



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0017174
(43) 공개일자 2017년02월15일

(51) 국제특허분류(Int. Cl.)
G06F 19/18 (2011.01) C12Q 1/68 (2006.01)
G01N 33/574 (2006.01)
(52) CPC특허분류
G06F 19/18 (2013.01)
C12Q 1/6811 (2013.01)
(21) 출원번호 10-2015-0110709
(22) 출원일자 2015년08월05일
심사청구일자 없음

(71) 출원인
엘지전자 주식회사
서울특별시 영등포구 여의대로 128 (여의도동)
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
최형석
서울특별시 서초구 양재대로11길 19
이재훈
서울특별시 서초구 양재대로11길 19
(뒷면에 계속)
(74) 대리인
정중옥, 조현동, 진천웅

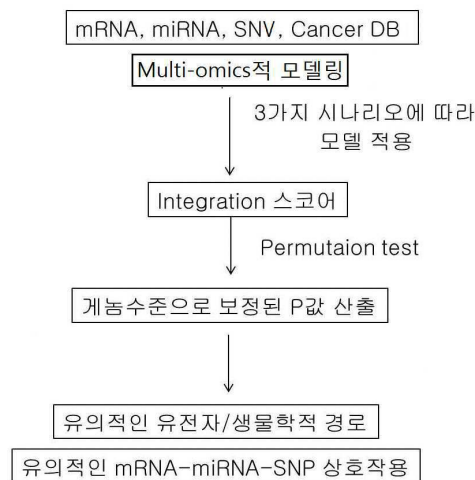
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 바이오마커의 추출방법, 췌장암 진단용 바이오마커 및 이를 이용한 췌장암의 진단 방법

(57) 요약

본 발명은 바이오마커의 추출방법, 췌장암 진단용 바이오마커 및 이를 이용한 췌장암의 진단 방법에 관한 것이다. 본 발명에 따른 알고리즘을 활용하면, 매우 신속하게 각종 질환에 대한 유의적인 관련성을 갖는 바이오마커를 발굴할 수 있다. 이를 통해 본 발명은 높은 특이성 및 민감성을 갖춘 췌장암의 진단용 바이오마커와 췌장암 진단 방법을 제공한다.

대표도 - 도1



(52) CPC특허분류

C12Q 1/6886 (2013.01)

G01N 33/57438 (2013.01)

(72) 발명자

허지연

서울특별시 서초구 양재대로11길 19

송시영

서울특별시 마포구 백범로 205, 104동 1106호(신공
덕동, 펜트하우스)

정다운

서울특별시 양천구 목동동로 437, 1007호(목동, 향
도드림네스트)

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

하기 수학적 식 1의 선형회귀모형에 따라 SNV와 miRNA의 상호작용에 따른 mRNA의 변화를 모델링하는 단계;

[수학적 식 1]

$$M = \beta_S S + \beta'_{MI} MI + \beta_{S:MI} S : MI + \varepsilon$$

(상기 수학적 식 1에서, M 은 mRNA를, S 는 SNP를, MI 는 miRNA 를, $S:MI$ 는 SNP와 miRNA의 상호작용을 의미하며, ε 는 잔여오류(Residual error)를 의미한다)

하기 수학적 식 2의 로지스틱회귀모형에 따라 miRNA와 mRNA의 변화에 따른 질병 또는 이의 유사상태의 발생 확률을 모델링하는 단계;

[수학적 식 2]

$$\text{logit } P(T = 1) = \beta_M M + \beta_{MI} MI$$

(상기 수학적 식 2에서, T 는 질병(암) 여부를 의미하며, β 는 회귀계수를 의미한다)

상기 모델에 따라 산출된 통계량을 다음의 1) 내지 3)의 시나리오로 나누고, 수학적 식 3 내지 5에 따라 각각 인테그레이션 스코어(Integration score)를 산출하는 단계;

1) Case/control 데이터가 있는 경우, SNP 돌연변이에 의해 mRNA와 miRNA의 음의 상관관계(Negative correlation)가 깨지고, 동시에 mRNA의 양이 증가하며 암발병(cancer)의 확률이 증가

[수학적 식 3]

$$G_1 = \sum_{all\ pairs} (t_M^2 + t_{S:MI}^2 + t_S^2) \cdot I(t_{S:MI} > 0 \& t'_{MI} < 0 \& t_M > 0 \& t_S > 0 \& p_{S:MI} < \alpha \& p'_{MI} < \alpha \& p_M < \alpha)$$

2) Case 데이터만 있는 경우, SNP 돌연변이에 의해 mRNA와 miRNA의 음의 상관관계(Negative correlation)가 깨지고, 동시에 mRNA의 양이 증가

[수학적 식 4]

$$G_2 = \sum_{all\ pairs} (t_{S:MI}^2 + t_S^2) \cdot I(t_{S:MI} > 0 \& t'_{MI} < 0 \& t_S > 0 \& p_{S:MI} < \alpha \& p'_{MI} < \alpha)$$

3) case 데이터만 있는 경우, SNP 돌연변이에 의해 mRNA와 miRNA의 음의 상관관계(Negative correlation)이 생기고, 동시에 mRNA의 양이 감소

[수학적 식 5]

$$G_3 = \sum_{all\ pairs} (t_{S:MI}^2 + t_S^2) \cdot I(t_{S:MI} < 0 \& t'_{MI} > 0 \& t_S < 0 \& p_{S:MI} < \alpha)$$

(상기 수학적 식 3 내지 5에서 $\alpha = 0.05$ or 0.1 이고, G_n 은 n 번 시나리오에 따른 인테그레이션 스코어를 의미하고, t 는 t통계량을 의미하며, t_M , p_M 은 각각 상기 수학적 식 2의 β_M 에 대응하는 t통계량 값과 p-value를 의미하며, $t_{S:MI}$, $p_{S:MI}$ 은 각각 상기 수학적 식 1의 $\beta_{S:MI}$ 에 대응하는 t통계량 값과 p-value를 의미하며, t_S , p_S 은 각각 상기 수학적 식 1의 β_S 에 대응하는 t통계량 값과 p-value를 의미하며, t'_{MI} , p'_{MI} 은 각각 상기 수학적 식 1의 β'_{MI} 에 대응하는 t통계량 값과 p-value를 의미한다)

상기에서 산출된 인테그레이션 스코어에 대해 permutation test를 수행하여 상기 mRNA를 코딩하는 유전자 발현

량의 변화가 질병의 발생과 유의적 관계가 있는지를 판단하는 단계를 포함하는 바이오마커의 추출 방법.

청구항 2

제1항에 있어서, 상기 permutation test의 수행은 mRNA의 발현량 값 (m_1, \dots, m_n)을 임의로 섞은 다음, 인테그레이션 스코어를 계산하여 $G_1^{(i)}, \dots, G_n^{(i)}$ 값을 얻는 단계;

상기 과정을 반복하여 인테그레이션 스코어들의 세트를 반복한 횟수(i번)만큼 얻는 단계;

상기에서 j번째 샘플의 P 값을 하기 수학적 식 6과 같이 구하는 단계를 포함하는 바이오마커의 추출방법:

[수학적 식 6]

$$p\text{-value}_j = \frac{\sum_{i=1}^{1000} I(G_j > G_j^{(i)})}{1000}$$

청구항 3

VPS37C(vacuolar protein sorting 37 homolog C(*S. cerevisiae*), Genbank ID: 55048), *ALS2CL*(ALS2 C-terminal like, Genbank ID:259173), *ZDHHC20*(zinc finger, DHHC-type containing 20, Genbank ID:253832), *SNX27*(sorting nexin family member 27, Genbank ID:81609), *CBRI*(carbonyl reductase 1, Genbank ID:873), *ZNRF1*(zinc and ring finger 1, E3 ubiquitin protein ligase, Genbank ID:84937), *CLIC5*(chloride intracellular channel 5, Genbank ID:53405), *TMEM92*(transmembrane protein 92, Genbank ID:162461), *EIF4E2*(eukaryotic translation initiation factor 4E family member 2, Genbank ID:9470), *MTM1*(myotubularin 1, Genbank ID:4534), *HMGB3*(high mobility group box 3, Genbank ID:3149), *TPO*(thyroid peroxidase, Genbank ID:7173), *SRCAP*(Snf2-related CREBBP activator protein, Genbank ID:10847), *ZNF267*(zinc finger protein 267, Genbank ID:10308) 중 적어도 하나의 mRNA 또는 이의 단백질의 수준을 측정하는 체제를 포함하는 체장암 진단용 조성물.

청구항 4

제3항에 있어서, 상기 유전자 mRNA의 수준을 측정하는 체제는 상기 유전자에 특이적으로 결합하는 프라이머를 포함하는 것인 체장암 진단용 조성물.

청구항 5

제3항에 있어서, 상기 단백질의 수준을 측정하는 체제는 상기 단백질에 특이적인 항체를 포함하는 것인 체장암 진단용 조성물.

청구항 6

제3항의 조성물을 포함하는 체장암 진단용 키트.

청구항 7

VPS37C(vacuolar protein sorting 37 homolog C(*S. cerevisiae*), Genbank ID: 55048), *ALS2CL*(ALS2 C-terminal like, Genbank ID:259173), *ZDHHC20*(zinc finger, DHHC-type containing 20, Genbank ID:253832), *SNX27*(sorting nexin family member 27, Genbank ID:81609), *CBRI*(carbonyl reductase 1, Genbank ID:873), *ZNRF1*(zinc and ring finger 1, E3 ubiquitin protein ligase, Genbank ID:84937), *CLIC5*(chloride intracellular channel 5, Genbank ID:53405), *TMEM92*(transmembrane protein 92, Genbank ID:162461), *EIF4E2*(eukaryotic translation initiation factor 4E family member 2, Genbank ID:9470), *MTM1*(myotubularin 1, Genbank ID:4534), *HMGB3*(high mobility group box 3, Genbank ID:3149), *TPO*(thyroid peroxidase, Genbank ID:7173), *SRCAP*(Snf2-related CREBBP activator protein, Genbank ID:10847), *ZNF267*(zinc finger protein 267, Genbank ID:10308) 중 적어도 하나의 유전자에 특이적으로 결합하는 프라이머를 사용하여 체장암 의심 환자의 생물학적 시료로부터 mRNA의 발현량을 측정하는 단계; 및

상기 mRNA의 발현량이 정상 대조구와 비교하여 유의적으로 증가한 경우를 체장암의 발병가능성이 높다고 판단하

는 단계를 포함하는 췌장암 진단을 위한 정보의 제공 방법.

발명의 설명

기술 분야

[0001] 본 발명은 바이오마커의 추출방법, 췌장암 진단용 바이오마커 및 이를 이용한 췌장암의 진단 방법에 관한 것이다.

배경 기술

[0002] 포스트 게놈시대와 NGS 등 기술의 발달로 다양한 멀티오믹스 데이터들의 축적이 촉진되고 있다. 유전체(genome)는 한 생물이 가지는 모든 유전 정보를 말하는데, 핵산 서열, 단백질의 집합체로부터 도출된 유전 정보들은 당뇨병, 암과 같은 질환의 조기진단에 사용될 수 있다.

[0003] 멀티 오믹스적 접근방법(Multi-omics approach)은 각기 다른 생물학적 정보들을 취합하고, 이들 간의 상호관계와 보다 상위개념으로서의 기능을 파악하는 것을 포함한다. 이를 통해 질병에 관여하는 새로운 유전자를 발굴하거나, 유전적 다양성과 개체의 발현 특성 간의 상관관계 등을 보다 정확하게 파악할 수 있다.

[0004] 췌장은 섭취한 음식물 중의 탄수화물, 지방, 단백질을 분해하는 소화 효소를 분비하는 외분비 기능과 혈당을 조절하는 인슐린과 글루카곤 등의 호르몬을 분비하는 내분비 기능을 갖는 기관으로, 이에 발생하는 췌장암은 특별한 초기 증상이 없어 조기발견이 어려운 특징이 있다. 또한 췌장은 두께가 2 cm 정도로 얇으며 피막만으로 싸여 있는데다가 소장내 산소를 공급하는 상장간막 동맥과 장에서 흡수한 영양분을 간으로 운반하는 간문맥 등과 밀착되어 있어 암의 침윤이 쉽게 일어난다. 또한 췌장 후면의 신경 다발과 임파선에도 조기에 전이가 발생하는 특징이 있다. 특히 췌장 암세포는 성장 속도가 빠르다. 발병시 4개월 내지 8개월밖에 살지 못하는 경우가 대부분으로, 예후가 좋지 않고, 수술에 성공하여 증상이 나아진다고 해도 5년 이상 생존하는 비율이 17~24% 정도로 낮은 편이다.

[0005] 췌장암의 진단은 초음파검사, 전산화단층촬영(CT), 자기공명영상(MRI), 내시경적 역행성담췌관조영술(ERCP), 내시경적 초음파검사(EUS) 및/또는 양성자방출단층촬영(PET)을 통해 이루어질 수 있으나, 이와 같은 방법으로는 조기에 췌장암을 발견할 수 없고, 비용 또한 높다. 췌장암의 바이오마커로는 단백질 바이오마커인 CA19-9, CEA 등이 알려져 있으나, 이들 단백질 바이오마커들은 민감도와 특이도가 60% 수준이어서 실제 진단에서 활용도가 매우 낮다. 특히, 조직 특이성이 부족하고, 루이스 항원을 발현하지 않는 혈액형에서는 증가하지 않는 문제점이 있다. 따라서, 민감도와 특이도가 높아서 신뢰할만한 진단을 내릴 수 있는 바이오마커의 개발이 필요하다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국공개특허 2014-0023607호

비특허문헌

[0007] (비특허문헌 0001) Nature reviews. Cancer / v.10 no.6, / 2010, pp.389

발명의 내용

해결하려는 과제

[0008] 본 발명은 바이오마커의 검출 방법; 및 췌장암의 진단이 가능한 바이오마커, 상기 바이오마커를 이용한 췌장암의 진단방법을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0009] 본 발명의 일 측면에 따르면, 하기 수학적 1의 선형회귀모형에 따라 SNV와 miRNA의 상호작용에 따른 mRNA의 변

화를 모델링하는 단계;

[수학식 1]

$$M = \beta_S S + \beta'_{MI} MI + \beta_{S:MI} S : MI + \varepsilon$$

(상기 수학식 1에서, M 은 mRNA를, S 는 SNP를, MI 는 miRNA 를, $S:MI$ 는 SNP와 miRNA의 상호작용을 의미하며, ε 는 잔여오류(Residual error)를 의미한다)

하기 수학식 2의 로지스틱회귀모형에 따라 miRNA와 mRNA의 변화에 따른 질병 또는 이의 유사상태의 발생 확률을 모델링하는 단계;

[수학식 2]

$$\text{logit } P(T = 1) = \beta_M M + \beta_{MI} MI$$

(상기 수학식 2에서, T 는 질병(암) 여부를 의미하며, β 는 회귀계수를 의미한다)

상기 모델에 따라 산출된 통계량을 다음의 1) 내지 3)의 시나리오로 나누고, 수학식 3 내지 5에 따라 각각 인테그레이션 스코어(Integration score)를 산출하는 단계;

1) Case/control 데이터가 있는 경우, SNP 돌연변이에 의해 mRNA와 miRNA의 음의 상관관계(Negative correlation)가 깨지고, 동시에 mRNA의 양이 증가하며 암발병(cancer)의 확률이 증가

[수학식 3]

$$G_1 = \sum_{all\ pairs} (t_M^2 + t_{S:MI}^2 + t_S^2) \cdot I(t_{S:MI} > 0 \& t'_{MI} < 0 \& t_M > 0 \& t_S > 0 \& p_{S:MI} < \alpha \& p'_{MI} < \alpha \& p_M < \alpha)$$

2) Case 데이터만 있는 경우, SNP 돌연변이에 의해 mRNA와 miRNA의 음의 상관관계(Negative correlation)가 깨지고, 동시에 mRNA의 양이 증가

[수학식 4]

$$G_2 = \sum_{all\ pairs} (t_{S:MI}^2 + t_S^2) \cdot I(t_{S:MI} > 0 \& t'_{MI} < 0 \& t_S > 0 \& p_{S:MI} < \alpha \& p'_{MI} < \alpha)$$

3) case 데이터만 있는 경우, SNP 돌연변이에 의해 mRNA와 miRNA의 음의 상관관계(Negative correlation)이 생기고, 동시에 mRNA의 양이 감소

[수학식 5]

$$G_3 = \sum_{all\ pairs} (t_{S:MI}^2 + t_S^2) \cdot I(t_{S:MI} < 0 \& t'_{MI} > 0 \& t_S < 0 \& p_{S:MI} < \alpha)$$

(상기 수학식 3 내지 5에서 $\alpha = 0.05$ or 0.1 이고, G_n 은 n 번 시나리오에 따른 인테그레이션 스코어를 의미하고, t 는 t통계량을 의미하며, t_M, p_M 은 각각 상기 수학식 2의 β_M 에 대응하는 t통계량 값과 p-value를 의미하며, $t_{S:MI}, p_{S:MI}$ 은 각각 상기 수학식 1의 $\beta_{S:MI}$ 에 대응하는 t통계량 값과 p-value를 의미하며, t_S, p_S 은 각각 상기 수학식 1의 β_S 에 대응하는 t통계량 값과 p-value를 의미하며, t'_{MI}, p'_{MI} 은 각각 상기 수학식 1의 β'_{MI} 에 대응하는 t통계량 값과 p-value를 의미한다)

상기에서 산출된 인테그레이션 스코어에 대해 permutation test를 수행하여 상기 mRNA를 코딩하는 유전자 발현량의 변화가 질병의 발생과 유의적 관계가 있는지를 판단하는 단계를 포함하는 바이오마커의 추출 방법이 제공된다.

일 실시예에 있어서, 상기 permutation test의 수행은 mRNA의 발현량 값 (m_1, \dots, m_n)을 임의로 섞은 다음, 인테그레이션 스코어를 계산하여 $G_1^{(i)}, \dots, G_n^{(i)}$ 값을 얻는 단계; 상기 과정을 반복하여 인테그레이션 스코어들의

세트를 반복한 횟수(i번)만큼 얻는 단계; 상기에서 j번째 샘플의 P 값을 하기 수학적 식 6과 같이 구하는 단계를 포함하는 것일 수 있다.

[수학적 식 6]

$$p\text{-value}_j = \frac{\sum_{i=1}^{1000} I(G_j > G_j^{(i)})}{1000}$$

본 발명의 다른 측면에 따르면, *VPS37C*(vacuolar protein sorting 37 homolog C(*S. cerevisiae*), Genbank ID: 55048), *ALS2CL*(ALS2 C-terminal like, Genbank ID:259173), *ZDHHC20*(zinc finger, DHHC-type containing 20, Genbank ID:253832), *SNX27*(sorting nexin family member 27, Genbank ID:81609), *CBRI*(carbonyl reductase 1, Genbank ID:873), *ZNRFI*(zinc and ring finger 1, E3 ubiquitin protein ligase, Genbank ID:84937), *CLIC5*(chloride intracellular channel 5, Genbank ID:53405), *TMEM92*(transmembrane protein 92, Genbank ID:162461), *EIF4E2*(eukaryotic translation initiation factor 4E family member 2, Genbank ID:9470), *MTMI*(myotubularin 1, Genbank ID:4534), *HMGB3*(high mobility group box 3, Genbank ID:3149), *TPO*(thyroid peroxidase, Genbank ID:7173), *SRCAP*(Snf2-related CREBBP activator protein, Genbank ID:10847), *ZNF267*(zinc finger protein 267, Genbank ID:10308) 중 적어도 하나의 mRNA 또는 이의 단백질의 수준을 측정하는 체제를 포함하는 체장암 진단용 조성물이 제공될 수 있다.

일 실시예에 있어서, 상기 유전자 mRNA의 수준을 측정하는 체제는 상기 유전자에 특이적으로 결합하는 프라이머를 포함할 수 있다.

일 실시예에 있어서, 상기 단백질의 수준을 측정하는 체제는 상기 단백질에 특이적인 항체를 포함할 수 있다.

본 발명의 또 다른 측면에 따르면, 상기의 조성물을 포함하는 체장암 진단용 키트가 제공된다.

본 발명의 또 다른 측면에 따르면, *VPS37C*(vacuolar protein sorting 37 homolog C(*S. cerevisiae*), Genbank ID: 55048), *ALS2CL*(ALS2 C-terminal like, Genbank ID:259173), *ZDHHC20*(zinc finger, DHHC-type containing 20, Genbank ID:253832), *SNX27*(sorting nexin family member 27, Genbank ID:81609), *CBRI*(carbonyl reductase 1, Genbank ID:873), *ZNRFI*(zinc and ring finger 1, E3 ubiquitin protein ligase, Genbank ID:84937), *CLIC5*(chloride intracellular channel 5, Genbank ID:53405), *TMEM92*(transmembrane protein 92, Genbank ID:162461), *EIF4E2*(eukaryotic translation initiation factor 4E family member 2, Genbank ID:9470), *MTMI*(myotubularin 1, Genbank ID:4534), *HMGB3*(high mobility group box 3, Genbank ID:3149), *TPO*(thyroid peroxidase, Genbank ID:7173), *SRCAP*(Snf2-related CREBBP activator protein, Genbank ID:10847), *ZNF267*(zinc finger protein 267, Genbank ID:10308) 중 적어도 하나의 유전자에 특이적으로 결합하는 프라이머를 사용하여 체장암 의심 환자의 생물학적 시료로부터 mRNA의 발현량을 측정하는 단계; 및 상기 mRNA의 발현량이 정상 대조구와 비교하여 유의적으로 증가한 경우를 체장암의 발병가능성이 높다고 판단하는 단계를 포함하는 체장암 진단을 위한 정보의 제공 방법이 제공된다.

발명의 효과

본 발명에 따른 알고리즘을 활용하면, 매우 신속하게 각종 질환에 대한 유의적인 관련성을 갖는 바이오마커를 발굴할 수 있다. 이를 통해 본 발명은 높은 특이성 및 민감성을 갖춘 체장암의 진단용 바이오마커와 체장암 진단 방법을 제공한다.

도면의 간단한 설명

도 1은 본 발명의 알고리즘을 나타내는 개요도이다.

도 2는 본 발명의 멀티 오믹스 데이터를 시나리오 1에 따라 분석한 결과를 그래프로 나타낸 것이다.

도 3은 본 발명의 멀티 오믹스 데이터를 시나리오 2, 3에 따라 분석한 결과를 그래프로 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

특정 질병 상태에 대한 바이오마커의 선별은 특정의 의학적 적용을 위하여 질병집단에서 주목할만하며 통계학적

으로 유의한 차이를 가지는 마커의 확인을 포함한다. 바이오마커는 질병의 발병 또는 유사한 상태의 진행에서 발병 조직 또는 주변 조직 및 종양에 대응하는 순환 세포로부터 혈류로 분산되는, 분비되거나 방출된 분자를 포함할 수 있다. 바이오마커는 소분자(small molecules), 펩티드(peptides), 단백질(proteins) 및 핵산(nucleic acid)을 포함할 수 있다. 바이오마커의 확인에 영향을 미치는 몇 가지 주요 문제들로는 이용가능한 데이터의 오버피팅(over-fitting) 및 데이터에서의 편향(bias)이 포함된다.

[0041] 바이오마커에 대한 용어 "민감도"는 질병에 걸린 개인을 정확하게 분류하는 성능(performance)을 나타내며, "특이도"는 질병에 걸리지 않은 개인을 정확하게 분류하는 것에 대한 바이오마커의 성능을 나타낸다.

[0042] 본원에서 "생물학적 샘플(biological sample)", "샘플(sample)" 및 "시료(test sample)"는 개인으로부터 유래된 임의의 물질, 생물학적 체액, 조직 또는 세포를 말하는 것으로, (전혈(whole blood), 백혈구(leukocytes), 말초혈액 단핵 세포(peripheral blood mononuclear cells), 백혈구 연층(buffy coat), 혈장(plasma) 및 혈청(serum)을 포함하는) 혈액, 객담(sputum), 눈물(tears), 점액(mucus), 세비액(nasal washes), 비강 흡인물(nasal aspirate), 호흡(breath), 소변(urine), 정액(semen), 침(saliva), 복강 세척액(peritoneal washings), 복수(ascites), 낭종액(cystic fluid), 뇌척수막 액(meningeal fluid), 양수(amniotic fluid), 선액(glandular fluid), 췌장액(pancreatic fluid), 림프액(lymph fluid), 흉수(pleural fluid), 유두 흡인물(nipple aspirate), 기관지 흡인물(bronchial aspirate), 활액(synovial fluid), 관절 흡인물(joint aspirate), 기관 분비물(organ secretions), 세포(cell), 세포 추출물(cell extract) 및 뇌척수액(cerebrospinal fluid)을 포함한다. 이것은 또한 모든 절차 중 실험적으로 분리된 단편(fractions)을 포함한다. 예를 들어, 혈액 샘플은 혈청 또는 적혈구 세포 또는 백혈구 세포(white blood cells; leukocytes)와 같은 특정 형태의 혈액 세포를 포함하는 단편으로 단편화될 수 있다. 만일 원한다면, 샘플은 조직 및 체액 샘플의 조합과 같은 개인으로부터의 샘플 조합일 수 있다.

[0043] 상기 용어 "생물학적 샘플(biological sample)"은 또한 예를 들어, 대변 샘플, 조직 샘플 또는 조직 생검으로부터의 샘플과 같은 균질화된 고체 물질을 포함하는 물질을 포함한다. 상기 용어 "생물학적 샘플"은 또한 조직 배양 또는 세포 배양으로부터 유래된 물질을 포함한다. 생물학적 샘플을 얻기 위한 임의의 적절한 방법이 사용될 수 있고; 대표적인 방법은 예를 들어, 채혈(phlebotomy), 면봉법(swab)(예를 들어, 구강상피세포 면봉법(buccal swab)) 및 세침 흡인 조직검사(fine needle aspirate biopsy procedure)를 포함한다. 세침 흡인에 적합한 대표적인 조직은 림프절, 폐, 폐 세척액, BAL(기관지폐포 세척액, bronchoalveolar lavage), 갑상선(thyroid), 유방(breast) 및 간(liver)을 포함한다. 샘플들은 또한 예를 들어, 미세절개(micro dissection)(예를 들어, 레이저 포획 미세절개(laser capture microdissection; LCM) 또는 레이저 미세 절개(laser micro dissection; LMD)), 방광 세척(bladder wash), 도말(smear)(예를 들어, PAP 도말) 또는 유즙도관 세척법(ductal lavage)에 의해 수집될 수 있다.

[0044] 본원에서 "진단하다(diagnose)", "진단하는 것(diagnosing)", "진단(diagnosis)" 및 이들의 변화형은 개인에게 관련된 하나 또는 그 이상의 징후, 증상, 데이터 또는 다른 정보에 기초하여, 개인의 건강 상태 또는 상황의 발견, 판단 또는 인지를 말한다. 개인의 건강 상태는 건강한(healthy)/정상(normal)(즉, 질병 또는 질환이 없음의 진단)으로 진단되거나, 건강이 나쁜(ill)/비정상(abnormal)(즉, 질병 또는 질환의 존재 또는 특성의 판단)으로 진단될 수 있다.

[0045] 상기 용어 "진단하다", "진단하는 것", "진단" 등은 특정 질병 또는 질환에 관하여 질병의 초기 발견; 질병의 특성 또는 분류; 질병의 진행(progress), 차도(remission) 또는 재발(recurrence)의 발견; 개인에 대한 처치 또는 치료의 시행 후 질병 반응의 발견을 포함한다. 췌장암의 진단은 암에 걸리지 않은 개인들로부터 암에 걸린 개인을 구별하는 것을 포함한다. 그것은 또한 췌장암으로부터 정상 대조군을 구별하는 것을 포함한다.

[0046] 본원에서 "예측하다(prognose)", "예측하는 것(prognosing)", "예측(prognosis)" 및 그것들의 변화형은 질병 또는 질환을 가지는 개인에서 질병 또는 질환의 진행 과정의 예측(예를 들어, 환자 생존율을 예측하는 것)을 말하며, 이러한 용어는 개인에 대한 처치 또는 치료의 시행 후 질병 반응의 평가를 포함한다.

[0047] 본원에서 "평가하다(evaluate)", "평가하는 것(evaluating)", "평가(evaluation)" 및 그것들의 변화형은 "진단하다(diagnose)" 및 "예측하다(prognose)" 모두를 포함하며, 또한 질병을 가지지 않은 개인에서의 질병 또는 질환의 진행 과정에 대한 판단 또는 예측뿐만 아니라, 질병이 분명히 치료된 개인에서의 질병 및 질환이 재발할 가능성에 관한 판단 또는 예측을 포함한다. 상기 용어 "평가하다"는 또한 예를 들어, 개인이 치료제에 대하여 순조롭게 반응할 것인지 아니면 치료제에 대하여 반응하지 않을 것인지(또는, 예를 들어, 독성을 경험하거나 다른 바람직하지 않은 부작용을 경험할 것인지)를 예측하는 것, 개인에게 투여하기 위한 치료제를 선택하는 것 또

는 개인에 대하여 시행된 치료에 대한 개인의 반응을 관찰하거나 발견하는 것과 같은 치료에 대한 개인의 반응을 평가하는 것을 포함한다. 따라서, 췌장암을 "평가하는 것"은 예를 들어 임의의 하기의 것: 개인에게서 췌장암의 진행 과정을 예측하는 것; 췌장암이 분명히 치료된 환자에게서 췌장암의 재발을 예측하는 것; 또는 췌장암 치료에 대한 개인의 반응을 판단하거나 예측하는 것 또는 개인의 생물학적 샘플로부터 유래된 바이오마커 값의 측정에 기초하여 개인에게 시행하기 위한 췌장암 치료를 선택하는 것을 포함할 수 있다.

[0048] 임의의 하기의 예들은 췌장암을 "진단하는 것" 또는 "평가하는 것" 중 하나: 췌장암의 존재 여부를 초기에 발견하는 것; 췌장암의 특정 단계, 형태 또는 서브-타입, 또는 다른 분류 또는 특성을 판단하는 것; 의심이 가는 종괴가 양성 병변 또는 악성 췌장 종양인지를 판단하는 것; 또는 췌장암의 진행(예를 들어, 췌장암의 성장 또는 전이 속도를 관찰하는 것), 차도 또는 재발을 발견/관찰하는 것으로서 칭해질 수 있다.

[0049] 본원에서 질병을 '예측', '진단', 또는 '평가'하는 행위는 생물학적 샘플에서 유전자의 발현량을 측정하는 것을 포함할 수 있다. 유전자의 발현량은 유전자의 mRNA 레벨 또는 단백질의 발현 레벨을 검출하여 측정할 수 있다.

[0050] 생물학적 샘플에서 messenger RNA (mRNA)를 측정하는 것은 생물학적 샘플에서 대응하는 단백질 레벨의 검출을 위한 대리물(surrogate)로서 사용될 수 있으며, 그 역도 마찬가지이다.

[0051] mRNA 발현 레벨은 역전사 중합효소연쇄반응(reverse transcription quantitative polymerase chain reaction; RT-PCR)(qPCR에 따른 RT-PCR)에 의해 측정될 수 있다. RT-PCR은 mRNA로부터 cDNA를 만드는데 사용된다. 상기 cDNA는 DNA 증폭 과정의 진행으로서 형광을 만들어내기 위하여 qPCR에서 사용될 수 있다. 평균 곡선과 비교하여, qPCR은 세포당 mRNA 복제의 수와 같은 절대 측정(absolute measurement)을 만들어낼 수 있다. 모세관 전기이동과 조합된 노던 블롯(Northern blots), 마이크로어레이(microarrays), 침입자 검정법(invader assay) 및 RT-PCR 모두 샘플에서 mRNA의 발현 레벨을 측정하는데 사용되고 있다(Gene Expression Profiling: Method and Protocols, Richard A. Shinkets, editor, Humana Press, 2004를 참조). 또한 NGS기반 대규모 RNA 서열 분석(High-throughput RNA Sequencing, RNA-seq)으로 mRNA의 발현 레벨이 측정될 수 있다.

[0052] 본원에서 “단백질 발현수준 측정”이란 질병을 진단하기 위하여 생물학적 시료에서 질병의 마커 유전자로부터 발현된 단백질의 존재 여부와 발현 정도를 확인하는 과정으로, 바람직하게는, 상기 유전자의 단백질에 대하여 특이적으로 결합하는 항체를 이용하여 단백질의 양을 확인할 수 있다. 이를 위한 분석 방법으로는 웨스턴 블랏, 엘라이자(enzyme linked immunosorbent assay, ELISA), 방사선면역분석(RIA: Radioimmunoassay), 방사 면역 확산법(radioimmunodiffusion), 오우크테로니(Ouchterlony) 면역 확산법, 로켓(rocket) 면역전기영동, 조직면역 염색, 면역침전 분석법(Immunoprecipitation Assay), 보체 고정 분석법(Complement Fixation Assay), 유세포분석(Fluorescence Activated Cell Sorter, FACS), 단백질 칩(protein chip) 등이 있으나 이로 제한되는 것은 아니다.

[0053] 본원에서 기재하고 있는 microRNA (miRNA)는 비암호화(non-coding)이지만 유전자 발현을 조절하는 소형 RNA이다. mRNA 발현 레벨의 측정에 적합한 임의의 방법들도 대응하는 miRNA에 대하여 사용될 수 있다. 또한 NGS 기반의 small-RNA sequencing (sRNA-seq) 방법으로 miRNA 레벨이 측정될 수 있다. 최근 많은 연구소들이 질병에 대한 바이오마커로서 miRNAs의 사용을 연구하고 있다. miRNA가 다양한 생물학적 프로세스에 관여하는 유전자의 전사후 (Post-transcription) 발현을 조절하기 때문에, 최근 다양한 질병에 대한 바이오마커로서 miRNA가 사용되고 있다.

[0055] 이하, 본 발명에 따른 알고리즘과 췌장암 바이오마커에 대하여 보다 상세하게 설명한다.

[0056] 본 발명에 따른 알고리즘은 멀티 오믹스 데이터로부터 도출되는 관련 정보들을 기반으로 한다.

[0057] 도 1을 참고하면, 본 발명의 췌장암 진단 방법에 대한 개요를 파악할 수 있다.

[0058] I. 오믹스 데이터의 통합적 분석

[0059] 본 발명의 알고리즘은 도 1에 나타나 있는 바와 같이 mRNA, miRNA, SNV(Single Nucleotide Variation) 및 Cancer 데이터베이스 각각을 통해 도출되는 정보들과 이들 상호간의 상관관계에 따른 정보를 멀티 오믹스적 접근법(Multi-omics approach)에 기반하여 특정 유전자와 췌장암과의 연관성을 추정한다. 보다 바람직하게는, 상기 SNV 데이터베이스는 SNP, 체세포변이(Somatic mutation), 생식세포변이(germline mutation)을 포함할 수 있다.

[0061] 상기 멀티 오믹스적 접근법은 다음의 측면들을 포함한다.

[0063] mRNA, miRNA 데이터베이스 간

[0064] miRNA-mRNA 간의 상호작용은 공개된 miRNA 표적 mRNA 예측 틀에서 miRNA와 이의 표적 유전자 사이의 상보 결합 정도를 수치화한 상호작용 스코어(Interaction Score)를 연산하여 나타낼 수 있다. 상호작용 스코어를 통해 miRNA와 이의 표적 유전자 사이의 상보 결합이 발생할 가능성의 높고 낮음을 판단할 수 있다. 즉, miRNA가 표적 유전자의 발현을 조절할 가능성의 높고 낮음을 판단할 수 있다.

[0065] 적어도 하나 이상의 miRNA 표적 예측(miRNA Target Prediction) 틀을 이용하여 miRNA와 유전자 사이의 예측 점수(Prediction score)를 정규화한 데이터 베이스를 획득할 수 있다. 상기miRNA 표적 예측 틀은 표적 유전자와, 상기 표적 유전자에 상보적으로 결합하여 표적 유전자가 단백질로 만들어지는 과정을 억제할 수 있는 정도 또는 miRNA-mRNA 페어(pair)의 결합의 정도를 수치로 예측하는 소프트웨어적 틀을 의미한다. miRNA-mRNA 페어들의 예측 점수를 획득하기 위한 miRNA 표적 예측 틀로는, TargetScan, miRDB, DIANAmicroT, PITA, miRanda MicroCosm, RNAhybrid, PicTar, RNA22 등이 포함될 수 있다.

[0066] 표적 예측 틀을 이용하면, miRNA와 상보 결합할 수 있는 다양한 유전자 사이의 예측 점수가 계산될 수 있다. 또한 miRNA와 상호 결합하는 위치가 두 개 이상인 경우 그 수치를 합하는 방식으로(Additive) 계산될 수 있다. 예측 점수가 작을수록 miRNA에 의한 표적 유전자의 발현 조절 가능성이 낮아지는 것을 의미한다.

[0067] miRNA-mRNA 페어(pair) 사이의 예측 점수의 신뢰성을 높이기 위해, 하나의 표적 예측 틀을 이용하는 것보다 복수의 표적 예측 틀을 이용하여 복수의 데이터베이스를 획득하는 것이 바람직하다. 복수의 표적 예측 틀을 이용하여, miRNA-mRNA 페어의 예측 점수를 통계화한 복수의 데이터 베이스를 획득한 경우, miRNA-mRNA 페어의 예측 점수의 순위를 기초로, 정규화된 점수를 연산할 수 있다. 하지만miRNA 표적 예측 틀이 이용하는 정보들이 다르고, 각 데이터베이스마다 예측 점수의 스케일과 단위가 서로 다르게 적용되고 있기 때문에, 복수의 데이터 베이스를 이용하고자 하는 경우 이의 정규화가 필수적이다. 상기 miRNA-mRNA 페어의 예측 점수를 정규화하는 방법의 일례로 한국공개특허 2015-0043789의 기재내용을 들 수 있다.

[0069] mRNA, miRNA 및 SNV 데이터베이스 간

[0070] miRNA는 표적유전자의 3' untranslated region (3' UTR) 부분과 결합하여 해당 mRNA의 발현을 억제하는 역할을 한다. 표적유전자의 3' UTR부분에 존재하는 SNP 및 돌연변이를 포함하는 SNV가 miRNA와 mRNA의 상호작용에 영향을 미칠 것으로 기대하여 이를 맵핑하고 이에 대한 결합가능성 수치를 얻을 수 있다.

[0072] Cancer 데이터베이스와 기타 데이터베이스 간

[0073] 상기 분석되는 결과들은 cancer DB를 참조함으로써, 기존에 알려져 있는 유전자 리스트와 비교하여 본 발명의 알고리즘을 통해 도출되는 결과의 신뢰성을 재고할 수 있고, miRNA, mRNA, SNV의 상호작용이 타겟 질환 발생에 통계적으로 얼마나 유의하게 영향을 미치는 지에 대한 통계치를 부여할 수 있다.

[0075] II. 인테그레이션 스코어(Integration score)의 계산

[0076] 본 발명의 알고리즘은 SNP와 miRNA의 상호작용과 이에 따른 mRNA와 질병에 걸릴 확률을 모델링하기 위해 먼저 상호작용에 대한 선형회귀모형 및 로지스틱회귀모형을 적합(fitting)한다.

[0078] 상호작용 회귀모형

[0079] [수학식 1]

[0080]
$$M = \beta_S S + \beta_{MI} MI + \beta_{S:MI} S:MI + \varepsilon$$

- [0081] [수학식 2]
- [0082] $\text{logit } P(T=1) = \beta_M M + \beta_{MI} MI$
- [0083] 상기 수학식 1은 SNP와 miRNA의 상호작용에 따른 mRNA의 발현량 변화를 모델링한 것이다.
- [0084] 상기 수학식 1에서,
- [0085] M 은 mRNA를 의미한다.
- [0086] S 는 SNP를 의미한다. $\beta_S S$ 는 SNP의 유무에 따른 mRNA의 발현량 변화 정도를 의미한다.
- [0087] MI 는 miRNA를 의미한다. $\beta'_{MI} MI$ 는 miRNA의 발현량에 따른 mRNA의 발현량 변화 정도를 의미한다.
- [0088] $S:MI$ 는 SNP와 miRNA의 상호작용(interaction)을 의미한다. (SNP:miRNA)
- [0089] $\beta_{S:MI} S:MI$ 는 SNP와 miRNA의 상호작용에 따른 mRNA발현량 변화 정도를 의미한다.
- [0090] ε 는 잔여오류(Residual error)를 의미한다.
- [0092] 상기 수학식2는 miRNA와 mRNA의 변화에 따른 질병의 확률을 모델링한 것이다.
- [0093] 상기 수학식 2에서,
- [0094] T 는 질병(암) 여부를 의미한다.
- [0095] β 는 회귀계수를 의미한다.
- [0097] miRNA-mRNA 상호작용에 대한 SNV의 효과를 세 가지 시나리오에 따라 분류하고 이에 따라 인테그레이션 스코어(Integration score)를 각각 다르게 계산한다.
- [0098] 시나리오 1) Case/control 데이터가 있는 경우, SNV에 의해 mRNA와 miRNA의 음의 상관관계(Negative correlation)가 깨지고, 동시에 mRNA의 양이 증가하며 암발병(cancer)의 확률이 증가함.
- [0099] [수학식 3]
- [0100]
$$G_1 = \sum_{\text{all pairs}} (t_M^2 + t_{S:MI}^2 + t_S^2) \cdot I(t_{S:MI} > 0 \& t'_{MI} < 0 \& t_M > 0 \& t_S > 0 \& p_{S:MI} < \alpha \& p'_{MI} < \alpha \& p_M < \alpha)$$
- [0101] 시나리오 2) Case 데이터만 있는 경우, SNV에 의해 mRNA와 miRNA의 음의 상관관계(Negative correlation)가 깨지고, 동시에 mRNA의 양이 증가함.
- [0102] [수학식 4]
- [0103]
$$G_2 = \sum_{\text{all pairs}} (t_{S:MI}^2 + t_S^2) \cdot I(t_{S:MI} > 0 \& t'_{MI} < 0 \& t_S > 0 \& p_{S:MI} < \alpha \& p'_{MI} < \alpha)$$
- [0104] 시나리오 3) case 데이터만 있는 경우, SNV에 의해 mRNA와 miRNA의 음의 상관관계(Negative correlation)이 생기고, 동시에 mRNA의 양이 감소함.
- [0105] [수학식 5]
- [0106]
$$G_3 = \sum_{\text{all pairs}} (t_{S:MI}^2 + t_S^2) \cdot I(t_{S:MI} < 0 \& t'_{MI} > 0 \& t_S < 0 \& p_{S:MI} < \alpha)$$
- [0107] 상기 식에서 $\alpha = 0.05$ or 0.1 이고,
- [0108] G_n 은 n 번 시나리오에 따른 인테그레이션 스코어를 의미한다.
- [0109] 이때, 상기 수학식 3 내지 5에서,
- [0110] t 는 t통계량을 의미한다

- [0111] t_M, p_M 은 각각 상기 수학적 식 2의 β_M 에 대응하는 t통계량 값과 p-value를 의미하며, 질병과 mRNA와의 유의적 연관성을 나타낸다.
- [0112] $t_{S:MI}, p_{S:MI}$ 은 각각 상기 수학적 식 1의 $\beta_{S:MI}$ 에 대응하는 t통계량 값과 p-value를 의미하며, mRNA와 SNP:miRNA의 유의적 연관성을 나타낸다.
- [0113] t_s, p_s 은 각각 상기 수학적 식 1의 β_s 에 대응하는 t통계량 값과 p-value를 의미하며, mRNA와 SNP의 유의적 연관성을 나타낸다.
- [0114] t'_{MI}, p'_{MI} 은 각각 상기 수학적 식 1의 β'_{MI} 에 대응하는 t통계량 값과 p-value를 의미하며, mRNA와 miRNA의 유의적 연관성을 나타낸다.

[0116] 시나리오 1의 경우에 GEO 데이터베이스(GSE32688)에서 얻은 췌장암 (PAAD; case) 25명과 정상 (normal; control) 7명에 대해 상기 선형회귀모형 (수학적 식 1)과 로지스틱회귀모형(수학적 식 2)을 적합하고 $\alpha=0.05$ 를 적용하여 인테그레이션 스코어(integration score)(상기 수학적 식 3)를 계산하였다. 각 회귀모형에 대한 모형적합도는 mRNA, miRNA와 SNP의 조합(이 데이터에서 총 조합의 수는 4,032개)에 따라 달라지게 되고 (R^2 : 0~0.8), mRNA, miRNA, SNP의 상호작용에 대한 최종 유의성은 모형적합도가 아닌 인테그레이션 스코어의 유의확률로 판단한다. 인테그레이션 스코어의 검정은 정규분포나 이산형분포 등의 분포가정이 없는 비모수 방법인 permutation test를 사용하였고, 유의성에 대한 평가 지수(metric)로 유의확률(p-value)이 구해진다. 그리고, 시나리오 2와 3의 경우는 TCGA(The Cancer Genome Atlas) 데이터베이스로부터 얻은 47명의 췌장암 환자데이터에 적용하였고 선형회귀모형(수학적 식 1)만이 적합되었고 $\alpha=0.1$ 을 적용하여 인테그레이션 스코어(수학적 식 4, 5)를 계산하였다. 사용된 TCGA데이터에서 상호작용의 총 조합의 수는 46개이고 조합마다 모형적합도가 다르며 최종 유의성은 인테그레이션 스코어(수학적 식 4, 5)의 유의확률로 판단한다.

[0118] III. 인테그레이션 스코어에 대한 유의확률값 산출

- [0119] 상기 얻어진 인테그레이션 스코어에 대하여 Permutation test를 수행함으로써, 유의확률(P-value)을 얻고 유전체 전체 수준에서 해석한다.
- [0120] Permutation test는 다음과 같은 방법으로 수행될 수 있다.
- [0121] 1) Case 샘플수가 n_1 개, control 샘플수가 n_2 개가 있는 경우, 상기 단계에 의해 $n (= n_1 + n_2)$ 개의 인테그레이션 스코어 (G_1, \dots, G_n)가 얻어짐
- [0122] 2) mRNA의 발현량 값 (m_1, \dots, m_n)을 임의로 섞고 (permutation) 다시 인테그레이션 스코어를 계산함 ($G_1^{(i)}, \dots, G_n^{(i)}$)
- [0123] 3) 2)과정을 10,000번 반복하여 인테그레이션 스코어들의 set을 얻음 ($i=1, \dots, 10,000$)
- [0124] 4) j번째 샘플의 p-value는 다음과 같이 구한다.
- [0125] [수학적 식 6]

$$p\text{-value}_j = \frac{\sum_{i=1}^{1000} I(G_j > G_j^{(i)})}{1000}$$

- [0126]
- [0127] 일반적으로 유의확률이 0.05보다 작거나 같으면 통계적으로 유의미하고, 0.05보다 크면 의미가 없는 것으로 해석하지만, 지금과 같이 상호작용의 조합의 수가 방대한 경우는 전체 게놈(genome-wide) 수준에서 p-value를 해석한다. 본 분석에서는 총 조합의 수가 많은 (4,032개) GEO데이터의 경우는 $p\text{-value} < 5.0E-04$ 인 경우에 유의하다고 판단하였고 조합의 수가 적었던(46개) TCGA 데이터의 경우는 $p\text{-value} < 0.05$ 인 경우에 유의하다고 판단하였다.

[0129] **IV. mRNA-miRNA-SNP 상관관계의 가시화 및 알고리즘의 적합성 테스트**

[0130] 도 2는 본 발명의 멀티 오믹스 데이터를 시나리오 1에 따라 분석한 결과를 그래프로 나타낸 것이다. Y축은 세 개의 그래프에서 동일하게 mRNA의 발현량이고 X축은 왼쪽 그래프부터 차례대로 miRNA의 발현량, SNP의 reference/mutation 여부, group의 case/control여부를 나타낸다.

[0131] 도 2의 CLDN4는 췌장암과의 상관관계가 이미 잘 알려져 있는 유전자이다. 도 2의 좌측 그래프 A는 miRNA의 발현량(X축)에 따른 mRNA의 발현량(Y축)의 변화를 나타내고 있는데, 점선은 miRNA와 연관된 SNP에 돌연변이가 없는 샘플들의 추세를 나타내며, miRNA의 발현량과 mRNA의 발현량이 음(-)의 상관관계를 나타냄을 알 수 있다.

[0132] 그래프 A의 굵은 실선은 SNP에 돌연변이가 있는 경우로서, miRNA의 발현량과 mRNA의 발현량이 양(+)의 상관관계를 나타냄을 알 수 있는데, 이와 같이 SNP의 돌연변이 여부에 따라 miRNA:mRNA의 발현량이 역상관관계를 갖게됨을 알 수 있다.

[0133] 도 2의 그래프 B는 SNP가 정상인 경우(Reference)와 돌연변이인 경우(Mutation)인 경우의 mRNA 레벨을 나타내는 것으로서, 그래프 B에 따르면, 본 발명의 시나리오 1에 따라 CLDN4 유전자의 mRNA의 발현량은 SNP_A-2031896의 돌연변이에 의해 유의적으로 증가함을 알 수 있다. 그래프 C에 따르면, 췌장암 환자의 경우(case) CLDN4의 mRNA 발현량이 증가함을 알 수 있다.

[0134] 도 3은 본 발명의 멀티 오믹스 데이터를 시나리오 2, 3에 따라 분석한 결과를 그래프로 나타낸 것이다. Y축은 두 개의 그래프에서 동일하게 mRNA의 발현량이고, 그래프 A의 X축은 miRNA의 발현량, 그래프 B의 X축은 SNP의 정상/돌연변이(reference/mutation) 여부를 나타내며, 상기 도 2와 마찬가지로 SNP의 정상/돌연변이(reference/mutation) 여부에 따라 mRNA가 역상관관계를 갖는다.

[0135] 여기서, 도 3의 MTM1은 본 발명의 알고리즘에 의해 새롭게 발견된 췌장암 관련 바이오마커 유전자이다. 이와 같이 본 발명에 따라 새롭게 발견된 바이오마커 유전자 및 이에 결합가능한 miRNA, 관련 SNV는 다음과 같다.

표 1

[0136]

유전자 ID	Description	Gene ID (NCBI)	결합가능한 miRNA	관련 SNV
<i>VPS37C</i>	vacuolar protein sorting 37 homolog C (S. cerevisiae)	55048	<i>miR-940</i> , <i>miR-506-3p</i> , <i>miR-612</i> , <i>miR-564</i> , <i>miR-885-5p</i>	SNP_A-8710767
<i>ALS2CL</i>	ALS2 C-terminal like	259173	<i>miR-510</i> , <i>miR-490-5p</i> , <i>miR-612</i> , <i>miR-939</i> , <i>miR-595</i> , <i>miR-1299</i> , <i>miR-635</i> , <i>miR-7-2-3p</i> , <i>miR-921</i> , <i>miR-598</i> , <i>miR-944</i> , <i>miR-361-3p</i> , <i>miR-575</i>	SNP_A-8691029, SNP_A-1838826, SNP_A-2071775
<i>ZDHHC20</i>	zinc finger, DHHC-type containing 20	253832	<i>miR-548a-5p</i>	SNP_A-1894930
<i>SNX27</i>	sorting nexin family member 27	81609	<i>miR-638</i> , <i>miR-665</i> , <i>miR-525-5p</i> , <i>miR-372</i> , <i>miR-1285-3p</i> , <i>miR-520b</i> , <i>miR-580</i> , <i>miR-892b</i> , <i>miR-940</i> , <i>miR-769-3p</i> , <i>miR-631</i> , <i>miR-517a-3p</i>	SNP_A-8545494, SNP_A-8532774
<i>CBR1</i>	carbonyl reductase 1	873	<i>miR-320c</i> , <i>miR-320b</i> , <i>miR-320d</i> , <i>miR-664-3p</i>	SNP_A-8339849

ZNRF1	zinc and ring finger 1, E3 ubiquitin protein ligase	84937	<i>miR-661</i> , <i>miR-499a-3p</i> , <i>miR-513c-5p</i> , <i>miR-940</i>	SNP_A-1817379, SNP_A-1817382
CLIC5	chloride intracellular channel 5	53405	<i>miR-607</i>	SNP_A-1878747
TMEM92	transmembrane protein 92	162461	<i>miR-555</i>	SNP_A-8394990, SNP_A-8567918
EIF4E2	eukaryotic translation initiation factor 4E family member 2	9470	<i>miR-198</i> , <i>miR-767-5p</i> , <i>miR-520b</i>	SNP_A-2049507
MTM1	myotubularin 1	4534	<i>mir-1284</i> , <i>mir-802</i> , <i>mir-217</i> , <i>mir-216b</i>	chrX:149809786, chrX:149826497, chrX:149839952
HMG3	high mobility group box 3	3149	<i>mir-1179</i> , <i>mir-375</i>	chrX:150156360
TPO	thyroid peroxidase	7173	<i>mir-935</i>	chr2:1497599, chr2:1497799, chr2:1520713
SRCAP	Snf2-related CREBBP activator protein	10847	<i>mir-891b</i>	chr16:30723229, chr16:30745907, chr16:30749384, chr16:30749733, chr16:30750367
ZNF267	zinc finger protein 267	10308	<i>mir-636</i>	chr16:31925867, chr16:31926313, chr16:31926727, chr16:31927189, chr16:31927534, chr16:31927755

[0138] 상기 시나리오 1의 경우는 GEO데이터베이스에서 얻은 Affy SNP데이터 (GSE32688)를 이용하였고, 상기 시나리오 2와 3의 경우는 TCGA데이터 베이스의 췌장암 환자들로부터 얻은 somatic mutation 데이터를 이용하였다(broad.mit.edu_Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf). Affy SNP 데이터의 경우는 상응하는 reference SNP ID가 존재하지만 TCGA의 somatic mutation은 SNV로서 해당 염색체와 염색체 상 위치로 표기하였다.

[0140] V. 췌장암 바이오마커에 대한 검증(Validation)

[0141] 본 발명에서는 췌장암 환자로부터 샘플을 채취하여 상기 알고리즘에 의해 발견된 새로운 유전자들의 단백질 발현양상을 검증하였다. 구체적인 실험방법은 다음과 같다.

[0142] 인체 췌장암 및 정상 췌장 세포주 배양

[0143] 인체 췌장암 세포주 8주(Hpac, Capan1, Capan2, Cfpac1, Aspc1, Bxpc3, Miapaca2, Panc1)는 모두 ATCC (American Type Culture Collection; Manassas, VA)로부터 구입하였으며, ATCC에 의해 제시된 프로토콜에 의해 세포주들을 성장 배지에서 배양하였다. 즉, AsPC-1 및 BxPC-3 세포는 10% 우태아혈청(FBS; Hyclone, Logan, UT)을 포함하는 RPMI1640(Invitrogen Gibco, Grand Island, NY)에서 배양하였고, Capan-1 및 CFPAC-1 세포는 10% FBS를 포함하는 IMDM(Invitrogen Gibco)에서 배양하였다. Capan-2 세포는 10% FBS를 포함한 맥코이 5a(Invitrogen Gibco)에서, 그리고 MIA PaCa-2 세포는 10% FBS 및 2.5% 마혈청(Hyclone)을 포함하는 DMEM(Invitrogen Gibco)에서 배양하였다. Hpac 세포는 10% FBS를 포함하는 DMEM/헴 F12(D/F12; Invitrogen Gibco)에서, 그리고 PANC-1 세포는 10% FBS를 포함하는 DMEM에서 배양하였다. 정상 췌장 세포주인 인간 췌장관상피(HPDE) 세포는 Dr. Ming Sound Tsao(Princess Margaret Cancer Centre)로부터 제공받아 0.2 ng/ml EGF 및 30 µg/ml BPE(bovine pituitary extract, Invitrogen Gibco)를 포함하는 케라티노사이트 무혈청(KSF) 배지에서 배양하였다.

[0145] 웨스턴 블롯(Western blot)을 통한 췌장암 줄기세포 관련 분비 단백질의 발현 분석

[0146] 단백질들의 발현 동정을 위해 세포를 긁어서 수집하여 동일 정량으로 25 µg 씩 10% SDS-폴리아크릴아미드 겔에 로딩(loading)한 후 전기영동하였다. 전기영동에 의해 크기별로 분획된 단백질들을 PVDF 멤브레인(Millipore corporation, Billerica, MA, USA)으로 이동시킨 후, 비특이적 반응을 감소시키기 위하여 5% 논-팻 밀크(non-

fat milk)를 첨가한 TBS-T(Tris-buffered saline/0.05% Tween-20)에 1시간 동안 블록킹하였다. 그 다음, 각각의 멤브레인을 프로테오믹스 결과 동정된 분비 단백질들에 해당하는 1차 항체와 4에서 하룻밤 반응시켰다. 1차 항체는 래빗 폴리클로날 항체 (santa cruz biotechnology Inc, santa cruz, CA) 을 사용하였다.

[0147] 1차 항체와 반응한 각각의 멤브레인을 TBS-T 버퍼에 세척시킨 후, HRP(horseradish peroxidase)-콘쥬게이트드 2차 항체(santa cruz biotechnology Inc, santa cruz, CA)와 1시간 동안 반응시켰다. 각각의 단백질을 Enhanced chemiluminescence system(PIERCE, Rockford, IL)을 이용하여 검출하였다.

[0148] 상기와 같이 웨스턴 블롯을 통해 분석된 단백질의 발현량은 밴드의 가시적 강도(intensity)를 정량화한 다음, t-test를 수행하여 대조군과 실험군간의 유의적 차이가 있는지의 여부를 검증하여 하기 표 2에 기재하였다.

[0150] **면역조직 화학염색을 통한 인체 췌장암 조직에서의 SNP에 의해 췌장암에서 특이적으로 증가하는 단백질들의 발현 분석**

[0151] 췌장 조직 슬라이드를 자일렌 내에서 탈파라핀화시키고 등급별 알코올 내에서 재수화시켰다. 내생성 퍼옥시다제 활성을 상온에서 20분 동안 0.3%의 과산화수소 포함 메탄올 용액으로 블록킹하였다. 마이크로웨이브 항원 검사를 시트레이트 완충액(0.01M, pH 6.0)에서 4분 동안 실시하였다. 그 다음, 슬라이드를 10% 노말 당나귀 혈청 용액으로 1시간 동안 배양시켜서 비-특이적인 배경 염색을 감소시켰다. 1: 200으로 희석하여 반응시킨 1차 항체는 래빗 폴리클로날 VPS37C, ALS2CL, ZDHHC20, SNX27, CBR1, CLIC5, EIF4E2, TPO, SRCAP (santa cruz biotechnology Inc, santa cruz, CA) 이다. 블록화된 섹션을 1차 항체와 4에서 하룻밤 배양시켰다. 그 이후의 반응은 Envision 키트(DakoCytomation, Carpinteria, CA, USA)에 포함된 권장 과정을 이용하여 실시하였다. 마지막으로, 슬라이드를 3,3'-디아미노벤지딘(3,3'-diaminobenzidine; DakoCytomation, Carpinteria, CA, USA)로 배양시키고, 변형된 해리스 헤마톡실린(Harris hematoxylin, Sigma-Aldrich, Inc., St. Louis, MO)용액으로 대조 염색하였다.

[0153] **췌장암 관련 유전자의 단백질 발현 검증 결과**

표 2

유전자 ID	웨스턴블롯 (정상vs 췌장암)	면역조직 화학염색
VPS37C	췌장암에서 유의적으로 발현증가	정상조직에서는 VPS37C가 정상췌관 및 췌도세포에서 발현하는데 반해 췌장암 조직에서는 암세포에서 확연한 발현을 나타냄
ALS2CL	췌장암에서 유의적으로 발현증가	정상조직에서는 ALS2CL이 췌도세포에서 발현하는데 반해 췌장암 조직에서는 암세포에서 확연한 발현을 나타냄
ZDHHC20	N/D	정상조직과 암조직에서 불특정하게 나타나지만 암세포에서 보다 많은 발현을 나타내고 있음
SNX27	췌장암에서 유의적으로 발현증가	정상조직과 암조직에서 불특정하게 나타나지만 암세포에서 보다 많은 발현을 나타내고 있음
CBR1	N/D	CBR1은 암세포에서만 발현을 나타냄
ZNRF1	N/D	N/D
CLIC5	췌장암에서 유의적으로 발현증가	정상조직에서는 CLIC5이 췌도세포에서 발현하는데 반해 췌장암 조직에서는 암세포에서 확연한 발현을 나타냄
TMEM92	N/D	N/D
EIF4E2	췌장암에서 유의적으로 발현증가	정상조직에서는 EIF4E2이 췌도세포에서 발현하는데 반해 췌장암 조직에서는 암세포에서 확연한 발현을 나타냄
MTM1	N/D	N/D
HMGB3	N/D	N/D
TPO	N/D	TPO가 암세포에서만 발현을 나타냄
SRCAP	N/D	정상조직 대비 SRCAP이 암세포에서 발현을 나타냄
ZNF267	N/D	N/D

[0155] 상기 표 2에서, HMGB3는 다른 여러 암에 관여할 가능성이 보고되고 있고 (<http://www.ncbi.nlm.nih.gov/gene/3149>), ZNF267도 종양세포의 증식과 이동(tumor cell proliferation과

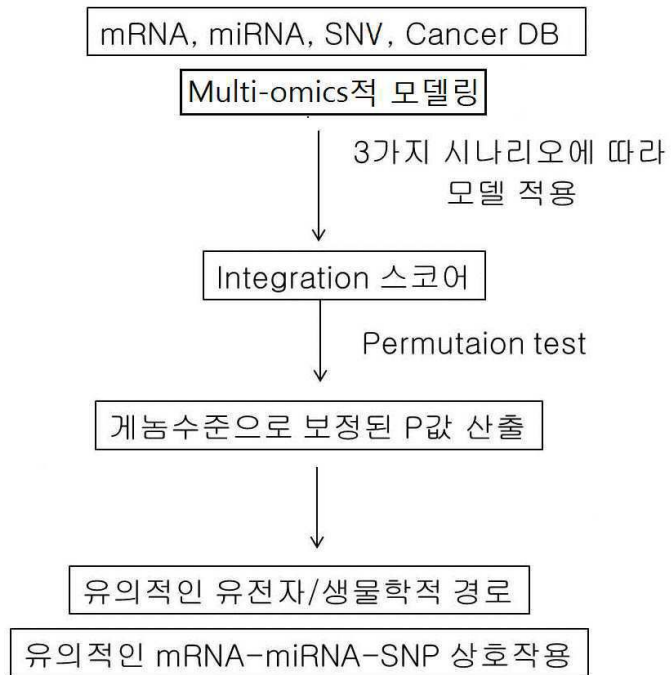
migration)에 대한 영향이 보고되어 있다(<http://www.ncbi.nlm.nih.gov/pubmed/21840307/>).

[0156]

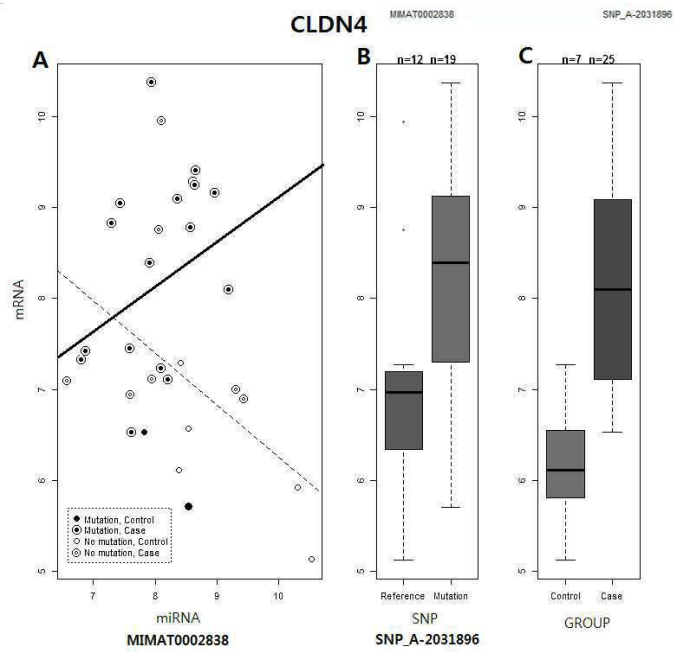
이상으로 본 발명 내용의 특정한 부분을 상세히 기술한 바, 당업계의 통상의 지식을 가진 자에게 있어서, 이러한 구체적 기술은 단지 바람직한 실시 양태일 뿐이며, 이에 의해 본 발명의 범위가 제한되는 것이 아닌 점은 명백할 것이다. 따라서 본 발명의 실질적인 범위는 첨부된 청구항들과 그것들의 등가물에 의하여 정의된다고 할 것이다.

도면

도면1



도면2



도면3

