



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0087344
(43) 공개일자 2022년06월24일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6881 (2018.01) C12Q 1/6827 (2018.01)
G16B 40/20 (2019.01) G16H 50/20 (2018.01)
(52) CPC특허분류
C12Q 1/6881 (2018.05)
C12Q 1/6827 (2018.05)
(21) 출원번호 10-2021-0058545
(22) 출원일자 2021년05월06일
심사청구일자 2021년05월06일
(30) 우선권주장
1020200177224 2020년12월17일 대한민국(KR)

(71) 출원인
주식회사 네오젠티씨
서울특별시 송파구 올림픽로43길 88, 11층(풍납동, 아산생명과학연구원융합연구관)
재단법인 아산사회복지재단
서울특별시 송파구 올림픽로43길 88 (풍납동)
울산대학교 산학협력단
울산광역시 남구 대학로 93(무거동)
(72) 발명자
이희진
서울특별시 강남구 압구정로39길 58, 63동 408호(압구정동, 구현대아파트)
김영애
서울특별시 송파구 풍성로6가길 14-3, 101호(풍납동)
(뒷면에 계속)
(74) 대리인
이대호, 박건홍

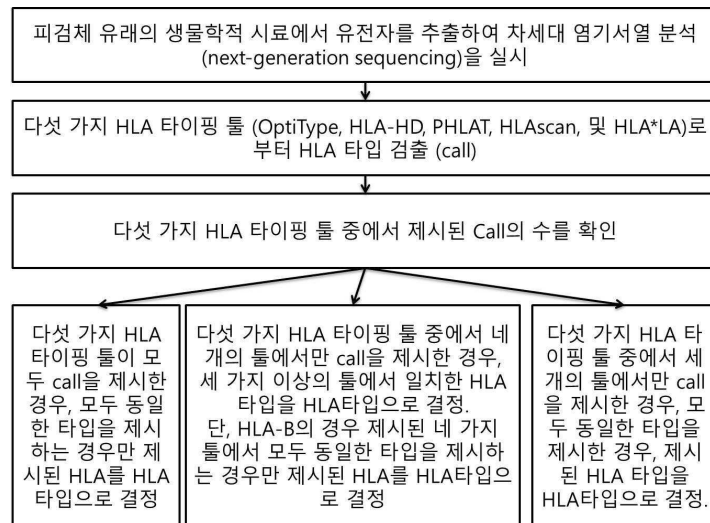
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 인간 백혈구 항원의 타입 결정 방법 및 이를 통한 의료 개인화 방법

(57) 요약

본 발명은 인간 백혈구 항원의 타입 결정 방법 및 이를 통한 의료 개인화 방법 등에 관한 것으로, 본 발명자들은 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 툴을 모두 이용하는 경우, HLA Class I 타이핑의 정확도가 현저히 향상됨을 확인하였는 바, 신뢰도가 향상된 최적의 HLA 타입 결정 방법임을 알 수 있었다. 따라서, 본 발명의 HLA 타입 결정 방법을 이용하여 이식 성공률을 높이기 위한 조직 적합성 테스트뿐만 아니라, PCR-SBT 검사법의 추가 수행 여부 판단 등 의료 개인화를 가능하게 할 수 있을 것으로 기대된다.

대표도 - 도6



(52) CPC특허분류

G16B 40/20 (2019.02)

G16H 50/20 (2018.01)

(72) 발명자

이미선

경상남도 창원시 진해구 중원동로 38, 104동 405
호(부흥동, 태영데시앙)

서정환

서울특별시 강동구 천호대로168나길 56-14, 302호
(성내동, 더존팰리스)

명세서

청구범위

청구항 1

하기의 단계를 포함하는, 인간 백혈구 항원(Human Leukocyte Antigen; HLA) 타입 결정 방법:

- (a) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 차세대 염기서열 분석(next generation sequencing)을 실시하는 단계; 및
- (b) 상기 서열분석 결과를 이용해 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA로 이루어진 군으로부터 선택된 세 가지 이상의 방법으로 HLA 타입을 분석 및 결정하는 단계.

청구항 2

제1항에 있어서,

상기 단계 (b)는, (b-1) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 틀에서 특정 타입을 모두 제시한 경우에는 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계를 포함하는 것을 특징으로 하는, 방법.

청구항 3

제1항에 있어서,

상기 단계 (b)는, (b-1) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법에서 특정 타입을 모두 제시한 경우에는 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계; 및

(b-2) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법 중 네 개의 방법에서만 특정 타입이 제시되었을 경우, 네 개의 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계;를 포함하는 것을 특징으로 하는, 방법.

청구항 4

제3항에 있어서,

상기 일치한 HLA가 HLA-A 또는 HLA-C인 경우,

상기 (b-2) 단계에서, OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA로 이루어진 군으로부터 선택된 세 가지 이상의 방법에서 일치한 HLA를 HLA 타입으로 결정하는 단계를 포함하는 것을 특징으로 하는, 방법.

청구항 5

제3항에 있어서,

상기 일치한 HLA가 HLA-B인 경우,

상기 (b-2) 단계에서, OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA로 이루어진 군으로부터 선택된 네 개의 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계; 및

(b-3) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법 중 세 개의 방법에서만 특정 타입이 제시되었을 경우, 세 개의 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계;를 포함하는 것을 특징으로 하는, 방법.

로 하는, 방법.

청구항 6

제1항에 있어서,

상기 염기서열 분석은 전장 유전체 염기서열 분석(whole genome sequencing), 전체 엑솜 염기서열 분석(whole exome sequencing) 또는 전체 전사체 염기서열 분석(whole transcriptome sequencing)인 것을 특징으로 하는, 방법.

청구항 7

제1항에 있어서,

상기 생물학적 시료는 조직, 세포, 전혈, 또는 혈액인 것을 특징으로 하는, 방법.

청구항 8

제1항에 있어서,

상기 단계 (b)에서 HLA 타입이 결정되지 못한 경우, (c) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 PCR-SBT (Polymerase Chain Reaction-Sequencing Based Typing)를 실시하는 단계를 더 포함하는 것을 특징으로 하는, 방법.

청구항 9

제1항에 있어서,

상기 HLA 타입은 HLA-A 타입, HLA-B 타입, 및 HLA-C 타입으로 이루어진 군으로부터 선택된 하나 이상인 것을 특징으로 하는, 방법.

청구항 10

개별화된 의료 개입을 결정하기 위한 시스템에 있어서, 상기 시스템은

- 호스트 서버;
- 데이터를 액세스하고 입력하기 위해 상기 호스트 서버를 액세스하기 위한 사용자 인터페이스;
- 입력된 데이터를 프로세싱하기 위한 프로세서; 및
- 상기 프로세서로 연결되는 메모리로써,
 - (i) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 차세대 염기서열 분석(next generation sequencing) 데이터를 액세스하기 위한 인스트럭션,
 - (ii) 상기 데이터로부터 제1항의 방법에 따라 결정된 HLA 타입에 액세스하기 위한 인스트럭션, 및
 - (iii) 프로세싱된 데이터가 저장된 메모리;를 포함하는 시스템.

청구항 11

제10항에 있어서,

상기 개별화된 의료 개입은 추가적인 PCR-SBT 검사법의 수행 여부를 결정하는 것을 특징으로 하는, 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 인간 백혈구 항원의 타입 결정 방법 및 이를 통한 의료 개인화 방법 등에 관한 것이다.

배경 기술

[0003] 인간 백혈구 항원(Human Leukocyte Antigen; HLA)은 인간의 주조직 적합성복합체(Major Histocompatibility Complex; MHC) 유전자에 의해 생성되는 당단백 분자로, 성숙한 적혈구에는 HLA 항원이 존재하지 않지만 미성숙 적아세포(erythroblast)에서는 발현이 되며 백혈구, 혈소판 등의 혈액세포를 포함한 인체 내 모든 조직세포의 표면에 발현된다. MHC 유전자는 모든 척추동물에 존재하며 인간의 MHC 유전자를 HLA 유전자, 이로부터 발현된 산물을 HLA 항원이라고 한다.

[0004] MHC 유전자들은 자기(Self) 및 비자기(Non-self)의 인지, 항원 자극에 대한 면역반응, 세포성 면역과 체액성 면역의 조절 및 질병에 대한 감수성 등에 관여한다. MHC 유전자의 산물인 HLA 항원은 고형 장기이식에서 이식된 장기의 생존에 있어서 ABO 혈액형 다음으로 중요한 항원이고, 골수이식에 있어서는 이식의 성패에 가장 중요한 역할을 한다고 알려져 있으며, HLA 항원의 차이를 면역학적으로 인지하는 것이 이식된 조직에 대한 거부반응(rejection)의 첫 단계라 볼 수 있다. 또한, 수혈요법에 있어서 HLA 항원과 항체는 혈소판수혈불응증, 발열성 비용혈성 수혈 부작용, 급성 폐손상, 수혈 후 이식편대숙주병과 같은 여러 가지 부작용의 발생에 중요한 역할을 한다.

[0005] HLA 항원은 크게 Class I 및 Class II로 분류되는데, Class I은 HLA-A, HLA-B, HLA-C로 분류되고 대부분의 유핵세포 및 혈소판에서 발현되며, 세포독성 T 세포가 바이러스에 감염된 세포나 종양세포를 인지하여 제거할 때 항원 인식(antigen recognition)에 필수적이다. HLA Class II는 HLA-DR, HLA-DQ, HLA-DP로 분류되고 B 세포, 단핵세포, 수지상세포, 활성화된 T 세포에서 발현되며, 헬퍼 T 세포의 항원 수용체(antigen receptor)와 작용하여 세포성 및 체액성 면역반응을 유발하고, 항원제시세포에 표현된 항원을 인지할 때 필수적인 것으로 알려져 있다. HLA는 인간이 가지고 있는 유전자 중에서 가장 큰 다형성(polymorphism)을 보이는 유전자이며 인종 및 민족 간에도 빈도 차이가 존재한다. 따라서 전술한 바와 같이 장기이식, 면역치료, 질병관련 연구, 친자감별과 같은 부친시험, 법의학적 이용, 유전학적 연구 등의 다양한 분야에서 정확한 HLA 타입을 결정하기 위한 HLA 타이핑이 필요하다. 이에 따라 현재까지 PCR-SSP, PCR-SSOP, 중합효소연쇄반응 시퀀싱 기반 타이핑(Polymerase chain reaction-sequencing based Typing; PCR-SBT) 기술과 차세대 염기서열 분석(Next Generation Sequencing; NGS) 방법에 기반한 다양한 HLA 타입 분석 방법들이 제시되어 왔다.

[0006] NGS 방법은 PCR-SBT 검사법에서 주로 시행해 오던 Class I (HLA-A, -B, -C)에 대한 엑손 2, 3, 4와 Class II (DRB1, DQB1, DPB1)에 대한 엑손 2 (또는 exon 2, 3)를 넘어 더 많은 유전자와 더 많은 엑손에 대한 분석을 동시에 시행할 수 있어 모호성 비율(ambiguity rate)을 줄이기 위해 노력해 왔던 추가적인 프라이머, 그룹 PCR 등의 요구를 없애는 효과가 있으며, 인덱싱을 이용하여 여러 표적 유전자 또는 여러 검체를 한 번에 pooling 검사하는 장점이 있다.

[0007] 그러나 IMGT-HLA database에서 제공하는 HLA 대립유전자에 기반하기 때문에 흔하게 발견되는 HLA 대립유전자에 대해서는 비교적 예측 정확도가 높으나 빈도가 낮은 대립유전자에 대해서는 정확도가 떨어지는 것으로 나타났다. 이에, HLA 타입을 보다 정확히 예측할 수 있는 신뢰도 높은 방법을 구축하는 것이 필요할 뿐만 아니라, 개체에 따라서는 NGS 후에도 PCR-SBT 검사법을 추가적으로 실시하여 정확한 타이핑이 필요하므로, 추가적인 PCR-SBT 검사법이 필요한 개체의 분류 방법이 필요한 실정이다.

[0008] 또한, 최근 의료기술의 발달로 돌연변이(somatic mutation)로 인한 종양에서 생성된 신생항원(neoantigen)을 기반으로 한 항암면역치료법이 개발되었다. 신생항원을 발굴한 후에는 이를 대상으로 종양백신을 생산하거나 신생항원에 반응하여 종양을 공격할 수 있는 T 세포 수용체(T cell receptor; TCR)를 가진 T 세포를 선별하여 체외에서 배양하거나 T 세포 수용체를 조작한 T 세포(TCR engineered T cell)를 제조하여 치료제로 사용할 수가 있다. 동일한 돌연변이를 가진 환자라고 하더라도 각 환자의 HLA 종류에 따라 해당 돌연변이 단백질이 HLA에 결합

되어 신생항원으로 제시되기도 하고, HLA에 결합되지 못 하는 경우에는 신생항원이 될 수 없다. 따라서 HLA를 정확히 확인하여 신생항원 여부를 결정하는 것이 이를 기반으로 한 항암면역치료법 개발에 매우 중요하다.

선행기술문헌

특허문헌

[0010] (특허문헌 0001) 한국공개특허 제2019-0001953호 (2019.01.07.)

발명의 내용

해결하려는 과제

- [0011] 이에 본 발명자들은 개인의 인간 백혈구 항원(Human Leukocyte Antigen; HLA) 타입을 보다 정확히 판정할 수 있는 방법을 예의 연구한 결과, 차세대 염기서열 분석(NGS) 데이터를 이용하여 보다 정확한 HLA 타입을 예측할 수 있는 최적의 방법을 개발하였는바, 이에 기초하여 본 발명을 완성하였다.
- [0012]
- [0013] 따라서, 본 발명의 목적은 하기의 단계를 포함하는, 인간 백혈구 항원(Human Leukocyte Antigen; HLA) 타입 결정 방법을 제공하는 것이다.
- [0014] (a) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 차세대 염기서열 분석(next generation sequencing)을 실시하는 단계; 및
- [0015] (b) 상기 서열분석 결과를 이용해 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA로 이루어진 군으로부터 선택된 세 가지 이상의 방법으로 HLA 타입을 분석 및 결정하는 단계.
- [0016] 본 발명의 다른 목적은 개별화된 의료 개입을 결정하기 위한 시스템에 있어서, 상기 시스템은
- [0017] - 호스트 서버;
- [0018] - 데이터를 액세스하고 입력하기 위해 상기 호스트 서버를 액세스하기 위한 사용자 인터페이스;
- [0019] - 입력된 데이터를 프로세싱하기 위한 프로세서; 및
- [0020] - 상기 프로세서로 연결되는 메모리로써,
- [0021] (i) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 차세대 염기서열 분석(next generation sequencing) 데이터를 액세스하기 위한 인스트럭션,
- [0022] (ii) 상기 데이터로부터 상기 방법에 따라 결정된 HLA 타입에 액세스하기 위한 인스트럭션, 및
- [0023] (iii) 프로세싱된 데이터가 저장된 메모리;를 포함하는 시스템을 제공하는 것이다.
- [0024]
- [0025] 그러나 본 발명이 이루고자 하는 기술적 과제는 이상에서 언급한 과제에 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0027] 상기 본 발명의 목적을 달성하기 위하여, 본 발명은 하기의 단계를 포함하는, 인간 백혈구 항원(Human Leukocyte Antigen; HLA) 타입 결정 방법을 제공한다.
- [0028] (a) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 차세대 염기서열 분석(next generation sequencing)을 실시하는 단계; 및
- [0029] (b) 상기 서열분석 결과를 이용해 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA로 이루어진 군으로부터 선택된

세 가지 이상의 방법으로 HLA 타입을 분석 및 결정하는 단계.

- [0030] 또한, 본 발명은 개별화된 의료 개입을 결정하기 위한 시스템에 있어서, 상기 시스템은
- [0031] - 호스트 서버;
- [0032] - 데이터를 액세스하고 입력하기 위해 상기 호스트 서버를 액세스하기 위한 사용자 인터페이스;
- [0033] - 입력된 데이터를 프로세싱하기 위한 프로세서; 및
- [0034] - 상기 프로세서로 연결되는 메모리로서,
- [0035] (i) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 차세대 염기서열 분석(next generation sequencing) 데이터를 액세스하기 위한 인스트럭션,
- [0036] (ii) 상기 데이터로부터 상기 방법에 따라 결정된 HLA 타입에 액세스하기 위한 인스트럭션, 및
- [0037] (iii) 프로세싱된 데이터가 저장된 메모리;를 포함하는 시스템을 제공한다.
- [0039] 본 발명의 일 구현예에서, 상기 단계 (b)는, (b-1) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 틀에서 특정 타입을 모두 제시한 경우에는 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계를 포함할 수 있으나, 이에 제한되는 것은 아니다.
- [0040] 본 발명의 다른 구현예에서, 상기 단계 (b)는, (b-1) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법에서 특정 타입을 모두 제시한 경우에는 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계; 및
- [0041] (b-2) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법 중 네 개의 방법에서만 특정 타입이 제시되었을 경우, 네 개의 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계를 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0042] 본 발명의 또 다른 구현예에서, 상기 일치한 HLA가 HLA-A 또는 HLA-C인 경우, 상기 (b-2) 단계에서, OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA로 이루어진 군으로부터 선택된 세 가지 이상의 방법에서 일치한 HLA를 HLA 타입으로 결정하는 단계를 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0043] 본 발명의 또 다른 구현예에서, 상기 일치한 HLA가 HLA-B인 경우,
- [0044] 상기 (b-2) 단계에서, OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA로 이루어진 군으로부터 선택된 네 개의 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계; 및
- [0045] (b-3) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법 중 세 개의 방법에서만 특정 타입이 제시되었을 경우, 세 개의 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계;를 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0046] 본 발명의 또 다른 구현예에서, 상기 염기서열 분석은 전장 유전체 염기서열 분석(whole genome sequencing), 전체 엑솜 염기서열 분석(whole exome sequencing) 또는 전체 전사체 염기서열 분석(whole transcriptome sequencing)인 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0047] 본 발명의 또 다른 구현예에서, 상기 생물학적 시료는 조직, 세포, 전혈, 또는 혈액인 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0048] 본 발명의 또 다른 구현예에서, 상기 단계 (b)에서 HLA 타입이 결정되지 못한 경우, (c) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 PCR-SBT (Polymerase Chain Reaction-Sequencing Based Typing)를 실시하는 단계를 더 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0049] 본 발명의 또 다른 구현예에서, 상기 HLA 타입은 HLA-A 타입, HLA-B 타입, 및 HLA-C 타입으로 이루어진 군으로부터 선택된 하나 이상인 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0050] 본 발명의 또 다른 구현예에서, 상기 개별화된 의료 개입은 추가적인 PCR-SBT 수행 여부를 결정하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0051] 또한, 본 발명의 인간 백혈구 항원 타입 결정 방법은 개별화된 의료 개입을 결정하기 위한 방법에 사용되는 것

을 특징으로 할 수 있으나, 이에 제한되는 것은 아니다.

발명의 효과

[0053] 본 발명자들은 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 툴을 모두 이용하는 경우 HLA Class I 타이핑의 정확도가 현저히 향상됨을 확인하였는 바, 신뢰도가 향상된 최적의 HLA 타입 결정 방법임을 알 수 있었다. 따라서 본 발명의 HLA 타입 결정 방법을 이용하여 이식 성공률을 높이기 위한 조직 적합성 테스트뿐만 아니라, PCR-SBT 추가 수행 여부 판단 등 의료 개인화를 가능하게 할 수 있을 것으로 기대된다.

도면의 간단한 설명

[0055] 도 1a 및 도 1b는 1000 유전체 프로젝트 데이터를 사용하여 8개의 알고리즘 성능을 비교한 결과를 나타낸 것으로, 도 1a는 총 2,619건에 대한 8개 툴 각각에 대한 가용성 및 정확도를 나타낸 것이며, 도 1b는 리드길이 100bp 이상에 해당하는 1,275건에 대한 8개 툴 각각의 가용성 및 정확도를 나타낸 것이다.

도 2a 내지 도 2c는 리드길이 100bp 이상에 해당하는 데이터에서 HLA 유전형 분석 툴을 결합하여 정확도를 분석한 결과를 나타낸 것이다.

도 3은 39개의 자체 샘플에 대한 8개의 알고리즘 성능을 비교한 결과를 나타낸 것이다.

도 4a 내지 도 4c는 자체 샘플 데이터에서 HLA 유전형 분석 툴을 결합하여 정확도를 분석한 결과를 나타낸 것이다.

도 5는 의료 개입을 결정하기 위한 시스템의 블록 다이어그램을 나타낸 것이다.

도 6은 의료 개입을 결정하기 위한 방법의 흐름도를 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0056] 이하, 본 발명에 대해 상세히 설명한다.

[0057]

[0058] 먼저, 본 발명은 하기의 단계를 포함하는, 인간 백혈구 항원(Human Leukocyte Antigen; HLA) 타입 결정 방법을 제공한다.

[0059] (a) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 차세대 염기서열 분석(next generation sequencing)을 실시하는 단계; 및

[0060] (b) 상기 서열분석 결과를 이용해 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법으로 HLA 타입을 분석 및 결정하는 단계.

[0061] 본 발명에서 사용되는 용어, "인간 백혈구 항원(Human Leukocyte Antigen; HLA)"은 인간의 주조직적합성복합체(Major Histocompatibility Complex: MHC) 유전자에 의해 생성되는 당단백 분자로, 인간이 가지고 있는 유전자 중에서 가장 큰 다형성(polymorphism)을 보이는 유전자이다. 이러한 개인 간의 HLA 타입을 결정하는 HLA 타이핑은 장기이식, 면역치료, 질병관련 연구, 친자감별과 같은 부친시험, 법의학적 이용, 유전학적 연구 등의 다양한 분야에서 매우 중요하게 이용되므로 HLA 타입을 정확히 예측하는 것이 무엇보다 중요하다.

[0062] 본 발명에 있어서, 상기 (a) 단계는 피검체의 HLA 타입 분석을 위해 시료를 수집하고 단계 (b)의 방법으로 HLA 타입 분석을 실시하기 위해 NGS 결과를 도출하는 단계이다.

[0063] 상기 OptiType은 2014년에 발표된 정수 선형 프로그래밍(integer linear programming)에 기반한 HLA 유전자형 결정 알고리즘으로 모든 주요 또는 소수의 HLA class I 대립유전자를 선택하여 NGS 데이터로부터 정확한 4자리의 HLA 유전자형의 예측이 가능한 방법이라고 알려져 있다. 그러나 OptiType은 정확도 및 유용성 면에서 문제점이 있는 것으로 알려져 있다(Szolek 등의 OptiType: precision HLA typing from next-generation sequencing data. Bioinformatic, Vol. 30 no. 23 2014, pages 3310-3316, 참조).

[0064] 상기 HLA-HD란 2017년에 일본에서 발표된 NGS 데이터에서 6자리 정밀도로 HLA 대립 유전자를 정확히 결정할 수

있는 알고리즘이다. 자세한 사항은 문헌(Hum Mutat. 2017 Jul;38(7):788-797. doi: 10.1002/humu.23230.)을 참조할 수 있다.

- [0065] 상기 PHLAT란 RNAseq 또는 엑솜 시퀀싱 데이터를 입력값으로 사용하여 주요 클래스 I 및 클래스 II HLA 유전자의 고해상도(6 자리) 타이핑을 위해 설계된 계산 툴이다. 자세한 사항은 문헌(Methods Mol Biol. 2018;1802:193-201.)을 참조할 수 있다.
- [0066] 상기 HLAscan이란 국제 ImMunoGeneTics 프로젝트/인간 백혈구 항원 (IMGT/HLA) 데이터베이스에서 HLA 시퀀스에 대한 읽기 정렬을 수행하는 툴이다. 자세한 사항은 문헌(BMC Bioinformatics volume 18, Article number: 258 (2017))을 참조할 수 있다.
- [0067] 상기 HLA*LA란 선형 정렬을 변이 그래프에 투영하는 것을 기반으로 HLA 유형을 추론하는 모델로써, 전체 엑솜 및 전체 게놈 Illumina 데이터에서 정확한 HLA 유전자형의 예측이 가능한 방법이라고 알려져 있다. 자세한 사항은 문헌(Bioinformatics, Volume 35, Issue 21, 1 November 2019, Pages 4394-4396)을 참조할 수 있다.
- [0068] 상기 1000 유전체 프로젝트 데이터란 26개 인종으로 구성된 인간 1,000명의 유전체를 해독하는 국제 프로젝트로부터 수득된 데이터로, 웹사이트 [<https://www.internationalgenome.org/>]에서 확인할 수 있다.
- [0069] 상기 각 방법에 따른 HLA 타입 판정에 대한 정확도는 상이하며 따라서 보다 정확도가 높은 최적의 HLA 타입 결정 방법이 요구되고 있다.
- [0070] 본 발명의 다른 구현예에서, 상기 단계 (b)는, (b-1) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법에서 특정 타입을 모두 제시한 경우에는 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계; 및
- [0071] (b-2) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법 중 네 개의 방법에서만 특정 타입이 제시되었을 경우, 네 개의 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계를 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0072] 본 발명의 또 다른 구현예에서, 상기 일치한 HLA가 HLA-A 또는 HLA-C인 경우, 상기 (b-2) 단계에서 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA로 이루어진 군으로부터 선택된 세 가지 이상의 방법에서 일치한 HLA를 HLA 타입으로 결정하는 단계를 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0073] 본 발명의 또 다른 구현예에서, 상기 일치한 HLA가 HLA-B인 경우,
- [0074] 상기 (b-2) 단계에서, OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA로 이루어진 군으로부터 선택된 네 개의 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계; 및
- [0075] (b-3) 상기 OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA 방법 중 세 개의 방법에서만 특정 타입이 제시되었을 경우, 세 개의 방법에서 모두 동일하게 제시된 HLA를 HLA 타입으로 결정하는 단계;를 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0076] 본 발명에서, 상기 (b) 단계의 HLA 타입을 분석하는 과정 중, OptiType, HLA-HD, PHLAT, HLAscan, 및 HLA*LA로 이루어진 군으로부터 선택된 하나 이상의 방법에서 다중 call을 나타내는 경우, 상기 방법을 제외할 수 있으나, 이에 제한되는 것은 아니다.
- [0077] 본 발명에서, 상기 일치한 HLA가 HLA-A 또는 HLA-C인 경우로써, 네 가지 방법에서 HLA 타입이 일치한 경우,
- [0078] OptiType, HLA-HD, PHLAT, 및 HLA*LA 조합;
- [0079] OptiType, HLA-HD, HLAscan, 및 HLA*LA 조합;
- [0080] OptiType, HLA-HD, PHLAT, 및 HLAscan 조합;
- [0081] OptiType, PHLAT, HLAscan, 및 HLA*LA 조합; 및
- [0082] HLA-HD, PHLAT, HLAscan, 및 HLA*LA 조합으로 이루어진 군으로부터 선택된 하나 이상의 조합을 사용하여 HLA 타입을 결정하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0083] 본 발명에서, 상기 일치한 HLA가 HLA-A 또는 HLA-C인 경우로써, 세 가지 방법에서 HLA 타입이 일치한 경우,
- [0084] OptiType, HLA-HD, 및 HLA*LA 조합;

- [0085] OptiType, HLA-HD, 및 PHLAT 조합;
- [0086] OptiType, HLA-HD, 및 HLAscan 조합;
- [0087] OptiType, PHLAT, 및 HLA*LA 조합;
- [0088] OptiType, HLAscan, 및 HLA*LA 조합;
- [0089] OptiType, PHLAT, 및 HLAscan 조합;
- [0090] HLA-HD, PHLAT, 및 HLA*LA 조합;
- [0091] HLA-HD, HLAscan, 및 HLA*LA 조합;
- [0092] HLA-HD, PHLAT, 및 HLAscan 조합; 및
- [0093] PHLAT, HLAscan, 및 HLA*LA 조합으로 이루어진 군으로부터 선택된 하나 이상의 조합을 사용하여 HLA 타입을 결정하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0094] 본 발명에서, 상기 일치한 HLA가 HLA-B인 경우로써, 네 가지 방법에서 HLA 타입이 일치한 경우,
- [0095] OptiType, HLA-HD, PHLAT, 및 HLA*LA 조합;
- [0096] OptiType, HLA-HD, HLAscan, 및 HLA*LA 조합;
- [0097] OptiType, HLA-HD, PHLAT, 및 HLAscan 조합;
- [0098] OptiType, PHLAT, HLAscan, 및 HLA*LA 조합; 및
- [0099] HLA-HD, PHLAT, HLAscan, 및 HLA*LA 조합으로 이루어진 군으로부터 선택된 하나 이상의 조합을 사용하여 HLA 타입을 결정하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0100] 본 발명에서, 상기 염기서열 분석은 전장 유전체 염기서열 분석(whole genome sequencing), 전체 엑솜 염기서열 분석(whole exome sequencing) 또는 전체 전사체 염기서열 분석(whole transcriptome sequencing)인 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0101] 본 명세서에서 사용된 용어 "개체" 또는 "대상체"는 HLA 타입을 결정하기 위한 피험자를 의미한다.
- [0102] 본 명세서에서 사용된 용어 "시료"는 HLA 타입을 결정하고자 하는 개체 또는 대상체로부터 채취된 것이라면 제한 없이 사용할 수 있으며, 예를 들어 생검 등으로 얻어진 세포나 조직, 혈액, 전혈, 혈청, 혈장, 타액, 뇌척수액, 각종 분비물, 소변, 대변 등일 수 있다. 바람직하게는 혈액, 혈장, 혈청, 타액, 비액, 객담, 복수, 질 분비물 및 소변으로 이루어진 군에서 선택될 수 있으며, 바람직하게는 혈액, 혈장 또는 혈청일 수 있다. 상기 시료는 검출 또는 진단에 사용하기 전에 전처리할 수 있다. 예를 들어, 균질화(homogenization), 여과, 증류, 추출, 농축, 방해 성분의 불활성화, 시약의 첨가 등을 포함할 수 있다.
- [0103] 본 발명에서, 상기 생물학적 시료는 조직, 세포, 전혈, 또는 혈액인 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0104] 본 발명에서, 상기 단계 (b)에서 HLA 타입이 결정되지 못한 경우, (c) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 PCR-SBT (Polymerase Chain Reaction-Sequencing Based Typing)를 실시하는 단계를 더 포함할 수 있으나, 이에 제한되는 것은 아니다. 본 발명의 일 실시예에서는 단계 (b)에서 HLA 타입이 결정되지 못한 경우, HLA 타입 정확도가 매우 떨어지게 되는 단점이 있어, 추가적인 PCR-SBT를 수행하는 것이 바람직함을 확인하였다.
- [0105] 본 발명에서, 상기 PCR-SBT는 추가적인 HLA 유전자형 정보가 필요하거나 PCR-SSP 검사 결과의 판정이 애매한 경우 보조검사에 이용되고 있다. PCR-SSO 또는 RFLP의 단점은 두 대립 유전자 간의 DNA 서열에서 단일 염기 차이를 감지할 수 있지만, 변이가 프로브에 의해 감지된 특정 부위에서 발생하지 않는 한 새로운 정의되지 않은 대립 유전자를 감지할 가능성이 적다는 것이다. PCR-SBT는 원래 수동 시퀀싱 방법으로 개발되었으며 염료 라벨이 붙은 프라이머와 형광 자동 시퀀싱의 도입으로 자동화가 가능해지고 정확도가 크게 향상되었다. PCR-SBT는 대립 유전자의 염기 서열을 직접 검출하는 기술로, 정확도가 우수하지만, 시간이 오래 걸리고 비용이 비싸며 검체 처리량이 적은 단점이 있다.
- [0106] 본 발명에서, 상기 PCR은 실시간 PCR일 수 있으며, 보다 구체적으로 단일 PCR 또는 멀티플렉스 PCR일 수

있으나, 이에 제한되는 것은 아니다. 본 발명에서 상기 단일 또는 멀티플렉스 실시간 PCR 반응 조건은 통상적인 조건을 취할 수 있다. 또한, 단일 실시간 PCR과 멀티플렉스 실시간 PCR 반응은 동일한 조건으로 수행될 수 있으며, 일 예로 초기 변성(initial denaturation)을 95℃에서 1분간 수행한 후, 변성(denaturation, 95℃에서 25초), 어닐링(annealing, 65℃에서 45초) 및 연장(extension, 72℃에서 30초)을 총 40회 실시하는 조건을 취할 수 있다. 본 발명에 사용 가능한 실시간 PCR 기기로는 AB사의 Real-time PCR 기기 7900, 7500, 7300, Roche사의 LightCycler 80, Stratagene사의 Mx3000p, 및 BioRad사의 Chromo 4 기기 등이 있으나, 이에 제한되는 것은 아니다.

[0107] 본 발명에서, 상기 PCR은 PCR-SSOP (PCR sequence-specific oligonucleotide probes) 또는 PCR-SBT (PCR sequencing-based typing)일 수 있으나, 이에 제한되는 것은 아니다.

[0108] 본 발명에서, 상기 HLA 타입은 HLA-A 타입, HLA-B 타입, 및 HLA-C 타입으로 이루어진 군으로부터 선택된 하나 이상인 것일 수 있으나, 이에 제한되는 것은 아니다.

[0110] 또한, 본 발명은 개별화된 의료 개입을 결정하기 위한 시스템에 있어서, 상기 시스템은

[0111] - 호스트 서버;

[0112] - 데이터를 액세스하고 입력하기 위해 상기 호스트 서버를 액세스하기 위한 사용자 인터페이스;

[0113] - 입력된 데이터를 프로세싱하기 위한 프로세서; 및

[0114] - 상기 프로세서로 연결되는 메모리로서,

[0115] (i) 피검체 유래의 생물학적 시료에서 DNA 또는 RNA를 추출하여 차세대 염기서열 분석(next generation sequencing) 데이터를 액세스하기 위한 인스트럭션,

[0116] (ii) 상기 데이터로부터 상기 방법에 따라 결정된 HLA 타입에 액세스하기 위한 인스트럭션, 및

[0117] (iii) 프로세싱된 데이터가 저장된 메모리;를 포함하는 시스템을 제공한다.

[0118] 본 발명에서, 상기 개별화된 의료 개입은 추가적인 PCR-SBT 수행 여부를 결정하는 것일 수 있으나, 이에 제한되는 것은 아니다.

[0119] 본 발명에서, "개별화된 의료 개입"이란 사용자(또는 환자)의 특성을 고려하여 진단, 치료 계획, 수술 계획, 치료 방법 등을 결정 또는 판단하는 것을 의미한다.

[0120] 간결함을 위해, 종래의 데이터 네트워킹, 애플리케이션 개발 및 시스템의 그 밖의 다른 기능적 양태(그리고 시스템의 개별적인 운영 구성요소의 구성요소)는 생략될 수 있다. 아울러, 본 발명에 포함된 다양한 도면에서 나타난 연결 선은 다양한 요소들 간의 바람직한 기능적 관계 및/또는 물리적 결합을 나타내기 위함이다. 다수의 대안적, 또는 추가적 기능적 관계, 또는 물리적 결합이 실제 시스템에서 제공될 수 있을 것이다.

[0121] 본 발명에서 논의되는 다양한 시스템 구성요소는, 디지털 데이터를 프로세싱하기 위한 프로세서를 포함하는 호스트 서버, 또는 그 밖의 다른 컴퓨팅 시스템과, 디지털 데이터를 저장하기 위해 프로세서로 연결되는 메모리와, 상기 프로세서로 연결되며, 디지털 데이터를 입력하기 위한 입력 디지털타이저(input digitizer)와, 메모리에 저장되고 프로세서에 의해 액세스될 수 있으며, 상기 프로세서에 의한 디지털 데이터의 프로세싱을 지시하기 위한 애플리케이션 프로그램과, 상기 프로세서와 메모리로 연결되며, 상기 프로세서에 의해 프로세싱된 디지털 데이터로부터 추출된 정보를 디스플레이하기 위한 디스플레이 장치와, 다수의 데이터베이스 중 하나 이상을 포함할 수 있다. 본 발명에서 사용되는 다양한 데이터베이스는, 가족 병력, 인구통계적 및 환경적 데이터, 생체 시료 데이터, 이전 치료 및 프로토콜 데이터, 환자 임상 데이터, 생체 시료의 분자 프로파일링 데이터, 치료약제(및/또는 임상실험용 약제)에 따른 데이터, 유전자 라이브러리(gene library), 질병 라이브러리, 약 라이브러리, 환자추적 데이터(patient tracking data), 파일 관리 데이터, 재정 관리 데이터, HLA 타입 분석 데이터, 청구 데이터 및/또는 시스템의 운영에 유용한 이와 유사한 데이터 등의 환자 데이터를 포함할 수 있다. 당해업계 종사자라면, 사용자 컴퓨터가 운영 체제(가령, Windows, OS/2, UNIX, Linux, Solaris, MacOS 등) 뿐 아니라, 컴퓨터와 연계되는 다양한 종래의 지원 소프트웨어 및 드라이버를 포함할 수 있음을 인지할 것이다. 상기 컴퓨터는 임의의 적합한 개인용 컴퓨터, 네트워크 컴퓨터, 워크스테이션, 미니컴퓨터, 메인프레임 등을 포함할 수 있다. 사용자 컴퓨터는 네트워크로의 액세스를 갖는 가정, 또는 의료/사업 환경에 위치할 수 있다. 바람

직한 실시예에서, 액세스는 상업적으로 이용가능한 웹-브라우저 소프트웨어 패키지를 통한 네트워크, 또는 인터넷을 통해 이뤄진다.

[0122] 본 발명에서, "송신(transmit)"은 전자 데이터를 하나의 시스템 구성요소에 네트워크 연결을 통해 또 하나의 시스템 구성요소로 보내는 것을 포함한다. 덧붙이자면, 본 발명에서 사용될 때, "데이터(data)"는 커맨드, 질의(query), 파일, 저장을 위한 데이터 및 디지털이나 그 밖의 다른 임의의 형태로 된 유사한 것 등의 포괄적인 정보를 포함할 수 있다.

[0123] 상기 시스템은 웹 서비스, 유틸리티 컴퓨팅, 편재형 및 개인화 컴퓨팅, 보안 및 신원 솔루션, 자동화 컴퓨팅, 코모디티 컴퓨팅(commodity computing), 이동 및 무선 솔루션, 오픈 소스, 생체측정, 그리드 컴퓨팅(grid computing), 메쉬 컴퓨팅 중 하나 이상과 연계되는 사용을 고려한다.

[0124] 본 발명에서 설명되는 임의의 데이터베이스는 관계형, 계층형, 그래픽형, 또는 객체 지향형 구조 및 임의의 그 밖의 다른 데이터 구조 중 하나 이상을 포함할 수 있다. 데이터베이스를 구현하기 위해 사용될 수 있는 통상적인 데이터베이스 제품으로는 IBM(뉴욕, 화이트 플레인즈 소재)의 DB2, Oracle사(캘리포니아, 레드우드 쇼어즈 소재)의 다양한 데이터베이스 제품들, Microsoft사(워싱턴, 레드몬드 소재)의 Microsoft Access, 또는 Microsoft SQL, 또는 그 밖의 다른 임의의 적합한 데이터베이스 제품이 있다. 덧붙이자면, 데이터베이스는 임의의 적합한 방식으로, 가령 데이터 테이블이나 룩업 테이블(lookup table)로써 조직될 수 있다. 각각의 레코드는 하나의 단일 파일, 또는 일련의 파일, 또는 링크된 일련의 데이터 필드, 또는 그 밖의 다른 임의의 데이터 구조일 수 있다. 특정 데이터의 연계는 임의의 요망 데이터 연계 기법, 예를 들면, 당해업계에서 알려지거나 실시되는 기법을 통해 이뤄질 수 있다. 예를 들어, 이러한 연계는 수동으로, 또는 자동으로 이뤄질 수 있다. 자동 연계 기법은, 예를 들어, 검색, 모든 테이블과 파일에 대한 순차 검색(sequential search), 룩업을 단순화시키기 위해 알려진 순서에 따라 파일의 레코드를 정렬하는 것 등 중 하나 이상의 속도를 빠르게 하기 위해 테이블의 키 필드(key field)를 이용하는, 데이터베이스 검색, 데이터베이스 통합, GREP, AGREP, SQL을 포함할 수 있다. 연계 단계는 데이터베이스 통합 기능, 가령 미리 선택된 데이터베이스, 또는 데이터 섹터에서의 "키 필드(key field)"를 이용함으로써 이뤄질 수 있다.

[0125] 당업계 종사자라면, 보안상의 이유로, 임의의 데이터베이스, 시스템, 장치, 서버, 또는 시스템의 그 밖의 다른 구성요소가 하나의 단일 위치에서, 또는 다수의 위치에서 임의의 조합을 구성할 수 있으며, 여기서 각각의 데이터베이스, 또는 시스템은 다양한 적합한 보안 특징, 가령, 방화벽, 액세스 코드, 암호화, 복호화, 압축, 압축-해제 등 중 하나 이상을 포함한다는 것을 인지할 것이다.

[0126] 본 발명에서 설명되는 통신, 입력, 저장, 데이터베이스, 또는 디스플레이 중 임의의 것은 웹 페이지를 갖는 웹사이트를 통해 촉진될 수 있다. 용어 "웹 페이지"가 본 발명에서 사용될 때, 사용자와 상호작용하기 위해 사용될 수 있는 문서와 애플리케이션의 타입을 제한하는 것을 의미하지 않는다. 예를 들어, 통상의 웹사이트는 표준 HTML 문서에 추가로, 다양한 형식, 가령, Java 애플릿, JavaScript, ASP (active server page), CGI (common gateway interface) script, XML (extensible markup language), 동적 HTML, CSS (cascading style sheet), 응용 소프트웨어(helper application), 플러그-인 등을 포함할 수 있다. 서버는 웹 서버로부터의 요청을 수신하는 웹 서비스를 포함할 수 있으며, 상기 요청은 URL(http://yahoo.com/stockquotes/ge)과 IP 어드레스(123.56.789.234)를 포함할 수 있다. 상기 웹 서버는 적정한 웹 페이지를 불러와서 상기 웹 페이지에 대한 데이터, 또는 애플리케이션을 IP 어드레스로 전송한다. 웹 서비스는 통신 수단, 가령 인터넷을 통해 다른 애플리케이션과 상호 대화할 수 있는 애플리케이션이다. 웹 서비스는 통상적으로, XML, XSLT, SOAP, WSDL 및 USSI 등의 표준이나 프로토콜을 기반으로 한다. 웹 서비스 방법은 당해업계에서 잘 알려져 있으며, 많은 표준 문서에서 커버된다. 예를 들어, ALEX NGHIEM, IT WEB SERVICE: A ROADMAP FOR THE ENTERPRISE (2003)를 참조하라. 상기 문서는 본 발명에서 참조로써 인용된다.

[0127] 본 발명에서, 메모리는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(RAM, Random Access Memory), SRAM (Static Random Access Memory), 롬(ROM, Read-Only Memory), EEPROM (Electrically Erasable Programmable Read-Only Memory), PROM (Programmable Read-Only Memory), 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입일 수 있으나, 이에 제한되는 것은 아니다.

[0128] 본 발명에서, 네트워크는 모든 통신 방식을 이용하여 설립 (또는 형성)된 연결을 의미한다. 통신 수단은 소정의 통신 규격, 소정의 주파수 대역, 소정의 프로토콜 또는 소정의 채널을 통한 통신, 근거리 통신, 원거리 통신, 무선 통신 및 유선 통신을 포함한다. 예를 들어, 통신 수단은 블루투스, BLE, Wi-Fi, Zigbee, 3G, LTE, 초음파

를 통한 통신 수단 등을 포함할 수 있으나, 이에 제한되지 않는다.

- [0129] 근거리 통신 수단은, 통신을 수행하는 디바이스들이 소정의 범위 내에 있을 때에만 서로 통신이 가능한 통신 수단을 의미할 수 있으며, 예를 들어, 블루투스 및 NFC 를 포함할 수 있으나, 이에 제한되지 않는다.
- [0130] 원거리 통신 수단은, 통신을 수행하는 디바이스들이 거리와 관계없이 서로 통신이 가능한 통신 수단을 의미할 수 있다. 예를 들어, 원거리 통신 수단은 AP와 같은 중계기를 통해 통신을 수행하는 두 디바이스가 소정의 거리 이상일 때에도 통신할 수 있는 수단, 및 문자 전송과 전화 통화에 사용되는 셀룰러 네트워크(3G, LTE)를 이용한 통신 수단을 포함할 수 있으나, 이에 제한되지 않는다.
- [0131] 본 발명에서, 인터페이스는 블루투스 통신 인터페이스, BLE (Bluetooth Low Energy) 통신 인터페이스, 근거리 자기장 통신 인터페이스(Near Field Communication), 와이파이(wireless fidelity: WIFI) 통신 인터페이스, 지그비 (Zigbee) 통신 인터페이스, 적외선(infrared Data Association: IrDA) 통신 인터페이스, WFD (Wi-Fi Direct) 통신 인터페이스, UWB (ultra wideband) 통신 인터페이스, Ant+ 통신 인터페이스, 3G 통신 인터페이스, LTE 통신 인터페이스, 및 5G 통신 인터페이스일 수 있으나, 이에 제한되지 않는다.
- [0132] 본 발명의 시스템은 디스플레이를 더 포함할 수 있으나, 이에 제한되는 것은 아니다. 본 발명에서 디스플레이는 프로세서에 의해 처리되는 정보를 가시화 (visualization) 하는 구성요소이다. 프로세서에 의해 디스플레이되는 정보는 디스플레이(2040)를 통해 가시화될 수 있다. 디스플레이는 액정 디스플레이(liquid crystal display), 박막 트랜지스터 액정 디스플레이(thin film transistor-liquid crystal display), 유기 발광 다이오드 (organic light-emitting diode), 플렉시블 디스플레이 (flexible display), 3차원 디스플레이(3D display), 전기영동 디스플레이 (electrophoretic display) 중에서 적어도 하나를 포함할 수 있다. 디스플레이와 터치패드가 레이어 구조를 이루어 터치 스크린으로 구성되는 경우, 디스플레이는 입력 인터페이스로도 기능할 수 있다.
- [0133] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU (arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA (field programmable gate array), PLU (programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(operating system: OS) 및 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 어플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소 (processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서 (parallel processor) 와 같은, 다른 처리 구성 (processing configuration) 도 가능하다.
- [0134] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embodiment)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.
- [0135] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk) 와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어

장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로써 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

- [0136] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.
- [0137] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.
- [0138] 본 발명에서 시스템 및 방법이 다양한 실시예에 따르는 방법, 장치(가령, 시스템) 및 컴퓨터 프로그램 프로덕트에 대한 스크린 샷, 블록 다이어그램 및 흐름도를 참조하여 서술된다. 블록 다이어그램 및 흐름도의 각각의 기능적 블록 및 기능적 블록들의 조합은 컴퓨터 프로그램 인스트럭션에 의해 구현될 수 있다.
- [0139] 본 실시예는 기능적인 블록 구성들 및 다양한 처리 단계들로 나타내어질 수 있다. 이러한 기능 블록들은 특정 기능들을 실행하는 다양한 개수의 하드웨어 또는/및 소프트웨어 구성들로 구현될 수 있다. 예를 들어, 실시예는 하나 이상의 마이크로프로세서들의 제어 또는 다른 제어 장치들에 의해서 다양한 기능들을 실행할 수 있는, 메모리, 프로세싱, 로직(logic), 룩 업 테이블(look-up table) 등과 같은 직접 회로 구성들을 채용할 수 있다. 구성 요소들이 소프트웨어 프로그래밍 또는 소프트웨어 요소들로 실행될 수 있는 것과 유사하게, 본 실시예는 데이터 구조, 프로세스들, 루틴들 또는 다른 프로그래밍 구성들의 조합으로 구현되는 다양한 알고리즘을 포함하여, C, C++, 자바(Java), 어셈블러(assembler) 등과 같은 프로그래밍 또는 스크립팅 언어로 구현될 수 있다. 기능적인 측면들은 하나 이상의 프로세서들에서 실행되는 알고리즘으로 구현될 수 있다. 또한, 본 실시예는 전자적인 환경 설정, 신호 처리, 및/또는 데이터 처리 등을 위하여 종래 기술을 채용할 수 있다. "매커니즘", "요소", "수단", "구성"과 같은 용어는 넓게 사용될 수 있으며, 기계적이고 물리적인 구성들로서 제한되는 것은 아니다. 상기 용어는 프로세서 등과 연계하여 소프트웨어의 일련의 처리들(routines)의 의미를 포함할 수 있다.
- [0140] 도 5는 개별화된 의료 개입을 결정하기 위한 시스템(10)의 바람직한 하나의 실시예의 블록 다이어그램이다. 시스템(10)은 사용자 인터페이스(12)와, 데이터를 프로세싱하기 위한 프로세서(16)를 포함하는 호스트 서버(14)와, 상기 프로세서로 연결되는 메모리(18)와, 상기 메모리(18)에 저장되며 상기 프로세서(16)에 의해 액세스가능하고 상기 프로세서(16)에 의한 데이터 프로세싱을 지시하기 위한 애플리케이션 프로그램(20)과, 다수의 내부 데이터베이스(22) 및 외부 데이터베이스(24)와, 유선(또는 무선) 통신 네트워크(26)(가령, 인터넷)와의 인터페이스를 포함한다. 또한 시스템(10)은 사용자 인터페이스(12)로부터 수신된 데이터로부터의 디지털 데이터를 입력하기 위해 프로세서(16)로 연결되는 입력 디지털타이저(input digitizer, 28)를 포함할 수 있다.
- [0141] 사용자 인터페이스(12)는 데이터를 시스템(10)으로 입력하고 상기 프로세서(16)에 의해 프로세싱된 데이터로부터 추출된 정보를 디스플레이하기 위한 입력 장치(30)와 디스플레이(32)를 포함한다. 또한 사용자 인터페이스(12)는 상기 프로세서(16)에 의해 프로세싱된 데이터로부터 추출된 정보, 가령, 표적에 대한 테스트 결과와 상기 테스트 결과를 기초로 하여 제안된 약 처방을 포함할 수 있는 환자 보고서를 인쇄하기 위한 프린터(34)를 포함할 수 있다.
- [0142] 내부 데이터베이스(22)는 환자의 생체 시료/표본 정보, 추적, 임상 데이터, 환자 데이터, 환자 추적, 파일 관리, 연구 프로토콜, 분자 프로파일링에서의 환자 테스트 결과 및 청구 정보 및 추적을 포함할 수 있다(그러나 이에 제한받지 않음). 외부 데이터베이스(24)는 예를 들어, 약 라이브러리(drug library), 유전자 라이브러리, 질병 라이브러리, 공공 및 사설 데이터베이스(가령, UniGene, OMIM, GO, TIGR, GeneBank, KEGG 및 Biocarta), PCR 방법, HLA 타입 결정 TOOL을 포함할 수 있다(그러나 이에 제한받지 않음).
- [0143] 다양한 방법이 시스템(10)에 따라서 사용될 수 있다. 도 6은 인간 백혈구 항원(Human Leukocyte Antigen; HLA) 타입 결정 방법 및 이를 이용한 의료 개입을 결정하기 위한 방법의 하나의 바람직한 실시예에 대한 흐름도를 나타낸다.
- [0144] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유

사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0145] 본 명세서에서 설명하는 특정 실행들은 예시들으로써, 어떠한 방법으로도 기술적 범위를 한정하는 것은 아니다. 명세서의 간결함을 위하여, 종래 전자적인 구성들, 제어 시스템들, 소프트웨어, 상기 시스템들의 다른 기능적인 측면들의 기재는 생략될 수 있다. 또한, 도면에 도시된 구성 요소들 간의 선들의 연결 또는 연결 부재들은 기능적인 연결 및/또는 물리적 또는 회로적 연결들을 예시적으로 나타낸 것으로써, 실제 장치에서는 대체 가능하거나 추가의 다양한 기능적인 연결, 물리적인 연결, 또는 회로 연결들로서 나타내어질 수 있다.

[0146] 본 발명에서의 발명 개념을 벗어남이 없이 이미 기재된 것들 이외의 더 많은 변형이 가능하다는 것이 당업자에게 명백할 것이다. 따라서, 본 발명의 대상은 첨부된 청구범위를 제외하고는 제한되지 않아야 한다. 더욱이, 명세서 및 청구범위 모두를 해석함에 있어서, 모든 용어는 문맥과 일치하는 가능한 가장 넓은 방식으로 해석되어야 한다. 특히, 용어 "포함하다" 및 "포함하는"은 비-배타적 방식으로 요소, 성분, 또는 단계를 지칭하는 것으로 해석되어야 하며, 이는 지칭된 요소, 성분, 또는 단계가 존재할 수 있거나, 이용될 수 있거나, 명시적으로 지칭되지 않은 다른 요소, 성분 또는 단계와 조합될 수 있음을 나타낸다.

[0148] 이하, 본 발명의 이해를 돕기 위하여 바람직한 실시예를 제시한다. 그러나 하기의 실시예는 본 발명을 보다 쉽게 이해하기 위하여 제공되는 것일 뿐, 하기 실시예에 의해 본 발명의 내용이 한정되는 것은 아니다.

[실시예]

실험방법

1. 1000 유전체 프로젝트의 NGS 데이터 세트

[0153] 알고리즘 도구의 예측 성능을 조사하기 위해 먼저 1000 유전체 프로젝트(<https://www.1000genomes.org>)에서 총 2,693명으로부터 NGS 기반의 WES 데이터를 얻었다. 또한, PolyPheMe 소프트웨어를 기반으로 한 HLA 클래스 I 유전자(HLA-A, -B, -C)에 대해 4자리 해상도의 추가 HLA 유전자형 정보를 얻어 HLA 유전자형 참조로 사용하였다. 모든 6개의 HLA 대립 유전자(HLA-A, -B 및 -C 각각에 대해 2개의 대립 유전자 x 3가지 유형)가 있는 샘플만 포함시켰다. 03/01/05/06/50과 같은 다중 call이 존재하는 샘플은 그 모호한 결과가 정확한 대립 유전자 예측이 필요한 향후 임상 적용에 적합하지 않기 때문에 사용하지 않았다. 다중 call 샘플을 제외하고 2,619개의 샘플을 사용하였다.

[0154] 2,619개 샘플의 시퀀싱 데이터는 Illumina의 플랫폼(HiSeq 2000, 2,407건, Genome Analyzer II, 212건)에서 얻었다. paired-end reads 길이는 76bp (819건), 77bp (1건), 79bp (18건), 90bp (506건), 100bp (843건), 101bp (432건)이었다.

2. 알고리즘의 HLA 유전형 분석 성능 평가

[0157] 1000 유전체 프로젝트에서 NGS를 기반으로 한 2,619명의 개인에 대한 WES 데이터를 사용하여 HLA-A, -B 및 -C에 대한 HLA 유전형 분석을 수행하였다. 알고리즘은 OptiType, HLA-HD, PHLAT, seq2HLA, arcashLA, HLAscan, HLA*LA 및 Kourami, 8가지 사용 가능한 알고리즘으로 각 톨의 지침에 따라 수행하였다.

[0158] 표 1에 나타난 바와 같이, 입력 데이터로써 FASTQ 파일은 OptiType, HLA-HD, PHLAT 및 seq2HLA에 사용되었으며, BAM 파일은 HLAscan, arcashLA, HLA*LA 및 Kourami에 사용되었다. HLA 유전형 분석에 대한 생성 call은 4자리 해상도로 평가되었으며 대립 유전자의 HLA 유전자형에 대한 call이 다중 call 없이 최소 4자리 해상도로 생성된 경우 사용 가능한 대립 유전자로 분류하였다. 다중 call의 경우에는 사용할 수 없는 것으로 간주하였다.

표 1

[0159]

Program	URL	Read type	MHC class	Resolution	Input format
OptiType	https://github.com/FRED-2/OptiType	WGS/WES/RNA-Seq	Class I	2 fields	FASTQ
HLA-HD	https://www.genome.med.kyoto-u.ac.jp/HLA-HD/	WGS/WES/RNA-Seq	Class I and II	3 fields	FASTQ
PHLAT	https://sites.google.com/site/projectphlat/Downloads	WES/RNA-Seq	Class I and II	3 fields	FASTQ
seq2HLA	https://github.com/TRON-Bioinformatics/seq2HLA	RNA-Seq	Class I and II	2 fields	FASTQ
arcashLA	https://github.com/RabadanLab/arcashLA	RNA-Seq	Class I and II	3 fields	BAM
HLAScan	http://genomekorea.com/display/tools/HLAScan	WGS/WES	Class I and II	4 fields	FASTQ / BAM
HLA*LA	https://github.com/DiltheyLab/HLA*LA	WGS, WES	Class I and II	3 fields	BAM / CRAM
Kourami	https://github.com/Kingsford-Group/Kourami	WGS	Class I and II	3 fields	BAM / CRAM

[0161]

그런 다음 HLA 유전형 분석을 위해 생성된 call을 1000 유전체 프로젝트의 HLA 일배체형(haplotype) 참조와 비교하고 4자리 해상도에서 정확도를 분석하였다.

[0162]

도 1a에 나타난 바와 같이, 1000 유전체 프로젝트를 참조로 하여, 일치하는 두 대립 유전자가 있는 샘플 수를 사용 가능한 두 대립 유전자가 있는 총 샘플 수로 나눈 값을 계산하여, HLA-A, -B 및 -C 각각에 대한 정확도를 정의하였다.

[0164]

총 2,619건의 사례에서 가용성과 정확도를 평가하였다. 또한, 최소 100bp paired-end read 길이로 1000 유전체 프로젝트 데이터만 분리하고 추가 분석을 수행하였다 (도 1b).

[0166]

3. 자체 샘플 준비

[0167]

다양한 암 환자의 39개 샘플로 구성된 사내 데이터(유방암 5건, 담관암 15건, 결장 직장암 4건, 폐암 5건, 위암 5건, 육종 5건)는 말초혈액 단핵세포 또는 포르말린 고정 파라핀 포매 조직으로부터 gDNA를 추출하였다. 추출된 gDNA는 SureSelectXT Library Prep kit를 이용하여 Illumina Hiseq 2500으로 101-bp paired-end read를 생산하여 WES 데이터를 얻었다. 또한, PCR-SBT를 수행하여 HLA 타이핑도 수행하였다.

[0168]

위에 기술한 모든 샘플은 서면 동의와 함께 기관 검토위원회 승인 하에 획득되었다.

[0170]

실시예 1. 종래 HLA 타이핑 툴의 가용성 및 정확도 확인

[0171]

HLA-A, HLA-B 및 HLA-C의 타이핑 과정에서, 총 2,619건의 4자리 해상도에 대하여, 8개의 HLA 타이핑 툴(OptiType, HLA-HD, PHLAT, seq2HLA, arcashLA, HLAScan, HLA*LA 및 Kourami)의 가용성 및 정확도를 계산하였다. 이 중 Kourami만 다중 call을 하였으며, 그 케이스 수는 HLA-A 398건, HLA-B 194건, HLA-C 233건이었다.

[0172]

도 1a에 나타난 바와 같이, HLA-A에 대한 HLA 유전형 분석의 가용성은 다음과 같은 순서로 이루어졌다(가장 높은 순서에서 낮은 순서로): OptiType (2,604/2,619, 99.4%), seq2HLA (2,567/2,619, 98.0%), HLA-HD (2,506/2,619, 95.7%), arcashLA (2,468/2,619, 94.2%), PHLAT (2,412/2,619, 92.1%), HLA*LA (2,385/2,619, 91.1%), Kourami (2,097/2,619, 80.1%), 및 HLAScan (1,591/2,619, 60.7%). HLA-A의 HLA 유전형 분석을 위한 가장 정확한 세 가지 툴은 OptiType (2,362/2,604, 90.7%), HLA*LA (1,976/2,385, 82.9%), 및 HLAScan (1,249/1,591, 78.5%) 이었다. OptiType은 다른 툴보다 HLA-A에 특이적으로 우수한 성능을 보이며 가장 높은

정확도와 가용성을 나타내었다.

[0173] 또한, HLA-B에 대한 HLA 유전형 분석의 가용성은 다음과 같은 순서로 이루어졌다(가장 높은 순서에서 낮은 순서로) : OptiType (2,611/2,619, 99.7%), seq2HLA (2,537/2,619, 96.9%), HLA-HD (2,520/2,619, 96.2%), PHLAT (2,491/2,619, 95.1%), arcasHLA (2,468/2,619, 94.2%), HLA*LA (2,356/2,619, 90.0%), Kourami (2,082/2,619, 79.5%), 및 HLAscan (1,622/2,619, 61.9%). HLA-B의 HLA 유전형 분석을 위한 가장 정확한 세 가지 툴은 HLA*LA (2,121/2,356, 90.0%), OptiType (2,261/2,611, 86.6%), 및 Kourami (1,649/2,082, 79.2%) 이었다.

[0174] 또한, HLA-C에 대한 HLA 유전형 분석의 가용성은 다음과 같은 순서로 이루어졌다(가장 높은 순서에서 낮은 순서로) : OptiType (2,612/2,619, 99.7%), seq2HLA (2,580/2,619, 98.5%), HLA-HD (2,516/2,619, 96.1%), PHLAT (2,483/2,619, 94.8%), arcasHLA (2,468/2,619, 94.2%), HLA*LA (2,344/2,619, 89.5%), Kourami (2,102/2,619, 80.3%), 및 HLAscan (1,614/2,619, 61.6%). 이는 HLA-B 유전형 분석 결과와 동일한 순서이며 유사한 가용성을 나타내었다. HLA-C의 HLA 유전형 분석을 위한 가장 정확한 세 가지 툴은 HLA*LA (2,075/2,344, 88.5%), OptiType (2,261/2,612, 86.6%), 및 PHLAT (2,110/2,483, 85.0%) 이었다.

[0176] 실시예 2. 리드 길이 100bp 이상인 케이스에서의 HLA 타이핑 툴의 가용성 및 정확도 확인

[0177] NGS 데이터의 최소 리드 길이가 100bp 이상인 경우만 선택하였으며, 그 결과 1,275건이 선별되었다. 4자리 해상도와 가용성에서 이러한 경우의 정확도를 추가로 분석하였다.

[0178] 도 1b에 나타난 바와 같이, HLA-A에 대한 HLA 유전형 분석의 가용성은 다음과 같은 순서로 이루어졌다(가장 높은 순서에서 낮은 순서로) : OptiType (1,275/1,275, 100.0%), seq2HLA (1,269/1,275, 99.5%), HLA-HD (1,265/1,275, 99.2%), PHLAT (1,238/1,275, 97.1%), arcasHLA (1,150/1,275, 90.2%), HLAscan (1,130/1,275, 88.6%), HLA*LA (1,100/1,275, 86.3%), 및 Kourami (974/1,275, 76.4%). HLA-A의 HLA 유전형 분석을 위한 가장 정확한 5가지 툴은 OptiType (1,232/1,275, 96.6%), HLA*LA (912/1,100, 82.9%), HLA-HD (1,044/1,265, 82.5%), HLAscan (911/1,130, 80.6%), 및 Kourami (732/974, 75.2%)이었다.

[0179] 또한, HLA-B에 대한 HLA 유전형 분석의 가용성은 다음과 같은 순서로 이루어졌다(가장 높은 순서에서 낮은 순서로) : OptiType (1,275/1,275, 100.0%), HLA-HD (1,269/1,275, 99.5%), PHLAT (1,269/1,275, 99.5%), seq2HLA (1,251/1,275, 98.1%), arcasHLA (1,150/1,275, 90.2%), HLAscan (1,127/1,275, 88.4%), HLA*LA (1,086/1,275, 85.2%), 및 Kourami (882/1,275, 69.2%)이었다.. HLA-B의 HLA 유전형 분석을 위한 가장 정확한 5가지 툴은 OptiType (1,203/1,275, 94.4%), HLA*LA (919/1,086, 84.6%), HLA-HD (1,068/1,269, 84.2%), PHLAT (1,067/1,269, 84.1%), 및 Kourami (686/882, 77.8%) 이었다.

[0180] 또한, HLA-C에 대한 HLA 유전형 분석의 가용성은 다음과 같은 순서로 이루어졌다(가장 높은 순서에서 낮은 순서로) : OptiType (1,275/1,275, 100.0%), seq2HLA (1,274/1,275, 99.9%), PHLAT (1,268/1,275, 99.5%), HLA-HD (1,257/1,275, 98.6%), arcasHLA (1,150/1,275, 90.2%), HLAscan (1,126/1,275, 88.3%), HLA*LA (1,081/1,275, 84.8%), 및 Kourami (911/1,275, 71.5%)이었다.

[0181] HLA-C의 HLA 유전형 분석을 위한 가장 정확한 5가지 툴은 OptiType (1,172/1,275, 91.9%), PHLAT (1,154/1,268, 91.0%), HLA-HD (1,077/1,257, 85.7%), HLA*LA (921/1,081, 85.2%), 및 HLAscan (937/1,126, 83.2%)이었다.

[0182] 실시예 1 및 2에서 OptiType은 다른 툴보다 성능이 뛰어났으며 세 가지 HLA 타입 모두에서 가장 높은 정확도와 가장 높은 가용성을 보여주었다. 한편, 나머지 툴들은 각 HLA 유형에 대해 툴의 가용성과 정확도의 순서는 거의 동일하였다.

[0184] 실시예 3. 5가지 툴을 사용한 새로운 알고리즘 개발

[0185] 실시예 1 및 2의 결과에 따라, 현재까지 사용 가능한 알고리즘을 결합하여 새로운 알고리즘을 개발하였다. 개발의 전제로써, 리드 길이가 100bp 이상인 경우의 가용성과 정확도를 고려하여 다중 call 및 최저 가용성을 보인 Kourami를 제외하였으며, 정확도가 가장 낮은 seq2HLA 및 arcasHLA를 제외하였다.

[0186] 이에, OptiType, HLA-HD, PHLAT, HLAscan 및 HLA*LA 5가지 툴을 결합하여 새로운 알고리즘을 개발하였다. 1000

유전체 프로젝트 사례를 5가지 툴의 call이 모두 사용 가능한지 여부에 따라 두 그룹으로 나누었다.

- [0187] 도 2a에 나타난 바와 같이, HLA-A에서는 5개 툴 모두에서 사용 가능한 call이 전체의 83.6% (1,066/1,275)를 차지하였다. 5개 툴 모두에서 사용 가능한 call이 있는 그룹에서는 37.9% (404/1,066)의 케이스에서 5개 툴의 call 사이에 일치점이 있었으며 정확도는 99.8% (403/404)였다. 5개 툴이 동일한 call을 제공하지 않는 나머지 케이스 중 57.3% (379/662)에서 4개의 call이 일치했으며, 일치한 call의 정확도는 98.7% (374/379)였다. 5개 툴 중 4개 이상에서 동일한 call을 얻지 못한 나머지 경우에는 73.9% (209/283)의 경우에서 3개 툴의 call이 일치하였으며 일치한 call의 정확도는 93.3% (195/209)였다.
- [0189] 5개 툴 모두에서 사용 가능한 call을 제시하지 않는 그룹의 경우 5개 툴 중 4개의 툴에서 사용가능한 call을 제시하는 경우는 39.7% (83/209) 이다.
- [0190] HLA-A에서 5개 툴 중 4개가 제시한 call의 일치도와 정확도는 다음과 같다 :
- [0191] OptiType, HLA-HD, PHLAT, 및 HLA*LA, 26.3% (5/19) 및 100% (5/5);
- [0192] OptiType, HLA-HD, HLAscan, 및 HLA*LA, 60.0% (3/5) 및 100% (3/3); 및
- [0193] OptiType, HLA-HD, PHLAT, 및 HLAscan, 52% (26/50) 및 100% (26/26).
- [0194] HLA-HD, PHLAT, HLAscan 및 HLA*LA call의 가용성은 0%였고 OptiType, PHLAT, HLAscan 및 HLA*LA call의 일치율은 0%였다.
- [0195] 4개 툴 call을 사용할 수 있지만 일치된 call을 제시하지 않은 나머지 케이스 중에서 4개 툴의 모든 조합에서 4개 툴 중 3개가 제공하는 call의 일치도와 정확도는 다음과 같다:
- [0196] OptiType, HLA-HD, PHLAT, 및 HLA*LA, 57.1% (8/14) 및 100% (8/8);
- [0197] OptiType, HLA-HD, HLAscan, 및 HLA*LA, 12.5% (1/8) 및 100% (1/1);
- [0198] OptiType, HLAscan, PHLAT, 및 HLA*LA, 55.6% (5/9) 및 100% (5/5); and OptiType, HLA-HD, PHLAT, 및 HLAscan, 79.2% (19/24) 및 100% (19/19).
- [0199] 5개 툴 중 4개 이상에서 call이 불가능한 나머지 126개 케이스 중 125개 케이스에 대해 5개 툴 중 3개에서는 call이 가능하였다.
- [0200] OptiType, HLA-HD 및 PHLAT에 대한 call의 일치도와 정확도는 각각 46.0% (57/124) 및 100% (57/57)였다. OptiType, PHLAT 및 HLA*LA의 일치도는 0% (0/1) 였고 5개 툴 중 3개 툴의 나머지 조합의 가용성은 0%였다.
- [0202] 도 2b에 나타난 바와 같이, HLA-B에서는 5개 툴 모두에서 사용 가능한 call을 제시하는 경우는 전체의 82.6% (1,053/1,275)를 차지하였다. 5개 툴 모두에서 사용 가능한 call이 있는 그룹에서는 44.4% (468/1,053)의 케이스에서 5개 툴의 call 사이에 일치점이 있었으며 정확도는 99.6% (466/468)였다. 5개 툴이 동일한 call을 제공하지 않는 나머지 케이스 중 58.6% (343/585)에서 4개의 call이 일치했으며 일치한 call의 정확도는 98.3% (337/343)였다. 5개 툴 중 4개 이상에서 동일한 call을 얻지 못한 나머지 경우에는 72.3% (175/242)의 경우에서 3개 툴의 call이 일치하였으며 일치한 call의 정확도는 92.6% (162/175)였다.
- [0203] 5개 툴 모두에서 사용 가능한 call을 제시하지 않는 그룹의 경우 5개 툴 중 4개의 툴에서 사용가능한 call을 제시하는 경우는 43.2% (96/222) 이다. HLA-HD, PHLAT, HLAscan 및 HLA*LA의 가용성은 0%였다.
- [0204] HLA-B에서 5개 툴 중 4개가 제시한 call의 일치도와 정확도는 다음과 같다 :
- [0205] OptiType, HLA-HD, PHLAT, 및 HLA*LA, 13.6% (3/22) 및 100% (3/3);
- [0206] OptiType, HLA-HD, HLAscan, 및 HLA*LA, 60.0% (3/5) 및 100% (3/3);
- [0207] OptiType, PHLAT, HLAscan, 및 HLA*LA, 40.0% (2/5) 및 100% (2/2); 및
- [0208] OptiType, HLA-HD, PHLAT, 및 HLAscan, 41.2% (27/64) 및 100% (27/27).
- [0209] 4개 툴 call을 사용할 수 있지만 일치된 call을 제시하지 않은 나머지 케이스 중에서 4개 툴의 모든 조합에서 4

개 툴 중 3개가 제공하는 call의 일치도와 정확도는 다음과 같다:

- [0210] OptiType, HLA-HD, PHLAT, 및 HLA*LA, 57.9% (11/19) 및 100% (11/11);
- [0211] OptiType, HLA-HD, HLAsan, 및 HLA*LA, 50.0% (1/2) 및 100% (1/1);
- [0212] OptiType, PHLAT, HLAsan, 및 HLA*LA, 100% (3/3) 및 100% (3/3); 및
- [0213] OptiType, HLA-HD, PHLAT, 및 HLAsan, 81.1% (30/37) 및 98.3% (28/30).
- [0214] 5개 툴 중 4개 이상에서 call이 불가능한 나머지 126개 케이스 중 125개 케이스에 대해 5개 툴 중 3개에서는 call이 가능하였다.
- [0215] OptiType, HLA-HD 및 PHLAT에 대한 call의 일치도와 정확도는 각각 79.8% (99/124) 및 98% (97/99)였다. OptiType, PHLAT 및 HLA*LA의 일치도는 0% (0/1) 였고 5개 툴 중 3개 툴의 나머지 조합의 가용성은 0%였다.
- [0217] 도 2c에 나타난 바와 같이, HLA-C에서는 5개 툴 모두에서 사용 가능한 call이 전체의 81.6% (1,040/1,275)를 차지하였다. 5개 툴 모두에서 사용 가능한 call이 있는 그룹에서는 58.3% (606/1,040)의 케이스에서 5개 툴의 call 사이에 일치가 있었으며 정확도는 99.8% (605/606)였다. 5개 툴이 동일한 call을 제공하지 않는 나머지 케이스 중 60.4% (262/434)에서 4개의 call이 일치했으며 일치한 call의 정확도는 97.3% (255/262)였다. 5개 툴 중 4개 이상에서 동일한 call을 얻지 못한 나머지 경우에는 79.1% (136/172)의 경우에서 3개 툴의 call이 일치하였으며 일치한 call의 정확도는 65.4% (89/136) 였다.
- [0218] 5개 툴 모두에서 사용 가능한 call을 제시하지 않는 그룹의 경우 5개 툴 중 4개의 툴에서 사용가능한 call을 제시하는 경우는 44.7% (105/235)이다. HLA-HD, PHLAT, HLAsan 및 HLA*LA의 가용성은 0%였다.
- [0219] HLA-C에서 5개 툴 중 4개가 제시한 call의 일치도와 정확도는 다음과 같다 :
- [0220] OptiType, HLA-HD, PHLAT, 및 HLA*LA, 20.0% (4/20) 및 100% (4/4);
- [0221] OptiType, HLA-HD, HLAsan, 및 HLA*LA, 60.0% (3/5) 및 100% (3/3);
- [0222] OptiType, PHLAT, HLAsan, 및 HLA*LA, 23.1% (3/13) 및 100% (3/3);
- [0223] OptiType, HLA-HD, PHLAT, 및 HLAsan, 65.7% (44/67) 및 100% (44/44).
- [0224] 4개 툴 call을 사용할 수 있지만 일치된 call을 제시하지 않은 나머지 케이스 중에서 4개 툴의 모든 조합에서 4개 툴 중 3개가 제공하는 call의 일치도와 정확도는 다음과 같다:
- [0225] OptiType, HLA-HD, PHLAT, 및 HLA*LA, 68.8% (11/16) 및 100% (11/11);
- [0226] OptiType, HLA-HD, HLAsan, 및 HLA*LA, 100% (2/2) 및 100% (2/2);
- [0227] OptiType, PHLAT, HLAsan, 및 HLA*LA, 50.0% (5/10) 및 100% (5/5);
- [0228] OptiType, HLA-HD, PHLAT, 및 HLAsan, 60.9% (14/23) 및 100% (14/14).
- [0229] 5개 툴 중 4개 이상에서 call이 불가능한 나머지 130개 케이스 중 128개 케이스에 대해 5개 툴 중 3개에서는 call이 가능하였다.
- [0230] OptiType, HLA-HD 및 PHLAT에 대한 call의 일치도와 정확도는 각각 83.1% (103/124) 및 100% (103/103)였다. OptiType, PHLAT 및 HLA*LA의 일치도는 33.3% (1/3) 였고 정확도는 100% (1/1)였다. OptiType, PHLAT 및 HLAsan의 일치도는 0%였고 5개 툴 중 3개 툴의 나머지 조합의 가용성은 0%였다.
- [0232] 본 발명자들은 상기 결과를 바탕으로 새로운 알고리즘을 구축하였다.
- [0233] 본 발명의 알고리즘은 1000 유전체 프로젝트 데이터에서 99% 이상의 정확도를 가진 사례를 기반으로 만들었다.
- [0235] 본 발명의 알고리즘은 5가지 툴이 제시한 call 중 사용 가능한 call 수에 따라 케이스를 세 그룹으로 나누는 것으로 시작한다 : 각각 5 calls, 4 calls, 3 calls 가능

- [0236] 5개의 call 모두 사용 가능한 경우, 5개의 call이 모두 일치했을 때 해당 call를 사용할 수 있다.
- [0237] 4개의 call만 사용 가능한 경우, 4개 call 중 4개 또는 3개가 일치하는 경우 해당 call들을 사용할 수 있다. 다만, HLA-B의 경우에는 4개 call들이 모두 일치하는 경우에만 해당 call들을 사용할 수 있다.
- [0238] 3개의 call만 사용 가능한 경우, 3개의 call이 일치하는 경우 해당 call들을 사용할 수 있다.
- [0240] 리드 길이가 100bp 이상인 1000 유전체 프로젝트 데이터의 1,275 세트에서, HLA-A 41.4% (528/1,275), HLA-B 39.5% (503/1,275), HLA-C 62.4% (796/1,275)가 상기 상황을 만족한 것으로 나타났다.
- [0242] **실시예 4. 자체 샘플에 본 발명의 알고리즘 적용 확인**
- [0243] 39개의 자체 샘플을 사용하여 새로운 알고리즘의 성능을 평가하였다.
- [0244] HLA-A, -B, -C의 39건 모두 PCR-SBT에 의해 실험적으로 얻은 대립 유전자 검출(call)이 가능하였다. 하지만 HLA-B에서 1개의 allele가 08:New로 call이 되어 추가 알고리즘 적용에는 제외하였다.
- [0245] 도 3에 나타난 바와 같이, 모든 케이스에서 모든 OptiType, HLA-HD, PHLAT, HLAsan 및 HLA*LA에서 call이 가능하였다.
- [0246] OptiType, HLA-HD, PHLAT, HLAsan 및 HLA*LA의 HLA-A에 대한 정확도는 각각 94.9%, 100%, 84.6%, 94.9% 및 87.2%였다.
- [0247] OptiType, HLA-HD, PHLAT, HLAsan 및 HLA*LA의 HLA-B에 대한 정확도는 각각 97.4%, 100%, 97.4%, 76.3% 및 100%였다. 마지막으로 OptiType, HLA-HD, PHLAT, HLAsan 및 HLA*LA에 대한 HLA-C의 정확도는 각각 92.3%, 89.7%, 89.7% 79.5% 및 100%였다.
- [0249] 도 4a에 나타난 바와 같이, HLA-A에서 5가지 call의 일치도는 69.2% (27/39)이고 일치한 call의 정확도는 100% (27/27)였다.
- [0250] 도 4b에 나타난 바와 같이, HLA-B의 경우 일치도는 71.1% (27/38), 일치한 call의 정확도는 100% (27/27)였다.
- [0251] 도 4c에 나타난 바와 같이, HLA-C의 경우 일치도는 76.9% (30/39), 일치한 call의 정확도는 100% (30/30)였다.
- [0253] **실시예 5. HLA 유전자 타이핑을 위한 알고리즘 제작**
- [0254] 상기 결과를 바탕으로 HLA 유형을 선택하기 위한 최적화된 지침을 개발하였으며, 아래와 같다 :
- [0255] 1. 케이스에 5가지 툴(OptiType, HLA-HD, PHLAT, HLAsan 및 HLA*LA) 중에서 몇 개의 툴에서 call이 제시되었는지 확인한다.
- [0256] 2. 5개 툴 모두에서 call이 제시되었을 경우, 5개 툴에서 일치된 HLA 타입을 제시한 경우만 그 타입을 HLA타입으로 선택한다.
- [0257] 3-1. 5개 툴 중에서 4개의 툴에서만 HLA 타입을 제시하였을 경우 HLA-A 및 HLA-C에서 3개 이상의 툴이 동일하게 제시하는 HLA타입을 선택한다.
- [0258] 3-2. HLA-B에서는 4개의 툴이 일치된 HLA 타입을 제시할 경우만 그 타입을선택한다.
- [0259] 4. 5개 툴 중에서 3개의 툴에서만 HLA타입을 제시하는 경우 3개의 툴에서 일치된 HLA 타입을 제시할 경우만 그 타입을 선택한다.
- [0260] 5. 2번, 3번, 4번이 적용되지 않는 경우, PCR-SBT를 통해 실험적으로 HLA 유형을 결정한다.
- [0262] 이상의 실시예를 종합하여, 본 발명자들은 1000 유전체 프로젝트 데이터를 이용하여 HLA타입을 결정할 때 OptiType, HLA-HD, PHLAT, HLAsan 및 HLA*LA 툴 모두를 사용하는 경우, HLA Class I 타이핑의 정확도가 현저

히 향상됨을 확인하였는 바, 신뢰도가 향상된 최적의 HLA 타입 결정 방법임을 알 수 있었다. 따라서, 본 발명의 HLA 타입 결정 방법을 이용하여 이식 성공률을 높이기 위한 조직 적합성 테스트뿐만 아니라, PCR-SBT 검사법의 추가 수행 여부 판단 등 의료 개인화를 가능하게 할 수 있을 것으로 기대된다.

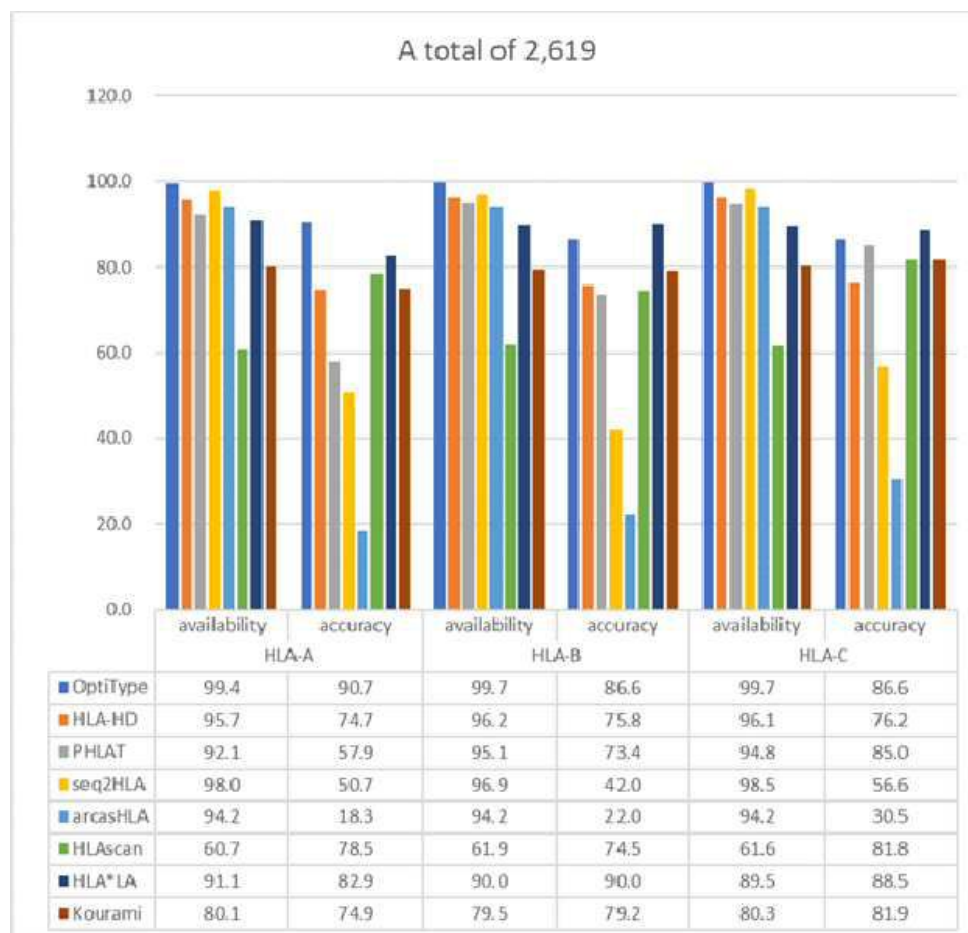
[0263]

[0264]

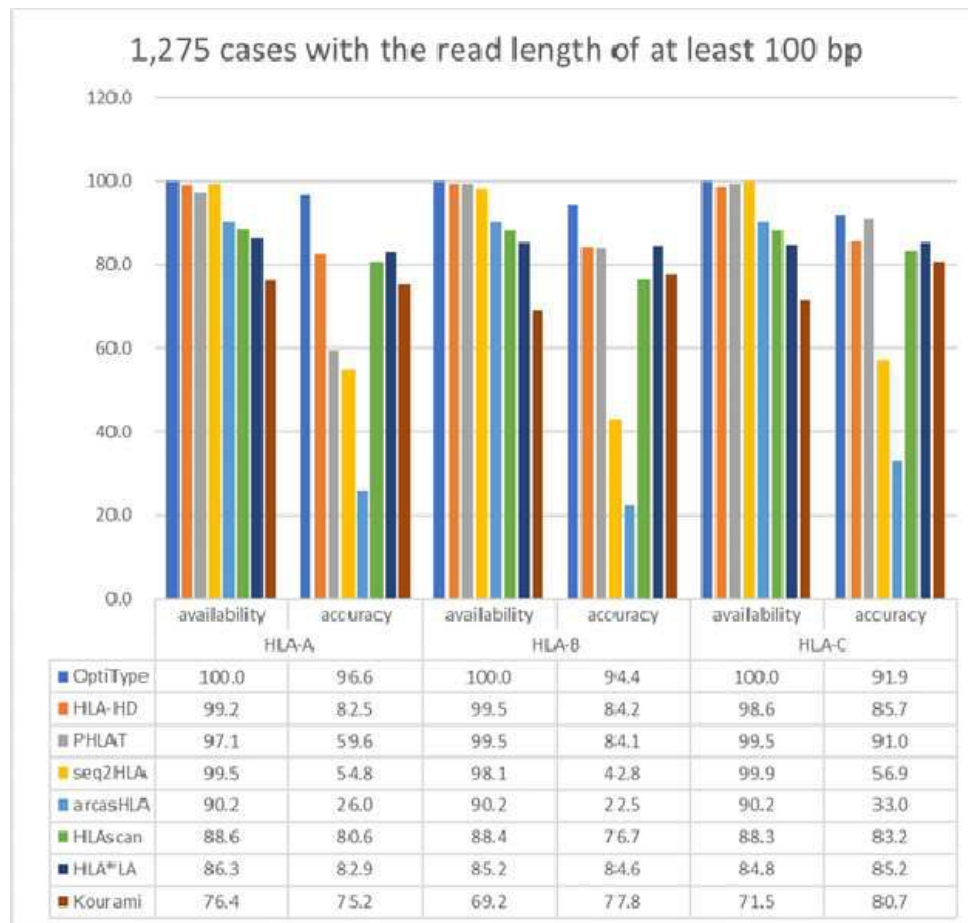
전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

도면

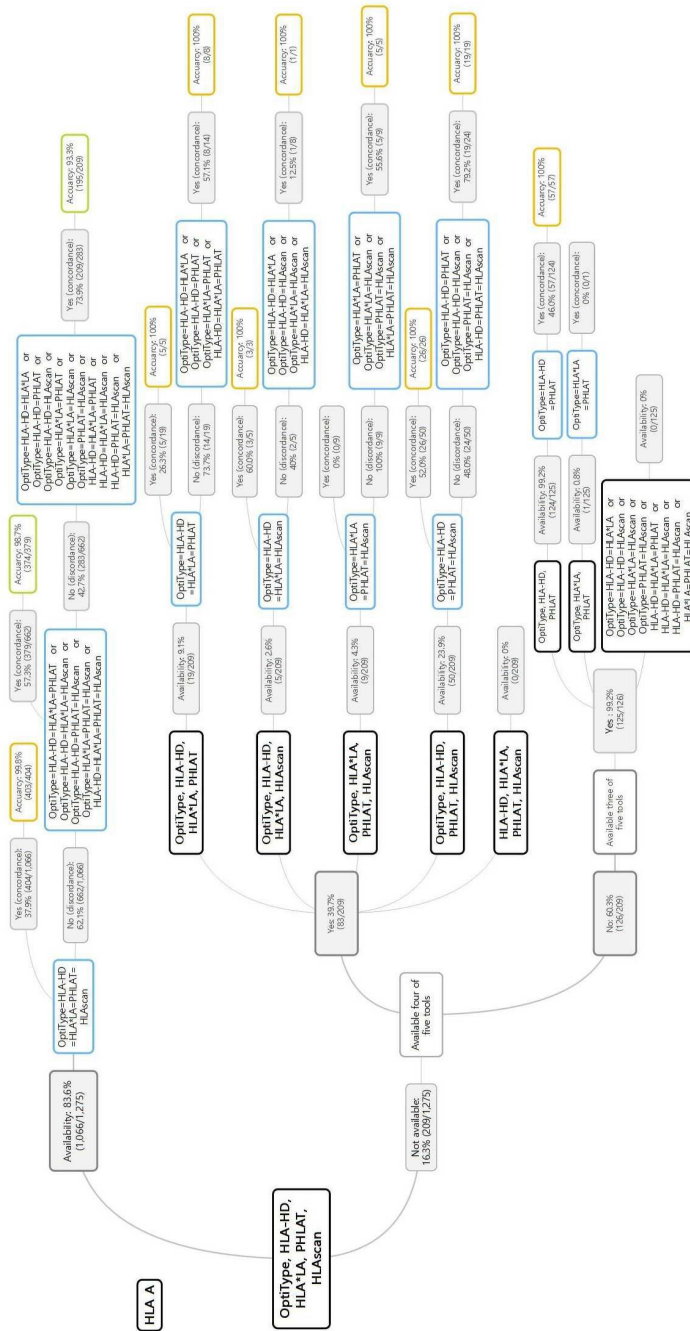
도면1a



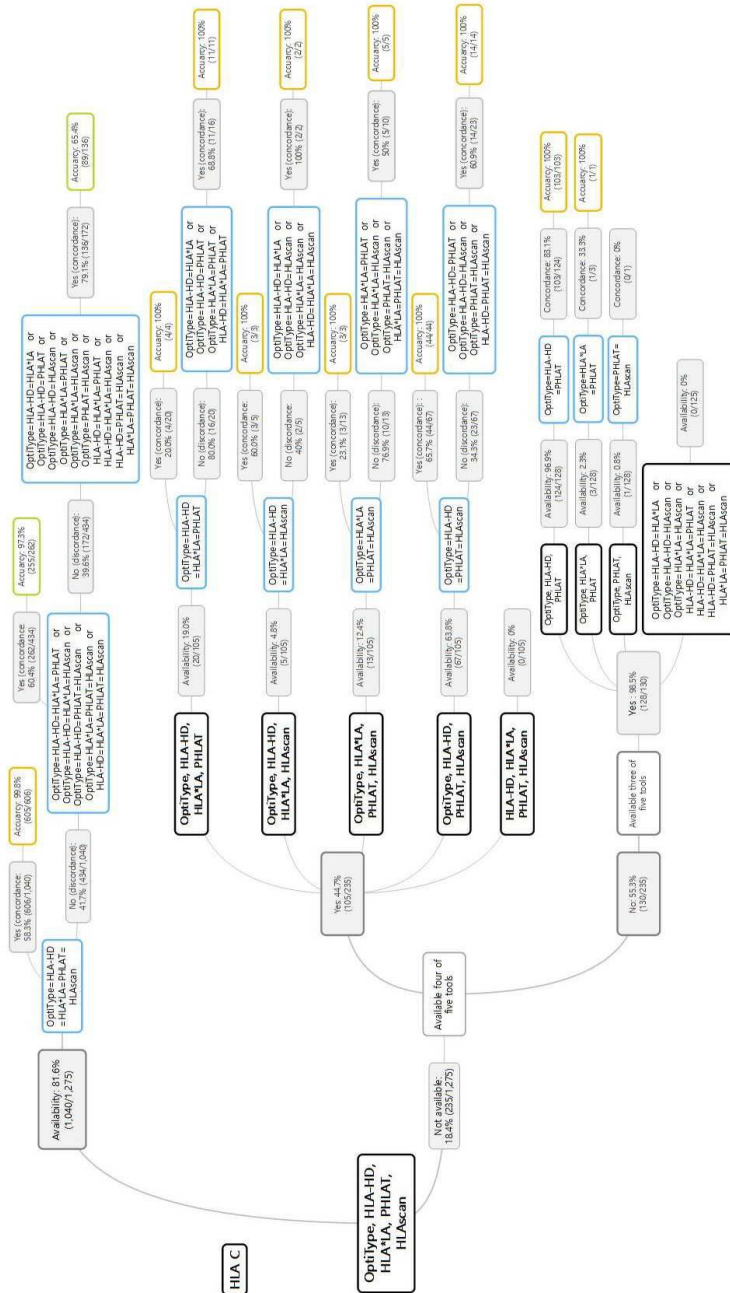
도면1b



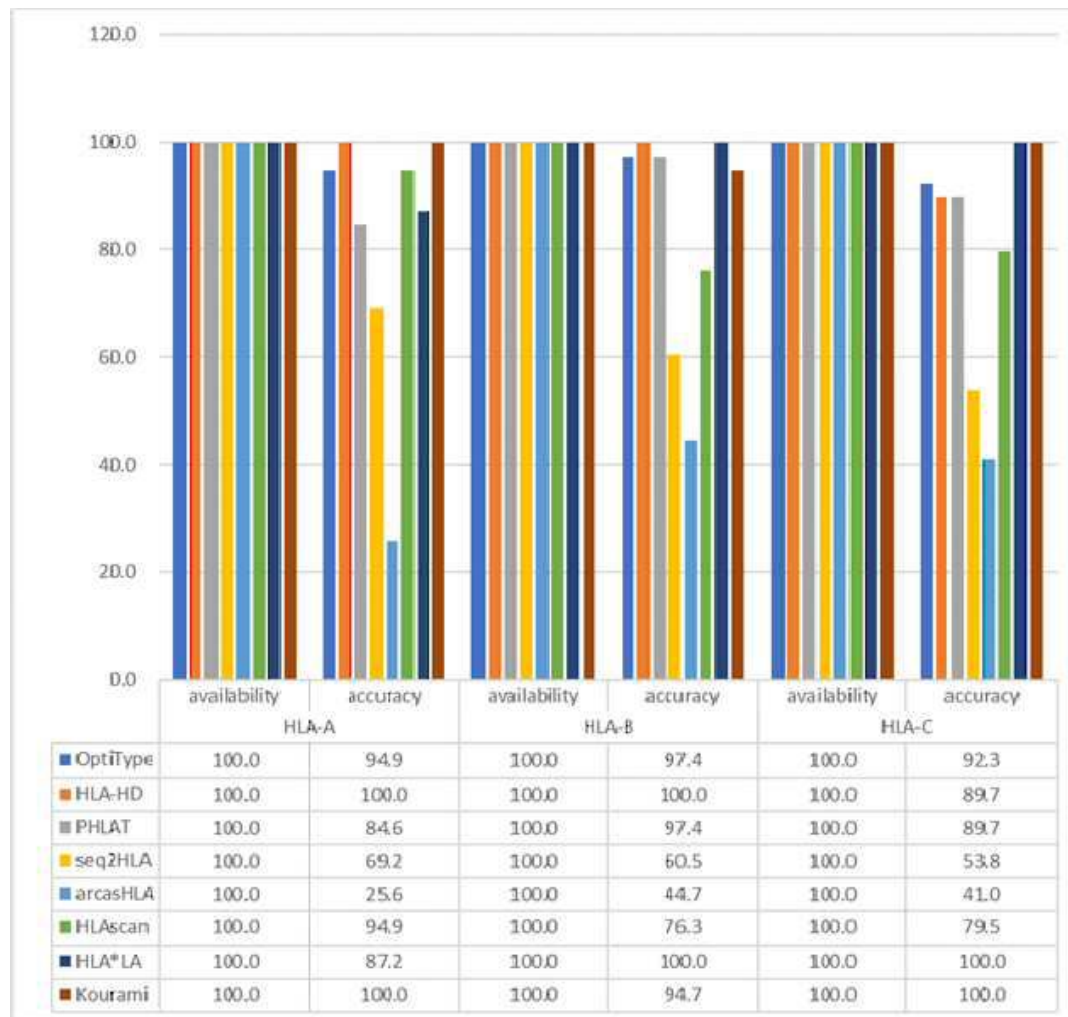
도면2a



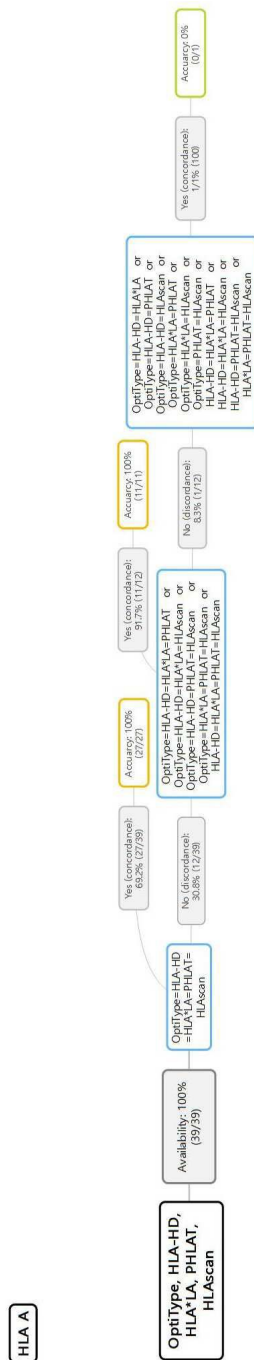
도면2c



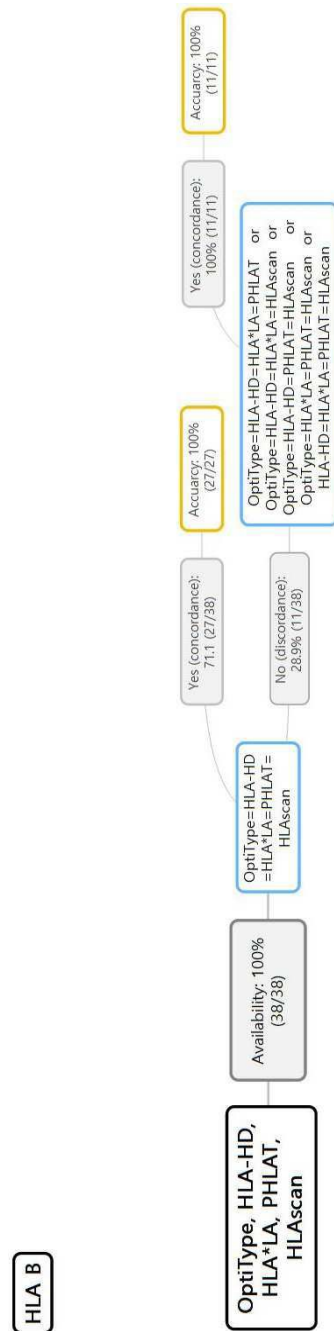
도면3



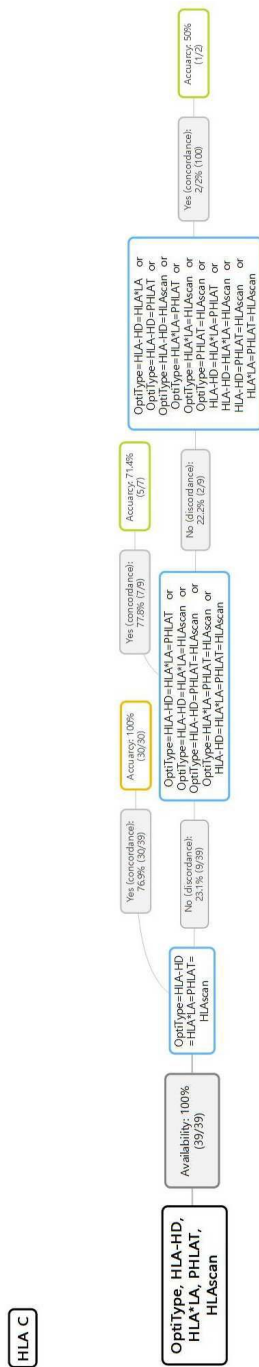
도면4a



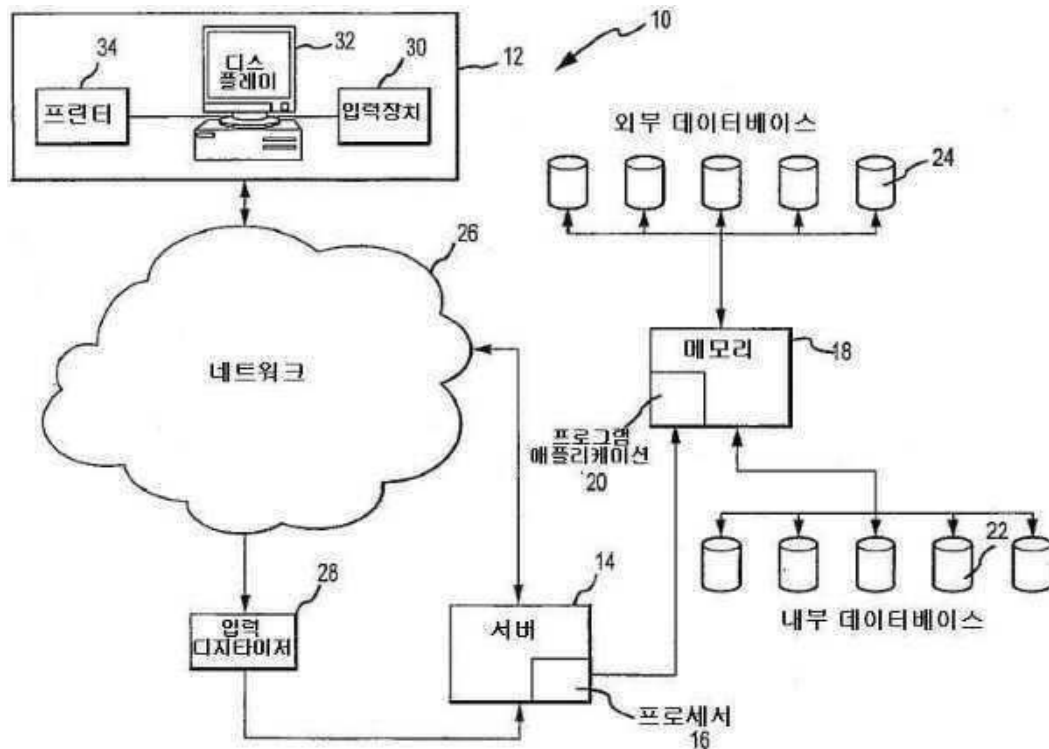
도면4b



도면4c



도면5



도면6

