



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2020년09월03일

(11) 등록번호 10-2151741

(24) 등록일자 2020년08월28일

(51) 국제특허분류(Int. Cl.)

C12Q 1/68 (2018.01) C12N 15/11 (2006.01)

(21) 출원번호 10-2013-0042328

(22) 출원일자 2013년04월17일

심사청구일자 2018년03월22일

(65) 공개번호 10-2014-0124600

(43) 공개일자 2014년10월27일

(56) 선행기술조사문헌

J Hepatol, 2009, 50(2): 358-369

(뒷면에 계속)

(73) 특허권자

연세대학교 산학협력단

서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

(72) 발명자

최형석

서울특별시 서초구 양재대로11길 19 LG전자 특허센터

허지연

서울특별시 서초구 양재대로11길 19 LG전자 특허센터

(뒷면에 계속)

(74) 대리인

특허법인 하나

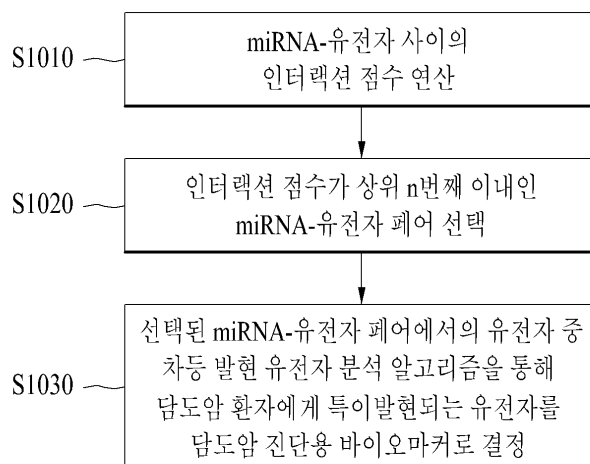
전체 청구항 수 : 총 7 항

심사관 : 이준혁

(54) 발명의 명칭 담도암 진단용 바이오마커, 이를 위한 컴퓨팅 장치 및 이의 제어 방법

**(57) 요약**

본 발명은 ACSM5, ADH6, ALDH1L1, APOA5, BHMT, CCL16, CYP1A2, CYP3A43, DAO, DDC, ESR1, F11, F13B, FETUB, GLYAT, GNMT, IGFALS, NAT2, PFKFB1, RDH16, SRD5A2, SULT2A1 및 THRSP를 포함하는 담도암 진단용 바이오마커를 개시한다.

**대표도** - 도10

(72) 발명자

**송시영**

서울특별시 마포구 백범로 205, 104동 1106호 (신공덕동, 마포펜트하우스)

**정다운**

서울특별시 마포구 월드컵북로 501, 916동 601호 (상암동, 상암월드컵파크9단지)

(56) 선행기술조사문헌

Plos one, 17 Aug 2011, 6(8):e2358

Int J Oncol, 2012, 41(6): 2213-2226 (2012.10.11.)\*

Ann Transl Med, 1(3):28 (2012.11.13.)\*

Nat Genet. 37(5): 495-500 (2005.04.03.)

Oncogene, 26(30): 4442-4452 (2007.01.22.)

\*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호 10040174

부처명 지식경제부

과제관리(전문)기관명 한국산업기술평가관리원

연구사업명 바이오의료기기 산업원천기술개발사업(바이오)

연구과제명 차세대 생명정보 통합분석을 이용한 다중 바이오마커 개발

기 여 율 1/1

과제수행기관명 (주)지노믹트리

연구기간 2011.06.01 ~ 2016.05.31

## 명세서

### 청구범위

#### 청구항 1

ACSM5, ADH6, ALDH1L1, APOA5, BHMT, CCL16, CYP1A2, CYP3A43, DAO, DDC, ESR1, F11, F13B, FETUB, GLYAT, GNMT, IGFALS, NAT2, PFKFB1, RDH16, SRD5A2, SULT2A1 및 THRSP를 포함하는 담도암 진단용 바이오마커 (biomarker) 조성물.

#### 청구항 2

마이크로 RNA와 유전자 사이 상보 결합이 발생할 가능성의 고저를 수치화한 인터랙션 점수를 연산하는 단계;

상기 인터랙션 점수가 기설정된 임계치 이상인 n 개의 마이크로 RNA 및 유전자 페어를 결정하는 단계; 및

상기 n 개의 마이크로 RNA 및 유전자 페어 중 담도암 환자에게서 특이발현되는 유전자와 공통되는 유전자를 추출하는 단계

를 포함하고,

상기 인터랙션 점수를 연산하는 단계는,

마이크로 RNA와 유전자 사이의 예측 점수를 통계화한 적어도 하나 이상의 데이터 베이스를 획득하는 단계; 마이크로 RNA와 유전자 사이의 상기 예측 점수를 기준으로 순위를 매긴 뒤, 이를 표준 점수로 변환하고, 변환된 표준 점수를 합하여 정규화된 점수를 연산하는 단계; 상기 정규화된 점수를 기초로 유전자별 마이크로 RNA의 결합 순위 및 마이크로 RNA별 유전자의 결합 순위를 연산하는 단계; 및 상기 마이크로 RNA의 결합 순위 및 상기 유전자의 결합 순위를 기초로 인터랙션 점수를 연산하는 단계를 포함하고,

상기 정규화된 점수는, 상기 하나 이상의 데이터 베이스에서의 마이크로 RNA 및 유전자 페어의 예측 점수를 기준으로 순위를 매긴 뒤, 이를 표준 점수로 변환하고, 변환된 표준 점수를 합하여 얻어지며, 하기 수학식에 의거 연산되는 것을 특징으로 하는 담도암 진단용 바이오마커 추출 방법.

[수학식]

$$\sum_{i=1}^n \frac{(T_i+1-R_{ij})}{T_i}$$

(i는 i번째 데이터 베이스, n은 데이터 베이스의 개수,  $T_i$ 는 i번째 데이터 베이스에서 마이크로 RNA-유전자 페어의 총 개수,  $R_{ij}$ 는 i번째 데이터베이스에서, j번째 마이크로 RNA-유전자 페어의 예측 점수 순위를 의미함)

#### 청구항 3

삭제

#### 청구항 4

제 2 항에 있어서,

상기 하나 이상의 데이터 베이스는 마이크로 RNA 타겟 예측 툴을 이용하여 생성되는 것을 특징으로 하는 담도암 진단용 바이오마커 추출 방법.

#### 청구항 5

제 4 항에 있어서,

상기 마이크로 RNA 타겟 예측 툴은 Targetscan, miRDB, DIANA-microT, PITA, miRanda MicroCosm, RNAhybrid,

PicTar 및 RNA22 중 적어도 하나를 포함하는 것을 특징으로 하는 담도암 진단용 바이오마커 추출 방법.

#### 청구항 6

삭제

#### 청구항 7

삭제

#### 청구항 8

제 2 항에 있어서,

상기 인터랙션 점수는, 상기 정규화된 점수를 기초로 유전자별 마이크로 RNA의 순위 및 마이크로 RNA별 유전자의 상보적인 결합력이 강한 정규화 점수가 기설정된 임계치 이상인 순위를 기초로 연산되는 것을 특징으로 하는 담도암 진단용 바이오마커 추출 방법.

#### 청구항 9

제 8 항에 있어서,

상기 인터랙션 점수는 하기 수학식에 의거 연산되는 것을 특징으로 하는 담도암 진단용 바이오마커 추출 방법.

[수학식]

$$\left( \frac{t_{mi} + 1 - r_{mi}}{t_{mi}} \right) \times \left( \frac{t_{gj} + 1 - r_{gj}}{t_{gj}} \right)$$

( $t_{mi}$ 는 i번째 마이크로 RNA의 유전자와의 페어의 수(number of miRNA<sub>i</sub>-gene),  $t_{gj}$ 는 j번째 유전자의 마이크로 RNA와의 페어의 수(number of gene<sub>j</sub>-miRNA),  $r_{mi}$ 는 j번째 유전자에 대한 i번째 마이크로 RNA의 정규화 점수 순위,  $r_{gj}$ 는 i번째 마이크로 RNA에 대한 j번째 유전자의 정규화 점수 순위를 의미함)

#### 청구항 10

데이터를 저장하기 위한 저장부; 및

연산을 위한 제어부를 포함하고,

상기 제어부는, 마이크로 RNA와 유전자 사이 상보 결합 정도를 수치화한 정규화 점수가 기설정된 임계치 이상인 인터랙션 점수를 연산하고, 상기 인터랙션 점수가 기설정된 임계치 이상인 n 개의 마이크로 RNA 및 유전자 페어를 결정하며, 상기 n 개의 마이크로 RNA 및 유전자 페어 중 담도암 환자에게서 특이발현되는 유전자와 공통되는 유전자를 추출하도록 구성되는 것을 특징으로 하고,

상기 제어부는,

마이크로 RNA와 유전자 사이의 예측 점수를 통계화한 적어도 하나 이상의 데이터 베이스를 획득하고, 마이크로 RNA와 유전자 사이의 상기 예측 점수를 기준으로 순위를 매긴 뒤, 이를 표준 점수로 변환하고, 변환된 표준 점수를 합하여 정규화된 점수를 연산하고, 상기 정규화된 점수를 기초로 유전자별 마이크로 RNA의 결합 순위 및 마이크로 RNA별 유전자의 결합 순위를 연산하고, 상기 마이크로 RNA의 결합 순위 및 상기 유전자의 결합 순위를 기초로 인터랙션 점수를 연산하고,

상기 정규화된 점수는, 상기 하나 이상의 데이터 베이스에서의 마이크로 RNA 및 유전자 페어의 예측 점수를 기준으로 순위를 매긴 뒤, 이를 표준 점수로 변환하고, 변환된 표준 점수를 합하여 얻어지며, 하기 수학식에 의거 연산되는 것을 특징으로 하는 컴퓨팅 장치.

[수학식]

$$\sum_{i=1}^n \frac{(T_i+1-R_{ij})}{T_i}$$

( $i$ 는  $i$ 번째 데이터 베이스,  $n$ 은 데이터 베이스의 개수,  $T_i$ 는  $i$ 번째 데이터 베이스에서 마이크로 RNA-유전자 페어의 총 개수,  $R_{ij}$ 는  $i$ 번째 데이터베이스에서,  $j$ 번째 마이크로 RNA-유전자 페어의 예측 점수 순위를 의미함)

## 발명의 설명

### 기술 분야

[0001] 본 발명은 담도암 진단용 바이오마커, 이를 위한 컴퓨팅 장치 및 이의 제어 방법에 관한 것이다.

### 배경 기술

[0002] 담관은 간에서 만들어지는 담즙을 십이지장으로 보내는 관으로서, 간 속에서 나뭇가지가 하나의 가지를 향해 모이듯이 서서히 합류하면서 굽어지며, 간에서 나올 때에 좌우의 담관이 대부분 하나로 합류하게 된다. 담관은 간 속을 지나는 간내 담관과 간을 벗어나 십이지장까지 이어지는 간의 담관으로 나뉜다. 간의 담관 중 담즙을 일시적으로 저장하여 농축하는 주머니를 담낭이라 부르며, 이들 간내외 담관과 담낭을 통틀어 담도라고 부른다.

[0003] 담도암은 담관암이라고도 하며, 담관의 상피에서 발생하는 악성종양으로서 발생 부위에 따라 간내 담도암과 간의 담도암의 두 종류로 나뉘는데, 일반적으로 담도암이라고 하면 주로 간의 담관에 발생한 암을 가리킨다. 본 명세서에서는 달리 지시되지 않는 한 간내 담도암 및 간의 담도암을 모두 지칭한다.

[0004] 담도암은 주위의 조직에 스며들듯이 퍼지는 일이 많고, 명료한 종양 덩어리를 형성하지 않으므로, 그 덩어리를 정확하게 확인하고 진단하는 것은 쉽지 않다. 최근에는 화상진단기술이 발달함에 따라 복부 초음파검사, 컴퓨터 단층촬영(CT), 자기공명영상(MRI), 경피경간담도조영(PTC), 경피경간담도배액술(PTBD), 내시경 역행성 담췌관 조영술(ERCP) 또는 혈관조영검사 등의 기술을 이용해 담도암을 진단하고 있다. 그러나 이러한 화상진단기술은 진단에 고비용이 들고 복잡하며, 사실상 조기 진단용으로는 소용이 없기 때문에, 특히 조기 진단을 위한 담도암 진단용 바이오마커의 개발이 절실하다.

[0005] 이와 관련하여, 지난 20년 동안 다른 암종에 대한 수십 개의 바이오마커가 나왔으나, 담도암의 바이오마커는 아직까지 상용화되어 있는 것이 없는 실정이다.

## 발명의 내용

### 해결하려는 과제

[0006] 본 발명에서는 높은 특이성 및 민감성을 갖춘 담도암의 진단용 바이오마커를 제공하고자 한다. 또한, 본 발명에서는 담도암 진단용 바이오마커를 발굴하는 방법을 제공하고자 한다.

[0007] 본 발명에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

### 과제의 해결 수단

[0008] 본 발명의 일예에 따른 담도암 진단용 바이오마커 군은 ACSM5, ADH6, ALDH1L1, APOA5, BHMT, CCL16, CYP1A2, CYP3A43, DAO, DDC, ESR1, F11, F13B, FETUB, GLYAT, GNMT, IGFALS, NAT2, PFKFB1, RDH16, SRD5A2, SULT2A1 및 THRSP를 포함한다.

[0009] 본 발명의 일예에 따른 담도암 진단용 바이오마커 추출 방법은, 마이크로 RNA와 유전자 사이 상보 결합 정도를 수치화한 인터랙션 점수를 연산하는 단계; 상기 인터랙션 점수가 높은  $n$  개의 마이크로 RNA 및 유전자 페어를 결정하는 단계; 및 상기  $n$  개의 마이크로 RNA 및 유전자 페어 중 담도암 환자에게서 특이발현되는 유전자와 공통되는 유전자를 추출하는 단계를 포함한다.

[0010] 본 발명에서 제시하는 해결 수단은 이상에서 언급한 해결 수단들로 제한되지 않으며, 언급하지 않은 또 다른 해결 수단들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

### 발명의 효과

[0011] 본 발명에서는 높은 특이성 및 민감성을 갖춘 담도암의 진단용 바이오마커를 제공할 수 있다. 또한, 본 발명에서는 담도암 진단용 바이오마커를 발굴하는 방법을 제공할 수 있다.

[0012] 본 발명에서 이루고자 하는 기술적 효과들은 이상에서 언급한 기술적 효과들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 효과들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

### 도면의 간단한 설명

- [0013] 도 1은 본 발명에 따른 컴퓨팅 장치의 블록도이다.
- 도 2는 miRNA와 유전자 사이의 인터랙션 점수를 연산하는 것의 일예를 설명하기 위한 개념도다.
- 도 3은 인터랙션 점수의 연산 방법의 흐름도이다.
- 도 4는 유사도 데이터베이스를 이용하여 유사 miRNA 및 특정 유전자 사이의 상관관계 값을 연산하는 방법을 설명하기 위한 개념도이다.
- 도 5는 유사도 데이터베이스를 이용하여 유사 miRNA 및 유전자 사이의 상관관계 값을 연산하는 방법의 흐름도이다.
- 도 6은 miRNA 클러스터 데이터베이스를 이용하여 인접 miRNA 및 특정 유전자 사이의 상관관계 값을 연산하는 방법을 설명하기 위한 개념도이다.
- 도 7은 miRNA 클러스터 데이터베이스를 이용하여 인접 miRNA 및 특정 유전자 사이의 가중치 값을 연산하는 방법의 흐름도이다.
- 도 8은 전사인자 데이터베이스를 이용하여 특정 miRNA 및 전사조절 유전자 사이의 상관관계 값을 연산하는 방법을 설명하기 위한 개념도이다.
- 도 9는 전사인자 데이터베이스를 이용하여 특정 miRNA 및 전사조절 유전자 사이의 가중치 값을 연산하는 방법의 흐름도이다.
- 도 10은 바이오마커 발굴용 통합 분석 알고리즘을 기초로 담도암 환자용 바이오마커를 발굴하는 방법의 흐름도이다.
- 도 11은 데이터 GSE26566을 이용한 주성분 분석 결과를 플롯(Plot) 한 것이다.
- 도 12는 데이터 GSE26566을 이용한 계층적군집분석 결과를 히트 맵(Heat Map)으로 도시한 것이다.

### 발명을 실시하기 위한 구체적인 내용

- [0014] 이하, 본 발명과 관련된 컴퓨팅 장치에 대하여 도면을 참조하여 보다 상세하게 설명한다.
- [0015] 이하의 설명에서 사용되는 구성요소에 대한 접미사 "모듈" 및 "부"는 명세서 작성의 용이함만이 고려되어 부여되거나 혼용되는 것으로서, 그 자체로 서로 구별되는 의미 또는 역할을 갖는 것은 아니다.
- [0016] 본 발명에서는 바이오마커 발굴용 통합 분석 알고리즘을 적용한 컴퓨팅 장치(100) 및 상기 컴퓨팅 장치(100)를 통해 발굴된 바이오마커를 개시한다. 여기서, 설명하는 컴퓨팅 장치(100)는 개인용 컴퓨터, 워크 스테이션, 슈퍼 컴퓨터 등 전자 회로를 이용한 고속의 연산 처리 장치를 포함할 수 있다. 비단, 컴퓨터, 워크 스테이션, 슈퍼 컴퓨터 등의 고정형 장치가 아니더라도, 중앙 처리 장치(Central Processing Unit)를 포함하고 연산 처리를 수행할 수 있는 스마트 폰, PDA, 랩탑 등의 이동형 장치도 컴퓨팅 장치에 포함될 수 있다.
- [0017] 도 1은 본 발명에 따른 컴퓨팅 장치의 블록도이다. 도 1을 참조하면, 본 발명에 따른 컴퓨팅 장치(100)는 저장부(110), 사용자 입력부(120), 통신부(130) 및 제어부(140)를 포함할 수 있다.

- [0018] 저장부(110)는 제어부(140)의 동작을 위한 프로그램을 저장할 수 있고, 입/출력되는 데이터들(예를 들어, 데이터베이스 등)을 임시 저장할 수도 있다. 나아가 저장부(110)는 통신부(130)가 통신을 수행하면서 송수신하는 데이터들을 저장할 수도 있다.
- [0019] 저장부(110)는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 저장부(110) 등), 램(Random Access Memory, RAM), SRAM(Static Random Access Memory), 롬(Read-Only Memory, ROM), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory), 자기 저장부(110), 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다.
- [0020] 사용자 입력부(120)는 사용자로부터 사용자 입력을 수신하는 역할을 수행한다. 사용자 입력부(120)는 키보드, 마우스 등을 포함할 수 있다.
- [0021] 통신부(130)는 통신을 통해 외부로부터 데이터를 수신하거나 외부로 데이터를 송신하는 역할을 수행한다. 본 발명에 의한 통신부(130)는 원격 서버로부터 각종 데이터베이스를 수신하는 역할을 수행할 수 있다.
- [0022] 제어부(140)는 컴퓨팅 장치(100)의 전반적인 동작을 제어하고, 각종 연산을 수행한다. 본 발명에 의한 제어부(140)는 후술할 인터랙션 점수 및 상관관계 값 등을 연산하고, 담도암 진단용 바이오마커를 추출하기 위한 연산을 수행할 수 있다.
- [0023] 본 발명에 의한 컴퓨팅 장치(100)는 정보의 출력을 위한 디스플레이부(150)를 더 포함할 수도 있다. 디스플레이부(150)는 사용자의 입력을 표시하고, 제어부(140)의 연산 결과를 출력하는 출력 장치로서의 역할을 수행할 수 있다. 디스플레이부(150)는 컴퓨팅 장치(100)를 보조하는 모니터 등의 장치일 수 있다.
- [0024] 상술한 컴퓨팅 장치(100)를 이용하여, 담도암 진단용 바이오마커 발굴 방법에 대해 상세히 설명하기로 한다.
- [0025] 본 발명에서 설명하는 바이오마커 발굴용 통합 분석 알고리즘은 차등 발현 유전자(Differentially Expressed Genes) 분석 알고리즘과 마이크로 RNA 표적 유전자 분석 알고리즘이 병합되어 구성된 형태일 수 있다.
- [0026] 먼저 차등 발현 유전자 알고리즘에 대해 설명하기로 한다. 차등 발현 유전자 알고리즘은 담도암 환자에게서 정상인과 다르게 과발현(over-expression)되거나, 저발현(low-expression)되는 유전자를 통계적으로 유의미하게 찾아내기 위한 알고리즘으로써, 다양한 요인들을 고려할 수 있는 고급 통계 방법 중 하나인 선형 모형(linear model)을 이용하여 정상인 그룹과 환자 그룹을 구분할 수 있는 유전자들을 찾는 것을 목적으로 한다(참고문헌 *Statistical Applications in Genetics and Molecular Biology*, Vol. 3, No. 1, Article 3).
- [0027] 차등 발현 유전자 분석 알고리즘은 크게 데이터 표준화(normalization) 단계와 통계 분석 단계로 구분될 수 있다. 데이터 표준화 단계는 정상인 그룹과 환자 그룹으로부터 얻어진 인간의 전체 유전자에 대한 마이크로 어레이 데이터를 통합하고 보정하는 단계이다. 데이터 표준화를 위해, 로버스트 멀티 칩 평균(Robust Multichip Average, RMA) 알고리즘이 이용될 수 있다(참고문헌 *Biostatistics*, Vol. 4, No. 2, 249-264).
- [0028] 통계분석 단계는 표준화된 데이터를 선형모형을 이용하여 두 그룹(즉, 정상인 그룹과 환자 그룹) 사이에서 통계적으로 유의미하게 발현량에 차이가 나는 유전자를 선별하는 단계이다. 통계적 유의 확률은 FDR (False Discovery) 방법(참고문헌 *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 57, No. 1, 289-300)을 이용하여 보정된 p-value인 q-value가 0.01 이하인 유전자를 선택할 수 있다.
- [0029] 본 발명의 컴퓨팅 장치(100)는 담도암 진단용 바이오마커를 발굴하기 위해, 차등 발현 유전자 분석 알고리즘을 이용한 담도암 환자에게서 특이발현(과발현 또는 저발현)되는 유전자 목록을 이용할 수 있다. 차등 발현 유전자 분석 알고리즘을 이용하여 담도암 환자에게서 특이발현되는 유전자 목록을 발굴하는 것은 이미 공지된 기술이므로, 이에 대한 상세한 설명은 생략한다.
- [0030] 다음으로 마이크로 RNA 표적 유전자 분석 알고리즘에 대해 설명하기로 한다. 본 발명에서 설명하는 마이크로 RNA 표적 유전자 분석 알고리즘은 기존의 마이크로 RNA 데이터베이스에서 얻어진 마이크로 RNA 표적 유전자 예측 계산값, 마이크로 어레이 실험을 통해 얻어진 마이크로 RNA와 유전자의 발현양상에 대한 상관관계 계산값 및 생물학적 메커니즘에 따른 가중치 계산값 중 적어도 하나를 이용하여 마이크로 RNA의 표적 유전자를 정확하게 찾아내는 통계 방정식을 제공하기 위한 것이다.
- [0031] 이하에서는, 마이크로 RNA 표적 유전자 예측 계산값(또는 인터랙션 점수), 상관관계 계산값 및 가중치 계산값의 연산 방법에 대해 상세히 설명하기로 한다. 설명의 편의를 위해, 본 발명에서 miRNA 및 gene이라 기재된 것은



각각 마이크로 RNA 및 유전자와 동등한 의미인 것으로 가정한다.

[0032] **마이크로 RNA 표적 유전자 예측값 연산**

[0033] 본 발명에 따른 컴퓨팅 장치(100)는 마이크로 RNA와 이의 표적 유전자 사이의 상보 결합 정도를 수치화한 인터랙션 점수(Interaction Score)를 연산할 수 있다. 인터랙션 점수를 통해 마이크로 RNA와 이의 표적 유전자 사이의 상보 결합이 발생할 가능성의 고저를 판단할 수 있다. 후술되는 도면을 참조하여 인터랙션 점수의 연산 방법에 대해 상세히 설명한다.

[0034] 도 2는 miRNA와 유전자 사이의 인터랙션 점수를 연산하는 것의 일예를 설명하기 위한 개념도이고, 도 3은 인터랙션 점수의 연산 방법의 흐름도이다.

[0035] 도 2 및 도 3을 참조하면, 먼저, 컴퓨팅 장치(100)는 적어도 하나 이상의 miRNA 타겟 예측(miRNA Target Prediction) 툴을 이용하여 miRNA와 유전자 사이의 예측 점수(Prediction score)를 통계화한 데이터 베이스를 획득할 수 있다(S310).

[0036] miRNA 타겟 예측 툴은 표적 유전자와, 상기 표적 유전자에 상보적으로 결합하여 표적 유전자가 단백질로 만들어지는 과정을 억제할 수 있는 miRNA 페어(pair)의 결합의 정도를 수치화한 소프트웨어적 툴을 의미할 수 있다. 유전자-miRNA 페어들의 예측 점수를 획득하기 위한 miRNA 타겟 예측 툴로는, Targetscan, miRDB, DIANA-microT, PITA, miRanda MicroCosm, RNAhybrid, PicTar, RNA22 등이 포함될 수 있다. 각각의 miRNA 타겟 예측 툴에 대한 간략한 설명이 표 1에 기재되어 있다.

**표 1**

[0037]

툴 이름	툴 설명(이용 정보)	참조 사이트
Targetscan	서열 유사성(sequence similarity) 정보와 보존(conservation) 정보 이용	<a href="http://www.ncbi.nlm.nih.gov/pubmed/18955434">http://www.ncbi.nlm.nih.gov/pubmed/18955434</a>
miRDB	서열 유사성 정보, 열역학적 안정성(thermodynamic stability) 정보 및, 보존(conservation) 정보 이용	<a href="http://www.ncbi.nlm.nih.gov/pubmed/18426918">http://www.ncbi.nlm.nih.gov/pubmed/18426918</a>
DIANA-microT	서열 유사성 정보 및 열역학적 안정성 정보 이용	<a href="http://www.ncbi.nlm.nih.gov/pubmed/15131085">http://www.ncbi.nlm.nih.gov/pubmed/15131085</a>
PITA	서열 유사성 정보 및 열역학적 안정성 정보 이용	<a href="http://www.ncbi.nlm.nih.gov/pubmed/17893677">http://www.ncbi.nlm.nih.gov/pubmed/17893677</a>
miRanda	열역학적 안정성 및 보존 정보 이용	<a href="http://www.ncbi.nlm.nih.gov/pubmed/14709173">http://www.ncbi.nlm.nih.gov/pubmed/14709173</a>
MicroCosm	열역학적 안정성 정보 및 보존 정보 이용	<a href="http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/info.html">http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/info.html</a>
RNAhybrid	열역학적 안정성 정보 이용	<a href="http://www.ncbi.nlm.nih.gov/pubmed/15383676">http://www.ncbi.nlm.nih.gov/pubmed/15383676</a>
PicTar	서열 유사성 정보와 보존 정보 이용	<a href="http://www.ncbi.nlm.nih.gov/pubmed/15806104">http://www.ncbi.nlm.nih.gov/pubmed/15806104</a>
RNA22	서열 패턴(sequence pattern) 정보 이용	<a href="http://www.ncbi.nlm.nih.gov/pubmed/16990141">http://www.ncbi.nlm.nih.gov/pubmed/16990141</a>

[0038] 타겟 예측 툴을 이용하면, miRNA와 상보 결합할 수 있는 다양한 유전자 사이의 예측 점수가 스코어링 될 수 있다. 예측 점수가 작을수록 miRNA와 유전자 사이의 상보 결합 가능성이 낮아지는 것을 의미할 수 있다.

[0039] 타겟 예측 툴은 본 발명에 의한 컴퓨팅 장치(100)에 의해 구동되고, 제어부(140)의 연산 처리에 의해 miRNA-유전자 페어의 예측 점수를 통계화한 데이터 베이스가 획득될 수도 있으나, 이에 한정되는 것은 아니다. 본 발명에 의한 컴퓨팅 장치(100)는 타겟 예측 툴을 이용하는 원격의 서버로부터 miRNA-유전자 페어의 예측 점수를 통계화한 데이터 베이스를 획득할 수도 있다.

[0040] miRNA-유전자 페어 사이의 예측 점수의 신뢰성을 높이기 위해, 하나의 타겟 예측 툴을 이용하는 것 보다 복수의 타겟 예측 툴을 이용하여 복수의 데이터베이스를 획득하는 것이 바람직하다. 도 2에서는 타겟 예측 툴로, PITA, DIANA-microT, TargetScan, MicroCosm, miRDB 및 miRanda가 이용된 것으로 예시되었다.

[0041] 복수의 타겟 예측 툴을 이용하여, miRNA-유전자 페어의 예측 점수를 통계화한 복수의 데이터 베이스를 획득한



경우, 데이터베이스의 정규화(normalization)하기 위해, 제어부(140)는 miRNA-유전자 페어의 예측 점수의 순위를 기초로, 정규화된 점수를 연산할 수 있다(S320).

[0042] 표 1에 도시된 예에서와 같이, miRNA 타겟 예측 틀이 이용하는 정보들이 다르고, 각 데이터베이스마다 예측 점수를 매기는데 서로 다른 단위가 적용될 수 있기 때문에, 복수의 데이터 베이스를 이용하고자 하는 경우 이의 정규화가 필수적이라 할 것이다. miRNA-유전자 페어의 예측 점수를 정규화 하기 위해, 제어부(140)는 각 데이터 베이스별로 miRNA-유전자 페어들의 예측 점수를 기준으로 순위를 매긴 뒤, 이를 표준 점수로 변환하고, 각 데이터베이스에서의 miRNA-유전자 페어들의 표준 점수를 합하여 정규화된 점수를 획득할 수 있다. 수학식 1은 정규화된 점수를 획득하는데 이용되는 수학적식을 예시한 것이다.

### 수학식 1

$$\sum_{i=1}^n \frac{(T_i+1-R_{ij})}{T_i}$$

[0043]

[0044] 수학식 1에서, i는 i번째 데이터 베이스, n은 데이터 베이스의 개수(일례로, 도 2에서는 6개의 예측 틀을 통해 6개의 데이터 베이스를 획득하였으므로, n은 6으로 설정될 수 있을 것임),  $T_i$ 는 i번째 데이터 베이스에서 miRNA-유전자 페어의 총 개수,  $R_{ij}$ 는 i번째 데이터베이스에서, j번째 miRNA-유전자 페어의 순위를 의미하는 것 일 수 있다.

[0045] 예컨대, 100개의 miRNA-유전자 페어가 존재하는 제 1 데이터베이스에서, miRNA1-gene1 페어의 예측 점수가 100 개의 페어 중 20위에 해당한다면, 제 1 데이터베이스에 miRNA1-gene1 페어의 표준 점수는  $(100+1-20)/100=0.81$  이 될 것이다. 제어부(140)는, 제 2 내지 제 n 데이터베이스에서 miRNA1-gene1 페어의 표준 점수를 합하여, miRNA1-gene1 페어의 정규화 점수를 연산할 수 있다.

[0046] 이후, 제어부(140)는 정규화 점수를 기초로, 특정 유전자에 대한 miRNA 들의 순위 및 특정 miRNA에 대한 유전자 들의 순위를 매길 수 있다(S330).

[0047] 예컨대, gene1과 상보결합 가능한 miRNA가 miRNA1, miRNA3, miRNA4 일때, 제어부(140)는 gene1-miRNA1, gene1-miRNA3 및 gene1-miRNA4 각각의 정규화 점수를 기초로, gene1과의 상보적인 결합력이 강한(즉, 정규화 점수가 높은) 순서대로 miRNA의 순위를 결정할 수 있다. 도 2에서는 miRNA1-gene1 사이의 정규화 점수가 0.4이고, miRNA3-gene1 사이의 정규화 점수가 0.6으로 설정되어, gene1에 대해 miRNA1은 두번째 순위를 갖고, miRNA3은 첫번째 순위를 갖는 것으로 예시되었다.

[0048] 이와 같은 방법으로 특정 miRNA에 대한 유전자들의 순위를 결정할 수 있다. 예컨대, miRNA1과 상보결합 가능한 gene이 gene1 및 gene3인 경우, 제어부(140)는 miRNA1-gene1 및 miRNA1-gene3 각각의 정규화 점수를 기초로, miRNA1과의 상보적인 결합력이 강한(즉, 정규화 점수가 높은) 순서대로 gene의 순위를 결정할 수 있다. 도 2에서는 miRNA1-gene1 사이의 정규화 점수가 0.4이고, miRNA1-gene3 사이의 정규화 점수가 0.5로 설정되어, miRNA1에 대해 gene1은 두번째 순위를 갖고, gene3은 첫번째 순위를 갖는 것으로 예시되었다.

[0049] 이후, 제어부(140)는 유전자 및 miRNA 들의 순위를 기초로, 유전자-miRNA 사이의 인터랙션 점수(Interaction Score)를 연산할 수 있다(S340). 수학식 2는 인터랙션 점수를 연산하기 위해 이용되는 수학적식을 예시한 것이다.

### 수학식 2

$$\left(\frac{t_{mi}+1-r_{mi}}{t_{mi}}\right) \times \left(\frac{t_{gj}+1-r_{gj}}{t_{gj}}\right)$$

[0050]

[0051]  $t_{mi}$ 는 i번째 miRNA의 유전자와의 페어의 수(number of miRNA<sub>i</sub>-gene),  $t_{gj}$ 는 j번째 gene의 miRNA와의 페어의 수

(number of gene<sub>j</sub>-miRNA),  $r_{mi}$ 는 j번째 유전자에 대한 i번째 miRNA의 정규화 점수 순위,  $r_{gj}$ 는 i번째 miRNA에 대한 j번째 유전자의 정규화 점수 순위를 의미할 수 있다.

[0052] **상관관계 연산**

[0053] 상술한 타겟 miRNA 예측 툴이 인체의 모든 miRNA 및 모든 유전자와 관련된 데이터베이스를 갖고 있지는 않다. 본 발명에서는 miRNA 사이의 유사도, miRNA 사이의 주변 영향력 및 유전자 들의 전사인자 등을 이용하여, 타겟 miRNA 예측 툴로부터 예측될 수 없는 다양한 miRNA 및 유전자 들의 인터랙션 점수를 획득할 수도 있다.

[0054] **실시예 1 - 상관관계에 기인한 가중치 연산**

[0055] 본 발명에 의한 컴퓨팅 장치(100)는 마이크로 어레이 실험을 통해 얻어진 특정 miRNA와 특정 유전자의 발현 양상에 대한 상관관계 값을 획득하고, 특정 miRNA와 유사한 유사 miRNA와 특정 유전자 사이의 상관관계 값을 예측할 수 있다. 후술되는 도면을 참조하여, 유사 miRNA와 특정 유전자 사이의 상관관계 값의 연산에 대해 상세히 설명한다.

[0056] 도 4는 유사도 데이터베이스를 이용하여 유사 miRNA 및 특정 유전자 사이의 상관관계 값을 연산하는 방법을 설명하기 위한 개념도이고, 도 5는 유사도 데이터베이스를 이용하여 유사 miRNA 및 유전자 사이의 상관관계 값을 연산하는 방법의 흐름도이다.

[0057] 먼저, 마이크로 어레이 실험을 통해 얻어진 유전자 발현 프로파일(Genes Expression Profiles) 및 miRNA 발현 프로파일(miRNAs Expression Profiles)을 포함하는 실험 데이터가 입력되면(S510), 제어부(140)는 입력된 실험 데이터를 기초로, 특정 miRNA와 특정 유전자 사이의 상관관계(Correlation)를 연산할 수 있다(S520).

[0058] 우선적으로 마이크로 어레이 실험에 대해 설명하면, 유전자 마이크로어레이(microarray)란, 생명체의 유전자 전체 또는 일부에 대해 유전자 발현 양을 측정할 수 있는 도구로서, DNA 마이크로어레이라고도 한다. 유전자 마이크로어레이의 도입으로 유전자의 관찰 능력이 개개 유전자 수준에서 생명체 전체로 확장되어, 생명체를 하나의 시스템으로 연구할 수 있다. 또한 유전자 마이크로어레이는 기본적으로 기존의 유전자 검출 기법을 병렬화해서 대규모로 수행하는 것이므로, 데이터를 처리하고 분석하는 방식에도 획기적인 변화를 가져왔다. 유전자 마이크로어레이의 수행 방법은 일반적으로 먼저 1cm<sup>2</sup> 정도의 슬라이드 표면에 수천 내지 수만 종류의 유전자 서열을 고정시킨 후, 다양한 실험 조건에서 채취한 세포의 RNA를 추출하고 DNA로 역전사하여 형광 물질로 표지한다. 이어서, 표지된 DNA를 마이크로어레이에 혼성화(hybridization)시키고, 이를 스캐닝하여 이미지로 형상화한 후 이미지 분석 프로그램으로 유전자 위치마다 형광 물질에 따른 발색 강도를 측정함으로써, 유전자 발현 여부 및 발현 정도를 수학, 통계 그리고 컴퓨터 공학과 같은 정보학을 이용하여 정량화된 유전자 발현 수치 데이터와 대조 분석한다.

[0059] 위와 같은 마이크로 어레이 실험을 통해, 특정 miRNA와 특정 유전자 사이의 발현 정도가 수치화 될 수 있다. 특정 miRNA와 특정 유전자 사이의 상관관계는 피어슨 상관관계(Pearson's Correlation)로, 특정 유전자의 발현량 증가에 따른, 특정 miRNA의 발현량의 증감의 상대량을 나타내는 것일 수 있다.

[0060] 이후, 컴퓨팅 장치(100)는 miRNA 유사도 데이터베이스(miRNA Similarity DataBase)를 이용하여 특정 miRNA와 유사한 유사 miRNA의 유사도값을 획득할 수 있다(S530). miRNA 유사도 데이터베이스는 miRNA 사이의 기능적 유사도를 수치화한 유사도 값을 포함할 수 있다. miRNA 유사도 데이터베이스는 이미 공지된 BLAST 또는 BLAT 툴을 통해 획득된 것일 수 있다.

[0061] 이후, 컴퓨팅 장치(100)는 유사도 값을 이용하여 유사 miRNA와 특정 유전자 사이의 상관관계를 연산할 수 있다(S540). 유사 miRNA와 유전자 사이의 가중치를 연산하는 데에는 유사도 값을 이용한 선형 회귀 분석(Linear regression model)이 사용될 수 있다.

[0062] **실시예 2 - miRNA 주변 영향력을 고려한 상관관계 연산**

[0063] 본 발명에 따른 컴퓨팅 장치(100)는 특정 miRNA와 클러스터를 형성하는 인접 miRNA에 대한 상관관계 값을 연산할 수도 있다. miRNA 들 사이의 영향력을 고려하여 상관관계 값을 연산하는 것에 대해서는 후술되는 도면을 참조하기로 한다.

[0064] 도 6은 miRNA 클러스터 데이터베이스를 이용하여 인접 miRNA 및 특정 유전자 사이의 상관관계 값을 연산하는 방법을 설명하기 위한 개념도이고, 도 7은 miRNA 클러스터 데이터베이스를 이용하여 인접 miRNA 및 특정 유전자 사이의 가중치 값을 연산하는 방법의 흐름도이다.

- [0065] 먼저, 마이크로 어레이 실험을 통해 얻어진 유전자 발현 프로파일(Genes Expression Profiles) 및 miRNA 발현 프로파일(miRNAs Expression Profiles)을 포함하는 실험 데이터가 입력되면(S710), 제어부(140)는 입력된 실험 데이터를 기초로, 특정 miRNA와 특정 유전자 사이의 상관관계(Correlation)를 연산할 수 있다(S720).
- [0066] 이후, 컴퓨팅 장치(100)는 miRNA 클러스터 데이터베이스(miRNA Cluster DataBase)를 이용하여 실험 데이터로 입력된 특정 miRNA와 영향력 있는 유효 거리 이내에 위치하는 인접 miRNA를 추출할 수 있다(S730). miRNA 클러스터 데이터베이스는 miRNA 들 사이의 거리 데이터를 포함하는 것으로, 컴퓨팅 장치(100)는 특정 miRNA와 10kb(kilobase) 이내에 위치하는 miRNA가 영향력 있는 유효 거리에 있는 것으로 판단할 수 있다. 다만, 유효 거리가 반드시 10kb로 설정되어야 하는 것은 아니며, 이는 선택에 따라 얼마든지 달라질 수 있다.
- [0067] 이후, 컴퓨팅 장치(100)는 특정 miRNA와 영향력 있는 유효 거리 이내에 인접 위치하는 인접 miRNA와 유전자 사이의 상관관계 값 연산할 수 있다(S740). 일례로, 도 6에 도시된 예에서,  $miRNA_i$ 이  $miRNA_i$ 의 인접 miRNA일 경우, 컴퓨팅 장치(100)는  $miRNA_i$ - $gene_m$ 에 대한 상관관계 값을 연산할 수 있다.
- [0068] **실시예 3 - 전사인자(Transcription Factor)를 고려한 상관관계 연산**
- [0069] 본 발명에 따른 컴퓨팅 장치(100)는 유전자들 사이의 전사인자를 고려하여 상관관계 값을 연산할 수 있다. 유전자들 사이의 전사인자를 고려하여 상관관계 값을 연산하는 것에 대해서는 후술되는 도면을 참조하기로 한다.
- [0070] 도 8은 전사인자 데이터베이스를 이용하여 특정 miRNA 및 전사조절 유전자 사이의 상관관계 값을 연산하는 방법을 설명하기 위한 개념도이고, 도 9는 전사인자 데이터베이스를 이용하여 특정 miRNA 및 전사조절 유전자 사이의 가중치 값을 연산하는 방법의 흐름도이다.
- [0071] 먼저, 마이크로 어레이 실험을 통해 얻어진 유전자 발현 프로파일(Genes Expression Profiles) 및 miRNA 발현 프로파일(miRNAs Expression Profiles)을 포함하는 실험 데이터가 입력되면(S910), 제어부(140)는 입력된 실험 데이터를 기초로, 특정 miRNA와 특정 유전자 사이의 상관관계(Correlation)를 연산할 수 있다(S920).
- [0072] 이후, 컴퓨팅 장치(100)는 전사인자 데이터베이스(Transcription Factor DataBase)로부터 특정 유전자의 전사조절 부위 DNA에 특이적으로 결합하여 특정 유전자의 전사를 활성화시키거나 억제하는 전사조절 유전자의 존부를 확인할 수 있다(S930).
- [0073] 특정 유전자의 전사조절 유전자가 존재하면, 컴퓨팅 장치(100)는 전사조절 유전자와 miRNA 사이의 상관관계 값을 연산할 수 있다(S940). 일례로, 도 8에 도시된 예에서,  $gene_m$ 의 전사조절 유전자가  $gene_n$ 인 경우, 컴퓨팅 장치(100)는  $miRNA_a$ - $gene_n$  사이의 상관관계 값을 기초로,  $miRNA_a$ - $gene_m$  사이의 상관관계 값을 연산할 수 있다.
- [0074] 실시예 1-3을 통해 연산된 상관관계 값을 기초로, 컴퓨팅 장치(100)는 유사 miRNA의 유전자에 대한 인터랙션 점수, 인접 miRNA의 유전자에 대한 인터랙션 점수, 전사조절 유전자의 miRNA에 대한 인터랙션 점수를 연산할 수 있다.
- [0075] 마이크로 RNA 표적 유전자 분석 알고리즘을 통해 miRNA-유전자 사이의 인터랙션 점수가 도출되면, 컴퓨팅 장치(100)는 차등 발현 유전자 분석 알고리즘을 이용한 담도암 환자의 특이 발현 유전자 목록을 이용하여 바이오마커를 검출할 수 있다.
- [0076] 상술한 바이오마커 발굴용 통합 분석 알고리즘을 기초로 담도암 환자용 바이오마커를 분석하는 방법에 대해 상세히 설명한다.
- [0077] 도 10은 바이오마커 발굴용 통합 분석 알고리즘을 기초로 담도암 환자용 바이오마커를 발굴하는 방법의 흐름도이다. 설명의 편의를 위해, 컴퓨팅 장치(100)는 이후, 컴퓨팅 장치(100)는 차등 발현 유전자 분석 알고리즘을 이용한 담도암 환자에게서 정상인과 다르게 특이발현(예컨대, 과발현 또는 저발현)되는 유전자 목록을 저장하고 있는 상태라 가정한다.
- [0078] 도 10을 참조하면, 컴퓨팅 장치(100)는 마이크로 RNA 표적 유전자 분석 알고리즘을 이용하여 miRNA-유전자 사이의 인터랙션 점수를 연산할 수 있다(S1010). 인터랙션 점수를 연산하는 단계는 앞서 도 4 내지 도 9를 통해 설명한 바와 같으므로, 이에 대한 상세한 설명은 생략한다.
- [0079] 이후, 컴퓨팅 장치(100)는 인터랙션 점수가 상위 n 번째 이내인 miRNA-유전자 페어를 선택하고(S1020), 선택된 miRNA-유전자 페어에서의 유전자와 차등 발현 유전자 분석 알고리즘을 기초로 담도암 환자에게서 정상인과 다르게 특이발현되는 유전자 목록의 교집합을 담도암 진단용 바이오마커로 결정할 수 있다(S1030). 즉, 인터랙션 점

수가 높으면서 차등 발현 유전자 분석 알고리즘에서도 담도암 환자에게서 정상인과 다르게 특이발현되는 유전자를 담도암 진단용 바이오마커로 결정할 수 있다.

- [0080] 또 다른 예로, 컴퓨팅 장치(100)는 miRNA-유전자 페어의 인터랙션 점수가 높은 순으로 m 개의 유전자를 결정하고, 차등 발현 유전자 분석 알고리즘을 기초로 담도암 환자에게서 정상인과 다르게 특이발현되는 유전자 목록의 교집합을 담도암 진단용 바이오마커로 결정할 수도 있다.
- [0081] miRNA 예측 툴로, Targetscan, miRDB, DIANA-microT, PITA, miRanda, MicroCosm 의 6개를 이용하고, miRNA-유전자 페어의 인터랙션 점수 중 상위 n 개의 유전자(q-value가 0.05 이하이면서 동시에 상관관계 값이 -0.5 이하)를 산정하였을 때, ACSM5, ADH6, ALDH1L1, APOA5, BHMT, CCL16, CYP1A2, CYP3A43, DAO, DDC, ESR1, F11, F13B, FETUB, GLYAT, GNMT, IGFALS, NAT2, PFKFB1, RDH16, SRD5A2, SULT2A1 및 THRSP이 담도암 진단용 바이오마커로 결정될 수 있다.
- [0082] 상술한 바이오마커의 개별적인 특징은 다음과 같다.
- [0083] ACSM5(acyl-CoA synthetase medium-chain family member 5)는 넓은 기질 특이성을 가진 중쇄 지방산:CoA 리가아제(라이케이즈) 활성을 갖는다. C(4) 내지 C(11)인 산 및 상응하는 3-하이드록시-불포화산 및 2,3-불포화산 또는 3,4-불포화산에 작용한다.
- [0084] ADH6(alcohol dehydrogenase 6 (class V))은 알코올 디하이드로게나아제 패밀리의 일원인 클래스 V 알코올 디하이드로게나아제(탈수화제)를 암호화한다. 상기 패밀리의 일원들은 에탄올, 레티놀, 다른 지방족 알코올, 하이드록시스테로이드, 및 지질 과산화 산물을 포함하는 다양한 기질을 대사시킨다. 이 유전자는 간뿐만 아니라, 위에서도 발현되고, 스테로이드 호르몬 수용체 결합 부위인 5' UTR의 상부에 글루코코르티코이드 반응 요소를 함유한다.
- [0085] ALDH1L1(aldehyde dehydrogenase 1 family, member L1)은 10-포르밀테트라하이드로폴레이트, 니코틴아마이드 아데닌 다이뉴클레오타이드 포스페이트(NADP+), 및 물을 테트라하이드로폴레이트, NADPH 및 이산화탄소로의 전환을 촉진한다. 이 유전자의 기능 또는 발현의 손실은 감소된 세포사멸, 증가된 세포 이동 및 종양 진행과 연관된다.
- [0086] APOA5(apolipoprotein A-V)는 HDL과 주로 관련된 마이너 아포지지방단백질이다. 킬로미크론(chylomicron)과 관련 있을 수 있다. apo-CII 지단백질 리파아제(LPL) TG 가수분해의 가능한 자극제 및 (혈중 중성지방 생성 속도에 영향 없이) 간 VLDL-TG 생성 속도의 억제제에 의해 혈중 중성지방(TG) 수준의 중요한 결정인자이다. 레시틴:콜레스테롤 아실트랜스퍼라아제(LCAT)를 저조하게 활성화시키고, 마크로파지로부터 콜레스테롤의 유출을 강화시키지 않는다.
- [0087] BHMT(betaine-homocysteine methyltransferase)는 호모시스테인 대사의 조절에 포함됨. 베타인 및 호모시스테인이 디메틸글리신 및 메티오닌으로 각각 전환됨. 이 반응은 또한 콜린의 비가역적 산화에 요구된다.
- [0088] CCL16(chemokine (C-C motif) ligand 16)은, 림프사이트(림프구) 및 모노사이트(단핵구)에 대한 화학주성적 활성을 나타낸다. 미엘로이드 간세포(myeloid progenitor cell)의 증식을 억제한다.
- [0089] CYP1A2(cytochrome P450, family 1, subfamily A, polypeptide 2)는 사이토크롬 P450은 헴(헴)-티올레이트 모노옥시게나아제의 그룹이다. 간 마이크로솜에서, 이 효소는 NADPH-의존 전자 수송 경로에 관여된다. 스테로이드, 지방산 및 생체이물(xenobiotic)을 포함하는 다양한 구조적으로 관련성이 없는 화합물을 산화시킨다. 2-하이드록실화반응을 촉진하는데 가장 활성이 있다. 카페인인 초기 N3-탈메틸화반응을 통해 간에서 사이토크롬 CYP1A2에 의해 주로 대사된다. 또한, 아플라톡신 B1 및 아세트아미노펜의 대사에 작용한다. 발암성 방향족 및 헤테로사이클 아민의 생활성에 관여한다. 헤테로사이클 아민의 N-하이드록실화반응 및 페나세틴의 O-탈에틸화반응을 촉진한다.
- [0090] CYP3A43(cytochrome P450, family 3, subfamily A, polypeptide 43)은 낮은 테스토스테론 6-베타-하이드록시라이아제 활성을 나타낸다.
- [0091] DAO(D-amino-acid oxidase)는 뇌에서 신경조절물질(neuromodulator)인 D-세린의 수준을 조절한다. D-DOPA에 대한 높은 활성을 가지고, 도파민 합성에 기여한다. 노화 동안 축적된 D-아미노산을 제거하는 독소 제거제로서 작용할 수 있다. 작은 소수성 측쇄에 이어 극성, 방향족, 및 염기성 기를 포함하는 D-아미노산에 선호도를 가지고 다양한 D-아미노산에 작용한다. 산성 아미노산에 작용하지 않는다.



- [0092] DDC(dopa decarboxylase (aromatic L-amino acid decarboxylase))는 L-3,4-디하이드록시페닐알라닌(DOPA)을 도파민으로, L-5-하이드록시트립토판을 세로토닌으로, 및 L-트립토판을 트립타민으로의 탈카복실화반응을 촉진시킨다.
- [0093] ESR1(estrogen receptor 1)은 핵 호르몬 수용체이다. 스테로이드 호르몬 및 그 수용체는 진행세포 유전자 발현의 조절에 관여되고, 세포 증식 및 표적 조직에서의 분화에 영향을 미친다. 리간드-의존성 핵 전사 촉진(ligand-dependent nuclear transactivation)은, 회귀성 에스트로겐 반응 요소(ERE, palindromic estrogen response element) 서열에 직접 호모다имер 결합, 또는 AP-1/c-Jun, c-Fos, ATF-2, Sp1 및 Sp3과 같은 다른 DNA-결합 전사인자와 결합에 관여하여, ERE-독립성 신호전달을 매개한다. 리간드 결합은 개별적인 구성요소의 LXXLL 모티프를 통하여 멀티단백질 공동활성화제(coactivator) 복합체와 서열상 또는 결합상의 결합을 허용하는 형태상의 변화를 유도한다. 상호 억제(mutual transrepression)가 세포 유형 특이적 방식으로 에스트로겐 수용체(ER) 및 NF-카파-B(NF-kappa-B) 사이에서 일어난다. NF-카파-B DNA-결합 활성을 감소시키고, IL6 프로모터로부터 NF-카파-B-매개된 전사를 억제시키고, 프로모터로부터 관련된 공동조절자(coregulator) 및 RELA/p65를 이동시킨다. CCL2 및 IL8 프로모터의 NF-카파-B 반응 요소로 모이고, CREBBP를 이동시킬 수 있다. ERE 서열상에 NF-카파-B 성분 RELA/p65 및 NFKB1/p50와 존재한다. 또한 NF-카파-B와 시너지를 일으켜 인접한 반응 요소들의 개별적인 모집을 포함하는 전사를 활성화시킬 수 있고, 기능은 CREBBP를 포함한다. TFF1의 전사 활성을 촉진할 수 있다. 또한, 다양한 키나아제 신호전달 단계연쇄반응(cascade)을 포함하는 막-개시(membrane-initiated) 에스트로겐 신호전달을 매개할 수 있다. 아이소폼(isoform) 3은 NOS3 및 내피 일산화질소 생성의 활성화에 관여된다. 하나 또는 여러 개의 기능성 도메인이 결합된 아이소폼은 전장 수용체와 경쟁적 리간드 또는 DNA 결합 및/또는 이질이량체화(heterodimerization)에 의해 전사 활성을 조정하는 것으로 여겨진다. 아이소폼 3은 ERE에 결합하고 아이소폼 1을 억제할 수 있다.
- [0094] F11(coagulation factor XI)은 인자 XI는 인자 IX를 활성화시킴으로써 혈액 응고의 내재적 경로의 중간 상(middle phase)을 촉진시킨다.
- [0095] F13B(coagulation factor XIII, B polypeptide)는 인자 XIII의 B 체인(chain)은 촉매적으로 활성은 아니지만, B 서브유닛을 안정화시키고, 트롬빈에 의해 트랜스글루타미나아제 형성 속도를 조절하는 것으로 여겨진다.
- [0096] FETUB(fetuin B)에 의해 인코딩되는 단백질은 시스테인 프로테아제 억제자(cysteine protease inhibitor)의 시스타틴(cystatin) 슈퍼패밀리의 일부인 페투인(fetuin) 패밀리의 일원이다. 페투인은 골 형성(osteogenesis) 및 골 재흡수(bone resorption), 인슐린 및 간 성장 인자 수용체의 조절, 및 전신 염증에 대한 반응을 포함하는 여러 개의 다양한 기능에 관련되어 있다.
- [0097] GLYAT(glycine-N-acyltransferase)는 아실기를 글리신의 N-말단으로 옮기는 미토콘드리아 아실트랜스퍼라아제(mitochondrial acyltransferase)이다. 다수의 기질을 접합시켜 다양한 N-아실글리신을 형성함으로써 벤조산 또는 살리신산과 같은 생체 이물 및 이소발레르산과 같은 내생성 유기산을 탈독소화시킬 수 있다.
- [0098] GNMT(glycine N-methyltransferase)는 S-아데노실메티오닌(AdoMet)을 사용함으로써 글리신의 메틸화를 촉진하여 S-아데노실호모시스테인(AdoHcy)의 부수적인 생성과 함께 N-메틸글리신(사르코신)을 형성한다. AdoMet의 조직 농도의 조절 및 메티오닌의 대사의 조절에서 중요한 역할을 할 수 있다.
- [0099] IGFALS(insulin-like growth factor binding protein, acid labile subunit)는 단백질 복합체, 수용체-리간드 결합 또는 세포 부착을 야기하는 단백질-단백질 상호작용에 관여한다.
- [0100] NAT2(N-acetyltransferase 2 (arylamine N-acetyltransferase))는 하이드라진 및 아릴아민 약물의 과잉시 탈독소화에 참여한다. 다양한 아릴아민 및 헤테로사이클릭 아민 기질의 N-아세틸화 또는 O-아세틸화를 촉진하고, 여러 개의 공지된 발암물질을 생활성화시킬 수 있다.
- [0101] PFKFB1(6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1)은 프럭토오스 2,6-비스포스페이트를 합성 및 분해한다.
- [0102] RDH16(retinol dehydrogenase 16 (all-trans))은 NAD에 대한 선호도를 가지는 산화환원효소(oxidoreductase)이다. 모두 트랜스인 레티놀 및 13-시스-레티놀을 산화시켜 상응하는 알데하이드로 산화시킨다. 자유 레티놀보다 CRBP-결합 레티놀에 대해 더 높은 활성을 갖는다. 3-알파-하이드록시스테로이드를 산화시킨다. 안드로스텐디올 및 안드로스테론을 디하이드로테스토스테론 및 안드로스텐디온으로 산화시킨다. 또한 역반응을 촉진시킬 수 있다.

- [0103] SRD5A2(steroid-5-alpha-reductase, alpha polypeptide 2 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 2))는 테스토스테론(T)을 5-알파-디하이드로테스토스테론(DHT)으로, 그리고 프로게스테론 또는 코르티코스테론을 상응하는 5-알파-3-옥소스테로이드로 전환시킨다. 성 분화 및 안드로겐 생리에서 중요한 역할을 한다.
- [0104] SULT2A1(sulfotransferase family, cytosolic, 2A, dehydroepiandrosterone (DHEA)-preferring, member 1)은, 3'-포스포-5'-아데닐일 설페이트(PAPS)를 설페이트 도너(donor)로 사용하여 간 및 부신에서 스테로이드 및 담즙의 설페화를 촉진시키는 설페트랜스퍼라아제이다.
- [0105] THRSP(thyroid hormone responsive (SPOT14 homolog, rat))는, 특히 젓 분비 유선에서, 지질 합성의 조절에서 역할을 한다. 중간 길이 지방산 사슬을 가진 트라이글리세라이드의 생합성에 중요하다. MID1IP1와 상호작용하고 ACACA와 그 상호작용을 예방함으로써 지질 합성을 조절할 수 있다. 전사 공활성자(coactivator)로서 기능할 수 있다. THRSP의 전사인자 활성을 조절할 수 있다.
- [0106] 상기 결과에 의해 획득된 담도암 진단용 바이오마커의 실험 과정 및 결과에 대해 상세히 설명한다.
- [0107] 담도암 환자 샘플 및 마이크로어레이 실험
- [0108] 간암 연구소 및 중산 병원(Liver Cancer Institute and Zhongshan Hospital, 중국 상하이 푸단 대학교)에서 2002년 내지 2003년, 및 카나자와 대학교 병원(Kanazawa University Hospital, 일본, 이시카와)에서 2008년 내지 2010년에 치유 절제를 거친 아시아계 환자로부터 간내 담도암종(intrahepatic cholangiocellular carcinoma, ICC) 및 복합 간세포 담도암종(combined hepatocellular cholangiocarcinoma, CHC) 조직을 사전 동의를 얻어 수득하였다.
- [0109] 샘플은 상응하는 연구소의 임상시험심사위원회(Institutional Review Board)에 의해 승인되었고, 인간 대상 연구(Human Subjects Research)의 미국 국립보건원(National Institutes of Health, NIH) 사무국에 의해 기록되었다. 총 23개의 ICC 및 CHC 사례가 mRNA 및 마이크로 RNA 특징(microRNA signatures)을 만들기 위해 사용되었다. 혈청 검사 및 이미징에 근거하여 조기 진단하였고, 병리학자들이 병리조직학적으로 이를 확인하였다. 독립적인 집단으로부터 68명의 백인 ICC 환자들의 특성화가 최근에 기술되었다(Hepatology, Vol. 56, No. 5, 1792-803).
- [0110] 본 발명의 바이오마커 세트의 검증
- [0111] 본 발명의 바이오마커 세트의 담도암 판정의 검증은 담도암 환자 104명 및 정상인 59명으로 총 163명을 대상으로 하였다. 상기 대상으로부터 채취한 혈액을 이용하여, GEO(Gene Expression Omnibus) 데이터 GSE26566을 사용하여 주성분 분석(principal component analysis) 및 계층적 군집분석 방법(hierarchical clustering, euclidean distance, complete method)으로 판정하였다.
- [0112] 그 결과, 담도암에 대한 민감도(sensitivity)가 82%(85/104) 및 특이도(specificity)가 97%(57/59)로 나타났다. 도 11은 주성분 분석 결과를 플롯(Plot) 한 도면이다. 도 11에서 가로축의 component 1은 제 1 주성분(PC 1)을 의미하고, 세로축의 component 2는 제 2 주성분(PC 2)을 의미할 수 있다. 나아가, 삼각형으로 표현된 개체는 암환자를 의미하고, 원으로 표현된 개체는 정상인을 의미하는 것일 수 있다. 도 12는 계층적 군집분석 결과를 히트 맵(Heat Map)으로 도시한 도면이다. 도 12에서 히트 맵 상단에 위치한 붉은색 바(Bar)는 암환자를 의미하고, 파란색 바는 정상인을 의미할 수 있다.
- [0113] 본 발명의 일실시예에 의하면, 전술한 방법들은, 프로그램이 기록된 매체에 프로세서가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 프로세서가 읽을 수 있는 매체의 예로는, ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장장치 등이 있으며, 캐리어 웨이브(예를 들어, 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다.
- [0114] 상기와 같이 설명된 컴퓨팅 장치(100)는 상기 설명된 실시예들의 구성과 방법이 한정되게 적용될 수 있는 것이 아니라, 상기 실시예들은 다양한 변형이 이루어질 수 있도록 각 실시예들의 전부 또는 일부가 선택적으로 조합되어 구성될 수도 있다.

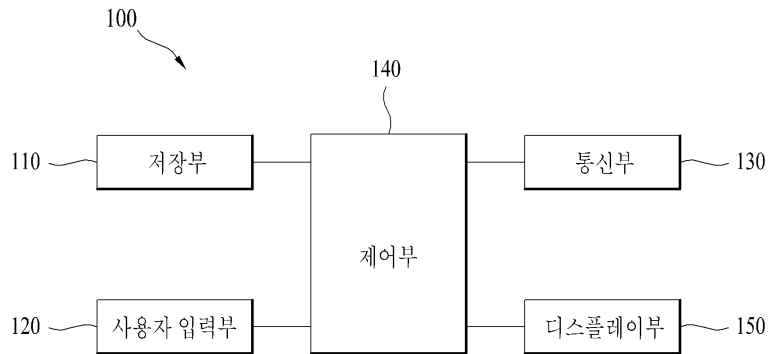
## 부호의 설명

- [0115] 100 : 컴퓨팅 장치  
110 : 저장부

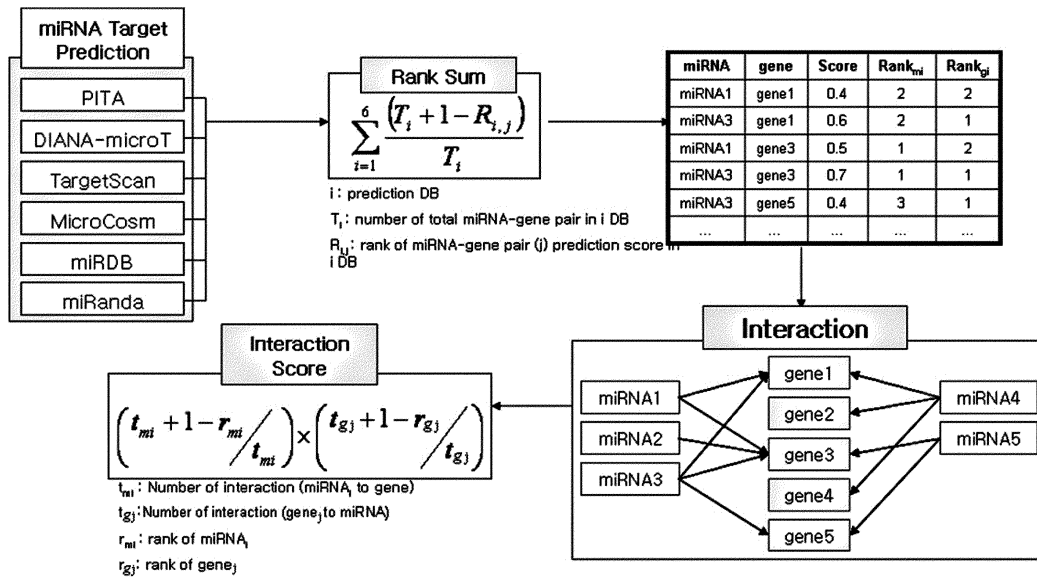
- 120 : 사용자 입력부  
130 : 통신부  
140 : 제어부  
150 : 디스플레이부

## 도면

### 도면1

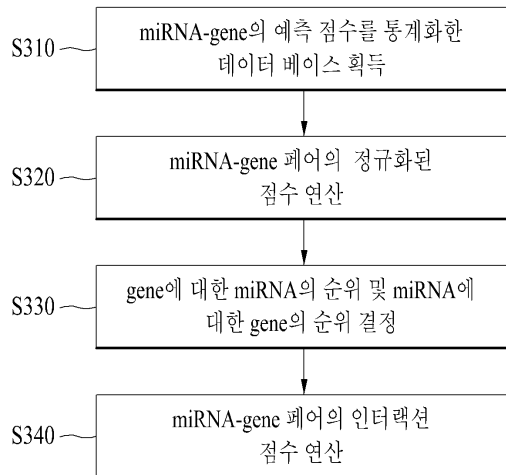


### 도면2

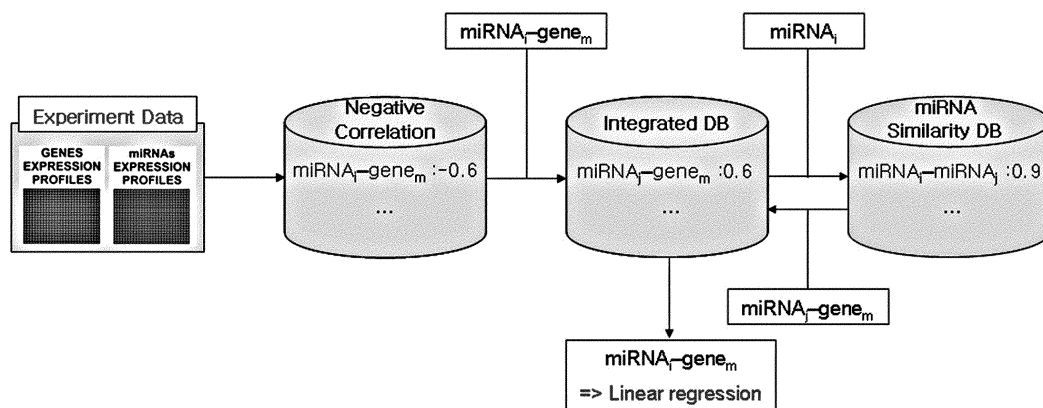




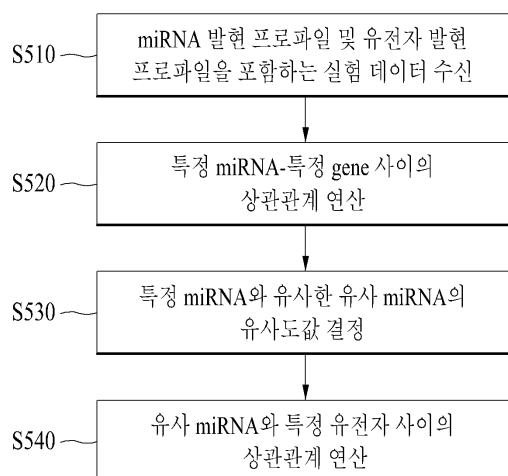
도면3



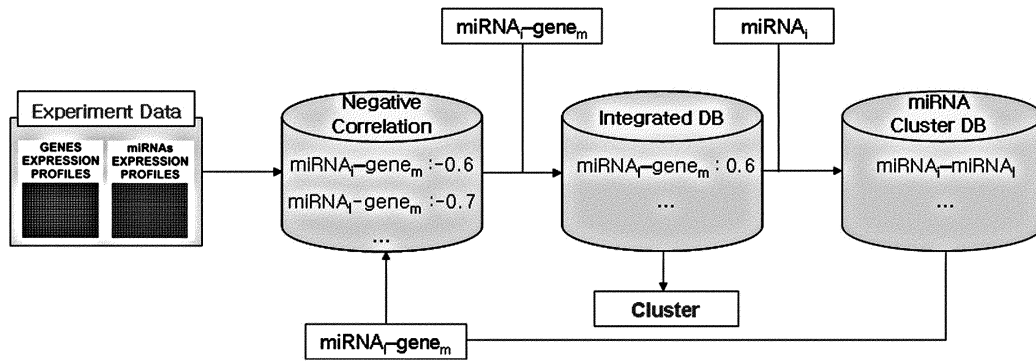
도면4



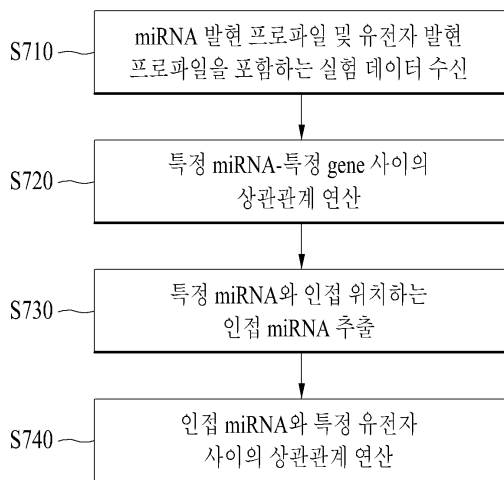
도면5



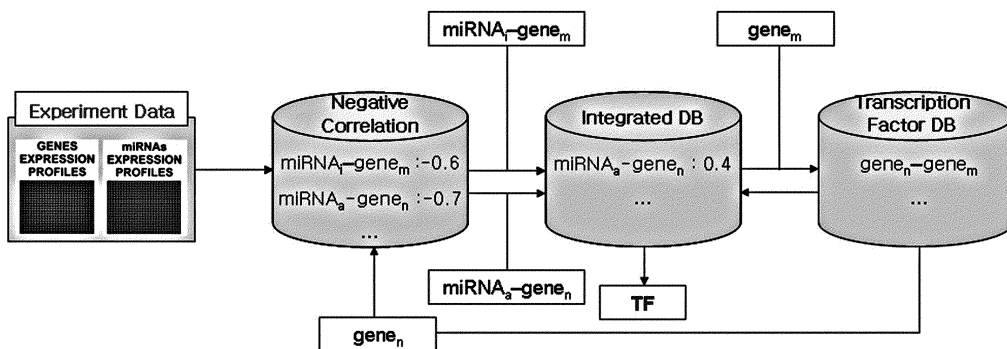
도면6



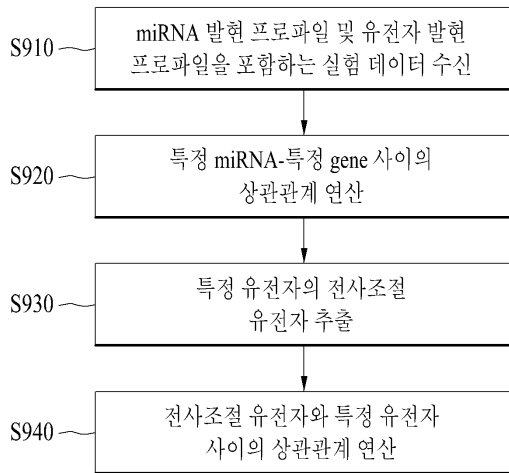
도면7



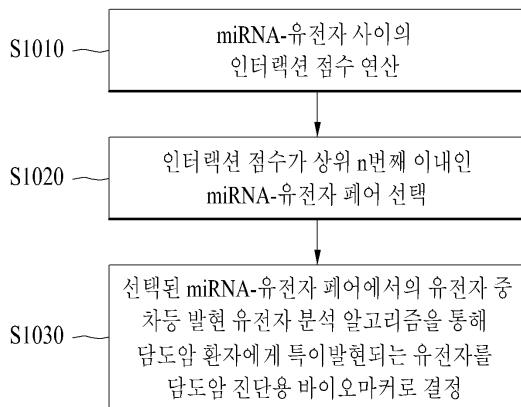
도면8



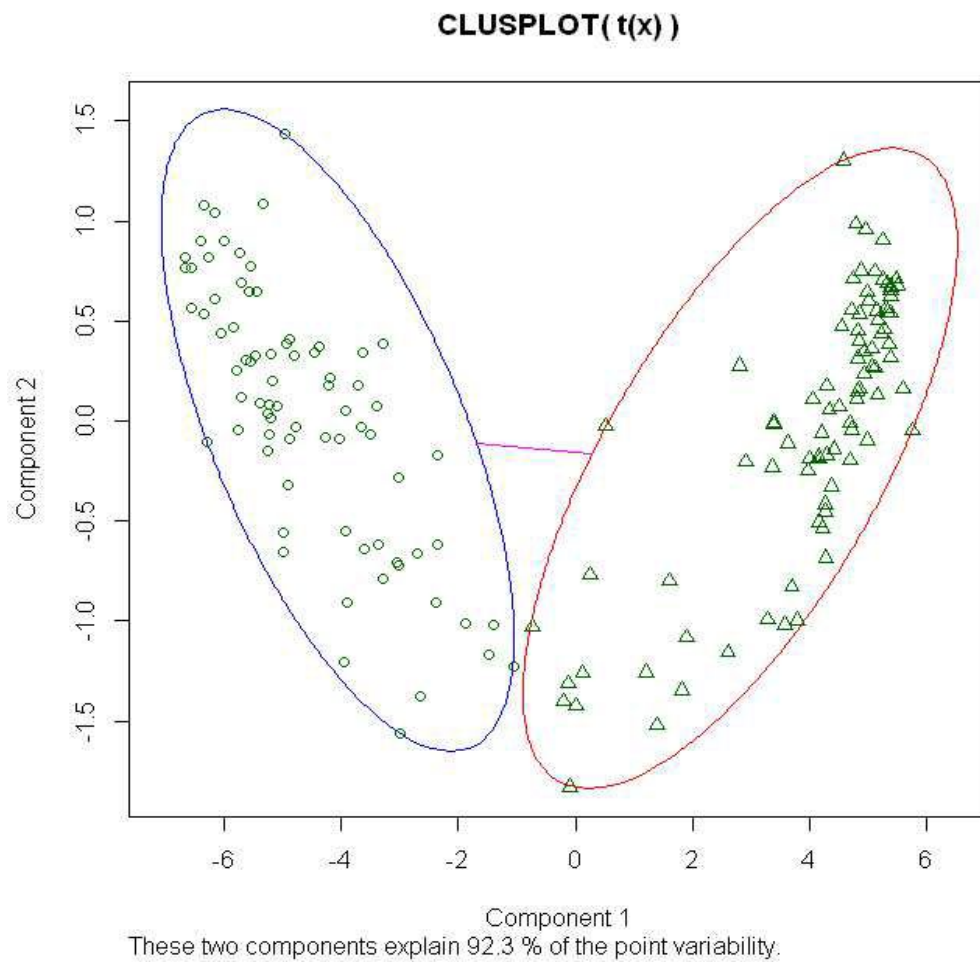
도면9



도면10



도면11



도면12

