

【명세서】**【발명의 명칭】**

조기 유방암 예후 예측 진단용 유전자 마커 및 이의 용도

【기술분야】

본 출원은 2013년 4월 18일에 출원된 대한민국 특허출원 제10-2013-0043160호(출원번호)를 우선권으로 주장하고, 상기 명세서 전체는 본 출원의 참고문헌이다.

본 발명은 조기 유방암 예후 예측 진단용 유전자 및 이의 용도에 관한 것으로, 보다 상세하게는 유방암 환자의 예후 예측 진단에 필요한 정보를 제공하기 위한 TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 또는 HLA-DPA1 (major histocompatibility complex, class II, DP alpha 1)의 유방암 예후 예측 진단 유전자 마커 및 이의 용도에 관한 것이다.

【배경기술】

인간 유전체정보가 활발하게 활용되면서 암연구는 유전체 수준에서 메카니즘을 밝히는 방향으로 나아가고 있다. 특히 마이크로어레이를 이용하여 수만 개의 유전자의 발현패턴이나 유전자 개수의 증가 혹은 감소에 대한 정보를 바탕으로 거시적인 관점에서 암세포의 특성을 규명할 수 있게 되었다. 이러한 유전체수준의 정보를 분석하는 것은 유기적이고 복잡한 생명현상을 이해하는데 매우 획기적인 방법으로, 앞으로 더욱더 활성화될 것이다. 특히 암과 같은 복합질병(complex disease)의 경우, 소수의 특정유전자에 대한 분석으로는 편협한 결과를 얻기 쉬우며, 암의 발생 및 발달에 대한 큰 행동패턴을 포착하는 것이 중요하기 때문에 유전체 정보 분석이 반드시 필요하다. 이처럼 암 연구에 기본이 되는 대부분의 유전체 정보는 마이크로어레이와 같은 유전체 칩을 이용하여 생성되는데, 수만 개의 유전자에 대한 정보를 한꺼번에 얻을 수 있는 기술은 날로 진화하고 있으며, 고비용의 단점에도 불구하고 마이크로어레이를 이용한 연구 활동이 활발하게 전개되면서 관련정보의 양도 폭발적으로 증가하고 있다. 2000년도 중반부터 이러한 유전체 정보가 수집되어 데이터베이스화되기 시작하였고, 이렇게 수집된 정보를 이용하여 2차 및 3차 분석을 수행하는 일은 생명현상 연구의 중심점이 되어가고 있다.

일반적인 발현(expression) 유전자 칩의 경우, 약 2만-3만개의 유전자를 나타내는 수만 개의 probe가 심어져 있고, SNP와 같은 정밀한 정보를 측정하는 마이

크로어레이는 백만 개 이상의 probe를 가지고 있는 경우도 있다. 이러한 마이크로어레이는 실험법이 비교적 간단하고 표준화가 되어있으며, 대량의 정보를 짧은 시간에 한꺼번에 얻어 매우 효율적이나, 얻어진 결과를 분석하는 일이 핵심이자 어려운 병목지점이 되었다. 기존의 소수의 유전자들 분석하는 것과는 비교가 되지 않는 수만 개의 유전자에 대한 종합적 분석은, 통계적 분석기술뿐 만 아니라 유전체에 대한 해박한 지식이 뒷받침되어야 비로소 유용한 정보를 캐낼 수 있는 것이다. 뿐만 아니라 대량의 정보를 저장하고 분석을 수행할 수 있는 고성능 전산장비도 필요하며, 관련 전산기술 역시 필수이다. 전통적인 생물학적 연구범위와 실험방법에만 익숙한 연구자가 수행하기 어렵기 때문에, 유전체정보가 엄청난 속도로 증가하더라도 이를 유용하게 활용하지 못하고 있는 것이 국내의 현실이다. 북미나 유럽에 비해 부족한 자본과 연구기술력에 대한 국내 사정을 감안한다면, 공개된 유전체 정보를 적극 활용하는 것이야말로 생물정보학에서 선두 지휘해야 할 부분이다. 특히 암에 대한 연구는 가장 활발하게 유전체 분석을 도입해 왔으며, 관련 정보가 상당한 양으로 축적되어 있다.

유방암은 자가진단이 가능하고 자가진단의 중요성이 많이 홍보되면서 초기에 발견되는 경우가 많다. 이러한 초기 유방암 환자들에 대해 수술 후 항암치료의 여부를 결정하기가 어려웠다. 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 투여량 환자가 겪는 고통이 매우 크고 경제적 지출이 요구되는데, 초기 유방암의 경우, 항암치료가 필요하지 않은 환자가 절반 이상일 것으로 추측된다. 따라서, 초기 유방암의 특성을 분석하여 환자의 예후를 예측하여 불필요한 항암치료를 줄인다면, 환자의 삶의 질에 큰 도움이 될 것이다. 마이크로어레이를 이용하여 유방암 수만 개의 유전자의 발현량에 대한 정보를 한 번에 얻을 수 있게 되면서, 분자수준에서 유방암을 분류하고 암의 발생과 발달에 대한 메커니즘을 밝히고자 하는 연구가 활발하게 수행되고 있다. 초기유방암 환자의 예후를 예측하는 것은 임상에서 중요한 일이고, 마이크로어레이를 이용하여 예후를 예측하는 유전자들 발굴하는 일은 이미 2000년대 초부터 시작되었다. 마이크로어레이를 이용한 연구가 고비용임에도 불구하고, 상당한 수의 유방암조직에 대한 발현 profiles이 생산되었고, 연구자들에게 공개되어왔다. 2002년, 78명의 초기유방암 조직과 10여년 동안 추적된 환자의 생존정보를 분석하여 70개의 예후예측유전자가 발굴된 것을 시작으로 하여, 이후 십여 가지의 예후예측 유전자들이 발표되었고, 그 중 몇 가지는 이미 상용화되어

임상에서 활용되고 있다(Chang, H.Y., et al., Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2(2): p. E7(2004); van de Vijver, M.J., et al., A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347(25):1999-2009(2002); van 't Veer, L.J., et al., Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871): 530-536(2002); Wang, Y., et al., Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365(9460): 671-679(2005); Buyse, M., et al., Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst*, 98(17):1183-92(2006); Paik, S., Development and clinical utility of a 21-gene recurrence score prognostic assay in patients with early breast cancer treated with tamoxifen. *Oncologist* 12(6):631-635(2007); Paik, S., et al., A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351(27) :2817-2826(2004); Sotiriou, C., et al., Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98(4):262-72(2006); Pawitan, Y., et al., Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 7(6):R953-964(2005); Miller, L.D., et al., An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA*, 102(38):13550-13555(2005); Bild, A.H., et al., Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439(7074):353-357(2006); Teschendorff, A.E., et al., A consensus prognostic gene expression classifier for ER positive breast cancer. *Genome Biol* 7(10):R101(2006); Desmedt, C., et al., Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* 13(11): 3207-3214(2007)). 대표적으로 mammaprint(Agendia)와 Oncotype DX(genomic health)가 있으며 임상에서 현재 활용되고 있지만, 여전히 예후에 대한 하나의 참고 자료로서 사용되는 경우가 많은 실정이다 (van de Vijver,

M.J., et al., A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347(25):1999-2009(2002); Paik, S., et al., A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351(27) :2817-2826(2004)).

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.

【발명의 상세한 설명】

【기술적 과제】

이에 본 발명자들은 환자의 암세포를 포함하는 조직의 FFPE 시료를 이용하여 초기 유방암 환자에 대한 예후예측 및 항암치료 여부 등을 진단할 수 있는 유전자 진단 시스템을 개발하기 위하여 예의 연구 노력한 결과, 초기 유방암 조직으로부터 얻은 마이크로어레이 데이터와 임상정보를 수집 및 분석하여 예후예측과 관련된 유전자를 발굴하고, 발굴된 유전자 중 FFPE 시료에 적용시키기 적합한 유전자 및 이의 세트를 선별하여 이의 유용성을 확인함으로써 본 발명을 완성하였다.

따라서, 본 발명의 목적은 유방암 환자를 예후 예측 진단하는 유전자 마커 및 이의 용도를 제공하는 것이다.

본 발명의 다른 목적은 유방암 환자를 예후 예측 진단하는 새로운 방법을 제공하는 것이다.

본 발명의 또다른 목적은 유방암 환자의 예후 예측 진단용의 키트를 제공하는 것이다.

본 발명의 또다른 목적은 유방암 환자의 예후 예측 진단에 필요한 정보를 제공하기 위하여, 환자의 시료로부터 mRNA 분리, 유전자 발현 수준 측정, 이의 표준화 및 예측치의 산출 단계를 포함하는 유방암 예후 예측치를 산출하는 방법을 제공하는 것이다.

【기술적 해결방법】

상기와 같은 목적을 달성하기 위하여, 본 발명은 유방암 환자를 예후 예측 진단하는 유전자 마커 및 이의 용도를 제공한다.

본 발명의 다른 목적을 달성하기 위하여, 본 발명은 유방암 환자를 예후 예측 진단하는 새로운 방법을 제공한다.

본 발명의 또다른 목적을 달성하기 위하여, 본 발명은 유방암 환자의 예후 예측 진단용의 키트를 제공한다.

본 발명의 또다른 목적을 달성하기 위하여, 본 발명은 유방암 환자의 예후 예측 진단에 필요한 정보를 제공하기 위하여, 환자의 시료로부터 mRNA 분리, 유전자 발현 수준 측정, 이의 표준화 및 예측치의 산출 단계를 포함하는 유방암 예후 예측치를 산출하는 방법을 제공한다.

본 발명의 또다른 목적을 달성하기 위하여, 본 발명은 TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 및 HLA-DPA1 (major histocompatibility complex, class II, DP alpha 1)로 이루어진 군에서 선택된 어느 하나의 유전자에 대한 프라이머쌍이며, 상기 프라이머쌍은 PCR 증폭에 의해서 대상유전자를 증폭할 수 있는 것임을 특징으로 하는 프라이머쌍을 제공한다.

본 발명의 또다른 목적을 달성하기 위하여, 본 발명은 TRBC1, BTN3A2 및 HLA-DPA1로 이루어진 군에서 선택된 어느 하나의 유전자에 대한 프라이머쌍이며, 상기 프라이머쌍은 PCR 증폭에 의해서 대상유전자를 증폭할 수 있는 것임을 특징으로 하는 프라이머쌍의 유방암 예후 예측용 제제의 제조를 위한 용도를 제공한다.

다른 정의가 없는 한, 본 명세서에 사용된 모든 기술적 및 과학적 용어는 당업자들에 의해 통상적으로 이해되는 동일한 의미를 가진다. 다음의 참고문헌은 본 발명의 명세서에 사용된 여러 용어들의 일반적인 정의를 갖는 기술(skill)의 하나를 제공한다: Singleton et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY (2ded.1994); THE CAMBRIDGE DICTIONARY OF SCIENCE AND TECHNOLOGY (Walkered.,1988); 및 Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY.

이하 본 발명의 내용을 보다 상세히 설명하기로 한다.

본 발명은 유방암 환자를 예후 예측 진단하는 유전자 마커 및 이의 용도를 제공한다. 보다 구체적으로 본 발명은 유방암, 특히 조기 유방암의 예후 예측 진단을 위해서 TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 또는 HLA-DPA1 (major histocompatibility complex,

class II, DP alpha 1)의 유전자 마커를 제공한다. 아울러, 본 발명은 (a) 시료로부터 mRNA를 분리하는 단계, (b) TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 및 HLA-DPA1 (major histocompatibility complex, class II, DP alpha 1)으로 이루어진 군에서 하나 이상 선택된 유전자의 mRNA 발현 수준을 측정하는 단계, (c) 상기 유전자의 mRNA 발현 수준을 표준화하는 단계 및 (d) 상기 유전자의 과발현은 유방암 예후가 좋은 것으로 판별하는 단계를 포함하는 유방암 예후 예측 방법을 제공한다.

본 발명에서 유전자 마커로 기능하는 것은 TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 또는 HLA-DPA1 (major histocompatibility complex, class II, DP alpha 1)일 수 있다. 이들은 각각 독립적으로 선택되거나 2개의 유전자의 조합 또는 3개의 유전자의 조합에 의해서 조기 유방암의 예후 예측 진단에 이용될 수 있다. 각 유전자는 당업계에서 공지된 각 유전자의 서열 또는 각 유전자의 동의어(synonym)의 서열, 바람직하게는 인간에서 유래된 각 유전자의 서열일 수 있으며, 더 바람직하게는 TRBC1은 Genbank Accession No. BC030533.1, BTN3A2는 Genbank Accession No. NM_007047.3 HLA-DPA1은 Genbank Accession No. NM_001242524.1, NM_001242525.1, NM_033554.3에 기재된 서열일 수 있다. 각 유전자에 대한 동의어 및 그 서열은 Genbank 또는 Swissprot에서 검색할 수 있다.

본 발명에서 유방암은 침윤 유방암, 또는 I기, II기 또는 III기 유방암일 수 있다. 또한 본 발명의 유방암은 에스트로젠 수용체 양성 (Estrogen receptor positive, ER+)일 수 있다.

본 발명에서 "예후(prognosis)"는 질병을 진단하여 판단된 장래의 증세 또는 경과에 대한 전망을 말한다. 암 환자에 있어서 예후는 통상적으로 암 발병 또는 외과적 시술 후 일정기간 내의 전이 여부 또는 생존기간을 뜻한다. 예후의 예측 (또는 예후의 진단)은 특히 초기유방암 환자의 화학치료 여부를 비롯하여 향후 유방암 치료의 방향에 대한 단서를 제시하므로 매우 중요한 임상적 과제이다. 예후 예측은 질환 치료제에 대한 환자의 반응, 치료 경과에 대한 예측도 포함한다.

본 발명에서 시료는 유방암 환자의 유방암 조직일 수 있다. 상기 유방암 조

직에는 일부 정상 세포도 포함되어 있을 수 있으며, 바람직하게는 환자의 암세포를 포함하는 유방암 조직의 포르말린 고정 파라핀 포매 (formalin-fixed paraffin-embedded, FFPE) 시료일 수 있다.

본 발명의 유방암 예후 예측 진단 마커의 검출은 대상 유전자에 대한 PCR (polymerase chain reaction) 증폭을 통해 수행될 수 있다. 본 발명의 대상 유전자의 검출은 바람직하게는 대상 유전자의 발현량의 검출, 더 바람직하게는 대상 유전자의 발현량의 정량적인 검출이다. 발현량의 검출을 위해서 시료 조직내에서의 mRNA 분리 및 mRNA에서의 cDNA 합성과정의 필요할 수 있다. mRNA의 분리를 위해서는 당업계에 공지된 시료에서의 RNA 분리 방법이 이용될 수 있으며, 바람직하게는 시료는 FFPE 시료이므로 FFPE 샘플에 적합한 mRNA의 분리방법일 수 있다. cDNA 합성과정은 mRNA를 주형으로 하여 이루어지는 당업계에 공지된 cDNA 합성 방법이 이용될 수 있다. 바람직하게 본 발명의 유방암 예후 예측 진단 마커의 검출은 FFPE 시료에서의 mRNA 발현의 정량적 검출이므로 FFPE 시료에 대한 mRNA 분리방법 및 RT-qPCR (reverse transcription quantitative polymerase chain reaction) 방법에 의한 검출일 수 있다.

본 발명에서 검출은 mRNA 발현 수준(expression level)을 측정일 수 있다. 발현 수준의 측정은 당업계에 공지된 방법에 따라 수행될 수 있으나, 리포터 형광 염료 및/또는 퀀처 (quencher) 형광 염료로 표지된 프로브를 사용한 광학적 정량 분석 시스템에 의해서 측정될 수 있다. 상기 측정은 상업적으로 판매되는 장비, 예를 들어, ABI PRISM 7700™ Sequence Detection System™, Roche Molecular Biochemicals Lightcycler 및 이에 부속되는 소프트웨어 등의 시스템에 의해서 이루어질 수 있다. 이와 같은 측정 데이터는 측정값 또는 역치 사이클 (Ct 또는 Cp)로서 표현될 수 있다. 측정된 형광값이 처음으로 통계학적으로 유의한 것으로 기록될 때의 지점이 역치 사이클이며, 이는 검출 대상이 PCR 반응의 주형으로써 존재하는 초기값에 반비례하여 나타나므로 역치 사이클 값이 작은 경우 정량적으로 더 많은 검출 대상이 존재하는 것을 나타낸다.

한편, 본 발명은 TRBC1, BTN3A2 및 HLA-DPA1로 이루어진 군에서 선택된 어느 하나의 유전자에 대한 프라이머쌍이며, 상기 프라이머쌍은 PCR 증폭에 의해서 대상 유전자를 증폭할 수 있는 것임을 특징으로 하는 프라이머쌍을 유효성분으로 포함하

는 유방암 예후 예측 진단용 조성물을 제공한다.

본 명세서에서 사용되는 용어 "프라이머"는 올리고뉴클레오타이드를 의미하는 것으로, 핵산쇄(주형)에 상보적인 프라이머 연장 산물의 합성이 유도되는 조건, 즉, 뉴클레오타이드와 DNA 중합효소와 같은 중합제의 존재, 그리고 적합한 온도와 pH의 조건에서 합성의 개시점으로 작용할 수 있다. 바람직하게는, 프라이머는 디옥시리보뉴클레오타이드이며 단일쇄이다. 본 발명에서 이용되는 프라이머는 자연(naturally occurring) dNMP(즉, dAMP, dGMP, dCMP 및 dTMP), 변형 뉴클레오타이드 또는 비-자연 뉴클레오타이드를 포함할 수 있다. 또한, 프라이머는 리보뉴클레오타이드도 포함할 수 있다.

본 발명의 프라이머는 타겟 핵산에 어닐링 되어 주형-의존성 핵산 중합효소에 의해 타겟 핵산에 상보적인 서열을 형성하는 연장 프라이머(extension primer)일 수 있으며, 이는 고정화 프로브가 어닐링 되어 있는 위치까지 연장되어 프로브가 어닐링 되어 있는 부위를 차지한다.

본 발명에서 이용되는 연장 프라이머는 타겟 핵산의 제1위치에 상보적인 혼성화 뉴클레오타이드 서열을 포함한다. 용어 "상보적"은 소정의 어닐링 또는 혼성화 조건하에서 프라이머 또는 프로브가 타겟 핵산 서열에 선택적으로 혼성화할 정도로 충분히 상보적인 것을 의미하며, 실질적으로 상보적(substantially complementary) 및 완전히 상보적(perfectly complementary)인 것을 모두 포괄하는 의미를 가지며, 바람직하게는 완전히 상보적인 것을 의미한다. 본 명세서에서, 프라이머 서열과 관련하여 사용되는 용어, "실질적으로 상보적인 서열"은 완전히 일치되는 서열뿐만 아니라, 특정 서열에 어닐링하여 프라이머 역할을 할 수 있는 범위 내에서, 비교 대상의 서열과 부분적으로 불일치되는 서열도 포함되는 의미이다.

프라이머는, 중합제의 존재 하에서 연장 산물의 합성을 프라이밍시킬 수 있을 정도로 충분히 길어야 한다. 프라이머의 적합한 길이는 다수의 요소, 예컨대, 온도, 응용분야 및 프라이머의 소스(source)에 따라 결정되지만 전형적으로 15-30 뉴클레오타이드이다. 짧은 프라이머 분자는 주형과 충분히 안정된 혼성 복합체를 형성하기 위하여 일반적으로 보다 낮은 온도를 요구한다. 용어 "어닐링" 또는 "프라이밍"은 주형 핵산에 올리고디옥시뉴클레오타이드 또는 핵산이 병치(apposition)되는 것을 의미하며, 상기 병치는 중합효소가 뉴클레오타이드를 중합시켜 주형 핵산 또는 그의 일부분에 상보적인 핵산 분자를 형성하게 한다.

프라이머의 서열은 주형의 일부 서열과 완전하게 상보적인 서열을 가질 필요는 없으며, 주형과 혼성화 되어 프라이머 고유의 작용을 할 수 있는 범위 내에서의 충분한 상보성을 가지면 충분하다. 따라서 본 발명에서 의 프라이머는 주형인 상술한 뉴클레오티드 서열에 완벽하게 상보적인 서열을 가질 필요는 없으며, 이 유전자 서열에 혼성화되어 프라이머 작용을 할 수 있는 범위 내에서 충분한 상보성을 가지면 충분하다. 이러한 프라이머의 디자인은 상술한 뉴클레오티드 서열을 참조하여 당업자에 의해 용이하게 실시할 수 있으며, 예컨대, 프라이머 디자인용 프로그램(예: PRIMER 3 프로그램)을 이용하여 할 수 있다.

본 발명은 본 발명의 프라이머쌍을 포함하는 유방암 예후 예측 진단 키트를 제공한다. 본 발명의 키트는 TRBC1, BTN3A2 및/또는 HLA-DPA1의 PCR에 의한 증폭이 가능한 프라이머쌍 이외에 PCR 반응, 시료에서의 RNA 분리 및 cDNA의 합성에 사용되는 당 업계에 공지된 도구 및/또는 시약을 추가로 포함할 수 있다. 본 발명의 키트는 필요에 따라 각 성분들을 혼합하는데 사용될 튜브, 웰 플레이트 및 사용방법을 기재한 지시자료 등을 추가로 포함할 수 있다.

본 발명은 유방암 환자의 예후 예측 진단에 필요한 정보를 제공하기 위하여, 환자의 시료로부터 하기 단계를 포함하는 유방암 예후 예측치를 산출하는 방법을 제공한다:

- (a) 시료로부터 mRNA를 분리하는 단계,
- (b) TRBC1, BTN3A2 및 HLA-DPA1으로 이루어진 군에서 하나이상 선택된 유전자의 mRNA 발현 수준을 측정하는 단계,
- (c) 상기 유전자의 mRNA 발현 수준을 표준화하는 단계 및
- (d) 표준화된 수치를 미리 정해진 계산식에 대입하여 수치를 계산하는 단계,
- (e) 상기 수치 값의 수준에 따라 유방암 예후가 좋은 것으로 산출하거나, 유방암 예후가 나쁜 것으로 산출하는 단계.

본 발명에서의 검출 대상의 발현 수준은 대상 환자 또는 시료에 따라 전체적인 유전자 발현량 또는 발현 수준에 차이가 있을 수 있으므로 표준화가 필요하다. 표준화는 기본 발현량 또는 발현 수준의 차이를 나타낼 수 있는 유전자의 발현량 또는 발현 수준과의 차이를 통해 이루어지며, 바람직하게는 CTBP1 (C-terminal-

binding protein 1), TBP (TATA-binding protein), HMBS (hydroxymethylbilane synthase), CUL1 (cullin 1) 및 UBQLN1 (Ubiquilin-1)에서 하나 내지 다섯의 유전자의 발현량 (또는 복수의 유전자가 선별된 경우 이들 발현량의 평균)을 측정하여 이에 대한 비를 산출하는 것에 의해서 수행될 수 있다.

한편, 본 발명은 TRBC1, BTN3A2 및 HLA-DPA1로 이루어진 군에서 선택된 어느 하나의 유전자에 대한 프라이머쌍이며, 상기 프라이머쌍은 PCR 증폭에 의해서 대상 유전자를 증폭할 수 있는 것임을 특징으로 하는 프라이머쌍을 이용하여 유방암 환자의 시료로부터 상기 선택된 유전자의 mRNA 발현 수준을 측정하는 단계를 포함하는 유방암 예후 예측 진단 방법을 제공한다.

또한 본 발명은 TRBC1, BTN3A2 및 HLA-DPA1로 이루어진 군에서 선택된 어느 하나의 유전자에 대한 프라이머쌍이며, 상기 프라이머쌍은 PCR 증폭에 의해서 대상 유전자를 증폭할 수 있는 것임을 특징으로 하는 프라이머쌍을 제공한다.

또한 본 발명은 TRBC1, BTN3A2 및 HLA-DPA1로 이루어진 군에서 선택된 어느 하나의 유전자에 대한 프라이머쌍이며, 상기 프라이머쌍은 PCR 증폭에 의해서 대상 유전자를 증폭할 수 있는 것임을 특징으로 하는 프라이머쌍의 유방암 예후 예측용 제제의 제조를 위한 용도를 제공한다.

참고로, 상기에서 언급한 뉴클레오티드 및 단백질 작업에는 다음의 문헌을 참조할 수 있다(Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.(1982); Sambrook et al., Molecular Cloning: A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press(1989); Deutscher, M., Guide to Protein Purification Methods Enzymology, vol. 182. Academic Press. Inc., San Diego, CA(1990); Ausubel et al., Current Protocols of Molecular Biology, John Wiley and Sons (1997); Rupp and Locker, Lab Invest. 56: A67 (1987); De Andres et al., BioTechniques 18: 42044 (1995); Held et al., Genome Research 6:986-994 (1996); T.E. Godfrey et al. J. Molec. Diagnostics 2: 84-91 (2000); K. Specht et al., Am. J. Pathol. 158: 419-29 (2001)).

【유리한 효과】

따라서, 본 발명은 초기 유방암의 예후 예측 진단용의 유전자 마커를 제공한

다. 본 발명의 유전자 마커는 유방암 환자의 예후 및 예측의 진단이 가능하므로 항암치료의 필요성의 판단을 비롯하여 향후 유방암 치료의 방향에 대한 단서를 제시하는 목적으로 유용하게 사용할 수 있다.

【도면의 간단한 설명】

도 1a는 유방암조직의 마이크로레이데이터의 큐레이션(curation) 및 전처리(pre-processing)에 의한 표준화 과정을 나타낸 모식도이다. 도 1b는 디스커버리데이터 세트로부터 예후 예측유전자를 발굴하는 과정을 나타낸 그림이다.

도 2는 예후예측모델(동결 시료)을 디스커버리데이터 세트에서 검증한 결과이다. 9a는 예후예측모델(동결 시료)을 이용한 전체환자의 예후예측지수를 4등분하여 4개의 예후집단으로 분류한 뒤, 각 예후집단의 관찰된 생존확률이 잘 분리되었는지 보는 것이다. 관찰된 생존확률과 예측된 생존확률도 비교하였다. 9b는 전체환자의 관찰된 생존확률과 예후예측모델을 이용하여 예측된 생존확률을 비교한 것이다. 9c는 가장 영향력이 높은 p.mean에 대해 전체 환자를 4개의 집단으로 나누는 뒤 각 집단의 관찰된 생존확률이 예후예측모델(동결 시료)을 이용하여 예측된 생존확률과 잘 일치하는 지 알아본 그림이다. 9d는 5년생존률에 대해 관찰된 생존확률과 예측된 생존확률이 얼마나 잘 일치하는지 알아본 그림이다.

도 3은 예후예측모델(동결 시료)을 검증세트 1에서 검증한 결과이다. 디스커버리데이터 세트에서 검증한 방법과 동일하다. 10a는 판별에 대한 검증결과이고, 10b는 전체 관찰된 시간에 대한 교정에 대한 검증결과이다. 10c는 5년 생존률에 대한 교정에 대한 검증결과이다.

도 4는 예후예측모델(동결 시료)을 검증세트 2에서 검증한 결과이다. 디스커버리데이터 세트에서 검증한 방법과 동일하다. 11a는 판별에 대한 검증결과이고, 11b는 전체 관찰된 시간에 대한 교정에 대한 검증결과이다. 11c는 5년 생존률에 대한 교정에 대한 검증결과이다.

도 5는 예후예측모델(동결 시료)을 검증세트 3에서 검증한 결과이다. 디스커버리데이터 세트에서 검증한 방법과 동일하다.

도 6은 선별된 p-gene에 대한 FFPE 시료 (siemens, 세로축)/동결 시료 (frozen, 가로축)간의 연관성 측정 결과이며, 유전자의 명칭 및 연관성 값 (cor)은 각각 기재된 것과 같다.

도 7은 선별된 i-gene에 대한 FFPE 시료 (siemens, 세로축)/동결 시료 (frozen, 가로축)간의 연관성 측정 결과이며, 유전자의 명칭 및 연관성 값 (cor)은

각각 기재된 것과 같다.

【발명의 실시를 위한 형태】

이하, 본 발명을 실시예에 의해 상세히 설명한다.

단, 하기 실시예는 본 발명을 예시하는 것일 뿐, 본 발명의 내용이 하기 실시예에 한정되는 것은 아니다.

본 명세서의 실시예는 대한민국 특허공개공보 제10-2012-0079295호 및 PCT 공개공보 WO2012093821A2에 개시된 내용이 그 전체로서 본 명세서에 참조로 삽입되어 본 발명이 속하는 기술 분야의 수준 및 본 발명의 내용이 보다 명확하게 설명된다.

<실험방법>

초기유방암조직의 발현 프로파일의 수집

초기 유방암 환자의 냉동 암 조직을 이용하여 얻은 발현 프로파일과 임상정보를 공개 데이터베이스인 GEO(<http://www.ncbi.nlm.nih.gov/geo>)에서 수집하였다. 총 9개의 독립된 발현 프로파일 세트들은 각각 100 개 이상의 샘플로 구성된 비교적 큰 데이터 세트이며 모두 초기유방암환자의 예후와 관련된 연구를 수행하기 위해서 만들어졌다(2, 4, 9, 10, 13, 25, 32, 33). 이중 8개의 데이터 세트는 Affymetrix U133A라는 마이크로어레이 플랫폼으로 만들었고, 나머지 하나만 Agilent Hu25K로 제작하였다. 대부분의 경우 환자의 중요 임상정보(나이, 성별, 암의 크기, 전이상태 및 암의 분화정도)와 생존정보가 함께 수집되어있다. 8개의 Affymetrix U133A로 제작된 데이터 세트들 중에서 6개의 데이터 세트는 생존정보가 외부조직으로의 전이(distant metastasis free survival)에 대한 것이며, 나머지 2개는 생존기간(overall survival)이었다. Agilent 데이터는 외부조직 전이에 대한 생존정보를 가지고 있었다. 외부조직의 전이가 예후 결정에 있어 가장 결정적인 사건인 점, 외부조직 전이는 암의 고유의 특성에 의해 결정된다는 점 및 수집된 데이터에 가장 많은 환자가 외부 조직전이에 대한 정보를 가지고 있다는 점을 기반으로 외부조직 전이여부를 기초로 하여 생존분석을 수행하기로 하였다. 수집된 모든 환자의 정보를 비교하여 중복된 186명의 환자의 발현 프로파일을 제거하였고, 총 1,861명의 유일한(unique) 환자들에 대해서 연구를 수행하였다. 동일한 플랫폼(Affymetrix U133A)으로 제작된 7개의 데이터 세트에 대해, 해당하는 모든 환자의

발현 프로파일의 원본파일(.CEL)을 모아서 한꺼번에 표준화를 시켰다. 표준화 방법은 `rma(background correction : rma, normalization : quantile, summarization : medianpolish)` 방법으로 수행하였다. 표준화 수행시 Manhong Dai 등이 개발한 custom CDF(<http://brainarray.mbni.med.umich.edu/Brainarray/>) ENTREZG version 13를 이용하였다(34). 표준화를 수행한 후, 각 프로브의 발현량은 디스커버리 데이터 세트 내의 프로브 별 평균값을 뺀으로써 1-색(color) 발현량을 2-색 발현량과 같은 형태로 변환시켰다. 총 8개의 표준화된 데이터세트에서 5개의 데이터세트는 하나로 묶어서 디스커버리 데이터 세트로 사용하였고, 2개는 따로 묶어서 검증(validation) 데이터 세트 1로, 나머지 1개는 검증 데이터 세트 2로 이용하였다. Agilent 데이터 세트도 검증 데이터 세트 3으로 사용하였다.

환자의 예후 및 ER 상태에 대한 정의 설정

환자의 예후와 관련된 유전자를 발굴하기 위해서 수집된 환자를 예후가 좋은 집단과 예후가 나쁜 집단으로 분류하였다. 일반적으로 임상에서는 5년 생존 혹은 전이 정보를 이용하여 분류한다. 즉, 5년 내에 전이가 발생하거나 사망을 할 경우 예후가 나쁘다고 말하고, 5년 이상 전이가 없거나 생존하였을 경우 예후가 좋다고 말한다. 디스커버리 데이터세트의 환자정보를 이용하여 전이가 일어난 환자들의 생존시간의 분포를 알아보았다. 전이가 발생한 환자의 73% 이상이 5년 이내에 전이가 발생하였으며, 10년 이후에 전이가 관찰된 경우는 7% 미만이었다. 이를 바탕으로 디스커버리 데이터세트의 환자 중에서 5년 이내에 전이가 발생한 217명의 환자를 '예후가 나쁜 집단'으로 10년 이상 전이가 발생하지 않은 281명의 환자를 '예후가 좋은 집단'으로 분류하였다. 분류 결과, 예후가 좋은 집단의 생존 시간 중앙값은 2.4년이었고, 예후가 나쁜 집단의 생존 시간 중앙값은 12.9년 이었다. 예후가 나쁜 집단과 좋은 집단을 명확하게 구분함으로써 불확실한 생존정보에 의한 오류를 최소한으로 줄일 수 있었다. 에스트로겐 수용체(Estrogen receptor, ER)의 발현 여부는 유방암 환자를 서브타입으로 분류할 때 가장 보편적으로 사용하는 기준이다. 보통 임상에서는 병리학자에 의한 ER IHC(immunohistochemistry)의 판독결과에 의해 ER+ 혹은 ER-로 나눈다. 수집된 디스커버리 데이터 세트에서 200여명의 환자가 ER IHC 정보가 없었고, 디스커버리 데이터 세트를 구성하는 5개의 데이터 세트마다 독립적으로 ER IHC의 결정이 이루어진 점을 고려하여, 환자별 발현 프로파일 내의 ESR1 유전자의 mRNA 발현량을 이용하여 ER 상태를 결정하였다. ER IHC 정보가 있는 환자에 대해, ER IHC 정보와 ESR1 mRNA 발현량을 이용하여 ROC(region of convergence)

분석을 수행하였다. ER IHC 결과와 ESR1 mRNA 발현량을 비교하여 가장 정확도 (0.88)가 높은 발현량 지점을 컷 오프로 잡았고, 컷 오프 이상의 발현량을 보이는 경우는 ER+로, 컷 오프 이하의 발현량을 보이는 경우는 ER-로 분류하였다. 디스커버리 데이터 세트에서 864명을 ER+로, 240명을 ER-로 배치하였다.

예후 예측 유전자의 선택

디스커버리 데이터 세트에서 예후가 좋은 집단과 예후가 나쁜 집단을 ER+, ER-의 경우로 나누었다. 예후가 좋은 환자는 총 275명 이었으며, 예후가 나쁜 환자는 218명 이었다. SAM(Significant Analysis of Microarray) 분석을 통해 예후집단 간 발현량이 차이가 나는 유전자를 알아보았다. SAM 분석결과와 q-값을 이용하여 예후가 좋은 집단에서 과발현된 유전자, 예후가 나쁜 집단에서 과발현된 유전자를 선택하였다. 선택된 유전자를 하나로 합친 결과 총 302개의 중복되지 않는 유전자 세트가 만들어졌고, 이 유전자들의 발현패턴을 알아보기 위한 군집분석을 주성분 분석(Principal Component Analysis, PCA) 방법을 이용하여 수행하였다. 2개의 주성분을 선택하여 각 주성분에 대해, 관련된 생물학적 기능을 알아보기 위하여, 군집별로 GO 기능분석을 수행하였다.

GO 분석결과 주성분 1은 증식에 집중되어 있고 주성분 2는 면역반응에 집중되어 있는 것으로 나타났다. 증식과 면역반응에 관여하는 2개의 주성분에 속하는 유전자를 대상으로 예후 집단 간 발현량이 가장 큰 유전자를 각각 선택하였다. 각 유전자세트는 유전자는 증식의 발현패턴을 대표하는 의미에서 p-gene, 면역반응의 발현패턴을 나타내는 i-gene으로 명명하였다.

모수적 생존분석을 이용한 예후예측모델 구성

모수형 생존모델 중 가속화 고장시간모델(accelerated failure time model, AFT)을 이용하여 p-gene과 i-gene의 발현량을 공변수로 하는 회귀분석을 수행하였다. 4개의 p-gene은 환자별로 평균값을 구하여 p.mean으로 변환하였고, 5개의 i-gene도 역시 환자별로 평균값을 구하여 i.mean으로 변환하여 적용하였다. 가속화 고장시간모델은

$$T_i = T_0 \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_q X_q) \varepsilon_i(1)$$

로서 여기서 T_i 는 i 번째 개체의 생존시간, T_0 는 기저선 생존시간, X_i 는 공변수의 벡터 ($j=1, 2, \dots, q$), β 는 대응하는 공변수의 계수이고 ε 는 오차이다. 이

모델에서는 공변수가 기저선 생존시간에 상승적인 영향을 미치기 때문에 이것을 자주 이용하는 산업계에서 가속화 고장시간 모델이라고 부른다. 생존시간에 상승적으로 작용하는 효과 $\Phi = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_q X_q$ 를 가속요인이라고 칭한다.

식(1)의 자연대수를 얻으면

$$\log T_i = \log T_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_q X_q + \varepsilon^* \quad (2)$$

이 되어 AFT모델은 일반 선형회귀 모델과 동일한 형태를 갖는다. 그러나 종속변수 $\log T$ 는 정규분포를 하지 않을 뿐더러 생존분석 자료에는 선형회귀 모델에서 용납되지 않는 중도절단예가 존재하기 마련이어서 식(2)를 선형회귀모델과 같이 처리할 수 없다. 식(2)의 ε^* 는 일반 선형회귀모델에서 정규분포를 가정하는 것과 달리 데이터세트에 따라 경우마다 분포가 다를 수 있기 때문에 실제적인 통계처리가 번거롭다. 이를 극복하기 위하여 $\log T_0$ 와 ε^* 를 변형하여 다음과 같이 표현한다.

$$\log T_i = \log T_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_q X_q + \sigma W \quad (3)$$

여기서 W 는 $\log T$ 의 분포를 따르며 그 분산은 표준화 분포의 값으로 고정되어 있다. σ 는 척도모수로서 상수인데 그 값은 다루는 데이터 세트에 따라 결정된다.

AFT모델을 이용하여 다양한 후보 예후예측모델에 대해, 와이블(weibull) 분포, 대수로지스틱(loglogistic)분포, 대수정규(lognormal) 분포에 맞추어 보고, 가장 적합한 모델을 선택하였다. AFT모델에 맞출 위험도 분포는 디스커버리 데이터 세트의 생존정보의 세대생명표를 작성하여 얻을 수 있는 해저드함수를 이용하였다. 세대생명표로 얻은 해저드함수는 단봉(unimodal)형태를 보이므로, 와이블, 대수로지스틱, 대수정규 분포가 잘 적합할 것으로 예측되었다. 최종 모델의 선택은 Akaike's information criterion(AIC)과 R square(R^2)를 고려하여 선택되었다.

예후예측모델의 검증

선택된 모델에 대한 검증은 '교정(calibration)'과 '판별(discrimination)'에 대해 수행하였다. '교정'은 만들어진 예후예측모델을 이용하여 예측된 생존확률과 실제 관찰된 생존확률이 얼마만큼 일치하는지를 알아보는 것이고, '판별'은 예후 예측모델에 의해 주어진 환자집단을 예후집단으로 분류하였을 때의 분리성을 알아보는 것이다. 여기서 말하는 실제 관찰되는 생존확률은 Kaplan-Meier 법에 의해 구해진 값을 뜻한다. AFT 기반의 예후예측 모델은 환자별 생존확률을 모든 시간대에 대해 구할 수 있다. 모델에 의해 예측된 생존확률과 Kaplan-Meier법에 의한 생

존확률을 비교하였다. Kaplan-Meier와 같이 전체 시간에 따른 예측 생존확률을 얻기 위해, 전체 환자들의 생존확률곡선은 0yr-25yrs까지 0.1 단위로 구하여 각 시간별 평균생존확률을 계산하여 구하였다. 전체 생존시간에 대한 생존확률비교와 함께, 5년 생존확률도 비교하였다. 주어진 데이터세트에서 환자들의 5년 생존확률을 예후 예측모델을 이용하여 예측한 생존확률을, 해저드회귀분석인 Hare를 이용하여 계산되는 5년 생존확률을 관측값으로 하여 비교하였다.

'판별'은 주어진 데이터세트의 모든 환자의 예후예측지수(prognostic index)를 4구간으로 나눈 뒤, 각 구간에 속하는 환자들의 생존확률을 KM 그래프로 비교하였다. 예후예측지수는 생존모델의 종속변수이다. 4개의 예측된 예후집단에 대한 KM 그래프가 분명하게 나뉠수록 판별의 기능이 좋은 모델이다.

디스커버리 데이터세트와 3개의 독립적인 검증 데이터 세트들에 대해 모두 '교정'과 '판별'을 알아보았다.

통계분석에 사용된 중요 R 패키지들은 다음과 같다:

affy : .CEL 파일에 대해 rma 알고리즘을 이용한 전 처리(pre-processing).

samr : 예후집단간 발현량에 차이가 있는 유전자 발굴.

GOstats : 선택된 유전자세트와 관련된 기능을 알아봄.

KMsurv : 디스커버리 데이터 세트의 생존자료를 이용하여 생명표를 작성함.

rma : AFT 모델을 이용하여 예후예측모델의 계수를 추정함. 모델에 대한교정 수행.

FFPE 샘플에 적용가능한 유전자 세트의 선별

FFPE(formalin-fixed paraffin embedded) 조직, 시료로부터 추출한 RNA는 조직 또는 시료의 처리과정에서 조직간의 cross-linking을 일으키는 고정 등 RNA의 안정성을 저해하는 많은 과정으로 인해 발현분석에 적합한 것으로 받아들여지지 않았다. 본 발명에서는 실제 유방암 치료과정에 근거하여 FFPE 샘플에 적합한 유방암 예후 예측 방법을 개발하기 위하여 상기 주성분 분석 및 주성분 분석에 대한 기능 분석 (GO 기능분석 또는 GO 분석)에 따라 얻어진 증식의 발현패턴을 대표하는 p-gene (증식관련 유전자 세트), 면역반응의 발현패턴을 나타내는 i-gene (면역반응 관련 유전자 세트)에서 각각 가장 기여도가 높은 유전자순으로 IQR(interquartile range)이 높은 것에, 그리고, 평균발현량이 높은 것에 우선순위를 두어 유전자 세트를 선별하였다.

본 발명에서 각각의 p-gene, i-gene의 동일한 패턴의 유전자를 여러 개 선택한 것은 microarray 데이터가 정확한 발현량을 측정하기에 한계가 있고, 패턴을 대표하는 유전자 1개의 발현량보다 패턴내의 여러 개의 유전자의 평균발현량이 실제 발현패턴을 더 대표할 수 있기 때문이다.

FFPE 시료와 Frozen 시료간 유전자간 연관성 (correlation) 측정 및 유전자 선별

같은 환자에서 채취한 FFPE 시료와 동결 (frozen) 시료를 27종을 확보하여 각각에 대해서 FFPE 또는 동결 시료에서의 RNA 추출방법에 따라 RNA를 추출하였다. 추출된 RNA를 주형으로 하여 선별된 32종의 유전자에 대해서 그 발현량을 각각 측정하였다.

유전자 발현량은 개체간에 차이가 있을 수 있으므로 이를 표준화 (normalization)할 필요가 있기 때문에 표준화용 유전자로 선별된 5종의 유전자 즉, CTBP1 (C-terminal-binding protein 1), TBP (TATA-binding protein), HMBS (hydroxymethylbilane synthase), CUL1 (cullin 1) 및 UBQLN1 (Ubiquilin-1)의 발현량으로 표준화하여 FFPE 시료와 동결 시료간의 유전자 발현의 연관성을 측정하였다.

연관성 측정 결과와 각 유전자별 시료에서의 발현량 결과를 바탕으로 연관성 값 (correlation rate)이 높고 샘플 간 유전자 발현 분포가 다양한 유전자를 초기 유방암 예후예측에 신뢰성이 높은 유전자를 최종 선별하였다.

<실험 결과>

예후 예측모델을 위한 예후유전자의 선택

초기 유방암 조직의 발현 프로파일로 이루어진 5개의 데이터 세트를 모두 합쳐서 1,104개 샘플의 디스커버리 데이터 세트를 구성하였다. 모든 환자들은 화학치료를 받지 않았고, 거의 대부분 액와절 전이가 전혀 없거나(N0 or N-) 유방암초기(1기 또는 2기)이다. 이 중, 외부조직 전이에 대한 생존정보를 가지는 1,072명을 대상으로 통계적 분석을 수행하였다. 예후와 관련된 유전자를 찾기 위하여 예후가 좋은 집단(10년이상 전이가 없는 경우)과 예후가 나쁜 집단(5년 이내에 전이가 있

는 경우)의 발현 프로파일로 나누어 비교하였다. 예후가 좋은 집단에서 높은 발현량을 보인 182개의 유전자와 예후가 나쁜 집단에서 높은 발현량을 보인 120개의 유전자를 선택하였다(결과미도시, $FDR < 0.001$).

선택된 302개의 유전자의 발현량에 대해 주성분 분석을 수행하였다. 주성분 1과 주성분 2에 대해 GO 기능분석을 수행하였다. 주성분 1은 매우 뚜렷하게 증식에 관련되어 있었고, 주성분 2는 면역반응과 관련이 강하게 나타났다. 이를 기반으로 주성분 1에 속하는 유전자를 선택하였고, 주성분 2에 속하는 유전자를 선택함으로써 2개의 발현패턴을 예후 예측모델에 반영하도록 하였다.

선택된 9개의 유전자들은 예후와 관련이 있을 뿐만 아니라, 예후집단간 발현 차이가 가장 큰 유전자들로 선택하였다. 증식을 나타내는 주성분 1에서 선택된 유전자는 p-gene로 면역반응을 나타내는 주성분 2에서 선택된 유전자는 i-genes로 명하였다.

ER+ 유방암과 ER- 유방암의 비교

에스트로겐 수용체(estrogen receptor, ER)의 발현유무는 유방암의 발생 및 발달과 밀접한 관련이 있는 것으로 알려져 있다. 예후와 관련하여 선택된 유전자들이 나타내는 2가지 기능, 즉 증식과 면역반응은 암의 메카니즘에 있어 흥미로운 기능이다. 선택된 16개의 유전자들(p-genes 및 i-genes)을 이용하여 ER- 유방암과 ER+ 유방암을 비교하여 보았다. 각 기능의 강도를 나타내기 위하여 p-genes 과 i-genes는 평균 발현량에 따라 3단계 (p1, p2, p3 또는 i1, i2, i3)로 층화하였다. p1은 p-gene의 발현량이 가장 낮은 집단이고 증식이 가장 느릴 것으로 가정하였다. p3는 p-genes의 발현량이 가장 높은 집단이고 증식이 가장 활발하게 일어날 것으로 가정하였다. p2는 중간 발현량을 보이고 중간수준의 증식을 가정하였다. i1은 i-genes가 가장 적게 발현하는 집단이고 약한 면역반응이 있다고 가정하였다. i3는 i-genes가 가장 많이 발현하는 집단이고 매우 강한 면역반응이 있다고 보았다. i2는 중간수준의 발현량과 활동을 보일 것으로 간주하였다.

디스커버리 데이터 세트 내의 1,072 명에 대해 p-gene과 i-gene의 발현량에 따라 분류를 하고 ER 상태별로 각 기능의 강도에 대한 구성을 살펴보았다. ER- 유방암은 ER+ 유방암에 비해 매우 활발히 증식하는 p3 타입의 비율이 매우 높았다. 약 62%의 ER- 유방암이 매우 높은 p-genes 발현량 (p3)을 보인 반면, 18%의 ER+ 유방암만이 높은 p-genes 발현량을 보임으로써, ER- 유방암이 ER+ 유방암보다 훨씬 공격적인 성향을 보인다고 알려진 바와 같았다. 약 35%의 ER+ 유방암이 약한 p-

genes (p1)을 보였고, ER-의 경우는 p1의 비율이 9%밖에 되지 않았다. 활발한 면역 반응 기능은 ER- 유방암의 또 다른 특징으로서 38% 이상의 ER- 유방암은 i-genes (i3)의 발현량이 매우 높았다. 반면 ER+ 유방암은 21% 정도가 높은 i-genes 발현량을 보였다. ER+와 ER- 모두 증식이 활발해질수록 면역반응 역시 활발해지는 것이 관찰되었지만, ER- 유방암이 면역반응을 더욱 적극적으로 보이는 것으로 나타났다.

이 외에, 유방암의 분화 (grade)정도도 증식과 밀접한 관계가 있는 것으로 나타났다. 분화가 잘 안되어 있는 유방암 (G3)일수록 빠른 증식을 보였고, 분화가 잘된 유방암 (G1)은 대부분 약한 증식을 보였다. 환자의 예후도 증식과 상관관계가 있는 것으로 나타났다. 5년 내에 전이가 일어난 예후가 나쁜 환자의 많은 수가 증식이 빠른 집단에 더 많이 몰려있는 것이 관찰되었다.

종합적으로 ER- 유방암은 증식과 면역반응 모두 ER+ 유방암에 비해 매우 활발하였고, ER의 발현량이 유방암의 발생 및 발달의 메카니즘에 영향을 주는 것으로 추측된다.

예후 예측모델의 확립

디스커버리 데이터 세트의 생존정보와 선택된 p-gene과 i-gene을 이용하여 초기 유방암 환자의 전이에 대한 AFT 예후예측모델을 만들었다. 디스커버리 데이터 세트의 생존정보를 이용하여 1년 단위의 세대생명표를 작성하여 대략적인 위험도를 계산하였다.

세대생명표로 얻은 사망확률은 단봉 (unimodal)형태를 보이므로, 와이블, 대수로지스틱, 대수정규 분포가 잘 적합할 것으로 예측되었다. 예후예측모델에 포함될 공변수는 p.mean와 i.mean이다. p.mean은 p-genes의 평균값이면, i.mean은 i-genes의 평균값이다.

3개의 모델에 대해 와이블, 대수로지스틱, 대수정규 분포에 대해 적용을 시킨 결과, 대수정규분포와 가장 잘 적합하였다. AIC(Akaike's information criterion)을 이용하여 대수정규분포에 따른 최종모델을 선택하였다.

$$\log(T) = -0.689 \times p.\text{mean} + 0.274 \times i.\text{mean} + 3.219$$

위의 추정된 모델에 의하면, p.mean, 즉 증식은 생존시간 (T)과 음의 상관관계(-0.689 , $p\text{-값} = 2.47 \times 10^{-17}$)를 가지므로 증식이 활발할수록 생존시간은 짧아지게 된다. 반대로, i.mean은 생존시간과 양의 상관관계 (0.274 , $p\text{-값} = 3.69 \times 10^{-11}$)를 가지는데, 면역반응이 활발할수록 생존시간이 길어지는 것을 뜻한다. 위의 추정된 변수들을 해석하면, 증식이 유방암의 예후에 결정적인 역할을 하며 활발할수록 예

후가 나쁜 반면, 면역반응이 빠른 증식에 대한 방어 메커니즘으로 활동하는 것으로 결론지을 수 있다.

예후예측모델의 검증

디스커버리 데이터세트의 1,072명의 초기유방암 환자의 발현 프로파일을 이용하여 만든 예후 예측모델에 대한 검증은 '교정' 과 '판별'에 대해 수행되었다. '교정'은 모델을 통해 예측된 생존확률이 실제 관찰된 생존확률과 얼마나 비슷한지를 알아보는 것인데, 이때 실제 관찰된 생존확률은 Kaplan-Meier 방법을 이용하여 얻은 생존확률을 말한다. '판별'은 모델을 이용하여 환자를 예후집단으로 잘 분류하는가 이다. 두 가지 성능에 대한 검증은 모델을 개발한 디스커버리 데이터세트와 3개의 독립된 검증 데이터세트에 대해 수행하였다.

예후 예측모델을 개발한 디스커버리 데이터 세트에 대해 예후예측지수 (prognostic index, PI)를 4등분하여 4개의 예후집단으로 분류하였다. 예후예측지수에 의해 분류된 4개의 예후그룹에 대해 관찰된 생존확률인 KM 그래프를 이용하여 비교하였다. 그 결과, 4개의 예후그룹이 매우 잘 분류된 것을 볼 수 있었으며, 각 예후집단의 예측된 생존확률과 관찰된 생존확률이 잘 일치하는 것을 볼 수 있다.

KM 생존확률과 예후예측모형에 의해 예측된 생존확률을 그래프를 이용하여 비교하였다. 예후 예측모델은 모든 환자에 대해 모든 시간별 생존확률을 구하기 때문에, KM 생존곡선과 같이 전체 생존시간에 대한 확률곡선을 얻기 위해, 각 환자들의 시간별(0년-25년, 0.1 간격) 평균 생존확률을 이용하여 생존확률그래프를 그렸다. 예측된 생존확률이 KM에 의한 생존확률보다 약간 높게 나오기는 했지만 전체적으로 비슷하였다. 전체 생존시간에 대한 생존확률 비교 외에 5년차 생존확률에 대해서도 비교하였다. 모델에 의한 5년 생존확률도 실제 관찰된 5년 생존확률과 유사하였고 특히 예측된 5년 생존확률이 높을수록 예측확률과 관찰확률이 잘 일치하였다.

보다 객관적인 검증을 위하여 3개의 독립적인 검증 데이터 세트들을 이용하여 예후예측모델을 검증하였다. 첫 번째 검증 데이터 세트는 Affymetrix U133A 플랫폼으로 생성된 2개의 데이터 세트를 합친 것이다. 두 번째 검증 데이터 세트는 Affymetrix U133A 플랫폼으로 생성된 데이터로서 모두 tamoxifen을 5년간 복용한 ER+ 환자이다. 세 번째 검증 데이터 세트는 70개의 예후 예측유전자(현재 mammaprint로 상용화)의 발굴 및 검증을 위하여 사용된 데이터세트인 Agilent Hu25K 플랫폼으로 생성되었다. 검증 데이터세트 1과 2의 경우 디스커버리 데이터

세트와 같은 Affymetrix U133A 플랫폼으로 제작된 것으로 디스커버리 데이터 세트와 함께 발현량을 표준화하였다. 검증 데이터 세트 1과 2는 교정과 판별의 성능을 평가하였고, 검증 데이터 세트 3은 발현량 표준화 문제로 판별의 성능만 평가하였다.

FFPE 샘플에 적용가능한 유전자 세트의 선별

1,104개 샘플의 디스커버리 데이터 세트 중에서 예후가 좋은 집단에서 높은 발현량을 보인 182개의 유전자와 예후가 나쁜 집단에서 높은 발현량을 보인 120개의 유전자에 대해서 각각 가장 기여도가 높은 유전자순으로 IQR (interquartile range)이 높은 것에, 그리고, 평균발현량이 높은 것에 우선순위를 두어 32종의 유전자를 선별하였다.

FFPE 시료와 Frozen 시료간 유전자간 연관성 (correlation) 측정 및 유전자 선별

FFPE 시료/동결 시료간의 연관성을 측정하기 위하여 환자 또는 암 조직에서 FFPE 시료와 동결 시료가 모두 확보된 샘플이 필요하다. 이와 같이 확보된 27쌍의 FFPE 시료 및 동결 시료에 대해서 선별된 32종의 유전자 각각의 발현량 및 FFPE 시료/동결 시료간의 연관성을 측정한 값을 얻었다. 그 결과 중에서 연관성 값 (correlation rate)가 높고 샘플 간 유전자 발현 분포가 다양한 유전자를 p-gene에서 12종, i-gene에서 15종을 초기유방암 예후예측에 신뢰성이 높은 유전자로 선별하였다.

상기 선별된 유전자 중 예후 진단용 키트에 포함시킬 유전자를 9종의 p-gene과 6종의 i-gene을 선별하였다.

각각의 유전자에 대한 연관성을 측정한 결과는 도 6 및 도 7에서와 같다. 도면에서의 실선은 가로축 및 세로축의 등가값 (기울기 1, 즉 가로축과 세로축의 같은 값을 이어 선으로 표시한 것)을 표시한다.

한편, 상기 각 유전자 중 TRBC1, BTN3A2, HLA-DPA1의 예후가 좋은 집단과 예후가 나쁜 집단에서의 발현 수준에 유의적인 차이가 있었다. 이와 같은 분석을 통해 TRBC1, BTN3A2, HLA-DPA1가 예후가 좋은 집단에서 유의적으로 발현이 증가되어 있음을 알 수 있었고, 아울러, 발현이 증가된 것 (과발현)이 좋은 유방암 예후를 나타냄을 확인할 수 있었다.

【산업상 이용가능성】

이상 살펴본 바와 같이, 본 발명은 초기 유방암의 예후 예측 진단용의 유전자 마커를 제공한다. 본 발명의 유전자 마커는 유방암 환자의 예후의 예측, 진단이 가능하므로 항암치료의 필요성의 판단을 비롯하여 향후 유방암 치료의 방향에 대한 단서를 제시하는 목적으로 유용하게 사용할 수 있다.

【청구의 범위】**【청구항 1】**

하기 단계를 포함하는 유방암 예후 예측 방법:

- (a) 시료로부터 mRNA를 분리하는 단계,
- (b) TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 및 HLA-DPA1 (major histocompatibility complex, class II, DP alpha 1)으로 이루어진 군에서 하나이상 선택된 유전자의 mRNA 발현 수준을 측정하는 단계,
- (c) 상기 유전자의 mRNA 발현 수준을 표준화하는 단계 및
- (d) 상기 유전자의 과발현은 유방암 예후가 좋은 것으로 판별하는 단계.

【청구항 2】

제1항에 있어서, 상기 발현 수준의 측정은 대상 유전자에 대한 PCR 증폭을 통해 수행되는 것을 특징으로 하는 방법.

【청구항 3】

제1항에 있어서, 상기 시료는 환자의 암세포를 포함하는 조직의 포르말린 고정 파라핀 포매 (formalin-fixed paraffin-embedded, FFPE) 시료인 것을 특징으로 하는 방법.

【청구항 4】

제1항에 있어서, 상기 표준화는 CTBP1 (C-terminal-binding protein 1), TBP (TATA-binding protein), HMBS (hydroxymethylbilane synthase), CUL1 (cullin 1) 및 UBQLN1 (Ubiquilin-1)로 이루어진 군에서 하나이상 선택된 표준유전자의 평균 발현량에 대한 비를 산출하는 것에 의해서 수행되는 것을 특징으로 하는 방법.

【청구항 5】

TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 및 HLA-DPA1 (major histocompatibility complex, class II, DP alpha 1)로 이루어진 군에서 선택된 어느 하나의 유전자에 대한 프라이머쌍이며, 상기 프라이머쌍은 PCR 증폭에 의해서 대상유전자를 증폭할 수 있는 것임을 특징으로 하는 프라이머쌍을 유효성분으로 포함하는 유방암 예후 예측 진단용 조성

물.

【청구항 6】

제5항의 조성물을 포함하는 유방암 예후 예측 진단 키트.

【청구항 7】

유방암 환자의 예후 예측 진단에 필요한 정보를 제공하기 위하여, 환자의 시료로부터 하기 단계를 포함하는 유방암 예후 예측치를 산출하는 방법:

- (a) 시료로부터 mRNA를 분리하는 단계,
- (b) TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 및 HLA-DPA1 (major histocompatibility complex, class II, DP alpha 1)으로 이루어진 군에서 하나이상 선택된 유전자의 mRNA 발현 수준을 측정하는 단계,
- (c) 상기 유전자의 mRNA 발현 수준을 표준화하는 단계,
- (d) 표준화된 수치를 미리 정해진 계산식에 대입하여 수치를 계산하는 단계,
- (e) 상기 수치 값의 수준에 따라 유방암 예후가 좋은 것으로 산출하거나, 유방암 예후가 나쁜 것으로 산출하는 단계.

【청구항 8】

제7항에 있어서, 상기 발현 수준의 측정은 대상 유전자에 대한 PCR 증폭을 통해 수행되는 것을 특징으로 하는 방법.

【청구항 9】

제7항에 있어서, 상기 시료는 환자의 암세포를 포함하는 조직의 포르말린 고정 파라핀 포매 (formalin-fixed paraffin-embedded, FFPE) 시료인 것을 특징으로 하는 방법.

【청구항 10】

제7항에 있어서, 상기 표준화는 CTBP1 (C-terminal-binding protein 1), TBP (TATA-binding protein), HMBS (hydroxymethylbilane synthase), CUL1 (cullin 1) 및 UBQLN1 (Ubiquilin-1)로 이루어진 군에서 하나이상 선택된 표준유전자의 평균 발현량에 대한 비를 산출하는 것에 의해서 수행되는 것을 특징으로 하는 방법.

【청구항 11】

TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 및 HLA-DPA1 (major histocompatibility complex, class II, DP alpha 1)로 이루어진 군에서 선택된 어느 하나의 유전자에 대한 프라이머쌍이며, 상기 프라이머쌍은 PCR 증폭에 의해서 대상유전자를 증폭할 수 있는 것임을 특징으로 하는 프라이머쌍을 이용하여 유방암 환자의 시료로부터 상기 선택된 유전자의 mRNA 발현 수준을 측정하는 단계를 포함하는 유방암 예후 예측 진단 방법.

【청구항 12】

TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 및 HLA-DPA1 (major histocompatibility complex, class II, DP alpha 1)로 이루어진 군에서 선택된 어느 하나의 유전자에 대한 프라이머쌍이며, 상기 프라이머쌍은 PCR 증폭에 의해서 대상유전자를 증폭할 수 있는 것임을 특징으로 하는 프라이머쌍.

【청구항 13】

TRBC1 (T cell receptor beta constant 1), BTN3A2 (butyrophilin, subfamily 3, member A2) 및 HLA-DPA1 (major histocompatibility complex, class II, DP alpha 1)로 이루어진 군에서 선택된 어느 하나의 유전자에 대한 프라이머쌍이며, 상기 프라이머쌍은 PCR 증폭에 의해서 대상유전자를 증폭할 수 있는 것임을 특징으로 하는 프라이머쌍의 유방암 예후 예측용 제제의 제조를 위한 용도.