-0122816 A (KNU-INDUSTRY COOPERATION FOUNDATION) 11 November 2013 See abstract, figures 1-5 and paragraphs [0023]-[0034]	1-21
A	KR 10-2013-0075559 A (MACROGEN INC. et al.) 05 July 2013 See abstract, figures 1-3 and paragraphs [0017]-[0050]	1-21
A	JP 2003-242154 A (CELESTAR LEXICO-SCIENCES INC.) 29 August 2003 See abstract, figure 3 and paragraphs [0164]-[0175]	1-21



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

08 OCTOBER 2015 (08.10.2015)

Date of mailing of the international search report

08 OCTOBER 2015 (08.10.2015)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 189 Seonsa-ro, Daejeon 302-701,
Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2015/006427

Patent document cited in search report	Publication date	Patent family member	Publication date
WO 2013-178801 A2	05/12/2013	AU 2013-269536 A1	18/12/2014
		AU 2013-269536 A1	05/12/2013
		CA 2874540 A1	05/12/2013
		CN 104520864 A	15/04/2015
		EP 2856375 A2	08/04/2015
		KR 10-2015-0016572 A	12/02/2015
		WO 2013-178801 A3	23/01/2014
		WO 2013-178801 A3	05/12/2013
KR 10-2013-0122816 A	11/11/2013	NONE	
KR 10-2013-0075559 A	05/07/2013	CN 104025101 A	03/09/2014
		EP 2798553 A1	05/11/2014
		JP 2015-503911 A	05/02/2015
		KR 10-1300475 B1	02/09/2013
		US 2014-0372046 A1	18/12/2014
		WO 2013-100572 A1	04/07/2013
JP 2003-242154 A	29/08/2003	AU 2003-235071 A1	25/03/2004
		AU 2003-235071 B2	20/07/2006
		EP 1477910 A1	17/11/2004
		JP 03-880417 B2	14/02/2007
		JP 2004-147640 A	27/05/2004
		US 2005-0107961 A1	19/05/2005
		WO 2003-081471 A1	02/10/2003

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06F 21/60(2013.01)i, G06F 19/10(2011.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06F 21/60; C12M 1/00; G06F 19/28; G06F 19/00; G06F 17/30; G06F 19/10

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: '핵산염기서열, 보안, 단위체, 위장'

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X A	WO 2013-178801 A2 (EUROPEAN MOLECULAR BIOLOGY LABORATORY) 2013.12.05 요약, 도2,7 및 문단[0036]-[0049] 참조	1, 8, 11, 12, 15, 16, 17, 20, 21 2-7, 9, 10, 13, 14, 18, 19
A	KR 10-2013-0122816 A (강원대학교산학협력단) 2013.11.11 요약, 도1-5 및 문단[0023]-[0034] 참조	1-21
A	KR 10-2013-0075559 A (주식회사 마크로젠 등) 2013.07.05 요약, 도면1-3 및 문단[0017]-[0050] 참조	1-21
A	JP 2003-242154 A (CELESTAR LEXICO-SCIENCES INC) 2003.08.29 요약, 도3 및 문단[0164]-[0175] 참조	1-21

☐ 추가 문헌이 C(계속)에 기재되어 있습니다.

☒ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

“A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“&” 동일한 대응특허문헌에 속하는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

국제조사의 실제 완료일

2015년 10월 08일 (08.10.2015)

국제조사보고서 발송일

2015년 10월 08일 (08.10.2015)

ISA/KR의 명칭 및 우편주소

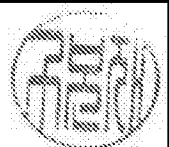
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-472-7140

심사관

구본재

전화번호 +82-42-481-8225



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
WO 2013-178801 A2	2013/12/05	AU 2013-269536 A1 AU 2013-269536 A1 CA 2874540 A1 CN 104520864 A EP 2856375 A2 KR 10-2015-0016572 A WO 2013-178801 A3 WO 2013-178801 A3	2014/12/18 2013/12/05 2013/12/05 2015/04/15 2015/04/08 2015/02/12 2014/01/23 2013/12/05
KR 10-2013-0122816 A	2013/11/11	없음	
KR 10-2013-0075559 A	2013/07/05	CN 104025101 A EP 2798553 A1 JP 2015-503911 A KR 10-1300475 B1 US 2014-0372046 A1 WO 2013-100572 A1	2014/09/03 2014/11/05 2015/02/05 2013/09/02 2014/12/18 2013/07/04
JP 2003-242154 A	2003/08/29	AU 2003-235071 A1 AU 2003-235071 B2 EP 1477910 A1 JP 03-880417 B2 JP 2004-147640 A US 2005-0107961 A1 WO 2003-081471 A1	2004/03/25 2006/07/20 2004/11/17 2007/02/14 2004/05/27 2005/05/19 2003/10/02