



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0043790  
(43) 공개일자 2015년04월23일

(51) 국제특허분류(Int. Cl.)

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

*G01N 33/48* (2006.01) *G01N 33/574* (2006.01)

(21) 출원번호 10-2013-0122635

(22) 출원일자 2013년10월15일

심사청구일자 없음

(71) 출원인

엘지전자 주식회사

서울특별시 영등포구 여의대로 128 (여의도동)

(72) 발명자

최형석

서울 서초구 바우피로 38, 전자기술원 (우면동, LG종합기술원)

허지연

서울 서초구 바우피로 38, 전자기술원 (우면동, LG종합기술원)

(74) 대리인

방해철, 김용인

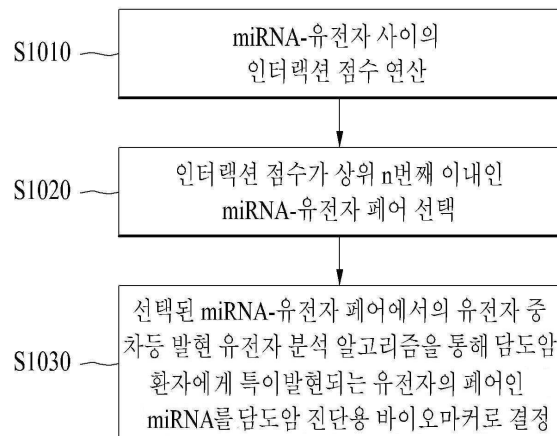
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 담도암 진단용 바이오마커의 추출 방법, 이를 위한 컴퓨팅 장치, 담도암 진단용 바이오마커 및 이를 포함하는 담도암 진단 장치

### (57) 요약

본 발명은 담도암 진단용 바이오마커의 추출 방법 및 이를 위한 컴퓨팅 장치, 그리고 담도암 진단용 바이오마커 및 이를 포함하는 담도암 진단 장치를 개시한다. 더욱 구체적으로 본 발명은 혈액 또는 조직에서 수득한 마이크로 RNA를 이용한 담도암 진단용 바이오마커의 추출 방법 및 이를 위한 컴퓨팅 장치, 그리고 담도암 진단용 바이오마커 및 이를 포함하는 담도암 진단 장치를 개시한다.

대표도 - 도10



이 발명을 지원한 국가연구개발사업

과제고유번호 10040174

부처명 산업통상자원부

연구관리전문기관 한국산업기술평가관리원

연구사업명 바이오의료기기 산업원천기술개발사업(바이오)

연구과제명 차세대 생명정보 통합분석을 이용한 다중 바이오마커 개발

기 여 율 1/1

주관기관 (주)지노믹트리

연구기간 2011.06.01 ~ 2016.05.31

---

## 명세서

### 청구범위

#### 청구항 1

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 유전자의 페어인 마이크로 RNA를 추출하는 단계

를 포함하는 담도암 진단용 바이오마커 추출 방법.

#### 청구항 2

제 1 항에 있어서,

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 점수로부터 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 단계

를 포함하는 담도암 진단용 바이오마커 추출 방법.

#### 청구항 3

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 4

제 3 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 5

제 2 항에 있어서,

상기 정규화된 점수는, 상기 하나 이상의 데이터 베이스에서의 마이크로 RNA 및 유전자 페어의 예측 점수 순위를 기초로 연산되는 것을 특징으로 하는 담도암 진단용 바이오마커 추출 방법.

#### 청구항 6

제 5 항에 있어서,

상기 정규화된 점수는 하기 수학적 식 1에 의거 연산되는 것을 특징으로 하는 담도암 진단용 바이오마커 추출 방법.

[수학식 1]

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

(상기 수학식 1에서,  $i$ 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)

#### 청구항 7

제 5 항에 있어서,

상기 인터랙션 점수는, 상기 정규화된 점수를 기초로 유전자별 마이크로 RNA의 순위 및 마이크로 RNA별 유전자의 순위를 기초로 연산되는 것을 특징으로 하는 담도암 진단용 바이오마커 추출 방법.

#### 청구항 8

제 7 항에 있어서,

상기 인터랙션 점수는 하기 수학식 2에 의거 연산되는 것을 특징으로 하는 담도암 진단용 바이오마커 추출 방법.

[수학식 2]

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

(상기 수학식 2에서,

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

#### 청구항 9

데이터를 저장하기 위한 저장부; 및

연산을 위한 제어부를 포함하고,

상기 제어부는, 마이크로 RNA와 유전자 사이 상보 결합 정도를 수치화한 인터랙션 점수를 연산하고, 상기 인터랙션 점수가 높은  $n$  개의 마이크로 RNA 및 유전자 페어를 결정하며, 상기  $n$  개의 마이크로 RNA 및 유전자 페어 중 담도암 환자에게서 특이발현되는 유전자와 공통되는 유전자의 페어인 마이크로 RNA를 추출하도록 구성되는 것을 특징으로 하는 컴퓨팅 장치.

#### 청구항 10

생물학적 시료로서 조직을 사용하는, hsa-miR-21-5p, hsa-miR-93-5p, hsa-miR-106b-5p, hsa-miR-155-5p, hsa-miR-181a-5p, hsa-miR-200a-3p, hsa-miR-200b-3p, hsa-miR-222-3p, 및 hsa-miR-331-3p를 포함하는 담도암 진단을 위한 바이오마커(biomarker).

#### 청구항 11

생물학적 시료로서 혈액을 사용하는, hsa-miR-21-5p, hsa-miR-93-5p, hsa-miR-106b-5p, hsa-miR-155-5p, hsa-miR-181a-5p, 및 hsa-miR-222-3p를 포함하는 담도암 진단을 위한 바이오마커(biomarker).

## 청구항 12

제 10 항 또는 제 11 항의 바이오마커를 포함하는 담도암 진단을 위한 장치.

## 청구항 13

제 12 항에 있어서,

상기 장치는 진단 칩, 진단 키트, 정량 PCR(qPCR) 장비, 현장검사(POCT) 장비 및 시퀀서 중 어느 하나인 것을 특징으로 하는 담도암 진단을 위한 장치.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 담도암 진단용 바이오마커의 추출 방법 및 이를 위한 컴퓨팅 장치, 그리고 담도암 진단용 바이오마커 및 이를 포함하는 담도암 진단 장치에 관한 것이다. 더욱 구체적으로 본 발명은 혈액 또는 조직에서 수득한 마이크로 RNA를 이용한 담도암 진단을 위한 바이오마커의 추출 방법 및 이를 위한 컴퓨팅 장치, 그리고 담도암 진단용 바이오마커 및 이를 포함하는 담도암 진단 장치에 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 간내의 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.

[0006] 한편, 마이크로 RNA(micro RNA, miRNA)란 약 17 내지 25개의 뉴클레오타이드(nucleotide)로 구성된 짧은 단일 가닥 비-암호화(non-coding) RNA 분자를 의미한다. 마이크로 RNA는 표적 mRNA(유전자)의 전사 과정을 방해하거나, mRNA를 분해하게 함으로써, 단백질 생산 유전자 발현을 조절하는 것으로 알려져 있다. 마이크로 RNA는 조직 내에서뿐만 아니라 혈액 속에도 존재하는 것으로 알려져 있다.

[0007] 또한, 취급 및 진단 상의 용이성을 위해 조직 또는 혈액 샘플을 이용한 바이오마커의 개발이 필요하다. 특히 혈액 샘플이 유리하다.

## 발명의 내용

### 해결하려는 과제

[0008] 상기 문제점을 해결하기 위하여, 본 발명에서는 혈액 또는 조직에서 수득한 마이크로 RNA를 이용한 담도암 진단을 위한 바이오마커의 추출 방법 및 이를 위한 컴퓨팅 장치를 제공하고자 한다. 또한, 본 발명에서는 담도암 진단용 바이오마커 및 이를 포함하는 담도암 진단용 장치를 제공하고자 한다.

[0009] 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 명확하

게 이해될 수 있을 것이다.

### 과제의 해결 수단

- [0010] 본 발명의 일예에 따른 담도암 진단용 바이오마커 추출 방법은, 마이크로 RNA와 유전자 사이 상호 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 유전자의 페어인 마이크로 RNA를 추출하는 단계를 포함한다.
- [0011] 본 발명의 일예에 따른 담도암 진단용 바이오마커는 생물학적 시료로서 조직을 사용하고, hsa-miR-21-5p, hsa-miR-93-5p, hsa-miR-106b-5p, hsa-miR-155-5p, hsa-miR-181a-5p, hsa-miR-200a-3p, hsa-miR-200b-3p, hsa-miR-222-3p, 및 hsa-miR-331-3p를 포함한다.
- [0012] 본 발명의 일예에 따른 담도암 진단용 바이오마커는 생물학적 시료로서 혈액을 사용하고, hsa-miR-21-5p, hsa-miR-93-5p, hsa-miR-106b-5p, hsa-miR-155-5p, hsa-miR-181a-5p, 및 hsa-miR-222-3p를 포함한다.
- [0013] 본 발명의 일예에 따른 담도암 진단을 위한 장치는 상술한 바이오마커를 포함한다.
- [0014] 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

- [0015] 본 발명에서는 높은 특이성 및 민감성을 갖춘 담도암의 진단용 바이오마커를 제공할 수 있다. 또한, 본 발명에서는 담도암 진단용 바이오마커를 발굴하는 방법을 제공할 수 있다.
- [0016] 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

- [0017] 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 GSE32957을 이용한 계층적 군집분석 결과를 히트 맵(Heat Map)으로 도시한 것이다.

도 12는 차세대 유전체 시퀀싱의 구체적인 활용에 중의 하나인 스몰 RNA 시퀀싱 데이터 분석에 관한 개념도이다.

도 13은 본 발명에 따라, 조직 샘플에서 발현되는 9개의 마이크로 RNA의 발현 양상을 나타내는 차세대 유전체 시퀀싱 데이터를 이용한 계층적 군집 분석 결과를 도시한 도면이다.

도 14는 본 발명에 따라, 혈액 샘플에서 발현되는 6개의 마이크로 RNA의 발현 양상을 나타내는 차세대 유전체 시퀀싱 데이터를 이용한 계층적 군집 분석 결과를 도시한 도면이다.

### 발명을 실시하기 위한 구체적인 내용

- [0018] 이하, 본 발명과 관련된 컴퓨팅 장치에 대하여 도면을 참조하여 보다 상세하게 설명한다.
- [0019] 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.
- [0020] 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.
- [0021] 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.
- [0022] 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.
- [0023] 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, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다.
- [0024] 사용자 입력부(120)는 사용자로부터 사용자 입력을 수신하는 역할을 수행한다. 사용자 입력부(120)는 키보드, 마우스 등을 포함할 수 있다.
- [0025] 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.
- [0026] 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.
- [0027] 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 장치로의 역할을 수행할 수 있다. 디스플레이부(150)는 컴퓨팅 장치(100)를 보조하는 모니터 등의 장치일 수 있다.
- [0028] 상기와 같이 설명된 컴퓨팅 장치(100)는 하술될 실시예들의 구성과 방법이 한정되어 적용될 수 있는 것이 아니라, 하기 실시예들의 다양한 변형이 이루어질 수 있도록 각 실시예들의 전부 또는 일부가 선택적으로 조합되어 구성될 수도 있다.
- [0029] 상술한 컴퓨팅 장치(100)를 이용하여, 담도암 진단용 바이오마커 발굴 방법에 대해 상세히 설명하기로 한다.
- [0030] 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.
- [0031] 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).

[0032] 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 마이크로어레이 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).

[0033] 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 Rate) 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.

[0034] 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.

[0035] 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, 마이크로어레이 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.

[0036] 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.

#### [0037] 마이크로 RNA 표적 유전자 예측값 연산

[0038] 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.

[0039] 도 2는 miRNA와 유전자 사이의 인터랙션 점수를 연산하는 것의 일예를 설명하기 위한 개념도이고, 도 3은 인터랙션 점수의 연산 방법의 흐름도이다.

[0040] 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).

[0041] 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**

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>

[0043] 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.

[0044] 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.

[0045] 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.

[0046] 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).

[0047] 표 1에 도시된 예에서와 같이, miRNA 타겟 예측 툴이 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

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

[0048]

[0049] 상기 수학식 1에서, i는 i번째 데이터 베이스, n은 데이터 베이스의 개수(일예로, 도 2에서는 6개의 예측 툴을 통해 6개의 데이터 베이스를 획득하였으므로, n은 6으로 설정될 수 있을 것임),  $T_i$ 는 i번째 데이터 베이스에서 miRNA-유전자 페어의 총 개수,  $R_{i,j}$ 는 i번째 데이터베이스에서, j번째 miRNA-유전자 페어의 순위를 의미하는 것 일 수 있다.

[0050] 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.

[0051] 이후, 제어부(140)는 정규화 점수를 기초로, 특정 유전자에 대한 miRNA 들의 순위 및 특정 miRNA에 대한 유전자

들의 순위를 매길 수 있다(S330).

[0052] 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 두 번째 순위를 갖고, miRNA3은 첫 번째 순위를 갖는 것으로 예시되었다.

[0053] 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 두 번째 순위를 갖고, gene3은 첫 번째 순위를 갖는 것으로 예시되었다.

[0054] 이후, 제어부(140)는 유전자 및 miRNA 들의 순위를 기초로, 유전자-miRNA 사이의 인터랙션 점수(Interaction Score)를 연산할 수 있다(S340). 수학적 2는 인터랙션 점수를 연산하기 위해 이용되는 수학적식을 예시한 것이다.

## 수학적식 2

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

[0055]

[0056] 상기 수학적식 2에서,  $t_{mi}$ 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.

## [0057] 상관관계 연산

[0058] 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.

## [0059] 실시예 1 - 상관관계에 기인한 가중치 연산

[0060] 본 발명에 의한 컴퓨팅 장치(100)는 마이크로어레이 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.

[0061] 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.

[0062] 먼저, 마이크로어레이 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).

[0063] 우선적으로 마이크로어레이 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. 이어

서, 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.

[0064] 위와 같은 마이크로어레이 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.

[0065] 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.

[0066] 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.

#### [0067] 실시예 2 - miRNA 주변 영향력을 고려한 상관관계 연산

[0068] 본 발명에 따른 컴퓨팅 장치(100)는 특정 miRNA와 클러스터를 형성하는 인접 miRNA에 대한 상관관계 값을 연산할 수도 있다. miRNA들 사이의 영향력을 고려하여 상관관계 값을 연산하는 것에 대해서는 후술되는 도면을 참조하기로 한다.

[0069] 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.

[0070] 먼저, 마이크로어레이 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).

[0071] 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 miRNA과 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.

[0072] 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<sub>i</sub>이 miRNA<sub>j</sub>의 인접 miRNA일 경우, 컴퓨팅 장치(100)는 miRNA<sub>i</sub>-gene<sub>m</sub>에 대한 상관관계 값을 연산할 수 있다.

#### [0073] 실시예 3 - 전사인자(Transcription Factor)를 고려한 상관관계 연산

[0074] 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.

[0075] 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.

[0076] 먼저, 마이크로어레이 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).

[0077] 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).

- [0078] 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.
- [0079] 실시예 1 내지 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.
- [0080] 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.
- [0081] 상술한 바이오마커 발굴용 통합 분석 알고리즘을 기초로 담도암 환자용 바이오마커를 분석하는 방법에 대해 상세히 설명한다.
- [0082] 도 10은 바이오마커 발굴용 통합 분석 알고리즘을 기초로 담도암 환자용 바이오마커를 발굴하는 방법의 흐름도이다. 설명의 편의를 위해, 컴퓨팅 장치(100)는 이후 차등 발현 유전자 분석 알고리즘을 이용한 담도암 환자에서 정상인과 다르게 특이발현(예컨대, 과발현 또는 저발현)되는 유전자 목록을 저장하고 있는 상태라 가정한다.
- [0083] 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.
- [0084] 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 환자에서 정상인과 다르게 특이발현되는 유전자 목록의 교집합의 페어인 miRNA의 세트를 담도암 진단용 바이오마커로 결정할 수 있다(S1030). 즉, 인터랙션 점수가 높으면서 차등 발현 유전자 분석 알고리즘에서도 담도암 환자에서 정상인과 다르게 특이발현되는 유전자의 페어인 miRNA의 세트를 담도암 진단용 바이오마커로 결정할 수 있다.
- [0085] 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 환자에서 정상인과 다르게 특이발현되는 유전자 목록의 교집합의 페어인 miRNA를 담도암 진단용 바이오마커로 결정할 수도 있다.
- [0086] miRNA 예측 툴로, Targetscan, miRDB, DIANA-microT, PITA, miRanda, MicroCosm 의 6개를 이용하고, 생물학적 샘플로 조직을 사용 시, miRNA-유전자 페어의 인터랙션 점수 중 상위  $n$  개의 유전자( $q$ -value가 0.05 이하이면서 동시에 상관관계 값이  $-0.5$  이하)의 페어인 miRNA 세트로, hsa-miR-21-5p, hsa-miR-93-5p, hsa-miR-106b-5p, hsa-miR-155-5p, hsa-miR-181a-5p, hsa-miR-200a-3p, hsa-miR-200b-3p, hsa-miR-222-3p, 및 hsa-miR-331-3p가 담도암 진단용 바이오마커로 결정될 수 있다.
- [0087] 또한, 생물학적 샘플로 혈액을 사용 시, hsa-miR-21-5p, hsa-miR-93-5p, hsa-miR-106b-5p, hsa-miR-155-5p, hsa-miR-181a-5p, 및 hsa-miR-222-3p가 담도암 진단용 바이오마커로 결정된다.
- [0088] 상술한 바이오마커에 속하는 각 miRNA의 염기서열은 하기 표 2와 같다.

표 2

Mature_id	miRNA_id	Sequence
hsa-miR-21-5p	hsa-mir-21	UAGCUUAUCAGACUGAUGUUGA
hsa-miR-93-5p	hsa-mir-93	CAAAGUGCUGUUCGUGCAGGUAG
hsa-miR-106b-5p	hsa-mir-106b	UAAAGUGCUGACAGUGCAGAU
hsa-miR-155-5p	hsa-mir-155	UUAUAGCUAAUCGUGAUAGGGGU
hsa-miR-181a-5p	hsa-mir-181a-1, hsa-mir-181a-2	AACAUUCAACGCUGUCGGUGAGU
hsa-miR-200a-3p	hsa-mir-200a	UACACUGUCUGGUAACGAUGU
hsa-miR-200b-3p	hsa-mir-200b	UAAUACUGCCUGGUAUGAUGA
hsa-miR-222-3p	hsa-mir-222	AGCUACAUCUGGCUACUGGGU
hsa-miR-331-3p	hsa-mir-331	GCCCCUGGGCCUAUCCUAGAA

- [0090] 상기 결과에 의해 획득된 담도암 진단용 바이오마커의 실험 과정 및 결과에 대해 상세히 설명한다.
- [0091] 담도암 환자 샘플 및 마이크로어레이 실험
- [0092] 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.
- [0093] 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).
- [0094] 본 발명의 바이오마커 세트의 검증
- [0095] 본 발명의 조직 샘플용 바이오마커 세트의 담도암 판정의 검증은 담도암 환자 25명 및 정상인 10명으로 총 35명을 대상으로 하였다. 상기 대상으로부터 채취한 혈액을 이용하여, GEO(Gene Expression Omnibus) 데이터 GSE32957을 사용하여 계층적 군집분석 방법(hierarchical clustering, euclidean distance, complete method)으로 판정하였다. 그 결과, 담도암에 대한 민감도(sensitivity)가 96%(24/25) 및 특이도(specificity)가 100%(10/10)로 매우 우수하게 나타났다. 도 11은 데이터 GSE32957을 사용하였을 때의 계층적 군집 분석 결과를 히트 맵(Heat Map)으로 도시한 도면이다. 도 11에서 히트 맵 상단에 위치한 붉은색 바(Bar)는 암환자를 의미하고, 파란색 바는 정상인을 의미한다.
- [0096] 또한, 본 발명의 조직 샘플용 및 혈액 샘플용 바이오마커의 담도암 판정의 검증을 별도로 실시하였다. 먼저 조직 샘플용 바이오마커는 담도암 환자 2명 및 정상인 2명으로 총 4명을 대상으로 하고, 혈액 샘플용 바이오마커는 담도암 환자 8명 및 정상인 2명을 대상으로 하였다. 상기 대상으로부터 채취한 조직 및 혈액을 각각 이용하여, 차세대 유전체 시퀀싱(Next Generation Sequencing, NGS) 방법인 스몰 RNA 시퀀싱 데이터(Small RNA sequencing data)를 사용하여 계층적 군집 분석 방법(hierarchical clustering, euclidean distance, complete method)으로 판정하였다. 이러한 스몰 RNA 시퀀싱 데이터 분석의 일반적인 설명은 도 12에 개시되어 있다. 그 결과, 조직 샘플에 대한 바이오마커의 담도암에 대한 민감도(sensitivity)가 100%(2/2) 및 특이도(specificity)가 100%(2/2)로 나타났고, 이 경우 Small RNA sequencing 데이터를 이용하였을 때의 계층적 군집 분석 결과를 도 13에 도시하였다. 또한 혈액 샘플에 대한 바이오마커의 담도암에 대한 민감도(sensitivity)가 75%(6/8) 및 특이도(specificity)가 50%(1/2)로 나타났고, 이 경우 Small RNA sequencing 데이터를 이용하였을 때의 계층적 군집 분석 결과를 도 14에 도시하였다. 도 13 및 14에서 히트 맵 상단에 위치한 붉은색 바(Bar)는 암환자를 의미하고, 파란색 바는 정상인을 의미한다.
- [0097] 한편, 상술한 바이오마커는 이를 포함한 담도암 진단용 장치로서 사용된다. 상기 담도암 진단용 장치로는 진단 칩, 진단 키트, 정량 PCR(qPCR) 장비, 현장검사(POCT) 장비, 시퀀서 등이 있다. 상기 진단 칩, 진단 키트, 정량 PCR(qPCR) 장비, 현장검사(POCT) 장비, 시퀀서에서, 바이오마커 세트를 제외한 부분은 공지된 것을 활용할 수 있다.
- [0098] 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.

## 부호의 설명

- [0099] 100 : 컴퓨팅 장치  
110 : 저장부



120 : 사용자 입력부

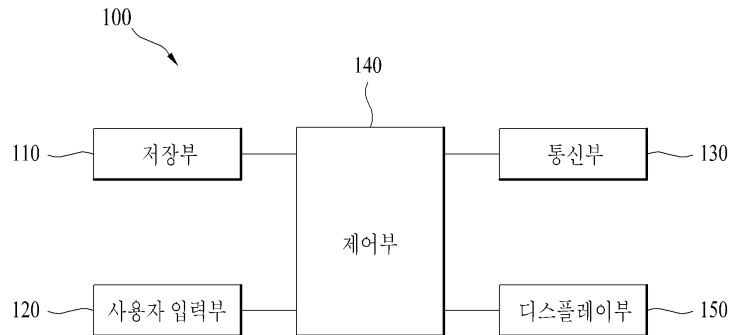
130 : 통신부

140 : 제어부

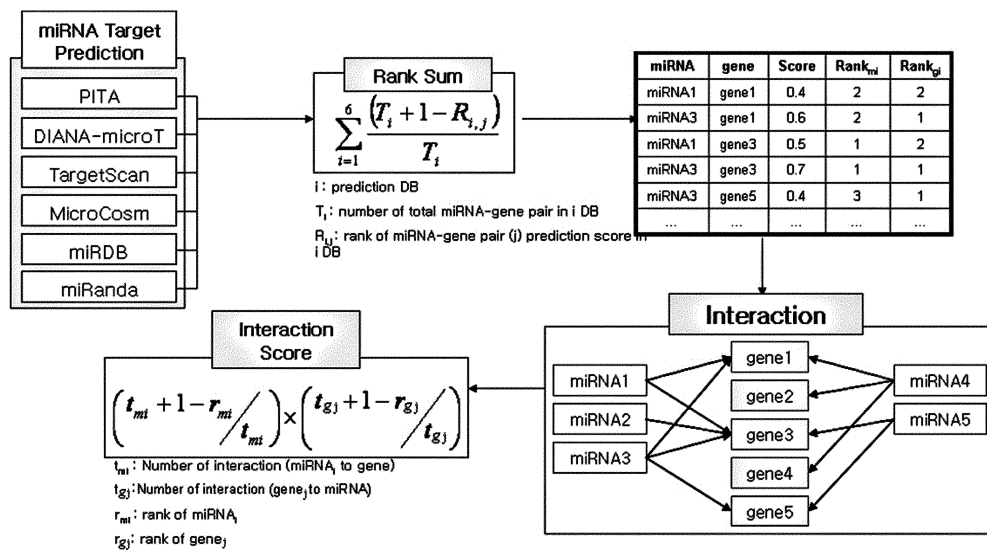
150 : 디스플레이부

## 도면

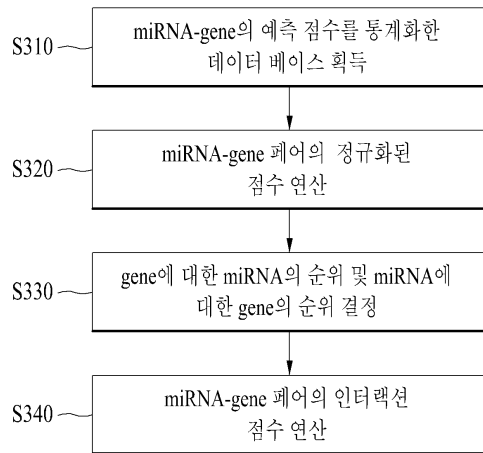
### 도면1



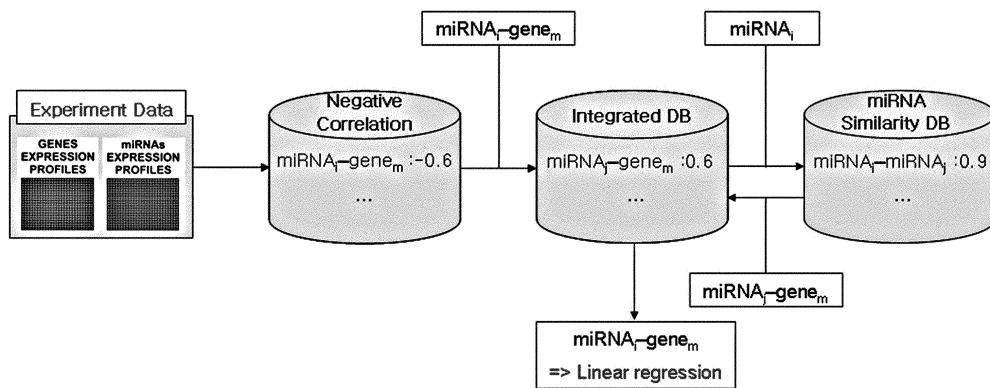
### 도면2



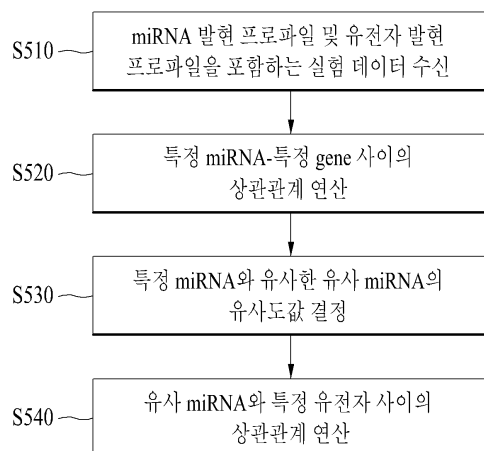
도면3



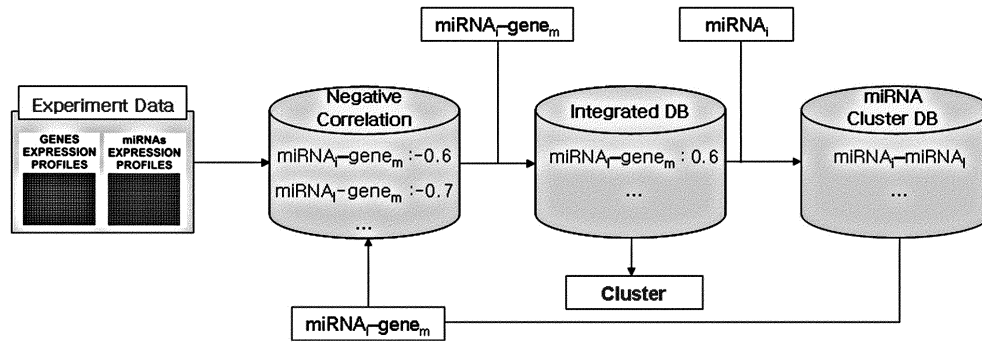
도면4



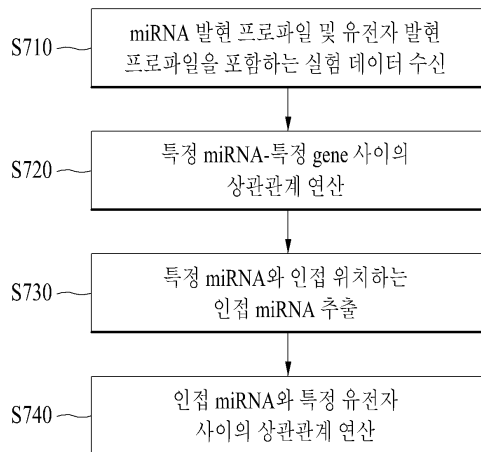
도면5



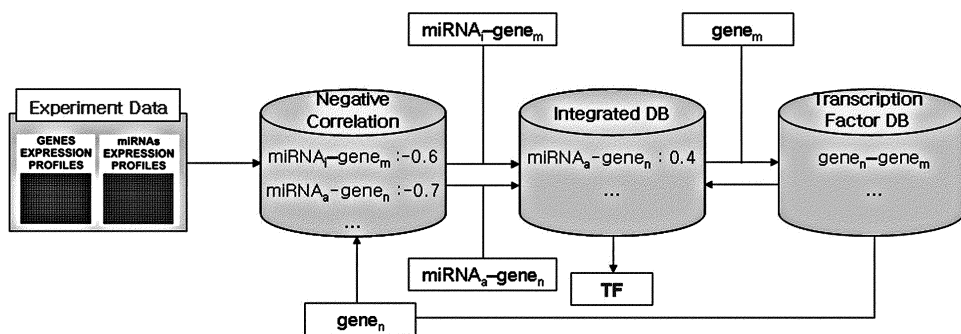
도면6



도면7

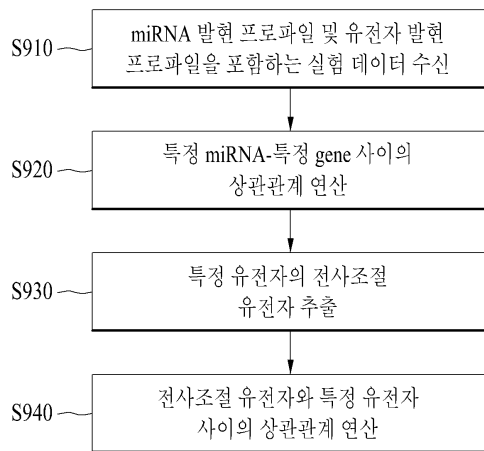


도면8

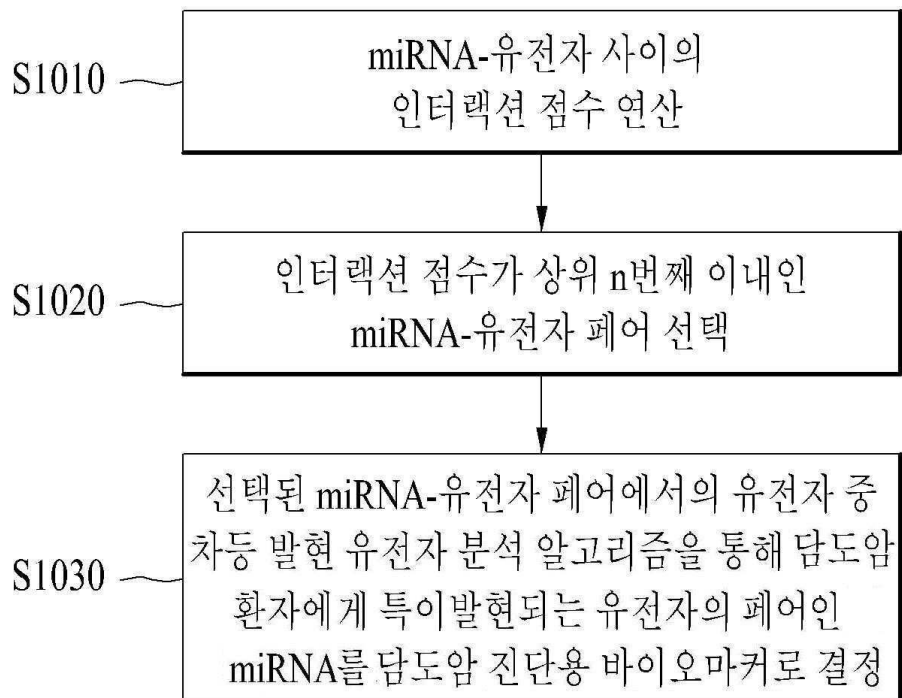




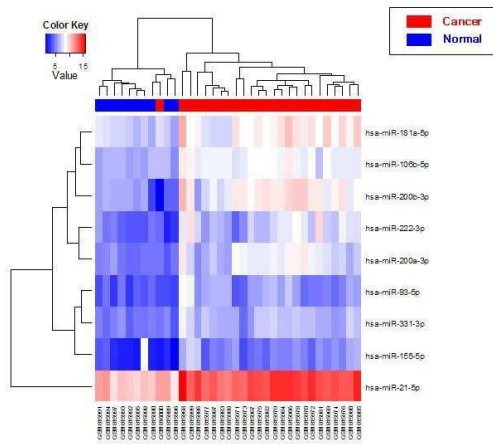
도면9



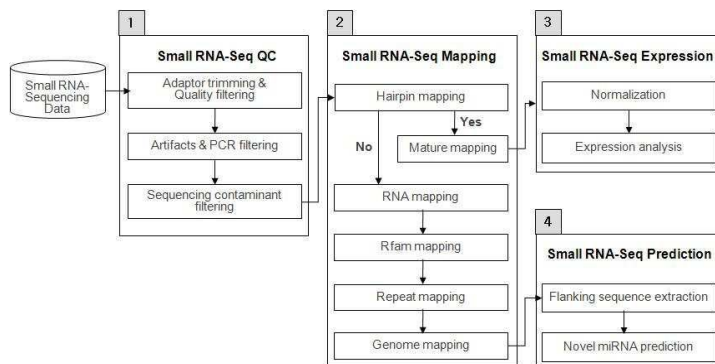
도면10



도면11



도면12



도면13

