



**(19) 대한민국특허청(KR)**  
**(12) 공개특허공보(A)**

(11) 공개번호 10-2019-0080832  
(43) 공개일자 2019년07월08일

<p>(51) 국제특허분류(Int. Cl.)  <i>G16B 50/00</i> (2019.01) <i>G16B 20/00</i> (2019.01)  <i>G16H 10/00</i> (2018.01) <i>G16H 20/10</i> (2018.01)  <i>G16H 50/20</i> (2018.01)</p> <p>(52) CPC특허분류  <i>G16B 50/00</i> (2019.02)  <i>G16B 20/00</i> (2019.02)</p> <p>(21) 출원번호 10-2019-0070025(분할)                  (22) 출원일자 2019년06월13일                  심사청구일자 없음                  (62) 원출원 특허 10-2017-0058507                  원출원일자 2017년05월11일                  심사청구일자 2017년05월11일</p> <p>(30) 우선권주장                  1020160065274 2016년05월27일 대한민국(KR)</p>	<p>(71) 출원인  <b>(주) 메디젠휴먼케어</b>                  서울특별시 강남구 테헤란로 223 , 20층(역삼동, 큰길타워빌딩)</p> <p>(72) 발명자  <b>신동직</b>                  서울특별시 영등포구 문래로4길 6 현대2차아파트 201-309  <b>정현경</b>                  서울특별시 솔샘로 174, 131동 1003호(미아동, SK 북한산시티아파트)  <b>배윤선</b>                  인천광역시 서구 가정로 387, 110동 2203호(신현동, 신현이편한세상하늘채)</p> <p>(74) 대리인  <b>이재영</b></p>
--	--

전체 청구항 수 : 총 21 항

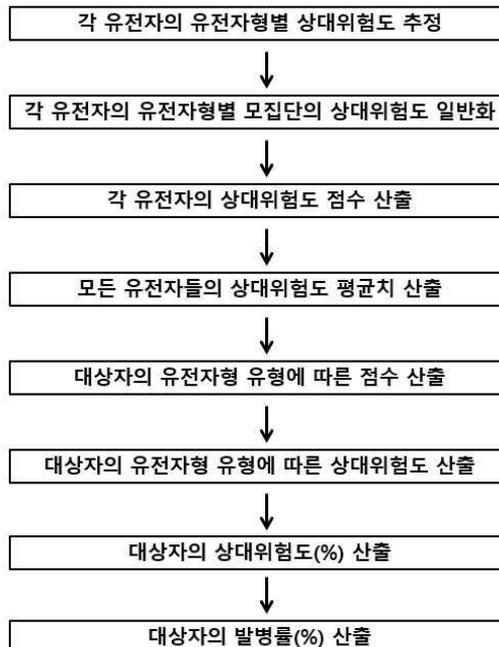
(54) 발명의 명칭 **SNP를 이용한 질병 관련 유전체 분석 시스템 및 장치**

**(57) 요약**

본 발명은 SNP를 이용한 질병 관련 유전체 분석 시스템 및 장치에 관한 것이다. 종래의 질병 및 약물반응 예측 유전자 검사 시스템은 의뢰자의 요구 및 의심 질환의 종류에 따른 사전상담, 접수, 개인 유전체 변이형 확인(실험), 예측, 결과보고서와 같은 일련의 과정 및 이와 관련된 정보 수집과 적용에 대한 기술로서 이루어지고 있는

(뒷면에 계속)

**대표도** - 도5



데, 광범위하게 발표되고 있는 데이터의 정확한 인식과 향후 의뢰자에게 전달하는 객관적이고 구체적인 보고서에 대한 논의가 미흡한 사항으로 인해 신뢰성 문제가 발생되고 있는 실정이다. 본 발명에 따른 유전체 분석 시스템 및 장치는 장치는 1단계로 질병 및 약물반응 관련 데이터베이스, 연구 데이터베이스, 및 유전자 데이터베이스로부터 개선된 알고리즘에 의해 객관적이고 구체적인 SNP-질병 연관성을 도출하고, 2단계로 상기 1단계에서 도출된 특정 질병 관련 SNP들의 복합성을 분석하여 최종 질병 위험도를 산출하므로, 질병 예측 결과의 정확도를 향상시키는 효과가 있을 것으로 기대된다.

(52) CPC특허분류

*G16H 10/00* (2018.01)

*G16H 20/10* (2018.01)

*G16H 50/20* (2018.01)

## 명세서

### 청구범위

#### 청구항 1

- (a) 각 유전자의 유전자형별 상대위험도를 추정하는 단계;
- (b) 각 유전자의 유전자형별 모집단의 상대위험도를 일반화하는 단계;
- (c) 각 유전자의 상대위험도 점수를 산출하는 단계;
- (d) 모든 유전자들의 상대위험도 평균치를 산출하는 단계;
- (e) 대상자의 유전자형에 따른 점수를 산출하는 단계; 및
- (f) 대상자의 유전자형에 따른 상대위험도를 산출하는 단계를 포함하는, 질병별 발생 위험 가중치를 산출하는 방법.

#### 청구항 2

제 1항에 있어서,

상기 (a) 단계에서의 상대위험도는  $\text{오즈비}/((1-\text{유병률})+(\text{유병률}*\text{오즈비}))$ 로 계산하는 것을 특징으로 하는, 질병별 발생 위험 가중치를 산출하는 방법.

#### 청구항 3

제 1항에 있어서,

상기 (b) 단계에서의 모집단의 상대위험도는  $\text{상대위험도}*\text{해당 유전자형 빈도}$ 로 계산하는 것을 특징으로 하는, 질병별 발생 위험 가중치를 산출하는 방법.

#### 청구항 4

제 1항에 있어서,

상기 (c) 단계에서의 상대위험도 점수는 각 유전자의 유전자형별 상대위험도 점수의 합으로 계산하는 것을 특징으로 하는, 질병별 발생 위험 가중치를 산출하는 방법.

#### 청구항 5

제 1항에 있어서,

상기 (d) 단계에서의 상대위험도 평균치는 각 유전자의 상대위험도 점수들의 곱으로 계산하는 것을 특징으로 하는, 질병별 발생 위험 가중치를 산출하는 방법.

#### 청구항 6

제 1항에 있어서,

상기 (e) 단계에서의 점수는 각 유전자의 대상자 유전자형에 따른 상대위험도의 곱으로 계산하는 것을 특징으로 하는, 질병별 발생 위험 가중치를 산출하는 방법.

**청구항 7**

제 1항에 있어서,

상기 (f) 단계에서의 상대위험도는 상기 (e)단계에서의 점수/ 상기 (d)단계에서의 평균치로 계산하는 것을 특징으로 하는, 질병별 발생 위험 가중치를 산출하는 방법.

**청구항 8**

(a) 의뢰된 검체에서 DNA를 추출하는 단계;

(b) 상기 DNA로부터 유전정보를 획득하는 단계;

(c) 상기 유전정보와 제 1 내지 제 3 데이터베이스에서의 질병-SNP 연관성 결과를 대조하여 특정 질병의 위험도를 측정하는 단계;

(d) 상기 의뢰된 검체의 유전정보에 특정 질병 관련된 SNP가 2개 이상 있을 경우에 제 1항 내지 제 7항 중 어느 하나 이상의 방법으로 산출된 가중치를 부과하여 유전자형에 따른 상대위험도를 구하는 단계;

(e) 의뢰된 검체의 상대위험도(%) 및 발병률(%)을 구하는 단계; 및

(f) 의뢰된 검체의 질병 발생 위험도를 판정하는 단계를 포함하는, 질병 예측 방법.

**청구항 9**

제 8항에 있어서,

상기 (c) 단계에서의 제 1 데이터베이스는 특정 SNP와 연관된 특정 질병의 증상, 처방 약물의 종류, 처방 약물의 농도, 약물 처방의 빈도, 약물 처방의 기간, 및 부작용에 관한 정보를 포함하는 질병 및 약물반응 관련 데이터베이스인, 질병 예측 방법.

**청구항 10**

제 8항에 있어서,

상기 (c) 단계에서의 제 2 데이터베이스는 특정 SNP와 연관된 특정 질병에 대한 연구 논문을 포함하고, 상기 논문의 논문고유번호(PMID), 연구대상, 연구방법, 연구기간, 연구결과, 저널정보 및 연구의 반복성 정보를 포함하는 연구 데이터베이스인, 질병 예측 방법.

**청구항 11**

제 8항에 있어서,

상기 (c) 단계에서의 제 3 데이터베이스는 특정 질병과 연관된 특정 SNP의 염색체 번호, 유전자위, 및 대립형질 정보를 포함하는 유전자 데이터베이스인, 질병 예측 방법.

**청구항 12**

제 8항에 있어서,

상기 (e) 단계에서의 상대위험도(%)는,

상기 (d) 단계에서의 상대위험도가 1 이상( $\geq 1$ )인 경우에는 (대상자의 위험도 평균 점수-1)\*100으로 계산하고,

상기 (d) 단계에서의 상대위험도가 1 미만(<1)인 경우에는 (1-대상자의 위험도 평균 점수)\*100으로 계산하는 것을 특징으로 하는, 질병 예측 방법.

### 청구항 13

제 8항에 있어서,

상기 (e) 단계에서의 발병률(%)은 상기 (d) 단계에서의 상대위험도\*유병률로 계산하는 것을 특징으로 하는, 질병 예측 방법.

### 청구항 14

제 8항에 있어서,

상기 (f)단계에서의 판정은,

상기 (e)단계에서의 상대위험도(%)가 1 이하( $\leq 1$ )이면 표준으로 판정하고, 1 초과(>1)이면 주의 또는 집중으로 판정하는 것을 특징으로 하는, 질병 예측 방법.

### 청구항 15

(a) 의뢰된 검체에서 DNA를 추출하는 추출부;

(b) 상기 DNA로부터 유전정보를 획득하는 입력부;

(c) 상기 유전정보와 제 1 내지 제 3 데이터베이스에서의 질병-SNP 연관성 결과를 대조하여 특정 질병의 위험도를 측정하는 비교부;

(d) 상기 의뢰된 검체의 유전정보에 특정 질병 관련된 SNP가 2개 이상 있을 경우에 제 1항 내지 제 7항 중 어느 하나 이상의 방법으로 산출된 가중치를 부과하여 유전자형에 따른 상대위험도를 구하는 연산부;

(e) 의뢰된 검체의 상대위험도(%) 및 발병률(%)을 구하는 산출부; 및

(f) 의뢰된 검체의 질병 발생 위험도를 판정하는 판정부를 포함하는, 질병 예측 장치.

### 청구항 16

제 15항에 있어서,

상기 (c) 단계에서의 제 1 데이터베이스는 특정 SNP와 연관된 특정 질병의 증상, 처방 약물의 종류, 처방 약물의 농도, 약물 처방의 빈도, 약물 처방의 기간, 및 부작용에 관한 정보를 포함하는 질병 및 약물반응 관련 데이터베이스인, 질병 예측 장치.

### 청구항 17

제 15항에 있어서,

상기 (c) 단계에서의 제 2 데이터베이스는 특정 SNP와 연관된 특정 질병에 대한 연구 논문을 포함하고, 상기 논문의 논문고유번호(PMID), 연구대상, 연구방법, 연구기간, 연구결과, 저널정보 및 연구의 반복성 정보를 포함하는 연구 데이터베이스인, 질병 예측 장치.

### 청구항 18

제 15항에 있어서,

상기 (c) 단계에서의 제 3 데이터베이스는 특정 질병과 연관된 특정 SNP의 염색체 번호, 유전자위, 및 대립형질 정보를 포함하는 유전자 데이터베이스인, 질병 예측 장치.

**청구항 19**

제 15항에 있어서,

상기 (e) 단계에서의 상대위험도(%)는,

상기 (d) 단계에서의 상대위험도가 1 이상( $\geq 1$ )인 경우에는 대상자의 위험도 평균 점수-1)\*100으로 계산하고,

상기 (d) 단계에서의 상대위험도가 1 미만( $< 1$ )인 경우에는 (1-대상자의 위험도 평균 점수)\*100으로 계산하는 것을 특징으로 하는, 질병 예측 장치.

**청구항 20**

제 15항에 있어서,

상기 (e) 단계에서의 발병률(%)은 상기 (d) 단계에서의 상대위험도\*유병률로 계산하는 것을 특징으로 하는, 질병 예측 장치.

**청구항 21**

제 15항에 있어서,

상기 (f)단계에서의 판정은,

상기 (e)단계에서의 상대위험도(%)가 1 이하( $\leq 1$ )이면 표준으로 판정하고, 1 초과( $> 1$ )이면 주의 또는 집중으로 판정하는 것을 특징으로 하는, 질병 예측 장치.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 SNP를 이용한 질병 관련 유전체 분석 시스템 및 장치에 관한 것이다.

**배경 기술**

[0003] 단일염기 다형성(single nucleotide polymorphism, SNP)이란 염색체의 단일부위에서 여러 가지 DNA 염기들 중의 하나에 나타나는 일반적인 돌연변이로 인간의 게놈(genome)에는 약 3백만 개의 SNP가 존재한다. SNP는 그 빈도가 높고 안정하며 유전체 전체에 분포되어 있고 이에 의하여 개인의 유전적 다양성이 발생한다. 이러한 SNP 차이는 서로 다른 질병에 대한 감수성의 차이를 만들어 낸다. 따라서 최근에는 SNP 정보를 바탕으로, 인간의 질병에 관여하는 적절한 유전자(질환 후보 유전자)들을 선정하고 이들의 변이형을 발굴하고 질환과의 연관성을 통계적으로 확인하는 과정을 포함하는 “질병 예측 유전자 검사 기술”이 개발되고 있다(US2008-0020484, KR1483284 등). 이러한 질병 및 약물반응 예측 유전자 검사 서비스는 특정 유전질환의 고위험 가족 및 친척들에게 시행되는 검사, 치료 및 처방하기 전에 개인적인 특성을 보고 관련 질병 및 약물반응 유전체 내의 변화를 검사해 발병 확률을 예측 및 예후를 목적으로 한다.

[0004] 따라서 정확한 질병 예측을 위해 “질병-유전정보 연관성 데이터베이스”의 정확성은 매우 중요하다 할 것이다. 그러나 종래의 질병 및 약물반응 예측 유전자 검사 시스템은 의뢰자의 요구 및 의심 질환의 종류에 따른 사전상담, 접수, 개인 유전체 변이형 확인(실험), 예측, 결과보고서와 같은 일련의 과정 및 이와 관련된 정보 수집과

적용에 대한 기술로서 이루어지고 있는데, 광범위하게 발표되고 있는 데이터의 정확한 인식과 향후 의뢰자에게 전달하는 객관적이고 구체적인 보고서에 대한 논의가 미흡한 사항으로 인해 신뢰성 문제가 발생되고 있는 실정이다.

[0005] 본 연구는 SNP를 이용한 질병 관련 유전체 분석 시스템 및 장치에 관한 것으로, 본 발명에 따른 시스템은 개선된 알고리즘으로서 정확도 높은 질병 예측 결과를 제공할 것으로 기대된다.

**발명의 내용**

**해결하려는 과제**

[0007] 본 발명은 상기와 같은 종래의 기술상의 문제점을 해결하기 위해 안출된 것으로, SNP를 이용한 질병 관련 유전체 분석 시스템 및 장치에 관한 것이다.

[0008] 그러나 본 발명이 이루고자 하는 기술적 과제는 이상에서 언급한 과제에 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 당 업계에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

**과제의 해결 수단**

[0010] 이하, 본원에 기재된 다양한 구체예가 도면을 참조로 기재된다. 하기 설명에서, 본 발명의 완전한 이해를 위해서, 다양한 특이적 상세사항, 예컨대, 특이적 형태, 조성물 및 공정 등이 기재되어 있다. 그러나, 특정의 구체예는 이들 특이적 상세 사항 중 하나 이상 없이, 또는 다른 공지된 방법 및 형태와 함께 실행될 수 있다. 다른 예에서, 공지된 공정 및 제조 기술은 본 발명을 불필요하게 모호하게 하지 않게 하기 위해서, 특정의 상세사항으로 기재되지 않는다. "한 가지 구체예" 또는 "구체예"에 대한 본 명세서 전체를 통한 참조는 구체예와 결부되어 기재된 특별한 특징, 형태, 조성 또는 특성이 본 발명의 하나 이상의 구체예에 포함됨을 의미한다. 따라서, 본 명세서 전체에 걸친 다양한 위치에서 표현된 "한 가지 구체예에서" 또는 "구체예"의 상황은 반드시 본 발명의 동일한 구체예를 나타내지는 않는다. 추가로, 특별한 특징, 형태, 조성, 또는 특성은 하나 이상의 구체예에서 어떠한 적합한 방법으로 조합될 수 있다.

[0012] 명세서에서 특별한 정의가 없으면 본 명세서에 사용된 모든 과학적 및 기술적인 용어는 본 발명이 속하는 기술 분야에서 당업자에 의하여 통상적으로 이해되는 것과 동일한 의미를 가진다.

[0014] 본 발명의 일 구체예에서 “유전정보(genetic information)”란, DNA의 염기배열로서 부호화되는 모든 정보를 총칭하는 광의의 개념으로서, 본 발명에 있어서 유전정보란 개체의 염기 다형성 정보를 포함한다.

[0016] 본 발명의 일 구체예에서 “단일염기 다형성(single nucleotide polymorphism, SNP)”이란, 염색체의 단일부위에서 여러 가지 DNA 염기들 중의 하나에 나타나는 일반적인 돌연변이로, 인간의 게놈(genome)에는 약 3백만 개의 SNP가 존재하여 약 500 ~ 1,000염기당 1개꼴로 나타나며, 그 중 약 20만개가 단백질을 만드는 유전자에 존재하는 cSNP일 것으로 추정된다. SNP는 그 빈도가 높고 안정하며 유전체 전체에 분포되어 있고 이에 의하여 개인의 유전적 다양성이 발생한다. 즉 DNA사슬의 특정부위에 어떤 사람은 아데닌(adenine; A)을 가지고 있는 반면 어떤 사람은 시토신(cytosine; C)을 가지고 있는 것이다. 이런 미세한 차이(SNP)에 의하여 각 유전자의 기능이 달라질 수 있고 이런 것들이 상호 작용하여 서로 다른 모양의 사람을 만들고 서로 다른 질병에 대한 감수성의 차이를 만들어 낸다. 즉 간염에 걸리는 사람과 걸리지 않는 사람간의 유전적 차이를 찾아낼 수 있다면 어떤 이유에서 간염에 대한 감수성이 달라지는지의 기능을 알아낼 수 있게 된다. 그렇게 된다면 이를 이용하여 간염의 예방이나 치료에 사용되는 약품을 개발할 수 있을 것이라는 것이 인간유전체 연구의 궁극적인 목적인 것이다. 이에 세계적인 거대 제약회사들과 게놈 연구기관들은 앞으로 SNP가 신약개발의 원천적인 정보를 제공할 수 있다고 판단하고 단일염기다형성 컨소시엄(The SNP consortium; TSC)을 형성하여 공동으로 인류의 영원한 이상이었던 무병장수의 꿈을 앞당기려 SNP 연구에 집중하고 있다. 그러나 수많은 SNP가 개발되어 있다고 하더라도 SNP

자체만으로는 아무런 의미가 없다. 즉 SNP를 비교 분석할 대상이 없다면 이는 무용지물인 것이다. 따라서 국내의 제약회사나 연구기관들은 그들이 많이 가지고 있는 심장병, 치매, 에이즈(AIDS) 등등의 질병에 대한 비교 대상(환자의 DNA와 임상자료)을 확보하고 어떤 SNP가 어떤 질병과 연관되어 있는지에 대한 데이터베이스를 구축하기 위한 노력을 기울이고 있다.

[0018] 본 발명의 일 구체예에서 “질병 및 약물반응 관련 데이터베이스”란, 질병-SNP 연관성 결과를 비교하여 특정 질병의 위험도를 측정하기 위한 데이터풀(Data pool)을 의미하며, 특정 SNP와 연관된 특정 질병의 증상, 처방 약물의 종류, 처방 약물의 농도, 약물 처방의 빈도, 약물 처방의 기간, 및 부작용에 관한 정보를 포함한다. 본 발명에서의 “질병 및 약물반응 관련 데이터베이스”는 국내의 식약청, 의료기관, 및 건강검진센터로부터 제공 받은 질병 및 약물 정보를 데이터풀에 포함시킬 수 있고, 특정 개체의 나이 및 성별 정보와 배우자, 자식, 부모, 사촌 등에 대한 병력 가계도 정보를 포함시킬 수 있다.

[0020] 본 발명의 일 구체예에서 “연구 데이터베이스”란, 질병-SNP 연관성 결과를 비교하여 특정 질병의 위험도를 측정하기 위한 데이터풀을 의미하며, 연구 자료는 임상 또는 학술 논문일 수 있으나, 이에 한정하는 것은 아니다. 데이터풀(Data pool) 내의 자료가 논문으로부터 도출된 자료일 경우에, 데이터풀은 논문의 논문고유번호(PMID), 연구대상, 연구방법, 연구기간, 연구결과, 저널정보 및 연구의 반복성 정보를 포함시킬 수 있고, 연구에 대상이 되는 개체의 나이, 및 성별 정보와 배우자, 자식, 부모, 사촌 등에 대한 병력 가계도 정보를 포함시킬 수 있다.

[0022] 본 발명의 일 구체예에서 “유전자 데이터베이스”란, 질병-SNP 연관성 결과를 비교하여 특정 질병의 위험도를 측정하기 위한 데이터풀(Data pool)을 의미하며, 특정 질병과 연관된 특정 SNP의 염색체 번호, 유전자위, 및 대립형질 정보를 포함하는 유전자 정보를 데이터풀에 포함시킬 수 있다. 특히, 유전자 데이터베이스 내의 저장된 자료에는 분석의 대상이 되는 개체의 인종 정보가 중요한 요소로 작용할 수 있으나, 이에 한정하는 것은 아니다.

[0024] 본 발명의 일 구체예에서, “오즈비(odds ratio)”란, 상대 위험도의 추정치라 불리는 지표로 환자-대조군 연구(case-control study)에서 추정하는 값이다. 상대위험도(relative risk)는 코호트 연구(Cohort study)에서 추정하는 값으로 위험 인자가 있는 경우 어떤 사건 발생 확률이 위험 인자가 없는 경우의 사건 발생할 확률의 비로 정의된다. 코호트 연구의 경우 위험인자를 미리 설정하고 시간의 흐름에 따라 사건의 발생 유무를 관찰하기 때문에 위험인자에 따른 사건 유무의 연관성을 나타내는 상대위험도를 신뢰할 수 있지만, 본 발명의 기술 분야인 환자-대조군 연구의 경우, 사건 발생 여부에 따라 집단을 구성한 뒤 위험인자의 여부를 분류하기 때문에 상대 위험도는 의미가 없으므로, 그 추정치인 오즈비를 사용한다.

[0026] 본 발명의 일 구체예에서 “알고리즘”이란, 주어진 문제를 해결하기 위해 컴퓨터 프로그래밍이 수행해야 할 과정들을 나타낸 것이다. 일정한 순서에 따라 기계적으로 처리하면 반드시 목적인 결과를 얻을 수 있을 때 그 일정한 순서를 목적에 대한 알고리즘이라고 한다. 일반적으로 알고리즘을 알고 있는 것은 컴퓨터의 프로그램으로 변환하여 처리할 수 있다. 본 명세서에 있어서 알고리즘은 특정 개체의 유전정보를 질병 및 약물반응 관련 데이터베이스, 연구 데이터베이스, 및 유전자 데이터베이스에 저장된 정보들과 대조하여 특정 질병이 발생할 확률(위험도)을 도출하거나, 특정 질병에 감수성이 높은 약물 후보군을 도출하거나, 또는 부작용 위험이 높은 약물 후보군을 도출하는 것이나, 이에 한정하는 것은 아니다.

[0028] 본 발명의 일 구체예에서 “타겟 질환”이란, 본 발명의 알고리즘으로 질병 발생 위험도 및 발병률을 예측 가능한 질환을 의미한다. 본 발명에 있어서, 상기 타겟 질환은 크게 만성질환, 암 질환, 약물반응 민감성 질환, 및 기타질환으로 분류할 수 있고, 구체적으로 상기 만성 질환은 제1형 당뇨병, 제2형 당뇨병, C형간염, 가와사키, 강직성척추염, 건선, 결핵, 고혈압, 골관절염, 골다공증, 관상동맥질환, 궤양성 대장염, 기면증, 녹내장, 뇌동맥류, 뇌졸중, 다낭성난소증후군, 다발성경화증, 담석증, 루게릭병, 루푸스, 류마티스관절염, 류마티스 심장질환, 만성신장질환, 무릎골관절염, 병적근시(고도근시), 베체트병, 백내장, 백반증, 비만, 비알코올성지방간, 심

근경색, 심방세동, 아스피린 과민성 만성 두드러기, 아토피피부염, 음식에 대한 알러지과민반응, 임신성당뇨병, 임신성중독증, 중성지질수치, 천식, 추간판탈출증, 치매, 크론병, 통풍, 파킨슨병, 폐쇄성폐질환, 피지만성질환, 관상동맥 심장질환, 편두통 및, 황반변성으로 구성되는 것이고, 상기 암질환은 간암, 갑상선암, 고환암, 구강암, 급성골수성 백혈병, 난소암, 담도계암, 대장암, 두경부암, 미만성 위암, 방광암, 소아백혈병, 식도암, 신장암, 위암, 유방암, 자궁경부암, 자궁내막암, 전립선암, 췌장암, 폐암, 및 피부암으로 구성되는 것이며, 상기 약물반응 민감성 질환은 메타암페타민 유발정신병, 안지오텐신 전환효소 억제제, 와파린 약 민감성, 및 프로포폴 마취제 민감성으로 구성되는 것이며, 상기 기타질환은 ADHD, 공황장애, 니코틴 중독성, 알코올의존성, 양극성장애, 우울증, 자폐증, 및 정신분열으로 구성되는 것이나, 이에 한정하는 것은 아니다.

[0030] 또한 상기 타겟 질환은 유병률 산출 가능 여부에 따라 구분 가능하다. 상기의 제1형 당뇨병, 제2형 당뇨병, C형 간염, 가와사키, 강직성척추염, 건선, 결핵, 고혈압, 골관절염, 골다공증, 관상동맥질환, 췌양성 대장염, 기면증, 녹내장, 뇌동맥류, 뇌졸중, 다낭성난소증후군, 다발성경화증, 담석증, 루게릭병, 루푸스, 류마티스관절염, 류마티스 심장질환, 만성신장질환, 무릎골관절염, 병적근시(고도근시), 베체트병, 백내장, 백반증, 비만, 비알코올성지방간, 심근경색, 심방세동, 아토피 피부염, 음식에 대한 알러지 과민반응, 임신성당뇨병, 임신성중독증, 천식, 추간판탈출증, 치매, 크론병, 통풍, 파킨슨병, 피지만성질환, 황반변성, 간암, 갑상선암, 고환암, 구강암, 급성골수성 백혈병, 난소암, 담도계암, 대장암, 두경부암, 미만성 위암, 방광암, 소아백혈병, 식도암, 신장암, 위암, 유방암, 자궁경부암, 자궁내막암, 전립선암, 췌장암, 폐암, 피부암, ADHD, 공황장애, 알코올의존성, 양극성장애, 우울증, 자폐증, 및 정신분열은 선행 공지된 데이터베이스에 의해 유병률 산출 가능이 용이한 질환군이고, 상기 아스피린 과민성 만성 두드러기, 중성지질수치, 메타암페타민 유발정신병, 안지오텐신 전환효소 억제제, 와파린 약 민감성, 프로포폴 마취제 민감성, 폐쇄성폐질환, 및 니코틴 중독성은 선행 공지된 데이터베이스에 의해 유병률 산출 가능이 용이하지 않은 질환군으로 분류 가능하다, 이에 한정하는 것은 아니다.

[0032] 본 발명에 있어서, 상기 유병률 산출이 용이한 질환군은 유병률과 해당 인종의 유전형 빈도를 이용하여 모집단 내의 모든 유전자들의 가중 평균 점수를 구하고, 모집단 기반으로 대상자의 상대 위험도 및 발병률을 예측하는 것으로 알고리즘을 설정한다. 이에 따른 판정기준은 위험도 평균점수와 유전자형(genotype) 빈도(%)를 고려하여 대상자의 위험도 평균 점수가  $\leq 1$ 인 경우 "표준단계", 대상자의 위험도 평균 점수  $> 1$ 인 경우에는 "주의관리단계"와 "집중관리단계"로 판정한다. 이때, 검사 대상자의 위험도 평균점수보다 높은 위험도 평균점수를 가지는 유전자형 조합의 빈도 합이 5% 이하이면 "집중관리단계", 5% 초과면 "주의관리단계"로 판정한다. 알고리즘의 보완 개전에 따라 주의관리단계와 집중관리단계를 구분하는 빈도에 대한 기준 값은 조절 가능하다. 상기 유병률 산출이 용이하지 않은 질환군에 대해서는 본 발명의 제 1 내지 제 3 데이터베이스로부터 확인된 유전형에 3단계(낮음/보통/높음)로 판정을 도출하여, 특정 질환 및 약물 반응에 대한 질병 발생 위험도 및 발병률 예측모델을 생성 할 수 있다.

[0034] 본 발명은 SNP를 이용한 질병 관련 유전체 분석 시스템 및 장치를 제공한다. 보다 구체적으로, 본 발명은 질병 및 약물 반응에 대한 유전자예측을 해 주는데 있어서, 전반적인 질병 및 약물 반응에 대한 원인, 진단, 예방 등에 대한 정보와 유전적 요인으로 이와 관련된 유전자 검색, 검색된 유전자의 SNP명, 위치 등 유전자 세부 정보, 통계적으로 관련 유전자임을 입증하거나 검증하기 위해 연구된 자료 들을 저장한다. 저장된 자료들을 검토하여 유전자의 우선순위와 유전자관련 연구의 우선순위를 대립형질 관련성 연구의 일반적인 특성과 유전자 관련 일반적인 관점으로 재검토하여 통계적인 예측 알고리즘에 적용할 SNP와 관련 값들을 선정한다. 따라서, 각 질병 및 약물 반응에 대해 통계적인 예측 모델이 생성된다.

[0035] 상기 통계적인 예측 모델 생성은, 관심 질병 및 약물반응에 대한 전반적인 정보를 검색하는 단계, 원인 중 유전적 요인으로 관련 유전자 검색하는 단계, 그리고 유전자 정보를 검색하는 단계로 구성된다.

[0036] 관심 질병 및 약물반응에 대한 전반적인 정보 검색하는 단계는 해당 질병 및 약물 반응에 대한 정보를 수집하는 단계로, 정의, 원인, 진단, 처치, 예방 및 관리 등을 확인하고 유전적인 요인으로 인해 발병될 수 있는 여부를 검토하는 과정이다.

[0037] 관련 유전자 검색하는 단계는 대립형질 관련성 연구를 검색하는 단계로, 유전자와의 관련성을 입증한 또는 입증

하러 했던 연구 모두 포함하여 유전적 요인으로 볼 수 있는지를 실험결과로 확인하는 과정이다. 유전자 정보를 검색하는 단계는 관련 유전자의 인종별 분포현황, 다른 유전자와의 LD관계 등을 확인하는 과정이다.

- [0038] 본 발명은 의뢰자가 의뢰한 질병 및 약물반응에 대해 상기에 선정된 유전자 검사 수행 후 해당 질병 및 약물반응에 따른 통계적인 예측 모델을 수행하여 개인 유전자형과 의뢰한 질병 및 약물반응과의 상관관계를 나타내는 결과값을 산출하는 시스템 및 장치를 제공한다.
- [0040] 보다 자세하게, 본 발명은 질병 및 약물 반응을 정확하게 예측하기 위하여 관련 정보를 수집하고, 관련 SNP에 대한 정보를 데이터베이스화하여 후보 SNP를 선별하고, 이를 이용하여 질병 등에 대한 예후를 정확하게 예측하기 위한 예측알고리즘을 적용하여 각각의 예측모델을 생성한다.
- [0041] 또한 질병 및 약물반응 관련 데이터베이스(본 발명에서의 제 1 데이터베이스)에서는 해당 질의어를 통한 검색 후에 질병과 관계가 있는 유전자와 질병에 대한 정보를 담고 있으며, 더불어, 연구대상에 따른 유병률/발생률, 임상정보와 헬스가이드 정보까지 근거출처를 포함하여 자료를 저장한다.
- [0042] 연구 데이터베이스(본 발명에서의 제 2 데이터베이스)에는 관련 유전자라 입증 또는 입증하려 한 논문의 논문고유번호(PubMed identifier, PMID), 연구대상, 연구방법, 연구기간, 연구결과, 문헌정보 등 자료를 저장한다.
- [0043] 유전자 데이터베이스(본 발명에서의 제 3 데이터베이스)에서는 질병 데이터베이스에서 검색된 유전자를 질의어로 검색 후에, 특정 질병과 연관된 특정 SNP의 특성(trait) 정보, 염색체 번호, 유전자위, 및 대립형질 정보 등 자료를 저장한다.
- [0045] 상기 제 1 내지 제 3의 데이터베이스로부터 도출된 질병-SNP 연관성 결과를 의뢰된 검체의 유전정보와 대조하여 의뢰된 검체를 제공한 개체에게서 특정 질병이 발생할 확률(위험도)을 도출하거나, 특정 질병에 감수성이 높은 약물 후보군을 도출하거나, 또는 부작용 위험이 높은 약물 후보군을 도출한다. 상기 대조에는 추가적으로 의뢰된 검체에 대하여 나이, 및 성별 정보와 배우자, 자식, 부모, 사촌 등에 대한 병력 가계도 정보, 질환관련 소재지와 관련된 환경성 요인 정보, 및/또는 습관 정보, 영양상태 정보, 생활 습관 정보 및 운동 수행 정보를 반영하여 결과를 도출할 수 있다. 이 때, 의뢰된 검체의 유전정보에 특정 질병의 발생과 연관도가 높은 SNP가 2개 이상 있을 경우에 질병 발생 위험도에 가중치를 부여할 수 있고, SNP들간의 상관성에 따라 차등적으로 부여할 수 있으며, 이 때에는 SNP들간의 상관성이 높을수록 높은 수치의 가중치를 부여할 수 있다.
- [0047] 상기에 기술한 의뢰된 검체를 제공한 개체에게서 특정 질병이 발생할 확률(위험도)을 도출하는 알고리즘은 제 1 내지 제 3의 데이터베이스로부터 질병-SNP 연관성 결과를 도출하는 과정에서 연구방법, 연구대상, 연구기간, 연구결과, 문헌정보 등에 정보 선정을 위한 우선순위를 부여할 수 있다. 예를 들어 데이터풀이 제 2 데이터베이스인 경우에, 우선순위는 연구가 GWAS분석을 수행하고 반복 연구되었는지, 최근에 게재된 논문인지, 문헌은 피인용지수가 높은 것인지, 결과값(위험도, 신뢰구간, p-value 등)은 어떠한지 등이 될 수 있다.
- [0048] 또한 제 1 내지 제 3의 데이터베이스를 통합하여 우선순위를 결정하는 경우에는, 이에 한정하는 것은 아니나, 우선순위는 하기와 같을 수 있다.
- [0049] 1. 대상질환의 유전력(heritability) 확인: 대상 질병에 대한 유전적 요인이 얼마나 작용하는지, 유전력에 따라 연구방법의 선정 및 연관성 연구의 성공 가능성들을 예측하여 연구를 진행하게 됨으로, 대상질환의 유병율/발병률을 함께 조사한다.
- [0050] 2. 연구대상 표현형(형질)의 선별 확인: 환자군을 나누는 기준으로 질병의 발생에 중요한 생화학적 표현형으로 선별하였는지 조사한다. 이는 질병에 관련된 중간 표현형(intermediate phenotype)이라 하며, 이는 복합적으로 발생하는 질병보다는 특정 유전자와의 연관성이 보다 강하게 나타날 것으로 기대하여 질병보다는 이 표현형으로 연구되었는지를 조사하는데 의미가 있다.
- [0051] 3. 분석에 사용된 연구대상의 수(sample size) 확인: 연구대상의 수는 검정력에 영향을 미치게 된다. 통계적 유의성은 유의수준으로 질환과 실제로 연관이 없는 유전변이형을 통계적 검정을 통해 유의성이 있다라고 판단할 확률(false positive)을 의미하고, 이와 반대로 질환과 관련이 있는 유전자를 유의성이 없다고 잘못 판단한 확률(false negative)라 하며 이를 확률 1에서 뺀 값을 검정력(power)라고 한다. 이는 유전모델, 유전자 빈도, 질

병에 대한 상대적 위험도, 연구대상의 수에 의해 결정이 되기 때문에 이를 확인하는데 의미가 있다.

- [0052] 4. 연구대상의 인증 확인: 다른 유전적 배경을 지닌 다수의 인증을 포함하는 경우, 인증에 따라 다른 유전형의 빈도가 질환과의 관련성보다는 인증의 차이에 의해 기인될 확률이 매우 높아지므로, 이질적인 인증집단의 연구인지, 동질적인 인증집단의 연구인지 확인하는데 의미가 있다. 만약 이질적인 인증집단인 경우는 집단간의 차이를 고려했는지 여부도 확인해야 한다.
- [0053] 5. 연구대상의 유전체 시료, 임상정보, 표현형 및 환경정보 수집 확인: 시료의 형태와 DNA 추출방법, 유전적 복잡성과 표현형적 복잡성으로 인해 정확한 임상정보의 수집과 질병의 세부유형 분류가 이루어졌는지, 복합질병이 대부분이므로 다양한 종류의 환경적 요인을 수집하여 분석에 적용하였는지 여부를 확인해야 한다.
- [0054] 6. Genotyping용 후보 유전자 및 후보 SNP 선정 방법 확인: 질병과의 연관성을 본 SNP 선정방법 및 유전자 빈도 등 확인하였는지 여부를 확인하여 통계적 검정력을 확인해야 한다.
- [0055] 7. 반복실험 여부 확인: 유의적인 결과가 나온 경우에 통계적 오류인지 아닌지를 증명하기 위해 반복실험을 해야 한다. 이 연구가 통계적 반복실험 또는 기능적 반복실험을 하여 통계적 유의성이 반복적으로 관측되는지를 조사하였는지 여부를 확인하여야 한다. 우선 연구하고자 하는 염기서열변이가 지난 혹은 현재 연구에서 어떻게 선별되었는지 확인하고 반복실험으로 인해 유의적인 결과를 확인해야 한다.
- [0056] 8. 상기 1 내지 7의 사항들로 추려진 연관성 연구에서의 통계분석 결과를 확인한다. SNP의 유전적 특성을 검증하는 통계적 분석은 유전자형 빈도(allele frequency) 및 유전형 빈도(genotype frequency)의 분포상태를 바탕으로 유의성 검정을 수행하게 되어, SNP에 대한 상대적 위험도(odds ratio)를 추정하게 된다. 유전형에 따라서는 유전 모델에 따라 연관성 분석을 수행하여 가능한 유전 유형을 추정하게 된다. 유전모델은 3가지 형태로 우성모델(dominant model), 열성모델(recessive model) 및 부가모델(additive model)이 있다. 이 모델을 통해 SNP가 특정조건에 대해 환자군과 대조군으로 유의하게 분류하는지에 대한 유의성 및 각 SNP의 유전적 특성의 검증하게 된다. 마찬가지로 상대적 위험도를 추정하게 되고, 추정된 상대적 위험도의 신뢰구간과 해당 p-value를 고려하게 된다. 통계분석결과의 오류점검사항으로 연구디자인 관련해서 다중비교에 의한 오류를 확인했는지를 검증하여야 한다.
- [0057] 9. 마지막으로 연관성 높은 연구 논문이 게재된 저널의 피인용지수, 연구발표년도, 같은 연구를 기반으로 반복 연구인지 등을 최종 확인한다. 게재 저널의 피인용지수가 높을수록, 연구발표 년도가 최신일수록, 반복연구를 수행할수록 우선순위를 부여한다.
- [0058] 상기 기술한 사항을 모두 고려하여 우선순위를 책정하고, 제 1 내지 제 3 데이터베이스로부터 정보를 산출하는 알고리즘을 설정한다. 이러한 예측 알고리즘을 적용하여 특정 질환 및 약물반응에 대한 예측모델을 생성하게 된다.
- [0059] 이와 같은 우선순위의 기준은 인간의 자의적 판단 또는 지식의 축적에 의하지 않은 객관적인 판단 기준에 의하기 위해 정보의 축적을 바탕으로 검증유무와 해당 질병, 약물, 특성과 관련 유무 등 객관적인 기준으로 사용한다. 이러한 기준으로 선정된 후보 유전자의 정보와 유병율/발병률을 고려하여 대상자의 예후를 예측하게 된다. 예측된 결과에 따라 대상자에게 질병 정보와 예방방법과 같은 헬스가이드를 제 1 내지 제 3 데이터베이스에서 선별하여 제공할 수 있다.
- [0061] 본 발명의 일 구체예에서, (a) 각 유전자의 유전자형별 상대위험도를 추정하는 단계; (b) 각 유전자의 유전자형별 모집단의 상대위험도를 일반화하는 단계; (c) 각 유전자의 상대위험도 점수를 산출하는 단계; (d) 모든 유전자들의 상대위험도 평균치를 산출하는 단계; (e) 대상자의 유전자형에 따른 점수를 산출하는 단계; 및 (f) 대상자의 유전자형에 따른 상대위험도를 산출하는 단계를 포함하는 질병별 발생 위험 가중치를 산출하는 방법을 제공하고, 상기 (a) 단계에서의 상대위험도는  $\frac{\text{오즈비}}{(1-\text{유병률})+(\text{유병률}*\text{오즈비})}$ 로 계산하는 것을 특징으로 하는 질병별 발생 위험 가중치를 산출하는 방법을 제공하며, 상기 (b) 단계에서의 모집단의 상대위험도는 상대위험도\*해당 유전자형 빈도로 계산하는 것을 특징으로 하는 질병별 발생 위험 가중치를 산출하는 방법을 제공하며, 상기 (c) 단계에서의 상대위험도 점수는 각 유전자의 유전자형별 상대위험도 점수의 합으로 계산하는 것을 특징으로 하는 질병별 발생 위험 가중치를 산출하는 방법을 제공하며, 상기 (d) 단계에서의 상대위험도 평균치는 각 유전자의 상대위험도 점수들의 곱으로 계산하는 것을 특징으로 하는 질병별 발생 위험 가중치를 산출하는 방법을 제공하며, 상기 (e) 단계에서의 점수는 각 유전자의 대상자 유전자형에 따른 상대위험도의 곱으로 계산하는 것을 특징으로 하는 질병별 발생 위험 가중치를 산출하는 방법을 제공하며, 상기 (f) 단계에서의 상대

위험도는 상기 (e)단계에서의 점수/ 상기 (d)단계에서의 평균치로 계산하는 것을 특징으로 하는 질병별 발생 위험 가중치를 산출하는 방법을 제공한다.

[0063] 본 발명의 다른 구체예에서, (a) 의뢰된 검체에서 DNA를 추출하는 단계; (b) 상기 DNA로부터 유전정보를 획득하는 단계; (c) 상기 유전정보와 제 1 내지 제 3 데이터베이스에서의 질병-SNP 연관성 결과를 대조하여 특정 질병의 위험도를 측정하는 단계; (d) 상기 의뢰된 검체의 유전정보에 특정 질병 관련된 SNP가 2개 이상 있을 경우에 제 1항 내지 제 9항 중 어느 하나 이상의 방법으로 산출된 가중치를 부과하여 유전자형에 따른 상대위험도를 구하는 단계; (e) 의뢰된 검체의 상대위험도(%) 및 발병률(%)을 구하는 단계; 및 (f) 의뢰된 검체의 질병 발생 위험도를 판정하는 단계를 포함하는 질병 예측 방법을 제공하고, 상기 (c) 단계에서의 제 1 데이터베이스는 특정 SNP와 연관된 특정 질병의 증상, 처방 약물의 종류, 처방 약물의 농도, 약물 처방의 빈도, 약물 처방의 기간, 및 부작용에 관한 정보를 포함하는 질병 및 약물반응 관련 데이터베이스인 질병 예측 방법을 제공하며, 상기 (c) 단계에서의 제 2 데이터베이스는 특정 SNP와 연관된 특정 질병에 대한 연구 논문을 포함하고, 상기 논문의 논문고유번호(PMID), 연구대상, 연구방법, 연구기간, 연구결과, 저널정보 및 연구의 반복성 정보를 포함하는 연구 데이터베이스인 질병 예측 방법을 제공하며, 상기 (c) 단계에서의 제 3 데이터베이스는 특정 질병과 연관된 특정 SNP의 염색체 번호, 유전자위, 및 대립형질 정보를 포함하는 유전자 데이터베이스인 질병 예측 방법을 제공하며, 상기 (e) 단계에서의 상대위험도(%)는 상기 (d) 단계에서의 상대위험도가 1 이상( $\geq 1$ )인 경우에는 (대상자의 위험도 평균 점수-1)\*100으로 계산하고, 상기 (d) 단계에서의 상대위험도가 1 미만( $< 1$ )인 경우에는 (1-대상자의 위험도 평균 점수)\*100으로 계산하는 것을 특징으로 하는 질병 예측 방법을 제공하며, 상기 (e) 단계에서의 발병률(%)은 상기 (d) 단계에서의 상대위험도\*유병률로 계산하는 것을 특징으로 하는 질병 예측 방법을 제공하며, 상기 (f)단계에서의 판정은 상기 (e)단계에서의 상대위험도(%)가 1 이하( $\leq 1$ )이면 표준으로 판정하고, 1 초과( $> 1$ )이면 주의 또는 집중으로 판정하는 것을 특징으로 하는 질병 예측 방법을 제공한다.

[0065] 본 발명의 또 다른 구체예에서, (a) 의뢰된 검체에서 DNA를 추출하는 추출부; (b) 상기 DNA로부터 유전정보를 획득하는 입력부; (c) 상기 유전정보와 제 1 내지 제 3 데이터베이스에서의 질병-SNP 연관성 결과를 대조하여 특정 질병의 위험도를 측정하는 비교부; (d) 상기 의뢰된 검체의 유전정보에 특정 질병 관련된 SNP가 2개 이상 있을 경우에 제 1항 내지 제 9항 중 어느 하나 이상의 방법으로 산출된 가중치를 부과하여 유전자형에 따른 상대위험도를 구하는 연산부; (e) 의뢰된 검체의 상대위험도(%) 및 발병률(%)을 구하는 산출부; 및 (f) 의뢰된 검체의 질병 발생 위험도를 판정하는 판정부를 포함하는 질병 예측 장치를 제공하고, 상기 (c) 단계에서의 제 1 데이터베이스는 특정 SNP와 연관된 특정 질병의 증상, 처방 약물의 종류, 처방 약물의 농도, 약물 처방의 빈도, 약물 처방의 기간, 및 부작용에 관한 정보를 포함하는 질병 및 약물반응 관련 데이터베이스인 질병 예측 장치를 제공하며, 상기 (c) 단계에서의 제 2 데이터베이스는 특정 SNP와 연관된 특정 질병에 대한 연구 논문을 포함하고, 상기 논문의 논문고유번호(PMID), 연구대상, 연구방법, 연구기간, 연구결과, 저널정보 및 연구의 반복성 정보를 포함하는 연구 데이터베이스인 질병 예측 장치를 제공하며, 상기 (c) 단계에서의 제 3 데이터베이스는 특정 질병과 연관된 특정 SNP의 염색체 번호, 유전자위, 및 대립형질 정보를 포함하는 유전자 데이터베이스인 질병 예측 장치를 제공하며, 상기 (e) 단계에서의 상대위험도(%)는 상기 (d) 단계에서의 상대위험도가 1 이상( $\geq 1$ )인 경우에는 (대상자의 위험도 평균 점수-1)\*100으로 계산하고, 상기 (d) 단계에서의 상대위험도가 1 미만( $< 1$ )인 경우에는 (1-대상자의 위험도 평균 점수)\*100으로 계산하는 것을 특징으로 하는 질병 예측 장치를 제공하며, 상기 (e) 단계에서의 발병률(%)은 상기 (d) 단계에서의 상대위험도\*유병률로 계산하는 것을 특징으로 하는 질병 예측 장치를 제공하며, 상기 (f)단계에서의 판정은 상기 (e)단계에서의 상대위험도(%)가 1 이하( $\leq 1$ )이면 표준으로 판정하고, 1 초과( $> 1$ )이면 주의 또는 집중으로 판정하는 것을 특징으로 하는 질병 예측 장치를 제공한다.

[0067] 이하 상기 본 발명을 단계별로 상세히 설명한다.

### 발명의 효과

[0069] 본 발명의 SNP를 이용한 질병 관련 유전체 분석 시스템 및 장치는 1단계로 질병 및 약물반응 관련 데이터베이스, 연구 데이터베이스, 및 유전자 데이터베이스로부터 개선된 알고리즘에 의해 객관적이고 구체적인

SNP-질병 연관성을 도출하고, 2단계로 상기 1단계에서 도출된 특정 질병 관련 SNP들의 복합성을 분석하여 최종 질병 위험도를 산출하므로, 질병 예측 결과의 정확도를 향상시키는 효과가 있을 것으로 기대된다.

**도면의 간단한 설명**

- [0071] 도 1은 본 발명의 일 구체예에 따른, SNP를 이용한 특정 질병 예측 방법의 모식도이다.
- 도 2는 본 발명의 일 구체예에 따른, SNP를 이용한 특정 약물 감수성 예측 방법의 모식도이다.
- 도 3은 본 발명의 일 구체예에 따른, SNP를 이용한 특정 약물 부작용 예측 방법의 모식도이다.
- 도 4는 본 발명의 일 구체예에 따른, 유전자형별 가중치 부과 및 질병 발생 위험도 예측 방법의 모식도이다.
- 도 5는 본 발명의 일 구체예에 따른, 대상자의 질병 발생 상대위험도(%) 및 발병률(%)을 산출하는 단계를 구체적으로 나타낸 모식도이다.

**발명을 실시하기 위한 구체적인 내용**

- [0072] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 이들 실시예는 오로지 본 발명을보다 구체적으로 설명하기 위한 것으로서, 본 발명의 요지에 따라 본 발명의 범위가 이들 실시예에 의해 제한되지 않는다는 것은 당업계에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.
- [0074] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

**[0076] 실시예 1. 자료수집**

- [0077] 먼저, 타겟 질병에 대한 유병률(prevalence), 오즈비(odds ratio), 및 유전형 빈도(genotype frequency)에 대한 자료를 수집한다.
- [0078] 구체적으로, 유병률은 국민건강영양조사, 보건복지부-정신질환실태 역학조사, 중앙치매센터 연차보고서, 건강보험심사평가원 정부공개 3.0 공개데이터, 및/또는 주민등록 영양인구 등 공인된 자료를 근거로 하여 수집하되, 출처를 명시한다.
- [0079] 오즈비는 해당 질병과 관련된 논문으로부터 환자(대조)군 연구, 메타분석, 가족연구, 코호트연구 등의 유형 정보를 수집하며, 상기 논문은 PubMed, Google scholar 등 학술논문을 검색할 수 있는 곳이면 제한하지 않는다. 상기 수집된 자료를 모집단자료, 유의성이 좋은 것 등의 순서에 의해 선정한다.
- [0080] 유전형 빈도는 Hapmap 3, 1000 Genomes project와 같은 공인된 자료를 근거로 하여 수집한다.

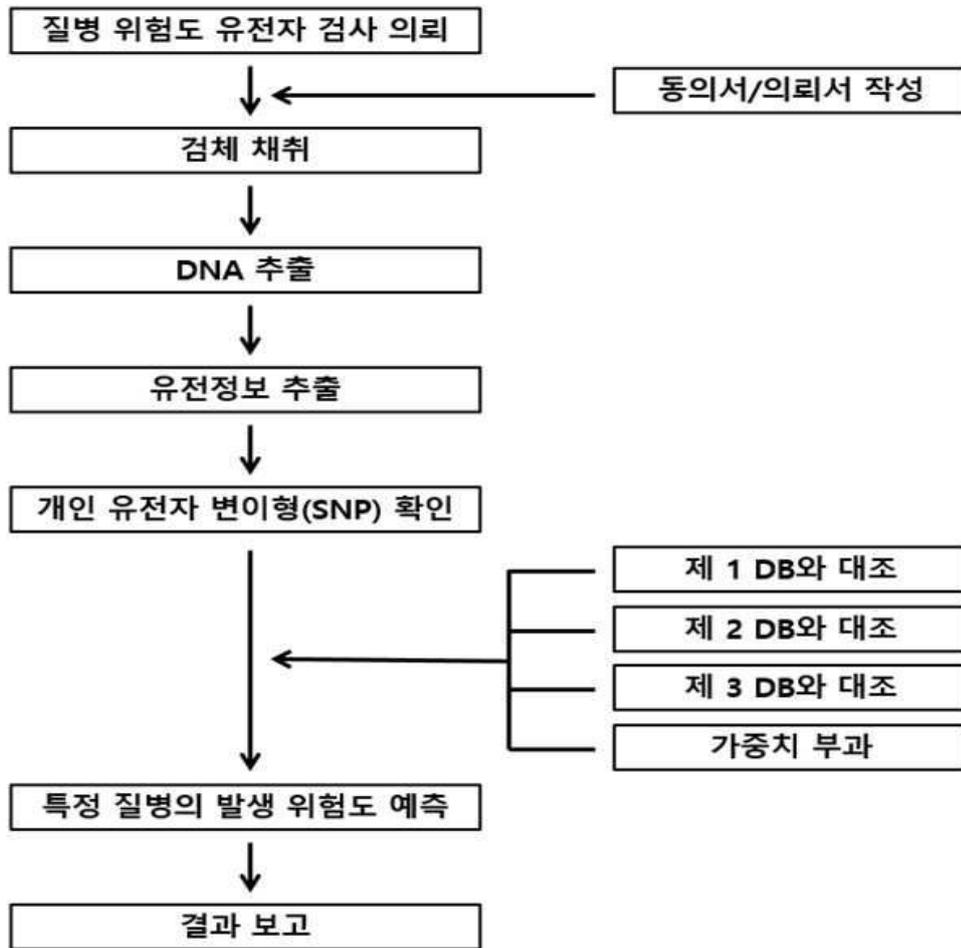
**[0082] 실시예 2. 질병별 발생 위험 가중치 산출**

- [0083] 질병별 위험도를 하기 1 내지 8단계의 과정으로 산출한다. 상기의 1 내지 4단계는 모집단에 따른 위험도를 산출하는 단계이고, 5단계는 대상자에 따른 위험도를 산출하는 단계이며, 6단계는 모집단에서의 대상자 평균 점수를 산출하는 단계이며, 7 내지 8단계는 모집단에서의 대상자 상대위험도(%) 및 발병률(%)을 산출하는 단계이다.
- [0084] 보다 자세하게,
- [0085] 1단계는 각 유전자의 유전자형별 상대위험도를 추정하는 단계이다. 즉, 해당 질병에 위험인자인 유전자마다 유전자형별로 오즈비로 상대 위험도를 추정하는 것으로, “상대위험도(relative risk)=오즈비(odds ratio)/((1-유병률(prevalence))+ (유병률(prevalence)\*오즈비(odds ratio)))” 로 계산한다.
- [0086] 2단계는 각 유전자의 유전자형별 상대위험도를 일반화하는 단계이다. 즉, 각 유전자의 유전자형별로 해당 국적, 인종에 맞게 위험도를 일반화하는 과정으로, “모집단의 상대위험도의 점수=상대위험도(relative risk)\*해당 유전자형 빈도(genotype frequency)” 로 계산한다.

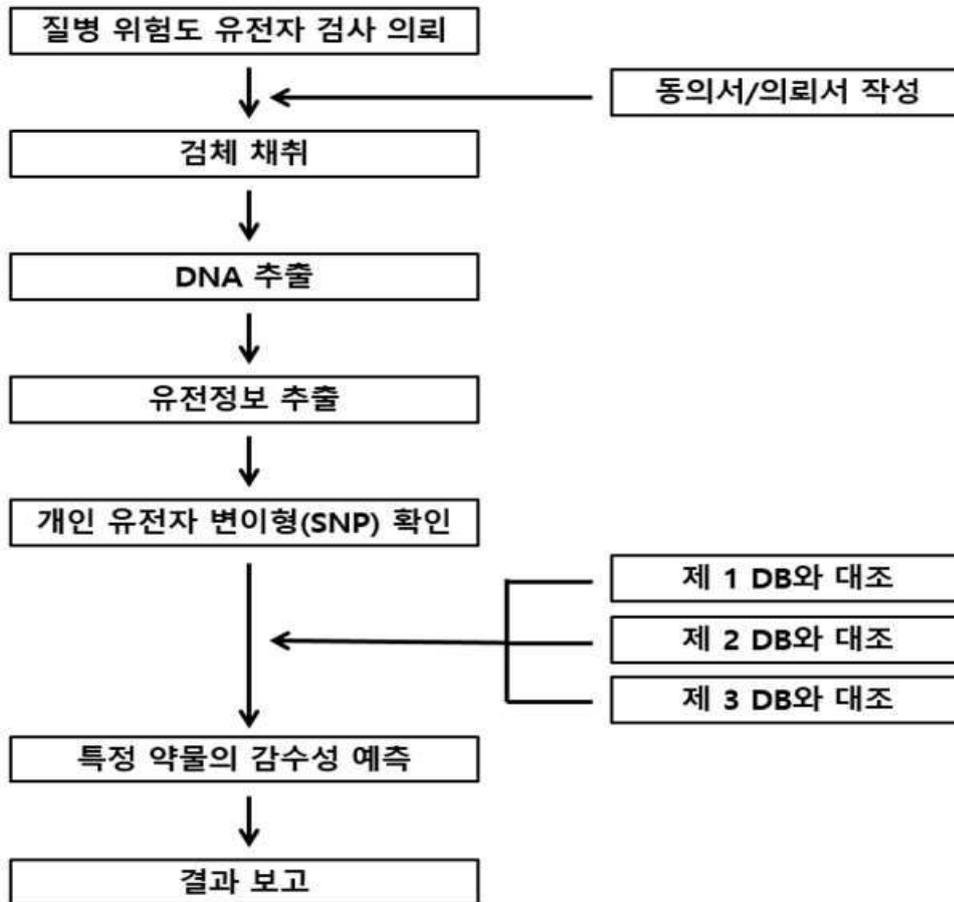
- [0087] 3단계는 각 유전자의 상대위험도 점수(weighted score)를 산출하는 단계로, 각 유전자의 유전자형별 상대위험도 점수의 합으로 계산한다.
- [0088] 4단계는 모든 유전자들의 상대위험도 평균치(weighted average score)를 산출하는 단계로, 각 유전자의 상대위험도 점수들(weighted scores)의 곱으로 계산한다.
- [0089] 5단계는 대상자의 유전자형에 따른 점수를 산출하는 단계로, 각 유전자의 대상자 유전자형에 따른 상대위험도의 곱으로 계산한다.
- [0090] 6단계는 대상자의 유전자형에 따른 상대위험도를 산출하는 단계로, “모집단 기반 대상자의 위험도 평균 점수=(대상자의 유전자형에 따른 점수(5 단계 점수))/(모든 유전자들의 상대위험도 평균치(4 단계 점수))” 로 계산한다.
- [0091] 7단계는 대상자의 상대위험도(%)를 산출하는 단계로, 대상자의 위험도 평균 점수가 1 이상( $\geq 1$ )인 경우에는 “(대상자의 위험도 평균 점수-1)\*100” 로 계산하고, 대상자의 위험도 평균 점수가 1 미만( $< 1$ )인 경우에는 “(1-대상자의 위험도 평균 점수)\*100” 로 계산한다.
- [0092] 8단계는 대상자의 발병률(%)을 산출하는 단계로, “모집단 기반 대상자의 위험도 평균 점수\*유병률(prevalence)” 로 계산한다.
- [0093] 상기 1 내지 8단계를 도 5에 도식화하였다.
- [0095] **실시예 3. 결과 판정**
- [0096] 모집단에서의 위험도 평균점수와 유전자형 빈도(%)를 기반으로 대상자의 질병 발생 위험도를 판정한다. 판정은 “표준/주의” 로 구분하는 2단계형으로 판정하거나, “표준/주의/집중” 으로 구분하는 3단계형 판정이 바람직하나, 이에 한정하는 것은 아니다.
- [0097] 상기에 2단계형 또는 3단계형 판정에 있어서, 표준은 대상자의 위험도 평균 점수가 1 이하( $\leq 1$ )인 경우이고, 주의 또는 집중은 대상자의 위험도 평균 점수가 1 초과( $> 1$ )인 경우를 기본으로 한다. 보다 구체적으로, 질환에 해당하는 유전자의 유전자형 조합별로 가지는 위험도 평균점수와 일반인구에서의 빈도(%)를 구하고, 대상자의 위험도 평균점수보다 높은 위험도 평균점수를 가지는 유전자형 조합의 빈도를 합하여 5% 이하이면 집중, 5% 초과하면 주의로 판정한다. 예외적으로 유병률 산출이 불가능한 질환 항목(예: 중성지방수치 등)에 대해서는 참고논문으로부터 확인된 유전형에 따라 3단계(낮음/보통/높음)로 판정을 도출한다.
- [0099] 이상으로 본 발명의 특정한 부분을 상세히 기술하였는바, 당업계의 통상의 지식을 가진 자에게 있어서 이러한 구체적인 기술은 단지 바람직한 구현 예일 뿐이며, 이에 본 발명의 범위가 제한되는 것이 아닌 점은 명백하다. 따라서 본 발명의 실질적인 범위는 첨부된 청구항과 그의 등가물에 의하여 정의된다고 할 것이다.

도면

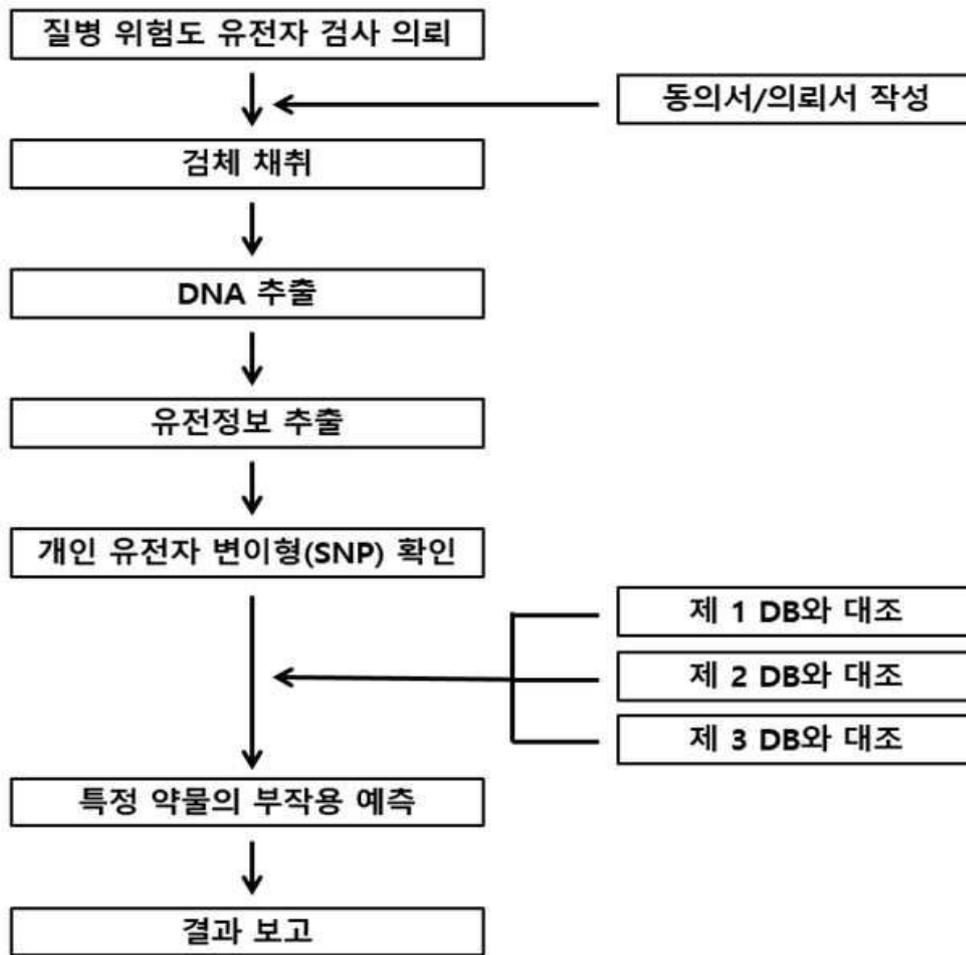
도면1



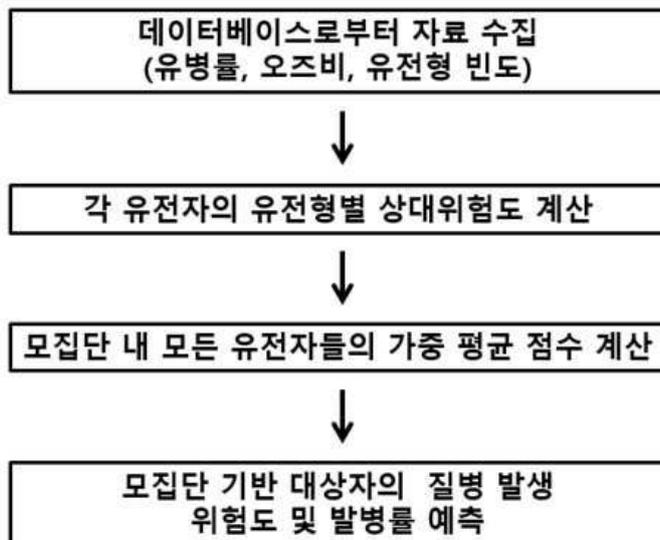
도면2



도면3



도면4



도면5

