

명세서

발명의 명칭: 암의 진단용 조성물

기술분야

- [1] 본 발명은 질량 분석을 이용한 유방암 진단 방법에 관한 것이다.
- [2]

배경기술

- [3] 유방암은 전 세계 여성 암 발병률 1위인 암이자, 여성 암 1위의 사망요인을 기록하는 치명적인 암이기도 하다. 종래 유방암을 선별하는 방법은 영상적 방법에 치중되어 있다. 특히 유방암의 영상 진단 방법인 맘모그래피(mammography)는 과량의 방사선 노출에 의한 위해 우려가 존재하고, 치밀형 유방에는 진단 정확도가 낮아지는 한계점이 존재하며, 신체 노출 및 압착에 따른 불편과 고통을 환자에게 주게 되는 단점이 있다. 또한 유방초음파 검사는 고가의 검사법으로 이용자의 접근성이 제한적이며, 검사자의 숙련도 또는 기기 노후 정도에 따라 결과의 판정이 달라질 수 있다. 이처럼 종래 기술의 단점 때문에 유방암 조기 진단 분야에서 쉽고 간편하면서 경제적인 검사법에 대한 높은 요구가 존재하며, 이에 가장 적합한 방법 중 하나로 혈액 검사기 꼽힌다.
- [4] 현존하는 유방암 혈액검사는 1980년대 개발되어 1997년 미국 FDA 승인을 받은 CA15-3 면역검사법이 존재하나, 초기 단계의 유방암에 대해서는 진단 정확도가 10~20% 대로 낮아 조기 진단용보다는 치료환자의 모니터링 용도로 사용하고 있는 실정이다.
- [5] 한편, 최근 임상 현장에서 단일마커에 의한 검사로는 정확한 진단을 내리기 어렵다는 인식이 있으며, 이를 해결하기 위한 다(多) 지표 마커가 대안으로 부상하고 있다. 이러한 인식 속에서, 질량 분석법은 동시에 많은 수의 마커를 측정할 수 있을 뿐 아니라 항체를 이용하지 않는다는 점에서 다 지표 마커를 이용하기에 적합한 방법에 해당한다. 다만, 지금까지 질량분석기를 통한 수천 개의 바이오마커 발굴 연구가 지속되어 왔으나 높은 가격과 낮은 재현성 및 긴 분석 시간의 이슈 등에 의해 여전히 임상에 적용되는 사례는 드문 상황이다.
- [6] 질량 분석법이 임상 현장에 잘 적용되기 위해서는 무엇보다 가격 경쟁력 및 재현성을 확보해야 하며, 분석 시간 또한 대폭 줄여야 한다. 특히 혈액 단백질을 전처리 과정에서 재현성과 경제성을 동시에 확보하기 위해 고농도 단백질을 제거(abundant proteins depletion)를 하지 않은 상태의 혈액을 그대로 전처리하는 것이 필요하며, 분석 시간 역시 기존 1-2시간에서 10-20분대로 대폭적인 축소가 필수적이다. 분석에 불필요한 고농도 단백질들(high abundant proteins)의 제거는 단백질 동정(profiling)의 수를 증가시켜 줄 수 있지만, 고농도 단백질들이 제거되는 과정에서 이들과 함께 일부의 피분석 물질들 또한 제거되며 샘플 간 또는 컬럼 간 단백질 제거 정도가 상이하기 때문에 재현성을 장담하기 어렵고, 여러 스텝을 거

쳐야 하는 속성상 핸들링 어려움이 증가하며, 분석 시간이 길어지게 되어 상당한 비용 증가의 요인이 된다.

- [7] 이에, 본 발명자들은 알부민, 면역글로불린, 트랜스페린 등의 고농도 단백질을 제거하지 않은 상태의 혈액(혈청 및 혈장)을 분석의 대상으로 하면서도 10분 이내로 분석 시간을 대폭 축소하여 궁극적으로 임상에서 사용하기에 충분한 경제성과, 재현성을 확보한 유방암 선별 바이오마커를 발굴하고자 하였다.

[8]

- [9] 본 명세서 전체에 걸쳐 다수의 논문 및 특허문헌이 참조되고 그 인용이 표시되어 있다. 인용된 논문 및 특허문헌의 개시 내용은 그 전체로서 본 명세서에 참조로 삽입되어 본 발명이 속하는 기술 분야의 수준 및 본 발명의 내용이 보다 명확하게 설명된다.

[10]

발명의 상세한 설명

기술적 과제

- [11] 본 발명자들은 기존 유방암에 대한 진단방법인 유방X선 촬영술이 가지는 과량의 방사선 노출 및 치밀형 유방에 대한 낮은 진단 정확도를 해결하기 위하여, 고농도의 단백질을 제거하지 않은 상태의 혈액을 분석의 대상으로 하여 간편하면서, 신속한 분석을 가능하게 하는 방법을 개발하기 위하여 예의 연구 노력하였다. 그 결과, 임상 체계에서 사용 가능한 수준의 정량성과 재현성이 확보된 바이오마커를 발굴하였고, 이를 통하여 신속하면서도 높은 신뢰도로 유방암을 진단할 수 있음을 확인함으로써, 본 발명을 완성하게 되었다.
- [12] 따라서 본 발명의 목적은 암의 진단용 조성물을 제공하는 데 있다.
- [13] 본 발명의 다른 목적은 암의 진단용 키트를 제공하는 데 있다.
- [14] 본 발명의 또 다른 목적은 암의 진단을 위한 정보 제공 방법을 제공하는 데 있다.
- [15] 본 발명의 또 다른 목적은 암의 예방 또는 치료용 조성물의 스크리닝 방법을 제공하는 데 있다.
- [16] 본 발명의 다른 목적 및 이점은 하기의 발명의 상세한 설명, 청구범위 및 도면에 의해 보다 명확하게 된다.

[17]

과제 해결 수단

- [18] 본 발명의 일 양태에 따르면, 본 발명은 APOC1(Apolipoprotein C1), CHL1(Neural cell adhesion molecule L1 like), MMP9(Matrix metalloproteinase-9), PRDX6(Peroxiredoxin-6), PRG4(Proteoglycan 4), PPBP(Platelet basic protein), FN1(Fibronectin), VWF(von Willebrand factor) 및 CLU(Clusterin)로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드 또는 이들의 일부 절편; 또는 이를 코

당하는 유전자의 발현 수준을 측정하는 제제를 포함하는, 암의 진단용 조성물을 제공한다.

- [19] 본 발명자들은 기존 유방암에 대한 진단방법인 유방X선 촬영술이 가지는 과량의 방사선 노출 및 치밀형 유방에 대한 낮은 진단 정확도를 해결하기 위하여, 고농도의 단백질들을 제거하지 않은 상태의 혈액을 분석의 대상으로 하여 간편하면서, 신속한 분석을 가능하게 하는 방법을 개발하기 위하여 예의 연구 노력하였다. 그 결과, 임상 체계에서 사용 가능한 수준의 정량성과 재현성이 확보된 바이오마커를 발굴하였고, 이를 통하여 신속하면서도 높은 신뢰도로 유방암을 진단할 수 있음을 확인함으로써, 본 발명을 완성하게 되었다.
- [20] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 아포리포단백질 C1(Apolipoprotein C1, APOC1)의 발현 수준을 측정하는 제제일 수 있다. 상기 APOC1은, 아포리포단백질 C(Apolipoprotein C)의 구성원이며, 인간에서는 *APOC1* 유전자에 의하여 암호화될 수 있다. 또한, 가장 처음에는 간에서 발현되고, 이후 단핵구가 대식세포로 분화할 때 활성화될 수 있다.
- [21] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 CHL1의 발현 수준을 측정하는 제제일 수 있다. 상기 CHL1은, “L1의 근접 상동(Close Homolog of L1)”을 뜻하며, 또한 신경세포부착분자-L1-유사-단백질(Neural Cell Adhesion Molecule L1 Like protein)이라고도 불릴 수 있으며, 인간에서는 *CHL1* 유전자에 의하여 암호화될 수 있다.
- [22] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 기질금속단백분해효소 9(Matrix Metalloproteinase-9, MMP9)의 발현 수준을 측정하는 제제일 수 있다. 상기 MMP-9은 92kDa 형 IV 콜로게나제(92kDa type IV Collagenase), 92kDa 젤라티나제(92kDa gelatinase) 또는 젤라티나제 B(Gelatinase B, GELB)로도 알려져 있다. MMP-9은 아연-금속단백분해효소 계열(Zinc-Metalloproteinases family) 계열의 구성원이며, 세포외기질(Extracellular matrix)을 분해하는 것에 관여한다고 알려져 있다. 인간에서는 *MMP9* 유전자에 의하여 암호화될 수 있다.
- [23] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 페로시레독신-6(Peroxiredoxin-6, PRDX6)의 발현 수준을 측정하는 제제일 수 있다. 상기 PRDX6는 항산화효소인 페레독신 패밀리(Peroxidoxin Family)의 구성원일 수 있으며, 인간에서는 *PRDX6* 유전자에 의하여 암호화될 수 있다.
- [24] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 프로테오글리칸 4(Proteoglycan 4, PRG4)의 발현 수준을 측정하는 제제일 수 있다. 상기 PRDG4는 루브리신(Lubricin)이라고도 불리며, 인간에서는 *Prp4* 유전자에 의하여 암호화될 수 있다.
- [25] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 전혈소판 기본 단백질(Pro-Platelet basic protein, PPBP)의 발현 수준을 측정하는 제제일 수 있다. 상기 PPBP는 케모카인(C-X-C 모티프) 리간드(케모카인 (C-X-C motif) ligand, CXCL7)이라고도 불리며, 인간에서는 *PPBP* 유전자에 의하여 암호화될 수 있다.

- [26] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 파이브로넥틴 1(Fibronectin 1, FN1)의 발현 수준을 측정하는 제제일 수 있다. 상기 FN1은 고분자량 당단백질로써 세포외기질의 막수용체 단백질인 인테그린(Integrin)에 부착될 수 있으며, 다른 세포외기질 단백질인 콜라겐, 파이브린, 황산헤파란 당단백질(Heparan Sulfate Proteoglycans)과 결합할 수 있다. 또한, 인간에서는 *FN1* 유전자에 의하여 암호화될 수 있다.
- [27] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 폰빌레브란드 인자(von Willebrand Factor, VWF)의 발현 수준을 측정하는 제제일 수 있다. 상기 VWF는 혈관내피세포 또는 골수의 거대핵세포에서 생산되는 접착인자로써 혈소판이 내피하조직과 결합 시 접착제 역할을 하거나, 응고인자 중 제 VIII인자의 보조체로써 혈액중에서 해당 인자와 결합하여 안정화를 유도하는 작용을 수행할 수 있다. 또한, 인간에서는 *VWF* 유전자에 의하여 암호화될 수 있다.
- [28] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 클러스터린(Clusterin, CLU)의 발현 수준을 측정하는 제제일 수 있다. 클러스터린 단백질은 75-80 kDa의 분자량을 가지는 디설파이드-연결 헤테로다имер 단백질로서, 세포 잔해 제거 및 세포 사멸과 관련이 있는 것으로 알려져 있다. 또한, 인간에서는 *CLU* 유전자에 의하여 암호화될 수 있다.
- [29]
- [30] 본 발명에서 진단의 대상이 되는 질환으로 상기 “암”은 포유류에서 전형적으로 조절되지 않는 세포 성장으로 특징지어진 생리적 상태를 나타내거나 가리킨다. 본 발명에서 진단의 대상이 되는 암은 구체적으로 유방암, 난소암, 대장암, 위암, 간암, 췌장암, 자궁경부암, 갑상선암, 부갑상선암, 폐암, 비소세포성폐암, 전립선암, 담낭암, 담도암, 비 호지킨 림프종, 호지킨 림프종, 혈액암, 방광암, 신장암, 흑색종, 결장암, 골암, 피부암, 두부암, 자궁암, 직장암, 뇌종양, 항문부근암, 나팔관암종, 자궁내막암종, 질암, 음문암종, 식도암, 소장암, 내분비선암, 부신암, 연조직 육종, 요도암, 음경암, 수뇨관암, 신장세포암종, 신장골반암종, 중추신경계(CNS, central nervous system) 종양, 1차 CNS 림프종, 척수 종양, 뇌간 신경교종 또는 뇌하수체 선종일 수 있으며, 보다 구체적으로는 유방암일 수 있다.
- [31]
- [32] 본 발명에서 상기 “진단”은 특정 질병 또는 질환에 대한 대상(subject)의 감수성(susceptibility)을 판정하는 것, 대상이 특정 질병 또는 질환을 현재 가지고 있는지를 판정하는 것, 특정 질병 또는 질환에 걸린 대상의 예후(prognosis)(예컨대, 전-전이성 또는 전이성 암 상태의 동정, 암의 단계 결정 또는 치료에 대한 암의 반응성 결정)를 판정하는 것, 또는 테라메트릭스(therametrics)(예컨대, 치료 효능에 대한 정보를 제공하기 위하여 객체의 상태를 모니터링하는 것)을 포함한다. 본 발명의 목적상, 상기 진단은 상기한 암의 발병 여부 또는 발병 가능성(위험성)을 확인하는 것이다.

- [33] 본 발명에서 상기 폴리펩타이드들의 발현 수준을 측정하는 제제는 특별히 제한하지는 않으나, 예를 들면 상기 폴리펩타이드들에 특이적으로 결합하는 항체, 올리고펩타이드, 리간드, PNA(peptide nucleic acid) 및 앵타머(aptamer)로 이루어진 군에서 선택된 1종 이상을 포함할 수 있다.
- [34] 본 발명에 상기 “항체”는 항원과 특이적으로 결합하여 항원-항체 반응을 일으키는 물질을 가리킨다. 본 발명의 목적상, 항체는 본 발명에서 언급하는 폴리펩타이드들에 대해 특이적으로 결합하는 항체를 의미한다.
- [35] 본 발명의 상기 항체는 다클론 항체, 단클론 항체 및 재조합 항체를 모두 포함한다. 상기 항체는 당업계에 널리 공지된 기술을 이용하여 용이하게 제조될 수 있다. 예를 들어, 다클론 항체는 상기 단백질의 항원을 동물에 주사하고 동물로부터 채혈하여 항체를 포함하는 혈청을 수득하는 과정을 포함하는 당업계에 널리 공지된 방법에 의해 생산될 수 있다. 이러한 다클론 항체는 염소, 토끼, 양, 원숭이, 말, 돼지, 소, 개 등의 임의의 동물로부터 제조될 수 있다. 또한, 단클론 항체는 당업계에 널리 공지된 하이브리도마 방법(hybridoma method; Kohler 및 Milstein (1976) *European Journal of Immunology* 6:511-519 참조), 또는 파지 항체 라이브러리 기술 (Clackson et al, *Nature*, 352:624-628, 1991; Marks et al, *J. Mol. Biol.*, 222:58, 1-597, 1991 참조)을 이용하여 제조될 수 있다. 상기 방법으로 제조된 항체는 겔 전기영동, 투석, 염 침전, 이온교환 크로마토그래피, 친화성 크로마토그래피 등의 방법을 이용하여 분리, 정제될 수 있다. 또한, 본 발명의 항체는 2개의 전장의 경쇄 및 2개의 전장의 중쇄를 갖는 완전한 형태뿐만 아니라, 항체 분자의 기능적인 단편을 포함한다. 항체 분자의 기능적인 단편이란, 적어도 항원 결합 기능을 보유하고 있는 단편을 의미하며, Fab, F(ab'), F(ab')₂ 및 Fv 등이 있다.
- [36]
- [37] 본 발명에 상기 “PNA(Peptide Nucleic Acid)”는 인공적으로 합성된, DNA 또는 RNA와 비슷한 중합체를 가리키며, 1991년 덴마크 코펜하겐 대학교의 Nielsen, Egholm, Berg와 Buchardt 교수에 의해 처음으로 소개되었다. DNA는 인산-리보스당 골격을 갖는데 반해, PNA는 펩타이드 결합에 의해 연결된 반복된 N-(2-아미노에틸)-글리신 골격을 가지며, 이로 인해 DNA 또는 RNA에 대한 결합력과 안정성이 크게 증가되어 분자 생물학, 진단 분석 및 안티센스 치료법에 사용되고 있다. PNA는 문헌[Nielsen PE, Egholm M, Berg RH, Buchardt O (December 1991). “Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide”. *Science* 254(5037): 1497-1500]에 상세하게 개시되어 있다.
- [38] 본 발명에서 상기 “앵타머”는 올리고핵산 또는 펩타이드 분자이며, 앵타머의 일반적인 내용은 문헌[Bock LC et al., *Nature* 355(6360):5646(1992); Hoppe-Seyler F, Butz K “Peptide aptamers: powerful new tools for molecular medicine”. *J Mol Med.* 78(8):42630(2000); Cohen BA, Colas P, Brent R. “An artificial cell- cycle inhibitor isolated from a combinatorial library”. *Proc Natl Acad Sci USA.* 95(24): 142727(1998)] 에 상세하게 개시되어 있다.

- [39] 본 발명에서 상기 폴리펩타이드들을 코딩하는 유전자의 발현 수준을 측정하는 제제는 상기 유전자에 특이적으로 결합하는 프라이머, 프로브 및 안티센스 뉴클레오티드로 이루어진 군에서 선택된 1종 이상을 포함할 수 있다.
- [40] 본 발명에서 상기 “프라이머”는 표적 유전자 서열을 인지하는 단편으로서, 정방향 및 역방향의 프라이머 쌍을 포함하나, 바람직하게는, 특이성 및 민감성을 가지는 분석 결과를 제공하는 프라이머 쌍이다. 프라이머의 핵산 서열이 시료 내 존재하는 비-표적 서열과 불일치하는 서열이어서, 상보적인 프라이머 결합 부위를 함유하는 표적 유전자 서열만 증폭하고 비특이적 증폭을 유발하지 않는 프라이머일 때, 높은 특이성이 부여될 수 있다.
- [41] 본 발명에서 상기 “프로브”란 시료 내의 검출하고자 하는 표적 물질과 특이적으로 결합할 수 있는 물질을 의미하며, 상기 결합을 통하여 특이적으로 시료 내의 표적 물질의 존재를 확인할 수 있는 물질을 의미한다. 프로브의 종류는 당업계에서 통상적으로 사용되는 물질로서 제한은 없으나, 바람직하게는 PNA(peptide nucleic acid), LNA(locked nucleic acid), 펩타이드, 폴리펩타이드, 단백질, RNA 또는 DNA일 수 있으며, 가장 바람직하게는 PNA이다. 보다 구체적으로, 상기 프로브는 바이오 물질로서 생물에서 유래되거나 이와 유사한 것 또는 생체 외에서 제조된 것을 포함하는 것으로, 예를 들어, 효소, 단백질, 항체, 미생물, 동식물 세포 및 기관, 신경세포, DNA, 및 RNA일 수 있으며, DNA는 cDNA, 게놈 DNA, 올리고뉴클레오타이드를 포함하며, RNA는 게놈 RNA, mRNA, 올리고뉴클레오타이드를 포함하며, 단백질의 예로는 항체, 항원, 효소, 펩타이드 등을 포함할 수 있다.
- [42] 본 발명에서 상기 “LNA(Locked nucleic acids)”란, 2'-O, 4'-C 메틸렌 브릿지를 포함하는 핵산 아날로그를 의미한다 [J Weiler, J Hunziker and J Hall Gene Therapy (2006) 13, 496.502]. LNA 뉴클레오사이드는 DNA와 RNA의 일반적 핵산 염기를 포함하며, Watson-Crick 염기 쌍 규칙에 따라 염기 쌍을 형성할 수 있다. 하지만, 메틸렌 브릿지로 인한 분자의 ‘locking’으로 인해, LNA는 Watson-Crick 결합에서 이상적 형상을 형성하지 못하게 된다. LNA가 DNA 또는 RNA 올리고뉴클레오티드에 포함되면, LNA는 보다 빠르게 상보적 뉴클레오티드 사슬과 쌍을 이루어 이중 나선의 안정성을 높일 수 있다. 본 발명에서 상기 “안티센스”는 안티센스 올리고머가 왓슨-크릭 염기쌍 형성에 의해 RNA 내의 표적 서열과 혼성화되어, 표적 서열 내에서 전형적으로 mRNA와 RNA:올리고머 헤테로 이중체의 형성을 허용하는, 뉴클레오티드 염기의 서열 및 서브유닛 간 백본을 갖는 올리고머를 의미한다. 올리고머는 표적 서열에 대한 정확한 서열 상보성 또는 근사 상보성을 가질 수 있다.
- [43] 본 발명에 따른 폴리펩타이드의 아미노산 서열 정보 및 이를 코딩하는 핵산 서열은 다양한 public data를 통해 개시되어 있으므로, 당업자라면 이를 바탕으로 상기 폴리펩타이드를 코딩하는 유전자에 특이적으로 결합하는 프라이머, 프로브 또는 안티센스 뉴클레오티드를 용이하게 디자인할 수 있을 것이다.

- [44]
- [45] 본 발명의 구체적인 구현예에 따르면,
- [46] 상기 APOC1 폴리펩타이드의 일부 절편은 서열번호 1(TPDVSSALDK)의 아미노산 서열을 가지며;
- [47] 상기 CHL1 폴리펩타이드의 일부 절편은 서열번호 2(VIAVNEVGR)의 아미노산 서열을 가지고;
- [48] 상기 MMP9 폴리펩타이드의 일부 절편은 서열번호 3(AVIDDAFAR)의 아미노산 서열을 가지며;
- [49] 상기 PRDX6 폴리펩타이드의 일부 절편은 서열번호 4(LSILYPATTGR)의 아미노산 서열을 가지고;
- [50] 상기 PRG4 폴리펩타이드의 일부 절편은 서열번호 5(AIGPSQTHIR)의 아미노산 서열을 가지며;
- [51] 상기 PPBP 폴리펩타이드의 일부 절편은 서열번호 6(TTSGIHPK)의 아미노산 서열을 가지고;
- [52] 상기 FN1 폴리펩타이드의 일부 절편은 서열번호 7(STTPDITGYR)의 아미노산 서열을 가지며;
- [53] 상기 VWF 폴리펩타이드의 일부 절편은 서열번호 8(ILAGPAGDSNVVK)의 아미노산 서열을 가지고;
- [54] 상기 CLU 폴리펩타이드의 일부 절편은 서열번호 9(TLLSNLEEAK)의 아미노산 서열을 가진다.
- [55]
- [56] 본 발명의 구체적인 구현예에 따르면, 본 발명의 진단용 조성물로 진단할 수 있는 암은 유방암이다.
- [57]
- [58] 본 발명의 구체적인 구현예에 따르면, 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 상기 폴리펩타이드 또는 이의 일부 절편에 특이적으로 결합하는 항체, 올리고펩타이드, 리간드, PNA(peptide nucleic acid) 및 앵타머(aptamer)로 이루어진 군에서 선택된 1 종 이상을 포함한다.
- [59]
- [60] 본 발명의 구체적인 구현예에 따르면, 상기 폴리펩타이드 또는 이의 일부 절편을 코딩하는 유전자의 발현 수준을 측정하는 제제는 상기 유전자에 특이적으로 결합하는 프라이머, 프로브 및 안티센스 뉴클레오티드로 이루어진 군에서 선택된 1종 이상을 포함한다.
- [61]
- [62] 본 발명의 또 다른 양태에 따르면, 본 발명은 본 발명의 진단용 조성물을 포함하는 진단용 키트를 제공한다.
- [63] 본 발명에서는 상기 진단용 키트를 이용하여 암 질환의 발병 여부, 발병 가능성, 치료 반응성, 예후, 병기, 재발 가능성 등을 진단할 수 있다.

- [64] 본 발명에서 상기 진단의 대상이 되는 상기 암에 대해서는 이미 상술하였으므로, 과도한 중복을 피하기 위하여 이하 그 자세한 기재를 생략한다.
- [65] 본 발명에서 상기 키트는 RT-PCR 키트, DNA 칩 키트, ELISA 키트, 단백질 칩 키트, 래피드(rapid) 키트 또는 MRM(Multiple reaction monitoring) 키트일 수 있으나, 이에 제한되는 것은 아니다.
- [66] 본 발명의 상기 암의 진단용 키트는 분석 방법에 적합한 한 종류 또는 그 이상의 다른 구성 성분 조성물, 용액 또는 장치를 더 포함할 수 있다.
- [67] 예를 들면, 본 발명에서 상기 암의 진단용 키트는 역전사 중합효소 반응을 수행하기 위해 필요한 필수 요소를 더 포함할 수 있다. 역전사 중합효소 반응 키트는 마커 단백질을 코딩하는 유전자에 대해 특이적인 프라이머 쌍을 포함한다. 프라이머는 상기 유전자의 핵산서열에 특이적인 서열을 가지는 뉴클레오티드로서, 약 7bp 내지 50bp의 길이, 보다 구체적으로는 약 10bp 내지 30bp의 길이를 가질 수 있다. 또한 대조군 유전자의 핵산 서열에 특이적인 프라이머를 포함할 수 있다. 그 외 역전사 중합 효소반응 키트는 테스트 튜브 또는 다른 적절한 용기, 반응 완충액(pH 및 마그네슘 농도는 다양), 데옥시뉴클레오타이드(dNTPs), Taq-폴리머라아제 및 역전사효소와 같은 효소, DNase, RNase 억제제 DEPC-수(DEPC-water), 멸균수 등을 포함할 수 있다.
- [68] 또한, 본 발명의 진단용 키트는 DNA 칩을 수행하기 위해 필요한 필수 요소를 포함할 수 있다. DNA 칩 키트는 유전자 또는 그의 단편에 해당하는 cDNA 또는 올리고뉴클레오티드(oligonucleotide)가 부착되어 있는 기판, 및 형광 표지 프로브를 제작하기 위한 시약, 제제, 효소 등을 포함할 수 있다. 또한 기판은 대조군 유전자 또는 그의 단편에 해당하는 cDNA 또는 올리고뉴클레오티드를 포함할 수 있다.
- [69] 또한, 본 발명의 진단용 키트는 ELISA를 수행하기 위해 필요한 필수 요소를 포함할 수 있다. ELISA 키트는 상기 단백질에 대해 특이적인 항체를 포함한다. 항체는 마커 단백질에 대한 특이성 및 친화성이 높고 다른 단백질에 대한 교차 반응성이 거의 없는 항체로, 단클론 항체, 다클론 항체 또는 재조합 항체이다. 또한 ELISA 키트는 대조군 단백질에 특이적인 항체를 포함할 수 있다. 그 외 ELISA 키트는 결합된 항체를 검출할 수 있는 시약, 예를 들면, 표지된 2차 항체, 발색단(chromophores), 효소(예: 항체와 컨주게이트됨) 및 그의 기질 또는 항체와 결합할 수 있는 다른 물질 등을 포함할 수 있다.
- [70] 본 발명의 진단용 키트에서 항원-항체 결합반응을 위한 고정체로는 니트로셀룰로오즈 막, PVDF 막, 폴리비닐 (polyvinyl) 수지 또는 폴리스티렌(polystyrene) 수지로 합성된 웰 플레이트(Well plate), 유리로 된 슬라이드 글래스 등이 사용될 수 있으나, 이에 제한되는 것은 아니다.
- [71] 또한, 본 발명의 진단용 키트에서 2차 항체의 표지체는 발색 반응을 하는 통상의 발색제가 바람직하며, HRP(horseradish peroxidase), 염기성 탈인산화효소(alkaline phosphatase), 콜로이드 골드(colloid gold), FITC(폴리 L-라이신-플루오르

세인 아이소티오시아네이트), RITC(로다민-B-아이소티오시아네이트) 등의 형광 물질 (fluorescein) 및 색소(dye) 등의 표지체가 사용될 수 있으나, 이에 제한되는 것은 아니다.

- [72] 또한, 본 발명의 진단용 키트에서 발색을 유도하기 위한 발색 기질은 발색 반응을 하는 표지체에 따라 사용하는 것이 바람직하며, TMB(3,3',5,5'-테트라메틸 베 지딘), ABTS[2,2'-아지노-비스(3-에틸벤조티아졸린-6-설폰산)], OPD(o-페닐렌다 이아민) 등을 사용할 수 있다. 이때, 발색기질은 완충용액(0.1 M NaAc, pH 5.5)에 용해된 상태로 제공되는 것이 더욱 바람직하다. TMB와 같은 발색기질은 이차항 체 접합체의 표지체로 사용된 HRP에 의해 분해되어 발색 침적체를 생성하고, 이 발색 침적체의 침적 정도를 육안으로 확인함으로써 상기 마커 단백질들의 존재 유무를 검출한다.
- [73] 본 발명의 진단용 키트에서 세척액은 인산염 완충용액, NaCl 및 트윈 20(Tween 20)을 포함하는 것이 바람직하며, 0.02M 인산염 완충용액, 0.13M NaCl, 및 0.05% 트윈 20으로 구성된 완충용액(PBST)이 더욱 바람직하다. 세척액은 항원-항체 결합반응 후 항원-항체 결합체에 2차 항체를 반응시킨 다음 적당량을 고정체에 첨 가하여 3 내지 6회 세척한다. 반응 정지용액은 황산 용액(H_2SO_4)이 바람직하게 사용될 수 있다.
- [74]
- [75] 본 발명의 구체적인 구현예에 따르면, 본 발명의 진단용 키트는 RT-PCR 키트, DNA 칩 키트, ELISA 키트, 단백질 칩 키트, 래피드(rapid) 키트 또는 MRM(Multiple reaction monitoring) 키트이다.
- [76]
- [77] 본 발명의 또 다른 양태에 따르면, 본 발명은 목적하는 개체로부터 분리된 생물학적 시료에서, APOC1(Apolipoprotein C1), CHL1(Neural cell adhesion molecule L1 like), MMP9 (Matrix metalloproteinase-9), PRDX6(Peroxiredoxin-6) , PRG4(Proteoglycan 4), PPBP(Platelet basic protein), FN1(Fibronectin), VWF(von Willebrand factor) 및 CLU(Clusterin)로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드 또는 이들의 일부 절편; 또는 이를 코딩하는 유전자의 발현 수준을 측정하는 단계를 포함한다.
- [78] 본 발명에서 측정의 대상이 되는 폴리펩타이드들에 대해서는 이미 상술하였으 므로, 과도한 중복을 피하기 위하여 이하 그 자세한 기재를 생략한다.
- [79] 본 발명에서 상기 “목적하는 개체”란, 상기 암의 발병 여부가 불확실한 개체로, 발병 가능성이 높은 개체를 의미한다.
- [80] 본 발명에서 상기 “생물학적 시료”는 개체로부터 얻어지거나 개체로부터 유래된 임의의 물질, 생물학적 체액, 조직 또는 세포를 의미하는 것으로, 예 를 들면, 전혈(whole blood), 백혈구(leukocytes), 말초혈액 단핵 세포 (peripheral blood mononuclear cells), 백혈구 연층(buffy coat), 혈장(plasma), 혈청(serum), 객 담(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), 뇌척수 막 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) 또는 뇌척수액(cerebrospinal fluid)을 포함할 수 있지만, 바람직하게는 발병 가능성이 높은 환자의 피부를 절개하지 않고 중공침 등을 생체 내 기관에 자입하여 병리조직학적 검사용으로 채취한 액체 생검(예를 들면, 환자의 조직, 세포, 혈액, 혈청, 혈장, 타액, 객담 또는 복수(ascites) 등)일 수 있다.

- [81] 본 발명에서는 상기와 같이 분리된 생물학적 시료에서 서열번호 1 내지 9로 표시되는 폴리펩타이드 또는 이를 코딩하는 유전자의 발현 수준을 측정하는 단계를 포함할 수 있다.
- [82] 본 발명에서 상기 발현 수준을 측정하는 단계는 APOC1(Apolipoprotein C1), CHL1(Neural cell adhesion molecule L1 like), MMP9(Matrix metalloproteinase-9), PRDX6(Peroxiredoxin-6), PRG4(Proteoglycan 4), PPBP(Platelet basic protein), FN1(Fibronectin), VWF(von Willebrand factor) 및 CLU(Clusterin)로 구성된 군으로부터 선택되는 하나 이상의 단백질(폴리펩타이드) 또는 이를 코딩하는 유전자의 발현 수준을 측정하는 단계일 수 있다.
- [83] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 특별히 제한하지는 않으나, 예를 들면 상기 폴리펩타이드에 특이적으로 결합하는 항체, 올리고펩타이드, 리간드, PNA(peptide nucleic acid) 및 앵타머(aptamer)로 이루어진 군에서 선택된 1종 이상을 포함할 수 있다.
- [84] 본 발명에 상기 폴리펩타이드의 발현 수준을 측정 또는 비교 분석 방법으로는 단백질 칩 분석, 면역측정법, 리간드 바인딩 어세이, MALDI-TOF(Matrix Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry) 분석, SELDI-TOF(Surface Enhanced Laser Desorption/Ionization Time of Flight Mass Spectrometry) 분석, 방사선 면역분석, 방사 면역 확산법, 오우크테로니 면역 확산법, 로케트 면역전기영동, 조직면역 염색, 보체 고정 분석법, 2차원 전기영동 분석, 액상 크로마토그래피-질량분석(liquid chromatography- Mass Spectrometry, LC-MS), LC-MS/MS(liquid chromatography-Mass Spectrometry/ Mass Spectrometry), 웨스턴 블랏팅 및 ELISA(enzyme linked immunosorbent assay) 등이 있으나, 이에 제한되는 것은 아니다.
- [85] 본 발명에서 상기 폴리펩타이드의 발현 수준을 측정 또는 비교 분석 방법으로는 다중 반응 모니터링(multiple reaction monitoring; MRM) 방법에 의할 수 있다.
- [86] 본 발명에서 상기 다중 반응 모니터링 방법은 질량분석기 (mass-spectrometry), 바람직하게는 삼중 사극자 질량 분석기 (triple quadrupole mass-spectrometry)를 이용하여 수행될 수 있다.

- [87] 본 발명에서 질량분석기 (mass-spectrometry)를 이용한 다중 반응 모니터링 (multiple reaction monitoring; MRM) 방법은 특정 분석물질을 선택적으로 분리하여 검출하고 정량하여 그 농도변화를 모니터링할 수 있는 분석 기술이다. MRM은 생체 시료 중에 존재하는 미량의 바이오마커와 같은 물질을 정량적으로 정확하게 다중 측정할 수 있는 방법으로 제 1 질량필터 (Q1)를 이용하여 이온화원에서 생성된 이온 단편들 중 어미이온을 선택적으로 충돌관으로 전달한다. 이어 충돌관에 도달한 어미이온은 내부 충돌기체와 충돌하여, 쪼개져 딸이온을 생성하여 제2 질량 필터 (Q2)로 보내지고, 여기서 특징적인 이온만이 검출부로 전달된다. 이런 방식으로 목적하는 성분의 정보만을 검출할 수 있는 선택성 및 민감도가 높은 분석방법이다. MRM은 작은 분자의 정량분석에 활용되어 특정 유전병을 진단하는데 쓰이고 있다. MRM 방법은 다수의 펩타이드를 동시에 측정하기에 용이하며, 항체가 없이 정상 인과 암환자 사이에서 단백질 진단 마커 후보들의 상대적 농도차를 확인할 수 있다는 장점이 있다. 또한 민감도와 선택성이 탁월하여 특히, 질량분석기를 이용한 프로테옴 분석에서 혈액 내에 있는 복잡한 단백질과 펩타이드의 분석을 위해 MRM 분석방법이 도입되고 있다(Anderson L. et al., Mol Cell Proteomics, 5: 375-88, 2006; DeSouza, L. V. et al., Anal. Chem., 81: 3462-70, 2009).
- [88] 본 발명에서 상기 다중 반응 모니터링 방법에 의해 본 발명에서 언급하는 폴리펩타이드의 발현 수준의 측정을 할 수 있다.
- [89] 본 발명에서 상기 다중 반응 모니터링 방법에 의해 본 발명에서 언급하는 폴리펩타이드의 발현 수준을 분석하기 위하여, 타겟 펩타이드의 질량/전하 값(m/z 값)을 사용할 수 있고, 이 때 m/z 값의 정보는 하기 표 1에 나타낸 바와 같으나, 이에 제한되는 것은 아니다.

[90]

[91] [표1]

No	Gene	Protein	Accession No.	Sequence	M+H
1	APOC1	Apolipoprotein C1	P02654	TPDVSSALDK	1032.5280
2	CHL1	Neural cell adhesion molecule L1 like	O00533	VIAVNEVGR	966.5524
3	MMP9	Matrix metalloproteinase-9	P14780	AVIDDAFAR	977.5051
4	PRDX6	Peroxiredoxin-6	P30041	LSILYPATTGR	1191.6732
5	PRG4	Proteoglycan 4	Q92954	AIGPSQTHIR	1180.6433
6	PPBP	Platelet basic protein	P02775	TTSGIHPK	840.4574
7	FN1	Fibronectin	P02751	STTPDITGYR	1110.5426

8	VWF	von Willebrand factor	P04275	ILAGPAG DSNVVK	1240.6896
9	CLU	Clusterin	P10909	TLLSNLEEAK	1117.6099

- [92] 본 발명에서 폴리펩타이드는 APOC1(Apolipoprotein C1), CHL1(Neural cell adhesion molecule L1 like), MMP9(Matrix metalloproteinase-9), PRDX6(Peroxiredoxin-6), PRG4(Proteoglycan 4), PPBP(Platelet basic protein), FN1(Fibronectin), VWF(von Willebrand factor) 또는 CLU(Clusterin)일 수 있다. 상기 폴리펩타이드들의 타겟 펩타이드를 이용하여 목적하는 개체의 생물학적 시료 속에 존재하는 각각의 폴리펩타이드들의 발현 수준을 측정할 수 있다.
- [93] 본 발명에 있어서 내부 표준 물질은 상기 다중 반응 모니터링 분석 시 일반적으로 사용되는 임의의 내부 표준 물질이 사용될 수 있으나, 예를 들어, 대장균 베타 갈락토시다아제가 사용될 수 있다.
- [94] 또한, 본 발명에서 상기 폴리펩타이드의 혈액 내 절대량을 측정하기 위하여 타겟 펩타이드의 일부 아미노산이 안정한 동위원소로 치환된 특정 펩타이드를 내부 표준 물질로서 합성하는 경우, 동위 원소로 치환된 아미노산은 리신(Lysine)이나 아르기닌(Arginine)일 수 있으나, 이에 제한되는 것은 아니다. 여기서 합성된 펩타이드는 95% 이상 순수 분리된 펩타이드를 사용할 수 있다.
- [95] 본 발명에서 내부 표준 물질은 ^2H , ^3H , ^{11}C , ^{13}C , ^{14}C , ^{13}N , ^{15}N , ^{15}O , ^{17}O 및 ^{18}O 로 구성된 군으로부터 선택되는 하나 이상의 방사성 동위원소를 포함할 수 있으나, 전술한 방사성 동위원소에 국한되지 않고, 폴리펩타이드의 절대량을 측정하는데 비교군으로 사용할 수 있는 모든 종류의 동위원소를 포함할 수 있다.
- [96] 한편, 본 발명에서 상기 폴리펩타이드를 코딩하는 유전자의 발현 수준을 측정하는 제제는 상기 유전자에 특이적으로 결합하는 프라이머, 프로브 및 안티센스 뉴클레오티드로 이루어진 군에서 선택된 1종 이상을 포함할 수 있다.
- [97] 본 발명에 상기 폴리펩타이드를 코딩하는 유전자의 존재 여부와 발현 정도를 확인하는 과정으로 mRNA의 양을 측정하는 분석 방법으로는 역전사 중합효소 반응(RT-PCR), 경쟁적 역전사 중합효소반응 (Competitive RT-PCR), 실시간 역전사 중합효소반응(Real-time RT-PCR), RNase 보호 분석법(RPA; RNase protection assay), 노던 블랏팅(Northern blotting), DNA 칩 등이 있으나 이에 제한되는 것은 아니다.
- [98] 본 발명에서 목적하는 개체의 생물학적 시료에 대하여 측정된 상기 폴리펩타이드 또는 이를 코딩하는 유전자의 발현 수준이 정상 대조군에 비하여 증가하거나 감소한 경우, 상기 암의 발병 가능성이 높은 것으로 예측할 수 있다.
- [99] 본 발명에서 목적하는 개체의 생물학적 시료에 대하여 측정된 상기 폴리펩타이드 또는 이를 코딩하는 유전자의 발현 수준을 측정하여 치료 반응성을 예측할 수 있다.

- [100] 본 발명에서 목적하는 개체의 생물학적 시료에 대하여 측정된 서열번호 1 내지 9로 표시되는 폴리펩타이드 또는 이를 코딩하는 유전자의 발현 수준을 측정하여 상기 개체의 예후, 바람직하게는 외과적 수술 후 예후를 예측할 수 있다. 여기서 상기 목적하는 개체는 암이 발병하여 외과적 절제 수술을 받은 개체일 수 있다.
- [101] 본 발명에서 목적하는 개체의 생물학적 시료에 대하여 측정된 상기 폴리펩타이드 또는 이를 코딩하는 유전자의 발현 수준을 측정하여 상기 개체에서 암의 병기를 예측할 수 있다.
- [102] 본 발명에서 상기 “병기(stage)”란 암세포가 퍼진 정도, 암의 진행 단계를 의미하는 것으로, 암의 진행상황에 따른 국제적 분류는 일반적으로 TNM 병기 분류에 따른다. 여기서 ‘T(Tumor Size)’는 원발 종양의 크기에 따른 분류이고, ‘N(Lymph Node)’은 림프절 전이 정도에 따른 분류이며, ‘M(Metastasis)’은 다른 장기로의 전이 여부에 따른 분류에 해당한다. T, N, M에 있어서 상세 분류는 하기 표 2와 같으며 이에 따른 암의 병기 분류는 하기 표 3과 같다.

[103]

[104] [표2]

TNM 병기		정의
원발 종양의 크기 (T 병기) Size of the primary tumor (T stage)	T0	종양세포의 형태가 악성종양의 모습을 보이거나 발생한 점막 또는 상피에 국한돼 있고 아직 기저막을 침윤하지 않은 종양
	T1	원발된 장기에 제한된 병변, 종양이 가동성이 있으며 인접 및 주위조직에 침범이 없음
	T2	종양의 크기가 2~5cm 정도
	T3	종양의 크기가 T2 보다 크나 장기 내에 국한됨
	T4	주변 조직과 유착 및 침윤한 상태
림프절 전이 여부 (N 병기) Lymph node status (N stage)	N0	림프절 병변의 증거가 없음
	N1	축지되고 가동성이 있으며 첫 번째 위치에 제한되어 있는 림프절(1~2cm 이상, 보통 3cm까지의 크기) 하나에 침범
	N2	축지되고 부분적으로 가동성이 있는 또는 단단하거나 딱딱한 림프절, 현미경적으로 침범의 증거가 있고 임상적으로 서로 엉켜있으며 반대측 또는 양측에서 나타남(3~5cm)
	N3	완전히 고정되어 있고 피막을 통과해 뼈나 큰 혈관, 피부, 신경 등에 완전히 고정되어 있으며 6cm 이상의 크기
원격전이 여부(M 병기) Distant metastasis	M0	원격전이가 없음
	M1	원격전이가 있음

(M stage)		
-----------	--	--

[105]

[106] [표3]

병기분류	T1	T2	T3	T4
N0	1기		2기	
N1	3기			
N2				
N3				
M1	4기			

[107] 본 발명에서 목적하는 개체의 생물학적 시료에 대하여 측정된 상기 폴리펩타이드 또는 이를 코딩하는 유전자의 발현 수준을 측정하여 암의 재발 가능성을 예측할 수 있다.

[108] 본 발명에서 암의 종류에 대해서는 이미 상술하였으므로, 과도한 중복을 피하기 위하여 이하 그 자세한 기재를 생략한다.

[109]

[110] 본 발명의 구체적인 구현예에 따르면,

[111] 상기 APOC1 폴리펩타이드의 일부 절편은 서열번호 1(TPDVSSALDK)의 아미노산 서열을 가지며;

[112] 상기 CHL1 폴리펩타이드의 일부 절편은 서열번호 2(VIAVNEVGR)의 아미노산 서열을 가지고;

[113] 상기 MMP9 폴리펩타이드의 일부 절편은 서열번호 3(AVIDDAFAR)의 아미노산 서열을 가지며;

[114] 상기 PRDX6 폴리펩타이드의 일부 절편은 서열번호 4(LSILYPATTGR)의 아미노산 서열을 가지고;

[115] 상기 PRG4 폴리펩타이드의 일부 절편은 서열번호 5(AIGPSQTHIR)의 아미노산 서열을 가지며;

[116] 상기 PPBP 폴리펩타이드의 일부 절편은 서열번호 6(TTSGIHPK)의 아미노산 서열을 가지고;

[117] 상기 FN1 폴리펩타이드의 일부 절편은 서열번호 7(STTPDITGYR)의 아미노산 서열을 가지며;

[118] 상기 VWF 폴리펩타이드의 일부 절편은 서열번호 8(ILAGPAGDSNVVK)의 아미노산 서열을 가지고;

[119] 상기 CLU 폴리펩타이드의 일부 절편은 서열번호 9(TLLSNLEEAK)의 아미노산 서열을 가진다.

[120]

- [121] 본 발명의 구체적인 구현예에 따르면, 상기 생물학적 시료는 전혈(whole blood), 백혈구(leukocytes), 말초혈액 단핵 세포(peripheral blood mononuclear cells), 백혈구 연층(buffy coat), 혈장(plasma), 혈청(serum), 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) 또는 뇌척수액(cerebrospinal fluid)이다.
- [122]
- [123] 본 발명의 구체적인 구현예에 따르면, 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 상기 폴리펩타이드에 특이적으로 결합하는 항체, 올리고펩타이드, 리간드, PNA(peptide nucleic acid) 및 앵타머(aptamer)로 이루어진 군에서 선택된 1종 이상을 포함한다.
- [124]
- [125] 본 발명의 구체적인 구현예에 따르면, 상기 폴리펩타이드의 발현 수준의 측정은 단백질 칩 분석, 면역측정법, 리간드 바인딩 어세이, MALDI-TOF(Matrix Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry) 분석, SELDI-TOF(Surface Enhanced Laser Desorption/Ionization Time of Flight Mass Spectrometry) 분석, 방사선 면역분석, 방사 면역 확산법, 오우크테로니 면역 확산법, 로케트 면역전기영동, 조직면역 염색, 보체 고정 분석법, 2차원 전기영동 분석, 액상 크로마토그래피-질량분석(liquid chromatography-Mass Spectrometry, LC-MS), LC-MS/MS(liquid chromatography-Mass Spectrometry/Mass Spectrometry), 웨스턴 블랏팅 또는 ELISA(enzyme linked immunosorbent assay)에 의해 수행된다.
- [126]
- [127] 본 발명의 구체적인 구현예에 따르면, 상기 폴리펩타이드의 발현 수준의 측정은 다중 반응 모니터링(multiple reaction monitoring; MRM) 방법에 의한다.
- [128]
- [129] 본 발명의 구체적인 구현예에 따르면,
- [130] 상기 서열번호 1로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z 값이 1일 때 1032.5280 또는 이 값의 ± 1 범위 내의 값이고;
- [131] 서열번호 2로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z 값이 1일 때 966.5524 또는 이 값의 ± 1 범위 내의 값이며;
- [132] 서열번호 3로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z 값이 1일 때 977.5051 또는 이 값의 ± 1 범위 내의 값이고;
- [133] 서열번호 4로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z 값이 1일 때 1191.6732 또는 이 값의 ± 1 범위 내의 값이며;

- [134] 서열번호 5로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 1180.6433 또는 이 값의 ± 1 범위 내의 값이고;
- [135] 서열번호 6로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 840.4574 또는 이 값의 ± 1 범위 내의 값이며;
- [136] 서열번호 7로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 1110.5426 또는 이 값의 ± 1 범위 내의 값이고;
- [137] 서열번호 8로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 1240.6896 또는 이 값의 ± 1 범위 내의 값이며;
- [138] 서열번호 9로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 1117.6099 또는 이 값의 ± 1 범위 내의 값이다.
- [139]
- [140] 본 발명의 구체적인 구현예에 따르면, 상기 다중 반응 모니터링 수행 시 내부 표준 물질로써 상기 각각의 폴리펩타이드를 구성하는 특정 아미노산의 특정 원소를 동위원소로 치환한 합성 펩타이드; 또는 대장균 베타 갈락토시다아제를 사용한다.
- [141]
- [142] 본 발명의 구체적인 구현예에 따르면, 상기 합성 펩타이드는 서열번호 1, 2, 3, 4, 5, 6, 7, 8, 또는 9로 표시되는 서열과 동일 서열을 가지며 안정 동위원소(stable isotope)를 포함한다.
- [143]
- [144] 본 발명의 구체적인 구현예에 따르면, 상기 안정 동위원소는 탄소 및 질소로 이루어진 군으로부터 선택되는 어느 하나 이상의 원소의 안정 동위원소이다.
- [145]
- [146] 본 발명의 구체적인 구현예에 따르면, 상기 폴리펩타이드를 코딩하는 유전자의 발현 수준의 측정은 역전사 중합효소반응(RT-PCR), 경쟁적 역전사 중합효소반응(Competitive RT-PCR), 실시간 역전사 중합효소반응(Real-time RT-PCR), RNase 보호 분석법(RPA; RNase protection assay), 노던 블랏팅(Northern blotting) 또는 DNA 칩에 의한다.
- [147]
- [148] 본 발명의 구체적인 구현예에 따르면,
- [149] 상기 목적하는 개체의 생물학적 시료에 대하여 측정된 상기 CHL1(Neural cell adhesion molecule L1 like), MMP9(Matrix metalloproteinase-9), PRG4(Proteoglycan 4), PPBP(Platelet basic protein), FN1(Fibronectin), VWF(von Willebrand factor) 및 CLU(Clusterin) 폴리펩타이드 또는 이를 코딩하는 유전자의 발현 수준이 정상 대조군에 비하여 증가하고;
- [150] APOC1(Apolipoprotein C1) 및 PRDX6(Peroxiredoxin-6) 폴리펩타이드 또는 이를 코딩하는 유전자의 발현 수준이 정상 대조군에 비하여 감소한 경우, 상기 암의 발병 가능성이 높은 것으로 예측한다.

[151]

[152] 본 발명의 구체적인 구현예에 따르면, 상기 정보 제공 방법은 목적하는 개체의 항암제에 대한 반응성을 예측한다.

[153]

[154] 본 발명의 또 다른 양태에 따르면, 본 발명은 다음의 단계를 포함하는 암의 예방 또는 치료용 조성물의 스크리닝 방법을 제공한다:

[155] (a) 서열번호 1(TPDVSSALDK), 서열번호 2(VIAVNEVGR), 서열번호 3(AVIDDAFAR), 서열번호 4(LSILYPATTGR), 서열번호 5(AIGPSQTHIR), 서열번호 6(TTSGIHPK), 서열번호 7(STTPDITGYR), 서열번호 8(ILAGPAGDSNVVK) 및 서열번호 9(TLLSNLEEAK)로 표시되는 폴리펩타이드로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드, 이들을 인코딩하는 유전자 또는 이들을 발현하는 세포를 포함하는 생물학적 시료에 후보물질을 접촉시키는 단계; 및

[156] (b) 상기 생물학적 시료 내 서열번호 1(TPDVSSALDK), 서열번호 2(VIAVNEVGR), 서열번호 3(AVIDDAFAR), 서열번호 4(LSILYPATTGR), 서열번호 5(AIGPSQTHIR), 서열번호 6(TTSGIHPK), 서열번호 7(STTPDITGYR), 서열번호 8(ILAGPAGDSNVVK) 및 서열번호 9(TLLSNLEEAK)로 표시되는 폴리펩타이드로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드 또는 이들을 인코딩하는 유전자의 발현량을 측정하는 단계;

[157] 상기 목적하는 개체의 생물학적 시료에 대하여 측정된 상기 서열번호 2, 서열번호 3, 서열번호 5, 서열번호 6, 서열번호 7, 서열번호 8 또는 서열번호 9 폴리펩타이드, 또는 이를 코딩하는 유전자의 발현 수준이 정상 대조군에 비하여 감소하고;

[158] 서열번호 1 또는 서열번호 4 폴리펩타이드, 또는 이를 코딩하는 유전자의 발현 수준이 정상 대조군에 비하여 증가한 경우, 상기 후보물질은 암의 예방 또는 치료용 조성물로 판정한다.

[159] 본 발명에서 상기 시료 및 상기 암에 관해서는 이미 상술하였으므로, 과도한 중복을 피하기 위하여 그 자세한 기재를 생략한다.

[160] 본 발명에서 이들 생물학적 시료를 조작하거나 조작하지 않은 상태로 암의 예방 또는 치료용 후보 약제와 반응시킬 수 있다.

[161] 본 발명에서의 용어 “후보 물질”은 본 발명의 유전자를 발현하는 세포를 포함하는 시료에 첨가되어 이들 유전자의 활성 또는 발현량에 영향을 미치는지 여부를 검사하기 위하여 스크리닝에서 이용되는 미지의 물질을 의미한다. 상기 시험 물질은 화합물, 뉴클레오타이드, 펩타이드 및 천연 추출물을 포함하나, 이에 제한되는 것은 아니다. 시험물질을 처리한 생물학적 시료에서 상기 유전자의 발현량 또는 활성을 측정하는 단계는 당업계에 공지된 다양한 발현량 및 활성 측정방법에 의해 수행될 수 있다.

[162]

[163] 본 발명의 또 다른 양태에 따르면, 본 발명은

- [164] 입력값을 입력받는 입력부;
- [165] 유방암의 발병 여부를 판독하도록 사전 학습된 기계학습 모델을 포함하는 판독부; 및
- [166] 유방암의 발병 여부를 출력하는 출력부를 포함하고;
- [167] 상기 입력값은 생물학적 시료 내의 1(TPDVSSALDK), 서열번호 2(VIAVNEVGR), 서열번호 3(AVIDDAFAR), 서열번호 4(LSILYPATTGR), 서열번호 5(AIGPSQTHIR), 서열번호 6(TTSGIHPK), 서열번호 7(STTPDITGYR), 서열번호 8(ILAGPAGDSNVVK) 및 서열번호 9(TLLSNLEEAK)로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드의 발현 수준의 측정값인 것을 특징으로 하는 시스템을 제공한다.
- [168] 본 발명에서의 용어 “기계학습”은 컴퓨터 시스템이 패턴과 추론에 의존하여, 외부에서의 명시적 지시 없이 업무를 수행하는데 사용하는 알고리즘 및 통계 모델을 의미한다. 본 발명에서 기계학습 모델은 구체적으로 딥러닝(Deep learning), 로지스틱 회귀(Logistic regression), 서포트 벡터 머신(Support vector machine, SVM), 랜덤 포레스트(Random forest), 경사 부스팅 알고리즘(Gradient boosting algorithm, GBM)를 사용할 수 있고, 보다 구체적으로 딥러닝(Deep learning)을 사용할 수 있으나, 이에 국한되지 않고 본 발명의 바이오마커의 발현량을 측정한 데이터를 이용하여 유방암을 진단할 수 있는 모든 종류의 기계학습 모델을 사용할 수 있다.
- [169] 본 발명의 구체적인 구현예에 따르면, 상기 기계학습 모델은 딥러닝(Deep Learning) 모델이다.
- [170] 본 발명의 구체적인 구현예에 따르면, 상기 생물학적 시료는 혈액이다.
- [171] 본 발명에서 기계학습 모델의 입력값은 생물학적 시료인 “혈액”에서 측정된 폴리펩타이드의 발현 수준의 측정값의 측정값일 수 있으나, 이에 국한되지 않고 개체로부터 얻어지거나 개체로부터 유래된 임의의 물질, 생물학적 체액, 조직 또는 세포를 의미하는 것으로, 예를 들면, 전혈(whole blood), 백혈구(leukocytes), 말초혈액 단핵 세포(peripheral blood mononuclear cells), 백혈구 연층(buffy coat), 혈장(plasma), 혈청(serum), 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) 또는 뇌척수액(cerebrospinal fluid)을 포함할 수 있지만, 바람직하게는 발병 가능성이 높은 환자의 피부를 절개하지 않고 중공침 등을 생체 내 기관에 자입하여 병리조직학적 검사용으로 채취한 액체 생검(예를 들면, 환자의 조직, 세포, 혈액, 혈청, 혈장, 타액, 객담 또는 복수(ascites) 등)일 수 있다.

- [172] 본 발명의 구체적인 구현예에 따르면, 상기 폴리펩타이드의 발현 수준의 측정값은 질량분석(Mass Spectrometry)에 따른 정량값이다.
- [173] 본 발명에서 폴리펩타이드의 발현 수준을 측정하기 위한 질량분석(Mass Spectrometry)에 대해서는 이미 상술하였으므로, 과도한 중복을 피하기 위하여 이하 그 자세한 기재를 생략한다.
- [174] 본 발명의 구체적인 구현예에 따르면, 상기 질량분석은 액체 크로마토그래피-다중 질량 분석(Liquid Chromatography-Tandem Mass Spectrometry, LC-MS/MS)이다.
- [175] 본 발명의 또 다른 양태에 따르면, 본 발명은 APOC1(Apolipoprotein C1), CHL1(Neural cell adhesion molecule L1 like), MMP9(Matrix metalloproteinase-9), PRDX6(Peroxiredoxin-6), PRG4(Proteoglycan 4), PPBP(Platelet basic protein), FN1(Fibronectin), VWF(von Willebrand factor) 및 CLU(Clusterin)로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드 또는 이들의 일부 절편; 또는 이를 코딩하는 유전자의 발현 수준을 측정하는 제제를 포함하는 조성물을 대상체에 투여하는 단계를 포함하는 암의 진단 방법을 제공한다.

[176]

발명의 효과

- [177] 본 발명의 특징 및 이점을 요약하면 다음과 같다:
- [178] (a) 본 발명은 암에서 특이적으로 발현되는 단백질(또는 이의 절편) 또는 이들을 인코딩하는 유전자의 발현량을 측정하는 제제를 발굴하고, 이를 이용하여 암을 진단하는 방법을 제공한다.
- [179] (b) 본 발명은 혈액 단백질 중 암의 진단에 유용하게 사용할 수 있는 특정 단백질 및 이의 절편을 발굴하여 이를 진단을 위한 바이오마커로 사용함으로써, 특히 유방암을 조기에 간편하고 정확하게 진단할 수 있으며, 이를 통해 관련 질환자의 사망률을 현저히 낮추는데 유용하게 이용될 수 있다.

[180]

도면의 간단한 설명

- [181] 도 1은 혈액 샘플 내의 펩타이드를 분석하여 구성한 라이브러리에 관한 그림이다. 도 1a는 혈액 샘플에 존재하는 펩타이드에 대한 질량분석 데이터로 구성된 라이브러리(PepQuant Library)를 구축 프로세스를 도시한 그림으로서, 샘플에 대한 설명 및 단백질 전처리와 질량분석 과정을 나타내고 있다. 도 1b는 삼중 사중극자(Triple quadrupole(MS/MS)) 장비에서 산출되는 크로마토그램과 표준품에 의한 크로마토그램을 비교하여 정량 가능한 타겟 펩타이드를 탐색하는 과정을 도시한 그림이다. 도 1c는 정량 가능한 펩타이드의 분포를 도시한 그림으로, PepQuant 라이브러리에서 정량 가능한 단백질 중 blood atlas에서 농도가 알려진 124개 혈액 단백질에 대한 로그-스케일 변환 분포를 도시한 그림이다. 도 1d는 도 1c의 혈

- 액 단백질에 대한 히스토그램 분포를 도시한 그림이다. 도 1e는 PepQuant 라이브러리에서 혈청/혈장(serum/plasma)별로 정량 가능한 마커 수를 도시한 그림이다.
- [182] 도 2는 유방암 진단 모델을 개발하기 위한 딥러닝(Deep learning) 모델의 구조를 도시한 그림이다.
- [183] 도 3은 본 발명에서 선정된 바이오마커 조합 및 딥러닝 알고리즘에 의한 진단 정확도를 확인한 실험의 결과를 도시한 그림이다. 도 3a는 본 발명에서 선정된 바이오마커 조합 및 딥러닝 알고리즘에 의한 진단 정확도를 분석한 ROC 곡선을 도시(ROC Curve)한 그림이다. 도 3b는 본 발명에서 선정된 바이오마커 조합 및 딥러닝 알고리즘에 의한 진단 정확도를 유방암의 기수 별로 분석한 결과 및 타암 결과를 표시한 박스 플롯(Box plot)이다.
- [184]
- 발명의 실시를 위한 형태**
- [185] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 이들 실시예는 오로지 본 발명을 보다 구체적으로 설명하기 위한 것으로, 본 발명의 요지에 따라 본 발명의 범위가 이들 실시예에 의해 제한되지 않는다는 것은 당업계에서 통상의 지식을 가진 자에 있어서 자명할 것이다.
- [186]
- [187] **실시예**
- [188] **실험방법**
- [189] *PepQuant library* 구축
- [190] 고농도 단백질 제거 없이 전체 혈청 및 혈장을 활용하여 10분이라는 매우 빠른 분석 시간 내에서 재현 가능하고 정량 가능한 펩타이드의 수를 최대한 확보하기 위해 MRM 라이브러리(PepQuant Library)를 구축하였다.
- [191] 이를 위해 Secretome Database(Science, 2019), Blood atlas DB 및 문헌 연구에서 혈액에 존재하는 것으로 알려진 3,338개의 서로 다른 단백질 목록을 수집했으며, 이 중 4,683개의 대표 펩타이드(surrogate peptide: 특정 단백질에만 존재하는 시퀀스를 갖고 있어, 특정 단백질을 대표할 수 있는 펩타이드를 의미함)를 자체 알고리즘 및 선정의 로직을 구축하여 인 실리코(in silico)로 선정하였다.
- [192] 선정된 4,683개 펩타이드를 모두 합성하였으며, 합성된 개별 펩타이드들은 실제 혈청 샘플에 존재하는 동일한 펩타이드들이 검출되는지 여부를 알기 위한 용도로 사용되었다. 즉, 합성 표준 펩타이드의 머무름 시간 및 단편화 패턴을 먼저 파악하고, 혈액 시료 내에서도 동일한 머무름 시간과 동일한 단편화 패턴이 관찰되면, 이 펩타이드는 혈액 내에서 검출이 된 것으로 판단하였다. MRM 라이브러리(PepQuant 라이브러리)는 3,338개의 단백질에서 생성되어 가장 안정적인 형태라고 판단되는 4,683개의 대리 펩타이드에 대한 포괄적인 LC-MSMS 크로마토그램 데이터베이스에 해당한다. 이 라이브러리에는 4,683개의 표준 펩타이드 크로마토그램과 내인성 펩타이드 크로마토그램이 모두 포함되어 있다. 총 148개의

혈액 샘플을 사용하여 'PepQuant Library'라는 MRM 라이브러리를 구축하였다. 이 샘플에는 6가지 다른 암 유형(유방암 40개, 췌장암 20개, 갑상선암 20개, 난소암 20개, 결장직장암 18개, 질병이 없는 정상 혈액 샘플 30개)이 포함되었다.

[193]

[194] *시료 수집 및 전처리법*

[195] *- 시료수집*

[196] 2019년부터 2020년까지 총 13개 기관에서 혈액 시료를 유방암 및 암 진단 마커 발굴을 위한 IRB를 통해 시료 수집이 진행되었다. 이중, 서울대 병원에서 수집한 암 샘플 40례 및 정상 샘플 30례를 이용하였다. 그 외, 유방암 외 타 암 마커 및 상관성 연구를 위해 2020년 수집한 혈청 시료 췌장암 20례, 갑상선암 20례, 난소암 20례, 폐암 20례, 대장암 18례를 이용하였다.

[197]

[198] *- MRM 분석 전 시료 전처리*

[199] 고농도 단백질(알부민, 트랜스페린, 면역글로불린 등)을 제거하지 않은 혈청/혈장 시료를 사용하였다. 혈청 5 μ l를 요소(urea)와 디티오프레이톨(DTT) 시약이 첨가된 버퍼에 로딩하였다. 이를 통해 시스테인의 이황화 결합을 끊어지면서 단백질 구조가 풀어지게 하였다. 이를 35°C에서 1시간 30분 동안 배양하였다. 배양 후, 실온으로 냉각하고, 요오도아세트아미드(IAA)를 넣고 실온 암실에서 30분 동안 배양하였다. IAA에 의해 이황화 결합의 끊어진 부분이 알킬화시켜 재결합이 없게 하였다. 이후 트립신 소화를 위해 중탄산암모늄(ammonium bicarbonate)버퍼로 희석시키고 트립신은 5 μ g (약 1:50 (w/w)) 로딩하고 37°C에서 16시간 배양하였다. 이후, 10% TFA를 이용하여 트립신 반응을 종결하고, 탈염을 수행하였다. 완전히 말려진 시료는 질량 분석 전까지 -80°C에서 보관하였다. 질량 분석 시 건조된 샘플을 0.1% 포름산(FA)으로 재현탁하였다.

[200]

[201] *- 질량분석*

[202] 질량분석기는 Qtrap 5500 및 Qtrap 5500 plus를 이용했으며(Sciex, USA), 분석 소프트웨어는 Analyst 1.7.2를 이용했고, 정량 프로그램은 Multiquant(3.0.2)를 사용하였다. LC 분리는 C18 역상 컬럼을 사용하였으며, MRM 모드에서 분석하였다.

[203]

[204] *시약*

[205] 본 개발에 사용된 각 바이오마커별 헤비 펩타이드 스탠다드(heavy peptide standard)는 (주)베르티스(Republic of Korea)의 GMP(Good Manufacturing Practice) 시설에서 합성하여 사용하였다. 각 헤비 펩타이드 스탠다드의 저장용 용액(stock solution)은 -80°C의 디프 프리저(deep freezer)에 보관하고, 필요할 때마다 희석하여 사용하였다.

[206]

[207] *- PepQuant library 구축 및 마커 발굴 결과*

[208] 고농도 단백질을 제거하지 않은 상태에서 10분의 빠른 농도구배 조건에서 검출 가능한 펩타이드를 선택하기 위해 'pepQuant-library'라는 MRM 라이브러리를 구축하였다. 이 라이브러리에는 4,683개의 표준 펩타이드 크로마토그램과 내인성 펩타이드 크로마토그램이 모두 데이터베이스화 되어 있다.

[209] 합성 표준품의 머무름 시간과 제2 생성물 패턴과 혈액 내의 펩타이드의 것과 일치하는 것 중 S/N 비율(S/N ratio)이 3 이상이면 이들 펩타이드를 검출 가능한 펩타이드로 정의하였다. 반대로 단편화된 패턴이나 머무름 시간이 일치하지 않으면 대상 펩타이드는 검출되지 않은 것으로 간주하였다. 4,683개의 합성 표준 펩타이드의 크로마토그램 정보를 기반으로 실제 혈액 샘플에 존재하는 452개의 단백질, 852개의 펩타이드를 검출하였다. 검출 가능한 펩타이드 중 농도를 알고 있는 124개 단백질의 분포를 파악한 결과, 40mg/ml 이상에서 1ng/ml 수준까지 광범위한 농도를 함유하고 있는 것을 확인하였다(도 1). 이러한 결과는 유사한 조건의 다른 연구와 비교하였을 때 훨씬 많은 단백질이 검출되었다는 것을 보여준다(표 4).

[210]

[211] [표4]

고농도 단백질 비제거 상태에서 측정 가능한 단백질 수 비교

	검출 (Detection)	단백질 분포 (Protein range)	질량분 석 종류 (Mass type)	구배시간 (Gradient time)	년도 (Year)
AJ Percy et al.	142	31mg~4 4ng/ml	LC-MSMS	45 min	2013
PE Geyer et al.	313	NA	Orbitrap	20 min	2016
A Kraut et al.	280	NA	Orbitrap	60 min	2019
PepQuant-library	452	40mg~1ng /ml	LC-MSMS	10 min	2022

[212]

[213] 알고리즘 개발 방법

[214] - 검체 이용

[215] 본 임상은 유방암 마커의 전향적 다기관 임상시험(KCT 0004847)에서 수집된 혈청 검체를 이용했다. 해당 임상은 13개 병원에서 2019년부터 2020년까지 총 649례의 혈장 및 혈청 검체를 수집하였다. 그러나 해당 임상시험의 목표 중 하나인 혈청 검체에서 추가적인 유방암 마커의 발굴과 개발을 위해 이용할 수 있는 충분한 양이 있는 혈청 검체는 12개 병원 402개의 혈청 검체에 한정되었다. 더불어 모두 성인 여성 검체이다. 해당 임상에서 이용한 혈청 검체는 정상 검체는 유방 영상에서 암이 발견되지 않는 BI-RADS 1 또는 2 범주에 속한다. 또한 5년 이

내에 다른 암이나 재발이 없었다는 조건을 만족해야 한다. 암 샘플은 생검 전에 채취한 유방암 환자의 샘플에 속한다. 각 병원의 검체수는 다음과 같다. 서울대 병원(정상 187명), 분당서울대병원(암 14명), 단국대병원(암 27명), 중앙대병원(암 26명), 한림대학교 강남성심병원(암 13명), 국립암센터(암 22명), 명지병원(암 25명), 한양대학교병원(암 9명), 가톨릭대학교, 서울 ST. 성모병원(암 11명), 고려대 안암병원(암 14명), 고려대 구로병원(암 29명), 경상대학교 병원(암 25명), (표 5).

[216] 수집 검체의 연령대는 20세 이상 여성이며, 40~50대가 전체 여성의 63%를 차지하고 있다. 정상군은 5년간 타 암을 포함하여 어떤 암도 발병하지 않았고, 영상학적으로 유방의 조직검사가 필요하지 않은 경우에 해당한다. 암 환자군은 유방 조직 검사를 통해 병리학적으로 확진된 케이스를 의미한다. 기수별 분포는 기수가 밝혀진 샘플에 대해 TNM병기 0기 9.5%, 1기 38.1%, 2기 35.7%, 3-4기 16.7%로 분포되었다.

[217] 유방암이 아닌 5대 암 시료(갑상선암 20례, 난소암 20례, 췌장암 20례, 폐암 18례, 대장암 20례)는 유방암 마커 발굴 및 검증을 위한 타 암 검증에 대한 후향적 임상 시험(서울대학교 IRB 승인 H-1911-085-1079)에서 98례의 혈청샘플을 수집하였다.

[218]

[219] [표5]

알고리즘 개발세트 : 총 샘플 정보

그룹	나이	수	BMI (AV G.)	농도 수준(Density_grade) (1~4)					TNM 병기 분류				
				1	2	3	4	NA	0	1	2	3-4	NA
정상	20 ~ 39	30	20.8	0	0	4	6	20	-	-	-	-	-
	40 ~ 49	68	23.1	1	0	26	28	13	-	-	-	-	-
	50 ~ 59	62	23.6	0	5	33	17	6	-	-	-	-	-
	60 ~ 69	26	24.3	1	5	17	2	1	-	-	-	-	-
	70 +	5	22.1	0	2	3	0	0	-	-	-	-	-
합계		187		2	12	81	52	40					
암 환자	20 ~ 39	9	21.5	0	2	5	1	0	0	2	5	1	1
	40 ~ 49	66	23.6	8	16	19	4	1	8	16	19	5	18
	50 ~ 59	85	24.0	3	29	18	11	2	3	29	18	14	21
	60 ~ 69	73	24.9	7	22	21	5	0	7	22	21	6	17
	70 +	28	25.1	2	10	10	2	0	2	10	10	2	4

합계	215		3	14	24	7	213		20	80	75	35	61
총계	402		2	91	155	75	43		20	80	75	35	61
			2										

[220]

[221] - 혈액분리, MRM 분석 전 시료 전처리, 질량분석

[222] 본 발명의 PepQuant library 구축에서 사용한 혈액분리, MRM 분석 전 시료 전처리 및 질량분석 방법과 동일한 방법을 사용하였다.

[223]

[224] - 분석 마커

[225] PepQuant-library에 의해 발굴된 최종 마커를 이용하였다.

[226]

[227] - 알고리즘 모델 개발

[228] 정상 샘플 70%와 암 샘플 70%는 랜덤하게 분류하여 알고리즘 모델 개발에 사용하였다. 무작위 샘플링은 총 5회 진행되고, 개발에 사용한 세트에서 사용하지 않은 30%로 검증(test)을 수행한다(표 6). 유방암 진단 모델을 개발하기 위해, 여러 알고리즘을 테스트하였다. 딥러닝(Deep learning), 로지스틱 회귀공식(Logistic regression), 랜덤 포레스트(Random forest), 경사 부스팅 알고리즘(Gradient boosting algorithm)을 사용하였고, 모든 알고리즘은 파이썬(Python)을 이용하였다. 로지스틱 회귀공식, 랜덤 포레스트는 사이킷-런(scikit-learn) 프레임워크에서 기본 파라미터(parameter)를 사용하여 개발하였고, 경사 부스팅 알고리즘은 XGboost 프레임워크의 기본 파라미터를 사용하였다. 딥러닝 모델은 토치(torch) 프레임워크를 사용해 개발하였다.

[229]

[230] [표6]

최적 알고리즘 모델 선정을 위한 입력 데이터(input data) 정보

	Algorithm development (70%)	Test (30%) external validation	Total (100%)
Normal	131	56	187
Age 20-39	20	8	
Age 40-59	92	36	
Age 60-79	19	11	
Age 80+	0	0	
Unknown	0	1	
Breast cancer	150	65	215

Age 20-39	3	3	
Age 40-59	82	35	
Age 60-79	51	19	
Age 80+	4	1	
Unknown	10	7	
Stage 0	14	2	
Stage 1	33	22	
Stage 2	41	16	
Stage 3+	10	3	
unknown	52	22	
Total	281	121	402

[231]

[232] 4가지 딥러닝/기계학습(deep learning/machine learning) 모델 중 딥러닝 모델이 가장 높은 AUC 값을 가진 것으로 조사되었다. 이에, 결론적으로 가장 높은 AUC 값을 보인 딥러닝 모델을 최종 알고리즘 모델로 선택하였다.

[233]

[234] - 딥러닝 모델 구조

[235] 모델(Model)의 구조는 100개의 에스티메이터(estimator)로 구성하였다. 100개의 에스티메이터는 완전 연결 계층(fully-connected layer)를 기반으로 하는 동일한 구조를 가진다. 각 에스티메이터는 총 4개의 계층으로 구성되어 있으며, 각 계층은 완전 연결 계층(fully-connected layer(linear layer)), 배치 정규화(batch normalization), Relu 함수(Relu function), 드롭-아웃(Drop-out)의 순서로 구성하였다. 각 에스티메이터의 마지막 층은 시그모이드 함수(Sigmoid function)을 통해 0과 1 사이의 값, 즉 암(cancer)일 확률을 나타내는 값을 출력한다. 100개 에스티메이터의 출력은 연결 계층(concatenation layer)을 통해 하나의 행렬로 합쳐졌으며, 해당 행렬은 마스킹 계층(masking layer)의 입력으로 사용되었다. 마스킹 계층(Masking layer)의 출력은 100개 에스티메이터 중, 중요 에스티메이터를 결정하는 마스크(mask) 행렬을 출력한다. 해당 마스크는 소프트맥스 함수(Softmax function)를 통해 각 에스티메이터에 대한 가중치를 가지며, 해당 마스크와 연결(concatenated)된 에스티메이터의 출력을 곱하여 최종 출력 값을 산출하였다(도 2).

[236]

[237] - 모델 검증

[238] 모델 훈련과 검증을 위해 사용된 검체는 타 암 샘플까지 포함하여 총 500샘플을 이용하였다. 이 중 암이 아닌 검체는 영상학적으로 유방암 가능성이 없고(BI-

RADS C1,C2), 5년 내 암 발병 경험이 없는 샘플 187개와 유방암이 아닌 타 암 질환자 샘플 98개가 이용되었으며, 유방암 확진자의 수술 전 샘플 215개가 사용되었다(표 7). 정상 샘플 70%와 암 샘플 70%는 랜덤하게 분류하여 알고리즘 모델 개발에 사용되고, 나머지 30%는 모델 검증에 사용하였다.

[239]

[240] [표7]

딥러닝 알고리즘 개발에 이용된 입력 데이터(input data) 정보

	Sample category	Training set (70%)	Test set (30%)	Total (100%)
Healthy controls	Age Total	131	56	187
	Age 20-39	20	8	
	Age 40-59	92	36	
	Age 60-79	19	11	
	Age 80+	0	0	
	Unknown	0	1	
Breast cancer	Age Total	150	65	215
	Age 20-39	3	3	
	Age 40-59	82	35	
	Age 60-79	51	19	
	Age 80+	4	1	
	Unknown	10	7	
	Stage Total	150	65	215
	Stage 0	14	2	
	Stage 1	33	22	
	Stage 2	41	16	
	Stage 3+	10	3	
	unknown	52	22	
Other cancers	Other cancers total	69	29	98
	Ovarian	14	6	
	Pancreas	14	6	
	Thyroid	14	6	

	Colon	14	6	
	lung	13	5	
	Total	350	150	500

[241]

[242] - 유방암 마커 선별

[243] Library에 사용된 시료 중 유방암 시료 50개와 정상 시료 50개를 이용하여 PepQuant library에 올라온 타겟 펩타이드 목록(target peptide list) 내에서 혈청 (Serum)에서 발견 가능한 418개에 대해 LC-MS/MS를 이용하여 MRM 분석(MRM analysis)을 수행했고 여기에서 유방암과 정상 간 농도 수치의 차이가 유의하게 발생하는 단백질을 마커로 선정했다. 정량을 위해 마커별로 합성 동위원소 표준품을 내부표준품(IS, internal standard)으로 이용했다. 분석물(A, analyte)의 농도는 $(A/IS \text{ ratio} \times IS \text{ quantity} = \text{quantification value})$ 으로 나타냈다. 내부표준품은 총 418개를 합성하였으며(Bertis, Korea) 전체 샘플 전처리법은 PepQuant library 구축 시 이용했던 방법과 동일하다.

[244] 일차 후보 마커 선별 기준은 유방암 시료 대 정상 시료의 평균 비율이 1.2배 이상 혹은 0.8배 이하인 마커를 선정했고, 30개의 후보 마커를 도출하였다(표 8). p-value의 도출은 t-test의 양측검정을 이용해 5% 미만으로 선정했으며, P-value 산정은 윌콕슨 순위합 검정(Wilcoxon rank sum test)을 이용하였다.

[245]

[246] [표8]

30개 후보 마커

번호	유전자	식별번호(Accession)	단백질 명칭(Protein name)	sequence
1	FN1	P02751	Fibronectin	STTPDITGYR
2	FN1	P02751	Fibronectin	VDVIPVNLPGEHGQR
3	VWF	P04275	von Willebrand factor	ILAGPAGDSNVVK
4	VWF	P04275	von Willebrand factor	VTVPFIGIGDR
5	MMP9	P14780	Matrix metalloproteinase 9	AVIDDAFAR
6	PRG4	Q92954	Proteoglycan 4	AIGPSQTHIR
7	PRG4	Q92954	Proteoglycan 4	DQYYNIDVPSR
8	THBS1	P07996	Thrombospondin 1	LVPNPDQK
9	APOC1	P02654	Apolipoprotein C1	TPDVSSALDK

10	CHL1	O00533	Neural cell adhesion molecule L1 like	VIADVNEVGR
11	B2M	P61769	Beta-2-microglobulin	IQVYSR
12	LYZ	P61626	Lysozyme C	STDYGIFQINSR
13	CTSD	P07339	Cathepsin D	VSTLPAITLK
14	PPBP	P02775	Platelet basic protein	TTSGIHPK
15	C4BPA	P04003	C4b-binding protein alpha chain	LSLEIEQLELQR
16	HBD	P02042	Hemoglobin subunit delta	TAVNALWGK
17	Igals3bp	Q08380	Galectin-3-binding protein	YSSDYFQAPSDYR
18	MASP1	P48740	Mannan-binding lectin serine protease 1	SDFSNEER
19	APOF	Q13790	Apolipoprotein F	SGVQQLIQYYQDQK
20	CPB2	Q96IY4	Carboxypeptidase B2	YSFTIELR
21	VCAM1	P19320	Vascular cell adhesion protein 1	SIDGAYTIR
22	GPLD1	P80108	Phosphatidylinositol-glycan-specific phospholipase D	GAVYVYFGSK
23	FCGBP	Q9Y6R7	IgGFc-binding protein	GNPAVSYVR
24	LTF	P02788	lactotransferrin	YYGYTGAFR
25	FCN2	Q15485	Ficolin-2	VDGSVDFYR
26	PRDX6	P30041	Peroxiredoxin-6	LSILYPATTGR
27	IGF1	P05019	Insulin-like growth factor1	GFYFNKPTGYGSSSR
28	CLU	P10909	Clusterin	TLLSNLEEAK
29	CHGA	P10645	Chromogranin-A	ILSILR
30	PIGR	P01833	Polymeric immunoglobulin receptor	VYTVDLGR

[247]

[248] 이후 30개 후보 마커를 암 시료 96개, 정상 시료 95개에서 비교 검증하였고, 유방암 시료 대 정상 시료의 평균 비율이 1.2배 이상 또는 0.8배 이하인 마커를 이용하였다(p-value <0.05). 이 과정에서 총 16개 마커가 추려졌으며, 질량분석기에서의 마커별 분석적 성능 테스트를 실시하여 최종 9개 마커를 선정하였다(표

9). 참고로 분석적 성능 시험은 유방암 후보 마커를 대상으로 질량분석기에서의 정량 결과의 신뢰성을 증명하기 위한 검증 시험으로 해당 펩타이드의 정량값을 LC-MS/MS로 측정하였고, 특이성 (선택성), 직선성, 일내 정밀성 / 일간 정밀성 (precision), 안정성 (stability), 매질효과의 항목을 중심으로 마커를 평가하였다.

[249]

[250] [표9]

최종 9종 마커 리스트

번호	유전자	단백질	서열	등록번호 (Accession No.)
1	APOC1	Apolipoprotein C1	TPDVSSALDK	P02654
2	CHL1	Neural cell adhesion molecule L1 like	VIAVNEVGR	O00533
3	MMP9	Matrix metalloproteinase-9	AVIDDAFAR	P14780
4	PRDX6	Peroxiredoxin-6	LSILYPATTGR	P30041
5	PRG4	Proteoglycan 4	AIGPSQHTIR	Q92954
6	PPBP	Platelet basic protein	TTSGIHPK	P02775
7	FN1	Fibronectin	STTPDITGYR	P02751
8	VWF	von Willebrand factor	ILAGPAGDSNVV K	P04275
9	CLU	Clusterin	TLLSNLEEAK	P10909

[251]

[252] 알고리즘 개발 결과

[253] 500개의 샘플 중 350개의 샘플(70%)로 개발된 deep learning 알고리즘에 나머지 150개의 샘플에서 검증할 때 최대 0.9207의 AUC 값이 산출되었으며, 5회 무작위 배정으로 학습 및 검증한 결과의 평균값 또한 0.9105임을 보여주었다(도 3a). 또한 유방암 0~1기에 대한 알고리즘 값이 2기~3기에 대한 값 대비 낮지 않음을 보여주고 있어, 0-1기의 유방암도 높은 정확도로 진단할 수 있음을 확인하였다(도 3b). 더불어 타 암 샘플에 대해서는 유방암이 아닌 상태인 정상으로 판정되어, 알고리즘이 유방암에 보다 특이적일 수 있음을 보여준다.

[254]

[255] 결론

[256] 지금까지 유방암의 조기 진단 시스템은 영상에 의존하는 시스템이어서, 유방 밀도, 기술자의 숙련도, 노후된 기기로 인해 진단 정확도가 떨어질 수 있다는 문제점이 존재하였다. 아울러, 방사선 위험, 불편함, 고통 등이 접근성을 낮추는 요인들도 종래 유방암 조기 진단 방법의 문제점으로 꼽힌다.

- [257] 따라서 고정밀 혈액 검사는 영상 진단의 근본적인 문제를 해결하는 대안이 될 수 있다. 유방암 혈액 검사인 기존 CA15-3 면역분석법은 발병 초기의 정확도가 낮은 단점이 존재한다. 따라서 정확도를 높이기 위해서는 멀티마커 조합으로 구성할 필요가 있었다.
- [258] 본 발명자들은 LC-MS/MS에서 고농도 단백질들을 제거하지 않은 상태에서 452개의 혈액단백질을 10분의 빠른 분석 시간 내 정량할 수 있음을 증명하였다. 이 정량 가능한 마커 중 분석적 성능 평가까지 통과한 최종 9개 마커를 유방암 선별 마커로 선정하였다. 또한, 알고리즘 공식을 개발 및 검증하여 AUC 기준으로 0.9 이상의 높은 정확도를 보여 임상에 직접 적용할 수 있을 정도의 정확도를 가짐을 확인하였다. 나아가, PepQuant 라이브러리는 유방암 마커뿐만 아니라 다른 유형의 암 및 기타 질병 마커 선택에도 적용할 수 있음을 확인하였다.
- [259]
- [260] 이상으로 본 발명의 특정한 부분을 상세히 기술하였는바, 당업계의 통상의 지식을 가진 자에게 있어서 이러한 구체적인 기술은 단지 바람직한 구현예일 뿐이며, 이에 본 발명의 범위가 제한되는 것이 아닌 점은 명백하다. 따라서, 본 발명의 실질적인 범위는 첨부된 청구항과 그의 등가물에 의하여 정의된다고 할 것이다.

청구범위

- [청구항 1] APOC1(Apolipoprotein C1), CHL1(Neural cell adhesion molecule L1 like), MMP9(Matrix metalloproteinase-9), PRDX6(Peroxiredoxin-6), PRG4(Proteoglycan 4), PPBP(Platelet basic protein), FN1(Fibronectin), VWF(von Willebrand factor) 및 CLU(Clusterin)로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드 또는 이들의 일부 절편; 또는 이를 코딩하는 유전자의 발현 수준을 측정하는 제제를 포함하는, 암의 진단용 조성물.
- [청구항 2] 제 1 항에 있어서,
 상기 APOC1 폴리펩타이드의 일부 절편은 서열번호 1(TPDVSSALDK)의 아미노산 서열을 가지며;
 상기 CHL1 폴리펩타이드의 일부 절편은 서열번호 2(VIAVNEVGR)의 아미노산 서열을 가지고;
 상기 MMP9 폴리펩타이드의 일부 절편은 서열번호 3(AVIDDAFAR)의 아미노산 서열을 가지며;
 상기 PRDX6 폴리펩타이드의 일부 절편은 서열번호 4(LSILYPATTGR)의 아미노산 서열을 가지고;
 상기 PRG4 폴리펩타이드의 일부 절편은 서열번호 5(AIGPSQTHIR)의 아미노산 서열을 가지며;
 상기 PPBP 폴리펩타이드의 일부 절편은 서열번호 6(TTSGIHPK)의 아미노산 서열을 가지고;
 상기 FN1 폴리펩타이드의 일부 절편은 서열번호 7(STTPDITGYR)의 아미노산 서열을 가지며;
 상기 VWF 폴리펩타이드의 일부 절편은 서열번호 8(ILAGPAGDSNVVK)의 아미노산 서열을 가지고;
 상기 CLU 폴리펩타이드의 일부 절편은 서열번호 9(TLLSNLEELAK)의 아미노산 서열을 가지는 것을 특징으로 하는, 진단용 조성물.
- [청구항 3] 제 1 항에 있어서, 상기 암은 유방암인 것을 특징으로 하는, 진단용 조성물.
- [청구항 4] 제 1 항에 있어서,
 상기 폴리펩타이드의 발현 수준을 측정하는 제제는 상기 폴리펩타이드 또는 이의 일부 절편에 특이적으로 결합하는 항체, 올리고펩타이드, 리간드, PNA(peptide nucleic acid) 및 앵타머(aptamer)로 이루어진 군에서 선택된 1 종 이상을 포함하는 것을 특징으로 하는, 진단용 조성물.
- [청구항 5] 제 4 항에 있어서,
 상기 폴리펩타이드 또는 이의 일부 절편을 코딩하는 유전자의 발현 수준을 측정하는 제제는 상기 유전자에 특이적으로 결합하는 프라이머, 프로

- 브 및 안티센스 뉴클레오티드로 이루어진 군에서 선택된 1종 이상을 포함하는 것을 특징으로 하는, 진단용 조성물.
- [청구항 6] 제 1 항 내지 제 5 항 중 어느 한 항의 진단용 조성물을 포함하는 진단용 키트.
- [청구항 7] 제 6 항에 있어서,
상기 키트는 RT-PCR 키트, DNA 칩 키트, ELISA 키트, 단백질 칩 키트, 래피드(rapid) 키트 또는 MRM(Multiple reaction monitoring) 키트인 것을 특징으로 하는 진단용 키트.
- [청구항 8] 목적하는 개체로부터 분리된 생물학적 시료에서, APOC1(Apolipoprotein C1), CHL1(Neural cell adhesion molecule L1 like), MMP9(Matrix metalloproteinase-9), PRDX6(Peroxiredoxin-6), PRG4(Proteoglycan 4), PPBP(Platelet basic protein), FN1(Fibronectin), VWF(von Willebrand factor) 및 CLU(Clusterin)로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드 또는 이들의 일부 절편; 또는 이를 코딩하는 유전자의 발현 수준을 측정하는 단계를 포함하는, 암의 진단을 위한 정보 제공 방법.
- [청구항 9] 제 8 항에 있어서,
상기 APOC1 폴리펩타이드의 일부 절편은 서열번호 1(TPDVSSALDK)의 아미노산 서열을 가지며;
상기 CHL1 폴리펩타이드의 일부 절편은 서열번호 2(VIAVNEVGR)의 아미노산 서열을 가지고;
상기 MMP9 폴리펩타이드의 일부 절편은 서열번호 3(AVIDDAFAR)의 아미노산 서열을 가지며;
상기 PRDX6 폴리펩타이드의 일부 절편은 서열번호 4(LSILYPATTGR)의 아미노산 서열을 가지고;
상기 PRG4 폴리펩타이드의 일부 절편은 서열번호 5(AIGPSQTHIR)의 아미노산 서열을 가지며;
상기 PPBP 폴리펩타이드의 일부 절편은 서열번호 6(TTSGIHPK)의 아미노산 서열을 가지고;
상기 FN1 폴리펩타이드의 일부 절편은 서열번호 7(STTPDITGYR)의 아미노산 서열을 가지며;
상기 VWF 폴리펩타이드의 일부 절편은 서열번호 8(ILAGPAGDSNVVK)의 아미노산 서열을 가지고;
상기 CLU 폴리펩타이드의 일부 절편은 서열번호 9(TLLSNLEEAK)의 아미노산 서열을 가지는 것을 특징으로 하는 방법.
- [청구항 10] 제 8 항에 있어서,
상기 생물학적 시료는 전혈(whole blood), 백혈구(leukocytes), 말초혈액 단핵 세포(peripheral blood mononuclear cells), 백혈구 연층(buffy coat), 혈장(plasma), 혈청(serum), 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) 또는 뇌척수액(cerebrospinal fluid)인 것을 특징으로 하는 방법.

- [청구항 11] 제 8 항에 있어서,
상기 폴리펩타이드의 발현 수준을 측정하는 제제는 상기 폴리펩타이드에 특이적으로 결합하는 항체, 올리고펩타이드, 리간드, PNA(peptide nucleic acid) 및 앵타머(aptamer)로 이루어진 군에서 선택된 1종 이상을 포함하는 것을 특징으로 하는 방법.
- [청구항 12] 제 8 항에 있어서,
상기 폴리펩타이드의 발현 수준의 측정은 단백질 칩 분석, 면역측정법, 리간드 바인딩 어세이, MALDI-TOF(Matrix Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry) 분석, SELDI-TOF(Surface Enhanced Laser Desorption/Ionization Time of Flight Mass Spectrometry) 분석, 방사선 면역분석, 방사 면역 확산법, 오우크테로니 면역 확산법, 로케트 면역전기영동, 조직면역 염색, 보체 고정 분석법, 2차원 전기영동 분석, 액상 크로마토그래피-질량분석(liquid chromatography-Mass Spectrometry, LC-MS), LC-MS/MS(liquid chromatography-Mass Spectrometry/ Mass Spectrometry), 웨스턴 블랏팅 또는 ELISA(enzyme linked immunosorbent assay)에 의해 수행되는 것을 특징으로 하는, 방법.
- [청구항 13] 제 8 항에 있어서,
상기 폴리펩타이드의 발현 수준의 측정은 다중 반응 모니터링(multiple reaction monitoring; MRM) 방법에 의하는 것을 특징으로 하는 방법.
- [청구항 14] 제 9 항에 있어서,
상기 서열번호 1로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 1032.5280 또는 이 값의 ± 1 범위 내의 값이고;
서열번호 2로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 966.5524 또는 이 값의 ± 1 범위 내의 값이며;
서열번호 3로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 977.5051 또는 이 값의 ± 1 범위 내의 값이고;
서열번호 4로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 1191.6732 또는 이 값의 ± 1 범위 내의 값이며;
서열번호 5로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 1180.6433 또는 이 값의 ± 1 범위 내의 값이고;

서열번호 6로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 840.4574 또는 이 값의 ± 1 범위 내의 값이며;
 서열번호 7로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 1110.5426 또는 이 값의 ± 1 범위 내의 값이고;
 서열번호 8로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 1240.6896 또는 이 값의 ± 1 범위 내의 값이며;
 서열번호 9로 표시되는 폴리펩타이드의 질량 대 전하비(m/z)는, z값이 1일 때 1117.6099 또는 이 값의 ± 1 범위 내의 값인 것을 특징으로 하는 방법.

- [청구항 15] 제 13 항에 있어서,
 상기 다중 반응 모니터링 수행 시 내부 표준 물질로써 상기 각각의 폴리펩타이드를 구성하는 특정 아미노산의 특정 원소를 동위원소로 치환한 합성 펩타이드; 또는 대장균 베타 갈락토시다아제를 사용하는 것을 특징으로 하는 방법.
- [청구항 16] 제 15 항에 있어서, 상기 합성 펩타이드는 서열번호 1, 2, 3, 4, 5, 6, 7, 8, 또는 9로 표시되는 서열과 동일 서열을 가지며 안정 동위원소(stable isotope)를 포함하는 것을 특징으로 하는 방법.
- [청구항 17] 제 16 항에 있어서, 상기 안정 동위원소는 탄소 및 질소로 이루어진 군으로부터 선택되는 어느 하나 이상의 원소의 안정 동위원소인 것을 특징으로 하는 방법.
- [청구항 18] 제 8 항에 있어서,
 상기 폴리펩타이드를 코딩하는 유전자의 발현 수준의 측정은 역전사 중합효소반응(RT-PCR), 경쟁적 역전사 중합효소반응(Competitive RT-PCR), 실시간 역전사 중합효소반응(Real-time RT-PCR), RNase 보호 분석법(RPA; RNase protection assay), 노던 블랏팅(Northern blotting) 또는 DNA 칩에 의하는 것을 특징으로 하는 방법.
- [청구항 19] 제 8 항에 있어서,
 상기 목적하는 개체의 생물학적 시료에 대하여 측정된 상기 CHL1(Neural cell adhesion molecule L1 like), MMP9(Matrix metalloproteinase-9), PRG4(Proteoglycan 4), PPBP(Platelet basic protein), FN1(Fibronectin), VWF(von Willebrand factor) 또는 CLU(Clusterin) 폴리펩타이드 또는 이를 코딩하는 유전자의 발현 수준이 정상 대조군에 비하여 증가하고;
 APOC1(Apolipoprotein C1) 또는 PRDX6(Peroxiredoxin-6) 폴리펩타이드 또는 이를 코딩하는 유전자의 발현 수준이 정상 대조군에 비하여 감소한 경우, 상기 암의 발병 가능성이 높은 것으로 예측하는 것을 특징으로 하는 방법.
- [청구항 20] 제 8 항에 있어서,
 상기 정보 제공 방법은 목적하는 개체의 항암제에 대한 반응성을 예측하는 것을 특징으로 하는 방법.

- [청구항 21] 다음의 단계를 포함하는 암의 예방 또는 치료용 조성물의 스크리닝 방법:
 (a) 서열번호 1(TPDVSSALDK), 서열번호 2(VIAVNEVGR), 서열번호 3(AVIDDAFAR), 서열번호 4(LSILYPATTGR), 서열번호 5(AIGPSQHTIR), 서열번호 6(TTSGIHPK), 서열번호 7(STTPDITGYR), 서열번호 8(ILAGPAGDSNVVK) 및 서열번호 9(TLLSNLEEAK)로 표시되는 폴리펩타이드로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드, 이들을 인코딩하는 유전자 또는 이들을 발현하는 세포를 포함하는 생물학적 시료에 후보물질을 접촉시키는 단계; 및
 (b) 상기 생물학적 시료 내 서열번호 1(TPDVSSALDK), 서열번호 2(VIAVNEVGR), 서열번호 3(AVIDDAFAR), 서열번호 4(LSILYPATTGR), 서열번호 5(AIGPSQHTIR), 서열번호 6(TTSGIHPK), 서열번호 7(STTPDITGYR), 서열번호 8(ILAGPAGDSNVVK) 및 서열번호 9(TLLSNLEEAK)로 표시되는 폴리펩타이드로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드 또는 이들을 인코딩하는 유전자의 발현량을 측정하는 단계;
 상기 목적하는 개체의 생물학적 시료에 대하여 측정된 상기 서열번호 2, 서열번호 3, 서열번호 5, 서열번호 6, 서열번호 7, 서열번호 8 또는 서열번호 9 폴리펩타이드, 또는 이를 코딩하는 유전자의 발현 수준이 정상 대조군에 비하여 감소하고;
 서열번호 1 또는 서열번호 4 폴리펩타이드, 또는 이를 코딩하는 유전자의 발현 수준이 정상 대조군에 비하여 증가한 경우, 상기 후보물질은 암의 예방 또는 치료용 조성물로 판정한다.
- [청구항 22] 유방암의 진단 시스템으로서,
 입력값을 입력받는 입력부;
 유방암의 발병 여부를 판독하도록 사전 학습된 기계학습 모델을 포함하는 판독부; 및
 유방암의 발병 여부를 출력하는 출력부를 포함하고;
 상기 입력값은 생물학적 시료 내의 서열번호 1(TPDVSSALDK), 서열번호 2(VIAVNEVGR), 서열번호 3(AVIDDAFAR), 서열번호 4(LSILYPATTGR), 서열번호 5(AIGPSQHTIR), 서열번호 6(TTSGIHPK), 서열번호 7(STTPDITGYR), 서열번호 8(ILAGPAGDSNVVK) 및 서열번호 9(TLLSNLEEAK)로 구성된 군으로부터 선택되는 하나 이상의 폴리펩타이드의 발현 수준의 측정값인 것을 특징으로 하는 시스템.
- [청구항 23] 제 22항에 있어서, 상기 기계학습 모델은 딥러닝(Deep Learning) 모델인 것을 특징으로 하는 시스템.
- [청구항 24] 제 22항에 있어서, 상기 생물학적 시료는 혈액인 것을 특징으로 하는 시스템.

- [청구항 25] 제 22항에 있어서, 상기 폴리펩타이드의 발현 수준의 측정값은 질량분석 (Mass Spectrometry)에 따른 정량값인 것을 특징으로 하는 시스템.
- [청구항 26] 제 25항에 있어서, 상기 질량분석은 액체 크로마토그래피-다중 질량 분석 (Liquid Chromatography-Tandem Mass Spectrometry, LC-MS/MS)인 것을 특징으로 하는 시스템.