



(19) 대한민국특허청(KR)
 (12) 등록특허공보(B1)

(45) 공고일자 2019년08월01일
 (11) 등록번호 10-2006417
 (24) 등록일자 2019년07월26일

- (51) 국제특허분류(Int. Cl.)
C12Q 1/68 (2018.01) *G06F 17/00* (2019.01)
 (21) 출원번호 10-2013-7023254
 (22) 출원일자(국제) 2012년01월30일
 심사청구일자 2017년01월24일
 (85) 번역문제출일자 2013년09월02일
 (65) 공개번호 10-2014-0040700
 (43) 공개일자 2014년04월03일
 (86) 국제출원번호 PCT/US2012/023195
 (87) 국제공개번호 WO 2012/106267
 국제공개일자 2012년08월09일
 (30) 우선권주장
 61/437,788 2011년01월31일 미국(US)

- (56) 선행기술조사문헌
Lancet Neurol., Vol. 8, No. 1, pp. 57-66
 (2009.01.)*
Bioinformatics, Vol. 22, No. 1, pp. 7-12
 (2005.11.02.)*

*는 심사관에 의하여 인용된 문헌

전체 청구항 수 : 총 3 항

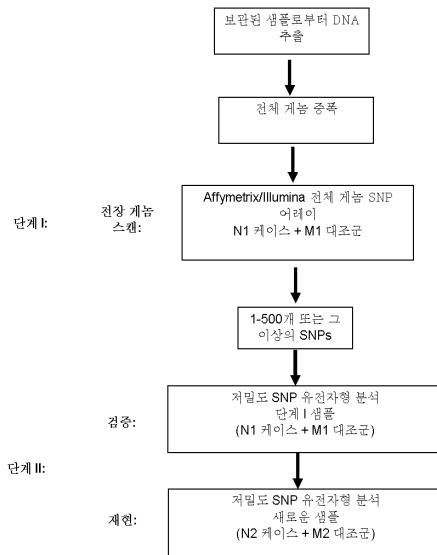
심사관 : 이준혁

(54) 발명의 명칭 게놈약학적 바이오마커 동정 방법

(57) 요 약

본 발명은 치료적 제제에 대하여 개인별로 다른 반응(효능, 부작용 및 기타 종말점)과 관련된 게놈약학적 바이오마커를 동정하는 방법에 관한 것이다. 본 발명은 신규한 게놈약학적 바이오마커를 식별하기 위하여, 보관된 임상 샘플을 이용하여 전장 게놈 연관 분석(genome-wide association study)을 수행하는 수단을 제공한다. 그 후, 새로이 발견된 바이오마커는 약물 반응을 예측하고, 치료가 이익이 될 대상들에게만 약물을 투여하거나 부작용이 있을 수도 있는 대상자들을 제외시키는 데 도움이 될 수 있는 동반 진단 테스트(companion diagnostic test)로 개발될 수 있다.

대 표 도 - 도1



명세서

청구범위

청구항 1

- a) 연관 표현형에 있어서 상이한 값을 나타내는 5명 또는 그 이상의 환자의, 종전에 시행된 임상 시험으로부터의 보관된 임상 샘플로부터 DNA를 단리하는 단계;
- b) 상기 단리된 DNA를 증폭시키는 단계로서, 상기 증폭은 전체 게놈 증폭 (WGA)이고, 그 결과물 DNA는 전체 게놈 증폭된 DNA (wgaDNA)인 것인 단계;
- c) 전장 게놈 유전자형 호출 알고리즘 (genome-wide genotype calling algorithm)을 이용함으로써 상기 증폭된 DNA의 제1 유전자형 분석 (genotyping) 데이터를 얻는 단계로서, 상기 전장 게놈 유전자형 호출 알고리즘의 호출률 컷-오프 값은 수회 조정되어 샘플들을 포함 또는 배제함으로써, 상기 증폭된 DNA로부터의 서열 정보에 기초하여 상기 호출률 컷-오프 값이 최적화되고, 상기 최적화된 호출률 컷-오프 값은 95% 미만이고, 유전자형들은 상기 서열 정보 및 상기 최적화된 호출률 컷-오프 값에 기초하여 계산되는 것인 단계;
- d) 상기 제1 유전자형 분석 데이터 및 상기 연관 표현형에 있어서 상이한 값을 기초로 관련성 분석을 수행하여 상기 연관 표현형의 하나 이상의 게놈약학적 바이오마커들을 식별하는 단계;
- e) 검증 유전자형 호출 알고리즘 (verification genotype calling algorithm)을 이용하여 종전에 시행된 임상 시험으로부터의 추가적인 환자들의 보관된 임상 샘플로부터 제2 유전자형 분석 데이터를 얻는 단계로서, 상기 검증 유전자형 호출 알고리즘의 호출률 컷-오프 값은 수회 조정되어 샘플들을 포함 또는 배제하는 것인 단계;
- f) 상기 검증 유전자형 호출 알고리즘을 이용하여 얻어진 상기 제2 유전자형 분석 데이터와 상기 전장 게놈 유전자형 호출 알고리즘을 이용하여 얻어지는 상기 유전자형 분석 데이터를 비교하는 단계; 및
- g) 상기 제2 유전자형 분석 데이터에 기초하여 관련성 분석을 수행하여 단계 d)에서 식별된 하나 이상의 게놈약학적 바이오마커들과 상기 연관 표현형 간의 관련성을 검증하는 단계를 포함하고,

종전에 시행된 임상 시험으로부터의 보관된 임상 샘플에 대한 소급적 연구를 위하여 단계 d)에서 식별된 상기 게놈약학적 바이오마커들의 하위세트 (subset)는 단계 g)에서 검증되며,

상기 보관된 임상 샘플은 혈장 샘플이고,

상기 단리된 DNA는 전혈로부터 단리된 게놈 DNA와 질적 및/또는 양적으로 하위인 차선의 (suboptimal) 게놈 DNA인 것인, 하나 이상의 게놈약학적 바이오마커들을 식별하는 방법.

청구항 2

삭제

청구항 3

제1항에 있어서, 상기 전장 게놈 유전자형 호출 알고리즘의 상기 호출률 컷-오프 값은 더 많은 상기 보관된 임상 샘플들을 포함시키기 위하여 조정되는 것인 방법.

청구항 4

제1항에 있어서, 상기 전장 게놈 유전자형 호출 알고리즘은 대치 알고리즘 (imputation algorithm)인 것인 방법.

청구항 5

삭제

청구항 6

삭제

청구항 7

삭제

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

발명의 설명

기술 분야

[0001] 본 특허 출원은 모든 도면 및 인용된 간행물 및 문헌을 포함하여 그 전체가 참조로서 본 출원에 포함되는 2011년 1월 31일자 출원된 미국 특허 가출원 제61/437,788호에 대한 우선권을 주장한다.

[0002] 본 발명은 치료적 제제에 대하여 개인별로 다른 반응 (효능, 부작용 및 기타)을 예측하는 동반 진단 테스트 (companion diagnostic test)로 개발될 수 있는 게놈약학적 바이오마커를 동정하는 방법에 관한 것이다.

배경 기술

[0003] 약학 산업은 수십 년간 "하나의 약물은 모두에게 적합하다"라는 패러다임 아래 진행되어 왔다. 그러나, 오로지 몇몇 약물만이 모든 환자군에서 보편적인 효능을 제공하고, 심지어 일부는 특정 환자 그룹에서 심각한 부작용을 야기한다. 하늘로 치솟는 R&D 비용과 엄격한 FDA 검토 기준에 더하여, 이러한 장애는 많은 새로 개발되는 약물들이 시장에 진입 실패하는 결과를 낳았다. 그러므로, 약물에 대한 잠재적인 반응을 예측할 수 있는 게놈약학적 바이오마커를 동정하는 것은 이들 다수의 실패 약물의 숨겨진 가치를 창출하는 이상적인 해결책이 될 것이다.

[0004] 한편, HapMap 프로젝트의 완성 및 마이크로어레이 기술의 급속한 발전으로, 유전자 다양성에 관한 전장 게놈 스캔이 일상적인 작업이 되었고, 수백의 전장 게놈 연관 분석 (genome-wide association study, GWAS)이 성공적으로 수행, 심혈관 질환, 당뇨 등의 통상적인 질병과 관련된 다수의 공지되거나 신규한 유전자 변이체의 발견을 이끌고 있다. 또한, 이러한 기술은 약물 반응과 관련된 유전자 다양성 식별에 대한 강력한 수단을 제공하고 있다. 그러나, Guessous 등이 약 300 건의 GWAS 연구를 검토한 결과, 이들 중 오직 12건만이 게놈약학적 연구였다 (Genome Med (2009) 1:46). 이들 12건의 연구 중, 오직 2건만이 임상적 시험에 사용된 GWAS이다. 명백히, 제약사들은 학문 연구자들과 비교하여 매우 느린 속도로 이 새로운 기술을 수용하였다. 가장 통상적인 원인 중 하나는 제약사들이 임상적 시험이 완료된 후에야 이러한 연구를 수행하여야 할 필요성을 깨닫는다는 것이고, 이것은 종종 적당한 샘플이 상기 시험 중에 수집되지 않기 때문에 그 시점에서는 상당히 늦게 된다. 이런 시나리오는 업계에서 게놈약물학의 개념을 널리 채택하고 임상적 시험을 시작하기 전에 바이오마커 설계를 포함하는 것을 개시하기까지는 가까운 미래에까지 변하지 않을 것으로 보인다.

[0005] 기술의 진보와 늘어나는 비용에도 불구하고, 대부분 전반적인 환자군에서의 유의한 효능 부족 또는 안정성 프로파일을 만족시키지 못한다는 이유로, 많은 수의 새로 개발되는 약물들이 임상 시험 단계 III에서 실패한다. 그러나, 실패한 약물 중 많은 수가 여전히 환자군의 하위집단에게는 유익할 수 있고 또는 소수의 환자에게서만 부작용을 일으키는 것이다. "구식의" 후보 유전자 접근과 비교하여, GWAS는 가설이 필요없고, 연구되는 임상적 종말점과 관계된 메커니즘에 대한 사전 지식을 요하지 않는다. 어레이 기술의 진보는 백만 또는 그 이상의 SNP 유전자형 분석 (genotyping)도 가능하게 만들었고, 이것은 단일 어레이 상에서 인간의 전체 유전자를 커버 할 수 있다. 따라서, GWAS는 약물 특이적 게놈약학적 바이오마커 연구를 위한 이상적인 옵션이다.

[0006] GWAS에서 통상적으로 사용되는 게놈 DNA원은 DNA가 풍부한 조직/세포, 예컨대 전혈이고, 이것은 적정한 양의 고품질 게놈 DNA를 제공할 수 있다. 그러나, 대부분의 제약사들은 그 임상적 시험에 게놈약학적 연구를 포함시키고 있지 않다. 따라서, 보통 GWAS 연구에 대하여 수집된 전용 샘플은 없다. 많은 경우, 다른 목적, 예컨대 조직검사 (병리용) 및 혈장 샘플 (약동학 연구용)을 위하여 수집된 인간 표본은 있을 수 있다. 그 나머지 임상 샘플을 잠재적으로 게놈 DNA를 얻는 데 사용할 수 있다. 몇몇 연구는 혈장 샘플로부터 게놈 DNA를 추출할 수 있고 소수의 SNP에 대한 성공적인 유전자형 분석을 이끌어 낼 수 있음을 보여주었다 (Sjoholm et al., Cancer Epidemiol Biomarkers Prev (2005) 14:251; Lu et al., Biotechniques (2005) 39:511; Park et al., Clin Chem (2005) 51:1520; Bergen et al., Hum Mutat (2005) 26:262). 그러나, 종종 부적절하게 처리되고

저장되는, 보관된 임상 샘플로부터의 게놈 DNA의 양과 질은 특히 GWAS에 사용하기 위하여 최적 상태와는 거리가 멀다. 사실, 고밀도 SNP 어레이에 혈장 샘플로부터 추출된 게놈 DNA를 사용하여 보고서를 발표한 몇몇 저자들은 이를 샘플로부터의 DNA 품질이 너무 "열악하여" 성공적으로 GWAS 결과를 낼 수 없었을 것이라고 시사하였다 (Croft et al., J Mol Diagn (2008) 10:249; Bucatas et al., BMC Genet (2009) 10:85). 유일하게 성공적으로 시행된 GWAS는 건조 혈액 스팟 샘플 (Hollegaard et al., BMC Genomics (2009) 10:297-302) 또는 환자의 혈액 (Singer et al., Nat Genet (2010) 42:711-714)을 사용하였다. 따라서, 우리가 아는 한, 누구도 보관된 임상 샘플로부터 게놈 DNA를 사용하여 성공적 GWAS를 수행 시도하지 않았다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0007]

발명의 요약

[0008]

[0007] 간행된 대부분의 GWAS가 전용 샘플, 예컨대 잘 설계된 연구의 전혈 등으로부터 고품질의 DNA를 사용하여 일반적인 질병의 원인이 되는 유전자 변이체를 발견하는 것에 초점을 두었던 반면, 강력한 기술인 본 출원은 신약 임상 시험에 거의 포함되지 않았으므로 게놈약학적 연구에 있어서 그 성공이 제한되었다. 현재 이루어지고 있는 공감대는 성공적인 GWAS는 오로지 충분량의 고품질 게놈 DNA를 사용하여서만 이루어질 수 있고, 차선의 게놈 DNA 및/또는 전혈 증폭된 DNA는 GWAS에 있어서 권장되지 않거나 포기된다. 본 발명은 차선의 게놈 DNA, 예컨대 보관된 임상 샘플로부터 추출된 게놈 DNA를 이용하여, 약물 반응을 예측하는 게놈약학적 바이오마커의 GWAS에 의한 동정 방법을 개시한다.

[0009]

[0008] 한 가지 측면에서, 본 발명은 하나 이상의 게놈약학적 바이오마커를 식별하는 방법을 제공하는데, 이 방법은 a) 연관된 표현형에 있어서 상이한 값을 나타내는 2 이상의 환자의 보관된 임상 샘플로부터 DNA를 단리하는 단계; b) 상기 단리된 DNA를 증폭시키는 단계; c) 상기 증폭된 DNA의 고밀도 유전자형 분석 결과를 얻는 단계; 및 d) 상기 유전자형 분석 결과 및 상기 연관된 표현형의 상이한 값에 기초하여 관련 분석을 수행하는 단계를 포함하고, 상기 게놈약학적 바이오마커(들)이 식별되는 방법이다.

[0010]

[0009] 몇 가지 실시 상태에 있어서, 보관된 임상 샘플은 혈장 샘플, 혈청 샘플, 건조 혈액 스팟, 소변 샘플, 조직 샘플, 종양 세포 및 구강 면봉 표본 (buccal swab)으로 이루어지는 군으로부터 선택될 수 있다. 몇 가지 실시 상태에 있어서, 상기 보관된 임상 샘플은 혈장 샘플일 수 있다. 몇 가지 실시 상태에 있어서, 상기 보관된 임상 샘플은 약 2~1000명 또는 그 이상의 환자로부터 유래한 것일 수 있다.

[0011]

[0010] 몇 가지 실시 상태에 있어서, 상기 단리된 DNA는 차선의 게놈 DNA일 수 있다. 몇 가지 실시 상태에 있어서, 그 증폭은 전체 게놈 증폭 (whole-genome amplification, WGA)일 수 있고, 그 결과적인 DNA는 전체 게놈 증폭 DNA (whole-genome amplified DNA, wgaDNA)일 수 있다. 몇 가지 실시 상태에 있어서, 고밀도 유전자형 분석은 전체 게놈 유전자형 분석일 수 있다. 몇 가지 실시 상태에 있어서, 상기 고밀도 유전자형 분석은 단일 염기 다형성 (SNP)을 이용함으로써 시행될 수 있다. 몇 가지 실시 상태에 있어서, 약 1000~5000000 또는 그 이상의, 좋기로는 약 1000000의 SNP가 고밀도 유전자형 분석에 사용될 수 있다. 몇 가지 실시 상태에 있어서, 고밀도 유전자형 분석은 어레이 기반인 것일 수 있다.

[0012]

[0011] 몇 가지 실시 상태에 있어서, 유전자형 분석 결과는 전장 게놈 유전자형 호출 알고리즘을 이용하여 얻을 수 있다. 몇 가지 실시 상태에 있어서, 이 방법은, e) 전장 게놈 유전자형 호출 알고리즘의 호출률 컷-오프 값을 조정하는 단계를 더 포함할 수 있다. 몇 가지 실시 상태에 있어서, 단계 d) 및 단계 e)는 수회 반복되어 샘플들을 포함 및/또는 재외시키고 전장 게놈 유전자형 호출 알고리즘을 최적화할 수 있다. 몇 가지 실시 상태에 있어서, 상기 전장 게놈 유전자형 호출 알고리즘에 대한 포함 기준을 식별할 수 있다. 몇 가지 실시 상태에 있어서, 상기 유전자형 호출은 고품질의 게놈 DNA의 전체 게놈 유전자형 분석용으로 사용되는 통상적인 호출률 컷-오프보다 낮은 호출률 컷-오프 값을 사용하여 이루어질 수 있고, 여기서 사용되는 상기 호출률 컷-오프 값은 50%, 60%, 70%, 80%, 90% 또는 95%일 수 있다. 몇 가지 실시 상태에 있어서, 유전자형 호출은 Affymetrix Genotyping Console™ 소프트웨어로 이루어질 수 있다. 몇 가지 실시 상태에 있어서, 상기 유전자형 호출은 BRLMM 알고리즘을 이용하여 이루어질 수 있다. 몇 가지 실시 상태에 있어서, 상기 유전자형 호출은 대치 알고리즘 (imputation algorithm)을 이용하여 이루어질 수 있고, 여기서 상기 대치 알고리즘을 위하여 HapMap이 사용

될 수 있다.

[0013] 몇 가지 실시 상태에 있어서, 관련 분석은 GWAS일 수 있다. 몇 가지 실시 상태에 있어서, 상기 관련 분석은 연관된 표현형을 갖는 각 SNP의 관련 p-값을 계산함으로써 수행될 수 있다. 몇 가지 실시 상태에 있어서, 상기 계산은 대립유전자 (allele) 빈도 및/또는 유전자형 기반 테스트에 기초할 수 있다. 몇 가지 실시 상태에 있어서, 연관 표현형은 범주적 특성, 양적 특성 또는 또 다른 연관 표현형일 수 있다.

[0014] 몇 가지 실시 상태에 있어서, 상기 방법은 식별된 게놈약학적 바이오마커를 이용한 추가적인 유전자형 분석 결과에 기초하여 관련 분석을 수행하는 단계를 더 포함할 수 있다. 몇 가지 실시 상태에 있어서, 약 1~500 또는 그 이상의 식별된 게놈약학적 바이오마커는 상기 추가적인 유전자형 분석에 사용될 수 있다. 몇 가지 실시 상태에 있어서, 단계 a)의 보관된 임상 샘플 전부 또는 일부 및/또는 추가적인 임상 샘플이 상기 추가적인 유전자형 분석을 위하여 사용될 수 있다. 몇 가지 실시 상태에 있어서, 추가적인 유전자형 분석 결과는 검증 유전자형 호출 알고리즘을 이용하여 얻을 수 있다. 몇 가지 실시 상태에 있어서, 상기 방법은 상기 검증 유전자형 호출 알고리즘의 호출률 컷-오프 값을 조정하는 단계를 더 포함할 수 있다. 몇 가지 실시 상태에 있어서, 유전자형 분석 및 호출률 컷-오프의 조정은 수회 반복되어 샘플을 포함 및/또는 제외시킬 수 있다. 몇 가지 실시 상태에 있어서, 최적의 포함 기준이 상기 검증 유전자형 호출 알고리즘에 대하여 식별될 수 있다. 몇 가지 실시 상태에 있어서, 상기 방법은 전장 게놈 유전자형 호출 알고리즘을 이용하여 얻은 유전자형 분석 결과를 상기 검증 유전자형 호출 알고리즘을 이용하여 얻은 추가적인 유전자형 분석 결과와 비교하는 단계를 더 포함할 수 있다. 몇 가지 실시 상태에 있어서, 단계 d)로부터 게놈약학적 바이오마커의 하위집합이 식별될 수 있다.

[0015] 몇 가지 실시 상태에 있어서, 상기 방법은 종전에 시행된 임상 시험으로부터의 보관된 임상 샘플에 대한 소급 연구용으로 사용될 수 있다. 몇 가지 실시 상태에 있어서, 상기 방법은 게놈약학적 바이오마커의 신생 (*de novo*) 식별을 위하여 사용될 수 있다.

[0016] 또 다른 측면에 있어서, 본 발명은 본 발명에 개시된 방법에 의하여 식별되는 게놈약학적 바이오마커 또는 1군의 게놈약학적 바이오마커를 제공하고, 여기서 상기 바이오마커는 하나 이상의 SNP일 수 있다. 몇 가지 실시 상태에 있어서, 상기 게놈약학적 바이오마커는 하나 이상의 추가적인 게놈약학적 바이오마커를 식별하기 위하여 사용될 수 있다. 몇 가지 실시 상태에 있어서, 상기 게놈약학적 바이오마커는 동반 진단 테스트를 개발하는 데 사용될 수 있다.

[0017] 또 다른 측면에서, 본 발명은 본 발명에 개시된 방법에 의하여 식별되는 게놈약학적 바이오마커를 이용하는 동반 진단 테스트를 제공한다. 또한, 본 발명은 본 발명에 개시되는 동반 진단 테스트를 이용한 치료에 대한 대상체의 반응을 예단하는 방법을 제공한다. 또한 본 발명은 본 발명에 개시되는 방법에 의하여 식별된 게놈약학적 바이오마커를 이용하여 신규한 약물 표적을 식별하는 방법을 제공한다. 본 발명의 방법은, 임상전문가가, 치료하기 위한 환자를 식별, 치료요법의 진행 과정 중 환자 선별에 도움, 특정한 치료법으로 각각의 환자를 치료하는 경우의 성공 가능성을 예측, 질병 진행을 평가 및 모니터링, 치료 효능을 모니터링 및 개개인 환자들에 대한 예후를 결정하는 데 유용하다. 이를 실시 상태들 중 임의의 것이 본 발명에 포함된다.

[0018] 추가적인 측면에 있어서, 본 발명은 본 발명에서 개시되는 방법에 의하여 식별되는 게놈약학적 바이오마커, 또는 1군의 게놈약학적 바이오마커 시험용 시약을 포함하는 키트를 제공한다. 몇 가지 실시 상태에 있어서, 상기 키트는 동반 진단 테스트를 시행하기 위하여 상기 게놈약학적 바이오마커 사용에 대한 지시사항을 더 포함할 수 있다.

[0019] 또 하나의 다른 측면에 있어서, 본 발명은 차선의 게놈 DNA 샘플을 이용한 유전자형 분석 방법을 제공하고, 이 방법은: a) 상기 차선의 게놈 DNA 샘플의 서열 정보를 받는 단계; b) 상기 서열 정보에 기초하여 포함 기준을 최적화하는 단계; 및 c) 상기 서열 정보 및 상기 최적화된 포함 기준에 기초하여 유전자형을 계산하는 단계를 포함한다. 몇 가지 실시 상태에 있어서, 상기 최적화는 수회 반복되어 샘플을 포함 및/또는 제외시킬 수 있다. 몇 가지 실시 상태에 있어서, 최적의 포함 기준이 식별될 수 있다. 몇 가지 실시 상태에 있어서, 상기 유전자형 분석 결과는 전장 게놈 유전자형 호출 알고리즘 및/또는 검증 유전자형 호출 알고리즘을 이용하여 얻어질 수 있다. 몇 가지 실시 상태에 있어서, 상기 포함 기준은 상기 유전자형 호출 알고리즘의 호출률 컷-오프 값을 일 수 있다. 몇 가지 실시 상태에 있어서, 상기 유전자형 호출은 고품질 또는 최적의 게놈 DNA의 전체 게놈 유전자형 분석을 위하여 사용되는 통상적인 호출률 컷-오프보다 낮은 호출률 컷-오프를 사용하여 이루어질 수 있고, 여기서 사용되는 상기 호출률 컷-오프 값은 약 50%, 60%, 70%, 80%, 90% 또는 95%일 수 있다. 몇 가지 실시 상태에 있어서, 유전자형 분석 결과는 다수의 유전자형 분석 플랫폼을 이용하여 얻을 수 있다. 몇 가지 실시 상태에 있어서, 상기 다수의 유전자형 분석 플랫폼으로부터의 유전자형 분석 결과는 최적화를 위하여 비교될 수

있다.

[0020] 또한, 차선의 게놈 DNA 샘플을 이용한 유전자형 분석 방법을 이용한 관련 분석을 수행하는 방법이 제공되고, 이 방법은 포함 기준을 최적화하는 단계를 포함한다. 몇 가지 실시 상태에 있어서, 상기 관련 분석은 최적화를 위하여 수회 반복될 수 있다. 또한, 본 발명은 차선의 게놈 DNA 샘플을 이용한 유전자형 분석 방법용의 복수의 지시사항을 포함하는 컴퓨터 분석가능한 매체를 제공하고, 이것은 a) 상기 차선의 게놈 DNA 샘플의 서열 정보를 받는 단계; b) 상기 서열 정보에 기초하여 포함 기준을 최적화하는 단계; 및 c) 상기 서열 정보 및 상기 최적화된 포함 기준에 기초하여 유전자형을 계산하는 단계를 포함한다.

[0021] 또 다른 측면에 있어서, 본 발명은 차선의 게놈 DNA를 이용하여 GWAS를 시행하는 방법을 제공한다. 몇 가지 실시 상태에 있어서, 상기 차선의 게놈 DNA는 보관된 샘플로부터 유래할 수 있다. 몇 가지 실시 상태에 있어서, 상기 차선의 게놈 DNA는 혈장 샘플로부터 유래할 수 있다. 몇 가지 실시 상태에 있어서, 상기 차선의 게놈 DNA는 증폭될 수 있다. 몇 가지 실시 상태에 있어서, 다수의 유전자형 분석 플랫폼이 사용될 수 있다. 몇 가지 실시 상태에 있어서, 동일하거나 상이한 샘플들이 다수의 유전자형 분석 플랫폼으로 사용될 수 있다. 몇 가지 실시 상태에 있어서, 상기 방법은 고품질 게놈 DNA를 제공하는 샘플을 더 이용할 수 있다.

도면의 간단한 설명

[0023] 도 1은 GWAS를 이용한 게놈약학적 바이오마커 동정의 예시적 흐름도를 보여준다. 게놈 DNA는 보관된 임상 샘플로부터 추출되고 WGA를 이용하여 증폭된다. 동정인 단계 I에서는, 케이스 그룹으로부터의 N1 (1 내지 1000 또는 그 이상 중 하나의 숫자) 샘플들 및 대조군으로부터의 M1 (1 내지 1000 또는 그 이상 중 하나의 숫자) 샘플들을 Affymetrix 및/또는 Illumina 전체 게놈 SNP 어레이를 이용하여 유전자형 분석한다. 케이스-대조군 상태 (즉, 반응 대 비반응)에 대한 각 SNP의 관련은 대립 유전자 빈도 및/또는 유전자형 기반 테스트에 기초하여 계산된다. 계속하여, 1 내지 500개 또는 그 이상의 대부분의 유의하게 관련된 SNP가 단계 II 연구용으로 선별된다. 단계 II에서는, 단계 I에서 얻어진 결과를 검증하기 위한 저밀도 SNP 유전자형 분석 플랫폼 (단계 I에서 사용된 것들로부터의 특유한 유전자형 분석 기술)을 이용하여 단계 I에서 사용된 동일한 샘플을 유전자형 분석한다. 추가적인 새로운 샘플들, 케이스 그룹으로부터의 N2 (1 내지 1000 또는 그 이상 중 하나의 숫자) 샘플 및 대조군으로부터의 M2 (1 내지 1000 또는 그 이상 중 하나의 숫자) 샘플을 유전자형 분석하여 알아낸 것을 재현한다 (replicate).

[0024] 도 2는 예시적인 전장 게놈 동정 단계 결과 분석의 흐름도를 보여준다.

[0025] 도 3은 예시적인 검증 단계 결과 분석의 흐름도를 보여준다.

발명의 상세한 설명

[0027] 본 발명은 하나 이상의 유전자형 분석 기술 또는 플랫폼을 적용함으로써 보관된 임상적 샘플로부터 단리된 게놈 DNA를 이용하는 유전자형 분석 결과의 잠재적으로 저조한 양을 극복하기 위한 신규한 접근법을 제공한다.

A. 일반적인 기법

[0029] 달리 나타내지 않는 한, 본 발명의 실시는 이 기술분야의 기술 범위 내인 문자 생물학 (재조합 기법 포함), 미생물학, 세포 생물학, 생화학, 및 면역학에 관한 종래 기법들을 채용할 것이다. 이러한 기법들은, 예컨대 문헌 [“Molecular Cloning: A Laboratory Manual”, second edition (Sambrook et al., 1989); “Oligonucleotide Synthesis” (M. J. Gait, ed., 1984); “Animal Cell Culture” (R. I. Freshney, ed., 1987); “Methods in Enzymology” (Academic Press, Inc.); “Current Protocols in Molecular Biology” (F. M. Ausubel et al., eds., 1987, and periodic updates); “PCR: The Polymerase Chain Reaction”, (Mullis et al., eds., 1994)]에 잘 설명되어 있다.

B. 정의

[0031] 달리 정의되지 않는 한, 본 명세서에서 사용되는 모든 기술적 및 과학적 용어는 본 발명이 속한 기술 분야에서 통상의 지식을 가진 자에 의하여 통상적으로 이해되는 것과 동일한 의미를 갖는다. 본 명세서에서 언급되는 모든 특허, 출원, 공개된 출원 및 기타 간행물들은 그 전체가 참조로서 본 명세서에 포함된다. 이 섹션에서 설명되는 정의가 상기 본 명세서에 참고로서 포함되는 특허, 출원, 공개된 출원 및 기타 간행물들에서 설명되는 정의와 반대되거나 또는 일치하지 않는다면, 본 섹션에서 설명되는 정의가 참조로서 본 명세서에 포함되는

정의에 우선한다.

[0032] [0027] 본 명세서에서 사용되는 바, 단수형 ("a", "an" 및 "the")은 달리 나타내지 않는 한 복수형을 포함한다. 예컨대, 이량체 ("a" dimer)는 하나 이상의 이량체를 포함한다.

[0033] [0028] 본 명세서에서 사용되는 "바이오마커" 또는 "마커"라는 용어는, 일반적으로 포유류의 조직 또는 세포 또는 분비물 중 또는 이를 상에서 그 발현이 공지의 방법 (또는 본 명세서에 개시된 방법)에 의하여 검출될 수 있고 예측되는 것이거나, 포유류 세포의 또는 조직의 치료법에 대한 민감도를 예측 (또는 예측을 보조)하는데 사용될 수 있는, 몇 가지 실시 상태에 있어서는 치료법에 대한 개개인의 반응을 예측 (또는 예측을 보조)하는 데 사용될 수 있는 유전자, 단백질, 탄수화물 구조 또는 당지질 등의 분자를 말한다.

[0034] [0029] 본 명세서에서 사용되는 "게놈약학적 바이오마커"라는 용어는 대상체에서 특정 임상적 약물 반응 또는 감수성과 관련된 목적하는 바이오마커이다 (예컨대, 문헌 [McLeod et al., *Eur. J. Cancer* (1999) 35:1650-1652] 참조). 이것은 생화학적 바이오마커, 임상적 정후 또는 증상 등일 수 있다. 게놈약학적 바이오마커의 존재 또는 양은 그 약물의 투여 이전에 특정 약물 또는 약물 부류에 대한 대상체의 예측되는 반응과 관련된다. 대상체에서 하나 이상의 게놈약학적 바이오마커의 존재 또는 양을 평가함으로써, 대상체에 대하여 가장 적합한 또는 성공의 확률이 더 높다고 예측되는 약물 요법을 선택할 수 있다. 예컨대, 대상체에서의 특정 종양 마커용 DNA, RNA 또는 단백질의 존재 또는 양에 기초하여, 그 대상체에 존재할 수도 있는 특정 종양의 치료에 최적인 약물 또는 치료 방식을 선별할 수 있다. 유사하게, 특성 서열 변이 또는 다형성의 존재 또는 부재는 약물 반응과 연관되어 있을 수 있다. 그러므로, 게놈약학적 바이오마커의 사용은 각각의 대상체에게 상기 요법을 시행하여 볼 필요 없이 가장 적합한 치료를 적용할 수 있도록 한다.

[0035] [0030] 본 명세서에서 사용되는 "샘플"이라는 용어는, 예컨대 물리적, 생화학적, 화학적 및/또는 생리학적 특성에 기반하여, 특징지워지고 및/또는 식별될 세포 성분 및/또는 기타 분자 성분을 함유하는, 목적하는 대상체로부터 유래하거나 그로부터 얻어지는 조성물을 말한다. 예컨대, "임상적 샘플" 또는 "질병 샘플" 및 이들의 변이라는 용어는, 특징지워질, 예컨대 바이오마커와 같은 세포 성분 및/또는 분자 성분을 함유할 것으로 기대되거나 알려진, 목적하는 대상체로부터 얻은 임의의 샘플을 말한다.

[0036] [0031] "조직 또는 세포 샘플"이라는 용어는 대상체 또는 환자의 조직으로부터 얻어지는 유사한 세포들의 수집체를 말한다. 상기 조직 또는 세포 샘플의 제공원은 신선한, 냉동의 및/또는 보존된 기관 또는 조직 샘플 또는 생검물 또는 흡인물; 혈액 또는 임의의 혈액 성분; 뇌척수액, 양수, 복막액 또는 세포간질액과 같은 체액; 대상체의 임신 또는 발생기 중 임의의 시기의 세포로부터의 고체 조직일 수 있다. 또한, 조직 샘플은 1차 세포 또는 배양 세포 또는 세포주일 수 있다. 필요에 따라, 상기 조직 또는 세포 샘플은 질병 조직/기관으로부터 얻어진다. 상기 조직 샘플은 자연적으로는 자연 상태의 상기 조직과 혼합되지 않는 화합물, 예컨대 방부제, 항응고제, 완충액, 고정액, 영양분, 항생제, 또는 등등을 함유할 수 있다.

[0037] [0032] 본 명세서에서 사용되는 "혈장", 또는 "혈액 혈장"은 세포 외액 (세포 바깥의 모든 체액)의 혈관 내 액체 부분을 말한다. 이것은 대부분 물이고 용해된 단백질, 글루코오스, 응고 인자, 미네랄 이온, 호르몬 및 이산화탄소 (배설되는 생성물 이동을 위한 주요 매질인 혈장)을 함유한다. 혈액 혈장은 혈액 세포들이 튜브의 바닥에 가라앉을 때까지 항응고제를 함유하는 신선한 혈액 튜브를 원심분리기에서 회전시켜 준비한다. 그 후, 이 혈액 혈장을 봇거나 따라낸다. "혈액 혈청"은 피브리노겐 또는 기타 응고 인자가 없는 혈액 혈장 (즉, 전혈에서 세포 및 응고 인자들을 모두 뺀 것)을 말한다.

[0038] [0033] 본 명세서에서 상호교환적으로 사용되는 "폴리뉴클레오티드" 또는 "핵산"은 임의의 길이의 뉴클레오티드의 중합체를 말하고, DNA 및 RNA를 포함한다. 뉴클레오티드는 디옥시리보뉴클레오티드, 리보뉴클레오티드, 개질된 뉴클레오티드 또는 염기, 및/또는 이들의 유사체, 또는 DNA 또는 RNA 중합효소에 의하여 중합체로 포함될 수 있는 임의의 기질일 수 있다. 폴리뉴클레오티드는 개질된 뉴클레오티드, 예컨대 메틸화 뉴클레오티드 및 이들의 유사체를 포함할 수 있다. 존재한다면, 뉴클레오티드 구조에의 개질은 중합체의 회합 전 또는 후에 이루어질 수 있다. 비(非) 뉴클레오티드 성분이 뉴클레오티드 서열에 끼어들 수 있다. 폴리뉴클레오티드는 중합 후에 더 개질될 수 있는데, 예컨대 표지 성분을 이용한 컨쥬게이션에 의할 수 있다. 다른 유형의 개질로는, 예컨대 "캡 (cap)", 자연적으로 존재하는 하나 이상의 뉴클레오티드의 유사체로의 치환, 뉴클레오티드간 개질, 예컨대 하전되지 않은 링키지를 이용한 개질 (예컨대, 메틸 포스포네이트, 포스포트리에스테르, 포스포아미레이트, 카바메이트 등) 및 하전된 링키지를 이용한 개질 (예컨대, 포스포로티오에이트, 포스포로디티오에이트 등), 웨던트 모이어티를 함유하는 개질, 예컨대 다액질 (예컨대, 뉴클레아제, 독소, 항체, 신호 웹타이드, 폴리-L-리신 등을 함유하는 개질, 인터칼레이터 (예컨대, 아크리딘, 소랄렌 등)을 이용한 개질, 칼레이터를 함유하는 개질 (예컨

대, 금속, 방사성 금속, 봉소, 산화 금속 등), 알킬레이터를 함유하는 개질, 개질된 링키지를 이용한 개질 (예컨대, 알파 아노머 핵산 등), 및 폴리뉴클레오티드(들)의 비개질 형태를 들 수 있다. 또한, 당에 일반적으로 존재하는 하이드록시기 중 임의의 것은, 예컨대 포스포네이트기, 포스페이트기로 대체될 수 있고, 표준적인 보호기로 보호될 수 있고, 또는 활성화되어 추가적인 뉴클레오티드와의 추가적인 링키지를 만들 수 있고, 또는 고체지지체에 컨쥬게이션될 수 있다. 5' 및 3' 말단 OH는 인산화될 수 있고 또는 아민 또는 탄소 원자 1 내지 20 개의 유기 캡핑기 (capping group) 모이어티로 치환될 수 있다. 또한, 다른 하이드록시기는 표준 보호기로 유도화될 수 있다. 또하느 폴리뉴클레오티드는 이 기술 분야에 일반적으로 알려진 리보오스 또는 디옥시리보오스 당의 유사형을 포함할 수 있는데, 예컨대 2'-0-메닐-2'-0-알릴, 2'-플루오로- 또는 2'-아지도-리보오스, 카르보시클릭 당 유사체, 알파-아노머 당, 에피머 당, 예컨대 아라비노오스, 자일로오스 또는 럭소오스, 피라노오스 당, 퓨라노오스 당, 세도헵톨로오스, 비고리 (acyclic) 유사체 및 비염기성 (abasic) 뉴클레오시드 유사체, 예컨대 메틸 리보사이드를 들 수 있다. 하나 이상의 포스포디에스테르 링키지는 대체적인 연결기로 대체될 수 있다. 이들 대체적인 연결기는 포스페이트가 P(O)S ("티오에이트"), P(S)S ("디티오에이트"), "(O)NR 2 ("아미데이트"), P(O)R, P(O)OR, CO 또는 CH 2 ('포르마세탈')로 치환된 실시 상태들을 포함하지만 이에 한정되는 것은 아니고, 여기서 각각의 R 또는 R'은 독립적으로 H 또는 필요에 따라 에테르 (--O--) 링키지, 아릴, 알케닐, 시클로알킬, 시클로알케닐 또는 아랄딜을 함유하는 치환 또는 비치환된 알킬 (1~20 C)이다. 폴리뉴클레오티드 내 모든 링키지들이 동일해야 하는 것은 아니다. 전술한 설명은 본 명세서에서 언급되는 모든 폴리뉴클레오티드, 예컨대 RNA 및 DNA에 적용된다.

[0039] [0034] 본 명세서에서 사용되는 "올리고뉴클레오티드"는 일반적으로 짧은, 일반적으로 단일 가닥의, 일반적으로 약 200 뉴클레오티드 길이 미만이지만 반드시 그럴 필요는 없는 일반적으로 합성 폴리뉴클레오티드를 말한다. "올리고뉴클레오티드 및 "폴리뉴클레오티드"라는 용어는 상호 배타적이지 않다. 폴리뉴클레오티드에 관한 상기 설명은 올리고뉴클레오티드에도 동일하게 전체적으로 적용 가능한 것이다.

[0040] [0035] 본 명세서에서 사용되는 "차선의 게놈 DNA"라는 용어는 전혈과 같은 DNA가 풍부한 조직/세포로부터 단리된 게놈 DNA와 질적 및/또는 양적으로 하위인 게놈 DNA를 말한다. 차선의 게놈 DNA는 생검물 또는 혈장과 같은 보관된 임상적 샘플로부터 단리될 수 있는데, 이들은 종종 부적절하게 처리되고 저장되어, 특히 GWAS에 사용하기 위하여는 질적 및/또는 양적으로 최적화는 거리가 있다. 예컨대, 차선의 게놈 DNA 샘플은 전체 게놈에 대한 온전한 커버를 제공할 수 없거나, 게놈 DNA의 짧은 단편을 함유할 수 있다. 몇 가지 실시 상태에 있어서, 고밀도 유전자형 분석에 적용되는 차선의 게놈 DNA 샘플은 약 99%, 95%, 90%, 80%, 70%, 60%, 50%보다 낮은 호출률 또는 그 미만을 가질 수 있다. 통상적으로, GWAS에 유용한 차선의 게놈 DNA를 위하여는, 종종 단계가 요구된다.

[0041] [0036] 본 명세서에서 사용되는 "증폭"은 일반적으로 원하는 서열의 카피를 다수 제조하는 방법을 말한다. "다수의 카피"는 2 이상의 카피를 의미한다. 하나의 "카피"가 반드시 주형 서열과의 완전한 서열 상보성 또는 동일성을 의미하는 것은 아니다. 예컨대, 카피는 뉴클레오티드 유사체를 포함할 수 있는데, 예컨대 디옥시이노신, 고의적인 서열 변경 (예컨대, 주형과 혼성화할 수 있으나 상보적인 것은 아닌 서열을 포함하는 프라이머를 통하여 도입되는 서열 변경) 및/또는 증폭 중 발생하는 서열 에러를 포함할 수 있다.

[0042] [0037] 본 발명에서 사용되는 "어레이" 또는 "마이크로어레이"라는 용어는 혼성 가능한 어레이 요소, 예컨대 폴리뉴클레오티드 프로브 (예컨대, 올리고뉴클레오티드), 또는 결합 체제 (예컨대, 항체)의 기질 상의 순서있는 배열을 말한다. 상기 기질은 고체 기질, 예컨대 유리 또는 실리카 슬라이드, 비드, 섬유 광학 바인더, 또는 반고체 기질, 예컨대 니트로셀룰로오스 멤브레인일 수 있다. 상기 뉴클레오티드 서열은 DNA, RNA, 또는 이들의 임의의 순열일 수 있다.

[0043] [0038] 본 명세서에서 사용되는, "표현형"이라는 용어는 개인들 간에 비교될 수 있는 특성, 예컨대 질환의 존재 또는 부재, 개인간 시각적으로 관찰될 수 있는 상이점, 대사적 차이, 생리학적 차이, 생물학적 분자 기능의 차이 등을 말한다. 표현형은 정성적이거나 정량적일 수 있다. 표현형의 예시는 치료, 예너대 약물에 대한 반응성이다.

[0044] [0039] "반응성"은 환자에게 이익을 나타내는 임의의 종말점을 이용하여 평가될 수 있고, 예컨대 (1) 질병 진행을 어느 정도 억제, 예컨대 둔화 및 완전 정지; (2) 질병 에피소드 및/또는 증상의 수적 감소; (3) 병변 크기의 감소; (4) 인점 주변 기관 및/또는 조직으로의 질병 세포 침윤 억제 (즉, 감소, 둔화 또는 완전한 중단); (5) 질병 확산의 억제 (즉, 감소, 둔화 또는 완전한 중단); (6) 질병과 관련된 하나 이상의 증상을 어느 정도 완화; (7) 치료 후 질병이 없는 기간의 증가; (8) 치료 후 주어진 시점에서 사망률 감소; 및/또는 (9) 치료 후 무(無)부작용 등을 들 수 있으나 이에 한정되는 것은 아니다. 또한, 반응성은 환자에 대한 부작용 및/또는 독성을

나타내는 임의의 종말점을 이용하여 평가될 수도 있다.

[0045] "치료하다" 또는 "치료" 또는 "경감"은 그 목적이 표적하는 병리학적 질환 또는 장애를 둔화 (감소)시키거나 그 질환의 재발을 방지하는 것인 치료적 처리를 말한다. 치료량의 치료제를 투약받은 후, 대상체가 특정 질병의 하나 이상의 징후 또는 증상에 있어서 눈에 띄는 및/또는 측정할 수 있을만한 감소나 이들의 부재를 나타낸다면 대상체는 성공적으로 "치료된" 것이다. 예컨대, 암 세포 수의 현저한 감소 또는 상기 암 세포의 부재; 종양 크기의 감소; 종양 전이의 억제 (즉, 어느 정도 둔화 및 좋기로는 중단); 종양 생장의 어느 정도의 억제; 특정 종양과 관련된 하나 이상의 증상을 어느 정도 완화 및 또는 호전 기간의 증가; 유병률 및 사망률의 감소 및 삶의 질 개선을 들 수 있다. 치료는 암의 모든 징후들이 사라지는 것으로 정의되는 완전한 반응, 또는 종양의 크기가 좋기로는 50% 이상, 더 좋기로는 75%까지 감소하는 부분적인 반응을 달성할 수 있다. 또한, 환자가 안정적인 질병을 경험한다면 그 환자는 치료된 것으로 간주된다. 몇 가지 실시 상태에 있어서, 치료제를 이용한 치료가 효과적이면 치료 후에 환자는 3 개월 무(無)질병 상태, 좋기로는 6 개월, 더욱 좋기로는 1 년, 더 더욱 좋기로는 2년 또는 그 이상 무질병 상태인 결과를 낳는다. 질병의 성공적인 치료 및 개선을 평가하는 이들 파라미터들은 이 기술 분야에서 적절한 기술을 갖춘 의사들에게 친숙한 일반적인 방법으로 쉽게 측정 가능하다.

[0046] 본 명세서에서 사용되는 "예측" 또는 "예단"이라는 용어는 환자가 약물 또는 일련의 약물들에 우호적으로 또는 비우호적으로 반응할지의 공산을 말한다. 한 가지 실시 상태에 있어서, 상기 예측은 환자가 치료 후, 예컨대 특정 치료제를 이용한 치료 후 질병 재발 없이 특정 기간 동안 생존 또는 개선될 것인지와 및/또는 그 확률에 관한 것이다. 본 발명의 예측 방법은 임의의 특정 환자를 위한 가장 적절한 치료 양상을 선택함으로써 치료 결정을 내리기 위하여 임상적으로 사용될 수 있다. 본 발명의 예측 방법은, 환자가 치료법, 예컨대 주어진 치료적 방법, 예컨대 주어진 치료제 또는 배합물의 투여, 외과적 개입, 스테로이드 치료 등에 우호적으로 반응할 것 같은지를 예측하는 가치있는 수단이다.

[0047] 본 명세서에서 사용되는 "특이적으로 결합하다"라는 용어는 특이적인 결합 쌍의 결합 특이성을 말한다. 다른 잠재적 표적의 존재 하에서 특정 타겟에 대한 항체의 인식이 이러한 결합의 한 특질이다. 특이적 결합은 2개의 상이한 문자와 관련이 있고, 여기서 상기 문자들 중 하나는 화학적 또는 물리적 수단을 통하여 제2의 문자와 특이적으로 결합한다. 이들 2개의 문자는 그들의 서로에 대한 결합이, 유사한 특징을 갖는 다른 시험 구성요소로부터 그들이 그들의 결합 파트너를 분간하는 것을 가능케 하도록 한다는 점에서 관련되어 있다. 결합 요소 쌍의 구성원은 리간드와 수용체 (항-리간드), 특이적 결합 쌍 (specific binding pair, SBP) 구성원 및 SBP 파트너 등등으로 일컬어진다. 또한 한 문자는 문자들의 응집에 대한 SBP 구성원일 수 있는데; 예컨대 제2 항체와 그 대응 항원으로 이루어진 면역 복합체에 대하여 발생한 항체는 상기 면역 복합체에 대한 SBP 구성원으로 간주될 수 있다.

[0048] 본 명세서에서 사용되는 "상동체 (homologue)"라는 용어는 자연적으로 존재하는 핵산 (즉, "프로토타입" 또는 "야생형" 핵산)에 경미한 개질을 가함으로써 자연적으로 존재하는 핵산과는 다르지만, 자연적으로 존재하는 형의 기본적인 뉴클레오티드 구조는 유지하고 있는 핵산을 말한다. 이러한 변화는 하나 또는 몇몇 뉴클레오티드의 변화, 예컨대 결실 (예컨대 핵산의 잘린 버전), 삽입 및/또는 치환과 같은 변화를 포함하지만 이에 한정되는 것은 아니다. 상동체는 자연적으로 존재하는 핵산과 비교하여 향상된, 감소된 또는 실질적으로 유사한 특성을 가질 수 있다. 상동체는 자연적으로 존재하는 핵산과 상보적 또는 매치될 수 있다. 상동체는 핵산 제조에 관한 이 기술 분야에 공지된 기법들을 사용하여 제조될 수 있고, 예컨대 재조합 DNA 기법, 화학적 합성 등을 들 수 있으나 이에 한정되는 것은 아니다.

[0049] 본 명세서에서 사용되는 "상보적 또는 매치되는"은 두 핵산 서열이 50% 이상의 서열 동일성을 가짐을 의미한다. 좋기로는 상기 두 핵산 서열은 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% 이상의 서열 동일성 또는 100% 서열 동일성을 갖는다. "상보적 또는 매치되는"은 또한 두 핵산 서열이 낮은, 중등의 및/또는 높은 스트린젠키 조건 하에서 혼성화할 수 있음을 의미한다.

[0050] 본 명세서에서 사용되는 "실질적으로 상보적이거나 실질적으로 매치되는"은 두 핵산 서열이 90% 이상의 서열 동일성을 가짐을 의미한다. 좋기로는 상기 두 핵산 서열은 95%, 96%, 97%, 98%, 99% 이상의 서열 동일성 또는 100%의 서열 동일성을 갖는다. 또는, "실질적으로 상보적인 또는 실질적으로 매치되는"은 두 핵산 서열이 높은 스트린젠키 조건 하에서 혼성화할 수 있음을 의미한다.

[0046] 일반적으로, 하이브리드의 안정성은 이온 농도 및 온도의 기능이다. 통상적으로 혼성화 반응은 낮은 스트린젠키 조건하에서 이루어지고, 다양한 그러나 높은 스트린젠키 하의 수세가 뒤따른다. 중등의 스트린젠키 혼성화는 프로브와 같은 핵산 문자가 상보적인 핵산 문자에 결합하게 하는 조건을 말한다. 혼성화된 핵산 문자는

일반적으로 60% 이상의 동일성을 갖고, 예컨대 70%, 75%, 80%, 85%, 90%, 또는 95%의 동일성을 갖는 임의의 것을 말한다. 중등의 스트린젠틱 조건은 42°C에서 50% 포름아마이드, 5x Denhardt 용액, 5x SSPE, 0.2% SDS로 혼성화, 이후 42°C에서 0.2x SSPE, 0.2% SDS에서 수세하는 것과 균등한 조건이다. 높은 스트린젠틱 조건은, 예컨대 42°C에서 50% 포름아마이드, 5x Denhardt 용액, 5x SSPE, 0.2% SDS에서의 혼성화, 이후 65°C에서 0.1x SSPE, 및 0.1% SDS에서 수세하는 것으로 제공될 수 있다.

[0052] 낮은 스트린젠틱 혼성화는 22°C에서 10% 포름아마이드, 5x Denhardt 용액, 6x SSPE, 0.2% SDS에서의 혼성화, 이후 37°C에서 1XSSPE, 0.2% SDS에서의 수세에 의하는 것과 균등한 조건을 말한다. Denhardt 용액은 1% Ficoll, 1% 폴리비닐파롤리돈, 및 1% 보바인 세럼 알부민 (BSA)을 함유한다. 20x SSPE (염화나트륨, 인산나트륨, 에틸렌 디아마이드 테트라아세트산 (EDTA))은 3 M 염화 나트륨, 0.2 M 인산나트륨 및 0.025 M 에틸렌 디아마이드 테트라아세트산 (EDTA)을 함유한다. 기타 적절한 중등의 스트린젠틱 및 높은 스트린젠틱 혼성화 완충액 및 조건은 이 기술 분야의 숙련자에게 잘 알려져 있다.

[0053] [0048] 본 명세서에 기술된 본 발명의 측면들 및 실시 상태들은 측면들 및 실시 상태들"로 이루어지는" 및/또는 "필수적으로 ~로 이루어지는" 것을 포함한다.

[0054] [0049] 본 발명의 기타 목적, 이점 및 특징은 동반되는 도면과 관계하여 설명되는 하기의 명세서로부터 뚜렷해질 것이다.

C. 게놈 DNA 유전자형 분석 방법

[0056] [0050] 본 발명은 체액, 조직, 혈액 또는 혈액 성분, 예컨대 혈장과 같은 다양한 제공원으로부터의 보관된 임상적 샘플을 이용하여 게놈약학적 바이오마커를 식별하는 신규한 방법을 제공한다. 일 측면에 있어서, 본 발명은 하나 이상의 게놈약학적 바이오마커를 식별하는 방법을 제공하는데, 이 방법은 a) 연관된 표현형에 있어서 상이한 값을 나타내는 2 이상의 환자의 보관된 임상 샘플로부터 DNA를 단리하는 단계; b) 상기 단리된 DNA를 증폭시키는 단계; c) 상기 증폭된 DNA의 고밀도 유전자형 분석 결과를 얻는 단계; 및 d) 상기 유전자형 분석 결과 및 상기 연관된 표현형의 상이한 값에 기초하여 관련 분석을 수행하는 단계를 포함하고, 상기 게놈약학적 바이오마커(들)이 식별되는 방법이다.

[0057] [0051] 몇 가지 실시 상태에 있어서, 상기 방법은 종전에 시행된 임상 시험으로부터의 보관된 임상 샘플에 대한 소급 연구용으로 사용될 수 있다. 몇 가지 실시 상태에 있어서, 상기 방법은 게놈약학적 바이오마커의 신생 식별을 위하여 사용될 수 있다.

[0058] [0052] 하나 이상의 게놈약학적 바이오마커를 식별하는 방법을 위하여, 연관된 표현형에 있어서 상이한 값을 나타내는 2 이상의 환자가 필요하다. 통상적으로, 관련성 연구를 수행하기 위하여 상당히 많은 수의 환자가 필요할 수 있다. 몇 가지 실시 상태에 있어서, 상기 보관된 임상적 샘플은 약 2, 5, 10, 20, 50, 100, 200, 500, 1,000명 또는 그 이상의 환자일 수 있다. 의약 치료에 대한 반응성과 같은 임의의 연관 표현형이 본 발명에 의하여 고려된다. 몇 가지 실시 상태에 있어서, 상기 연관 표현형은 범주적 특성, 정량적 특성 또는 또 다른 연관 표현형일 수 있다. 통상적으로, 상기 환자들은 임상 시험 명부에 있고 그들의 표현형 데이터가 이용 가능할 수 있다. 또는, 다수의 임상 시험에 올라있는 환자들은 관련성 연구를 위한 풀 (pool)을 구성할 수 있다. 좋기로는, 상기 다수의 임상 시험은 유사한 의약 치료에 관한 것, 예컨대 동일한 치료제에 관한 것이고, 상기 연관 표현형에 관한 데이터는 모든 임상 시험에 있어서 이용 가능하다.

샘플 준비

[0060] [0053] 샘플 준비를 위하여, 포유류 (통상적으로 인간 환자)로부터의 조직 또는 세포 샘플이 사용될 수 있다. 샘플의 예로는 조직 생검물, 혈액, 폐 흡인물, 가래, 림프액 등을 들 수 있으나 이에 한정되는 것은 아니다. 샘플은 이 기술 분야에 알려져 있는 다양한 방법에 의하여 얻을 수 있고, 예컨대 외과적 절제, 흡인 또는 생검을 들 수 있으나 이에 한정되는 것은 아니다. 상기 샘플은 신선한 것 또는 동결된 것일 수 있고, 예컨대 보관된 임상 샘플일 수 있다. 몇 가지 실시 상태에 있어서, 상기 보관된 임상 샘플은 혈장 샘플, 혈청 샘플, 건조 혈액 스팟, 소변 샘플, 조직 샘플, 종양 샘플 및 구강 면봉 표본으로 이루어지는 군으로부터 선택될 수 있다. 몇 가지 실시 상태에 있어서, 상기 보관된 임상 샘플은 혈장 샘플일 수 있다. 몇 가지 실시 상태에 있어서, 상기 샘플은 파라핀 또는 그 유사한 것에 고정되고 내장될 수 있다.

[0061] [0054] 보관된 혈장 샘플이 본 발명을 설명하기 위한 실시예로서 사용된다. 환자 또는 건강한 자원자들로부터 수집된 보관된 혈장 샘플은 임의의 적절한 방법을 이용하여, 예컨대 QIAGEN QIAamp MinElute Virus Spin Kit (Valencia, CA)를 이용하여 DNA를 추출하는 데 사용될 수 있다. 이 키트는 일부 변경을 가하여 사용될 수도 있

다. 예를 들어, 1 ml의 혈장을 간단히 불택성하고 30 μg tRNA와 완전히 혼합한다. 이 혼합물을 200 μl 엘리퀴트로 나누어 용해 완충액을 첨가하기 전 1 시간 동안 반응시킨다. 그 후, 용해물을 96°C에서 5 분간 끓이고 각각의 엘리퀴트를 동일한 컬럼을 통하여 여과한다. DNA는 멸균수에 녹인 진공 건조된 10 mM Tris-HCl (pH 8.5)에서 용리된다. 대부분의 경우, 혈장으로부터 추출된 게놈 DNA의 양은 너무 적어서 후속하는 유전자형 분석용으로는 부적당하다. 그러므로, 몇 가지 실시 상태에 있어서, 상기 단리된 DNA는 차선의 게놈 DNA일 수 있다. 후속하는 유전자형 분석용으로 충분한 양의 DNA를 얻기 위하여 상기 단리된 DNA의 증폭이 필요할 수 있다. 몇 가지 실시 상태에 있어서, 상기 증폭은 WGA일 수 있고, 그 결과물 DNA는 wgaDNA이다. 예컨대, DNA 샘플은 Amersham Bioscience GenomiPhi DNA Amplification Kit (Piscataway, NJ) 또는 균등한 제제를 이용하여 증폭될 수 있고, 이 방법은 통상적으로 유전자형 분석에 충분한, 수 마이크로그램의 DNA를 생산한다.

[0062] SNP를 이용한 유전자형 분석

[0055] 상기 단리된 게놈 DNA로부터 유전자형 분석 결과를 얻기 위하여 임의의 적절한 방법들이 사용될 수 있다. 유전자형 분석 방법은 하나 이상의 다형좌 (polymorphic loci)에서 동시에, 1인 이상의 개체에 대한 정보를 얻을 수 있다. 몇 가지 실시 상태에 있어서, 유전자형 분석은 주어진 유전자좌에서 단일 뉴클레오티드 분해능 (single nucleotide resolution)으로 대립 유전자를 분별하는 것으로 정의될 수 있다. 유전자좌는 유전자 또는 DNA 마커의 염색체 위치로서 정의된다. 그러므로, 본 발명에 따른 방법은 의약 결정의 기초로 사용될 수 있는 개개인에 대한 스크리닝 및 진단 정보를 제공하기 위하여 필요한 정밀성을 갖는다.

[0056] 본 명세서에서 사용되는 "유전자형 분석되다"라는 용어는 1 이상의 개체의 유전자형을 결정하는 방법을 말하며, "유전자형"은 군 (population) 내의 하나 이상의 다형 변이를 나타낸다. 통상적으로, 유전자형 분석은 하나 이상의 다형좌에서 다형 변이의 존재 또는 부재를 평가하는 것을 수반한다. 몇 가지 실시 상태에 있어서, 고밀도 유전자형 분석이 사용될 수 있다. 몇 가지 실시 상태에 있어서, 상기 고밀도 유전자형 분석은 전체 게놈 유전자형 분석이다.

[0057] 본 명세서에서 사용되는 "다형좌"라는 용어는 개체군의 핵산 샘플 중 상당수에서 2 이상의 택일적 뉴클레오티드 서열이 관찰되는 핵산의 부위를 말한다. 다형좌는, 예컨대 2 이상의 뉴클레오티드로 이루어지는 뉴클레오티드 서열, 삽입된 뉴클레오티드 또는 뉴클레오티드 서열, 결실된 뉴클레오티드 또는 뉴클레오티드 서열, 또는 마이크로세틸라이트의 카페수의 변화일 수 있다. 2 이상의 뉴클레오티드 길이의 다형좌는 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15개 또는 그 이상, 20개 또는 그 이상, 30개 또는 그 이상, 50개 또는 그 이상, 75개 또는 그 이상, 100개 또는 그 이상, 500개 또는 그 이상, 또는 약 1000개 뉴클레오티드 길이일 수 있고, 여기서 상기 뉴클레오티드 서열들의 전부 또는 일부는 부위 내에서 상이하다. 다형좌는 종종 1개 뉴클레오티드 길이이고, 이것은 본 명세서에서 "단일 뉴클레오티드 다형성" 또는 "SNP"로 일컬어진다. 몇 가지 실시 상태에 있어서, 고밀도 유전자형 분석은 SNP를 이용하여 시행될 수 있다. 몇 가지 실시 상태에 있어서, 약 1,000~5,000,000개 또는 그 이상, 종기로는 약 1,000,000개의 SNP가 사용될 수 있다. 몇 가지 실시 상태에 있어서, 고밀도 유전자형 분석은 어레이 기반일 수 있다. 몇 가지 실시 상태에 있어서, 상기 고밀도 유전자형 분석은 시퀀싱, 예컨대 하이-쓰루풋 시퀀싱 (high-throughput sequencing)에 의하여 시행될 수 있다.

[0058] 다형좌에 2개, 3개, 또는 4개의 택일적인 뉴클레오티드 서열이 있는 경우, 각각의 뉴클레오티드 서열은 "다형 변이체" 또는 "핵산 변이체"라고 일컬어진다. 예컨대, 2개의 다형 변이체들이 존재하는 경우, 군 중 소수의 샘플로 대표되는 다형 변이체는 때때로 "소수 대립 유전자"로 일컬어지고, 더 우세하게 대표되는 다형 변이체는 때때로 "주 대립 유전자"로 일컬어진다. 많은 생물들은 각각의 염색체를 1 카페를 갖고 (예컨대, 인간), 2 개의 주 대립 유전자 또는 2개의 소수 대립 유전자를 갖는 개체들은 그 다형성과 관련하여 종종 "동형접합"으로 일컬어지고, 하나의 주 대립 유전자와 하나의 소수 대립 유전자를 갖는 개체들은 그 다형성과 관련하여 일반적으로 "이형접합"이라고 일컬어진다. 하나의 대립 유전자에 관하여 동형접합은 개체들은 때때로 이형접합 또는 다른 대립 유전자에 관하여 동형접합인 개체와 비교하여 상이한 표현형인 경향이 있다.

[0059] 하나 이상의 게놈약학적 바이오마커를 식별하는 유전적 분석에 있어서, 관련 표현형인 경향에 있어서 상이한 값을 갖는 개체들로부터의 샘플은 종종 대립 유전자형 분석되고 (allelotyped) 및/또는 유전자형 분석된다. 본 명세서에서 사용되는 "대립 유전자형 분석"이라는 용어는 케이스 및 대조군으로부터 모든 DNA 샘플 중의 다형 변이체에 대하여 대립 유전자 빈도를 결정하는 방법을 말한다. 각 그룹으로부터 DNA를 모음으로써, 각각의 그룹 내 각각의 좌에 대한 대립 유전자 빈도가 계산된다. 이들 대립 유전자 빈도는 그 후 다른 것과 비교된다.

[0060] 유전자형 또는 다형 변이체는 "하플로타입"이라는 용어로 표현될 수 있으며, 본 명세서에서 사용되는 바 이것은 함께 유전되는 경향이 있는 DNA 변이 또는 다형성의 한 세트를 말한다. 하플로타입은 동일한 염색체 상

에서 발견되는 대립 유전자 또는 한 세트의 SNP의 조합을 말할 수 있다. 예컨대, 각 SNP 위치가 시토신 변이 및 아데닌 변이를 포함하는 2개의 SNP가 하나의 유전자 내에 존재할 수 있다. 집단 내 특정 개체는 각 SNP 위치에서 시토신을 함유하는 유전자를 갖는 하나의 대립 유전자 (이형접합) 또는 2개의 대립 유전자 (동형접합)을 포함할 수 있다. 상기 유전자 내의 각 SNP에 해당하는 2개의 시토신은 이들 개체들에서 하나의 대립 유전자 또는 두 대립 유전자 양자 모두에서 함께 이동하기 때문에 상기 개체들은 상기 유전자 내 2개의 SNP에 관하여 시토신/시토신 하플로타입을 갖는 것으로 특징지워질 수 있다.

[0069] 때때로, 연구자들은 그 변이체가 한 집단에서 유의미한 분율로 나타나는지를 결정함이 없이 다형 변이체를 데이터베이스에 보고한다. 이들 보고된 다형 변이체들의 하위집합은 그 집단에서 통계적으로 유의미한 분율로 나타나는 것이 아니기 때문에, 이들 중 일부는 시くん성 에러 및/또는 생물학적으로 무연관이다. 그러므로, 보고된 다형 변이체가 통계적으로 유의미하거나 생물학적으로 연관이 있는 것인지는 상기 변이체의 존재가 개체군에서 검출되고 그 변이체의 빈도가 결정되기까지 종종 알려지지 않는다. 다형 변이체는 집단 중 1% 또는 그 이상, 때로는 5% 또는 그 이상, 10% 또는 그 이상, 15% 또는 그 이상, 20% 또는 그 이상, 25% 또는 그 이상, 30% 또는 그 이상, 35% 또는 그 이상, 40% 또는 그 이상, 45% 또는 그 이상, 50% 또는 그 이상으로 나타난다면 통계적으로 유의미하고 종종 생물학적으로 연관인 것이다.

[0070] [0062] 다형 변이체는 이중 가닥 핵산의 어느 하나의 가달 또는 양 가달 모두에서 검출될 수 있다. 또한, 다형 변이체는 유전자의 인트론 또는 엑손 내에 위치하거나 조절 부위 부분 내에, 예컨대 프로모터, 5' 미번역 부위 (UTR), 3' UTR, 번역 부위, 유전자간 부위, 또는 미공지된 유전자를 함유하는 게놈 부위에 위치할 수 있다. 다형 변이는 유전자 발현, 폴리펩타이드 구조 또는 폴리펩타이드 기능에 있어서 검출 가능한 차이점이라는 결과를 낳거나 낳지 않을 수도 있다.

파라미터 최적화

[0072] 보관된 샘플로부터 얻어진 DNA로부터 도출된 유전자형 결과들은 상이한 특성을 나타내고, 이것은 표준 실시에서 사용되는 고품질의 DNA로부터의 결과보다 현저히 낮은 호출률을 나타낼 수도 있다. 통상적으로 95% 또는 심지어 그 이상인 표준 GWAS에서 사용되는 통상적인 호출률 컷 오프를 적용하는 것과는 반대로, 본 발명에서 사용되는 분석은 분석에 더 많은 샘플들을 포함시키기 위하여 얻어진 유전자형 결과들에 기초하여 컷 오프 호출률을 조정하는 비관행적인 접근을 채용한다. 그 결과로, 사용되는 컷 오프 값은 판매자에 의하여 권장되는 표준 컷 오프 값보다 실질적으로 낮을 수 있다. 본 발명으로부터 생성되는 결과의 또 다른 연관 특징은 많은 양의 데이터가 손실될 수 있고, 이것은 표준 또는 "공지의" 분석 알고리즘을 이용하는 경우에 큰 도전일 수 있다. 따라서, 본 발명은 유전자형 분석된 다형좌 중에서 연관비평형 (linkage disequilibrium, LD)에 기초하여 손실 데이터를 대체할 수 있는 대치 알고리즘 역시 사용한다.

[0073] [0064] 유전자형 분석 데이터를 얻기 위하여 포함 기준의 최적화가 이루어질 수 있다 (도 2). 몇 가지 실시 상태에 있어서, 상기 유전자형 분석 데이터는 전장 게놈 유전자형 호출 알고리즘을 사용함으로써 얻는다. 혈장 샘플로부터 생성된 차선의 게놈 DNA의 품질 및 유전자형 분석 이전에 적용되는 전체 게놈 증폭 때문에, 일부 샘플들의 호출률은 고품질의 게놈 DNA를 사용하는 경우의 통상적인 호출률 (>95%)보다 실질적으로 낮을 수 있다. 본 발명에 있어서, 컷-오프 호출률은 가능한 많은 샘플들을 포함할 수 있도록 종래 표준보다 현저히 낮게 조정된다. 몇 가지 실시 상태에 있어서, 상기 전장 게놈 유전자형 호출 알고리즘의 호출률 컷-오프 값은 조정될 수 있다. 몇 가지 실시 상태에 있어서, 유전자형 호출은 고품질 게놈 DNA의 전체 게놈 유전자형 분석에 사용되는 통상적인 호출률 컷-오프 값보다 낮은 호출률 컷-오프 값을 이용함으로써 이루어지고, 여기서 상기 사용되는 호출률 컷-오프 값은 약 50~95%, 약 80~90% 또는 약 90%이다. 몇 가지 실시 상태에 있어서, 유전자형 호출 알고리즘의 호출률 컷-오프 값의 조정 및 조정된 호출률 컷-오프 값에 기초한 유전자형 호출은 수회 반복되어 샘플을 포함 및/또는 배제하기 위한 기준을 식별할 수 있다. 몇 가지 실시 상태에 있어서, 최적의 포함 기준, 예컨대 호출률 컷-오프 값은 조정 사이클의 반복 이후 식별된다. 몇 가지 실시 상태에 있어서, 상기 유전자형 호출은 대치 알고리즘을 이용하여 이루어지고, 여기서 HapMap이 상기 대치 알고리즘으로 사용된다.

[0074] [0065] 몇 가지 실시 상태에 있어서, DNA 샘플들은 Affymetrix (Santa Clara, CA) 및/또는 Illumina (San Diego, CA)에 의하여 제조된 전체 게놈 SNP 어레이를 이용하여 유전자형 분석된다. Affymetrix 500K 어레이가 본 발명을 설명하기 위한 실시예로서 사용된다. Affymetrix 어레이 뿐 아니라, Illumina 칩 및 Sequenom MassArray 역시 하나의 플랫폼에 의하여 생성된 결과를 확인하기 위하여 사용된다.

[0066] 몇 가지 실시 상태에 있어서, 유전자형 호출은 Affymetrix Genotyping Console™ 소프트웨어로 생성된다. 몇 가지 실시 상태에 있어서, 유전자형 호출은 Mahalanobis Distance Classifier의 Robust Linear

Model (RLMM) 알고리즘, Bayesian step의 RLMM (BRLMM) 알고리즘, Axiom™ GT1 알고리즘, 완벽 매치 프로브 이용 BRLMM (BRLMM-P) 알고리즘, 또는 Birdseed 알고리즘 (Rabbee et al., Bioinformatics (2006) 22:7-12; Korn et al., Nat Genet (2008) 40:1253-60)을 이용하여 이루어진다.

[0076] D. 관련성 분석에 의한 게놈약학적 바이오마커의 식별

[0067] 얻어진 유전자형 분석 데이터는 게놈약학적 바이오마커를 식별하기 위하여 연관 표현형에 기초한 관련성 분석을 수행하기 위하여 사용된다 (도 1). 서포트 벡터 머신 (SVM) 및 로지스틱 회귀분석을 포함하지만 이에 한정되지는 않는 분류 알고리즘이 데이터세트에 적용되고 최적의 바이오마커 및 점수화 알고리즘을 식별할 수 있다. 몇 가지 실시 상태에 있어서, 상기 관련성 분석은 GWAS이다. 몇 가지 실시 상태에 있어서, 상기 관련성 분석은 각각의 다형좌, 예컨대 SNP의 관련 p-값을 연관 표현형과 함께 계산함으로써 이루어진다. 몇 가지 실시 상태에 있어서, 상기 계산은 대립 유전자 빈도 및/또는 유전자형 기반 테스트로 기초한다.

[0068] 게놈약학적 바이오마커 식별을 위한 연관 표현형은 일반적으로 치료법에 대한 개체들의 반응성과 관련된다. 연관 표현형은 정성적이거나 정량적일 수 있다. 반응성은 1차적이거나, 예컨대 항암 약물에 대한 반응으로 암 매쓰가 감소한다거나, 또는 2차적일 수 있는데, 예컨대 백사로틴에 대한 반응으로 중성지방고혈증 (HTC)이 일어난다거나 할 수 있다. 상기 1차적 및 2차적 반응은 서로 관련된 것일 수도 아닐 수도 있다. 반응은 양성적 이거나 음성적일 수 있다. 음성적 반응은 효과적인 반응의 부재 또는 유독한 부작용의 존재 둘 중의 하나로 정의될 수 있다. 하나 이상의 연관 표현형은 게놈약학적 바이오마커를 식별하기 위한 관련성 분석에 사용될 수 있다.

[0069] 몇 가지 실시 상태에 있어서, 상기 관련성 분석은 상이한 호출률 컷-오프 값을 사용하여 수회 반복되어 사용되는 기준을 최적화할 수 있다 (도 2). 몇 가지 실시 상태에 있어서, 유전자형 결과는 PLINK와 같은 유전자 데이터 분석 소프트웨어를 이용하여 분석되어 (Purcell et al., *Am J Hum Genet* (2007) 81:559-5752) 연관 표현형의 각각의 다형좌의 관련 p-값을 계산할 수 있다. 다형좌는 그 연관 표현형과의 계산된 관련성에 의하여 정렬될 수 있다. 가장 유의미하게 관련된 다형좌가 그 연관 표현형에 대한 게놈약학적 바이오마커로서 식별될 수 있다.

[0080] 검증 및/또는 재현

[0081] 고밀도 유전자형 분석으로 식별된 게놈약학적 바이오마커는 추가로 관련성 분석, 또는 "단계 II" 분석 진행될 수 있다 (도 1). 추가의 관련성 분석은 고밀도 유전자형 분석을 통하여 식별된 게놈약학적 바이오마커의 검증 및/또는 재현 및 관련성 분석, 또는 단계 I 분석을 위하여 사용될 수 있다. 단계 I의 보관된 임상 샘플의 일부 또는 전부 및/또는 추가적인 임상 샘플들은 추가적인 유전자형 분석용으로 사용될 수 있다.

[0082] 몇 가지 실시 상태에 있어서, 추가적인 관련성 분석은 식별된 게놈약학적 바이오마커를 이용한 추가적인 유전자형 분석 데이터에 기초할 수 있다. 몇 가지 실시 상태에 있어서, 약 1, 10, 20, 50, 100, 200, 500개 또는 그 이상의 식별된 게놈약학적 바이오마커가 추가적인 유전자형 분석을 위하여 사용될 수 있다.

[0083] 몇 가지 실시 상태에 있어서, 단계 I 분석으로부터 식별된 고도로 관련된 SNP는 단계 I에서 사용된 것들과는 차별화될 수 있는 저밀도 유전자형 분석 플랫폼, 예컨대 Sequenom iPLEX MassArray 기법을 이용하여 재현될 수 있다.

[0084] 단계 II에서, 단계 I에서 사용되지 않았던 새로운 DNA 샘플들이 단계 I에서 식별된 게놈약학적 바이오마커를 재현할 목적으로 유전자형 분석 될 수 있다. 추가적인 임상 샘플들은 새로운 임상 시험 환자들로부터 유래할 수 있고 신선한 임상 샘플들일 수 있다. 그러므로, 추가적인 임상 샘플들로부터 단리된 게놈 DNA는 중폭없이 고품질이고 다양이며, 고밀도 유전자형 분석에 적합한 것일 수 있다.

[0085] 포함 기준의 최적화는 추가적인 유전자형 분석 데이터를 얻기 위하여 수행될 수 있다 (도 3). 몇 가지 실시 상태에 있어서, 상기 추가적인 유전자형 분석 데이터는 검증 유전자형 호출 알고리즘을 사용하여 얻어질 수 있다. 몇 가지 실시 상태에 있어서, 추가적인 관련성 분석은 검증 유전자형 호출 알고리즘의 호출률 컷-오프 값을 조정하는 단계를 포함할 수 있다. 몇 가지 실시 상태에 있어서, 유전자형 분석 및 호출률 컷-오프를 조정하는 것은 수회 반복되어 샘플들을 포함 및/또는 배제시킬 수 있다. 몇 가지 실시 상태에 있어서, 최적의 포함 기준이 식별될 수 있다. 몇 가지 실시 상태에 있어서, 상기 방법은 검증 유전자형 호출 알고리즘을 이용하여 얻어진 추가적인 유전자형 분석 데이터를 전장 게놈 유전자형 호출 알고리즘을 이용하여 얻어진 유전자형 분석 데이터와 비교하는 단계를 더 포함할 수 있다 (도 3).

- [0086] [0075] 두 단계 연구 후, 가장 유의미하게 관련된 게놈약학적 바이오마커가 식별될 수 있다. 몇 가지 실시 상태에 있어서, 관련성 분석은 여러 가지 흐출률 컷-오프 값을 사용하여 수회 반복되어 사용되는 기준을 최적화할 수 있다 (도 3). 몇 가지 실시 상태에 있어서, 상기 게놈약학적 바이오마커는 단계 I에서 식별된 게놈약학적 바이오마커의 하위그룹이다. 몇 가지 실시 상태에 있어서, 상기 게놈약학적 바이오마커는 하나 이상의 SNP를 포함한다. 식별된 다수의 게놈약학적 바이오마커는 게놈상에서 서로 가까이 위치할 수 있고, 유전자의 인트론/엑손, 유전자간 부위 또는 게놈상 공지되지 않은 유전자를 포함하는 부위에 위치할 수 있다.
- [0087] E. 보관된 샘플을 이용한 전장 게놈 관련성 연구
- [0088] [0076] 또 다른 측면에 있어서, 본 발명은 차선의 게놈 DNA를 이용하여 GWAS를 시행하는 방법을 제공한다. 몇 가지 실시 상태에 있어서, 상기 차선의 게놈 DNA는 보관된 샘플로부터 유래한 것일 수 있다. 몇 가지 실시 상태에 있어서, 상기 차선의 게놈 DNA는 혈장 샘플로부터 유래한 것일 수 있다. 몇 가지 실시 상태에 있어서, 상기 차선의 게놈 DNA는 증폭된 것일 수 있다. 본 발명은 상기 차선의 게놈 DNA를 이용하여 GWAS를 시행하는 방법을 사용하여 게놈약학적 바이오마커를 식별하는 방법을 더 제공하고, 이 방법은 소급적인 방법일 수 있다.
- [0089] [0077] 전술된 2 단계 데이터 분석은 차선의 게놈 DNA를 이용하는 GWAS를 위하여 사용될 수 있다. 몇 가지 실시 상태에 있어서, 다수의 유전자형 분석 플랫폼이 사용될 수 있다. 몇 가지 실시 상태에 있어서, 이 다수의 유전자형 분석 플랫폼으로 동일하거나 상이한 샘플들이 사용될 수 있다. 몇 가지 실시 상태에 있어서, 이 방법은 차선의 게놈 DNA로부터 얻어지는 데이터를 재현하기 위하여 사용될 수 있는 고품질의 게놈 DNA를 제공하는 샘플을 더 사용할 수 있다. 몇 가지 실시 상태에 있어서, 고품질의 게놈 DNA를 제공하는 샘플들은 전혈 샘플들일 수 있다.
- [0090] F. 게놈약학적 바이오마커의 활용
- [0091] [0078] 게놈약학은 대상체의 유전자형에 따라 차등적인 효과를 발휘할 수 있는 특정한 치료법으로서 환자에 대한 치료를 대상체의 유전자형에 따라 맞춤화하는 것을 수반한다. 예컨대, 예후 테스트 결과에 기초하여 임상전문가 또는 의사는 그 정보 또는 치료에 의하여 이익을 받을 대상체에 대하여 타당한 정보와 예방 또는 치료적 치료를 표적화할 수 있고 이익이 되지 않을 대상체에게 (예컨대, 이 치료가 치료적 효과가 없고 및/또는 환자가 부작용을 경험하는) 이러한 정보 및 치료를 지시하는 것을 피할 수 있다. 본 발명에서 설명하는 방법을 이용하여 게놈약학적 바이오마커로부터 생성된 정보는 개체에 대한 적절한 투여량 및 치료법을 결정하는데 사용될 수 있다. 약물의 투여량 적용 또는 선택시, 이러한 지식은 부작용이나 치료 실패를 피할 수 있게 하고 따라서 치료적 조성물의 투여시 치료적 효과를 증대시킬 수 있다. 몇 가지 실시 상태에 있어서, 게놈약학적 바이오마커는 동반 진단 테스트를 개발하는데 사용될 수 있다.
- [0092] 그러므로, 또 다른 측면에 있어서, 본 발명은 본 발명에 개시되는 방법에 의하여 식별되는 게놈약학적 바이오마커를 이용한 동반 진단 테스트를 제공한다. 예컨대, 한 가지 실시 상태에 있어서, 의사 또는 임상전문가는 대상체에게 약학적 조성물을 투여할지를 결정할 때, 본 발명에 개시되는 방법을 사용하여 게놈약학적 바이오마커에서 얻은 지식을 적용할 것을 고려할 수 있다. 또 다른 실시 상태에 있어서, 의사 또는 임상전문가는 환자에게 투여되는 치료 투여량, 예컨대 치료당 투여량 또는 치료의 빈도 결정시 이러한 지식을 적용할 것을 고려 할 수 있다.
- [0093] [0080] 본 발명은 치료에 대한 대상체의 반응성을 평가 또는 평가를 보조하는 방법을 제공한다. 또한, 본 발명은 대상체에서 치료에 대한 반응성을 예측하거나 치료/반응성을 모니터링하는 방법을 제공한다. 본 발명은 치료 대상체를 선별하고 그 대상체를 치료하는 방법을 제공한다. 몇 가지 실시 상태에 있어서, 상기 방법은 상기 대상체로부터 얻은 샘플에서 하나 이상의 게놈약학적 바이오마커를 평가하는 단계; 및 상기 하나 이상의 게놈약학적 바이오마커의 유전자형에 기초하여 치료에 대한 상기 대상체의 반응성을 예측, 평가, 또는 평가 보조하는 단계를 포함한다. 몇 가지 실시 상태에 있어서, 상기 반응성은 SVM, 로지스틱 회귀 분석, 또는 K-최근접 이웃 분석 (K-nearest neighbors analysis)과 같은 알고리즘을 이용하여 상기 대상체를 분류함으로써 예측 또는 평가된다.
- [0094] [0081] 이하는 게놈약학적 실시 상태에 관한 예시이다. 특정한 치료법은 대상체의 유전자형에 따라 차별적인 효과를 발휘할 수 있다. 후보인 치료가 주 대립 유전자와 유의미한 상호작용을 하고, 소수 대립 유전자와는 비교적 약한 상호작용을 한다면 (예컨대, 상호작용에 있어서 한 자리수 이상의 차이), 이러한 치료는 통상적으로 소수 대립 유전자의 동형접합 유전자형의 대상체에게는 투여하지 않고, 이 소수 유전자의 이형접합 유전자형의 대상체에게도 때때로 투여하지 않을 것이다. 다른 예시에 있어서, 후보인 치료가 주 대립 유전자 동형접합인 대상

체에게 투여시에는 그다지 독성이지 않지만 이형접합이거나 소수 대립 유전자의 동형접합인 대상체에게 투여시에는 비교적 독성인 경우라면, 이 후보 치료는 통상적으로 상기 소수 대립 유전자에 관하여 이형접합 또는 동형접합인 유전자형의 대상체에게는 투여하지 않는다.

[0095] [0082] 본 발명에 개시되는 방법은 대사 질환, 심혈관질환, 암 등의 질환을 예방, 경감 또는 치료하는 계놈약학적 방법에 적용된다. 예컨대, 한 개체로부터의 혁산 샘플은 본 명세서에 개시되는 예후 테스트를 거칠 수 있다. 유형 II 당뇨의 증가 위험과 관련된 하나 이상의 다형 변이가 대상체에게서 식별된다면, 그 후 유형 II 당뇨를 예방 또는 치료하기 위한 정보 및/또는 하나 이상의 유형 /II 당뇨 치료법이 그 대상체에게 처방될 수 있다.

[0096] [0083] 특정 실시 상태에 있어서, 치료법은 본 명세서에 개시되는 방법에 의하여 평가되는, 치료법에 대한 그들의 반응 공산에 기초하여 그로부터 가장 이익을 받을 개체들에게 특이적으로 처방 및/또는 투여된다. 따라서, 치료법에 대하여 반응할 공산이 높은 대상체를 선별한 후 이러한 치료법을 반응 공산이 높은 것으로 식별되는 개체들에게 처방하는 방법이 제공된다. 따라서, 특정 실시 상태는 대상체를 치료하는 방법에 관한 것이고, 이 방법은: 대상체로부터의 혁산 샘플에서 본 발명에 설명된 뉴클레오티드 서열 중 치료법에 대한 반응성과 관련된 계놈약학적 바이오마커의 존재 또는 부재를 검출하는 단계와, 그 뉴클레오티드 서열 중에서 상기 치료법에 대한 반응성과 관련된 계놈약학적 바이오마커의 존재가 검출된 샘플이 유래한 대상체에게 상기 치료법을 처방하거나 투여하는 단계를 포함한다.

[0097] [0084] 때때로 상기 치료는 예방적 (예컨대, 질병이 발병 또는 진행할 확률을 감소시키기 위하여 처방 또는 투여)이고, 때로는 치료적이며, 때로는 질병의 진행을 지연, 경감 또는 중단시킨다. 장애의 발생을 경감 또는 예방하기 위한 임의의 알려진 예방 또는 치료적 치료가 처방 및/또는 투여될 수 있다.

[0098] [0085] 또한, 계놈약학적 방법은 약물에 대한 반응을 분석하고 예측하는 데 사용될 수 있다. 예컨대, 계놈약학적 분석이 한 개체가 특정한 약물을 이용한 치료에 대하여 양성으로 반응할 공산을 나타낸다면, 이 약물은 그 개체에게 투여될 수 있다. 반대로, 상기 분석이 개체가 특정 약물을 이용한 치료에 대하여 음성으로 반응할 것 같다고 나타낸다면, 대체적 코스의 치료가 처방될 수 있다. 치료적 치료에 대한 반응은 하기 군들 중 임의의 군의 대상체들이 유전자형 분석되는 배경 연구에서 예측될 수 있다: 치료법에 대하여 우호적으로 반응하는 군, 치료법에 대하여 크게 반응하지 않는 군 및 치료법에 대하여 부정적으로 반응하는 군 (예컨대, 하나 이상의 부작용을 나타내는 군). 이들 군들은 예시로서 제공되고 다른 군들 및 하위군들이 분석될 수 있다. 이러한 분석 결과에 기초하여, 대상체는 유전자형 분석되어 그 또는 그녀가 치료법에 대하여 우호적으로 반응할지, 치료법에 대하여 크게 반응하지 않을지, 또는 치료법에 대하여 부정적으로 반응할지를 예측하게 된다.

[0099] [0086] 분류/예측 알고리즘은 검증 및/또는 재현 데이터세트를 이용하여 개발될 수 있다. 유전자형 분석된 다형좌 중에서 LD에 기초한 결손 데이터의 일부를 대신할 수 있는 대치 알고리즘이 사용될 수 있다. SNP가 유전자형 분석에 사용된 실시 상태에 있어서, Hapmap과 같은 SNP 데이터베이스가 대치 알고리즘용으로 사용될 수 있다. 분류/예측 알고리즘의 개발을 위하여, 상기 검증 데이터세트는 훈련 데이터세트로서 사용될 수 있다. 일단 분류/예측 알고리즘이 개발되면, 재현 데이터세트가 그 알고리즘을 시험하기 위하여 사용될 수 있다.

[0100] [0087] 몇 가지 실시 상태에 있어서, 본 발명의 방법은 대상체의 샘플 및 알려진 부류의 레퍼런스 샘플들의 계놈약학적 바이오마커의 유전자형에 기초한 K-최근접 이웃 분석을 이용하여 상기 대상체를 반응성 또는 비반응성 대상체로 분류하는 단계를 포함한다. 몇 가지 실시 상태에 있어서, K-최근접 이웃 분석을 이용한 대상체 분류는 (1) 파라미터 K (즉, 최근접 이웃의 수)를 결정; (2) 분류될 새로운 샘플 중의 마커 유전자들의 측정된 발현 수준과, 각각의 레퍼런스 샘플 중 각각의 마커 유전자들의 발현 수준간 차이를 계산; (3) 새로운 샘플과 레퍼런스 샘플간 절대적 차이에 대한 가중 평균 (weighted average of the absolute differences, WAAD)이 가장 작은 샘플들을 선별함으로써 가장 최근접한 레퍼런스 샘플들을 결정; 및 (4) K 최근접 레퍼런스 샘플의 알려진 부류에 기초하여 새로운 샘플의 부류를 결정함으로써 수행된다. 중량 및/또는 파라미터 K는 알려진 부류의 임상적 시험 샘플을 이용한 교차 유효성 검사를 이용하여 결정된다. 예컨대, 5배 (예컨대, 5배, 6배, 7배, 8배, 9배 또는 10배) 내지 N배의 교차 유효성 검사가 가중 K-최근접 이웃 분류 에러를 최소화하기 위하여 사용될 수 있고, 여기서 N은 샘플의 크기이다. 몇 가지 실시 상태에 있어서, K는 4 내지 13 (예컨대, 4, 5, 6, 7, 8, 9, 10, 11, 12, 및 13) 사이의 정수이다. 몇 가지 실시 상태에 있어서, 상기 최근접 레퍼런스 샘플들 (최근접 이웃)은 분류될 새로운 샘플의 발현 수준과 각각의 계놈약학적 바이오마커에 대한 각각의 레퍼런스 샘플의 발현 수준 사이의 절대적 차이의 가중 평균이 가장 작은 것들이다.

[0101] [0088] 예측, 평가 또는 평가 보조를 위한 비교 및/또는 계산은 문제의 계놈약학적 바이오마커에 대하여 측정되는 값 및/또는 레퍼런스 값의 유형에 적절한 임의의 편리한 방식으로 수행될 수 있다. 비교 또는 계산의 방법은

수동일 수 있고 또는 자동 (예컨대, 컴퓨터 기반 기계를 포함하는 기계에 의한 것)일 수 있다. 이 기술 분야의 숙련자에게 자명할 바와 같이, 유전자형 분석 재현은 게놈약학적 바이오마커에 대하여 이루어질 수 있다.

[0102] 또한, 본 발명은 본 발명에 개시된 동반 진단 테스트를 사용하여 치료에 대한 대상체의 반응성을 예단하는 방법을 제공한다. 본 발명에서 개시되는 테스트는 임상적 약물 시험에도 역시 적용될 수 있다. 몇 가지 실시 상태에 있어서, 게놈약학적 바이오마커는 임상 시험용 대상체 군을 계층화하거나 선별하는 데 사용될 수 있다. 몇 가지 실시 상태에 있어서, 게놈약학적 바이오마커는 치료에 대하여 독성의 반응을 나타낼 수 있는 개체들을 그렇지 않을 개체들로부터 계층화하는 데 사용될 수 있다. 다른 실시 상태에 있어서, 게놈약학적 바이오마커는 비반응일 개체들을 반응성일 개체들로부터 분리해내는 데 사용될 수 있다. 본 발명에서 개시되는 게놈약학적 바이오마커는 게놈약학계 설계 및 임상 시험의 시행 관리에 사용될 수 있다.

[0103] 치료 제제에 대한 반응 또는 치료제에 대한 부작용의 표지자인 하나 이상의 게놈약학적 바이오마커는 본 발명에 개시되는 방법을 이용하여 식별될 수 있다. 이후, 이러한 제제에 대한 임상 시험의 잠재적 참여자들은 스크리닝되어 상기 약물에 대하여 가장 우호적으로 반응할 것 같은 개체들을 식별하고 부작용을 일으킬 것 같은 개체들을 배제할 수 있다. 이러한 방식으로, 연구에 있어서 양성으로 반응할 것 같지 않은 개체들을 포함한 결과로 측정치를 낮추지 않고 원치않는 안전 문제의 위험없이 약물에 대하여 양성으로 반응하는 개체들 중에서 약물 치료의 효율이 측정될 수 있다.

[0091] 따라서, 또 하나의 실시 상태는 치료 또는 약물의 임상 시험에 포함될 개체를 선택하는 방법으로서, (a) 개체로부터 핵산 샘플을 얻는 단계; (b) 상기 핵산 샘플에서 상기 치료 또는 약물에 대하여 양성의 반응과 관련된 다형 변이를 확인, 또는 상기 치료 또는 약물에 대하여 음성의 반응과 관련된 하나 이상의 다형 변이를 확인 결정하는 단계; 및 (c) 상기 핵산 샘플이 상기 치료 또는 약물에 대한 양성의 반응과 관련된 상기 다형 변이를 함유하거나, 상기 핵산 샘플이 상기 치료 또는 약물에 대한 음성의 반응과 관련된 상기 다형 변이를 결손하고 있다면 그 개체를 상기 임상 시험에 포함시키는 단계를 포함하는 방법이다. 또한, 본 발명에 개시된, 치료 또는 약물의 임상 시험에 포함될 개체를 선택하는 방법은 본 명세서에서 개시된 임의의 추가적인 한정 사항을 갖는 방법, 또는 하기의 단독으로 또는 임의의 조합으로 구체화되는 방법들을 포함한다. 단계 (c)는 필요에 따라, 상기 핵산 샘플이 상기 치료 또는 약물에 대한 양성의 반응과 관련된 상기 다형 변이를 함유하고, 상기 핵산 샘플이 상기 치료 또는 약물에 대한 음성의 반응과 관련된 상기 2-대립 유전자형 (biallelic) 마커를 결손하고 있다면 그 개체에 상기 약물을 투여하거나 치료하는 단계를 포함한다.

G. 추가의 게놈약학적 바이오마커 또는 약물 표적

또한, 식별된 게놈약학적 바이오마커에 가까운 (proximal) 다형 변이를 식별하는 방법이 제공된다. 따라서, 본 발명에서 특징적인 것은 식별된 게놈약학적 바이오마커에 가까운 다형 변이를 식별하는 방법이다. 또 다른 실시 상태에 있어서, 상기 식별된 가까운 다형 변이체는 때때로 공개적으로 개시된 다형 변이체이고, 예컨대 이것은 때때로 공중에 이용가능한 데이터베이스에 공개된 것이다. 다른 실시 상태에 있어서, 상기 식별된 다형 변이체는 공개적으로 개시되지는 않고 공지의 방법을 이용하여 발견된 것, 예컨대 핵산 샘플들의 그룹에서 식별된 게놈약학적 바이오마커를 둘러싼 부위를 시퀀싱하는 방법을 이용하여 발견된 것이지만 이에 한정되는 것은 아니다. 따라서, 식별된 게놈약학적 바이오마커와 가까운 다수의 다형 변이체들은 이 방법을 이용하여 식별된다.

[0093] 상기 가까운 다형 변이체는 종종 상기 식별된 게놈약학적 바이오마커를 둘러싼 부위에서 확인된다. 특정한 실시 상태에 있어서, 이러한 둘러싼 부위는 상기 식별된 게놈약학적 바이오마커 측면의 약 50 kb (예컨대, 제1 다형 변이체의 5' 약 50 kb 및 상기 제1 다형 변이체의 3' 약 50 kb)이고, 이 부위는 때로 더 짧은 측면 서열로 구성되는데, 예컨대 상기 식별된 게놈약학적 바이오마커의 5' 및 3' 약 40 kb, 약 30 kb, 약 25 kb, 약 20 kb, 약 15 kb, 약 10 kb, 약 7 kb, 약 5 kb, 또는 약 2 kb이다. 다른 실시 상태에 있어서, 이 부위는 더 긴 측면 서열로 구성되는데, 예컨대 상기 식별된 게놈약학적 바이오마커의 5' 및 3' 약 55 kb, 약 60 kb, 약 65 kb, 약 70 kb, 약 75 kb, 약 80 kb, 약 85 kb, 약 90 kb, 약 95 kb, 또는 약 100 kb이다.

[0094] 몇 가지 실시 상태에 있어서, 상기 게놈약학적 바이오마커는 하나 이상의 추가적인 게놈약학적 바이오마커를 식별하는 데 사용될 수 있다. 예컨대, 상기 게놈약학적 바이오마커에 인접하여 위치한 다른 다형자는 연관 표현형과의 관련성에 대하여 분석될 수 있다. 또한, 상기 게놈약학적 바이오마커와 인접한 유전자들이 식별되어 그들의 기능이 분석될 수 있다. 상기 연관 표현형과 직접 또는 간접으로 관련된 기능을 가지는 유전자들, 또는 동일한 세포 경로의 다른 유전자들이 상기 연관 표현형을 이용한 추가 분석의 표적이 될 수 있고 새로운 게놈약학적 바이오마커가 식별될 수 있다.

- [0109] [0095] 특정한 실시 상태에 있어서, 다형 변이체들은 반복적으로 식별된다. 예컨대, 첫번째 가까운 다형 변이체가 전술된 방법을 이용하여 식별된 후, 상기 첫번째 가까운 다형 변이체와 가까운 다른 다형 변이체가 식별되고 (예컨대, 공개적으로 개시 또는 발견된) 상기 첫번째 가까운 다형 변이체에 가까운 하나 이상의 다른 다형 변이체의 관련성의 존재 또는 부재가 결정된다.
- [0110] [0096] 본 발명에 개시된 방법은 질환, 질병 또는 장애와 관련된 유전자, 부위 또는 위치를 추가로 특징짓는 데 유용할 수 있는 추가적인 다형 변이체를 식별 또는 동정하는 데 유용하다. 예컨대, 추가적인 다형 변이체로부터의 대립 유전자형 분석 또는 유전자형 분석 데이터는 기능적 돌연변이 또는 연관비평형 부위를 식별하는 데 사용할 수 있다. 특정한 실시 상태에 있어서, 상기 식별된 게놈약학적 바이오마커를 포함하는 한 부위 내에서 식별 또는 동정된 다형 변이체들은 본 발명에 개시된 유전자형 방법 및 샘플 선택 기법을 이용하여 유전자형 분석되고, 이들 다형 변이체들이 상기 식별된 게놈약학적 바이오마커와 연관비평형에 있는지 여부가 결정될 수 있다. 상기 식별된 게놈약학적 바이오마커와의 연관비평형에 있어서 상기 부위의 크기 역시 이들 유전자형 분석 방법을 이용하여 평가될 수 있다. 따라서, 본 발명은 다형 변이체가 식별된 게놈약학적 바이오마커와 연관비평형에 있는지를 결정하는 방법을 제공하고, 이러한 정보는 본 발명에 개시된 예후/진단 방법에 사용될 수 있다.
- [0111] [0097] 또한, 본 발명은 본 발명에 개시된 방법에 의하여 식별된 게놈약학적 바이오마커를 이용하여 신규한 약물 표적을 식별하는 방법을 제공한다. 몇 가지 실시 상태에 있어서, 상기 바이오마커와 그들의 관련 SNP 또는 유전자는 연관 표현형을 기본으로 하는 근본적인 생물학적 경로 또는 메커니즘, 예컨대 효능, 부작용 또는 기타 종말점에 관한 통찰력을 부여한다. 이들의 발견은 더 나은 진단 또는 치료적 제제의 개발을 위한 수단일 수 있다.
- [0112] H. 키트
- [0113] [0098] 전술한 게놈약학적 바이오마커에 기초한 진단 키트가 개발될 수 있고, 이들은 해당 약물에 대한 개체의 반응을 예측하는 데 사용될 수 있다. 이러한 테스트 키트는 대상체가 샘플, 예컨대 구강 세포 또는 혈액을 헬스케어 공급자의 도움없이도 얻기 위하여 사용할 수 있는 장치 및 지시사항을 포함할 수 있다.
- [0114] [0099] 상기에 설명 또는 제시된 활용법으로 사용하기 위한 키트 또는 제품 역시 본 발명에 의하여 제공된다. 이러한 키트는 본 발명에서 개시된 게놈약학적 바이오마커를 유전자형 분석하기 위한 특정한 제제를 하나 이상 포함할 수 있고, 본 발명에 개시된 방법을 수행하기 위한 지시사항을 더 포함할 수 있다.
- [0115] [0100] 몇 가지 실시 상태에 있어서, 본 발명은 프라이머 및 프라이머 쌍과 프로브를 포함하는 조성물 및 키트를 제공하는데, 이들 프라이머 및 프라이머 쌍은 본 발명의 폴리뉴클레오티드 또는 이들의 임의의 특정 부분의 특이적 증폭을 가능하게 하며, 프로브는 본 발명의 핵산 또는 이들의 임의의 부분에 선택적으로 또는 특이적으로 혼성화한다. 프로브는 검출 가능한 마커, 예컨대 방사선 동위원소, 형광 화합물, 바이오발광 화합물, 화학발광 화합물, 금속 킬레이터 또는 효소 등으로 표지될 수 있다.. 이러한 프로브 및 프라이머는 샘플 내의 폴리뉴클레오티드의 존재를 검출하는 데 사용될 수 있고, 상기 폴리뉴클레오티드로 인코딩되는 세포 발현 단백질들을 검출하는 수단으로서 사용될 수 있다. 이 기술 분야의 숙련자에 의하여 이해되겠지만, 대단히 많은 수의 상이한 프라이머와 프로브가 본 발명에서 제공되는 서열에 기초하여 준비되어 효과적으로 게놈 DNA의 존재 및/또는 수준을 증폭, 클로닝 및/또는 그 존재를 결정하는 데 사용될 수 있다.
- [0116] [0101] 몇 가지 실시 상태에 있어서, 상기 키트는 폴리펩타이드의 존재를 검출하기 위한 제제를 포함할 수 있다. 이러한 제제들은 항체 또는 폴리펩타이드에 특이적으로 결합하는 기타 결합 분자들일 수 있다. 몇 가지 실시 상태에 있어서, 이러한 항체 또는 결합 분자들은 다형성의 결과로서 상기 폴리펩타이드에 일어난 구조적 변이를 구별할 수 있으며, 따라서 유전자형 분석에 사용될 수 있다. 상기 항체 또는 결합 분자들은 검출 가능한 마커, 예컨대 방사선 동위원소, 형광 화합물, 바이오발광 화합물, 화학발광 화합물, 금속 킬레이터, 효소 또는 입자로 표지될 수 있다. 결합 실험, 예컨대 ELISA를 수행하기 위한 기타 제제들이 상기 키트 내에 포함될 수 있다.
- [0117] [0102] 몇 가지 실시 상태에 있어서, 상기 키트는 2개 이상, 3개 이상, 5개 이상, 10개 이상, 또는 15개 이상의 게놈약학적 바이오마커를 유전자형 분석하기 위한 제제를 포함한다. 몇 가지 실시 상태에 있어서, 상기 키트는 증폭된 핵산을 검출하기 위한 프로브 포획용 표면 또는 기판 (예컨대, 마이크로어레이)을 더 포함할 수 있다.
- [0118] [0103] 상기 키트는 하나 이상의 컨테이너 수단, 예컨대 바이알, 튜브, 등등으로 나누어 밀폐하여 담아 가두기 위한 캐리어 수단을 더 포함할 수 있고, 상기 컨테이너 수단 각각은 이 방법에서 사용되는 별개의 요소들 중 하나를 포함한다. 예컨대, 상기 컨테이너 수단 중 하나는 검출 가능하게 표지된 또는 표지될 수 있는 프로브를 포

함할 수 있다. 이러한 프로브는 게놈약학적 바이오마커에 대하여 특이적인 폴리뉴클레오티드일 수 있다. 상기 키트가 표적 핵산을 검출하는 데 핵산 혼성화를 이용한다면, 상기 키트는 표적 핵산 서열의 증폭을 위한 뉴클레오티드(들)을 포함하는 컨테이너 및/또는 리포터 수단, 예컨대 비오틴 결합 단백질, 예컨대 애비딘 또는 스트렙트애비딘, 리포터 분자에 결합하는, 예컨대 효소적, 형광적 또는 방사선 동위원소 표지를 포함하는 컨테이너 역시 포함할 수 있다.

- [0119] [0104] 본 발명의 키트는 통상적으로 전술한 컨테이너 및 상업적 및 사용자적 관점에서 바람직한 물질들, 예컨대 완충액, 희석제, 필터, 바늘, 주사기 및 사용을 위한 지시사항을 포함하는 패키지 삽입물을 포함하는 하나 이상의 다른 컨테이너를 포함할 것이다. 컨테이너 상에는 이 조성물이 특이적인 요법용이라거나 비치료적으로 적용되는 것임을 표시하기 위한 표지가 존재할 수 있고, 전술한 바와 같이 생체 내 (*in vivo*) 또는 시험관 내 (*in vitro*) 사용 중 어느 하나를 위한 것이라는 지침을 표시할 수도 있다.
- [0120] [0105] 상기 키트는 조직 또는 세포 샘플을 준비 및 상기 샘플로부터 핵산 (예컨대 게놈 DNA)을 준비하기 위한 한 세트의 지시사항과 물질들을 더 포함할 수 있다.
- [0121] [0106] 본 발명은, 본 발명의 방법을 수행하는 데 사용하기에 적합한, 키트에 사용할 수 있는 다양한 조성물을 제공한다. 예컨대, 본 발명은 표면, 예컨대 이러한 방법에 사용될 수 있는 어레이 등을 제공한다. 몇 가지 실시 상태에 있어서, 본 발명의 어레이는 본 발명의 게놈약학적 바이오마커를 검출하기 위하여 유용한 핵산 분자 각각 또는 그 모음을 포함한다. 예컨대, 본 발명의 어레이는 표적 핵산을 포함하는 샘플에 혼성화할 수 있는, 분산하여 위치하는 일련의 개별적인 핵산 올리고뉴클레오티드 또는 핵산 올리고뉴클레오티드 조합의 세트를 포함할 수 있고, 이로써 이러한 혼성화는 본 발명의 게놈약학적 바이오마커의 유전자형을 표시하게 된다.
- [0122] [0107] 핵산을 고체 기판, 예컨대 유리 슬라이드에 부착시키는 몇 가지 기법이 이 기술 분야에 잘 알려져 있다. 한가지 방법은 고체 기판에 부착 가능한 모이어티, 예컨대 아민기, 아민기의 유도체 또는 양전하를 갖는 기타 기능기를 함유하는 개질된 염기 또는 유사체들을 합성되는 핵산 분자에 혼입시키는 것이다. 합성되는 생성물을 그 후 고체 기판, 예컨대 증폭된 생성물 상에서 반응성기와 공유 결합을 형성하고 유리 슬라이드에 공유적으로 부착될 알데히드 또는 기타 반응성기로 코팅된 유리 슬라이드와 접촉시킨다. 다른 방법, 예컨대 아미노 프로프릴 실리카 표면 화학을 이용하는 방법 등이 또한 웹 [cmt.corning.com and cmgm.stanford.edu/pbrown1]에 개시된 바와 같이 이 기술 분야에 알려져 있다.
- [0123] [0108] 이 기술 분야에 알려진 방법을 사용하여, 나중에 반응성기로 전환될 수 있는 기능기를 올리고뉴클레오티드에 부착하는 것 역시 가능하다. 올리고뉴클레오티드의 뉴클레오티드에 이루어지는 임의의 부착이 올리고뉴클레오티드의 일부가 될 것이고, 이것은 이후 마이크로어레이의 고체 표면에 부착될 수 있다. 이 기법에서 필요로 되고 및/또는 허용하는 바, 증폭된 핵산은, 상기 고체 기판에의 부착 이전 또는 이후에, 예컨대 단편으로의 분해 또는 검출 가능한 표지의 부착을 통하여 더 개질될 수 있다.
- [0124] [0109] 본 발명은 생물의학 분야에서 광범위하게 적용되고 현대적인 약물 개발 및 개인별 의학이라는 신흥 분야에 큰 이점을 제공할 수 있다. 이를 유익은 임상 개발 약물에 대한 바이오마커를 식별하는 데 요구되는 시간을 단축시키고 비용을 절감하며, 연구 약물의 성공 확률을 크게 향상시키고, 임상 시험에서 환자를 계층화하지 않고 포기되었을 약물을 구제하는 것 등을 포함하지만 이에 국한되는 것은 아니다.
- [0125] I. 컴퓨터 가독성 매체
- [0126] [0110] 또 하나의 다른 측면에 있어서, 본 발명은 차선의 게놈 DNA 샘플을 이용하는 유전자형 분석 방법에 대한 다수의 지시사항을 포함하는 컴퓨터 가독성 매체를 제공하고, 상기 방법은 a) 상기 차선의 게놈 DNA 샘플의 서열 정보를 받는 단계; b) 상기 서열 정보에 기초하여 포함 기준을 최적화하는 단계; 및 c) 상기 서열 정보 및 상기 최적화된 포함 기준에 기초하여 유전자형을 계산하는 단계를 포함한다.
- [0127] [0111] 본 발명은 또한 차선의 게놈 DNA를 이용하는 유전자형 분석 방법을 제공하는데, 이 방법은 포함 기준을 최적화하는 단계를 포함한다. 몇 가지 실시 상태에 있어서, 상기 최적화는 수회 반복되어 샘플들을 포함 및/또는 배제시킬 수 있다. 몇 가지 실시 상태에 있어서, 최적의 포함 기준이 식별될 수 있다. 몇 가지 실시 상태에 있어서, 상기 유전자형 분석 데이터는 전장 게놈 유전자형 호출 알고리즘 및/또는 검증 유전자형 호출 알고리즘을 이용하여 얻어질 수 있다. 몇 가지 실시 상태에 있어서, 상기 유전자형 호출은 고품질의 게놈 DNA의 전체 게놈 유전자형 분석용으로 사용되는 통상적인 호출률 컷-오프보다 낮은 호출률 컷-오프를 사용하여 이루어질 수 있고, 여기서 사용되는 상기 호출률 컷-오프 값은 50%, 60%, 70%, 80%, 90% 또는 95%일 수 있다. 몇 가지 실시 상태에 있어서, 상기 유전자형 분석 데이터는 다수의 유전자형 분석 플랫폼을 이용하여 얻어질 수 있다. 몇 가

지 실시 상태에 있어서, 다수의 유전자형 분석 플랫폼으로부터 얻은 유전자형 분석 데이터는 상기 최적화를 위하여 비교될 수 있다.

[0128] [0112] 차선의 게놈 DNA 샘플을 이용한 유전자형 분석 방법을 이용하는 관련성 분석 수행 방법 역시 제공되고, 이 방법은 포함 기준을 최적화하는 단계를 포함한다. 몇 가지 실시 상태에 있어서, 상기 관련성 분석은 최적화를 위하여 수회 반복될 수 있다.

J. 실시예

[0130] [0113] 하기의 실시예들은 본 발명에 구체화하기 위하여 제공되는 것이지 본 발명을 한정하는 것이 아니다.

실시예 1

임상 시험의 보관된 혈장 샘플을 이용한 게놈약학적 바이오마커의 소급적

신생 식별

[0134] [0114] 환자. 임상 시험에 등록되고 약물로 치료된 환자들 중, 400명 개체의 혈장 샘플이 이용가능하다. 케이스는 상기 약물 치료로부터 양성적으로 반응했던 개체들로 정의되고, 대조군은 상기 약물 치료로부터 아무 반응이 없거나 음성적으로 반응했던 개체들이다. 연구 이전에, 환자 식별 및 개별적으로 식별할 수 있는 정보는 제거되었고, 환자의 신원을 보호하기 위하여 모든 샘플들은 제3자에 의하여 재표지되었다.

[0135] [0115] DNA 준비. 약간의 변경을 가하여 QIAGEN QIAamp MinElute Virus Spin Kit (Valencia, CA, USA)를 이용, 혈장 샘플들로부터 DNA를 추출한다. 간략히, 1 ml의 혈장을 간단히 볼텍싱하고 30 µg tRNA와 완전히 혼합한다. 이 혼합물을 200 µl 엘리퀴트로 나누어 용해 완충액을 첨가하기 전 1 시간 동안 반응시킨다. 그 후, 용해물을 96°C에서 5 분간 끓이고 각각의 엘리퀴트를 동일한 컬럼을 통하여 여과한다. DNA는 멸균수에 녹인 진공 건조된 10 mM Tris-HCl (pH 8.5)에서 용리된다. 대부분의 경우, 혈장으로부터 추출된 게놈 DNA의 양은 너무 적어서 후속하는 유전자형 분석용으로는 부적당하고, 그 후 DNA 샘플은 Amersham Bioscience GenomiPhi DNA Amplification Kit (Piscataway, NJ, USA)를 이용하여 증폭된다.

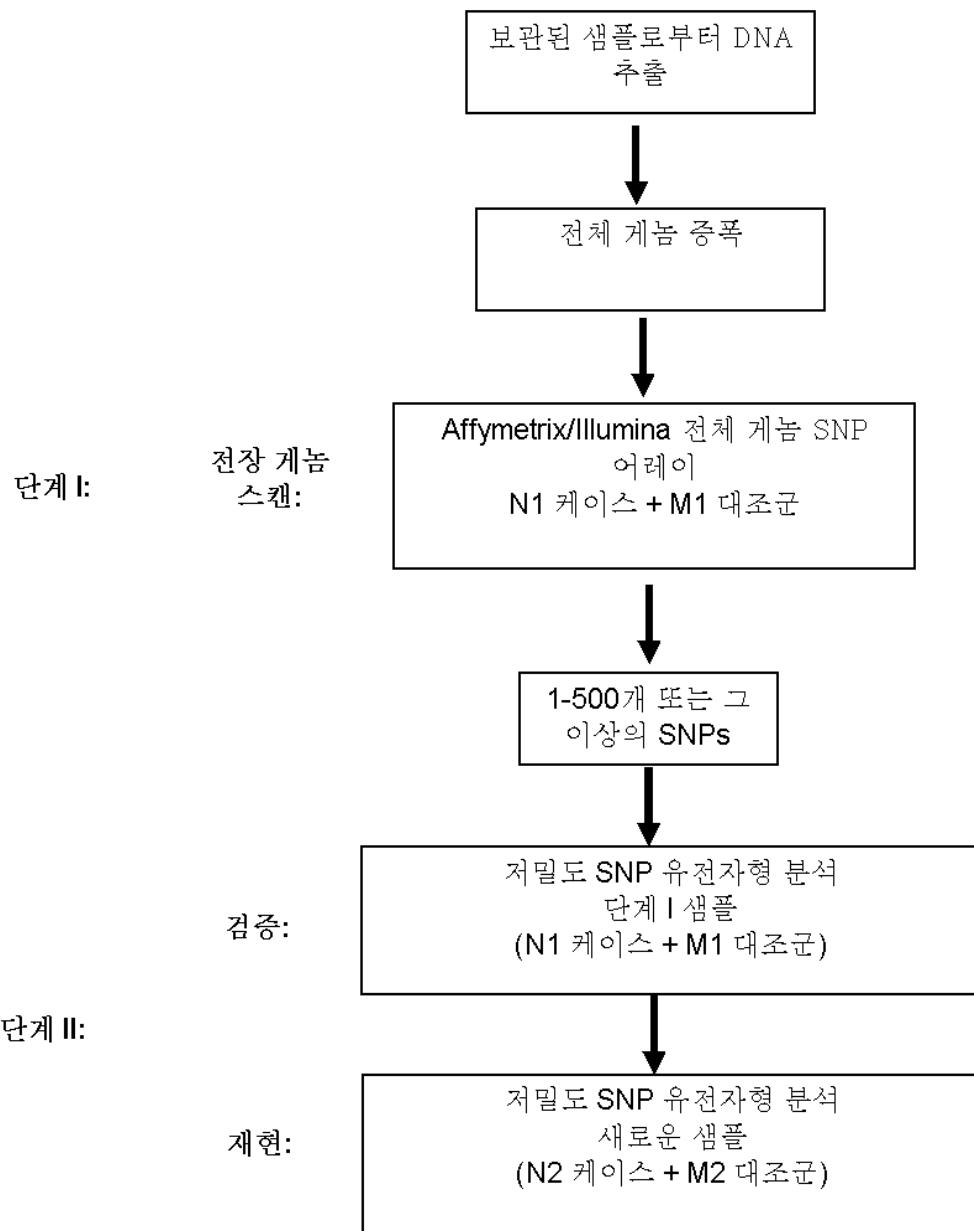
[0136] [0116] SNP 유전자형 분석 및 데이터 분석. 단계 I에서 (도 1), Affymetrix 표준 프로토콜에 따라 (Santa Clara, CA, USA), 150개의 샘플 (75개 케이스 및 75개 대조군)이 500,000 SNP를 포함하는 Affymetrix GeneChip 500K Mapping Array Set을 이용하여 유전자형 분석된다. 혈장 샘플로부터 생성된 차선의 게놈 DNA의 품질 및 유전자형 분석 이전에 적용되는 전체 게놈 증폭 때문에, 일부 샘플들의 호출률은 고품질의 게놈 DNA를 사용하는 경우의 통상적인 호출률 (>95%)보다 실질적으로 낮을 수 있다. 그러므로, 컷-오프 호출률은 가능한 많은 샘플들을 포함할 수 있도록 종래 표준보다 현저히 낮게 조정된다. 유전자형 호출 알고리즘의 호출률 컷-오프 값의 조정 및 상기 조정된 호출률 컷-오프 값에 기초하여 유전자형 호출을 시행하는 것은 수회 반복되어 최적의 기준을 식별하여 샘플들을 포함 및/또는 배제한다 (도 2). 최적의 컷-오프 기준보다 낮은 호출률을 갖는 샘플들을 제거한 후, 유전자형 결과를 유전자 데이터 분석 소프트웨어 PLINK (Purcell et al., *Am J Hum Genet* (2007) 81:559-5752)를 이용하여 분석하여 각 다형좌의 연관 표현형과의 관련 p-값을 계산한다. 다형좌를 상기 연관 표현형과의 계산된 관련성에 의하여 정렬한다. 단계 II 연구를 위하여 가장 유의미하게 관련된 200개의 SNP가 Sequenom iPLEX 실험 (Sequenom, San Diego, CA, USA)을 이용하여 선택된다. 이를 실험은 상기 임상 시험으로부터의 400개의 모든 DNA 샘플들을 유전자형 분석하는 데 사용된다. 이들 중에서, 단계 I에서 사용되었던 150개의 샘플들이 검증 그룹으로 선택되고, 다른 250개의 샘플은 재현 그룹으로 사용된다 (도 1). 최종 유전자형 호출은 MassARRAY Typer suite (Sequenom, San Diego, CA, USA)의 Sequenom Typer Analyzer에 의하여 생성된다. 검증 그룹의 유전자형 분석된 결과를 단계 I에서 생성된 결과와 비교한다. 일부 샘플들로부터의 호출률은 고품질의 게놈 DNA를 사용하는 경우의 통상적인 호출률 (>95%)보다 실질적으로 낮을 수 있고, 컷-오프 호출률은 수회 조정되어 샘플들을 포함 및/또는 배제시킨다. 최적 컷-오프 기준보다 낮은 호출률을 갖는 샘플들 및 두 단계 간에 불일치가 매우 큰 샘플들을 제거한 후 (도 3), PLINK 프로그램을 이용하여 각 SNP의 연관 표현형과의 관련 P-값을 계산함으로써 연관성 분석이 수행된다. 상기 계산은 대립 유전자 빈도 및/또는 유전자형 분석에 기반한 테스트에 기초할 수 있고, 상기 연관 표현형은 범주적 특성, 정량적 특성 또는 다른 연관 표현형일 수 있다. 약물 반응과 유의미한 관련성을 나타내는 게놈약학적 바이오마커가 식별되고, 이를 게놈약학적 바이오마커는 하나 이상의 SNP를 포함할 수 있다. 식별된 다수의 게놈약학적 바이오마커는 게놈 상에서 서로 가깝게 위치할 수 있고, 유전자의 인트론/엑손, 유전자간 부위, 또는 게놈 상 알려지지 않은 유전자를 포함하는 부위에 위치할 수 있다. 분류 알고리즘, 예컨대 서포트 벡터 머신 (SVM) 및 로지스틱 회귀분석이 데이터세트에 적용되어 최적의 바이오마커 및 점수화 알고리즘을 식별할 수 있으므로, 환자의 약물 치료에 대한 반응이 자격절히 예측될 수 있

다.

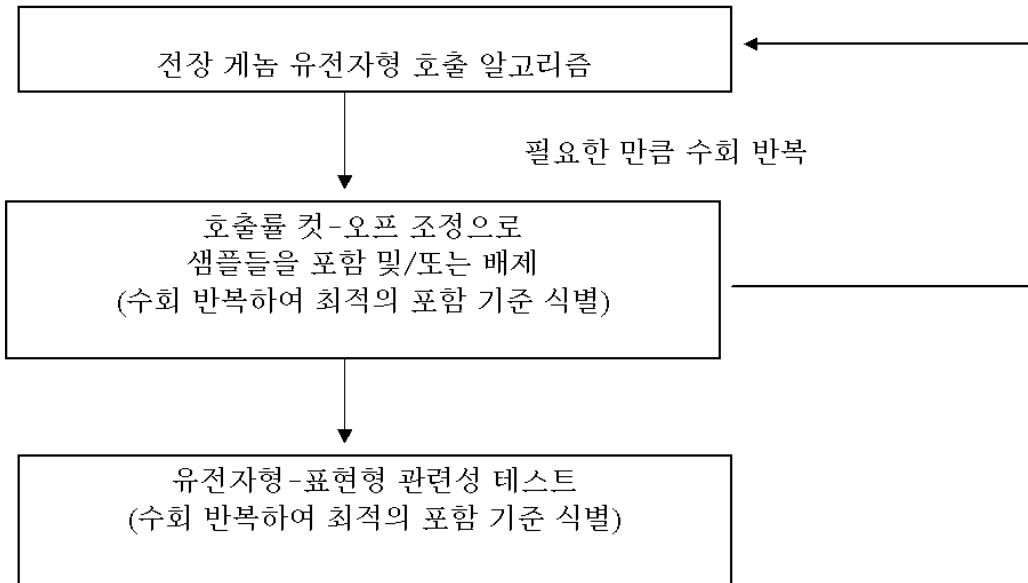
[0137] 상기 실시예들은 오로지 설명의 목적으로만 포함되고 본 발명의 범위를 제한하려는 의도가 아니다. 상기 개시된 것들에 대한 많은 변형이 가능하다. 절술된 실시예들에 대한 변경 및 변형이 이 기술 분야의 숙련자에게 명확하므로, 본 발명은 첨부된 청구항들의 범위에 의하여만 제한되는 것을 목적한다.

도면

도면1



도면2



도면3

