

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(10) 국제공개번호

(43) 국제공개일
2018년 4월 26일 (26.04.2018) WIPO | PCT

WO 2018/074865 A2

(51) 국제특허분류:
C12Q 1/68 (2018.01)

(21) 국제출원번호: PCT/KR2017/011606

(22) 국제출원일: 2017년 10월 19일 (19.10.2017)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2016-0137278 2016년 10월 21일 (21.10.2016) KR

(71) 출원인: 서울대학교병원 (SEOUL NATIONAL UNIVERSITY HOSPITAL) [KR/KR]; 03080 서울시 중로구 대학로 101, Seoul (KR). 서울대학교산학협력단 (SEOUL NATIONAL UNIVERSITY R&DB FOUNDATION) [KR/KR]; 08826 서울시 관악구 관악로 1, Seoul (KR). 주식회사 셀레믹스 (CELEMICS, INC.) [KR/KR]; 08506 서울시 금천구 가산디지털1로 131, 에이동 19층 1902, 1903, 1904, 1905-나호, Seoul (KR).

(72) 발명자: 한원식 (HAN, Won Shik); 05572 서울시 송파구 올림픽로4길 15, 3동 501호, Seoul (KR). 이한별 (LEE, Han Byoel); 06091 서울시 강남구 삼성로 649, 2동 1101호, Seoul (KR). 윤성로 (YOON, Sung Roh); 08826 서울시 관악구 관악로 1 서울대학교 공과대학 301동 908호, Seoul (KR). 김선 (KIM, Sun); 08826 서울시 관악구 관악로 1 서울대학교 공과대학 301동 316호, Seoul (KR). 권선영 (KWON, Sun Young); 08826 서울시 관악구 관악로 1 서울대학교 공과대학 301동 908호, Seoul (KR). 김민수 (KIM, Min Su); 08826 서울시 관악구 관악로 1 서울대학교 공과대학 301동 316호, Seoul (KR). 김효기 (KIM, Hyo Ki); 08506 서울시 금천구 가산디지털1로 131, BYC하이시티 19층, Seoul (KR). 한효준 (HAN, Hyo Jun); 08506 서울시 금천구 가산디지털1로 131, BYC하이시티 19층, Seoul (KR).

(74) 대리인: 콰현규 (KWAK, Hyun Kyu); 06236 서울시 강남구 테헤란로 22길 15 A&C빌딩 4층, Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

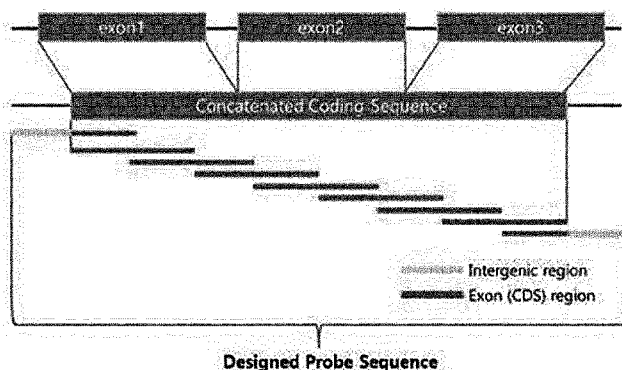
(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

- 국제조사보고서 없이 공개하며 보고서 접수 후 이를 별도 공개함 (규칙 48.2(g))
- 명세서의 서열목록 부분과 함께 (규칙 5.2(a))

(54) Title: COMPOSITION AND METHOD FOR BREAST CANCER PROGNOSIS PREDICTION

(54) 발명의 명칭: 유방암 예후 예측용 조성물 및 방법



(57) Abstract: The present invention relates to a composition for breast cancer prognosis prediction comprising a formulation for measuring the expression level of genes for breast cancer prognosis prediction, a kit for breast cancer prognosis prediction comprising the composition, and a method for providing information for the prediction of breast cancer prognosis by measuring the expression level of the genes.

(57) 요약서: 유방암 예후 예측을 위한 유전자의 발현 수준을 측정하기 위한 제제를 포함하는, 유방암의 예후 예측용 조성물, 상기 조성물을 포함하는 유방암의 예후 예측용 키트, 및 상기 유전자의 발현 수준을 측정하여 유방암의 예후 예측을 위한 정보를 제공하는 방법에 관한 것이다.



WO 2018/074865 A2

명세서

발명의 명칭: 유방암 예후 예측용 조성물 및 방법

기술분야

- [1] 본 발명은 유방암 예후 예측을 위한 유전자의 발현 수준을 측정하기 위한 제제를 포함하는, 유방암의 예후 예측용 조성물, 상기 조성물을 포함하는 유방암의 예후 예측용 키트, 및 상기 유전자의 발현 수준을 측정하여 유방암의 예후를 예측하기 위한 정보를 제공하는 방법에 관한 것이다.

배경기술

- [2] 유방암은 병리학적 관찰로 대략적인 예후를 예측할 수 있으나, 관찰결과에 대한 표준화와 정량화가 어렵고 예후예측에 대한 신뢰성이 낮아, 실제 임상에서는 대부분의 초기 유방암 환자에게 항암치료를 권하고 있다. 항암치료의 특성상 환자가 겪는 고통이 매우 크고 경제적 지출이 요구된다. 따라서, 환자의 예후를 예측하여 불필요한 항암치료를 줄인다면, 환자의 삶의 질에 큰 도움이 될 것이다.
- [3] 현재 유전자 분석을 이용하여 구성된 유방암 예후예측 도구 중 표준으로 사용되고 있는 Oncotype DX® 등은 실시간 PCR 기반의 검사로서 비용과 효율성의 측면에서 많은 유전자를 동시에 분석하기 어렵다. 단일 마커에 의한 예측이나 임상병리학적 요인에 의한 분석에 비해 다유전자 마커의 분석에 의한 유방암 예후 예측이 우월하다. 차세대 서열분석법(NGS)은 이전의 기술에 비해 효율성이 월등한 대용량 서열분석기술로, 급격한 검사가격의 하락으로 현재 모든 유전체 연구의 표준 플랫폼으로 빠르게 자리잡고 있다. 임상에 적용할 경우 치료를 결정하기 위한 정보를 기존 PCR 방식보다 빠르고 민감하고 재현성 높게 얻을 수 있어 유용성이 높다.
- [4] 이에 본 발명에서는 에스트로겐 수용체 양성 유방암의 발현량을 이용하여 예후와 관련된 유전자 세트를 구성하고, NGS 기반으로 분석하고자 하였다.

발명의 상세한 설명

기술적 과제

- [5] 일 양상은 유방암의 예후 예측용 조성물을 제공한다.
- [6] 다른 양상은 유방암의 예후 예측용 키트를 제공한다.
- [7] 또 다른 양상은 유방암의 예후 예측을 위한 정보를 제공하는 방법을 제공한다.

과제 해결 수단

- [8] 일 양상은 하기 표 1에 기재된 유전자의 mRNA 발현 수준을 측정하기 위한 제제를 포함하는, 유방암의 예후 예측용 조성물을 제공한다.

[9] [표1]

번호	유전자명	번호	유전자명	번호	유전자명	번호	유전자명	번호	유전자명
1	ACTB	31	CDKN3	61	FANCI	91	LRRC59	121	RAI2
2	APOBEC3B	32	CENPA	62	FBXO5	92	MAD2L1	122	RFC4
3	ASF1B	33	CENPE	63	FEN1	93	MARCH8	123	RPLP0
4	ASPM	34	CENPF	64	FOXN1	94	MCM10	124	RRM2
5	AURKA	35	CENPM	65	GAPDH	95	MCM2	125	SCUBE2
6	AURKB	36	CENPN	66	GIN51	96	MCM6	126	SETBP1
7	BAG1	37	CEP55	67	GRB7	97	MELK	127	SF3B3
8	BCL2	38	CHEK1	68	GSTM1	98	MKI67	128	SHCBP1
9	BIRC5	39	CIRBP	69	GTSE1	99	MLF1IP	129	SHMT2
10	BLM	40	CKS2	70	GUSB	100	MMP11	130	SLC25A12
11	BUB1	41	CRIM1	71	HJURP	101	MYBL2	131	SLC7A5
12	BUB1B	42	CTSL2	72	HMMR	102	NCAPG	132	SPAG5
13	C14orf45	43	CX3CR1	73	HN1	103	NCAPG2	133	SPC25
14	C16orf61	44	CYBRD1	74	IFT46	104	NCAPH	134	SQLE
15	C7orf63	45	DBF4	75	KIF11	105	NDC80	135	STARD13
16	CACNA1D	46	DDX39	76	KIF13B	106	NEK2	136	STIL
17	CCNA2	47	DLGAP5	77	KIF14	107	NUP93	137	STMN1
18	CCNB1	48	DNMT3B	78	KIF15	108	NUSAP1	138	SYNC
19	CCNB2	49	DONSON	79	KIF18A	109	OIP5	139	TACC3
20	CCNE1	50	DTL	80	KIF18B	110	PBK	140	TFRC
21	CCNE2	51	E2F1	81	KIF20A	111	PDSS1	141	TK1
22	CCT5	52	E2F8	82	KIF23	112	PGR	142	TOP2A
23	CD68	53	ECHDC2	83	KIF2C	113	PKMYT1	143	TPX2
24	CDC20	54	ERBB2	84	KIF4A	114	PLK1	144	TRIP13
25	CDC25A	55	ERCC6L	85	KIFC1	115	PLK4	145	TROAP
26	CDC45	56	ESPL1	86	KPNA2	116	PRC1	146	TTK
27	CDC6	57	ESR1	87	LMNB1	117	PTTG1	147	UBE2C
28	CDCA3	58	EXO1	88	LMNB2	118	RACGAP1	148	UBE2S
29	CDCA8	59	EZH2	89	LRIG1	119	RAD51	149	ZWINT
30	CDK1	60	FAM64A	90	LRRC48	120	RAD51AP1		

[10] 상기 유전자 symbol에 해당하는 명칭 및 accession No.에 대해서는 공개된 문헌 또는 데이터베이스를 통해 통상의 기술자에게 알려진 정보를 참조한다. 상기 공개 데이터베이스로서 HUGO Gene Nomenclature Committee(HGNC)를 참조할 수 있다. 상기 149개의 유전자 중 일부를 예로서 설명하면 다음과 같다: ACTB에 대한 명칭은 actin beta이고 뉴클레오티드 서열에 대한 accession No.는 NM_001101이다. APOBEC3B에 대한 명칭은 Homo sapiens apolipoprotein B mRNA editing enzyme catalytic subunit 3B이고 뉴클레오티드 서열에 대한 accession No.는 NM_004900이다. ASF1B에 대한 명칭은 anti-silencing function 1B histone chaperone이고 뉴클레오티드 서열에 대한 accession No.는

NM_018154이다.

- [11] 본 명세서에서 용어 "mRNA 발현 수준 측정"은 유방암의 예후를 예측하기 위하여 생물학적 시료에서 해당 유전자의 mRNA 존재 여부 및 발현 정도를 확인하는 과정으로, mRNA의 양을 측정함으로써 알 수 있다. 상기 발현 수준은 RT-PCR, 경쟁적 RT-PCR(Competitive RT-PCR), 실시간 RT-PCR(Real-time RT-PCR), RNase 보호 분석법(RNase protection assay), 노던 블롯팅(Northern blotting), DNA 마이크로어레이, 또는 초병렬 시퀀싱에 의해 측정되는 것일 수 있다. 상기 조성물은, 초병렬 시퀀싱을 이용하여 상기 유전자의 mRNA의 존재 여부 및 발현 정도를 확인하기 위한 분석에서 사용되는 것일 수 있다. 상기 조성물은 예를 들면, RNA-시퀀싱 분석 또는 표적화(targeted) RNA-시퀀싱 분석에서 사용되는 것일 수 있다.
- [12] 용어 "초병렬 시퀀싱"은, 차세대 염기서열 분석법(Next Generation Sequencing: NGS) 또는 고용량 시퀀싱(high-throughput sequencing)으로도 명명될 수 있다. 상기 초병렬 시퀀싱은 현재 차세대 염기서열 분석법으로 알려진 방법 및 장래 개발될 수 있는 방법을 포함한다. 상기 초병렬 시퀀싱은 합성에 의한 시퀀싱(sequencing by synthesis), 이온 토렌트(Ion-Torrent) 시퀀싱, 파이로시퀀싱(pyrosequencing), 라이게이션에 의한 시퀀싱, 나노포어 시퀀싱, 및 단일-분자 실시간 시퀀싱으로 이루어진 군으로부터 선택되는 것일 수 있다.
- [13] 상기 발현 수준을 측정하기 위한 제제는 상기 유전자의 mRNA 또는 그의 cDNA에 특이적으로 혼성화하는 프라이머 쌍 또는 프로브를 포함할 수 있다. 상기 유전자의 염기서열은 당해 분야에 알려진 상태이므로, 통상의 기술자는 상기 염기서열을 바탕으로 각 유전자의 특정 영역을 특이적으로 증폭할 수 있는 프라이머 쌍 또는 특정 영역과 특이적으로 혼성화할 수 있는 프로브를 디자인할 수 있다. 상기 프로브는 단일가닥 DNA 프로브 또는 RNA 프로브일 수 있다.
- [14] 상기 제제는 상기 유전자의 mRNA 또는 그의 cDNA에 특이적으로 혼성화하는 올리고뉴클레오타이드의 세트일 수 있다. 상기 올리고뉴클레오타이드의 세트는 상기 각 유전자에 대해서 그의 mRNA 또는 cDNA에 특이적으로 혼성화하는 하나 이상의 올리고뉴클레오타이드를 포함할 수 있다. 상기 제제는 상기 각 유전자의 mRNA 또는 cDNA에 특이적으로 혼성화하는 10종 이상, 20종 이상, 30종 이상, 50종 이상, 70종 이상, 90종 이상, 100종 이상, 또는 150종 이상의 상이한 올리고뉴클레오타이드 세트를 포함할 수 있다. 예를 들면, 상기 제제는 ACTB 유전자에 대해서 그의 mRNA 또는 cDNA에 특이적으로 혼성화하는 30개 내지 40개의 올리고뉴클레오타이드를 포함할 수 있다.
- [15] 상기 제제는 DNA 올리고뉴클레오타이드의 세트일 수 있다. 상기 DNA 올리고뉴클레오타이드는 화학적으로 합성된 것일 수 있다. 상기 DNA 올리고뉴클레오타이드는 전사됨으로써 상기 각 유전자의 mRNA 또는 그의 cDNA에 특이적으로 혼성화하는 RNA 올리고뉴클레오타이드를 생성하는 것일 수 있다. 상기 전사는 인 비트로 전사일 수 있다.

- [16] 상기 제제는, 상기 각 유전자에 대하여 그의 mRNA 또는 cDNA에 특이적으로 혼성화하는 하나 이상의 올리고뉴클레오티드로 전사되는 하나 이상의 DNA 올리고뉴클레오티드를 포함할 수 있다. 상기 제제는 상기 149개의 유전자 각각에 대해 그와 같은 하나 이상의 DNA 올리고뉴클레오티드를 포함할 수 있다. 상기 제제는 서열번호 1 내지 9241의 뉴클레오티드 서열을 갖는 올리고뉴클레오티드의 세트일 수 있다. 예를 들면, 서열번호 1 내지 서열번호 316의 뉴클레오티드 서열을 갖는 올리고뉴클레오티드의 세트는 전사됨으로써 ACTB, APOBEC3B, ASF1B, ASPM, 및 AURKA 유전자의 mRNA 또는 그의 cDNA에 특이적으로 혼성화하는 올리고뉴클레오티드의 세트를 생성할 수 있다.
- [17] 상기 유방암의 예후 예측용 조성물에서, 상기 유방암은 에스트로겐 수용체 양성(estrogen receptor-positive, ER+)을 나타내는 것일 수 있다.
- [18]
- [19] 다른 양상은, 상기 조성물을 포함하는, 유방암의 예후 예측용 키트를 제공한다. 상기 키트는 상기 유전자의 mRNA 발현 수준을 측정하기 위한 프라이머 쌍, 프로브, 또는 올리고뉴클레오티드 외에도, mRNA 발현 수준의 측정방법에 적합한 조성물, 용액, 또는 장치를 포함할 수 있다.
- [20] 예를 들면, 상기 키트는 RT-PCR을 수행하는데 요구되는 구성요소를 포함하는 키트일 수 있다. RT-PCR 키트는 상기 프라이머 쌍 외에도 반응 완충액, dNTPs, Taq-중합효소 및 역전사효소와 같은 효소, DNase, RNase 억제제, DEPC-물 등을 포함할 수 있다. 상기 키트는 상기 프로브 또는 올리고뉴클레오티드가 기판 상에 고정화된 마이크로어레이 형태일 수 있다.
- [21] 또한, 상기 키트의 프로브 또는 올리고뉴클레오티드는 표적화 RNA-시퀀싱을 위해 제조된 라이브러리와 혼성화될 수 있다. 표적화 RNA-시퀀싱을 위해, 상기 프로브 또는 올리고뉴클레오티드는 비오틴과 같은 물질과 결합될 수 있다. 이 경우, 스트렙타비딘 자성 비드에 의해 상기 프로브 또는 올리고뉴클레오티드와 혼성화된 라이브러리의 일부가 분리될 수 있다.
- [22]
- [23] 다른 양상은, 상기 DNA 올리고뉴클레오티드의 세트 또는 서열번호 1 내지 9241의 뉴클레오티드 서열을 갖는 올리고뉴클레오티드의 세트를 인 비트로 전사하여 RNA 올리고뉴클레오티드의 세트를 생성하는 단계를 포함하는, 유방암의 예후 예측을 위한 프로브를 제조하는 방법을 제공한다.
- [24] 상기 생성된 RNA 올리고뉴클레오티드의 세트는 RNA-시퀀싱, 예를 들면, 표적화 RNA-시퀀싱을 위해 시료 중의 mRNA 또는 cDNA와 인큐베이션될 수 있다.
- [25]
- [26] 또 다른 양상은, 유방암의 예후를 예측하고자 하는 개체에서 상기 표 1의 유전자의 mRNA 발현 수준을 측정하는 단계를 포함하는, 유방암의 예후 예측을 위한 정보를 제공하는 방법을 제공한다.

- [27] 상기 방법은, 상기 유전자의 mRNA 발현 수준을 측정하기 위하여, 상기 개체로부터 분리된 mRNA 또는 그의 cDNA와 상기 유전자의 mRNA 발현 수준을 측정하기 위한 제제를 인큐베이션하는 것을 포함할 수 있다. 상기 제제에 대해서는 전술된 바와 같다.
- [28] 상기 인큐베이션에서, 상기 제제는 상기 유전자의 mRNA 또는 그의 cDNA에 특이적으로 결합하는 올리고뉴클레오타이드일 수 있다. 상기 제제는 상기 각 유전자의 mRNA 또는 cDNA에 특이적으로 혼성화하는 10종 이상, 20종 이상, 30종 이상, 50종 이상, 70종 이상, 90종 이상, 100종 이상, 또는 150종 이상의 상이한 올리고뉴클레오타이드 세트를 포함할 수 있다. 상기 인큐베이션에 의해 상기 유전자의 mRNA 또는 그의 cDNA와 상기 올리고뉴클레오타이드의 혼성 복합체(hybridized complex)가 형성될 수 있다. 상기 혼성 복합체 중 상기 유전자의 mRNA 또는 그의 cDNA는 포획된(captured) 표적 전사체로 명명될 수 있다.
- [29] 상기 제제는 DNA 올리고뉴클레오타이드를 인 비트로 전사하여 수득되는 것일 수 있다. 상기 DNA 올리고뉴클레오타이드는 화학적으로 합성된 것일 수 있다. 상기 제제는, 상기 각 유전자에 대하여 그의 mRNA 또는 cDNA에 특이적으로 혼성화하는 하나 이상의 올리고뉴클레오타이드로 전사되는 하나 이상의 DNA 올리고뉴클레오타이드를 포함할 수 있다. 상기 제제는 상기 149개의 유전자 각각에 대해 그와 같은 하나 이상의 DNA 올리고뉴클레오타이드를 포함할 수 있다. 상기 DNA 올리고뉴클레오타이드는 서열번호 1 내지 9241의 뉴클레오타이드 서열을 갖는 올리고뉴클레오타이드의 세트일 수 있다.
- [30] 상기 방법은, 상기 혼성 복합체를 분리하는 단계를 더 포함할 수 있다. 상기 혼성 복합체의 분리는 자성 입자에 의해 이루어지는 것일 수 있다. 상기 자성 입자에 의한 분리를 위하여, 상기 올리고뉴클레오타이드는 비오틴과 같은 물질과 결합된 것일 수 있다. 비오틴과 결합된 상기 올리고뉴클레오타이드는 그에 특이적인 mRNA 또는 cDNA와 혼성 복합체를 형성하고, 상기 혼성 복합체가 스트렙타비딘이 코팅된 자성 입자에 의해 분리될 수 있다. 또한, 인큐베이션 반응물 중 상기 올리고뉴클레오타이드와 혼성 복합체를 형성하지 않은 비표적(nontargeted) 전사체를 워싱함으로써(wash away) 상기 혼성 복합체를 분리할 수 있다.
- [31] 상기 발현 수준의 측정은 RT-PCR, 경쟁적 RT-PCR, 실시간 RT-PCR, RNase 보호 분석법, 노던 블롯팅, DNA 마이크로어레이, 또는 초병렬 시퀀싱에 의해 수행될 수 있다. 상기 방법은 다수 유전자를 분석 대상으로 하므로, 초병렬 시퀀싱에 의한 발현 수준의 측정이 효율적일 수 있다. 상기 초병렬 시퀀싱은 합성에 의한 시퀀싱, 이온 토렌트 시퀀싱, 파이로시퀀싱, 라이게이션에 의한 시퀀싱, 나노포어 시퀀싱, 및 단일-분자 실시간 시퀀싱으로 이루어진 군으로부터 선택될 수 있다. 상기 발현 수준의 측정은 예를 들면, 표적화 RNA-시퀀싱 분석을 통해 수행될 수 있다.

- [32] 상기 유방암은 에스트로겐 수용체 양성을 나타내는 것일 수 있다.

발명의 효과

- [33] 일 양상에 따른 유방암의 예후 예측용 조성물 또는 키트, 또는 유방암의 예후 예측을 위한 정보제공방법은, 정확한 예측을 위해 구성된 유전자 세트의 발현 정보에 기초하므로 신뢰성 높은 결과를 제공한다. 상기 조성물 또는 키트, 또는 정보제공방법은 다수 유전자에 대한 발현 데이터를 정확하고 효과적으로 생성할 수 있도록 최적화된 프로브를 이용하고, 초병렬 시퀀싱을 통해 분석될 수 있어 신속하고 민감하며 재현성 높은 예후 진단을 가능하게 한다. 또한, 유방암의 재발 가능성에 따라 적절한 치료법을 선택할 수 있게 하고 환자의 생존율을 높이는데 기여할 수 있다.

도면의 간단한 설명

- [34] 도 1은 GSE2034 마이크로어레이 데이터에 대해 카플란-마이어 생존 분석 곡선을 비교한 결과를 나타낸다.
- [35] 도 2는 TCGA 데이터에 대해 카플란-마이어 생존 분석 곡선을 비교한 결과를 나타낸다.
- [36] 도 3은 프로브 디자인의 모식도를 나타낸다.

발명의 실시를 위한 형태

- [37] 이하, 본 발명을 하기 실시예에 의해 더욱 구체적으로 설명한다. 그러나, 이들 실시예는 본 발명에 대한 이해를 돕기 위한 것일 뿐, 어떤 의미로든 본 발명의 범위가 이들에 의해 제한되는 것은 아니다.

[38]

- [39] 실시예 1: 유방암 예후예측 유전자 선정

- [40] 1.1. 후보 유전자의 선정

- [41] Oncotype DX®는 21종의 유전자로 재발 점수(Recurrent Score: RS)를 계산하여 0 내지 100의 값으로 유방암 환자의 예후 예측을 점수화한다. Oncotype DX의 RS 값을 공개 데이터 세트에 적용하여, RS와 개별 유전자간의 상관관계가 높은 새로운 후보 유전자를 선정하고자 하였다.

- [42] 사용된 공개 데이터는 ER+, Lymph node- 조건을 만족하는 Affymetrix platform의 GSE7390, GSE2034, GSE2990, GSE4922, GSE12093, GSE3494, 및 GSE6532의 7개 데이터였다. 7개 데이터의 각 샘플마다 논문에 기재된 하기 식의 재발-점수 알고리즘을 이용하여 RS 값을 산출하였다(Paik S et al. "A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer." New England Journal of Medicine 351.27 (2004): 2817-2826).

- [43]
$$RS = +0.47 \times \text{HER2 group score} - 0.34 \times \text{ER group score} + 1.04 \times \text{proliferation group score} + 0.10 \times \text{invasion group score} + 0.05 \times \text{CD68} - 0.08 \times \text{GSTM1} - 0.07 \times \text{BAG1}$$

- [44] 상기 알고리즘은 reference 유전자 5종을 제외한 16종의 유전자를 이용하고 각

그룹 점수의 산출식과 범위 한계값은 상기 논문의 도 1에 기재된 바에 따른다. 최종적으로 RS 값은 0 내지 100으로 재조정되었다.

- [45] RS와 개별 유전자 간 상관관계를 구하기 위하여, 각 샘플의 RS 값과 유전자의 발현값 간 평균 상관관계를 Pearson's correlation과 Spearman's rank correlation의 평균값을 이용하여 구하였다. 7개 데이터에 대한 평균 상관관계수의 절대값이 0.5 이상인 경우 RS와 개별 유전자 간 상관관계가 있는 것으로 판단하였다. 그 결과, RS와 상관관계가 있는 135개의 유전자를 선정하고 이들 유전자에 상기 알고리즘에 포함된 16개의 유전자를 더하여, 상기 표 1에 기재된 총 149개의 유전자로 이루어진 후보 유전자군을 얻었다.

[46]

[47] 1.2. 후보 유전자 검증

- [48] SurvExpress에서 제공되는 유방암 데이터 중 마이크로어레이 데이터 Wang Foekens Breast GSE2034 (Affymetrix) 및 Breast Invasive Carcinoma TCGA (Agilent) 데이터 세트에 대해, Oncotype DX 재발-점수 알고리즘에 포함된 16개 유전자(reference 유전자 제외) 및 1.1.에서 RS와 상관관계가 있는 것으로 분석된 135개의 후보 유전자의 카플란-마이어 플롯을 그려 비교하였다. SurvExpress는 생존 분석을 이용하여 암 유전자 발현 데이터에 대하여 바이오마커의 유효성을 온라인 상에서 검증하는 도구이자 데이터베이스이다(Raul Aguirre-Gamboa et al., PLoS One. 2013; 8(9): e74250).

- [49] 도 1은 GSE2034 마이크로어레이 데이터에 대해 카플란-마이어 생존 분석 곡선을 비교한 결과를 나타낸다. 도 2는 TCGA 데이터에 대해 카플란-마이어 생존 분석 곡선을 비교한 결과를 나타낸다. 각 그래프에서 가로축은 시간을 나타내고 세로축은 생존 확률을 나타낸다.

- [50] 도 1 및 2에 나타낸 바와 같이, 기존 Oncotype DX 알고리즘의 16종의 유전자에 비해 새로 선정된 유전자들로 그려진 곡선에서 보다 뚜렷한 차이가 나타났다. 이를 통해, 새로 선정된 유전자들을 이용하여 유방암의 예후를 예측하는 것이 더욱 효과적이라는 것을 확인할 수 있었다.

[51]

[52] 실시예 2: 선정된 유전자에 대한 프로브 제작

[53] 2.1. 엑손 서열의 재조합

- [54] 전술된 실시예 1.에서 선정된 유전자에 대해 RefSeq 데이터베이스를 참조하여 각 유전자의 엑손 영역 정보를 확인하였다. 이와 함께, 각 유전자의 전사 관련 인자에 의한 영향을 확인할 수 있도록 프로모터, 리보솜 결합 부위(RBS), 및 전사인자 결합 부위 등을 포함한 intergenic 영역에 대한 정보도 확인하였다.

- [55] 각 유전자에 대하여 엑손 서열을 이어 붙여 하나로 연결된 암호화 서열(concatenated coding sequence)을 제작하였다. 이때, 선택적 스플라이싱(alternative splicing)의 가능성을 배제하고 각 유전자의 모든 엑손 영역이 연결되어 하나의 전사체(transcript)를 형성한다고 가정하였다. 또한,

유전자의 전사 수준에 영향을 미칠 수 있는 전사인자 결합 부위의 서열도 전사체에 포함되도록 하였다.

[56]

[57] 2.2. 프로브 제작

[58] 각 유전자에 대해 엑손 서열이 제조합된 서열 중 암호화 서열의 시작지점을 기준으로 레퍼런스 게놈 상에서 upstream 방향으로 일정 부분 확장하여 intergenic 영역에서 프로브의 시작지점을 설정하였다. 이 시작지점으로부터 프로브의 일부가 다음 순서의 프로브와 겹치도록 순차적으로 프로브를 디자인하였다. 프로브 간 겹쳐지는 정도는 선행 연구에서의 해당 영역의 포획(capture) 효율에 기반하여 조절하였다. 프로브가 결합되는 부위의 서열이 높은 GC ratio를 지니거나 특정 서열이 반복적으로 나타날 경우, 결합되는 부위의 서열과 다르게 프로브 서열을 조정하여 포획 효율을 증가시켰다.

[59] 도 3은 프로브 디자인의 모식도를 나타낸다.

[60] 유방암 예후예측 유전자로 선정된 149개 유전자에 대한 프로브 제작 결과, 각 유전자에 대한 프로브 서열은 하기 표 2 내지 표 4에 기재된 바와 같다.

[61] [표2]

번호	유전자명	프로브 서열번호	번호	유전자명	프로브 서열번호
1	ACTB	1 내지 38	31	CDKN3	1695 내지 1708
2	APOBEC3B	39 내지 63	32	CENPA	1709 내지 1734
3	ASF1B	64 내지 91	33	CENPE	1735 내지 1877
4	ASPM	92 내지 274	34	CENPF	1878 내지 2048
5	AURKA	275 내지 316	35	CENPM	2049 내지 2085
6	AURKB	317 내지 336	36	CENPN	2086 내지 2173
7	BAG1	337 내지 402	37	CEP55	2174 내지 2217
8	BCL2	403 내지 516	38	CHEK1	2218 내지 2282
9	BIRC5	517 내지 567	39	CIRBP	2283 내지 2344
10	BLM	568 내지 653	40	CKS2	2345 내지 2354
11	BUB1	654 내지 723	41	CRIM1	2355 내지 2447
12	BUB1B	724 내지 785	42	CTSL2	2448 내지 2492
13	C14orf45	786 내지 846	43	CX3CR1	2493 내지 2543
14	C16orf61	847 내지 857	44	CYBRD1	2544 내지 2615
15	C7orf63	858 내지 928	45	DBF4	2616 내지 2680
16	CACNA1D	929 내지 1066	46	DDX39	2681 내지 2706
17	CCNA2	1067 내지 1112	47	DLGAP5	2707 내지 2757
18	CCNB1	1113 내지 1146	48	DNMT3B	2758 내지 2839
19	CCNB2	1147 내지 1171	49	DONSON	2840 내지 2899
20	CCNE1	1172 내지 1207	50	DTL	2900 내지 2972
21	CCNE2	1208 내지 1252	51	E2F1	2973 내지 3016
22	CCT5	1253 내지 1322	52	E2F8	3017 내지 3081
23	CD68	1323 내지 1352	53	ECHDC2	3082 내지 3142
24	CDC20	1353 내지 1382	54	ERBB2	3143 내지 3236
25	CDC25A	1383 내지 1443	55	ERCC6L	3237 내지 3308
26	CDC45	1444 내지 1477	56	ESPL1	3309 내지 3420
27	CDC6	1478 내지 1527	57	ESR1	3421 내지 3536
28	CDCA3	1528 내지 1581	58	EXO1	3537 내지 3589
29	CDCA8	1582 내지 1621	59	EZH2	3590 내지 3658
30	CDK1	1622 내지 1694	60	FAM64A	3659 내지 3689

[62] [표3]

번호	유전자명	프로브 서열번호	번호	유전자명	프로브 서열번호
61	FANCI	3690 내지 3794	91	LRRC59	5555 내지 5601
62	FBXO5	3795 내지 3835	92	MAD2L1	5602 내지 5624
63	FEN1	3836 내지 3872	93	MARCH8	5625 내지 5697
64	FOXMI	3873 내지 3932	94	MCM10	5698 내지 5773
65	GAPDH	3933 내지 3961	95	MCM2	5774 내지 5833
66	GIN51	3962 내지 4017	96	MCM6	5834 내지 5895
67	GRB7	4018 내지 4058	97	MELK	5896 내지 5937
68	GSTM1	4059 내지 4077	98	MKI67	5938 내지 6145
69	GTSE1	4078 내지 4147	99	MLF1IP	6146 내지 6187
70	GUSB	4148 내지 4187	100	MMP11	6188 내지 6231
71	HJURP	4188 내지 4237	101	MYBL2	6232 내지 6278
72	HMMR	4238 내지 4289	102	NCAPG	6279 내지 6355
73	HN1	4290 내지 4317	103	NCAPG2	6356 내지 6457
74	IFT46	4318 내지 4347	104	NCAPH	6458 내지 6540
75	KIF11	4348 내지 4431	105	NDC80	6541 내지 6576
76	KIF13B	4432 내지 4615	106	NEK2	6577 내지 6622
77	KIF14	4616 내지 4736	107	NUP93	6623 내지 6680
78	KIF15	4737 내지 4822	108	NUSAP1	6681 내지 6720
79	KIF18A	4823 내지 4884	109	OIP5	6721 내지 6740
80	KIF18B	4885 내지 4955	110	PBK	6741 내지 6775
81	KIF20A	4956 내지 5012	111	PDSS1	6776 내지 6805
82	KIF23	5013 내지 5077	112	PGR	6806 내지 7043
83	KIF2C	5078 내지 5131	113	PKMYT1	7044 내지 7089
84	KIF4A	5132 내지 5210	114	PLK1	7090 내지 7125
85	KIFC1	5211 내지 5255	115	PLK4	7126 내지 7190
86	KPNA2	5256 내지 5293	116	PRC1	7191 내지 7249
87	LMNB1	5294 내지 5343	117	PTTG1	7250 내지 7262
88	LMNB2	5344 내지 5420	118	RACGAP1	7263 내지 7319
89	LRIG1	5421 내지 5511	119	RAD51	7320 내지 7359
90	LRRC48	5512 내지 5554	120	RAD51AP1	7360 내지 7399

[63] [표4]

번호	유전자명	프로브 서열번호
121	RAI2	7400 내지 7445
122	RFC4	7446 내지 7477
123	RPLP0	7478 내지 7497
124	RRM2	7498 내지 7553
125	SCUBE2	7554 내지 7616
126	SETBP1	7617 내지 7798
127	SF3B3	7799 내지 7959
128	SHCBP1	7960 내지 8012
129	SHMT2	8013 내지 8055
130	SLC25A12	8056 내지 8123
131	SLC7A5	8124 내지 8198
132	SPAG5	8199 내지 8261
133	SPC25	8262 내지 8283
134	SQLE	8284 내지 8332
135	STARD13	8333 내지 8453
136	STIL	8454 내지 8567
137	STMN1	8568 내지 8628
138	SYNC	8629 내지 8686
139	TACC3	8687 내지 8743
140	TFRC	8744 내지 8831
141	TK1	8832 내지 8859
142	TOP2A	8860 내지 8954
143	TPX2	8955 내지 9016
144	TRIP13	9017 내지 9063
145	TROAP	9064 내지 9115
146	TTK	9116 내지 9164
147	UBE2C	9165 내지 9190
148	UBE2S	9191 내지 9209
149	ZWINT	9210 내지 9241

청구범위

[청구항 1] 하기 표에 기재된 유전자의 mRNA 발현 수준을 측정하기 위한 제제를 포함하는, 유방암의 예후 예측용 조성물:

번호	유전자명	번호	유전자명	번호	유전자명	번호	유전자명	번호	유전자명
1	ACTB	31	CDKN3	61	FANCI	91	LRRC59	121	RAI2
2	APOBEC3B	32	CENPA	62	FBXO5	92	MAD2L1	122	RFC4
3	ASF1B	33	CENPE	63	FEN1	93	MARCH8	123	RPLP0
4	ASPM	34	CENPF	64	FOXM1	94	MCM10	124	RRM2
5	AURKA	35	CENPM	65	GAPDH	95	MCM2	125	SCUBE2
6	AURKB	36	CENPN	66	GIN51	96	MCM6	126	SETBP1
7	BAG1	37	CEP55	67	GRB7	97	MELK	127	SF3B3
8	BCL2	38	CHEK1	68	GSTM1	98	MKI67	128	SHCBP1
9	BIRC5	39	CIRBP	69	GTSE1	99	MLF1IP	129	SHMT2
10	BLM	40	CKS2	70	GUSB	100	MMP11	130	SLC25A12
11	BUB1	41	CRIM1	71	HJURP	101	MYBL2	131	SLC7A5
12	BUB1B	42	CTSL2	72	HMMR	102	NCAPG	132	SPAG5
13	C14orf45	43	CX3CR1	73	HN1	103	NCAPG2	133	SPC25
14	C16orf61	44	CYBRD1	74	IFT46	104	NCAPH	134	SQLE
15	C7orf63	45	DBF4	75	KIF11	105	NDC80	135	STARD13
16	CACNA1D	46	DDX39	76	KIF13B	106	NEK2	136	STIL
17	CCNA2	47	DLGAP5	77	KIF14	107	NUP93	137	STMN1
18	CCNB1	48	DNMT3B	78	KIF15	108	NUSAP1	138	SYNC
19	CCNB2	49	DONSON	79	KIF18A	109	OIP5	139	TACC3
20	CCNE1	50	DTL	80	KIF18B	110	PBK	140	TFRC
21	CCNE2	51	E2F1	81	KIF20A	111	PDSS1	141	TK1
22	CCT5	52	E2F8	82	KIF23	112	PGR	142	TOP2A
23	CD68	53	ECHDC2	83	KIF2C	113	PKMYT1	143	TPX2
24	CDC20	54	ERBB2	84	KIF4A	114	PLK1	144	TRIP13
25	CDC25A	55	ERCC6L	85	KIFC1	115	PLK4	145	TROAP
26	CDC45	56	ESPL1	86	KPNA2	116	PRC1	146	TTK
27	CDC6	57	ESR1	87	LMNB1	117	PTTG1	147	UBE2C
28	CDCA3	58	EXO1	88	LMNB2	118	RACGAP1	148	UBE2S
29	CDCA8	59	EZH2	89	LRIG1	119	RAD51	149	ZWINT
30	CDK1	60	FAM64A	90	LRRC48	120	RAD51AP1		

- [청구항 2] 청구항 1에 있어서, 상기 제제는 상기 mRNA 또는 그의 cDNA에 특이적으로 혼성화하는 프라이머 쌍 또는 프로브인 것인 조성물.
- [청구항 3] 청구항 1에 있어서, 상기 제제는 상기 mRNA 또는 그의 cDNA에 특이적으로 혼성화하는 올리고뉴클레오티드의 세트인 것인 조성물.
- [청구항 4] 청구항 1에 있어서, 상기 제제는 전사됨으로써 상기 mRNA 또는 그의 cDNA에 특이적으로 혼성화하는 올리고뉴클레오티드를 생성하는 DNA 올리고뉴클레오티드의 세트인 것인 조성물.
- [청구항 5] 청구항 1에 있어서, 상기 제제는 서열번호 1 내지 9241의 뉴클레오티드

서열을 갖는 올리고뉴클레오타이드의 세트인 것인 조성물.

[청구항 6] 청구항 1에 있어서, 상기 유방암은 에스트로겐 수용체 양성을 나타내는 것인 조성물.

[청구항 7] 청구항 1 내지 6 중 어느 한 항의 조성물을 포함하는, 유방암의 예후 예측용 키트.

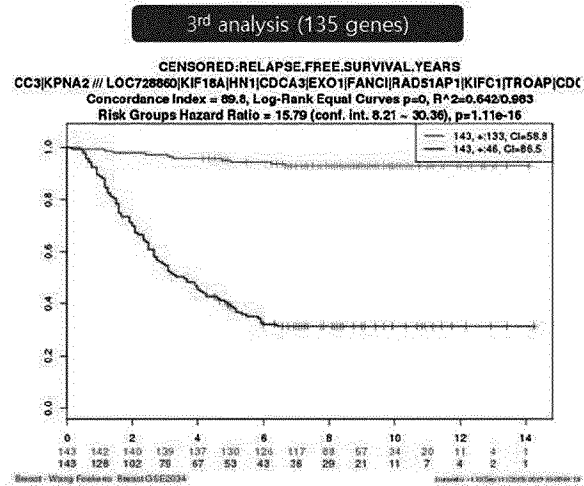
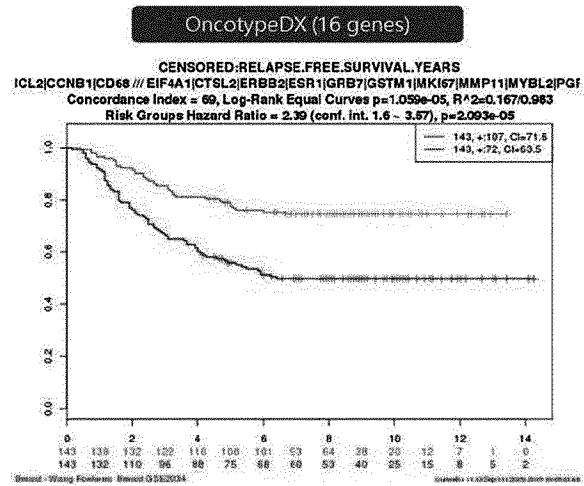
[청구항 8] 청구항 4 또는 5의 제제를 인 비트로(in vitro) 전사하여 RNA 올리고뉴클레오타이드의 세트를 생성하는 단계를 포함하는, 유방암의 예후 예측을 위한 프로브를 제조하는 방법.

[청구항 9] 유방암의 예후를 예측하고자 하는 개체에서 하기 표에 기재된 유전자의 mRNA 발현 수준을 측정하는 단계를 포함하는, 유방암의 예후 예측을 위한 정보를 제공하는 방법:

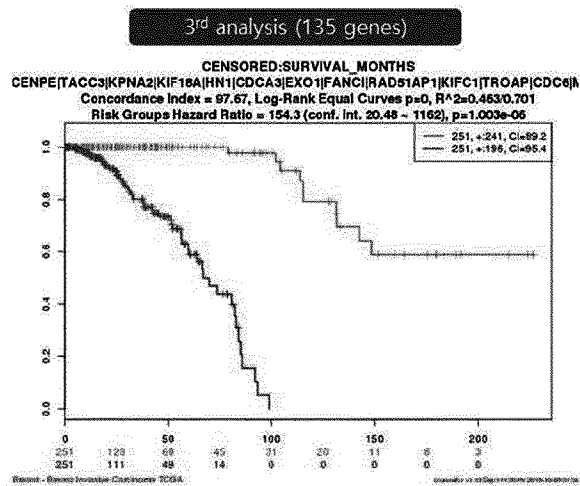
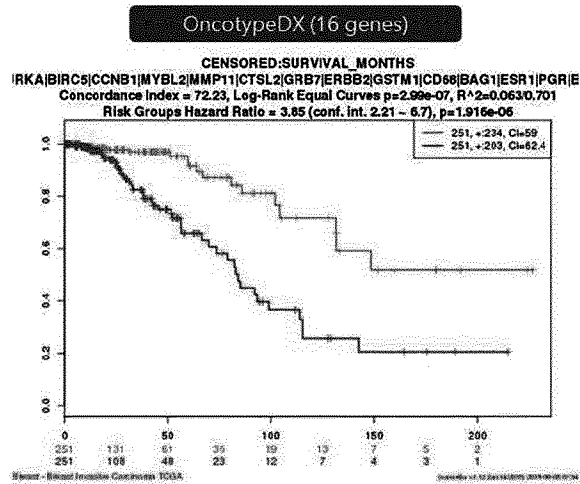
번호	유전자명	번호	유전자명	번호	유전자명	번호	유전자명	번호	유전자명
1	ACTB	31	CDKN3	61	FANCI	91	LRRC59	121	RAI2
2	APOBEC3B	32	CENPA	62	FBXO5	92	MAD2L1	122	RFC4
3	ASF1B	33	CENPE	63	FEN1	93	MARCH8	123	RPLP0
4	ASPM	34	CENPF	64	FOXM1	94	MCM10	124	RRM2
5	AURKA	35	CENPM	65	GAPDH	95	MCM2	125	SCUBE2
6	AURKB	36	CENPN	66	GIN51	96	MCM6	126	SETBP1
7	BAG1	37	CEP55	67	GRB7	97	MELK	127	SF3B3
8	BCL2	38	CHEK1	68	GSTM1	98	MKI67	128	SHCBP1
9	BIRC5	39	CIRBP	69	GTSE1	99	MLF1IP	129	SHMT2
10	BLM	40	CKS2	70	GUSB	100	MMP11	130	SLC25A12
11	BUB1	41	CRIM1	71	HJURP	101	MYBL2	131	SLC7A5
12	BUB1B	42	CTSL2	72	HMMR	102	NCAPG	132	SPAG5
13	C14orf45	43	CX3CR1	73	HN1	103	NCAPG2	133	SPC25
14	C16orf61	44	CYBRD1	74	IFT46	104	NCAPH	134	SQLE
15	C7orf63	45	DBF4	75	KIF11	105	NDC80	135	STARD13
16	CACNA1D	46	DDX39	76	KIF13B	106	NEK2	136	STIL
17	CCNA2	47	DLGAP5	77	KIF14	107	NUP93	137	STMN1
18	CCNB1	48	DNMT3B	78	KIF15	108	NUSAP1	138	SYNC
19	CCNB2	49	DONSON	79	KIF18A	109	OIP5	139	TACC3
20	CCNE1	50	DTL	80	KIF18B	110	PBK	140	TFRC
21	CCNE2	51	E2F1	81	KIF20A	111	PDSS1	141	TK1
22	CCT5	52	E2F8	82	KIF23	112	PGR	142	TOP2A
23	CD68	53	ECHDC2	83	KIF2C	113	PKMYT1	143	TPX2
24	CDC20	54	ERBB2	84	KIF4A	114	PLK1	144	TRIP13
25	CDC25A	55	ERCC6L	85	KIFC1	115	PLK4	145	TROAP
26	CDC45	56	ESPL1	86	KPNA2	116	PRC1	146	TTK
27	CDC6	57	ESR1	87	LMNB1	117	PTTG1	147	UBE2C
28	CDCA3	58	EXO1	88	LMNB2	118	RACGAP1	148	UBE2S
29	CDCA8	59	EZH2	89	LRIG1	119	RAD51	149	ZWINT
30	CDK1	60	FAM64A	90	LRRC48	120	RAD51AP1		

- [청구항 10] 청구항 9에 있어서, 상기 개체로부터 분리된 mRNA 또는 그의 cDNA와 상기 유전자의 mRNA 발현 수준을 측정하기 위한 제제를 인큐베이션하는 것을 포함하는 것인 방법.
- [청구항 11] 청구항 10에 있어서, 상기 제제는 상기 유전자의 mRNA 또는 그의 cDNA에 특이적으로 결합하는 올리고뉴클레오티드이고, 인큐베이션에 의해 상기 유전자의 mRNA 또는 그의 cDNA와 상기 올리고뉴클레오티드의 혼성 복합체(hybridized complex)가 형성되는 것인 방법.
- [청구항 12] 청구항 10에 있어서, 상기 제제는 DNA 올리고뉴클레오티드를 인 비트로 전사하여 수득되는 것인 방법.
- [청구항 13] 청구항 12에 있어서, 상기 DNA 올리고뉴클레오티드는 서열번호 1 내지 9241의 뉴클레오티드 서열을 갖는 올리고뉴클레오티드의 세트인 것인 방법.
- [청구항 14] 청구항 11에 있어서, 상기 혼성 복합체를 분리하는 단계를 더 포함하는 것인 방법.
- [청구항 15] 청구항 14에 있어서, 상기 혼성 복합체는 자성 입자에 의해 분리되는 것인 방법.
- [청구항 16] 청구항 9에 있어서, 상기 발현 수준은 초병렬 시퀀싱에 의해 측정되는 것인 방법.
- [청구항 17] 청구항 16에 있어서, 상기 초병렬 시퀀싱은 합성에 의한 시퀀싱, 이온 토렌트 시퀀싱, 파이로시퀀싱, 라이게이션에 의한 시퀀싱, 나노포어 시퀀싱, 및 단일-분자 실시간 시퀀싱으로 이루어진 군으로부터 선택되는 것인 방법.
- [청구항 18] 청구항 9에 있어서, 상기 유방암은 에스트로겐 수용체 양성을 나타내는 것인 방법.

[도 1]



[도2]



[도3]

