



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0007440  
(43) 공개일자 2020년01월22일

(51) 국제특허분류(Int. Cl.)  
C12Q 1/6883 (2018.01) G16B 30/00 (2019.01)  
(52) CPC특허분류  
C12Q 1/6883 (2018.05)  
G16B 30/00 (2019.02)  
(21) 출원번호 10-2018-0081563  
(22) 출원일자 2018년07월13일  
심사청구일자 2018년07월13일

(71) 출원인  
(주) 메디젠휴먼케어  
서울특별시 강남구 테헤란로 223, 20층(역삼동, 큰길타워빌딩)  
(72) 발명자  
신동직  
서울특별시 영등포구 문래로4길 6 현대2차아파트 201-309  
남지호  
서울특별시 서초구 사평대로 240, 503동 709호(반포동, 반포미도2차아파트)  
(74) 대리인  
이재영

전체 청구항 수 : 총 24 항

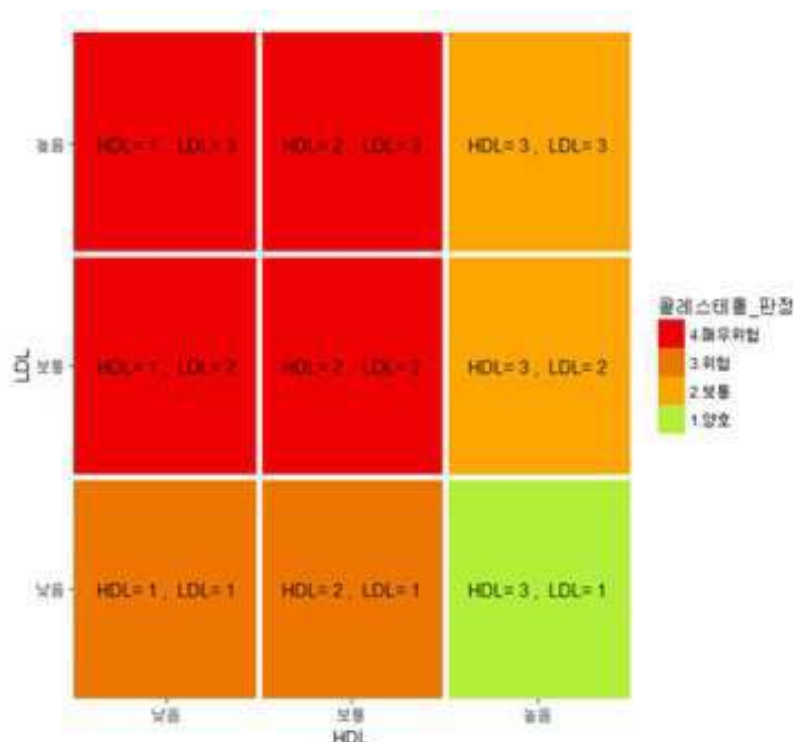
(54) 발명의 명칭 단일염기다형성을 이용한 대사 표현형 예측 방법

(57) 요약

본 발명은 단일염기다형성을 이용한 대사 표현형 예측 방법에 관한 것이다. 대사증후군은 증상이 없거나 혹은 한 가지 위험인자만 보유하고 있더라도 상호 복합적으로 작용하여 만성 질환인 고혈압이나 당뇨, 고지혈증, 비만을 발현한 가능성이 크므로, 복합적인 대사 관련 인자의 관리가 필요하나, 적절히 이루어지지 않고 있는 실정이다.

(뒷면에 계속)

대표도 - 도4



이에 대하여, 최근 대사 관련 인자들 각각이 개인의 유전형질에 의거하여 기초적인 대사 레벨에 차이를 보인다는 보고에 따라, 개인별 유전자의 단일염기다형성을 분석하여 개인별 대사증후군 관리에 이용하고자 하는 노력이 이루어지고 있다.

본 발명은 상기와 같은 종래의 기술상의 문제점을 해결하기 위해 안출된 것으로, 단일염기다형성을 이용한 대사 표현형 예측 방법에 관한 것이다. 본 발명의 대사 표현형 예측 방법은 정확한 개인별 대사 레벨의 진단을 가능하게 하므로, 의학 및 미용 분야에서 크게 이용될 것으로 기대된다.

(52) CPC특허분류

*C12Q 2537/165* (2013.01)

*C12Q 2600/156* (2013.01)

*C12Q 2600/158* (2013.01)

## 명세서

### 청구범위

#### 청구항 1

- (a) 특정 유전자의 유전자형별 빈도를 산출하는 제 1 산출부;
- (b) 표현형 예측을 위한 타겟 표현형 특성지표를 산출하는 제 2 산출부;
- (c) 사용되는 모든 유전자들로부터 대상자의 각 유전자형 조합에 따른 타겟 표현형 특성지표를 산출하는 제 3 산출부;
- (d) 사용되는 모든 유전자들로부터 대상자의 각 유전자형 조합에 따른 빈도를 산출하는 제 4 산출부;
- (e) 가중치 타겟 표현형 특성지표를 산출하는 제 5 산출부;
- (f) 모집단 기반 대상자의 유전자형 조합에 따른 평균 타겟 표현형 특성지표를 산출하는 제 6 산출부; 및
- (g) 타겟 표현형을 판정하는 판정부;를 포함하는, 단일염기다형성을 이용한 표현형 예측 장치.

#### 청구항 2

제 1항에 있어서,

상기 (a) 단계에서의 산출은 유전자형별 빈도의 총합이 1이 되는 것을 특징으로 하는, 장치.

#### 청구항 3

제 1항에 있어서,

상기 (b) 단계에서의 산출은 연구 데이터 베이스, 또는 유전자 데이터 베이스로부터 도출하는 것을 특징으로 하는, 장치.

#### 청구항 4

제 1항에 있어서,

상기 (c) 단계에서의 산출은 유전자형 조합을 구성하는 각 유전자형별 타겟 표현형 특성지표를 모두 합산하는 것을 특징으로 하는, 장치.

#### 청구항 5

제 1항에 있어서,

상기 (d) 단계에서의 산출은 유전자형 조합을 구성하는 각 유전자형별 빈도를 모두 곱하는 것을 특징으로 하는, 장치.

#### 청구항 6

제 1항에 있어서,

상기 (e) 단계에서의 산출은 모든 유전자형 조합의 “유전자형 조합에 따른 타겟 표현형 특성지표\*유전자형 조합에 따른 빈도”를 합산하거나, “(-1)\*유전자형 조합에 따른 타겟 표현형 특성지표\*유전자형 조합에 따른 빈

도”를 합산하여 계산하는 것을 특징으로 하는, 장치.

#### 청구항 7

제 1항에 있어서,

상기 (f) 단계에서의 산출은 "(유전자형 조합에 따른 타겟 표현형 특성지표-가중치 타겟 표현형 특성지표)/|가중치 타겟 표현형 특성지표|\*100"으로 계산하는 것을 특징으로 하는, 장치.

#### 청구항 8

제 1항에 있어서,

상기 타겟 표현형은 혈압, 혈당, 체질량지수, 중성지방, 또는 콜레스테롤인 것을 특징으로 하는, 장치.

#### 청구항 9

제 8항에 있어서,

상기 타겟 표현형이 혈압인 경우에, 상기 단일염기다형성은 ATP2B1(ATPase plasma membrane Ca<sup>2+</sup> transporting 1)유전자의 다형성 rs17249754, NT5C2(5'-nucleotidase, cytosolic II)유전자의 다형성 rs11191593, 및 CYP17A1(cytochrome P450 family 17 subfamily A member 1)의 유전자 다형성 rs3824755로 구성되는 그룹으로부터 선택되는 어느 하나 이상의 다형성인 것을 특징으로 하는, 장치.

#### 청구항 10

제 9항에 있어서,

상기 (g)단계에서의 판정은,

상기 (f)단계에서의 산출값이 0을 기준으로 수치가 높을수록 혈압이 높은 유형, 수치가 낮을수록 혈압이 낮은 유형으로 판정하는 것을 특징으로 하는, 장치.

#### 청구항 11

제 8항에 있어서,

상기 타겟 표현형이 혈당인 경우에, 상기 단일염기다형성은 G6PC2(glucose-6-phosphatase catalytic subunit 2)유전자의 다형성 rs560887, MTNR1B(melatonin receptor 1B)유전자의 다형성 rs10830963, 및 7번 염색체 내 다형성 rs4607517로 구성되는 그룹으로부터 선택되는 어느 하나 이상의 다형성인 것을 특징으로 하는, 장치.

#### 청구항 12

제 11항에 있어서,

상기 (g)단계에서의 판정은,

상기 (f)단계에서의 산출값이 0을 기준으로 수치가 높을수록 혈당 유지가 원활한 유형, 수치가 낮을수록 혈당 유지가 어려운 유형으로 판정하는 것을 특징으로 하는, 장치.

#### 청구항 13

제 8항에 있어서,

상기 타겟 표현형이 체질량지수인 경우에, 상기 단일염기다형성은 BDNF(brain derived neurotrophic factor)유전자의 다형성 rs10767664, 또는 18번 염색체 내 다형성 rs17782313인 것을 특징으로 하는, 장치.

#### 청구항 14

제 13항에 있어서,

상기 (g)단계에서의 판정은,

상기 (f)단계에서의 산출값이 -50 미만이면 체질량지수가 낮은 유형, -50 내지 50이면 체질량지수가 보통인 유형, 50 초과이면 체질량지수가 높은 유형으로 판정하는 것을 특징으로 하는, 장치.

#### 청구항 15

제 8항에 있어서,

상기 타겟 표현형이 중성지방인 경우에, 상기 단일염기다형성은 GCKR(glucokinase regulator)유전자의 다형성 rs2293571, 또는 rs780092인 것을 특징으로 하는, 장치.

#### 청구항 16

제 15항에 있어서,

상기 (g)단계에서의 판정은,

상기 (f)단계에서의 산출값이 -20 미만이면 중성지방이 낮은 유형, -20 내지 20이면 중성지방이 보통인 유형, 50 초과이면 중성지방이 높은 유형으로 판정하는 것을 특징으로 하는, 장치.

#### 청구항 17

제 8항에 있어서,

상기 타겟 표현형이 콜레스테롤인 경우에, 고밀도 지단백, 및 저밀도 지단백으로 구분하여 산출하는 것을 특징으로 하는, 장치.

#### 청구항 18

제 17항에 있어서,

상기 타겟 표현형이 고밀도 지단백인 경우에, 상기 단일염기다형성은 MYL2(myosin light chain 2)유전자의 다형성 rs3782889, LIPG(lipase G, endothelial type)유전자의 다형성 rs3813082, 및 CETP(Cholesteryl Ester Transfer Protein)유전자의 다형성 rs6499861로 구성되는 그룹으로부터 선택되는 어느 하나 이상의 다형성인 것을 특징으로 하는, 장치.

#### 청구항 19

제 18항에 있어서,

상기 (g)단계에서의 판정은,

상기 (f)단계에서의 산출값이 -1 미만이면 고밀도 지단백이 낮은 유형, -1 내지 1이면 고밀도 지단백이 보통인

유형, 1 초과이면 고밀도 지단백이 높은 유형으로 판정하는 것을 특징으로 하는, 장치.

#### 청구항 20

제 17항에 있어서,

상기 타겟 표현형이 저밀도 지단백인 경우에, 상기 단일염기다형성은 MYL2(myosin light chain 2)유전자의 다형성 rs3782889, 또는 LIPG(lipase G, endothelial type)유전자의 다형성 rs3813082인 것을 특징으로 하는, 장치.

#### 청구항 21

제 20항에 있어서,

상기 (g)단계에서의 판정은,

상기 (f)단계에서의 산출값이 -0.1 미만이면 저밀도 지단백이 낮은 유형, -0.1 내지 0.1이면 저밀도 지단백이 보통인 유형, 0.1 초과이면 저밀도 지단백이 높은 유형으로 판정하는 것을 특징으로 하는, 장치.

#### 청구항 22

(a) 특정 유전자의 유전자형별 빈도를 산출하는 단계;

(b) 표현형 예측을 위한 타겟 표현형 특성지표를 산출하는 단계;

(c) 사용되는 모든 유전자들로부터 대상자의 각 유전자형 조합에 따른 타겟 표현형 특성지표를 산출하는 단계;

(d) 사용되는 모든 유전자들로부터 대상자의 각 유전자형 조합에 따른 빈도를 산출하는 단계;

(e) 가중치 타겟 표현형 특성지표를 산출하는 단계;

(f) 모집단 기반 대상자의 유전자형 조합에 따른 평균 타겟 표현형 특성지표를 산출하는 단계; 및

(g) 타겟 표현형을 판정하는 단계;를 포함하는, 단일염기다형성을 이용한 표현형 예측 방법.

#### 청구항 23

제 22항에 있어서,

상기 타겟 표현형은 혈압, 혈당, 체질량지수, 중성지방, 또는 콜레스테롤인 것을 특징으로하는, 방법.

#### 청구항 24

제 23항에 있어서,

상기 타겟 표현형이 콜레스테롤인 경우에, 고밀도 지단백, 및 저밀도 지단백으로 구분하여 산출하는 것을 특징으로 하는, 방법.

### 발명의 설명

### 기술 분야

[0001] 본 발명은 유전자 다형성을 이용한 대사 표현형 예측 방법에 관한 것이다.

## 배경 기술

[0003] 1988년 Reaven GM에 의해 대사증후군(metabolic syndrome)이 명명된 이후, 대사증후군은 일차의료에서 흔하게 마주치는 건강 문제가 되었다. 상기 일차의료에서 대사증후군의 유병률은 약 30%에 달하는 것으로 알려져 있으며 연령이 높을수록, 비만의 정도가 심할수록 유병률이 증가되었다. 상기 대사증후군의 의학적 중요성은 대사증후군이 있는 경우가 대사증후군이 없는 경우에 비해 심혈관질환의 사망률이 높다는 사실로부터 드러난다. 국내에서도 비만 인구가 점차 증가하고 있고, 고혈압이나 당뇨, 고지혈증 등과 같은 심혈관 질환의 유병률이 급격하게 증가하고 있으므로, 심혈관 질환에 의한 사망률이 남녀에서 2위에 해당하는 점 등을 고려하여, 최근 대사증후군 발명 전에 신체 대사 지표를 관리하고자 하는 노력이 지속되고 있다. 그러나 증상이 없거나 혹은 한 가지 위험인자만 보유하고 있더라도, 대사증후군은 상호 복합적으로 작용하여 만성 질환인 고혈압이나 당뇨, 고지혈증, 비만을 발현한 가능성이 크므로, 복합적인 대사 관련 인자의 관리가 필요하나, 적절히 이루어지지 않고 있는 실정이다. 이에 대하여, 최근 대사 관련 인자들 각각이 개인의 유전형질에 의거하여 기초적인 대사 레벨에 차이를 보인다는 보고에 따라, 개인별 유전자의 단일염기다형성(Single-nucleotide polymorphism; SNP)을 분석하여 개인별 대사증후군 관리에 이용하고자 하는 노력이 이루어지고 있다.

[0004] 한편, 개인별 유전자의 단일염기다형성이란, DNA 염기서열에서 하나의 염기서열(A,T,G,C)의 차이를 보이는 유전적 변화 또는 변이를 의미하는 것으로, 단일 염기 다형현상은 각 개인마다 많은 변이를 보이는 부분이므로 DNA 지문 분석에 주로 이용된다. 단일염기다형성 분석을 이용하면 개인별 유전자에 내재되어 있는 기초적인 대사 능력의 차이를 고려할 수 있으므로, 복합적인 대사증후군을 보다 효과적으로 관리 가능하다.

[0005] 본 발명은 상기와 같은 종래의 기술상의 문제점을 해결하기 위해 안출된 것으로, 단일염기다형성을 이용한 대사 표현형 예측 방법에 관한 것이다. 본 발명의 대사 표현형 예측 방법은 정확한 개인별 대사 레벨의 진단을 가능하게 하므로, 의학 및 미용 분야에서 크게 이용될 것으로 기대된다.

## 발명의 내용

### 해결하려는 과제

[0007] 본 발명은 상기와 같은 종래의 기술상의 문제점을 해결하기 위해 안출된 것으로, 단일염기다형성을 이용한 대사 표현형 예측 방법에 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)이란, DNA의 염기배열로서 부호화되는 모든 정보를 총칭하는 광의의 개념으로서, 본 발명에 있어서 "유전정보"란 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 "연구 데이터베이스"란, 표현형-SNP 연관성 결과를 비교하여 특정 표현형의 발현을 예측하기 위한 데이터풀을 의미하며, 이에 한정하는 것은 아니나, 연구 자료는 임상 또는 학술 논문일 수 있다. 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, 성별 정보와 배우자, 자식, 부모, 사촌 등에 대한 가계도 정보, 및 환경 정보를 포함시킬 수 있다.
- [0020] 본 발명의 일 구체예에서 "유전자 데이터베이스"란, 표현형-SNP 연관성 결과를 비교하여 특정 표현형의 발현을 예측하기 위한 데이터풀(Data pool)을 의미하며, 특정 표현형과 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.
- [0022] 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 처리 할 수 있다. 본 명세서에 있어서 알고리즘은 특정 개체의 유전정보를 연구 데이터베이스, 및 유전자 데이터베이스에 저장된 정보들과 대조하여 특정 표현형이 발현될 확률을 도출하는 것이나, 이에 한정하는 것은 아니다.
- [0024] 본 발명의 일 구체예에서 "대사(metabolism)"란, 생물체 내에서 일어나는 물질의 분해나 합성과 같은 모든 물질적 변화를 총칭하는 것으로서, 주위로부터 흡수한 저분자 유기물이나 무기물을 이용해 고분자 화합물을 합성하는 동화작용과 반대로 고분자 화합물을 저분자 유기물이나 무기물로 분해하는 이화작용으로 구분된다. 생물체 내의 대사 과정에 이상이 생겨, 생물체의 에너지 균형이 깨지는 상태를 대사증후군(metabolic syndrome)이라



한다. 보다 구체적으로, 대사증후군이란, 뇌심혈관질환 및 당뇨병의 위험을 높이는 체지방 증가, 혈압 상승, 혈당 상승, 혈중 지질 이상 등의 이상 상태들의 집합을 의미한다. 대사증후군이 있는 경우에는 심혈관 질환의 위험을 두 배 이상 상승하며, 당뇨병의 발병을 10배 이상 증가된다. 대사증후군은 단일한 질병이 아니라 유전적 소인과 환경적 인자가 더해져 발생하는 포괄적 질병으로 알려져 있다.

[0026] 본 발명의 일 구체예에서, (a) 특정 유전자의 유전자형별 빈도를 산출하는 제 1 산출부; (b) 표현형 예측을 위한 타겟 표현형 특성지표를 산출하는 제 2 산출부; (c) 사용되는 모든 유전자들로부터 대상자의 각 유전자형 조합에 따른 타겟 표현형 특성지표를 산출하는 제 3 산출부; (d) 사용되는 모든 유전자들로부터 대상자의 각 유전자형 조합에 따른 빈도를 산출하는 제 4 산출부; (e) 가중치 타겟 표현형 특성지표를 산출하는 제 5 산출부; (f) 모집단 기반 대상자의 유전자형 조합에 따른 평균 타겟 표현형 특성지표를 산출하는 제 6 산출부; 및 (g) 타겟 표현형을 판정하는 판정부;를 포함하는, 단일염기다형성을 이용한 표현형 예측 장치를 제공하고, 상기 (a) 단계에서의 산출은 유전자형별 빈도의 총합이 1이 되는 것을 특징으로 하는 단일염기다형성을 이용한 표현형 예측 장치를 제공하며, 상기 (b) 단계에서의 산출은 연구 데이터 베이스, 또는 유전자 데이터 베이스로부터 도출하는 것을 특징으로 하는 단일염기다형성을 이용한 표현형 예측 장치를 제공하며, 상기 (c) 단계에서의 산출은 유전자형 조합을 구성하는 각 유전자형별 타겟 표현형 특성지표를 모두 합산하는 것을 특징으로 하는 단일염기다형성을 이용한 표현형 예측 장치를 제공하며, 상기 (d) 단계에서의 산출은 유전자형 조합을 구성하는 각 유전자형별 빈도를 모두 곱하는 것을 특징으로 하는 단일염기다형성을 이용한 표현형 예측 장치를 제공하며, 상기 (e) 단계에서의 산출은 모든 유전자형 조합의 “유전자형 조합에 따른 타겟 표현형 특성지표\*유전자형 조합에 따른 빈도”를 합산하거나, “(-1)\*유전자형 조합에 따른 타겟 표현형 특성지표\*유전자형 조합에 따른 빈도”를 합산하여 계산하는 것을 특징으로 하는 단일염기다형성을 이용한 표현형 예측 장치를 제공하며, 상기 (f) 단계에서의 산출은 “(유전자형 조합에 따른 타겟 표현형 특성지표-가중치 타겟 표현형 특성지표)/|가중치 타겟 표현형 특성지표|\*100”으로 계산하는 것을 특징으로 하는 단일염기다형성을 이용한 표현형 예측 장치를 제공한다.

[0028] 상기 단일염기다형성을 이용한 표현형 예측 장치의 타겟 표현형은 이에 제한하는 것은 아니나, 혈압, 혈당, 체질량지수, 중성지방, 또는 콜레스테롤인 것이 바람직하다.

[0030] 상기 타겟 표현형이 혈압인 경우에, 상기 단일염기다형성은 ATP2B1(ATPase plasma membrane Ca<sup>2+</sup> transporting 1)유전자의 다형성 rs17249754, NT5C2(5'-nucleotidase, cytosolic II)유전자의 다형성 rs11191593, 및 CYP17A1(cytochrome P450 family 17 subfamily A member 1)의 유전자 다형성 rs3824755로 구성되는 그룹으로부터 선택되는 어느 하나 이상의 다형성인 것을 특징으로 하고, 상기 (g)단계에서의 판정은, 상기 (f)단계에서의 산출값이 0을 기준으로 수치가 높을수록 혈압이 높은 유형, 수치가 낮을수록 혈압이 낮은 유형으로 판정한다.

[0032] 상기 타겟 표현형이 혈당인 경우에, 상기 단일염기다형성은 G6PC2(glucose-6-phosphatase catalytic subunit 2)유전자의 다형성 rs560887, MTNR1B(melatonin receptor 1B)유전자의 다형성 rs10830963, 및 7번 염색체 내 다형성 rs4607517로 구성되는 그룹으로부터 선택되는 어느 하나 이상의 다형성인 것을 특징으로 하고, 상기 (g)단계에서의 판정은, 상기 (f)단계에서의 산출값이 0을 기준으로 수치가 높을수록 혈당 유지가 원활한 유형, 수치가 낮을수록 혈당 유지가 어려운 유형으로 판정한다.

[0034] 상기 타겟 표현형이 체질량지수인 경우에, 상기 단일염기다형성은 BDNF(brain derived neurotrophic factor)유전자의 다형성 rs10767664, 또는 18번 염색체 내 다형성 rs17782313인 것을 특징으로 하고, 상기 (g)단계에서의 판정은, 상기 (f)단계에서의 산출값이 -50 미만이면 체질량지수가 낮은 유형, -50 내지 50이면 체질량지수가 보통인 유형, 50 초과이면 체질량지수가 높은 유형으로 판정한다.

[0036] 상기 타겟 표현형이 중성지방인 경우에, 상기 단일염기다형성은 GCKR(glucokinase regulator)유전자의 다형성 rs2293571, 또는 rs780092인 것을 특징으로 하고, 상기 (g)단계에서의 판정은, 상기 (f)단계에서의 산출값이 -20 미만이면 중성지방이 낮은 유형, -20 내지 20이면 중성지방이 보통인 유형, 50 초과이면 중성지방이 높은

유형으로 판정한다.

- [0038] 상기 타겟 표현형이 콜레스테롤인 경우에, 고밀도 지단백, 및 저밀도 지단백으로 구분하여 산출하는 것을 특징으로 한다.
- [0039] 상기 타겟 표현형이 고밀도 지단백인 경우에, 상기 단일염기다형성은 MYL2(myosin light chain 2)유전자의 다형성 rs3782889, LIPG(lipase G, endothelial type)유전자의 다형성 rs3813082, 및 CETP(Cholesteryl Ester Transfer Protein)유전자의 다형성 rs6499861로 구성되는 그룹으로부터 선택되는 어느 하나 이상의 다형성인 것을 특징으로 하고, 상기 (g)단계에서의 판정은, 상기 (f)단계에서의 산출값이 -1 미만이면 고밀도 지단백이 낮은 유형, -1 내지 1이면 고밀도 지단백이 보통인 유형, 1 초과이면 고밀도 지단백이 높은 유형으로 판정하는 것을 특징으로 한다.
- [0040] 상기 타겟 표현형이 저밀도 지단백인 경우에, 상기 단일염기다형성은 MYL2(myosin light chain 2)유전자의 다형성 rs3782889, 또는 LIPG(lipase G, endothelial type)유전자의 다형성 rs3813082인 것을 특징으로 하고, 상기 (g)단계에서의 판정은, 상기 (f)단계에서의 산출값이 -0.1 미만이면 저밀도 지단백이 낮은 유형, -0.1 내지 0.1이면 저밀도 지단백이 보통인 유형, 0.1 초과이면 저밀도 지단백이 높은 유형으로 판정하는 것을 특징으로 한다.
- [0042] 본 발명의 다른 구체예에서, (a) 특정 유전자의 유전자형별 빈도를 산출하는 단계; (b) 표현형 예측을 위한 타겟 표현형 특성지표를 산출하는 단계; (c) 사용되는 모든 유전자들로부터 대상자의 각 유전자형 조합에 따른 타겟 표현형 특성지표를 산출하는 단계; (d) 사용되는 모든 유전자들로부터 대상자의 각 유전자형 조합에 따른 빈도를 산출하는 단계; (e) 가중치 타겟 표현형 특성지표를 산출하는 단계; (f) 모집단 기반 대상자의 유전자형 조합에 따른 평균 타겟 표현형 특성지표를 산출하는 단계; 및 (g) 타겟 표현형을 판정하는 단계;를 포함하는 단일염기다형성을 이용한 표현형 예측 방법을 제공하고, 상기 타겟 표현형은 혈압, 혈당, 체질량지수, 중성지방, 또는 콜레스테롤인 것을 특징으로 하는 단일염기다형성을 이용한 표현형 예측 방법을 제공하며, 상기 타겟 표현형이 콜레스테롤인 경우에, 고밀도 지단백, 및 저밀도 지단백으로 구분하여 산출하는 것을 특징으로 하는 단일염기다형성을 이용한 표현형 예측 방법을 제공한다.
- [0044] 이하 상기 본 발명을 단계별로 상세히 설명한다.

### 발명의 효과

- [0046] 개인별 유전자의 단일염기다형성(Single-nucleotide polymorphism; SNP)을 분석하여 개인별 대사 상태를 측정하면, 개인별 유전자에 내재되어 있는 기초적인 대사 능력의 차이를 고려할 수 있으므로, 복합적인 대사증후군을 보다 효과적으로 관리 가능하다. 본 발명은 단일염기다형성을 이용한 대사 표현형 예측 방법에 관한 것으로, 본 발명의 대사 표현형 예측 방법은 정확한 개인별 대사 레벨의 진단을 가능하게 하므로, 의학 및 미용 분야에서 크게 이용될 것으로 기대된다.

### 도면의 간단한 설명

- [0048] 도 1은 본 발명의 일 구체예에 따른, 단일염기다형성을 이용한 혈압 및 혈당 표현형 예측 알고리즘의 판정 분류를 나타낸 도이다.
- 도 2는 본 발명의 일 구체예에 따른, 단일염기다형성을 이용한 BMI 및 중성지방 표현형 예측 알고리즘의 판정 분류를 나타낸 도이다.
- 도 3은 본 발명의 일 구체예에 따른, 단일염기다형성을 이용한 HDL 및 LDL 표현형 예측 알고리즘의 판정 분류를 나타낸 도이다.
- 도 4는 본 발명의 일 구체예에 따른, 단일염기다형성을 이용한 콜레스테롤 위험도 판정 분류를 나타낸 도이다.

## 발명을 실시하기 위한 구체적인 내용

[0049] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 이들 실시예는 오로지 본 발명을 보다 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.

[0051] 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.

### [0053] 실시예 1. 단일염기다형성을 이용한 대사 표현형 예측 알고리즘 설계

#### [0055] 실시예 1-1. 자료수집

[0056] 먼저, 연구 데이터베이스로부터 타겟 표현형 특성지표를 선정한다. 상기의 연구 데이터베이스는 PubMed, Google scholar 등 학술논문을 검색할 수 있는 곳이면 제한하지 않는다. 수집된 논문으로부터 대조군 연구, 메타분석, 가족연구, 코호트연구 등의 유형 정보를 도출하고, 자료를 모집단자료, 유의성이 좋은 것 등의 순서에 의해 선정한다.

[0057] 유전형 빈도는 Hapmap 3, 1000 Genomes project와 같은 공인된 자료를 근거로 하여 수집한다.

#### [0059] 실시예 1-2. 유전자 표현형 인덱스 모델 설계

[0060] 유전자 표현형 인덱스를 하기 1 내지 5단계의 과정으로 산출한다.

[0062] 1단계는 대상자의 유전자형(genotype) 조합에 따른 빈도(frequency)를 산출하는 과정이다. 사용되는 모든 유전자들의 유전자형 조합별 빈도의 총합이 1이 되도록 계산한다. 구체적인 계산식을 표 1에 나타내었다.

표 1

$$genotype \text{ 조합 } frequency = \prod_{i=1}^n \text{유전자 } i \text{의 } genotype \text{ frequency}$$

[0063]

[0065] 2단계는 대상자의 유전자형 조합에 따른 타겟 표현형 특성지표를 산출하는 과정이다. 유전자형 조합을 구성하는 각 유전자형별 타겟 표현형 특성지표를 모두 합산한다. 구체적인 계산식을 표 2에 나타내었다.

표 2

$$genotype \text{ 조합 타겟 표현형 특성지표} = \sum_{i=1}^n \text{유전자 } i \text{의 } genotype \text{ 타겟 표현형 특성지표}$$

[0066]

[0068] 3단계는 모집단의 대사특성지표를 산출하는 과정이다. 사용되는 모든 유전자들의 유전자형 조합에 따른 해당 국적과 인종별로 대사특성지표를 일반화한다. 구체적인 계산식을 표 3에 나타내었다.

표 3

$$genotype \text{ 조합 모집단 대사특성지표} = genotype \text{ 조합 } frequency * genotype \text{ 조합 대사특성지표}$$

[0069]

[0071] 4단계는 가중치(weighted average) 타겟 표현형 특성지표를 산출하는 과정이다. 모든 유전자들의 유전자형 조합별 "유전자형 조합에 따른 타겟 표현형 특성지표\*유전자형 조합에 따른 빈도"를 합산한다. 경우에 따라서, "-1"을 추가로 곱하는 것도 가능하다.

[0073] 5단계는 모집단 기반 대상자의 유전자형 조합에 따른 평균 타겟 표현형 특성지표를 산출하는 과정이다. 구체적인 계산식을 표 4에 나타내었다.

표 4

$$\text{모집단 기반 대상자의 genotype조합에 따른 average 타겟 표현형 특성지표} = \frac{\text{대상자의 genotype조합에 따른 타겟 표현형 특성지표} \times \text{weighted average 타겟 표현형 특성지표}}{|\text{weighted average 타겟 표현형 특성지표}|} \times 100$$

[0074]

[0076] 상기 1 내지 5단계의 산출 과정으로부터 대상자의 대사 표현형 예측 결과를 도출한다. 대사 표현형 예측을 위한 타겟 표현형 특성지표는 각 유전자형의 G값(value)로 설정한다. G값은 표현형 특성에 영향을 미치는 모든 loci에서 모든 유전자의 결합된 효과로서, 실제로 동일한 유전자형을 가진 개체의 표현형 평균값으로 해석된다. G값은 연구 데이터 베이스, 또는 유전자 데이터 베이스로부터 도출 가능하다. 종합적인 대사 표현형 예측을 위한 세부 측정 요소는 "혈압", "혈당", "체질량지수(BMI)", "중성지방", 및 "콜레스테롤"로 설정한다. 산출된 결과에 따른 판정은 혈압과 혈당은 2개의 단계로 구분하고, 체질량지수(body mass index; BMI)와 중성지방은 3개의 단계로 구분한다. 상기 요소별 단계 구분을 위한 가이드를 도 1, 도2, 및 표 5에 나타내었다.

표 5

[0077]

판정분류	단계 구분을 위한 가이드		
	제 1단계	제 2단계	제 3단계
혈압	혈압이 낮음		혈압이 높음
혈당	혈당 유지 어려움		혈당 유지 원활
체질량지수(BMI)	BMI가 낮음	BMI가 보통	BMI가 높음
중성지방	중성지방이 낮음	중성지방이 보통	중성지방이 높음

[0078] 상기 표 5의 가이드를 적용하는 판정 기준은, 구체적으로, 혈압의 경우, 유전자형에 따른 PP(펄스압력)수치가 높으면 혈압이 높고, 수치가 낮으면 혈압이 낮음으로 판정한다. 판정 기준값은 0으로 한다. 혈당의 경우에는 유전자형에 따른 혈당수치가 높을수록 혈당유지가 잘되는 정상군, 낮을수록 혈당유지가 어려운 위험군으로 판정한다. 판정 기준값은 0으로 한다. 체질량지수(body mass index; BMI)의 경우에는 유전자형에 따른 BMI수치에 따라 판정한다. 판정 기준값은 수치가 -50보다 작으면 낮음, -50 내지 50이면 보통, 50보다 크면 높음으로 판정한다. 중성지방의 경우에는 유전자형에 따른 중성지방 농도에 따라 판정한다. 판정 기준값은 수치가 -20보다 작으면 낮음, -20 내지 20이면 보통, 20보다 크면 높음으로 판정한다.

[0079] 콜레스테롤의 경우에는 고밀도지단백질(high density lipoprotein; HDL) 및 저밀도지단백질(low density lipoprotein; LDL)을 먼저 도출하고, 이로부터 콜레스테롤 결과를 판정한다. HDL과 LDL의 최종 결과는 3개의 단계로 구분하고, 콜레스테롤 판정시 HDL과 LDL의 결과를 종합적으로 고려하여, HDL수치가 높고 LDL수치가 낮을수록 양호, HDL수치가 낮고 LDL수치가 높을수록 위험군으로 판정한다. HDL과 LDL 단계 구분을 위한 가이드를 도 3, 및 표 6에 나타내었다.

표 6

[0080]

판정분류	단계 구분을 위한 가이드		
	제 1단계	제 2단계	제 3단계

HDL	HDL 수치 낮음	HDL 수치 보통	HDL 수치 높음
LDL	LDL 수치 낮음	LDL 수치 보통	LDL 수치 높음

[0081] 상기 표 6의 가이드를 적용하는 판정 기준은, 구체적으로, HDL의 경우, 유전자형에 따른 HDL수치에 따라 판정한다. 판정 기준값은 수치가 -1보다 작으면 낮음, -1 내지 1이면 보통, 1보다 크면 높음으로 보통으로 판정한다. LDL의 경우에는 유전자형에 따른 LDL수치에 따라 판정한다. 판정 기준값은 수치가 -0.1보다 작으면 낮음, -0.1 내지 0.1이면 보통, 0.1보다 크면 높음으로 판정한다. 콜레스테롤은 HDL과 LDL판정 결과에 따라 도 4, 및 표 7과 같이 판정한다.

표 7

[0082]

양호	보통	위험	매우 위험
HDL=높음, LDL=낮음	HDL=높음, LDL=높음 HDL=보통, LDL=보통	HDL=보통, LDL=낮음 HDL=낮음, LDL=낮음	HDL=보통, LDL=높음 HDL=보통, LDL=보통 HDL=낮음, LDL=높음 HDL=낮음, LDL=보통

[0084] 실시예 2. 단일염기다형성을 이용한 대사 표현형 예측

[0085] 상기 실시예 1에서 설계된 알고리즘을 이용하여 대상자의 대사 표현형을 판정하였다.

[0087] 실시예 2-1. 혈압 표현형 예측

[0088] 혈압 표현형 예측을 위해서는 ATP2B1(ATPase plasma membrane Ca<sup>2+</sup> transporting 1)유전자의 다형성 rs17249754(89666809, G>A), NT5C2(5'-nucleotidase, cytosolic II)유전자의 다형성 rs11191593(103179458, T>C), 및 CYP17A1(cytochrome P450 family 17 subfamily A member 1)의 유전자 다형성 rs3824755(102836092, G>A)을 이용하였다.

[0089] 연구 데이터베이스, 및 유전자 데이터로부터 도출한 상기 유전자들의 유전자형별 빈도(frequency)와 G값(value)은 하기 표 8에 나타내었다.

표 8

[0090]

RS 번호	유전자형	빈도 (frequency)	G값 (value)
rs17249754	AA	0.78251716	53.1
	AC	0.20416568	54.4
	CC	0.01331716	63.2
rs11191593	TT	0.662596	51.3
	CT	0.302808	50.1
	CC	0.034596	49.5
rs3824755	GG	0.008464	-9.328
	CG	0.167072	-4.664
	CC	0.824464	0

[0091] 상기 표 8의 유전자형별 빈도와 G값을 실시예 1의 1 내지 5 단계에 대입하여 혈압 표현형을 예측하였다. 구체적인 도출 과정은 표 9에 나타내었다.

표 9

[0092]

rs11191593	rs17249754	rs3824755	조합 frequency	조합 value	(-1)* 조합 * value frequency	COMPARE TO GLOBAL MEAN (%)	결과
TT	AA	CC	0.00686909	-0.6	0.004121457	-481.3953488	1단계

TT	AA	GC	0.02919365	-0.32	0.009341968	-210.0775194	1단계
TT	AA	GG	0.03101825	-0.04	0.00124073	61.24031008	3단계
TT	GA	CC	0.02551378	-0.2	0.005102756	-93.79844961	1단계
TT	GA	GC	0.10843356	0.08	-0.008674685	177.5193798	3단계
TT	GA	GG	0.11521066	0.36	-0.041475837	448.8372093	3단계
TT	GG	CC	0.02369137	0.2	-0.004738273	293.7984496	3단계
TT	GG	GC	0.10068831	0.48	-0.048330387	565.1162791	3단계
TT	GG	GG	0.10698133	0.76	-0.081305808	836.4341085	3단계
TC	AA	CC	0.00482693	-0.98	0.004730393	-849.6124031	1단계
TC	AA	GC	0.02051446	-0.7	0.01436012	-578.2945736	1단계
TC	AA	GG	0.02179661	-0.42	0.009154577	-306.9767442	1단계
TC	GA	CC	0.0179286	-0.58	0.010398589	-462.0155039	1단계
TC	GA	GC	0.07619656	-0.3	0.022858967	-190.6976744	1단계
TC	GA	GG	0.08095884	-0.02	0.001619177	80.62015504	3단계
TC	GG	CC	0.01664799	-0.18	0.002996638	-74.41860465	1단계
TC	GG	GC	0.07075395	0.1	-0.007075395	196.8992248	3단계
TC	GG	GG	0.07517607	0.38	-0.028566906	468.2170543	3단계
CC	AA	CC	0.00084797	-1.36	0.001153245	-1217.829457	1단계
CC	AA	GC	0.00360389	-1.08	0.003892202	-946.5116279	1단계
CC	AA	GG	0.00382913	-0.8	0.003063308	-675.1937984	1단계
CC	GA	CC	0.00314962	-0.96	0.003023634	-830.2325581	1단계
CC	GA	GC	0.01338588	-0.68	0.009102399	-558.9147287	1단계
CC	GA	GG	0.0142225	-0.4	0.005689	-287.5968992	1단계
CC	GG	CC	0.00292465	-0.56	0.001637802	-442.6356589	1단계
CC	GG	GC	0.01242975	-0.28	0.003480329	-171.3178295	1단계
CC	GG	GG	0.01320661	0	0	100	1단계
				TOTAL AVERAGE	-0.1032		

[0093] 표 9에서 "조합 frequency"는 "해당 열의 각 유전자형별 빈도의 곱"을 나타내고, "조합 value"는 "해당 열의 각 유전자형별 value의 합"을 나타낸다. "(-1)\*조합 value\*조합 frequency"는 "(-1)\*해당 열의 조합 frequency\*해당 열의 조합 value"이고, "COMPARE TO GLOBAL MEAN (%)"은 "(해당 열의 조합 value - total average) / total average의 정수 절댓값 \* 100"으로 계산된다.

[0095] 실시예 2-2. 혈당 표현형 예측

[0096] 혈당 표현형 예측을 위해서는 G6PC2(glucose-6-phosphatase catalytic subunit 2)유전자의 다형성 rs560887(168906638, T>C), MTNR1B(melatonin receptor 1B)유전자의 다형성 rs10830963(92975544, C>G), 및 7번 염색체에 포함되나 특정유전자에 속하지는 않고, position 44196069으로 특정되는 다형성 rs4607517(44196069, G>A)을 이용하였다.

[0097] 연구 데이터베이스, 및 유전자 데이터로부터 도출한 상기 유전자들의 유전자형별 빈도(frequency)와 G값(value)은 하기 표 10에 나타내었다.

표 10

[0098]

RS 번호	유전자형	빈도 (frequency)	G값 (value)
rs560887	CC	0.49	-0.084
	CT	0.42	-0.042
	TT	0.09	0
rs10830963	GG	0.09	-0.068
	GC	0.42	-0.034
	CC	0.49	0



rs4607517	AA	0.0256	-0.068
	AG	0.2688	-0.034
	GG	0.7056	0

[0099] 상기 표 10의 유전자형별 빈도와 G값을 실시예 1의 1 내지 5 단계에 대입하여 혈당 표현형을 예측하였다. 구체적인 도출 과정은 표 11에 나타내었다.

표 11

[0100]

rs560887	rs10830963	rs4607517	조합 frequency	조합 value	조합value * 조합 frequency	COMPARE TO GLOBAL MEAN (%)	결과
CC	GG	AA	0.00112896	-0.22	-0.000248371	-144.2273535	1단계
CC	GG	AG	0.01185408	-0.186	-0.002204859	-106.4831261	1단계
CC	GG	GG	0.03111696	-0.152	-0.004729778	-68.73889876	1단계
CC	GC	AA	0.00526848	-0.186	-0.000979937	-106.4831261	1단계
CC	GC	AG	0.05531904	-0.152	-0.008408494	-68.73889876	1단계
CC	GC	GG	0.14521248	-0.118	-0.017135073	-30.9946714	1단계
CC	CC	AA	0.00614656	-0.152	-0.000934277	-68.73889876	1단계
CC	CC	AG	0.06453888	-0.118	-0.007615588	-30.9946714	1단계
CC	CC	GG	0.16941456	-0.084	-0.014230823	6.74955595	3단계
CT	GG	AA	0.00096768	-0.178	-0.000172247	-97.60213144	1단계
CT	GG	AG	0.01016064	-0.144	-0.001463132	-59.85790409	1단계
CT	GG	GG	0.02667168	-0.11	-0.002933885	-22.11367673	1단계
CT	GC	AA	0.00451584	-0.144	-0.000650281	-59.85790409	1단계
CT	GC	AG	0.04741632	-0.11	-0.005215795	-22.11367673	1단계
CT	GC	GG	0.12446784	-0.076	-0.009459556	15.63055062	3단계
CT	CC	AA	0.00526848	-0.11	-0.000579533	-22.11367673	1단계
CT	CC	AG	0.05531904	-0.076	-0.004204247	15.63055062	3단계
CT	CC	GG	0.14521248	-0.042	-0.006098924	53.37477798	3단계
TT	GG	AA	0.00020736	-0.136	-2.8201E-05	-50.97690941	1단계
TT	GG	AG	0.00217728	-0.102	-0.000222083	-13.23268206	1단계
TT	GG	GG	0.00571536	-0.068	-0.000388644	24.51154529	3단계
TT	GC	AA	0.00096768	-0.102	-9.87034E-05	-13.23268206	1단계
TT	GC	AG	0.01016064	-0.068	-0.000690924	24.51154529	3단계
TT	GC	GG	0.02667168	-0.034	-0.000906837	62.25577265	3단계
TT	CC	AA	0.00112896	-0.068	-7.67693E-05	24.51154529	3단계
TT	CC	AG	0.01185408	-0.034	-0.000403039	62.25577265	3단계
TT	CC	GG	0.03111696	0	0	100	3단계
				TOTAL AVERAGE	-0.09008		

[0101] 표 11에서 "조합 frequency"는 "해당 열의 각 유전자형별 빈도의 곱"을 나타내고, "조합 value"는 "해당 열의 각 유전자형별 value의 합"을 나타낸다. "조합 value\*조합 frequency"는 "해당 열의 조합 frequency\*해당 열의 조합 value"이고, "COMPARE TO GLOBAL MEAN (%)"은 "(해당 열의 조합 value - total average) / total average의 정수 절댓값 \* 100"으로 계산된다.

[0103] 실시예 2-3. 체질량지수(BMI) 표현형 예측

[0104] 체질량지수 표현형 예측을 위해서는 BDNF(brain derived neurotrophic factor)유전자의 다형성 rs10767664(27704439, T>A), 및 18번 염색체에 포함되고 MC4R(melanocortin 4 receptor)유전자 근처에 위치하나 특정유전자에 속하지는 않고, position 60183864으로 특정되는 다형성 rs17782313(60183864, T>C)을 이용하였다.

[0105] 연구 데이터베이스, 및 유전자 데이터로부터 도출한 상기 유전자들의 유전자형별 빈도(frequency)와 G값(value)은 하기 표 8에 나타내었다.

표 12

[0106]

RS 번호	유전자형	빈도 (frequency)	G값 (value)
rs17782313	CC	0.058564	0.576
	TC	0.366872	0.288
	TT	0.574564	0
rs10767664	TT	0.3364	0
	TA	0.4872	0.14
	AA	0.1764	0.28

[0107]

상기 표 12의 유전자형별 빈도와 G값을 실시예 1의 1 내지 5 단계에 대입하여 혈압 표현형을 예측하였다. 구체적인 도출 과정은 표 13에 나타내었다.

표 13

[0108]

rs17782313	rs10767664	조합 frequency	조합 value	조합value * 조합 frequency	COMPARE TO GLOBAL MEAN (%)	결과
CC	TT	0.01970093	0.576	0.011347735	124.1314905	3단계
CC	TA	0.02853238	0.716	0.020429185	178.6078944	3단계
CC	AA	0.01033069	0.856	0.00884307	233.0842983	3단계
TC	TT	0.12341574	0.288	0.035543733	12.06574524	2단계
TC	TA	0.17874004	0.428	0.076500736	66.54214917	3단계
TC	AA	0.06471622	0.568	0.036758813	121.0185531	3단계
TT	TT	0.19328333	0	0	-100	1단계
TT	TA	0.27992758	0.14	0.039189861	-45.52359607	2단계
TT	AA	0.10135309	0.28	0.028378865	8.95280787	2단계
			TOTAL AVERAGE	0.256992		

[0109]

표 13에서 "조합 frequency"는 "해당 열의 각 유전자형별 빈도의 곱"을 나타내고, "조합 value"는 "해당 열의 각 유전자형별 value의 합"을 나타낸다. "조합 value\*조합 frequency"는 "해당 열의 조합 frequency\*해당 열의 조합 value"이고, "COMPARE TO GLOBAL MEAN (%)"은 "(해당 열의 조합 value - total average) / total average 의 정수 절대값 \* 100"으로 계산된다.

[0111]

#### 실시예 2-4. 중성지방 표현형 예측

[0112]

중성지방 표현형 예측을 위해서는 GCKR(glucokinase regulator)유전자의 다형성 rs2293571(27506613, G>A)와 rs780092(27520287, A>G)를 이용하였다.

[0113]

연구 데이터베이스, 및 유전자 데이터로부터 도출한 상기 유전자들의 유전자형별 빈도(frequency)와 G값(value)은 하기 표 14에 나타내었다.

표 14

[0114]

RS 번호	유전자형	빈도 (frequency)	G값 (value)
rs2293571	CC	0.7744	0.04
	TC	0.2112	0.04
	TT	0.0144	0
rs780092	TT	0.459684	0.1
	CT	0.436632	0.05
	CC	0.103684	0

[0115]

상기 표 14의 유전자형별 빈도와 G값을 실시예 1의 1 내지 5 단계에 대입하여 혈당 표현형을 예측하였다. 구체



적인 도출 과정은 표 15에 나타내었다.

표 15

rs2293571	rs780092	조합 frequency	조합 value	조합value * 조합 frequency	COMPARE TO GLOBAL MEAN (%)	결과
CC	TT	0.35597929	0.14	0.049837101	30.56778333	3단계
CC	CT	0.33812782	0.09	0.030431504	-16.06356786	2단계
CC	CC	0.08029289	0.04	0.003211716	-62.69491905	1단계
TC	TT	0.09708526	0.14	0.013591937	30.56778333	3단계
TC	CT	0.09221668	0.09	0.008299501	-16.06356786	2단계
TC	CC	0.02189806	0.04	0.000875922	-62.69491905	1단계
TT	TT	0.00661945	0.1	0.000661945	-6.73729762	2단계
TT	CT	0.0062875	0.05	0.000314375	-53.36864881	1단계
TT	CC	0.00149305	0	0	-100	1단계
			TOTAL AVERAGE	0.107224		

표 15에서 "조합 frequency"는 "해당 열의 각 유전자형별 빈도의 곱"을 나타내고, "조합 value"는 "해당 열의 각 유전자형별 value의 합"을 나타낸다. "조합 value\*조합 frequency"는 "해당 열의 조합 frequency\*해당 열의 조합 value"이고, "COMPARE TO GLOBAL MEAN (%)"은 "(해당 열의 조합 value - total average) / total average 의 정수 절대값 \* 100"으로 계산된다.

#### 실시예 2-5. 콜레스테롤 표현형 예측

##### 실시예 2-5-1. 고밀도지단백질(HDL) 표현형 예측

고밀도지단백질(HDL) 표현형 예측을 위해서는 MYL2(myosin light chain 2)유전자의 다형성 rs3782889(110912851, A>G), LIPG(lipase G, endothelial type)유전자의 다형성 rs3813082(49561673, A>G), 및 CETP(Cholesteryl Ester Transfer Protein)유전자의 다형성 rs6499861(56957583, C>G)을 이용하였다.

연구 데이터베이스, 및 유전자 데이터로부터 도출한 상기 유전자들의 유전자형별 빈도(frequency)와 G값(value)은 하기 표 16에 나타내었다.

표 16

RS 번호	유전자형	빈도 (frequency)	G값 (value)
rs3813082	AA	0.78251716	53.1
	AC	0.20416568	54.4
	CC	0.01331716	63.2
rs3782889	TT	0.662596	51.3
	CT	0.302808	50.1
	CC	0.034596	49.5
rs6499861	GG	0.008464	-9.328
	CG	0.167072	-4.664
	CC	0.824464	0

상기 표 16의 유전자형별 빈도와 G값을 실시예 1의 1 내지 5 단계에 대입하여 해당 표현형을 예측하였다. 구체 적인 도출 과정은 표 17에 나타내었다.

표 17

rs3813082	rs3782889	rs6499861	조합 frequency	조합 value	조합value * 조합 frequency	COMPARE TO GLOBAL MEAN (%)	결과
AA	TT	GG	0.00438852	95.072	0.417225616	-8.15728208	1단계
AA	TT	CG	0.08662562	99.736	8.639692745	-3.651702768	1단계
AA	TT	CC	0.4274786	104.4	44.62876568	0.853876544	2단계
AA	CT	GG	0.00200557	93.872	0.188266453	-9.316522041	1단계
AA	CT	CG	0.03958812	98.536	3.900855067	-4.810942728	1단계
AA	CT	CC	0.19535877	103.2	20.16102505	-0.305363416	2단계
AA	CC	GG	0.00022914	93.272	0.021372076	-9.896142021	1단계
AA	CC	CG	0.00452297	97.936	0.442961307	-5.390562708	1단계
AA	CC	CC	0.02231986	102.6	2.29001758	-0.884983396	2단계
AC	TT	GG	0.001145	96.372	0.110346376	-6.90143879	1단계
AC	TT	CG	0.02260139	101.036	2.283554416	-2.395859478	1단계
AC	TT	CC	0.11153296	105.7	11.78903436	2.109719835	3단계
AC	CT	GG	0.00052327	95.172	0.049800641	-8.06067875	1단계
AC	CT	CG	0.01032889	99.836	1.031195308	-3.555099438	1단계
AC	CT	CC	0.05097084	104.5	5.326452664	0.950479874	2단계
AC	CC	GG	5.9784E-05	94.572	0.005653884	-8.64029873	1단계
AC	CC	CG	0.00118008	99.236	0.117106648	-4.134719418	1단계
AC	CC	CC	0.00582345	103.9	0.605056419	0.370859894	2단계
CC	TT	GG	7.4685E-05	105.172	0.00785482	1.599654252	3단계
CC	TT	CG	0.00147423	109.836	0.161923099	6.105233564	3단계
CC	TT	CC	0.00727499	114.5	0.832985825	10.61081288	3단계
CC	CT	GG	3.4131E-05	103.972	0.003548714	0.440414292	2단계
CC	CT	CG	0.00067372	108.636	0.073190784	4.945993604	3단계
CC	CT	CC	0.00332469	113.3	0.376686945	9.451572916	3단계
CC	CC	GG	3.8995E-06	103.372	0.000403103	-0.139205688	2단계
CC	CC	CG	7.6973E-05	108.036	0.008315908	4.366373624	3단계
CC	CC	CC	0.00037985	112.7	0.042808806	8.871952936	3단계
				TOTAL AVERAGE	103.5161003		

표 17에서 "조합 frequency"는 "해당 열의 각 유전자형별 빈도의 곱"을 나타내고, "조합 value"는 "해당 열의 각 유전자형별 value의 합"을 나타낸다. "조합 value\*조합 frequency"는 "해당 열의 조합 frequency\*해당 열의 조합 value"이고, "COMPARE TO GLOBAL MEAN (%)"은 "(해당 열의 조합 value - total average) / total average의 정수 절댓값 \* 100"으로 계산된다.

#### 실시예 2-5-2. 저밀도지단백질(LDL) 표현형 예측

저밀도지단백질(LDL) 표현형 예측을 위해서는 MYL2(myosin light chain 2)유전자의 다형성 rs3782889(110912851, A>G), 및 LIPG(lipase G, endothelial type)유전자의 다형성 rs3813082(49561673, A>G)를 이용하였다. 연구 데이터베이스, 및 유전자 데이터로부터 도출한 상기 유전자들의 유전자형별 빈도(frequency)와 G값(value)은 하기 표 18에 나타내었다.

표 18

RS 번호	유전자형	빈도 (frequency)	G값 (value)
rs3813082	AA	0.78251716	117.7
	AC	0.20416568	118.3
	CC	0.01331716	99

rs3782889	TT	0.662596	116.5
	CT	0.302808	116.5
	CC	0.034596	117.8

[0132] 상기 표 18의 유전자형별 빈도와 G값을 실시예 1의 1 내지 5 단계에 대입하여 혈당 표현형을 예측하였다. 구체적인 도출 과정은 표 19에 나타내었다.

표 19

rs3813082	rs3782889	조합 frequency	조합 value	조합value * 조합 frequency	COMPARE TO GLOBAL MEAN (%)	결과
AA	TT	0.51849274	234.2	121.4309997	0.034835651	2단계
AA	CT	0.23695246	234.2	55.49426524	0.034835651	2단계
AA	CC	0.02707196	235.5	6.375447444	0.590110144	3단계
AC	TT	0.13527936	234.8	31.76359441	0.291116186	3단계
AC	CT	0.061823	234.8	14.51604069	0.291116186	3단계
AC	CC	0.00706332	236.1	1.667648876	0.84639068	3단계
CC	TT	0.0088239	215.5	1.901549792	-7.952574369	1단계
CC	CT	0.00403254	215.5	0.869012927	-7.952574369	1단계
CC	CC	0.00046072	216.8	0.099884197	-7.397299876	1단계
			TOTAL AVERAGE	234.1184433		

[0134] 표 19에서 "조합 frequency"는 "해당 열의 각 유전자형별 빈도의 곱"을 나타내고, "조합 value"는 "해당 열의 각 유전자형별 value의 합"을 나타낸다. "조합 value\*조합 frequency"는 "해당 열의 조합 frequency\*해당 열의 조합 value"이고, "COMPARE TO GLOBAL MEAN (%)"은 "(해당 열의 조합 value - total average) / total average의 정수 절대값 \* 100"으로 계산된다.

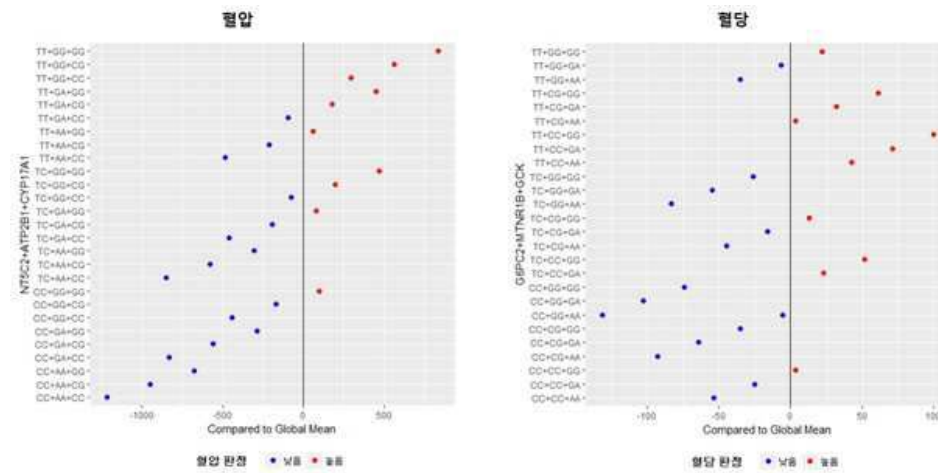
[0136] 실시예 2-5-3. 콜레스테롤 표현형 예측

[0137] 콜레스테롤은 HDL과 LDL판정 결과에 따라 상기 도 4, 및 표 7과 같이 판정한다.

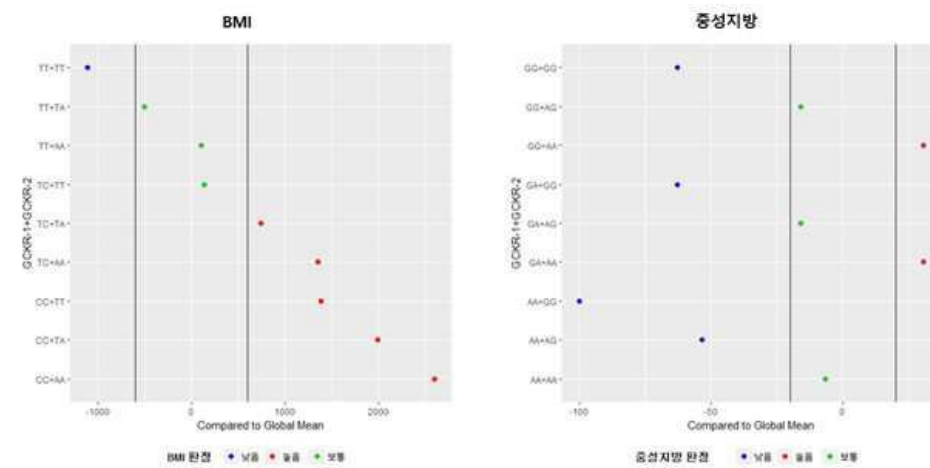
[0139] 이상으로 본 발명의 특정한 부분을 상세히 기술하였는 바, 당업계의 통상의 지식을 가진 자에게 있어서 이러한 구체적인 기술은 단지 바람직한 구현 예일 뿐이며, 이에 본 발명의 범위가 제한되는 것이 아닌 점은 명백하다. 따라서 본 발명의 실질적인 범위는 첨부된 청구항과 그의 등가물에 의하여 정의된다고 할 것이다.

## 도면

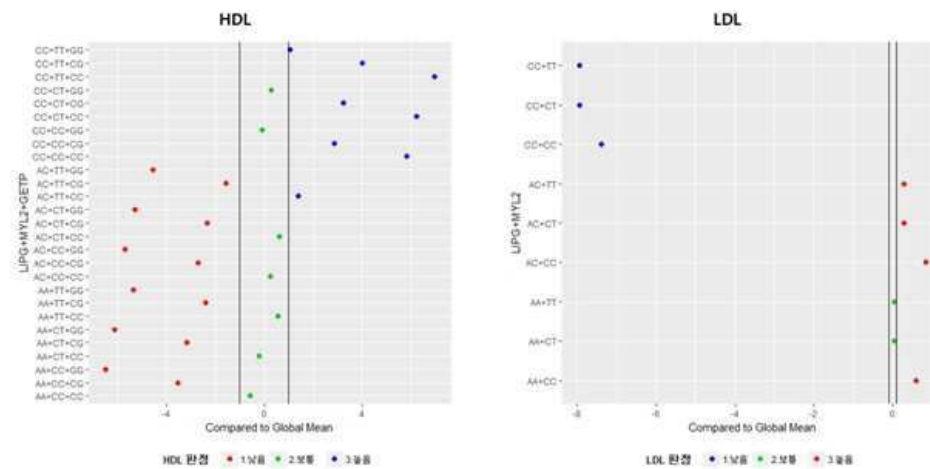
### 도면1



### 도면2



### 도면3



도면4

