



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0057188  
(43) 공개일자 2011년05월31일

(51) Int. Cl.

*G01N 33/68* (2006.01)   *C12Q 1/68* (2006.01)  
*G06F 19/28* (2011.01)   *G06F 17/18* (2006.01)

(21) 출원번호 10-2011-7007024

(22) 출원일자(국제출원일자) 2009년08월27일  
시사처구이자 양우

(85) 번역문제총정리 2011년03월25일

(85) 민국군세율 증자 2011년03월23일  
(86) 국제총회법률 RCT/US2000/055144

(86) 국제출원번호 PCI/US2009/053144  
(87) 그림 10

(87) 국제공개번호 WO 2010/025216  
국내공개번호 2010-002-00000

국제공개일자 2010년03월04일

### (30) 우선권주장

61/092,270 2008년08월27일 미국(US)

(71) 출원인

하. 룬트벡 아크티에 셀스카브

덴마크 디케이-2500 코펜하겐 발비 오텔리아베이  
9

(72) 발명자

안토니예빅 이리나

미국 07430 뉴저지주 마와 리오 비스타 드라이브  
1201

탐조센

미국 07506 뉴저지주 호손 부에나 비스타 애비뉴  
103

(뒷면에 계속)

### (74) 대리인

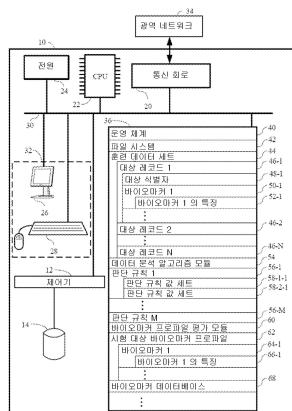
전체 청구항 수 : 총 35 항

#### (54) 바이오마커 프로파일 측정 시스템 및 방법

(57) 요약

본 발명은 정동 장애가 있는 환자를 진단하는 방법 및 시스템에 관한 것이다. 본 방법은 또한 대상 내 정동 장애에 대한 간수성을 예측하는데 유용하다.

대표도 - 도1



(72) 발명자  
**아티미신 로만**  
미국 08822 뉴저지주 플레밍턴 스텐튼 로드 114  
**제랄드 크리스토фф 퍼.지.**  
미국 07430 뉴저지주 마와 리오 비스타 드라이브  
1201

**비스티센 얀 바스톨름**  
덴마크 디케이-9600 아르스 아토프텐 6

## 특허청구의 범위

### 청구항 1

하기 단계를 포함하는, 시험 대상에서 정동 장애를 진단하는 방법:

시험 대상의 바이오마커 프로파일 중 다수의 바이오마커의 다수의 특징이 값 세트를 만족하는지를 평가하는 단계로서, 값 세트를 만족하는 것이 시험 대상이 상기 정동 장애를 갖는지를 예측하며, 다수의 특징은 다수의 바이오마커의 측정가능한 양상으로, 상기 다수의 바이오마커는 표 1A에 열거된 2 개 이상의 바이오마커를 포함하는 단계.

### 청구항 2

제 1 항에 있어서, 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 인터페이스 소자, 모니터, 유형(有形) 컴퓨터 판독가능 저장 매체, 또는 지역적 또는 원격 컴퓨터 시스템으로 출력하는 것; 또는 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 판독가능 형태로 표시하는 것을 추가로 포함하는 방법.

### 청구항 3

제 1 항에 있어서, 상기 다수의 바이오마커가 표 1A에 열거된 2 내지 29 개의 바이오마커로 이루어지는 방법.

### 청구항 4

제 1 항에 있어서, 상기 다수의 바이오마커가 표 1A에 열거된 3 내지 20 개의 바이오마커로 이루어지는 방법.

### 청구항 5

제 1 항에 있어서, 상기 다수의 바이오마커가 표 1A에 열거된 2 개 이상의 바이오마커를 포함하는 방법.

### 청구항 6

제 1 항에 있어서, 상기 다수의 바이오마커가 표 1A에 열거된 3 개 이상의 바이오마커를 포함하는 방법.

### 청구항 7

제 1 항에 있어서, 상기 다수의 바이오마커가 표 1A에 열거된 4 개 이상의 바이오마커를 포함하는 방법.

### 청구항 8

제 1 항에 있어서, 상기 다수의 특징이 표 1A에 열거된 2 내지 29 개의 바이오마커에 상응하는 2 내지 29 개의 특징으로 이루어지는 방법.

### 청구항 9

제 1 항에 있어서, 상기 다수의 특징이 표 1A에 열거된 3 내지 15 개의 바이오마커에 상응하는 3 내지 15 개의 특징으로 이루어지는 방법.

### 청구항 10

제 1 항에 있어서, 상기 다수의 특징이 표 1A에 열거된 2 개 이상의 바이오마커에 상응하는 2 개 이상의 특징을 포함하는 방법.

### 청구항 11

제 1 항에 있어서, 상기 다수의 바이오마커가 ERK1 및 MAPK14를 포함하는 방법.

### 청구항 12

제 1 항에 있어서, 상기 다수의 바이오마커가 Gi2 및 IL-1b를 포함하는 방법.

**청구항 13**

제 1 항에 있어서, 상기 다수의 바이오마커가 ARRB1 및 MAPK14 를 포함하는 방법.

**청구항 14**

제 1 항에 있어서, 상기 다수의 바이오마커가 ERK1 및 IL1b 를 포함하는 방법.

**청구항 15**

제 1 항에 있어서, 상기 다수의 바이오마커가 ARRB1, IL6 및 CD8a 를 포함하는 방법.

**청구항 16**

제 1 항에 있어서, 상기 다수의 바이오마커가 ARRB1, ODC1 및 P2X7 을 포함하는 방법.

**청구항 17**

제 1 항에 있어서, 상기 다수의 바이오마커 중 각각의 바이오마커가 핵산인 방법.

**청구항 18**

제 1 항에 있어서, 상기 다수의 바이오마커 중 각각의 바이오마커가 DNA, cDNA, 증폭된 DNA, RNA, 또는 mRNA 인 방법.

**청구항 19**

제 1 항에 있어서, 시험 대상의 바이오마커 프로파일 중 상기 다수의 특징 중 특징이 다수의 바이오마커 중 바이오마커의 측정가능한 양상이고, 상기 특성에 대한 특성 값이 상기 시험 대상으로부터 취해진 생물학적 샘플을 사용하여 측정되는 방법.

**청구항 20**

제 19 항에 있어서, 상기 특징이 생물학적 샘플 내 상기 바이오마커의 풍부함이고, 생물학적 샘플이 전혈인 방법.

**청구항 21**

제 1 항에 있어서, 평가 단계 전에, 상기 제 1 값 세트를 구축하는 것을 추가로 포함하는 방법.

**청구항 22**

제 21 항에 있어서, 구축 단계가 데이터 분석 알고리즘을 집단의 일원으로부터 수득된 특징에 적용하는 것을 포함하는 방법.

**청구항 23**

제 22 항에 있어서, 상기 집단이 정동 장애를 갖지 않은 제 1 의 다수의 대조군 대상으로부터의 제 1 의 다수의 생물학적 샘플 및 정동 장애를 갖는 제 2 의 다수의 대상으로부터의 제 2 의 다수의 생물학적 샘플을 포함하는 방법.

**청구항 24**

제 22 항에 있어서, 상기 데이터 분석 알고리즘이 결정 트리, 마이크로어레이의 예측 분석, 다중 합산적 회귀 트리, 뉴럴 네트워크, 클러스터링 알고리즘, 주요 성분 분석, 최근접 이웃 분석, 선형 판별식 분석, 이차 판별식 분석, 지원 벡터 머신, 진화 방법, 관련성 벡터 머신, 유전학적 알고리즘, 투사 지향, 또는 가중치 투표인 방법.

**청구항 25**

제 21 항에 있어서, 구축 단계가 판단 규칙을 생성하고, 상기 평가 단계가 제 1 값 세트를 만족하는지를 측정

하기 위해 상기 판단 규칙을 다수의 특징에 적용하는 것을 포함하는 방법.

#### 청구항 26

제 25 항에 있어서, 상기 판단 규칙이 상기 집단에서 대상을 (i) 정동 장애를 갖지 않는 대상 및 (ii) 정동 장애를 갖는 대상으로서 70% 이상의 정확성으로 분류하는 방법.

#### 청구항 27

제 25 항에 있어서, 상기 판단 규칙이 상기 집단에서 대상을 (i) 정동 장애를 갖지 않는 대상 및 (ii) 정동 장애를 갖는 대상으로서 90% 이상의 정확성으로 분류하는 방법.

#### 청구항 28

제 1 항에 있어서, 정동 장애가 양극성 장애 I, 양극성 장애 II, 기분저하 장애, 또는 우울 장애인 방법.

#### 청구항 29

제 1 항에 있어서, 정동 장애가 경증 우울증, 중등 우울증, 중증 우울증, 비정형 우울증, 우울성 우울증, 또는 경계 인격 장애인 방법.

#### 청구항 30

컴퓨터 판독가능 저장 매체 및 그곳에 저장된 컴퓨터 프로그램 메카니즘을 포함하고, 상기 컴퓨터 프로그램 메카니즘이 제 1 항의 방법을 수행하기 위한 지침을 포함하는 컴퓨터 프로그램 제품.

#### 청구항 31

제 30 항에 있어서, 컴퓨터 프로그램 메카니즘이, 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 인터페이스 소자, 모니터, 유형 컴퓨터 판독가능 저장 매체, 또는 지역적 또는 원격 컴퓨터 시스템으로 출력하는 것; 또는 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 판독가능 형태로 표시하는 것을 위한 지침을 추가로 포함하는 컴퓨터 프로그램 제품.

#### 청구항 32

하나 이상의 프로세서;

하나 이상의 프로세서에 연결된 메모리를 포함하고, 상기 메모리는 제 1 항의 방법을 수행하기 위한 지침을 저장하는 컴퓨터.

#### 청구항 33

제 32 항에 있어서, 메모리가, 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 인터페이스 소자, 모니터, 유형 컴퓨터 판독가능 저장 매체, 또는 지역적 또는 원격 컴퓨터 시스템으로 출력하는 것; 또는 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 판독가능 형태로 표시하는 것을 위한 지침을 추가로 포함하는 컴퓨터.

#### 청구항 34

하기 단계를 포함하는, 시험 대상이 정동 장애의 증상을 나타내는 가망성을 측정하는 방법:

시험 대상의 바이오마커 프로파일 중 다수의 바이오마커의 다수의 특징이 값 세트를 만족하는지를 평가하는 단계로서, 값 세트를 만족하는 것이 시험 대상이 정동 장애의 증상을 나타내는 가망성을 제공하며, 다수의 특징은 다수의 바이오마커의 측정가능한 양상인 단계 (다수의 바이오마커는 표 1A 에 열거된 2 개 이상의 바이오마커를 포함함).

#### 청구항 35

제 34 항에 있어서, 시험 대상이 정동 장애의 증상을 나타내는 가망성을 사용자 인터페이스 소자, 모니터, 유형 컴퓨터 판독가능 저장 매체, 또는 지역적 또는 원격 컴퓨터 시스템으로 출력하는 것; 또는 시험 대상이 정동 장애의 증상을 나타내는 가망성을 사용자 판독가능 형태로 표시하는 것을 추가로 포함하는 방법.

## 명세서

### 기술분야

[0001] 본 출원은 2009년 8월 25일에 작성된, 148,658 바이트의 크기의 제목이 71021-WO-PCT\_SequenceListing\_ST25.txt인 전자적 형태로 제출된 서열 목록을 담고 있다. 서열 목록은 전체가 참조로서 본원에 인용되어 있다.

### 1. 기술 분야

[0003] 본 발명은 질환이 있는 대상과 비교하여 대조군 대상 내 유전자의 mRNA 발현 수준을 프로파일링 및 비교함으로써 장애를 겪고 있는 대상 내 전사 프로파일을 확인하는 방법 및 조성물을 제공한다. 본 발명은 추가로, 대상 내 바이오마커와 관련된 전사 프로파일을 측정함으로써 정동 장애와 같은 장애를 예측하고 진단하기 위한 방법 및 조성물을 제공한다.

### 배경기술

#### 2. 배경 기술

[0005] 본 명세서 전반에 걸쳐 다양한 공개 문헌이 괄호 내의 인용구로써 언급된다. 상기 문헌의 명세서는 전체가, 본 발명이 속하는 상태를 더욱 완전히 기술하기 위해 본 출원에 참조로서 인용된다.

[0006] 현재의 정신과 진단 분류, 특히 정동 장애에 대한 진단 분류에는, 뚜렷한 임상적 기술이 결핍되고, 하나의 진단 실체를 또 다른 것으로부터 기술하는 생물학적 특징이 포함되지 않는다. 오늘날의 분류가 정동 장애, 예를 들어, 주요 우울 장애의 임상적 특징을 더욱 구체화하는 것을 허용함에도 불구하고, 기준은 중요한 논쟁의 소지로 남아 있고, 생물학적 이론적 근거를 반드시 따를 필요는 없다 (Parker, et al. *Am. J Psychiatry* 2000, 157(8): 1195-1203).

[0007] 정동 장애 중에서, 예컨대 양극성 장애 I 및 II, 기분저하증, 및 정신병적 우울증, 중증 대 경증 또는 중등 우울증, 우울성 대 비정형 우울증 등을 비롯한 주요 우울 장애와 같은 많은 임상적 구분이 존재한다. 그러므로, 상기 구분에 대해 뚜렷한 생물학적 마커 또는 바이오마커가 설명되지 않는다. 게다가, 구체적인 장애에 대한 구분의 결핍은 치료 암시를 가질 수 있다. 또한, 동반이환은 2 가지 장애의 존재를 기술할 수 없는 의사에 있어 문제가 된다.

[0008] 대체로, 정신의학에서의 임상 평가 및 비특이적 임상적 진단 기준은 유사한 생물학을 공유하는 환자를 인지하기 위해 생물학적 마커에 대한 필요성을 강조한다. 이것은 임상적 차이 및 뚜렷한 생물학적 특징을 나타내는 부차유형의 존재에 대한 최근의 증거가 있으므로, 정동 장애에 대해서는 특정한 딜레마인 것 같다 (Gold and Chrousos, *Mol. Psychiatry* 2002, 7(3): 254-275). 그러나 지금까지 정동 장애와 관련하여 환자 집단의 구분을 기술하기 위해 일관되게 보여지는 생물학적 마커가 없었다.

[0009] 이전 연구는 우울증이 있는 대상 대 대조군 대상, 또는 텍사메타손/코르티코트로핀 방출 호르몬 (DEX/CRH) 시험과 같은 처리 전후의 대상에서 생물학적 변화를 측정하는 시험을 조사해왔다. 그러나, 이러한 시험은 소수의 환자에서 시험되었고, 재현되지 않았고/거나 생물학적 판독값이 특이적 표현형과 연결되지 않았다 (Ising, M. et al., *Biol. Psychiatry*, 2006 Nov 20, e-pub ahead of print; Kunugi, H. et al., *Neuropsychopharmacology*, 2006, 31(1): 212-20). 이것은 임상적으로 관련있는 바이오마커가 특이적 생물학 및 특이적 표현형과 관련되어야만 하고, 이상적으로는 치료에 의해 정상 수준으로 되돌아가야만 하므로 적절하다.

[0010] 단백질 바이오마커가 당뇨병, 알츠하이머 질환, 및 암에 대해 확인되었다 (예를 들어, 미국 특허 제 7,125,663 호; 제 7,097,989 호; 제 7,074,576 호; 및 제 6,925,389 호 참조). 그러나, 질량 분석법 및 항체에 대한 특이적 결합과 같은 단백질 바이오마커의 검출 방법이 종종 재현불가능한 데이터를 산출하고, 상기 방법은 고처리량 사용에 바람직하지 않다.

[0011] 마이크로어레이를 사용하는 고 처리량 발현 분석 방법은, 유전자 발현 변화를 혼합된 결과 또는 관련 없는 산출물로 평가하기 위해 사용되어왔다 (Brenner, S. et al *Nat Biotechnol.* 2000, 18(6):597-8; Schena et al. *Science*. 1995, 270(5235):467-70; Velculescu, V.E. et al, *Science*. 1995, 270(5235):484-7). 측정된 유전자 발현 대 대상의 수의 큰 비, 및 우울 장애의 이질성 제공으로 인해, 다수의 가 양성이 마이크로어레이 데이터로 예상되는 것이다 (리뷰를 위해, Iwamoto K, and Kato T., *Neuroscientist* 2006, 12(4):349-61;

Bunney WE, et al., *Am J Psychiatry* 2003, 160(4):657-66; 및 Iga J, Ueno S, and Ohmori T., *Ann Med* 2008, 40(5):336-42 참조). Sibille 등 (*Neuropsychopharmacology* 2004, 29(2):351-61) 은 대규모 계놈 분석을 수행하였으나, 우울증 및 자살과 연관성이 있는 분자적 차이점에 대한 증거를 찾지 못했고, 우울증과 연관된 것으로 이전에 발견되었던 유전자에 대한 발현 수준 변화를 재현하지 못했다. 이러한 어려움으로 인해, 일관성 있는 프로파일이 확인되지 않았다.

[0012] 복합적 관련 유전자에 대한 집중 어레이 및 qPCR 은 스트레스 관련 유전자의 확인을 위해 사용되나, 이들 연구는 우울증과 관련된 진단 프로파일을 확인하지는 못했다 (Rokutan et al, *J. Med. Invest.* 2005, 52(3-4): 137-44; Ohmori et al., *J. Med. Invest.* 2005, 52 (Suppl):266-71). 래트 뇌 영역에서, 특정 유전자의 전사적 변화는 기분 및 불안의 조절에 연루되어 있으나, 이러한 변화는 인간 혈액 샘플과는 연관성이 없다 (WO2007106685 A2).

## 발명의 내용

### 과제의 해결 수단

#### 3. 발명의 요약

[0014] 본 발명은 하기 단계를 포함하는, 시험 대상에서 정동 장애를 진단하는 방법을 제공한다: 시험 대상의 바이오마커 프로파일 중 다수의 바이오마커의 특징이 값 세트를 만족하는지를 평가하는 단계로서, 값 세트를 만족하는 것이 시험 대상이 상기 정동 장애를 갖는지를 예측하며, 다수의 특징은 다수의 바이오마커의 측정가능한 양상으로, 상기 다수의 바이오마커는 표 1A 에 열거된 2 개 이상의 바이오마커를 포함하는 단계.

[0015] 본 발명은 또한 컴퓨터 판독가능 저장 매체 및 그곳에 저장된 컴퓨터 프로그램 메카니즘을 포함하고, 상기 컴퓨터 프로그램 메카니즘은 진단 방법을 수행하기 위한 지침을 포함하는 컴퓨터 프로그램 제품을 제공한다.

[0016] 본 발명의 하나의 양상은 하나 이상의 프로세서 및 하나 이상의 프로세서에 연결된 메모리를 포함하고, 상기 메모리는 진단 방법을 수행하기 위한 지침을 저장하는 컴퓨터를 제공한다.

[0017] 본 발명의 또 다른 양상은 하기 단계를 포함하는, 시험 대상이 정동 장애의 증상을 나타내는 가망성을 측정하는 방법을 제공한다: 시험 대상의 바이오마커 프로파일 중 다수의 바이오마커의 특징이 값 세트를 만족하는지를 평가하는 단계로서, 값 세트를 만족하는 것이 시험 대상이 정동 장애의 증상을 나타내는 가망성을 제공하며, 다수의 특징은 다수의 바이오마커의 측정가능한 양상으로, 상기 다수의 바이오마커는 표 1A 에 열거된 2 개 이상의 바이오마커를 포함하는 단계.

[0018] 본 발명은 또 다른 양상에서, 다수의 대조군 대상으로부터 수집된 각각의 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 예를 들어, 본 발명은 다수의 우울증, 중증 우울증, 또는 양극성 대상으로부터 수집된 각각의 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 또한 다수의 경계 인격 장애 대상으로부터 수집된 각각의 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 또한 다수의 PTSD 대상으로부터 수집된 각각의 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다.

[0019] 본 발명은 또한 예를 들어 데이터베이스에 저장된, 제 1 의 다수의 대조군 대상의 집합적 측정값을 포함하는 전사 프로파일을 제공한다. 제 2 의 다수의 대상, 예를 들어, 질환이 있는 대상의 집합적 측정값을 포함하는 전사 프로파일을 분류 알고리즘을 사용하는 제 1 의 다수의 대조군 대상의 전사 프로파일과 비교한다. 분류 알고리즘은 각각의 대상을 분류하는 출력물을 제공한다.

[0020] 본 발명은 환자에서 전사 프로파일을 확인하고, 이러한 전사 프로파일과 대조군 대상 또는 대조군 대상 그룹의 프로파일을 비교하여, 전사 프로파일 변화의 존재 또는 부재에 근거하여 환자의 정동 장애를 진단함으로써 정동 장애를 진단하는 방법을 제공한다.

[0021] 본 발명의 하나의 양상은 하기 단계를 포함하는, 정동 장애를 가진 대상을 진단하는 방법을 제공한다:

[0022] (a) 다수의 대조군 대상 및 다수의 질환이 있는 대상으로부터 생물학적 샘플을 수득하는 단계;

[0023] (b) 다수의 대조군 대상 및 다수의 질환이 있는 대상의 샘플에서 유전자의 mRNA 발현 수준을 측정하는 단계로, 유전자가 ADA, ARRB1, ARRB2, CD8a, CD8b, CREB1, CREB2, DPP4, ERK1, ERK2, Gi2, Gs, GR, IL1b, IL6, IL8, INDO, MAPK14, MAPK8, MKP1, MR, ODC1, P2X7, PBR, PREP, RGS2, S100A10, SERT 및 VMAT2 로 이루어지는 군으

로부터 선택되는 단계;

- [0024] (c) 다수의 대조군 대상 및 다수의 질환 있는 대상으로부터의 각각의 유전자에 대한 mRNA 발현 수준을 mRNA 데이터로서 컴퓨터 매체에 수집하고 저장하는 단계;
- [0025] (d) 분류 알고리즘에 의해 이러한 mRNA 데이터를 프로세싱하는 단계; 및
- [0026] (e) 대상을 분류하는 출력물 데이터를 제공하는 단계에 의해,
- [0027] 정동 장애를 가진 대상을 진단하는 방법.
- [0028] 본 발명은 추가로 ADA, ARRB1, ARRB2, CD8a, CD8b, CREB1, CREB2, DPP4, ERK1, ERK2, Gi2, Gs, GR, IL1b, IL6, IL8, INDO, MAPK14, MAPK8, MKP1, MR, ODC1, P2X7, PBR, PREP, RGS2, S100A10, SERT 및 VMAT2로 이루어지는 군으로부터 선택되는 유전자의 대상의 전사 프로파일을 다수의 대조군 대상의 유전자의 전사 프로파일과 비교함으로써 대상의 정동 장애에 대한 감수성을 예측하는 방법을 제공한다.
- [0029] 본 발명의 하나의 양상은 하기 단계를 포함하는, 정동 장애의 증상을 나타내는 대상의 가망성을 예측하는 방법을 제공한다:
- [0030] (a) 다수의 대조군 대상 및 다수의 질환 있는 대상으로부터 생물학적 샘플을 수득하는 단계;
- [0031] (b) 다수의 대조군 대상 및 다수의 질환 있는 대상의 샘플에서 유전자의 mRNA 발현 수준을 측정하는 단계로, 유전자가 ADA, ARRB1, ARRB2, CD8a, CD8b, CREB1, CREB2, DPP4, ERK1, ERK2, Gi2, Gs, GR, IL1b, IL6, IL8, INDO, MAPK14, MAPK8, MKP1, MR, ODC1, P2X7, PBR, PREP, RGS2, S100A10, SERT 및 VMAT2로 이루어지는 군으로부터 선택되는 단계;
- [0032] (c) 다수의 대조군 대상 및 다수의 질환 있는 대상으로부터의 각각의 유전자에 대한 mRNA 발현 수준을 mRNA 데이터로서 컴퓨터 매체에 수집하고 저장하는 단계;
- [0033] (d) 분류 알고리즘에 의해 이러한 mRNA 데이터를 프로세싱하는 단계; 및
- [0034] (e) 대상을 분류하는 출력물 데이터를 제공하는 단계에 의해,
- [0035] 정동 장애의 증상을 나타내는 대상의 가망성을 예측하는 방법.

### 도면의 간단한 설명

#### 4. 도면의 간단한 설명

도 1은 본 발명의 구현예에 따른 컴퓨터 시스템의 도식이다.

도 2A 및 2B. qPCR 방법에 의해 카피/ng cDNA로 측정된, 대조군 대상 대 우울증 대상에서 ARRB1(베타-아레스틴 1) 및 Gi2(구아닌 뉴클레오티드 결합 단백질 알파 i2) 각각의 상대적인 mRNA 수준을 나타내는 산점도 ( $p<0.001$ ; Mann Whitney 검정).

도 3A 및 3B. qPCR 방법에 의해 카피/ng cDNA로 측정된, 대조군 대상 대 우울증 대상에서 MAPK14(p38 미토젠-활성화 단백질 키나아제 14) 및 ODC1(오르니틴 디카르복실라아제 1) 각각의 상대적인 mRNA 수준을 나타내는 산점도 ( $p<0.001$ ; Mann Whitney 검정).

도 4A, 4B 및 4C. qPCR 방법에 의해 카피/ng cDNA로 측정된, 대조군 대상 대 중증 우울증 대상에서 ERK1(세포외부 신호-조절 키나아제 1), Gi2(구아닌 뉴클레오티드 결합 단백질 알파 i2), 및 MAPK14(p38 미토젠-활성화 단백질 키나아제 14) 각각의 상대적인 mRNA 수준을 나타내는 산점도 ( $p<0.001$ ; Mann Whitney 검정).

도 5A, 5B 및 5C. qPCR 방법에 의해 카피/ng cDNA로 측정된, 대조군 대상 대 중증 우울증/양극성 대상에서 Gi2(구아닌 뉴클레오티드 결합 단백질 알파 i2), GR(알파-글루코코르티코이드 수용체), 및 MAPK14(p38 미토젠-활성화 단백질 키나아제 14) 각각의 상대적인 mRNA 수준을 나타내는 산점도 ( $p<0.001$ ; Mann Whitney 검정).

도 6A, 6B 및 6C. qPCR 방법에 의해 카피/ng cDNA로 측정된, 대조군 대상 대 경계 인격 장애 대상에서 Gi2(구아닌 뉴클레오티드 결합 단백질 알파 i2), MAPK14(p38 미토젠-활성화 단백질 키나아제 14), 및 MR(미네랄로코르티코이드 수용체) 각각의 상대적인 mRNA 수준을 나타내는 산점도 ( $p<0.001$ ; Mann Whitney 검정).

도 7A, 7B 및 7C. qPCR 방법에 의해 카피/ng cDNA에 의해 측정된 바와 같이 196명의 대조군 대상 대 66명의 급성 PTSD 대상에서 ARRB2(베타-아레스틴 2), ERK2(세포외부 신호-조절 키나아제 2), 및 RGS2(G-단백질

신호 2 의 조절자) 각각의 상대적인 mRNA 수준을 나타내는 산점도 ( $p<0.001$ ; Mann Whitney 검정).

도 8A 및 8B. 도 8A 는 우울증 대상 대 대조군의 분류 내 93% 의 정확성, PPV = 93%, 및 NPV = 94% 의 점수를 얻은, 유전자 선별 및 훈련을 모두 수행하는 SLR 알고리즘수행의 도식이다. RF 유전자 선별에 앞서 지원 벡터 머신 (SVM) 분류기가 우울증 대상 대 대조군의 분류 내 88% 의 정확성, PPV = 89% 및 NPV = 88% 를 점수화한다. 도 8B 는 우울증 대상 대 대조군의 분류 내 각각의 방법의 통계적 파라미터에 근거하여 표 1A 로부터 14 개의 유전자를 선별하는 랜덤 포레스트 (Random Forest: RF) 및 17 개의 유전자를 선별하는 단계식 로지스틱 회귀 (Stepwise Logistic Regression: SLR) 를 나타낸다. 분류 프로세스의 선별 단계에서 RF 및 SLR 방법 모두에 의해 선별된 중복 유전자가 회색으로 제시된다.

도 9. 도 9 는 평균 발현 수준 (전사체 값) 이 중증 우울증 환자와 대조군 사이에서 유의하게 상이한 ( $p<0.05$ ) 유전자를 기술한다. 상기 유전자는 표 5A 에서 제시된 바와 같이 계산된  $-\text{Log}(p)$  값의 규모에 따라 순위가 매겨진다.

도 10. 도 10 은 각각의 대상에 대해 ERK1 및 MAPK14 로 이루어지는 전사 프로파일에 따른 중증 우울증 대상 및 대조군 대상의 분포를 나타낸다. 중증 우울증 대상은 개방 원형 (○) 으로 표시되고, 대조군 대상은 폐쇄 삼각형 (▲) 으로 표시된다. X 및 Y 축은 ERK1 및 MAPK14 각각에 대한 전사체 값 (카피/ng cDNA) 을 기술한다.

도 11. 도 11 은 각각의 대상에 대해 Gi2 및 IL1b 로 이루어지는 전사 프로파일에 따른 중증 우울증 대상 및 대조군 대상의 분포를 나타낸다. 중증 우울증 대상은 개방 원형 (○) 으로 표시되고, 대조군 대상은 폐쇄 삼각형 (▲) 으로 표시된다. X 및 Y 축은 Gi2 및 IL1b 각각에 대한 전사체 값 (카피/ng cDNA) 을 기술한다.

도 12. 도 12 는 각각의 대상에 대해 ERK1 및 IL1b 로 이루어지는 전사 프로파일에 따른 중증 우울증 대상 및 대조군 대상의 분포를 나타낸다. 중증 우울증 대상은 개방 원형 (○) 으로 표시되고, 대조군 대상은 폐쇄 삼각형 (▲) 으로 표시된다. X 및 Y 축은 ERK1 및 IL1b 각각에 대한 전사체 값 (카피/ng cDNA) 을 기술한다.

도 13. 도 13 은 각각의 대상에 대해 ARRB1 및 MAPK14 로 이루어지는 전사 프로파일에 따른 중증 우울증 대상 및 대조군 대상의 분포를 나타낸다. 중증 우울증 대상은 개방 원형 (○) 으로 표시되고, 대조군 대상은 폐쇄 삼각형 (▲) 으로 표시된다. X 및 Y 축은 ARRB1 및 MAPK14 각각에 대한 전사체 값 (카피/ng cDNA) 을 기술한다.

## 발명을 실시하기 위한 구체적인 내용

[0037]

### 5. 발명의 상세한 설명

[0038]

본 발명은 바이오마커 프로파일 내 바이오마커 특징을 평가함으로써 정동 장애의 신속하고 정확한 진단을 가능하게 한다. 상기 바이오마커 프로파일은 대상의 생물학적 샘플로부터 구축된다.

[0039]

#### 5.1 정의

[0040]

본원에서 사용되는 바와 같은, "정동 장애" 는 기분, 충격적인 생각, 감정 및 행동의 일관적인, 만연하는 변화를 특징으로 하는 정신적 장애를 의미할 것이다. 정동 장애의 예에는 우울 장애, 불안 장애, 양극성 장애, 기분저하증 및 정신분열정동 장애가 포함되나 이에 제한되는 것은 아니다. 불안 장애에는 범불안 장애, 공황 장애, 강박반응성 장애, 공포증, 및 외상후 스트레스 장애가 포함되나 이에 제한되는 것은 아니다. 우울 장애에는 주요 우울 장애 (MDD), 긴장성 우울증, 우울성 우울증, 비정형 우울증, 정신병적 우울증, 산후 우울증, 양극성 우울증 및 경증, 중등 또는 중증 우울증이 포함되나 이에 제한되는 것은 아니다. 인격 장애에는 편집, 반사회적 및 경계 인격 장애가 포함되나 이에 제한되는 것은 아니다.

[0041]

"바이오마커" 는 생물학적 샘플에 존재하는 또는 이로부터 유래되는, 사실상 임의의 검출가능한 화합물, 예컨대 단백질, 웨티드, 프로테오글리칸, 당단백질, 지단백질, 탄수화물, 지질, 핵산 (예를 들어, DNA, 예컨대 cDNA 또는 증폭된 DNA, 또는 RNA, 예컨대 mRNA), 유기 또는 무기 화학물질, 천연 또는 합성 중합체, 소형 분자 (예를 들어, 대사물질), 또는 상기 중 임의의 것의 식별 분자 또는 식별 조각, 또는 정상적인 생물학적 과정, 발병 과정 또는 치료 개입에 대한 약리학적 반응의 지표자로서 객관적으로 측정되고 평가되는 임의의 기타 특성, 또는 이의 지표이다. 문헌 [Atkinson, A.J., et al. Biomarkers and Surrogate Endpoints: Preferred

Definitions and Conceptual Framework, Clinical Pharm. & Therapeutics, 2001 March; 69(3): 89-95] 를 참조한다. 본 문맥에서 사용되는 바와 같은 "~로부터 유래되는" 이라는 구절은 검출되는 경우, 특정 분자가 생물학적 샘플에 존재한다는 지표인 화합물을 말한다. 예를 들어, 특정 cDNA 의 검출은 생물학적 샘플 내 특정 RNA 전사체의 존재의 지표일 수 있다. 또 다른 예로서, 특정 항체의 검출 또는 이에 대한 결합은 생물학적 샘플 내 특정 항원 (예를 들어, 단백질) 의 존재의 지표일 수 있다. 여기서, 식별 분자 또는 조각은 검출되는 경우, 상기 확인된 화합물의 존재 또는 풍부함을 나타내는 분자 또는 조각이다.

[0042] 바이오마커는 예를 들어, 생물학적 샘플로부터 단리되고, 생물학적 샘플에서 직접 측정되고, 또는 생물학적 샘플에서 검출되거나 샘플 내에 있는 것으로 측정될 수 있다. 바이오마커는 예를 들어, 기능적, 부분적으로 기능적, 또는 비-기능적일 수 있다. 하나의 구현예에서, 바이오마커는 예를 들어, 다양한 진단 어세이에서 바이오마커 검출을 용이하게 할 수 있는 특이적으로 결합하는 항체를 생성하기 위해 단리되고 사용된다. 임의의 면역어세이는 바이오마커 분자 (예를 들어, Fab, F(ab')<sub>2</sub>, Fv, 또는 scFv 조각) 에 결합할 수 있는 임의의 항체, 항체 조각 또는 이의 유도체를 사용할 수 있다. 이러한 면역어세이는 당업계에 잘 알려져 있다. 또한, 바이오마커가 단백질 또는 이의 조각인 경우, 잘 성립된 기술을 사용하여 이것은 서열분석될 수 있고, 이의 코딩 유전자는 클로닝될 수 있다.

[0043] 본원에서 사용되는 바와 같은, "바이오마커의 종" 이라는 용어는 본원에 기재된 바이오마커의 임의의 식별 부분 또는 식별 조각, 예컨대 본원에 기재된 특정 유전자 (예를 들어, 이하 표 1A 에 열거된 유전자) 의 스플라이싱 변이체를 말한다. 여기서, 식별 부분 또는 식별 조각은 검출되는 경우, 상기 확인된 전사체, cDNA, 증폭된 핵산, 또는 단백질의 존재 또는 풍부함을 나타내는 분자의 부분 또는 조각이다.

[0044] "바이오마커 프로파일" 은 바이오마커의 특징, 예컨대 측정가능한 양상 (예를 들어, 풍부함) 와 함께, 다수의 하나 이상의 종류의 바이오마커 (예를 들어, mRNA 분자, cDNA 분자, 단백질 및/또는 탄수화물, 또는 이의 지표 등) 를 포함한다. 바이오마커 프로파일은 2 개 이상의 이러한 바이오마커를 포함하며, 바이오마커는 동일 또는 상이한 부류, 예컨대, 예를 들어, 핵산 및 탄수화물일 수 있다. 바이오마커 프로파일은 또한 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 또는 100 개 이상, 또는 그 초과의 바이오마커를 포함할 수 있다. 하나의 구현예에서, 바이오마커 프로파일은 수백, 또는 심지어 수천 개의 바이오마커를 포함한다. 바이오마커 프로파일은 하나 이상의 대조군 또는 내부 표준을 추가로 포함할 수 있다. 하나의 구현예에서, 바이오마커 프로파일은 내부 표준으로서 담당하는 하나 이상의 바이오마커를 포함한다. 본 문맥에서 본원에서 사용되는 바와 같은 "지표" 라는 용어는 단지, 바이오마커 프로파일이 핵산, mRNA 분자, cDNA 분자, 단백질 및/또는 탄수화물, 또는 바이오마커 분자 실제 그 자체 이외의 임의의 다른 형태의 바이오마커에 대한 기호, 데이터, 약어 또는 기타 유사한 지표를 함유하는 상황을 말한다. 예를 들어, 본 발명의 예시적 바이오마커 프로파일은 표 1A 에 있는 유전자의 명칭을 포함한다.

[0045] 바이오마커 프로파일 내의 각각의 바이오마커에는 상응하는 "특징" 이 포함된다. 본원에서 사용되는 바와 같은 "특징" 이라는 용어는 바이오마커의 측정가능한 양상을 말한다. 특징에는 예를 들어, 예시적 바이오마커 프로파일 1 에 설명된 바와 같은 대상으로부터의 생물학적 샘플 내 바이오마커의 존재 또는 부재가 포함될 수 있다:

#### 예시적 바이오마커 프로파일 1

|             |               |
|-------------|---------------|
| 바이오마커       | 특징<br>샘플 내 존재 |
| 유전자 A 의 전사체 | 존재            |
| 유전자 B 의 전사체 | 부재            |

[0046]

[0047] 예시적 바이오마커 프로파일 1 에서, 유전자 A 의 전사체에 대한 특징 값은 "존재" 이고, 유전자 B 의 전사체에 대한 특징 값은 "부재" 이다.

[0048] 특징에는 예를 들어, 예시적 바이오마커 프로파일 2 에 설명된 바와 같은 대상으로부터의 생물학적 샘플 내 바이오마커의 풍부함이 포함될 수 있다:

## 예시적 바이오마커 프로파일 2

|             |                         |
|-------------|-------------------------|
| 바이오마커       | 특징<br>샘플 내 풍부함 (상대적 단위) |
| 유전자 A 의 전사체 | 300                     |
| 유전자 B 의 전사체 | 400                     |

[0049]

[0050] 예시적 바이오마커 프로파일 2에서, 유전자 A의 전사체에 대한 특징 값은 300 단위이고, 유전자 B의 전사체에 대한 특징 값은 400 단위이다.

[0051]

특징은 또한 예시적 바이오마커 프로파일 3에 설명된 바와 같은 바이오마커의 2 개 이상의 측정가능한 양상의 비일 수 있다:

## 예시적 바이오마커 프로파일 3

|             |  |
|-------------|--|
| 바이오마커       | 특징<br>유전자 A 의 전사체/유전자 B 의 전사체의<br>풍부함의 비 |
| 유전자 A 의 전사체 | 300                                      |
| 유전자 B 의 전사체 | 400                                      |

[0052]

[0053] 예시적 바이오마커 프로파일 3에서, 유전자 A의 전사체에 대한 특징 값 및 유전자 B의 전사체에 대한 특징 값은 0.75 (300/400) 이다.

[0054]

일부 구현예에서, 상기 예시적 바이오마커 프로파일 1에 설명된 바와 같이 바이오마커 프로파일에서 특징과 바이오마커 사이의 1 대 1 상응성이 있다. 일부 구현예에서, 상기 예시적 바이오마커 프로파일 3에 설명된 바와 같이 본 발명의 바이오마커 프로파일에서 특징과 바이오마커 사이의 관계는 더욱 복잡하다.

[0055]

당업자는 특징의 다른 계산 방법이 고안될 수 있고 모든 이러한 방법은 본 발명의 범주 내에 있다는 것을 인지 할 것이다. 예를 들어, 특징은 2 회 이상의 시점에서 대상으로부터 수집된 생물학적 샘플에 걸친 바이오마커의 풍부함의 평균을 나타낼 수 있다. 게다가, 특징은 단일 시점에서 대상으로부터 수득된 생물학적 샘플로부터 2 개 이상의 바이오마커의 풍부함의 비 또는 차이일 수 있다. 바이오마커 프로파일은 또한 2, 3, 4, 5, 10, 20, 30 개 이상, 또는 그 초과의 특징을 포함할 수 있다. 하나의 구현예에서, 바이오마커 프로파일은 수백, 또는 심지어 수천 개의 특징을 포함한다.

[0056]

일부 구현예에서, 바이오마커의 특징은 정량적 PCR (qPCR) 을 사용하여 측정된다. 유전자 전사체 풍부함을 측정하기 위한 qPCR의 사용은 잘 알려져 있다. 일부 구현예에서, 바이오마커의 특징은 마이크로어레이를 사용하여 측정된다. 풍부함 데이터를 수득하기 위한 마이크로어레이의 구축 및 마이크로어레이를 프로세싱 하는 데 사용되는 기술은 잘 알려져 있고, 예를 들어, 문헌 [Draghici, 2003, *Data Analysis Tools for DNA Microarrays*, Chapman & Hall/CRC], 및 국제 공개 번호 WO 03/061564 호에 기재되어 있다. 마이크로어레이 는 다수의 탐침을 포함한다. 일부 예에서, 각각의 탐침은 상이한 바이오마커를 인지, 예를 들어 이에 결합 한다. 일부 예에서, 마이크로어레이 상의 2 개 이상의 상이한 탐침은 동일한 바이오마커를 인지, 예를 들어 이에 결합한다. 그러므로, 전형적으로는, 마이크로어레이 상의 탐침 스팟과 대상 바이오마커 사이의 관계는 2 대 1 상응성, 3 대 1 상응성, 또는 일부 다른 형태의 상응성이다. 그러나, 이것은 마이크로어레이 상의 탐침과 바이오마커 사이의 독특한 1 대 1 상응성이 있는 경우일 수 있다.

[0057]

핵산 서열 (예를 들어, 본원에 기재된 유전자를 코딩하는 뉴클레오티드 서열)의 문맥에서 본원에 사용되는 바와 같은 "상보적"이라는 용어는, 수소 결합 특성의 결과로서 특이적인 질소 염기들 사이의 화학적 친화성을 말 한다. 예를 들어, 구아닌 (G) 은 오직 시토신 (C) 과 수소 결합을 형성하는 반면, 아데닌은 DNA의 경우 오직 티민 (T) 과, RNA의 경우 오직 우라실 (U) 과 수소 결합을 형성한다. 상기 반응은 염기 짹짓기로서 설명되며, 상기 짹지워진 염기 (G 와 C, 또는 A 와 T/U) 가 상보적이라고 불려진다. 그러므로, 2 개의 핵산 서열은 이들의 질소 염기가 수소 결합을 형성할 수 있다면 상보적일 수 있다. 이러한 서열은 서로 "상보체"로서 언급된다. 이러한 상보체 서열은 자연적으로 발생할 수 있고 또는, 이들은 예를 들어, DNA 분자 또는 RNA 분자 (예를 들어, mRNA 전사체)의 센스 가닥에 상보적인 안티센스 핵산 분자의 경우에서와 같이 당업자에게 공지된 임의의 방법에 의해 화학적으로 합성될 수 있다. 예를 들어, 문헌 [Lewin, 2002, *Genes VII*. Oxford University Press Inc., New York, NY] 을 참조한다.

[0058]

본원에서 사용되는 바와 같은, "데이터 분석 알고리즘"은 훈련 집단 내 대상의 바이오마커 프로파일을 사용하는 판단 규칙을 구축하기 위해 사용되는 알고리즘이다. 대표적인 데이터 분석 알고리즘이 하기에 기술된다.

"판단 규칙"은 데이터 분석 알고리즘의 최종 생성물이고, 하나 이상의 값 세트를 특징으로 하고, 상기 값 세트의 각각은 정동 장애의 양상, 정동 장애 발병, 대상에게 정동 장애가 있을 것이라는 예측, 또는 대상이 정동 장애의 증상을 나타내는 가망성의 지표이다. 하나의 특정 구현예에서, 값 세트는 대상이 정동 장애가 전개될 것이라는 예측을 나타낸다. 또 다른 예에서, 값 세트는 대상이 정동 장애가 전개되지 않을 것이라는 예측을 나타낸다.

[0059]

"판단 규칙"은 바이오마커 프로파일을 평가하기 위해 사용되는 방법이다. 이러한 판단 규칙은 문헌 [Hastie et al., 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York]에 예시된 바와 같이 당업계에 공지된 하나 이상의 형태를 취할 수 있다. 판단 규칙은 특히, 정동 장애의 존재, 또는 대상이 정동 장애의 증상을 나타내는 가망성, 또는 정동 장애가 전개될 감수성을 나타내는 가망성을 예측하기 위한 특징의 데이터 세트에 작용하기 위해 사용될 수 있다. 본 발명의 일부 구현예에서 사용될 수 있는 예시적 판단 규칙은 하기에 더욱 상세히 기재된다.

[0060]

본원에 사용되는 바와 같은, "중간표현형 (endophenotype)"이라는 용어는 개인에게 증상이 있는지의 여부와 관계 없이 특성이 존재하는 질병과 연관된 유전성 특성, 예컨대 바이오마커를 의미할 것이다 (리뷰를 위해, Lenox et al., 2002, *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 114:391-406 참조).

[0061]

본원에 사용되는 바와 같은, "유전자 발현 프로파일" 및 "전사 프로파일"이라는 용어는 선택된 유전자의 메신저 리보핵산 (mRNA) 수준의 상대적인 측정값에 의해 측정된 바이오마커 프로파일이다. 전사 프로파일은 대상 또는 환자의 생물학적 샘플로부터 유전자의 전사 분석에 의해 측정된다.

[0062]

본원에 사용되는 바와 같은, "건강한 대조군 대상", "건강한 대조군", 및 "대조군 대상"은 주요한 현재의 의료적 또는 정신적 문제가 없으나, 예를 들어, 두통은 앓을 수 있는 대상을 의미할 것이다. 대조군 대상은 바람직하게는 낮은 체질량 지수 (BMI, 30 미만)를 갖고, 지난 3 개월 동안 약물 사용을 하지 않으며, 적은 또는 0의 스트레스 점수, 가족력 점수, 및 증상 점수를 갖는다. 대조군 대상은 자가 관리 질문에 의해 측정된 바와 같은 임의의 정신적 질환 이력, 임의의 약물 남용 이력, 임의의 정신적 질환 가족력, 임의의 어린 시절 스트레스 요인 또는 임의의 최근의 스트레스 요인이 없을 수 있다. 대조군 대상은 생물학적 샘플을 수득하기 전 담당의에 의해 추가로 평가될 필요는 없을 것이다.

[0063]

본원에 사용되는 바와 같은, "수득하다" 및 "수득하는"이라는 용어는, 각각 "~의 소유가 된다" 또는 "~의 소유가 되는"을 의미한다. 이것은 예를 들어, 컴퓨터 시스템 내 데이터 저장물로부터 데이터를 검색함으로써 수행될 수 있다. 이것은 또한 예를 들어, 직접 측정에 의해 수행될 수 있다.

[0064]

본원에 사용되는 바와 같은, "표현형"이라는 용어는 대상의 유전자형 및 환경의 결과인, 측정가능한 및/또는 관찰가능한 생물학적, 임상적 또는 행동적 특성을 의미할 것이다.

[0065]

본원에 사용되는 바와 같은, "단백질", "펩티드", 및 "폴리펩티드"라는 용어는 다르게 표시되지 않는다면 상호 교환적이다.

[0066]

본원에 사용되는 바와 같은, "PTSD 대조군 대상"은 극한 트라우마성 스트레스 요인에 적용되지 않고, 임의의 신경정신적 질환이 없는 것으로 담당의에 의해 평가된 대상을 의미할 것이다. 본 발명의 PTSD 대조군 대상은 일반적으로 대상, 예를 들어, 동일한 지리학적 지역으로부터의 대상 및 질환을 나타내는 대상과 동일한 성별의 대상과 매치된다.

[0067]

본원에 사용되는 바와 같은, "특이적으로"라는 용어 및 유사한 용어는 항체의 문맥에서, 항원 또는 조각에 특이적으로 결합하고, 다른 항원 또는 다른 조각에 특이적으로 결합하지 않는 웨პ티드, 폴리웨პ티드, 및 항체 또는 이의 조각을 말한다. 항원에 특이적으로 결합하는 웨პ티드 또는 폴리웨პ티드는 표준 실험 기술에 의해, 예를 들어, 당업자에게 잘 알려진 임의의 면역어세이에 의해 측정되는 바와 같이 낮은 친화성으로 다른 웨პ티드 또는 폴리웨პ티드에 결합할 수 있다. 이러한 면역어세이에는 방사능면역어세이 (RIA) 및 효소 연결 면역흡착 어세이 (ELISA)가 포함되나 이에 제한되지 않는다. 항원에 특이적으로 결합하는 항체 또는 조각은 관련 항원과 교차 반응성일 수 있다. 바람직하게는, 항원에 특이적으로 결합하는 항체 또는 이의 조각은 다른 항원과 교차 반응하지 않는다. 항원-항체 상호작용, 특이성 및 교차 반응성, 및 상기 모두의 측정 방법과 관련된 논의에 대해서는 예를 들어, 문헌 [Paul, ed., 2003, *Fundamental Immunology*, 5th ed., Raven Press, New York, 페이지 69-105]를 참조한다.

[0068] 본원에 사용되는 바와 같은, "대상"은 동물, 바람직하게는 포유류, 더욱 바람직하게는 비-인간 영장류, 가장 바람직하게는 인간이다. "대상," "개체," "후보자," 및 "환자"라는 용어는 본원에서 상호교환적으로 사용된다. 일부 구현예에서, 대상은 동물이다. 다른 구현예에서, 대상은 포유류이다.

[0069] 본원에 사용되는 바와 같은, "시험 대상"은 전형적으로, 판단 규칙을 구축하는데 사용되는 훈련 집단이 아닌 임의의 대상이다. 시험 대상은 임의로, 정동 장애 또는 정동 장애가 전개될 가능성을 갖는 것으로 의심될 수 있다.

[0070] 본원에 사용되는 바와 같은, "훈련 집단"은 정동 장애를 가질 위험이 있는 대상의 바이오마커 프로파일을 평가하기 위해 데이터 분석 알고리즘을 사용하는 판단 규칙을 구축하는데 사용되는 대상의 집단으로부터의 샘플 세트이다. 바람직한 구현예에서, 훈련 집단에는 정동 장애를 갖는 대상 및 정동 장애를 갖지 않는 대상으로부터의 샘플이 포함된다.

[0071] 본원에 사용되는 바와 같은, "검증 집단"은 판단 규칙의 정확성, 또는 다른 성능 미터를 측정하는데 사용되는 대상의 집단으로부터의 샘플의 세트이다. 바람직한 구현예에서, 검증 집단에는 정동 장애를 갖는 대상 및 정동 장애를 갖지 않는 대상으로부터의 샘플이 포함된다. 바람직한 구현예에서, 검증 집단에는 정확성, 또는 다른 성능 미터가 추구되는 판단 규칙을 훈련하는데 사용되는 훈련 집단의 일부인 대상이 포함되지 않는다.

[0072] 본원에 사용되는 바와 같은, "값 세트"는 바이오마커 프로파일 내 특징에 대한 값의 조합, 또는 값의 범위이다. 상기 값 세트의 특성 및 그곳의 값은 값 세트를 구슬하는 판단 규칙을 구축하는데 사용되는 데이터 분석 알고리즘 및 바이오마커 프로파일에 존재하는 특징의 유형에 따라 다르다. 설명을 위해, 예시적 바이오마커 프로파일 2를 다시 고려한다:

### 예시적 바이오마커 프로파일 2

|            |                         |
|------------|-------------------------|
| 바이오마커      | 특정<br>샘플 내 풍부함 (상대적 단위) |
| 유전자 A의 전사체 | 300                     |
| 유전자 B의 전사체 | 400                     |

[0073]

[0074] 본 예에서, 훈련 집단의 각각의 일원의 바이오마커 프로파일이 수득된다. 각각의 이러한 바이오마커 프로파일에는 유전자 A의 전사체에 대해 측정된 특징(여기서는 풍부함), 및 유전자 B의 전사체에 대해 측정된 특징(여기서는 풍부함)이 포함된다. 상기 특징 값(여기서는 풍부함 값)이 판단 규칙을 구축하기 위해 데이터 분석 알고리즘에 의해 사용된다. 본 예에서, 데이터 분석 알고리즘은 하기 기재된 결정 트리이고, 상기 데이터 분석 알고리즘의 최종 생성물인 판단 규칙은 결정 트리이다. 판단 규칙이 값 세트를 규정한다. 하나의 이러한 값 세트가 정동 장애의 예측값이다. 바이오마커 특징 값이 상기 값 세트를 만족시키는 대상은 정동 장애를 갖는다. 상기 부류의 예시적 값 세트는 예시적 값 세트 1이다:

### 예시적 값 세트 1

|            |                                |
|------------|--------------------------------|
| 바이오마커      | 값 세트 성분<br>(샘플 내 풍부함 (상대적 단위)) |
| 유전자 A의 전사체 | <400                           |
| 유전자 B의 전사체 | <600                           |

[0075]

[0076] 또 다른 이러한 값 세트는 정동 장애가 없는 상태의 예측값이다. 대상의 바이오마커 특징 값이 상기 값 세트를 만족시키는 대상은 정동 장애를 갖는 것으로 진단되지 않는다. 상기 부류의 예시적 값 세트는 예시적 값 세트 2이다:

### 예시적 값 세트 2

|            |                                |
|------------|--------------------------------|
| 바이오마커      | 값 세트 성분<br>(샘플 내 풍부함 (상대적 단위)) |
| 유전자 A의 전사체 | >400                           |
| 유전자 B의 전사체 | >600                           |

[0077]

[0078] 데이터 분석 알고리즘이 뉴럴 네트워크 분석이고 상기 뉴럴 네트워크 분석의 최종 생성물이 적합하게 가중치된

뉴럴 네트워크인 경우, 하나의 값 세트는 대상이 정동 장애를 갖는 것을 나타내는 가중치된 뉴럴 네트워크를 야기할 바이오마커 프로파일 특정 값의 범위이다. 또 다른 값 세트는 대상이 정동 장애를 갖지 않는 것을 나타내는 가중치된 뉴럴 네트워크를 야기할 바이오마커 프로파일 특정 값의 범위이다.

[0079]

본원에 사용되는 바와 같은, 마이크로어레이의 문맥에서 "탐침 스팟"이라는 용어는 샘플 내 특정 핵산의 풍부함을 측정하는데 사용되는 본원에서 "탐침"으로서 불리는 단일 가닥 DNA 분자 (예를 들어, 단일 가닥 cDNA 분자 또는 합성 DNA 올리고머)를 말한다. 예를 들어, 탐침 스팟은 시험 대상으로부터 생물학적 샘플 (예를 들어, 세포의 수집물) 내 mRNA의 수준을 측정하기 위해 사용될 수 있다. 구체적인 구현예에서, 전형적인 마이크로어레이에는 격자 상의 공지된 위치에 유리 슬라이드 (또는 다른 기판) 상에 놓여진 복합 탐침 스팟을 포함한다. 각각의 탐침 스팟에 대한 핵산은 유전자의 서열 또는 관심의 유전자의 단일 가닥 인접 부분 (예를 들어, 10 랑체, 11 랑체, 12 랑체, 13 랑체, 14 랑체, 15 랑체, 16 랑체, 17 랑체, 18 랑체, 19 랑체, 20 랑체, 21 랑체, 22 랑체, 23 랑체, 24 랑체, 25 랑체 또는 그 초과)이고, 특정 유전자 또는 관심의 유전자에 의해 코딩되는 mRNA에 대한 탐침이다. 각각의 탐침 스팟은 단일 핵산 서열을 특징으로 하고, 오직 그의 상보성 DNA 가닥 또는 mRNA 분자에만 혼성화하도록 하는 조건하에서 혼성화된다. 이와 같이, 기판 상에 많은 탐침 스팟이 있을 수 있고, 각각은 독특한 유전자 또는 관심의 서열을 나타낼 수 있다. 또한, 2 개 이상의 탐침 스팟이 동일한 유전자 서열을 나타낼 수 있다. 일부 구현예에서, 표지된 핵 샘플은 탐침 스팟에 혼성화되고, 탐침 스팟에 특이적으로 혼성화되는 표지된 핵산의 양은 특정 생물학적 샘플 내 특이적 핵산 (예를 들어, 특정 유전자의 mRNA 전사체)의 수준을 측정하기 위해 정량화될 수 있다. 탐침, 탐침 스팟, 및 마이크로어레이는 일반적으로 문헌 [Draghici, 2003, *Data Analysis Tools for DNA Microarrays*, Chapman & Hall/CRC, Chapter, 2]에 기재되어 있다.

[0080]

## 5.2 대상 스크리닝 방법

[0081]

본 발명은 개체로부터의 생물학적 샘플 내 정동 장애를 갖는 것으로 의심되는 시험 개체의 바이오마커 프로파일의 2 개 이상의 특징의 검출을 통해 정동 장애의 정확하고, 빠른 예측 및/또는 진단을 가능하게 한다.

[0082]

본 발명의 특정 구현예에서, 정동 장애를 갖는 것으로 의심되는 대상을 본 발명의 방법을 사용하여 스크리닝한다. 상기 구현예에 따르면, 본 발명의 방법은 예를 들어, 정신과 병동에 입원할 대상 및/또는 일종의 심리학적 트라우마를 경험하는 대상을 스크리닝하는데 사용될 수 있다.

[0083]

특정 구현예에서, 예를 들어 혈액과 같은 생물학적 샘플이 채취된다. 일부 구현예에서, 생물학적 샘플은 혈액, 뇌척수액, 복막액, 간질액, 적혈구 세포, 백혈구 세포 또는 혈소판이다. 백혈구 세포 (백혈구)에는 중성구, 호염기구, 호산구, 림프구, 단핵구 및 대식 세포가 포함되나 이에 제한되지 않는다. 일부 구현예에서 생물학적 샘플은 전혈의 일부 성분이다. 하나의 구현예에서, 본 발명은 RNA 안정화제 또는 방부제를 함유하는 즉석 사용 수집튜브로 전혈 샘플링을 이용한다. 이러한 프로토콜은 입증되고, 적절한 샘플 취급 과정이 뒤따른다면 매우 적은 가변성을 보장한다. 본 발명은 대규모 샘플 세트에 대한 고 처리량 분석에서 사용될 수 있는 믿을 수 있고 확고한 전사 마커를 제공한다. 이러한 믿을 수 있는 방법은 대조군과 환자를 식별하는데 나타난다. 일부 구현예에서 혈액의 세포 분획 내 또는 액체 (예를 들어, 혈장 또는 혈청 분획) 내의 단백질, 핵산 및/또는 기타 분자 (예를 들어, 대사 산물)의 혼합물의 일부분이 바이오마커 프로파일로서 분석된다. 이는 바이오마커 프로파일 내 바이오마커의 특징을 측정함으로써 달성될 수 있다. 일부 구현예에서, 생물학적 샘플은 전혈이나, 바이오마커 프로파일은 발현된 바이오마커로부터 분석되거나 다르게는 전혈로부터 단리된 백혈구 세포에서 발견된다. 일부 구현예에서, 생물학적 샘플은 전혈이나, 바이오마커 프로파일은 발현된 바이오마커로부터 분석되거나 다르게는 전혈로부터 단리된 적혈구 세포에서 발견된다.

[0084]

바이오마커 프로파일은 2 개 이상의 바이오마커를 포함할 수 있으며, 바이오마커는 동일하거나 상이한 부류, 예를 들어 핵산 및 탄수화물일 수 있다. 일부 구현예에서, 바이오마커 프로파일은 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 96, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195 또는 200 개 이상, 또는 그 초과의 바이오마커를 포함한다. 하나의 구현예에서, 바이오마커 프로파일은 수백, 또는 심지어 수천 개의 바이오마커를 포함한다. 일부 구현예에서, 바이오마커 프로파일은 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 개 이상, 또는 그 초과의 바이오마커를 포함한다. 하나의 예에서, 일부 구현예의 경우 바이오마커 프로파일은 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 개 이상, 또는 그 초과의 표 1A로부터 선택되는 바이오마커를 포함한다.

[0085]

전형적인 구현예에서, 바이오마커 프로파일 내 각각의 바이오마커는 특징으로 나타내어진다. 다시 말해서, 바이오마커와 특정 사이의 상응성이 있다. 일부 구현예에서, 바이오마커와 특정 사이의 상응성을 1:1 이고, 이는 각각의 바이오마커에 대해 특징이 있다는 것을 의미한다. 일부 구현예에서, 각각의 바이오마커에 대해 1 개 초과의 특징이 있다. 일부 구현예에서, 바이오마커 프로파일 내 1 개의 바이오마커에 상응하는 특징의 수는 바이오마커 프로파일 내 또다른 바이오마커에 상응하는 특징의 수와 상이하다. 그러므로, 바이오마커 프로파일은 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 96, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195 또는 200 개 이상, 또는 그 초과의 특징을 포함할 수 있고, 단 바이오마커 프로파일 중 2, 3, 4, 5, 6 또는 7 개 이상, 또는 그 초과의 바이오마커가 있다. 일부 구현예에서, 바이오마커 프로파일은 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 개 이상, 또는 그 초과의 특징을 포함할 수 있다. 구현예와 상관 없이, 이러한 특징은 임의의 재현가능한 측정 기술 또는 측정 기술의 조합의 사용을 통해 측정될 수 있다. 이러한 기술에는 본원에 기재된 임의의 기술을 비롯하여 당업계에 잘 공지된 기술, 또는 예를 들어 하기 섹션 5.4 에 기재된 임의의 기술이 포함된다. 전형적으로, 이러한 기술은 시간의 단일 지점에서 대상으로부터 취득된 생물학적 샘플 또는 시간의 다중 지점에서 선택된 다중 샘플을 사용하여 특징 값을 측정하는데 사용된다. 하나의 구현예에서, 대상으로부터 취득된 샘플로부터 바이오마커 프로파일을 수득하기 위한 예시적인 기술은 cDNA 마이크로어레이 (예를 들어 하기 섹션 5.4.1.2 참조)이다. 또다른 구현예에서, 대상으로부터 취득된 샘플로부터 바이오마커 프로파일을 수득하기 위한 예시적인 기술은 문현 [BD Cytometric Bead Array (CBA) Human Inflammation Kit Instruction Manual (BD Biosciences)]에 기재된 바와 같은 단백질-기반 어세이 또는 단백질-기반 기술의 다른 형태 또는 미국 특허 제 5,981,180 호에 기재된 비이드 (bead) 어세이이고, 이를 각각은 전체, 특히 생물학적 샘플 내 단백질 농도의 다양한 어세이 방법의 교시가 본원에 참조로서 인용된다. 게다가 또다른 구현예에서, 바이오마커 프로파일은 혼합되며, 이것이 핵산인 일부 바이오마커 또는 이의 지표, 및 단백질인 일부 바이오마커 또는 이의 지표를 포함한다는 것을 의미한다. 이러한 구현예에서, 단백질-기반 및 핵산-기반 기술은 모두, 대상으로부터 취득된 하나 이상의 샘플로부터 바이오마커 프로파일을 수득하기 위해 사용된다. 다시 말해서, 핵산인 바이오마커 프로파일에서 바이오마커와 관련된 특징에 대한 특징 값은 핵산-기반 측정 기술 (예를 들어, 핵산 마이크로어레이)에 의해 수득되고, 단백질인 바이오마커 프로파일에서 바이오마커와 관련된 특징에 대한 특징 값은 단백질-기반 측정 기술에 의해 수득된다. 일부 구현예에서, 바이오마커 프로파일은 하기 섹션 5.3 에 기재된 키트와 같은 키트를 사용하여 수득될 수 있다.

[0086]

### 5.3 키트

[0087]

본 발명은 또한 대상에서 정동 장애를 진단하는데 유용한 키트를 제공한다. 일부 구현예에서, 본 발명의 키트는 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 96, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195 또는 200 개 이상, 또는 그 초과의 바이오마커 및/또는 이러한 바이오마커의 존재 또는 풍부함을 검출하기 위한 시약을 포함한다. 다른 구현예에서, 본 발명의 키트는 2 개 이상, 그러나 수백 또는 그 이상의 많은 바이오마커를 포함한다. 일부 구현예에서, 본 발명의 키트는 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 개 이상, 또는 그 초과의 표 1A로부터 선택되는 바이오마커, 또는 이러한 바이오마커의 존재 또는 풍부함을 검출하기 위한 시약을 포함한다. 섹션 5.1 에 제공된 바이오마커의 정의에 따르면 일부 예에서, 바이오마커는 실제로, 예를 들어 유전자, mRNA 또는 단백질 그 자체 이외의 유전자, mRNA 또는 단백질의 식별 분자이다. 따라서, 바이오마커는 실제의 유전자, mRNA 또는 단백질 그 자체 이외에 표 1A에서 확인되는 특정 유전자, mRNA 또는 단백질, 또는 이의 조각의 존재 또는 풍부함을 나타내는 분자일 수 있다. 일부 구현예에서, 본 발명의 키트는 2 개 이상, 그러나 수백 또는 그 이상의 많은 바이오마커를 포함한다. 일부 구현예에서, 25% 이상, 30% 이상, 35% 이상, 40% 이상, 60% 이상, 80% 이상의 바이오마커 및/또는 바이오마커의 존재 또는 풍부함을 검출하기 위한 시약은 표 1A로부터의 바이오마커 및/또는 표 1A로부터 선택되는 바이오마커의 존재 또는 풍부함을 검출하기 위한 시약으로부터 선택된다.

[0088]

본 발명의 키트의 바이오마커는 본 발명에 따른 바이오마커 프로파일을 생성하는데 사용될 수 있다. 키트의 화합물의 부류의 예에는 단백질 및 이의 조각, 웨티드, 프로테오글리칸, 당단백질, 지단백질, 탄수화물, 지질, 핵산 (예를 들어, DNA, 예컨대 cDNA 또는 증폭된 DNA, 또는 RNA, 예컨대 mRNA), 유기 또는 무기 화학물질, 천연 또는 합성 중합체, 소형 분자 (예를 들어, 대사물질), 또는 상기 중 임의의 것의 식별 분자 또는 식별 조각이 포함되나 이에 제한되는 것은 아니다. 특정 구현예에서, 바이오마커는 특정의 크기 (예를 들어 10, 15, 20,

25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 1000, 2000, 3000, 5000, 10k, 20k, 100k 달톤 이상, 또는 그 초과) 의 것이다. 바이오마커(들) 은 어레이의 일부일 수 있거나, 바이오마커(들) 은 개별적으로 및/또는 독립적으로 패키지될 수 있다. 키트는 또한 본 발명의 바이오마커 프로파일을 생성하는데 사용될 하나 이상의 내부 표준을 포함할 수 있다. 마찬가지로, 내부 표준 또는 표준은 상기 기재된 화합물의 중 임의의 부류일 수 있다.

[0089] 하나의 구현예에서, 본 발명은 예를 들어 마이크로어레이에서 발견되는 것과 같이 기판 상의 지정 가능한 위치에 고정되거나 고정되지 않을 수 있는 탐침 및/또는 프라이머를 포함하는 키트를 제공한다. 특정 구현예에서, 본 발명은 이러한 마이크로어레이를 제공한다.

[0090] 본 발명의 일부 구현예에서, 키트는 앱타머와 같은 특이적 바이오마커 결합 성분을 포함할 수 있다. 바이오마커가 핵산을 포함하는 경우, 키트는 바이오마커 또는 바이오마커의 상보적 가닥과 중복 부위를 형성할 수 있는 올리고뉴클레오티드 탐침을 제공할 수 있다. 올리고뉴클레오티드 탐침은 검출 가능하게 표지될 수 있다. 이러한 구현예에서, 탐침은 본 발명의 범주 내에 있는 바이오마커 그 자체이다.

[0091] 본 발명의 키트에는 또한 바이오마커 프로파일을 구축하는데 사용될 수 있는 완충액과 같은 부가적인 조성물이 포함될 수 있다. 미생물의 작용 예방은 다양한 항박테리아제 및 항진균제, 예를 들어, 파라벤, 클로로부탄올, 폐놀 소르브산 등의 포함에 의해 확보될 수 있다. 당, 염화나트륨 등과 같은 등장성체를 포함하는 것이 바람직할 수 있다.

[0092] 본 발명의 일부 키트는 마이크로어레이를 포함한다. 하나의 구현예에서, 상기 마이크로어레이는 다수의 탐침 스팟을 포함하고, 다수의 탐침 스팟 내 20% 이상의 탐침 스팟이 표 1A 의 바이오마커에 상응한다. 일부 구현예에서, 다수의 탐침 스팟 내 25% 이상, 30% 이상, 35% 이상, 40% 이상, 60% 이상, 또는 80% 이상의 탐침 스팟이 표 1A 의 바이오마커 및/또는 표 1A 의 바이오마커의 존재 또는 풍부함을 검출하기 위한 시약에 상응한다. 이러한 탐침 스팟은 본 발명의 범주 내의 바이오마커이다. 일부 구현예에서, 마이크로어레이는 기판 상에 약 2 내지 약 100 개의 탐침 스팟으로 이루어진다. 일부 구현예에서, 마이크로어레이는 기판 상에 약 2 내지 약 100 개의 탐침 스팟으로 이루어진다. 본 문맥에서 사용되는 바와 같은, "약" 이라는 용어는 언급된 값의 5% 이내, 언급된 값의 10% 이내, 또는 언급된 값의 25% 이내를 의미한다. 일부 구현예에서, 이러한 마이크로어레이는 상호-마이크로어레이 검정곡선에 대한 또는 당업자에게 공지된 기술을 사용하는 참조 마이크로어레이와 같은 다른 마이크로어레이로의 검정곡선에 대한 하나 이상의 탐침 스팟을 함유한다. 일부 구현예에서, 이러한 마이크로어레이는 핵산 마이크로어레이이다. 일부 구현예에서, 이러한 마이크로어레이는 단백질 마이크로어레이이다.

[0093] 본 발명의 일부 키트는 컴퓨터 판독 가능 저장 매체에 저장된 컴퓨터 프로그램 메카니즘을 포함하는 컴퓨터 프로그램 제품으로서 시행된다. 또한, 본 발명의 임의의 방법은 하나 이상의 컴퓨터 또는 다른 형태의 장치에서 실행될 수 있다. 장치의 예에는 컴퓨터, 및 분광 측정 소자 (예를 들어, 마이크로어레이 판독기 또는 마이크로어레이 스캐너) 가 포함되나 이에 제한되는 것은 아니다. 더욱 또한, 본 발명의 임의의 방법은 하나 이상의 컴퓨터 프로그램 제품에서 실행될 수 있다. 본 발명의 일부 구현예는 본원에 기재된 방법 중 임의의 것 또는 모두를 코딩하는 컴퓨터 프로그램 제품을 제공한다. 이러한 방법은 CD-ROM, DVD, 마그네틱 디스크 저장 제품, 또는 임의의 기타 유형 (有形) 컴퓨터-판독 가능 데이터 또는 유형 프로그램 저장 제품에 저장될 수 있다. 이러한 방법은 또한 영구적 저장소, 예컨대 ROM, 하나 이상의 프로그램 가능 칩, 또는 하나 이상의 적용 특이적 통합 회로 (ASIC) 에 저장될 수 있다. 이러한 영구적 저장소는 서버, 802.11 접근 지점, 802.11 무선 브릿지/스테이션, 리피터, 라우터, 휴대폰, 또는 기타 전자 소자에 위치할 수 있다. 컴퓨터 프로그램 제품에 저장된 이러한 방법은 인터넷을 통해 전자적으로 또는 다르게 배포될 수 있다.

[0094] 본 발명의 일부 키트는 본 발명의 방법 중 임의의 것을 개별적으로 또는 집합적으로 수행하는 하나 이상의 프로그램을 함유하는 컴퓨터 프로그램 제품을 제공한다. 상기 프로그램 모듈은 CD-ROM, DVD, 마그네틱 디스크 저장 제품, 또는 임의의 기타 유형 컴퓨터-판독 가능 데이터 또는 프로그램 저장 제품에 저장될 수 있다. 프로그램 모듈은 또한 영구적 저장소, 예컨대 ROM, 하나 이상의 프로그래밍 가능 칩, 또는 하나 이상의 적용 특이적 통합 회로 (ASIC) 에 저장될 수 있다. 이러한 영구적 저장소는 서버, 802.11 접근 지점, 802.11 무선 브릿지/스테이션, 리피터, 라우터, 휴대폰, 또는 기타 전자 소자에 위치할 수 있다. 컴퓨터 프로그램 제품 내의 소프트웨어 모듈은 또한 인터넷을 통해 전자적으로 또는 다르게 배포될 수 있다.

[0095] 본 발명의 일부 키트는 하나 이상의 프로세싱 유닛을 갖는 컴퓨터 및 하나 이상의 프로세싱 유닛에 연결된 메모

리를 포함한다. 메모리는 정동 장애를 가질 위험이 있는 시험 대상의 바이오마커 프로파일 내 다수의 특징이 값 세트를 만족하는지의 여부를 평가하기 위한 지침을 저장한다. 일부 구현예에서, 값 세트를 만족하는 것은 대상이 정동 장애를 갖는 것으로서 진단한다. 일부 구현예에서, 값 세트를 만족하는 것은 대상이 정동 장애를 갖지 않는 것으로서 진단한다. 하나의 구현예에서, 다수의 특징은 표 1A에 열거된 바이오마커에 상응한다.

[0096] 도 1은 상기 기술된 기능을 지지하는 예시적 시스템을 제시한다. 시스템은 바람직하게는 하기를 갖는 컴퓨터 시스템 (10)이다:

[0097] · 중앙 프로세싱 유닛 (22);

[0098] · 소프트웨어 및 데이터를 저장하기 위한 주요 비-휘발성 저장 유닛 (14), 예를 들어, 하드 디스크 드라이브, 저장 유닛 (14)는 저장 제어기 (12)에 의해 조절됨;

[0099] · 비-휘발성 저장 유닛 (14)로부터 로딩된 데이터 및 프로그램을 포함하는 저장 시스템 조절 프로그램, 데이터, 및 적용 프로그램용의 시스템 메모리 (36), 바람직하게는 고속 랜덤-접근 메모리 (RAM); 시스템 메모리 (36)에는 또한 판독 전용 메모리 (ROM)가 포함될 수 있다;

[0100] · 하나 이상의 입력 소자 (예를 들어, 키보드 (28)) 및 디스플레이 (26) 또는 기타 출력 소자를 포함하는 사용자 인터페이스 (32);

[0101] · 임의의 유선 또는 무선 통신망 (34) (예를 들어, 인터넷과 같은 광역 네트워크)를 연결하기 위한 네트워크 인터페이스 카드 (20);

[0102] · 시스템의 상기 언급된 요소를 서로 연결하기 위한 내부 버스 (30); 및

[0103] · 상기 언급된 요소에 전력을 공급하기 위한 전원 (24).

[0104] 컴퓨터 (10)의 작동은 중앙 프로세싱 유닛 (22)에 의해 실행되는 운영 체계 (40)에 의해 주로 조절된다. 운영 체계 (40)은 시스템 메모리 (36)에 저장될 수 있다. 운영 체계 (40)에 더하여, 전형적인 실행에서, 시스템 메모리 (36)에는 하기가 포함된다:

[0105] · 본 발명에 의해 사용되는 데이터 구조 및 다양한 파일에 대한 접근을 조절하기 위한 파일 시스템 (42);

[0106] · 본 발명에 따른 하나 이상의 판단 규칙을 구축하는데 사용하기 위한 훈련 데이터 세트 (44);

[0107] · 훈련 데이터 프로세싱 및 판단 규칙 구축을 위한 데이터 분석 알고리즘 모듈 (54);

[0108] · 하나 이상의 판단 규칙 (56);

[0109] · 시험 대상의 바이오마커 프로파일 내 다수의 특징이 제 1 값 세트 또는 제 2 값 세트를 만족하는지의 여부를 측정하기 위한 바이오마커 프로파일 평가 모듈 (60);

[0110] · 바이오마커 (64) 및, 각각의 이러한 바이오마커에 대한 특징 (66)을 포함하는 시험 대상 바이오마커 프로파일 (62); 및

[0111] · 본 발명의 선택된 바이오마커 (예를 들어, 표 1A) 및/또는 상기 선택된 바이오마커 각각에 대한 하나 이상의 특징의 데이터베이스 (68).

[0112] 훈련 데이터 세트 (46)은 다수의 대상 (46)에 대한 데이터를 포함한다. 각각의 대상 (46)에 대해, 대상 확인자 (48) 및 다수의 바이오마커 (50)이 있다. 각각의 바이오마커 (50)에 대해, 하나 이상의 특징 (52)가 있다. 도 1에서는 제시되지 않았지만, 각각의 특징 (52)에 대해, 특징 값이 있다. 데이터 분석 알고리즘을 사용하여 구축된 각각의 판단 규칙 (56)에 대해, 하나 이상의 판단 규칙 값 세트 (58)이 있다.

[0113] 도 1에 설명된 바와 같이, 컴퓨터 (10)은 소프트웨어 프로그램 모듈 및 데이터 구조를 포함한다. 컴퓨터 (10)에 저장된 데이터 구조에는 훈련 데이터 세트 (44), 판단 규칙 (56), 시험 대상 바이오마커 프로파일 (62), 및 바이오마커 데이터베이스 (68)이 포함된다. 상기 데이터 구조 각각은 플랫 ASCII 또는 바이너리 파일, 엑셀 (Excel) 스프레드시트, 상관 데이터베이스 (SQL), 또는 온라인 분석 프로세싱 (OLAP) 데이터베이스 (MDX 및/또는 이의 변형체)를 포함하나 이에 제한되지 않는 임의의 형태의 데이터 저장 시스템을 포함할 수 있다. 일부 특정 구현예에서, 이러한 데이터 구조는 각각, 계급 구조를 포함하는 하나 이상의 데이터베이스의 형태 (예를 들어, 별형 스키마)로 있다. 일부 구현예에서, 이러한 데이터 구조는 각각, 명쾌한 계급을 갖

지 않는 데이터베이스의 형태 (예를 들어, 계급적으로 배열되지 않은 차원 표) 이다.

[0114] 일부 구현예에서, 시스템 (10)에 저장되거나 접근 가능한 각각의 데이터 구조는 단일 데이터 구조이다. 다른 구현예에서, 이러한 데이터 구조는 사실상 동일한 컴퓨터 (10)에 의해 호스트되거나 모두 호스트되지 않을 수 있는 다수의 데이터 구조 (예를 들어, 데이터베이스, 파일, 아르키브)를 포함한다. 예를 들어, 일부 구현예에서, 훈련 데이터 세트 (44)는 컴퓨터 (10) 및/또는 광역 네트워크 (34)를 통해 컴퓨터 (10)에 의해 다룰 수 있는 컴퓨터에 저장되는 다수의 엑셀 스프레드시트를 포함한다. 또 다른 예에서, 훈련 데이터 세트 (44)는 컴퓨터 (10)에 저장되거나 광역 네트워크 (34)를 통해 컴퓨터 (10)에 의해 다룰 수 있는 하나 이상의 컴퓨터에 분포된 데이터베이스를 포함한다.

[0115] 도 1에 설명된 많은 모듈 및 데이터 구조는 하나 이상의 원격 컴퓨터에 위치할 수 있다는 것이 명백할 것이다. 예를 들어, 본 발명의 일부 구현예는 웹 서비스 유형 실행이다. 이러한 구현예에서, 바이오마커 프로파일 평가 모듈 (60) 및/또는 기타 모듈은 네트워크 (34)를 통해 컴퓨터 (10)과 통신하는 클라이언트 컴퓨터에 위치할 수 있다. 일부 구현예에서, 예를 들어, 바이오마커 프로파일 평가 모듈 (60)은 상호작용 웹 페이지 일 수 있다.

[0116] 일부 구현예에서, 도 1에 설명된 훈련 데이터 세트 (44), 판단 규칙 (56), 및/또는 바이오마커 데이터베이스 (68)은 단일 컴퓨터 (컴퓨터 (10)) 상에 있고, 다른 구현예에서, 하나 이상의 이러한 데이터 구조 및 모듈은 하나 이상의 원격 컴퓨터 (제시되지는 않음)에 의해 호스트된다. 하나 이상의 컴퓨터 상의, 도 1에 설명된 데이터 구조 및 소프트웨어 모듈의 배열은 상기 데이터 구조 및 소프트웨어 모듈이 다른 전자 수단 또는 네트워크 34를 통해 서로에 대해 다룰 수 있는 한 본 발명의 범주 내에 있다. 그러므로, 본 발명은 컴퓨터 시스템의 광범위한 어레이를 완전히 포함한다.

[0117] 본 발명의 또 다른 구현예는 대상이 정동 장애를 갖는 거의 여부를 측정하기 위한 그래픽 사용자 인터페이스를 제공한다. 그래픽 사용자 인터페이스는 원격 컴퓨터로부터 받은 운반체 파동 상에 저장된 디지털 신호에서 코딩되는 결과를 디스플레이 하기 위한 디스플레이 필드를 포함한다. 다수의 특징은 다수의 바이오마커의 측정 가능한 양상이다. 다수의 바이오마커는 표 1A에 열거된 2 개 이상의 바이오마커를 포함한다. 결과는 시험 대상의 바이오마커 프로파일 내 다수의 특징이 제 1 값 세트를 만족하는 경우 제 1 값을 갖는다. 결과는 시험 대상의 바이오마커 프로파일 내 다수의 특징이 제 2 값 세트를 만족하는 경우 제 2 값을 갖는다.

#### 5.4 바이오마커 프로파일의 생성

[0119] 하나의 구현예에 따르면, 본 발명의 방법은 대상으로부터 취득된 생물학적 샘플로부터 바이오마커 프로파일을 생성하는 것을 포함한다. 생물학적 샘플 예를 들어, 말초 조직, 전혈, 뇌척수액, 복막액, 간질액, 적혈구 세포, 백혈구 세포 또는 혈소판일 수 있다.

##### 5.4.1 핵산 바이오마커의 검출 방법

[0121] 본 발명의 특정 구현예에서, 바이오마커 프로파일 중 바이오마커는 핵산이다. 이러한 바이오마커 및 바이오마커 프로파일의 상응하는 특징은 예를 들어, 본원에 기재된 하나 이상의 유전자 (예를 들어, 표 1A에 열거된 유전자)의 발현 생성물 (예를 들어, 폴리뉴클레오티드 또는 폴리펩티드)을 검출하여 생성될 수 있다. 특정 구현예에서, 바이오마커 및 바이오마커 프로파일 내 상응하는 특징은 혼성화, 마이크로어레이 분석, RT-PCR, 뉴클레아제 보호 어세이 및 노던 블롯 (Northern blot) 분석을 포함하나 이에 제한되지 않는 수단에 의해 당업자에게 잘 알려진 임의의 방법을 사용하는 본원에 기재된 유전자 (예를 들어, 표 1A에 열거된 유전자)로부터 발현된 하나 이상의 핵산을 검출 및/또는 분석하여 수득된다.

[0122] 특정 구현예에서, 본 발명의 방법 및 조성물에 의해 검출 및/또는 분석된 핵산에는 RNA 분자, 예를 들어, 메신저 RNA (mRNA) 분자, mRNA 스플라이싱 변이체 뿐 아니라 조절 RNA, cRNA 분자 (예를 들어, 시험관 내에서 전사된 cDNA 분자로부터 제조된 RNA 분자) 또는 이의 식별 조각을 포함하는 발현된 RNA 분자가 포함된다. 본 발명의 방법 및 조성물에 의해 검출 및/또는 분석된 핵산에는 또한 예를 들어, DNA 분자, 예컨대 게놈 DNA 분자, cDNA 분자, 및 이의 식별 조각 (예를 들어, 올리고뉴클레오티드, EST, STS, 등)이 포함될 수 있다.

[0123] 본 발명의 방법 및 조성물에 의해 검출 및/또는 분석된 핵산 분자는 자연 발생적 핵산 분자, 예컨대 샘플로부터 단리되는 게놈 또는 엑스트라게놈 DNA 분자, 또는 RNA 분자, 예컨대 생물학적 샘플에 존재하는, 그로부터 단리되는 또는 그로부터 유도되는 mRNA 분자일 수 있다. 본 발명의 방법 및 조성물에 의해 검출 및/또는 분석된 핵산 샘플은 예를 들어, DNA의 분자, RNA, 또는 DNA 및 RNA의 공중합체를 포함한다. 일반적으로, 상기 핵산은 특정 유전자 또는 유전자의 대립유전자에, 또는 특정 유전자 전사체 (예를 들어, 특이적 세포 유형에서 발

현되는 특정 mRNA 서열 또는 이러한 mRNA 서열로부터 유도된 특정 cDNA 서열)에 상응한다. 본 발명의 방법 및 조성물에 의해 검출 및/또는 분석된 핵산은 예를 들어, 유전자의 상이한 스플라이싱 변이체가 검출 및/또는 분석될 수 있도록 동일한 유전자의 상이한 엑손에 상응할 수 있다.

[0124] 특정 구현예에서, 핵산은 생물학적 샘플에 존재하는, 또는 이로부터 단리되는 또는 부분적으로 단리되는 핵산으로부터 시험관 내에서 제조된다. 예를 들어, 하나의 구현예에서, RNA는 샘플로부터 추출되고 (예를 들어, 총 세포 RNA, 폴리(A)<sup>+</sup> 메신저 RNA, 이의 분획), 메신저 RNA는 총 추출된 RNA로부터 정제된다. 총 및 폴리(A)<sup>+</sup> RNA의 제조 방법은 당업계에 잘 알려져 있고, 일반적으로 예를 들어, 문헌 [Sambrook et al., 2001, Molecular Cloning: A Laboratory Manual. 3<sup>rd</sup> ed. Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York)]에 기재되어 있다.

#### 5.4.1.1 핵산 어레이

[0126] 본 발명의 특정 구현예에서, 핵산 어레이는 본원에 기재된 유전자 중 임의의 하나 이상 (예를 들어, 표 1A에 열거된 유전자)의 발현을 검출하여 바이오마커 프로파일 내 바이오마커의 특징을 생성하는데 사용된다. 본 발명의 하나의 구현예에서, 마이크로어레이, 예컨대 cDNA 마이크로어레이는 바이오마커 프로파일 내 바이오마커의 특징 값을 측정하기 위해 사용된다. cDNA 어레이의 진단적 용도는 당업계에 잘 알려져 있다 (예를 들어, Zou et. al., 2002, *Oncogene* 21:4855-4862; 및 Draghici, 2003, *Data Analysis Tools for DNA Microarrays*, Chapman & Hall/CRC 참조). cDNA 마이크로어레이 분석을 위한 예시적 방법이 하기에 기재되어 있다.

[0127] 특정 구현예에서, 바이오마커 프로파일 내 바이오마커에 대한 특징 값은 하나 이상의 탐침 스팟을 포함하는 마이크로어레이에 대해 생물학적 샘플에 존재하는 mRNA 전사체 내 핵산 서열 (예를 들어, 샘플로부터 합성된 형광 표지된 cDNA)을 대표하거나 이에 상응하는 검출 표지된 핵산에 혼성화시켜 수득한다.

[0128] 핵산 어레이, 예를 들어, 마이크로어레이에는 여러 방식으로 제조될 수 있으며, 이 중 여러 개가 하기 본원에 기재된다. 바람직하게는, 어레이는 재현성이 있어, 제공된 어레이의 다수의 카페가 생성되게 하고, 상기 마이크로어레이로부터의 결과를 서로 비교한다. 바람직하게는, 어레이는 결합 (예를 들어, 핵산 혼성화) 조건 하에서 안정한 물질로부터 제조된다. 당업자는 어레이 상의 탐침 스팟에 시험 탐침을 혼성화시키기 위한 적합한 지지체, 기판 또는 운반체를 알고 있거나, 또는 정규적 실험을 사용하여 동일한 것을 알아낼 수 있을 것이다.

[0129] 사용되는 어레이, 예를 들어, 마이크로어레이에는 하나 이상의 시험 탐침이 포함될 수 있다. 일부 구현예에서, 각각의 이러한 시험 탐침은 검출되는 RNA 또는 DNA의 서열에 상보적인 핵산 서열을 포함한다. 각각의 탐침은 전형적으로 상이한 핵산 서열을 갖고, 어레이의 고체 표면 상의 각각의 탐침의 위치는 통상적으로 알려져 있거나 측정될 수 있다. 본 발명에 따라 유용한 어레이에는 예를 들어, 본원에 기재된 유전자 (예를 들어, 표 1A에 열거된 유전자)의 발현의 질적, 양적 또는 반-양적 측정을 제공할 수 있는 올리고뉴클레오티드 마이크로어레이, cDNA 기반 어레이, SNP 어레이, 스플라이싱 변이체 어레이 및 임의의 기타 어레이가 포함될 수 있다. 마이크로어레이의 일부 유형은 다를 수 있는 어레이이다. 더욱 구체적으로는, 일부 마이크로어레이는 위치적으로 다를 수 있는 어레이이다. 일부 구현예에서, 어레이의 각각의 탐침은 고체 지지체 상의 공지된, 미리 결정된 위치에 위치하여 각각의 탐침의 정체 (예를 들어, 서열)를 어레이 (예를 들어, 지지체 또는 표면 상에) 상의 위치로부터 측정할 수 있도록 한다. 일부 구현예에서, 어레이는 주문 어레이이다. 마이크로어레이는 일반적으로 문헌 [Draghici, 2003, *Data Analysis Tools for DNA Microarrays*, Chapman & Hall/CRC]에 기재되어 있다.

[0130] 본 발명의 일부 구현예에서, 발현된 전사체 (예를 들어, 본원에 기재된 유전자의 전사체)가 핵산 어레이에서 제시된다. 이러한 구현예에서, 결합 사이트 세트에는 발현된 전사체의 상이한 서열 분절과 상보적인 상이한 핵산과의 탐침이 포함될 수 있다. 상기 부류의 예시적 핵산은 15 내지 200 개의 염기, 20 내지 100 개의 염기, 25 내지 50 개의 염기, 40 내지 60 개의 염기 또는 일부 다른 범위의 염기 길이일 수 있다. 각각의 탐침 서열은 또한 표적 서열에 상보적인 서열 외에도 하나 이상의 연결자 서열을 포함할 수 있다. 본원에서 사용되는 바와 같이, 연결자 서열은 지지체의 표면과 표적 서열에 상보적인 서열 사이의 서열이다. 예를 들어, 본 발명의 핵산 어레이에는 각각의 표적 유전자 또는 엑손에 특이적인 하나의 탐침을 포함할 수 있다. 그러나, 바람직한 경우, 핵산 어레이에는 일부 발현된 전사체 (예를 들어, 본원에, 예를 들어, 표 1A에 기술된 유전자의 전사체)에 특이적인 2, 5, 10, 100, 또는 1000 개 이상 또는 그 이상의 탐침을 함유할 수 있다. 예

를 들어, 어레이는 유전자의 가장 긴 mRNA 이소형의 서열을 가로질러 타일로 된 탐침을 함유할 수 있다.

[0131] 세포, 예를 들어, 생물학적 샘플 내 세포의 RNA에 상보적인 cDNA가 제조되고, 적합한 혼성화 조건 하에서 마이크로어레이에 혼성화되는 경우, 본원에 기재된 유전자(예를 들어, 표 1A에 열거된 유전자)에 상응하는 어레이 내 부위에 대한 혼성화 수준은 유전자로부터 전사된 mRNA 또는 mRNA의 세포 내 보급을 반영할 것임이 인지될 것이다. 대안적으로는, 특정 유전자에 의해 생성되는 이소형 복합 또는 대안적인 스플라이싱 변이체가 구별되어야 하는 예에서는, 총 세포 mRNA에 상보적인 검출가능하게 표지된(예를 들어, 형관단으로) cDNA는 마이크로어레이에 혼성화될 수 있고, 세포 내 RNA 스플라이싱 동안 전사되지 않거나 제거되는 유전자의 엑손에 상응하는 어레이 상의 부위는 신호(예를 들어, 형광 신호)를 거의 또는 전혀 가지지 않을 것이고, 엑손을 발현하는 코딩된 mRNA가 보급된 유전자의 엑손에 상응하는 부위는 비교적 강한 신호를 가질 것이다. 그 다음 대안적인 스플라이싱에 의해 동일한 유전자로부터 생성되는 상이한 mRNA의 상대적인 풍부함을 유전자에 대해 모니터링된 엑손의 전체 세트에 걸쳐 신호 강도 패턴에 의해 측정한다.

[0132] 하나의 구현예에서, 상이한 혼성화 시간에서의 혼성화 수준을 상이한, 동일한 마이크로어레이 상에서 별도로 측정한다. 각각의 이러한 측정을 위해, 혼성화 시간에서 혼성화 수준을 측정할 때, 마이크로어레이를, 바람직하게는 실온에서, 모든 결합된 또는 혼성화된 핵산을 남겨놓으나, 모든 미결합된 핵산을 제거하는 조건 하에서 고 내지 중간 염 농도(예를 들어, 0.5 내지 3 M 염 농도)의 수용액에서 간단히 세정한다. 각각의 탐침에 대해 남아 있는, 혼성화된 핵산 분자 상의 검출가능한 표지를 사용된 측정 표지 방법에 적합한 방법에 의해 측정한다. 그 다음 수득된 혼성화 수준을 조합하여 혼성화 곡선을 형성한다. 또다른 구현예에서, 혼성화 수준을 단일 마이크로어레이를 사용하여 실시간 측정한다. 본 구현예에서, 마이크로어레이에는 중단 없이 샘플에 혼성화되는 것이 허용될 수 있고, 마이크로어레이에는 각각의 혼성화 시간에 비-침습성 방식으로 탐문된다.

더욱 또다른 구현예에서, 하나의 어레이를 사용하고, 단시간 동안 혼성화시키고, 세정 및 혼성화 수준을 측정하고, 동일한 샘플로 되돌려 놓고, 또다른 시간 동안 혼성화시키고, 세정 및 다시 측정하여 혼성화 시간 곡선을 얻을 수 있다.

[0133] 일부 구현예에서, 분석될 핵산 바이오마커가 상보적 DNA가 위치하는 어레이의 상보적 핵산 서열에, 전형적으로는 특이적인 어레이 부위에 특이적으로 결합하거나 특이적으로 혼성화하도록 핵산 혼성화 및 세정 조건이 선택된다.

[0134] 위치시킨 이중 가닥 탐침 DNA를 함유하는 어레이는 표적 핵산 분자와 접촉 전에 DNA 단일 가닥이 되도록 변성 조건에 적용될 수 있다. 단일 가닥 탐침 DNA(예를 들어, 합성 올리고데옥시리보핵산)를 함유하는 어레이는 예를 들어, 자가 상보 서열로 인해 형성되는 헤어핀 또는 이량체를 제거하기 위해 표적 핵산 분자와 접촉하기 전 변성될 필요가 있다.

[0135] 최적 혼성화 조건은 탐침 및 표적 핵산의 길이(예를 들어, 올리고머 대 200 개 염기 초과의 폴리뉴클레오티드) 및 유형(예를 들어, RNA, 또는 DNA)에 따라 다를 것이다. 핵산에 대한 특이적(즉, 업격함) 혼성화 조건을 위한 일반적인 파라미터는 문헌 [Sambrook et al., (supra)], 및 [Ausubel et al., latest edition, *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York]에 기재되어 있다. Shena et al의 cDNA 마이크로어레이가 사용되는 경우, 전형적인 혼성화 조건은 5 X SSC + 0.2% SDS, 65°C에서 4시간 동안 혼성화 후, 25°C, 낮은 업격함 세정 완충액(1 X SSC + 0.2% SDS)에서의 세정 후, 25°C, 높은 업격함 세정 완충액(0.1 X SSC + 0.2% SDS)에서의 10분이다(Shena et al., 1996, *Proc. Natl. Acad. Sci. U.S.A.* 93:10614). 유용한 혼성화 조건은 또한 예를 들어 문헌 [Tijesssen, 1993, *Hybridization With Nucleic Acid Probes*, Elsevier Science Publishers B.V.; Kricka, 1992, Nonisotopic DNA Probe Techniques, Academic Press, San Diego, CA; 및 Zou et. al., 2002, *Oncogene* 21:4855-4862; 및 Draghici, *Data Analysis Tools for DNA Microanalysis*, 2003, CRC Press LLC, Boca Raton, Florida, pp. 342-343]에 제공된다.

[0136] 특정 구현예에서, 마이크로어레이에는 예를 들어, 하기 섹션 5.4.1.2에 기재된 방법에 의해 생성되는 RT-PCR 생성물을 정리하는데 사용될 수 있다.

#### 5.4.1.2 RT-PCR

[0138] 특정 구현예에서, 본 발명의 바이오마커 프로파일 내 바이오마커의 특징 값을 측정하기 위해, 하나 이상의 본원에 기재된 유전자(예를 들어, 표 1A에 열거된 유전자)의 발현 수준을 폴리머라아제 연쇄 반응(PCR)과 함께 역전사(RT)를 사용하여 샘플로부터 RNA를 증폭시켜 측정한다. 본 구현예에 따르면, 역전사는 정량적 또

는 반-정량적일 수 있다. 본원에 교시된 RT-PCR 방법은 상기 기재된 마이크로어레이 방법 (예를 들어, 섹션 5.4.1.1) 과 함께 사용될 수 있다. 예를 들어, 벌크 PCR 반응이 수행될 수 있고, PCR 생성물은 분해될 수 있고 마이크로어레이 상의 탐침 스팟으로서 사용될 수 있다.

[0139] 샘플로부터의 총 RNA, 또는 mRNA 가 주형으로서 사용되고, 유전자(들)의 전사된 부분에 측정적인 프라이머가 역전사를 개시하기 위해 사용된다. RNA 를 cDNA 로 역전사하는 방법은 잘 알려져 있고, 문헌 [Sambrook et al., 2001, supra] 에 기재되어 있다. 프라이머 디자인은 GenBank 와 같은 임의의 공개적으로 이용가능한 서열 데이터베이스로부터 이용가능하거나 공개된 공지된 뉴클레오티드 서열에 근거하여 달성될 수 있다. 예를 들어, 프라이머는 본원에 (예를 들어, 표 1A 참조) 기재된 유전자 중 임의의 것에 대해 디자인될 수 있다. 또한, 프라이머 디자인은 시판되는 소프트웨어 (예를 들어, Primer Designer 1.0, Scientific Software 등) 를 이용하여 달성될 수 있다. 역전사의 생성물은 이후에 PCR 에 대한 주형으로서 사용된다.

[0140] PCR 은 관심의 표적 서열을 증폭시키기 위해 열안정성, DNA-의존성 DNA 폴리머라아제에 의해 촉매화된 DNA 복제의 복합 사이클을 사용하여 특정 핵산 서열을 빠르게 증폭시키는 방법을 제공한다. PCR 은 증폭될 핵산, 증폭될 서열의 측면에 위치하는 2 개의 단일 가닥 올리고뉴클레오티드 프라이머, DNA 폴리머라아제, 테옥시리보뉴클레오티드 트리포스페이트, 완충제 및 염의 존재를 필요로 한다. PCR 방법은 당업계에 잘 알려져 있다. PCR 은 예를 들어, 문헌 [Mullis and Faloona, 1987, Methods Enzymol. 155:335] 에 기재된 바와 같이 수행된다.

[0141] PCR 은 주형 DNA 또는 cDNA (1 fg 이상; 더욱 유용하게는, 1-1000 ng) 및 25 pmol 이상의 올리고뉴클레오티드 프라이머를 사용하여 수행될 수 있다. 전형적인 반응 혼합물에는 2  $\mu$ l 의 DNA, 25 pmol 의 올리고뉴클레오티드 프라이머, 2.5  $\mu$ l 의 10 M PCR 완충제 1 (Perkin-Elmer, Foster City, CA), 0.4  $\mu$ l 의 1.25 M dNTP, 0.15  $\mu$ l (또는 2.5 유닛) 의 Taq DNA 폴리머라아제 (Perkin Elmer, Foster City, CA) 및 총 부피 25  $\mu$ l 까지의 틸이 온수가 포함된다. 광유로 덮어씌우고 PCR 은 프로그램 가능 열 사이클러를 사용하여 수행된다.

[0142] PCR 사이클의 각 단계의 길이 및 온도, 뿐 아니라 사이클 횟수는 실제로 염격함 필요조건에 따라 조절된다. 어닐링 온도 및 타이밍은 프라이머가 주형에 대해 어닐링되는 것을 예상하는 효율 및 용인되는 미스매치 정도 모두에 의해 측정된다. 프라이머 어닐링 조건의 염격함을 최적화시키는 능력은 당업자의 지식 범위 내에 있다. 30°C 내지 72°C 의 어닐링 온도가 사용된다. 주형 분자의 초기 변성은 통상적으로 92°C 내지 99°C 에서 4 분 동안, 그 후 변성 (94-99°C, 15 초 내지 1 분), 어닐링 (상기 논의된 바와 같이 측정되는 온도; 1-2 분), 및 신장 (72°C, 1 분) 으로 이루어지는 사이클 20-40 회로 일어난다. 최종 연장 단계는 일반적으로 4 분 동안 72°C 에서 수행되며, 4°C 에서 무기한 (0-24 시간) 단계가 뒤따를 수 있다.

[0143] 사실상 정량적인, 정량적 RT-PCR ("QRT-PCR") 은 또한 유전자 발현 수준의 정량적 측정치를 제공하기 위해 수행될 수 있다. QRT-PCR 에서 역전사 및 PCR 은 2 단계로 수행될 수 있고, 또는 PCR 과 조합된 역전사가 동시에 수행될 수 있다. 상기 기술 중 하나 (이를 위해 시판 키트, 예컨대 Taqman (Perkin Elmer, Foster City, California) 또는 Applied Biosystems (Foster City, California) 에 의해 제공되는 것이 있음) 는 전사-특이적 안티센스 탐침으로 수행된다. 상기 탐침은 PCR 생성물 (예를 들어, 유전자로부터 유도된 핵산 조각) 에 대해 특이적이고, 올리고뉴클레오티드의 5' 말단과 복합체를 형성한 형광 리포터 탐침 및 켄쳐로 제조된다. 상이한 형광 마커가 상이한 리포터에 부착되어, 하나의 반응에서 2 개의 생성물의 측정을 가능하게 한다. Taq DNA 폴리머라아제가 활성화되는 경우, 이것은 이것의 5'-에서-3' 엑소뉴클레아제 활성에 의해 주형에 결합된 탐침의 형광 리포터를 분할한다. 켄쳐의 부재시, 리포터는 이제 형광을 냈다. 리포터의 색 조 변화는 각각의 특이적 생성물의 양에 비례하고, 형광계에 의해 측정된다; 그러므로, 각각의 색조의 양을 측정하고, PCR 생성물을 정량한다. PCR 반응을 96 웰 플레이트에서 수행하여, 많은 개인으로부터 유래한 샘플을 동시에 처리하고 측정하였다. Taqman 시스템은 젤 전기영동을 필요로 하지 않는 부가적인 장점을 가지며, 표준 곡선으로 사용되는 경우 정량을 가능하게 한다.

[0144] 정량적으로 PCR 생성물을 검출하는데 유용한 제 2 기술은 시판되는 Quant iTect SYBR Green PCR (Qiagen, Valencia California) 과 같은 인터콜레이트 (intercalating) 염료를 사용하는 것이다. RT-PCR 은 PCR 공정 동안 PCR 생성물 내로 흔입되는 형광 표지로서 SYBR 녹색을 사용하여 수행되고, PCR 생성물의 양과 비례하는 형광을 생성한다.

[0145] Taqman 및 Quant iTect SYBR 시스템 모두는 RNA 의 역전사에 후속하여 사용될 수 있다. 역 전사는 PCR 단계 (1 단계 프로토콜) 와 동일한 반응 혼합물에서 수행될 수 있고, 또는 역전사는 PCR (2 단계 프로토콜) 을 사용하는 증폭 전에 먼저 수행될 수 있다.

[0146] 부가적으로는, mRNA 발현 생성물을 정량적으로 측정하기 위한 다른 시스템이 형광 분자 및 켄처 분자를 갖는 탐침을 사용하는 MOLECULAR BEACONS® 을 비롯하여 알려져 있고, 상기 탐침은 헤어핀 형태로 있는 경우, 형광 분자는 켄칭되고, 혼성화되는 경우, 형광 증가가 유전자 발현의 정량적 측정을 산출하는 식으로 헤어핀 구조를 형성할 수 있다.

[0147] RNA 발현을 정량적으로 측정하기 위한 부가적인 기술에는 폴리미라아제 연쇄 반응, 리가아제 연쇄 반응, Qbeta 레플리카아제 (예를 들어, 국제 출원 번호 PCT/US87/00880 호 참조), 등온 증폭 방법 (예를 들어, Walker et al., 1992, PNAS 89:382-396 참조), 가닥 대체 증폭 (SDA), 보수 연쇄 반응, 비대칭적 정량적 PCR (예를 들어, U.S. 공개 번호 US 2003/30134307A1 호 참조) 및 문헌 [Fuja et al., 2004, Journal of Biotechnology 108:193-205] 에 기재되어 있는 멀티플렉스 마이크로스피어 비이드 어세이가 포함되나 이에 제한되는 것은 아니다.

#### 5.4.2 단백질 검출 방법

[0149] 본 발명의 특정 구현예에서, 바이오마커 프로파일 내 바이오마커의 특징 값은 단백질을 검출함으로써, 예를 들어, 하나 이상의 본원에 기재된 유전자 (예를 들어, 표 1A 에 열거된 유전자) 의 발현 생성물 (예를 들어, 핵산 또는 단백질) 을 검출함으로써, 또는 번역 후 변형되고, 또는 다르게 변형되고, 또는 이러한 단백질의 형태를 프로세싱할 수 있다. 특정 구현예에서, 바이오마커 프로파일은 단백질 마이크로어레이 분석, 번역조직화학 및 질량 분석법을 포함하나 이에 제한되지 않는 단백질 검출을 위한 당업자에게 알려진 임의의 방법을 사용하여 본원에 기재된 유전자 (예를 들어, 표 1A 에 열거된 유전자) 로부터 발현된 하나 이상의 단백질 및/또는 이의 식별 조각을 검출 및/또는 분석함으로써 생성된다.

[0150] 표준 기술은 샘플 내에 존재하는 관심의 단백질(들) (예를 들어, 표 1A 에 열거된 유전자로부터 발현된 단백질) 의 양을 측정하는데 유용할 수 있다. 예를 들어, 표준 기술은 샘플 내에 존재하는 관심의 단백질(들) 의 양을 측정하기 위해 예를 들어, 면역어세이, 예컨대, 웨스턴 블롯, 나트륨 도데실 술페이트 폴리아크릴아미드 젠전기영동 (SDS-PAGE) 후 면역침전, 면역세포화학 등을 이용할 수 있다. 관심의 단백질을 측정하기 위한 하나의 예시적 작용제는 관심의 단백질에 특이적으로 결합할 수 있는 항체, 바람직하게는 직접 또는 간접적으로 검출가능하게 표지된 항체이다.

[0151] 이러한 검출 방법을 위해, 바람직한 경우 분석될 샘플로부터의 단백질은 당업자에게 잘 알려진 기술을 사용하여 쉽게 단리될 수 있다. 단백질 단리 방법은 예를 들어, 문헌 [Harlow and Lane, 1988, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York)] 에 기재되어 있는 방법일 수 있다.

#### 5.5 데이터 분석 알고리즘

[0153] 상응하는 특징 값이 정동 장애를 진단할 수 있는 바이오마커가 본 발명에서 확인된다. 이들 바이오마커의 확인 및 이들의 상응하는 특징들 (예를 들어, 발현 수준) 은, 정동 장애를 갖는 대상과 정동 장애를 갖지 않는 대상을 구별하는 판단 규칙 또는 다수의 판단 규칙을 개발하는데 사용될 수 있다. 일단, 이들의 예시적인 데이터 분석 알고리즘 또는 당업계에 알려진 다른 기술을 이용하여 판단 규칙이 구축되면, 이 결정 규칙은 시험 대상을 2 개 이상의 표현형 부류 (예를 들어, 정동 장애를 갖는 것, 정동 장애를 갖지 않는 것) 중 하나로 분류하는데 이용될 수 있다. 이것은 시험 대상으로부터 수득된 바이오마커 프로파일에 결정 규칙을 적용함으로써 달성된다. 그러므로, 이러한 결정 규칙은 진단 지표로서 막대한 값을 갖는다.

[0154] 본 발명은, 하나의 양상에서, 훈련 집단으로부터 수득된 바이오마커 프로파일들에 대한 시험 대상으로부터의 바이오마커 프로파일의 평가를 제공한다. 일부 구현예에서, 시험 대상뿐만 아니라 훈련 집단에서의 대상으로부터 수득된 각각의 바이오마커 프로파일은, 다수의 상이한 바이오마커의 각각에 대한 특징을 포함한다. 일부 구현예에서, 이 비교는 (i) 훈련 집단으로부터의 바이오마커 프로파일들을 이용하여 판단 규칙을 개발하는 것 및 (ii) 이 판단 규칙을 시험 대상으로부터의 바이오마커 프로파일에 적용하는 것에 의해 달성된다. 이와 같이, 본 발명의 일부 구현예에 적용된 이 판단 규칙은 시험 대상이 정동 장애를 갖는지의 여부를 결정하는데 이용된다.

[0155] 본 발명의 일부 구현예에서, 대상이 정동 장애를 갖는다고 판단 규칙의 적용의 결과가 지시하는 경우, 대상은 "정동 장애" 대상으로서 진단된다. 대상이 장애를 갖지 않는다고 판단 규칙의 적용의 결과가 지시하는 경우, 대상은 "비 정동 장애" 대상으로서 진단된다. 따라서, 일부 구현예에서, 전술한 2 가지 결정 상황에서

의 결과는 다음의 4 가지 가능한 결과들을 갖는다:

[0156] (i) 대상이 정동 장애를 갖는다고 판단 규칙이 지시하고 있으며 사실상 대상이 정동 장애를 갖는, 정동 장애를 참으로 갖는 것 (진 양성 (true positive, TP));

[0157] (ii) 대상이 정동 장애를 갖는다고 판단 규칙이 지시하고 있지만 사실상 대상이 정동 장애를 갖지 않는, 정동 장애를 거짓으로 갖는 것 (가 양성 (false positive, FP));

[0158] (iii) 대상이 정동 장애를 갖지 않는다고 판단 규칙이 지시하고 있으며 사실상 대상이 정동 장애를 갖지 않는, 정동 장애를 참으로 갖지 않는 것 (진 음성 (true negative, TN)); 또는

[0159] (iv) 대상이 정동 장애를 갖지 않는다고 판단 규칙이 지시하고 있으며 사실상 대상이 정동 장애를 갖는, 정동 장애를 거짓으로 갖지 않는 것 (가 음성 (false negative, FN)).

[0160] TP, FP, TN, FN에 대해 다른 정의들이 내려질 수 있다는 점은 이해될 것이다. 본 발명의 이해를 용이하게 하기 위한 이러한 모든 대안적인 정의들은 본 발명의 범주 내에 있기는 하지만, 다른 언급이 없다면, 정의 (i) 내지 (iv)에 의해 주어진 TP, FP, TN, 및 FN에 대한 정의가 본 명세서에서 사용될 것이다.

[0161] 당업자에 의해 이해될 바와 같이, 시험 바이오마커 프로파일과 참조 바이오마커 프로파일 사이에서 이루어지는 비교들의 수행 (예를 들어, 시험 대상으로부터의 바이오마커 프로파일에 대한 판단 규칙의 적용)을 통신하기 위해 수많은 정량적인 기준이 이용될 수 있다. 이들은 양성 예측값 (PPV), 음성 예측값 (NPV), 특이성, 감수성, 정확성, 및 확실성을 포함한다. 또한, 이러한 ROC (receiver operator curve)를 구성하는 다른 것이 판단 규칙 성능을 평가하는데 이용될 수 있다. 본 명세서에 이용되는 바는 다음과 같다:

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

$$\text{특이성} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{감수성} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{정확성} = \text{확실성} = \frac{\text{TP} + \text{TN}}{\text{N}}$$

[0167] 식 중, N은 비교되는 샘플의 수 (예컨대, 시험 샘플의 수)이다. 예를 들어, 정동 장애 분류가 발견되는 10개의 대상이 존재하는 경우를 고려해본다. 10개의 시험 대상의 각각에 대한 바이오마커 프로파일이 구축된다. 그 후, 판단 규칙을 적용함으로써 바이오마커 프로파일 각각이 평가되며, 이 판단 규칙은 훈련 집단으로부터 수득된 바이오마커 프로파일에 기초하여 개발되었다. 이 예에서, 상기 식들로부터 N은 10이다. 통상적으로, N은 샘플의 수이며, 각각의 샘플은 집단의 상이한 일원으로부터 수집되었다. 이 집단은, 사실상, 2개의 상이한 유형으로 이루어진다. 하나의 유형에서, 집단은 대상을 포함하며, 여기서 이 대상들의 샘플 및 표현형 데이터 (예를 들어, 바이오마커의 특징 값 및 대상이 정동 장애를 갖는지 아닌지의 표시)는 판단 규칙을 구성 또는 개선하는데 이용되었다. 이러한 집단은 본 명세서에서 훈련 집단으로서 지칭된다. 다른 유형에서는, 집단이 판단 규칙을 구성하는데 이용되지 않는 대상들을 포함한다. 이러한 집단은 본 명세서에서 검증 집단으로서 지칭된다. 다른 언급이 없으면, 2개의 집단 유형의 혼합과는 대조적 으로, N에 의해 표현되는 집단은 오로지 훈련 집단 또는 오로지 검증 집단 중 어느 하나이다. 점수가 검증 집단과는 대조적으로 훈련 집단에 기초하는 경우, 정확성과 같은 점수는 더 높을 (일치 (unity)에 더 근접할) 것이라는 점은 이해될 것이다. 그럼에도 불구하고, 본 명세서에서 명시적으로 다르게 언급하지 않는다면, 확실성 (정확성)을 포함하는 판단 규칙의 수행을 평가하는데 이용되는 모든 기준 (또는 시험 대상으로부터의 바이오마커 프로파일의 평가의 다른 형식들)은, 이 기준에 상응하는 판단 규칙을 훈련 집단 또는 검증 집단 중 어느 하나에 적용함으로써 측정되었던 기준을 의미한다. 또한, 상기 정의된 PPV, NPV, 특이성, 감수성, 및 정확성에 대한 정의들은 Draghici, *Data Analysis Tools for DNA Microanalysis*, 2003, CRC Press LLC, Boca Raton, Florida, pp. 342-343에서도 또한 찾을 수 있다.

[0168]

일부 구현예에서, N은 1 초과, 5 초과, 10 초과, 20 초과, 10 내지 100, 100 초과, 또는 1000 미만의 대상들이다. 일부 구현예에서, 훈련 집단 또는 검증 집단에 비하여, 판단 규칙(또는 비교의 다른 형식들)이 약 99% 이상의 확실성 또는 심지어 그 이상을 가질 수 있다. 다른 구현예에서, 훈련 집단 또는 검증 집단에 비하여(즉, 임상 환자와 같은 훈련 집단의 일부가 아닌 단일 대상에 비하여), 확실성이 약 97% 이상, 약 95% 이상, 약 90% 이상, 약 85% 이상, 약 80% 이상, 약 75% 이상, 약 70% 이상, 약 65% 이상, 또는 약 60% 이상이다.

유용한 확실성 정도는, 본 발명의 특정 방법에 따라 변화될 수도 있다. 본 명세서에서 사용되는 바와 같이, "확실성"은 "정확성"을 의미한다. 하나의 구현예에서는, 훈련 집단 또는 검증 집단에 비하여, 감수성 및/또는 특이성은 약 97% 이상, 약 95% 이상, 약 90% 이상, 약 85% 이상, 약 80% 이상, 약 75% 이상, 또는 약 70% 이상이다. 일부 구현예에서, 규정된 정확성으로 대상이 정동 장애를 갖는지를 예측하는데 이러한 규칙이 사용된다. 일부 구현예에서, 규정된 정확성으로 정동 장애를 진단하기 위해 이러한 판단 규칙이 사용된다. 일부 구현예에서는, 규정된 정확성으로 대상이 정동 장애의 증상을 갖게 되는 가망성을 결정하기 위해, 이러한 규칙이 사용된다.

[0169]

적합한 확실성을 갖는 시험 대상을 분류하기 위해 판단 규칙에 의해 이용될 수 있는 특징의 수는 2 개 이상이다. 일부 구현예에서는, 3 개 이상, 4 개 이상, 10 개 이상, 또는 10 개 내지 200 개이다. 그러나, 발견된 확실성 정도에 따라, 판단 규칙에서 이용되는 특징들의 수는 더 많거나 더 적을 수 있지만, 모든 경우들에 있어서 2 개 이상이다. 하나의 구현예에서, 시험 대상을 분류하기 위해 판단 규칙에 의해 이용될 수 있는 특징의 수는, 높은 확실성으로 시험 대상의 분류를 허용하도록 최적화된다.

[0170]

판단 규칙을 개발하는 관련 데이터 분석 알고리즘들은 선형 판별 기술, 로지스틱 판별 기술, 그리고 더욱 유연한 판별 기술들을 포함하는 판별 분석(예를 들어, Gnanadesikan, 1977, *Methods for Statistical Data Analysis of Multivariate Observations*, New York Wiley 1977 참조); CART(classification and regression trees: 분류 및 회귀 트리) 및 변형형과 같은 트리-기반 알고리즘(예를 들어, Breiman, 1984, *Classification and Regression Trees*, Belmont, California Wadsworth International Group 참조); 일반화된 추가적 모델(예를 들어, Tibshirani, 1990, *Generalized Additive Models*, London Chapman and Hall 참조); 및 뉴럴 네트워크(예를 들어, Neal, 1996, *Bayesian Learning for Neural Networks*, New York Springer-Verlag; and Insua, 1998, *Feedforward neural networks for nonparametric regression* In: *Practical Nonparametric and Semiparametric Bayesian Statistics*, pp 181-194, New York Springer, 뿐만 아니라 이하의 섹션 5.5.2 참조)를 포함하지만, 이에 한정되는 것은 아니다.

[0171]

하나의 구현예에서, 훈련 집단으로부터 수득된 바이오마커 프로파일에 대한 시험 대상의 바이오마커 프로파일의 비교가 수행되고, 이 비교는 판단 규칙을 적용하는 것을 포함한다. 컴퓨터 패턴 인식 알고리즘과 같은 데이터 분석 알고리즘을 이용하여, 판단 규칙이 구성된다. 판단 규칙을 구성하는 다른 적합한 데이터 분석 알고리즘들은 특징 값의 분포에 있어서 상이함을 검출하는 로지스틱 회귀 알고리즘 또는 비-파라미터 알고리즘(예를 들어, (조정되지 않은 및 조정된) Wilcoxon Signed Rank Test)을 포함하지만, 이에 한정되는 것은 아니다.

판단 규칙은, 1 개, 2 개, 3 개, 4 개, 5 개, 10 개, 20 개, 또는 그보다 큰 수의 바이오마커로부터의 측정된 가측치들에 대응하는, 2 개, 3 개, 4 개, 5 개, 10 개, 20 개, 또는 그보다 큰 수의 특징들에 기초할 수 있다. 하나의 구현예에서, 판단 규칙은 수 백개 이상의 특징에 기초한다. 판단 규칙은 또한 분류 트리 알고리즘을 이용하여 구축될 수도 있다. 예를 들어, 훈련 집단으로부터의 각 바이오마커 프로파일은 3 개 이상의 특징을 포함할 수 있으며, 이 특징은 분류 트리 알고리즘에서 예측변수(predictor) 들이다(이하의 섹션 5.5.1 참조). 판단 규칙은 약 70% 이상, 약 75% 이상, 약 80% 이상, 약 85% 이상, 약 90% 이상, 약 95% 이상, 약 97% 이상, 약 98% 이상, 약 99% 이상, 또는 약 100%의 정확성으로 집단(또는 부류) 내의 구성원 수를 예측한다.

[0172]

적합한 데이터 분석 알고리즘은 당업계에 공지되어 있으며, 그 중 일부는 Hastie *et al.*, *supra*에서 검토되어 있다. 특정 구현예에서, 본 발명의 데이터 분석 알고리즘은 CART(이하의 섹션 5.5.1), MART(Multiple Additive Regression Tree), PAM(Prediction Analysis for Microarrays) 또는 랜덤 포레스트 분석(이하의 섹션 5.5.1)을 포함한다. 이러한 알고리즘들은 혈액 샘플과 같은 생물학적 물질로부터 복합 스펙트럼들을 분류하여, 대상들을, 특정 질병 상태의 정상적인 바이오마커 발현 수준들 특성 또는 특정 질병 상태의 프로세싱 바이오마커 발현 수준들 특성으로서 구별한다. 다른 구현예에서, 본 발명의 데이터 분석 알고리즘은 ANOVA 및 비-파라미터 등가물들, 선형 판별식 분석, 로지스틱 회귀 분석, 최근접 이웃 분류기 분석, 뉴럴 네트워크(이하의 섹션 5.5.2 참조), 주요 성분 분석, 이차 판별식 분석, 회귀 분류기 및 지원 벡터 머신(이하의 섹션 5.5.4), 관련성 벡터 머신 및 유전학적 알고리즘(이하의 섹션 5.5.5)을 포함한다. 이러한 알고리즘들은

판단 규칙을 구성하고/하거나 판단 규칙의 적용 속도와 효율을 증가시키고, 조사자 편견을 회피하는데 이용될 수도 있지만, 본 발명의 방법을 수행하기 위해 컴퓨터 기반의 알고리즘들이 필요하지 않는다는 점을 당업자는 인식할 것이다.

[0173] 판단 규칙은, 바이오마커 프로파일을 생성하는데 이용되었던 방법과는 상관없이, 바이오마커 프로파일을 평가하는데 이용될 수 있다. 예를 들어, 기체 크로마토그래피를 이용하여 생성된 바이오마커 프로파일들을 평가하는데 이용될 수 있는 적합한 판단 규칙은 Harper, "Pyrolysis and GC in Polymer Analysis," Dekker, New York (1985)에서 논의되는 바와 같다. 또한, Wagner *et al.*, 2002, *Anal. Chem.* 74:1824-1835는 정적 비행시간 이차이온질량분광법 (time-of-flight secondary ion mass spectrometry; TOF-SIMS)에 의해 수득된 스펙트럼들에 기초하여 대상들을 분류하는 능력을 향상시키는 판단 규칙을 개시하고 있다. 또한, Bright *et al.*, 2002, *J. Microbiol. Methods* 48:127-38은 MALDI-TOF-MS 스펙트럼들의 분석에 의해 높은 확실성 (79-89% 올바른 분류 비율)으로 박테리아 균주를 구별하는 방법을 개시하고 있다. Dalluge, 2000, *Fresenius J. Anal. Chem.* 366:701-711은, 복잡한 생물학적 샘플에서 바이오마커의 프로파일을 분류하기 위해, MALDI-TOF-MS 및 액체 크로마토그래피-전자분무 이온화 질량분석법 (liquid chromatography-electrospray ionization mass spectrometry; LC/ESI-MS)의 이용을 논하고 있다.

#### 5.5.1 결정 트리

[0175] 본 발명에서 확인되는 바이오마커의 특징 값을 이용하여 구성될 수 있는 판단 규칙의 하나의 유형은 결정 트리이다. 여기서, "데이터 분석 알고리즘"은 결정 트리를 구축할 수 있는 임의의 기술이며, 반면, 최종 "결정 트리"는 판단 규칙이다. 훈련 집단 및 특정 데이터 분석 알고리즘을 이용하여 결정 트리가 구성된다. 결정 트리들은, Duda, 2001, *Pattern Classification*, John Wiley & Sons, Inc., New York. pp. 395-396에 의해 일반적으로 설명된다. 트리 기반의 방법들은 특정 공간을 한 세트의 직사각형들로 구획하고, 그 후 각 하나에서 (상수와 같이) 모델을 피팅 (fitting) 한다.

[0176] 훈련 집단 데이터는, 훈련 세트 집단에 걸친 본 발명의 바이오마커에 대한 특징 (예를 들어, 발현 값 또는 일부 다른 관찰가능한 것)을 포함한다. 결정 트리를 구성하는데 이용될 수 있는 하나의 특정 알고리즘은 분류 및 회귀 트리 (CART)이다. 다른 특정 결정 트리 알고리즘들은 ID3, C4.5, MART, 및 랜덤 포레스트를 포함하지만, 이에 한정되는 것은 아니다. CART, ID3, 및 C4.5는 Duda, 2001, *Pattern Classification*, John Wiley & Sons, Inc., New York. pp. 396-408 및 pp. 411-412에서 설명된다. CART, MART, 및 C4.5는 Hastie *et al.*, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York, Chapter 9에서 설명된다. 랜덤 포레스트는, Breiman, 1999, "Random Forests - Random Features," Technical Report 567, Statistics Department, U.C.Berkeley, September 1999에서 설명된다.

[0177] 본 발명의 일부 구현예에서, 결정 트리는 본 발명의 바이오마커의 조합에 대한 특징을 이용하여 대상을 분류하는데 이용된다. 결정 트리 알고리즘은 감독 습득 알고리즘의 부류에 속한다. 결정 트리의 목표는 현실 세계의 예시 데이터로부터 분류기 (트리)를 유도하는 것이다. 이 트리는 결정 트리를 추론하는데 이용되지 않는 보이지 않는 예시를 분류하는데 이용될 수 있다. 이와 같이, 결정 트리는 훈련 데이터로부터 추론된다. 예시적인 훈련 데이터는 다수의 대상을 (훈련 집단)에 대한 데이터를 포함한다. 각각의 각 대상에 있어서, (예를 들어, 정동 장애를 갖는/정동 장애를 갖지 않는) 각각의 대상의 부류에 다수의 특징이 존재한다. 본 발명의 하나의 구현예에서, 훈련 데이터는 훈련 집단에 걸친 바이오마커의 조합에 대한 발현 데이터이다.

[0178] 일반적으로, 수많은 상이한 결정 트리 알고리즘이 존재하며, 이에 대한 상당 부분은 Duda, *Pattern Classification*, Second Edition, 2001, John Wiley & Sons, Inc.에서 설명된다. 결정 트리 알고리즘은 특징 프로세싱, 불순도 측정, 정지 기준, 및 프루닝 (pruning)에 대한 고려를 종종 필요로 한다. 특정 결정 트리 알고리즘은 분류 및 회귀 트리 (CART), 다변량 결정 트리, ID3, 및 C4.5을 포함하지만, 이에 한정되는 것은 아니다.

[0179] 하나의 접근법에서, 결정 트리가 이용되는 경우, 훈련 집단에 걸쳐 본 발명에서 설명된 유전자의 선택 조합에 대한 유전자 발현 데이터는, 평균 0 및 단위 변량을 갖도록 표준화된다. 훈련 집단의 일원은 무작위로 훈련 세트 및 시험 세트로 나눈다. 예를 들어, 하나의 구현예에서, 훈련 집단의 구성원들 중 2/3는 훈련 세트에 배치되고, 훈련 집단의 구성원들 중 1/3은 시험 세트에 배치된다. 본 발명에서 설명된 바이오마커의 선택 조합에 대한 발현 값은 결정 트리를 구성하는데 이용된다. 그 후, 결정 트리가 시험 세트 내의 일원을 정확히 분류하는 능력이 결정된다. 일부 구현예에서, 이 계산은 바이오마커의 소정 조합에 대해 얼마의 시간 동

안 수행된다. 각 계산 반복에서, 훈련 집단의 일원은 훈련 세트 및 시험 세트에 무작위로 나눈다. 그 후, 바이오마커의 조합의 품질은 결정 트리 계산의 이러한 각각의 반복의 평균으로서 취급된다.

[0180] 각 스플릿이 상응하는 바이오마커에 대한 특징 값에 기초하는 보편적인 결정 트리 외에도, 본 발명의 바이오마커의 세트, 또는 이러한 2 개의 바이오마커의 상대적인 특징 값 중에서, 보편적인 결정 트리가 판단 규칙으로서 구현될 수 있다. 이러한 보편적인 결정 트리에서, 결정의 일부 또는 전부는 실제로 본 발명의 다수의 바이오마커에 대한 특징 값의 선형 조합을 포함한다. 이러한 선형 조합은 분류 상의 경사하강 (gradient descent) 과 같은 공지의 기술을 이용하여 또는 합-제곱-에러 (sum-squared-error) 기준의 사용에 의해 훈련될 수 있다. 이러한 결정 트리를 설명하기 위해, 다음의 식을 고려한다:

$$0.04 x_1 + 0.16 x_2 < 500$$

[0181] 식 중,  $x_1$  및  $x_2$  는 본 발명의 바이오마커 중 2 개의 상이한 바이오마커에 대한 2 개의 상이한 특징을 의미한다. 판단 규칙을 모집하기 위해, 특징  $x_1$  및  $x_2$  의 값은 분류되지 않은 대상으로부터 수득된 측정값으로부터 수득된다. 그 후, 이 값을 식에 삽입한다. 500 미만의 값이 컴퓨팅된다면, 결정 트리에서의 제 1 브랜치가 취해진다. 그렇지 않으면, 결정 트리에서의 제 2 브랜치가 취해진다. 보편적인 결정 트리들은 Duda, 2001, *Pattern Classification*, John Wiley & Sons, Inc., New York, pp. 408-409 에서 설명된다.

[0183] 본 발명에서 사용될 수 있는 다른 접근법은 MARS (multivariate adaptive regression splines) 이다. MARS 는 회귀에 대한 적응성 있는 절차이며, 본 발명에 의해 언급되는 고차원 문제에 충분히 적합하다. MARS 는 CART 방법의 점진적인 선형의 회귀의 일반화 또는 회귀 세팅에서 CART 의 성능을 향상시키기 위해 CART 방법의 변형으로서 간주될 수 있다. MARS 는 Hastie *et al.*, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York, pp. 283-295 에서 설명된다.

### 5.5.2 뉴럴 네트워크

[0185] 일부 구현예에서, 본 발명의 선택 바이오마커에 대해 측정된 특징 데이터 (예컨대, RT-PCR 데이터, 질량 분석 데이터, 마이크로어레이 데이터) 는 뉴럴 네트워크를 훈련하는데 이용될 수 있다. 뉴럴 네트워크는 2 스텝 이지의 회귀 또는 분류 판단 규칙이다. 뉴럴 네트워크는 가중치의 층에 의해 출력 유닛들의 층에 접속된 입력 유닛 (및 바이어스) 의 층을 포함하는 층구조를 갖는다. 회귀에 있어서, 출력 유닛의 층은 통상적으로 단지 하나의 출력 유닛만을 포함한다. 그러나, 뉴럴 네트워크는 다수의 양적 반응들을 심리스 방식 (seamless fashion) 으로 핸들링할 수 있다.

[0186] 다층 뉴럴 네트워크에는, 입력 유닛 (입력 층), 숨겨진 유닛 (숨겨진 층), 및 출력 유닛 (출력 층) 이 존재한다. 또한, 입력 유닛 이외의 각각의 유닛에 접속되는 단일의 바이어스 유닛이 존재한다. 뉴럴 네트워크는 Duda *et al.*, 2001, *Pattern Classification*, Second Edition, John Wiley & Sons, Inc., New York; 및 Hastie *et al.*, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York 에서 설명된다.

뉴럴 네트워크는 또한 Draghici, 2003, *Data Analysis Tools for DNA Microarrays*, Chapman & Hall/CRC; 및 Mount, 2001, *Bioinformatics: sequence and genome analysis*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 에서 설명된다. 아래에 개시되는 것은 몇몇 뉴럴 네트워크의 예시적인 형식이다.

[0187] 뉴럴 네트워크의 사용에 대한 기본적인 접근법은 비훈련 네트워크와 함께 시작하는 것이고, 입력 층에 대한 훈련 패턴을 제공하는 것이며, 네트를 통해 신호들을 통과시켜서 출력 층에서의 출력을 결정하는 것이다. 그 후, 이를 출력물을 타겟 값과 비교하며; 임의의 차이점이 에러에 상응한다. 이 에러 또는 기준 함수는 가중치의 스칼라 함수이며, 네트워크 출력들이 원하는 출력들에 매칭할 때 최소화된다. 따라서, 가중치들은 이 에러의 측정을 감소시키도록 조정된다. 회귀에 있어서, 이 에러는 합-제곱 에러일 수 있다. 분류에 있어서, 이 에러는 제곱 에러 또는 크로스-эн트로피 (편차) 중 어느 일방일 수 있다. 예컨대, Hastie *et al.*, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York 을 참조한다.

[0188] 공통적으로 사용되는 3 개의 훈련 프로토콜은 확률적, 일괄적, 및 온라인이다. 확률적인 훈련에서, 패턴은 훈련 세트로부터 무작위로 선택되며, 네트워크 가중치가 각 패턴 표현에 대해 업데이트된다. 확률적 역전달 (back-propagation) 과 같은 경사하강법에 의해 훈련되는 다층 비선형 네트워크들은, 네트워크 토플로지에 의해 규정된 분류기에서의 가중된 값들에 대한 최대 우도 평가를 수행한다. 일괄적인 훈련에서, 습득이 발생하기 전에 네트워크에 모든 패턴들이 제공된다. 통상적으로, 일괄적인 훈련에서, 몇몇 통과는 훈련 데이터를 통

해 이루어진다. 온라인 훈련에서, 각 패턴은 한번 그리고 네트에게 오직 한번만 (once and only once) 제공된다.

[0189] 일부 구현예에서는, 가중치에 대한 시작 값들을 고려한다. 가중치가 거의 0이라면, 뉴럴 네트워크의 숨겨진 층에서 공통적으로 사용되는 S자형의 동작부가 개략 선형 (roughly linear) 이므로 뉴럴 네트워크는 대략 선형 분류기로 이탈한다 (예컨대, Hastie *et al.*, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York 참조). 일부 구현예에서, 가중치의 시작값은 0에 가까운 무작위 값이 되도록 선택된다.

따라서, 분류기는 거의 선형을 나타내며, 가중치가 증가됨에 따라 비선형으로 된다. 개별 유닛은 방향으로 지역화되고 필요로 하는 곳에 비선형성을 도입한다. 정확한 0 가중치의 사용은 0 유도체들 및 완전한 대칭으로 인도하고, 알고리즘은 결코 움직이지 않는다. 이와 달리, 큰 가중치와 함께 시작하는 것은 종종 열악한 솔루션들로 인도한다.

[0190] 입력들의 스케일링 (scaling)은 하부 층에서의 가중치의 효과적인 스케일링을 결정하므로, 최종 솔루션의 품질에 큰 영향을 미칠 수도 있다. 따라서, 일부 구현예에서는, 처음에는 모든 발현 값이 하나의 평균 0 및 표준 편차로 표준화된다. 이것은 모든 입력들이 정규화 프로세스에서 동등하게 처리되는 것을 확보하며, 어느 하나로 하여금 무작위 시작 가중치에 대한 의미있는 범위를 선택하게 한다. 표준화 입력으로는, 범위 [-0.7, +0.7]에 걸쳐 무작위의 균일한 가중치를 취하는 것이 통상적이다.

[0191] 3 층 네트워크의 사용에서 반복되는 문제는 네트워크에서 사용될 숨겨진 유닛의 최적의 수이다. 3 층 네트워크의 입력 및 출력의 수는 해결될 문제에 의해 결정된다. 본 발명에서, 소정의 뉴럴 네트워크에 대한 입력의 수는 훈련 집단으로부터 선택된 바이오마커의 수와 같을 것이다. 뉴럴 네트워크에 대한 출력의 수는 통상적으로 딱 1 개일 것이다.

[0192] 그러나, 일부 구현예에서, 딱 2 개보다 큰 수의 상태들이 네트워크에 의해 규정될 수 있도록, 하나보다 많은 출력이 이용된다. 예를 들어, 다출력 (multi-output) 뉴럴 네트워크는 건강한 표현형, 정동 장애의 각종 스테이지를 구별하는데 이용될 수 있다. 너무 많은 숨겨진 유닛들이 뉴럴 네트워크에서 이용되는 경우, 네트워크는 너무 많은 자유도를 가질 것이고, 너무 오래 훈련되고, 네트워크가 데이터를 오버핏 (overfit) 할 위험이 있다. 너무 적은 숨겨진 유닛들이 존재하는 경우, 훈련 세트는 습득될 수 없다. 그러나, 일반적으로 말하면, 너무 적은 것보다는 너무 많은 숨겨진 유닛을 갖는 것이 더 유리하다. 너무 적은 숨겨진 유닛을 이용하는 경우에는, 분류기는 데이터에서의 비선형성을 캡처하기에 충분히 유동적이지 않으며; 너무 많은 숨겨진 유닛들을 이용하는 경우에는, 이하에서 설명되는 바와 같이, 적절한 정규화 또는 프루닝이 이용된다면 여분 가중치가 0으로 감소하게 될 수 있다. 통상적인 구현예에서, 입력의 수와 훈련 케이스의 수와 함께 증가하는 수를 이용하면, 숨겨진 유닛의 수는 5 내지 100의 범위 중 어딘가에 해당한다.

[0193] 사용할 숨겨진 유닛의 수를 결정하는 하나의 일반적인 접근법은 정규화 접근법을 적용하는 것이다. 정규화 접근법에서, 전통적인 훈련 에러뿐만 아니라 분류기 복잡성에 의존하는 새로운 기준 함수가 구성된다. 특히, 새로운 기준 함수는 아주 복잡한 분류기에 패널티를 주며; 이 기준에서의 최소치를 검색하는 것은 솔루션들의 제약들 또는 바람직한 특성들을 표현하는 정규화 기간에 훈련 세트 상에 에러를 갖는 훈련 세트 상의 에러를 더하여 밸런싱 (balancing) 하는 것이다:

$$J = J_{\text{pat}} + \lambda J_{\text{reg}}$$

[0195] 파라미터  $\lambda$ 는 정규화를 다소 강력하게 부과하도록 조정된다. 환언하면,  $\lambda$ 에 대한 더 큰 값들은 0을 향하여 가중치를 감소시키는 경향이 있을 것이다: 통상적으로 검증 세트를 갖는 크로스-검증은  $\lambda$ 를 평가하는데 이용된다. 이 검증 세트는 훈련 집단의 무작위의 부차세트를 무시함으로써 획득될 수 있다. 다른 형식들의 패널티, 예를 들어 가중치 제거 패널티 (예컨대, Hastie *et al.*, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York 참조)가 제안되어 왔다.

[0196] 사용할 숨겨진 유닛의 수를 결정하는 다른 접근법은, 가장 적게 필요로 되는 가중치를 제거-프루닝하는 것이다. 하나의 접근법에서, 가장 작은 크기를 갖는 가중치가 제거 (0으로 설정) 된다. 이러한 크기 기반의 프루닝이 작동할 수 있지만, 이것이 최선은 아니고; 때때로 작은 크기들을 갖는 가중치들이 데이터를 습득하고 훈련하는데 중요하다. 일부 구현예에서, 크기 기반의 프루닝 접근법을 이용하는 것 외에도, Wald 통계가 컴퓨팅된다. Wald 통계에서의 근본적인 아이디어는, 이들이 분류기에서의 숨겨진 유닛 (가중치)의 중요성을 평가하는데 이용될 수 있다는 것이다. 그 후, 가장 적은 중요성을 갖는 숨겨진 유닛은 (이들 입력 및 출력 가중치들을 0으로 설정함으로써) 제거된다. 이 관점에서 볼 때 2 개의 알고리즘은, 훈련 에러가 가중치에 의

존하는 방법을 예측하기 위해 2 차 근접법 (second-order approximation) 을 이용하고 훈련 에러에서 가장 작은 증가로 인도하는 가중치를 제거하는 OBD (Optimal Brain Damage) 알고리즘 및 OBS (Optimal Brain Surgeon) 알고리즘이다.

[0197] OBD 와 OBS 는, 네트워크를 가중치  $\mathbf{w}$  에서의 국소 최소 에러로 훈련하고 나서 훈련 에러에서의 가장 작은 증가로 인도하는 가중치를 프루닝하는 동일 기반의 접근법을 공유한다. 완전 가중 벡터  $\delta\mathbf{w}$  에서의 변경에 대한 에러에 있어서 예측된 함수적 증가는 다음과 같다:

$$\delta J = \left( \frac{\partial J}{\partial \mathbf{w}} \right)' \cdot \delta\mathbf{w} + \frac{1}{2} \delta\mathbf{w}' \cdot \frac{\partial^2 J}{\partial \mathbf{w}^2} \cdot \delta\mathbf{w} + O(\|\delta\mathbf{w}\|^3)$$

[0198] 여기서,  $\frac{\partial^2 J}{\partial \mathbf{w}^2}$  는 헤시안 행렬 (Hessian matrix) 이다. 제 1 기간은 에러에서의 국소 최소치에서 0 이 되고 (vanish); 3 번째와 더 높은 순위의 기간들은 무시된다. 하나의 가중치를 삭제하는 제약이 주어진 이 함수를 최소화하는 일반적인 솔루션은 다음과 같다.

$$\delta\mathbf{w} = -\frac{\mathbf{w}_q}{[\mathbf{H}^{-1}]_{qq}} \mathbf{H}^{-1} \cdot \mathbf{u}_q \quad \text{및} \quad L_q = \frac{1}{2} - \frac{\mathbf{w}_q^2}{[\mathbf{H}^{-1}]_{qq}}$$

[0200] 여기서, 가중치  $q$  가 프루닝되고 다른 가중치들이  $\delta\mathbf{w}$  업데이트되는 경우,  $\mathbf{u}_q$  는 가중 공간에서의  $q$  번째 방향을 따르는 단위 벡터이고,  $L_q$  는 가중치  $q$  의 현출성으로의 근접-훈련 에러에서의 증가이다. 이들 식들은  $\mathbf{H}$  의 역 (inverse) 을 요청한다. 이 역행렬을 산출하는 하나의 방법은 작은 값,  $H_0^{-1} = \alpha^{-1} \mathbf{I}$  과 함께 시작하는 것이며, 여기서  $\alpha$  는 사실상 가중 상수인 작은 파라미터이다. 다음으로, 행렬은 아래의 식에 따른 각 패턴을 이용하여 업데이트되며,

$$\mathbf{H}_{m+1}^{-1} = \mathbf{H}_m^{-1} - \frac{\mathbf{H}_m^{-1} \mathbf{X}_{m+1} \mathbf{X}_{m+1}^T \mathbf{H}_m^{-1}}{\frac{n}{\alpha_m} + \mathbf{X}_{m+1}^T \mathbf{H}_m^{-1} \mathbf{X}_{m+1}}$$

식 1

[0201] 여기서, 아래의 기재는 제공되는 패턴에 대응하며,  $\alpha_m$  은  $m$  에 따라 감소한다. 완전 훈련 세트가 제공된 후, 역 헤시안 행렬 (inverse Hessian matrix) 은  $\mathbf{H}^{-1} = \mathbf{H}_n^{-1}$  로 주어진다. 알고리즘 형식에서, OBS 방법은 다음과 같다.

```

begin initialize n_H, w, θ
    train a reasonably large network to minimum error
    do compute H-1 by Eqn. 1
        q* ← arg minq wq2 / (2[H-1]qq) (saliency Lq)
        w ← w -  $\frac{\mathbf{w}_q}{[H^{-1}]_{qq}}$  H-1 eq. (saliency Lq)
    until J(w) > θ
return w
end

```

[0205] 3 번째 줄에서 역 해시안 행렬의 산출은 대각 행렬 (diagonal matrix)에 있어서 특히 단순하기 때문에 OBD 방법은 계산적으로 더 단순화된다.  $\Theta$ 로 초기화되는 기준보다 예러가 더 큰 경우 상기 알고리즘은 종결된다.

다른 접근법은 6 째줄을 변경하여, 가중치의 제거로 인한  $J(\mathbf{w})$ 에서의 변경이 일부 기준 값보다 더 큰 경우에 종결하는 것이다. 일부 구현예에서는, 역전달 뉴럴 네트워크이다. 예를 들어, Abdi, 1994, "A neural network 라이브러리," J. Biol System. 2, 247-283 을 참조한다.

### [0206] 5.5.3 클러스터링 (Clustering)

[0207] 일부 구현예에서, 본 발명의 선택 바이오마커에 대한 특징들은 훈련 세트를 클러스터링하는데 이용된다. 예를 들어, 본 발명에서 설명된 (10 개의 바이오마커에 대응하는) 10 개의 특징들이 이용되는 경우를 고려해본다.

훈련 집단의 각 구성원  $\mathbf{m}$ 은 10 개의 바이오마커의 각각에 대한 특징 값 (예컨대, 발현 값)을 가질 것이다. 훈련 집단에서의 구성원  $\mathbf{m}$ 으로부터의 이러한 값들은 다음과 같은 벡터를 정의하며:

$$\mathbf{X}_{1m} \quad \mathbf{X}_{2m} \quad \mathbf{X}_{3m} \quad \mathbf{X}_{4m} \quad \mathbf{X}_{5m} \quad \mathbf{X}_{6m} \quad \mathbf{X}_{7m} \quad \mathbf{X}_{8m} \quad \mathbf{X}_{9m} \quad \mathbf{X}_{10m}$$

[0208] 여기서,  $X_{im}$ 은 유기체  $\mathbf{m}$ 에서의 i 번째 바이오마커의 발현 수준이다. 훈련 세트에 유기체  $\mathbf{m}$ 이 존재하는 경우, i 바이오마커의 선택은  $\mathbf{m}$  벡터들을 정의할 것이다. 본 발명의 방법들은 벡터들에서 이용되는 매 단일 바이오마커의 각 발현값이 매 단일 벡터  $\mathbf{m}$ 마다 표시되도록 요청되지 않는다는 점에 유의한다. 환언하면, i 번째 바이오마커 중 하나가 발견되는 대상으로부터의 데이터는 여전히 클러스터링하는데 이용될 수 있다. 이러한 경우, 빠진 발현값은 "0" 또는 일부 다른 표준화된 값 중 어느 일방으로 할당된다. 일부 구현예에서, 클러스터링 전, 특징 값은 0 의 평균값 및 단위 변위를 갖도록 정규화된다.

[0209] 훈련 그룹에 걸쳐 유사한 발현 패턴들을 나타내는 훈련 집단의 이를 일원은 함께 클러스터링되는 경향이 있을 것이다. 벡터들이 훈련 집단에서 발견되는 특별 그룹들로 클러스터링하는 경우, 본 발명의 유전자들의 특정 조합은 본 발명의 이 양태에서 우수한 분류기임이 고려된다. 예를 들어, 훈련 집단이 클래스 a: 연구 하에서 정동 장애를 갖지 않는 대상들, 및 클래스 b: 연구 하에서 정동 장애를 갖는 대상들을 포함하는 경우, 이상적인 분류기는 집단을, 클래스 a 를 고유하게 나타내는 일 클러스터 그룹 및 클래스 b 를 고유하게 나타내는 다른 클러스터 그룹을 갖는 2 개의 그룹들로 클러스터링할 것이다.

[0210] 클러스터링은, Duda and Hart, *Pattern Classification and Scene Analysis*, 1973, John Wiley & Sons, Inc., New York (이하 "Duda 1973")의 페이지 211-256에 설명되어 있다. Duda 1973의 섹션 6.7에서 설명되는 바와 같이, 클러스터링 문제는 데이터 세트에서 발견한 자연적 그룹화 중 하나로서 설명된다. 뉴럴 그룹화들을 식별하기 위해 2 개의 이슈들이 제기된다. 첫째로, 2 개의 샘플들 사이에서의 유사성 (또는 비유사성)을 측정하는 방법이 결정된다. 이 측정 기준 (metric) (유사성 측정법)은 하나의 클러스터에서의 샘플들이 다른 클러스터들에서의 샘플들에 대한 것보다는 오히려 서로에 더 가깝게 되는 것을 확보하는데 이용된다. 둘째로, 유사성 측정법을 이용하여 데이터를 클러스트들로 구획하는 메커니즘이 결정된다.

[0211] 유사성 측정법은 Duda 1973의 섹션 6.7에서 논의되고 있으며, 여기서 클러스터링 조사를 시작하는 하나의 방법은 거리 함수를 정의하여 데이터 세트에서의 모든 샘플 쌍들 사이의 거리들의 행렬을 컴퓨팅하는 것이라고 기술되어 있다. 거리가 유사성의 우수한 측정법이라고 한다면, 동일한 클러스터에서의 샘플들 사이의 거리는 상이한 클러스터들에서의 샘플들 사이의 거리보다 상당히 작을 것이다. 그러나, Duda 1973의 페이지 215에 기술된 바와 같이, 클러스터링은 거리 측정 기준의 이용을 요청하지 않는다. 예를 들어, 비계량 (nonmetric) 유사성 함수  $s(x, x')$ 는 2 개의 벡터들  $x$  및  $x'$ 를 비교하는데 이용될 수 있다. 종래에는,  $s(x, x')$ 가 대칭 함수이며, 이 대칭 함수의 값은  $x$  및  $x'$ 가 그런대로 "유사" 한 경우에 크다. 비계량 유사성 함수  $s(x, x')$ 의 예는 Duda 1973의 페이지 216에서 제공된다.

[0212] 일단 데이터 세트에서의 지점들 간에 "유사성" 또는 "비유사성"을 측정하는 방법이 선택된다면, 클러스터링은 데이터의 임의의 구획의 클러스터링 품질을 측정하는 기준 함수를 요청한다. 기준 함수를 극대화하는 데이터 세트의 구획들은 데이터를 클러스터링하는데 이용된다. Duda 1973의 페이지 217을 참조한다. 기준 함수들은 Duda 1973의 섹션 6.8에서 논의되고 있다.

[0213] 보다 최근에는, Duda et al., *Pattern Classification*, 2<sup>th</sup> edition, John Wiley & Sons, Inc. New York 이 발행되었다. 페이지 537-563은 클러스터링을 상세히 설명하고 있다. 클러스터링 기술들에 대한 더 많은

정보는 Kaufman and Rousseeuw, 1990, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, NY; Everitt, 1993, *Cluster analysis* (3d ed.), Wiley, New York, NY; 및 Backer, 1995, *Computer-Assisted Reasoning in Cluster Analysis*, Prentice Hall, Upper Saddle River, New Jersey에서 찾을 수 있다. 본 발명에서 이용될 수 있는 특정의 예시적인 클러스터링 기술들은 계층적 클러스터링 (hierarchical clustering) (최근접 이웃 알고리즘, 가장 먼 이웃 알고리즘 (farthest-neighbor algorithm), 평균 연결 알고리즘, 중심 알고리즘 (centroid algorithm), 또는 제곱합 알고리즘 (sum-of-squares algorithm)), k-평균 클러스터링 (k-means clustering), 퍼지 k-평균 클러스터링 (fuzzy k-means clustering) 알고리즘 및 Jarvis-Patrick 클러스터링을 포함하지만, 이에 한정되는 것은 아니다.

#### [0215] 5.5.4 지원 벡터 머신

본 발명의 일부 구현예에서, 지원 벡터 머신 (SVM)은 본 발명에서 설명된 유전자들의 특징 값을 이용하여 대상들을 분류하는데 이용된다. SVM은 상대적으로 새로운 유형의 습득 알고리즘이다. 예를 들어, Cristianini and Shawe-Taylor, 2000, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge; Boser *et al.*, 1992, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5<sup>th</sup> Annual ACM Workshop on Computational Learning Theory*, ACM Press, Pittsburgh, PA, pp. 142-152; Vapnik, 1998, *Statistical Learning Theory*, Wiley, New York; Mount, 2001, *Bioinformatics: sequence and genome analysis*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, Duda, *Pattern Classification*, Second Edition, 2001, John Wiley & Sons, Inc.; and Hastie, 2001, *The Elements of Statistical Learning*, Springer, New York; and Furey *et al.*, 2000, *Bioinformatics* 16, 906-914을 참조한다. 분류를 위해 SVM이 이용되는 경우, SVM은 소정 세트의 바이너리 라벨링 데이터인 훈련 데이터를, 이들로부터의 최대 거리인 초평면 (hyper-plane)으로 분리한다. 선형 분리가 가능하지 않은 경우에 있어서, SVM은 특정 공간에 대한 비선형 매핑을 자동으로 실현하는 "커널 (kernel) 들"의 기술과 조화되어 동작할 수 있다. 특정 공간에서 SVM에 의해 발견되는 초평면은 입력 공간에서의 비선형 결정 경계에 대응한다.

[0217] 일 접근법에서, SVM이 이용되는 경우, 특정 데이터가 평균 0 및 단위 변위를 갖도록 표준화되고 훈련 집단의 일원은 훈련 세트 및 시험 세트로 무작위로 분할된다. 예를 들어, 일 구현예에서, 훈련 집단의 구성원들 중 2/3은 훈련 세트에 배치되고 훈련 집단의 구성원들 중 1/3은 시험 세트에 배치된다. 본 발명에서 설명된 유전자들의 조합에 대한 발현값들은 SVM을 훈련하는데 이용된다. 그 후, 시험 세트에서 구성원들을 정확히 분류하는 훈련된 SVM의 능력이 결정된다. 일부 구현예에서, 이 계산이 문자 마커의 소정 조합으로 수행된다. 계산의 각 반복에서, 훈련 집단의 일원은 훈련 세트 및 시험 세트에 무작위로 배정된다. 그 후, 바이오마커들의 조합의 품질은 SVM 계산의 이러한 각각의 반복의 평균으로서 취급된다.

#### [0218] 5.5.5 관련성 벡터 머신 및 유전자 알고리즘

[0219] 관련성 벡터 머신 (RVM)은, 감독 멀티클래스 분류 문제들뿐만 아니라 회귀에서도 이용가능한 커널 (kernel) 기반의 베이지안 (Bayesian) 통계 모델이다 (참조: M: *Sparse Bayesian Learning and the Relevance Vector Machine*, Journal of Machine Learning Research 1, 2001, 211-244). 훈련된 RVM이 분류 툴로서 이용되는 경우, 훈련된 RVM은 새로운 데이터 지점들의 클래스 구성원 수에 대한 확률적 예측을 행할 수 있다. RVM 모델에서, 미리 규정된 설명적인 변수들 (즉, 유전자들 또는 바이오마커들)의 세트가 로지스틱 링크 (logistic link) 함수를 통해 클래스 구성원 수 확률에 영향을 미친다고 가정해본다. 수많은 후보 변수들로부터 선택된 최적의 설명적인 변수들의 세트를 결정하기 위해, RVM 모델은, 후보 변수들의 상이한 서브세트들 상에서 훈련되고 시험되는 매우 많은 RVM들을 평가하는 유전자 최적화 알고리즘 (Deb, K: *Multi-Objective Optimization using Evolutionary Algorithms*, Wiley, 2001) 내에서 동작하고 있다. 각각의 가변 서브세트의 성능은 크로스-검증을 통해 평가된다.

#### [0220] 5.5.6 다른 데이터 분석 알고리즘

[0221] 전술한 데이터 분석 알고리즘들은 전환체 (converter)들을 비전환체 (nonconverter)들로부터 판별하는 판단 규칙을 구성하는데 이용될 수 있는 방법들의 유형들에 대한 단순 예시에 불과하다. 또한, 전술한 기술들의 조합들이 이용될 수 있다. 결정 트리들과 부스팅 (boosting)의 조합의 이용과 같은 일부 조합들은 설명되었다. 그러나, 다른 많은 조합들이 가능하다. 또한, 사영 추적 (Projection Pursuit)와 같은 기술분야에서의 다른 기술들에서, 각종 투표가 판단 규칙을 구성하는데 이용될 수 있다.

## 5.6 바이오마커

특정 구현예에서, 바이오마커 프로파일은 표 1A에 열거된 2 개 이상의 상이한 바이오마커를 포함한다. 바이오마커 프로파일은 2 개 이상의 바이오마커에 대한 각각의 상응하는 특징을 추가로 포함한다. 이러한 바이오마커는 예를 들어, mRNA 전사체, cDNA 또는 일부 다른 핵산, 예를 들어 증폭된 핵산, 또는 단백질일 수 있다. 일반적으로, 2 개 이상의 바이오마커는 2 개 이상의 상이한 유전자로부터 유도된다. 2 개 이상의 상이한 바이오마커 내 바이오마커가 표 1A에 열거된 경우에서, 바이오마커는 예를 들어, 열거된 유전자로 만들어진 전사체, 이의 보체, 또는 이의 식별 조각 또는 보체, 또는 이의 cDNA, 또는 cDNA의 식별 조각, 또는 전사체 또는 이의 보체의 모두 또는 일부에 상응하는 식별 증폭된 핵산 분자, 또는 유전자에 의해 코딩되는 단백질, 또는 단백질의 식별 조각, 또는 상기 중 임의의 것의 지표일 수 있다. 이러한 구현예에 따르면, 본 발명의 바이오마커 프로파일은 바이오마커를 측정하기 위해 당업자에게 알려진 임의의 표준 어세이, 또는 본원에 기재된 어세이를 사용하여 수득될 수 있다. 이러한 어세이는 예를 들어, 관심의 유전자 (예를 들어, 표 1A에 열거된 유전자)의 특정 유전자 또는 대립유전자의 발현 생성물 (예를 들어, 핵산 및/또는 단백질)을 검출할 수 있다. 하나의 구현예에서, 이러한 어세이는 핵산 마이크로어레이를 이용한다.

일부 구현예에서, 바이오마커 프로파일은 표 1A에 열거된 2 내지 29 개의 바이오마커를 갖는다. 일부 구현예에서, 바이오마커 프로파일은 표 1A에 열거된 3 내지 20 개의 바이오마커를 갖는다. 일부 구현예에서, 바이오마커 프로파일은 표 1A에 열거된 4 내지 15 개의 바이오마커를 갖는다. 일부 구현예에서, 바이오마커 프로파일은 표 1A에 열거된 2 개 이상의 바이오마커를 갖는다. 일부 구현예에서, 바이오마커 프로파일은 표 1A에 열거된 3 개 이상의 바이오마커를 갖는다. 일부 구현예에서, 바이오마커 프로파일은 표 1A에 열거된 4 개 이상의 바이오마커를 갖는다. 일부 구현예에서, 바이오마커 프로파일은 표 1A에 열거된 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25 개 이상, 또는 그 초과의 바이오마커를 갖는다. 일부 구현예에서, 각각의 이러한 바이오마커는 핵산이다. 일부 구현예에서, 각각의 이러한 바이오마커는 단백질이다. 일부 구현예에서, 바이오마커 프로파일 내 바이오마커 중 일부는 핵산이고, 바이오마커 프로파일 내 바이오마커 중 일부는 단백질이다.

## 5.7 특정 구현예

본 발명의 하나의 양상은 정동 장애의 증상을 나타낼 것 같은 대상의 유전자 전사 프로파일을 확인하는 방법에 관한 것이다. 이러한 유전자 전사 프로파일은 표 1A로부터 선택된 유전자와 같은 대상의 생물학적 샘플로부터 선택된 유전자의 전사 분석에 근거한다.

본 발명을 사용하면, 단일 프로파일로 종합될 수 있는 개체의 바이오마커의 풍부함 (예를 들어, 발현 수준)을 확인 및 분석하는 것이 가능하다. 이러한 풍부함 프로파일은 질환 분류에 대한 서명으로서 사용된다. 여기에서 논의되는 바와 같이, 전사 분석은 대조군 대상 및 질환이 있는 대상의 전혈 샘플에서 유전자 발현 프로파일을 측정하기 위해 수행되었다. 표 1A로부터 선택된 유전자의 풍부함은 표 4, 표 5, 및 표 6에 예시된다. 표 4, 표 5, 및 표 6 각각은 대조군과 비교하여 우울증 대상, 중증 우울증 대상, 및 양극성 대상 각각에 대한 유전자 전사 프로파일의 대표적인 예이다. 하나의 구현예에서, 표 4에서 제시된 바와 같은 우울증 유전자 전사 프로파일을 갖는 대상은 우울증을 갖는 것으로 진단된다. 또 다른 구현예에서, 표 5에서 제시된 바와 같은 중증 우울증 유전자 전사 프로파일을 갖는 대상은 중증 우울증을 갖는 것으로 진단된다. 또 다른 구현예에서, 표 6에서 제시된 바와 같은 양극성 유전자 전사 프로파일을 갖는 대상은 양극성 장애를 갖는 것으로 진단된다. 유전자 전사 프로파일의 추가의 대표적인 예는 표 4A 및 5B에 제시된다.

하나의 예에서, 유전자 발현 프로파일을 측정하기 위해 사용된 바이오마커는 표 1A에 기재된 유전자로부터 선택되었다. 대표적인 전사 바이오마커 탐침 세트는 또한 표 1A에 기재된다. 잘 알려진 방법에 의해 정량적 PCR (qPCR) 을 수행하기 위해 탐침 세트가 사용되었다.

본 발명의 양상은 표 1A로부터 선택된 유전자의 전사 분석에 의해 측정된 바와 같이 각각의 대상에 대한 전사 프로파일을 제공한다.

전사 분석은 당업계에 잘 알려진 방법에 의해 수행될 수 있다. 예를 들어, 메신져 RNA (mRNA) 를 비롯한 RNA 는 동물 신체, 특히 인체의 세포 물질, 또는 세포 물질을 함유하는 유액으로부터 단리될 수 있다. 세포 물질이 mRNA 를 비롯한 세포 내용물을 함유한다는 것으로 이해된다. 본 발명에 사용되는 생물학적 샘플은 예를 들어, 말초 조직, 전혈, 뇌척수액, 복막액 및 간질액으로부터 선택될 수 있다.

본 발명의 다른 구현예에서, 생물학적 샘플은 전혈, 뇌척수액, 및 말초 조직으로 이루어지는 군으로부터 선택된

다. 본 발명은 또한 적혈구 세포 (RBC), 백혈구 세포 및 혈소판으로 이루어지는 군으로부터 선택되는 전혈의 분획을 사용하여 수행될 수 있다. 백혈구 세포 (백혈구)에는 중성구, 호염기구, 호산구, 림프구, 대식세포 및 단핵구가 포함되나 이에 제한되지 않는다.

[0232] 샘플 내 유전자 발현을 측정하기 위해, 그 샘플 내의 RNA 또는 mRNA를 역전사에 적용하여 카피 DNA를 제작한 다음, DNA 서열에 근거하여 탐침, 또는 프라이머 서열을 사용하는 표준 방법에 의해 분석할 수 있다. 각각의 개별 유전자는 폴리머라아제 연쇄 반응 (PCR), 정량적 PCR, 제자리 혼성화, 노던 블롯 분석, 고체-지지체 고정 어세이, 예컨대 비이드-기반 어세이 또는 유전자 어레이, 및 당업계에 잘 알려진 기타 방법에 의해 분석될 수 있다.

[0233] 본원에 기재된 본 발명의 양상에 따르면, 정량적 PCR (qPCR)은 mRNA 수준을 측정하는데 사용된다. 하나 이상의 핵산 탐침이 생물학적 샘플로부터 mRNA 수준을 측정하는데 사용되었다. 탐침 또는 프라이머는 관심의 유전자에 상보적인 뉴클레오티드 (nt) 서열이고, 이러한 탐침/프라이머의 선별 및 합성은 당업자에게 잘 알려진 방법에 의해 수행된다. 본 발명의 탐침/프라이머는 표 1A에 기재된 뉴클레오티드 서열에 제한되지 않는다.

[0234] 본 발명은 또한, 대상으로부터 수득된 생물학적 샘플로부터 분석되는 바와 같이 이러한 대상의 전사 프로파일을 측정하여 대조군 대상과 비교하는 질환이 있는 대상의 분류 방법을 제공한다.

[0235] 본 발명은 표 1A로부터 선택된 유전자의 전사 분석에 의해 측정된 구별되는 전사 프로파일을 제공한다. 이러한 전사 프로파일은, 공지된 건강한 대조군 대상 또는 공지된 질환이 있는 대상의 전사 프로파일과 유사한 것으로 측정되는 경우, 대상에서 구별되는 것으로 측정된다. 공지된 건강한 대조군 대상 또는 공지된 질환이 있는 대상의 전사 프로파일과의 유사성은 분류 방법, 예컨대 본원에 기재된 바와 같은 분류 알고리즘에 의해 측정된다.

[0236] 일부 구현예에서, 전사 데이터는 본원에 기재된 바와 같은 다수의 대조군 대상으로부터 수집된다. 전사 데이터는 본원에 기재된 바와 같은 질환 또는 장애, 예컨대 정동 장애를 앓고 있는 다수의 대상으로부터 수집된다. 데이터 분석 알고리즘은 각각의 전사 데이터 세트에 함유되어 있는 분류 유전자를 식별 또는 구별하기 위해 입력 값으로서 각각의 전사 데이터 세트를 사용한다. 이러한 알고리즘은 전형적으로 분류 알고리즘으로서 기술되며, 또한 "분류기"로서도 알려져 있다. 상기 과제를 수행하는데 사용되는 데이터 분석 알고리즘은 당업자에게 잘 알려져 있고, 하기 예가 사용될 수 있다: 랜덤 포레스트 (Breiman, L., 2001, *Machine Learning* 45(1):5-32), 지원 벡터 머신 (Support Vector Machine: SVM) (Cortes, C. and Vapnik, V. 1995, *Machine Learning*, 20(3):273-97), 단계식 로지스틱 회귀 (Stepwise Logistic Regression: SLR) (Ersbøll, B.K. and Conradien, K. (2005) *An Introduction to Statistics*. 7th ed. IMM; Draper, N. and Smith, H. (1981) *Applied Regression Analysis*, 2d Edition, New York: John Wiley & Sons, Inc.), 반복 분할 (recursive partitioning: RPART) (James K.E. et al, 2005, *Statistics in Medicine*, 24 (19): 3019-35), 폐널화 로지스틱 회귀 분석 (Penalized Logistic Regression Analysis: PELORA) (Dettling, M., 2003, Proceedings of the 3<sup>rd</sup> International Workshop on Distributed Statistical Computing, March 20-22, Vienna Austria, Hornick, Leisch and Seilis, eds.), Neural Networks, Relevance Vector Machines (RVM), LogitBoost (Friedman, J., Hastie, T. and Tibshirani, R. 2000, *Annals of Statistics* 28(2):337-407), 마이크로어레이 예측 분석 (Prediction Analysis of Microarrays: PAM), 및 기타 (V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998 참조). 이러한 분류 알고리즘, 또는 "분류기"는 전사 데이터에 근거한 환자의 분류와 관련된 출력물을 제공하기 위해 조정 및 훈련된다.

[0237] 훈련된 분류 알고리즘에 의해 선별된 유전자 또는 바이오마커 분류는 특정 데이터 세트가 속한 부류, 예를 들어, 대조군 데이터와 관련된 부류 또는 질환 데이터와 관련된 부류와 연관되어 있는 전사 데이터의 예측 측정 값을 산출한다.

[0238] 임의의 특정 이론에 구애됨 없이, 랜덤 포레스트 알고리즘은 다수의 결정 트리로부터의 출력물에 근거하여 대상을 분류하는 총체적 습득 방법으로 고려된다. 각각의 결정 트리는 이용 가능한 데이터의 부트스트랩 샘플 상에 훈련되고, 결정 트리 내 각각의 마디는 최고의 설명 변수 (즉, 유전자 또는 바이오마커)에 의해 나뉜다. 랜덤 포레스트는 자동 변수 선택을 제공하고 선택된 변수 간의 비-선형 상호작용을 설명할 수 있다.

[0239] 단계식 로지스틱 회귀 (SLR)는 데이터 입력값을 로지스틱 곡선에 피팅시켜 사건 발생 가능성을 예측하는 통계 모델로 고려된다. 로지스틱 모델에서 미리 정의된 세트의 설명 변수 (즉, 유전자 또는 바이오마커)가 로지스틱 연계 함수를 통해 확률에 영향을 주는 것으로 가정한다. 다수의 후보자 변수로부터 선택된 설명 변수

의 최적 세트를 결정하기 위해, 다수의 로지스틱 회귀 모델은 단계식 방식의 초기 모델로부터 설립되고, 가장 정확한 모델을 측정하기 위해 아카이케 정보 기준 (Akaike Information Criteria: AIC) 평가를 통해 비교한다 (Burnham, K. P., and D. R. Anderson, 2002. *Model Selection and Multimodel Inference: A Practical-Theoretic Approach*, 2nd ed. Springer-Verlag).

[0240] 지원 벡터 머신 (SVM) 은 일반화된 선형 분류기 계열에 속하는 것으로 고려된다. 2 그룹 분류에서 입력값 데이터를 n 평면 공간에서 2 개 세트의 벡터로서 보면, SVM 은 2 개 세트의 벡터 사이의 가장자리를 최대화하는, 초평면에 의해 데이터를 분류한다. 최대 초평면에 대해 최소 거리를 갖는 벡터를 지지체 벡터라고 부른다. SVM 은 자동 변수 (즉, 유전자 또는 바이오마커) 선택을 제공하지 않는다.

[0241] 관련성 벡터 머신 (RVM) 은 미리 정의된 세트의 설명 변수 (즉, 유전자 또는 바이오마커) 가 로지스틱 연계 함수를 통해 부류 소속 확률에 영향을 주는 것으로 가정한다. RVM 은 다수의 후보자 변수로부터 선택된 설명 변수의 최적 세트를 결정하는 것을 추구한다. RVM 은 많은 RVM 을 평가하고 교차 입증하고, 후보자 변수 (즉, 유전자 또는 바이오마커) 의 최적 세트를 선택하는 유전학적 최적화 알고리즘으로 작동될 수 있다.

[0242] 분류 알고리즘으로 설립된 전사 프로파일은 상기 언급된 데이터 분석 알고리즘 중 하나를 사용하여 추가로 훈련된다. 분류 에러는 훈련된 분류 알고리즘이 부류 내 소속을 예측하는 것에 대한 정확성의 측정값이다. 분류 에러는 교차 검증법, 예컨대 1 개를 남기는 (leave-one-out) 교차 검증 (LOOCV), K 배 검증, 또는 10 배 검증에 의해 측정될 수 있다 (Devijver, P. A., and J. Kittler, 1982, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, London).

[0243] 미리 기술된 전사 프로파일을 가진 알고리즘의 정확성은 훈련 동안 알고리즘에 의해 예측되는 진 양성 (TP), 진 음성 (TN), 가 양성 (FP), 및 가 음성 (FN) 의 수를 측정하여 측정될 수 있다. 정확성은 다음과 같이 측정된다:

$$\text{정확성} = (TP + TN) / (TP + TN + FP + FN)$$

[0245] 양성 예측 값 (PPV), 또는 알고리즘에 의해 양성으로 점수를 받은 질환이 있는 대상의 % 는 다음과 같이 측정된다:

$$\text{PPV} = TP / (TP + FP)$$

[0247] 음성 예측 값 (NPV), 또는 알고리즘에 의해 음성으로 점수를 받은 대조군 대상 (질환이 없는) 의 % 는 다음과 같이 측정된다:

$$\text{NPV} = TN / (TN + FN)$$

[0249] 분류 알고리즘의 수행은 또한 분류가 올바른 변수 (즉, 유전자) 를 얼마나 잘 확인하는지를 평가하는 잭카드 (Jaccard) 유사 계수 (Jaccard Index) 에 의해 측정된다. 훈련된 분류 알고리즘의 정확성은 약 60%, 65%, 70%, 75%, 80%, 85%, 90% 또는 95% 초과일 수 있다. 훈련된 분류 알고리즘의 잭카드 지수 (Jaccard Index) 는 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 또는 95% 초과일 수 있다. 훈련된 분류 알고리즘의 PPV 및 NPV 는 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 또는 95% 초과일 수 있다.

[0250] 대상의 분류는 정동 장애를 갖는 또는 정동 장애의 증상을 나타낼 것 같은 대상의 진단에 유용할 수 있다. 대상의 분류를 위한 유전자 전사 프로파일은 표 1A 의 유전자의 전사 분석에 근거한다. 본원에 기재된 방법에 의해 분석되는 바와 같은 대상의 전사 프로파일은 대상이 질환이 있는 대상의 부류에 속하는 지의 여부에 대한 지표일 것이다.

[0251] 일부 구현예에서, 본 발명은 시험 대상의 바이오마커 프로파일 내 다수의 바이오마커의 다수의 특징이 값 세트를 만족하는지의 여부를 평가하는 것을 포함하며, 값 세트를 만족하는 것이 시험 대상이 상기 정동 장애를 갖는다는 것을 예측하고, 다수의 특징은 다수의 바이오마커의 측정 가능한 양상으로, 상기 다수의 바이오마커는 표 1A 에 열거된 2 개 이상의 바이오마커를 포함하는, 시험 대상 내 정동 장애의 진단 방법을 제공한다. 본 발명은 또한, 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 인터페이스 소자, 모니터, 유형 컴퓨터 판독 가능 저장 매체, 또는 지역적 또는 원격 컴퓨터 시스템으로 출력하는 것; 또는 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 판독 가능 형태로 표시하는 것을 포함한다.

[0252] 본 발명의 일부 구현예에서, 다수의 바이오마커는 표 1A 에 열거된 2 내지 29 개의 바이오마커로 이루어진다. 다른 구현예에서, 다수의 바이오마커는 표 1A 에 열거된 3 내지 20 개의 바이오마커로 이루어진다. 또

다른 구현예에서, 다수의 바이오마커는 표 1A에 열거된 2, 3, 4 또는 5 개 이상의 바이오마커를 포함한다.

[0253] 일부 구현예에서, 다수의 특징은 표 1A에 열거된 2 내지 29 개의 바이오마커에 상응하는 2 내지 29 개의 특징으로 이루어진다. 다른 구현예에서, 다수의 특징은 표 1A에 열거된 3 내지 15 개의 바이오마커에 상응하는 3 내지 15 개의 특징으로 이루어진다. 또 다른 구현예에서, 다수의 특징은 표 1A에 열거된 2 개 이상의 바이오마커에 상응하는 2 개 이상의 특징을 포함한다.

[0254] 다른 구현예에서, 다수의 바이오마커는 ERK1 및 MAPK14를 포함한다. 다른 구현예에서, 다수의 바이오마커는 Gi2 및 IL-1b를 포함한다. 다른 구현예에서, 다수의 바이오마커는 ARRB1 및 MAPK14를 포함한다. 다른 구현예에서, 다수의 바이오마커는 ERK1 및 IL1b를 포함한다.

[0255] 본 발명의 일부 양상에서, 상기 다수의 바이오마커 내 각각의 바이오마커는 핵산이다. 다른 양상에서, 상기 다수의 바이오마커 내 각각의 바이오마커는 DNA, cDNA, 증폭된 DNA, RNA, 또는 mRNA이다. 또 다른 양상에서, 상기 다수의 바이오마커 내 각각의 바이오마커는 단백질이다.

[0256] 다른 구현예에서, 시험 대상의 바이오마커 프로파일 내 상기 다수의 특징 내 특징은 다수의 바이오마커 내 바이오마커의 측정가능한 양상이고, 상기 특징에 대한 특징 값은 상기 시험 대상으로부터 취득된 생물학적 샘플을 사용하여 측정된다. 다른 구현예에서, 특징은 생물학적 샘플 내 상기 바이오마커의 풍부함이다. 또 다른 구현예에서, 생물학적 샘플은 말초 조직, 전혈, 뇌척수액, 복막액, 간질액, 적혈구 세포, 백혈구 세포 또는 혈소판이다.

[0257] 또 다른 구현예에서, 상기 다수의 특징 내 특징은 상기 바이오마커 프로파일 내 바이오마커의 측정가능한 양상이고, 상기 특징에 대한 특징 값은 상기 시험 대상으로부터 채취된 샘플을 사용하여 측정된다. 일부 구현예에서, 바이오마커 프로파일 내 바이오마커는 핵산의 지표 또는 단백질의 지표이다. 다른 구현예에서, 바이오마커 프로파일 내 바이오마커는 mRNA 분자의 지표 또는 cDNA 분자의 지표이다. 일부 구현예에서, mRNA 분자 또는 cDNA 분자의 지표는 카페/ng cDNA와 같은 전사체 값이다. 다른 구현예에서, 바이오마커 프로파일 내 제 1 바이오마커는 핵산의 지표이고, 바이오마커 프로파일 내 제 2 바이오마커는 단백질의 지표이다.

[0258] 본 발명의 일부 양상에서, 값 세트는 표 4에 언급된 바와 같은 바이오마커의 풍부함을 포함하고, 표 4의 값 세트를 만족하는 것이 대상이 우울증을 갖는다는 것을 예측한다. 다른 양상에서, 값 세트는 표 5에 언급된 바와 같은 바이오마커의 풍부함을 포함하고, 표 5의 값 세트를 만족하는 것이 대상이 중증 우울증을 갖는다는 것을 예측한다. 다른 양상에서, 값 세트는 표 6에 언급된 바와 같은 바이오마커의 풍부함을 포함하고, 표 6의 값 세트를 만족하는 것이 대상이 양극성 우울증을 갖는다는 것을 예측한다. 또한, 본 발명은 표 4A에서와 같은 우울증의 진단을 위한 값 세트 및 표 5B에서와 같은 중증 우울증의 진단을 위한 값 세트를 제공한다.

[0259] 표 4, 5 및 6에 묘사된 값 세트는 바이오마커의 풍부함 (카페/ng cDNA), 즉, 바이오마커 유전자의 전사체에 의해 나타내진다. 예를 들어, 표 4의 바이오마커 ARRB1에 대한 우울증 대상에 대한 전사체 값의 범위는 189062 ± 62727 카페/ng cDNA이고, 이것은 126335 내지 251789 카페/ng cDNA 범위와 동등하다. 표 4의 바이오마커 CD8a에 대한 우울증 대상에 대한 전사체 값의 범위는 8304 ± 5825 카페/ng cDNA이고, 이것은 2479 내지 14129 카페/ng cDNA 범위와 동등하다. 본 발명의 일부 양상에서, 값 세트를 만족하는 것은 각각의 바이오마커에 대한 제시된 범위 내에 값을 갖는다는 것을 의미한다.

[0260] 일부 구현예에서, 15148 내지 35504 카페/ng cDNA의 범위 내의 ERK1의 풍부함 및 39241 내지 107071 카페/ng cDNA의 범위 내의 MAPK14의 풍부함을 포함하는 값 세트는 대상이 우울증을 갖는다는 것을 예측한다. 다른 구현예에서, 61734 내지 168500 카페/ng cDNA의 범위 내의 Gi2의 풍부함 및 15939 내지 43323 카페/ng cDNA의 범위 내의 IL1b의 풍부함을 포함하는 값 세트는 대상이 우울증을 갖는다는 것을 예측한다. 다른 구현예에서, 126335 내지 251789 카페/ng cDNA의 범위 내의 ARRB1의 풍부함 및 39241 내지 107071 카페/ng cDNA의 범위 내의 MAPK14의 풍부함을 포함하는 값 세트는 대상이 우울증을 갖는다는 것을 예측한다. 다른 구현예에서, 15148 내지 35504 카페/ng cDNA의 범위 내의 ERK1의 풍부함 및 15939 내지 43323 카페/ng cDNA의 범위 내의 IL1b의 풍부함을 포함하는 값 세트는 대상이 우울증을 갖는다는 것을 예측한다.

[0261] 다른 구현예에서, 범위 0.25 내지 0.45 내의, MAPK14의 풍부함으로 나뉘어진 ERK1의 풍부함의 비를 포함하는 값 세트는 대상이 우울증을 갖는다는 것을 예측한다. 다른 구현예에서, 범위 0.16 내지 0.36 내의, IL1b의 풍부함으로 나뉘어진 Gi2의 풍부함의 비를 포함하는 값 세트는 대상이 우울증을 갖는다는 것을 예측한다. 다른 구현예에서, 범위 0.29 내지 0.49 내의, ARRB1의 풍부함으로 나뉘어진 MAPK14의 풍부함의 비를 포함하는

값 세트는 대상이 우울증을 갖는다는 것을 예측한다. 다른 구현예에서, 범위 0.0.75 내지 0.95 내의, IL1b 의 풍부함으로 나뉘어진 ERK1 의 풍부함의 비를 포함하는 값 세트는 대상이 우울증을 갖는다는 것을 예측한다.

[0262] 다른 구현예에서, 범위 0.19 내지 0.39 내의, MAPK14 의 풍부함으로 나뉘어진 ERK1 의 풍부함의 비를 포함하는 값 세트는 대상이 중증 우울증을 갖는다는 것을 예측한다. 다른 구현예에서, 범위 0.18 내지 0.38 내의, IL1b 의 풍부함으로 나뉘어진 Gi2 의 풍부함의 비를 포함하는 값 세트는 대상이 중증 우울증을 갖는다는 것을 예측한다. 다른 구현예에서, 범위 0.32 내지 0.52 내의, ARRB1 의 풍부함으로 나뉘어진 MAPK14 의 풍부함의 비를 포함하는 값 세트는 대상이 중증 우울증을 갖는다는 것을 예측한다. 다른 구현예에서, 범위 0.60 내지 0.80 내의, IL1b 의 풍부함으로 나뉘어진 KRK1 의 풍부함의 비를 포함하는 값 세트는 대상이 중증 우울증을 갖는다는 것을 예측한다.

[0263] 상기 방법의 다른 양상에서, 방법은 평가 단계 전에, 상기 바이오마커 프로파일을 구축하는 것을 추가로 포함한다. 다른 구현예에서, 구축 단계는 상기 시험 대상의 생물학적 샘플로부터 상기 다수의 특징을 수득하는 것을 포함한다. 일부 양상에서, 바이오마커 프로파일은 제 2 바이오마커의 특징 값에 의해 제 1 바이오마커의 특징 값을 나누어 바이오마커의 풍부함의 비를 측정함으로써 구축된다. 이러한 바이오마커 프로파일은 표 4, 표 5 또는 표 6에 제시된 값을 사용하여 구축될 수 있다. 다른 구현예에서, 샘플은 말초 조직, 전혈, 뇌척수액, 복막액, 간질액, 적혈구 세포, 백혈구 세포 또는 혈소판이다.

[0264] 상기 방법의 또 다른 양상에서, 방법은 평가 단계 전에, 상기 제 1 값 세트를 구축하는 것을 추가로 포함한다. 다른 구현예에서, 구축 단계는 데이터 분석 알고리즘을 집단의 일원으로부터 수득된 특징에 적용하는 것을 포함한다.

[0265] 일부 양상에서, 특징은 ERK1 및 MAPK14 를 포함하는 바이오마커의 측정가능한 양상이고, 특징 값은 상기 시험 대상으로부터 채취된 혈액 샘플을 사용하여 측정된다.

[0266] 다른 구현예에서, 집단은 정동 장애를 갖지 않은 제 1 의 다수의 대조군 대상으로부터의 제 1 의 다수의 생물학적 샘플 및 정동 장애를 갖는 제 2 의 다수의 대상으로부터의 제 2 의 다수의 생물학적 샘플을 포함한다. 또 다른 구현예에서, 데이터 분석 알고리즘은 결정 트리, 마이크로어레이의 예측 분석, 다중 합산적 회귀 트리, 뉴럴 네트워크, 클러스터링 알고리즘, 주요 성분 분석, 최근접 이웃 분석, 선형 판별식 분석, 이차 판별식 분석, 지원 벡터 머신, 진화 방법, 관련성 벡터 머신, 유전학적 알고리즘, 투사 지향, 또는 가중치 투표이다.

[0267] 또 다른 구현예에서, 구축 단계는 판단 규칙을 생성하고, 상기 평가 단계는 제 1 값 세트를 만족하는지를 측정하기 위해 상기 판단 규칙을 다수의 특징에 적용하는 것을 포함한다. 일부 구현예에서, 판단 규칙은 상기 집단에서 대상을 (i) 정동 장애를 갖지 않는 대상 및 (ii) 정동 장애를 갖는 대상으로서 70% 이상의 정확성으로 분류한다. 다른 구현예에서, 판단 규칙은 상기 집단에서 대상을 (i) 정동 장애를 갖지 않는 대상 및 (ii) 정동 장애를 갖는 대상으로서 90% 이상의 정확성으로 분류한다.

[0268] 본 발명의 특정 양상에서, 정동 장애는 양극성 장애 I, 양극성 장애 II, 기분저하 장애, 또는 우울 장애이다. 다른 양상에서, 정동 장애는 경증 우울증, 중등 우울증, 중증 우울증, 비정형 우울증, 우울성 우울증, 또는 경계 인격 장애이다. 또 다른 양상에서, 정동 장애는 (i) 외상 후 스트레스 장애 또는 (ii) 외상 후 스트레스 장애가 없는 트라우마이다. 일부 양상에서, 정동 장애는 급성 외상 후 스트레스 장애 또는 완화된 외상 후 스트레스 장애이다.

[0269] 본 발명은 시험 대상에서 정동 장애를 진단하기 위해 사용되는 키트를 제공하며, 키트는 시험 대상의 바이오마커 프로파일 내 다수의 바이오마커의 다수의 특징이 값 세트를 만족하는지를 평가하기 위한 시약 및 지침을 포함하고, 값 세트를 만족하는 것은 시험 대상이 상기 정동 장애를 갖는지를 예측하고, 다수의 특징은 다수의 바이오마커의 측정가능한 양상이고, 다수의 바이오마커는 표 1A에 열거된 2 개 이상의 바이오마커를 포함한다.

일부 양상에서, 시약은 표 1A로부터 선택된 바이오마커의 뉴클레오티드 서열을 인지하는 탐침 및/또는 프라이머를 포함한다. 본 발명의 키트는 본 발명에 따른 바이오마커 프로파일을 생성하는데 사용된다. 일부 양상에서, 본 발명의 키트는 표 1A에 열거된 2 개 이상의 바이오마커를 포함하는 다수의 바이오마커로부터 시험 대상의 바이오마커 프로파일을 시험 및 평가하기 위한 지침을 제공한다. 다른 양상에서, 본 발명의 키트는 시험 대상의 바이오마커 프로파일이 이러한 값 세트를 만족하는지를 측정하기 위해 값 세트를 함유하는 지침을 제공한다.

[0270] 본 발명은 또한 컴퓨터 관독가능 저장 매체 및 그곳에 저장된 컴퓨터 프로그램 메카니즘을 포함하고, 상기 컴퓨터 프로그램 메카니즘은 상기 방법 중 임의의 것을 수행하기 위한 지침을 포함하는 컴퓨터 프로그램 제품을 제

공한다. 일부 구현예에서, 컴퓨터 프로그램 메카니즘은 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 인터페이스 소자, 모니터, 유형 컴퓨터 판독가능 저장 매체, 또는 지역적 또는 원격 컴퓨터 시스템으로 출력하는 것; 또는 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 판독가능 형태로 표시하는 것을 위한 지침을 추가로 포함한다.

[0271] 본 발명은 또한 하나 이상의 프로세서; 하나 이상의 프로세서에 연결된 메모리를 포함하고, 상기 메모리는 상기 방법 중 임의의 것을 수행하기 위한 지침을 저장하는 컴퓨터를 제공한다. 본 발명의 일부 양상에서, 메모리는 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 인터페이스 소자, 모니터, 유형 컴퓨터 판독가능 저장 매체, 또는 지역적 또는 원격 컴퓨터 시스템으로 출력하는 것; 또는 시험 대상이 정동 장애를 갖는 지의 진단을 사용자 판독가능 형태로 표시하는 것을 위한 지침을 추가로 포함한다.

[0272] 본 발명은 시험 대상이 정동 장애의 증상을 나타내는 가망성을 측정하는 방법을 추가로 제공하고, 상기 방법은, 시험 대상의 바이오마커 프로파일 내 다수의 바이오마커의 특징이 값 세트를 만족하는지를 평가하는 것을 포함하고, 값 세트를 만족하는 것이 시험 대상이 정동 장애의 증상을 나타내는 가망성을 제공하고, 다수의 특징이 다수의 바이오마커의 측정가능한 양상이고, 다수의 바이오마커는 표 1A에 열거된 2 개 이상의 바이오마커를 포함한다.

[0273] 일부 구현예에서, 다수의 바이오마커는 ERK1 및 MAPK14를 포함한다. 다른 구현예에서, 다수의 바이오마커는 Gi2 및 IL-1b를 포함한다. 다른 구현예에서, 다수의 바이오마커는 ARRB1 및 MAPK14를 포함한다. 다른 구현예에서, 다수의 바이오마커는 ERK1 및 IL1b를 포함한다.

[0274] 본 발명의 일부 구현예에서, 다수의 바이오마커는 ERK1, PBR 및 MAPK14를 포함한다. 또 다른 구현예에서, 다수의 바이오마커는 PBR, Gi2 및 IL1b를 포함한다. 다른 구현예에서, 다수의 바이오마커는 ERK1, ARRB1 및 MAPK14를 포함한다. 일부 구현예에서, 다수의 바이오마커는 MAPK14, ERK1 및 CD8b를 포함한다. 다른 구현예에서, 다수의 바이오마커는 MAPK14, ERK1 및 P2X7을 포함한다. 또 다른 구현예에서, 다수의 바이오마커는 ARRB1, IL6 및 CD8a를 포함한다. 특정 구현예에서, 다수의 바이오마커는 ARRB1, ODC1 및 P2X7을 포함한다.

[0275] 또 다른 구현예에서, 방법은 시험 대상이 정동 장애의 증상을 나타내는 가망성을 사용자 인터페이스 소자, 모니터, 유형 컴퓨터 판독가능 저장 매체, 또는 지역적 또는 원격 컴퓨터 시스템으로 출력하는 것; 또는 시험 대상이 정동 장애의 증상을 나타내는 가망성을 사용자 판독가능 형태로 표시하는 것을 추가로 포함한다.

[0276] 본 발명은 다수의 대조군 대상으로부터 수집된 각각의 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 다수의 우울증 대상, 중증 우울증 대상, 또는 양극성 대상으로부터 수집된 각각의 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 또한, 다수의 경계 인격 장애 대상으로부터 수집된 각각의 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 다수의 PTSD 대상으로부터 수집된 각각의 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다.

[0277] 본 발명은 또한 예를 들어 데이터베이스에 저장된, 제 1의 다수의 대조군 대상의 집합적 측정값을 포함하는 전사 프로파일을 제공한다. 제 2의 다수의 대상, 예를 들어, 질환이 있는 대상의 집합적 측정값을 포함하는 전사 프로파일을 데이터 분석 알고리즘, 특히 훈련된 분류 알고리즘을 사용하는 제 1의 다수의 대조군 대상의 전사 프로파일과 비교한다. 훈련된 분류 알고리즘은 각각의 세트의 대상을 분류한다. 훈련된 분류 알고리즘은 분류를 진단 및 배정하는데 유용한 예측 값을 제공한다. 훈련된 분류 알고리즘은 대상이 장애의 질환을 나타낼 것인 가망성을 예측하는데 유용한 예측 값을 제공한다.

[0278] 본 발명의 또 다른 구현예는 질환 또는 장애에 대한 대상의 감수성을 진단 또는 예측하는 것 또는 건강한 대조군 대상 및 질환이 있는 대상과 비교하여 대상의 구별되는 전사 프로파일에 근거한 장애의 증상을 나타내는 가망성을 예측하는 것에 관한 것이다. 진단 용도를 위한 유전자 전사 프로파일은 표 IA로부터 선택된 유전자의 전사 분석에 근거한다.

[0279] 본 발명의 하나의 양상은 상이한 유형의 정동 장애, 특히 주요 우울 장애, 양극성 장애, 경계 인격 장애, 및 외상 후 스트레스 장애의 진단에 관한 것이다.

[0280] 본 발명의 또 다른 양상은 전사 프로파일을 확인하여 환자 집단을 차별화시키는 것에 관한 것이다. 예를 들어, 보통 주요 우울증으로 진단받을 것 같은 환자는, 전사 프로파일에 의해 우울증의 부차유형, 예를 들어 우울성 및 비정형 우울증으로 세분화될 수 있다. 상기 부차유형의 우울증에 대한 상이한 치료 반응에 대한 증거

가 있다. 동반이환을 나타내는, 즉, 1 개 초과의 장애에 대해 DSM-IV® 기준을 충족하는 환자는, 전사 프로파일의 확인으로부터 이득을 볼 것이다. 전사 프로파일은 하나의 장애에 대한 공통의 생물학적 근거를 확인 할 수 있다.

[0281] 상기 방법의 예로서, 본 발명은 하나의 구현예에서, 다수의 건강한 대조군 대상으로부터 수집된 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 또한 다수의 정동 장애 대상으로부터 수집된 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 예를 들어, 본 발명은 또한 다수의 우울증, 중증 우울증, 또는 양극성 대상으로부터 수집된 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 표 4 의 다수의 우울증 대상으로부터 수집된 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 표 5 의 다수의 중증 우울증 대상으로부터 수집된 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 또한 표 6 의 다수의 양극성 대상으로부터 수집된 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 또한 다수의 경계 인격 장애 대상으로부터 수집된 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다. 본 발명은 다수의 PTSD 대상으로부터 수집된 생물학적 샘플에 대한 전사 분석의 측정값인 전사 프로파일을 제공한다.

[0282] 본 발명의 하나의 구현예에서, 생물학적 샘플은 전혈이다.

[0283] 본 발명은 또한 예를 들어 데이터베이스에 저장된, 제 1 의 다수의 대조군 대상의 집합적 측정값을 포함하는 전사 프로파일을 제공한다. 제 2 의 다수의 대상, 예를 들어, 질환이 있는 대상의 집합적 측정값을 포함하는 전사 프로파일을 분류 알고리즘을 사용하는 제 1 의 다수의 대조군 대상의 전사 프로파일과 비교한다. 분류 알고리즘은 각각의 대상을 분류하는 출력물을 제공한다.

[0284] 본 발명의 일부 양상에서, 전사 프로파일은 ADA, ARRB1, ARRB2, CD8a, CD8b, CREB1, CREB2, DPP4, ERK1, ERK2, Gi2, Gs, GR, IL1b, IL6, IL8, INDO, MAPK14, MAPK8, MKP1, MR, ODC1, P2X7, PBR, PREP, RGS2, S100A10, SERT 및 VMAT2 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다.

[0285] 또 다른 구현예에서, 전사 프로파일은 ADA, ARRB1, ARRB2, CD8a, CD8b, CREB1, CREB2, DPP4, ERK1, ERK2, Gi2, Gs, GR, IL1b, IL6, IL8, INDO, MAPK14, MAPK8, MKP1, MR, ODC1, P2X7, PBR, PREP, RGS2, S100A10, SERT 및 VMAT2 로 이루어진 군으로부터 선택되는 3 개 이상의 유전자의 전사 분석으로부터 측정된다.

[0286] 일부 구현예에서, 전사 프로파일은 ARRB1, ARRB2, CD8a, CREB1, CREB2, ERK2, Gi2, MAPK14, ODC1, P2X7, 및 PBR 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다.

[0287] 또 다른 구현예에서, 전사 프로파일은 CD8a, ERK1, MAPK14, P2X7, 및 PBR 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다.

[0288] 또 다른 구현예에서, 전사 프로파일은 Gi2, GR, 및 MAPK14 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다.

[0289] 또 다른 구현예에서, 전사 프로파일은 Gi2, GR, MAPK14, 및 MR 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다.

[0290] 또 다른 구현예에서, 전사 프로파일은 ARRB1, ARRB2, CD8b, ERK2, IDO, IL-6, MR, ODC1, PREP 및 RGS2 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다.

[0291] 또 다른 구현예에서, 전사 프로파일은 ARRB1, CREB1, ERK2, Gs, IL-6, MKP1, 및 RGS2 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다.

[0292] 또 다른 구현예에서, 전사 프로파일은 ERK1 및 MAPK14 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다. 또 다른 구현예에서, 전사 프로파일은 Gi2 및 IL1b 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다. 또 다른 구현예에서, 전사 프로파일은 ARRB1 및 MAPK14 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다. 또 다른 구현예에서, 전사 프로파일은 ERK1 및 IL1b 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다.

[0293] 또 다른 구현예에서, 전사 프로파일은 ERK1, MAPK14, 및 P2X7 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다. 또 다른 구현예에서, 전사 프로파일은 Gi2, IL1b, 및 PBR 로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다. 또 다른 구현예에서, 전사 프로파일은 ARRB1, ODC1, 및

P2X7로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다. 또 다른 구현예에서, 전사 프로파일은 ARRB1, CD8a, 및 IL6으로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다. 또 다른 구현예에서, 전사 프로파일은 CD8b, ERK1, 및 MAPK14로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다. 또 다른 구현예에서, 전사 프로파일은 ARRB1, ERK1, 및 MAPK14로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다. 또 다른 구현예에서, 전사 프로파일은 ERK1, MAPK14, 및 PBR로 이루어진 군으로부터 선택되는 유전자의 전사 분석으로부터 측정된다.

- [0294] 본 발명의 양상은 대상에서 전사 프로파일을 확인하고, 이러한 전사 프로파일을 대조군 대상 또는 건강한 대조군 대상의 그룹의 프로파일과 비교하여, 전사 프로파일 내 변화 또는 차이의 존재 또는 부재에 근거하여 대상이 정동 장애를 나타내는 지의 여부를 진단하는 것을 포함하는, 대상에서의 정동 장애의 진단 방법을 제공한다.
- [0295] 본 발명의 일부 구현예에서, 정동 장애는 우울증, 중증 우울증, 양극성 장애, 경계 인격 장애로 이루어진 군으로부터 선택된다. 일부 구현예에서, 정동 장애는 외상 후 스트레스 장애 또는 외상 후 스트레스 장애가 없는 트라우마로부터 선택된다. 다른 구현예에서, 정동 장애는 급성 외상 후 스트레스 장애 또는 완화된 외상 후 스트레스 장애로부터 선택된다.
- [0296] 본 발명의 하나의 양상은 하기를 포함하는, 대상이 정동 장애를 나타내는 지의 여부를 진단하는 방법을 제공한다:
- (a) 정동 장애를 갖는 것으로 의심되는 대상으로부터 생물학적 샘플을 수득하는 단계;
  - (b) 생물학적 샘플에서 mRNA 수준을 측정하는 단계로, mRNA 수준이 ADA, ARRB1, ARRB2, CD8a, CD8b, CREB1, CREB2, DPP4, ERK1, ERK2, Gi2, Gs, GR, IL1b, IL6, INDO, MAPK14, MAPK8, MKP1, MR, ODC1, P2X7, PBR, PREP, RGS2, S100A10, SERT 및 VMAT2로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준인 단계;
  - (c) mRNA 수준을 mRNA 데이터로서 컴퓨터 매체에 수집하고 저장하는 단계;
  - (d) 분류 알고리즘을 통해 이러한 mRNA 데이터를 프로세싱하여, 프로세싱이 mRNA 데이터가 건강한 대조군 대상의 mRNA 데이터와 동일 또는 상이한 지의 여부를 결정하는 단계; 및
  - (e) 대상을 분류하는 출력물 데이터를 제공하는 단계에 의해,
- [0302] 대상이 정동 장애를 나타내는 지의 여부를 진단하는 방법.
- [0303] 본 발명은 또한, ADA, ARRB1, ARRB2, CD8a, CD8b, CREB1, CREB2, DPP4, ERK1, ERK2, Gi2, Gs, GR, IL1b, IL6, IL8, INDO, MAPK14, MAPK8, MKP1, MR, ODC1, P2X7, PBR, PREP, RGS2, S100A10, SERT 및 VMAT2로 이루어지는 군으로부터 선택되는 유전자의 대상의 전사 프로파일을, 다수의 건강한 대조군 대상의 상기 유전자의 전사 프로파일과 비교함으로써 정동 장애에 대한 대상의 감수성을 예측하는 방법을 제공한다.
- [0304] 본 발명의 하나의 양상은 하기를 포함하는, 대상이 정동 장애의 증상을 나타내는 가망성을 예측하는 방법을 제공한다:
- (a) 대상으로부터 생물학적 샘플을 수득하는 단계;
  - (b) mRNA 수준이 ADA, ARRB1, ARRB2, CD8a, CD8b, CREB1, CREB2, DPP4, ERK1, ERK2, Gi2, Gs, GR, IL1b, IL6, IL8, INDO, MAPK14, MAPK8, MKP1, MR, ODC1, P2X7, PBR, PREP, RGS2, S100A10, SERT 및 VMAT2로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준인, mRNA 수준을 측정하는 단계;
  - (c) mRNA 수준을 mRNA 데이터로서 컴퓨터 매체에 수집하고 저장하는 단계;
  - (d) 분류 알고리즘을 통해 이러한 mRNA 데이터를 프로세싱하여, 프로세싱이 mRNA 데이터가 건강한 대조군 대상의 mRNA 데이터와 동일 또는 상이한 지의 여부를 결정하는 단계; 및
  - (e) 대상을 분류하는 출력물 데이터를 제공하는 단계에 의해,
- [0310] 대상이 정동 장애의 증상을 나타내는 가망성을 예측하는 방법.
- [0311] 또 다른 구현예에서, 본 방법은 ADA, ARRB1, ARRB2, CD8a, CD8b, CREB1, CREB2, DPP4, ERK1, ERK2, Gi2, Gs, GR, IL1b, IL6, IL8, INDO, MAPK14, MAPK8, MKP1, MR, ODC1, P2X7, PBR, PREP, RGS2, S100A10, SERT 및 VMAT2로 이루어진 군으로부터 선택되는 2개 이상의 유전자의 mRNA 수준을 측정하는 것을 포함할 수 있다.
- [0312] 다른 구현예에서, 본 방법은 표 1A에 열거된 임의의 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,

18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 또는 28 개의 유전자의 mRNA 수준을 측정하는 것을 포함한다.

[0313] 다른 구현예에서, 본 방법은 ARRB1, ARRB2, CD8a, CREB1, CREB2, ERK2, Gi2, MAPK14, ODC1, P2X7, 및 PBR로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다.

[0314] 또다른 구현예에서, 본 방법은 CD8a, ERK1, MAPK14, P2X7, 및 PBR로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다.

[0315] 또다른 구현예에서, 본 방법은 Gi2, GR, 및 MAPK14로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다.

[0316] 또다른 구현예에서, 본 방법은 Gi2, GR, MAPK14, 및 MR로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다.

[0317] 또다른 구현예에서, 본 방법은 ARRB1, ARRB2, CD8b, ERK2, IDO, IL-6, MR, ODC1, PREP 및 RGS2로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다.

[0318] 또다른 구현예에서, 본 방법은 ARRB1, CREB1, ERK2, Gs, IL-6, MKP1, 및 RGS2로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다.

[0319] 또다른 구현예에서, 본 방법은 ERK1 및 MAPK14로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다. 또다른 구현예에서, 본 방법은 Gi2 및 IL1b로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다. 또다른 구현예에서, 본 방법은 ARRB1 및 MAPK14로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다. 또다른 구현예에서, 본 방법은 ERK1 및 IL1b로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다.

[0320] 또다른 구현예에서, 본 방법은 ERK1, MAPK14, 및 P2X7로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다. 또다른 구현예에서, 본 방법은 Gi2, IL1b, 및 PBR로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다. 또다른 구현예에서, 본 방법은 ARRB1, ODC1, 및 P2X7로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다. 또다른 구현예에서, 본 방법은 ARRB1, CD8a, 및 IL6으로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다.

또다른 구현예에서, 본 방법은 CD8b, ERK1, 및 MAPK14로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다. 또다른 구현예에서, 본 방법은 ARRB1, ERK1, 및 MAPK14로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다. 또다른 구현예에서, 본 방법은 ERK1, MAPK14, 및 PBR로 이루어진 군으로부터 선택되는 유전자의 mRNA 수준을 측정하는 것을 포함한다.

[0321] 본 발명의 일부 구현예에서, 정동 장애는 우울증, 중증 우울증, 양극성 장애, 경계 인격 장애로 이루어진 군으로부터 선택된다. 일부 구현예에서, 정동 장애는 외상 후 스트레스 장애 또는 외상 후 스트레스 장애가 없는 트라우마로부터 선택된다. 다른 구현예에서, 정동 장애는 급성 외상 후 스트레스 장애 또는 완화된 외상 후 스트레스 장애로부터 선택된다.

[0322] 일부 구현예에서, 상기 방법은 컴퓨터-보조 방법이다.

### 5.7 정동 장애

[0324] 본원에 기재된 정식의학적 또는 정신적 장애, 및 이들의 임상적 징후는 정신과 개업의에게 알려져 있다. 각각의 장애의 구체적인 증상은 대부분의 정신과 의사에 의해 인정될 수 있다.

[0325] 미국 정신과 협회 (American Psychiatric Association)에 의해 발행된 문헌 [*The Diagnostic and Statistical Manual of Mental Disorders*, 제 4 판, 본문 개정 (DSM-IV-TR<sup>®</sup>)] (1994년 10월, 본문 개정 2000년 5월)은 미국에서 의사에 의해 사용되는 정신적 장애의 임상적 분류에 대한 표준이다. 정신적/정신의학적 장애에 대한 징후학 및 진단 기준은 DSM-IV-TR<sup>®</sup> 가이드라인에 정리되어 있다.

#### 5.7.1 우울 장애

[0327] DSM-IV-TR<sup>®</sup>에는 우울증 및 주요 우울 장애 (MDD)에 대한 구체적인 진단 기준이 열거되어 있다.

[0328] DSM-IV-TR<sup>®</sup>에는 주요 우울성 에피소드를, 동일한 2주 기간 동안 하기 증상 중 5 가지 이상이 존재하고, 잘 기능하던 이전 상태로부터의 변화로서 그 자체를 잘 드러내는 (더욱이, 증상에는 하기 (1) 또는 (2)가 반드시

포함되어야만 한다) 증상으로서 정의하고 있다:

[0329] 1. 우울한 기분

[0330] 2. 관심 또는 기쁨의 감소

[0331] 3. 현저한 체중 감소 또는 증가

[0332] 4. 불면증 또는 수면파ing

[0333] 5. 정신운동 불안 또는 지체

[0334] 6. 피곤 또는 에너지 손실

[0335] 7. 무가치한 느낌

[0336] 8. 생각하거나 집중하는 능력이 줄어듬; 우유부단

[0337] 9. 죽음, 자살 관념, 자살 시도 또는 자살에 대한 구체적인 계획에 대한 되풀이되는 생각

[0338] DSM-IV-TR®에는 또한, 우울증의 다양한 부차유형에 존재해야만 하는 증상에 대한 설명이 포함된다. 우울증은 정신병적 증상이 있는 또는 없는 것으로 주지될 수 있고, 우울성 또는 긴장성 특징을 가질 수 있거나 비정형 우울증으로서 분류될 수 있다.

[0339] 환자에 의해 나타내어지는 증상의 수 및 경중도에 따라, 우울 에피소드는 경증, 중등 또는 중증으로서 명시될 수 있다. 임상의는 또한 환자가 전형적인 (우울성), 비정형, 긴장성, 또는 정신병적 우울증을 앓고 있는지를 결정할 수 있다.

[0340] 임상적으로, 우울증은 매우 이질적인 질환인 것으로 고려된다. 우울증 환자의 유전자 발현 프로파일은 상기 이질성을 반영할 수 있다. 본 발명에 근거하여, 환자를 더욱 잘 분류 또는 진단하기 위해 유전자 발현 프로파일에 근거한 우울증의 상기 부차유형을 더욱 잘 정의하는 것이 가능하다. 따라서, 약물의 개발 및 투여는 우울증의 부차유형을 앓고 있는 환자에 대해 맞춰질 수 있다.

[0341] 대조군으로부터 임상적 이력 및 증상 정보를 수득 및 분석함으로써, 유전자 발현 프로파일은 또한 대상이 본원에 기재된 장애의 증상을 나타내는 가망성을 예측하기 위해 사용된다.

[0342] 우울 장애, 양극성 장애 및 기분저하 장애는 기분 장애의 카테고리의 일부로 고려된다.

[0343] 본 발명은 경증, 중등 또는 중증 우울증과 같은 우울 장애의 전사 프로파일 지표의 객관적인 측정값을 제공한다. 본 발명은 또한 우울 장애의 부차유형의 분류를 위한 전사 프로파일을 제공한다. 본 발명은 또한 경증, 중등 또는 중증 우울증과 같은 우울 장애를 갖는 대상을 진단하는 방법을 제공한다.

#### 5.7.2 양극성 장애

[0345] 우울증에 대해 설명할 때, 양극성 장애 (BD)는 이질성 질환이고, 양극성 I, 양극성 II 및 순환기분장애를 포함하는 부차카테고리 또는 부차유형으로 나뉘어진다. 조울증으로도 알려져 있는 양극성 장애는 사람의 기분, 에너지 및 기능하는 능력에 있어서 통상적이지 않은 변위를 일으키는 뇌 장애이다. 모든 개인이 경험하는 정상적인 "업 앤 다운 (ups and downs)" 과는 다르게, 양극성 장애의 증상은 심각하며, 관계를 망치고, 회사 또는 학교 생활을 잘 못하며, 심지어 자살에 이를 수 있다.

[0346] BD는 사람의 일생에 걸쳐 전형적으로 재발하는 조증 및 우울증의 간헐적인 에피소드로서 나타난다. 에피소드 사이에, 양극성 장애를 가진 대부분의 사람은 증상이 없거나, 일부 잔존하는 증상을 가질 수 있다. 우울 에피소드는 종종 존재하고, 주요 또는 중증일 수 있다. 조증 에피소드는 환자 또는 다른 사람들에게 위험을 주거나 일에 장애를 일으키기에 충분하고, 약물 남용 또는 의료상의 문제의 결과가 아닌 심오한 기분 병폐, 수면에 대한 필요 감소, 과도한 수다 또는 강압적인 말투, 및/또는 급반전하는 생각 또는 싸움 생각, 및 더욱은 DSM-IV-TR®에 따라 기재된 바와 같은 증상을 특징으로 한다.

[0347] 본 발명은 양극성 장애를 갖는 대상을 진단하는 방법을 제공한다. BD 환자는 양극성 장애의 전사 프로파일 지표의 객관적인 측정값으로부터 이득을 볼 것이다.

#### 5.7.3 경계 인격 장애

- [0349] 경계 인격 장애 (BPD) 는 현저한 충동성을 가진, 자가-이미지, 대인 관계 및 감정의 불안정 패턴을 포함한다. 상기 불안정은 종종 가족 및 직업 생활 및 개인의 자가-정체성을 망친다.
- [0350] The DSM-IV-TR® 은 하기 중 5 가지 이상에 의해 표시되는 바와 같은 BPD 를 특징으로 한다:
- [0351] 1. 극한 과대 이상화와 평가절하 사이가 교차하는 것을 특징으로 하는 불안정하고 극심한 대인 관계 패턴.
  - [0352] 2. 잠재적으로 자가-손상을 일으키는, 예를 들어, 소비, 섹스, 약물 사용, 좀도둑질, 무모한 운전 또는 폭식인 2 가지 이상의 영역에서의 충동성.
  - [0353] 3. 기분의 현저한 반응성으로 인한 정서 불안정.
  - [0354] 4. 부적합한, 극심한 화 또는 화 통제 결핍, 예를 들어, 잦은 성미의 드러냄, 계속되는 화 또는 재발성 신체적 싸움.
  - [0355] 5. 재발성 자살 위협, 제스처, 또는 행동 또는 자기-훼손 행동.
  - [0356] 6. 정체성 혼란; 현저하고 집요한 불안정한 자가-이미지.
  - [0357] 7. 공허함 또는 지루함의 만성적 느낌.
  - [0358] 8. 실제적 또는 상상적 포기를 피하기 위한 발광적인 노력.
  - [0359] 9. 일시적인, 스트레스와 관련된 편집성 사고 또는 중증 해리 증상.
- [0360] BPD 환자는 정신치료에서 보이는 가장 저항적이고 치료에 내성이 있는 환자이다.
- [0361] 본 발명은 BPD 을 가진 대상을 진단하는 방법을 제공한다. BPD 환자는 경계 인격 장애의 전사 프로파일 지표의 객관적인 측정값으로부터 이득을 볼 것이다.
- [0362] 5.7.4 외상 후 스트레스 장애 (PTSD)
- [0363] DSM-IV-TR® 에는 외상 후 스트레스 장애를, 실제적인 또는 위협적인 죽음 또는 심각한 죽음과 관련된 사건의 직접적인 개인적 경험을 포함하는 극한 트라우마성 스트레스 요인에 노출된 후 특징적인 증상을 발달시키는 것으로 기재하고 있다. 사람은 다른 사람의 죽음, 상해, 또는 신체적 온전성에 대한 위협을 포함하는 사건을 목격할 수 있다. 사건에 대한 인간의 반응은 강렬한 두려움, 무기력 또는 공포를 포함한다. 사람은 이미지, 생각 또는 지각을 포함하여 사건의 영속적인 기억을 가질 수 있거나, 사건의 재발적인 고통스러운 꿈을 꾸 수 있다.
- [0364] 본 발명은 급성 PTSD, 완화된 PTSD, 또는 PTSD 가 없는 트라우마를 갖는 대상을 진단하는 방법을 제공한다. 환자/대상은 급성 PTSD, 완화된 PTSD, 또는 PTSD 가 없는 트라우마의 전사 프로파일 지표의 객관적인 측정값으로부터 이득을 볼 것이다.
- [0365] 상기 기재된 방법에 의해 확인된 전사 프로파일에 근거하여 정상, 또는 건강한, 대상 및 정동 장애를 앓고 있는 대상 사이를 측정, 차별화 및/또는 구별하는 것이 가능하다. 예를 들어, 본 발명은 하기 실험 사항에 의해 더욱 잘 이해될 것이다. 당업자는 본원에 논의되는 특정 방법 및 결과가 하기 청구항에 더욱 자세히 기재되는 바와 같이 본 발명을 단지 설명하는 것으로 쉽게 인지할 것이다.
- [0366] 6 실험 사항
- [0367] 총 RNA 단리. 인간 혈액을 PAXgene™ 혈액 RNA 튜브 (*PreAnalytiX, Hombrechtkon, CH*) 내로 수집하고, 수회 역위에 의해 혼합하고, RNA 단리가 이루어질 때까지 -20°C 또는 -80°C 에서 저장하였다. 샘플을 실온에서 밤새 인큐베이션후, 3000 × G 에서 10 분 동안 원심분리에 의해 처리를 시작하였다. 상청액을 따르고, 펠렛을 5 ml 물에 재현탁한 후, 또다른 원심분리 단계를 수행하였다. 세정 및 원심분리 단계를 다시 반복하고, 펠렛을 튜브에 남아 있는 잔류 물 (약 100 µl) 에 재현탁하였다. 상기 용액에, 941 µl 의 Ambion ToTALLY RN A™ Lysis / Denaturation Solution (*Ambion, Austin, TX*) 및 59 µl 3M 나트륨 아세테이트, pH 5.5 (*Ambion*) 를 첨가한 후, 혼합하였다. 실온에서 15 분 동안 인큐베이션 후, 770 µl 의 산 페놀/클로로포름 (*Ambion*) 을 첨가하고, 튜브를 볼텍스에 의해 혼합하였다. 용액을 2 ml 플라스틱 스크류 뚜껑 튜브에 옮기고, 5 분 동안 실온에서 인큐베이션하였다. 페놀 추출을 마이크로퓨즈 (대략 13,000 × G) 에서 전 속력으로 1 분 동안 스핀다운시키고, 수성층 (1100 µl) 을 550 µl 의 100% 에탄올을 함유하는 새로운 튜브로 제

거하였다. 혼합 후, 용액을 Ambion RNAqueous<sup>®</sup> -96 Automated Kit 필터 플레이트의 하나의 웰에 적용하고, RNA 를 제조자의 프로토콜에 따라 정제하였다. RNA 용리 후, 샘플을 DNase I (Invitrogen, Carlsbad, CA) 로 다시 처리하여 잔류 게놈 DNA 를 제거하였다. RNA 를 1× DNase 소화 완충액과 3 유닛의 효소에서 1 시간 동안 실온에서 인큐베이션하였다. 효소를 13 mM 의 최종 농도로 EDTA 를 첨가한 후 68°C 에서 10 분 동안 가열하여 비활성화시켰다. 혼합물을 Multiscreen<sup>®</sup> PCR micro96 서플레이트 (Millipore, Billerica, MA) 에 통과하여 탈염시키고, 50 μl 의 물에 용리하였다. 1 μl 분취액의 RNA 를 Agilent 2100 Bioanalyzer (Agilent, Waldbronn, Germany) 상에서 분석하고, 나머지를 -80°C 에 저장하였다. RNA 샘플의 품질을 Bioanalyzer 소프트웨어에 의해 계산된 RIN 값을 사용하여 평가하였다.

## [0368]

cDNA 합성

## [0369]

대략 1 μg 의 총 RNA 와 1.5 μl 랜덤 6 량체 (Invitrogen, 500 ng/μl) 를 16.5 μl 의 최종 부피로 혼합하여 cDNA 의 합성을 달성하였다. 75°C 에서 10 분 동안, 25°C 에서 10 분 동안 인큐베이션 후, 6 μl 의 제 1 가닥 완충액 (Invitrogen), 1.5 μl 의 10 mM dNTP (Invitrogen, 10 mM 각각의 dNTP), 1.25 μl Superscript II<sup>™</sup> (Invitrogen, 200 유닛/μl), 및 4 μl 물을 첨가하였다. 최종 반응 부피는 30 μl 이었고, 인큐베이션을 25°C 에서 10 분 동안, 42°C 에서 1 시간 동안, 95°C 에서 10 분 동안 수행하였다. 70 μl 의 물을 첨가한 후 MultiScreen<sup>®</sup> PCR micro96 플레이트로 정제할 때까지 반응물을 4°C 로 냉각시켰다. 100 μl 의 물로 cDNA 의 용리를 수행하고, 수득된 물질을 정량할 때까지 -20°C 에 저장하였다. 일부 경우에서, 물질의 수율을 증가시키기 위해 cDNA 반응 부피를 2 배로 하였다.

## [0370]

cDNA 의 정량

## [0371]

염료 개입 어세이를 사용하여 cDNA 수율을 측정하였다. 5 μl 의 cDNA 를 7 μl 의 0.5 N NaOH, 50 mM EDTA 와 47 μl 의 최종 부피로 혼합하였다. 혼합물을 65°C 에서 1 시간 동안 인큐베이션하여 RNA 를 가수분해한 다음, 10 μl 의 1M Tris, pH7 을 첨가하여 중화시켰다. 제조자의 지침에 따라 Quant-it<sup>™</sup> Oligreen<sup>®</sup> ssDNA 시약 (Invitrogen) 을 사용하여 가수분해 반응의 25 μl 분취물 내 cDNA 농도를 측정하였다. 미공지된 샘플을 공지된 농도의 단일 가닥 DNA 를 사용하여 생성된 표준 곡선과 비교하였다. Fusion<sup>™</sup> 알파 기구 (Packard, Meridian, CT) 를 사용하여 모든 형광 판독값을 측정하였다. 2 회 반복 가수분해 반응으로부터 수득된 값을 각각의 미공지된 cDNA 샘플에 대해 평균을 내었다. 2 회 반복이 서로의 15% 내에 있지 않은 경우, 제 3 샘플을 실행하여 이전의 2 회 측정값과 비교하고, 2 개의 가장 유사한 값으로 평균을 내었다.

## [0372]

정량적 폴리머라아제 연쇄 반응 (qPCR)

## [0373]

모든 qPCR 실행을 표 1A 및 1B 에서 제시된 프라이머/탐침 세트를 사용하여 Applied Biosystems 7900HT Fast Real Time PCR System (Applied Biosystems, Foster City, CA) 또는 MX3000P<sup>®</sup> (Stratagene, La Jolla, CA) 에서 수행하였다. 모든 탐침을 5' 말단에는 FAM<sup>™</sup> (Applera, Norwalk, CT) 로, 3' 말단에는 BHQ-1<sup>®</sup> 켄처로 표지하고, Biosearch (Novato, CA) 에 의해 합성하였다. 각각의 프라이머/탐침 세트를 체크하여 PCR 증폭 효율이 어세이의 발현 범위에 대해 대략 100% 였음을 확인하였다. 복제본 플레이트 (96 웰 형태) 를 각각의 인간 기증자로부터 웰 당 1 ng 또는 10 ng 의 cDNA 를 함유하도록 구축하였다. 또한 플레이트는 2 개의 음성 대조군 웰 ("NTC", 오직 물만) 및 10 명의 개인의 혈액으로부터 유래된 모아진, 시판 cDNA (참조 cDNA) 가 있는 3 개의 웰을 함유한다. 각각의 qPCR 반응은 25 μl (최종 부피) 였고, 하기 성분: 12.5 μl Brilliant QPCR Master Mix<sup>®</sup> (Stratagene), 400 nM 정방향 프라이머, 400 nM 역방향 프라이머, 50 nM 탐침, 및 60 nM/300 nM ROX<sup>™</sup> (Applera) (MX3000P<sup>®</sup> 7900HT 기구) 를 함유하였다. 사이클 조건은 95°C, 10 분 후; 95°C, 15 초의 40 회 사이클; 60°C, 1 분이었다. 각각의 유전자에 대해 2 회 반복 qPCR 실행을 수행하였다. 드물게, 유전자에 대한 복제본 플레이트가 충분히 일치하지 않는 경우, 제 3 qPCR 플레이트를 실행하였다. 수득된 Ct 값에 따라, 모든 3 개의 플레이트로부터의 값을 평균을 내거나, 이상한 플레이트는 추가 분석에서 배제하였다.

## [0374]

qPCR 실행에 사용되는 기구는 예비 데이터 분석 단계를 필요로 한다. 그러나, 각각의 경우 목표는, 동일한 역치가 제공된 플레이트 상에 모든 샘플에 대해 사용되는 종종 곡선의 중간 점 근처에 종종 역치에 설정하였다. 동일한 유전자에 대해 2 회 반복 플레이트 실행에 대해서는 역치가 유사하였다 (반드시 동일할 필요는 없음). MX3000P<sup>®</sup> 의 경우, 역치를 처음 측정하기 위해 하기 설정을 사용하였다: 평활 파라미터 = 5, MX4000

알고리즘을 사용하는 기준선 계산, 및 20 의 시그마 승수 (multiplier) 로의 사이클 6 ~ 14 를 사용하는 배경-기반 역치. 역치를 중폭 플롯의 중간에 대략적으로 두기 위해 역치의 미미한 조정은 필요다하면 수동으로 이루어졌다. 7900HT 상의 플레이트 실행의 경우, 역치를 처음 설정하기 위해 기구의 디폴트 세팅을 사용하였다. 필요하다면 수동 조정이 이후에 이루어졌다.

표 1A: 선별된 유전자/바이오마커에 대한 프라이머 / 탐침 서열.

| 유전자 명칭               | 약어    | 유전자 접근 번호<br>(SEQ ID NO:)   | 대표적인<br>프라이머/탐침 서열 (5'에서 3') <sup>†</sup>   |
|----------------------|-------|-----------------------------|---|
| 아데노신 테아미나이체          | ADA   | NM_000022<br>(SEQ ID NO:88) | F = GGTGGTGGAGCTGTTAAGAAAGTAC (SEQ ID NO:1)<br>R = CTTCTGGATGGTCATCTC (SEQ ID NO:2)<br>P = CAGCAGACCGTGGTAGCCATTGACCT (SEQ ID NO:3)       |
| 베타-아레스틴 1            | ARRB1 | L04685<br>(SEQ ID NO:89)    | F = AGACACGAACCTGGCCTTAGC (SEQ ID NO:4)<br>R = TTGTTAGGAAACAATGATCCCCAG (SEQ ID NO:5)<br>P = TTGAGGGAAAGGTGCCAACCGTGAGAT (SEQ ID NO:6)    |
| 베타-아레스틴 2            | ARRB2 | B0007427<br>(SEQ ID NO:90)  | F = TCTTCCCATGCTCCGTACAC (SEQ ID NO:7)<br>R = CGAATCTCAAAGTCTACGCCG (SEQ ID NO:8)<br>P = AGCCAGGGCCAGAGGATAACAGGAAA (SEQ ID NO:9)         |
| CD8 알파-              | CD8a  | M12824<br>(SEQ ID NO:91)    | F = TTCCGCCAGAGAACGAG (SEQ ID NO:10)<br>R = AAGACCGGGCACGAAATGG (SEQ ID NO:11)<br>P = TCGGCCCTGAGCAACTCCATCATGTA (SEQ ID NO:12)           |
| CD8 베타-              | CD8b  | M37601<br>(SEQ ID NO:92)    | F = TGACAGTCACCACGAGTTCCCTG (SEQ ID NO:13)<br>R = TCTCCTGTTCCACCTCTCACC (SEQ ID NO:14)<br>P = CTCTGGGATTCCGCAAAAGGGACTAT (SEQ ID NO:15)   |
| cAMP 반응성 요소 결합 단백질 1 | CREB1 | NM_134442<br>(SEQ ID NO:93) | F = CTGGCTAACAAATGGTACCGATG (SEQ ID NO:16)<br>R = GTGGTCTGTGCATACTGTAGAATGG (SEQ ID NO:17)<br>P = CATGACCAATGCAGCAGCCACTCA (SEQ ID NO:18) |

| 유전자-명칭                  | 약어    | 유전자 접근<br>번호<br>(SEQ ID NO:) | 대표적인<br>프라이머/탐침 서열 (5'에서 3') <sup>†</sup>  |
|-------------------------|-------|------------------------------|--|
| cAMP 반응성 요소 결합 단백질 2    | CREB2 | M86842<br>(SEQ ID NO:94)     | F = CACGTTGGATGACACTTGTGATC (SEQ ID NO:19)<br>R = CTGGGAGATGGCCAATTGG (SEQ ID NO:20)<br>P = ACTAATAAGCAGCCCCAGACGGT (SEQ ID NO:21)       |
| 디펜티딜 펩티다아제 IV 키나이제 1    | DPP4  | M74777<br>(SEQ ID NO:95)     | F = GTGTCATTCACTAAAGAGGCCGAAG (SEQ ID NO:22)<br>R = CTCAGCCCTTATCATTCACGC (SEQ ID NO:23)<br>P = TTCCGGTCCCTGGCTGCCCTCTATA (SEQ ID NO:24) |
| 세포외부 신호-조절 키나이제 1       | ERK1  | M84490<br>(SEQ ID NO:96)     | F = TGACGGGAGTATGGGCTACGC (SEQ ID NO:25)<br>R = CCACAGCCAGATGTCGATGG (SEQ ID NO:26)<br>P = CTGGTACCGGGCCCCAGAGATCAT (SEQ ID NO:27)       |
| 세포외부 신호-조절 키나이제 2       | ERK2  | M84489<br>(SEQ ID NO:97)     | F = TAACGTTCTGCACCGTGACC (SEQ ID NO:28)<br>R = CAGGCCAAAGTCACAGATCTTG (SEQ ID NO:29)<br>P = ACCTGCTGCTCAAACACCCCTGTGAT (SEQ ID NO:30)    |
| 구아닌 뉴클레오티드 결합 단백질 알파 i2 | Gi2   | X04828<br>(SEQ ID NO:98)     | F = AGGCGTGTCCCTGATGAC (SEQ ID NO:31)<br>R = GCTCCAGGTCGTTCAAGGTAGTAG (SEQ ID NO:32)<br>P = AGGCCTGCTTGGCCGCTCAA (SEQ ID NO:33)          |

| 유전자 명칭                      | 약어   | 유전자 접근<br>(SEQ ID NO:)       | 대표적인<br>변호 | 대표적인<br>프라이머/탈착 소열 (5'에서 3') <sup>†</sup>   |
|-----------------------------|------|------------------------------|------------|---|
| 구아닌 뉴클레오티드 결합 단백질 알파-s (긴것) | Gs   | AF493897<br>(SEQ ID NO:99)   |            | F = GACTATGTCGCCAGCGATCAG (SEQ ID NO:34)<br>R = GTCCACCTGGAAACTGGTCTCA (SEQ ID NO:35)<br>P = CTGCTTCGCTGCCGTCCCTGA (SEQ ID NO:36)             |
| 알파-글루코코리티코이드 수용체            | GR   | X03225<br>(SEQ ID NO:100)    |            | F = TCCCTGGTCGAACAGTTTTTC (SEQ ID NO:37)<br>R = TTTGGGAGGTGGTCTGTG (SEQ ID NO:38)<br>P = TGTAAGCTCTCCCATCCAGCTCCCAA (SEQ ID NO:39)            |
| 인터류킨 1, 베타                  | IL1b | NM_000576<br>(SEQ ID NO:101) |            | F = GATGGCCCTAACACAGATGAAGTG (SEQ ID NO:40)<br>R = CCTGAAAGGCCCTTGCTGTAGTG (SEQ ID NO:41)<br>P = ATGGCGGCATCCAGCTACGAATCTC (SEQ ID NO:42)     |
| 인터류킨 6                      | IL6  | M14584<br>(SEQ ID NO:102)    |            | F = AGCCACTCACCTCTCAGAACG (SEQ ID NO:43)<br>R = CATGTCCTCTCAGGGCTG (SEQ ID NO:44)<br>P = CAAATTGGTACATCCTCGACGGCAT (SEQ ID NO:45)             |
| 인터류킨 8                      | IL8  | M28130<br>(SEQ ID NO:103)    |            | F = CTGCTAGCCAGGATCCACAAAG (SEQ ID NO:46)<br>R = CTGTGAGGTAAGATGGTGGCTAATAC (SEQ ID NO:47)<br>P = CTGTGTCACACTGTGCCTTCTCCCT (SEQ ID NO:48)    |
| 인톨아민-파리를 2,3 디옥시제나이제        | INDO | NM_002164<br>(SEQ ID NO:104) |            | F = GCTTCGAGAAAAGAGTTGACAAGTTAAC (SEQ ID NO:49)<br>R = GACCTTGGCCCCACACATATG (SEQ ID NO:50)<br>P = CTCACAGACCAAAAGTCACAGCGCCCT (SEQ ID NO:51) |

| 유전자 명칭                  | 약어                 | 유전자 접근<br>번호<br>(SEQ ID NO:) | 대표적인<br>프리이머/탈진 서열 (5'에서 3') <sup>†</sup>  |
|-------------------------|--------------------|------------------------------|--|
| p38 미토센 활성화 단백질 카나아제 14 | MAPK1 <sub>4</sub> | L35253<br>(SEQ ID NO:105)    | F = CGGAGGACTACACAGGCCATAG (SEQ ID NO:52)<br>R = AGCAGCACACACAGGCCATAG (SEQ ID NO:53)<br>P = CCGAGGGTTACCAAAACCTGTCTCCA (SEQ ID NO:54)   |
| 미토센 활성화 단백질 카나아제 8      | MAPK8              | AY893269<br>(SEQ ID NO:106)  | F = CCAACACCCGTACATCAATGTC (SEQ ID NO:55)<br>R = CACTCTTCTATTGTGTCCCTTC (SEQ ID NO:56)<br>P = CACCACAAAGATCCCTGACAAGCAGTT (SEQ ID NO:57) |
| map 카나아제 포스파티아제 1       | MKP1               | X68277<br>(SEQ ID NO:107)    | F = GCCAGGCAGGCCATTTTC (SEQ ID NO:58)<br>R = ATGCTTCGCCTCTGCTTCAC (SEQ ID NO:59)<br>P = TCAGGCCACCATCTGCCTTGCTTACCT (SEQ ID NO:60)       |
| 미네랄로코르티코이드 수용체          | MR                 | M16801<br>(SEQ ID NO:108)    | F = AGCCCCAGAGGAAGGGACAAAC (SEQ ID NO:61)<br>R = TGTGAGGGCTCGTGAGATG (SEQ ID NO:62)<br>P = CTCCTGCAAAAGAACCCCTCGGTCAACA (SEQ ID NO:63)   |
| 오르니틴 테카로복실라아제 1         | ODC1               | NM_002539<br>(SEQ ID NO:109) | F = CCATGTAGGAAGGGCTGTAC (SEQ ID NO:64)<br>R = TCAGGCCCTCATGTCAAAAAC (SEQ ID NO:65)<br>P = ATCCCTGAGACCTCGTGCAGGGCAATCT (SEQ ID NO:66)   |
| 퓨린성 수용체 P2X7            | P2X7               | NM_002562<br>(SEQ ID NO:110) | F = GCTGTGGCTCCCATTTTATCC (SEQ ID NO:67)<br>R = CACAATGGACTCGCACTTCTTC (SEQ ID NO:68)<br>P = CTGTCAGGCCCTGTGGTCAACGAAATAC (SEQ ID NO:69) |

| 유전자 명칭                      | 약어          | 유전자 접근<br>(SEQ ID NO:)          | 대표적인<br>프라이머/탐침 서열 (5'에서 3') <sup>†</sup>   |
|-----------------------------|-------------|---------------------------------|---|
| 벤조디아자핀 수용체<br>(말초-유형)       | PBR         | BC001110<br>(SEQ ID<br>NO:111)  | F = CTGGTCTGGAAAGAGCTGGG (SEQ ID NO:70)<br>R = CAGCAGGAGATCCACCAAGG (SEQ ID NO:71)<br>P = CCCCCATCTCTTGGTGCCTCGAC (SEQ ID NO:72)          |
| 프롤레일 엔도펩티다이-체<br>조절체        | PREP        | D21102<br>(SEQ ID<br>NO:112)    | F = GGGAAATATGACTACGTGACCAATG (SEQ ID NO:73)<br>R = GGATCCCTGAAGTCATGTGATC (SEQ ID NO:74)<br>P = CATTCAAGACGAATGCCAGTCTCCC (SEQ ID NO:75) |
| G-단백질 신호 2 의<br>A10 (p11)   | RGS2        | NM_002923<br>(SEQ ID<br>NO:113) | F = GATTGGAAAGACCCGTTGAGC (SEQ ID NO:76)<br>R = CAGGAGAAGGCTGATGAAAGC (SEQ ID NO:77)<br>P = CTGGGAAGCCCCAAACCGGCAA (SEQ ID NO:78)         |
| S100 칼슘 결합 단백질<br>A10 (p11) | S100A1<br>0 | NM_002966<br>(SEQ ID<br>NO:114) | F = AGGAGTCCCTGGATTTTGG (SEQ ID NO:79)<br>R = GCCCACTTTGCCATCTCACAC (SEQ ID NO:80)<br>P = CAAAAAGACCCCTGGCTGTGGACAAAAA (SEQ ID NO:81)     |
| 세로토닌 수송체                    | SERT        | NM_001045<br>(SEQ ID<br>NO:115) | F = CATGGCTGAGATGAGGAATGAAG (SEQ ID NO:82)<br>R = GCTGGCATGTTGGCTATCG (SEQ ID NO:83)<br>P = ACGCAGGTCCCAGCCTCCCTTCAT (SEQ ID NO:84)       |
| 소낭 모노아민 수송체 2               | VMAT2       | L23205<br>(SEQ ID<br>NO:116)    | F = TGGATTTCGTCAAATGATGCCTATC (SEQ ID NO:85)<br>R = ATGCCACATCCGCAATGG (SEQ ID NO:86)<br>P = AGACCTGGGCACGTGTCGTCTA (SEQ ID NO:87)        |

<sup>†</sup> F = 정방향 프라이머 서열; R=역방향 프라이머 서열; P = 탐침 서열

[0379]

#### [0380] 유전자 발현의 표준화

상이한 샘플 간의 유전자 발현 프로파일을 효과적으로 비교하기 위해서는, 임의의 균원적인 생물학적 변화를 차폐할 수 있는 변수에 대해 통제하는 것이 바람직하다. 예를 들어, 효소 반응 효율, 기기 장치 성능, 및 파이펫팅에서의 날마다 다른 차이가 제시된 날짜에 수득된 신호에 모두 영향을 줄 것이다. 이러한 변수의 영향을 최소화하기 위한 바람직한 방식은 복합 표준화 유전자의 사용을 통한 것이다 (Andersen, CL. et al., *Cancer Res*, 2004, 64:5245-5250; Jin, P. et al., *BMC Genomics*, 2004, 5:55; Huggett, J. et al., *Genes and Immunity*, 2005, 6:279-284). 이상적인 표준화 유전자는 편리하게 측정된 수준으로 발현되고, 실험 디자인의 일부인 조작에 의해 변하지 않는다. 표준화 유전자의 사용이 흔한 것이라도, 연구자는 그들이 사용하는 유전자가 실험 시스템에서 안정적으로 발현되는 지의 여부를 종종 입증하지 못한다. 이 문제를 피하기 위해, 시판 소프트웨어 프로그램 GeNorm™ (PrimerDesign Ltd., Southampton, UK) 을 사용하였다. 이 방법은 [Vandesompele, J. et al., *Genome Biol*, 2002, 3(7):RESEARCH0034.1- 0034.11 (Epub June 18, 2002)]

에 의해 출판된 논문에 근거한 것이고, 후보자 표준화 유전자가 안정적으로 발현되는지 또는 그렇지 않은지를 측정할 수 있게 한다. 표준화 유전자를 선택하기 위해, 문헌에서는 실험이 혈액 샘플로 수행되었음을 강조하며, 인간에서 유전자 발현을 표준화하기 위해 조사자에 의해 이전에 사용되었던 유전자를 확인하기 위해 일차로 스캔한다 (Vandesompele, J. et al. *Genome Biol*, Epub June 18, 2002, 3(7):RESEARCH0034.1- 0034.11, 특히 페이지 0034.5, 표 3; Applied Biosystems Application Note 2006, publication 127AP08-01, 특히 페이지 3, 도 1). 이러한 조사로부터, 표 1B에 제시된 유전자가 확인되었다. 상기 유전자들이 본 실험에서의 표준화에 대해 타당한지를 확인하기 위해, 7 개의 유전자의 발현 프로파일을 정상 대상, 약물 처리를 받지 않은 우울증 환자 및 약물 처리를 받은 우울증 환자를 비롯하여 상이한 실험 세트로부터 유도된 혈액 샘플을 사용하여 Genorm™으로 분석하였다. 모든 세트에서, 0.15 이하의 쌍 방식 변화 값 (V)에 의해 측정되는 바와 같이, 7 개의 유전자의 조합은 양호한 표준화를 달성하였다 (Vandesompele, J. et al., *Genome Biol*, Epub June 18, 2002, 3(7):RESEARCH0034.1- 0034.11).

[0382]

Genorm™에는 표준화를 위해 2 개 또는 3 개의 최적 유전자를 사용하는 것만이 필요하다고 언급되어 있긴 하지만, 3 개 초과의 표준화 유전자의 조합이 여러 가지 이유로 고려되어야만 한다. 먼저, 더 많은 표준화 유전자를 사용하면 신규 약물 처리, 유전적 배경, 또는 질환 상태가 표준화 유전자의 발현에 영향을 줄 수 있다는 예측 고려에 도움을 줄 것이다. 3 개 초과의 표준화 유전자는 특정 실험에서 안정적으로 발현되지 않는 임의의 유전자의 영향을 약화시킴으로써 과정을 개선시키는 것으로 예상된다. 또한, 발현 데이터를 표준화하기 위해 3 개 초과의 유전자를 일관적으로 사용함으로써, 발현 결과를 시간에 걸쳐 수행되는 모든 연구와 비교할 수 있다. 임상적 샘플이 항상 적합한 대조군과 매치되는 것은 아니기 때문에, 3 개 초과의 표준화 유전자를 사용하는 것이 중요하게 고려된다. 3 개 초과의 유전자로의 표준화가 상이한 실험 간에 유전자 발현을 비교할 때 바람직한 방법이긴 하지만, 비교되는 모든 샘플이 동일한 방식으로 처리된다면 임의의 특정 실험 내의 2 개 또는 3 개의 유전자를 사용하는 것이 여전히 유효하다.

표 1B: 표준화 유전자.

| 유전자 명칭                      | 약어     | 유전자 접근 번호<br>(SEQ ID NO:)    |
|-----------------------------|--------|------------------------------|
| 베타-액틴                       | ACTB   | NM_001101<br>(SEQ ID NO:117) |
| 베타 -2-마이크로글로불린              | B2M    | NM_004048<br>(SEQ ID NO:118) |
| 글리세르알데하이드-3-포스페이트 테하이드로제나아제 | GAPD   | NM_002046<br>(SEQ ID NO:119) |
| 글루쿠로니다아제, 베타                | GUSB   | NM_000181<br>(SEQ ID NO:120) |
| 히드록시메틸-빌란 신타아제              | HMBS   | NM_000190<br>(SEQ ID NO:121) |
| 하이포잔틴 포스포리보실-트랜스페라아제 I      | HPRT1  | NM_000194<br>(SEQ ID NO:122) |
| 포스포글리세레이트 키나아제              | PGK1   | NM_000291<br>(SEQ ID NO:123) |
| 펩티딜프로필 이소머라아제 A (시클로필린 A)   | PPIA   | NM_021130<br>(SEQ ID NO:124) |
| 리보솜 단백질, 대형, P0             | RPLP0  | NM_001002<br>(SEQ ID NO:125) |
| 리보솜 단백질 L13a                | RPL13A | NM_012423<br>(SEQ ID NO:126) |

[0383]

| 유전자 명칭   | 약어    | 유전자 접근 번호<br>(SEQ ID NO:)                |
|--|-------|--|
| 숙시네이트 테히드로게나아제 복합체, 서브유닛 A                       | SDHA  | NM_004168<br>(SEQ ID NO:127)             |
| TATA 박스 결합 단백질 (전사 인자 IID)                       | TBP   | NM_003194<br>(M34960)<br>(SEQ ID NO:128) |
| 수송 수용체 (p90, CD71)                               | TFRC  | NM_003234<br>(SEQ ID NO:129)             |
| 유비퀴틴 C   | UBC   | NM_021009<br>(M26880)<br>(SEQ ID NO:130) |
| 티로신 3-모노옥시게나아제/트립토판 5-모노옥시게나아제 활성화 단백질, 제타 폴리펩티드 | YWHAZ | NM_003406<br>(SEQ ID NO:131)             |
| 진핵세포 18S 리보솜 RNA                                 | 18S   | X03205<br>(SEQ ID NO:132)                |

[0384]

상기 섹션 5.4.1.2 에 기재된 바와 같이, 프라이머는 본원에 기재된 유전자 중 임의의 것에 대해 디자인될 수 있다. 표 1A 및 표 1B 에 확인된 유전자에 대한 공개적으로 이용가능한 서열은 유전자 접근 번호 (Gene Accession Number, GenBank 데이터베이스) 에 의해 나타내지고, 본원에 전체가 참조로서 인용된다. 표 1A 및 표 1B 에 확인된 유전자에 대한 서열은 표에 제시된 적합한 SEQ ID NO 에 의해 열거된 바와 같은 첨부 서열 목록에 기재되어 있다.

[0386]

#### 전사 데이터 분석

[0387]

이중 PCR 플레이트로부터 유도된 각각의 미공지된 샘플에 대한 평균 Ct (사이클 역치) 값을 각각의 유전자에 대해 측정하였다. 실시간 PCR 어세이에서, 형광 신호의 축적에 의해 양성 반응을 검출하였다. Ct 는 역치를 넘는 (즉, 배경 수준을 초과함) 형광 신호를 필요로 하는 사이클의 수로서 정의된다. Ct 수준은 샘플 내 표적 핵산의 양에 반비례한다 (즉, Ct 수준이 낮으면 샘플 내 표적 핵산의 양은 커진다).

[0388]

각각의 미공지된 cDNA 샘플, 및 참조 cDNA 에 대한 상대적인 발현 수준을 7 개의 표준화 유전자로부터의 평균 Ct 를 사용하는  $2^{-\Delta \text{Ct}}$  방법 (Livak, K. and Schmittgen, T., Methods, 2001, 25:402-408) 에 의해 계산하였다. 그 다음, 참조 cDNA 의 상대적인 발현 수준을 100% 로 놓고, 그 다음 모든 다른 샘플을 참조% 로서 표현하였다. 마지막으로, 상기 % 를 참조 cDNA 에 함유된 각각의 유전자의 카피 수에 % 를 곱하여, 카피/ng cDNA 로 전환하였다.

[0389]

#### 단변량 통계 분석 및 그래프

[0390]

유전자 발현 값과 환자/대상 설문지로부터 유도된 임상적인 파라미터 사이의 관련성을 R 통계 패키지를 사용하여 조사하였다. 비교를 용이하게 하기 위해 필요하다면 설문지 데이터를 코드화하였다. 유전자 발현 데이터를 분석 전에 로그 전환시키고, 파라미터 및 비-파라미터 분석을 수행하였다. 유의값에 대한 역치를  $p < 0.05$  에 설정하였다. 예를 들어, 표 3 을 참조한다. 특정 유전자가 제시된 대상 집단에 대해 일관적으로 업- 또는 다운-조절되는지를 측정하기 위해 단변량 검정을 사용하였다.

[0391]

대조군 대상과 우울증 환자 사이의 발현 수준을 비교하는 산점도 및 관련 단변량 통계 분석을 GraphPad Prism4® (GraphPad Software, Inc, San Diego, CA) 를 사용하여 각각의 유전자에 대해 생성하였다. 유전자 발현 값이 반드시 정규 분포하지는 않기 때문에, 그룹을 비교하는데 비-파라미터 Mann-Whitney 검정을 사용하였다. 유의성 역치를  $p < 0.05$  에 설정하였다. 특정 유전자, 및 혈액 내 이들의 상대적인 발현 수준을 도 2 내지 7 에 예시한다.

[0392]

#### 다변량 분석

[0393]

건강한 대조군 대상으로부터 질환이 있는 환자를 구별하기 위해, 분류 알고리즘을 사용하였다. 분류 알고리즘, 전형적으로 머신 습득 알고리즘은 하기 2 단계를 통해 실시한다: (1) 유전자 발현 수준이 집합적으로 가장 정보적인 것으로 발견되는, mRNA 전사 데이터 세트로부터 유전자의 부차세트를 선택함; (2) 단계 (1) 에서 확인된 바와 같은 유전자의 부차세트로 훈련된 미리 선택된 유형의 분류 알고리즘을 훈련 및 되돌림.

[0394] (1) 유전자의 선별

[0395] 제 1 단계에서, 건강한 대조군 대상 및 질환이 있는 대상, 또는 다른 질환이 있는 대상으로부터의 mRNA 전사 데이터 세트를 랜덤 포레스트 알고리즘에 대한 입력값으로서 집합적으로 사용하였다 (Breiman, L., 2001, *Machine Learning* 45(1):5-32)). 각각의 대상의 혈액 샘플로부터의 mRNA 전사 데이터를 나타내는 각각의 데이터 세트는 표 1A 에 열거된 유전자 및 본원에 기재된 방법에 근거한다. 가장 덜 중요한 유전자를 성공적으로 제거함으로써, 랜덤 포레스트 알고리즘은 아웃오브백 (out-of-bag: OOB) 오류 최소화 기준을 사용하여 가장 중요한 유전자를 함유하는 목록을 되돌려 놓는다 (Liaw, A., and Wiener, M. December 2002, *Classification and regression by randomForest. R News Vol. 2/3: 18-22*).

[0396] (2) 훈련 및 분류

[0397] 제 2 단계에서, 지원 벡터 머신 분류 알고리즘 (Cortes, C. and Vapnik, V. 1995, *Machine Learning*, 20(3):273-97) 등을 단계 (1) 에서와 같이 확인되고 교차-검증에 근거하여 훈련된 가장 중요한 유전자와 조합된 전사 프로파일을 사용하여 조율하였다.

[0398] 또 다른 방법에서, 단계식 로지스틱 회귀를 가장 중요한 또는 해설적 유전자를 선별하는 단계 (1), 및 교차-검증을 통한 분류를 위해 알고리즘을 훈련하는 단계 (2) 모두에 대해 사용하였다.

[0399] 다른 분석에서, 유전학적 알고리즘에 따라 RVM 분류기를 사용하였다. 데이터 세트를 RVM 알고리즘으로 훈련하고, 가능한 유전자-상호작용을 확인하기 위해 후보자 변수의 상이한 부차세트에서 훈련되고 시험된 다수의 RVM 을 유전학적 알고리즘으로 평가하였다. 각각의 변수 부차세트의 수행을 교차 검증을 통해 평가하였다.

[0400] 훈련 단계 동안, 1 개를 남기는 (leave-one-out) 교차 검증 (LOOCV) 또는 10 배 교차 검증과 같은 교차 검증 방법을 알고리즘에 의해 수행하였다. 교차 검증은 단일 부차세트에 대해 먼저 분석을 수행하는 반면, 다른 부차세트(들) 은 초기 분석의 확인 및 검증에 후속 사용하기 위해 남겨놓는 식으로, 데이터 샘플을 구별되는 부차세트로 분리하는 통계적 실시이다. 데이터의 초기 부차세트는 훈련 세트이고; 다른 부차세트(들) 은 분류를 결정하기 위해 미공지된 것으로 취급된 검증 또는 검정 세트이다.

[0401] 예를 들어, 모든 샘플로부터의 데이터 ( $N$ ) 을 2 개의 구별되는 부차세트 (데이터의 하나의 부차세트 ( $m$ ) 은 샘플의 검증을 위해 사용하고, 즉, 부차세트  $m$  은 미공지의 세트로서 사용된다) 로 나눈다. 나머지 부차세트 ( $N-m$ ) 은 분류 알고리즘을 훈련한다. 이러한 5 교차-검증 (CV) 방법을 모든 데이터 세트를 미공지된 것으로 취급할 때까지 반복한다. 정확성 값 및 예측 값은 미공지된 것으로 취급된 각각의 샘플이 올바르게 분류되는지에 근거하여 계산될 수 있다.

[0402] 하나의 이러한 교차 검증 방법에서, 분류 알고리즘을 샘플 데이터 세트의 90% 로 훈련하고, 샘플 데이터의 나머지 10% 의 분류는 훈련된 알고리즘에 의해 예측된다. 이러한 10-배 CV 를 10 회 반복한다. 교차 검증은 "작업 곡선" 을 설명할 수 있으며, 즉, 훈련된 분류 알고리즘이 더 나은 일부 랜덤 선택 프로세스, 예를 들어 더 나은 가능성을 수행한다. 상기 (1) 및 (2) 에 제시된 처방에 따라 성립된 분류 알고리즘의 분류 오류를 추정하기 위해, 훈련된 분류 알고리즘이 얼마나 잘 수행되는지를 특정하기 위한 정확성, 양성 예측값 (PPV), 및 음성 예측값 (PPV) 을 계산하였다.

[0403] 훈련된 분류 알고리즘의 정확성은 샘플의 총 갯수 중의 올바른 분류의 총 갯수이다.

[0404] 상기 방법에 의해, "질환에 걸린" 부류로 올바르게 점수화된 데이터 세트의 수 (즉, 대상) 는 양성 예측값 (PPV) 의 측정값을 산출한다. 또한 정확성 비율, 또는 질환의 검정 후 확률로도 불리는 PPV 는 올바르게 진단된 양성 검정 결과를 가진 환자의 비율이다.

[0405] 또한 상기 방법에 의해, "건강한" 또는 "대조군" 그룹으로 올바르게 점수화된 데이터 세트의 수 (즉, 대상) 는 음성 예측값 (NPV) 의 측정값을 산출한다. 음성 예측값은 올바르게 진단된 음성 검정 결과를 가진 환자의 비율이다.

[0406] 랜덤화 (치환된) 데이터 세트의 분석.

[0407] SLR 또는 SVM 을 사용하여 수득된 분류 정확성이 유의미한지 (즉, 더 나은 가능성) 를 측정하기 위해, 각각의 데이터 세트를 하기와 같이 추가 분석하였다:

[0408] a) 본래 데이터 세트에 대한 정확성을 본원에 상기 설명된 방법에 의해 수득하였다.

- [0409] b) 3 개의 새로운 치환된 데이터 세트를 생성하고, 각각의 개별 샘플에 대한 배정을 랜덤하게 배정하지만, 본래 데이터 세트와 동일한 환자%를 여전히 유지하였다.
- [0410] c) 그 다음 정확성을 각각의 랜덤화 데이터 세트에 대해 계산하였다.
- [0411] d) 10 개의 정확성 (본래 데이터 세트의 10-배 CV로부터) 을 Mann Whitney 검정을 사용하는 30 치환된 정확성 (10-배 CV를 겪는 3 개의 랜덤 세트)과 비교하였다.
- [0412] e) 0.01 미만의 p 값을 산출하는 비교는 본래 데이터 세트로부터의 정확성이 랜덤 선택성으로 인한 것이 아닌, 즉, 대조군 및 환자 그룹이 분리될 수 있다는 것을 의미하는 것으로 해석된다. 0.01 초과의 p 값을 산출하는 비교는 랜덤한 것으로 간주되며, 환자 및 대조군 그룹이 납득성 있게 분리되지 못한다는 것을 의미한다.
- [0413] 전사 프로파일 확인에 사용되는 환자/대상
- [0414] 본 연구의 하나의 목표는 정상 기증자의 혈액에서 확인되는 전사 프로파일을 정동 장애와 같은 신경정신병적 장애에 대한 위험이 있는 표현형을 확인하는 것을 도울 수 있는 부차그룹으로 정의, 연관 및 연결시키는 것이었다. 일단 정상 지원자의 기준선 전사 프로파일이 성립되면, 정상 집단과 임상적으로 진단된 우울증, 중증 우울증, 양극성 장애, BPD 또는 PTSD를 가진 환자 사이를 비교하였다. 본 연구의 또 다른 목표는 대상을 정상 대조군 또는 우울증, 중증 우울증, 양극성 장애, BPD 또는 PTSD와 같은 정동 장애를 갖는 환자로 분류할 수 있을 프로파일을 확인하는 것이었다.
- [0415] 정상 집단 내 부차그룹의 존재, 예를 들어, 위험 프로파일을 가진 대상을 측정하기 위해, 그리고 부차그룹을 전혈 내 전사 프로파일과 연관시킬 수 있도록, 정상 지원자의 기준선 데이터베이스를 성립하였다.
- [0416] 대조군 환자/대상 (미국)
- [0417] 남동부 펜실베니아와 델라웨어 지역을 담당하는 혈액 은행에서 혈액을 기부한 정상 지원자로부터 500 개의 혈액 샘플을 수집하였다. 모든 기증자로부터 정보 동의를 얻었다. 개인적인 정보는 불가역적으로 익명이었다.
- [0418] 집단 내 변수를 최소화하기 위해 기증자를 코카시안 인종 (Caucasian)에 국한시켰다. 집단 내에서 기증자를 성별로 대등하게 나누었다. 기증자에 대한 혈액 은행에 의해 사용되는 상기 인자 외에 부가적인 배제 인자는 없었다. 모든 기증자는 일반적인 신체적 상태, 의료적 문제, 약물 사용 및 남용, 가족력, 및 정신의학적 문제의 특징화를 돋기 위한 설문지를 채워넣는 것이 요구되었다. 설문지의 항목은 공개 도메인에서 이용 가능한 표준 정신의학적 측정항목에 근거하였다. 설문지에 대한 답은 스스로 적어 넣었고, 기증자는 의료적 또는 정신의학적 평가를 받지 않았다. 설문지는 표 2에 카테고리화된 인자를 포함하는 복합 인자를 포함하였다.
- [0419] 표 2
- | 인구통계학적 | 일반적인 의료항목    | 가족력          | 정신의학적 이력/삶의 경험             | 우울증 증상     |
|--------|--------------|--------------|----------------------------|------------|
| 인종     | 키/체중         | 친척의 자살       | 과거/최근의 스트레스성 일생 사건의 존재/경중도 | 영양적 기능의 변화 |
| 성별     | 현재/과거 약물 치료  | 가족의 정신의학적 이력 | 정신의학적 질병에 대한 이전의 진단        | 인지적 기능의 변화 |
| 결혼 상태  | 현재/과거 의료적 문제 |              |                            | 불안/공황 업무   |
| 고용 상태  | 수술           |              |                            |            |
| 직업     | 흡연           |              |                            |            |
| 식사 빈도  | 음주           |              |                            |            |
|        | 약물 남용        |              |                            |            |
- [0420]
- [0421] 장래의 정신의학적 장애의 위험을 증가시킬 수 있는 기증자의 이력 또는 현재 의료적 상태에서 복합 인자에 대

한 데이터를 수득하고 독특한 전사 프로파일을, 설문지를 사용하여 확인된 구체적인 표현형과 연관시키기 위해 폭넓은 설문지를 사용하였다. 정상 5 개 집단으로 나누고, 현재 이용되는 방법론을 사용하는 것보다 더욱 용이하고 일관적으로 우울증 환자 내 구획을 확인하기 위해 상기 데이터를 사용하였다. 평가되는 요인에는 최근의 스트레스성 일생 사건의 경중도, 초기 일생 스트레스의 존재 및 경중도, 정신의학적 장애의 가족력 및 식욕 및 수면 패턴의 변화를 포함하는 전-우울성 영양적 증상의 그룹을 포함한다 (그러나 이에 제한되지 않는다). 필요한 경우, 질문의 복합 그룹으로부터의 점수를 복합 음성 인자의 영향을 평가하기 위해 조합하였다 (즉, 증상 점수).

[0422] 혈액 전사 프로파일에 잠재적으로 영향을 줄 수 있는 극단으로 고려될 수 있는 흡연 또는 체질량 지수 (BMI) 와 같은 공통 인자의 혼동을 피하기 위해, 설문지 데이터를 인구통계학적, 개인적 또는 의료적 속성 내의 인식 가능한 패턴에 의해 기증자를 그룹화하는데 사용하였다. 전사 프로파일에 대한 영향을 가늠하기 위해 이들 인자를 독립적으로 평가하였다. 기증자의 확인 및 구획화는, 이들이 전-우울성 표현형의 확인에 혼동을 줄 수 있으므로 전사 프로파일에 대한 영향을 평가하기 위해 비-정신의학적 인자에 따랐고, 이러한 인자에는 BMI, 흡연, 알코올 남용, 약물 사용 (및 남용) 이 포함된다. 다른 인자의 영향도 또한 평가하였다.

[0423] 대조군 환자/대상 (덴마크)

[0424] 지리학적으로 덴마크를 포함하고 덴마크인 (Danish) 인종 기원 (2 세대를 거슬러) 에 근거하여 대략 1000 명의 건강한 지원자 (대조군 대상)로부터 초기의 혈액 수집으로부터 200 명의 대상을 선별하였다. 그러므로, 출생지 (및 부모 및 조부모의 출생지) 와 관련된 데이터를 수득하였다. 일반적인 건강 상태 및 정신의학적 이력을 먼저 수득하였다. 정신의학적 이력 정보를 우울증의 이전 에피소드에 대한 간단한 스크리닝으로 보충하였다. 200 명의 대조군 대상의 코호트는 대략 40 세 (범위 18-65 세) 의 평균 연령의 동등한 분포의 남성 및 여성을 산출하였다. 각각의 대상은 복부 및 영덩이 둘레 측정 및 EKG 의 평가를 포함하는 간단한 신체 검진에 노출되었다. 각각의 대상은 성격 및 더욱 철저한 의료적 및 정신의학적 질병의 가족력의 특정한 특성과 관련하여 특징화하는 상세화된 설문지를 완결하였다 (표 2 참조).

[0425] 상기 언급된 바와 같은 대조군 대상에 의해 제공되는 데이터를 사용하여, 정상 집단을 구획화하고, 특이적 표현형을 말초 혈액에서 확인된 전사 프로파일의 변화와 연관시켰다. 표 3A 및 3B 를 참조한다.

[0426] 대조군 환자/대상 (영국)

[0427] 영국에서 통제된 임상 연구에 참여하는 건강한 지원자로부터 혈액 샘플을 수집하였다. 모든 기증자로부터 정보 동의를 얻었다. 남성 및 여성의 본 연구에 포함되었다. 허용되는 피임 방법 (이중-장벽 피임) 을 사용하고, 수술로 불임 시술을 받거나, 폐경기 (2 년 동안 생리가 없는 것으로 정의됨) 인 경우 여성은 포함시켰다 (경구 피임은 허용되지 않았다). 포함된 대상은 연령이 18 세 이상, 45 세 이하, 그러나 65 세 이상 미만이었다. 연구에 포함되는 각각의 대상은 조사자의 의견으로는, 연구 전 신체 검사, 의료적 이력, 바이탈 사인, ECG, 및 혈액 생화학, 혈액학, 및 혈청학 검사 결과, 및 소변검사에 근거하여 건강상태가 양호하다.

[0428] 우울증 환자에서의 전사 프로파일의 확인

[0429] 우울증 환자에서의 전사 프로파일 변화를 평가하기 위해, 우울증 환자, 즉, 주요 우울 장애 (MDD) 를 앓고 있는 환자로부터의 혈액을 통제된 임상 연구에서 수득하였다. 모든 기증자로부터 정보 동의를 얻었다.

[0430] 환자 선별 기준:

[0431] 연구에 자격이 있는 환자/대상은 기준선 방문에서 MADRS 총 점수가 26 이상이고, CGI-S 점수가 4 이상인 중등 MDD 를 앓고 있는 남성 또는 여성 외래환자였다. MDD 의 일차적 진단은 DSM-IV-TR® 기준을 따라야만 한다.

환자의 연령은 18 내지 65 세 (극단값도 포함) 이고, 정신의학 외래환자 클리닉 및 일반적인 개업의로부터 모집되었다. 강박반응성 장애 (Obsessive-Compulsive Disorder: OCD), 외상후 스트레스 장애 (Post-traumatic Stress Disorder: PTSD), 또는 공황 장애 (Panic Disorder: PD) 를 제외한 2 차 동반이환 불안 장애 (DSM-IV-TR® 기준) 를 앓고 있는 환자를 연구에 포함시킬 수 있다. 또한, 환자는 조사자의 의견으로는, 신체 검사, 의료적 이력 및 바이탈 사인에 근거하여 그 외에는 건강하였다. 조사자의 의견으로, 임상 연구 프로토콜에 응하지 않을 것 같거나 어떠한 이유로도 부적합한 환자는 연구로부터 배제될 수 있다.

[0432] 우울증 환자에서의 전사 프로파일의 확인

[0433] 중증 주요 우울 장애 (SMDD) 를 앓고 있는 환자에서의 전사 프로파일 변화를 평가하기 위해, 상기 환자로부터의

혈액을 통제된 임상 연구에서 수득하였다. 모든 기증자로부터 정보 동의를 얻었다.

[0434] 환자 선별 기준:

[0435] 연구에 자격이 있는 환자/대상은 SMDD 를 앓고 있는 외래환자였고, 정신의학 외래환자 클리닉에서 모집되고, 남성 또는 여성이었고, 연령은 18 내지 65 세 (극단값도 포함) 이었다. 본 연구에 포함되는 모든 환자는 MADRS 총 점수가 30 이상 (즉, 더욱 심각한 우울증 환자) 이어야만 한다. 선택된 환자는 DSM IV-TR® 기준 (Mini International Neuropsychiatric Interview (MINI)로 평가된 현행 에피소드)에 따라 일차적 진단으로서 주요 우울증 에피소드 (MDE) 를 앓고 있다. 현행 MDE 의 보고된 기간은 기준선에서 3 개월 이상 12 개월 미만이다. 중등 우울증 환자에 대해 상기 설명된 바와 같은 기준에 근거하여 연구로부터 환자를 포함/배제한다. 조사자의 의견으로, 임상 연구 프로토콜에 응하지 않을 것 같거나 어떠한 이유로도 부적합한 환자는 연구로부터 배제될 수 있다.

[0436] 양극성 환자에서의 전사 프로파일의 확인

[0437] 양극성 환자에서의 전사 프로파일의 변화를 측정하기 위해, 양극성 환자로부터 혈액을 수득하였다. 상기 환자는 정신과 의사에 의해 광범위한 평가를 거쳤고, 의료적 치료를 받았다. 모든 기증자로부터 정보 동의를 얻었다.

[0438] 환자 선별 기준:

[0439] 환자/대상이 상기 프로토콜 하에 혈액을 기증할 수 있기 전에, 하기 기준을 충족해야만 한다:

[0440] a) 환자는 DSM IV-TR®에 따라 중등 또는 중증 주요 우울증 또는 양극성 I로 진단받았다. 환자의 87% 는 양극성 I 장애에 대한 DSM IV-TR® 기준을 충족시켰다.

[0441] b) 환자는 임의의 정신약리학적 약물을 섭취하지 않으며, 혈액 수집시, 2 주 이상 동안 임의의 정신약리학적 약물을 섭취하지 않았다. 또한, 환자 중 누구도 2 개월 이상 동안 플루옥세틴, 비가역적 MAOI 또는 데포 (depot) 신경이완제로 치료되지 않았다.

[0442] c) 다른 급성 정신의학적 증상 (예를 들어, 약물 남용) 을 앓고 있지 않다.

[0443] d) 가능할 때마다, 여성 환자로부터의 혈액 샘플은 생리 시작 2 주 내에 수집해야만 한다. 임의의 경우에서, 마지막 생리 주기의 제 1 일 날짜를 기록할 것이다.

[0444] e) 환자는 지난 6 개월 동안 임의의 불법 약물/남용하는 약물을 섭취하지 않았다.

[0445] f) 환자는 지난 6 개월 동안 알코올을 남용하지 않았다.

[0446] g) 여성 환자는 임신하지 않고 수유하지 않는다.

[0447] h) 환자는 현재 (마지막 주 포함) 임의의 다른 급성 일반적 의료적 상태 (미미한 상태, 예를 들어, 통상의 감기 포함) 를 앓고 있지 않다.

[0448] i) 환자는 현재 (마지막 주 포함) 임의의 정규적 의약 (경구 피임약, 허브 테라피, 영양 보충제, 비타민 포함) 을 섭취하지 않았다.

[0449] j) 환자는 혈액 샘플 수집 전 1 주일 내에는 임의의 의약 (경구 피임약, 허브 테라피, 영양 보충제, 비타민 포함) 을 섭취하지 않아야만 한다. 예를 들어 급성 두통을 위해 약물을 섭취한 경우, 혈액 샘플 수집은 1 주일 연기해야만 한다.

[0450] k) 환자가 흡연하는 경우, 1 일 당 평균량에 대한 정보를 제공할 필요가 있다.

[0451] l) 환자가 남용하지 않고 음주한 경우, 1 주일 당 평균량에 대한 정보를 제공할 필요가 있다.

[0452] m) 환자는 혈액 샘플 수집을 동반한 설문지로 되돌아 간다.

[0453] n) 환자는 환자 정보를 읽고 이해하였다.

[0454] o) 환자는 정보 동의에 사인했다.

[0455] 상기 프로토콜 하에서 혈액을 기증하는 모든 환자로부터 하기 정보: 상세한 정신의학적 및 일반적 의료적 이력,

정신의학적 가족력, 현행 증상의 상세한 임상적 기술, 지난 3 개월 이상 동안 의약 이력, 및 지난 6 개월 이상 내 불법 및 합법적 약물의 남용에 대한 정보가 수득되어야만 한다.

[0456] 경계 인격 장애 환자에서의 전사 프로파일의 확인

[0457] 경계 인격 장애 (BPD) 환자에서의 전사 프로파일의 변화를 측정하기 위해, 경계 인격 장애 환자로부터 혈액을 수득하였다. 상기 환자는 정신과 의사에 의해 광범위한 평가를 거쳤고, 의료적 치료를 받았다. 모든 기증자로부터 정보 동의를 얻었다.

[0458] BPD 연구를 위한 환자/대상 선별 기준:

[0459] 환자가 상기 프로토콜 하에 혈액을 기증할 수 있기 전에, 하기 기준을 충족해야만 한다:

[0460] a) 환자는 DSM-IV®에 따라 경계 인격 장애로 진단받았다.

[0461] b) 미처리된 환자 그룹에 대해, 환자는 임의의 정신약리학적 약물을 섭취하지 않으며, 혈액 수집시, 2 주 이상 동안 임의의 정신약리학적 약물을 섭취하지 않았다. 과거에 플루옥세틴, 비가역적 MAOI 또는 데포 신경이완제로 치료된 환자는, 혈액 수집 전 4 주 이상 동안 상기 의약 중 임의의 것을 섭취하지 않았다.

[0462] c) 소수의 환자 코호트 (대략 25 명의 환자)로부터, 혈액 샘플은 일차적 정신의학적 장애 (경계 인격 장애)의 급성 정신의학적 악화 동안 수집될 것이다. 모든 다른 환자는 혈액 수집시 급성 정신의학적 악화를 앓지 않을 것이다. 급성 악화 동안 혈액이 수집되는 환자에서만, 진정 동안 제 2 샘플을 수집할 것이다. 의료적으로 가능할 때마다, 2 회 시점에서의 처리는 동일할 것이다.

[0463] d) 환자는 다른 급성 정신의학적 증상 (예를 들어, 약물 남용) 을 앓고 있지 않다.

[0464] e) 가능할 때마다, 여성 환자로부터의 혈액 샘플은 생리 시작 2 주 내에 수집해야만 한다. 임의의 경우에서, 마지막 생리 주기의 제 1 일 날짜를 기록할 것이다.

[0465] f) 환자는 지난 6 개월 동안 임의의 불법 약물/남용하는 약물을 섭취하지 않았다.

[0466] g) 환자는 지난 6 개월 동안 알코올을 남용하지 않았다.

[0467] h) 여성 환자는 임신하지 않고 수유하지 않는다.

[0468] i) 환자는 현재 (마지막 주 포함) 임의의 다른 급성 일반적 의료적 상태 (미미한 상태, 예를 들어, 통상의 감기 포함) 를 앓고 있지 않다.

[0469] j) 환자는 현재 (마지막 주 포함) 처방받은 벤락파신 또는 둘록세틴 외에는 임의의 정규적 의약 (경구 피임약, 허브 테라피, 영양 보충제, 비타민 포함) 을 섭취하지 않았다.

[0470] k) 환자가 벤락파신 또는 둘록세틴으로 치료받은 경우, 치료는 3 개월 이상 동안 현행 투여량으로 제공되어야만 한다.

[0471] l) 환자는 혈액 샘플 수집 전 1 주일 내에는 임의의 의약 (경구 피임약, 허브 테라피, 영양 보충제, 비타민 포함) 을 섭취하지 않아야만 한다. 예를 들어 급성 두통을 위해 약물을 섭취한 경우, 혈액 샘플 수집은 1 주일 연기해야만 한다.

[0472] m) 환자가 흡연하는 경우, 1 일 당 평균량에 대한 정보를 제공할 필요가 있다.

[0473] n) 환자가 남용하지 않고 음주한 경우, 1 주일 당 평균량에 대한 정보를 제공할 필요가 있다.

[0474] o) 환자는 혈액 샘플 수집을 동반한 설문지로 되돌아 간다.

[0475] p) 환자는 환자 정보를 읽고 이해하였다.

[0476] q) 환자는 정보 동의에 사인했다.

[0477] 상기 프로토콜 하에서 혈액을 기증하는 모든 환자로부터, 가족력, 임상적 기술 및 의약 및 약물 기록을 비롯한 상세한 정신의학적 이력을 수득한다.

[0478] 환자는 전사 프로파일, 예를 들어, 약물 사용, 일반적 의료적 상태가 혼동될 수 있는 인자를 구체적으로 고심하기 위해 개발된 설문지를 완료하였다. 환자는 조사자에게 설문지로 되돌아 간다. 설문지는 혈액 샘플 및 다른 임상 데이터와 동일한 코드로 코드화하여, 전사 분석 현장에서 환자의 신분이 직원에게 노출되지 않도록

록 하였다. 설문지는 혈액 샘플과 함께 전사 분석 현장으로 옮겨졌다.

[0479] 외상후 스트레스 장애 (PTSD) 환자에서의 전사 프로파일

[0480] PTSD 환자에서의 전사 프로파일의 변화를 측정하기 위해, PTSD 환자로부터 혈액을 수득하였다. 상기 환자는 정신과 의사에 의해 광범위한 평가를 거쳤고, 의료적 치료를 받았다. 모든 기증자로부터 정보 동의를 얻었다.

[0481] PTSD 연구를 위한 환자/대상 선별 기준:

[0482] 본 연구를 위한 대상은 하기 기준을 충족한 남성이었다:

[0483] a) 대상은 급성 PTSD, 또는 완화된 PTSD (DSM-IV® 에 따름)로 진단받았거나, 트라우마에 노출되고 PTSD 가 전개되지 않거나, 대조군으로서 카테고리가 나뉜다. 트라우마에 노출되지 않은 대조군이 본 연구를 위해 선택되었고, 동일한 지리학적 영역으로부터 기원하였다.

[0484] b) 환자는 임의의 정신약리학적 약물을 섭취하지 않으며, 혈액 수집시 2 주 이상 동안 임의의 정신약리학적 약물을 섭취하지 않았다. 과거에 플루옥세틴, 비가역적 MAOI 또는 데포 신경이완제로 치료된 환자는, 혈액 수집 전 4 주 이상 동안 상기 의약 중 임의의 것을 섭취하지 않았다.

[0485] c) 환자는 다른 급성 정신의학적 증상 (예를 들어, 약물 남용)을 앓고 있지 않다.

[0486] d) 환자는 지난 6 개월 동안 임의의 불법 약물/남용하는 약물을 섭취하지 않았다.

[0487] e) 환자는 지난 6 개월 동안 알코올을 남용하지 않았다.

[0488] f) 환자는 현재 (마지막 주 포함) 임의의 다른 급성 일반적 의료적 상태 (미미한 상태, 예를 들어, 통상의 감기 포함)를 앓고 있지 않다.

[0489] g) 환자는 혈액 샘플 수집 전 1 주일 내에는 임의의 의약 (허브 테라피, 영양 보충제, 비타민 포함)을 섭취하지 않아야만 한다. 예를 들어 급성 두통을 위해 약물을 섭취한 경우, 혈액 샘플 수집은 1 주일 연기해야만 한다.

[0490] h) 환자가 흡연하는 경우, 1 일 당 평균량에 대한 정보를 제공할 필요가 있다.

[0491] i) 환자가 남용하지 않고 음주한 경우, 1 주일 당 평균량에 대한 정보를 제공할 필요가 있다.

[0492] j) 환자는 현재 (마지막 주 포함) 허브 테라피, 영양 보충제, 비타민을 포함하는 임의의 정규적 의약을 섭취하지 않았다.

[0493] 상기 기재한 바와 같은 모든 임상적 및 인구 통계학적 데이터를, 상기 정보를 전사 분석 현장 (Lundbeck Research USA, Inc., Paramus, NJ)에 전달하기 전 혈액 채취 현장에서 수집하였다. 임상적 특징과 전사 프로파일 사이의 임의의 관련성의 탐색적 분석을, Lundbeck Research USA에서 환자의 신분에 대한 지식 없이 수행하였다.

[0494] 결과 및 토의

[0495] 대조군 대상에서의 전사 프로파일의 확인

[0496] 표 1A에 열거된 29 개의 유전자에 대한 유전자 발현 수준을, 2 개의 대조군 (U.S. 및 DK)으로부터의 대상을 포함하는 대조군 대상으로부터의 혈액 샘플에서 측정하였다.

[0497] 이러한 개인들이 모두 건강함에도 불구하고, 설문지 항목에 대한 특정 응답과 상관관계가 있는 유전자 발현의 경향을 확인하였다. 확인된다면 이러한 경향은 우울증 환자 집단에서 과장될 수 있다.

[0498] 통계 분석을 위한 설문지 응답의 코드화된 값으로의 전환.

[0499] US 및 덴마크 대조군 대상에 의해 채워진 자가-평가 설문지는 유사한 항목을 담고 있었으나, 동일하지는 않았다. 응답과 유전자 발현 데이터 사이의 가능한 연관성을 조사하기 위해 설문지로부터의 정보를 사용하기 위해, 통계 분석 전 정보를 코드화하는 것이 필요하였다.

[0500] 코드화 전략의 예는 하기와 같다:

[0501] a) 연령 및 BMI 와 같은 연속 변수를 대상에 의해 보고된 바와 같이 사용하였다. 대안적으로는, 분석 전에

미가공 점수를 2 또는 3 개의 칸 (높은 값, 중간값, 낮은 값) 내에 조합시켰다.

[0502] b) 성별을 이진법 응답 (0, 1) 으로 전환하였다.

[0503] c) 우울증과 연결된 증상의 빈도에 대한 질문, 예컨대 수면 장애, 에너지 결핍, 또는 무기력함을 단어 대답 (전혀 없음, 가끔 있음, 대부분 그러함, 매일 그러함) 으로부터 수치 값 (0, 1, 2, 3) 으로 전환하였다.

[0504] d) 복합 점수를 생성하기 위해, 증상의 특이적 조합에 대한 값을 첨가함으로써 조합된 증상 점수를 생성하였다. 이후 복합 점수를 칸에 넣었다.

[0505] e) 대상의 우울증/불안의 가족력과 관련된 질문을 단어 대답 (없음, 2 차 친척만, 1 차 친척) 으로부터 수치 값 (0, 1, 2) 으로 전환하였다.

[0506] f) 대상의 우울증/불안 또는 우울증/불안에 대한 약리학적 치료의 개인력과 관련된 질문을 단어 대답 (없음, 1 회 이상) 을 이진법 응답 (0, 1) 으로 전환하였다.

[0507] 코드화 후, 스피어만 (Spearman) 상관관계 분석, t-검정 및 ANOVA 를 포함하는 다양한 통계적 검정을 사용하여 유전자 발현 수준과 특이적 임상 변수 사이의 연관성을 탐색하였다.

[0508] 적절히 통계적 검정을 사용하여, 각 유전자의 발현을 자가-평가 설문지에 대해 대상에 의해 제공된 코드화된 대답과 비교하여 상관관계를 확인하였다. 총 377 개의 비교가 이루어졌기 때문에 (29 개의 유전자 × 13 가지 설문지 응답), 유의성에 대한 역치는 유형 1 오류의 가능성성이 최소화되도록  $p<0.01$  에서 설정된 한편, 여전히 다수의 통계적으로 유의한 결과가 보유되었다.

[0509] 표 3A 및 3B 는, 분석한 설문지 응답에 근거하여 대조군 집단 내에 유의한 차이점을 갖는 (표 1A로부터의) 29 개의 유전자 중 단지 15 개에 대한 상관관계 데이터를 나타낸다. 나머지 유전자에 대해서는 유의한 차이점이 검출되지 않았다. 표 3A 및 3B 는 13 가지의 설문지 응답 중 11 개에 대한 데이터를 나타내지만, 유의하게 상이하지 않기 때문에 BMI 및 연령에 대한 상관관계 데이터는 제시되지 않았다. 유의한 유전자 발현 프로파일과 상관관계가 있는 일부 임상 파라미터는 일생 경험, 일생 치료 및 증상 점수이다.

[0510] 표 3A 및 3B. 두 대조군에서의 임상적 변수 및 유전자 발현 사이의 상관관계

[0511] ( $** = p<0.01$  기준;  $*** = p<0.001$  기준)

[0512]

표 3A 1) US 대상 2) DK 대상

|                 | CREB2   | DPP4    | ERK1   | ERK2    | GR       | Gs     | MAPK8   | MAPK14 |
|-----------------|---------|---------|--------|---------|----------|--------|---------|--------|
| 1) 가족력 (D/A/S)  |         | Inc **  |        |         |          |        | Inc **  |        |
| 2) 가족력 (D/A/S)  |         |         |        |         |          |        |         |        |
| 1) 흡연           |         |         |        |         |          |        |         |        |
| 2) 흡연           |         |         |        |         |          |        |         |        |
| 1) 일생 경험 (D/A)  | Inc *** | Inc *** |        | Inc **  | Inc ***  |        | Inc *** |        |
| 2) 일생 경험 (D/A)  |         | Inc *** |        |         | trend up |        |         |        |
| 1) 일생 치료 (D/A)  | Inc **  | Inc *** |        |         | Inc **   |        | Inc *** |        |
| 2) 일생 치료 (D/A)  |         |         |        |         | trend up | Inc ** |         |        |
| 1) 식욕 변화        |         | Inc **  |        |         |          |        |         |        |
| 2) 식욕 변화        |         |         |        |         |          |        |         |        |
| 1) 수면 장애        |         | Inc **  |        |         |          |        |         |        |
| 2) 수면 장애        |         | Inc **  |        |         |          |        |         |        |
| 1) 10 증상 점수 (*) |         | Inc *** |        |         |          |        | Inc *** |        |
| 2) 10 증상 점수 (*) |         |         | Inc ** | Inc *** |          |        |         |        |
| 1) 영양적 증상       |         | Inc **  |        |         |          |        |         |        |
| 2) 영양적 증상       |         |         |        |         |          |        |         |        |
| 1) 최근의 스트레스     |         | Inc **  |        |         |          |        |         |        |
| 2) 최근의 스트레스     |         |         |        |         |          |        |         |        |
| 1) 어린시절 스트레스    |         |         |        |         |          |        |         |        |
| 2) 어린시절 스트레스    |         |         |        |         |          |        |         |        |
| 1) 섹스에 대한 관심    |         |         |        |         | Inc **   |        |         |        |
| 2) 섹스에 대한 관심    |         |         |        |         |          |        |         | Inc ** |

(D/A/S = 우울증/불안/자살; D/A = 우울증/불안)

[0513]

[0514]

표 3B 1) US 대상 2) DK 대상

|                 | MKP1        | MR     | PBR | RGS2          | S100<br>A10   | SERT | VMAT2  |
|-----------------|-------------|--------|-----|---------------|---------------|------|--------|
| 1) 가족력 (D/A/S)  |             |        |     |               |               |      |        |
| 2) 가족력 (D/A/S)  |             |        |     |               |               |      |        |
| 1) 흡연           |             |        |     |               | Dec<br>***    |      |        |
| 2) 흡연           |             |        |     |               | trend<br>down |      |        |
| 1) 일생 경험 (D/A)  |             |        |     | Inc<br>**     |               |      | Inc ** |
| 2) 일생 경험 (D/A)  |             |        |     | trend<br>up   |               |      |        |
| 1) 일생 치료 (D/A)  | Inc<br>**   |        |     |               |               |      |        |
| 2) 일생 치료 (D/A)  | trend<br>up |        |     |               |               |      |        |
| 1) 식욕 변화        |             |        |     |               | Dec **        |      |        |
| 2) 식욕 변화        |             |        |     |               | trend<br>down |      |        |
| 1) 수면 장애        |             |        |     |               |               |      |        |
| 2) 수면 장애        |             |        |     |               |               |      |        |
| 1) 10 증상 점수 (*) | trend<br>up |        |     | Dec<br>**     |               |      |        |
| 2) 10 증상 점수 (*) | Inc<br>**   |        |     | Inc<br>**     |               |      |        |
| 1) 영양적 증상       |             |        |     |               |               |      |        |
| 2) 영양적 증상       |             |        |     |               |               |      |        |
| 1) 최근의 스트레스     |             |        |     |               |               |      |        |
| 2) 최근의 스트레스     |             |        |     |               |               |      |        |
| 1) 어린시절 스트레스    |             |        |     |               |               |      |        |
| 2) 어린시절 스트레스    |             |        |     |               | Inc<br>***    |      |        |
| 1) 섹스에 대한 관심    |             |        |     | trend<br>down |               |      |        |
| 2) 섹스에 대한 관심    |             | Dec ** |     | Inc<br>**     |               |      |        |

(D/A/S = 우울증/불안/자살; D/A = 우울증/불안)

[0515]

분석한 377 개의 총 조합 중, 23 개 조합 (6%) 이 분석한 두 대조군 사이에 유의한 차이점을 나타낸다. 그러나, 345 개 조합 (94%) 은 동일한 프로파일을 나타낸다. 표 3A 및 3B 에서의 음영 상자에 의해 나타내는 바와 같이, 이들 조합 중 9 개는 연구한 대조군 모두에 대해 동일한 방향 (즉, 유전자의 업- 또는 다운-조절) 으로 유전자 발현에서의 변화를 나타낸다. 전체적으로, 분석은 분석에 사용된 두 대조군이 매우 유사한 유전자 발현 경향 또는 프로파일을 나타낸다는 것을 보여준다.

[0517]

임상적 파라미터에 관련된 유전자 발현 프로파일은 또한 본원에 기재된 다변량 알고리즘에 의해 분석될 수 있다. 따라서, 전사 데이터와 조합된 임상적 변수는 단계적 로지스틱 회귀 또는 PELORA 와 같은 당업자에게 알려져 있는 임의의 적합한 알고리즘에 적용될 수 있다.

[0518]

우울증 환자에서의 전사 프로파일의 확인

[0519]

먼저 항우울제 치료를 받지 않은 174 명의 중등 우울증 환자/대상으로부터 수득한 혈액 샘플을 단변량 방법에 의해 분석하였다. 표 1A 에서 선택된 유전자에 대한 전사 수준을 측정하고 196 명의 건강한 대조군 대상에서의 상기 유전자의 발현 수준과 비교하였다. 대조군과 비교한 바와 같은 우울증 환자에서의 대표적 유전자의 발현 프로파일을 도 2A-2B 및 3A-3B 에 제시한다.

[0520]

RF (선별) 및 SVM (훈련) 을 사용하는 중등 우울증 환자 대 대조군의 분류로, 도 8A 에서 나타낸 바와 같이 88%

의 높은 정확성이 야기되었다 (PPV = 89%; NPV = 88%). 유전자 선별 및 훈련을 모두 수행하는 SLR 알고리즘을 사용하는 중등 우울증 환자 대 대조군의 분류로, 도 8A에 제시된 바와 같이 93%의 높은 정확성이 야기되었다 (PPV = 93%; NPV = 94%).

[0521] 알고리즘 모두는 도 8B에 제시된 전체 데이터 세트를 기준으로 선택된 유전자에 있어서 양호한 일치를 나타내었다. 각 방법의 통계적 파라미터를 기준으로 분류를 위해 가장 중요한 유전자로서, 랜덤 포레스트로 14개의 유전자가 선택되고 SLR로 17개의 유전자가 선별되었다. ARRB1, ARRB2, CD8a, CREB1, CREB2, ERK2, Gi2, MAPK14, ODC1, P2X7 및 PBR을 포함하여 11개의 유전자가 두 방법 모두에 의해 선별되었다.

[0522] 데이터 세트를 랜덤화하고, 즉 환자 또는 대조군으로서의 샘플의 배정을 랜덤화하고, 상기와 같은 동일한 다변량 분석에 적용하였다. 랜덤화 후, 두 분류 알고리즘 (RF/SVM 및 SLR)에서 실제 데이터로 수득된 것들과 통계학적으로 상이한 정확성 값이 생성되었는데, 이는 상기 열거한 값 (도 8A)이 더 나은 가능성이 있으며, 그룹을 통계학적으로 분리할 수 있다는 것을 나타낸다.

[0523] 대상은 프로파일링될 수 있으며 표 1A에서의 유전자를 기준으로 한 이의 전사 데이터는 중등 우울증의 진단이 얻어지도록 상기 본원에 기재된 바와 같은 파라미터로 훈련된 분류 알고리즘에 적용될 수 있다.

[0524] 표 1A에서 선택된 유전자에 대한 우울증 대상의 전사 프로파일을, 각각의 바이오마커의 풍부함 (즉, 유전자 전사체)을 기준으로 표 4에 제시한다. 대조군 대상 전사체 값을 비교를 위해 나타낸다.

[0525]

표 4

| 바이오마커<br>(유전자 약어) | 우울증 대상 그룹<br>특징:<br>풍부함 =<br>바이오마커의 평균<br>전사체 값 ( $\pm$ SD) | 대조군 대상 그룹<br>특징: 풍부함 =<br>바이오마커의<br>평균 전사체 값<br>( $\pm$ SD) |
|-------------------|---|---|
| ADA               | 4691 $\pm$ 2453   | 4511 $\pm$ 1710   |
| ARRB1             | 189062 $\pm$ 62727  | 297143 $\pm$ 91094  |
| ARRB2             | 84195 $\pm$ 31728   | 114780 $\pm$ 39962  |
| CD8a              | 8304 $\pm$ 5825   | 14693 $\pm$ 8416  |
| CD8b              | 8145 $\pm$ 4394   | 8687 $\pm$ 3880   |
| CREB1             | 71743 $\pm$ 20237   | 63725 $\pm$ 16022   |
| CREB2             | 63732 $\pm$ 14463   | 77059 $\pm$ 15755   |
| DPP4              | 6649 $\pm$ 2331   | 7169 $\pm$ 2890   |
| ERK1              | 25326 $\pm$ 10178   | 39016 $\pm$ 12900   |
| ERK2              | 58338 $\pm$ 18813   | 54137 $\pm$ 18660   |
| Gi2               | 115117 $\pm$ 53383  | 226358 $\pm$ 87609  |
| Gs                | 262885 $\pm$ 112989   | 303930 $\pm$ 139837   |
| GR                | 73224 $\pm$ 23517   | 80610 $\pm$ 26544   |
| IL1b              | 29631 $\pm$ 13692   | 21006 $\pm$ 9313  |
| IL6               | 348 $\pm$ 523   | 182 $\pm$ 221   |
| IL8               | 45487 $\pm$ 106224  | 28024 $\pm$ 19993   |
| INDO              | 6031 $\pm$ 10133  | 5596 $\pm$ 4418   |
| MAPK14            | 73156 $\pm$ 33915   | 51632 $\pm$ 20341   |
| MAPK8             | 12906 $\pm$ 3836  | 12162 $\pm$ 3500  |
| MKP1              | 525383 $\pm$ 268053   | 499308 $\pm$ 220665   |
| MR                | 2565 $\pm$ 1110   | 2830 $\pm$ 887  |
| ODC1              | 71892 $\pm$ 32249   | 58670 $\pm$ 40801   |
| P2X7              | 1095 $\pm$ 432  | 1542 $\pm$ 563  |
| PBR               | 70854 $\pm$ 30278   | 64439 $\pm$ 29328   |
| PREP              | 6715 $\pm$ 2072   | 7072 $\pm$ 2102   |
| RGS2              | 632976 $\pm$ 262593   | 477280 $\pm$ 165907   |
| S100A10           | 32173 $\pm$ 9530  | 35819 $\pm$ 10568   |
| SERT              | 1400 $\pm$ 1164   | 1711 $\pm$ 1317   |
| VMAT2             | 3469 $\pm$ 1602   | 2792 $\pm$ 1344   |

(SD = 표준 편차)

[0526]

[0527] 또한 우울증 대상 대 대조군 대상에 대한 전사체 값의 비를 비교하여 2 개 유전자 조합을 평가하였다. 표 4A 에서와 같이 우울증 대상과 대조군 대상 사이에서 특정 바이오마커의 풍부함의 비에 있어서의 현저한 차이점이 나타난다.

[0528]

표 4A

| 바이오마커  | 우울증 대상 그룹에 대한 전사체의 풍부함의 비 | 대조군 그룹에 대한 전사체의 풍부함의 비 |
|--------|---------------------------|------------------------|
| ERK1   |                           |                        |
| MAPK14 | 0.35                      | 0.76                   |
| IL1b   |                           |                        |
| Gi2    | 0.26                      | 0.09                   |
| MAPK14 |                           |                        |
| ARRB1  | 0.39                      | 0.17                   |
| ERK1   |                           |                        |
| IL1b   | 0.85                      | 1.86                   |

[0529]

[0530]

더욱 중증인 우울증 환자 집단에서의 전사 프로파일에 있어서의 변화를 평가하기 위해, 120 명의 중증 우울증 환자로부터의 혈액을 수득하고 표 1A 에서 선택된 유전자에 대해 유전자 발현을 측정하였다. 단변량 방법에 의해 유전자 발현 데이터를 통계학적으로 평가하였다. 환자 전사 데이터를 196 명의 대조군의 것과 비교하고, 개별적 유전자 데이터에 대한 대표적 산점도를 도 4A-4C 에 나타낸다.

[0531]

RF/SVM 을 사용하는 분류로, 92% 의 높은 정확성이 야기되었다 (PPV = 89%; NPV = 94%). 유전자 선별 및 훈련을 모두 수행하는 SLR 알고리즘의 분류로, 93% 의 높은 정확성이 야기되었다 (PPV = 91%; NPV = 95%).

[0532]

알고리즘 모두는 전체 데이터 세트를 기준으로 선택된 유전자에 있어서 양호한 일치를 나타내었다. 각각의 방법의 통계적 파라미터를 기준으로 분류를 위해 가장 중요한 유전자로서, 랜덤 포레스트 분류로 총 7 개의 유전자가 선택되고 SLR 로 총 12 개의 유전자가 선별되었다. CD8a, ERK1, MAPK 14, P2X7, 및 PBR 을 포함하여 5 개의 유전자가 두 방법 모두에 의해 선별되었다.

[0533]

환자/대조군 배정의 랜덤화 후, 두 분류 알고리즘 (RF/SVM 및 SLR) 에서 실제 데이터로 수득된 것들과 통계학적으로 상이한 정확성 값이 생성되었는데, 이는 상기 열거한 값이 더 나은 가능성이 있으며, 그룹을 통계학적으로 분리할 수 있다는 것을 나타낸다.

[0534]

대상은 프로파일링될 수 있으며 표 1A 에 포함된 유전자를 기준으로 한 이의 전사 데이터는 중증 우울증의 진단이 얻어지도록 상기 본원에 기재된 바와 같은 파라미터로 훈련된 분류 알고리즘에 적용될 수 있다.

[0535]

표 1A 에서 선택된 유전자에 대한 중증 우울증 대상의 전사 프로파일을, 각각의 바이오마커의 풍부함 (즉, 유전자 전사체) 을 기준으로 표 5 에 제시한다. 대조군 대상 전사체 값을 비교를 위해 나타낸다.

[0536]

표 5

| 바이오마커<br>(유전자 약어) | 중증 우울증 대상<br>그룹 특징:<br>풍부함 =<br>바이오마커의 평균<br>전사체 값 ( $\pm SD$ ) | 대조군 대상 그룹<br>특징: 풍부함 =<br>바이오마커의<br>평균 전사체 값<br>( $\pm SD$ ) |
|-------------------|---|--|
| ADA               | 3812 $\pm$ 1365   | 4511 $\pm$ 1710  |
| ARRB1             | 161284 $\pm$ 47341  | 297143 $\pm$ 91094   |
| ARRB2             | 79487 $\pm$ 22860   | 114780 $\pm$ 39962   |
| CD8a              | 7666 $\pm$ 4603   | 14693 $\pm$ 8416   |
| CD8b              | 6897 $\pm$ 3320   | 8687 $\pm$ 3880  |
| CREB1             | 64463 $\pm$ 18736   | 63725 $\pm$ 16022  |
| CREB2             | 71534 $\pm$ 12311   | 77059 $\pm$ 15755  |
| DPP4              | 5873 $\pm$ 2194   | 7169 $\pm$ 2890  |
| ERK1              | 19389 $\pm$ 7612  | 39016 $\pm$ 12900  |
| ERK2              | 48236 $\pm$ 17894   | 54137 $\pm$ 18660  |
| Gi2               | 97344 $\pm$ 42195   | 226358 $\pm$ 87609   |
| Gs                | 185642 $\pm$ 82731  | 303930 $\pm$ 139837  |
| GR                | 75411 $\pm$ 24542   | 80610 $\pm$ 26544  |
| IL1b              | 27643 $\pm$ 12046   | 21006 $\pm$ 9313   |
| IL6               | 153 $\pm$ 100   | 182 $\pm$ 221  |
| IL8               | 38817 $\pm$ 29253   | 28024 $\pm$ 19993  |
| INDO              | 5735 $\pm$ 5467   | 5596 $\pm$ 4418  |
| MAPK14            | 67519 $\pm$ 29094   | 51632 $\pm$ 20341  |
| MAPK8             | 11446 $\pm$ 3231  | 12162 $\pm$ 3500   |
| MKP1              | 615915 $\pm$ 307961   | 499308 $\pm$ 220665  |
| MR                | 2023 $\pm$ 893  | 2830 $\pm$ 887   |
| ODC1              | 55085 $\pm$ 30043   | 58670 $\pm$ 40801  |
| P2X7              | 769 $\pm$ 331   | 1542 $\pm$ 563   |
| PBR               | 67863 $\pm$ 24974   | 64439 $\pm$ 29328  |
| PREP              | 5186 $\pm$ 1620   | 7072 $\pm$ 2102  |
| RGS2              | 571284 $\pm$ 270572   | 477280 $\pm$ 165907  |
| S100A10           | 21812 $\pm$ 7985  | 35819 $\pm$ 10568  |
| SERT              | 795 $\pm$ 553   | 1711 $\pm$ 1317  |
| VMAT2             | 3073 $\pm$ 1715   | 2792 $\pm$ 1344  |

(SD = 표준 편차)

[0537]

[0538] 중증 우울증 환자와 대조군 사이에서 평균 발현 수준 (전사체 값) 이 유의하게 상이하였던 유전자 ( $p<0.05$ ) 는: ADA, ARRB1, ARRB2, CD8a, CD8b, CREB2, DPP4, ERK1, Gi2, Gs, IL1b, IL8, MAPK 14, MKP1, MR, P2X7, PREP, RGS2, S100A10, 및 SERT (표 5A) 이다.

[0539]

표 5A:  $p$ -값 ( $p<0.05$ ) 에 근거하여, 대조군 대상과 비교하였을 때 중증 우울증 대상에서 유의하게 상이한 유전자.

| 바이오마커<br>(유전자 약어) | p-값                       |
|-------------------|---------------------------|
| ADA               | $3.2673 \times 10^{-6}$   |
| ARRB1             | $4.40419 \times 10^{-60}$ |
| ARRB2             | $1.61434 \times 10^{27}$  |
| CD8a              | $1.92916 \times 10^{38}$  |
| CD8b              | $3.13307 \times 10^8$     |
| CREB2             | 0.0000507671              |
| DPP4              | $1.25015 \times 10^7$     |
| ERK1              | $1.12946 \times 10^{-72}$ |
| Gi2               | $3.27538 \times 10^{-64}$ |
| Gs                | $1.98625 \times 10^{35}$  |
| IL1b              | $2.13924 \times 10^{-11}$ |
| IL8               | $2.00073 \times 10^{-6}$  |
| MAPK14            | $5.2042 \times 10^{-15}$  |
| MKP1              | $1.25421 \times 10^{-6}$  |
| MR                | $1.73784 \times 10^{-23}$ |
| P2X7              | $3.7121 \times 10^{-67}$  |
| PREP              | $2.72022 \times 10^{-26}$ |
| RGS2              | 0.0000152985              |
| S100A10           | $2.3756 \times 10^{-53}$  |
| SERT              | $4.36216 \times 10^{-26}$ |

[0540]

[0541] 상기 유전자는 계산된 -Log(p) 값의 크기에 따라 순위를 매겼고 (도 9), 그러므로 ERK1, P2X7, Gi2, ARRB1 및 S100A10 과 같은 여러 유전자에 대해 환자 전사체 값과 대조군 값 사이의 현저한 차이를 나타낸다.

[0542]

전사체 값 간의 선형 및 비선형 상호작용에 대해 탐색하기 위해, 관련성 벡터 머신 (RVM) 분류 알고리즘을 수행한 다음, 가능한 유전자-유전자 상호작용 간격을 통해 탐색하고 가장 강력하고 의미있는 상호작용을 선택하기 위해 유전학적 알고리즘을 사용하였다. 또한 단일-유전자 해법을 상기 세트의 알고리즘에 의해 시험하였고, 대조군과 환자를 분리하기 위한 단일-유전자 해법의 타당성을 확인시켜준다. ARRB1 (정확성 = 0.86) 및 ERK1 (정확성 = 0.85)은 단일-유전자 분석에서 고도로 정보적인 것으로 측정되고, 그 후 P2X7 (정확성 = 0.82) 및 Gi2 (정확성 = 0.81)가 측정된다. 또한, 예를 들어, 정보적 유전자 발현 데이터를 중등 우울증, 중증 우울증 및 양극성 환자 대 대조군에 대해 묘사하고 있는 도 2 ~ 5 를 참조한다.

[0543]

여러 개의 2 개-유전자 해법은 우울증 환자와 대조군을 90% 이상의 정확성으로 분류하기 위해 확인되었다. 우울증 환자 대 대조군을 분류하기 위해 HRK1 및 MAPK14 전사체 값이 92% 의 정확성으로 제시된다. 도 10 에는 오로지 ERK1 및 MAPK14 의 전사체 값에 근거한 중증 우울증 대상 및 대조군의 분포가 묘사된다. 우울증 대상의 분류 (표 4 에서와 같은 프로파일로) 는 중증 우울증 대상의 결과와 일치한다. 도 11, 12 및 13 에는 다른 2 개-유전자 전사 프로파일, IL1b/Gi2, MAPK14/ARRB1, 및 ERK1/IL1b 각각의 전사체 값에 근거한 중증 우울증 대상 및 대조군의 분포가 묘사된다. 2 개-유전자 조합은 도한 중증 우울증 대상 대 대조군 대상에 대한 전사체 값의 비를 비교함으로써 평가되었다. 중증 우울증 대상과 대조군 대상 사이에 풍부함의 비의 현저한 차이는 표 5B 에 제시된다.

[0544]

표 5B

| 바이오마커  | 증증 우울증 대상 그룹에 대한 전사체의 풍부함의 비 | 대조군 그룹에 대한 전사체의 풍부함의 비 |
|--------|------------------------------|------------------------|
| ERK1   | 0.29                         | 0.76                   |
| MAPK14 |                              |                        |
| IL1b   | 0.28                         | 0.09                   |
| Gi2    |                              |                        |
| MAPK14 | 0.42                         | 0.17                   |
| ARRB1  |                              |                        |
| ERK1   | 0.70                         | 1.86                   |
| IL1b   |                              |                        |

[0545]

양극성 장애 환자에서의 전사 프로파일의 확인

[0546]

양극성 장애 환자에서의 전사 프로파일의 변화를 측정하기 위해, 23 명의 우울증 환자 (20 명의 환자는 DSM-IV 기준에 따라 양극성 장애로 명확하게 진단되었음)로부터의 혈액을 수득하고 표 1A에서 선택된 유전자에 대해 유전자 발현을 측정하였다. 단변량 방법에 의해 유전자 발현 데이터를 통계학적으로 평가하였다. 환자 전사 데이터를 196 명의 대조군의 것과 비교하고, 개별적 유전자 데이터에 대한 대표적 산점도를 도 5A-5C에 나타낸다.

[0547]

RF/SVM 을 사용하는 분류로, 94% 의 높은 정확성이 야기되었다 (PPV = 86%; NPV = 95%). 유전자 선별 및 훈련을 모두 수행하는 SLR 알고리즘의 분류로, 97% 의 높은 정확성이 야기되었다 (PPV = 90%; NPV = 99%).

[0548]

알고리즘 모두는, 각각의 방법의 통계적 파라미터를 기준으로 분류를 위해 가장 중요한 유전자로서, 랜덤 포레스트 분류로 총 3 개의 유전자가 선택되고 SLR로 총 5 개의 유전자가 선별되어, 전체 데이터 세트를 기준으로 선택된 유전자에 있어서 양호한 일치를 나타내었다. Gi2, GR, 및 MAPK14 를 포함하여 3 개의 유전자가 두 방법 모두에 의해 선별되었다.

[0549]

환자/대조군 배정의 랜덤화 후, 두 분류 알고리즘 (RF/SVM 및 SLR)에서 실제 데이터로 수득된 것들과 통계학적으로 상이한 정확성 값이 생성되었는데, 이는 상기 열거한 값이 더 나은 가능성이 있으며, 그룹을 통계학적으로 분리할 수 있다는 것을 나타낸다.

[0550]

대상은 프로파일링될 수 있으며 표 1A에 포함된 유전자를 기준으로 한 이의 전사 데이터는 양극성 장애의 진단이 얻어지도록 상기 본원에 기재된 바와 같이 훈련된 분류 알고리즘에 적용될 수 있다.

[0551]

각각의 유전자에 대한 양극성 대상의 전사 프로파일을, 각각의 바이오마커의 풍부함 (즉, 유전자 전사체) 을 기준으로 표 6에 제시한다. 대조군 대상 전사체 값을 비교를 위해 나타낸다.

[0553]

표 6

| 바이오마커<br>(유전자 약어) | 양극성 대상 그룹<br>특징:<br>풍부함 =<br>바이오마커의 평균<br>전사체 값 ( $\pm SD$ ) | 대조군 대상 그룹<br>특징: 풍부함 =<br>바이오마커의<br>평균 전사체 값<br>( $\pm SD$ ) |
|-------------------|--|--|
| ADA               | 4775 $\pm$ 1508  | 4511 $\pm$ 1710  |
| ARRB1             | 292298 $\pm$ 89272   | 297143 $\pm$ 91094   |
| ARRB2             | 111023 $\pm$ 39397   | 114780 $\pm$ 39962   |
| CD8a              | 11668 $\pm$ 5573   | 14693 $\pm$ 8416   |
| CD8b              | 7998 $\pm$ 3841  | 8687 $\pm$ 3880  |
| CREB1             | 62347 $\pm$ 18282  | 63725 $\pm$ 16022  |
| CREB2             | 79456 $\pm$ 16778  | 77059 $\pm$ 15755  |
| DPP4              | 7618 $\pm$ 3077  | 7169 $\pm$ 2890  |
| ERK1              | 34901 $\pm$ 15116  | 39016 $\pm$ 12900  |
| ERK2              | 57832 $\pm$ 21427  | 54137 $\pm$ 18660  |
| Gi2               | 192417 $\pm$ 98987   | 226358 $\pm$ 87609   |
| Gs                | 304202 $\pm$ 171505  | 303930 $\pm$ 139837  |
| GR                | 124054 $\pm$ 42231   | 80610 $\pm$ 26544  |
| IL1b              | 21577 $\pm$ 13468  | 21006 $\pm$ 9313   |
| IL6               | 173 $\pm$ 78   | 182 $\pm$ 221  |
| IL8               | 24568 $\pm$ 19226  | 28024 $\pm$ 19993  |
| INDO              | 5428 $\pm$ 3847  | 5596 $\pm$ 4418  |
| MAPK14            | 66946 $\pm$ 25751  | 51632 $\pm$ 20341  |
| MAPK8             | 12584 $\pm$ 3060   | 12162 $\pm$ 3500   |
| MKP1              | 501068 $\pm$ 251853  | 499308 $\pm$ 220665  |
| MR                | 3409 $\pm$ 1094  | 2830 $\pm$ 887   |
| ODC1              | 67672 $\pm$ 50925  | 58670 $\pm$ 40801  |
| P2X7              | 1322 $\pm$ 418   | 1542 $\pm$ 563   |
| PBR               | 64761 $\pm$ 29660  | 64439 $\pm$ 29328  |
| PREP              | 6806 $\pm$ 1677  | 7072 $\pm$ 2102  |
| RGS2              | 499864 $\pm$ 264854  | 477280 $\pm$ 165907  |
| S100A10           | 42063 $\pm$ 12765  | 35819 $\pm$ 10568  |
| SERT              | 1435 $\pm$ 710   | 1711 $\pm$ 1317  |
| VMAT2             | 2736 $\pm$ 1050  | 2792 $\pm$ 1344  |

(SD = 표준 편차)

[0554]

경계 인격 장애 환자에서의 전사 프로파일의 확인

[0555]

경계 인격 장애 환자에서의 전사 프로파일의 변화를 측정하기 위해, 21 명의 경계 인격 장애 환자로부터의 혈액을 수득하고 표 1A에서 선택된 유전자에 대해 유전자 발현을 측정하였다. 단변량 방법에 의해 유전자 발현 데이터를 통계학적으로 평가하였다. 환자 전사 데이터를 196 명의 대조군의 것과 비교하고, 개별적 유전자 데이터에 대한 대표적 산점도를 도 6A-6C에 나타낸다.

[0556]

RF (선별) 및 SVM (훈련)을 사용하는 분류로, 97%의 높은 정확성이 야기되었다 (PPV = 87%; NPV = 98%). 유전자 선별 및 훈련을 모두 수행하는 SLR 알고리즘의 분류로, 98%의 높은 정확성이 야기되었다 (PPV = 90%; NPV = 100%).

[0557]

알고리즘 모두는, 각각의 방법의 통계적 파라미터를 기준으로 분류를 위해 가장 중요한 유전자로서, 랜덤 포레스트 분류로 총 5 개의 유전자가 선택되고 SLR로 총 4 개의 유전자가 선별되어, 전체 데이터 세트를 기준으로 선택된 유전자에 있어서 양호한 일치를 나타내었다. Gi2, GR, MAPK14, 및 MR을 포함하여 4 개의 유전자가 두 방법 모두에 의해 선별되었다.

[0558]

환자/대조군 배정의 랜덤화 후, 두 분류 알고리즘 (RF/SVM 및 SLR)에서 실제 데이터로 수득된 것들과 통계학적

으로 상이한 정확성 값이 생성되었는데, 이는 상기 열거한 값이 더 나은 가능성이 있으며, 그룹을 통계학적으로 분리할 수 있다는 것을 나타낸다.

[0560] 대상은 프로파일링될 수 있으며 표 1A에 포함된 유전자를 기준으로 한 이의 전사 데이터는 경계 인격 장애의 진단이 얻어지도록 상기 본원에 기재된 바와 같이 훈련된 분류 알고리즘에 적용될 수 있다.

[0561] PTSD 환자에서의 전사 프로파일의 확인

[0562] 전사 프로파일을 급성 PTSD 환자, 완화된 PTSD 환자, 및 PTSD 를 발달시키지 않은 트라우마 사건에 적용되었던 개인 그룹에서 평가하였다. 상기 그룹의 조합된 평가는 급성 PTSD 와 관련된 발현 가능성을 확인할 뿐 아니라, 질환으로부터의 회복 또는 질환에 대한 내성과 연관될 수 있는 차이를 규정하기 위한 기회를 제공한다. 단변량 방법에 의해 유전자 발현 데이터를 통계학적으로 평가하였다. 66 명의 급성 PTSD 환자로부터의 환자 전사 데이터를 196 명의 대조군의 것과 비교하고, 개별적 유전자 데이터에 대한 대표적 산점도를 도 7A-7C에 나타낸다.

[0563] RF (선별) 및 SVM (훈련) 을 사용하는 대조군 대상과 비교한 급성 PTSD 환자의 분류로, 77% 의 정확성이 야기되었다 (PPV = 64%; NPV = 82%). 유전자 선별 및 훈련을 모두 수행하는 SLR 알고리즘으로의 분류로, 84% 의 정확성이 야기되었다 (PPV = 77%; NPV = 87%). SLR 알고리즘은 상기 세트의 시험 데이터를 사용하는 SVM 알고리즘을 능가한다. 각각의 분류 알고리즘을 데이터 세트의 랜덤화 (치환된) 데이터 세트 버전과 비교하였고, SLR 은 치환된 데이터 세트를 사용하여 73% 의 정확성 값을 산출하였다 (PPV = 39%; NPV = 75%). 통계학적 분석은 실제 대 랜덤화 데이터로 수득된 SLR 정확성 값이 상이하다는 것을 나타내며, 그룹을 분리할 수 있다는 것을 나타낸다.

[0564] 치환된 데이터 세트를 사용하여, SVM 은 73% 의 정확성 값을 산출하였고 (PPV = 10%; NPV = 75%), 이것은 치환된 (랜덤화) 데이터에 대해 하향하는 경향을 나타낸다. SVM 알고리즘에서의 실제 데이터를 사용하는 PPV (질환이 있는 환자를 긍정적으로 예측하는 능력) 는 치환된 데이터로의 10% 정밀도와 비교하여 60% 양호하여, 실제 데이터를 사용하여 훈련된 알고리즘이 랜덤 예측을 능가한다는 것을 나타낸다.

[0565] SLR 은 급성 PTSD 환자 대 대조군의 전체 데이터 세트에 근거한 분류를 위해 가장 중요한 유전자로서 총 10 개의 유전자: ARRB1, ARRB2, CD8b, ERK2, IDO, IL-6, MR, ODC1, PREP 및 RGS2 를 선별하였다.

[0566] 대상은 프로파일링될 수 있으며 표 1A에 포함된 유전자를 기준으로 한 이의 전사 데이터는 급성 PTSD 의 진단이 얻어지도록 상기 본원에 기재된 바와 같이 훈련된 분류 알고리즘에 적용될 수 있다.

[0567] RF (선별) 및 SVM (훈련) 을 사용하는 대조군 대상과 비교한 완화된 PTSD 환자의 분류로, 81% 의 정확성이 야기되었다 (PPV = 59%; NPV = 85%). 유전자 선별 및 훈련을 모두 수행하는 SLR 알고리즘의 분류로, 80% 의 정확성이 야기되었다 (PPV = 33%; NPV = 86%). 그러나, 분류 알고리즘이 상기 데이터 세트의 랜덤화 버전에 대해 실행되는 경우, SVM 및 SLR 은 각각 82% 및 81% 의 정확성 값을 산출하였다. 상기 값은 실제 데이터로 수득된 것들과 통계학적으로 상이하지 않으며, 이것은 알고리즘이 상기 그룹을 신뢰성있게 분리할 수 없다는 것을 나타낸다. 분리의 결핍 때문에, 유전자 목록은 상기 비교를 위해 보고되지 않는다. 임상적 관점에서, 대조군과 완화된 환자를 구별하기 위한 알고리즘의 무능력은 상기 그룹 사이에 생물학적 차이의 결핍으로 인한 것으로 예상된다. 완화된 환자는 더이상 질병의 증상을 나타내지 않으므로, 그들의 유전자 발현 수준이 정상 수준으로 되돌아가서, 알고리즘이 그룹을 효과적으로 분리하는 것을 방해한다고 추정하는 것이 합리적이다.

[0568] RF (선별) 및 SVM (훈련) 을 사용하는 대조군 대상과 비교한 PTSD 를 전개하지 않으나 트라우마가 있는 대상의 분류로, 74% 의 정확성이 야기되었다 (PPV = 61%; NPV = 79%). 유전자 선별 및 훈련을 모두 수행하는 SLR 알고리즘의 분류로, 73% 의 정확성이 야기되었다 (PPV = 59%; NPV = 80%). 다변량 분석이 랜덤화 데이터 세트에 대해 수행되는 경우, RF/SVM 및 SLR 분류 알고리즘 모두는 실제 데이터로 수득된 것과 통계학적으로 상이한 정확성 값을 산출하고, 이것은 상기 보고된 바와 같은 값이 더 나은 가능성이 있으며, 그룹을 분리할 수 있다는 것을 나타낸다.

[0569] 트라우마 환자 및 대조군으로부터의 전체 데이터 세트를 사용하고 각각의 방법의 통계적 파라미터를 기준으로 분류를 위해 가장 중요한 유전자로서, 랜덤 포레스트 분류로 총 14 개의 유전자가 선택되고 SLR 로 총 13 개의 유전자가 선별되었다. ARRB2, CREB1, ERK2, Gs, IL-6, MKP1, 및 RGS2 를 포함하여 7 개의 유전자가 두 방법 모두에 의해 선별되었다.

[0570] 상기 개인이 PTSD로 진단되지 않는다 하더라도, 알고리즘은 본원에 제시된 다른 비교 중 일부에 대해 PPV, 및 NPV 값보다는 비록 정확성이 떨어질지라도 여전히 이를 대조군과 구별할 수 있다. 흥미롭게도, 급성 PTSD 환자로부터의 SLR 유전자 목록 상의 유전자 중 6 개는 PTSD가 없는 트라우마 환자에 대한 상응하는 목록 상의 유전자와 매치한다 (ARRB2, CD8b, ERK2, MR, IL-6, 및 RGS2). 트라우마 환자에게 아직 질병이 발병하지 않았더라도, 이들이 환자가 가지고 있는 일부 유전자 발현 프로파일을 공유한다면, 이들에게 위협이 있을 수 있다는 것을 나타낸다.

[0571] 대상은 프로파일링될 수 있으며 표 1A에 포함된 유전자를 기준으로 한 이의 전사 데이터는 PTSD가 없는 트라우마의 진단이 얻어지도록 상기 본원에 기재된 바와 같이 훈련된 분류 알고리즘에 적용될 수 있다.

## 7. 참조 언급

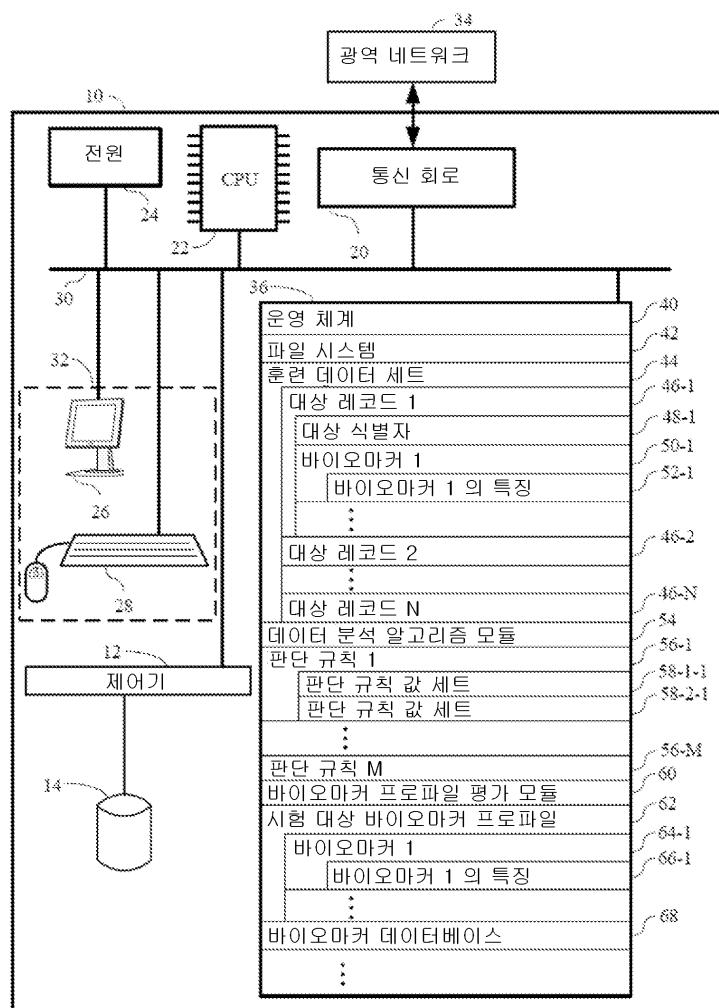
[0573] 본원에 언급된 모든 참조는, 각각의 개별 공개 또는 특허 또는 특허 출원이 모든 목적을 위해 본원에 전체가 참조로서 인용되는 것으로 구체적이고 개별적으로 표시된다면 동일한 범위로 모든 목적을 위해 그 전체가 참조로서 본원에 인용된다.

## 8. 변형

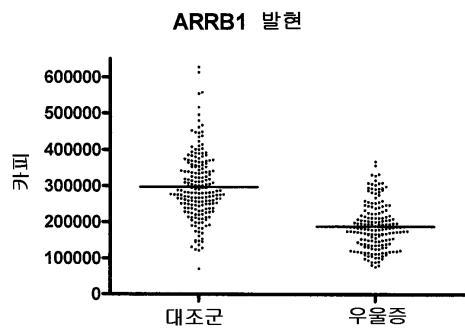
[0575] 당업자에게 명백할 것인 바와 같이, 본 발명의 많은 변형 및 변이가 본 발명의 취지 및 범주로부터 벗어남 없이 이루어질 수 있다. 본원에 기재된 구체적인 구현에는 오직 예로서 제공되며, 본 발명은 이러한 청구항 표제와 동등한 전체 범주와 함께 특허청구범위에 관해서만 제한되는 것이다.

### 도면

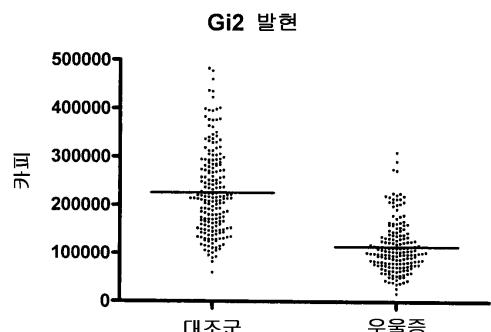
#### 도면1



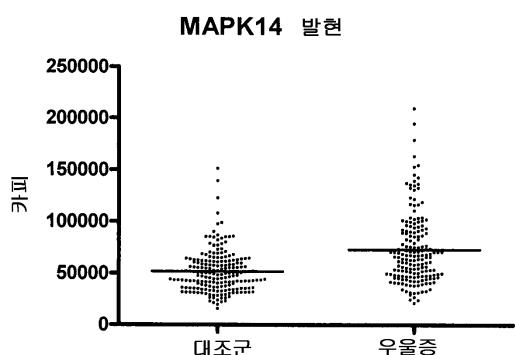
도면2a



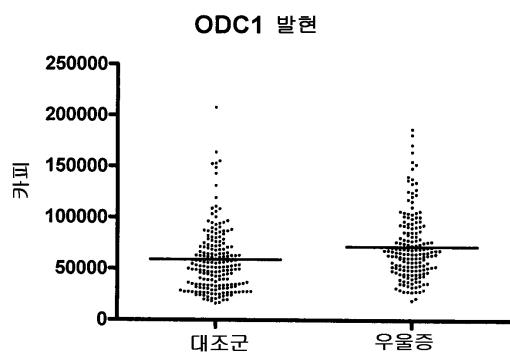
도면2b



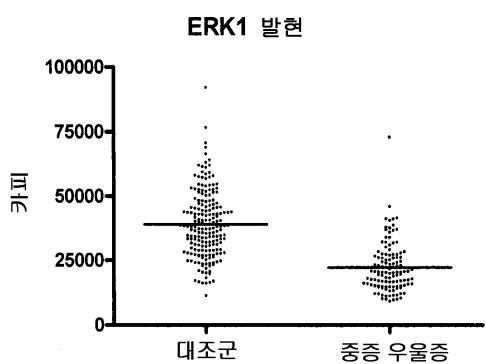
도면3a



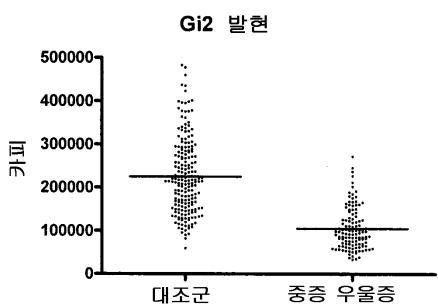
도면3b



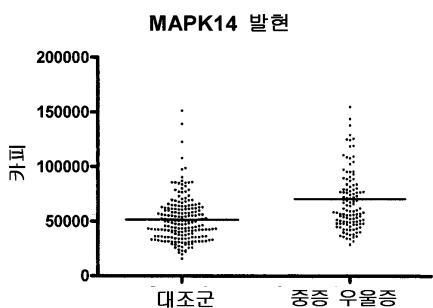
도면4a



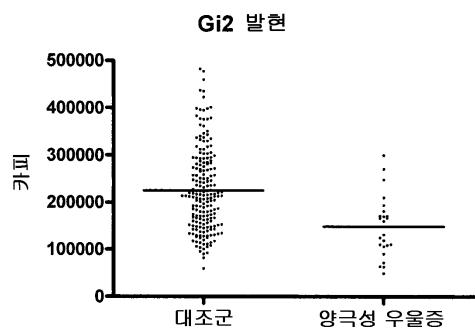
도면4b



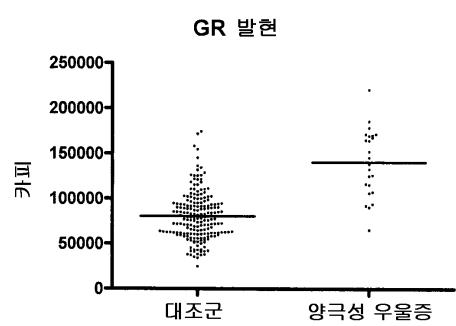
도면4c



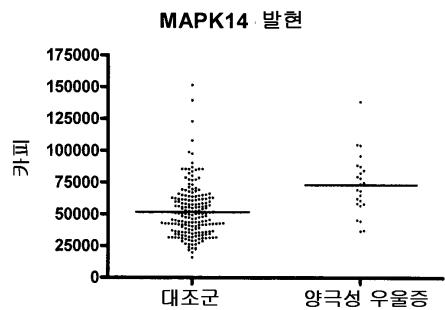
도면5a



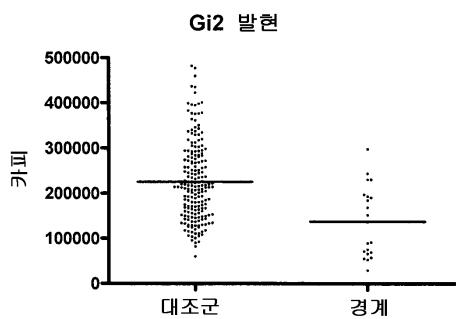
도면5b



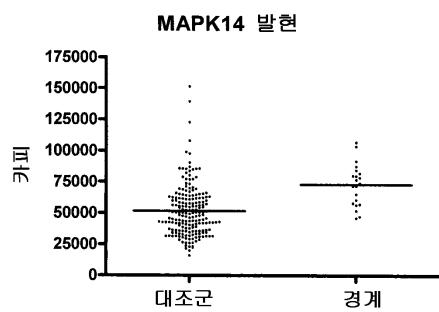
도면5c



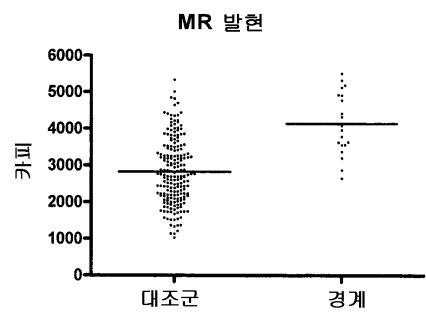
도면6a



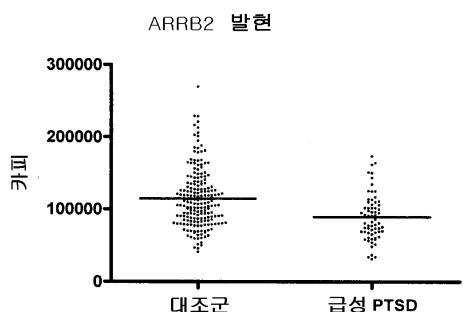
도면6b



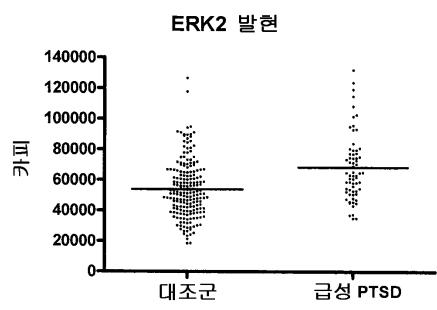
도면6c



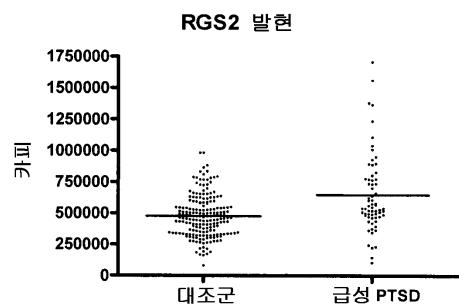
도면7a



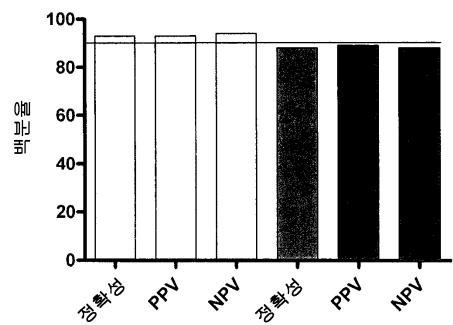
도면7b



도면7c



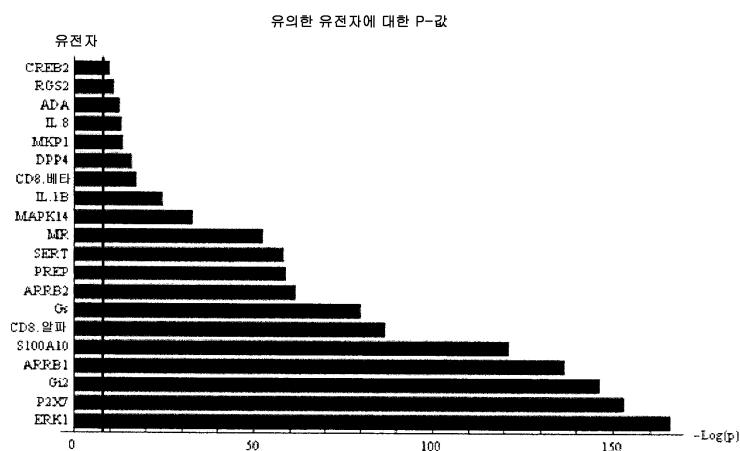
도면8a

**SLR****SVM**

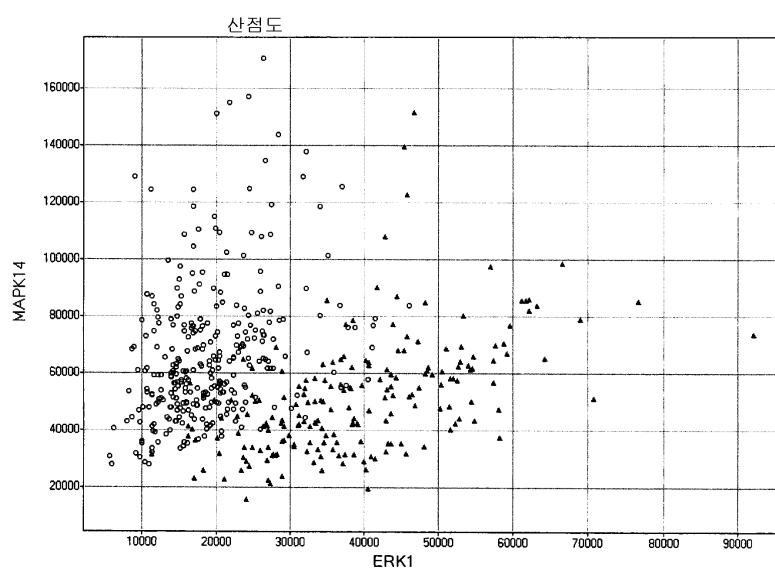
도면8b

| SLR     | RF     |
|---------|--------|
| ARRB1   | ARRB1  |
| ARRB2   | ARRB2  |
| CD8a    | CD8a   |
| CREB1   | CREB1  |
| CREB2   | CREB2  |
| ERK2    | ERK2   |
| Gi2     | Gi2    |
| MAPK14  | MAPK14 |
| ODC1    | ODC1   |
| P2X7    | P2X7   |
| PBR     | PBR    |
| ADA     | ERK1   |
| CD8b    | IL-1b  |
| IDO     | RGS2   |
| MAPK8   |        |
| SERT    |        |
| S100A10 |        |

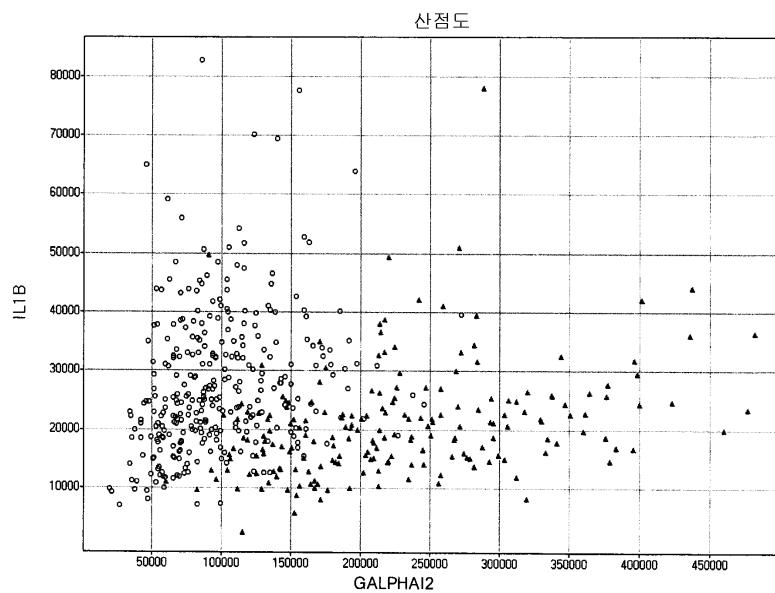
## 도면9



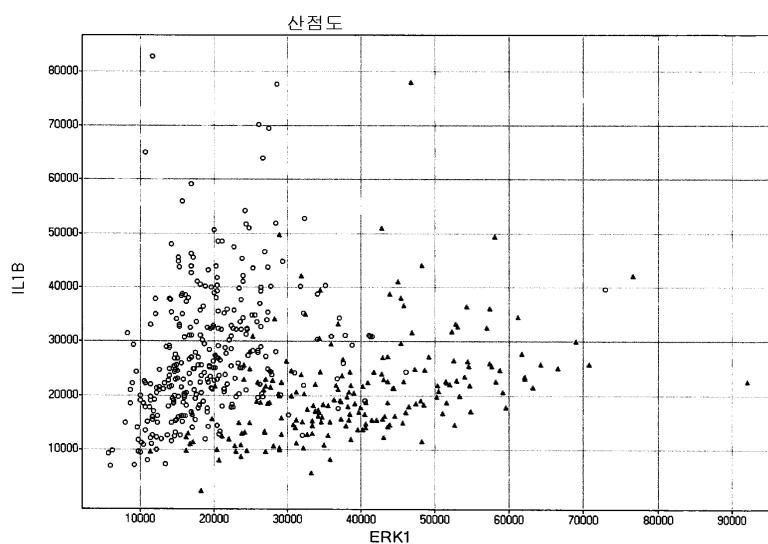
## 도면10



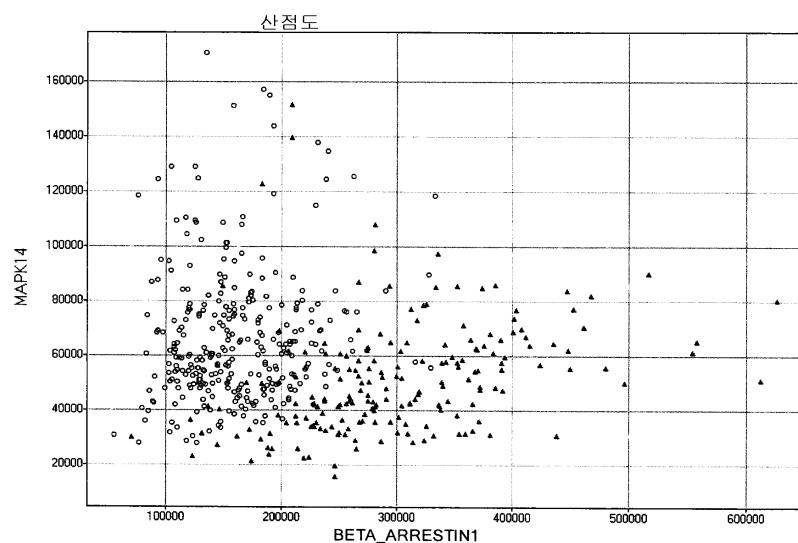
도면11



도면12



## 도면13



## 서 열 목 록

- <110> H. Lundbeck A/S  
 Antonijevic, Irina  
 Tamm, Joseph  
 Artymyshyn, Roman  
 Gerald, Christophe P.G.
- <120> SYSTEM AND METHODS FOR MEASURING BIOMARKER PROFILES
- <130> 71021-WO-PCT
- <150> US 61/092,270
- <151> 2008-08-27
- <160> 132
- <170> PatentIn version 3.4
- <210> 1
- <211> 25
- <212> DNA
- <213> Homo sapiens
- <400> 1
- ggtggtagag ctgtgtttaaga agtac 25
- <210> 2
- <211> 22
- <212>
- > DNA
- <213> Homo sapiens

|                               |              |    |
|-------------------------------|--------------|----|
| <400>                         | 2            |    |
| cttcctggga tggctcata tc       |              | 22 |
| <210>                         | 3            |    |
| <211>                         | 26           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 3            |    |
| cagcagaccg tggttagccat tgacct |              | 26 |
| <210>                         | 4            |    |
| <211>                         | 22           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 4            |    |
| agacacgaac ttggcctcta gc      |              | 22 |
| <210>                         | 5            |    |
| <211>                         | 23           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         |              |    |
| >                             | 5            |    |
| ttgttagaaaa caatgatccc cag    |              | 23 |
| <210>                         | 6            |    |
| <211>                         | 26           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 6            |    |
| ttgagggaaag gtgccaaccg tgagat |              | 26 |
| <210>                         | 7            |    |
| <211>                         | 20           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 7            |    |
| tcttccatgc tccgtcacac         |              | 20 |
| <210>                         | 8            |    |
| <211>                         | 21           |    |

|                               |              |    |
|-------------------------------|--------------|----|
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 8            |    |
| cgaatctcaa agtctacgcc g       |              | 21 |
| <br>                          |              |    |
| <210>                         | 9            |    |
| <211>                         | 25           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 9            |    |
| agccaggccc agaggataca ggaaa   |              | 25 |
| <210>                         | 10           |    |
| <211>                         | 18           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 10           |    |
| ttccgcccag agaacgag           |              | 18 |
| <210>                         | 11           |    |
| <211>                         | 18           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 11           |    |
| aagaccggca cgaagtgg           |              | 18 |
| <210>                         | 12           |    |
| <211>                         | 26           |    |
| <212>                         | DNA          |    |
| <br>                          |              |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 12           |    |
| tcggccctga gcaactccat catgtta |              | 26 |
| <210>                         | 13           |    |
| <211>                         | 22           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 13           |    |

|                                  |    |
|----------------------------------|----|
| tgacagt cac cacgagttcc tg        | 22 |
| <210> 14                         |    |
| <211> 22                         |    |
| <212> DNA                        |    |
| <213> Homo sapiens               |    |
| <400> 14                         |    |
| <br>tctcctgttc cacctttca cc      | 22 |
| <210> 15                         |    |
| <211> 26                         |    |
| <212> DNA                        |    |
| <213> Homo sapiens               |    |
| <400> 15                         |    |
| <br>ctctggatt ccgcaaaagg gactat  | 26 |
| <210> 16                         |    |
| <211> 22                         |    |
| <212> DNA                        |    |
| <213> Homo sapiens               |    |
| <400> 16                         |    |
| <br>ctggctaaca atggtaaccga tg    | 22 |
| <210> 17                         |    |
| <211> 25                         |    |
| <212> DNA                        |    |
| <213> Homo sapiens               |    |
| <400> 17                         |    |
| <br>gtggtctgtg catactgttag aatgg | 25 |
| <210> 18                         |    |
| <211> 24                         |    |
| <212> DNA                        |    |
| <213> Homo sapiens               |    |
| <400> 18                         |    |
| <br>catgaccaat gcagcagcca ctca   | 24 |
| <210> 19                         |    |
| <211> 23                         |    |

|                               |              |    |
|-------------------------------|--------------|----|
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 19           |    |
| cacgtggat gacacttgtc atc      |              | 23 |
| <210>                         | 20           |    |
| <211>                         | 19           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 20           |    |
| ctggagatg gccaaattgg          |              | 19 |
| <210>                         | 21           |    |
| <211>                         | 26           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 21           |    |
| actaataaagc agccccccca gacggt |              | 26 |
| <210>                         | 22           |    |
| <211>                         | 24           |    |
| <212>                         |              |    |
| DNA                           |              |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 22           |    |
| gtgtcattca gtaaagaggc gaag    |              | 24 |
| <210>                         | 23           |    |
| <211>                         | 22           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 23           |    |
| ctcagccctt tatcattcac gc      |              | 22 |
| <210>                         | 24           |    |
| <211>                         | 26           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 24           |    |
| ttccggtcct ggtctgcccc tctata  |              | 26 |

<210> 25  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 25  
tgacggagta tgtggctacg c 21

<210> 26  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 26

ccacagacca gatgtcgatg g 21

<210> 27  
<211> 24  
<212> DNA  
<213> Homo sapiens  
<400> 27

ctggtaccgg gccccagaga tcat 24

<210> 28  
<211> 20  
<212> DNA  
<213> Homo sapiens  
<400> 28

taacgttctg caccgtgacc 20

<210> 29  
<211> 22  
<212> DNA  
<213> Homo sapiens  
<400> 29

caggccaaag tcacagatct tg 22

<210> 30  
<211> 26  
<212> DNA

|                              |              |    |
|------------------------------|--------------|----|
| <213>                        | Homo sapiens |    |
| <400>                        | 30           |    |
| acctgctgct caacaccacc tgttat |              | 26 |
| <210>                        | 31           |    |
| <211>                        | 19           |    |
| <212>                        | DNA          |    |
| <213>                        | Homo sapiens |    |
| <400>                        | 31           |    |
| aggcgtgctc cctgatgac         |              | 19 |
| <210>                        | 32           |    |
| <211>                        | 23           |    |
| <212>                        |              |    |
| DNA                          |              |    |
| <213>                        | Homo sapiens |    |
| <400>                        | 32           |    |
| gctccaggta gttcaggtat tag    |              | 23 |
| <210>                        | 33           |    |
| <211>                        | 21           |    |
| <212>                        | DNA          |    |
| <213>                        | Homo sapiens |    |
| <400>                        | 33           |    |
| aggcctgctt tggccgctca a      |              | 21 |
| <210>                        | 34           |    |
| <211>                        | 21           |    |
| <212>                        | DNA          |    |
| <213>                        | Homo sapiens |    |
| <400>                        | 34           |    |
| gactatgtgc cgagcgatca g      |              | 21 |
| <210>                        | 35           |    |
| <211>                        | 22           |    |
| <212>                        | DNA          |    |
| <213>                        | Homo sapiens |    |
| <400>                        | 35           |    |
| gtccacacctgg aacttggatca     |              | 22 |

|                                 |              |    |
|---------------------------------|--------------|----|
| <210>                           | 36           |    |
| <211>                           | 22           |    |
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 36           |    |
| ctgcttcgct gccgtgtcct ga        |              | 22 |
| <210>                           | 37           |    |
| <211>                           | 22           |    |
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 37           |    |
| tccctggtcg aacagtttt tc         |              | 22 |
| <210>                           | 38           |    |
| <211>                           | 20           |    |
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 38           |    |
| tttgggaggt ggtcctgttg           |              | 20 |
| <210>                           | 39           |    |
| <211>                           | 29           |    |
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 39           |    |
| tgtaaagctct cctccatcca gtcctcaa |              | 29 |
| <210>                           | 40           |    |
| <211>                           | 23           |    |
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 40           |    |
| gatggcccta aacagatgaa gtg       |              | 23 |
| <210>                           | 41           |    |
| <211>                           | 21           |    |
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |

|                              |              |    |
|------------------------------|--------------|----|
| <400>                        | 41           |    |
| cctgaagccc ttgctgtagt g      |              | 21 |
| <210>                        | 42           |    |
| <211>                        | 25           |    |
| <212>                        |              |    |
| DNA                          |              |    |
| <213>                        | Homo sapiens |    |
| <400>                        | 42           |    |
| atggcggcat ccagctacga atctc  |              | 25 |
| <210>                        | 43           |    |
| <211>                        | 22           |    |
| <212>                        | DNA          |    |
| <213>                        | Homo sapiens |    |
| <400>                        | 43           |    |
| agccactcac ctcttcagaa cg     |              | 22 |
| <210>                        | 44           |    |
| <211>                        | 22           |    |
| <212>                        | DNA          |    |
| <213>                        | Homo sapiens |    |
| <400>                        | 44           |    |
| catgtctcct ttctcaggc tg      |              | 22 |
| <210>                        | 45           |    |
| <211>                        | 26           |    |
| <212>                        | DNA          |    |
| <213>                        | Homo sapiens |    |
| <400>                        | 45           |    |
| caaattcggt acatcctcga cggcat |              | 26 |
| <210>                        | 46           |    |
| <211>                        | 21           |    |
| <212>                        | DNA          |    |
| <213>                        | Homo sapiens |    |
| <400>                        | 46           |    |
| ctgcttagcca ggatccacaa g     |              | 21 |
| <210>                        | 47           |    |

|                                |              |    |
|--------------------------------|--------------|----|
| <211>                          | 26           |    |
| <212>                          | DNA          |    |
| <213>                          | Homo sapiens |    |
| <400>                          | 47           |    |
| ctgtgaggta agatggtggc taatac   |              | 26 |
| <210>                          | 48           |    |
| <211>                          | 29           |    |
| <212>                          | DNA          |    |
| <213>                          | Homo sapiens |    |
| <400>                          | 48           |    |
| cttgtccac tgtgccttgg tttctcctt |              | 29 |
| <br>                           |              |    |
| <210>                          | 49           |    |
| <211>                          | 28           |    |
| <212>                          | DNA          |    |
| <213>                          | Homo sapiens |    |
| <400>                          | 49           |    |
| gcttcgagaa agagttgaga agttaaac |              | 28 |
| <210>                          | 50           |    |
| <211>                          | 21           |    |
| <212>                          | DNA          |    |
| <213>                          | Homo sapiens |    |
| <400>                          | 50           |    |
| gaccttgcc ccacacatat g         |              | 21 |
| <210>                          | 51           |    |
| <211>                          | 27           |    |
| <212>                          | DNA          |    |
| <213>                          | Homo sapiens |    |
| <400>                          | 51           |    |
| ctcacagacc acaagtaca gcgcctt   |              | 27 |
| <210>                          | 52           |    |
| <211>                          | 20           |    |
| <212>                          |              |    |
| DNA                            |              |    |
| <213>                          | Homo sapiens |    |

|                               |              |    |
|-------------------------------|--------------|----|
| <400>                         | 52           |    |
| cggcaggagc tgaacaagac         |              | 20 |
| <210>                         | 53           |    |
| <211>                         | 22           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 53           |    |
| agcagcacac acagagccat ag      |              | 22 |
| <210>                         | 54           |    |
| <211>                         | 26           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 54           |    |
| ccgagcgta ccagaacctg tctcca   |              | 26 |
| <210>                         | 55           |    |
| <211>                         | 22           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 55           |    |
| ccaacacccg tacatcaatg tc      |              | 22 |
| <210>                         | 56           |    |
| <211>                         | 26           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 56           |    |
| cactttcta ttgtgtgttc ctttc    |              | 26 |
| <210>                         | 57           |    |
| <211>                         | 28           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 57           |    |
| caccaccaaa gatccctgac aagcagt |              | 28 |
| <210>                         | 58           |    |
| <211>                         | 17           |    |

|                               |              |    |
|-------------------------------|--------------|----|
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 58           |    |
| gccaggcagg cattcc             |              | 17 |
| <br>                          |              |    |
| <210>                         | 59           |    |
| <211>                         | 20           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 59           |    |
| atgttcgcc tctgttac            |              | 20 |
| <210>                         | 60           |    |
| <211>                         | 27           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 60           |    |
| tcagccacca tctgccttgc ttacctt |              | 27 |
| <210>                         | 61           |    |
| <211>                         | 20           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 61           |    |
| agcccgagg aaggacaac           |              | 20 |
| <210>                         | 62           |    |
| <211>                         | 20           |    |
| <212>                         |              |    |
| DNA                           |              |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 62           |    |
| tgtgagcgct cgtgagatttg        |              | 20 |
| <210>                         | 63           |    |
| <211>                         | 27           |    |
| <212>                         | DNA          |    |
| <213>                         | Homo sapiens |    |
| <400>                         | 63           |    |

|                                   |    |
|-----------------------------------|----|
| ctcctgcaaa agaacccctcg gtcaaca    | 27 |
| <210> 64                          |    |
| <211> 21                          |    |
| <212> DNA                         |    |
| <213> Homo sapiens                |    |
| <400> 64                          |    |
| <br>ccatgttagga agcggctgta c      | 21 |
| <210> 65                          |    |
| <211> 20                          |    |
| <212> DNA                         |    |
| <213> Homo sapiens                |    |
| <br><400> 65                      |    |
| <br>tcagccccca tgtcaaaaac         | 20 |
| <210> 66                          |    |
| <211> 27                          |    |
| <212> DNA                         |    |
| <213> Homo sapiens                |    |
| <400> 66                          |    |
| <br>atcctgagac ctgcgtgcag gcaatct | 27 |
| <210> 67                          |    |
| <211> 22                          |    |
| <212> DNA                         |    |
| <213> Homo sapiens                |    |
| <400> 67                          |    |
| <br>gctgtcgctc ccatatttat cc      | 22 |
| <210> 68                          |    |
| <211> 22                          |    |
| <212> DNA                         |    |
| <213> Homo sapiens                |    |
| <400> 68                          |    |
| <br>cacaatggac tcgcacttct tc      | 22 |
| <210> 69                          |    |
| <211> 28                          |    |

|                                 |              |    |
|---------------------------------|--------------|----|
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 69           |    |
| ctgtcagccc tgitgttgtca acgaatac |              | 28 |
| <210>                           | 70           |    |
| <211>                           | 20           |    |
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 70           |    |
| ctggcttgaa aagagctggg           |              | 20 |
| <210>                           | 71           |    |
| <211>                           | 20           |    |
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 71           |    |
| cagcaggaga tccacccaagg          |              | 20 |
| <210>                           | 72           |    |
| <211>                           | 23           |    |
| <212>                           |              |    |
| DNA                             |              |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 72           |    |
| ccccatcttc ttgggtgcc gac        |              | 23 |
| <210>                           | 73           |    |
| <211>                           | 24           |    |
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 73           |    |
| ggaatatgta ctacgtgacc aatg      |              | 24 |
| <210>                           | 74           |    |
| <211>                           | 24           |    |
| <212>                           | DNA          |    |
| <213>                           | Homo sapiens |    |
| <400>                           | 74           |    |
| ggatccctgaa agtcaatgtt gatc     |              | 24 |

<210> 75  
<211> 26  
<212> DNA  
<213> Homo sapiens

<400> 75  
cattcaagac gaatgccag tctccc

26

<210> 76  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 76

gatttgaaga cccgttttag c

21

<210> 77  
<211> 22  
<212> DNA  
<213> Homo sapiens  
<400> 77

caggagaagg ctgtatgaaa gc

22

<210> 78  
<211> 22  
<212> DNA  
<213> Homo sapiens  
<400> 78

ctggaaagcc caaaaccggc aa

22

<210> 79  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 79

aggagttccc tggattttg g

21

<210> 80  
<211> 22  
<212> DNA

|                                |              |    |
|--------------------------------|--------------|----|
| <213>                          | Homo sapiens |    |
| <400>                          | 80           |    |
| gccccacttgc ccatctctac ac      |              | 22 |
| <210>                          | 81           |    |
| <211>                          | 28           |    |
| <212>                          | DNA          |    |
| <213>                          | Homo sapiens |    |
| <400>                          | 81           |    |
| caaaaagacc ctctggctgt ggacaaaa |              | 28 |
| <210>                          | 82           |    |
| <211>                          | 23           |    |
| <212>                          |              |    |
| DNA                            |              |    |
| <213>                          | Homo sapiens |    |
| <400>                          | 82           |    |
| catggctgag atgaggaatg aag      |              | 23 |
| <210>                          | 83           |    |
| <211>                          | 19           |    |
| <212>                          | DNA          |    |
| <213>                          | Homo sapiens |    |
| <400>                          | 83           |    |
| gctggcatgt tggctatcg           |              | 19 |
| <210>                          | 84           |    |
| <211>                          | 25           |    |
| <212>                          | DNA          |    |
| <213>                          | Homo sapiens |    |
| <400>                          | 84           |    |
| acgcagggcc cagccctc ttcatt     |              | 25 |
| <210>                          | 85           |    |
| <211>                          | 23           |    |
| <212>                          | DNA          |    |
| <213>                          | Homo sapiens |    |
| <400>                          | 85           |    |
| tggattcgta aatgatgcct atc      |              | 23 |

|   |              |     |
|---|--------------|-----|
| <210>   | 86           |     |
| <211>   | 18           |     |
| <212>   | DNA          |     |
| <213>   | Homo sapiens |     |
| <400>   | 86           |     |
| atgccacatc cgcaatgg   |              | 18  |
| <210>   | 87           |     |
| <211>   | 24           |     |
| <212>   | DNA          |     |
| <213>   | Homo sapiens |     |
| <400>   | 87           |     |
| agacctgcgg cacgtgtccg tcta  |              | 24  |
| <210>   | 88           |     |
| <211>   | 1566         |     |
| <212>   | DNA          |     |
| <213>   | Homo sapiens |     |
| <400>   | 88           |     |
| ggccgttaa gaagagcgtg gccggccgca gcccaccgtg gccccaggaa aagccgagcg  |              | 60  |
| gccacggagc cggcagagac ccacccgagcg gcggcggagg gagcagcgcc gggcgcacg |              | 120 |
| agggcaccat gcccagacg cccgcctcg acaagccaa atgagaactg catgtccacc    |              | 180 |
| tagacggatc catcaaggct gaaaccatct tatactatgg caggaggaga gggatcgccc |              | 240 |
| tcccagctaa cacagcagag gggctgctga acgtcattgg catggacaag ccgctcaccc |              | 300 |
| ttccagactt cctggccaag ttgactact acatgcctgc tatcgccggc tgccggagg   |              | 360 |
| ctatcaaag gatgcctat gagttttag agatgaaggc caaagagggc gtggtgtatg    |              | 420 |
| tggaggtgcg gtacagtccg cacctgctgg ccaactccaa agtggagcca atcccctgga |              | 480 |
| accaggctga aggggacctc accccagacg aggtggtggc cctagtgcc cagggcctgc  |              | 540 |
| aggaggggaa gcgagacttc ggggtcaagg cccggccat cctgtgtgc atgcgccacc   |              | 600 |
| agcccaactg gtcccccaag gtgggggac tgtgtaaagaa gtaccagcag cagaccgtgg |              | 660 |
| tagccattga cctggctgga gatgagacca tcccaaggaa cagcctttg cctggacatg  |              | 720 |
| tccaggccta ccaggaggct gtggggggc gcattcacgg tactgtccac gccggggagg  |              | 780 |
| tgggctggc cgaagtagta aaagaggctg tggacatact caagacagag cggctggac   |              | 840 |
| acggctacca caccctggaa gaccaggccc ttataacag gctgcggcag gaaaacatgc  |              | 900 |

|  |      |
|--|------|
| acttcgagat ctgccctgg tccagctacc tcactgggc    | 960  |
| atgcagtcat tcggctcaaa aatgaccagg ctaactactc  | 1020 |
| gtccacccctg gacactgatt accagatgac caaacggac  | 1080 |
| atgaagagga gttaaaaagg ctgaacatca atgcggcaa   | 1140 |
| atgaaaagag ggagcttctc gacctgctc ataaagccta   | 1200 |
| ctgcaggcca gaacctctga agacgccact cctccaagcc  | 1260 |
| caactctgtg gggctgagca acattttac atttattcct   | 1320 |
| aatagtcagt tactgatgct cctgaaccct atgtgtccat  | 1380 |
| ttctgcacac acgtataacct cggtatggcc            | 1440 |
| cggtatggcc ggttcacttc tctgattatg tgccctggcc  | 1500 |
| agggaccagc gcccttgcac                        |      |
| atggcatgg ttgaatctga aaccctcctt ctgtggcaac   | 1560 |
| ttgtactgaa aatctggc                          |      |
| tcaataaaga ageccatggc tggtgccatg caaaaaaaaaa | 1566 |
| aaaaaaaaaaaaaaaaaaaa                         |      |
| <210> 89                                     |      |
| <211> 1254                                   |      |
| <212> DNA                                    |      |
| <213> Homo sapiens                           |      |
| <400> 89                                     |      |
| atggcgaca aaggacgacg ggtgttcaag aaggccagtc   | 60   |
| caaatggaaa gtcaccgtc                         |      |
| tacctggaa agcggactt tgtggaccac atcgacctcg    | 120  |
| tggaccctgt ggatgggttg                        |      |
| gtcctggtgg atcctgagta tctcaaagag cggagactt   | 180  |
| atgtgacgct gacctgcgc                         |      |
| ttccgctatg gcccggagga cctggatgtc ctggcctga   | 240  |
| ccttcgcac ggacctgttt                         |      |
| gtggccaacg tacagtcgtt cccaccggcc cccgaggaca  | 300  |
| agaagccccct gacgcggctg                       |      |
| caggaacgcc tcatcaagaa gctggcgag cacgcttacc   | 360  |
| ctttcacctt tgagatccct                        |      |
| ccaaacccctc catgttctgt gacactgcag ccggggcccg | 420  |
| aagacacggg gaaggcttgc                        |      |
| ggtgtggact atgaagccaa agcctctgc gcggagaatt   | 480  |
| tggaggagaa gatccacaag                        |      |
| cggaattctg tgggtctggt catccggaag gttcagtagt  | 540  |
| ccccagagag gcctggcccc                        |      |
| cagcccacag ccgagaccac caggcagtcc ctcatgtcg   | 600  |
| acaagccctt gcacctagaa                        |      |
| gcctctctgg ataaggagat ctattaccat ggagaaccca  | 660  |
| tcaagatct cagtcgtcaa cgtccacgtc              |      |
| accaacaaca ccaacaagac ggtggagaag atcaagatct  | 720  |
| cagtgcgcac gtatgcagac                        |      |
| atctgcctt tcaacacagc tcaagtacaag tgccctgtt   | 780  |
| ccatggaaaga ggctgtatgac                      |      |
| actgtggcac ccagctcgac gttctgcaag gtctacacac  | 840  |
| tgacccctt cctagccat                          |      |

|   |      |
|---|------|
| aaccgagaga agcggggcct cgccttggac ggaaagctca agcacgaaga cacgaacttg   | 900  |
| <br>  |      |
| gcctcttagca ccctgtttag ggaagggtgcc aaccgtgaga tcctgggat cattgtttcc  | 960  |
| tacaaagtga aagtgaagct ggtgggtct cggggcgcc tgggggaga tcttgcattcc     | 1020 |
| agcagctgg ccgtggaact gcccttcacc ctaatgcacc ccaagccaa agaggAACCC     | 1080 |
| ccgcattcgg aagttccaga gaacgagacg ccagtagata ccaatctcat agaacttgac   | 1140 |
| acaatgtacgacattgt attttagggat ttgctgcc agagactgaa aggcatggag        | 1200 |
| gatgacaagg aggaagagga ggatggtacc ggctctccgc ggctcaacga caga         | 1254 |
| <br>  |      |
| <210> 90  |      |
| <211> 1770  |      |
| <212> DNA   |      |
| <br>  |      |
| <213> Homo sapiens  |      |
| <400> 90  |      |
| <br>  |      |
| tggcagcggg cgaggaggct gcgagcggc cgcaaccga gggggcgcc ggcgcgcga       | 60   |
| ccatggggga gaaacccggg accagggtct tcaagaagtc gagccctaac tgcaagtc     | 120  |
| ccgtgtactt gggcaagcgg gacttcgtat atcacctgga caaatggac cctgttagatg   | 180  |
| gcgtgggtct tggggacctt gactacctga aggaccgaa agtgggtgt accctcacct     | 240  |
| gccccttcgg ctatggccgt gaagacctgg atgtgctgg ctgtccttc cgcaaagacc     | 300  |
| tgttcatcgc cacctaccag gccttcccc cggtgccaa cccacccgg ccccccaccc      | 360  |
| gcctgcagga cggctgctg aggaagctgg gccagcatgc ccaccccttc ttcttcacca    | 420  |
| <br>  |      |
| taccccgaaa tctccatgc tccgtcacac tgcagccagg cccagaggat acaggaaagg    | 480  |
| cctgcggcggt agacttttag attcgagct tctgtgctaa atcactagaa gagaaaagcc   | 540  |
| acaaaaggaa ctctgtcgg ctggtgatcc gaaaggtgca gttggcccg gagaaacccg     | 600  |
| gccccccagcc ttccggccaa accacacgccc acttccatgt gtctgaccgg tccctgcacc | 660  |
| tccgggttcc ctggacaag gagctgtact accatgggaa gcccctcaat gtaatgtcc     | 720  |
| acgtcaccaa caactccacc aagaccgtca agaagatcaa agtctctgt agacagtacg    | 780  |
| ccgacatctg cctttcagc accggccagt acaagtgtcc tggctcaat ctcgaacaag     | 840  |
| <br>  |      |
| atgaccaggt atctcccagc tccacattct gtaagggtta caccataacc ccactgtca    | 900  |
| gtgacaaccg ggaaagcgg ggtctcgcc tggatggaa actcaagcac gaggaccca       | 960  |
| acctggcttc cagcaccatc gtgaaggagg gtgccaacaa ggaggtgctg ggaatctgg    | 1020 |
| tgtcctacag ggtcaaggtg aagctggatgg tggctcgagg cggggatgtc tctgtggagc  | 1080 |
| tgcctttgt tcttatgcac ccaagcccc acgaccacat cccctcccc agaccccgat      | 1140 |

|   |      |
|---|------|
| cagccgctcc ggagacagat gtccctgtgg acaccaacct cattgaattt gataccaact   | 1200 |
| atgccacaga ttagtgcatt gtgttgagg actttgcccg gtttcggctg aaggggatga    | 1260 |
| <br>  |      |
| aggatgacga ctatgtat caactctgtc aggaagcggg gtggaaagaa gggagggat      | 1320 |
| gggggtggga gaggtgaggg caggattaag atccccactg tcaatgggg attgtcccag    | 1380 |
| ccccctttcc ctccccctca cctgaaagct tcttcaacca atccctcac actctctccc    | 1440 |
| ccatcccccc aagatacaca ctggaccctc tcttgcgaa tgtggcatt aatttttga      | 1500 |
| ctgcagctct gcttctccag ccccgccgtg ggtggcaagc tgtgttcata cctaaattt    | 1560 |
| ctggaaaggaa acagtgaaaaa gaggagtgac aggaggaaa gggggagaca aaactctac   | 1620 |
| tctcaacctc acaccaacac ctccccattat cactctct gccccattc cttcaagagg     | 1680 |
| <br>  |      |
| agacccttg gggacaaggc cgtttcttg tttctgagca taaagaagaa aataaatctt     | 1740 |
| ttactaagca tgaaaaaaaaaaaaaaa  | 1770 |
| <210> 91  |      |
| <211> 1975  |      |
| <212> DNA   |      |
| <213> Homo sapiens  |      |
| <400> 91  |      |
| <br>  |      |
| actttccccct ctcggcgccc caccggctcc cgccgcgcctc ccctcgcgcc cgagcttcga | 60   |
| gccaaggcgcg gtcctggga gcgcgtcatg gccttaccag tgaccgcctt gctctgcgg    | 120  |
| ctggccttgc tgctccacgc cgccaggccg agccagttcc gggtgtcgcc gctggatcg    | 180  |
| acctggaaacc tgggcgagac agtggagctg aagtgcagg tgctgctgtc caacccgacg   | 240  |
| <br>  |      |
| tcgggctgct cgtggctt ccagccgcgc ggcgcgcgcg ccagtcccac cttctcccta     | 300  |
| tacctctccc aaaacaagcc caaggcggcc gaggggctgg acacccagcg gttctcgccc   | 360  |
| aagaggttgg gggacacccctt cgtcctcacc ctgagcgact tccggcggaga gaacgaggc | 420  |
| tactatttct gctcgccctt gagcaactcc atcatgtact tcagccactt cgtccggc     | 480  |
| ttcctgccag cgaagccac cacgacgcca ggcgcgcgcac caccaacacc ggcgcacc     | 540  |
| atcgcgtcgc agccctgtc cctgcgcaca gagggcgtgcc ggcgcgcgc ggggggcga     | 600  |
| gtgcacacga gggggctgga cttcgctgt gatatctaca tctggcgcgc cctggccggg    | 660  |
| <br>  |      |
| acttgtgggg tccttctct gtcactggtt atcaccctt actgcaacca caggaaccga     | 720  |
| agacgtgttt gcaaattgtcc cggcctgtg gtcaaattgg gagacaagcc cagccttcg    | 780  |
| gcgcgatagc tctaaccctgt tgcaacagcc actacattac ttcaactga gatccttct    | 840  |
| tttggggag caagtccttc ctttcattt ttccagctc ttccctgt gtattcattt        | 900  |

|  |      |
|--|------|
| tcatgattat tattttagtg gggcggggt gggaaagatt acttttctt tatgtgttg     | 960  |
| acggaaaca aaacttagta aaatctacag tacaccacaa gggtcacaat actgttgtgc   | 1020 |
| gcacatcgcg gttaggcgtg gaaaggca gccagagct acccgagat ttctcagaat      | 1080 |
| <br>   |      |
| catgctgaga gagctggagg cacccatgcc atctcaacct ctcccccc cgtttacaa     | 1140 |
| agggggaggc taaagccag agacagctg atcaaaggca cacagcaagt cagggttgg     | 1200 |
| gcagtagctg gagggaccc ttctcccagc tcagggctct ttctccaca ccattcagg     | 1260 |
| cttccttcc gaggccctg tctcagggtg aggtgttga gtctccaacg gcaagggAAC     | 1320 |
| aagtacttct tgatacctgg gatactgtgc ccagagcc tcaggaggtaa tgaattaaag   | 1380 |
| aagagaactg ccittggcag agttctataa tgtaaacaat atcagacttt ttttttttat  | 1440 |
| aatcaagcct aaaattgtat agacctaAAA taaaatgaag tggtgagctt aaccctggaa  | 1500 |
| <br>   |      |
| aatgaatccc tctatctcta aagaaaaatct ctgtgaaacc cctacgttga ggcggatttg | 1560 |
| ctctcccaagc cttgcattt cagaggggcc catgaaagag gacaggttac cccttacaa   | 1620 |
| atagaatttg agcatcagtg aggttaaact aaggccctct tgaatctctg aatttgagat  | 1680 |
| acaaacatgt tcctggatc actgatgact ttttatactt tgtaaagaca attgttgag    | 1740 |
| agccctcac acagccctgg cctccgctca actagcagat acaggatgt ggcagacctg    | 1800 |
| actctttaa ggaggctgag agcccaaact gctgtccaa acatgcactt cttgtttaa     | 1860 |
| ggtatggtac aagcaatgcc tgcccatgg agagaaaaaa cttaaatgaga taagaaata   | 1920 |
| <br>   |      |
| agaaccactc ataattcttc accttaggaa taatctcctg ttaatatggt gtaca       | 1975 |
| <210> 92   |      |
| <211> 1411   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 92   |      |
| <br>   |      |
| gcgactgtct ccggcgagcc cccggggcca ggtgtccgg ggcgcacg atgcggccgc     | 60   |
| ggctgtggct cctctggcc ggcgcacgtga cagtttcca tggcaactca gtcctccagc   | 120  |
| agacccctgc atacataaag gtgaaacca acaagatggt gatgtgtcc tgcgaggcta    | 180  |
| aaatctccct cagtaacatg cgcacatctact ggctgagaca ggcgcaggca ccgagcagt | 240  |
| acagtcacca cgagttcctg gcccctggg attccgaaa agggactatc cacggtgaag    | 300  |
| <br>   |      |
| agggtgaaaca ggagaagata gctgtgtttc gggatgcaag ccgggttatt ctcaatctca | 360  |
| caagcgtgaa gcccggaaagac agtggcatct acttctgcat gatgtcggt agcccccggc | 420  |
| tgaccttcgg gaaggaaact cagctgagtg tgggtgattt cttccacc actgcccagc    | 480  |

|             |              |            |            |             |            |      |
|-------------|--------------|------------|------------|-------------|------------|------|
| ccaccaagaa  | gtccaccctc   | aagaagagag | tgtgccggtt | acccaggcca  | gagacccaga | 540  |
| agggcccact  | tttagcccc    | atcacccctg | gcctgcttgt | ggctggcg    | tcgttctgc  | 600  |
| tggttccct   | gggagtggcc   | atccacctgt | gctgccggcg | gaggagagcc  | cggcttcgtt | 660  |
| tcatgaaaca  | atittacaaa   | taagcagaga | atacggttt  | ggtgtcctgc  | tacaaaaga  | 720  |
|             |              |            |            |             |            |      |
| catcggtcag  | taatgagcac   | gatgtggaaa | aatgagagaa | gggacacatt  | caaccctgga | 780  |
| gagttcaatg  | gctgctgaag   | ctgcctgctt | ttcactgctg | caaggcctt   | ctgtgtgtga | 840  |
| tgtgcatggg  | agcaacttgt   | tcgtgggtca | tcgggaatac | tagggagaag  | gtttcatgtc | 900  |
| ccccaggcga  | cttcacagag   | tgtgctggag | gactgagtaa | gaaatgctgc  | ccatgccacc | 960  |
| ctttccggct  | ccigtgcctt   | ccctgaactg | ggacctttag | tggtggccat  | ttagccacca | 1020 |
| tcttcgcagg  | ttgcttgcc    | ctggtagggc | agtaacattg | ggtcctgggt  | ctttcatggg | 1080 |
| gtgatgctgg  | gctggctccc   | tctggtctt  | cccaggctgg | ggctgacctt  | cctcgagag  | 1140 |
|             |              |            |            |             |            |      |
| aggccaggtg  | caggttgga    | atgaggctg  | ctgagagggg | ctgtccagtt  | cccagaaggc | 1200 |
| atatcagtct  | ctgagggctt   | ccttgggc   | cggaaacttg | cgggttgag   | gataggagtt | 1260 |
| cacttcatct  | tctcagctcc   | catttctact | cttaagtttc | tcccccatttc | tactcttaag | 1320 |
| tttctcagct  | cccatttcta   | ctctccatg  | gcttcatgct | tcttcattt   | ttctgtttgt | 1380 |
| tttataacaa  | tgtgttagtt   | gtacaaataa | a          |             |            | 1411 |
| <210>       | 93           |            |            |             |            |      |
| <211>       | 9794         |            |            |             |            |      |
| <212>       | DNA          |            |            |             |            |      |
| <213>       | Homo sapiens |            |            |             |            |      |
| <400>       | 93           |            |            |             |            |      |
|             |              |            |            |             |            |      |
| tgttccgtg   | cgcggccgct   | gcmcactcg  | cactgggggg | cgctggctgg  | ctccctggct | 60   |
|             |              |            |            |             |            |      |
| gcggctcc    | agtccggcgc   | ggctgctgt  | gcctgtggcc | cggcggctg   | ggagaagcgg | 120  |
| agtgttgt    | agtgcgcgg    | cggaggtgt  | gttgacgcg  | gtgtgttacg  | tggggagag  | 180  |
| aataaaactc  | cagcgagatc   | cggccgtga  | acgaaagcag | tgacggagga  | gcttgtacca | 240  |
| ccggttaacta | aatgaccatg   | aatctggag  | ccgagaacca | gcagagtgg   | gatgcagctg | 300  |
| taacagaagc  | tgaaaaccaa   | caaatacag  | ttcaagccca | gccacagatt  | gccacattag | 360  |
| cccaggtatc  | tatgccagca   | gctcatgcaa | catcatctgc | tcccaccgt   | actctagtag | 420  |
| agctgcccaa  | tggcagaca    | gttcaagtcc | atggagtcat | tcaggcggcc  | cagccatcag | 480  |
|             |              |            |            |             |            |      |
| ttattcagtc  | tccacaagtc   | caaacagttc | agtcttcctg | taaggactta  | aaaagacttt | 540  |
| tctccggAAC  | acagatttca   | actattgcag | aaagtgaaga | ttcacaggag  | tcagtggata | 600  |

|   |      |
|---|------|
| gtgtaactga ttccaaaag cgaaggaaa ttcttcaag gaggccttc tacaggaaa          | 660  |
| ttttgaatga ctatcttct gatgcaccag gagtgccaag gattgaagaa gagaagtctg      | 720  |
| aagaggagac tttagcacct gccatcacca ctgtaacggt gccaactcca atttaccaa      | 780  |
| cttagcagtgg acagtatatt gccattaccc agggaggagc aatacagctg gctaacaatg    | 840  |
| gtaccgatgg ggtacaggc ctgcaaacat taaccatgac caatgcagca gccactcagc      | 900  |
| <br>  |      |
| cgggtactac cattctacag tatgcacaga ccactgatgg acagcagatc ttagtgc当地      | 960  |
| gcaaccaagt tgggttcaa gtcgcctcg gagacgtaca aacataccag attgc当地          | 1020 |
| caccactag cactattgcc cctggagttt ttatggcatc ctccccagca cttcc当地         | 1080 |
| agccgtgtga agaaggcagca cgaaagagag aggtccgtct aaigaagaac agggaaagcag   | 1140 |
| ctcgagagtgc tcgtagaaag aagaaagaat atgtgaaatg tttagaaaac agagtggc当地    | 1200 |
| tgctgaaaa tcaaacaag acattgattt aggagctaaa agcacttaag gacctt当地         | 1260 |
| gccacaaatc agattaattt gggattttaa tttcacctg ttaagggtgaa aaatggactg     | 1320 |
| <br>  |      |
| gcttggccac aacctgaaag acaaataaa catttatatt tctaaacatt tctttttc        | 1380 |
| tatgc当地 aactgc当地 agcaactaca gaatttcatt catttgct tttgc当地               | 1440 |
| actgtgaatg ttccaacacc tgc当地act tctccc当地 agaaatttc aacgccagga          | 1500 |
| atcatgaaga gacttctgct ttcaacccc caccctc当地 aagaagtaat aattgtt当地        | 1560 |
| ctt当地attt gatgggagaa atgaggaaaa gaaaatctt ttaaaaatgaa tttcaagg        | 1620 |
| tgtgctgagc tc当地tattt ccttagggac agaattaccc cagcctt当地 agctgaa          | 1680 |
| atgtgtggc cgc当地tgcata aagtaagtaa ggtgcaatgaa agaagtgtt当地 attgccaat    | 1740 |
| <br>  |      |
| tgacatgttg tc当地atttctc attgtgaaatt atgtaaagtt gt当地agagac ataccctcta   | 1800 |
| aaaaagaact ttagcatggt attgaaggaa ttagaaatgaa attggagtg ct当地ttatgt     | 1860 |
| atgtgtctt ct当地ataact gaaaatttgc cttgggtct taaaagcatt ctgtaactat       | 1920 |
| acagcttccataggccag ttgttgc当地 ttaattcattt ctgtatgtt当地 ttcaacattt       | 1980 |
| ttgaatacat taaaagaagt aaccaactgaa acgacaaagc atggattttt当地 aattttaat   | 2040 |
| taaagc当地aaag taaaatggat attttagtta gtactaaattt ct当地tagaaaa            | 2100 |
| tgctgatcag taaaccaatc ctttagtta tataacaaga ttttaataa aatgtt当地         | 2160 |
| <br>  |      |
| tc当地tacattt caaaaatatt tatattgtca ct当地ttaacg taaaagata tttctaa        | 2220 |
| actgttgc当地 attgc当地tta cataccacca ccaaggaaagc ct当地caagatg tcaatataa    | 2280 |
| caaagtgata tatatttgc当地 ttagaaatgt tacatgttaga aaaatactgaa ttttaat     | 2340 |
| tttccatattt aacaatatttca cagagaatct ct当地tgaattt ttttaatgaa aagaagtgtt | 2400 |
| aaggatataa aaagtacagt gtttagatgtg cacaaggaaa gtttatttca gacatatttgc当地 | 2460 |

|  |      |
|--|------|
| aatgactgct gtactgaat attggatt tcattttac aaaacattt tttgttct         | 2520 |
| tgtaaaaaga gtagttatta gttctgcctt agcttccaa tatgctgtat agccttgct    | 2580 |
| <br>   |      |
| atttataat tttaattcct gattaaaaca gtctgtatTTT gtgtatatca tacattgtt   | 2640 |
| tcaataaccac tttaattgt tactcatTTT attcactaag ctgcataaat ctaacagtta  | 2700 |
| ctctaaaaaa aaaaaaaaaa agactaaggT ggatTTTaaa aattggaaac tgacataatg  | 2760 |
| ttagttata atttctcatt tggagccggg cgcaGTggct cacgcctgta atcccagcac   | 2820 |
| tttgggaggc caaggTgggt ggatcacTg tggTcaagag ttcaagacca gcctggccat   | 2880 |
| catggtaaa cccatctct actaaaaata caaaaattag ccaggcgtgg tggctggcgc    | 2940 |
| ctgtaatccc agtactcag gaggttgagg cagcagaatt gcttgaaccc aggaggcaga   | 3000 |
| <br>   |      |
| gggtTgcagt gagccgagat agcaccattg cactccagcc tggcgactc catctaaaa    | 3060 |
| aataaaaata aaaaaaatgt ctcatttggg aaggaaattc ctTTTaaaaa agagttgaga  | 3120 |
| cacttagaaa actaatgttt tatatttagt caagagtta ttaagaagt caagcttgtt    | 3180 |
| taacaacaaa atatgaagat ttaagtgtt attgcTggat ccattttaaa ataagatTTT   | 3240 |
| attaaacatt tggtaatgtt atatTTTcgt ttgttaacaaa ccattgtctt tttcaagga  | 3300 |
| tgaacagagt ttatgaagga gcatcattct aagaattaag tggatgttagtc ttatgtttg | 3360 |
| gacagttcac cagattctca agaaggctt caaacaacta taaagttga tggctgtcct    | 3420 |
| <br>   |      |
| gctgagctaa tgggaaagt tatagcataa aaattgtgtt accgcataaa tatgtcattt   | 3480 |
| ttaaaaactg gttiacaga aatcaagcaa agtcacaaat atgttccacaa gttggaaatta | 3540 |
| tttattgagt caaaatgtcg aatcgaacat ttgtatgaa gtaagtgtt taaatgaaaa    | 3600 |
| attgcctgat gtttagcagt ttgtattctc taaagcttt tttcaaaaagt tcaggcttc   | 3660 |
| tacttactgg gaaggTgggt gtcctcttag tccctgataa atcaaggcaa tcacattcat  | 3720 |
| gtgagctgga tgaatttata agttataaag accttatcct tcataccttggaggatgattt  | 3780 |
| cactggTTT gaagtcaGTT gcttaatgtt gaggtgagaa atgtatccTtggctaaatc     | 3840 |
| <br>   |      |
| tgtcttagac cttgggtgaa acttgaagat ttcaGTTTaa aagataaaaa tcaagcatct  | 3900 |
| tttgcagt ttTttttttt taatgcaga atggTgggaa ggTTTtttggtaaagcatgaa     | 3960 |
| actttgagaa tctttattaa gaaaatgaca taatTTTaa aaacctgtt gccaagaaca    | 4020 |
| tatgtggcca cattaccagt aataaatgtt ttTctttta tattggccaa aagggaaataa  | 4080 |
| aaatgtcatc ataggaattt gtacatatgc tactgatttgcctgaaaat agcaagtgg     | 4140 |
| atattgtca ctttgcataat atagggccat gtggcactt tatctatagg acagattaat   | 4200 |



|   |      |
|---|------|
| ttaacatcaa agaaatgctg cctaaattt atttcagatg aggaaggaga aagtaaagtg    | 6000 |
| tgcatagtaa ggctgttagt gaagagttt gagataata gttcactcg ttgtacaag       | 6060 |
| cacaactaga actttttgtt gggaggctt catacatctt gaatattctt aatgtataa     | 6120 |
| tgttactat taagttggct acacagtac tttatgtact aggaacttgtt ttcctigaca    | 6180 |
| ttctagaatc aatggctagg agaggcatta atcttgagg ggctgaacat atcatgaagc    | 6240 |
| tgagttagta tgaaaattt tcaaataaac agggtgtga agttccatct gtctcatctg     | 6300 |
| cttatgataa gttcttattt attagtgaat gtagcttaag cctttgtatg tgtcctcagg   | 6360 |
|   |      |
| ggcagaccg acttaagag ggaccagata acgttgaat ggagggatta tatttcaggt      | 6420 |
| gttttagctt gaaatttattt tttaaaaaa agaaaaattt aaaaatata taaaataaaat   | 6480 |
| agaacaaagc cggtgatgca agttgatatt ataaacaggc agtttagca cagaaagaaa    | 6540 |
| atactgacct gttgcattc tggtaacgggt ggtgcaggc ccagctgggt atgacatgat    | 6600 |
| acattttaa ttattctcac cagcaagtaa aaggaaaatg aacaatctt tgaaattgtc     | 6660 |
| tttggaaaagg atcaaagagt agggaaattca catttgacct aacattactt gcctatagaa | 6720 |
| gtatggcatt tccaagctt tgcgtgagga gcatctcaga gaagtgagag taaatctgag    | 6780 |
|   |      |
| ttagctaaa aattggtagg gaggaagaaa atctctgcaa ataatgattt tatgtttgtt    | 6840 |
| ggccaagtga aatgatctat cattgtgtt gggagggtt atttcttat gttttaaaaa      | 6900 |
| ttggtaatg ctatagat gtatTTTat ccaagtgcc ctcatttgc tttatgtat          | 6960 |
| aaaattttt atattaaaag tggaaataa ttgtcaacat ttttttgag tatagattt       | 7020 |
| tttagggtgg caaagaagag tgcttagttag cagtttcca tggaaatgt tccttgactg    | 7080 |
| atTTGTCAC atgtcagttt taactcccc actccctgca aaaggaattt tttctaacc      | 7140 |
| agatgtatca ctgaaactt tttagaaagca aaataatcag ggaagttcct agaaagggt    | 7200 |
|   |      |
| ttggctttt gtttttgag gttgggtt aagaagactt ccccaacaac tgcgtcaca        | 7260 |
| aaacagggtt ttgatTTTA actctgtatgt ttctatttga gttgaatact aaataataa    | 7320 |
| ctataatgag gaaaatacat ttctaataaa attccctaca ttctagaaac atccctgttt   | 7380 |
| taattttttt atctaaatct tttgtgttt tatgtgtaaa gaaaaaaatg tactgagtt     | 7440 |
| caatgcattt tattaacact atgtacataa tagctgttt gtgttcagaa tagtagcgt     | 7500 |
| tgctttgtat attaaagtga tccttgcgaa ttgtgaaat attgtataa agtgcTTTT      | 7560 |
| cttactgtaa tctttgtgtt atcaactgtc ataatgtct tttacacaa acatttatgt     | 7620 |
|   |      |
| gcagtacat aaacatgctt taaaaactc tgtaagtctc tttttgggg atggatctc       | 7680 |
| tatattttgt tggTTTTT ttgcttagtag tggaaagcca tggTTTATTG gacttaaagt    | 7740 |
| tacaatataat tacaagcttg tggaaagg cagcaaaact aattcagaca acaacatgtc    | 7800 |

|   |      |
|---|------|
| ttcagttact ggatccctaa tttcaggac aaaacctgtt tttcaataag attgaacagt        | 7860 |
| gcctatttgt ggatttgag atgttactgt caagatgact aatggagaca tacgaccagc        | 7920 |
| tgtgtctgat gtcataaaaac acgtgtcac taaaaaggaca ataagactat atacccctc       | 7980 |
| aggtccccctt gcaattctaa aactctgtga tcataataat tggaaaggaaa ggggaggggaa    | 8040 |
| <br>  |      |
| tatggtaat ctggcttaa gctgtaagaa taaaaaagtt atccctata ctattaactt          | 8100 |
| ctgaaataag ttctgagacg agacatctga aaataagcag ctgcattatt tgtatgttcc       | 8160 |
| ttcactgcca agatgtgttc aagcctgcta tacctgcccatttgtattggaa ggcttaatga      | 8220 |
| atttcattta tttctgcaa caacgattac agaatttatt gcacaaaatg agacatttg         | 8280 |
| agagtgatataaattacatg agggacaata ggcataact aggattgttc taagcaaatc         | 8340 |
| ggaatcggtt caccctgcca cggtcagggt cttggacctt cagggaaaaga ttgcccatct      | 8400 |
| tgtcatttga ccaggcactg aagtgacaag accatccttga agaagtca a tccaaagata      | 8460 |
| <br>  |      |
| aaattctgat ccatttcttag ttttagtgtt tcgcccactga agacttaaca tatgtcttt      | 8520 |
| acactcaggt tgcaaaaacac aggcccaaga caaacttaac ttctccccca aatcttcctt      | 8580 |
| ccgctggttt ttccatctcg taagtggcactt cactatccat ctgttaattt gtttagggaa     | 8640 |
| aacctagaaa agcactaccc taatcgtgt tatccttctt cttactgtg cgtccattt          | 8700 |
| tctccacatc ttcttaagt gcagtgcacca aaccggatga gaattcta acgggcctga         | 8760 |
| catcaaatgg aaaggaagga taatgtccag gagttgaaat gttatcctt ttttaatta         | 8820 |
| agatgcaatt cacataaattt aactttttaa gtgaacaattt aagtggtagt acatccacaa     | 8880 |
| <br>  |      |
| tgggttacaa ccaccacttc tatctagctc caaaaacattc tcatcactcc aaaagtaaag      | 8940 |
| tcccgttact ctccattttc tcctccacc gccctgtcc ctggcaacca ccaatctgct         | 9000 |
| tcctgtttct ttggatttac atccgggtat ttcatgtgag actcatcac tttgttattac       | 9060 |
| ttcttcgtc tagcttaat gtgttgtga ggttgcata ttgtacatg ttatcactac            | 9120 |
| ttcatttcctt ttatagcta agtataacttt ttatagtaag tatgcattt tagatata         | 9180 |
| ccacaagttt atcgattcat ccagttgagt tgttctact gttggctaa tggtcatagt         | 9240 |
| gctgttatga atgttcgtgt acaagtattt gagtccgtt tttcaattt ttgggtata          | 9300 |
| <br>  |      |
| tgcctggag tggagttgtt gggcatgtt gaaatcgac atttaacttt ttgaggaact          | 9360 |
| gtcaaacttt ccctcagcag ctgtaccgtt ttacccatca ccattgtatgt atgagggttc      | 9420 |
| caatttcac acacccatca caacacttat ttgcattttaaaaattt agccatcctc            | 9480 |
| atgggtgtgg tctctcatttgg tgggtttgtt ttgcatttcc ctgattacta atgatgtgga     | 9540 |
| gcatcttttgg tttcttgg ccatctgcgt atcttcttgc aagaaatgtc tggtgagggtc       | 9600 |
| ctttgttcat tggaaatttttgg tttttgggtt ctgagttccctt tataatttctt gggtactagg | 9660 |

|             |              |             |             |            |            |      |
|-------------|--------------|-------------|-------------|------------|------------|------|
| cccttataat  | atttcgctt    | ataagtttt   | gctttataat  | gtcctcattt | ttttcaaact | 9720 |
| <br>        |              |             |             |            |            |      |
| tactttatgt  | aatatgtaca   | cttctaaaaa  | aaagaaaacat | ggaaaaggc  | aaactgtaaa | 9780 |
| <br>        |              |             |             |            |            |      |
| aaaaaaaaaa  | aaaa         |             |             |            |            | 9794 |
| <br>        |              |             |             |            |            |      |
| <210>       | 94           |             |             |            |            |      |
| <211>       | 1241         |             |             |            |            |      |
| <212>       | DNA          |             |             |            |            |      |
| <213>       | Homo sapiens |             |             |            |            |      |
| <400>       | 94           |             |             |            |            |      |
| <br>        |              |             |             |            |            |      |
| gaattcgcgg  | ccgcccgttc   | tcacggcatt  | cagcagcagc  | gttgctgttt | ccgacaaaga | 60   |
| caccccgaa   | ttaagcacat   | tcctcgattt  | cagcaagca   | ccgcaacatg | accgaaatga | 120  |
| gtttctgag   | cagcgagggt   | ttgggtgggg  | acttggatgt  | ccccctcgac | ccgtcgggtt | 180  |
| tggggcgtga  | agaaagccta   | ggtcttttag  | atgattacct  | ggaggtggcc | aagcacttca | 240  |
| <br>        |              |             |             |            |            |      |
| aacccatgg   | gttctccagc   | gacaaggcta  | aggcgggctc  | ctccgaatgg | ctggctgtgg | 300  |
| atgggttgtt  | cagtccctcc   | aacaacagca  | aggaggatgc  | cttctccggg | acagattgga | 360  |
| tgttggagaa  | aatggatttg   | aaggagttcg  | acttggatgc  | cctgttgggt | atagatgacc | 420  |
| tggaaaccat  | gccagatgac   | cttctgacca  | cgttggatga  | cacttgttat | ctcttgc    | 480  |
| cccttagtcca | ggagactaat   | aagcagcccc  | cccagacggt  | gaacccaatt | ggccatctcc | 540  |
| cagaaagttt  | aacaaaaccc   | gaccaggttg  | cccccttac   | cttcttacaa | cctttcccc  | 600  |
| tttccccagg  | ggtcctgtcc   | tccactccag  | atcattcctt  | tagtttagag | ctggcagtgt | 660  |
| <br>        |              |             |             |            |            |      |
| aagtggatat  | cactgaagga   | gataggaagc  | cagactacac  | tgcttacgtt | gccatgatcc | 720  |
| ctcagtgcatt | aaaggaggaa   | gacaccctt   | cagataatga  | tagggcatc  | tgtatgagcc | 780  |
| cagagtctta  | tctgggtct    | cctcagcaca  | gcccctctac  | cagggctct  | ccaaatagga | 840  |
| gcctcccatc  | tccaggttt    | ctctgtgggt  | ctgcccgtcc  | caaaccctac | gatcccttg  | 900  |
| gagagaagat  | gttagcagca   | aaagtaaagg  | gtgagaaact  | ggataagaag | ctgaaaaaaa | 960  |
| tggagcaaaa  | caagacagca   | gccacttaggt | accgccagaa  | gaagaggcgc | gagcaggagg | 1020 |
| ctcttactgg  | tgagtgcaaa   | gagctggaaa  | agaagaacga  | ggctctaaaa | gagagggcgg | 1080 |
| <br>        |              |             |             |            |            |      |
| attccctggc  | caaggagatc   | cagtacctga  | aagatttgc   | agaagaggc  | cgcaaggcaa | 1140 |
| gggggaagaa  | aagggtcccc   | tagttgagga  | tagtcaggag  | cgtcaatgt  | cttgcata   | 1200 |
| gagtgcgtta  | gctgtgtgtt   | ccaataattt  | attttgcagg  | g          |            | 1241 |
| <210>       | 95           |             |             |            |            |      |
| <211>       | 2924         |             |             |            |            |      |

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 95

|  |      |
|--|------|
| gacccgacg atgaagacac cgtggagg tcttctggc ctgcgggtc ctgcgtgc           | 60   |
| tgtcaccatc atcacccgtgc ccgtggttct gctgaacaaa ggcacagatg atgctacagc   | 120  |
| tgacagtgcg aaaacttaca ctctaactga ttacttaaaa aatacttata gactgaagg     | 180  |
| <br>   |      |
| atactcctta agatggattt cagatcatga atatctctac aaacaagaaa ataataatctt   | 240  |
| ggtattcaat gctgaatatg gaaacagctc agtttcttg gagaacagta catttgatga     | 300  |
| gtttggacat tcitatcaatg attattcaat atctcctgtat gggcaggttt ttctctttaga | 360  |
| atacaactac gtgaagcaat ggaggcattc ctacacagct tcataatgaca tttatgattt   | 420  |
| aaataaaagg cagctgatta cagaagagag gattccaaac aacacacagt gggtcacatg    | 480  |
| gtcaccagtg ggtcataaat tggcatatgt ttggaaacaat gacatttatg taaaattga    | 540  |
| accaaattta ccaagttaca gaatcacatg gacggggaaa gaagatataa tatataatgg    | 600  |
| <br>   |      |
| aataactgac tgggtttatg aagaggaagt cttcagtgcc tactctgctc tgtgggtgtc    | 660  |
| tccaaacggc acttttttag catatgccca atttaacgac acagaagtcc cacttattga    | 720  |
| atactccttc tactctgtat agtcaatgca gtacccaaag actgtacggg ttccatatcc    | 780  |
| aaaggcagga gctgtgaatc caactgtaaa gttctttgtt gtaaatacag actctctcag    | 840  |
| ctcagtcacc aatgcaactt ccatacaaat cactgctcct gttctatgt tgatagggaa     | 900  |
| tcactacttg tgitgatgtga catggcaac acaagaaaga atttcttgc agtggctcag     | 960  |
| gaggattcag aactattcgg tcatggatat ttgtgactat gatgaatcca gtggaaagatg   | 1020 |
| <br>   |      |
| gaactgctta gtggcacggc aacacattga aatgagtact actggctggg ttggaaagatt   | 1080 |
| taggccttca gaacctcatt ttacccttga tggtaatagc ttctacaaga tcatcagcaa    | 1140 |
| tgaagaaggt tacagacaca ttgtctat ccaaataatgaaaaaagact gcacatttat       | 1200 |
| tacaaaaggc acctggaaatg tcatggat agaagctcta accagtgatt atctatacta     | 1260 |
| catttagataa gaatataaaatg gaatgccagg aggaaggaat cttataaaa tccaaacttag | 1320 |
| tgactataca aaagtgcacat gcctcagtt tgagctgaat ccggaaaggt gtcagacta     | 1380 |
| ttctgtgtca ttcaatggaaatggcgaagta ttatcagctg agatgttccg gtcctggct     | 1440 |
| <br>   |      |
| gccctctat actctacaca gcagcgtgaa tgataaaggg ctgagagtcc tggaagacaa     | 1500 |
| ttcagcttg gataaaatgc tgcagaatgt ccagatgcc tccaaaaaac tggacttcat      | 1560 |
| tatgttgaat gaaacaaaat ttggatca gatgtatgt ctcctcatt ttgataatc         | 1620 |
| caagaaatataat cctctactat tagatgtgta tgcaggccc tggatcaaa aagcagacac   | 1680 |

|  |      |
|--|------|
| tgtttcaga ctgaaactggg ccacttacct tgcaaggaca gaaaacatta tagtagctag    | 1740 |
| ctttgatggc agaggaagtg gttaccaagg agataagatc atgcatgcaa tcaacagaag    | 1800 |
| actggaaaca tttaaattt aagatcaaat tgaagcagcc agacaatttt caaaaatggg     | 1860 |
| <br>   |      |
| atttgtggac aacaaacgaa ttgcaattt gggctggta tatggagggt acgttaacctc     | 1920 |
| aatggtcctg ggatcaggaa gtggcggtt caagtgtgga atagccgtgg cgccgtatc      | 1980 |
| ccgggtggag tactatgact cagtgtacac agaacgttac atgggtctcc caactccaga    | 2040 |
| agacaacctt gaccattaca gaaattcaac agtcatgagc agagctgaaa atttaaaca     | 2100 |
| agttgagttac ctccattttc atggaacagc agatgataac gttcacttc agcagtcagc    | 2160 |
| tcagatctcc aaagccctgg tcgatgtgg agtggatttc caggcaatgt ggtatactga     | 2220 |
| tgaagaccat ggaatagcta gcagcacagc acaccaacat atatatacc acatgagcca     | 2280 |
| <br>   |      |
| cttcataaaa caatgtttct ctttaccta gcacctaata accatgcc atttaaagct       | 2340 |
| tattaaaact cattttgtt ttcattatct caaaaactgca ctgtcaagat gatgatgatc    | 2400 |
| tttaaaatac acactcaaata caagaaactt aaggtaacct ttgttccaa atttcatacc    | 2460 |
| tatcatctta agtagggact tctgtctca caacagatata ttaccttaca gaagttgaa     | 2520 |
| ttatccggtc gggtttatt gttaaaatc atttctgcat cagctgctga aacaacaat       | 2580 |
| aggaattgtt ttatggagg ctttgcatag attccctgag caggatttt atcttttct       | 2640 |
| aactggactg gttcaaatgt ttttttttcc tttaaaggga tggcaagatg tggcagtga     | 2700 |
| <br>   |      |
| tgtcactagg gcagggacag gataagaggg attaggaga gaagatagca gggcatggct     | 2760 |
| gggaacccaa gtccaaagcat accaacacga ccaggtact gtcagctccc ctcggagaaa    | 2820 |
| actgtgcagt ctgcgtgtga acagctttc tccttagag cacaatggat ctcggggat       | 2880 |
| tttcataacc taccaggta tcgcctcgag gccgcgactc taga                      | 2924 |
| <br>   |      |
| <210> 96   |      |
| <211> 1745   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 96   |      |
| <br>   |      |
| ccccgttagaa ccgaggggggt gggcccgaaaa gtcccgaaaa aggtggagat ggtgaagggg | 60   |
| cagccgttcg acgtggccccc ggcctacacg cagttgcagt acatcggcga gggcgcgtac   | 120  |
| <br>   |      |
| ggcatggtaa gtcggccta tgaccacgtg cgcaagactc gcgtggccat caagaagatc     | 180  |
| agcccttcg aacatcagac ctactgccag cgacgttcc gggagatcca gatcgtctg       | 240  |
| cgcttccccc atgagaatgt catcggcata cgagacattc tgcggcgta caccctggaa     | 300  |

|  |      |
|--|------|
| gccatgagag atgtctacat tgtgcaggac ctgatggaga ctgacctgta caagtgtcg   | 360  |
| aaaagccagc agctgagcaa tgaccatatac tgctacttcc tctaccagat cctgcggggc | 420  |
| ctcaagtaca tccactccgc caacgtgctc caccgagatc taaagccctc caacctgctc  | 480  |
| atcaacacca ccigcacct taagatttgt gatttcggcc tggccggat tgccgatct     | 540  |
| <br>   |      |
| gagcatgacc acaccggctt cctgacggag tatgtggcta cgcgctggta cggggccccca | 600  |
| gagatcatgc tgaactccaa gggctatacc aagtccatcg acatctggtc tgtggctgc   | 660  |
| attctggctg agatgctctc taaccggccc atctccctg gcaagcacta cctggatcag   | 720  |
| ctcaaccaca ttctggcat cctggctcc ccattccagg aggacctgaa ttgtatcatc    | 780  |
| aacatgaagg cccgaaacta cctacagtct ctgcctcca agaccaaggt ggcttggcc    | 840  |
| aagctttcc ccaagtcaga ctccaaagcc cttgacactgc tggaccggat gttaacctt   | 900  |
| aaccccaata aacggatcac agtggaggaa gcgctggctc acccctacct ggagcagtag  | 960  |
| <br>   |      |
| tatgaccgca cggatgagcc agtggccgag gagccctca cttcgccat ggagctggat    | 1020 |
| gacctaccta aggagcggct gaaggagctc atctccagg agacagcactg cttccagccc  | 1080 |
| ggagtgtctgg aggcccccta gcccagacag acatctctgc accctgggc ctggacctgc  | 1140 |
| ctcctgectg cccctctccc gccagactgt tagaaaatgg acactgtgcc cagccggac   | 1200 |
| cttggcagcc caggccgggg tggagcatgg gcctggccac ctctctccct tgctgaggcc  | 1260 |
| tccagcttca ggcaggccaa ggcttctct ccccacccgc cctccccacg ggcctggga    | 1320 |
| cctcaggtgg gcccagttca atctcccgt gctgctgtc cgccttacc ttccccagcg     | 1380 |
| <br>   |      |
| tcccagtc tggcagttt ggaatgaaag ggttctggct gccccaacct gctgaaggc      | 1440 |
| agaggtggag ggtggggggc gctgagtagg gactcacggc catgcctgcc cccctcatct  | 1500 |
| cattcaaacc ccaccctagt ttccctgaag gaacattct tagtctcaag ggctagcatc   | 1560 |
| cctgaggagc caggccgggc cgaatccct ccctgtcaaa gctgtcactt cgcgtgccct   | 1620 |
| cgctgcttct gtgtgtggtg agcagaagtg gagctggggg gcgtggagag cccggctgcc  | 1680 |
| cctgccacct ccctgaccgc tctaataat aaatatagag atgtgtctat ggctaaaaaa   | 1740 |
| aaaaaa   | 1745 |
| <br>   |      |
| <210> 97   |      |
| <211> 1611   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 97   |      |
| acataatttc tggagccctg taccaacgtg tggccacata ttctgtcagg aaccctgtgt  | 60   |

|   |      |
|---|------|
| gatcatggtc tggatctgca acacgggcca ggccaaagtc acagatctt agatcacagg    | 120  |
| tgggttgag caggcaggcag gcaggcaatc ggtccgagtg gctgtcggt cttagctct     | 180  |
| ccgcctggcg tttcccttcc tctccggc agcgtcgccg gctgcaccgg cgccggcag      | 240  |
| tcctgcggga gggcgacaa gagctgaggc gcggccgccc agcgtcgagc tcagcgcggc    | 300  |
| ggaggcggcg gcggcccgcc agccaacatg gcggcggcgg cgccggcgg cgccggcccg    | 360  |
| <br>  |      |
| gagatggtcc gcgggcaggt gttcgacgtg gggccgcgt acaccaacct ctcgtacatc    | 420  |
| ggcgaggcgc cttacggcat ggtgtgcctt gcttatgata atgtcaacaa agttcgagta   | 480  |
| gctatcaaga aaatcagccc cttttagac cagacactact gccagagaac cttgaggag    | 540  |
| ataaaaaatct tactgcgtt cagacatgaa aacatcatg gaatcaatga cattattcg     | 600  |
| gcaccaacca tcgagcaaat gaaagatgta tatatagtagc aggacccat gaaacagat    | 660  |
| ctttacaagc ttgttaagac acaacaccc agcaatgacc atatctgcta ttttctctac    | 720  |
| cagatcctca gagggttaaa atatatccat tcagctaacg ttctgcaccg tgacctcaag   | 780  |
| <br>  |      |
| ccttccaacc tgctgctcaa caccacctgt gatctcaaga tctgtgactt tggcctggcc   | 840  |
| cgtgttcag atccagacca tgatcacaca gggttctga cagaatatgt ggccacacgt     | 900  |
| tggtagcaggc ctccagaaat tatgttgaat tccaaaggctt acaccaagt cattgatatt  | 960  |
| tggctgttag gtcgttccat ggcagaaatg ctttccaaaca ggcccatctt tccaggaaag  | 1020 |
| cattatctt accagctgaa tcacatttt ggtattctt gatccccatc acaagaagac      | 1080 |
| ctgaattgtta taataaattt aaaagctagg aactatttgc ttctttcc acacaaaaat    | 1140 |
| aagggtccat ggaacaggct gttccaaat gctgactcca aagctctgga cttatggac     | 1200 |
| <br>  |      |
| aaaatgttga cattcaaccc acacaagagg attgaagtag aacaggctt ggcccaccca    | 1260 |
| tatctggagc agtattacga cccgagtgtac gagcccatcg ccgaaggcacc attcaagttc | 1320 |
| gacatggaaat tggatgactt gcctaaggaa aagctaaaag aactaatttt tgaagagact  | 1380 |
| gcttagattcc agccaggata cagatctttaa atttgtcagg acaagggttc agaggactgg | 1440 |
| acgtgctcag acatcggtt tttttttcc agtttttgc ccctggctt gtctccagcc       | 1500 |
| cgtttggct tatccactt gactccttgg agccgttgg agggccgggtt tctggtagtt     | 1560 |
| gtggctttta tgcttcaaa gaatttcttcc agtccagaga attcaactggc c           | 1611 |

|       |              |
|-------|--------------|
| <210> | 98           |
| <211> | 1702         |
| <212> | DNA          |
| <213> | Homo sapiens |
| <400> | 98           |

|   |      |
|---|------|
| ccggcagtcc cgagtgcgttcc cgcagagggtt ctgggtgg gagcggagtggatcgggcg          | 60   |
| ggccgaagc cggccgtgg gcgttagatgg gggccggcg gcccggagc ggcggaaacgc           | 120  |
| ggatggct gcaccgtgag cgcgaggac aaggcggcgccgagcgc taagatgtac                | 180  |
| gacaagaacc tgcgggagga cggagagaag gcggcgcgg aggtgaagtt gctgcgttg           | 240  |
| ggtgctgggg agtcaggaa gagcaccatc gtcaagcaga tgaagatcat ccacgaggat          | 300  |
| ggctactccg aggaggaatg cggcagtac cggccgttg tctacagcaa caccatccag           | 360  |
| <br>  |      |
| tccatcatgg ccattgtcaa agccatggaa aacctgcaga tcgactttgc cgaccctcc          | 420  |
| agagcggacg acgccaggca gctatttgcctgtca ctgtccgtca ccggcggagga gcaaggcgtg   | 480  |
| ctccctgatg acctgtccgg cgtcatccgg aggctctggg ctgaccatgg tgtgcaggcc         | 540  |
| tgcttggcc gctcaaggga ataccagctc aacgactcag ctgcctacta cctgaacgac          | 600  |
| ctggagcgtt tgacacagatg tgactacatc cccacacagc aagatgtgt acggaccgc          | 660  |
| gtaaagacca cggggatcgtt ggagacacac ttcaccttca aggacctaca cttcaagatg        | 720  |
| tttgatgtgg ttggtcagcg gtctgagcgg aagaagtggatccactgttg tgagggcg            | 780  |
| <br>  |      |
| acagccatca tcttctgcgtt agccttgagc gcctatgact tgggtcttagc tgaggacgag       | 840  |
| gagatgaacc gcatgcata gaggatgttgcata gcatctgca caacaagtgg                  | 900  |
| ttcacagaca cgtccatcat cctttccctt aacaagaagg acctgttgcgtt ggagaagatc       | 960  |
| acacacagtc ccctgaccat ctgcttccctt gaggatcacag gggccaacaa atatgttag        | 1020 |
| gcagccagctt acatccatgtt tagttttagt gacccatgttgc acgcacaaaga caccaaggag    | 1080 |
| atctacacgc acttcacgtt cggccaccgc accaagaacgtt tgcagttcgtt gtttgcgc        | 1140 |
| gtcaccatgtt tcacatcatcaaa gaacaacatgtt aaggactgcgcgc gcctcttctt agggcagcg | 1200 |
| <br>  |      |
| gggcctggcg ggtatggcca cgcggaaatt ttgtacccccc aaccccttagtggatgggg          | 1260 |
| gcaagaagat cacgctcccccc gcctgttccc cgcgcgtttt tctcctctt cctcttttgc        | 1320 |
| ttctcagctc cccctgtccc cttagtccaa aacgttagggg aggggttcgc acaggcctcc        | 1380 |
| ctgtttgaag cctgccttgc tctgatgttgc tggtaatggc catggatccc cttctggc          | 1440 |
| atctgttctt gtttttaacc attgtttgtt tctgtatgttgc ggggaggggg gcacatgttgc      | 1500 |
| agtctccaa ggctgcgttgc ggagggggccctt cttgcgttcc accctggacc cccagtttgc      | 1560 |
| cccaacacca gcccctgccc cagcccaagt ccaaattttt acgggagccctt cctgcccagt       | 1620 |
| <br>  |      |
| cccccaaccc cagccgctcg gaggcccaa aggaaaaagc acaagaagcgc tgagacgc           | 1680 |
| ccatccctgg aaaccacagt cc  | 1702 |
| <br>  |      |
| <210> 99  |      |
| <211> 1185  |      |

<212> DNA  
<213> Homo sapiens  
<400> 99

|             |              |             |            |            |            |      |
|-------------|--------------|-------------|------------|------------|------------|------|
| atggcgtgcc  | tcggaacag    | taagaccgag  | gaccagcgca | acgaggagaa | ggcgcagcgt | 60   |
| gaggccaaca  | aaaagatcga   | gaagcagctg  | cagaaggaca | agcaggctca | ccggccacg  | 120  |
| caccgcctgc  | tgctgctggg   | tgctggagaa  | tctgtaaaa  | gcaccattgt | gaagcagatg | 180  |
| aggatcctgc  | atgttaatgg   | gtttaatgga  | gagggcggcg | aagaggaccc | gcaggctgca | 240  |
| <br>        |              |             |            |            |            |      |
| aggagcaaca  | gcgatggtga   | gaaggcaacc  | aaagtgcagg | acatcaaaaa | caacctgaaa | 300  |
| gaggcgattg  | aaaccattgt   | ggccgccaatg | agcaacctgg | tgcccccggt | ggagctggcc | 360  |
| aaccccgaga  | accagttcag   | agtggactac  | atcctgagtg | tgtgaacgt  | gcctgacttt | 420  |
| gactccctc   | cggattcta    | ttagcatgcc  | aaggcttgt  | gggaggatga | aggagtgcgt | 480  |
| gcctgctacg  | aacgctccaa   | cgagtaccag  | ctgattgact | gtgcccagta | cttcctggac | 540  |
| aagatcgacg  | tgtcaagca    | ggctgactat  | gtgccgagcg | atcaggacct | gtttcgctgc | 600  |
| cgtgcctga   | cttctgaaat   | ctttgagacc  | aagttccagg | tggacaaagt | caacttccac | 660  |
| <br>        |              |             |            |            |            |      |
| atgttgacg   | tgggtggcca   | gcmcgtgaa   | cgccgcaagt | ggatccagtg | cttcaacgat | 720  |
| gtgactgcca  | tcatcttcgt   | ggtgccagc   | agcagctaca | acatggtcat | ccgggaggac | 780  |
| aaccagacca  | accgcctgca   | ggaggctctg  | aacctttca  | agagcatctg | gaacaacaga | 840  |
| tggctgcga   | ccatctctgt   | gatctgttc   | ctcaacaagc | aagatctgct | cgctgagaaa | 900  |
| gtccttgctg  | gaaaatcgaa   | gattgaggac  | tacttccag  | aatttgcctg | ctacactact | 960  |
| cctgaggatg  | ctactccgaa   | gcccgagag   | gaccacgca  | tgacccggc  | caagtacttc | 1020 |
| attcgagatg  | agtttctgag   | gatcagcact  | gccagtggag | atggcgta   | ctactgtcac | 1080 |
| <br>        |              |             |            |            |            |      |
| cctcatttca  | cctgcgtgt    | ggacactgag  | aacatccgcc | gtgtgtcaa  | cgactgccgt | 1140 |
| gacatcattc  | agcgcatgca   | cttcgtag    | tacgagctgc | tctaa      |            | 1185 |
| <210>       | 100          |             |            |            |            |      |
| <211>       | 4788         |             |            |            |            |      |
| <212>       | DNA          |             |            |            |            |      |
| <213>       | Homo sapiens |             |            |            |            |      |
| <400>       | 100          |             |            |            |            |      |
| tttttagaaa  | aaaaaaatat   | atttccctcc  | tgctcttct  | gcgttaccaa | gctaagtgt  | 60   |
| ttatctggc   | tgcggcggga   | actgcggacg  | gtggcggcgc | agcggctct  | ctgcccaggt | 120  |
| tgtatattcac | tgtatggactc  | caaagaatca  | ttaactcctg | gtagagaaga | aaaccccagc | 180  |

|  |      |
|--|------|
| agtgtgcttg ctcaaggagag gggagatgtg atggacttct ataaaaccct aagaggagga | 240  |
| <br>   |      |
| gctactgtga aggtttctgc gtcttcaccc tcactggctg tcgcttctca atcagactcc  | 300  |
| aagcagcgaa gacttttgtt tgatttcca aaaggctcag taagcaatgc gcagcagccca  | 360  |
| gatctgtcca aagcagtttc actctcaatg ggactgtata tggagagac agaaacaaaa   | 420  |
| gtgatggaa atgacctggg attcccacag cagggccaaa tcagccccc ctcggggaa     | 480  |
| acagacttaa agctttgga agaaagcatt gcaaacctca ataggtcgac cagtgttcca   | 540  |
| gagaacccca agagttcagc atccactgct gtgtctgctg ccccacaga gaaggagttt   | 600  |
| ccaaaaactc actctgatgt atcttcagaa cagcaacatt tgaaggccca gactggcacc  | 660  |
| <br>   |      |
| aacggtgcca atgtgaaatt gtataccaca gaccaagca ctttgacat tttcaggat     | 720  |
| ttggagttt ctctgggtc cccaggtaaa gagacgaatg agagtcctt gagatcagac     | 780  |
| ctgttgatag atgaaaactg ttgccttct cctctggcgg gagaagacga ttcattcctt   | 840  |
| ttgaaaggaa actcgaatga ggactgcaag cctctcattt taccggacac taaacccaaa  | 900  |
| attaaggata atggagatct gttttgtca agccccagta atgtaacact gccccaagtg   | 960  |
| aaaacagaaaa aagaagattt catcgaactc tgcacccctg ggttaattaa gcaagagaaa | 1020 |
| ctggcacag ttactgtca ggcaagctt cctggagcaa atataattgg taataaaatg     | 1080 |
| <br>   |      |
| tctgccattt ctgttcatgg tgtgagtacc tctggaggac agatgtacca ctatgacatg  | 1140 |
| aatacagcat cccttctca acagcaggat cagaaggcta ttttaatgt cattccacca    | 1200 |
| attcccggt gttccgaaaa ttgaaatagg tgccaaggat ctggagatga caacttgact   | 1260 |
| tctctggga ctctgaactt ccctggcga acagttttt ctaatggcta ttcaagcccc     | 1320 |
| agcatgagac cagatgtaa ctctcctca tccagctctt caacagcaac aacaggacca    | 1380 |
| cctccaaac tctgcctggt gtgtctgtat gaagctttagt gatgtcatta tggagtcata  | 1440 |
| acttgtggaa gctgtaaagt ttcttcaaa agagcagtgg aaggacagca caattaccta   | 1500 |
| <br>   |      |
| tgtgctggaa ggaatgattt catcatcgat aaaattcgaa gaaaaaactg cccagcatgc  | 1560 |
| cgctatcgaa aatgtttca ggctggaaatg aacctggaaag ctggaaaaac aaagaaaaaa | 1620 |
| ataaaaaggaa ttccggc cactacagga gtctcacaag aaacctctga aaatcctggt    | 1680 |
| aacaaaacaa tagttctgc aacgttacca caactcaccc ctaccctggt gtcactgttgc  | 1740 |
| gaggttattt aacctgaagt gttatatgca ggatgtata gctctgttcc agactcaact   | 1800 |
| tggggatca tgactacgct caacatgtt ggagggcggc aagtgttgc agcgtgaaa      | 1860 |
| tggccaaagg caataccagg ttcaggaac ttacacctgg atgaccaaatt gaccctactg  | 1920 |

|  |      |
|--|------|
| cagtaactcct ggatgtttct tatggcattt gctctgggtt ggagatcata tagacaatca   | 1980 |
| agtgc当地 acc tgc当地 gat ctgattatta atgagcagag aatgactcta               | 2040 |
| ccctgc当地 gt acgaccaatg taaacacatg ctgtatgtt cctctgagtt acacaggctt    | 2100 |
| caggtaatctt atgaagagta tctctgtatg aaaaccciac tgc当地 ctc ttc当地 gctt    | 2160 |
| aaggacggtc tgaagagcca agagctattt gatgaaattha gaatgaccta catcaaagag   | 2220 |
| ctagaaaaag ccattgtcaa gagggaaagga aactccagcc agaactggca gc当地 ttat    | 2280 |
| caactgacaa aactcttgaa ttctatgcat gaagtgggtt aaaaatctc当地 taactattgc   | 2340 |
| <br>   |      |
| ttccaaacat tttggataa gaccatgagt attgaattcc cc当地 agatgaaatc           | 2400 |
| atc当地 accatc agataccaaa atattcaat ggaaatatca aaaaacttct gtttcatcaa   | 2460 |
| aagtgactgc ctaataaga atggttgc当地 taaagaaagt cgaattaata gcttttattt     | 2520 |
| tataaaactat cagtttgc当地 ttttagggtt ttgttgc当地 atttttattt gtttcatct     | 2580 |
| gttgc当地 tt当地 aataacg cactacatgt ggttataga gggcaagac tt当地 acag        | 2640 |
| aagcaggtaa gtc当地 catca cttttagt atggagagt agatggtaa atttattatg       | 2700 |
| taataatcc cagaaattttag aaacccataat atgtggacgt aatctccaca gt当地 aagaag | 2760 |
| <br>   |      |
| gatggcacct aaaccaccag tgccaaatg ctgtgtgatg aacttctct tc当地 actttt     | 2820 |
| ttt当地 cacagtt ggctggatga aatttcttag acttctgtt ggttatccc cccctgtat    | 2880 |
| agttaggata gc当地 ttttgc当地 tttatgc当地 gaaacctgaa aaaaagtttca aagtgta    | 2940 |
| tcagaaaagg gaagttgc当地 ct当地 tattactgtc tggttttaac aatttccccc当地        | 3000 |
| atatttagt aactacgctt gtc当地 ttttgc当地 tcttacataa ttttatttca aagtattgt  | 3060 |
| acagctgtt aagatggca gctagttcgat agcttccca aataactctt aacatataat      | 3120 |
| caatcatctg tttgaaaatg gtttgc当地 tcttacactg tggcacttag ctatcagaag      | 3180 |
| <br>   |      |
| accacaaaaa ttgactcaaa tctccagtt tcttgc当地 aaaaaaaaaa aaaaagctca       | 3240 |
| tattingtata atatctgctt cagttggaaat ttatataatgt tttgcaattt aacatgttca | 3300 |
| actgtatag agcacctgtt cc当地 gctggtaaa ctgtggatga tggtaatgc当地           | 3360 |
| agactaattt aaaaataac taccaagagg cc当地 tctgtctgtt acctaaacgcc ctat     | 3420 |
| aatggctata tggcaagaaa gctggtaaaat tatttgc当地 tcaggacctt ttgaatgt      | 3480 |
| ttgtataact tcttaaaatg tttgatttca gataaccagg ttttgc当地 ctgagagact      | 3540 |
| tttaatcaga caaagtaattt cctctacta aacttaccc aaaaactaaa tctctaatat     | 3600 |
| <br>   |      |
| ggcaaaaatg gctagacacc cattttccata ttcccatctg tc当地 accatgtt gttt      | 3660 |
| cctgatggta cagggaaatgtt cagttactgtt ttttgc当地 tt当地 aactgttca          | 3720 |
| tccatgtttt gtaaaactaca catccctaat gtttgc当地 gagtttaca caagtc当地        | 3780 |

|  |      |
|--|------|
| gaatttcttc actgttgaaa attatttaa acaaaataga agctgttagta gcccttctg | 3840 |
| tgtgcaccc accaacttcc tgtaaactca aaacttaaca tatttactaa gccacaagaa | 3900 |
| atttgatttc tattcaaggt ggc当地atta ttttgtata agaaaactga aaatctaata  | 3960 |
| ttaaaaaat ggaacttcta atatatttt atatttagtt atagttcag atataatca    | 4020 |

|  |      |
|--|------|
| tattggatt cactaatctg ggaagggaag ggctactgca gcttacatg caatttatta    | 4080 |
| aaatgattgt aaaatagctt gtatagtgtaa aaataagaat gatfffftga tgagattgtt | 4140 |
| ttatcatgac atgttatata tttttgttag gggtaaaga aatgctgatg gataaccat    | 4200 |
| atgatttata gtttgatcat gcattcatac aggcccgat ggtctcagaa accaaacagt   | 4260 |
| ttgctctagg ggaagaggga gatggagact ggtccgtgt gcagtgaagg ttgctgaggc   | 4320 |
| tctgaccagg tgagattaca gaggaagtta tcctctgcct cccattctga ccaccctct   | 4380 |
| cattccaaca gtgagtcgt cagcgcagg ttagttact caatctcccc ttgcactaaa     | 4440 |

|   |      |
|---|------|
| gtatgtaaag tatgtaaaca ggagacagga aggtgggtct tacatcctta aaggcaccat | 4500 |
| ctaatacgcc gttacttca catacagccc tccccagca gttgaatgac aacagaagct   | 4560 |
| tcagaagttt ggcataatgtt tgcatagagg taccagcaat atgtaatag tgcagaatct | 4620 |
| cataggttgc caataataca ctaattcctt tctatcctac aacaagagtt tattccaaa  | 4680 |
| taaaatgagg acatgtttt gtttcttg aatgctttt gaatgttatt tgttatssc      | 4740 |
| agtatttgg agaaattatt taataaaaaa acaatcattt gcttttg                | 4788 |

&lt;210&gt; 101

&lt;211&gt; 1498

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 101

|  |     |
|--|-----|
| accaaacctc ttggggcac aaggcacaac aggctgctct gggattctct tcagccatc    | 60  |
| ttcattgctc aagtgtctga agcagccatg gcagaagtac ctgagctcgc cagtgaaatg  | 120 |
| atggcttatt acagtggcaa tgaggatgac ttgttcttg aagctgatgg ccctaaacag   | 180 |
| atgaagtgt cttccagga cttggaccc tggccctgg atggccgat ccagctacga       | 240 |
| atctccgacc accactacag caagggcttc aggccggccg cgtcaggatgt tgtggccatg | 300 |
| gacaagctga ggaagatgct gttccctgc ccacagacct tccaggagaa tgacctgagc   | 360 |

|  |     |
|--|-----|
| accttcttc cttcatctt tgaagaagaa cctatcttct tcgacacatg ggataacgag    | 420 |
| gtttatgtgc acgatgcacc tgtacgatca ctgaactgca cgctccggaa ctcacagcaa  | 480 |
| aaaagcttgg tggatgtctgg tccatatgaa ctgaaagctc tccacctcca gggacaggat | 540 |

|   |      |
|---|------|
| atggagcaac aagtgggtt ctccatgtcc tttgtacaag gagaagaaag taatgacaaa    | 600  |
| atacctgtgg ccttgggcct caaggaaaag aatctgtacc tgtcctgcgt gttgaaagat   | 660  |
| gataagccca ctctacagct ggagagtgt a gatccaaaa attacccaaa gaagaagatg   | 720  |
| gaaaagcgat ttgtttcaa caagatagaa atcaataaca agctgaaatt tgagtctgcc    | 780  |
|   |      |
| cagttccccca actggatcat cagcacctct caagcagaaa acatgcccgt cttctggga   | 840  |
| gggaccaaaag gcggccagga tataactgac ttcaccatgc aatttgtgtc ttcctaaaga  | 900  |
| gagctgtacc cagagagtcc tgtgctaat gtggactcaa tccttagggc tggcagaaag    | 960  |
| ggaacagaaa gtttttgag tacggctata gcctggactt tcctgtgtc tacaccaatg     | 1020 |
| cccaactgcc tgccttaggg tagtgctaag aggatctcct gtccatcagc caggacagtc   | 1080 |
| agctctctcc tttcagggcc aatccccagc cttttgtt agccaggcct ctctcacctc     | 1140 |
| tcctactcac ttaagcccg cctgacagaa accacggcca catttggttc taagaaaccc    | 1200 |
|   |      |
| tctgtcatc gctccacat tctgatgagc aaccgttcc ctatttattt atttatttgt      | 1260 |
| ttgttgttt tattcattgg tctaatttat tcaaaggggg caagaagtag cagtgtctgt    | 1320 |
| aaaagagcct agttttaat agctatggaa tcaattcaat ttggactggt gtgctcttt     | 1380 |
| taaatcaagt ccttaatta agactgaaaa tatataagct cagattattt aaatggaaat    | 1440 |
| atttataaat gagcaaatat catactgttc aatggttctg aaataaactt cactgaag     | 1498 |
| <210> 102   |      |
| <211> 1128  |      |
| <212> DNA   |      |
| <213> Homo sapiens  |      |
| <400> 102   |      |
| attctgcctt cgagcccacc gggAACGAAA gagaagctct atctccctc caggagccca    | 60   |
|   |      |
| gctatgaact ccttctccac aagcgccccc ggtccagttt ctttctccct ggggctgtc    | 120  |
| ctgggtttgc ctgtgcctt ccctggccca gtaccccccag gagaagatgc caaagatgt    | 180  |
| gccggccccc acagacagcc actcacctct tcagaacgaa ttgacaaaca aattcggtac   | 240  |
| atccctcgacg gcatctcagc cctgagaaag gagacatgtt acaagagttt catgtgtt    | 300  |
| agcagcaaag aggcactggc agaaaacaac ctgaacccctc caaagatggc tgaaaaagat  | 360  |
| ggatgctcc aatctggatt caatgaggag acttgcctgg tgaaaatcat cactggctt     | 420  |
| ttggagttt aggtataacct agagtacccctc cagaacagat ttgagagtag tgaggaacaa | 480  |
|   |      |
| gccagagctg tccagatgag tacaaaagtc ctgatccagt tcctgcagaa aaaggcaag    | 540  |
| aatcttagatg caataaccac ccctgacccca accacaaatg ccagcctgct gacgaagctg | 600  |

|  |      |
|--|------|
| caggcacaga accagtggt gcaggacatg acaactcatc tcattctgct cagcttaag    | 660  |
| gagttccgc agtccagcct gagggctt cggcaaatgt agcatggca cctcagattg      | 720  |
| ttgttgtaa tggcattcc ttcttctgtt cagaaacctg tccactggc acagaactta     | 780  |
| tgttgttctc taiggagaac taaaagtatg agcgttagga cactattta attatittta   | 840  |
| atttataat attaaatat gtgaagctga gttaaattat gtaagtata ttttatattt     | 900  |
| <br>   |      |
| ttaagaagta ccacttgaaa catttatgt attagtttg aaataataat ggaaagtggc    | 960  |
| tatgcagttt gaatatcctt tggttcagag ccagatcatt tcttgaaag tgtaggctt    | 1020 |
| cctcaaataa atggctaact ttatacatat tttaaagaa atatttatat tgtatttata   | 1080 |
| taatgtataa atggtttta taccaataaa tggcattttt aaaaattt                | 1128 |
| <210> 103  |      |
| <211> 5191   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 103  |      |
| gaattcagta acccaggcat tatttatcc tcaagtctt ggtgggtgg agaaagataa     | 60   |
| caaaaaagaaa catgattgtg cagaaacaga caaaccttt tggaaagcat ttgaaaatgg  | 120  |
| <br>   |      |
| cattccccct ccacagtgtg ttcacagtgt gggcaaattt actgctctgt cgtactttct  | 180  |
| aaaaatgaag aactgttaca ccaagggtaa ttatttataa attatgtact tgcccagaag  | 240  |
| cgaacagact ttactatca taagaaccct tccttgggtgt gcttttatac tacagaatcc  | 300  |
| aagaccttca aagaaagggtc ttggattttt ttcttcagga cactaggaca taaagccacc | 360  |
| tttttatgtat ttgttgaat ttctcactcc atccctttt ctgtatgtca tgggtcctca   | 420  |
| gaggtcagac ttgggtgcct tggataaaga gcatgaagca acagtggctg aaccagatgtt | 480  |
| ggaacccaga tgcttttcc actaagcata caacttcca ttagataaca cctccccc      | 540  |
| <br>   |      |
| accccaacca agcagctcca gtgcaccact ttctggagca taaacataacc ttaactttac | 600  |
| aacttgatgt gccttgaata ctgttccat ctggaatgtg ctgttcttt tcatcttct     | 660  |
| ctattgaagc cctccttattt ctaatgect tgctccaact gccttggaa gattctgctc   | 720  |
| ttatgcctcc actggattt atgttttagt accacttgc tattctgcta tatagtcgt     | 780  |
| ccttacattt cttttttt ctgatagacc aaacttttta aggacaagta cctagtttta    | 840  |
| tctatttcta gatccccac attactcaga aagttactcc ataaatgtt gtggaaactga   | 900  |
| tttctatgtg aagacatgtg ccccttcaact ctgttaacta gcattagaaa aacaaattt  | 960  |
| <br>   |      |
| ttgaaaagtt gtagtatgcc cctaagagca gtaacagtcc ctagaaactc tctaaatgc   | 1020 |

|   |      |
|---|------|
| ttagaaaaag atttattta aattacctcc ccaataaaat gattggctgg cttatctca         | 1080 |
| ccatcatgat agcatctgta attaactgaa aaaaataat tatgccatta aaagaaaatc        | 1140 |
| atccatgatc ttgttctaact acctgccact ctagtactat atctgtcaca tggctatga       | 1200 |
| taaagttatc tagaataaaa aaagcataca attgataatt caccaatig tggagctca         | 1260 |
| gtatTTaaa tgtatattaa aattaaatta tttaaagat caaagaaaac tttcgtcata         | 1320 |
| ctccgtattt gataaggaac aaataggaag tgtgtgact caggttgcc ctgagggat          | 1380 |
|   |      |
| ggccatcag ttgcaaatcg tggaaattcc tctgacataa tgaaaagatg aggggtgcata       | 1440 |
| agtctctag tagggtgatg atataaaaag ccaccggagc actccataag gcacaaactt        | 1500 |
| tcagagacag cagagcacac aagcttctag gacaagagcc aggaagaaac caccggagg        | 1560 |
| aaccattctc actgtgtgta aacatgactt ccaagctggc cgtggctctc ttggcagcct       | 1620 |
| tcctgatttc tgcaagctcg tgtgaaggta agcacattt tctgacactac agcgtttcc        | 1680 |
| tatgtctaaa tgtgtcctt agatagcaaa gctattctt atgcttggt aacaacatc           | 1740 |
| cttttattt agaaacagaa tataatcttta gcagtcaatt aatgttaaat tgaagattta       | 1800 |
|   |      |
| aaaaaaacta tatataaacac ttaggaataa taaaggttt atcaatatacg atattctgt       | 1860 |
| tttataattt ataccaggta gcatgcataat attaacgtt aataagtaat ttatgtatg        | 1920 |
| tcctattgag aaccacggtt accttatatttttgttataa ttgagttgag caaggtaact        | 1980 |
| cagacaattt cactcctgtt agtatttcat tgacaaggctt cagattgtc attaattct        | 2040 |
| gtctggttt aagataccctt gattatagac caggcatgta taacttattt atatattct        | 2100 |
| gttaattctt tctgaaggca atttctatgc tggagatct tagttgcctt actataaata        | 2160 |
| acactgtggtt atcacagagg attatgcaat attgaccaga taaaatacc atgaagatgt       | 2220 |
|   |      |
| tgatattgtt caaaaagaac tctaactttt atataggaag ttgttcaatg ttgtcagtta       | 2280 |
| tgactgtttt ttaaaacaaa gaactaactg aggtcaaggg cttaggataat ttcaggaatg      | 2340 |
| agttcactag aaacatgatg cttccatag tctccaaata atcatattgg aattagaagg        | 2400 |
| aagtagctgg cagagctgtt cctgttataa aatcaatcc ttaatcactt tttccccaa         | 2460 |
| caggtgcagt ttgcaggaa agtgctaaag aacttagatg tcagtgcata aagacatact        | 2520 |
| ccaaacctttt ccacccaaa ttatcaaag aactgagatg gattgagatg ggaccacact        | 2580 |
| gcgccaacac agaaattatgtt taagtactttt aaaaagattt agatattttt ttttagcaaa    | 2640 |
|   |      |
| cttaaaatatttta aggaagggtgg aaatattttt gaaagttcca ggtgttagga ttacagtatgt | 2700 |
| aaatgaaaca aaacaaaataaaaatatttgc tctacatgac atttaatattt ggtgttcc        | 2760 |
| acaactacta taaatgttat ttggactta gactttatgc ctgacttaag gaatcatgat        | 2820 |
| ttgaatgcaaa aaactaaata ttaatctgaa ccatttctt ctttatttcag tgtaaagctt      | 2880 |

|   |      |
|---|------|
| tctgtatggaa gagagctcg tctggacccc aaggaaaact gggtgcagag ggttgtggag     | 2940 |
| aagttttga agaggtaagt tatatattt ttaatttaaa ttttcattt atcctgagac        | 3000 |
| ataataatcca aagtccagct ataatttct ttctgttgct aaaaatcgac attaggtatc     | 3060 |
| <br>  |      |
| tgcccttttg gttaaaaaaaaa aaggaatagc atcaatagtg agtttgtgt acttatgacc    | 3120 |
| agaaagacca tacatagttt gcccaggaaa ttctgggtt aagttgtgt cctatactct       | 3180 |
| tagtaaagtt ctgtcact cccagtagtg tcctatTTT gatgataatt tctttgatct        | 3240 |
| ccctatttt agttgagaat atagagcatt tctaacaat gaatgtcaaa gactatattg       | 3300 |
| actttcaag aaccctactt tccttcttat taaacatagc tcatctttt attttaatt        | 3360 |
| ttatTTTtagg gctgagaatt cataaaaaaa ttcatctct gtggtatcca agaatcagt      | 3420 |
| aagatgccag tgaaacttca agcaaatttca cttcaacact tcatgtattt tttgggtctg    | 3480 |
| <br>  |      |
| ttttaggggtt gccagatgca atacaagatt cctggtaaa ttgaatttca agtaaacaat     | 3540 |
| gaatagttt tcattgtacc atgaaatatc cagaacatac ttatgtaa agtattttt         | 3600 |
| atttgaatct acaaaaaaca acaaataatt tttaaatata aggattttcc tagatattgc     | 3660 |
| acgggagaat atacaatag caaaattggg ccaaggccca agagaatatc cgaactttaa      | 3720 |
| tttcaggaat tgaatgggtt tgctagaatg tgatatttga agcatcacat aaaaatgtat     | 3780 |
| ggacaataaa ttttgcata aagtcaatt tagctggaaa tcctggattt tttctgtta        | 3840 |
| aatctggcaa ccctagtctg ctggccagga tccacaagtc cttgttccac tttgtcccttgg   | 3900 |
| <br>  |      |
| tttctcctt atttctaagt ggaaaaagta ttagccacca tcttacctca cagtgtatgtt     | 3960 |
| gtgaggacat gtggaaagcac tttaagttt ttcatcataa cataaatttt tttcaagtgt     | 4020 |
| aacttattaa cctattttt attttagtattt ttatTTTaaatc atcaaatttt tttgtcaagaa | 4080 |
| tttggaaaaa tagaagatga atcattgattt gaatagttt aaagatgtt tagtaaattt      | 4140 |
| attttatttt agatattttaa tgatgtttt ttagataat ttcaatcagg gtttttagat      | 4200 |
| taaacaaaca aacaattggg tacccagttt aattttcattt tcagatatac aacaataat     | 4260 |
| tttttagtat aagtacattt ttgtttatct gaaattttaa ttgaactaac aatccatgtt     | 4320 |
| <br>  |      |
| tgtactccc agtcttgc tttgcagctg tttggtagt gctgtgttga attacggaaat        | 4380 |
| aatgagttttag aacttattaa acagccaaaa ctccacagtc aatatttagta atttcttgc   | 4440 |
| ggttgaaact ttgtttttt gtacaaatag attttatTTT tattatTTTaaatgactgtat      | 4500 |
| ttttaataac aaggctttt atttttaact tttagtggttt tatgtgtctt ccaaattttt     | 4560 |
| tttactgtttt ctgattgtat gggaaatataa aagtaaatat gaaacatttta aatataatt   | 4620 |
| ttttgtcaaa gtaatcaagt gtttgcctt ttttagttt tagtttattt ggtttctt         | 4680 |

|  |      |
|--|------|
| tgttatatt taaaattata ctttgattt gaaaacataa atgcccccc tttagcattt     | 4740 |
| <br>   |      |
| gttatggaaa attacaaact tttatTTTA gaaaacagaa ctccTTCCA gaaataggTT    | 4800 |
| acaacagta gtgcCTCCA cagaatgtt gaaatgttt caactcccc CTGTATACTA       | 4860 |
| tcttgctaat aagtctgtct tcagatttcg attaaccggT ttgtatgtct gtgcACTTTA  | 4920 |
| gcatacgTGG acatTTAAAGA ggaaAGAGAG tacatattat aagtgtCTTA tcagtaACTG | 4980 |
| aggagtaaaa ctgataaATG tgaggcaAG aagttaaaa tatggtaaa gcctaAGCAT     | 5040 |
| atTTGCAAAC aaatcaaACA atactCTGAG aagtAAAAC ataattATTt aattaacAAA   | 5100 |
| tttcagtGGA taaaTTTAT aacaattAG acacAGTGA aaataAAATT agaaaACTAG     | 5160 |
| <br>   |      |
| aaaatagaac aaaagaaaact tctggaaattc a                               | 5191 |
| <210> 104  |      |
| <211> 1572   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 104  |      |
| <br>   |      |
| aatttctcac tgcccctgtg ataaactgtg gtcactggct gtggcagcaa ctattataag  | 60   |
| atgcTCTGAA aactttcag acactgaggg gcaccagagg agcagactac aagaatggca   | 120  |
| cacgctatgg aaaactcctg gacaatcagt aaagagtacc atattgtga agaagtgggc   | 180  |
| tttgctctgc caaatccaca ggaaaatcta cctgattttt ataatgactg gatttcatt   | 240  |
| gctaaacatc tgccTgatct catagagtct ggccagcttc gagaaagagt tgagaagtta  | 300  |
| <br>   |      |
| aacatgctca gcattgatca tctcacagac cacaagtcac agcgcTTGC acgtctagTT   | 360  |
| ctgggatgca tcaccatggc atatgtgtgg ggcaaaggTC atggagatgt ccgtaaggTC  | 420  |
| ttgccaagaa atattgctgt tcTTTACTGC caactctcca agaaactggA actgcctcT   | 480  |
| atTTTGGTTT atgcagactg tgtctggca aactggAAGA aaaaggatcc taataagccc   | 540  |
| ctgacttatg agaacatggA cgTTTGTTC tcattcgtg atggagactg cagtaaAGGA    | 600  |
| ttcttcTCTGG TCTCTCTT GGTGGAATA gcagctgCTT ctgcaatcaa agtaattcCT    | 660  |
| actgtattca aggcaatgca aatgcaagaa cgggacactt tgctaaaggc gctgttgaa   | 720  |
| <br>   |      |
| atagcttctt gcttggagaa agccCTTCAA gtgttCACC aaatCCACGA tcatgtGAAC   | 780  |
| ccaaaAGCAT tttcagtgt tcttcgcata tattttgtctg gctggaaagg caacCCCCAG  | 840  |
| ctatcagacg gtctggTGTa tgaagggttc tgggaAGACC caaaggAGTT tgcaggGGGC  | 900  |
| agtgcaggCC aaagcagcgt ctTTCAGTGC ttTGACGTCC tgctggcat ccagcagact   | 960  |
| gctggTggag gacatgctgc tcaGTTCTC caggacatga gaagatataat gccaccagCT  | 1020 |

|  |      |
|--|------|
| cacaggaact tcctgtgctc attagagtca aatccctcag tccgtgagtt tgtccttca     | 1080 |
| aaagggtgatg ctggcctgcg ggaagcttat gacgcctgtg taaaagctct ggtccctcg    | 1140 |
| <br>   |      |
| aggagctacc atctgcaaat cgtgactaag tacatcctga ttccctgcaag ccagcagccca  | 1200 |
| aaggagaata agacctctga agacccttca aaactggaag ccaaaggAAC tggaggcact    | 1260 |
| gattaatga attccctgaa gactgttgaaga agtacaactg agaaatccct tttgaaggaa   | 1320 |
| ggtaatgtt accaacaacag agcacatTTT atcatagcag agacatctgt atgcattcct    | 1380 |
| gtcatttaccc attgttacacag agccacaaac taatactatg caatgtttt ccaataatgc  | 1440 |
| aatacaaaag acctaataat acctgttcat ttcttgttagg aaaacaacaa aaggttattt   | 1500 |
| tgtttaattt tactagaagt ttgttattt gtatcttac attgttataa aatgacattt      | 1560 |
| <br>   |      |
| aataaataaa aa  | 1572 |
| <210> 105  |      |
| <211> 1539   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 105  |      |
| ggaattccgg gccgggtctt tcctcccggcc gccggccggcc tggtcccggt gactggcctc  | 60   |
| cacgtccgac tcgtccgagc tgaagcccgag cagcactttg ctgccagccg cggggggccgc  | 120  |
| ggaggcgccc cggggccctc ccaggaggct ctctggcca gaggccgaga ttccggcacag    | 180  |
| gccccccagga gtccgttaagt aggagggttc gcccggagacc ggccggaccc ccatccccgc | 240  |
| ggccggccgc gcccgtggc cccggctgc gaccgtggcg gctggccgtg gaaaaatgtct     | 300  |
| <br>   |      |
| caggagaggc ccacgttcta cccggcaggag ctgaacaaga caatctggga ggtggccgag   | 360  |
| cgttaccaga acctgtctcc agtgggtctt ggcgcctatg gctctgtgt tgctgtttt      | 420  |
| gacacaaaaa cgggggttacg tggcgttgc aagaagctct ccagaccatt tcagtccatc    | 480  |
| attcatgcga aaagaaccta cagagaactg cggttactta aacatatgaa acatgaaaat    | 540  |
| gtgattggc tttggacgt ttgttacacct gcaagggtctc tggaggaatt caatgtatgt    | 600  |
| tatctggtga cccatctcat gggggcagat ctgaacaaca ttgtgaaatg tcagaagctt    | 660  |
| acagatgacc attttcagtt ctttatctac caaattctcc gaggtctaaa gtatatacat    | 720  |
| <br>   |      |
| tcagctgaca taatttcacag ggacctaaaa cctagtaatc tagctgtgaa tgaagactgt   | 780  |
| gagctgaaga ttctggattt tggactggct cggcacacag atgtgaaat gacaggctac     | 840  |
| gtggccacta ggtggtacag ggctcctgag atcatgctga actggatgca ttacaaccag    | 900  |
| acagttgata ttgggtcagt gggatgcata atggccgagc tggactgg aagaacattt      | 960  |

|  |      |
|--|------|
| tttcctggta cagaccatat tgatcagttt aagctcattt taagactcgt tggaacccca  | 1020 |
| ggggctgagc ttttgaagaa aatctcctca gagtctgcaa gaaaactataat tcagtcttg | 1080 |
| actcagatgc cgaagatgaa ctttgcgaat gtatttattt gtgccaatcc cctggctgtc  | 1140 |

|  |      |
|--|------|
| gacttgctgg agaagatgct tgtattggac tcagataaga gaattacagc ggcccaagcc    | 1200 |
| cttgcacatg cctactttgc tcagtaccac gatcctgatg atgaaccagg ggccgatcct    | 1260 |
| tatgatcagt ctttgaag cagggacctc cttatagatg agtgaaaaag cctgacccat      | 1320 |
| gatgaagtca tcagcttgc gcccaccctt cttgaccaag aagagatgga gtcctgagca     | 1380 |
| cctggtttct gttctgttga tcccactca ctgtgagggg aaggcccttt cacggaaact     | 1440 |
| ctccaaatat tattcaagtgc cctcttgttgc cagagatttc ctccatggtg gaaggggttgc | 1500 |
| tgcggtcgctg tgcgtgcgtt ttagtgtgtg tgcatgtgt                          | 1539 |

&lt;210&gt; 106

&lt;211&gt; 1155

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 106

|  |     |
|--|-----|
| atgagcagaa gcaagcgtga caacaatttt tatagtgtag agattggaga ttctacattc  | 60  |
| acagtcctga aacgatataca gaatttaaaa cctataggct caggagctca aggaatagta | 120 |
| tgcgcagctt atgatgccat tcttggaaaga aatgttgcaa tcaagaagct aagccgacca | 180 |
| tttcagaatc agactcatgc caagcgggcc tacagagagc tagttcttat gaaatgttt   | 240 |
| aatcacaaaa atataattgg ctttttgaat gttttcacac cacagaaatc cctagaagaa  | 300 |
| tttcaagatg ttacatagt catggagctc atggatgcaa atcttgcca agtgattcag    | 360 |

|   |     |
|---|-----|
| atggagctag atcatgaaag aatgtcctac cttctctatc agatgctgtg tggaatcaag     | 420 |
| caccttcatt ctgcgtggaaat tattcatcggt gacttaaagc ccagtaataat agtagtaaaa | 480 |
| tctgattgca ctttgaatg tcttgacttc ggtctggcca ggactgcagg aacgagtttt      | 540 |
| atgatgacgc cttatgtatg gactcgctac tacagagcac ccgaggtcat cttggcatg      | 600 |
| ggctacaagg aaaacgtgga tttatggct gtggggtgca ttatggaga aatggttgc        | 660 |
| cacaaaatcc tcttccagg aaggactat attgatcagt ggaataaagt tattgaacag       | 720 |
| cttggAACAC catgtcctga attcatgaag aaactgcaac caacagtaag gacttacgtt     | 780 |

|   |     |
|---|-----|
| gaaaacagac ctaaatatgc tggatatagc tttgagaaac tttccctga tgtcctttc   | 840 |
| ccagctgact cagaacacaa caaactaaa gccagtcagg caagggattt gttatccaa   | 900 |
| atgctggtaa tagatgcatac taaaaggatc tctgtatgt aagctctcca acacccgtac | 960 |

|  |      |
|--|------|
| atcaatgtct ggtatgatcc ttctgaagca gaagctccac caccaaagat ccctgacaag  | 1020 |
| cagtttagatg aaagggaaca cacaatagaa gagtgaaaag aattgatata taaggaagtt | 1080 |
| atggacttgg aggagagaac caagaatgga gttatacggg ggcagccctc tccttagca   | 1140 |
| cagggtgcagc agtga  | 1155 |
| <br>   |      |
| <210> 107  |      |
| <211> 2000   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 107  |      |
| <br>   |      |
| tttgggttgt gtgtgcgacg cgggtcgag gggcagtgg gggAACCGCG aagaagccga    | 60   |
| ggagccccgga gcccccgcgtg acgctctct ctcagtccaa aagcggctt tggttggcg   | 120  |
| cagagagacc cgggggtcta gctttcctc gaaaagcgcc gcctgccct tggccccgag    | 180  |
| aacagacaaa gacgaccgca gggccgatca cgctggggc gctgaggccg gccatggta    | 240  |
| tggaagtggg caccctggac gctggaggcc tgcgggcgt gctgggggag cgagcggcgc   | 300  |
| aatgcctgtc gctggactgc cgctcttct tcgcttcaa cgccggccac atcggcgct     | 360  |
| <br>   |      |
| ctgtcaacgt gcgttcagc accatcgtgc ggccgcgggc caagggcgcc atggcctgg    | 420  |
| agcacatcg gccaaacgccc gagctccgcg gccgcgtct ggccggcgcc taccacgcgg   | 480  |
| tgggttgct ggacgagcgc agcgcggccc tggacggcgc caagcgcac ggcaccctgg    | 540  |
| ccctggcgcc cggcgccctc tgccgcgagg cgccgcgcgc gcaagtcctc ttccctaaag  | 600  |
| gaggatacga acgcgttcg gttctctgc cggagctgt cagcaaacag tcgaccccca     | 660  |
| tggggctcag cttcccttg agtactagcg tccctgacag cgccgaatct gggtgcagtt   | 720  |
| cctgcagttac cccactctac gatcagggtg gcccgggtga aatcctgccc tttctgtacc | 780  |
| <br>   |      |
| tgggcagtgc gtatcacgct tcccgcaagg acatgctgga tgccttggc ataactgcct   | 840  |
| tgatcaacgt ctacgcatt tgccttgcacc attttgggg tcaactaccag tacaagagca  | 900  |
| tccctgtgga ggacaaccac aaggcagaca tcagctctg gttcaacgag gccattgact   | 960  |
| tcatagactc catcaagaat gctggaggaa ggggtttgt ccactgccag gcaggcattt   | 1020 |
| cccggtcagc caccatctgc cttgcttacc ttatgaggac taatcgagtc aagctggacg  | 1080 |
| aggccttga gtttgtaaag cagaggcga gcatcatctc tcccaacttc agcttcatgg    | 1140 |
| gccagctgct gcagttttag tccctagggtc tggctccgca ctgttggca gaggctggga  | 1200 |
| <br>   |      |
| gccccggccat ggctgtgctc gaccgaggca cctccaccac caccgtttc aacttccccg  | 1260 |
| tctccatccc tgtccactcc acgaacagtgcgctgagacta cttcagagc cccattacga   | 1320 |

|   |      |
|---|------|
| cctctccag ctgctgaaag gccacggag gtgaggctc tcacatccc ttggactcc      | 1380 |
| atgctcctg agaggagaaa tgcaataact ctggagggg ctcgagaggg ctggccta     | 1440 |
| tttatttaac ttcacccgag ttcctctggg tttctaagca gttatggta tgacttagcg  | 1500 |
| tcaagacatt tgctgaactc agcacatcg ggaccaatat atagtggta catcaagtcc   | 1560 |
| atctgacaaa atgggcaga agagaaagga ctcagtgtg gatccggtt ctttgctc      | 1620 |
|   |      |
| gcccctgtt tttttagaat ctttcatgc ttgacatacc taccagtatt attccgacg    | 1680 |
| acacatatac atatgagaat ataccttatt tatttttgt tagtgtctg cttcacaaa    | 1740 |
| tgtcattgtc tactcctaga agaaccaa atacctaaat ttgttttga gtactgtact    | 1800 |
| atccgtaaa tatacttaa gcaggttgtt tttcagcact gaigggaaat accagtttgc   | 1860 |
| ggtttttt tagttgccaa cagttgtatg tttgctgatt atttatgacc tgaataata    | 1920 |
| tattttct tctaagaaga cattttgtt cataaggatg actttttat acaatggaaat    | 1980 |
| aaattatggc atttctatttgc   | 2000 |
|   |      |
| <210> 108   |      |
| <211> 5749  |      |
| <212> DNA   |      |
| <213> Homo sapiens  |      |
| <400> 108   |      |
|   |      |
| cgcgggagcc aacttcaggc tgctcagagg aagccgtgc agtcagtc ac ctgggtgcaa | 60   |
| gagcggtgct gcctcggtct ctcccgctgc agggagagcg gcactcgctg gcctggatgt | 120  |
| ggttggattt aggggggctc cgcagcagg gtttctggc ggtggcaagc gctgcaacag   | 180  |
| gtagacggcg agagacggac cccggccgag gcagggatgg agaccaaagg ctaccacagt | 240  |
| ctccctgaag gtctagatgg gaaagacgg tgggtcaag ttctcaggc tgtggagcgt    | 300  |
| tcttccctgg gacctacaga gaggaccat gagaataact acatggagat tgtcaacgt   | 360  |
|   |      |
| agctgtgtt ccgggtctat tccaaacaac agtactcaag gaagcagcaa agaaaaacaa  | 420  |
| gaactactcc ctgccttca gcaagacaat aatcgccctg ggattttaac atctgatatt  | 480  |
| aaaactgagc tggatctaa ggaacttca gcaactgttag ctgagttccat gggtttat   | 540  |
| atggattctg taagagatgc tgactattcc tatgagcagc agaaccaaca aggaagcatg | 600  |
| agtccagcta agatttatca gaatgttcaa cagctggta aatttacaa aggaaatggc   | 660  |
| catcgctt ccactctaag ttgtgtaac acgccttga gatcatttat gtctgactct     | 720  |
| gggagctccg tgaatggtg cgtcatgcgc gccattgtt aaaggccat catgtgtcat    | 780  |
|   |      |
| gagaaaagcc cgtctgtttg cagccctctg aacatgacat ctgcgtttt cagccctgt   | 840  |

|  |      |
|--|------|
| ggaatcaact ctgtgcctc caccacagcc agcttgca gtttccagt gcacagccca        | 900  |
| atcacccagg gaactcctct gacatgctcc cctaattgtc aaaatcgagg ctccaggctcg   | 960  |
| cacagccctg cacatgttagt caatgtggc tctccctct caagtccgtt aagtagcatg     | 1020 |
| aaatcctcaa ttccagccc tccaagtac tgcatgtaa aatctccagt ctccagtc         | 1080 |
| aataatgtca ctctgagatc ctctgtgtc agccctgcaaa atattaacaa ctcaagggtgc   | 1140 |
| tctgtttcca gcccttcgaa cactaataac agatccacgc ttccagtc ggcagccagt      | 1200 |
|  |      |
| actgtggat ctatctgttag ccctgtaaac aatgcctca gctacactgc ttctggcacc     | 1260 |
| tctgctggat ccagtacatt gcgggatgtg gttcccagt cagacacgca ggagaaaaggt    | 1320 |
| gctcaagagg tccctttcc taagacttagt gaagtagaga gtgccatctc aaatgggtg     | 1380 |
| actggccagc ttaatattgt ccagtacata aaaccagaac cagatggagc ttttagcagc    | 1440 |
| tcatgtctag gaggaaatag caaaaataat tcggattttt cattctcagt accaataaaag   | 1500 |
| caagaatcaa ccaaggattc atgttcaggc acctcttta aaggaaatcc aacagtaaac     | 1560 |
| ccgtttccat ttatggatgg ctcgtatTT tccttatgg atgataaaaga ctattattcc     | 1620 |
|  |      |
| ctatcaggaa ttttaggacc acctgtgccg ggcttgatg gtaactgtga aggccagcgga    | 1680 |
| ttcccaagtgg gtatcaaaca agaaccagat gacgggagct attaccaga ggccagcatc    | 1740 |
| ccttcctctg ctattgtgg ggtgaattca ggtggacagt cttccacta caggattgg       | 1800 |
| gctcaaggta caaatcttt atcacgatcg gctagagacc aatcttcca acacctgagt      | 1860 |
| tccttcctc ctgtcaatac tttatggag tcatggaaat cacacggcga cctgtcgct       | 1920 |
| agaagaagtg atgggtatcc ggtcttagaa tacattccag aaaatgtatc aagctctact    | 1980 |
| ttacgaagtg ttctactgg atcttcaaga cttcaaaaa tatgtttggt gtgtgggat       | 2040 |
|  |      |
| gaggcttcag gatgccatta tgggttagtc acctgtggca gctgcaaagt tttttcaaa     | 2100 |
| agagcagtgg aaggcaaca caactatTTA tgtgctggaa gaaatgattt catcattgtat    | 2160 |
| aagattcgcac gaaagaattt tcctgcttgc agacttcaga aatgtttca agctggaaatg   | 2220 |
| aatttaggag cacgaaagtgc aagaagtgg ggaaagttaa aaggattca cgaggagcag     | 2280 |
| ccacagcgc acgagcccccc accccccccccc ccaccccccgc aaagcccaga ggaagggaca | 2340 |
| acgtacatcg ctctgcaaa agaaccctcg gtcaacacag cactggttcc tcagctctcc     | 2400 |
| acaatctcac gagcgcac accttcccccc gttatggtcc ttgaaaacat tgaacctgaa     | 2460 |
|  |      |
| attgtatatg caggctatga cagtcaaaa ccagatacag ccgaaaatct gctctccacg     | 2520 |
| ctcaaccgct tagcaggcaaa acagatgtac caagtctgtca agtggcaaa ggtacttcca   | 2580 |
| ggatttaaaa acttgcctct tgaggacca attaccctaa tccagtattt ttggatgtgt     | 2640 |
| ctatcatcat ttgccttgag ctggagatcg tacaaacata cgaacagccaa atttcttat    | 2700 |

|   |      |
|---|------|
| tttgaccag acctagtctt taatgaagag aagatgcac agtctgccat gtatgaacta     | 2760 |
| tgcagggaa tgaccaaata cagccttcag ttgcgtcgac tgcaagtcac ctttgaagaa    | 2820 |
| tacaccatca taaaagttt gctgctacta agcacaattc caaaggatgg cctcaaaagc    | 2880 |
|   |      |
| caggctgcat ttgaagaaat gaggacaaat tacatcaaag aactgaggaa gatggtaact   | 2940 |
| aagtgtccca acaattctgg gcagagctgg cagaggatctt accaactgac caagctgctg  | 3000 |
| gactccatgc atgacctggt gagcgacctg ctggattct gttctacac cttccgagag     | 3060 |
| tcccatgcgc tgaaggtaga gttcccgca atgcgtgg agatcatcg cgaccagctg       | 3120 |
| cccaagggtgg agtcgggaa cgccaagccg ctctacttcc accggaagtg actgcccgt    | 3180 |
| gcccagaaga actttgcctt aagttccct gtgttgtcc acacccagaa ggacccaaga     | 3240 |
| aaacctgttt ttaacatgtg atggttgatt cacacttgtt caacagttc tcaagttaa     | 3300 |
|   |      |
| agtcatgtca gaggtttgga gccgggaaag ctgttttcc gtggatttg cgagaccaga     | 3360 |
| gcagtcgtcaa ggattccccca cctccaatcc cccagcgctt agaaacatgt tcctgttcc  | 3420 |
| cgggatgaaa agccatatct agtcaataac tctgatttg atatttcac agatggaaga     | 3480 |
| agtttaact atgcgtgtta gtttctggta tcgttcgtt gtttaaaag gttcaagga       | 3540 |
| ctaacgaacg tttaaagct tacccttggt ttgcacataa aacgtatagt caatatgggg    | 3600 |
| cattaatatt ctttgttat taaaaaaaca caaaaaaata ataaaaaaat atatacagat    | 3660 |
| tcctgttgta taataacaga actcgtggcg tgggcagca gctgcctctg agccctcgct    | 3720 |
|   |      |
| cgtccacggc cttctgcatac actggtatac acactcgta gctccattt cttatataat    | 3780 |
| tagaatggat aagatgatgt taaatgcctt gtttgattt ctgtatcta ttgtgttgc      | 3840 |
| tttacaata atttttgca gtctttgtct gtgtgtaca ttactgtatg tataaattat      | 3900 |
| gaaggacctg aaataaggta taaggatctt ttgtaaatga gacacataca aaaaaaatct   | 3960 |
| ttaatggta ataggatgaa tggaaagta ttttgaaag aattctat tgcgtggagac       | 4020 |
| tatataagta ctatcttgc taaacaagg taatttttt ttgtaaatgt caatgtccgt      | 4080 |
| catgcataat gaaccgttta cagtgtattt aagaaaggaa aagctgtgcc ttttttagct   | 4140 |
|   |      |
| tcatatctaa ttaccattta tttagtgc tctgttgtaa ataaccacac tggaaacctct    | 4200 |
| tcggttgtct tggaaaccttt ctacttttc tgcgtttttt gttttgttct tggctcccg    | 4260 |
| cttggggcat ttgtggact ccagcacgtt ttctggcttc tgcttcatacc tgctccatcg   | 4320 |
| gggaatgaca cactgcgggtg tctgcagtc ctggaaagggtg tcatttgaca acacatgtgg | 4380 |
| gagaggaggt cttggagtg ctgcagctt gggaaagcct gcctcggtt cctttccctc      | 4440 |
| tagaaggcaga accagctcta cgagagttagt actggaaact tgatggctca gagagcatct | 4500 |

|   |      |
|---|------|
| tttcctccca ttttagaaaa tcagatttc tcctgtggaa aaaaaaaaaatt ccatgcactc  | 4560 |
| <br>  |      |
| tctctctgtt aaagatcagc tattcccttc tgatcttgga aagagggtct gcactcctgg   | 4620 |
| aaccggcac aggaacgcac agatcatggc aggatgcgct gggacggccc atctggcaa     | 4680 |
| ggttcagtct gaatggcatg gagaccggga gatagagggg ttttagattt taaaaggtta   | 4740 |
| gttttaaaa ataagttta tacataaaca gttttggaga aaaattacag atcatataag     | 4800 |
| caagacagtgcactaaaat gttaattca ttaatctgtt tggtggcac tgatgcaatg       | 4860 |
| tatggcttt ctcttgcccc aaatcacaaa catatgtatc ttggggaaa ctaacaatat     | 4920 |
| gattgcacta aataaactac ttgaataga gccaaatata atctttaaa aatgatgata     | 4980 |
| <br>  |      |
| atcatcaggt ttactcagtg aaatcatatt aattatttc caaaatctaa aagctgtagc    | 5040 |
| tggagaagcc catggccacg aggaagcagc aattaattag atcaacactt ttctccaggg   | 5100 |
| ttcaccatgc aggcaacatt accttgtt tcaaaagaca cctgccttag tgcaagggga     | 5160 |
| aacctgtgaa agctgcactc agagggagga gtcttctta cataattgc aatttcagga     | 5220 |
| attnaattta taggcagatc tttaataca gtcaacttac ggtgcacagt aatatgaaag    | 5280 |
| ccacacttt aaggtataaa atacacagca tgcagactgg gagttgctag caaacaaatg    | 5340 |
| gcttacttac aaaaggcagct tttagttcag acttagttt tataaaatga gaattctgac   | 5400 |
| <br>  |      |
| ttacttaacc aggttgaaa tggagatggc ctgcattcagc tttttgtatt aacaaatgtt   | 5460 |
| ctggctttt gtgtgtctcc aggttaactt gcttigattaa acagcaaaagc catattctaa  | 5520 |
| attcaactgtt gaatgcctgt cccagtc当地 attgtctgtc tgcatttatt ttgtaccat    | 5580 |
| attgctctta aaaaatctgg ttggtagatc ttccataattc accaaaaagt tcatataatt  | 5640 |
| taaagaaaca ctaaattatgt taaaatgaa gcaatttata tctttatgca aaaacatatg   | 5700 |
| tctgtctttt caaaggactg taaggcattt acaataaatc cttaacttt               | 5749 |
| <br>  |      |
| <210> 109   |      |
| <211> 2062  |      |
| <212> DNA   |      |
| <br>  |      |
| <213> Homo sapiens  |      |
| <400> 109   |      |
| gtcagtcctt cctgttagccg ccggccgcgc cgcccgccgc cccctgtccca gcagctccgg | 60   |
| cgccacctcg ggccggcgtc tccggcgccc gggagccagg cgctgacggg cgccggcgccc  | 120  |
| ggccggcggc gctctgtcg ctgcactca ggctccggcg tctgcgttccc cccatggggc    | 180  |
| tggcctgcgg cgccctggcg ctctgagatt gtcactgctg ttccaaggc acacgcagag    | 240  |
| ggatttgaa ttccctggaga gttgccttg tgagaagctg gaaatatttcc ttcaattcc    | 300  |

|  |      |
|--|------|
| atctcttagt ttccatagg aacatcaaga aatcatgaac aactttggta atgaagagt      | 360  |
| <br>   |      |
| tgactgccac ttccatcgatg aaggtttac tgccaaggac attctggacc agaaaattaa    | 420  |
| tgaagtttct tcttcgtatg ataaggatgc cttctatgtg gcagacctgg gagacattct    | 480  |
| aaagaaacat ctgaggtggt taaaagctct ccctcgatc accccccttt atgcagtcaa     | 540  |
| atgtatgtatg agcaaagcca tcgtgaagac ccttgcgtct accgggacag gatttgactg   | 600  |
| tgcttagcaag actgaaatac agttggtgca gagtctgggg gtgcctccag agaggattat   | 660  |
| ctatgcaaat cttgtaaac aagtatctca aattaagtat gctgctaata atggagtcca     | 720  |
| gatgtactt tttatgtt aagttgatgtt gatgaaagtt gccagagcac atccaaagc       | 780  |
| <br>   |      |
| aaagttgggtt ttgcggattt ccactgatga ttccaaagca gtcgtcgatc tcagtgtgaa   | 840  |
| attcgggtgcc acgctcagaa ccagcaggct cctttggaa cggcgaaag agctaaatat     | 900  |
| cgttgtt ggtgcagct tccatgttagg aagcggctgt accgatcctg agacattcgt       | 960  |
| gcaggcaatc tctgtatgccc gctgtgttt tgacatgggg gctgagggtt gttcagcat     | 1020 |
| gtatctgctt gatattggcg gtggcttcc tggatctgag gaigtgaaac ttaaatttga     | 1080 |
| agagatcacc ggcgtaatca acccagcgatgg gacaaatac tttccgtcactg actctggagt | 1140 |
| gagaatcata gctgagcccg gcagatacta tggatcatca gcttcacgc ttgcagttaa     | 1200 |
| <br>   |      |
| tatcattgcc aagaaaattt tattaaagga acagacgggc tctgtatgacg aagatgagtc   | 1260 |
| gagtgagcag acctttatgtt attatgtaa tggatggcgatc tttatgtatgtt ttaatttgc | 1320 |
| actctatgac cacgcacatg taaagccct tctgcaaaag agacctaatac cagatgagaa    | 1380 |
| gtattattca tccagcatat ggggaccaac atgtgtatggc ctgcgtcgatc ttgttgcgc   | 1440 |
| ctgtgacactg cctgaaatgc atgtgggtga ttggatgtct tttgaaaaca tggcgctta    | 1500 |
| cactgttgc tgcgttgcacta cgttcaatgg cttccagagg ccgacgatct actatgtatgt  | 1560 |
| gtcaggccct gcgtggcaac tcatgcagca attccagaac cccgacttcc caccgcgat     | 1620 |
| <br>   |      |
| agaggaacag gatgccagca ccctgcgtgt gtcttgcgtcc tgggagagtg ggatgaaacg   | 1680 |
| ccacagagca gcctgtgtt cggctgttat taatgtgttag atagcactct ggttagcttt    | 1740 |
| aactgcaagt ttagcttggaa ttaagggtt tggggggacc atgtacttta attactgtct    | 1800 |
| gttttggaaat gtcttgcgtt gatgtgggtc gcatgtatgc agccatgttgg aagacttagga | 1860 |
| tatgggtcac acttatgtt gttcctatgg aaactatttg aatatttgc ttatgttgc       | 1920 |
| tttttatttgc ttttgcata cgttacttca gatgtgggtc cagctgtatgc acaagcattt   | 1980 |
| gtatgttgc ttttgcata cgttacttca gatgtgggtc cagctgtatgc acaagcattt     | 2040 |

|   |      |
|---|------|
| aagtatcttg aaataattag gc  | 2062 |
| <210> 110   |      |
| <211> 3155  |      |
| <212> DNA   |      |
| <213> Homo sapiens  |      |
| <400> 110   |      |
| gttcattt gcagttactg ggagggggct tgctgtggcc ctgtcaggaa gagtagagct       | 60   |
| ctggccagg tccgcgcagg gaggaggct gtcaccatgc cggcctgctg cagctgcagt       | 120  |
| gatgtttcc agtatgagac gaacaaagtc actcgatcc agagcatgaa ttatggcacc       | 180  |
| attaagtggt tcctcacgt gatcatctt tcctacgtt gcttgctct ggtgagtgac         | 240  |
| aagctgtacc agcggaaaga gcctgtcatc agttctgtgc acaccaaggt gaagggata      | 300  |
| <br>gcagaggtga aagaggagat cgtggagaat ggagtgaaga agttggtgca cagtgtctt  | 360  |
| gacaccccgactacacacctt cccttgcag gggactt tttcgatgat gacaaacttt         | 420  |
| ctcaaaacag aaggccaaga gcagcggtt tgcccggat atccaccccg caggacgctc       | 480  |
| tgttctctg accgagggtt taaaaaggga tggatggacc cgagagcaa aggaattcag       | 540  |
| accggaaggt gtgtatgtta tgaaggaaac cagaagacct gtgaagtctc tgccctggc      | 600  |
| cccatcgagg cagtggaaaga ggccccccgg cctgctct tgaacagtgc cgaaaacttc      | 660  |
| actgtgtca tcaagaacaa tatcgactt cccggccaca actacaccac gagaaacatc       | 720  |
| <br>ctgccaggtaaaacatcac ttgtacccctt cacaagactc agaatccaca gtgtccatt   | 780  |
| ttccgacttag gagacatctt ccgagaaaca ggcgataatt tttcagatgt ggcaattcag    | 840  |
| ggcggataaa tggcattga gatctactgg gactgcaacc tagaccgtt gttccatcac       | 900  |
| tgcgtccca aatacagttt ccgtgcctt gacgacaaga ccaccaacgt gtccttgac        | 960  |
| cctggctaca actttagata cgccaaatgtac tacaaggaaa acaatgtta gaaacggact    | 1020 |
| ctgataaaatcc tttcggtt ccgttttgc acctcgat tttggcaccgg aggaaaattt       | 1080 |
| gacattatcc agctgggtgt gtacatcgcc tcaaccctct cctacttcgg tctggccgt      | 1140 |
| <br>gtgttcatcg acttcctcat cgacacttac tccagtaact gctgtcgctc ccatatttat | 1200 |
| ccctgggtca agtgcgtca gcccgtgtg gtcaacgaat actactacag gaagaagtgc       | 1260 |
| gagttccatttggagccaaa gccgacatta aagtatgtt ctttggta tgaatccac          | 1320 |
| ataggatgg tgaaccagca gctacttaggg agaagtctgc aagatgtcaa gggccaaagaa    | 1380 |
| gtcccaagac ctgcgtggat cttcacatgtt tgccctggc cctccatgac                | 1440 |
| acaccccgat ttcctggaca accagaggat atacagctgc tttagaaagga ggcgactcct    | 1500 |

|  |      |
|--|------|
| agatccaggg atagccccgt ctgggccag tgtggaagct gcctcccatc tcaactccct   | 1560 |
|  |      |
| gagagccaca ggtgcctgga ggagctgtgc tgccggaaaa agccggggc ctgcatcacc   | 1620 |
| acctcagagc tttcaggaa gctggctcg tccagacacg tcctgcagtt cctctgctc     | 1680 |
| taccaggagc ctttgcgtgc gctggatgtg gattccacca acagccggct gcggcactgt  | 1740 |
| gcctacagggt gctacggcac ctggcgcttc ggctcccagg acatggctga ctggccaaac | 1800 |
| ctgcccagct gctggcgctg gaggatccgg aaagagttc cgaagagtga agggcagttac  | 1860 |
| agtggcttca agagtcccta ctgaagccag gcaccgtggc tcacgtctgt aatcccagcg  | 1920 |
| ctttgggagg ccgaggcagg cagatcacct gaggtcgaaa gttggagacc cgcctggcta  | 1980 |
|  |      |
| acaaggcgaa atcctgtctg tactaaaaat acaaaaatca gccagacatg gtggcatgca  | 2040 |
| cctgcaatcc cagctactcg ggaggcttag gcacaagaat cacttgaacc cgggaggcag  | 2100 |
| aggttgttagt gagcccgat tggccactg ctctccagcc tgggaggcac agcaaactgt   | 2160 |
| cccccaaaaa aaaaaaagag tccttaccaa tagcagggc tgcaatggcc atgttaacat   | 2220 |
| gacatttacc agcaacttga acttcacctg caaagctctg tggccacatt ttcagccaaa  | 2280 |
| ggaaatatg ctttcatctt ctgttgctct ctgtgtctga gagcaaaatgt acctggtaa   | 2340 |
| acaaaccaga atccctctac atggacttag agaaaagaga ttgagatgtta agtctcaact | 2400 |
|  |      |
| ctgtccccag gaagttgtgt gacccttaggc ctctcaccc tggccctctg tctccctgtt  | 2460 |
| cccccaactac tatctcagag atattgttag gacaaattga gacagtgcac atgaactgtc | 2520 |
| ttttaatgtg taaagatcta catgaatgca aaacattca ttatgaggc agactaggat    | 2580 |
| aatgtccaac taaaaacaaa ccctttcat cctggctgga gaatgtggag aactaaaggt   | 2640 |
| ggccacaatat tcttgacac tcaagtcccc caagacctaa gggtttatac tcctccctt   | 2700 |
| gaatatgggt ggctctgatt gctttatcca aaagtggaaag tgacatgtg tcagttcag   | 2760 |
| atcctgatct taaggcgtg acagttcta cttgctgtcc cttggaaactc ttgctatcg    | 2820 |
|  |      |
| ggaagccaga cggcattaa aagtctgcct atcctggcca ggtgtgggtgg ctcacacctg  | 2880 |
| taatccage acttgggag accaaggcgg gggatcaact taaagttagg agtccaagac    | 2940 |
| cagactcgcc aacatggta aaccgtatct ctaataaaaa tacaaaaatt agctggcat    | 3000 |
| ggtgccggca cctgttagtcc tagctatcaa gaggctgaga caggagaaac acttgaacct | 3060 |
| gggaggtgga ggttgcatgg agctgagatc gtgccactgc actccaggct gggtagacaga | 3120 |
| gcgagactcc atctaaaaaa aaaaaaaaaaag aaaaaa                          | 3155 |
| <210> 111  |      |
| <211> 871  |      |

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 111

|  |     |
|--|-----|
| ctgccaggca gtgcccttcc cggagcgtgc cctcgccgct gagctcccct             | 60  |
| gaacagcagc   |     |
| tgcagcagcc atggccccgc cctgggtgcc cgccatggc ttcacgctgg cgccagcct    | 120 |
| gggggtgcttc gtgggctccc gctttgtcca cggcgagggt ctccgctggt acgcccggct | 180 |
| gcagaagccc tcgtggcacc cgcggactg ggtgctggc cctgtctgg gcacgctcta     | 240 |
| ctcagccatg ggtaacggct cctacctggt ctggaaagag ctggaggagct tcacagagaa | 300 |
| ggctgttgtt cccctggcc tctacactgg gcagctggc ctgaactggg catggcccc     | 360 |

|   |     |
|---|-----|
| catttcttt ggtgcccggac aaatgggctg ggccttggtg gatctcctgc tggtcagtgg   | 420 |
| ggcggcggca gccactaccc tggcctggta ccaggtgagc ccgctggccg cccgcctgct   | 480 |
| ctaccctac ctggcctggc tggccttcgc gaccacactc aactactgca tatggcggga    | 540 |
| caaccatggc tggcatgggg gacggcggtt gccagagtga gtgcccggcc caccaggac    | 600 |
| tgcagctgca ccagcaggtt ccatcacgt tgtatgtgg tggccgtcac gcttcatga      | 660 |
| ccactgggcc tgcttagtctg tcagggcattt ggcccagggg tcagcagagc ttcagaggaa | 720 |
| gccccacactg agccccacc cgggagcagt gtcctgtgtt ttctgcatgc ttagagcatg   | 780 |

|   |     |
|---|-----|
| ttcttggAAC atggAAATTT ataagctgaa taaagTTTT gacttcTTT aaaaaaaaaa | 840 |
| aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaa a                          | 871 |

&lt;210&gt; 112

&lt;211&gt; 2133

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 112

|  |     |
|--|-----|
| atgcgtgtcct tccagtaccc cgacgtgtac cgcgacgaga ccgcataca ggatttatcat | 60  |
| ggtcataaaa ttgtgaccc ttacgcctgg cttgaagacc ccgacagtga acagactaag   | 120 |
| gcctttgtgg aggcccagaa taagattact gtgccatttc ttgagcagtg tcccatcaga  | 180 |
| gttttataca aagagagaat gactgaacta tatgattatc ccaagtatag ttgccacttc  | 240 |

|   |     |
|---|-----|
| aagaaaggaa aacggtattt ttatTTTAC aatacagggtt tgcagaacca gcgagtatta   | 300 |
| tatgtacagg attccttaga gggtgaggcc agagtgtcc tggaccccaa catactgtct    | 360 |
| gacgatggca cagtggcaact ccgaggatgtat gcgttcagcg aagatggtga atatTTGCC | 420 |
| tatggctgtga gtgccagtgg ctcagactgg gtgacaatca agttcatgaa agttgtatgg  | 480 |

|   |      |
|---|------|
| gccaaagagc ttccagatgt gcttcaaaga gtcaaggtaa gctgtatggc ctggaccat        | 540  |
| gatggaaagg gaatgttcta caactcatac cctcaacagg atggaaaag tcatggcaca        | 600  |
| gagacatcta ccaatctcca ccaaagctc tactaccatg tcttggAAC cgatcagtca         | 660  |
| <br>  |      |
| gaagatattt tgtgtgctga gtttcctgtat gaacctaaat ggatgggtgg agctgagtt       | 720  |
| tctgtatgt gctgttatgt ctgttatca ataaggaaag gatgtatcc agtaaacgaa          | 780  |
| ctctggtaact gtgacctaca gcaggaatcc agtggcatcg cggaatcct gaagtggta        | 840  |
| aaactgattt acaacttga agggaaatgactacgtga ccaatgaggg ggcgggtttc           | 900  |
| acattcaaga cgaatcgcca gtctccaaac tatcgctga tcaacattga cttcaggat         | 960  |
| cctgaagagt ctaagtggaa agtacttgtt cctgagcatg agaaagatgt cttagaatgg       | 1020 |
| atagcttgc tcaggtccaa cttcttggtc ttatgttacc tccatgacgt caagaacatt        | 1080 |
| <br>  |      |
| ctgcagctcc atgacctgac tactgggtct ctccttaaga cttcccgct cgtatgtcgcc       | 1140 |
| agcattgttag ggtacagcggtcagaagaag gacactgaaa tcttctatca gtttacttcc       | 1200 |
| tttttatctc caggtatcat ttatctgtt gatctgacca aagaggagct ggagccaaga        | 1260 |
| gtctcccgcg aggtgaccgt gaaaggaaatt gatgttctg actaccagac agtccagtt        | 1320 |
| ttctacccta gcaaggatgg tacgaagatt ccaatgttca ttgtgcataaa aaaaagcata      | 1380 |
| aaattggatg gctctcatcc agctttcttata gatgttgcataa gatgttgcataa catatccatc | 1440 |
| acacccaact acagtgttccaggcttattttgtgagac acatgggtgg tatcctggca           | 1500 |
| <br>  |      |
| gtggccaaca tcagaggagg tggcgaatat ggagagacgt ggcataaagg tggtatcttgc      | 1560 |
| cccaacaac aaaaactgctt tgatgacttt cagtggtctg ctgagtatct gatcaaggaa       | 1620 |
| ggttacacat ctcccaagag gctgactatt aatggaggtt caaatggagg cctcttagtg       | 1680 |
| gctgcttgc caaatcagag acctgaccc tttgggttg ttattgccc agttggagta           | 1740 |
| atggacatgc tgaagttca taaatatacc atcgccatg ctggaccac tgattatgg           | 1800 |
| tgctcgacca gcaaacaaca ctttgaatgg ctgtcaaat actctccatt gcataatgtg        | 1860 |
| aagtaccatc aagcagatga catccagttac ccgtccatgc tgctccatc tgctgaccat       | 1920 |
| <br>  |      |
| gatgaccgacg tggtcccgct tcactccctg aagttcatttccacccttca gtacatcg         | 1980 |
| ggccgcacca ggaagcaaag caacccctg cttatccacg tggacacccaa ggcggccac        | 2040 |
| ggggcgggaa agcccacagc caaagtgtata gaggaagtct cagacatgtt tgcgttcatc      | 2100 |
| gcgcgggtgcc tgaacatcgatcgattccg taa                                     | 2133 |
| <210> 113   |      |
| <211> 1375  |      |
| <212> DNA   |      |

&lt;213&gt; Homo sapiens

&lt;400&gt; 113

|  |     |
|--|-----|
| gcaaacagcc gggctccag cgggagaacg ataatgaaa gtgctatgtt cttggcttt   | 60  |
| caacacgact gcagacccat ggacaagagc gcaggcagt gccacaagag cgaggagaag | 120 |

|  |     |
|--|-----|
| cgagaaaaga tgaaacggac cttttaaaaa gattggaaa cccgttgag ctacttta      | 180 |
| caaaattcct ctactcctgg gaagccccaa accggcaaaa aaagcaaca gcaagcttc    | 240 |
| atcaagcctt ctccctgagga agcacagctg tggtcagaag catttgacga gctgttagcc | 300 |
| agcaaatacg gtcttgctgc attcaggctt tttttaagt cgaaattctg tgaagaaaat   | 360 |
| attgaattct ggctggcctg tgaagactc aaaaaaccca aatcacccca aaagctgtcc   | 420 |
| tcaaaagcaa gaaaaatata tactgacttc atagaaaagg aagctccaaa agagataaac  | 480 |
| atagattttc aaacccaaac tctgattgcc cagaatatac aagaagctac aagtggctgc  | 540 |

|  |     |
|--|-----|
| tttacaactg cccagaaaag ggtatacagc ttgtatggaga acaactctta tcctcgttc  | 600 |
| ttggagttag aattctacca ggacttgtgt aaaaagccac aaatcaccac agaggctcat  | 660 |
| gctacatgaa atgtaaaagg gagcccgaaa atggaggaca tttcattttt tttcctgagg  | 720 |
| ggaaggactg tgacctgcca taaagactga ctttgaattc agcctgggtg ttcagggaaac | 780 |
| atcaactcaga actattgatt caaatgtgg tagtgaatca ggaagccagt aactgactag  | 840 |
| gagaagctgg tatcagaaca gttccctca ctgtgtacag aacgcaagaa gggataggt    | 900 |
| ggctgtacg tgggtctca ctctgaaaag caggaatgta agatgtgaa agagacaatg     | 960 |

|   |      |
|---|------|
| taatactgtt ggtccaaaag catttaaaat caatagatct gggattatgt ggccttaggt | 1020 |
| agctgggtgt acatcttcc ctaaatcgat ccatgttacc acatagtagt ttttagtttag | 1080 |
| gattcgttaa cagtgaagtg ttactatgt gcaacggtat tgaagtctt atgaccacag   | 1140 |
| atcatcgat ctgttgtctc atgtaatgct aaaactgaaa tggccgtgt ttgcattttt   | 1200 |
| aaaaatgtatg tggaaatag aatgagtgtc atgggtttaa aaactgcagt gtccgttatg | 1260 |
| agtgcacaaa atctgtctt aaggcagcta cacttgaag tggctttga atactttaa     | 1320 |
| taaattttt ttgataaata atattgaaca aaaaaaaaaa aaaaaaaaaa aaaaa       | 1375 |

&lt;210&gt; 114

&lt;211&gt; 1069

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 114

|  |    |
|--|----|
| aagtaattcc tagacccgta ggtggccgca gagccggta cctctggttc tgcgccagcg | 60 |
|--|----|

|            |              |            |             |            |            |      |
|------------|--------------|------------|-------------|------------|------------|------|
| tgccccaccc | gcaggacggc   | cgggttctt  | gatttgtaca  | cttctaaaa  | ccaaaccga  | 120  |
| gaggaagggc | aggctcaggg   | tggggatgcc | ctgaataatt  | cgagagcagg | accgttcta  | 180  |
| ctgaagagaa | gttacaaga    | acgctctgtc | tggggcgggc  | gaggcctctg | cgaggcgggt | 240  |
| ccgggagcga | ggcaggcgc    | tggccgcgc  | gcccgggtc   | ggggagtcg  | ggggcagaa  | 300  |
| gagggggagg | agacaggcgt   | ggggagcgc  | cctgcccagc  | gccgcaggc  | ctcccccgc  | 360  |
|            |              |            |             |            |            |      |
| tcccgcccg  | cctccctcta   | cccacccgcc | gcacgtacta  | aggaaggcgc | acagccgc   | 420  |
| gcgcgcgcct | ctccgccccg   | cgtccagctc | gcccaagctcg | cccagcgtcc | gccgcgcctc | 480  |
| ggccaaggct | tcaacggacc   | acaccaaaat | gccatctcaa  | atgaaacacg | ccatggaaac | 540  |
| catgatttt  | acattcaca    | aattcgctgg | ggataaaggc  | tacttaacaa | aggaggacct | 600  |
| gagagtactc | atggaaaagg   | agttccctgg | attttgaa    | aatcaaaaag | accctctggc | 660  |
| tgtggacaaa | ataatgaagg   | acctggacca | gtgttagagat | ggcaaagtgg | gcttccagag | 720  |
| cttctttcc  | ctaattgcgg   | gcctcaccat | tgcatgcaat  | gactatttg  | tagtacacat | 780  |
|            |              |            |             |            |            |      |
| gaagcagaag | gaaaagaagt   | aggcagaaat | gagcagttcg  | ctctccctg  | ataagagttg | 840  |
| tcccaaagg  | tcgcttaagg   | aatctgcccc | acagttccc   | ccatagaagg | atttcatgag | 900  |
| cagatcagga | cacttagcaa   | atgtaaaaat | aaaatcta    | tctcatttga | caagcagaga | 960  |
| aagaaaagtt | aaataccaga   | taagctttt  | attttgat    | tgttgcatc  | ccctgcct   | 1020 |
| caataaataa | agttttttt    | tagtccaaa  | tttggaaaaa  | aaaaaaa    | aaaaaaa    | 1069 |
| <210>      | 115          |            |             |            |            |      |
| <211>      | 4535         |            |             |            |            |      |
| <212>      | DNA          |            |             |            |            |      |
| <213>      | Homo sapiens |            |             |            |            |      |
| <400>      | 115          |            |             |            |            |      |
| ggtccctccc | ctccctggctc  | tgggtcggt  | cgcgcacccc  | gccccgtagc | gcggccctc  | 60   |
|            |              |            |             |            |            |      |
| cctggcgagc | gcaacccat    | ccagcggag  | cgcggagccg  | cggccgcgg  | gaagcattaa | 120  |
| gtttattcgc | ctcaaagtga   | cgcaaaaatt | cttcaagac   | tctttggcgg | cggctatcta | 180  |
| gagatcagac | catgtgaggg   | ccgcggta   | caaatacggc  | cgcgcggcg  | ccctccgca  | 240  |
| cagccagcgc | cgcgggtgc    | ctcgaggcgc | cgaggccagc  | ccgcctgc   | agcccgac   | 300  |
| cagcctcccc | gcmcgcctg    | gcaggtctcc | tggaggcaag  | gcgcaccc   | ttgcctctc  | 360  |
| ttgcagaata | acaagggct    | tagccacagg | agttgtggc   | aagtggaaag | aagaacaat  | 420  |
| gagtcaatcc | cgcacgtgtca  | atccgcacga | tagagagctc  | ggaggtgtac | cacaatcca  | 480  |
|            |              |            |             |            |            |      |
| agcacccaga | gatcaattgg   | gatccttggc | agatggacat  | cagtgtcatt | tactaaccag | 540  |

|   |      |
|---|------|
| caggatggag acgacgcct tgaattctca gaagcagcta tcagcgtgt aagatggaga       | 600  |
| agattgtcag gaaaacggag ttctacagaa ggttgtccc accccaggaa acaaagtgga      | 660  |
| gtccggcaa atatccaatg ggtactcagc agttccaagt cctggtgcgg gagatgacac      | 720  |
| acggcactct atcccagcga ccaccaccc cctagtggtc gagcttcata aagggaaacg      | 780  |
| ggagacctgg ggcaagaagg tggatttcct tctctcagtg attggctatg ctgtggacct     | 840  |
| ggccaatgtc tggcgcttcc cctacatatg ttaccagaat ggaggggggg cattcctct      | 900  |
|   |      |
| ccccctacacc atcatggcca tttttgggg aatccgcctc ttttacatgg agctcgact      | 960  |
| gggacagtac caccgaaatg gatgcatttc aatatggagg aaaatctgcc cgatttcaa      | 1020 |
| agggtatggta taigccatct gcatcatatgc cttttacatt gtttccatc acaacaccat    | 1080 |
| catggcctgg gcgtataact acctcatctc ctccttcacg gaccagctgc cctggaccag     | 1140 |
| ctgcaagaac tcctggaaaca ctggcaactg caccaattac ttctccgagg acaacatcac    | 1200 |
| ctggaccctc cattccacgt cccctgctga agaattttac acgcgcacg tcctgcagat      | 1260 |
| ccaccggctc aaggggctcc aggacctggg gggcatcagc tggcagctgg ccctctgcat     | 1320 |
|   |      |
| catgtgtatc ttcaactgtta tctacttcag catctggaaa gggtcaaga cctctggca      | 1380 |
| ggtgtgtgg gtgacagcca cttccctta tatcatcctt tctgtcctgc tggtgagggg       | 1440 |
| tgcaccctc cttggagcct ggaggggtgt tctttctac ttgaaaccca attggcagaa       | 1500 |
| actcctggag acaggggtgt ggatagatgc agccgctcag atcttcttct ctcttggcc      | 1560 |
| gggccttggg gtcctgctgg cttttcttag ctacaacaag ttcaacaaca actgtacca      | 1620 |
| agatgcctg gtgaccagcg tggtaactg catgacgagc ttcttcgg gatttgtcat         | 1680 |
| ttcacatgt ctgcgttaca tggctgagat gaggaatgaa gatgtgtctg aggtggccaa      | 1740 |
|   |      |
| agacgcaggt cccagcctcc ttcatcac gtatgcagaa gcgtatgcc acatgccagc        | 1800 |
| gtccactttc ttgcctca tttttttct gatgttaatc acgtggct tggacagcac          | 1860 |
| gtttgcaggc ttggaggggg tgcacggc tggcgtggat gagttccac acgtctggc         | 1920 |
| caagcgcgg gacgcgttcg tgctgcgtt ggtcatcacc tggttttg gatccctgg          | 1980 |
| caccctgact ttggaggggg cttacgtgtt gaagctgttg gaggagatg ccacggggcc      | 2040 |
| cgcagtgtc actgtcgccg tgcacggc agtgcgttg tcttggttct atggcatcac         | 2100 |
| tcaatgtc agggacgtga agaaatgtt cggcttcagc cgggggtgt tctggaggat         | 2160 |
|   |      |
| ctgctgggtt gccatcagcc ctgtttctt cctgttcatc atttgcatgtt ttctgtatg      | 2220 |
| ccgcaccaa ctacgacttt tccatataa ttatccttac tggatgtatca tcttgggtt       | 2280 |
| ctgcataatggc acctcatctt tcatttgcattt cccacatata atagtttata ggttgcatt  | 2340 |
| cactccaggc acattnaaag agcgtatttata taaaagtattt accccagaaaa caccaacaga | 2400 |

|  |      |
|--|------|
| aattccttgt gggacatcc gcttgaatgc tgtgtAACAC actcaccgag agaaaaagg    | 2460 |
| cttctccaca acctcctcct ccagttctga tgaggcacgc ctgccttc ccctccaagt    | 2520 |
| gaatgagttt ccagctaAGC ctgatgatgg aaggccttc tccacaggga cacagtctgg   | 2580 |
| <br>   |      |
| tgcccagact caaggcctcc agccacttat ttccatggat tcccctggac atattccat   | 2640 |
| ggttagactgt gacacagctg agctggccta ttttggacgt gtgaggatgt ggtggaggt  | 2700 |
| gatgaaaacc accatatcat cagtaggat tagtttaga atcaagtctg tgaaagtctc    | 2760 |
| ctgtatcatt tcttggatg atcattggta tctgataatct gtttgcttct aaaggttca   | 2820 |
| ctgttcatga atacgtaaac tgcgttaggag agaacaggga tgctatctcg ctggccatat | 2880 |
| attttctgag tagcatatat aattttatig ctggaatcta ctagaacctt ctaatccatg  | 2940 |
| tgctgctgtg gcatcaggaa aggaagatgt aagaagctaa aatgaaaaat agtgtgtcca  | 3000 |
| <br>   |      |
| tgcaagcttg tgagtctgtg tatatttttg tttcagtgtat ttcttatctc tagtccaata | 3060 |
| ttttggccc attacaata tatgaattcc ccaaattttt cttacattaa caaattctac    | 3120 |
| caactcaatt gtgtatggag gttattttt gaagggtaca atcactacaa catgctctgc   | 3180 |
| caccactcc tttccagtg acactacttg agccacacac tttcctttac aggccagcct    | 3240 |
| ctggcgttt ctgcacctca ttgccacctt cctgtctctc tgtgctaaac attcaggaca   | 3300 |
| gtgttccaca ggcagatctg gcctattca ttagtcacca tggcttgct gtgaagtacg    | 3360 |
| ttgaagggtgg atcttgcac atgccccttc agtgttacc tggccctctg gtttaagttc   | 3420 |
| <br>   |      |
| tgtctgcctt acgtgactga gtttgactgt ccaggttgc ttgctcggtg aagagaggag   | 3480 |
| ggtaaatcg attctcgttt agcactgggt tatacagatc tggcaccccta acctaaacca  | 3540 |
| aggcatcttc actccaagag cagttggaga gtctgggtta gccttacgtg gacctcgccg  | 3600 |
| ctcgctggcg gtcacgattt tgagccctcc agataatttt taaggttgag tctaagtaag  | 3660 |
| gctgcttggg aaatggtcag ctaagtaat caccttcat ttcacataag gcccttaata    | 3720 |
| tagataagta aatttggcct ttgggtctc gtgactctca gaggcgtagg tagaggagca   | 3780 |
| aattaatatt tgacgcatgg gaattcctta tcagaatttt gagggaaata aatccatc    | 3840 |
| <br>   |      |
| agagacaAAA ggacttaatc atctggccac ctatcacttc agttctctgt ataaatgaaa  | 3900 |
| tttaattcta acaaccttat aaaaagaagg tccagacagc agaggaaaca tcctgtccaa  | 3960 |
| ttcttaggttt tcctcccttg gcctcccttc cccagcatgt tctaccctgg cccacttcct | 4020 |
| gcattctccc catgcctgc tatttctgtat tctttgcttc tcttagcgag atacttcct   | 4080 |
| tatatgatag ctgctgagaa gttcccaga actgctagag gaaaagaagt ggggaattta   | 4140 |
| ggaatatacc ctcaactgacc taactccatt atcttactc tttccttc cctggccacct   | 4200 |

|  |      |
|--|------|
| catgccatt ctcttactg tctagcatgc tgaaaagg aagtgatcta aatgccagcg      | 4260 |
| <br>   |      |
| tgttcagtgg taaatattag ttggtgcaaa agaaaaacca tgattacttt tgcaactaacc | 4320 |
| taatagcttt gcaaattttt agaacttgc ttatgaagat attcggatat ggattctccc   | 4380 |
| caccccacat acttagacat tgttcaaata tactacttt aaaaaaacac ctttcaaac    | 4440 |
| agaattagcg tttgccaag tctggattt atgaaattgt acaggagctt tgaaagttt     | 4500 |
| caaactttat taaactaaaa aaaaaaaatc gaaaa                             | 4535 |
| <210> 116  |      |
| <211> 1800   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 116  |      |
| <br>   |      |
| actgcgaccc ggagccgccc ggactgacgg agcccactgc ggtgcggcgc ttggcgcccc  | 60   |
| <br>   |      |
| cacggaggac ccgggcaggc atcgcaagcg accccgagcg gagcccccgg accatggccc  | 120  |
| tgagcgagct ggcgtggc cgctggctgc aggagagccg ccgctcgccgg aagctcatcc   | 180  |
| tgttcatcggt gtctctggcg ctgctgctgg acaacatgct gctactgtc gtggtcccc   | 240  |
| tcatcccaag ttatctgtac agcattaaggc atgagaagaa tgctacagaa atccagacgg | 300  |
| ccagggccagt gcacactgcc tccatctcg acagcttcca gagcatcttc tcctattatg  | 360  |
| ataactcgac tatggtcacc gggaatgcta ccagagacct gacacttcat cagaccgcca  | 420  |
| cacagcacat ggtgaccaac gcgtccgcgtt ttcctccga ctgtcccgat gaagacaag   | 480  |
| <br>   |      |
| acctcctgaa tgaaaacgtg caagttggc tgggtttgc ctgcggaaagcc accgtccagc  | 540  |
| tcatcaccaa cccttcata ggactactga ccaacagaat tggctatcca attccatat    | 600  |
| ttgcgggatt ctgcatcatg ttgtctcaa caattatgtt tgccttctcc agcagctatg   | 660  |
| ccttcctgct gattgccagg tcgctgcagg gcatcggtc gcctgtcc tctgtggctg     | 720  |
| ggatggccat gcttgcaggat gtctacacag atgatgaaga gagaggcaac gtcatggaa  | 780  |
| tgccttggg aggctggcc atgggggtct tagtggccc ccccttggg agtgtgtct       | 840  |
| atgagttgt gggaaagacg gtcgggttcc tgggtgtggc cggccctggta ctcttggatg  | 900  |
| <br>   |      |
| gagctattca gcttttgc ctccagccgt cccgggtgca gccagagagt cagaagggga    | 960  |
| cacccctaac cacgctgctg aaggacccgt acatctcat tgctgcaggc tccatctgtc   | 1020 |
| ttgcaaacat gggcatcgcc atgctggagc cagccctgcc catctggatg atggagacca  | 1080 |
| tgtgtcccg aaagtggcag ctgggcgttgc ccttcttgc agcttagtata tcttatctca  | 1140 |
| ttggaaccaa tattttggg atacttgcac aaaaaatggg gaggtggctt tgtgtcttc    | 1200 |

|  |      |
|--|------|
| tgggaatgat aattgttgga gtcagcatt tatgtattcc attgcaaaa aacattatg     | 1260 |
| gactcatagc tccgaactt ggagttgggt ttgcaattgg aatggtggat tcgtcaatga   | 1320 |
| <br>   |      |
| tgcctatcat gggcacctc gtagacctgc ggcacgtgtc cgcttatggg agtgtgtacg   | 1380 |
| ccattgcgga tgtggcattt tgtatgggt atgctatagg tccttcgtc ggtggtgcta    | 1440 |
| ttgcaaaggc aattggattt ccatggctca tgacaattat tggataatt gatattttt    | 1500 |
| ttgcccctct ctgtttttt ctgcgaagtc cacctgcca agaagaaaaa atggctattc    | 1560 |
| tcatggatca caactgccct attaaaacaa aaatgtacac tcagaataat atccagtcat  | 1620 |
| atccgatagg tgaagatgaa gaatctgaaa gtgactgaga tgagatctc aaaaatcatc   | 1680 |
| aaagtgttta atigtataaa acagtgtttc cagtgacaca actcatccag aactgttta   | 1740 |
| <br>   |      |
| gtcataccat ccatccctgg taaaagagta aaaccaaagg ttattatttc cttccatgg   | 1800 |
|  | 1800 |
| <210> 117  |      |
| <211> 1852   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 117  |      |
| accggcgaga ccgcgtccgc cccgcgagca cagagcctcg ccttgcga tccggcccc     | 60   |
| gtccacaccc gcccgcagct caccatggat gatgatatcg ccgcgcgtcgt cgtcgacaac | 120  |
| ggctccggca tggcaaggc cggcttcgat ggcgacgatg ccccccggc cgtttcccc     | 180  |
| tccatgtgg ggcgcggccag gcaccaggc gtgatgggtt gcatgggtca gaaggattcc   | 240  |
| <br>   |      |
| tatgtggcg acgaggccca gagcaagaga ggcattctca ccctgaagta ccccatcgag   | 300  |
| cacggcatcg tcaccaactg ggacgacatg gagaaaatct ggcaccacac ctttacaat   | 360  |
| gagctgcgtg tggctccgat ggagcacccc gtgcgtcgt ccgaggcccc cctgaacccc   | 420  |
| aaggccaaacc gcgagaagat gacccagatc atgtttgaga cttcaacac cccagccatg  | 480  |
| tacgttgcata tccaggctgt gctatccctg tacgcctcg gccgttaccac tggcatcgat | 540  |
| atggactccg gtgacgggtt cacccacact gtgcccacatc acgaggggtt tgccctcccc | 600  |
| catgccatcc tgcgtctgga cctggctggc cgggacactga ctgactacat catgaagatc | 660  |
| <br>   |      |
| ctcacggcgc gggctacag cttcaccacc acggccgagc gggaaatcgt gcgtgacatt   | 720  |
| aaggagaagc tgtgtacgt cgcctggac ttgcgacaa agatggccac ggctgcttcc     | 780  |
| agctccctccc tggagaagag ctacgagctg cctgcggcc agtcatcac cattggaat    | 840  |
| gagcggttcc gctgcccgtga ggcactttc cagccttctt tcctggcat ggagtctgt    | 900  |

ggcatccacg aaactacctt caactccatc atgaagtgt acgtggacat ccgcaaagac 960  
 ctgtaccca acacagtgcgt gtctggcgcc accaccatgt accctggcat tgccgacagg 1020  
 atgcagaagg agatcaactgc cctggcaccc agcacaatga agatcaagat cattgctct 1080

cctgagcgca agtactccgt gtggatcgcc ggctccatcc tggctcgct gtccacacctc 1140  
 cagcagatgt ggtcagcaaa gcaggagttt gacgagtcggcc gcccctccat cgccacccgc 1200  
 aaatgcttctt aggccggacta tgacttagtt gcgttacacc ctttcttgac aaaacctaacc 1260  
 ttgcgcagaa aacaagatga gattggcatg gctttatttg tttttttgt tttttttgg 1320  
 tttttttttt tttttggct tgactcagga tttaaaaact ggaacggta aggtgacagc 1380  
 agtcggttgg agcgagcatc ccccaaaggta cacaatgtgg ccgaggactt tgatgtcaca 1440  
 ttgttgtttt tttaatagtc attccaaata tgagatgcgt tggtacagga agtccctgc 1500

catcctaaaa gccacccac ttctctctaa ggagaatggc ccagtcctct cccaaatcca 1560  
 cacaggggag gtgatagcat tgcttcgtg taaattatgt aatgaaaaat ttttttaatc 1620  
 ttcgccttaa tacttttta ttttggatga tgacccctcg tgccccccct 1680  
 tccccctttt ttgtccccca acttgagatg tatgaaggct tttggctcc ctgggagtgg 1740  
 gtggaggcag ccagggctta cctgtacact gacttgagac cagttgaata aaagtgcaca 1800  
 ccttaaaaat gaaaaaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aa 1852

&lt;210&gt; 118

&lt;211&gt; 987

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 118

aatataagtg gaggcgctc gctggcgccg attccctgaag ctgacagcat tcgggccgag 60  
 atgtctcgct ccgtggccctt agctgtgtc gcgtactct ctctttctgg cctggaggct 120  
 atccagcgta ctccaaagat tcaggtttac tcacgtcatc cagcagagaa tggaaatgtca 180  
 aatttcctga attgttatgt gtctgggttt catccatccg acattgaagt tgacttactg 240  
 aagaatggag agagaattga aaaagtggag cattcagact tgtcttcag caaggactgg 300  
 tctttctatc tcttgtacta cactgaattc acccccactg aaaaagatga gtatgcctgc 360

cgtgtgaacc atgtgacttt gtcacagccc aagatagttt agtggatcg agacatgtaa 420  
 gcagcatcat ggagggttta agatgccca tttggatgg atgaattcca aatttgcctt 480  
 gcttgctttt taatattgtatgccttatac acttacactt tatgcacaaa atgtagggtt 540  
 ataataatgt taacatggac atgatcttct ttataattct actttggatgc ctgtctccat 600

|   |      |
|---|------|
| gtttgatgta tctgagcagg ttgctccaca ggtagctca ggagggctgg caacttagag    | 660  |
| gtggggagca gagaattctc ttatccaaca tcaacatctt ggtcagattt gaactcttca   | 720  |
| atctcttgca ctcaaagctt gttaagatag ttaagcgtgc ataagtaac ttccaattta    | 780  |
| <br>  |      |
| catactctgc ttagaatttg gggaaaatt tagaaatata attgacagga ttattggaaa    | 840  |
| tttgttataa tgaatgaaac atttgtcat ataagattca tattnacttc ttatacattt    | 900  |
| gataaagtaa ggcattgtt ggttaatct ggttatttt tttccacaa gttaataaaa       | 960  |
| tcataaaact tcatgtgtt tctctta  | 987  |
| <br>  |      |
| <210> 119   |      |
| <211> 1310  |      |
| <212> DNA   |      |
| <213> Homo sapiens  |      |
| <400> 119   |      |
| <br>  |      |
| aaattgagcc cgccgcctcc cgcttcgctc tctgctcc tcgttcgaca gtcagccgca     | 60   |
| tcttcttttgc cgtcgccagc cgagccacat cgctcagaca ccatgggaa ggtgaaggc    | 120  |
| <br>  |      |
| ggagtcaacg gatttggctg tattggcgc ctggcacca gggctgctt taactctgg       | 180  |
| aaagtggata ttgttgcattt caatgacccc ttcatggacc tcaactacat ggttacatg   | 240  |
| ttccaatatg attccaccca tggcaattt catggcaccc tcaaggctga gaacggaaag    | 300  |
| cttgcatca atggaaatcc catcaccatc ttccaggagc gagatccctc caaaatcaag    | 360  |
| tggggcgttgc ctggcgctga gtacgtcg ggttccactg ggttcttac caccatggag     | 420  |
| aaggctgggg ctcatggca ggggggagcc aaaagggtca tcatctcgc cccctctgct     | 480  |
| gatccccca ttttcgtcat ggggtgtgaac catgagaagt atgacaacag cctcaagatc   | 540  |
| <br>  |      |
| atcagcaatg cctcctgcac caccaactgc ttagcacccc tggccaaagg catccatgac   | 600  |
| aacttggta tcgtggagg actcatgacc acagtccatg ccatcactgc caccagaag      | 660  |
| actgtggatg gcccctccgg gaaactgtgg cgtgtggcc gggggctct ccagaacatc     | 720  |
| atccctgcct ctactggcgc tgccaaaggct gtggcaagg tcatccctga gctgaacggg   | 780  |
| aagctcaactg gcatggcctt ccgtgtcccc actgccaacg tgcgtggatgg accatggacc | 840  |
| tgcgtctag aaaaacctgc caaatatgt gacatcaaga aggtggtgaa gcaggcgtcg     | 900  |
| gaggggcccc tcaagggcat cctggctac actgaggacc aggtggctc ctctgacttc     | 960  |
| <br>  |      |
| aacagcgaca cccactcctc caccttgac gctggggctg gcattgcct caacgaccac     | 1020 |
| tttgtcaagc tcatttcctg gtatgacaac gaatttgctt acagcaacag ggtggtgac    | 1080 |
| ctcatggccc acatggcctc caaggagtaa gaccctgga ccaccagccc cagcaagagc    | 1140 |

acaagaggaa gagagagacc ctcactgctg gggagtccct gccacactca gtccccacc 1200  
 acaactgaatc tcccctcctc acagttgccatgtttagacccc ttgaagaggg gaggggccta 1260  
 gggagccgca ccttgtcatg taccatcaat aaagtaccct gtgctcaacc 1310  
 <210> 120  
 <211> 2245  
 <212> DNA

<213> Homo sapiens

<400> 120

acgcgacccg ccctacggc acctcccg cgccgcagac ggtggccgag 60  
 cgggggaccc ggaagcatgg cccgggggtc ggccgttgcc tggcgccgc tcggccgtt 120  
 gttgtgggc tgccgcgtgg ggctgcaggc cgggatgcgt taccccccagg agagccgc 180  
 gcgggagtgc aaggagctgg acggcctctg gagctccgc gccgacttct ctgacaaccg 240  
 acgcggggc ttccaggagc agtggtaaccg gcccgcgtg tggagtcag gcccaccgt 300  
 ggacatgcca gttccctcca gttcaatga catcagccag gactggcg 360

tgtcggtgg gtgtggtaacg aacgggaggt gatcctgcg gagcgatgga cccaggacct 420  
 ggcgcacaaga gtgggtgtga ggattggcag tgcccatcc tatgccatcg tgtgggtgaa 480  
 tggggtcac acgctagagc atgagggggg ctacccccc ttccggccg acatcagcaa 540  
 cctgggtccag gtggggcccc tgcctcccg gctccgaatc actatgc 600  
 actcaccccc accaccctgc caccaggac catccaatac ctgactgaca cctccaagta 660  
 tcccaagggt tactttgtcc agaacacata ttttgactt ttcaactacg ctggactgca 720  
 gcggtctgtta ctctgtaca cgacacccac cacctacatc gatgacatca ccgtcaccac 780

cagcgtggag caagacagtggctggaa ttaccagatc tctgtcaagg gcagtaacct 840  
 gttcaagttg gaagtgcgtc tttggatgc agaaaacaaa gtcgtggcga atggactgg 900  
 gaccggc caacttaagg tgccagggt cagcctctgg tggcgatcc tgcacgac 960  
 acgcctgtcc tatctgtatt cattggagggt gcagctgact gcacagacgt cactggggcc 1020  
 tgtgtctgac ttctacacac tccctgtgg gatccgact gtggctgtca ccaagagcca 1080  
 gttccctatc aatggaaac ctttctatcc acgggtgtc aacaagcatg aggatgccc 1140  
 catccgaggg aagggttcg actggccgt gctggtaag gacttcaacc tgctcgctg 1200

gcttgggtgcc aacgcttcc gtaccagcca ctaccctat gcagaggaag tgcacatgc 1260  
 gtgtgaccgc tatggatttggatgc tgcgtgtccc ggcgtggcc tggcgatcc 1320  
 gcagttttc aacaacgttt ctctgtatca ccacatgcag gtgcgtggaaag aagtggcg 1380

|  |      |
|--|------|
| tagggacaag aaccaccccg cggtcgtat gtggctgtg gccaacgagc ctgcgtccca    | 1440 |
| cctagaatct gctggctact acttgaagat ggtgatcgct cacaccaaat ccttggaccc  | 1500 |
| ctcccgccct gtgaccttg tgagcaactc taactatgca gcagacaagg gggctccgta   | 1560 |
| tgtggatgtg atctgttgta acagctacta ctcttgtat cacgactacg ggcacctgga   | 1620 |
| <br>   |      |
| gttgattcag ctgcagctgg ccacccagg ttgagaactgg tataagaagt atcagaagcc  | 1680 |
| cattattcag agcgagtatg gagcagaaac gattgcaggg tttcaccagg atccacctct  | 1740 |
| gatttcact gaagagtacc agaaaagtct gctagagcag taccatctgg gtctggatca   | 1800 |
| aaaacgcaga aaatacgtgg ttggagagct catttggaat tttgccatt tcatgactga   | 1860 |
| acagtcccg acgagagtgc tgggaataa aaaggggatc ttcaactcggc agagacaacc   | 1920 |
| aaaaagtgca gcgttcctt tgcgagagag atactggaag attgccaatg aaaccaggtt   | 1980 |
| tccccactca gtggccaagt cacaatgttt ggaaaacagc ccgttactt gagcaagact   | 2040 |
| <br>   |      |
| gataccacct gcgtgtccct tcctccccga gtcagggcga cttccacagc agcagaacaa  | 2100 |
| gtgcctctg gactgttac ggcagaccag aacgtttctg gcctgggtt tgtggatc       | 2160 |
| tattcttagca gggAACACTA aaggtggaaa taaaagattt tctattatgg aaataaagag | 2220 |
| ttggcatgaa agtggctact gaaaa  | 2245 |
| <br>   |      |
| <210> 121  |      |
| <211> 1526   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 121  |      |
| ccggaagtga cgcgaggctc tggggagacc aggagtca cttgttaggacg acctcggtc   | 60   |
| ccacgtgtcc ccggtaactcg ccggccggag ccccccgtt cccggggccg ggggacctta  | 120  |
| <br>   |      |
| gcggcaccca cacacaggct actttccaag cggagccatg tctggtaacg gcaatcgcc   | 180  |
| tgcaacggcg gaagaaaaca gcccaaagat gagagtgtt cgcgtggta cccgcaagag    | 240  |
| ccagcttgct cgcatacaga cggacagtgt ggtggcaaca ttgaaagcct cgtaccctgg  | 300  |
| cctgcagttt gaaatcattt cttatgtccac cacagggac aagattttt atactgcact   | 360  |
| ctctaagatt ggagagaaaa gcctgtttac caaggagctt gaacatgccc tggagaagaa  | 420  |
| tgaagtggac ctgggtgttc actccttga ggacactccc actgtgcttc ctccggctt    | 480  |
| caccatcgga gccatctgca agcgggaaaa ccctcatgtat gctgttgtt ttcacccaa   | 540  |
| <br>   |      |
| atttgttggg aagaccctag aaaccctgcc agagaagagt gtgggtggaa ccagctccct  | 600  |
| gcgaagagca gcccagctgc agagaaagtt cccgcatctg gagttcagga gtattcgggg  | 660  |

|             |              |             |             |             |             |      |
|-------------|--------------|-------------|-------------|-------------|-------------|------|
| aaaccta     | ccccggcttc   | ggaagctgga  | cgagcagcg   | gagttcagtgc | ccatcatcct  | 720  |
| ggcaacagct  | ggcctgcagc   | gcatgggctg  | gcacaaccgg  | gtggggcaga  | tcctgcaccc  | 780  |
| tgaggaatgc  | atgtatgtc    | tggccagggg  | ggcctgggc   | gtggaagtgc  | gagccaaggaa | 840  |
| ccaggacatc  | ttggatctgg   | tgggtgtgct  | gcacgatccc  | gagactctgc  | ttcgctgcat  | 900  |
| cgctgaaagg  | gccttcctga   | ggcacctgga  | aggaggctgc  | agtgtgccag  | tagccgtgca  | 960  |
|             |              |             |             |             |             |      |
| tacagctatg  | aaggatggc    | aactgtacct  | gactggagga  | gtctggagtc  | tagacggctc  | 1020 |
| agatagcata  | caagagacca   | tgcaggctac  | catccatgtc  | cctgcccagc  | atgaagatgg  | 1080 |
| ccctgaggat  | gaccacagt    | tggtaggcat  | cactgctcgt  | aacattccac  | gagggccccca | 1140 |
| gttggctgcc  | cagaacttgg   | gcatcagcct  | ggccaacttgc | ttgctgagca  | aaggagccaa  | 1200 |
| aaacatcctg  | gatgttgac    | ggcagctaa   | cgatccccat  | taactggtt   | gtggggcaca  | 1260 |
| gatgcctgg   | ttgctgtgt    | ccagtgccta  | catccggc    | ctcagtgc    | cattctact   | 1320 |
| gctatctgg   | gagtgttac    | cccgaggagac | tgaactgcag  | ggttcaagcc  | ttccaggat   | 1380 |
|             |              |             |             |             |             |      |
| ttgcctcacc  | ttggggcctt   | gatgactgcc  | ttgcctcctc  | agtatgtggg  | ggcttcatct  | 1440 |
| cttttagagaa | gtccaaagcaa  | cagccttga   | atgtaaacc   | tcctactaat  | aaaccagttc  | 1500 |
| tgaaggtgt   | aaaaaaaaaa   | aaaaaaa     |             |             |             | 1526 |
| <210>       | 122          |             |             |             |             |      |
| <211>       | 1435         |             |             |             |             |      |
| <212>       | DNA          |             |             |             |             |      |
| <213>       | Homo sapiens |             |             |             |             |      |
| <400>       | 122          |             |             |             |             |      |
| ggcggggcct  | gcttctcctc   | agcttcaggc  | ggctgcgacg  | agccctcagg  | cgaacctc    | 60   |
| ggcttcccg   | cgcggcgccg   | cctcttgctg  | cgcctccgccc | tcctcctctg  | ctccgcacc   | 120  |
| ggcttccccc  | tcctgagcag   | tcagccccgc  | cgccggccgg  | ctccgttatg  | gcgaccgc    | 180  |
|             |              |             |             |             |             |      |
| gccctggcgt  | cgtgattagt   | gatgtgaac   | caggttatga  | ctttgattta  | ttttgcatac  | 240  |
| ctaatcatta  | tgctgaggat   | ttggaaaggg  | tgtttattcc  | tcatggacta  | attatggaca  | 300  |
| ggactgaacg  | tcttgcgcga   | gatgtatga   | aggagatggg  | aggccatcac  | attgtagccc  | 360  |
| tctgtgtgt   | caagggggc    | tataaattct  | ttgctgaccc  | gctggattac  | atcaaagcac  | 420  |
| tgaatagaaa  | tagtgataga   | tccattccta  | tgactgtaga  | ttttatcaga  | ctgaagagct  | 480  |
| attgtaatga  | ccagtcaaca   | ggggacataa  | aagtaattgg  | tggagatgt   | ctctcaactt  | 540  |
| taactggaaa  | aatgtcttg    | attgtgaaag  | atataattga  | cactggcaa   | acaatgcaga  | 600  |
|             |              |             |             |             |             |      |
| cttgcttcc   | cttggtcagg   | cagtataatc  | caaagatggt  | caaggtcgca  | agcttgctgg  | 660  |

|   |      |
|---|------|
| tgaaaaggac cccacgaagt gttggatata agccagactt tggattt gaaattccag    | 720  |
| acaagtttgt tggatata gcccttgact ataatgaata ctgcaggat ttgaatcatg    | 780  |
| tttgtgtcat tagtgaaact ggaaaagcaa aatacaaagc ctaagatgag agttcaagtt | 840  |
| gagtttgaa acatctggag tcctattgac atcgccagta aaattatcaa tggtagtt    | 900  |
| ctgtggccat ctgcttagta gagcttttg catgtatctt ctaagaattt tatctgttt   | 960  |
| gtactttaga aatgtcagtt gctgcattcc taaactgttt attgcacta tgagcctata  | 1020 |

|   |      |
|---|------|
| gactatcagt tcccttggg cgattgttg tttaacttgt aatgaaaaa attctttaa       | 1080 |
| accacagcac tattgagtga aacattgaac tcatactgt aagaataaaa gagaagatat    | 1140 |
| attagtttt taattggat tttaattttt atatatgcag gaaagaatag aagtgttga      | 1200 |
| atattgttaa ttataccacc gtgtgttaga aaagtaagaa gcagtcaatt ttcacatcaa   | 1260 |
| agacagcatc taagaagttt tgttctgtcc tggaattatt ttagtagtgtt ttcagtaatg  | 1320 |
| ttgactgtat ttccaacctt gttcaaattt ttaccagtga atctttgtca gcagttccct   | 1380 |
| tttaaatgca aatcaataaaa ttcccaaaaaa tttaaaaaaaaaa aaaaaaaaaaaa aaaaa | 1435 |

<210> 123

<211> 2439

<212> DNA

<213> Homo sapiens

<400> 123

gagagcagcg gccggaaagg ggcgggtcg ggaggccgggt gtggggcggt agtgtggcc 60  
ctgttcctgc ccgcgcggtg ttccgcattc tgcaagcctc cggagcgcac gtcggcagtc 120  
ggctccctcg ttgaccgaat caccgacctc tctccccagc tgtatttcca aaatgtcgct 180  
ttcttaacaag ctgacgctgg acaagctgga cgttaaaggg aagcgggtcg ttatgagagt 240  
cgacttcaat gttcctatga agaacaacca gataacaaac aaccagagga ttaaggctgc 300  
tgtcccaagc atcaaattct gcttgacaa tgaggccaaag tcggtagtcc ttatgagccca 360

|   |      |
|---|------|
| cccagagcga cccttcctgg ccatcctggg cgagctaaa gttgcagaca agatccagct  | 840  |
| catcaataat atgctggaca aagtcaatga gatgattatt ggtggtgaa tggctttac   | 900  |
| cttccttaag gtgtcaaca acatggagat tggcacttct ctgttgatg aagagggagc   | 960  |
| caagaatgtc aaagaccaa tgcctaaagg tgagaagaat ggtgtgaaga ttaccitgcc  | 1020 |
| tgttgacttt gtcactgctg acaagttga tgagaatgcc aagactggcc aagccactgt  | 1080 |
| ggcttcggc atacctgctg gctggatggg cttggactgt ggtcctgaaa gcagcaagaa  | 1140 |
| gtatgtcgag gctgtcactc gggctaagca gattgtgtgg aatggcctg tggggattt   | 1200 |
| <br>  |      |
| tgaatggaa gctttgcc ggggaaccaa agctctcatg gatgaggtgg taaaagccac  | 1260 |
| ttcttagggc tcacatcacca tcataaggtagg tggagacact gccacttgtt gtgc当地atg<br>gaacacggag gataaagtca gccatgttag cactgggggt ggtgccagtt tggagctct | 1320 |
| ggaaggtaaa gtccttcctg ggggtggatgc tctcagcaat atttagtact ttccctgcctt   | 1380 |
| tttagttcctg tgacacagccc ctaagtcaac tttagcattt ctgc当地tcc acttggcatt  | 1440 |
| agctaaaacc ttccatgtca agattcagct agtggccaag agatgcagtg ccaggaaccc   | 1500 |
| ttaaacagtt gcacagcatc tcagctcatc ttcaactgcac cctggatttgc catacattct   | 1560 |
| <br>  |      |
| tcaagatccc atttgaattt tttagtgact aaaccattgt gcattctaga gtgc当地tat  | 1620 |
| ttatattttc cctgttaaaa agaaagttag cagtgtagc tttagttctt tttgtatgttag  | 1680 |
| gttattatga tttagcttgc cactgttca ctactcagca tggaaacaag atgaaattcc  | 1740 |
| attttaggt agtgagacaa aatttgatgtat ccattaagta aacaataaaa gtgtccattt  | 1800 |
| aaaccgttat tttttttttt ttcctgtcat actttgttag gaagggttag aatagaatct   | 1860 |
| tgaggaacgg atcagatgtc tatattgtcg aatgcaagaa gtggggcagc agcagtgag  | 1920 |
| agatgggaca attagataaa tgtccatttct ttatcaaggg cctactttat ggcagacatt  | 1980 |
| <br>  |      |
| gtgcttagtgc ttttatttcta acttttattt ttatcagttt cacatgtatca taatttaaaa  | 2040 |
| agtcaaggct tataacaaaaa aagccccagc ccattccccc cattcaagat tcccactccc  | 2100 |
| cagaggtgac cacttcaac tcttgatgtt ttcaggtata tacctccatg tttcttaagta   | 2160 |
| atatgtttat attgttact tctttttttt ttatttttta aagaaatctta tttcataccat  | 2220 |
| tggaggaagg ctctgttcca catatatttca cacttccatg ttctctcggt atagttttgt  | 2280 |
| cacaattata gattagatca aaagtctaca taactaatac agctgagctt tggatgtgc  | 2340 |
| tatgattaaa ttactttagt taaaaaaaaaaaaaaaa   | 2400 |
|   | 2439 |

&lt;210&gt; 124

&lt;211&gt; 2276

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 124

|   |      |
|---|------|
| gaacgtggta taaaaggggc gggaggccag gctcgccg tttgcagac gccaccgccc    | 60   |
| aggaaaaaccg tgtactatta gccatggca accccaccgt gtcttcgac attgccgtcg  | 120  |
| acggcgagcc ctggggccgc gtctccttg agctgttgc agacaaggc ccaaagacag    | 180  |
| cagaaaattt tcgtgctctg agcactggag agaaaggatt tggttataag gtttcgtct  | 240  |
| ttcacagaat tattccaggg ttatgtgtc agggtggta ctacacacgc cataatggca   | 300  |
| ctggtggcaa gtccatctat ggggagaaat ttgaagatga gaacttcata ctaaagcata | 360  |
| <br>  |      |
| cgggtcttgg catcttgtcc atggcaaaatg ctggacccaa cacaatggt tccagttt   | 420  |
| tcatctgcac tgccaagact gagtggttgg atggcaagca tgtgggttt ggcaagtga   | 480  |
| aagaaggcat gaatattgtg gaggccatgg agcgtttgg gtccaggaat ggcaagacca  | 540  |
| gcaagaagat caccattgtc gactgtggac aactcgaata agttgactt gtgtttatc   | 600  |
| ttaaccacca gatcatcct tctgtagtc aggagagcac ccctccaccc cattgtctcg   | 660  |
| cagtatccta gaatctttgt gctctcgctg cagttccctt tgggtccat gtttccttg   | 720  |
| ttccctccca tgccatctg gattgcagag ttaagtttat gattatgaaa taaaaactaa  | 780  |
| <br>  |      |
| ataacaattt tcctcgttt agttaagagt gttgatgttag gttttttttt aagcagtaat | 840  |
| gggttacttc tgaacatca ctgtttgtc taattctaca cagacttag atttttttta    | 900  |
| ctttccagtc ccaggaagtg tcaatgtttt ttgagtggaa tattgaaaat gtggcagca  | 960  |
| actggccatg gtggctact gtctgtatg tattacctga ggcagaagac cacctgaggg   | 1020 |
| taggagtcaa gatcgcctg ggcaacatag tgagacgctg tctctacaaa aaataattag  | 1080 |
| cctggcctgg tggtgcatgc ctgccttag ctgatctgga ggctgacgtg ggaggattgc  | 1140 |
| ttgagcctag agttagctat tatcatgcca ctgtacagcc tgggtgtca cagatctgt   | 1200 |
| <br>  |      |
| gtctcaaagg taggcagagg cagaaaaagc aaggagccag aattaagagg ttgggtcagt | 1260 |
| ctgcagttag ttcatgcatt tagaggtttt cttcaagatg actaatgtca aaaattgaga | 1320 |
| catctgtgc gttttttttt tttttttttt cccctggaat gcagtggcgt gatctcagct  | 1380 |
| cactgcagcc tccgcctct gggttcaagt gattcttagt ctcgcctc ctgagtagct    | 1440 |
| gggataatgg gcgtgtgcca ccatgcccag ctaattttt tatttttagt atagatgggg  | 1500 |
| tttcatcatt ttgaccaggc tggctcaaa ctctgaccc cagctgatgc gcctgccttg   | 1560 |
| gcctccaaa ctgcgtgatg tacagatgtg agccaccgca ccctaccta tttctgtaa    | 1620 |
| <br>  |      |
| caaagctaag cttaaacact gttgatgttc ttgaggaaag catattggc ttttaggtgt  | 1680 |
| aggtaagtt tatacatctt aattatgggtt gaattcctat gtagagtcta aaaagccagg | 1740 |

|            |              |             |            |              |            |      |
|------------|--------------|-------------|------------|--------------|------------|------|
| tactgggc   | tacagtca     | ctccctgcag  | agggttaagg | cgcagactac   | ctgcagttag | 1800 |
| gaggta     | tgc          | tttgcata    | tagagcctc  | ccctagctt    | ggttatggag | 1860 |
| tttgcaaa   | acc          | tgaccaattt  | aagccataag | atctggtcaa   | aggataccc  | 1920 |
| gactigtt   | tctcaggaaa   | ttatatgtac  | agtgcgtgc  | ggcagttaga   | tgtcaggaca | 1980 |
| atctaagctg | agaaaacccc   | ttctctgcc   | accttaacag | acctctaggg   | ttcttaaccc | 2040 |
| <br>       |              |             |            |              |            |      |
| agcaatcaag | tttgccat     | ctagagggtgg | cggtttgtat | cattgggtgt   | gttgggcaat | 2100 |
| ttttgttta  | ctgtctgg     | cttctgcgt   | gaattaccac | caccaccact   | tgtgcac    | 2160 |
| agtcttgtgt | gttgtctgg    | tacgtattcc  | ctgggtgata | ccattcaatg   | tcttaatgt  | 2220 |
| cttggg     | tc           | agac        | gtgtgg     | gaaataaacatc | aaacatctt  | 2276 |
| <210>      | 125          |             |            |              |            |      |
| <211>      | 1229         |             |            |              |            |      |
| <212>      | DNA          |             |            |              |            |      |
| <213>      | Homo sapiens |             |            |              |            |      |
| <400>      | 125          |             |            |              |            |      |
| gtctgacggg | cgtggcgca    | gccaatagac  | aggagcgcta | tccgggttt    | ctgattggct | 60   |
| actttgttcg | cattataaaa   | ggcacgcg    | ggcgcgaggc | ccttctctcg   | ccaggcg    | 120  |
| <br>       |              |             |            |              |            |      |
| tcgtgaa    | gt           | gacatcg     | ttaaaccctg | cgtggcaatc   | cctgacgcac | 180  |
| cccagg     | gaag         | acaggcgac   | ctggaa     | gtcc         | cgccgtgat  | 240  |
| gatgattatc | cgtt         | cgaaatgttt  | cattgtgg   | gcagacaatg   | tggctccaa  | 300  |
| cagatccgca | tgtcc        | ctcg        | cg         | gtgg         | gtctga     | 360  |
| cgcaaggc   | tccg         | agggca      | cctgg      | aaaac        | aacc       | 420  |
| atccggg    | gg           | atgtgg      | tt         | tgtgtt       | cacc       | 480  |
| ttgctgg    | cc           | ataagg      | gtgc       | cc           | ctgtggat   | 540  |
| <br>       |              |             |            |              |            |      |
| actgtgccag | cccaga       | acac        | tgg        | tctcg        | ccc        | 600  |
| ggtatcacca | ctaaat       | ctc         | cagg       | ggcacc       | cc         | 660  |
| aagactggag | aca          | aaatggg     | agcc       | agc          | cc         | 720  |
| cccttctct  | t            | ttgg        | ggct       | catcc        | cc         | 780  |
| gaagtgc    | tt           | at          | atc        | acaga        | cc         | 840  |
| gttgc      | cc           | at          | atc        | actctg       | cc         | 900  |
| atcaacgg   | gt           | aca         | aaac       | cgagt        | cc         | 960  |
| <br>       |              |             |            |              |            |      |
| gctgaaaagg | tcaagg       | cctt        | ttgg       | ctgat        | ccat       | 1020 |

|  |      |
|--|------|
| gctgctgccca ccacagctgc tcctgctgt gctgcagccc cagctaagg tgaagccaag       | 1080 |
| gaagagtccgg aggagtccgga cgaggatatg ggatttggtc tctttgacta atcaccaaaa    | 1140 |
| agcaaccaac tttagccagtt ttatggcaa aacaaggaaa taaaggctta ctctttaaa       | 1200 |
| aagtaaaaaaa aaaaaaaaaa aaaaaaaaaa                                      | 1229 |
| <210> 126  |      |
| <211> 1142   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 126  |      |
| <br>cttttccaag cgctgcccga agatggcgga ggtgcaggc ctggtgcttg atggtcgagg   | 60   |
| <br>ccatctcctg ggccgcctgg cggccatcgt ggctaaacag gtactgctgg gccggaagggt | 120  |
| ggtgtcgta cgctgtgaag gcatcaacat ttctggcaat ttctacagaa acaagttgaa       | 180  |
| gtacacctggct ttccctccgca agcggatgaa caccaaccct tcccgaggcc cctaccactt   | 240  |
| ccggggcccccc agccgcattt tctggcggac cgtgcgaggt atgctgcccc acaaaaccaa    | 300  |
| gcgaggccag gccgcctctgg accgtctcaa ggtgtttgac ggcattccac cgcctacga      | 360  |
| caagaaaaag cgatgggtgg ttccctgctgc cctcaaggc gtgcgtctga agcctacaag      | 420  |
| aaagtttgcc tatctggggc gcctggctca cgaggttggc tggaagtacc aggcaagtac      | 480  |
| <br>agccaccctg gaggagaaga ggaaagagaa agccaagatc cactaccgga agaagaaaca  | 540  |
| gctcatgagg ctacggaaac aggccgagaa gaacgtggag aagaaaattt acaaatacac      | 600  |
| agaggcttc aagacccacg gactcctggt ctgagccaa taaagactgt taattcctca        | 660  |
| tgcgttgcct gcccttcctc cattgttgcc ctggaatgta cgggaccctag gggcagcagc     | 720  |
| agtccaggtg ccacaggcag ccctgggaca taggaagctg ggagcaagga aagggtctta      | 780  |
| gtcactgcct cccgaagttt ctgaaagca ctggagaat tgtgcaggtg tcatttatct        | 840  |
| atgaccaata ggaagagcaa ccagttacta tgagtgaaag ggagccagaa gactgattgg      | 900  |
| <br>agggccctat ctgtgagtg gggcatctgt tggactttcc acctggcat atactctgca    | 960  |
| gctgttagaa tgtgcaagca cttggggaca gcatgagctt gctgttgtac acagggtatt      | 1020 |
| tctagaagca gaaatagact ggaaagatgc acaaccaagg gttacaggc atcgcccatg       | 1080 |
| ctcctcacct gtatggta atcagaaata aattgtttt aaagaaaaaa aaaaaaaaaa         | 1140 |
| aa   | 1142 |
| <210> 127  |      |
| <211> 2405   |      |
| <212> DNA  |      |

&lt;213&gt; Homo sapiens

&lt;400&gt; 127

tccggcgtgg tgcgaggcg cggtatcccc cctcccccgc cagctcgacc ccggtgtggt 60

gchgaggcgc agtctgcgca gggactggcg ggactgcgcg gggcaacag cagacatgtc 120

gggggtccgg ggcctgtcg gcgtcgtag cgctcgccgc ctggcgctgg ccaaggcgtg 180

gccaacagtg ttgcaaacag gaacccgagg ttttacttc actgttgatg ggaacaagag 240

ggcatctgct aaagtttcag attccatttc tgctcagtat ccagtagtgg atcatgaatt 300

tgcgtcgatg gtggtaggcg ctggagggc aggctgcga gctgcatttg gccttctga 360

ggcagggttt aatacagcat gtgttaccaa gctgttcctt accaggtcac acactgtgc 420

agcacagggaa ggaatcaatg ctgctctggg gaacatggag gaggacaact ggaggtggca 480

tttctacgac accgtgaagg gctccgactg gctggggac caggatgcc tccactacat 540

gacggagcag gccccggccg ccgtggtcga gctagaaaat tatggcatgc cgtttagcag 600

aactgaagat gggaaagattt atcagcgtgc atttggtgaa cagagccta agttggaaa 660

ggcgccggcag gccatcggt gctgtgtgt ggctgtatcg actggccact cgctattgca 720

caccttatat ggaaggcttc tgcgatatga taccagctat tttgtggagt atttgcctt 780

ggatctcctg atggagaatg gggagtgcgg tgggtgtcatc gcactgtgc tagaggacgg 840

gtccatccat cgccataagag caaagaacac tggatgtgcc acaggaggct acggcgcac 900

ctacttcagc tgcacgtctg cccacaccag cactggcgac ggcacggcca tgcgtaccag 960

ggcaggcctt ccttgccagg accttagagtt tggatgttc caccctacag gcatatatgg 1020

tgctgggtgt ctcattacgg aaggatgtcg tggagaggga ggcattctca ttaacagtca 1080

aggcgaaagg ttatggagc gatacgcccc tgcgcaag gacctggcgt cttagatgt 1140

gggtgtctcg tccatgactc tggagatccg agaaggaaaga ggctgtggcc ctgagaaaga 1200

tcacgtctac ctgcagctgc accacctacc tccagagcag ctggccacgc gcctgcctgg 1260

catttcagag acagccatga tcttcgtgg cgtggacgtc acgaaggagc cgatccctgt 1320

cctcccccacc gtgcattata acatggcggtt cattccacc aactacaagg ggcaggcct 1380

gaggcacgtg aatggccagg atcagatgtt gcccggctg tacgcctgtg gggaggccgc 1440

ctgtgcctcg gtacatggtg ccaaccgcct cggggcaaac tgcgtttgg acctgggtgt 1500

ctttggctgg gcatgtgcgg tggatgtcg agatgtatgc aggcctggag ataaagtccc 1560

tccaattaaa ccaaaccgtg gggagaatac tgcgtatgtt ctgcataaaat tgagatgtc 1620

tgcgtggaaatc ataaaggatc cggaaactgcg actcagcatg cagaagtcaa tgcaaatca 1680

|  |      |
|--|------|
| tgctgccgtg ttccgtgtgg gaagcgtgtt gcaagaagggt tgtggaaaaa tcagcaagct | 1740 |
| <br>   |      |
| ctatggagac ctaaagcacc tgaagacgtt cgaccggga atggctcgaa acacggacct   | 1800 |
| ggtggagacc ctggagctgc agaacctgtat gctgtgtcgctgcagacca tctacggagc   | 1860 |
| agaggcacgg aaggagtacac ggggcgcga tgccaggaa gactacaagg tgccgattga   | 1920 |
| tgagtagcat tactccaagg ccatccaggaa gcaacagaag aagccctttg aggagactg  | 1980 |
| gaggaagcac accctgtcct atgtggacgt tggcactgg aaggtcactc tggaatata    | 2040 |
| acccgtgatc gacaaaactt tgaacgaggc tgactgtgcc accgtcccgc cagccatcg   | 2100 |
| ctcctactga tgagacaaga tgtggtgatg acagaatcag cttttaat tatgtataat    | 2160 |
| <br>   |      |
| agctcatgca tgggtccatg tcataactgt cttcatacgc ttctgcactc tgggaaagaa  | 2220 |
| ggagtagcatt gaaggagat tggcacctag tggctggag ctggccagga acccagtggc   | 2280 |
| cagggagcgt ggcacttacc ttgtccctt gcttcattct tgtgagatga taaaactggg   | 2340 |
| cacagctttt aaataaaata taatgaaca aactttttt tatttccaaa aaaaaaaaaaa   | 2400 |
| aaaaaa   | 2405 |
| <210> 128  |      |
| <211> 1867   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 128  |      |
| ggttcgctgt ggcggcgcc tggccgcgg gctgttaac ttgcgttccg ctggccata      | 60   |
| <br>   |      |
| gtgatcttg cagtgaccca gcagcatcac tgttcttgg cgtgtgaaga taacccaagg    | 120  |
| aattgaggaa gttgctgaga agagtgtgct ggagatgctc tagaaaaaa ttgaatagt    | 180  |
| agacgagttc cagcgcaagg gttctgggt tgccaagaag aaagtgaaca tcatggatca   | 240  |
| gaacaacagc ctgccaccc acgctcaggc ttggccctcc ctcagggtt ccattactcc    | 300  |
| cggaaatccct atcttagtc caatgatgcc ttatggact ggactgaccc cacgcctat    | 360  |
| tcagaacacc aatagtctgt ctatttggaa agagcaacaa aggcagcagc agcaacaaca  | 420  |
| acagcagcag cagcagcagc agcagcaaca gcaacagcagc cagcagcagc agcagcagca | 480  |
| <br>   |      |
| gcagcagcag cagcagcagc agcagcagc gcaacaggca gtggcagctg cagccgttca   | 540  |
| gcagtcacg tcccgaggc caacacaggaa acacccatgc caggcaccac agctttcca    | 600  |
| ctcacagact ctcacaactg cacccttgcg gggcaccact ccactgtatc cctccccat   | 660  |
| gactcccatg acccccatca ctcctgccac gccagttcg gagagttctg ggattgtacc   | 720  |
| gcagctgcaa aatattgtat ccacagtcaa tcttggttgt aaacttgacc taaagaccat  | 780  |

|   |      |
|---|------|
| tgcaactcgt gcccggaaacg ccgaatataa tcccaagcgg tttgctgcgg taatcatgag  | 840  |
| gataagagag ccacgaacca cggcactgat ttcagttct gggaaaatgg tgtgcacagg    | 900  |
| <br>  |      |
| agccaagagt gaagaacagt ccagactggc agcaagaaaa tatgctagag ttgtacagaa   | 960  |
| gttgggttt ccagctaagt tctggactt caagatttag aatatggtgg ggagctgtga     | 1020 |
| tgtgaagttt cctataaggt tagaaggcct tgtgctcacc caccaacaat ttagtagtta   | 1080 |
| tgagccagag ttatttcctg gttaatcta cagaatgatc aaacccagaa ttgttctcct    | 1140 |
| tatTTTgtt tctggaaaag ttgtattaac aggtgtaaa gtcagagcag aaatttatga     | 1200 |
| agcatttgaa aacatctacc ctattctaaa gggattcagg aagacgacgt aatggcttc    | 1260 |
| atgtaccctt gcctccccca ccccttctt tttttttt taaacaaatc agtttgttt       | 1320 |
| <br>  |      |
| ggtacctta aatgggtgt ttgtgagaag atggatgtt agttgcaggg tgtggcacca      | 1380 |
| ggtgatgccc ttctgttaagt gcccaccgcg ggatgccggg aagggcatt atttgcac     | 1440 |
| tgagaacacc gcgcagcgtg actgtgagtt gctcatccg tgctgctatc tggcagcgc     | 1500 |
| tgcccattha ttatatgtt gatTTTaaac actgtgttg acaagtgggt ttgagggaga     | 1560 |
| aaacttaag tttaaagcc acctctataa ttgattggac ttttaattt taatgtttt       | 1620 |
| ccccatgaac cacagttttt atatttctac cagaaaagta aaaatctttt taaaagtgt    | 1680 |
| tgttttcta attataact cctagggtt atttctgtc cagacacatt ccacctctcc       | 1740 |
| <br>  |      |
| agtatigcag gacagaatat atgtgttaat gaaaatgaat ggctgtacat atttttct     | 1800 |
| ttcttcagag tactctgtac aataaatgca gtttataaaa gtgttaaaaa aaaaaaaaaa   | 1860 |
| aaaaaaaa  | 1867 |
| <210> 129   |      |
| <211> 5241  |      |
| <212> DNA   |      |
| <213> Homo sapiens  |      |
| <400> 129   |      |
| <br>  |      |
| agagcgctcg gatatcggtt ggccgcctcg gacggaggac gcgcgtatgtt gagtgccggc  | 60   |
| ttctagaact acaccgaccc tcgtgtcctc cttcatcct gcggggctgg ctggagcggc    | 120  |
| cgcctccgtt ctgtccagca gccataggga gccgcacggg gagcgggaaa gcggctcg     | 180  |
| <br>  |      |
| ccccaggcgg ggccggccggg atggagcggg gccgcgagcc tggggaaag gggctgtggc   | 240  |
| ggccgcctcg gccgcgtcgag gttttctgtt gtggcgttc agaatgtgg atcaagctag    | 300  |
| atcagcattc tctaacttgtt ttgggtggaga accattgtca tatacccggt tcagcctggc | 360  |
| tcggcaagta gatggcgata acagtcatgt ggagatgaaa cttgctgttag atgaagaaga  | 420  |

|  |      |
|--|------|
| aaatgctgac aataacacaa aggccaatgt cacaaaacca aaaaggtgta gtggaagtat  | 480  |
| ctgctatgg actattgctg tgatcgctt tttcttgatt ggatttatga ttggctactt    | 540  |
| gggcatttgt aaagggttag aaccaaaaac tgagtgtgag agactggcag gaaccgagtc  | 600  |
| <br>   |      |
| tccagtgagg gaggagccag gagaggactt ccctgcagca cgtcgcctat attggatga   | 660  |
| cctgaagaga aagttgtcg agaaaactgga cagcacagac ttccacggca ccatcaagct  | 720  |
| gctgaatgaa aattcatatg tccctcgta ggctggatct caaaaagatg aaaatctgc    | 780  |
| gttgtatgtt gaaaatcaat ttcgtgaatt taaactcagc aaagtctggc gtgatcaaca  | 840  |
| ttttgttaag attcaggtca aagacagcgc tcaaaaactcg gtgatcatag ttgataagaa | 900  |
| cggtagactt gttiacctgg tggagaatcc tggggttat gtggcgatata gtaaggctgc  | 960  |
| aacagttact gtaaaactgg tccatgtaa ttttgtact aaaaaagatt ttgaggattt    | 1020 |
| <br>   |      |
| atacactcct gtgaatggat ctatagtat tgtcagagca gggaaaatca ccttgcaga    | 1080 |
| aaagggtgca aatgctgaaa gcttaaatgc aattgggtgt ttgatataca tggaccagac  | 1140 |
| taaattccc atgttaacg cagaacttc attcttgga catgctcatc tggggacagg      | 1200 |
| tgacccttac acacctggat tcccttcctt caatcacact cagttccac catctcggtc   | 1260 |
| atcaggattt cctaataaac ctgtccagac aatctccaga gctgctgcag aaaagctttt  | 1320 |
| tggaaatatg gaaggagact gtccctctga ctggaaaaca gactctacat gtaggatggt  | 1380 |
| aacctcagaa agcaagaatg tgaagctcac tgtgagcaat gtgctgaaag agataaaaat  | 1440 |
| <br>   |      |
| tcttaacatc ttggagtt ttaaggctt tgtagaacca gatcactatg ttgttagttgg    | 1500 |
| ggcccgaga gatgcatggg gccctggagc tgcaaaatcc ggtgtaggca cagctctcct   | 1560 |
| attgaaacct gcccagatgt tctcagatgt ggtcttaaaa gatgggtttc agcccagcag  | 1620 |
| aagcattatc ttgccagtt ggagtgcgtt agactttgga tcggttgggt ccactgaatg   | 1680 |
| gcttagaggaa taccttcgtt ccctgcattt aaaggcttc acttatatta atctggataa  | 1740 |
| agcggttctt ggtaccagca acttcaaggt ttctggccagc ccactgttgt atacgcttat | 1800 |
| tgagaaaaca atgaaaatg tgaagcatcc ggtaactggg caatttctat atcaggacag   | 1860 |
| <br>   |      |
| caactgggcc agcaaagttt agaaaactcac tttagacaat gctgcttc ctttccttgc   | 1920 |
| atattctgga atcccagcag tttttttctg tttttgcgag gacacagatt atccttattt  | 1980 |
| gggttaccacc atggacacct ataaggaact gattgagagg attcctgagt tgaacaaagt | 2040 |
| ggcacgagca gtcgcagagg tcgctggta gttcgtgtt aaactaaccat atgatgtga    | 2100 |
| attgaacctg gactatgaga ggtacaacag ccaactgctt tcatttgtga gggatctgaa  | 2160 |
| ccaatacaga gcagacataa aggaaatggg cctgagttt cagtggtgtt attctgctcg   | 2220 |

|   |      |
|---|------|
| tggagacttc ttccgtgcta cttccagact aacaacagat ttcgggaatg ctgagaaaac     | 2280 |
| <br>  |      |
| agacagattt gtcataaga aactcaatga tcgtgtcatg agagtggagt atcacttcct      | 2340 |
| ctctccctac gtatctccaa aagagtctcc tttccgacat gtcttctgg gctccggctc      | 2400 |
| tcacacgctg ccagcttac tggagaacctt gaaactgcgt aaacaaaata acggtgctt      | 2460 |
| taatgaaacg ctgttcagaa accagtttgc tctagactact tgacttattc agggagctgc    | 2520 |
| aatgccctc tctggtgacg ttggacat tgacaatgag tttaaatgt gataccata          | 2580 |
| gcttccatga gaacagcagg gtagtctgg ttttagactt gtgtgtatcg tgctaaattt      | 2640 |
| tcagtagggc tacaaaacctt gatgttaaaa ttccatccca tcatcttg actactagat      | 2700 |
| <br>  |      |
| gtcttaggc agcagctttt aatacagggt agataacctg tacttcaagt taaagtgaat      | 2760 |
| aaccacttaa aaaatgtcca tcatggataa ttcccctatc tctagaattt taagtgttt      | 2820 |
| gtaatggaa ctgcctttt cctgttgtt ttaatgaaaa tgtcagaaac cagttatgt         | 2880 |
| aatgatctct ctgaatccta agggctggc tctgctgaag gttgttaagt gtcgttact       | 2940 |
| ttgagtgtatc ctccaaacttc atttgatgtt aaataggaga taccaggtt aaagaccc      | 3000 |
| tccaaatgag atctaaggct ttccataagg aatgttagctg gttccatcat tcctgaaaga    | 3060 |
| aacagttAAC tttcagaaga gatgggcttg tttcttgcc aatgaggtct gaaatggagg      | 3120 |
| <br>  |      |
| tccttctgct ggataaaatg aggttcaact gttgattgca ggaataaggc cttatatgt      | 3180 |
| taacctcagt gtcatttatg aaaagagggg accagaagcc aaagacttag tatatittct     | 3240 |
| tttcctctgt ccctcccccc ataagcctcc atttagttct ttgttatttt tgttctcc       | 3300 |
| aaagcacatt gaaagagaac cagtttcagg ttgttagttt cagactcagt ttgtcagact     | 3360 |
| ttaaagaata atatgctgcc aaattttggc caaagtgtta atcttagggg agagcttct      | 3420 |
| gtccctttgg cactgagata ttattttttt atttatcagt gacagagttc actataatg      | 3480 |
| gtgtttttt aatagaatat aattatcgga agcagtgcct tccataatta tgacagttat      | 3540 |
| <br>  |      |
| actgtcggtt ttttttaat aaaaggcagca tctgctaata aaacccaaca gatactggaa     | 3600 |
| gtttgcatt tatggtaaac acttaagggt tttagaaaaac agccgtcage caaatgtat      | 3660 |
| tgaataaagt tgaagctaaatg atttagagat gaattaaattt taatttagggg ttgctaagaa | 3720 |
| gcgagcactg accagataag aatgctggtt ttccataatg cagtgattt tgaccaagtt      | 3780 |
| ataaatcaat gtcacttaaa ggctgtggta gtactcctgc aaaattttt agctcagttt      | 3840 |
| atccaaaggta taactctaatt tccatatttgg caaaattttc agtacccattt tcacaatcct | 3900 |
| aacacattat cgggagcagt gtctccata atgtataaag aacaaggtag ttttaccta       | 3960 |

|  |      |
|--|------|
| ccacagtgtc tgtatcgag acagtgtat ccatatgtta cactaagggt gtaagtaatt    | 4020 |
| atcggaca gtgttccca taattttctt catgcaatga catttcaaa gcttgaagat      | 4080 |
| cgttagtac taacatgtat cccaactct ataattccct atcttttagt tttagttgca    | 4140 |
| gaaacattt gtggtcatta agcattgggt gggtaaatc aaccactgtaaatgaaatt      | 4200 |
| actacaaaat ttgaaatttgcgttgggtttttgttaccctttatggtttctccaggtcctc     | 4260 |
| tacttaatga gatagtagca tacattata atgttgcta ttgacaagtc atttaactt     | 4320 |
| tatcacatta ttgcgtt acctctata aacttagtgc ggacaagttt taatccagaa      | 4380 |
| <br>   |      |
| ttgaccttt gacttaaagc agagggactt tgtatagaag gtttggggc tgtgggaag     | 4440 |
| gagagtcccc tgaaggtctg acacgtctgc ctaccatc gtggatca attaaatgtat     | 4500 |
| ggtatgaata agttcgaagc tccgtgatgt aaccatcatt ataaacgtgtatcagctg     | 4560 |
| tttgtcatag ggcagttgaa aacggcctcc tagggaaaatgttcatagggctcttcaggt    | 4620 |
| tcttagtgc acttacccatg atttacagcc tcacttgaat gtgtcaactac tcacagtctc | 4680 |
| ttaatcttc agtttatct ttaatctcttcttggactgacat ttgcgttagc             | 4740 |
| taagtgaaaa ggtcatagct gagattccctg gttcgggtgt tacgcacacgtacttaatg   | 4800 |
| <br>   |      |
| aaagcatgtg gcatgttcat cgataaacac aatatgaata cagggcatgc atttgcagc   | 4860 |
| agttagtctc tttagaaaac cttttctac agtttagggtt gagttacttc ctatcaagcc  | 4920 |
| agtacgtgct aacaggctca atattcctga atgaaatatc agactagtgtaaatgtca     | 4980 |
| gtcttgatgt gtcttcgtttaaggagatg ggcctttgg aggtaaaggtaaatgtat        | 5040 |
| gagttctgatc atgattcaacttcttagaaac ttgcgtgacc ttactgtgt tagcttttg   | 5100 |
| aatgttcttggaaattttaga cttttttgtt aaacaaatgtatgtcctta tcattgtata    | 5160 |
| aaagctgtta tggcaacag tggagatt cttgtctgtttataaaaa tacttaaaca        | 5220 |
| <br>   |      |
| ctgaaaaaaaaaaaaaaa a   | 5241 |
| <br>   |      |
| <210> 130  |      |
| <211> 2602   |      |
| <212> DNA  |      |
| <213> Homo sapiens   |      |
| <400> 130  |      |
| <br>   |      |
| gagccgcggc taaggaacgc gggccgccc cccgctcccg gtgcagcggc ctccgcggc    | 60   |
| gttttggcg cctccgcgg ggcggccctt cctcaegcg agcgctgccatcgacgat        | 120  |
| agggcgcagc gagcgtctgtt atccttcgc ccggacgctc aggacagcgg cccgctgtc   | 180  |
| ataagactcg gccttagaac cccagttatca gcagaaggac attttaggac gggacttggg | 240  |

|  |      |
|--|------|
| tgactctagg gcactggttt tctttccaga gagcggaaaca ggcgaggaaa agtagtccct     | 300  |
| <br>   |      |
| tctcgccgat tctcgccgagg gatctccgtg gggcggtgaa cgccgatgtatataagga        | 360  |
| cgcgccgggt gtggcacagc tagttccgtc gcagccggga ttgggtcgc agttttgttt       | 420  |
| tgtggatcgc tgtgtatcgtc acttgacaat gcagatctc gtgaagactc tgactggtaa      | 480  |
| gaccatcacc ctgcgagggtt agcccagtga caccatcgag aatgtcaagg caaagatcca     | 540  |
| agataaggaa ggcatccctc ctgaccagca gaggctgatc ttgctggaa aacagctgga       | 600  |
| agatgggcgc accctgtctg actacaacat ccagaaagag tccaccctgc acctgggtgct     | 660  |
| ccgtctcaga ggtgggatgc aaatcttcgt gaagacactc actggcaaga ccatcaccct      | 720  |
| <br>   |      |
| tgaggtcgag cccagtgaca ccatcgagaa cgtcaaagca aagatccagg acaaggaagg      | 780  |
| cattcctcct gaccagcaga ggttgatctt tgccggaaag cagctggaag atggcgac        | 840  |
| cctgtctgac tacaacatcc agaaagagtc taccctgcac ctggtgctcc gtctcagagg      | 900  |
| tgggatcag atttcgtga agaccctgac tggtaagacc atcaccctcg aggtggagcc        | 960  |
| cagtgacacc atcgagaatg tcaaggcaaa gatccaagat aaggaaggca ttcctctga       | 1020 |
| tcagcagagg ttgatcttgc cgaaaaaca gctgaaagat ggtcgatccc tgtctgacta       | 1080 |
| caacatccag aaagagtcca cttgcaccc ggtactccgt ctgcgagggtg ggtgcaaat       | 1140 |
| <br>   |      |
| cttcgtgaag acactcactg gcaagaccat cacccttgag gtgcgaccgcgtgacactat       | 1200 |
| cgagaacgtc aaagcaaaga tccaagacaa ggaaggcatt ctcctgacc agcagaggtt       | 1260 |
| gatcttgcc gaaaagcagc tggaagatgg ggcacccctg tctgactaca acatccagaa       | 1320 |
| agagtctacc ctgcacctgg tgctccgtct cagaggtggg atgcagatct tcgtgaagac      | 1380 |
| cctgactggt aagaccatca ctctcgaagt ggagccgagt gacaccattt agaatgtcaa      | 1440 |
| ggcaaagatc caagacaagg aaggcatccc tcctgaccag cagaggttga tcttgccgg       | 1500 |
| aaaacagctg gaagatggc gtaccctgtc tgactacaac atccagaaag agtccaccc        | 1560 |
| <br>   |      |
| gcacctgggt ctccgtctca gaggtggat gcagatctc gtgaagaccc tgactggtaa        | 1620 |
| gaccatcaact ctgcgagggtgg agcccgagtga caccatttgag aatgtcaagg caaagatcca | 1680 |
| agacaaggaa ggcatccctc ctgaccagca gaggttgatc ttgctggaa aacagctgga       | 1740 |
| agatggacgc accctgtctg actacaacat ccagaaagag tccaccctgc acctgggtgct     | 1800 |
| ccgtctttaga ggtgggatgc agatcttcgt gaagaccctg actggtaaga ccatcactct     | 1860 |
| cgaagtggag ccgagtgaca ccattgagaa tgtcaaggca aagatccaag acaaggaagg      | 1920 |
| catccctcct gaccagcaga ggttgatctt tgctggaaa cagctggaag atggacgcac       | 1980 |

|            |              |            |            |             |             |            |     |
|------------|--------------|------------|------------|-------------|-------------|------------|-----|
| cctgtctgac | tacaacatcc   | agaaagagtc | caccctgcac | ctgggtgctcc | gtcttagagg  | 2040       |     |
| tgggatcgag | atttcgtga    | agaccctgac | tggtaagacc | atcactctcg  | aagtggagcc  | 2100       |     |
| gagtgacacc | attgagaatg   | tcaaggcaaa | gatccaagac | aaggaaggca  | tccctcctga  | 2160       |     |
| ccagcagagg | ttgatcttg    | ctggaaaca  | gctggaagat | ggacgcaccc  | tgtctgacta  | 2220       |     |
| caacatccag | aaagagtcca   | ccctgcacct | ggtgctccgt | ctcagaggtg  | ggatgcaa    | 2280       |     |
| ttcgtgaag  | accctgactg   | gtaagaccat | caccctcgag | gtggagccc   | gtgacaccat  | 2340       |     |
| cgagaatgtc | aaggcaaaga   | tccaa      | gataa      | ggaaggcatc  | cctcctgatc  | 2400       |     |
| <br>       |              |            |            |             |             |            |     |
| gatcttgct  | ggaaacagc    | tggaagatgg | acgcaccc   | tctgactaca  | acatccagaa  | 2460       |     |
| agagtccact | ctgcacttgg   | tcctgcgtt  | gaggggggt  | gtctaagttt  | cccc        | 2520       |     |
| ggtttcaaca | aatttcattt   | cacttcctt  | tcaataaa   | gttgcattc   | ccaaaaaaa   | 2580       |     |
| aaaaaaaaaa | aaaaaaaaaa   | aa         |            |             |             | 2602       |     |
| <210>      | 131          |            |            |             |             |            |     |
| <211>      | 3003         |            |            |             |             |            |     |
| <212>      | DNA          |            |            |             |             |            |     |
| <213>      | Homo sapiens |            |            |             |             |            |     |
| <400>      | 131          |            |            |             |             |            |     |
| ctttctc    | cccttcttc    | cgggctcccg | tcccgctca  | tcacccggcc  | tgtggccac   | 60         |     |
| tcccaccc   | agctggAAC    | ctggggacta | cgacgtccct | caaacc      | ttcttaggaga | 120        |     |
| <br>       |              |            |            |             |             |            |     |
| taaaaaaa   | atccagtcat   | ggataaaaat | gagctgg    | tc          | agaaggccaa  | actggccgag | 180 |
| caggctgagc | gatatgatga   | catggcagcc | tgcatga    | ctgtactga   | gcaaggagct  | 240        |     |
| gaattatcca | atgaggagag   | aatcttctc  | tcaatgtt   | ataaaaatgt  | tgttaggagcc | 300        |     |
| cgttagtcat | cttggagggt   | cgtctca    | attgaacaaa | agacgg      | tgctgagaaa  | 360        |     |
| aaacacgaga | tggctcgaga   | atacagagag | aaaattgaga | cgagacta    | agatatctgc  | 420        |     |
| aatgatgtac | tgtctttt     | ggaaaagt   | ttgatccc   | atgc        | tca         | agcagagagc | 480 |
| aaagtcttct | attgaaaat    | gaaaggagat | tactacc    | tt          | acttggctga  | ggttggccgt | 540 |
| <br>       |              |            |            |             |             |            |     |
| ggtgatgaca | agaaaggat    | tgtcgatc   | tcacaaca   | gatccaaga   | agctttgaa   | 600        |     |
| atcagcaaaa | aggaaatgca   | accaacacat | cctatc     | agac        | tgggtctggc  | 660        |     |
| tctgtttct  | attatgagat   | tctgaactcc | ccagagaa   | cctgctct    | tgcaaa      | 720        |     |
| gctttigatg | aagccattgc   | tgaacttgat | acatta     | agagt       | cata        | caaagacagc | 780 |
| acgctaataa | tgcattact    | gagagacaac | ttgacattgt | ggacatcg    | tacca       | agg        | 840 |
| gacgaagctg | aagcaggaga   | aggagggaa  | aattaacc   | ggttgc      | cttccaa     | ct         | 900 |

tcattctaaa atttacacag tagaccattt gtcatccatg ctgtcccaca aatagtttt 960  
  
tgtttacgat ttatgacagg tttatgttac ttctatttga atttctatat ttcccatgtg 1020  
gttttatgt ttaatattag gggagtagag ccagttaaaca tttagggagt tatctgttt 1080  
catcttgagg tggccaatat ggggatgtgg aattttata caagttataa gtgttggca 1140  
tagtacttt ggtacattgt ggcttcaaaa gggccagtgt aaaactgctt ccatgtctaa 1200  
gcaaagaaaa ctgcctacat actggttgt cctggcgggg aataaaaggg atcattggtt 1260  
ccagtcacag gtgttagtaat tgtggtact ttaaggtttg gagcacttac aaggctgtgg 1320  
tagaatcata ccccatggat accacatatt aaaccatgta tatctgtgga atactcaatg 1380  
  
  
tgtacacctt tgactacagc tgcagaagtg ttcctttaga caaagtgtg acccatttt 1440  
ctctggataa gggcagaaac ggttcacatt ccattattttaaagttacc tgctgttagc 1500  
tttcattatt tttgctacac tcattttatt tgtatttaaa tgttttaggc aacctaagaa 1560  
caaatgtaaa agtaaagatg cagaaaaat gaattgcttgcatttgcattatgtat 1620  
atcaagcaca gcagtaaaac aaaaacccat gtatttaact ttttttagg attttgctt 1680  
ttgtgatttt ttttttttg atactgcct aacatgcattg tgctgtaaaa atagttaaaca 1740  
gggaaataac ttgagatgat ggctagctt gttaatgtc ttatgaaatt ttcatgaaca 1800  
  
  
atccaagcat aattgttaag aacacgtgta ttaaattcat gtaagtggaa taaaagttt 1860  
atgaatggac ttttcaacta ctttctctac agctttcat gtaaatttagt ctgggtctg 1920  
aaacttctct aaagggaaatt gtacattttt tgaaatttat tccttattcc ctctggcag 1980  
ctaatggct cttaccaagt ttaaacacaa aatttatcat aacaaaaata ctactaatat 2040  
aactactgtt tccatgtccc atgateccct ctttcctcc ccaccctgaa aaaaatgagt 2100  
tccttatttt tctggagag ggggggattt attagaaaaa aatgtgtgt gttccattta 2160  
aaattttggc atatggcatt ttctacttta ggaagccaca atgttcttgg cccatcatga 2220  
  
  
cattgggttag cattaactgt aagtttgtt cttccaaatc acttttgggt tttttaagaat 2280  
ttcttgatac tcttatacgcc tgcctcaat tttgatcctt tattttctt atttgtcagg 2340  
tgcacaagat taccttccctt ttttagcctt ctgtttgtc accaaccatt ctacttgggt 2400  
ggccatgtac ttggaaaaag gccgcatgtat cttctggctt ccactcagtg tctaaggcac 2460  
cctgcttctt tgcattgtat cccacagactt attccctca tccttatttac tgcagcaat 2520  
ctctccttag ttgtatgagac tttgtttatc tccctttaaa accctaccta tcctgaatgg 2580  
tctgtcatgt tctgccttta aaatccttcc tcttttttcc tccttatttac tcttaataat 2640

|  |      |
|--|------|
| gatggggcta agttataccca aaagctcaact ttacaaaata ttcctcagt actttgcaga | 2700 |
| aaacaccaaa caaaaatgcc atttaaaaa aggtgtattt tttcttttag aatgtaaatc   | 2760 |
| cctcaagagc agggacaatg tttctgtat gttctattgt gcctagtaca ctgtaaatgc   | 2820 |
| tcaataaata ttgtatgtgg gaggcagtgta gtctigatga taagggttag aaactgaaat | 2880 |
| cccaaacact gtttgtgc ttgtttatt atgacctcag attaaattgg gaaatattgg     | 2940 |
| ccctttgaa taattgtccc aaatattaca ttcaaataaa agtgcaatgg agaaaaaaaaa  | 3000 |
| aaa  | 3003 |

<210> 132

<211> 1869

<212> DNA

<213> Homo sapiens

<400> 132

|   |     |
|---|-----|
| tacctggttg atccgtccag tagcatatgc ttgtctaaa gattaagcca tgcatgtcta  | 60  |
| agtacgcacg gccggtagac taaaaactgca aatggctcat taaatcgtt atgggtccct | 120 |
| tggcgctcg ctccctcccc acttggataa ctgtggtaat tctagagcta atacatgccc  | 180 |
| acggcgctg acccccttcg cggggggat gcgtgcattt atcagatcaa aaccaacccg   | 240 |
| gtcagccct ctccggcccc ggccgggggg cggcgccgg cggcttttgt gactctagat   | 300 |
| aacctcgggc cgatgcacg ccccccgtgg cggcgacgac ccattcgaac gtctgcccta  | 360 |

|   |     |
|---|-----|
| tcaactttcg atggtagtcg ccgtgcctac catggtagacc acgggtgacg gggaaatcagg | 420 |
| gttcgattcc ggagagggag cctgagaaac ggctaccaca tccaaggaag gcagcaggcg   | 480 |
| cgcaaattac ccactcccga cccggggagg tagtgacgaa aaataacaat acaggactct   | 540 |
| ttcgaggccc tgaatttggaa atgagtccac tttaaatcct ttaacgagga tccattggag  | 600 |
| ggcaagtctg gtgccagcag ccgcggtaat tccagctcca atagcgtata ttaaagtgc    | 660 |
| tgcagttaaa aagtcgttag ttggatcttggagggcg ggcggccgc cgcgaggcga        | 720 |
| gccaccggcc gtcggccccc ctgccttcg ggcggccctt cgatgcctt agctgagtg      | 780 |

|  |      |
|--|------|
| cccgcggggc cggaaagcggt tactttgaaa aaatttagagt gttcaagca ggcccgagcc | 840  |
| gcctggatac cgcaagctagg aataatggaa taggaccgcg gttctatTTT gttggTTTC  | 900  |
| ggaactgagg ccatgattaa gagggacggc cggggcatt cgtattgcgc cgctagaggt   | 960  |
| gaaattcttggacccggcga agacggacca gagcggaaagc atttgccaaatgtttca      | 1020 |
| ttaatcaaga acgaaagtgcg gaggttcgaa gacgatcaga taccgtcgta gttccgacca | 1080 |
| taaacgtgc cgaccggcga tgccggcggcgttattccat gaccggccgg gcagcttcgg    | 1140 |

|  |      |
|--|------|
| gaaaacccaaa gtcttgggt tcgggggga gtaggttgc aaagctgaaa cttaaaggaa    | 1200 |
| <br>   |      |
| ttgacggaag ggaccacca ggagtggagc ctgcggctta atttgactca acacggaaa    | 1260 |
| cctcacccgg cccggacacg gacaggattg acagattat agctttct cgattccgt      | 1320 |
| ggtgtggtg catggcggtt cttagttggt ggagcgattt gtctggtaa ttccgataac    | 1380 |
| gaacgagact ctggcatgct aactagttac gcgaccccg agcggtcgcc gtccccaa     | 1440 |
| ttcttagagg gacaagtggc gttcagccac ccgagattga gcaataacag gtctgtatg   | 1500 |
| cccttagatg tccgggctg cacgcgcgt acactgactg gctcagcgtg tgcctaccct    | 1560 |
| acgcggcag gcgcggtaa cccgttgaac cccattcgtg atgggatcg gggattgcaa     | 1620 |
| <br>   |      |
| ttattccccca tgaacgagga attcccagta agtgcgggtc ataagcttgc gttgattaag | 1680 |
| tccctgcctt ttgtacacac cgcccgtcgc tactaccat tggatggttt agtgaggccc   | 1740 |
| tcggatcgcc cccgcgggg tcggccacg gccctggcgg agcgctgaga agacggtcga    | 1800 |
| acttgactat cttagaggaag taaaagtctt aacaaggttt ccgtaggtga acctgcggaa | 1860 |
| ggatcatta  | 1869 |