



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0083114
(43) 공개일자 2019년07월11일

(51) 국제특허분류(Int. Cl.)
G16B 50/00 (2019.01)

(52) CPC특허분류
G16B 50/00 (2019.02)

(21) 출원번호 10-2018-0000661

(22) 출원일자 2018년01월03일

심사청구일자 2018년01월03일

(71) 출원인

울산대학교 산학협력단

울산광역시 남구 대학로 93(무거동)

재단법인 아산사회복지재단

서울특별시 송파구 올림픽로43길 88 (풍납동)

(72) 발명자

박유량

서울특별시 송파구 올림픽로 435, 304동 104호 (신천동, 파크리오아파트)

김규표

서울특별시 송파구 올림픽로 435, 224동 1204호 (잠실4동, 파크리오아파트)

김덕훈

경기도 광주시 초월읍 경충대로1127번길 15, 104동 306호 (동광모닝스카이라파트)

(74) 대리인

특허법인 무한

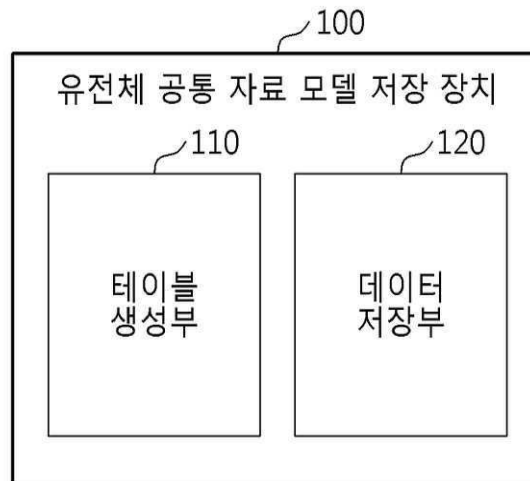
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 유전체 공통 자료 모델 저장 장치 및 상기 장치의 동작 방법

(57) 요약

유전체 공통 자료 모델 저장 장치 및 상기 장치의 동작 방법이 제공된다. 일 실시예에 따른, 유전체 공통 자료 모델 저장 장치는, 유전체 파일에 관한 정보를 기초로 코어 엘리먼트들을 분리하여 유전체 데이터 테이블을 생성하는 테이블 생성부; 및 공통 자료 모델을 이용하여 상기 생성된 유전체 데이터 테이블을 데이터 저장소에 저장하는 데이터 저장부를 포함할 수 있다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호 1711058081
 부처명 과학기술정보통신부
 연구관리전문기관 한국연구재단
 연구사업명 바이오·의료기술개발
 연구과제명 임상·오믹스 정보 통합 개방형 플랫폼 구축 및 다기관 검증
 기 여 율 50/100
 주관기관 서울아산병원
 연구기간 2017.06.30 ~ 2022.06.29

이 발명을 지원한 국가연구개발사업

과제고유번호 1465023214
 부처명 보건복지부
 연구관리전문기관 한국보건산업진흥원
 연구사업명 임상연구인프라조성
 연구과제명 임상/영상/유전체 정보 통합 플랫폼 구축
 기 여 율 20/100
 주관기관 서울아산병원
 연구기간 2017.04.01 ~ 2018.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호 1345266442
 부처명 교육부
 연구관리전문기관 한국연구재단
 연구사업명 개인기초연구(교육부)
 연구과제명 자연어처리 기반 비정형 텍스트 임상자료 처리 방법 개발 및 검증
 기 여 율 30/100
 주관기관 서울아산병원
 연구기간 2017.06.01 ~ 2020.05.31

명세서

청구범위

청구항 1

유전체 공통 자료 모델 저장 장치에 있어서,

유전체 파일에 관한 정보를 기초로 코어 엘리먼트들을 분리하여 유전체 데이터 테이블을 생성하는 테이블 생성부; 및

공통 자료 모델을 이용하여 상기 생성된 유전체 데이터 테이블을 데이터 저장소에 저장하는 데이터 저장부를 포함하는 장치.

청구항 2

제1항에 있어서,

상기 저장된 유전체 데이터 테이블을 이용하여 공동 연구를 위한 통합 서비스를 제공하는 서비스 제공부를 더 포함하는 장치.

청구항 3

제2항에 있어서,

상기 서비스 제공부는,

분산형으로 복수의 기관에서 연구하기 위한 네트워크에 관한 서비스를 제공하는 장치.

청구항 4

제1항에 있어서,

상기 테이블 생성부는,

상기 유전체 파일에서의 돌연변이에 관한 정보, 유전자 복제수 변이에 관한 정보, 유전자 구조 변이에 관한 정보, 유전자 단일 염기 변이에 관한 정보 중 적어도 하나를 기초로 유전체 데이터 테이블을 생성하는 장치.

청구항 5

제1항에 있어서,

상기 테이블 생성부는,

인구 통계 또는 시료와 같은 환자 정보를 이용하여 유전체 데이터 테이블을 생성하는 장치.

청구항 6

제1항에 있어서,

상기 테이블 생성부는,

임상 데이터 또는 유전체 데이터를 이용하여 유전체 데이터 테이블을 생성하는 장치.

청구항 7

제4항에 있어서,

상기 테이블 생성부는,

상기 유전자 복제수 변이에 관한 정보, 상기 유전자 구조 변이에 관한 정보, 상기 유전자 단일 염기 변이에 관한 정보 중 적어도 하나를 기초로 생성한 제1 유전체 데이터 테이블과, 시료에 관한 정보를 기초로 생성한 제2

유전체 데이터 테이블을 연결함으로써 유전체 데이터 테이블을 생성하는 장치.

청구항 8

제1항에 있어서,

상기 테이블 생성부는,

대상(person), 프로바이더(provider), 프로시저(procedure), 컨셉(concept), 보캐블러리(vocabulary), 도메인(domain) 중 적어도 하나를 기초로 유전체 데이터 테이블을 생성하고,

상기 생성한 유전체 데이터 테이블 중 하나인 제3 유전체 데이터 테이블과 상기 생성한 유전체 데이터 테이블 중 다른 하나인 제4 유전체 데이터 테이블을 연결하는 장치.

청구항 9

제1항에 있어서,

상기 테이블 생성부는,

대상, 시료, 프로시저, 메타정보, 돌연변이, 복제수 변이, 구조 변이, 발현, 관찰, 컨셉에 관한 테이블들 중 적어도 하나를 기초로 유전체 데이터 테이블을 생성하고,

상기 생성한 유전체 데이터 테이블 중 하나인 제5 유전체 데이터 테이블과 상기 생성한 유전체 데이터 테이블 중 다른 하나인 제6 유전체 데이터 테이블을 연결하거나 선택적으로 연결하는 장치.

청구항 10

제2항에 있어서,

상기 서비스 제공부는,

상기 저장된 유전체 데이터 테이블 중 적어도 하나의 유전체 데이터 테이블의 데이터를 외부물에 내보내기하는 장치.

청구항 11

유전체 공통 자료 모델 저장 장치의 동작 방법에 있어서,

유전체 파일에 관한 정보를 기초로 코어 엘리먼트들을 분리하여 유전체 데이터 테이블을 생성하는 단계; 및
공통 자료 모델을 이용하여 상기 생성된 유전체 데이터 테이블을 데이터 저장소에 저장하는 단계를 포함하는 방법.

청구항 12

제11항에 있어서,

상기 저장된 유전체 데이터 테이블을 이용하여 공동 연구를 위한 통합 서비스를 제공하는 단계를 더 포함하는 방법.

청구항 13

제11항 또는 제12항의 방법을 수행하는 프로그램을 수록한 컴퓨터 판독 가능 기록 매체.

발명의 설명

기술 분야

아래 설명은, 유전체 공통 자료 모델 저장 장치 및 상기 장치의 동작 방법에 관한 것이다.

[0001]

배경 기술

- [0002] 데이터베이스는 컴퓨터화된 정보 저장 및 검색 시스템으로서, 관계형 데이터베이스 관리 시스템(relational database management system; RDBMS)은 데이터를 저장하고 검색하는 관계형 기술을 사용하는 컴퓨터 데이터베이스 관리 시스템이다. "테이블"은 정식으로는 "속성(attributes)"으로 명명되는 복수 개의 열을 스패닝(spanning)하는 행의 세트를 포함한다.
- [0003] 임상 게놈 애플리케이션은, 다양한 데이터 소스로부터의 의학 정보를 데이터 베이스 등에 저장될 수 있게 할 수 있다.
- [0004] 대한민국 등록특허공보 10-0681795호를 참조하면, 그리드 컴퓨팅 환경에서 유전체 서열을 효율적으로 정렬하는 방법에 대하여 기술되어 있다.

발명의 내용

해결하려는 과제

- [0005] 유전체 공통 자료 모델 저장 장치는, 기관 통합 정보 저장 시스템을 제공하는 것을 목적으로 한다.
- [0006] 유전체 공통 자료 모델 저장 장치는, 유전체 데이터 공동연구에 있어서 분산형 네트워크 기반 공통 자료 모델에 연관되는 다국가적 플랫폼을 제공하는 것을 다른 목적으로 한다.

과제의 해결 수단

- [0007] 일 실시예에 따른, 유전체 공통 자료 모델 저장 장치는, 유전체 파일에 관한 정보를 기초로 코어 엘리먼트(core element)들을 분리하여 유전체 데이터 테이블을 생성하는 테이블 생성부; 및 공통 자료 모델을 이용하여 상기 생성된 유전체 데이터 테이블을 데이터 저장소에 저장하는 데이터 저장부를 포함할 수 있다.
- [0008] 일 실시예에 따른, 유전체 공통 자료 모델 저장 장치는, 상기 저장된 유전체 데이터 테이블을 이용하여 공동 연구를 위한 통합 서비스를 제공하는 서비스 제공부를 더 포함할 수 있다.
- [0009] 일 실시예에 따른, 상기 서비스 제공부는, 분산형으로 복수의 기관에서 연구하기 위한 네트워크에 관한 서비스를 제공할 수 있다.
- [0010] 일 실시예에 따른, 상기 테이블 생성부는, 상기 유전체 파일에서의 돌연변이에 관한 정보를 기초로 유전체 데이터 테이블을 생성할 수 있다.
- [0011] 일 실시예에 따른, 상기 테이블 생성부는, 유전자 복제수 변이에 관한 정보를 이용하여 유전체 데이터 테이블을 생성할 수 있다.
- [0012] 일 실시예에 따른, 상기 테이블 생성부는, 유전자 구조 변이에 관한 정보를 이용하여 유전체 데이터 테이블을 생성할 수 있다.
- [0013] 일 실시예에 따른, 상기 테이블 생성부는, 유전자 단일 염기 변이에 관한 정보를 이용하여 유전체 데이터 테이블을 생성할 수 있다.
- [0014] 일 실시예에 따른, 상기 테이블 생성부는, 인구 통계 또는 시료와 같은 환자 정보를 이용하여 유전체 데이터 테이블을 생성할 수 있다.
- [0015] 일 실시예에 따른, 상기 테이블 생성부는, 임상 데이터 또는 유전체 데이터를 이용하여 유전체 데이터 테이블을 생성할 수 있다.
- [0016] 일 실시예에 따른, 상기 테이블 생성부는, 상기 유전자 복제수 변이에 관한 정보, 상기 유전자 구조 변이에 관한 정보, 상기 유전자 단일 염기 변이에 관한 정보 중 적어도 하나를 기초로 생성한 제1 유전체 데이터 테이블과, 시료에 관한 정보를 기초로 생성한 제2 유전체 데이터 테이블을 연결함으로써 유전체 데이터 테이블을 생성할 수 있다.
- [0017] 일 실시예에 따른, 상기 테이블 생성부는, 대상(person), 프로바이더(provider), 프로시저(procedure), 컨셉(concept), 보캐블러리(vocabulary), 도메인(domain) 중 적어도 하나를 기초로 유전체 데이터 테이블을 생성하고, 상기 생성한 유전체 데이터 테이블 중 하나인 제3 유전체 데이터 테이블과 상기 생성한 유전체 데이터 테이블

블 중 다른 하나인 제4 유전체 데이터 테이블을 연결할 수 있다.

- [0018] 일 실시예에 따른, 상기 테이블 생성부는, 대상, 시료, 프로시저, 메타정보, 돌연변이, 복제수 변이, 구조 변이, 발현, 관찰, 컨셉에 관한 테이블들 중 적어도 하나를 기초로 유전체 데이터 테이블을 생성하고, 상기 생성한 유전체 데이터 테이블 중 하나인 제5 유전체 데이터 테이블과 상기 생성한 유전체 데이터 테이블 중 다른 하나인 제6 유전체 데이터 테이블을 연결하거나 선택적으로 연결할 수 있다.
- [0019] 일 실시예에 따른, 상기 서비스 제공부는, 상기 저장된 유전체 데이터 테이블 중 적어도 하나의 유전체 데이터 테이블의 데이터를 외부물에 내보내기할 수 있다.
- [0020] 일 실시예에 따른, 유전체 공통 자료 모델 저장 장치의 동작 방법은, 유전체 파일에 관한 정보를 기초로 코어 엘리먼트들을 분리하여 유전체 데이터 테이블을 생성하는 단계; 및 공통 자료 모델을 이용하여 상기 생성된 유전체 데이터 테이블을 데이터 저장소에 저장하는 단계를 포함할 수 있다.
- [0021] 일 실시예에 따른, 유전체 공통 자료 모델 저장 장치의 동작 방법은, 상기 저장된 유전체 데이터 테이블을 이용하여 공동 연구를 위한 통합 서비스를 제공하는 단계를 더 포함할 수 있다.
- [0022] 일 실시예에 따른, 컴퓨터 판독 가능 기록 매체는 상기 방법을 수행하는 프로그램을 수록할 수 있다.

도면의 간단한 설명

- [0023] 도 1은 일 실시예에 따른 유전체 공통 자료 모델 저장 장치를 나타내는 도면이다.
- 도 2는 일 실시예에 따른 전체 시스템을 나타내는 도면이다.
- 도 3은 일 실시예에 따른 유전체 공통 자료 모델 저장 장치에 의해 생성된 테이블의 구성을 나타내는 도면이다.
- 도 4는 일 실시예에 따른 유전체 공통 자료 모델 저장 장치의 동작 방법을 나타내는 도면이다.
- 도 5는 일 실시예에 따른 유전체 공통 자료 모델 저장 장치의 동작 방법을 나타내는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0024] 이하에서, 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다. 그러나, 권리범위는 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.
- [0025] 아래 설명에서 사용되는 용어는, 연관되는 기술 분야에서 일반적이고 보편적인 것으로 선택되었으나, 기술의 발달 및/또는 변화, 관례, 기술자의 선호 등에 따라 다른 용어가 있을 수 있다. 따라서, 아래 설명에서 사용되는 용어는 기술적 사상을 한정하는 것으로 이해되어서는 안 되며, 실시예들을 설명하기 위한 예시적 용어로 이해되어야 한다.
- [0026] 또한 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 설명 부분에서 상세한 그 의미를 기재할 것이다. 따라서 아래 설명에서 사용되는 용어는 단순한 용어의 명칭이 아닌 그 용어가 가지는 의미와 명세서 전반에 걸친 내용을 토대로 이해되어야 한다.
- [0027] 도 1은 일 실시예에 따른 유전체 공통 자료 모델 저장 장치를 나타내는 도면이다.
- [0028] 먼저, 도 1을 참조하면, 유전체 공통 자료 모델 저장 장치(100)는 테이블 생성부(110), 데이터 저장부(120)를 포함할 수 있다.
- [0029] 테이블 생성부(110)는 유전체 파일에 관한 정보를 기초로 코어 엘리먼트들을 분리하여 유전체 데이터 테이블을 생성할 수 있다. 테이블 생성부(110)는 유전체 파일에서의 돌연변이에 관한 정보를 기초로 유전체 데이터 테이블을 생성할 수 있다.
- [0030] 테이블 생성부(110)는 유전자 복제수 변이에 관한 정보를 이용하여 유전체 데이터 테이블을 생성할 수 있다. 테이블 생성부(110)는 유전자 구조 변이에 관한 정보를 이용하여 유전체 데이터 테이블을 생성할 수도 있다. 테이블 생성부(110)는 유전자 단일 염기 변이에 관한 정보를 이용하여 유전체 데이터 테이블을 생성할 수도 있다.
- [0031] 데이터 저장부(120)는 공통 자료 모델을 이용하여 생성된 유전체 데이터 테이블을 데이터 저장소에 저장할 수 있다.
- [0032] 유전체 공통 자료 모델 저장 장치(100)는 서비스 제공부를 더 포함할 수 있다. 서비스 제공부는 저장된 유전체

데이터 테이블을 이용하여 공동 연구를 위한 통합 서비스를 제공할 수 있다. 서비스 제공부는 분산형으로 복수의 기관에서 연구하기 위한 네트워크에 관한 서비스를 제공할 수 있다.

- [0033] 도 2는 일 실시예에 따른 전체 시스템을 나타내는 도면이다.
- [0034] 도 2를 참조하면, 전체 시스템은 기관 데이터베이스(210, Institutional Database), 유전체 실험실 정보 시스템(221, Genomic Lab Information System), 임상 정보 시스템(222, Clinical Information System), 의료 행정 시스템(223, Administrative System), 데이터 시각화(231, Data Visualization), 공통 데이터 모델(232, Common Data Model)을 포함할 수 있다. 유전체 공통 자료 모델 저장 장치는 기관 데이터베이스(210, Institutional Database)를 이용하거나 시스템의 다른 구성을 이용할 수 있다. 또한, 유전체 공통 자료 모델 저장 장치는 기관 데이터베이스(210, Institutional Database)를 포함할 수도 있고, 시스템의 다른 구성을 포함할 수도 있다.
- [0035] 유전체 실험실 정보 시스템(221)에서 이용하거나 획득되는 정보는 차세대 염기서열 분석(NGS, next-generation sequencing) 결과(results)일 수 있다. 임상 정보 시스템(222)에서는 실험실 및 방사선 결과(lab and radiologic results)와 의료 노트(MD notes) 등의 정보가 이용되거나 획득될 수 있다. 의료 행정 시스템(223)에서는 방문, 치료, 클레임(visits, treatments, claims) 등의 정보가 이용되거나 획득될 수 있다. 데이터 시각화(231)는 cBioPortal(www.cbioportal.org)과 같은 바이오 관련 포털 사이트 등이 이용될 수 있다.
- [0036] 유전체 공통 자료 모델 저장 장치는 임상 데이터 및 유전체 데이터(clinical and genomic data)를 추출(extracting) 및 로딩(loading)하기 위한 새로운 모델로서, 인구 통계 및 시료(demographic and specimen)과 같은 환자 정보(patient information)를 나타내는 환자 모듈(Patient module)을 포함하거나 이용할 수 있다.
- [0037] 유전체 공통 자료 모델 저장 장치는 수술(surgery), 화학 요법(chemotherapy), 방사선 치료(radiation therapy) 및 방사선 반응 평가(radiological response evaluation)와 같은 임상 사건(clinical events)을 나타내는 임상 데이터 모듈(Clinical Data module), 유전체 데이터 모듈(Genomic Data module)을 포함하거나 이용할 수도 있다.
- [0038] 도 3은 일 실시예에 따른 유전체 공통 자료 모델 저장 장치에 의해 생성된 테이블의 구성을 나타내는 도면이다.
- [0039] 도 3을 참조하면, 유전체 공통 자료 모델 저장 장치는 복수의 코어 엘리먼트들을 분리하여 유전체 데이터 테이블을 생성할 수 있다. 유전체 공통 자료 모델 저장 장치는 유전자 복제수 변이, 유전자 구조 변이, 유전자 단일 염기 변이 등에 관하여 genomic_copy_number_variation, genomic_structural_variation, genomic_single_nucleotide_variants 등의 테이블을 생성할 수 있다. genomic_copy_number_variation, genomic_structural_variation, genomic_single_nucleotide_variants 테이블 각각은 유전체 공통 자료 모델 저장 장치가 생성한 specimen(시료) 테이블에 연결될 수 있다.
- [0040] 유전체 공통 자료 모델 저장 장치는 person, provider, procedure_occurrence, concept, concept_class, vocabulary, domain 등의 테이블도 생성할 수 있다. person 테이블은 specimen, provider, concept, procedure_occurrence 테이블에 연결될 수 있다. specimen 테이블은 person, concept 테이블에 연결될 수 있다. provider 테이블은 person, concept, procedure_occurrence 테이블에 연결될 수 있다. concept 테이블은 person, provider, procedure_occurrence, concept_class, vocabulary, domain 테이블에 연결될 수 있다.
- [0041] 각각의 유전체 공통 자료 모델 저장 장치에 의해 생성된 테이블은 속성들(attributes) 또는 컬럼들(columns)을 포함할 수 있다. 이때, 각각의 유전체 공통 자료 모델 저장 장치에 의해 생성된 테이블은 기본 키(Primary Key) 또는 식별 정보로서 각 테이블의 데이터에 대한 id를 속성 또는 컬럼으로 포함할 수 있다.
- [0042] 누적된 임상 유전체 데이터의 활용도를 높이기 위하여, 자료 모델 저장 장치는 NGS(Next-Generation Sequencing) 암 패널 테스트(cancer panel test) 결과를 저장하는 과정을 확장하여 적용할 수 있다.
- [0043] 자료 모델 저장 장치는 유전체 데이터 테이블(genomic_data table)을 생성함으로써, 유전체 데이터를 포함하거나 이용 또는 생성할 수 있다. 유전체 테이블의 구조는 다양한 유전체 파일들(genomic files) 중 돌연변이(mutation)에 대한 주석(annotation) 정보를 포함하는 MAF(Mutation Annotation Format) 파일에서 참조될 수 있다. AMC(Asan Medical Center, 아산병원) NGS 암 패널의 유전자는 컨셉 테이블(Concept table)에서 조직 방어적 컨셉(organization defended concept)으로서 로드(load)될 수 있다. 컨셉 테이블에 등록된 모든 유전자는 HGNC 명명법(HUGO Gene Nomenclature Committee nomenclature)에 따라서 명명(named)될 수 있다.
- [0044] 확장 유전체 테이블에는 아산 병원 등의 기관의 임상 연구 데이터 웨어 하우스(CRDW, clinical research data warehouse)에 누적된 유전체 데이터가 로딩될 수 있다. NGS 암 패널 테스트의 순서 또는 명령은 프로시저 테이블

블(Procedure table)에 로딩될 수 있다. 임상 및 유전체 데이터는 임상 및 유전체 데이터의 통합 시각화(integrated visualization)와 분석을 위하여, 오픈 소스 유전체 데이터 API(open source genomic data API)로서 cBioPortal과 같은 바이오 포털 사이트에 로딩될 수 있다.

[0045] 도 4는 일 실시예에 따른 유전체 공통 자료 모델 저장 장치의 동작 방법을 나타내는 도면이다.

[0046] 도 4를 참조하면, 유전체 공통 자료 모델 저장 장치는 대상(410, person), 시료(421, specimen), 프로시저(422, procedure), 메타정보(430, omics_meta), 돌연변이(441, mutation), 복제수 변이(442, CNV), 구조 변이(443, SV), 발현(444, expression), 관찰(451, observation), 컨셉(452, concept)의 테이블들을 생성하거나 관리할 수 있다.

[0047] 대상(410) 테이블은 시료(421), 프로시저(422) 테이블과 각각 연결될 수 있다. 시료(421), 프로시저(422) 테이블 각각은 메타정보(430) 테이블과 연결될 수 있다. 메타정보(430) 테이블은 돌연변이(441), 복제수 변이(442), 구조 변이(443), 발현(444) 테이블 각각과 연결될 수 있다. 돌연변이(441), 복제수 변이(442), 구조 변이(443), 발현(444) 테이블 각각은 컨셉(452)의 테이블과 선택적으로 연결될 수 있다. 돌연변이(441)은 관찰(451) 테이블과 선택적으로 연결될 수 있다.

[0048] 유전체 공통 자료 모델 저장 장치가 생성하거나 관리하는 테이블들 중 적어도 하나의 테이블의 데이터는 cBioportal, DNANexus, GeneGo 등과 같은 외부툴(460, External tools)에 내보내기(export)될 수도 있다.

[0049] 메타정보(430) 테이블의 이름(name)은 omics_meta일 수 있다. 메타정보(430) 테이블은 Specimen_id, Test_id, Source_type, Common values, Assay_type, common values, Allowed values, Panel_name, enrichment_methods, Instrument, Center, Sequence_read_length, Sequencing_layout, Reference_genome_build, Analysis_tool, Gene_db, Tumor_cellularity(%) 중 적어도 하나를 속성(attribute) 또는 컬럼(column)으로서 포함할 수 있다. 메타정보(430) 테이블은 각각의 속성 또는 컬럼에 대응하는 데이터들을 포함할 수 있다.

[0050] Specimen_id는 시료 아이디로서 시료 테이블(specimen table)에 매치(matched)될 수 있다. Test_id는 테스트 아이디로서 AMC_1, AMC_2 등과 같이 기관_번호(“institute_number”) 등과 동일해야 될 수 있다. Source_type은 실험에 사용된 시료 타입일 수 있다. Common values는 DNA(deoxyribonucleic acid, 데옥시리보핵산), RNA(ribo nucleic acid, 리보핵산), 단백질(protein) 등일 수 있다. Assay_type은 진행한 시험 타입일 수 있다. common values는 WGS(Whole genome sequencing, 전체 유전체 염기서열분석), WXS(whole exome sequencing, 전체 엑솜 염기서열분석), RNA-seq(sequencing), Targeted-Seq 등일 수 있다. Allowed values는 Sanger sequencing, IHC(immunohistochemistry, 면역조직화학법), qPCR, FISH(fluorescence in situ hybridization, 형광동소혼성화법), CISH(Chromogenic in situ hybridization, 발색동소혼성화법) 등일 수 있다.

[0051] Panel_name은 Targeted-Seq(targeted sequencing)일 경우, 선택적(optional)일 수 있고, OP_AMCv3 등의 패널(panel) 정보일 수 있다. enrichment_methods는 선택적일 수 있고, 캡처(capture), 증폭(amplification) 등일 수 있고, WXS 또는 targeted sequencing일 경우 어떤 방법으로 인리치먼트(enrichment)했는지에 관한 정보일 수 있다. Instrument는 Illumina Miseq, HiSeq2000 등의 사용된 장비일 수 있다. Center는 Asan Medical Cent(아산병원), SNUH(서울대학교병원) 등의 실험이 진행된 장소일 수 있다. Sequence_read_length는 선택적일 수 있고, NGS(Next-Generation Sequencing)일 경우 사용된 염기서열분석용 시약(sequencing reagent)의 한 번에 읽을 수 있는 길이(read length)일 수 있다.

[0052] Sequencing_layout은 선택적일 수 있고, single-end, paired-end, mate-pair 등일 수 있고, NGS일 경우 사용된 시퀀싱 레이아웃(sequencing layout)일 수 있다. Reference_genome_build는 NGS기반 테스트(NGS-based test)의 경우 필수적(mandatory)일 수 있고, hg19, b37, hg38 등 분석에 사용된 대조 유전체(reference genome)일 수 있다.

[0053] Analysis_tool은 NGS기반 테스트의 경우 필수적일 수 있고, 툴 네임:버전((tool name):(version)) 등의 정보로서 BWA(Burrows-Wheeler Aligner):0.7.10, GATK(Genome Analysis Tool Kit):3.8.0, PICARD:2.14.0, Mutect:1.1.7 등의 분석에 사용된 분석 툴에 관한 정보일 수 있다. Gene_DB는 NGS기반 테스트의 경우 필수적일 수 있고, 데이터베이스:버전((DB):(version)) 등의 정보로서 Ensembl:90 등의 유전자 어노테이션(gene annotation)에 사용된 유전자 데이터베이스(gene DB)에 관한 정보일 수 있다. Tumor_cellularity(%)는 종양 샘플(tumor sample)의 경우 선택적일 수 있고, 70 등 사용된 암조직의 종양 세포질(tumor cellularity)일 수 있다.

- [0054] 돌연변이(441) 테이블의 이름은 mutation일 수 있다. 돌연변이(441) 테이블의 정보는 단일염기서열변이(SNV), 삽입(INS)-결실(DEL) 등에 관한 정보일 수 있다. 돌연변이(441) 테이블은 Test_id, Chromosome, Start_Position, End_Position, Strand, Variant_Classification, Variant_Type, Reference_Allele, Variant allele, Mutation_Status, HGVSP(Human Genome Variation Society Protein), T_TOTAL_DEPTH, T_REF_DEPTH, T_VAR_DEPTH, N_TOTAL_DEPTH, N_REF_DEPTH, N_VAR_DEPTH, ALLELE_FREQUENCY, EXON, INTRON, Transcript id 중 적어도 하나를 속성 또는 컬럼으로서 포함할 수 있다. 돌연변이(441) 테이블은 각각의 속성 또는 컬럼에 대응하는 데이터들을 포함할 수 있다.
- [0055] Test_id는 테스트 아이디이고, Chromosome는 염색체 번호이고, Start_Position는 변이가 생긴 곳 시작 지점이고, End_Position는 변이가 생긴 곳 끝나는 지점일 수 있다. Strand는 +, - 등의 가닥(strand)일 수 있고, Variant_Classification는 SNP(Single Nucleotide Polymorphism, 단일염기다형성), INS(삽입), DEL(결실), TNP, MNP(Multi-Nucleotide Polymorphism, 다염기다형성), ... 등의 변이 종류일 수 있다. Variant_Type은 silent, missense, nonsense, splice-site 등의 변이 타입일 수 있다. Reference_Allele는 A,G,C,T 등의 대조 대립유전자일 수 있고, Variant allele는 A,G,C,T 등의 변이 대립유전자일 수 있다.
- [0056] Mutation_Status는 somatic(체세포성), germline(생식세포성), unknown(미지) 등의 변이의 상태일 수 있고, HGVSC는 c.755C>A 등 HGVS 등의 명명법에 의한 cDNA 변이 표기일 수 있고, HGVSP는 p.T790M 등 HGVS 등의 명명법에 의한 아미노산(amino acid) 변이 표기일 수 있다.
- [0057] T_TOTAL_DEPTH는 체세포 및 생식세포 테스트 모두에 관한 정보일 수 있고, 종양에서의 토탈 시퀀싱 템스이고, T_REF_DEPTH는 체세포 및 생식세포 테스트 모두에 관한 정보일 수 있고, 종양에서의 레퍼런스 염기서열의 시퀀싱 템스일 수 있다. T_VAR_DEPTH는 체세포 및 생식세포 테스트 모두에 관한 정보일 수 있고, 종양에서의 변이 염기서열 시퀀싱 템스일 수 있다.
- [0058] N_TOTAL_DEPTH는 신체 테스트(somatic test)에 대해 정상적(normal)으로 매치되어 사용될 경우, 선택적일 수 있고, 정상(normal)에서의 토탈 시퀀싱 템스일 수 있다. N_REF_DEPTH는 신체 테스트에 대해 정상적으로 매치되어 사용될 경우, 선택적일 수 있고, 정상에서의 레퍼런스 염기서열 시퀀싱 템스일 수 있다. N_VAR_DEPTH는 신체 테스트에 대해 정상적으로 매치되어 사용될 경우, 선택적일 수 있고, 정상에서의 변이 염기서열 시퀀싱 템스일 수 있다.
- [0059] ALLELE_FREQUENCY는 $0 < AF \leq 1$ 등의 변이염기서열 시퀀싱 템스 /레퍼런스 염기서열의 템스일 수 있다. EXON는 3/10 등의 변이가 생긴 엑손(exon)일 수 있고, INTRON는 3/10 등의 변이가 생긴 인트론(intron)일 수 있고, Transcript id는 ENST..., NM..., 등의 전사물 식별 정보(transcript ID)일 수 있다.
- [0060] 복제수 변이(442) 테이블의 이름은 CNV 또는 copy number variations에 관한 정보를 포함할 수 있다. 복제수 변이(442) 테이블은 Test_id, Chromosome, Gene or genes, Log2_ratio, Estimated_CN, Segment_start, Segment_end 중 적어도 하나를 속성 또는 컬럼으로서 포함할 수 있다. 복제수 변이(442) 테이블은 각각의 속성 또는 컬럼에 대응하는 데이터들을 포함할 수 있다.
- [0061] Test_id는 테스트 아이디이고, Chromosome는 염색체 번호이고, Gene or genes는 유전자일 수 있다. Log2_ratio는 log2_ratio 등의 비율일 수 있고, Estimated_CN은 예측되는 복제수(copy number)일 수 있다. Segment_start는 세그먼트(segment) 시작 지점이고, Segment_end는 세그먼트 종말 지점일 수 있다.
- [0062] 구조 변이(443) 테이블의 이름은 SV 또는 구조 변이(structural variations)에 관한 정보를 포함할 수 있다. 구조 변이(443) 테이블은 Test_id, Sequence_Source, Event_type, Gene, Breakpoint1, Breakpoint2, Strands, Supporting_reads, Contig_sequence 중 적어도 하나를 속성 또는 컬럼으로서 포함할 수 있다. 구조 변이(443) 테이블은 각각의 속성 또는 컬럼에 대응하는 데이터들을 포함할 수 있다.
- [0063] Test_id는 테스트 아이디이고, Sequence_Source는 WGS, WXS, RNA-seq 등의 결과 소스일 수 있다. Event_type은 SV, fusion, virus_integration 등의 이벤트 타입일 수 있다. Gene은 GENE1-GENE2 등의 유전자이고, Breakpoint1은 앞 유전자의 끝단(breakpoint)이고, Breakpoint2는 뒤 유전자의 끝단일 수 있다.
- [0064] Strands는 가닥(strand)이고, Supporting_reads는 SV 이벤트(event)를 지원(supporting)하는 리드(read) 수일 수 있다. Contig_sequence는 이벤트(event)를 지원(supporting)하는 시퀀스일 수 있다.
- [0065] 발현(444) 테이블의 이름은 발현(expression)에 관한 정보를 포함할 수 있다. 발현(444) 테이블은 Test_id, Expression_unit, Gene, Expression_value 중 적어도 하나를 속성 또는 컬럼으로서 포함할 수 있다. 발현(444)

테이블은 각각의 속성 또는 컬럼에 대응하는 데이터들을 포함할 수 있다.

- [0066] Test_id은 테스트 아이디이고, Expression_unit은 RPKM(Reads Per Kilobase Million), TPM(Transcripts Per Million), FPKM(Fragments Per Kilobase Million), ... 등의 발현을 보여주는 유니트일 수 있다. Gene은 유전자이고, Expression_value는 발현 값일 수 있다.
- [0067] 도 5는 일 실시예에 따른 유전체 공통 자료 모델 저장 장치의 동작 방법을 나타내는 흐름도이다.
- [0068] 도 5를 참조하면, 유전체 공통 자료 모델 저장 장치가 수행하는, 유전체 공통 자료 모델 저장 장치의 동작 방법은 하기와 같은 단계를 포함할 수 있다. 유전체 공통 자료 모델 저장 장치의 동작 방법은 유전체 공통 자료 모델 저장 장치의 테이블 생성부, 데이터 저장부, 서비스 제공부 중 적어도 하나의 구성 요소 또는 유전체 공통 자료 모델 저장 장치의 다른 구성 요소를 이용하여 수행될 수 있다.
- [0069] 단계(510)에서, 유전체 공통 자료 모델 저장 장치는 유전체 파일에 관한 정보를 기초로 코어 엘리먼트들을 분리하여 유전체 데이터 테이블을 생성할 수 있다. 유전체 공통 자료 모델 저장 장치는 유전체 파일에서의 돌연변이에 관한 정보를 기초로 유전체 데이터 테이블을 생성할 수 있다.
- [0070] 유전체 공통 자료 모델 저장 장치는 유전자 복제수 변이에 관한 정보를 이용하여 유전체 데이터 테이블을 생성할 수 있다. 유전체 공통 자료 모델 저장 장치는 유전자 구조 변이에 관한 정보를 이용하여 유전체 데이터 테이블을 생성할 수도 있다. 유전체 공통 자료 모델 저장 장치는 유전자 단일 염기 변이에 관한 정보를 이용하여 유전체 데이터 테이블을 생성할 수도 있다.
- [0071] 단계(520)에서, 유전체 공통 자료 모델 저장 장치는 공통 자료 모델을 이용하여 생성된 유전체 데이터 테이블을 데이터 저장소에 저장할 수 있다.
- [0072] 유전체 공통 자료 모델 저장 장치는 저장된 유전체 데이터 테이블을 이용하여 공동 연구를 위한 통합 서비스를 제공할 수도 있다. 유전체 공통 자료 모델 저장 장치는 분산형으로 복수의 기관에서 연구하기 위한 네트워크에 관한 서비스를 제공할 수 있다.
- [0073] 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, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 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.
- [0074] 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(embody)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.
- [0075] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 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

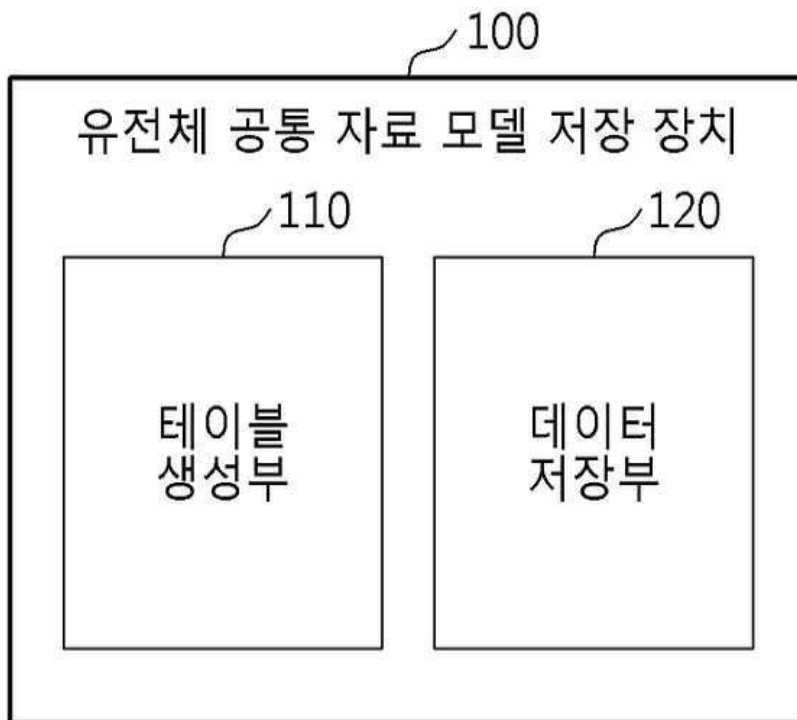
(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0076] 실시예들이 비록 한정된 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.

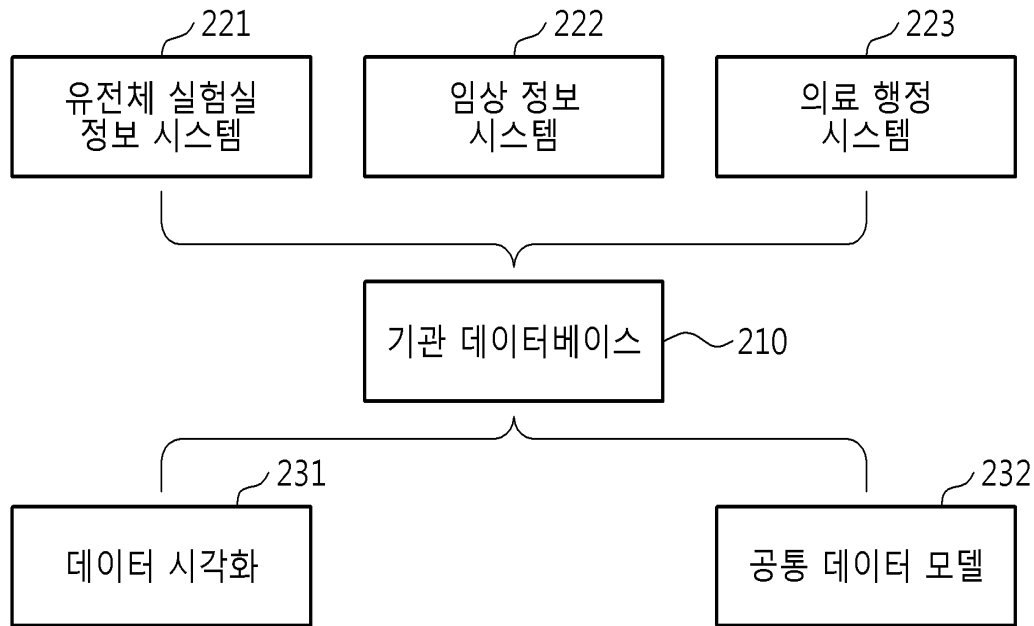
[0077] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

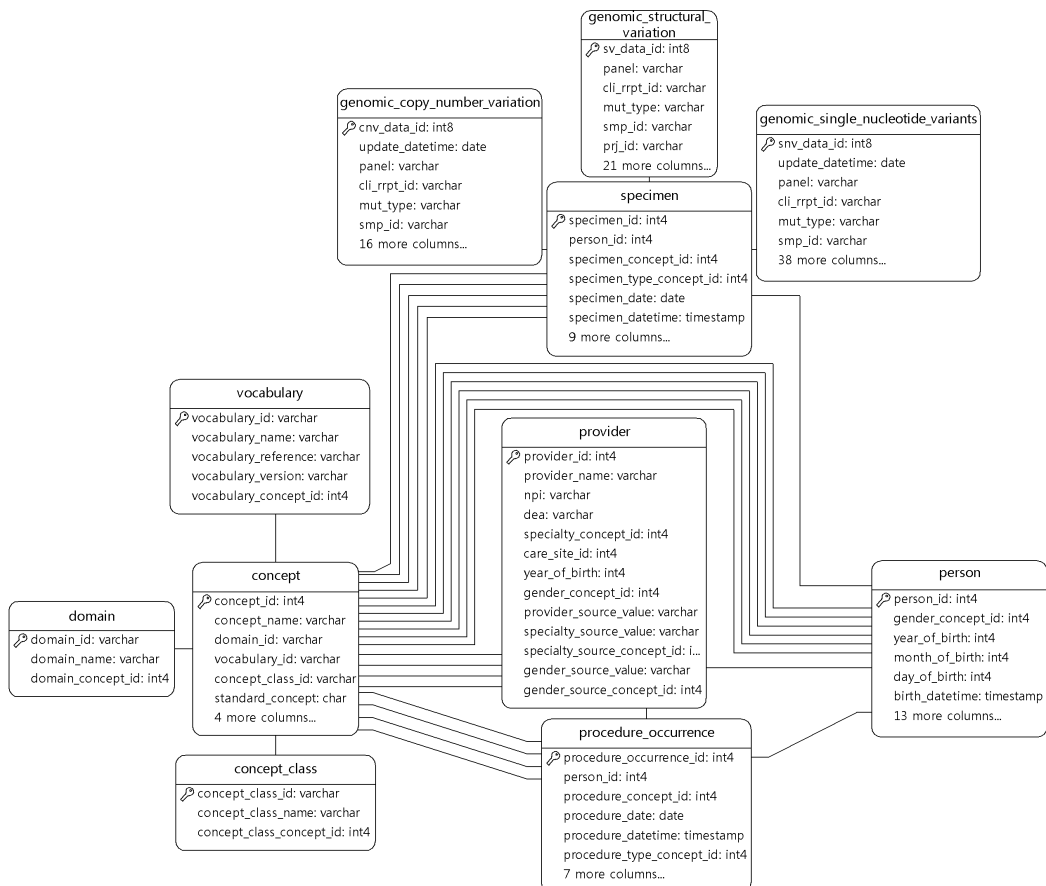
도면1



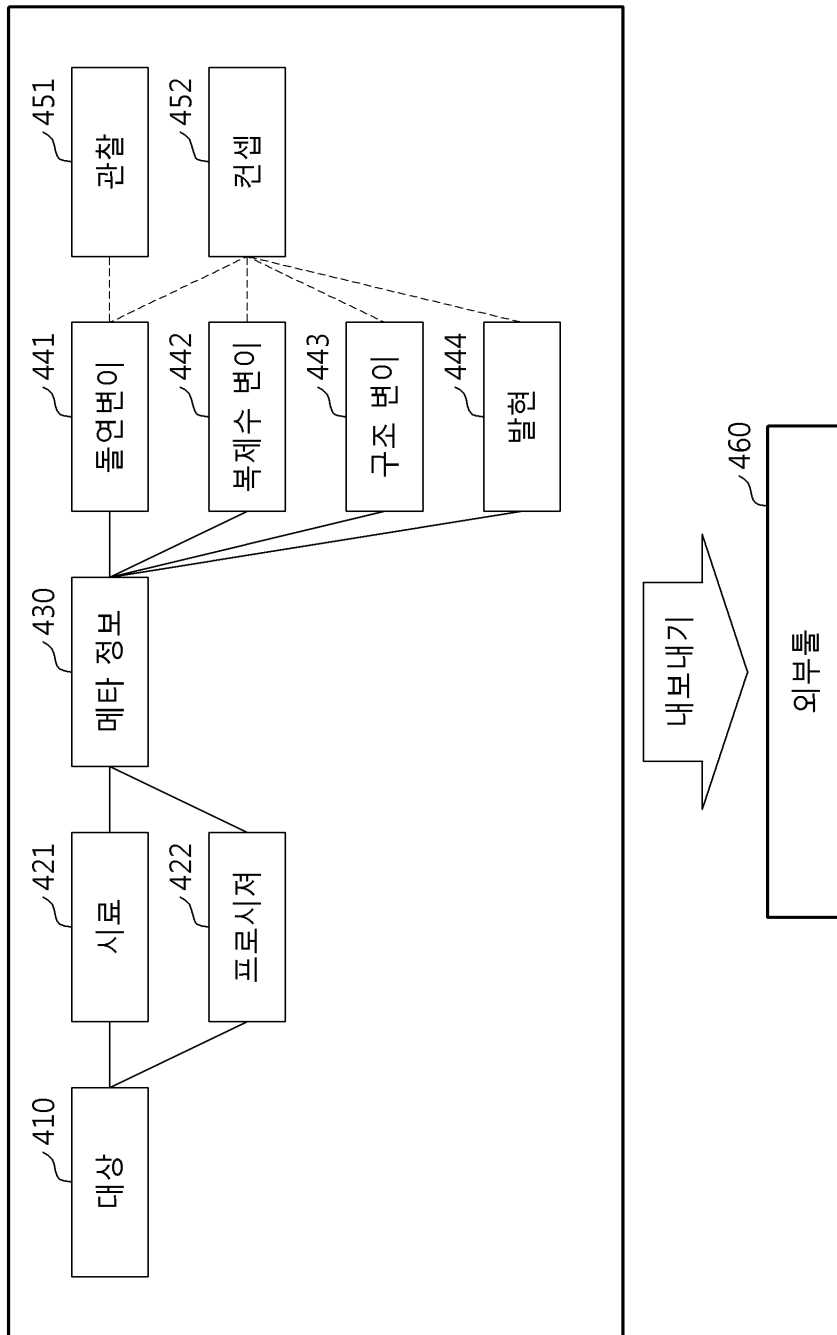
도면2



도면3



도면4



도면5

