



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0143457
(43) 공개일자 2014년12월16일

- | | |
|---|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>G01N 33/574</i> (2006.01) <i>G01N 33/533</i> (2006.01)
 <i>G01N 33/534</i> (2006.01) <i>C12Q 1/68</i> (2006.01)
 <i>C40B 30/04</i> (2006.01) <i>C40B 40/04</i> (2006.01)</p> <p>(21) 출원번호 10-2014-7031565
 (22) 출원일자(국제) 2013년04월10일
 심사청구일자 없음
 (85) 번역문제출일자 2014년11월10일
 (86) 국제출원번호 PCT/IB2013/052858
 (87) 국제공개번호 WO 2013/153524
 국제공개일자 2013년10월17일
 (30) 우선권주장
 1206323.6 2012년04월10일 영국(GB)</p> | <p>(71) 출원인
 이뮤노비아 에이비
 스웨덴 에스이-223 81 룬드 메디콘 빌리지</p> <p>(72) 발명자
 보레바예크 칼 아르네 크리스테르
 스웨덴 스-223 63 룬드 헬고나베겐 21
 빈그렌 크리스테르 라르스 베르틸
 스웨덴 산드뷔 세-247 32 쇠드라 외스토크라베겐 23</p> <p>(74) 대리인
 리엔목특허법인</p> |
|---|---|

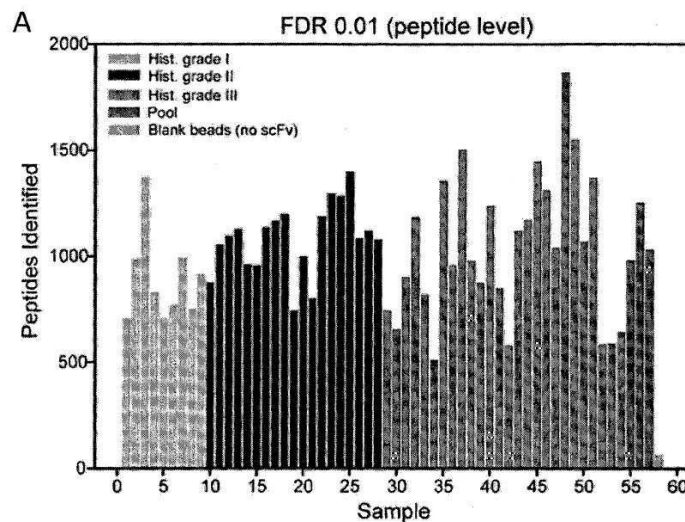
전체 청구항 수 : 총 68 항

(54) 발명의 명칭 **유방암-관련 질환 상태를 결정하는 방법 및 그에 사용하기 위한 어레이**

(57) 요약

본 발명은 a) 시험할 샘플을 제공하는 단계; 및 b) 표 1에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하여 시험 샘플의 바이오마커 시그니처를 결정하는 단계;를 포함하고 상기 표 1에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양은 유방암-관련 질환 상태의 표시인 것인 유방암-관련 질환 상태를 결정하는 방법을 제공한다. 본 발명은 그에 사용하기 위한 어레이 및 키트를 더 제공한다.

대표도 - 도1a



특허청구의 범위

청구항 1

a) 시험될 샘플을 제공하는 단계; 및

b) 표 1A, 표 1B 및/또는 표 1C에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하여 시험 샘플의 바이오마커 시그니처(biomarker signature)를 결정하는 단계;를 포함하고,

상기 표 1A, 표 1B 및/또는 표 1C에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양은 유방암-관련 질환 상태임을 나타내는 것인 유방암-관련 질환 상태를 결정하는 방법.

청구항 2

청구항 1에 있어서, 상기 유방암-관련 질환 상태는 조직학적 등급 및/또는 무전이 생존 시간(metastasis-free survival time)인 것인 방법.

청구항 3

청구항 1 또는 2에 있어서, 유방암-관련 질환 상태는 유방암 세포의 조직학적 등급인 것인 방법.

청구항 4

청구항 3에 있어서, 상기 방법은

c) 조직학적 등급 1 유방암 세포, 조직학적 등급 2 유방암 세포 및/또는 조직학적 등급 3 유방암 세포를 포함하거나 조직학적 등급 1 유방암 세포, 조직학적 등급 2 유방암 세포 및/또는 조직학적 등급 3 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 및

d) 단계 (b)에서 측정된 하나 이상의 바이오마커의 상기 대조군 샘플(들) 중 존재 및/또는 양을 측정하여 상기 대조군 샘플(들)의 바이오마커 시그니처를 결정하는 단계;를 더 포함하고,

단계 (b)에서 측정된 상기 하나 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양이:

i) (존재하는 경우) 제1 조직학적 등급의 유방암 세포를 포함하거나 제1 조직학적 등급의 유방암 세포로 구성된 대조군 샘플 중 상기 존재 및/또는 양에 해당하는 경우;

ii) (존재하는 경우) 제2 조직학적 등급의 유방암 세포를 포함하거나 제2 조직학적 등급의 유방암 세포로 구성된 대조군 샘플 중 상기 존재 및/또는 양과 다른 경우; 및/또는

iii) (존재하는 경우) 제3 조직학적 등급의 유방암 세포를 포함하거나 제3 조직학적 등급의 유방암 세포로 구성된 대조군 샘플 중 상기 존재 및/또는 양과 다른 경우, 유방암 세포의 상기 존재가 확인되는 것인 방법.

청구항 5

청구항 4에 있어서, 각 대조군 샘플은 유방암 세포의 단일 조직학적 등급을 포함하거나 유방암 세포의 단일 조직학적 등급으로 구성된 것인 방법.

청구항 6

청구항 4 또는 5에 있어서, 상기 단계 (c)는

i) 조직학적 등급 1 유방암 세포를 포함하거나 조직학적 등급 1 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 조직학적 등급 2 유방암 세포를 포함하거나 조직학적 등급 2 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 및 조직학적 등급 3 유방암 세포를 포함하거나 조직학적 등급 3 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계;

ii) 조직학적 등급 1 유방암 세포를 포함하거나 조직학적 등급 1 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 및 조직학적 등급 2 유방암 세포를 포함하거나 조직학적 등급 2 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계;

iii) 조직학적 등급 1 유방암 세포를 포함하거나 조직학적 등급 1 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 및 조직학적 등급 3 유방암 세포를 포함하거나 조직학적 등급 3 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계;

iv) 조직학적 등급 2 유방암 세포를 포함하거나 조직학적 등급 2 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 및 조직학적 등급 3 유방암 세포를 포함하거나 조직학적 등급 3 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계;

v) 조직학적 등급 1 유방암 세포를 포함하거나 조직학적 등급 1 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계;

vi) 조직학적 등급 2 유방암 세포를 포함하거나 조직학적 등급 2 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 또는

vii) 조직학적 등급 3 유방암 세포를 포함하거나 조직학적 등급 3 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계를 포함하거나 이들 단계로 구성된 것인 방법.

청구항 7

청구항 1 또는 2에 있어서, 상기 유방암-관련 질환 상태는 개체의 무전이 생존 시간 (metastasis-free survival time)인 것인 방법.

청구항 8

청구항 7에 있어서, 상기 방법은

c) 10 년 무전이 생존 미만을 갖는 개체로부터 유래된 유방암 세포를 포함하거나 상기 유방암 세포로 구성된 하나 이상의 제1 대조군 샘플 및/또는 10년 이상의 무전이 생존을 갖는 개체로부터 유래된 유방암 세포를 포함하거나 상기 유방암 세포로 구성된 하나 이상의 제2 대조군 샘플을 제공하는 단계; 및

d) 단계 (b)에서 측정된 상기 하나 이상의 바이오마커의 상기 대조군 샘플(들) 중 상기 존재 및/또는 양을 측정하여 상기 대조군 샘플(들)의 바이오마커 시그니처를 결정하는 단계를 더 포함하고,

상기 개체의 무전이 생존 시간은 단계 (b)에서 측정된 상기 하나 이상의 바이오마커의 상기 존재 및/또는 양이 (존재하는 경우) 제1 대조군 샘플의 상기 존재 및/또는 양에 해당하는 경우 및/또는 (존재하는 경우) 제2 대조군 샘플의 상기 존재 및/또는 양과 다른 경우, 10년 미만인 것으로서 확인되고; 및

상기 개체의 무전이 생존 시간은 단계 (b)에서 측정된 상기 하나 이상의 바이오마커의 상기 존재 및/또는 양이 (존재하는 경우) 제1 대조군 샘플의 상기 존재 및/또는 양과 다른 경우 및/또는 (존재하는 경우) 제2 대조군 샘플의 상기 존재 및/또는 양에 해당하는 경우, 10년 이상인 것으로서 확인되는 것인 방법.

청구항 9

청구항 8에 있어서, 상기 하나 이상의 제1 및/또는 제2 대조군 샘플은 상기 시험될 샘플과 동일한 조직학적 등급인 것인 방법.

청구항 10

청구항 3 내지 6 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1A에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1A에 정의된 군으로부터 선택된 2개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 11

청구항 3 내지 6 및 10 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1B에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1B에 정의된 군으로부터 선택된 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 12

청구항 3 내지 6, 10 및 11 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1C에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1C에 정의된 군으로부터 선택된 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 또는 28개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 13

청구항 3 내지 6 및 10 내지 12 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1D에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1D에 정의된 군으로부터 선택된 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 14

청구항 3 내지 6 및 10 내지 13 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1E에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1E에 정의된 군으로부터 선택된 2, 3, 4, 5, 6, 7, 8 또는 9개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 15

청구항 3 내지 6 및 10 내지 14 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1에 정의된 모든 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 16

청구항 7 내지 9 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1A에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1A에 정의된 군으로부터 선택된 2개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 17

청구항 7 내지 9 및 16 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1B에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1B에 정의된 군으로부터 선택된 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 18

청구항 7 내지 9, 16 및 17 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1D에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1D에 정의된 군으로부터 선택된 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 19

청구항 7 내지 9, 16 내지 18 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1A, 표 1B 및 표 1D에 정의된 모든 것의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 20

청구항 7 내지 9, 16 내지 19 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1C에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1C에 정의된 군으로부터 선택된 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27 또는 28개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 21

청구항 7 내지 9, 16 내지 20 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1E에 정의된 군으로부터 선택된 하

나 이상의 바이오마커, 예를 들면 표 1E에 정의된 군으로부터 선택된 2, 3, 4, 5, 6, 7, 8 또는 9개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 22

청구항 7 내지 9, 16 내지 21 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1C 및 표 1E에 정의된 상기 바이오마커의 모든 것의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 23

청구항 7 내지 9, 16 내지 22 중 어느 한 항에 있어서, 상기 단계 (b)는 표 1에 정의된 상기 바이오마커의 모든 것의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 상기 단계로 구성된 것인 방법.

청구항 24

선행하는 청구항 중 어느 한 항에 있어서, 상기 단계 (b)는 하나 이상의 바이오마커(들)을 코딩하는 핵산 분자의 발현을 측정하는 단계를 포함하는 것인 방법.

청구항 25

청구항 24에 있어서, 상기 핵산 분자는 cDNA 분자 또는 mRNA 분자인 것인 방법.

청구항 26

청구항 25에 있어서, 상기 핵산 분자는 mRNA 분자인 것인 방법.

청구항 27

청구항 25 또는 26에 있어서, 상기 단계 (b) 중 하나 이상의 바이오마커(들)의 발현을 측정하는 단계는 서던 혼성화(Southern hybridisation), 노던 혼성화(Northern hybridisation), 중합효소 연쇄 반응 (PCR), 역전사효소 PCR (RT-PCR), 정량적 실시간 PCR (qRT-PCR), 나노어레이, 마이크로어레이, 마크로어레이, 자기방사선화법 (autoradiography) 및 인 situ 혼성화(*in situ* hybridisation)로 구성된 군으로부터 선택된 방법을 사용하여 수행되는 것인 방법.

청구항 28

청구항 25 내지 27 중 어느 한 항에 있어서, 상기 단계 (b) 중 하나 이상의 바이오마커(들)의 발현을 측정하는 단계는 DNA 마이크로어레이를 사용하여 결정되는 것인 방법.

청구항 29

선행하는 청구항 중 어느 한 항에 있어서, 상기 단계 (b) 중 하나 이상의 바이오마커(들)의 발현을 측정하는 단계는 표 1에 확인된 상기 바이오마커들 중 하나를 코딩하는 핵산 분자에 선택적으로 각각 결합할 수 있는, 하나 이상의 결합 모이어티를 사용하여 수행되는 것인 방법.

청구항 30

청구항 29에 있어서, 상기 하나 이상의 결합 모이어티는 각각 핵산 분자를 포함하거나 상기 핵산 분자로 구성된 것인 방법.

청구항 31

청구항 30에 있어서, 상기 하나 이상의 결합 모이어티는 각각 DNA, RNA, PNA, LNA, GNA, TNA 또는 PMO를 포함하거나 이로 구성된 것인 방법.

청구항 32

청구항 30 또는 31에 있어서, 상기 하나 이상의 결합 모이어티는 각각 DNA를 포함하거나 이로 구성된 것인

방법.

청구항 33

청구항 30 내지 32 중 어느 한 항에 있어서, 상기 하나 이상의 결합 모이어티는 길이가 5 내지 100개 뉴클레오타이드인 것인 방법.

청구항 34

청구항 30 내지 33 중 어느 한 항에 있어서, 상기 하나 이상의 핵산 분자는 길이가 15 내지 35개 뉴클레오타이드인 것인 방법.

청구항 35

청구항 30 내지 34 중 어느 한 항에 있어서, 상기 결합 모이어티는 검출가능한 모이어티를 포함하는 것인 방법.

청구항 36

청구항 1 내지 23 중 어느 한 항에 있어서, 상기 단계 (b)는 상기 하나 이상의 바이오마커(들)의 단백질 또는 폴리펩티드의 발현을 측정하는 단계를 포함하는 것인 방법.

청구항 37

청구항 36에 있어서, 단계 (b) 중 상기 하나 이상의 바이오마커(들)의 발현을 측정하는 단계는 표 1에서 확인된 하나 이상의 바이오마커에 각각 선택적으로 결합할 수 있는 하나 이상의 결합 모이어티를 사용하여 수행되는 것인 방법.

청구항 38

청구항 37에 있어서, 상기 하나 이상의 결합 모이어티는 항체 또는 그의 항원-결합 단편을 포함하거나 이로 구성된 것인 방법.

청구항 39

청구항 38에 있어서, 상기 항체 또는 그의 단편은 단일클론 항체 또는 그의 단편인 것인 방법.

청구항 40

청구항 38 또는 39에 있어서, 상기 항체 또는 그의 항원-결합 단편은 온전한 항체, Fv 단편 (예, 단쇄 Fv 및 이항화-결합 Fv), Fab-유사 단편 (예, Fab 단편, Fab' 단편 및 F(ab)₂ 단편), 단일 가변 도메인 (예, V_H 및 V_L 도메인) 및 도메인 항체 (dAb, 단일 및 이중 포맷 [즉, dAb-링커-dAb] 포함)으로 구성된 군으로부터 선택되는 것인 방법.

청구항 41

청구항 40에 있어서, 상기 항체 또는 항원-결합 단편은 단쇄 Fv(scFv)인 것인 방법.

청구항 42

청구항 41에 있어서, 상기 하나 이상의 결합 모이어티는 항체-유사 결합제, 예를 들면 아피바디(affibody) 또는 압타머를 포함하거나 이로 구성되는 것인 방법.

청구항 43

청구항 37 내지 42 중 어느 한 항에 있어서, 상기 하나 이상의 결합 모이어티는 검출가능한 모이어티를 포함하는 것인 방법.

청구항 44

청구항 35 내지 43 중 어느 한 항에 있어서, 상기 검출가능한 모이어티는 형광성 모이어티, 발광성 모이어티, 화학발광성 모이어티, 방사성 모이어티 및 효소 모이어티로 구성된 군으로부터 선택되는 것인 방법.

청구항 45

청구항 44에 있어서, 상기 검출가능한 모이어티는 방사성 원자를 포함하거나 이로 구성된 것인 방법.

청구항 46

청구항 45에 있어서, 상기 방사성 원자는 테크네튬-99m, 요오드-123, 요오드-125, 요오드-131, 인듐-111, 플루오린-19, 탄소-13, 질소-15, 산소-17, 인-32, 황-35, 중수소, 트리튬, 레늄-186, 레늄-188 및 이트륨-90으로 구성된 군으로부터 선택되는 것인 방법.

청구항 47

청구항 45에 있어서, 상기 결합 모이어티의 상기 검출가능한 모이어티는 형광성 모이어티인 것인 방법.

청구항 48

선행하는 청구항 중 어느 한 항에 있어서, 단계 (a) 및/또는 단계 (c)에서 제공되는 상기 샘플을 각각 단계 (b) 및/또는 단계 (d) 이전에 처리되어, 상기 샘플에 존재하는 바이오마커를 비오틴으로 표지하고, 상기 단계 (b) 및/또는 단계 (d)는 형광성 검출가능한 모이어티 및 스트렙타비딘을 포함한 검출제를 사용하여 수행하는 것인 방법.

청구항 49

선행하는 청구항 중 어느 한 항에 있어서, ROC AUC 값에 의해 결정되는 상기 방법의 예측 정확도(predicative accuracy)는 0.50, 예를 들면 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 0.96, 0.97, 0.98 또는 0.99이상인 것인 방법.

청구항 50

청구항 49에 있어서, ROC AUC 값에 의해 결정되는 상기 방법의 예측 정확도는 0.80 이상인 것인 방법.

청구항 51

선행하는 청구항 중 어느 한 항에 있어서, 상기 단계 (b)는 어레이를 사용하여 수행되는 것인 방법.

청구항 52

청구항 51에 있어서, 상기 어레이는 비드-기반 어레이인 방법.

청구항 53

청구항 51에 있어서, 상기 어레이는 표면-기반 어레이인 것인 방법.

청구항 54

청구항 51 내지 53 중 어느 한 항에 있어서, 상기 어레이는 마이크로어레이, 마이크로어레이, 나노어레이로 구성된 군으로부터 선택되는 것인 방법.

청구항 55

선행하는 청구항 중 어느 한 항에 따른 방법에 사용하기 위한 어레이로서, 상기 어레이는 청구항 29 내지 35 및 37 내지 48 중 어느 한 항에 정의된 하나 이상의 제1 결합제를 포함하는 것인 어레이.

청구항 56

청구항 55에 있어서, 표 1A에 정의된 바이오마커의 모든 것에 총괄적으로 결합할 수 있는 결합제를 포함하는 것인 어레이.

청구항 57

청구항 55 또는 56에 있어서, 표 1B에 정의된 바이오마커의 모든 것에 총괄적으로 결합할 수 있는 결합제를 포

합하는 것인 어레이.

청구항 58

청구항 55 내지 57 중 어느 한 항에 있어서, 표 1C에 정의된 바이오마커의 모든 것에 총괄적으로 결합할 수 있는 결합제를 포함하는 것인 어레이.

청구항 59

청구항 55 내지 58 중 어느 한 항에 있어서, 표 1D에 정의된 바이오마커의 모든 것에 총괄적으로 결합할 수 있는 결합제를 포함하는 것인 어레이.

청구항 60

청구항 55 내지 59 중 어느 한 항에 있어서, 표 1E에 정의된 바이오마커의 모든 것에 총괄적으로 결합할 수 있는 결합제를 포함하는 것인 어레이.

청구항 61

청구항 55 내지 60 중 어느 한 항에 있어서, 표 1에 정의된 바이오마커의 모든 것에 총괄적으로 결합할 수 있는 결합제를 포함하는 것인 어레이.

청구항 62

청구항 55 내지 61 중 어느 한 항에 있어서, 상기 제1 결합제는 고정된 것인 방법.

청구항 63

표 1A, 표 1B 및/또는 표 1C에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 유방암-관련 질환 상태를 결정하기 위한 용도.

청구항 64

청구항 63에 있어서, 표 1A, 표 1B, 표 1C, 표 1D 및 표 1E에 정의된 바이오마커의 모든 것은 유방암-관련 질환 상태를 결정하기 위해 총괄적으로 사용되는 것인 용도.

청구항 65

C) 청구항 55 내지 62 중 어느 한 항에 따른 어레이; 및

D) (선택적) 청구항 1 내지 55 중 어느 한 항에서 정의된 방법을 수행하기 위한 지시(instruction)를 포함하는 청구항 1 내지 54 중 어느 한 항에 따른 방법에 사용하기 위한 분석 키트.

청구항 66

청구항 65에 있어서, 하나 이상의 대조군 샘플을 더 포함하는 것인 분석 키트.

청구항 67

실질적으로 본 명세서에 기재된 방법 또는 용도.

청구항 68

실질적으로 본 명세서에 기재된 어레이 또는 키트.

명세서

기술분야

본 발명은 유방암-관련 질환 상태를 결정하는 방법, 및 이러한 방법에 사용하기 위한 어레이 및 키트를 제공한다.

[0001]

배경 기술

[0002] 유방암은 가장 자주 진단되는 암이고 여성 중에서 암 사망의 가장 큰 원인이고, 전체 암 사례의 23% 및 암 관련 사망의 14%를 차지한다 (Jemal et al., 2011). 조직학적 등급화(histological grading), 종양 크기, 연령, 림프관 침범(lymph node involvement), 및 호르몬 수용체 상태와 같은 전통적 임상 병리학 파라미터가 치료 및 예후 추정을 결정하기 위해 사용된다 (Ciocca and Elledge, 2000; Elston and Ellis, 1991; Hondermarck et al., 2008; Hudis, 2007; Slamon et al., 2001). 가장 통상적으로 사용되는 예후 인자(prognostic factor) 중 하나인, 조직학적 등급화는 종양 세포의 형태학적 및 세포학적 특징의 현미경적 평가에 기반하고, 종양의 공격성을 반영하는, 종양의 조합된 점수(combined score)이다. 그 후 이 조합된 점수는 유방암 종양을 다음으로 계층화하기 위해 사용된다; 등급 1 - 느리게 성장하고(slow growing) 잘 분화된 것(well differentiated), 등급 2 - 적절하게 분화된 것(moderately differentiated), 및 등급 3 - 고도로 증식하고 저조하게 (poorly) 분화된 것 (Elston and Ellis, 1991). 그러나, 환자 예후에 대한 조직학적 등급의 임상값은 의문이었으며, 종양 등급과 관련된 현재의 도전을 주로 반영한다 (Frierson et al., 1995; Robbins et al., 1995). 또한, 종양의 30-60%는 조직학적 등급 2로 분류되고, 이는 매우 불균일(heterogeneous) 환자 집단을 나타내고 임상 의사 결정에 대하여 덜 정보를 제공한 것으로 보였다 (Sotiriou et al., 2006). 분명하게, 전형적 임상적 실험 파라미터(clinical laboratory parameter)는 적절한 예후 및 위험-군 구별 및 주어진 치료가 성공적일지 여부를 예측하는 것에는 여전히 충분하지 않다. 그 결과, 일부 환자는 과대-치료, 과소-치료, 또는 어떠한 이득을 제공하지 않을 요법으로 치료될 것이다. 따라서, 예후 및 치료 결정을 개별화하는(individualize) 개선된 방식과 조합된, 유방암 생물학 및 종양 진행의 더 심오한 분자적 이해가 예후적 결과 및 결과적으로 치료 결과를 더 전진시키기 위해 요구된다 (Dowsett et al., 2007).

발명의 내용

해결하려는 과제

[0003] 지금까지, 게놈 노력(genomic effort)의 세트가 유방암 종류를 서브그룹핑(subgrouping)하고 (Ivshina et al., 2006; Perou et al., 2000; Sorlie et al., 2001), 및 유방암 예후(prognostics) 및 위험 계층화(risk stratification)하기 위한 분자 시그니처를 생성하였다 (Paik et al., 2004; van 't Veer et al., 2002; van de Vijver et al., 2002). 반면에, 프로테옴 발견(proteomic finding)은 핵심 발견을 임상 실시로 이행을 가속시킬 것으로 예상되었다 (Hanash, 2003). 이러한 문맥에서, 전통적 질량 분석법 (MS)-기반 프로테오믹스는 주로 세포주 및 약간의 조직 샘플을 표적화하는 유방암 프로테옴의 소중한 목록(inventory)을 생성하였고 (Bouchal et al., 2009; Geiger et al., 2010; Geiger et al., 2012; Gong et al., 2008; Kang et al., 2010; Strande et al., 2009; Sutton et al., 2010), 더 최근에, 친화성 프로테오믹스(affinity proteomics) 노력은 유방암 진단 및 재발의 위험을 예측하기 위한 제1 다중다수의 혈청 묘사(first multiplexed serum portrait)를 산출하였다 (Carlsson et al., 2008; Carlsson et al., 2011). 그러나 최근 기술적 발전에도 불구하고, 조 프로테옴(curde proteome), 예를 들면 조직 추출물의 큰 집단(cohorts)의 상세적 단백질 발현 프로파일을 민감성 있고 (sensitive) 재현성있는 방식으로 생성하는 것은 전통적 프로테옴 기술 (Aebersold and Mann, 2003) 또는 친화성 프로테오믹스 (Borrebaeck and Wingren, 2011)를 사용하는 도전을 남겨놓았다.

[0004] 이들 문제를 해결하기 위해, 본 발명자는 친화성 프로테오믹스와 MS의 가장 좋은 특징을 조합한, 글로벌 프로테옴 조사(global proteom survey, GPS) 기술 플랫폼을 최근에 개발하였다 (Wingren et al., 2009). GPS는 민감성 있고 정량적 방식으로 조 프로테옴을 재현성있게 관측하는 발견 노력(discovery endeavour)에 적합하였다 (Olsson et al., 2012; Olsson et al., 2011).

[0005] 이러한 연구에서, 본 발명자는 GPS를 사용하여 종양 진행을 반영하는 조직학적 등급화된 유방암 조직의 상세한 분자 조직 묘사를 기술하였다. 이것을 위하여, 52개의 유방암 조직 프로테옴이 프로파일되어, 본 발명자가 알고 있는 한 가장 큰 무-표지 LC-MS/MS-기반 유방암 조직 연구 중 하나를 나타내었다. 상기 단백질 발현 프로파일은 정통적 방법(orthogonal method)을 사용해 성공적으로 검증되었다. 장기간 실시에서, 이들 조직 바이오마커 묘사는 개선된 분류 및 예후를 위한 방식을 개척할 수 있다.

[0006] 따라서, 본 발명의 제1 양태는

[0007] a) 시험될 샘플을 제공하는 단계; 및

[0008] b) 표 1에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하여 상

기 시험 샘플의 바이오마커 시그니처(biomarker signature)를 결정하는 단계;를 포함하고,

- [0009] 표 1에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 상기 존재 및/또는 양은 유방암-관련 질환 상태임을 나타내는 것인 유방암-관련 질환 상태를 결정하는 방법을 제공한다. 따라서, 사실상, 단계 (b)는 표 1에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 사용하거나 이에 기반하여 유방암 관련 질환 상태를 결정하는 단계 ((b)(i))의 추가적 단계를 포함한다.
- [0010] "유방암-관련 질환 상태(breast cancer-associated disease state)"는 유방암 세포의 조직학적 등급 및/또는 유방암 세포를 포함하는 개체의 무전이 생존 시간(metastasis-free survival time)을 의미한다.
- [0011] 유방암-관련 질환 상태는 (유방암 세포의) 조직학적 등급 및/또는 (개체의) 무전이 생존 시간일 수 있다.
- [0012] "바이오마커(biomarker)"는 자연적으로 존재하는 생물학적 분자, 또는 그의 구성 요소(component) 또는 단편을 의미하고, 그의 측정은 유방암의 예후에 유용한 정보를 제공할 수 있다. 예를 들면, 바이오마커는 자연적으로 존재하는 단백질 또는 탄수화물 모이어티, 또는 그의 항원성 구성요소 또는 단편일 수 있다.
- [0013] 바람직하게 시험될 샘플은 포유동물로부터 제공된다. 포유동물은 가축(domestic animal) 또는 농장 동물(farm animal)일 수 있다. 바람직하게, 상기 포유동물은 랫트, 마우스, 기니피그, 고양이, 개, 말 또는 영장류이다. 가장 바람직하게, 상기 포유동물은 인간이다. 바람직하게 상기 샘플은 유방암 세포를 포함하거나 유방암 세포로 구성된 세포 또는 조직 샘플 (또는 그 파생물) 또는 동등하게 바람직하게는, 유방암 세포를 포함하거나 이로 구성된 세포 또는 조직 샘플로부터 유래된 단백질 또는 핵산이다. 바람직하게 시험 샘플 및 대조군 샘플은 동일한 종으로부터 유래된다.
- [0014] 유방암-관련 질환 상태가 유방암 세포의 조직학적 등급이거나 상기 등급을 포함하는 경우, 상기 방법은
- [0015] c) 조직학적 등급 1 유방암 세포, 조직학적 등급 2 유방암 세포 및/또는 조직학적 등급 3 유방암 세포를 포함하거나 조직학적 등급 1 유방암 세포, 조직학적 등급 2 유방암 세포 및/또는 조직학적 등급 3 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 및
- [0016] d) 단계 (b)에서 측정된 하나 이상의 상기 바이오마커의 대조군 샘플(들) 중 존재 및/또는 양을 측정하여 상기 대조군 샘플(들)의 바이오마커 시그니처를 결정하는 단계;를 더 포함할 수 있고,
- [0017] 단계 (b)에서 측정된 상기 하나 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양이:
- [0018] i) (존재하는 경우) 제1 조직학적 등급의 유방암 세포를 포함하거나 제1 조직학적 등급의 유방암 세포로 구성된 대조군 샘플 중 상기 존재 및/또는 양에 해당하는 경우;
- [0019] ii) (존재하는 경우) 제2 조직학적 등급의 유방암 세포를 포함하거나 제2 조직학적 등급의 유방암 세포로 구성된 대조군 샘플 중 상기 존재 및/또는 양과 다른 경우; 및/또는
- [0020] iii) (존재하는 경우) 제3 조직학적 등급의 유방암 세포를 포함하거나 제3 조직학적 등급의 유방암 세포로 구성된 대조군 샘플 중 상기 존재 및/또는 양과 다른 경우, 유방암 세포의 상기 존재가 확인된다.
- [0021] 따라서, 제1 조직학적 등급이 Elston 등급 1인 경우, (존재하는 경우) 제2 및 제3 조직학적 등급은 Elston 등급 2 및 Elston 등급 3 (또는 그 반대)일 것이다. 제1 조직학적 등급이 Elston 등급 2인 경우, (존재하는 경우) 제2 및 제3 조직학적 등급은 Elston 등급 1 및 Elston 등급 3 (또는 그 반대)일 것이다. 제1 조직학적 등급이 Elston 등급 3인 경우, (존재하는 경우) 제2 및 제3 조직학적 등급은 Elston 등급 1 및 Elston 등급 2 (또는 그 반대)일 것이다.
- [0022] "제1 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플 중 상기 존재 및/또는 양에 해당하는 (corresponds to the presence and/or amount in a control sample comprising or consisting breast cancer cells of a first histological grade)"은 상기 존재 및/또는 양이 제1 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플의 그것과 동일하거나, 제2 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플 및/또는 제3 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플의 그것 (또는 동일한 것을 나타내는 소정의 참조값)과 더 가깝다는 것을 의미한다. 바람직하게 상기 존재 및/또는 양은 제1 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플의 그것의 60% 이상, 예를 들면 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 이상이다.

- [0023] "제3 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플 중 상기 존재 및/또는 양과 다른 (is different to the presence and/or amount in a control sample comprising or consisting breast cancers of a third histological grade)"은 상기 존재 및/또는 양이 제1 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플의 그것과 다르고, 제2 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플 및/또는 제3 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플의 그것(또는 동일한 것을 나타내는 소정의 참조값)보다 크다는 것을 의미한다. 바람직하게 상기 존재 및/또는 양은 제2 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플 및/또는 제3 조직학적 등급의 유방암 세포를 포함하거나 이로 구성된 대조군 샘플의 그것과 40% 이하, 예를 들면 39%, 38%, 37%, 36%, 35%, 34%, 33%, 32%, 31%, 30%, 29%, 28%, 27%, 26%, 25%, 24%, 23%, 22%, 21%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1% 또는 0% 이하이다.
- [0024] 바람직하게 조직학적 등급 대조군 샘플은 유방암 세포의 단일 조직학적 등급을 포함하거나 이로 구성된다. 바람직하게, 단계 (c)는 하기 단계를 포함하거나 하기 단계로 구성된다:
- [0025] i) 조직학적 등급 1 유방암 세포를 포함하거나 조직학적 등급 1 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 조직학적 등급 2 유방암 세포를 포함하거나 조직학적 등급 2 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 및 조직학적 등급 3 유방암 세포를 포함하거나 조직학적 등급 3 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계;
- [0026] ii) 조직학적 등급 1 유방암 세포를 포함하거나 조직학적 등급 1 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 및 조직학적 등급 2 유방암 세포를 포함하거나 조직학적 등급 2 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계;
- [0027] iii) 조직학적 등급 1 유방암 세포를 포함하거나 조직학적 등급 1 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 조직학적 등급 3 유방암 세포를 포함하거나 조직학적 등급 3 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계;
- [0028] iv) 조직학적 등급 2 유방암 세포를 포함하거나 조직학적 등급 2 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 조직학적 등급 3 유방암 세포를 포함하거나 조직학적 등급 3 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계;
- [0029] v) 조직학적 등급 1 유방암 세포를 포함하거나 조직학적 등급 1 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계;
- [0030] vi) 조직학적 등급 2 유방암 세포를 포함하거나 조직학적 등급 2 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계; 또는
- [0031] vii) 조직학적 등급 3 유방암 세포를 포함하거나 조직학적 등급 3 유방암 세포로 구성된 하나 이상의 대조군 샘플을 제공하는 단계.
- [0032] 상기 유방암-관련 질환 상태는 개체의 무전이 생존 시간 (etastasis-free survival time)이고, 상기 방법은
- [0033] c) 10 년 무전이 생존 미만을 갖는 개체로부터 유래된 유방암 세포를 포함하거나 상기 유방암 세포로 구성된 하나 이상의 제1 대조군 샘플 및/또는 10년 이상의 무전이 생존을 갖는 개체로부터 유래된 유방암 세포를 포함하거나 상기 유방암 세포로 구성된 하나 이상의 제2 대조군 샘플을 제공하는 단계; 및
- [0034] d) 단계 (b)에서 측정된 하나 이상의 바이오마커의 상기 대조군 샘플(들) 중 상기 존재 및/또는 양을 측정하여 상기 대조군 샘플(들)의 바이오마커 시그니처를 결정하는 단계를 더 포함할 수 있고,
- [0035] 상기 개체의 무전이 생존 시간은 단계 (b)에서 측정된 하나 이상의 바이오마커의 상기 존재 및/또는 양이 (존재하는 경우) 제1 대조군 샘플의 상기 존재 및/또는 양에 해당하는 경우 및/또는 (존재하는 경우) 제2 대조군 샘플의 상기 존재 및/또는 양과 다른 경우, 10년 미만인 것으로서 확인되고; 및
- [0036] 상기 개체의 무전이 생존 시간은 단계 (b)에서 측정된 하나 이상의 바이오마커의 상기 존재 및/또는 양이 (존재하는 경우) 제1 대조군 샘플의 상기 존재 및/또는 양과 다른 경우 및/또는 (존재하는 경우) 제2 대조군 샘플의 상기 존재 및/또는 양에 해당하는 경우, 10년 이상인 것으로서 확인된다.
- [0037] "하나 이상의 제1 대조군 샘플의 상기 존재 및/또는 양에 해당하는(corresponds to the presence and/or amount of the one or more first control sample)"은 상기 하나 이상의 제1 대조군 샘플의 것과 동일하거나

상기 하나 이상의 제2 대조군 샘플 (또는 동일한 것을 나타내는 소정의 참조값)에 대한 것보다 제1 대조군 샘플의 그것에 더 가깝다는 것을 의미한다. 바람직하게 상기 존재 및/또는 양은 제1 대조군 샘플의 그것과 60% 이상, 예를 들면, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 이상이다.

[0038] "상기 하나 이상의 제2 대조군 샘플의 상기 존재 및/또는 양과 다른 (is different to the presence and/or amount of the one or more a second control sample)"은 제2 대조군 샘플의 그것 (또는 동일한 것을 나타내는 소정의 참조값)과 다르다는 것을 의미한다. 바람직하게 상기 존재 및/또는 양은 제2 대조군 샘플의 그것의 40% 이하, 예를 들면 39%, 38%, 37%, 36%, 35%, 34%, 33%, 32%, 31%, 30%, 29%, 28%, 27%, 26%, 25%, 24%, 23%, 22%, 21%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1% 또는 0% 이하이다.

[0039] 바람직하게는, 상기 하나 이상의 제1 및/또는 제2 무전이 생존 시간 대조군 샘플은 시험될 샘플과 동일한 조직학적 등급의 것이다.

[0040] 바람직하게는, 상기 하나 이상의 대조군 샘플은 시험될 개체에 대하여 연령- 및/또는 성별- 일치(match)된다. 즉, 건강한 개체는 시험될 개체와 대략 동일한 연령 (예, 5살 이내)이고 동일한 성별이다.

[0041] 바람직하게는, 단계 (b)에서 측정되는 상기 하나 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양은 소정의 참조값과 비교된다.

[0042] 따라서, 단계 (b)에서 측정되는 상기 하나 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양은 단계 (d)에서 측정되는 상기 하나 이상의 바이오마커의 상기 존재 및/또는 양 또는 소정의 참조값과 유의성있게 다르다 (즉, 통계적으로 다르다). 예를 들면, 수반하는 실시예에 논의된 바와 같이, 상기 시험 샘플과 대조군 샘플 중 특정 바이오마커의 존재 및/또는 양 간의 유의성있는 차이는 $p < 0.05$ (예, $p < 0.04$, $p < 0.03$, $p < 0.02$ 또는 $p < 0.01$ 인 것)인 것으로 분류될 수 있다.

[0043] 따라서, 본 발명의 제1 양태의 방법은 (동시에 또는 연속적으로) 유방암 세포의 조직학적 등급 및 개체의 무전이 생존 시간을 결정하는 단계를 포함하거나 이로 구성될 수 있다.

[0044] 바람직하게는 단계 (b)는 표 1에서 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1에서 정의된 군으로부터 선택된 2 이상, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78 또는 79개 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성된다.

[0045] 따라서, 본 발명의 제1 양태는 하기 단계를 포함하는, 유방암 세포의 조직학적 등급을 결정하기 위한 방법 (즉, 조직학적 등급을 결정하기 위한 유방암 샘플의 단계매김(staging))을 포함하거나 이로 구성될 수 있다:

[0046] a) 시험될 샘플을 제공하는 단계;

[0047] b) 표 1에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양을 측정하여 시험 샘플의 바이오마커 시그니처를 결정하는 단계;

[0048] 상기 표 1에 정의된 군으로부터 선택된 상기 하나 이상의 바이오마커의 상기 시험 샘플 중 상기 존재 및/또는 양은 유방암 세포의 조직학적 등급인 것을 나타낸다.

[0049] "유방암 세포의 조직학적 등급 결정 (determining the histological grade of breast)"은 유방암 세포의 샘플을 참조에 의해 본 명세서에 포함되는 Elston, C. W., and Ellis, I. O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 19, 403-410에 정의된 바와 같이 조직학적 등급 1 (즉, Elston 등급 1), 조직학적 등급 2 (즉, Elston 등급 2) 또는 조직학적 등급 3 (즉, Elston 등급 3)으로 분류되는 것을 의미한다.

[0050] 상기 방법이 유방암 세포의 조직학적 등급을 결정하는 단계를 포함하거나 이로 구성되는 경우, 단계 (b)는 표 1A에서 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1A에 정의된 군으로부터 선택된 2개

이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성될 수 있다. 바람직하게 단계 (b)는 표 1B 에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양, 예를 들면 표 1B에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30개 이상의 바이오마커를 측정하는 단계를 포함하거나 이로 구성된다. 바람직하게 단계 (b)는 표 1C에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양, 예를 들면 표 1C에 정의된 군으로부터 선택된 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27 또는 28개 이상의 바이오마커를 측정하는 단계를 포함하거나 이로 구성된다. 덜 바람직하게는, 단계 (b)는 표 1D 에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양, 예를 들면 표 1D에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8, 9 또는 10개 이상의 바이오마커를 측정하는 단계를 포함하거나 이로 구성된다. 또한 덜 바람직하게는, 단계 (b)는 표 1E 에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양, 예를 들면 표 1E에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8 또는 9개 이상의 바이오마커를 측정하는 단계를 포함하거나 이로 구성된다. 따라서, 단계 (b)는 표 1에 정의된 바이오마커의 모든 것의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성될 수 있다.

[0051] 따라서, 본 발명의 제1 양태는 하기 단계를 포함하는 개체의 무전이 생존 시간을 결정하는 방법을 포함하거나 이로 구성될 수 있다:

[0052] a) 시험될 샘플을 제공하는 단계;

[0053] b) 표 1에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하여 시험 샘플의 바이오마커 시그니처를 결정하는 단계;

[0054] 상기 표 1에 정의된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양은 개체의 무전이 생존 시간을 나타낸다.

[0055] "개체의 무전이 생존 시간 결정 (determining the metastasis-free survival time of an individual)"은 수득된 시험 샘플로부터의 개체가 초기 진단으로부터 10년 미만 또는 10년 초과 무전이 생존 시간 (원격 전이-무생존(distant metastasis-free survival)/DMFS)을 갖는 것으로 예후된다는 것을 의미한다.

[0056] 상기 방법이 개체의 무전이 생존 시간을 결정하는 단계를 포함하거나 이로 구성되는 경우 단계 (b)는 표 1A에서 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1A에 정의된 군으로부터 선택된 2개 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성될 수 있다. 바람직하게 단계 (b)는 표 1B 에서 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1B에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30개 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성된다. 바람직하게는 단계 (b)는 표 1D에서 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1D에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8, 9 또는 10개 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성된다. 바람직하게는 단계 (b)는 표 1A, 표 1B 및 표 1D에서 정의된 모든 것의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성된다.

[0057] 상기 방법은 개체의 무전이 생존 시간을 결정하는 단계를 포함하거나 이로 구성되고, 비록 덜 바람직하지만, 단계 (b)는 표 1C에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1C에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27 또는 28개 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성될 수 있다. 또한 덜 바람직하게는, 단계 (b)는 표 1E에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면 표 1E에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8 또는 9개 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성될 수 있다. 또한 덜 바람직하게 단계 (b)는 표 1C 및 표 1E에 정의된 바이오마커의 모든 것의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성될 수 있다.

[0058] 따라서, 비록 덜 바람직하지만, 본 발명의 제 1 양태의 방법은 개체의 무전이 생존 시간을 결정하는 단계를 포함하거나 이로 구성될 수 있고 상기 단계 (b)는 표 1에 정의된 바이오마커의 모든 것의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성된다.

[0059] 따라서, 본 발명의 제1 양태에 따른 방법은 SPON1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 KERA 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 APCS 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 ATP6V1G1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 RPS27L 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 DPYSL3 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 ERP44 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 RAPGEF1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 ACLY 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 CMA1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 MCM3 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 ANGPTL2 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 AEBP1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 UBE2V2 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 MIS18BP1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 CLCF1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 ABAT 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 SLC25A5 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 STIP1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 OLFML3 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 CD3G 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 MCM7 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 SLC25A11 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 NOP56 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 RRP8 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 SLTM 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 TSN 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 ECH1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 PRELP 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 SARS 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 RPS25 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 ESYT1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 PODN 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 RPRD1B 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 RPLP0P6 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 CD300LG 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 SUGT1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 POTEF 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 KARS 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 NDUFS2 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 HNRNP2 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 CALU 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 EIF3B 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 SLC4A1AP 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 RPS5 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 PLXDC2 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 KIAA1324 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 MRC1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 RPRD1A 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 SHMT2 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 CCT4 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 TSSC1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 IKZF3 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 UBE2Q1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 PSMD9 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 SNRNP70 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 RALB 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 ACO2 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 MYO18A 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 QARS 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 PABPC4 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 SCGB1D2 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 PFKP 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 SLC3A2 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 ASPN 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 CD38 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 MXRA5 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 CDK1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 STC2 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 CTSC 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 NOP58 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 PGK1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 FKBP3 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 GSTM3 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 CALML5 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 PML 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 ADAMTS4 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 THBS1 발현을 측정하는 것을 포함할 수 있다. 상기 방법은 FN1 발현을 측정하는 것을 포함할 수 있다.

[0060] 일 바람직한 구체예에서, 단계 (b)는 MCM7, NOP56, MCM3, PABPC4, MXRA5, STC2, SCGB1D2 및 ANGPTL2로 구성된 군으로부터 선택된 하나 이상의 바이오마커의 시험 샘플 중 존재 및/또는 양을 측정하는 단계를 포함하거나 이로 구성된다. 예를 들면, 단계 (b)는 이들 바이오마커의 2, 3, 4, 5, 6, 7 또는 8개의 시험 샘플 중 존재 및/또는 양의 측정을 포함하거나 이로 구성될 수 있다. 바람직하게, 본 구체예에서, 상기 유방암-관련 질환 상태는 조직학적 등급이고, 그러나, 덜 바람직하게는 상기 유방암-관련 질환 상태는 무전이 생존 시간이거나 또한 이를 포함한다.

[0061] 또 다른 바람직한 구체예에서, 단계 (b)는 OLFML3, SPON1, PODN 및 ASPN으로 구성된 군으로부터 선택된 하나 이

상의 바이오마커의 시험 샘플 중 존재 및/또는 양의 측정을 포함하거나 이로 구성된다. 예를 들면, 단계 (b)는 이들 바이오마커의 2, 3 또는 4개의 시험 샘플 중 존재 및/또는 양의 측정을 포함하거나 이로 구성될 수 있다. 바람직하게, 본 구체예에서, 상기 유방암-관련 질환 상태는 조직학적 등급이고, 그러나, 덜 바람직하게는 상기 유방암-관련 질환 상태는 무전이 생존 시간이거나 또한 이를 포함한다.

- [0062] "발현(expression)"은 mRNA 또는 단백질과 같은 유전자 산물의 수준 또는 양을 의미한다.
- [0063] 단백질 및/또는 핵산의 농도를 검출 및/또는 측정하는 방법은 당해 기술분야에서 통상의 기술자에게 잘 알려져 있고, 예를 들면 Sambrook and Russell, 2001, Cold Spring Harbor Laboratory Press를 참고한다.
- [0064] 단백질의 검출 및/또는 측정하는 바람직한 방법은 웨스턴 블롯, 노스-웨스턴 블롯, 면역흡착 분석법 (ELISA), 항체 마이크로어레이, 조직 마이크로어레이 (TMA), 면역침전(immunoprecipitation), 인 시투 혼성화(*in situ* hybridisation) 및 기타 면역조직화학(immunohistochemistry) 기법, 방사면역측정법(radioimmunoassay, RIA), 면역방사계측법(immunoradiometric assays, IRMA) 및 단일클론 및/또는 폴리클론 항체를 이용한 샌드위치 분석(sandwich assay)을 포함한 면역효소분석(immunoassays, IEMA)을 포함한다. 예시적인 샌드위치 분석은 참조에 의해 본 명세서에 포함된 미국 특허 제4,376,110호 및 제4,486,530호 중 David *et al.*,에 의해 기재된다. 슬라이드 상의 세포의 항체 염색은 세포학 진단 시험에 잘 알려진 방법, 및 당해 기술분야에서의 통상의 기술자에게 잘 알려진 방법에 사용될 수 있다.
- [0065] 통상적으로, ELISA는 보통 고체상 분석에서, 착색된 반응 산물을 제공하는 효소의 사용을 포함한다. 호스레디쉬 퍼옥시다제(horseradish peroxidase) 및 포스파타제(phosphatase)와 같은 효소가 광범위하게 사용되었다. 포스파타제 반응을 증폭하는 방식은 기질로서 NADP를 사용하여 제2 효소 시스템을 위한 보조효소로서 작용하는 NAD를 생성시키는 것이다. 에스케키아 콜라이 (*Escherichia coli*)로부터 유래된 포스포타제는 안정하고 우수한 반응 색깔을 제공하고, 이는 상기 효소가 조직에 존재하지 않기 때문이다. 루시페라제와 같은 효소에 기반한 화학-발광계(chemi-luminescent system)가 또한 사용될 수 있다.
- [0066] 비타민 비오틴과의 접합(conjugation)은 큰 특이성 및 친화성으로 결합하는 효소-연결된 아비딘 또는 스트렙타비딘과의 반응에 의해 용이하게 검출될 수 있기 때문에 자주 사용된다.
- [0067] 핵산(예, mRNA)의 검출 및/또는 측정을 위한 바람직한 방법은 서던 블롯, 노던 블롯, 중합효소 연쇄 반응(PCR), 역전사효소 PCR (RT-PCR), 정량적 실시간 PCR (qRT-PCR), 나노어레이, 마이크로어레이, 매크로어레이, 자기방사선화법 및 인 시투 혼성화(*in situ* hybridisation)를 포함한다.
- [0068] 본 발명의 제1 양태의 일 구체예에서, 단계 (b)는 하나 이상의 바이오마커(들)를 코딩하는 핵산 분자의 발현의 측정을 포함한다. 상기 핵산 분자는 cDNA 분자 또는 mRNA 분자일 수 있다. 바람직하게 상기 핵산 분자는 mRNA 분자이다. 또한 바람직하게 상기 핵산 분자는 cDNA 분자이다.
- [0069] 따라서, 단계 (b) 중 하나 이상의 바이오마커(들)의 발현의 측정은 서던 혼성화, 노던 혼성화, 중합효소 연쇄 반응 (PCR), 역전사효소PCR (RT-PCR), 정량적 실시간 PCR (qRT-PCR), 나노어레이, 마이크로어레이, 매크로어레이, 자기방사선화법(autoradiography) 및 인 시투 혼성화로 구성된 군으로부터 선택된 방법을 사용하여 수행될 수 있다. 바람직하게 단계 (b) 중 하나 이상의 바이오마커(들)의 발현 측정은 DNA 마이크로어레이를 사용하여 결정된다. 따라서, 상기 방법은 표 1에서 확인된 바이오마커의 하나를 코딩하는 핵산 분자에 각각 선택적으로 결합할 수 있는, 하나 이상의 결합 모이어티를 사용하여 단계 (b) 중 하나 이상의 바이오마커(들)의 발현의 측정을 포함하거나 이로 구성될 수 있다.
- [0070] 바람직하게 하나 이상의 결합 모이어티는 각각 DNA, RNA, PNA, LNA, GNA, TNA 또는 PMO (바람직하게 DNA)와 같은 핵산 분자를 포함하거나 이로 구성된다. 바람직하게 하나 이상의 결합 모이어티는 길이가 5 내지 100 뉴클레오티드이다. 더 바람직하게, 하나 이상의 핵산 분자는 길이가 15 내지 35 뉴클레오티드이다. 결합 모이어티는 검출가능한 모이어티를 포함할 수 있다.
- [0071] 적합한 결합제 (또한 결합 분자로 지칭)는 하기 논의되는 바와 같이, 주어진 핵산, 단백질 또는 아미노산 모티프에 결합하는 능력에 기반한 라이브러리로부터 선택 또는 스크린(screen)될 수 있다.
- [0072] 본 발명의 제1 양태의 또 다른 구체예에서 단계 (b)는 하나 이상의 바이오마커의 단백질 또는 폴리펩티드, 또는 그의 단편 또는 그의 파생물(derivative)의 발현 측정을 포함한다. 바람직하게 단계 (b) 중 하나 이상의 바이오마커(들)의 발현 측정은 표 1에서 확인된 바이오마커 중 하나에 각각 선택적으로 결합할 수 있는 하나 이상의 결합 모이어티를 사용하여 수행된다.

- [0073] 상기 하나 이상의 결합 모이어티는 항체 또는 그의 항원-결합 단편을 포함하거나 이로 구성될 수 있다.
- [0074] 용어 "항체(antibody)"는 합성 항체, 재조합 항체 또는 항체 하이브리드 (antibody hybrid), 예를 들면 면역글로불린 경쇄 및/또는 중쇄 가변 및/또는 불변 영역의 파지-디스플레이(phage-display)에 의해 생성된 단쇄 항체 분자, 또는 통상의 기술자에게 알려진 면역 분석(immunoassay format) 포맷에서 항원과 결합할 수 있는 기타 면역상호작용(immunointeractive) 분자를 포함하나 이에 제한되지 않는다. 본 발명자는 또한 아피바디 또는 압타머와 같은 항체-유사 결합체의 용도를 포함시킨다.
- [0075] 특이적 결합 부위를 보유하는 항체 단편의 합성에 관련된 기법의 일반적 검토는 Winter & Milstein (1991) *Nature* **349**, 293-299에서 발견된다.
- [0076] 추가적으로, 또는 대안적으로, 하나 이상의 제1 결합 분자는 압타머일 수 있다 (Collett *et al.*, 2005, *Methods* **37**:4-15 참조).
- [0077] 분자 라이브러리 예를 들면 항체 라이브러리 (Clackson *et al.*, 1991, *Nature* **352**, 624-628; Marks *et al.*, 1991, *J Mol Biol* **222**(3): 581-97), 펩티드 라이브러리(Smith, 1985, *Science* **228**(4705): 1315-7), 발현된 cDNA 라이브러리(Santi *et al.* (2000) *J Mol Biol* 296(2): 497-508), 아피바디(affibodies)와 같은 항체 프레임워크(framework)가 아닌 스카폴드(scaffold)에 대한 라이브러리 (Gunneriusson *et al.*, 1999, *Appl Environ Microbiol* **65**(9): 4134-40) 또는 압타머에 기반한 라이브러리 (Kenan *et al.*, 1999, *Methods Mol Biol* **118**, 217-31)가 이로부터 주어진 모티브에 특이적인 결합 분자가 본 발명의 방법에서의 사용을 위해 선택되는 공급원(source)으로 사용될 수 있다.
- [0078] 분자 라이브러리는 원핵세포 (Clackson *et al.*, 1991, *op. cit.*; Marks *et al.*, 1991, *op. cit.*) 또는 진핵세포 (Kieck *et al.*, 1999, *Proc Natl Acad Sci USA*, **96**(10):5651-6)에서 인 비보에서 발현될 수 있거나, 세포의 관여 없이 인 비트로에서 발현될 수 있다 (Hanes & Pluckthun, 1997, *Proc Natl Acad Sci USA* **94**(10):4937-42; He & Taussig, 1997, *Nucleic Acids Res* **25**(24):5132-4; Nemoto *et al.*, 1997, *FEBS Lett.*, **414**(2):405-8).
- [0079] 단백질 기반 라이브러리가 사용될 경우, 잠재적 결합 분자의 라이브러리를 코딩하는 유전자가 종종 바이러스에 패키징되고(packaged) 상기 잠재적 결합 분자는 바이러스의 표면에 표시된다 (Clackson *et al.*, 1991, *op. cit.*; Marks *et al.*, 1991, *op. cit.*; Smith, 1985, *op. cit.*).
- [0080] 아마도 가장 흔하게 사용되는 표시 시스템(display system)은 그의 표면에 항체 단편을 표시하는 필라멘트성 박테리오파지(filamentous bacteriophage)이고, 상기 항체 단편은 박테리오파지의 소수 외피 단백질(minor coat protein)로의 융합체(fusion)로서 발현된다 (Clackson *et al.*, 1991, *op. cit.*; Marks *et al.*, 1991, *op. cit.*). 그러나, 표시하기 위한 기타 적합한 시스템은 기타 바이러스 (EP 39578), 박테리아 (Gunneriusson *et al.*, 1999, *op. cit.*; Daugherty *et al.*, 1998, *Protein Eng* **11**(9):825-32; Daugherty *et al.*, 1999, *Protein Eng* **12**(7):613-21), 및 효모 (Shusta *et al.*, 1999, *J Mol Biol* **292**(5):949-56) 사용을 포함한다.
- [0081] 또한, 표시 시스템은 소위 리보솜 표시 시스템에서 폴리펩티드 산물과 이를 코딩하는 mRNA와의 연결(linkage) (Hanes & Pluckthun, 1997, *op. cit.*; He & Taussig, 1997, *op. cit.*; Nemoto *et al.*, 1997, *op. cit.*), 또는 대안적으로 폴리펩티드 산물과 이를 코딩하는 DNA와의 연결(미국 특허 제5,856,090호 및 WO 98/37186 참조)을 이용하여 개발되었다.
- [0082] 항체의 가변 중쇄(V_H) 및 가변 경쇄 (V_L) 도메인은 항원 인식에 관련되며, 이는 초기 프로테아제 소화 실험에 의해 처음으로 인식된 사실이다. 설치류 항체의 "인간화(humanisation)"에 의해 추가 확인이 이루어졌다. 설치류 기원의 가변 도메인은 인간 기원의 불변 도메인과 융합될 수 있어서, 그 결과로 생긴 항체는 설치류 부모인 (rodent parented) 항체의 항원 특이성을 보유한다 (Morrison *et al.* (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6851-6855).
- [0083] 항원 특이성은 가변 도메인에 의해 주어지고, 불변 도메인과 독립적이라는 점은 모두가 하나 이상의 가변 도메인을 포함하는, 항체 단편의 박테리아 발현을 수반하는 실험으로부터 알려졌다. 이러한 분자는 Fab-유사 분자 (Better *et al.* (1988) *Science* **240**, 1041); Fv 분자 (Skerra *et al.* (1988) *Science* **240**, 1038); V_H 및 V_L 파트너 도메인이 유연한 올리고펩티드를 통해 연결되는 단쇄 Fv(ScFv) 분자(Bird *et al.* (1988) *Science* **242**, 423; Huston *et al.* (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5879) 및 단리된 V 도메인을 포함하는 단일 도메인 항체 (dAbs) (Ward *et al.* (1989) *Nature* **341**, 544)를 포함한다. 특이적 결합 부위를 보유한 항체 단편의 합성에 관련된 기술의 일반적인 검토는 Winter & Milstein (1991) *Nature* **349**, 293-299에서 발견된다.

- [0084] 항체 또는 항원-결합 단편은 온전한 항체, Fv 단편 (예, 단쇄 Fv 및 이황화-결합 Fv), Fab-유사 단편 (예, Fab 단편, Fab' 단편 및 F(ab)₂ 단편), 단일 가변 도메인 (예, V_H 및 V_L 도메인) 및 도메인 항체 (dAb, 단일 및 이중 포맷 [즉, dAb-링커-dAb] 포함)으로 구성된 군으로부터 선택될 수 있다. 바람직하게, 상기 항체 또는 항원-결합 단편은 단쇄 Fv (scFv)이다.
- [0085] 상기 하나 이상의 결합 모이어티는 대안적으로 항체-유사 결합제, 예를 들면 아피바디 또는 압타머를 포함하거나 이로 구성될 수 있다.
- [0086] "ScFv 분자(ScFv molecules)"은 V_H 파트너 영역과 V_L 파트너 도메인이 유연한 올리고펩티드를 통해 연결된 것인 분자를 의미한다.
- [0087] 전체 항체(whole antibody)보다 항체 단편을 사용하는 장점은 여러 가지이다. 단편의 더 작은 크기는 고정 조직(solid tissue)의 더 나은 침투(penetration)와 같은, 향상된 약리적 특성을 가져올 수 있다. 보체 결합(complement binding)과 같은, 전체 항체의 효과기 기능(effector function)은 제거된다. Fab, Fv, ScFv 및 dAb 항체 단편은 모두 대장균(*E. coli*)에서 발견되고 대장균으로부터 분비될 수 있고, 따라서 많은 양의 상기 단편의 용이한 생산(facile production)을 가능하게 한다.
- [0088] 전체 항체 및 F(ab')₂ 단편은 "이가(bivalent)"이다. "이가"는 상기 항체 및 F(ab')₂ 단편이 두 개의 항원 결합 부위를 가짐을 의미한다. 반대로, Fab, Fv, ScFv 및 dAb 단편은 일가(monovalent)이고, 오직 하나의 항원 결합 부위를 갖는다.
- [0089] 항체는 단일클론 또는 폴리클론일 수 있다. 적절한 단일클론 항체는 공지된 기법, 예를 들어 모두 참조로서 본 명세서에 포함된, "Monoclonal Antibodies: A manual of techniques" H Zola (CRC Press, 1988) 및 "Monoclonal Hybridoma Antibodies: Techniques and applications", J G R Hurrell (CRC Press, 1982)에 개시된 것에 의해 제조될 수 있다.
- [0090] 잠재적 결합 분자(potential binding molecule)가 라이브러리로부터 선택될 경우, 정의된 모티프를 갖는 하나 이상 셀렉터 펩티드(selector peptide)가 일반적으로 채용된다. 구조, 상기 펩티드에서 감소하는 유연성(flexibility) 또는 상기 결합 분자와의 상호작용을 허용하는 하전된, 극성 또는 소수성 측쇄를 제공하는 아미노산 잔기가 셀렉터 펩티드를 위한 모티프의 설계에 사용될 수 있다. 예를 들면:
- [0091] (i) 프롤린은 그의 측쇄가 알파 탄소 및 질소 모두와 결합되어 있어서 펩티드 구조를 안정화할 수 있고;
- [0092] (ii) 페닐알라닌, 티로신 및 트립토판은 방향족 측쇄를 갖고 높은 소수성인 반면, 류신 및 이소류신은 지방족 측쇄를 갖고 또한 소수성이고;
- [0093] (iii) 리신, 아르기닌 및 히스티딘은 염기성 측쇄를 갖고 중성 pH에서 양으로 하전될 것인 반면, 아스파르테이트 및 글루타메이트는 산성 측쇄를 갖고 중성 pH에서 음으로 하전될 것이고;
- [0094] (iv) 아스파라긴 및 글루타민은 중성 pH에서 중성이지만, 수소 결합에 참여할 수 있는 아미노기를 포함하고;
- [0095] (v) 세린, 트레오닌 및 티로신 측쇄는 히드록실기를 포함하고, 이는 수소 결합에 참여할 수 있다.
- [0096] 통상적으로, 결합제의 선택은 결합 분자의 타입에 해당하는 스팟(spot)에의 결합을 분석하기 위한 어레이 기술 및 시스템의 용도를 포함할 수 있다.
- [0097] 따라서, 바람직하게 항체 또는 그의 단편은 단일클론 항체 또는 그의 단편이다. 바람직하게 항체 또는 항원-결합 단편은 온전한 항체, Fv 단편 (예, 단쇄 Fv 및 이황화-결합 Fv), Fab-유사 단편 (예, Fab 단편, Fab' 단편 및 F(ab)₂ 단편), 단일 가변 도메인 (예, V_H 및 V_L 도메인) 및 도메인 항체 (dAbs, 단일 및 이중 포맷 [즉, dAb-링커-dAb] 포함)으로 구성된 군으로부터 선택된다. 따라서, 항체 또는 항원-결합 단편은 단쇄 Fv (scFv)일 수 있다. 대안적으로, 하나 이상의 결합 모이어티는 항체-유사 결합제, 예를 들면 아피바디 또는 압타머를 포함하거나 이로 구성된다. 하나 이상의 결합 모이어티는 검출가능한 모이어티를 포함한다.
- [0098] "검출가능한 모이어티 (detectable moiety)"는 모이어티의 존재 및/또는 상대적 양 및/또는 위치(예를 들어, 어레이 상의 위치)가 직접적으로 또는 간접적으로 결정될 수 있도록 하는 모이어티를 포함한다
- [0099] 적합한 검출가능한 모이어티는 해당 기술분야에 잘 알려져 있다.
- [0100] 예를 들면, 상기 검출가능한 모이어티는 특정 조건에 노출될 경우, 검출될 수 있는, 형광성(fluorescent) 및/또

는 발광성(luminescent) 및/또는 화학발광성(chemiluminescent) 모이어티일 수 있다. 예를 들어, 형광 모이어티는 형광성 모이어티의 여기를 야기하는 특정 파장 및 강도에서 방사선(radiation)(즉, 빛)에 노출될 필요가 있을 수 있고, 이에 의해 검출될 수 있는 특정 파장에서 검출가능한 형광을 발산하게 할 수 있다.

[0101] 대안적으로, 상기 검출가능한 모이어티는 (바람직하게는 검출될 수 없는) 기질을 시각화 및/또는 검출될 수 있는 검출가능한 생성물로 전환할 수 있는 효소일 수 있다. 적절한 효소의 예는 예를 들어, ELISA 분석과 관련하여 하기에 상술된다.

[0102] 따라서, 상기 검출가능한 모이어티는 형광성 모이어티; 발광성 모이어티; 화학발광성 모이어티; 방사성 모이어티 (예, 방사성 원자); 또는 효소 모이어티로 구성된 군으로부터 선택될 수 있다. 바람직하게, 상기 검출가능한 모이어티는 방사성 원자를 포함하거나 이로 구성된다. 상기 방사성 원자는 테크네튬-99m, 요오드-123, 요오드-125, 요오드-131, 인듐-111, 플루오린-19, 탄소-13, 질소-15, 산소-17, 인-32, 황-35, 중수소, 트리튬, 레늄-186, 레늄-188 및 이트륨-90으로 구성된 군으로부터 선택될 수 있다.

[0103] 명백하게, (예를 들면, 본 명세서에 상기 기재된 시험 샘플 및/또는 대조군 샘플 중 하나 이상의 바이오마커 및/또는 선택된 단백질 검출에 사용하기 위한 항체 분자와 같은) 검출될 수 있는 작용제(agent)는 검출가능한 모이어티가 용이하게 검출가능하도록 적절한 동위 원자(atomic isotope)를 충분히 가져야만 한다.

[0104] 대안적 바람직한 구체예에서, 상기 결합 모이어티의 검출가능한 모이어티는 형광성 모이어티이다.

[0105] 방사성- 또는 기타 표지는 알려진 방식으로 본 발명의 방법의 샘플에 존재하는 바이오마커 및/또는 본 발명의 결합 모이어티에 혼입될 수 있다. 예를 들면, 결합체가 폴리펩티드인 경우, 이는 생합성되거나 또는 예를 들어, 수소를 대신하여 불소-19(fluorine-19)를 수반하는 적절한 아미노산 전구체를 사용하여 화학적 아미노산 합성에 의해 합성될 수 있다. ^{99m}Tc, ¹²³I, ¹⁸⁶Rh, ¹⁸⁸Rh 및 ¹¹¹In 과 같은 표지는 예를 들어, 결합 모이어티에서 시스테인 잔기를 통해 부착될 수 있다. 이트륨-90(Yttrium-90)은 리신 잔기를 통해 부착될 수 있다. IODOGEN 방법(Fraker et al (1978) *Biochem. Biophys. Res. Comm.* 80, 49-57)은 ¹²³I 를 혼입시키기 위해 사용될 수 있다. 참조 문헌 ("Monoclonal Antibodies in Immunoscinigraphy", J-F Chatal, CRC Press, 1989)은 상세하게 다른 방법을 기재한다. 다른 검출가능한 모이어티(예를 들면, 효소, 형광, 발광, 화학발광 또는 방사성 모이어티)를 단백질에 접합하는 방법은 해당 기술분야에 잘 알려져 있다.

[0106] 상기 시험될 샘플(들) 중 바이오마커가 상기 단백질의 존재, 양 및/또는 위치를 결정하는 것을 간접적으로 돕는 모이어티로 표지될 수 있다는 것은 당해 기술분야에서의 통상의 기술자는 이해할 것이다. 따라서, 상기 모이어티는 다중성분 검출가능한 모이어티의 일 성분을 구성할 수 있다. 예를 들면, 상기 시험될 샘플(들) 중 바이오마커는 비오틴으로 표지될 수 있는데, 이는 융합된 스트렙타비딘을 사용하여 뒤이어 검출하게 하거나 검출가능한 표지에 연결되도록 한다.

[0107] 검출가능한 모이어티는 형광성 모이어티, 발광성 모이어티, 화학발광성 모이어티, 방사성 모이어티 및 효소 모이어티로 구성된 군으로부터 선택될 수 있다.

[0108] 따라서, 상기 검출가능한 모이어티는 방사성 원자를 포함하거나 또는 이로 구성될 수 있다. 상기 방사성 원자는 테크네튬-99m, 요오드-123, 요오드-125, 요오드-131, 인듐-111, 플루오린-19, 탄소-13, 질소-15, 산소-17, 인-32, 황-35, 중수소, 트리튬, 레늄-186, 레늄-188 및 이트륨-90으로 구성된 군으로부터 선택될 수 있다.

[0109] 대안적으로, 상기 결합 모이어티의 검출가능한 모이어티는 형광성 모이어티일 수 있다.

[0110] 본 발명의 제1 양태에 따른 방법에서 단계 (a) 및/또는 단계 (c)에서 제공되는 샘플은 각각 단계 (b) 및/또는 단계 (d) 이전에 처리되어, 상기 샘플에 존재하는 바이오마커를 비오틴으로 표지되도록 한다. 단계 (b) 및/또는 단계 (d)는 스트렙타비딘 및 검출가능한 모이어티 (예를 들면 형광성 모이어티)를 포함한 검출제를 사용하여 수행할 수 있다.

[0111] 따라서, 시험될 샘플 중 관심대상의 단백질은 처음으로 제1 결합제(들)을 사용해 단리 및/또는 고정될 수 있고, 그 후 상기 바이오마커의 존재 및/또는 상대적 양이 제2 결합제(들)을 사용하여 결정될 수 있다.

[0112] 일 구체예에서, 제2 결합제는 항체 또는 그의 항원-결합 단편, 통상적으로 재조합 항체 또는 그의 단편이다. 편리하게는, 항체 또는 그의 단편은 scFv; Fab; 면역글로불린의 결합 도메인 분자로 구성된 군으로부터 선택된다. 적합한 항체 및 단편, 및 이를 제조하기 위한 방법은 상기에 구체적으로 기재된다.

- [0113] 대안적으로, 제2 결합제는 아피바디 또는 압타머와 같은 항체-유사 결합제일 수 있다.
- [0114] 대안적으로, 시험될 샘플 중 단백질 상의 상기 검출가능한 모이어티가 특이적 결합쌍(specific binding pair)의 멤버(예, 비오틴)를 포함하거나 이로 구성되고, 제2 결합제는 특이적 결합쌍의 상보적 멤버(complimentary member) (예, 스트렙타비딘)를 포함하거나 이로 구성될 수 있다.
- [0115] 검출 분석이 사용되는 경우, 상기 검출가능한 모이어티는 형광성 모이어티; 발광성 모이어티; 화학발광성 모이어티; 방사성 모이어티; 효소 모이어티로 구성된 군으로부터 선택되는 것이 바람직하다. 본 발명의 방법에 사용하기 위한 적합한 검출가능한 모이어티의 예는 위에서 기재한 바와 같다.
- [0116] 혈청 또는 혈청 단백질을 검출하기 위한 바람직한 분석은 단일클론 및/또는 폴리클론 항체를 사용한 샌드위치 분석법 (sandwich assay)을 포함한 효소 결합 면역 흡착 측정법(enzyme linked immunosorbent assays, ELISA), 방사면역측정법 (radioimmunoassay, RIA), 면역방사측정 분석법 (immunoradiometric assays, IRMA) 및 면역효소 분석법 (immunoenzymatic assays, IEMA)을 포함한다. 예시적인 샌드위치 분석법은 미국 특허 제4,376,110호 및 제4,486,530호 중에 David *et al*에 의하여 기재되어 있고, 이들은 참조에 의해 본 명세서에 포함된다. 슬라이드 상의 항체 염색은 해당 기술분야의 통상의 기술자에게 잘 알려진 바와 같이, 세포학 실험 진단 시험 (cytology laboratory diagnostic test)에서 잘 알려진 방법에 사용될 수 있다.
- [0117] 따라서, 일 구체예에서 상기 분석법은, 보통 고체상(solid phase) 분석에서, 유색 반응 산물(coloured reaction product)을 제공하는 효소의 사용을 통상적으로 포함하는 ELISA이다. 호스라디시 피옥시다제 및 포스파타제와 같은 효소가 널리 사용되고 있다. 포스파타제 반응을 증폭하는 방식은 제2 효소 시스템을 위한 조효소로서 기능하는 NAD를 생성하는 기질로서 NADP를 사용하는 것이다. 에스케리키아 콜라이 유래의 피로포스파타제는 그 효소가 조직에 존재하지 않기 때문에 우수한 접합체(conjugate)를 제공하고, 안정하며, 우수한 반응색(reaction colour)을 제공한다. 루시페라제와 같은 효소에 기반한 화학발광 시스템이 또한 사용될 수 있다.
- [0118] 비타민 비오틴과의 접합은 비오틴의 우수한 특이성 및 친화성을 갖고 결합하는 효소-결합 아비딘 또는 스트렙타비딘과의 반응에 의해 용이하게 검출될 수 있기 때문에, 자주 사용된다.
- [0119] 대안적 구체예에서, 단백질 검출에 사용되는 분석은 편리하게는 형광적 분석이다. 따라서, 상기 제2 결합제의 검출가능한 모이어티는 Alexa 형광단 (예를 들면 Alexa-647)와 같은 형광성 모이어티일 수 있다.
- [0120] 바람직하게 ROU AUC에 의해 결정된, 본 발명의 예측(predicative) 정확도는 0.50 이상, 예를 들면 0.55 이상, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 0.96, 0.97, 0.98 또는 0.99 이상이다. 더 바람직하게 ROU AUC에 의해 결정된, 본 발명의 예측 정확도는 0.80 이상 (가장 바람직하게 1)이다.
- [0121] 본 발명의 제1 양태의 방법에서 단계 (b)는 비드-기반 어레이 또는 표면-기반 어레이와 같은 어레이를 사용하여 수행될 수 있다. 바람직하게 어레이는 마이크로어레이; 마이크로어레이; 나노어레이로 구성된 군으로부터 선택된다.
- [0122] 유방암-관련 질환 상태를 결정하는 방법은 <http://cran.r-project.org/web/packages/e1071/index.html> (e.g. e1071 1.5-24)로부터 이용가능한 것과 같은 서포트 벡터 머신 (support vector machine, SVM)을 사용하여 수행될 수 있다. 그러나, 다른 적절한 방법이 또한 사용될 수 있다. SVM은 또한 본 명세서에 정의된 하나 이상의 표 1 바이오마커를 포함하거나 이로 구성된 바이오마커 시그니처의 ROC AUC를 결정하는데 사용될 수 있다.
- [0123] 서포트 벡터 머신 (SVMs)은 분류(classification) 및 회귀(regression)를 위해 사용되는 관련 감독 학습법 (supervised learning method)의 한 세트이다. 각각 둘 중 하나의 카테고리에 속하는 것으로 표시되는, 한 세트의 훈련 예(training example)가 주어지면, SVM 훈련 알고리즘은 새로운 예가 하나의 카테고리 또는 다른 카테고리에 속하는지 여부를 예측하는 모델을 정립한다. 직관적으로, SVM 모델은 매핑된, 공간의 포인트로서 예를 표시하고, 따라서 개별적인 카테고리의 예는 가능한 넓은 명확한 갭으로 나뉜다. 그 다음 새로운 예가 동일한 공간으로 매핑되고 그들이 속하는 갭의 쪽에 기반한 카테고리에 속하는 것으로 예측된다.
- [0124] 더욱 형식적으로, 서포트 벡터 머신은 높은 또는 무한 차원 공간(dimensional space)에서 초평면(hyperplane) 또는 초평면 세트를 구성하는데, 이는 분류, 회귀 또는 다른 작업을 위해 사용될 수 있다. 직관적으로, 우수한 분리(separation)는 클래스의 가장 가까운 훈련 데이터 포인트(training data point)에서 가장 먼 거리에 있는 초평면(소위 기능적 여유도(functional margin))에 의해 달성되는데, 이는 일반적으로 여유도가 더 클수록 분류자(classifier)의 일반화 에러(generalization error)는 더 낮기 때문이다. SVM에 관한 더 많은 정보에 대해서는 예를 들어 Burges, 1998, 데이터 Mining and Knowledge Discovery, 2:121-167을 참조한다.

- [0125] 발명의 일 구체예에서, SVM은 알려진 시약(agent) (즉, 알려진 조직학적 등급의 유방암 세포 또는 알려진 원격 무전이 생존(distant metastasis-free survival)을 갖는 유방암 환자로부터의 유방암 세포)의 바이오마커 프로파일을 사용하여 본 발명의 방법을 수행하기에 앞서 '훈련(train)'된다. 이러한 훈련 샘플을 실시하여, SVM은 어떤 바이오마커 프로파일이 특정 특성과 연관되었는지를 알 수 있다. 일단 훈련 과정이 완료되면, SVM은 시험된 바이오마커 샘플이 특정 유방암 샘플 종류로부터 유래된 것인지 아닌지를 알 수 있다.
- [0126] 그러나, 이러한 훈련 방법은 SVM을 필요한 훈련 파라미터로 예비-프로그래밍(pre-programming)함에 의해 우회될 수 있다. 예를 들면, 특정 유방암-관련 질환 상태에 속하는 세포는 본 명세서에서 상세한 값 및/또는 조절 패턴(regulation pattern)을 사용하여, 표 1에 열거된 바이오마커의 측정에 기반하여, 표 4에 상세히 기재된 SVM 알고리즘을 사용하여 알려진 SVM 파라미터에 따라 수행될 수 있다.
- [0127] 적절한 SVM 파라미터는 데이터의 적절한 선택으로 SVM 머신을 훈련시킴에 의해 표 1에 열거된 바이오마커의 어떠한 조합에 대해서도 결정될 수 있다는 것을 통상의 기술자는 이해할 것이다(즉, 알려진 조직학적 등급 및/또는 알려진 무전이 생존 시간을 갖는 개체로부터의 세포로부터 바이오마커 측정).
- [0128] 대안적으로, 표 1 데이터는 주성분 분석(Principal Component Analysis, PCA) 및 기타 다변수 통계 분석 (예, 백워드(backward) 단계적 로지스틱 회귀 모델)과 같은, 해당 기술분야에 알려진 기타 적절한 통계 방법에 따라 특정 유방암-관련 질환 상태를 결정하는데 사용될 수 있다. 다변수 통계 분석을 위해, 예를 들면 참조에 의해 본 명세서에 포함된 Schervish, Mark J. (November 1987). "A Review of Multivariate Analysis". *Statistical Science* 2 (4): 396-413를 참조하라.
- [0129] 바람직하게, 본 발명의 방법은 65% 이상의 정확도, 예를 들면 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 정확도를 갖는다.
- [0130] 바람직하게, 본 발명의 방법은 65% 이상의 민감도(sensitivity), 예를 들면 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 민감도를 갖는다.
- [0131] 바람직하게, 본 발명의 방법은 65% 이상의 특이성(specificity), 예를 들면 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 특이성을 갖는다.
- [0132] "정확도"는 방법의 올바른 결과의 비율을 의미하고, "민감도"는 양성으로 올바르게 분류된 모든 양성 화학물질의 비율을 의미하며, "특이성"은 음성으로 올바르게 분류된 모든 음성 화학물질의 비율을 의미한다.
- [0133] 본 발명의 제1 양태의 방법은 하기 단계를 더 포함할 수 있다:
- [0134] e) 이전의 단계에서 결정된 유방암 관련 질환 상태에 기반하여 시험될 개체에 치료를 제공하는 단계.
- [0135] 따라서, 본 발명의 방법은 그 유방암의 조직학적 등급 및/또는 그 예측된 무전이 생존 시간에 따라 환자를 치료하는 것을 포함한다. 예를 들면, 더 공격적인 치료가 더 높은 등급 유방암에 대하여 제공될 수 있고, 및/또는 무전이 생존 기간은 상대적으로 낮음 (예, 10년 미만) 대 상대적으로 높음 (예, 10년 초과)으로 예측된다. 적절한 치료 시도가 그 시점의 우세한 안내(prevaling guidance)에 따라 통상의 기술자에 의해 결정될 수 있고, 예를 들면 참조에 의해 본 명세서에 포함된 NICE 임상의 Guideline 80 "Early and locally advanced breast cancer: Diagnosis and treatment" (available here: <http://www.nice.org.uk/nicemedia/pdf/CG80NICEGuideline.pdf>)을 참고하라.
- [0136] 따라서, 본 발명은 유방암 치료에 사용하기 위한 항신생물제(antineoplastic agent)를 포함하고, 투여량 요법(dosage regime)은 본 발명의 제1 양태의 방법의 결과에 기반하여 결정된다.
- [0137] 본 발명은 유방암 치료에서의 항신생물제의 용도를 포함하고, 투여량 요법은 본 발명의 제1 양태의 방법의 결과에 기반하여 결정된다.
- [0138] 본 발명은 유방암을 치료하기 위한 약제의 제조에서의 항신생물제의 용도를 포함하고, 투여량 요법은 본 발명의 제1 양태의 방법의 결과에 기반하여 결정된다.
- [0139] 본 발명은 충분한 양의 항신생물제를 제공하는 단계를 포함하는 유방암을 치료하는 방법을 포함하고, 상기 유방

암을 치료하기에 충분한 항신생물제의 양은 본 발명의 제1 양태의 방법의 결과에 기반하여 결정된다.

- [0140] 일 구체예에서, 항신생물제는 알킬화제 (ATC code L01a), 항대사물질 (ATC code L01b), 식물 알칼로이드 또는 기타 천연 산물 (ATC code L01c), 세포독성 항생제(cytotoxic antibiotic) 또는 관련된 물질 (ATC code L01d), 또는 기타 항신생물제 (ATC code L01x)이다.
- [0141] 따라서, 일 구체예에서 항신생물제는 질소 머스타드 유사체(nitrogen mustard analogue) (예, 시클로포스파미드, 클로람부실, 멜팔란, 클로르메틴, 이포스파미드, 트로스포스파미드, 프레드니무스틴 또는 벤다무스틴), 알킬 술포네이트(예, 부술판, 트레오술판, 또는 만노술판), 에틸렌 이민 (예, 티오테파, 트리아지쿠온(triaziquone) 또는 카르보쿠온(carboquone)), 니트로소우레아 (예, 카르무스틴, 로무스틴, 세무스틴, 스트랩토조신, 포테무스틴, 니무스틴 또는 라니무스틴), 에폭시드 (예, 에토글루시드) 또는 기타 알킬화제 (ATC code L01ax, 예, 미토브로니톨(mitobronitol), 피포브로만, 테모졸로미드 또는 다카르바진)로 구성된 군으로부터 선택된 알킬화제이다.
- [0142] 또 다른 구체예에서 항신생물제는 폴산 유사체(folic acid analogue) (예, 메토티렉세이트, 라티트렉세드, 페메트렉세드 또는 프랄라트렉세이트(pralatrexate)), 퓨린 유사체 (예, 메르캅토피린, 티오구아닌, 클라드리빈, 플루다라빈, 클로파라빈 또는 넬라라빈) 또는 피리미딘 유사체 (예, 시타라빈, 플루오로우라실, 테가푸르, 카르모푸르, 켄시타빈, 카페시타빈, 아자시티딘 또는 데시타빈)로부터 선택된 항대사물질이다.
- [0143] 추가적 구체예에서 항신생물제는 빈카 알칼로이드 또는 빈카 알칼로이드 유사체 (예, 빈블라스틴, 빈크리스틴, 빈데신, 비노렐빈 또는 빈플루닌), 포도필로톡신 유사체(podophyllotoxin derivative) (예, 에토포시드 또는 테니포시드), 콜히친 유사체(colchicine derivative) (예, 데메콜신), 타산 (예, 파클리탁셀, 도세탁셀 또는 파클리탁셀 폴리글루맥스(paclitaxel poliglumex) 또는 기타 식물 알칼로이드 또는 천연 산물 (ATC code L01cx, 예, 트라벡테딘(trabectedin))로 구성된 군으로부터 선택된 식물 알칼로이드 또는 기타 천연 산물이다.
- [0144] 일 구체예에서 항신생물제는 악티노마이신 (예, 닥티노마이신), 안트라시클린 또는 관련 물질 (예, 독소루비신, 다우노루비신, 에피루비신, 아클라루비신, 조루비신, 이다루비신, 미톡산트론, 피라루비신, 발루비신, 암루비신 또는 픽산트론) 또는 기타 (ATC code L01dc, 예, 블레오마이신, 플리카마이신, 미토마이신 또는 익사베필론)로 구성된 군으로부터 선택된 세포독성 항생제 또는 관련된 물질이다.
- [0145] 추가적 구체예에서, 항신생물제는 백금 화합물 (예, 시스플라틴, 카르보플라틴, 옥살리플라틴, 사트라플라틴 또는 폴리플라틸렌) 메틸히드라진 (예, 프로카르바진) 단일클론 항체 (예, 에드레콜로맙, 리톡시맙, 트라스투주맙, 알렘투주맙, 켄투주맙, 세톡시맙, 베바시주맙, 파니투무맙, 카투막소맙 또는 오파투무맙), 광역동/방사 요법에 사용되는 증감제(sensitizer)(예, 포르피머 소듐, 메틸 아미노레볼리네이트, 아미노레볼릭산, 테모포르핀 또는 에파프록시랄) 또는 단백질 키나아제 억제제 (예, 이마티닙, 게피티닙, 에프로티닙, 서니티닙, 소라페닙, 다사티닙, 라파티닙, 닐로티닙, 템시롤리무스, 에베롤리무스, 파조파닙, 반데타닙, 아파티닙, 마시티닙 또는 토세라닙)로부터 선택된 기타 항신생물제이다.
- [0146] 추가적 구체예에서, 항신생물제는 암사크린, 아스파라기나제, 알트레타민, 히드록시카르바미드, 로니다민, 펜토스타틴, 밀테포신, 마소프로콜, 에스트라무스틴, 트레티노인, 미토구아존, 토포테칸, 티아조퓨린, 이리노테칸, 알리트레티노인, 미토탄, 페가스파르가제, 백사로텐, 삼산화비소(arsenic trioxide), 테너루킨 티프티톡스, 보르테조밋, 셀레코픽, 아나그렐리드, 오블리메르센, 시티마젠 세라데노백, 보리노스타트, 로미렙신, 오마세탁신 메페속시네이트 또는 에리블린으로 구성된 군으로부터 선택된 기타 신생물제이다.
- [0147] 따라서, 본 발명의 제2 양태는 본 발명의 제1 양태에 따른 방법에 사용하기 위한 어레이를 제공하고, 상기 어레이는 본 발명의 제1 양태와 관련하여 위에서 정의된 바와 같은 하나 이상의 제1 결합제를 포함한다.
- [0148] 상기 어레이 결합제는 표 1A에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면, 표 1A에 정의된 군으로부터 선택된 2개 이상의 바이오마커에 총괄적으로 결합할 수 있는 결합제를 포함하거나 이로 구성될 수 있다. 바람직하게 어레이 결합제는 표 1B에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면, 표 1B에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30개 이상의 바이오마커에 총괄적으로 결합할 수 있는 결합제를 포함하거나 이로 구성될 수 있다. 바람직하게 어레이 결합제는 표 1C에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면, 표 1C에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27 또는 28개 이상의 바이오마커에 총괄적으로 결합할 수 있는 결합제를 포함하거나 이로 구성될 수 있다. 바람직하게 어레이 결합제는 표 1D에 정의된 군으로부터 선택된

택된 하나 이상의 바이오마커, 예를 들면, 표 1D에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8, 9 또는 10개 이상의 바이오마커에 총괄적으로 결합할 수 있는 결합체를 포함하거나 이로 구성될 수 있다. 바람직하게 어레이 결합체는 표 1E에 정의된 군으로부터 선택된 하나 이상의 바이오마커, 예를 들면, 표 1E에 정의된 군으로부터 선택된 2개 이상, 3, 4, 5, 6, 7, 8 또는 9개 이상의 바이오마커에 총괄적으로 결합할 수 있는 결합체를 포함하거나 이로 구성될 수 있다.

[0149] 따라서, 상기 어레이 결합체는 표 1A에 정의된 바이오마커 모든 것에 총괄적으로 결합할 수 있는 결합체를 포함하거나 이로 구성될 수 있다. 어레이 결합체는 표 1 B에 정의된 바이오마커 모든 것에 총괄적으로 결합할 수 있는 결합체를 포함하거나 이로 구성될 수 있다. 어레이 결합체는 표 1 C에 정의된 바이오마커 모든 것에 총괄적으로 결합할 수 있는 결합체를 포함하거나 이로 구성될 수 있다. 어레이 결합체는 표 1 D에 정의된 바이오마커 모든 것에 총괄적으로 결합할 수 있는 결합체를 포함하거나 이로 구성될 수 있다. 어레이 결합체는 표 1 E에 정의된 바이오마커 모든 것에 총괄적으로 결합할 수 있는 결합체를 포함하거나 이로 구성될 수 있다. 바람직하게 어레이 결합체는 표 1에 정의된 바이오마커 모든 것에 총괄적으로 결합할 수 있는 결합체를 포함하거나 이로 구성될 수 있다.

[0150] 상기 어레이의 제1 결합체는 고정됨(immobilised) 수 있다.

[0151] 상기 어레이 자체는 해당 기술분야에서 잘 알려져 있다. 통상적으로 각각은 고체 지지체의 표면에 형성된 유한 영역(finite area)을 갖는 간격을 두고 떨어진(즉, 별개의) 영역("스팟(spot)")을 갖는 선형 또는 2차원 구조로 형성된다. 어레이는 각 비드(bead)가 분자 코드 또는 컬러 코드에 의해 확인되거나 연속 흐름(continuous flow)에서 확인될 수 있는 것인 비드 구조체(bead structure)일 수 있다. 각각이 용액으로부터 분자의 클래스(class of molecule)를 흡착하는 일련의 스팟 위로 샘플을 통과시키는 것인 분석이 순차적으로 수행될 수도 있다. 고체 지지체는 통상적으로 유리 또는 폴리머이고, 가장 흔하게 사용되는 폴리머는 셀룰로스, 폴리아크릴아미드, 나일론, 폴리스티렌, 염화 폴리비닐 또는 폴리프로필렌이다. 고체 지지체는 튜브, 비드, 디스크, 실리콘 칩, 마이크로플레이트, 폴리비닐리덴 디플루오라이드(PVDF) 멤브레인, 니트로셀룰로스 멤브레인, 나일론 멤브레인, 기타 다공성 멤브레인, 비다공성 멤브레인(예를 들면, 특히, 플라스틱, 폴리머, 퍼스펙스, 실리콘), 복수의 폴리머 핀(polymer pin), 또는 복수의 마이크로타이터 웰, 또는 단백질, 폴리뉴클레오티드 및 다른 적합한 분자를 고정화시키고 및/또는 면역분석법을 실행하기에 적합한 다른 표면의 형태일 수 있다. 결합 과정은 당해 기술 분야에 잘 알려져 있고 일반적으로 고체 지지체에 단백질 분자, 폴리뉴클레오티드 등을 공유적으로 결합 또는 물리적으로 흡착시키는 가교로 구성된다. 대안적으로, 친화성-태그를 통한 프로브의 친화성 커플링(affinity coupling) 또는 유사한 구조체가 이용될 수 있다. 접촉 또는 비접촉 프린팅, 마스크 또는 포토리소그래피와 같은, 잘 알려진 기술을 이용하여, 각 스팟의 위치가 정의될 수 있다. 검토를 위하여, Jenkins, R.E., Pennington, S.R. (2001, *Proteomics*, **2**,13-29) 및 Lal 등 (2002, *Drug Discov Today* **15**:7(18 Suppl):S143-9)를 참조한다.

[0152] 통상적으로, 어레이는 마이크로어레이이다. "마이크로어레이"는 적어도 약 $100/\text{cm}^2$ 및 바람직하게는 적어도 약 $1000/\text{cm}^2$ 의 별개의 영역(discrete region)의 밀도를 갖는 영역의 어레이의 의미를 포함한다. 마이크로어레이의 영역은 전형적인 크기(typical dimension), 예를 들어, 약 10-250 μm 사이의 범위의 직경을 갖고, 대략 동일 거리에 의해 어레이의 다른 영역과 분리된다. 어레이는 또한 매크로어레이 또는 나노어레이일 수 있다.

[0153] 적절한 결합 분자(상기에서 논의됨)가 확인되고 단리되면, 통상의 기술자는 분자생물학 분야에서 잘 알려진 방법을 사용하여 어레이를 제조할 수 있다; 하기 실시예 참조.

[0154] 본 발명의 제3 양태는 표 1A, 표 1B, 표 1C, 표 1D, 및/또는 표 1D에서 정의된 군으로부터 선택된 하나 이상의 바이오마커의 유방암-관련 질환 상태를 결정하기 위한 용도를 제공한다.

[0155] 일 구체예에서 표 1A, 표 1B, 표 1C, 표 1D 및 표 1D에서 정의된 바이오마커의 모든 것은 유방암-관련 질환 상태를 결정하기 위해 총괄적으로 사용된다.

[0156] 본 발명의 제4 양태는 하기를 포함하는 본 발명의 제1 양태에 따른 방법에 사용하기 위한 분석 키트를 제공한다:

[0157] A) 본 발명의 제2 양태에 따른 어레이 또는 본 발명의 제1 양태에서 정의된 바와 같은 어레이; 및

[0158] B) (선택적) 본 발명의 제1 양태에 정의된 방법을 수행하기 위한 지시.

[0159] 상기 분석 키트는 본 발명의 제1 양태에 정의된 하나 이상의 대조군 샘플을 포함할 수 있다.

[0160] 바람직하게, 본 발명의 특정 양태를 이용하는 비제한적 실시예가 하기 표를 참조하여 기재될 것이다:

도면의 간단한 설명

[0161] **도 1.** 펩티드 및 단백질 통계. (A) 샘플당 확인된 고유 펩티드 서열(unique peptide sequence)의 총 개수 (FDR 0.01, Mascot + X!Tandem을 사용). (B) 샘플당 확인된 조립된 단백질군(assembled protein group)의 총 개수 (FDR 0.01, 단백질 수준에서 설정됨, Mascot + X!Tandem을 사용). (C) 전체 연구에서 2140개의 단백질군의 총 단백질 커버리지를 가져오는, 단백질군 당 고유 펩티드의 수 (FDR 0.01, 단백질 수준에서 설정, Mascot + X!Tandem을 사용). (복제(replicates), 풀 실시(pool runs) 및 누락된 임상 파라미터를 가진 샘플을 포함한, 모든 샘플 및 실시(run)에 기반한 데이터). (D) 펩티드아틀라스(PeptideAtlas)(버전 2011-08 Ens62, human)에 대하여 정량화된 펩티드의 평가 (Progenesis LC-MS 소프트웨어, FDR 0.01을 사용한 Mascot 점수화된(scored) 펩티드에 한정). 또한, 상기 펩티드아틀라스에 존재하지 않는 펩티드의 경우, 대응하는 단백질이 보고되었는지 평가하기 위해 제2 비교가 수행되었다. 다수 단백질 가입(accession)의 경우, 모든 것이 평가되었다. (E) 펩티드 길이의 비교. (F) 펩티드아틀라스에서 관찰된 펩티드의 빈도.

도 2. 대표적인 샘플 (샘플 7267개) 및 기준 (풀(pool)) 샘플에 대한 조합된 데이터 및 개별 혼합물 모두를 위해 도시된 전체 GPS 설정(setup)의 재현성 (즉, 포획(capture) + LC-MS/MS). (플롯) 데이터 포인트를 포함시키기 위해, 단백질이 모두 삼중(triplicate) 실시로 정량화되어야만 했다 (정규화된 존재빈도(abundance) > 0). 이러한 요건은 패널 A-E에 플롯된 모든 데이터에 대하여 사용되었다. (A) (1264개의 단백질에 기반한) 조합된 모든 데이터에 대하여 도시됨. (B) (315개의 단백질에 기반한) CIMS-mix 1에 대하여 도시됨 (C) (661개의 단백질에 기반한) CIMS-mix 2에 대하여 도시됨 (D) (452개의 단백질에 기반한) CIMS-mix 3에 대하여 도시 (E) (370개의 단백질)에 기반한 CIMS-mix 4에 대하여 도시됨.

도 3. 조직학적 등급에 기반하여, 유의성 있게 차등적으로 발현되는 단백질, 에스트로겐 수용체 상태(status), 및 HER2 상태. 차등적으로 발현되는 분석물이 히트맵에 보인다 (적색 - 상향-조절, 녹색 - 하향-조절). (A) 조직학적 등급 1, 등급 2 및 등급 3 샘플의 PCA-플롯 및 관련 히트맵(분산 0.2, p-값 <0.01, q-값 <0.25에서 여과된 데이터). 또한 ROC-면적 값으로 표시된, SVM을 이용한 리브-원 아웃 교차 검증 시도(leave-one out cross validation approach)로부터의 결과. (B) ER-양성 및 ER-음성 샘플의 PCA-플롯 및 관련 히트맵 (분산 0.2, p-값 <0.01, q-값 <0.32에 여과된 데이터). 또한 ROC-곡선으로 도시된 리브-원 아웃 교차 검증 시도로부터의 결과. (C) HER2-양성 및 음성 샘플의 PCA-플롯 및 관련 히트맵. (분산 0.2, p-값 <0.01, q-값 <0.9에서 여과된 데이터). ROC-곡선으로 도시된 리브-원 아웃 교차 검증 시도로부터의 결과.

도 4. IPA를 사용하여 3개의 조직학적으로 등급화된 종양 타입 간의 차등적으로 발현된 분석물의 생물학적 연관성. (A) 단백질의 세포 위치(cellular localization)에 대하여 매핑된 3개의 종양 집단 간에서 유의성있게 차등적으로 발현된 단백질로 확인된 49개의 단백질. 적색은 상향-조절을 나타내고 녹색은 하향-조절을 나타내는 것인 유색 log₂-비율 (중간값 등급 3 / 중간값 등급 1). 종양형성과 연관된 것으로 알려진 단백질을 표시하였다. (B) DNA 복제, 재조합, 세포 주기, 및 자유 라디칼 청소(scavenging)와 관련된 것으로 확인된 최고로 보고된 네트워크(top reported network). (C) 유전자 발현, 감염성 질환, 및 암과 관련된 것으로 확인된 두번째로 보고된 네트워크.

도 5. 정통적(orthogonal) 방법을 사용한 단백질 발현 프로파일의 검증. 이를 위하여, 1141개의 조직학적으로 등급화된 종양 샘플로부터의 데이터에 기반한 mRNA 발현 프로파일을 사용하였다. 조직학적 등급 1, 2 및 3 중 49개의 차등적으로 발현된 단백질의 42개를 GOBO-데이터베이스에 (Gene Entrez ID를 사용하여) 성공적으로 매핑하였다. (A) 조직학적 등급 3 종양(조직학적 등급 1과 비교된 중간값 비율(median ratio))에서 감소된 단백질 발현을 보이는 것으로 확인된 단백질에 대한 mRNA 발현 프로파일로서, (총 16개 중) 15개 분석물이 GOBO-도구로 매핑될 수 있다. 또한, 상이한 유전자 세트 모듈 발현 패턴(different gene set module expression pattern)에 대한 15개 유전자의 상관관계가 표시되어 있다. 회색 점은 실제 상관관계 값을 나타낸다. (B) (조직학적 등급 1에 비하여) 조직학적 등급 3 종양에서 증가된 발현을 디스플레이하는 것으로 확인된 단백질에 대한 mRNA 발현 프로파일로서, (총 33개 중) 27개가 GOBO-도구로 매핑될 수 있다. 또한, 상이한 유전자 세트 모듈 발현 패턴에 대한 27개의 유전자의 상관관계가 표시되어 있다. 회색 점은 실제 상관관계 값을 나타낸다.

도 6. 표 1 바이오마커의 다양한 길이 및 조합의 예시적 바이오마커 시그니처의 카플란 마이어(kaplan meier)

분석.

도 S1. 연구에 사용된 작업흐름의 도식적인 개요. (A) 종양 샘플 제조 및 (B) CIMS-항체를 사용한 펩티드 포획, LC-MS/MS의 실시 스케줄 및 데이터 분석. 하나의 CIMS-결합제 믹스(CIMS-binder mix)로부터 유래된 모든 용출액의 분석은 다음 CIMS-결합제 믹스 유래된 용출액으로 이동하기 전에 최종화하였다. 결과적으로, 모든 CIMS-결합제 믹스 분석을 집합된 샘플로부터의 용출액의 분석으로 시작하고, 조직학적 등급에 따른 임의적 서열 중 개별 샘플의 반으로 계속하였다. 믹스를 통한 중간에(halfway), 또 다른 집합 샘플(pool sample)을 주입하고, 그 후 잔여 샘플 및 끝에 제3 집합 샘플로 마무리하였다. 완료된 후, 다음 CIMS-결합제 믹스로부터의 용출액의 분석을 개시하였다. 결합제 믹스 2와 결합제 믹스 3 간에, 블랭크(blank) 비드의 2개의 주입을 실시하였다. 블랭크 비드는 부착된 항체를 포함하지 않아서, 오직 자성 비드에 결합한 백그라운드 비드가 용출되어야 한다. 데이터를 펩티드 및 단백질의 확인 및 정량을 수득하기 위한 Proteios SE 및 Progenesis에서 분석하였다.

도 S2. 벤다이어그램으로 도시된, 전체 GPS 설정(즉, 포획 + LC-MS/MS)의 확인 재현성. (A) 샘플 7267개에 대한 복제 포획 실시(replicate capture run) 간의 펩티드(모든 고유 서열)의 중첩. 샘플 7267개의 총 커버리지(상단 다이어그램) 및 개별 믹스(더 작은 4개의 벤다이어그램)에 대한 통계가 보인다. Proteios SE로부터 생성된 데이터(즉 Mascot 및 X!Tandem 점수화된 펩티드). (B) 집합 샘플에 대한 복제 포획 실시 간의 펩티드(모든 고유 서열)의 중첩. 전체(상단 다이어그램) 및 개별 믹스 모두에 대한 통계가 보인다. Proteios SE로부터 생성된 데이터(즉 Mascot 및 X!Tandem 점수화된 펩티드).

도 S3. 정량화된 단백질에 대한 log2 MS 강도의 분포. (A) 1364개의 단백질에 대하여 도시된(임상 기록을 갖는 50개의 샘플에 기반한) 중간값 정규화된 존재빈도(median normalized abundance) (0의 중간값 log2 강도값을 갖는 24개 단백질은 제외되었다). 막대(bar)는 MS 강도에 따라 유색으로 표시되고, 연한 노랑(낮은 MS 강도)으로부터 진한 빨강(높은 MS 강도)의 범위에 존재한다. (B) 선택된 단백질 카테고리에 대한 GO 생물학적 과정에 기반한 log2 MS 강도값의 분포. 분석물을 Generic Gene Ontology (GO) Term Mapper tool (<http://go.princeton.edu/cgi-bin/GOTermMapper>)를 사용하여 주요 생물학적 과정에 의해 그룹핑하였다.

도 S4. 조직학적 등급 3 종양에서 가장 높은 발현을 보여주는, 3개의 조직학적 등급 간에 8개의 차등적으로 발현된 단백질에 대한 개별 강도 박스플롯(boxplot).

도 S5. 조직학적 등급 1 종양에서 가장 높은 발현을 보여주는, 3개의 조직학적 등급 간에 8개의 차등적으로 발현된 단백질에 대한 개별 강도 박스플롯.

도 S6. 조직학적 등급 간의 연장된 비교 (A) 조직학적 등급 2 (H2)와 조직학적 등급 1 (H1) 간, 조직학적 등급 3 (H3)과 조직학적 등급 1 (H1) 간, 및 조직학적 등급 3 (H3)과 조직학적 등급 2 (H2) 간의 log2-배 변화. 톱(top) 49개의 도시된 분석물은 3개의 등급 간에 차등적으로 발현되어 확인된 단백질 시그니처이다. 따라서, 모든 비교가 계산되고 표시되어 있다. 더 낮은 47개의 분석물은 제3 등급이 제외된, 2개의 등급 간의 SVM-계산으로부터 도출된다. 이 계산을 모든 3개의 비교에 대하여 하였고, 유의성있는 분석물의 목록이 결과적으로 기록되었다. 매트릭스 색 수치를 Matrix2png (Pavlidis and Noble, 2003) 사용하여 생성하였다. (B) 2개의 그룹 비교로부터 SVM 계산으로부터 도출된 ROC AUC 값. 여과되지 않은, (전체 데이터세트) 및 여과된 데이터 모두로부터의 ROC AUC 값이 기재되어있다(분산 0.2 및 p-값 <0.01). (C) 조직학적 등급 1 및 등급 3의 히트맵(분산 0.2, p-값 <0.01, q-값 <0.25으로 여과된 데이터). 차등적으로 발현된 분석물이 적색이 상향-조절을 도시하고 녹색이 하향-조절을 도시하는 것인, 히트맵에 보인다. (D) 등급 1 및 등급 3 간의 50개의 차등적으로 발현된 단백질을 사용한 조직학적 등급 1, 등급 2 및 등급 3의 PCA-플롯(도 S8C).

도 S7. ER-상태 비교 또는 HER2/*neu*-상태 비교에서 차등적으로 발현된 것으로 확인된 단백질의 서브세트에 대하여 예시된 개별 강도 박스플롯. (A) ER-양성 및 ER-음성 종양 간의 차등적으로 발현된 분석물. (B) HER2-음성 및 HER2-양성 종양 간의 차등적으로 발현된 분석물.

도 S8. Ki67-양성(25% cut off) 및 Ki67-음성 단계의(staged) 종양의 평가. 차등적으로 발현된 분석물이 적색이 상향-조절을 도시하고 녹색이 하향-조절을 도시하는 것인, 히트맵에 보인다. (A) Ki67-양성 및 Ki67-음성 단계의 종양의 PCA-플롯. 상응하는 분석물 및 샘플의 히트맵. (분산 0.2, p-값 <0.01, q-값 <0.27으로 여과된 데이터). (B) ROC-곡선으로 도시된 SVM을 갖는 리브-원 아웃 교차 검증 시도로부터의 결과.

도 S9. 조직학적 등급 또는 ER-상태를 반영하는 차등적으로 발현된 분석물에 대하여 IPA를 사용한 전사 인자 연관 네트워크 분석. 분자를 연결하는 선은 분자적 관계를 나타내고 화살표의 스타일은 특이적 분자적 관계 및 상호작용의 방향성을 나타낸다. (A) 다군(multi-group) 조직학적 등급 비교에서 확인된 49개의 단백질을 입력

(input)으로서 사용하였다. 측정된 분석물에 색상 코드를 부여하기 위해 조직학적 등급 3 vs 조직학적 등급 2에 대한 중간값의 log2 비율이 사용되었다. 적색은 상향 조절을 도시하고 녹색은 하향-조절을 도시한다. (B) ER-상태 비교에서 확인된 39개의 단백질을 입력으로서 사용하였다. 측정된 분석물에 색상 코드를 부여하기 위해, 상기 중간값의 log2 비율이 사용되었다. 적색은 ER-음성 샘플 중 상향 조절을 도시하고 녹색은 하향-조절을 도시한다.

도 S10. 조직학적 등급 간 유의성있게 차등적 단백질 발현을 보이는 것으로 확인된 분석물의 서브세트에 대하여 예시된 1141개의 조직학적 등급화된 종양 샘플로부터의 데이터에 기반한 개별 mRNA 발현 프로파일. (A-E) 조직학적 등급 3 종양에서 증가된 발현을 보이는 것으로 확인된 5개의 단백질에 대한 mRNA 발현 수준. (F-J) 조직학적 등급 3 종양에서 감소된 발현을 보이는 것으로 확인된 5개의 단백질에 대한 mRNA 발현 수준.

도 S11. 1620개의 ER-상태 정의된 유방 종양 샘플로부터의 데이터에 기반한 mRNA 발현 프로파일. 39개의 차등적으로 발현된 단백질 중 32개가 gene entrez ID를 사용한 GOBO-데이터베이스로 성공적으로 매핑되었다. (A) ER-양성 종양에서의 증가된 단백질 발현을 나타내는 것으로 확인된 10개의 단백질에 대한 mRNA 발현 프로파일. 또한 상이한 유전자 세트 모듈 발현 패턴에 대한 10개의 유전자의 상관관계가 도시된다. 회색 점은 실제의 상관관계 값을 나타낸다. (B) ER-양성 종양에서의 감소된 발현을 나타내는 것으로 확인된 단백질에 대한 mRNA 발현 프로파일. 또한 상이한 유전자 세트 모듈 발현 패턴에 대한 22개의 유전자의 상관관계가 보일 수 있다. 회색 점은 실제의 상관관계 값을 나타낸다. (C-D) ER-양성 종양에서의 증가된 발현을 나타내는 것으로 확인된 2개의 단백질에 대하여 예를 들은 개별 mRNA 발현 프로파일. (E-F) ER-양성 종양에서의 감소된 발현을 나타내는 것으로 확인된 2개의 단백질에 대하여 예를 들은 개별 mRNA 발현 프로파일.

도 S12. HER2/neu 비교에서 유의성있는 차등적 단백질 발현을 보이는 것으로 확인된 3개의 분석물에 대하여 예를 들어진, GOBO-데이터베이스 도구를 사용하여, 매핑된 개별 mRNA 발현 프로파일. 1881개의 이용가능한 종양 샘플에 기반한 데이터. (A) HER2/neu, (B) S100A9 (C) GRB7. 앞서 3개의 분석물에 더하여 제4 단백질(accession P22392)이 매핑되기 위해 테스트되었다. 그러나, Gene Entrez ID 4831 또는 654364를 사용한 GOBO-데이터베이스 도구로부터의 에러 메시지로 인해, 데이터가 빠졌다.

도 S13. 10년 종점(endpoint)으로서 DMFS를 사용한 카플란-마이어 분석. 조직학적 등급을 구별하는 49개의 단백질 중 42개를 Entrez Gene ID를 사용한 유전자 발현 데이터베이스에 성공적으로 매핑하였다 (swissprot ID로 변환한 후). 분석물을 (관찰된 단백질 발현 수준에 대하여 조직학적 등급 3 및 등급 1 샘플 간의 비율을 사용하여, 상향- 또는 하향-조절에 기반하여) 2개의 군으로 나누었고, 15개의 하향-조절된 분석물 및 27개의 상향-조절된 분석물을 가져왔다. 이들 2개의 군을 그 후 유전자 발현 데이터 세트를 사용하여 원격 전이 무생존(distant metastasis free survival, DMFS)의 잠재적 위험을 평가하기 위해 사용하였다. 카플란-마이어 분석을 유전자 발현 데이터를 조직학적 등급화된 종양(n = 1379)에 대한 10년 종점으로서의 DMFS를 사용하여, 3개의 분위수(quantile)로 유전자 발현 데이터를 계층화함으로써 수행하였다. 또한, (2개의 하향-조절된 및 2개의 상향-조절된) 단일 유전자에 기반한 10년 종점으로서의 DMFS를 사용한 4개의 개별 카플란-마이어 분석을 생성하고 GOBO-도구를 사용하여 유사한 방식으로 표시하였다.

도 7. 유방암 조직 샘플을 6개의 등급 1 샘플, 9개의 등급 2 샘플 및 6개의 등급 9 샘플을 포함한, 52개의 샘플의 동일한 원래의 집단으로부터 선택하였다. 샘플을 용액 중 소화(트립신화)하고 선택적 모니터링 반응(Selective Monitoring Reaction, SRM) 세트-업(set-up) (확립된 질량 분석법 기반 시도)을 사용하여 분석하였다. 명시된 목록의 바이오마커로부터의 8개 단백질에 상응하는 9개의 펩티드를 표적화하고 정량화하였다. 샘플을 3중으로 실시하였다. 데이터를 Anubis 및 뒤이어 P-값 여과 ($p < 0.01$) 및 q-값 여과 ($q < 0.11$)를 사용해 분석하였다. 데이터는 유방암 조직 샘플이 마커의 절단된(truncated) 목록을 사용한 등급에 따라 구별될 수 있다는 것을 보여주었다.

도 8. 유방암 조직 샘플을 등급 1, 2 및 3 중 펼쳐진(spread) (기술적 복제(technical replicate)를 갖는) 47개의 샘플을 본 명세서에 포함한, 52 샘플의 동일한 원래 집단으로부터 선택하였다. 샘플을 겔에서 소화(트립신화)하고, 선택적 모니터링 반응 (SRM) 세트-업 (확립된 질량 분석법 기반 시도)을 사용하여 분석하였다. 바이오마커의 명시된 목록으로부터 4개의 단백질에 대응하는 8개의 펩티드를 표적화하고 정량화하였다. 샘플을 2중(duplicate)으로 실시하였다. 데이터를 Anubis 및 뒤이어 P-값 여과 ($p < 0.01$) 및 q-값 여과 ($q < 0.009$)를 사용해 분석하였다. 데이터는 유방암 조직 샘플이 마커의 절단된(truncated) 목록을 사용하여 등급에 따라 구별될 수 있다는 것을 보여주었다.

발명을 실시하기 위한 구체적인 내용

[0162] 실시예

[0163] 도입

[0164] 유방암 환자 중 종양 진행 및 예후는 현재 임상 및 실험 파라미터를 사용하여 평가하기 어렵고, 다중다수의 (multiplex) 조직 바이오마커 시그니처가 존재하지 않는다. 이러한 임상의 충족되지 않는 요구를 해결하기 위한 시도에서, 본 발명자는 글로벌 프로테오믹스 조사 (global proteome survey)로 나타낸, 최근에 개발된 프로테오믹스 발견 도구를 적용하였다. 따라서, 단 9개만의 항체에 기반한 친화성 프로테오믹스 및 무-표지 LC-MS/MS를 조합함으로써, 본 발명자는 52개의 유방암 조직 샘플을 프로파일하여, 가장 큰 유방암 조직 프로테오믹스 연구 중 하나를 나타내고, 1388개의 단백질을 나타내는 상세화된 정량화된 프로테오믹스 맵을 성공적으로 생성하였다. 그 결과는 본 발명자가 종양 진행을 반영하는 조직학적 등급화된 유방암 종양의 깊이 있는(in-depth) 분자적 묘사를 밝혔다는 것을 보여주었다. 더 상세하게는, 높은 정확도로 조직학적 등급 1 내지 3 유방암 종양을 구별하는 49-플렉스(49-plex) 조직 바이오마커 시그니처 ($p < 0.01$ 임) 및 79-플렉스 조직 바이오마커 ($p < 0.02$ 임) 시그니처가 정의되었다. 고도로 생물학적으로 관련있는 단백질이 확인되고, 차등적으로 발현된 단백질이 종양 진행에 대한 종양 미세환경의 리모델링에 관한 현재 가설을 지지하였다. 또한, 원격 전이 무 생존(distant metastasis free survival)의 위험을 평가하기 위한 마커 사용이 또한 보여졌다. 또한, ER-, HER2-, 및 Ki67-상태를 반영하는 유방암 관련 바이오마커 시그니처가 각각 기술되었다. 바이오마커 시그니처를 각각, 독립적 방법 (mRNA 프로파일링) 및 환자 집단을 사용하여 협조하였다. 종합하면, 이들 분자 묘사(molecular portrait)는 개선된 유방암의 분류 및 예후를 제공한다.

[0165] 실험 과정(Experimental Procedures)

[0166] 임상의 샘플

[0167] 이 연구는 스웨덴, Lund에서 지역 윤리 리뷰 위원회(regional ethics review board)에 의해 승인되었다. 52명의 유방암 환자를 Department of Oncology (SUS, Lund)로부터 모집하였다. 완전한 임상 기록은 조직 샘플의 50개에 대하여 입수가가능하였다. 샘플을 조직학적 등급 1 ($n=9$), 등급 2 ($n=17$), 및 등급 3 ($n=24$)에 기반하여 세분화하였다.

[0168] 트립신-소화된 인간 유방암 조직 샘플의 제조

[0169] 단백질을 52개의 유방암 조직 조각으로부터 추출하고 그 후 환원, 알킬화, 트립신 소화, 및 최종적으로 추가 사용때까지 -80°C 에서 저장하였다. 또한, 참조 샘플로서 사용된, 통합 샘플(pooled sample)을 모든 소화된 샘플로부터 $5\ \mu\text{L}$ 분취량(aliquot)을 합해 생성하고, 추가 사용 때까지 -80°C 에서 저장하였다. 샘플 제조에 대한 상세사항은 보충 실험 절차(Supplemental Experimental Procedures)에 제공된다.

[0170] CIMS-scFv 항체의 제조 및 자성 비드에의 커플링

[0171] 6개의 짧은 C-터미널 아미노산 펩티드 모티프에 대하여 지향된 9개의 CIMS scFv 항체 (표 S2)를 대장균(*E. coli*) 배양에서 생산하였고, Ni^{2+} -NTA 친화성 크로마토그래피를 사용하여 정제하였다. 그 후, 정제된 항체를 자성 비드에 결합하였다. scFv 제조 및 커플링에 대한 상세사항은 보충 실험 절차에 제공된다.

[0172] 무-표지 정량적 GPS 실험

[0173] 항체-접합된 비드의 4개의 상이한 풀(pool) (CIMS-결합제 믹스 1 내지 4를 나타냄)을 2개 또는 3개의 상이한 결합제(binder)의 동일한 양을 혼합하여 제조하였다 (표 S2). 항체 믹스를 트립신 작용의 샘플(tryptic sample)에 노출하고, 세척하고, 마지막으로 아세트산과 함께 인큐베이션하여 포획된 펩티드를 용출하였다. 용출액을 그 후 추가적 클린 업(clean up) 없이 MS-분석을 위하여 직접적으로 사용하였다. 완전한 연구는 26일의 MS-기기화 시간(MS-instrumentation time)을 사용하여 실시하고, 6.5일의 4개의 블록으로 나누었다 (하나의 CIMS-결합제 믹

스/블록). 모든 샘플을 CIMS-binder mix 당 1회 개별적으로 분석하였다. 또한, 선택된 샘플의 3중(triplicate) 포획을 백투백(back-to-back) LC-MS/MS 실시로서 각각의 블록 내 수행하였다. 기준 샘플을 4 블록 내 및 사이에 걸쳐 반복적으로 분석하였다 (도 S1). 총 238 LC-MS/MS 실시를 수행하고 펩티드 포획 및 관련 질량 분석법에 대한 상세 사항을 보충 실험 과정(Supplemental Experimental Procedures)에 제공한다.

[0174] 단백질 확인 및 정량

[0175] 생성된 데이터를 두개의 소프트웨어 패키지, Proteios SE (Hakkinen et al., 2009) 및 Progenesis LC-MS (Nonlinear Dynamics, UK)에 의해 분석하였다. 검색을 수행하였다. 검색은 펩티드 확인을 생성하기 위해 확인된 리버스 히트(reverse hit)의 수에 근거하여 평가된 0.01의 오류발견율(false discovery rate, FDR)로 포워드 및 리버스 조합된 데이터베이스(*Homo Sapiens* Swiss-Prot, Aug-2011, 총 71324 데이터베이스 엔트리를 초래)에 대하여 수행하였다. Progenesis-LC-MS 소프트웨어 (v 4.0)을 특징 정렬, 확인 (Mascot), 및 정량값 생성을 위해 사용하였다. 검색 파라미터 및 데이터 가공에 관한 세부사항은 보충 실험 과정에 제공된다.

[0176] 통계 및 바이오인포메틱스 분석

[0177] Qlucore Omics Explorer v (2.2) (Qlucore AB, Lund, Sweden)를 일 방향(one-way) ANOVA를 사용하여 상향- 또는 하향-조절된 단백질을 유의성있게 ($p < 0.01$) 확인하기 위해 사용하였다. q-값을 Benjamini 및 Hochberg 방법 (Benjamini and Hochberg, 1995)에 기반하여 생성하였다. 주성분 분석(Principal component analysis, PCA) 플롯 및 히트맵을 Qlucore에서 생성하였다. 서포트 벡터 머신 (SVM)은 리브-원-아웃 교차 검증(leave-one-out cross-validation) 과정을 사용해 샘플을 분류하기 위해 사용되었던 학습(learning) 방법이다 (Cortes and Vapnik, 1995)고, 분석물을 여과되지 않은 데이터 및 p-값 여과된 데이터 모두에 대하여 수행하였다. SVM 결정값(decision value)을 사용하여 구축된 수용자 운영 특성(receiver operating characteristics, ROC) 곡선 (Lasko et al., 2005) 및 곡선하면적 (AUC)을 분류자의 수행(performance of the classifier)의 척도로서 사용하였다. 또한, Ingenuity Systems Pathway Analysis (IPA) (v 11904312, www.ingenuity.com)를 단백질 위치(localisation), 잠재적 네트워크 상호작용, 전사 인자 연관, 및 종양형성(tumorigenesis)과의 관련과 같은 정보를 추출하기 위해, 유의성있게 차등적으로 발현되는 단백질에 대하여 사용하였다. 실험적으로 도출된 단백질 시그니처를 GOBO 검색 도구(Ringner et al., 2011)를 사용하여 조직학적 등급 1, 2 및 3, ER-상태 또는 HER2-상태와 같은 임상적 파라미터로 유방암 조직에 대한 공개된 유전자 발현 데이터의 큰 집단에 대하여 mRNA 수준에서 최종적으로 검증하였다.

[0178] 보충 실험 절차

[0179] 트립신-소화된 인간 유방암 조직 샘플의 제조

[0180] 단백질을 유방암 조직 조각으로부터 추출하고, 사용할 때까지 -80°C 에서 저장하였다. 간단한게는, 조직 조각 (약 50 mg/샘플)을 2개의 흔들기 라운드(shaking round)간의 액체 질소에서의 신속한 냉각으로 2 x 30초 동안 진탕기에서 봄베(bomb)를 고정하여, Teflon 용기에 균질화시키고, 액체 질소에서 예비-냉각시켰다(pre-cool). 균질화된 조직 분말을 8 M 요소, 30 mM Tris, 5 mM 마그네슘 아세테이트 및 4% (w/v) CHAPS (pH 8.5)을 함유한 분해(lysis) 버퍼 (2 mg 조직/30 μl 버퍼)에서 수집하였다. 튜브를 단시간동안(briefly) 볼텍싱하고 매 5분 샘플의 단시간 볼텍싱하며, 40분 동안 얼음에서 인큐베이션하였다. 인큐베이션 후, 샘플을 13000 rpm에서 원심분리하고, 상청액을 새로운 튜브로 이동하고 뒤이어 제2 원심분리를 하였다. 단백질 농도를 총 단백질 키트, Micro Lowry (Sigma, St. Louis, MO, USA)를 사용하여 결정하기 전에, 버퍼를 Zeba desalting spin column (Pierce, Rockford, IL, USA)을 사용하여 0.15 M HEPES, 0.5 M Urea (pH 8.0)로 교체하였다. 최종적으로, 샘플을 분취하였고 추가 사용 때까지 -80°C 에서 보관하였다. 단백질 추출물을 해동하고, 환원하고, 알킬화하고 트립신 소화하였다. 처음, SDS 및 TCEP-HCl (Thermo Scientific, Rockford, IL, USA)을 각각 0.02% (w/v) 및 5 mM으로 첨가하고, 샘플을 56°C 에서 60분 동안 환원하였다. 샘플을 상온으로 냉각시킨 후 요오드아세트아미드를 10 mM에 첨가하고 그 후 상온에서 30분 동안 알킬화시켰다. 다음, 시퀀싱-등급 변형된 트립신 (Promega, Madison, Wisconsin, USA)을 37°C 에서 16시간 동안 20 $\mu\text{g}/\text{mg}$ 의 단백질에 첨가하였다. 완전한 소화를 확인하기 위해, 제2 분취량의 트립신 (10 $\mu\text{g}/\text{mg}$ 단백질)을 첨가하고 튜브를 37°C 에서 추가적 3시간 동안 인큐베이션하였다. 최종적으로, 소화된 샘플을 분취하고 추가 사용때까지 -80°C 에서 저장하였다. 또한, 모든 소화된 샘플로부터의 5 μl

분취량을 조합하여 생성된, 분리된 집합적 샘플을 제조하고 추가적 사용 때까지 -80℃에서 저장하였다. 잠재적 잠정적 프로테오믹 커버리지를 증가시키기 위해, 한정된 임상의 데이터가 가까이 표 S1에 있는 2개의 샘플을 여전히 개별적으로 분석하고 집합적 샘플에 포함시켰다.

[0181] CIMS-scFv 항체의 제조 및 자성 비드에의 접합

[0182] 9개의 CIMS scFv 항체 (6개의 짧은 C-말단 아미노산 펩티드 모티프(M-1, M-15, M-31, M-32, M-33, 및 M-34 표시)에 대하여 지향된, 클론 1-B03, 15-A06, 17-C08, 17-E02, 31-001-D01, 32-3A-G03, 33-3C-A09, 33-3D-F06 및 34-3A-D10)는 n-CoDeR (Soderlind et al., 2000) 라이브러리로부터 선택되었고, BioInvent International AB, Lund, Sweden(표 S2)에 의해 친절하게(kindly) 제공되었다. CIMS 항체 중 6개에 대한 특이성 및 해리 상수 (낮은 μM 범위)가 최근 결정되었다 (Olsson et al., 2011). 상기 항체를 100 ml 대장균(*E. coli*) 배양에서 생산하고 Ni^{2+} -NTA 아가로스 상의 친화성 크로마토그래피를 사용하여 정제하였다 (Qiagen, Hilden, Germany). 결합된 분자를 250 mM의 이미다졸로 용출하고, 72시간 동안 PBS(pH 7.4)에 대하여 투석하고 그 후 사용시까지 + 4℃에서 저장하였다. 단백질 농도를 280 nm에서의 흡광도를 측정하여 결정하였다. scFv 항체의 보전성(integrity) 및 순도를 Agilent Bioanalyzer (Agilent, Waldbronn, Germany) 상의 단백질 80 칩을 실시(run)하여 확인하였다. 상기 정제된 scFv를 개별적으로 이전에 기재된 자성 비드(M-270 카복실산-활성화된 것, Invitrogen Dynal, Oslo)에 접합하였다 (Olsson et al., 2011). 간단하게는, 180-250 μg 정제된 scFv의 배치(batch)를 ~9 mg (300 μl)의 자성 비드에 공유결합으로 접합하고(EDC-NHS 화학), 추가 사용시까지 4℃에서 PBS 중 0.005% (v/v) Tween-20 중에 저장하였다. 또한 블랭크 비드의 배치를 생성하였다 (즉, scFv 추가 없는 접합 프로토콜에 의해 생성된 비드).

[0183] 표지가 없는 정량적 GPS 실험

[0184] 접합된 비드의 4개의 상이한 집합(pool) (CIMS-결합제 믹스(CIMS-binder mix) 1 내지 4라고 표시)를 하기에 따라 2 내지 3개의 상이한 결합제의 동일한 양을 혼합함으로써 제조하였다: 믹스 1 (CIMS-33-3D-F06 및 CIMS-33-3C-A09), 믹스 2 (CIMS-17-C08 및 CIMS-17-E02), 믹스 3 (CIMS-15-A06 및 CIMS-34-3A-D10) 및 믹스 4 (CIMS-1-B03, CIMS-32-3A-G03, 및 CIMS-31-001-D01) (표 S2). 각각의 포획을 위하여, 50 μl 의 집합된 비드 용액(pooled bead solution)을 사용하고 scFv-비드를 전혀 재사용하지 않았다. 상기 비드를 350 μl PBS로 예비세척한 후 35 μl 의 최종 부피(PBS로 희석하고 1 mM의 최종 농도까지 페닐메틸설포닐 플루오리드 (PMSF)의 첨가) 중에 트립신 샘플 소화물(trypsin sample digest)에 노출하고, 그 후 온화한(gentle) 혼합과 함께 20분 동안 상기 비드와 함께 인큐베이션하였다. 다음으로, 상기 튜브를 자석 상에 놓고, 상청액을 제거하고, 상기 비드를 각각 100 및 90 μl PBS로 세척하였다 (비드를 세척 단계 사이에 새로운 튜브로 이동하고 전체 세척 시간은 5분이었다). 최종적으로, 비드를 포획된 펩티드를 용출하기 위해 2분 동안 9.5 μl 의 5% (v/v) 아세트산 용액과 함께 인큐베이션하였다. 그 후, 용출액을 추가적 클린 업 없이 질량분석법(mass spectrometry) 분석을 위하여 직접적으로 사용하였다.

[0185] Eksigent nanoLC 2DTM plus HPLC system (Eksigent technologies, Dublin, CA, USA)와 접속된 ESI-LTQ-Orbitrap XL 질량 분석계(Thermo Electron, 브레멘, 독일)를 모든 샘플에 대하여 사용하였다. 오토샘플러(auto-sampler)는 6 μl 의 GPS-생성된 용출액을 주입하였다. 블랭크 LC-MS/MS 실험을 각각의 분석된 샘플 사이에 사용하였다. 펩티드를 프리-컬럼 (PepMap 100, C18, 5 μm , 5 mm x 0.3 mm, LC Packings, Amsterdam, Netherlands) 상에 15 $\mu\text{l}/\text{min}$ 의 일정한 유속으로 로딩하였다. 펩티드를 뒤이어 Reprosil-Pur C18-AQ 레진 (3 μm Dr. Maisch, GmbH, 독일)이 내부에 패킹된, 용융 실리카 방출기(fused silica emitter), 75 μm x16 cm (PicoTip™ Emitter, New Objective, Inc. Woburn, MA, USA) 상에서 분리하였다. 펩티드를 300 nl/min의 유속으로, 0.1% (v/v) 포름산을 함유한, 물 중 3 내지 35% (v/v)의 아세트오니트릴의 35분 선형 농도구배(linear gradient)로 용출하였다. LTQ-Orbitrap을 Orbitrap-MS (m/z 400 내지 2000) 및 LTQ-MS/MS 취득(acquisition) 간의 자동화된 전환(automatically switch)으로의 데이터-종속 모드(data-dependent mode)에서 작동하였다. 4개의 MS/MS 스펙트럼을 각각의 FT-MS 스캔 당 선형 이온 트랩(linear ion trap per each FT-MS scan) 중에서 취득하고, 이는 내부 보정을 위한 lock mass option (m/z 445.120025)을 사용한 60,000 FWHM 명목상 분해능 설정(nominal resolution setting)에서 취득하였다. 동적 제외 목록(dynamic exclusion list)은 20초의 반복 지속기간(repeat duration) 및 120초의 최고 보유 기간(maximum retention period)을 갖는 두 개의 반복 카운트

(repeat count)를 사용한 500 엔트리에 한정되었다. 전구체 이온 전하 상태 스크리닝(precursor ion charge state screening)은 2 이상의 전하를 갖는 이온을 위해 선택하고 결정되지 않은 전하 상태를 갖는 이온을 버릴 수 있게 한다. 정규화된 충돌 에너지를 35%에 설정하고, 하나의 마이크로 스캔을 각각의 스펙트럼에 대하여 취득하였다. 모든 샘플을 CIMS-결합제 믹스 당 한번 개별적으로 분석하였다. 또한, (상기 연구 중 모든 샘플에 기반한) 집합된 샘플의 삼중 포획을 각각의 CIMS-결합제 믹스에 대하여 수행하였고 더 긴 시간 기간(결합제 믹스 당 LC-MS 시퀀스 실시 순서(LC-MS sequence run order per binder mix)의 개시, 중간 및 마지막)에 걸쳐 MS-분석을 위하여 분배하였다 (도 S1). 이는 CIMS-결합제 믹스 1 및 4를 위하여 가능하였다. 그러나, CIMS-결합제-믹스 2 및 3을 위한 시퀀스 실시의 중간(halfway)보다 많은 것에서, 분석 LC-컬럼은 (2번) 교체될 필요 있고 새로 교체된 컬럼에 대하여 직접적으로 계획된 후속 풀 런(last pool run)을 실시하는 것을 결정하였고 일부 (11개 각각 9 샘플(11 respectively 9 sample))를 상기 풀 런 후에 분석하였다. 또한, 3중 포획을 각각의 CIMS-결합제 믹스에 대한 샘플 (7267, 8613)에 대하여 수행하였다. 잠재적 비드 백그라운드 결합 펩티드를 평가하기 위해, 블랭크 비드, 즉 집합된 항체를 갖지 않는 비드를 집합된 소화물에 노출시켰다. 2개의 블랭크 비드 "포획"으로부터 확인된 백그라운드 결합 펩티드의 낮은 수에 기반하여, 생성된 모든 데이터를 달리 명시되지 않으면 여과되지 않은 상태로 두었다.

[0186] 단백질 확인 및 정량

[0187] Mascot 및 X!Tandem을 사용하여 확인을 생성하기 위해, 생성된 데이터를 Proteios SE를 사용하여 처음(first) 분석하였다. 간단한게는, 모든 파일을 가공하고 Proteios (v 2.17) 플랫폼을 사용하여 mzML 및 mgf 포맷으로 전환하였고 이후 검색 파라미터를 Mascot 및 X!Tandem 효소: 트립신; 누락된 절단(missed cleavages) 1; 고정된 변형(fixed modification): 카르바미도메틸 (C); 가변 변형(variable modification): 메티오닌 산화 (O)을 위하여 사용하였다. 또한, 가변 N-아세틸을 X!Tandem (www.thegpm.org/tandem/)에서 수행되는 검색을 위하여 가능하게 하였다. 3 ppm의 펩티드 질량 허용값(tolerance) 및 0.5 Da의 단편 질량 허용값을 사용하고 검색을 포워드 및 리버스 조합된 데이터베이스에 대하여 수행되었다 (*Homo Sapiens* Swiss-Prot, Aug-2011, 총 71324개의 데이터베이스 엔트리를 초래함). Mascot 및 X!Tandem에서의 자동화된 데이터베이스 검색 및 결과적으로 조합 (0.01의 오류발견율 (FDR)임)이 펩티드 확인을 발생하기 위해 사용되었다 (확인된 리버스 히트(reverse hit)의 수에 기반하여 확립됨). Proteios SE를 사용하여 각각의 샘플에 대하여 단백질 확인을 생성하는 경우, 단백질 수준 상에 0.01의 FDR이 적용되었다. 모든 가공되지 않은(raw) 데이터는 Proteios SE 내에 저장된다.

[0188] 분석의 시점에 Proteios SE는 정량적 무-표지 플러그-인 분석 모듈(quantitative label-free plug-in analyzing module) (개발 진행중)을 제공하지 않기 때문에, Progenesis-LC-MS 소프트웨어 (v 4.0)를 모든 양적인 값을 생성하기 위해 사용하였다. 간단한게는, Progenesis-LC-MS 소프트웨어를 사용하기 전에 가공되지 않은 데이터 파일을 ProteoWizard 소프트웨어 패키지를 사용하여 mzXML로 전환하였다. 디폴트 설정(default setting) 및 최소 입력(minimal input)을 갖는, 내장(built-in) 특징 검색 도구(finding tool), Mascot 검색 도구 및 조합된 분획 도구(fractions tool) (CIMS-binder-mix 1, 2, 3 및 4)가 사용되었다. 최적 특징 정렬을 위해, 개별적 CIMS-결합제 믹스에 대하여, 집합된 샘플(pooled sample)의 제1 주입 실시(injection run)는 중간 풀 실시(halfway pool run)가 기준 정렬 파일로 사용되었던, CIMS-믹스 3 실시를 제외하고, 기준 정렬 파일로서 사용되었다. CIMS-결합제 믹스 1 및 2에 대한 보유시간 10-50분간 및 CIMS-결합제 믹스 3 및 4에 대한 10-49분간에, 정렬되고 검출된 특징이 정량을 위해 포함되었다. 생성된 정규화된 존재빈도값(abundance value)을 추출하고 통계 및 바이오인포메틱스 분석을 위해 사용하였다. Progenesis 소프트웨어에 의한 제한으로 인해, 상기 확인은 Mascot 검색에만 제한되었고, 이는 Proteios SE로부터의 X!Tandem 생성된 확인이 다운스트림 정량 분석에 대하여 포함되지 않았다는 것을 의미한다. 상기 언급한 바와 같이 동일한 데이터베이스 (*Homo Sapiens* Swiss-Prot, Aug-2011, 포워드 및 리버스 조합된 데이터베이스) 및 검색 파라미터를 사용하였고, 0.01의 컷-오프(cut-off) FDR 값을 적용하였다.

[0189] 결과

[0190] 본 연구에서, 52개의 조(crude) 유방암 조직 추출물의 세미-글로벌 단백질 발현 프로파일 (확인 및 정량)을 GPS를 사용하여 판독하였다. 조직적 등급뿐만 아니라, 에스트로겐 수용체 (ER), HER2, 및 Ki-67과 같은 다른 핵심적 임상 실험 파라미터를 반영하는 조직 바이오마커 시그니처가 기술되었다. 실험 설계를 보여주는 전체적 작업 흐름이 도 S1에 보인다.

- [0191] 단백질 커버리지, 동적 범위(Dynamic Range), 및 분석 성능
- [0192] GPS를 사용하여, 총 2,140의 단백질군을 확인하였다 (도. 1A-C). 확인 재현성 (identification reproducibility)은 높았고, 54.7% 펩티드 중첩을 초래하였다 (도 S2A). 비교에서, 전체 프로젝트를 통해 반복적으로 분석되었던 기준 샘플은 43.9%의 펩티드 확인 중첩(overlap)을 보였다 (도 S2B). 확인된 단백질 중, 전체 중 1388개가 성공적으로 정량되었고 (도 S3), 뒤이어 질환-관련 마커를 위한 검색에 사용되었다. 7267-샘플에 대한 정량에 대한 전체 중간값 CV값이 10.8%임을 확인하였고 (도 2A), 반면에 기준 샘플에 대한 대응하는 전체 중간값 CV값이 22.8%이었다 (도 2A). 주목할만하게, 61개의 단백질에 대응하는, 정량화된 단백질의 약 38% (833개의 펩티드)는 이전에 펩티드아틀라스에 보고되지 않았고 (도 1D), 실질적으로 새로운 커버리지에 대하여 나타낸다. 이는 검출된 펩티드의 유의성있는 부분이 이전에 보고된 것보다, 9 대 11 아미노산의 중간값 길이 (median length of 9 versus 11 amino acid)로, 더 짧다는 것에 의해 더 강조되었다 (도. 1E).
- [0193] 모든 정량화된 단백질에 대하여 측정된 \log_2 -MS 강도 정규화된 존재빈도(abundance)를 평가하고 $\sim 10^6$ 의 동적 범위를 나타내었다 (도 S3A). GPS 분석에 의해 생성된 상세한(in-depth) 커버리지는 펩티드아틀라스에서 종종 보고된 것으로부터 거의 보고되지 않은 것의 범위로(ranging from frequently reported in the PeptideAtlas to rarely reported), 펩티드가 용이하게 검출된다는 사실에 의해 더 도시되었다 (도 1F). 상기 검출된 단백질을 그 이후 주요 생물학적 과정에 의해 그룹핑되고, 여러 군 간에 분포되는 것이 확인되었다 (도 S3B). 흥미롭게, 예상된 바와 같이, 번역과 같은 과정에 의해 그룹핑된 단백질 (예, 60S 리보솜 단백질)이 예를 들면 유사분열에 관련된 다른 단백질(예, CDK1) 보다 더 높은 전체 존재빈도(abundance)를 표시하는 것으로 확인되었다. 종합하면, 상기 데이터는 재현할 수 있는 방식으로 신규하고 깊은(deep) 커버리지를 제공하는 GPS의 능력을 보여주었다.
- [0194] 조직학적 등급을 반영한 단백질 발현 프로파일
- [0195] 우선, 본 발명자는 조직학적 등급을 반영한 조직 바이오마커 시그니처가 판독될 수 있는지 여부를 조사하였다. 다변수(multivariate) 분석 (3군 비교)을 사용하여, 49개의 유의성있게 ($p < 0.01$, q -값 < 0.25) 차등적으로 발현된 단백질을 등급 1, 등급 2, 및 등급 3 집단 간에 확인하였다. 이 시그니처에 기반하여, PCA-플롯은 조직학 등급 1 및 등급 3 종양이 잘 분리될 수 있는 반면, 조직학 등급 2 종양은 더 불균일이고 다른 군의 모두 간에 스프레드(spread)되었다 (도 3A). 증가하는 조직학 등급을 갖는 상향- 및 하향-조절된 분석물의 패턴이 관찰될 수 있다. 예를 들면, 시클린-의존 키나아제 1 (CDK1), MCM3(minichromosome maintenance complex component 3), DNA 복제 라이선싱 요인(DNA replication licensing factor) MCM7, ATP-시트레이트 합성효소(ATP-citrate synthase, ACLY), 폴리아데닐레이트-결합 단백질 4(polyadenylate-binding protein 4, PABPC4), 및 6-포스포프럭토키나아제 타입 C(6-phosphofructokinase type C, PFKP)는 상향-조절된 조직 마커 중에 있다 (도 3A 및 도 S4). 반면에, 케라토칸(keratocan, KERA), 스폰딘(spondin, SPON1), 아스포린(asperin, ASPB), 아디포사이트 인핸서-결합 단백질 1(adipocyte enhancer-binding protein 1, AEBP1), 키마제(chymase, CMA1), 및 올팩토메딘-유사 단백질 3(olfactomedin-like protein 3, OLFML3)과 같은 분석물은 즉, 조직학적 등급 1 종양에서 더 높은 발현 수준이 표시된, 하향-조절된 분석물에 있다 (도 3A 및 도 S5).
- [0196] 본 발명자는 그 후 49개의 p -값 여과된 ($p < 0.01$) 바이오마커 목록이 조직학적 등급에 기반하여 조직을 분류하는데 사용될 수 있는지를 검사하였다. 이를 위하여, 본 발명자는 모든 샘플에 대하여 SVM으로 교차-검증된 리브-원-아웃 (leave-one-out)을 실시하고 모든 샘플에 대하여 결정값을 수집하였다. 예측값은 그 후 ROC 곡선을 구축하는데 사용되고, AUC 값을 산출하였다 (도 3A). 그 결과는 등급 2는 다시 더 불균일인 것으로 보였지만, 조직학 등급 종양 서브그룹이 잘 분리될 수 있다는 ($AUC=0.75-0.93$) 것을 보여주었다.
- [0197] 다음으로, 본 발명자는 차등적으로 발현된 마커를 정의하기 위해 다변수 시도(multivariate approach) 대신에 2-군 비교(two-group comparison) 사용의 영향을 조사하였다 (도 S6). 예상된 바와 같이, 데이터는 개별 조직학 서브그룹의 분류가 AUC-값 ($AUC=0.91-0.92$)에 의해 판단된 바와 같이 개선되었다는 것을 보여주었다. 조직학 등급 1 대 등급 3에 집중하여, 50개의 유의성있게 ($p < 0.01$) 차등적으로 분석된 분석물이 기술되었고, 그 중 31개가 이전 49개의 바이오마커 시그니처와 중첩되었다 (cfs. 도 3 및 S6C). 조직학적 등급 2가 등급 1 대 등급 1의 동결된 50개의 바이오마커 비교에 매핑되는 경우, 이는 불균일 특징을 다시 나타내고 모든 집단 중에 스프레

드되었다 (cfs. 도 3A 및 S6D).

[0198] *ER-상태의 영향*

[0199] 24개의 조직학적 등급 3 종양의 14개가 ER-음성으로 분류되고 17개의 ER-음성 샘플의 14개가 사실상 등급 3 종양에 존재하여, 본 발명자는 발현 프로파일에 대한 ER-상태의 직접적 영향을 조사하였다. 이 가설을 테스트하기 위해, 종양을 ER-양성 샘플 (n=33)만을 사용하여 재조사하였다. 다변량 접근법을 적용하여, 그 결과는 18개의 유의성있게 차등적으로 발현된 단백질 ($p < 0.01$, q-값 < 0.51)을 핀-포인트하고 조직학적 등급 1 대 등급 3 조직이 잘 분류될 수 있다 (0.9의 AUC-값, 데이터는 보이지 않음). 주목할만하게, 18개 분석물 중의 16개 (예, ASPN, SPON1, KERA, ACLY, APCS 및 PABPC4)가 처음 관독된 49개의 바이오마커 시그니처와 중첩하는 것으로 확인되었다 (도 3A). 따라서, 데이터는 49개의 바이오마커 시그니처가 조직학적 등급을 반영하였다는 관찰을 더 지지하였다.

[0200] 또한, 본 발명자는 또한 ER-관련 조직 바이오마커 시그니처가 풀릴 수 있는지(unravellled) 검사하였다. 그 결과는 ER-양성 및 ER-음성 유방암 조직이 잘 분류될 수 있다는 것을 보여주었고 (AUC=0.82) (도 3B), 39개 차등적으로 발현된 분석물($p < 0.01$, q-값 < 0.32)이 확인되었다 (예, GREB1 도 3B 및 S7A). 따라서, 데이터는 ER-관련 조직 바이오마커 시그니처가 검출되었다는 것을 보여주었다.

[0201] *HER2/neu-상태 및 Ki67-상태를 반영한 단백질 발현 프로파일*

[0202] 리브-윈-아웃 교차 검증을 사용한 HER2/neu-상태에 기반한 52개의 유방암 조직 추출물을 비교하는 경우, 데이터는 2개의 집단은 구별될 수 있고(AUC=0.98) 5개의 차등적으로 발현된 마커($p < 0.01$, q-값 < 0.9)가 확인된다는 것을 보여주었다 (도 3C). 가장 중요하게, 수용체 티로신-단백질 키나아제 erbB-2 (HER2)는 상향-조절된 단백질 중에서 확인되었다 (도 3C 및 도 S7B).

[0203] 또한, 동일한 방식으로, (Ki67-양성 암 핵의 25%가 컷-오프로서 사용된) Ki67-상태를 반영한 조직 단백질 시그니처가 또한 해독될 수 있다. 전체로서, 45개의 단백질이 차등적으로 발현되는 것으로 확인되었다 ($p < 0.01$, q-값 < 0.27) (도 S8A). 상기 데이터는 Ki67-양성 대 Ki67-음성 종양이 분리될 수 있음을 나타내었다 (AUC=0.84) (도 S8B). 따라서, 그 결과는 HER2/neu 상태 및 Ki67-상태 모두를 반영한 단백질 발현 프로파일이 정확히 찾아졌음을(pin-pointed) 보여주었다.

[0204] *생물학적 연관성*

[0205] 조직학적 등급 1 내지 3로 구별한 49개의 조직 바이오마커 시그니처의 생물학적 연관성을 그 후 조사하였다. 이것을 위하여, 각각의 개별 단백질의 세포 위치(cellular localization)를 IPA 소프트웨어를 사용하여 매핑하고 (도 4A), 네트워크 관련된 기능 및 잠재적 관계(potential relationship)를 조사하였다 (도 4B-4C). 세포 위치를 반영한 주로 하향-조절된 단백질(세포외 기질 (ECM)) 및 상향-조절된 분석물(원형질 막, 세포질, 및 핵)의 패턴이 드러났다. 더 중요하게, 상위에 랭크된 네트워크(top ranked network)가 DNA 복제, 재조합, 및 수선, 세포 주기 및 자유 라디칼 청소와 관련된 반면, 두번째로 가장 높게 랭크된 네트워크는 유전자 발현, 감염성 질환 및 암과 관련된 것으로 확인되었다. 주목할만하게도, 상위에 랭크된 네트워크 내 일부의 단백질은 NF-kB 및 VEGF와 직접적 또는 간접적으로 관련되었다 (도 4B). 또한, ECM 단백질의 대다수가 제2 네트워크 내에서 핀-포인트(pin-point)되고, 일부는 직접적으로 또는 간접적으로 형질전환 증식 인자(transforming growth factor) β (TGF β 1)와 관련되었다 (도 4C). 따라서, 그 결과는 조직학적 등급을 반영한, 생물학적으로 고도로 관련된 조직 바이오마커가 확인되었다.

[0206] 또한, 49개의 조직 바이오마커 시그니처 및 전사 인자 네트워크 간의 관계를 또한 IPA를 사용하여 평가하였다 (도 S9). 주목할 것 중, Rb 및 E2F2는 가장 관련된 전사 조절자 중에서 확인되었다 (도 S9A). 비교하여, 에스트로젠 수용체 2 (estrogen receptor 2, ESR2) 및 프로그스테론 수용체 (PGR)가 조직 바이오마커 시그니처가 ER-양성 대 ER-음성 종양을 구별하는 경우 가장 관련있는 조절자인 것으로 확인되었다 (도 S9B).

[0207] *후보 유방암 진행 시그니처의 검증*

- [0208] 조직학적 등급 1 내지 3을 구별하는 49개의 조직 바이오마커 시그니처를 검증하는 시도에서, 상기 데이터를 공개적으로 이용가능한 정통적(orthogonal) 유방암 mRNA 프로파일링 데이터 세트와 비교하였다. 검증 집단은 1,881개의 샘플로 구성되고, 그 중 등급 1 (n=239), 등급 2 (n=677), 및 등급 3 (n=495)을 포함한 할당된 조직학적 등급을 갖는 1,411개이다. 49개의 조직 바이오마커 중 42개가 gene entrez ID를 사용한 유전자 발현 데이터 베이스에 매핑될 수 있고, 그 후 검증 테스트에 사용되었다. 42개의 조직 마커를 그 후 등급 3 대 등급 1에 대하여 관찰된 하향- (15개의 분석물) 또는 상향-조절된 (27개의 분석물) 단백질 발현 프로파일에 기반하여, 2개의 그룹으로 나누고, 대응하는 mRNA 발현 프로파일을 비교하였다 (도 5). 하향-조절된 (예, SPON1 및 KERA) (도 5A, S5, S10I, 및 S11J) 및 상향-조절된 단백질 (예, CDK1 및 MCM3) (도 5B, S4, S10A, 및 S10B)의 단백질 발현 프로파일은 mRNA 프로파일 수준과 잘 검증되는(corroborate) 것으로 확인되었다. 흥미롭게도, 상기 상향-조절된 마커는 체크포인트 및 M-기 유전자 모듈과 높은 상관관계를 갖는 mRNA를 나타내는 반면 (도 5A), 하향-조절된 마커의 군은 기질 유전자 세트 모듈과 높은 상관관계를 갖는 mRNA 프로파일을 나타내는 것으로 확인되었다 (도 5B).
- [0209] *ER- 및 HER2-관련 조직 바이오마커 시그니처의 검증*
- [0210] 유사한 방식으로, 상술한 바와 같이 동일하게 공개적으로 이용가능한 정통적 (orthogonal) 유방암 mRNA 프로파일링 데이터 세트를 사용한, 시도가 ER-상태 (도 3B), 및 HER2-상태(도 3C)를 반영한 조직 바이오마커 시그니처에 그 이후에 만들어졌다.
- [0211] ER의 경우, 검증 세트는 395개의 ER-음성 및 1225 ER-음성 샘플을 포함한, 할당된 ER-상태를 갖는 1,620 샘플로 구성되었다. 39개의 조직 바이오마커 중 32개가 유전자 발현 데이터에 매핑될 수 있고, 이후 검증에서 사용되었다. 상기 32개의 마커는 그 후 관찰된 단백질 발현 프로파일에 기반하여 단백질 2개의 그룹(10개의 상향-조절 및 22개의 하향-조절됨)으로 나누어졌고, 상응하는 mRNA 발현 프로파일과 비교하였다 (도 S11). 일부 예외(예, 보체(complement)와 함께(도 S11F 및 도 S7A), 관찰된 단백질 발현 프로파일은 상응하는 mRNA 발현 프로파일과 잘 제공하였다(corroborate) (cfs. 도 S 3B, S7A, and S11). 이 맥락에서, ER-양성 종양 중 상향-조절된 단백질의 군은 스테로이드 반응 유전자 모듈과 높은 상관관계를 갖는 mRNA 프로파일을 나타내는 것으로 보이는 반면. 하향-조절된 단백질의 군은 면역-반응 및 기저 유전자 세트 모듈(basal gene set module)을 갖는 mRNA 프로파일을 나타내는 것으로 보였다 (도 S11A-S11B).
- [0212] HER2에 대한 검증 세트는 1,881개의 샘플로 구성되었고, HER2-양성 (n=152), basal (n=357), luminal-A (n=483), luminal-B (n=289), normal like (n=257), 및 미분류 (n=344)로 나누어졌다. 5개의 조직 마커 중 3개는 검증 데이터 세트에 매핑될 수 있고, 후속 평가에 사용되었다 (도 S12). 그 결과는 단백질 발현 프로파일 및 유전자 프로파일이 잘 연관되고 (cfs. 도 3C, S7B 및 S12), 더 관찰을 검증한다는 것을 나타낸다.
- [0213] *원격 전이 무 생존 평가*
- [0214] 최종적으로, 본 발명자는 조직학적 등급을 반영한 49개의 조직 바이오마커 시그니처가 원격 전이 무 생존(distant metastasis free survival, DMFS)의 위험을 동일하게 공개적으로 이용가능한 유전자 발현 데이터 세트를 다시 사용하여 평가하는데 사용될 수 있는지를 조사하였다. 49개의 조직 바이오마커 중 42개가 10년 종점 생존 데이터(endpoint survival data)를 갖는 1379개의 샘플에 매핑될 수 있다. 마커는 등급 3 대 등급 1 중, 하향-조절된 (n=15) 및 상향-조절된 (n=27) 마커를 반영하는 2개의 그룹으로 나누어지고, 카플란-마이어 분석이 그 후 이들 분석물의 발현 수준에 기반하여 3개의 변위치(quantile) (저(low), 중(intermediate), 및 고(high))로 유전자 발현 데이터를 계층화하여 10년 종점을 갖는 DMFS에 수행된다 (도 S13). 상기 데이터는 특히 하향-조절된 분석물 (주로 ECM-관련 분석물)의 집단이 DMFS의 위험을 예측하였다는 것을 보여주었다. 사실, 이는 단일 하향-조절된 (예, KERA 및 OLFM3) 또는 상향-조절된 (예, CDK1) 바이오마커를 표적화하는 것에 의해 달성될 수 있다.
- [0215] 논의(discussion)
- [0216] 본 연구에서 본 발명자는 유방암 중 종양 진행을 반영하고, 유방암에서의 맞춤형 의료(personalized medicine)를 향한 다음 단계로 취하는, 제1 면밀한(in-depth), 다중다수의(multiplexed) 조직 바이오마커 시그니처를 판독하였다. 이 달성은 본 발명자의 최근에 내부에서(in-house) 개발된 GPS 기술을 사용하여 달성되었다 (Olsson

et al., 2012; Olsson et al., 2011; Wingren et al., 2009). 따라서, 단지 9개의 항체만에 기반한, 친화성 프로테오믹스 및 무-표지 LC-MS/MS를 조합함으로써, 본 발명자는 가장 큰 유방암 조직 프로테وم 연구 중 하나를 나타내는 58개의 유방암 조직 샘플을 프로파일하였고, 1388개 단백질을 반영한 상세화된 정량화된 프로테وم 지도를 성공적으로 생성하였다.

[0217] 더 구체적으로, 높은 특이성 및 민감도를 갖는 조직학적 등급 1 내지 3의 유방암 종양을 구별하는 제1의 49-플렉스(49-plex) 조직 바이오마커 시그니처를 기술하였다. 이 목록은 p-값 기준을 $p < 0.02$ 로 설정한 79개의 차등적으로 발현된 마커로 확대될 수 있으나, 본 명세서에서 논의는 상위(top) 49개의 분석물 ($p < 0.01$)에 집중된다. 분자적 프로파일, 또는 단백질 핑거프린트는 등급 1 및 등급 3 종양이 더 구분적인(distinct) 반면에, 등급 2 종양이 더 불균일이라는 현재 관점에 의해 뒷받침된다(Sotiriou et al., 2006). 시그니처를 선형적이고 알려진 마커로 나누는 경우, 유방암과 관련된 것으로 알려진, 신규한 후보 바이오마커가 확인되었다. 관점의 기술적 요점으로부터, 이 신규한 커버리지가 정량화된 펩티드의 큰 부분 (~38%)이 펩티드아틀라스 데이터 베이스에 이전에 보고되지 않은 사실에 의해 반영되었다(Deutsch et al., 2008). GPS 설정에 의해 제공된 신규한 커버리지는 인간 단백질 아틀라스 프로젝트에 대하여 이들 49개의 분석물에 대하여 검색하는 경우 명백해졌다(Uhlen et al., 2010). 인간 단백질 아틀라스 프로젝트가 쓸모없지 않은(non-redundant) 인간 프로테오믹스의 50% 초과를 현재 커버함에도 불구하고, 어떠한 항체 및 조직학 염색도 49개 중 13개의 차등적으로 발현된 단백질을 보고하지 않았다.

[0218] ER-상태 단독은 유방암에서 유전자의 10% 초과 발현에 영향을 미치는 것으로 보이고, 생존에 영향을 갖는 것으로 일반적으로 생각된다. ER-음성 유방암은 일반적으로 더 공격적이고 항-에스트로겐 기반 용법은 비효율적이고, 추가적 표적화된 요법이 급히 요구된다(Rochefort et al., 2003). 본 발명자는 적절한 특이성 및 민감도로 ER-양성 종양과 ER-음성 종양을 구별할 수 있는 39개의 단백질 시그니처를 확인하였다. 주목할만하게도, 39개의 마커 중 11개는 GPS 기술에 의해 제공된 신규한 신규 커버리지를 다시 나타내는, 인간 단백질 아틀라스(Human Protein Atlas) 프로젝트에 의해 포함되지 않는다(Uhlen et al., 2010). 39개의 마커 중 하나, GREB1은 내분비 요법에 대한 반응 및 잠재적 치료 표적을 위한 후보 임상적 마커로 제안되었다(Hnatyszyn et al., 2010; Rae et al., 2005). GREB1은 에스트로겐-활성화된 세포 증식을 매개하는 에스트로겐-조절된 유전자이고, ER에 대한 대리 마커(surrogate marker)로서의 가능성을 보여주는, ER-양성 유방암 세포 및 정상 유방 조직에서 발현되지만, ER-음성 샘플에서는 발현되지 않는 것으로 최근에 보고되었다.(Hnatyszyn et al., 2010). GPS로 생성된 단백질 프로파일은 이 개념을 더 지지하였다(도 S7A).

[0219] 또한, 임상적으로 정의된 HER2-양성 및 HER2-음성 샘플을 구별할 수 있는 5개의 단백질 시그니처를 판독하였다(도 3C). 사실, 미량 존재 수용체(low abundant receptor) 티로신-단백질 키나아제 erbB-2(HER2-단백질)를 확인하고, 정량하고, 차등적으로 발현된 마커 중 하나인 것을 확인하였다. 따라서, 임상 현장(clinical setting)에서의 GPS를 사용한 HER2 측정의 가능성(potential)은 현재 사용되는 종래 면역조직화학(immunohistochemistry) 또는 FISH(fluorescence in situ hybridization) 기반 검출 시스템에서 현재 사용된 보완책으로서 전망이 될 수 있다. 최근 연구는 5개의 HER2-기반 테스트 중 하나가 부정확한 결과를 생성시킬 수 있다고 나타내었다(Phillips et al., 2009). 또한, S100-A9 및 성장 인자 수용체-결합 단백질 7(growth factor receptor-bound protein 7, GRB7)은 또한 HER2-양성 정의된 샘플의 대다수에서 증가된 발현을 나타내는 것으로 확인되었다(도 S7B). 높은 GRB7 발현은 높은 HER2-발현과 관련된 것으로 최근에 보고되었고, 감소된 생존을 갖는 유방암 환자의 서브셋을 정의하는데 사용되었다(Nadler et al., 2010). S100 유전자군(S100 gene family)은 저분자량-칼슘 결합 단백질을 코딩하고, 특정 S100 멤버는 암 진행, 전이와 관련되고, 유방암을 갖는 환자에서의 약물 내성의 예측 마커로서 가능성을 갖는다(McKiernan et al., 2011; Yang et al., 2011).

[0220] 가장 중요하게는, 조직학적 등급을 반영한 바이오마커 시그니처 및 ER-상태 및 HER2-상태를 반영한 바이오마커 시그니처를 GOBO-툴을 사용한, 독립적 데이터 세트 및 정형적(orthogonal) 방법(mRNA 발현 수준)을 사용하여 평가하였다(Ringner et al., 2011). 이는 종종 유전자 시그니처에 의해 포획된 기능적 프로세스이고 중요한 개별 유전자가 아니기 때문에, 상향- 및 하향-조절된 단백질의 군을 알려진 유전자 세트 모듈에 대한 연관관계에 기반하여 평가하였다(Wirapati et al., 2008). 기질(stroma), 체크포인트, 및 스테로이드 반응을 위한 유전자-세트 모듈(gene-set module)에 대한 유의한 연관관계가 특히 주목할만하다(도 5 및 도 S11). 또한, 종점(endpoint)으로서 DMFS를 평가하는 경우, 조직학적 도출된 단백질 분석물을 사용하여, 데이터는 특히 하향-조절된 ECM 단백질을 사용하는 경우 더 악화된 임상 결과를 분명하게 나타내었다. 따라서, 독립적 mRNA 검증이 보고된 후보 바이오마커 시그니처 및 장래 유방 조직 종양 분류에 대한 가능성에 대한 강한 지지를 추가하였다.

[0221] 종합하면, 본 발명자는 최근 개발된 임상의 프로테오믹스 발견 프로파일링 노력에 대한 GPS 기술 플랫폼의 적용가능

성을 보여주었다. 조직학적 등급, 즉, 종양 진행 및 다른 주요한 임상적 실험 파라미터, 예를 들면 ER-, HER2-, 및 Ki67-상태(status)를 반영한 조직 바이오마커 시그니처가 본 연구에 보고되었고, 이들 신규한 조직 바이오마커 묘사(portrait)는 개선된 유방암의 분류 및 예후를 가능하게 한다.

[0222] 참조문헌

- [0223] Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198-207.
- [0224] Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57, 289-300.
- [0225] Bergamaschi, A., Tagliabue, E., Sorlie, T., Naume, B., Triulzi, T., Orlandi, R., Russnes, H. G., Nesland, J. M., Tammi, R., Auvinen, P., et al. (2008). Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. *The Journal of pathology* 214, 357-367.
- [0226] Bierie, B., and Moses, H. L. (2006). Tumour microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer. *Nature reviews Cancer* 6, 506-520.
- [0227] Borrebaeck, C. A., and Wingren, C. (2011). Recombinant antibodies for the generation of antibody arrays. *Methods Mol Biol* 785, 247-262.
- [0228] Bouchal, P., Roumeliotis, T., Hrstka, R., Nenutil, R., Vojtesek, B., and Garbis, S. D. (2009). Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. *Journal of proteome research* 8, 362-373.
- [0229] Carlsson, A., Wingren, C., Ingvarsson, J., Ellmark, P., Baldertorp, B., Ferno, M., Olsson, H., and Borrebaeck, C. A. (2008). Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays. *Eur J Cancer* 44, 472-480.
- [0230] Carlsson, A., Wingren, C., Kristensson, M., Rose, C., Ferno, M., Olsson, H., Jernstrom, H., Ek, S., Gustavsson, E., Ingvar, C., et al. (2011). Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. *Proceedings of the National Academy of Sciences of the United States of America* 108, 14252-14257.
- [0231] Ciocca, D. R., and Elledge, R. (2000). Molecular markers for predicting response to tamoxifen in breast cancer patients. *Endocrine* 13, 1-10.
- [0232] Cortes, C., and Vapnik, V. (1995). Support-Vector Networks. *Machine Learning* 20, 273-297.
- [0233] Cortez, D., Glick, G., and Elledge, S. J. (2004). Minichromosome maintenance proteins are direct targets of the ATM and ATR checkpoint kinases. *Proceedings of the National Academy of Sciences of the United States of America* 101, 10078-10083.
- [0234] Desmedt, C., Haibe-Kains, B., Wirapati, P., Buyse, M., Larsimont, D., Bontempi, G., Delorenzi, M., Piccart, M., and Sotiriou, C. (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clinical cancer research : an official journal of the American Association for Cancer Research* 14, 5158-5165.
- [0235] Deutsch, E. W., Lam, H., and Aebersold, R. (2008). PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO reports* 9, 429-434.
- [0236] Dowsett, M., Goldhirsch, A., Hayes, D. F., Senn, H. J., Wood, W., and Viale, G. (2007). International Web-based consultation on priorities for translational breast cancer research. *Breast cancer research : BCR* 9, R81.
- [0237] Elston, C. W., and Ellis, I. O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up.

Histopathology 19, 403-410.

- [0238] Fata, J. E., Werb, Z., and Bissell, M. J. (2004). Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes. *Breast cancer research : BCR* 6, 1-11.
- [0239] Frierson, H. F., Jr., Wolber, R. A., Berean, K. W., Franquemont, D. W., Gaffey, M. J., Boyd, J. C., and Wilbur, D. C. (1995). Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. *American journal of clinical pathology* 103, 195-198.
- [0240] Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., and Mann, M. (2010). Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature methods* 7, 383-385.
- [0241] Geiger, T., Madden, S. F., Gallagher, W. M., Cox, J., and Mann, M. (2012). Proteomic portrait of human breast cancer progression identifies novel prognostic markers. *Cancer research*.
- [0242] Gong, Y., Wang, N., Wu, F., Cass, C. E., Damaraju, S., Mackey, J. R., and Li, L. (2008). Proteome profile of human breast cancer tissue generated by LC-ESI-MS/MS combined with sequential protein precipitation and solubilization. *Journal of proteome research* 7, 3583-3590.
- [0243] Ha, S. A., Shin, S. M., Namkoong, H., Lee, H., Cho, G. W., Hur, S. Y., Kim, T. E., and Kim, J. W. (2004). Cancer-associated expression of minichromosome maintenance 3 gene in several human cancers and its involvement in tumorigenesis. *Clinical cancer research : an official journal of the American Association for Cancer Research* 10, 8386-8395.
- [0244] Hakkinen, J., Vincic, G., Mansson, O., Warell, K., and Levander, F. (2009). The proteios software environment: an extensible multiuser platform for management and analysis of proteomics data. *Journal of proteome research* 8, 3037-3043.
- [0245] Hanahan, D., and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell* 100, 57-70.
- [0246] Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646-674.
- [0247] Hanash, S. (2003). Disease proteomics. *Nature* 422, 226-232.
- [0248] Hnatyszyn, H. J., Liu, M., Hilger, A., Herbert, L., Gomez-Fernandez, C. R., Jorda, M., Thomas, D., Rae, J. M., El-Ashry, D., and Lippman, M. E. (2010). Correlation of GREB1 mRNA with protein expression in breast cancer: validation of a novel GREB1 monoclonal antibody. *Breast cancer research and treatment* 122, 371-380.
- [0249] Hondermarck, H., Tastet, C., El Yazidi-Belkoura, I., Toillon, R. A., and Le Bourhis, X. (2008). Proteomics of breast cancer: the quest for markers and therapeutic targets. *Journal of proteome research* 7, 1403-1411.
- [0250] Hudis, C. A. (2007). Trastuzumab--mechanism of action and use in clinical practice. *The New England journal of medicine* 357, 39-51.
- [0251] Ivshina, A. V., George, J., Senko, O., Mow, B., Putti, T. C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H., et al. (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer research* 66, 10292-10301.
- [0252] Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. *CA: a cancer journal for clinicians* 61, 69-90.
- [0253] Johnson, N., Li, Y. C., Walton, Z. E., Cheng, K. A., Li, D., Rodig, S. J., Moreau, L. A., Unitt, C., Bronson, R. T., Thomas, H. D., et al. (2011). Compromised CDK1 activity sensitizes BRCA-proficient cancers to PARP inhibition. *Nature medicine* 17, 875-882.
- [0254] Kang, S., Kim, M. J., An, H., Kim, B. G., Choi, Y. P., Kang, K. S., Gao, M. Q., Park, H., Na, H. J., Kim, H. K., et al. (2010). Proteomic molecular portrait of interface zone in breast cancer. *Journal of*

proteome research 9, 5638-5645.

- [0255] Kuhn, U., and Wahle, E. (2004). Structure and function of poly(A) binding proteins. *Biochimica et biophysica acta* 1678, 67-84.
- [0256] Lasko, T. A., Bhagwat, J. G., Zou, K. H., and Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of biomedical informatics* 38, 404-415.
- [0257] Malumbres, M., and Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. *Nature reviews Cancer* 9, 153-166.
- [0258] Mangus, D. A., Evans, M. C., and Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome biology* 4, 223.
- [0259] McKiernan, E., McDermott, E. W., Evoy, D., Crown, J., and Duffy, M. J. (2011). The role of S100 genes in breast cancer progression. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 32, 441-450.
- [0260] Moon, J. S., Kim, H. E., Koh, E., Park, S. H., Jin, W. J., Park, B. W., Park, S. W., and Kim, K. S. (2011). Kruppel-like factor 4 (KLF4) activates the transcription of the gene for the platelet isoform of phosphofructokinase (PFKP) in breast cancer. *The Journal of biological chemistry* 286, 23808-23816.
- [0261] Nadler, Y., Gonzalez, A. M., Camp, R. L., Rimm, D. L., Kluger, H. M., and Kluger, Y. (2010). Growth factor receptor-bound protein-7 (Grb7) as a prognostic marker and therapeutic target in breast cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 21, 466-473.
- [0262] Olsson, N., James, P., Borrebaeck, C. A., and Wingren, C. (2012). Quantitative proteomics targeting classes of motif-containing peptides using immunoaffinity-based mass spectrometry. Submitted.
- [0263] Olsson, N., Wingren, C., Mattsson, M., James, P., D, O. C., Nilsson, F., Cahill, D. J., and Borrebaeck, C. A. (2011). Proteomic analysis and discovery using affinity proteomics and mass spectrometry. *Molecular & cellular proteomics : MCP* 10, M110 003962.
- [0264] Olsson, N., Wingren, C., Mattsson, M., James, P., D, O. C., Nilsson, F., Cahill, D. J., and Borrebaeck, C. A. (2011). Proteomic analysis and discovery using affinity proteomics and mass spectrometry. *Molecular & cellular proteomics : MCP* 10, M110 003962.
- [0265] Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M. G., Watson, D., Park, T., et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine* 351, 2817-2826.
- [0266] Pavlidis, P., and Noble, W. S. (2003). Matrix2png: a utility for visualizing matrix data. *Bioinformatics* 19, 295-296.
- [0267] Pei, D. S., Qian, G. W., Tian, H., Mou, J., Li, W., and Zheng, J. N. (2012). Analysis of human Ki-67 gene promoter and identification of the Sp1 binding sites for Ki-67 transcription. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 33, 257-266.
- [0268] Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747-752.
- [0269] Phillips, K. A., Marshall, D. A., Haas, J. S., Elkin, E. B., Liang, S. Y., Hassett, M. J., Ferrusi, I., Brock, J. E., and Van Bebbber, S. L. (2009). Clinical practice patterns and cost effectiveness of human epidermal growth receptor 2 testing strategies in breast cancer patients. *Cancer* 115, 5166-5174.
- [0270] Place, A. E., Jin Huh, S., and Polyak, K. (2011). The microenvironment in breast cancer progression: biology and implications for treatment. *Breast cancer research : BCR* 13, 227.
- [0271] Rae, J. M., Johnson, M. D., Scheys, J. O., Cordero, K. E., Larios, J. M., and Lippman, M. E. (2005). GREB 1 is a critical regulator of hormone dependent breast cancer growth. *Breast cancer research and*

treatment 92, 141-149.

- [0272] Ringner, M., Fredlund, E., Hakkinen, J., Borg, A., and Staaf, J. (2011). GOBO: gene expression-based outcome for breast cancer online. *PloS one* 6, e17911.
- [0273] Robbins, P., Pinder, S., de Klerk, N., Dawkins, H., Harvey, J., Sterrett, G., Ellis, I., and Elston, C. (1995). Histological grading of breast carcinomas: a study of interobserver agreement. *Human pathology* 26, 873-879.
- [0274] Rochefort, H., Glondu, M., Sahla, M. E., Platet, N., and Garcia, M. (2003). How to target estrogen receptor-negative breast cancer? *Endocrine-related cancer* 10, 261-266.
- [0275] Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., et al. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *The New England journal of medicine* 344, 783-792.
- [0276] Soderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N., Alexeiva, V., Aberg, A. M., Nilsson, A., Jansson, B., Ohlin, M., Wingren, C., *et al.* (2000). Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nature biotechnology* 18, 852-856. _
- [0277] Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America* 98, 10869-10874.
- [0278] Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B., et al. (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *Journal of the National Cancer Institute* 98, 262-272.
- [0279] Strande, V., Canelle, L., Tastet, C., Burlet-Schiltz, O., Monsarrat, B., and Hondermarck, H. (2009). The proteome of the human breast cancer cell line MDA-MB-231: Analysis by LTQ-Orbitrap mass spectrometry. *Proteomics Clinical applications* 3, 41-50.
- [0280] Sutton, C. W., Rustogi, N., Gurkan, C., Scally, A., Loizidou, M. A., Hadjisavvas, A., and Kyriacou, K. (2010). Quantitative proteomic profiling of matched normal and tumor breast tissues. *Journal of proteome research* 9, 3891-3902.
- [0281] Turashvili, G., Bouchal, J., Baumforth, K., Wei, W., Dziechciarkova, M., Ehrmann, J., Klein, J., Fridman, E., Skarda, J., Srovnal, J., et al. (2007). Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. *BMC cancer* 7, 55.
- [0282] Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., et al. (2010). Towards a knowledge-based Human Protein Atlas. *Nature biotechnology* 28, 1248-1250.
- [0283] van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *The New England journal of medicine* 347, 1999-2009.
- [0284] van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-536.
- [0285] Wang, T. H., Chao, A., Tsai, C. L., Chang, C. L., Chen, S. H., Lee, Y. S., Chen, J. K., Lin, Y. J., Chang, P. Y., Wang, C. J., et al. (2010). Stress-induced phosphoprotein 1 as a secreted biomarker for human ovarian cancer promotes cancer cell proliferation. *Molecular & cellular proteomics : MCP* 9, 1873-1884.

- [0286] Wingren, C., James, P., and Borrebaeck, C. A. (2009). Strategy for surveying the proteome using affinity proteomics and mass spectrometry. *Proteomics* 9, 1511-1517.
- [0287] Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., Schutz, F., et al. (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast cancer research : BCR* 10, R65.
- [0288] Yang, W. S., Moon, H. G., Kim, H. S., Choi, E. J., Yu, M. H., Noh, D. Y., and Lee, C. (2011). Proteomic Approach Reveals FKBP4 and S100A9 as Potential Prediction Markers of Therapeutic Response to Neoadjuvant Chemotherapy in Patients with Breast Cancer. *Journal of proteome research*.
- [0289] Zhang, L., Reidy, S. P., Bogachev, O., Hall, B. K., Majdalawieh, A., and Ro, H. S. (2011). Lactation defect with impaired secretory activation in AEBP1-null mice. *PloS one* 6, e27795.

[0290] 표 1: 유방암-관련 질병 상태를 결정하기 위한 바이오마커

[0291]

A) 코어 바이오마커 (H-등급 및 DMSF)							개별 비교에서의 상향- 또는 하향 조절			
Prot acc.	p-값	q-값	F-통계	이름	H-등급	DMSF	H1 vs H2	H1 vs H3	H2 vs H3	DMSF*
1 O60938	3.54E-06	0.00148377	16.5884017	9 KERA	yes	yes	Down H2	Down H3	Down H3	Down H3
2 Q9HCB6	3.27E-06	0.00148377	16.7249374	4 SPON1	yes	yes	Down H2	Down H3	Down H3	Down H3
B) 바람직한 바이오마커 (H-등급 및 DMSF)							개별 비교에서의 상향- 또는 하향 조절			
Prot acc.	p-값	q-값	F-통계	이름	H-등급	DMSF	H1 vs H2	H1 vs H3	H2 vs H3	DMSF
3 P02743	1.74E-06	0.00148377	17.8235511	8 APCS	yes	yes	Down H2	Down H3	Down H3	Down H3
4 O75348	0.00079994	0.12338118	8.33115768	ATP6V1G	yes	yes	Down H2	Up H3	Up H3	Down H3
5 Q71UM5	0.00098264	0.12338118	8.05375766	1 RPS27L	yes	yes	Up H2	Up H3	Up H3	Up H3
6 Q14195	0.00115970	0.12338118	7.83207941	1 DPYSL3	yes	yes	Down H2	Down H3	Down H3	Down H3
7 Q9BS26	0.00117786	0.12338118	7.81137371	1 ERP44	yes	yes	Down H2	Up H3	Up H3	Up H3
8 Q13905	0.00191064	0.18474455	7.17342996	6 RAPGEF1	yes	yes	Down H2	Up H3	Up H3	Up H3
9 P53396	0.00235178	0.19466547	6.90347814	6 ACLY	yes	yes	Up H2	Up H3	Up H3	Up H3
10 P23946	0.00243670	0.19466547	6.85762119	3 CMA1	yes	yes	Down H2	Down H3	Down H3	Down H3
11 P25205	0.00264061	0.19466547	6.75398016	MCM3	yes	yes	Up H2	Up H3	Up H3	Up H3

12	Q9UKU9	0.00278757	2	0.19466547	1	6.68433713	9	ANGPTL2	yes	yes	Down H2	Down H3	Down H3
13	Q8IUX7	0.00329872		0.21039261	2	6.46885728	8	AEBP1	yes	yes	Down H2	Down H3	Down H3
14	Q15819	0.00334753	6	0.21039261	2	6.45012950	9	UBE2V2	yes	yes	Up H2	Up H3	Up H3
15	Q6P0N0	0.00367042	3	0.21554090	3	6.33300256	7	MIS18BP1	yes	yes	Up H2	Up H3	Up H3
16	Q9UBD9	0.00382176	2	0.21554090	3	6.28175306	3	CLCF1	yes	yes	Up H2	Up H3	Up H3
17	P80404	0.00428322	8	0.22009741	5	6.13763570	8	ABAT	yes	yes	Up H2	Down H3	Down H3
18	P05141	0.00480072	5	0.22009741	5	5.99413442	6	SLC25A5	yes	yes	Up H2	Up H3	Up H3
19	P31948	0.00501211	8	0.22009741	5	5.94010114	7	STP1	yes	yes	Up H2	Up H3	Up H3
20	Q9NFRN5	0.00549968		0.22009741	5	5.82403421	4	OLFML3	yes	yes	Down H2	Down H3	Down H3
21	P09693	0.00633343	9	0.22009741	5	5.64451599	1	CD3G	yes	yes	Up H2	Up H3	Up H3
22	P33993	0.00650666	6	0.22009741	5	5.61497592	9	MCM7	yes	yes	Up H2	Up H3	Up H3
23	Q02978	0.00675539	5	0.22009741	5	5.56853532	8	SLC25A11	yes	yes	Down H2	Up H3	Up H3
24	O00567	0.00694376	6	0.22009741	5	5.53453540	8	NOP56	yes	yes	Up H2	Up H3	Up H3
25	O43159	0.00698571	2	0.22009741	5	5.52709531	8	RRP8	yes	yes	Up H2	Down H3	Up H3
26	Q9NWH	0.00768360	7	0.22009741	5	5.40971517	6	SLTM	yes	yes	Up H2	Up H3	Up H3
27	Q15631	0.00774940		0.22009741	5	5.39922761		TSN	yes	yes	Up H2	Up H3	Up H3

Prot acc.	p-값	q-값	F-통계	이름	H-등급	DMSF	개별 비교에서의 상향- 또는 하향 조절			
							H1 vs H2	H1 vs H3	H2 vs H3	DMSF*
28 Q13011	0.00787938	0.22009741	5.37877941	1 ECH1	yes	yes	Down H2	Up H3	Up H3	Up H3
29 P51888	0.00846166	0.22908640	5.29129695	9 PRELP	yes	yes	Down H2	Down H3	Down H3	Down H3
30 P49591	0.00856568	0.22908640	5.27633237	8 SARS	yes	yes	Up H2	Up H3	Down H3	Up H3
31 P62851	0.00954485	0.24995586	5.14409637	5 RPS25	yes	yes	Up H2	Up H3	Up H3	Up H3
32 Q9BSJ8	0.00987108	0.25322346	5.10316181	2 ESYT1	yes	yes	Down H2	Up H3	Up H3	Up H3
C) 바람직한 바이오마커 (H-등급)										
33 Q7Z5L7	0.00051876	0.12338118	8.92324543	PODN	yes	T.B.D.	Up H2	Down H3	Down H3	T.B.D.
34 Q9NQG5	0.00488782	0.22009741	5.97157764	4 RPRD1B	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
35 Q8NHW	0.00505076	0.22009741	5.93047952	7 RPLP0P6	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
36 Q6UJG3	0.00526947	0.22009741	5.87743854	5 CD300LG	yes	T.B.D.	Down H2	Up H3	Up H3	T.B.D.
37 Q9Y2Z0	0.00586541	0.22009741	5.74380445	5 SUGT1	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
38 A5A3E0	0.00721476	0.25770143	5.05727291	9 POTEF	yes	T.B.D.	Up H2	Up H3	Down H3	T.B.D.
39 Q15046	0.01025065	0.25770143	5.05727291	1 KARS	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
40 O75306	0.01061354	0.26159273	5.01502752	3 NDUF52	yes	T.B.D.	Up H2	Up H3	Down H3	T.B.D.

41	P55795 O43852	0.01129597	0.26790630	4.93951463	7	HNRNPH2	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
42	-2	0.01160955 0.01208899	0.26933089	4.90639638	9	CALU	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
43	P55884 Q9BWU	6 0.01234520	0.26933089	4.85752201	1	EIF3B	yes	T.B.D.	Down H2	Up H3	Up H3	T.B.D.
44	0	9	0.26933089	4.8322258	P	SLC4A1A	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
45	P46782	0.01242736 0.01277219	0.26933089	4.82423067	1	RPS5	yes	T.B.D.	Up H2	Up H3	Down H3	T.B.D.
46	Q6UX71	4	0.27746514	4.79126119	6	PLXDC2	yes	T.B.D.	Up H2	Down H3	Down H3	T.B.D.
47	Q6UXG2	0.01324416 0.01470293	5	4.74761056	9	KIAA1324	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
48	P22897	5	0.29954613	4.62228918	1	MRC1	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
49	Q96P16	0.01479683 0.01524876	4	4.61467218	4	RPRD1A	yes	T.B.D.	Down H2	Up H3	Up H3	T.B.D.
50	P34897	9	0.29954613	4.57870149	6	SHMT2	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
51	P50991	0.01538449 0.01683115	3	4.56811571	1	CCT4	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
52	Q53HC9	1	0.29954613	4.46097946	2	TSSC1	yes	T.B.D.	Down H2	Up H3	Up H3	T.B.D.
53	Q9UKT9	0.01695365 0.01706097	6	4.45235204	7	IKZF3	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
54	Q7Z7E8	8	0.29954613	4.44484710	7	UBE2Q1	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
55	O00233	0.01796363 0.01824483	5	4.38360738	8	PSMD9	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
56	P08621	0.01824483	3	4.36518335	70	SNRNP70	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.

70	P53634	0.00725501	1	0.22009741	5	5.48041105	3	CTSC	yes	yes	Up H2	Up H3	Down H3	Up H3
E) 선택적 바이오마커 (H-등급)														
Prot acc.		p-값		q-Value		F-값		이름	H-등급	DMSF	H1 vs H2	H1 vs H3	H2 vs H3	DMSF*
71	Q9Y2X3	0.0078144		0.22009741	5	5.38895797		NOP58	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
72	P00558	0.01119282	2	0.26790630	6	4.95061826	7	PGK1	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
73	Q00688	0.01194368	8			4.87211799	6	FKBP3	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
74	P21266	0.01630190	6	0.29954613	4	4.49901962	3	GSTM3	yes	T.B.D.	Down H2	Down H3	Down H3	T.B.D.
75	Q9NZT1	0.016391		0.29954613	4	4.49252653	1	CALML5	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
76	P29590	0.01680743	2	0.29954613	4	4.46265745	2	PML	yes	T.B.D.	Up H2	Up H3	Up H3	T.B.D.
77	O75173	0.01681119	1	0.29954613	4	4.46239137	6	ADAMTS4	yes	T.B.D.	Down H2	Down H3	Down H3	T.B.D.
78	P07996	0.01715777	4	0.29954613	4	4.43812036	5	THBS1	yes	T.B.D.	Down H2	Down H3	Down H3	T.B.D.
79	P02751	0.01824138	2	0.30578347	3	4.36540794	4	FN1	yes	T.B.D.	Down H2	Down H3	Down H3	T.B.D.

*종간값H1/H3 비율에 기반함

T.B.D. = 미결

[0297] 표2: 유방암-관련 질환 상태를 결정하기 위한 바이오마커의 추천 명칭

Prot acc.	Name	Recommended name
Q9HCB6	SPON1	Spondin-1
O60938	KERA	Keratocan
P02743	APCS	Serum amyloid P-component
Q7Z5L7	PODN	Podocan
O75348	ATP6V1G1	V-type proton ATPase subunit G 1
Q71UM5	RPS27L	40S ribosomal protein S27-like
Q14195	DPYSL3	Dihydropyrimidinase-related protein 3
Q9BS26	ERP44	Endoplasmic reticulum resident protein 44
Q13905	RAPGEF1	Rap guanine nucleotide exchange factor 1
P53396	ACLY	ATP-citrate synthase
P23946	CMA1	Chymase
P25205	MCM3	DNA replication licensing factor MCM3
Q9UKU9	ANGPTL2	Angiopoietin-related protein 2
Q8IUX7	AEBP1	Adipocyte enhancer-binding protein 1
Q15819	UBE2V2	Ubiquitin-conjugating enzyme E2 variant 2
Q6P0N0	MIS18BP1	Mis18-binding protein 1
Q9UBD9	CLCF1	Cardiotrophin-like cytokine factor 1
P80404	ABAT	4-aminobutyrate aminotransferase, mitochondrial
P05141	SLC25A5	ADP/ATP translocase 2
Q9NQG5	RPRD1B	Regulation of nuclear pre-mRNA domain-containing protein 1B
P31948	STIP1	Stress-induced-phosphoprotein 1
Q8NHW5	RPLP0P6	60S acidic ribosomal protein P0-like
Q6UXG3	CD300LG	CMRF35-like molecule 9
Q9NRN5	OLFML3	Olfactomedin-like protein 3
Q9Y2Z0	SUGT1	Suppressor of G2 allele of SKP1 homolog
P09693	CD3G	T-cell surface glycoprotein CD3 gamma chain
P33993	MCM7	DNA replication licensing factor MCM7
Q02978	SLC25A11	Mitochondrial 2-oxoglutarate/malate carrier protein
O00567	NOP56	Nucleolar protein 56
O43159	RRP8	Ribosomal RNA-processing protein 8
A5A3E0	POTEF	POTE ankyrin domain family member F
Q9NWH9	SLTM	SAFB-like transcription modulator
Q15631	TSN	Translin
Q13011	ECH1	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial
P51888	PRELP	Prolargin
P49591	SARS	Serine--tRNA ligase, cytoplasmic
P62851	RPS25	40S ribosomal protein S25
Q9BSJ8	ESYT1	Extended synaptotagmin-1
Q15046	KARS	Lysine--tRNA ligase
O75306	NDUFS2	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial
P55795	HNRNPH2	Heterogeneous nuclear ribonucleoprotein H2
O43852-2	CALU	Calumenin

[0298]

P55884	EIF3B	Eukaryotic translation initiation factor 3 subunit B
Q9BWU0	SLC4A1AP	Kanadaptin
P46782	RPS5	40S ribosomal protein S5
Q6UX71	PLXDC2	Plexin domain-containing protein 2
Q6UXG2	KIAA1324	UPF0577 protein KIAA1324
P22897	MRC1	Macrophage mannose receptor 1
Q96P16	RPRD1A	Regulation of nuclear pre-mRNA domain-containing protein 1A
P34897	SHMT2	Serine hydroxymethyltransferase, mitochondrial
P50991	CCT4	T-complex protein 1 subunit delta
Q53HC9	TSSC1	Protein TSSC1
Q9UKT9	IKZF3	Zinc finger protein Aiolos
Q7Z7E8	UBE2Q1	Ubiquitin-conjugating enzyme E2 Q1
O00233	PSMD9	26S proteasome non-ATPase regulatory subunit 9
P08621	SNRNP70	U1 small nuclear ribonucleoprotein 70 kDa
P11234	RALB	Ras-related protein Ral-B
Q99798	ACO2	Aconitate hydratase, mitochondrial
Q92614	MYO18A	Unconventional myosin-XVIIIa
P47897	QARS	Glutamine--tRNA ligase
Q13310	PABPC4	Polyadenylate-binding protein 4
O95969	SCGB1D2	Secretoglobin family 1D member 2
Q01813	PFKP	6-phosphofructokinase type C
P08195	SLC3A2	4F2 cell-surface antigen heavy chain
Q9BXN1	ASPN	Asporin
P28907	CD38	ADP-ribosyl cyclase 1
Q9NR99	MXRA5	Matrix-remodeling-associated protein 5
P06493	CDK1	Cyclin-dependent kinase 1
O76061	STC2	Stanniocalcin-2
P53634	CTSC	Dipeptidyl peptidase 1
Q9Y2X3	NOP58	Nucleolar protein 58
P00558	PGK1	Phosphoglycerate kinase 1
Q00688	FKBP3	Peptidyl-prolyl cis-trans isomerase FKBP3
P21266	GSTM3	Glutathione S-transferase Mu 3
Q9NZT1	CALML5	Calmodulin-like protein 5
P29590	PML	Protein PML
O75173	ADAMTS4	A disintegrin and metalloproteinase with thrombospondin motifs 4
P07996	THBS1	Thrombospondin-1
P02751	FN1	Fibronectin

[0299]

[0300] 표 3: 예시적인 바이오마커 조합에 대한 ROC AUC 값

Biomarker signature	ROC AUC value		
	H1 vs H3	H2 vs H3	H1 vs H2
Core-1	0.94	0.82	0.69
Core-2	0.82	0.82	0.56
Core-1+core-2	0.94	0.83	0.50
Core-1+core-2+(marker 3-12)	0.90	0.85	0.69
Core-1+core-2+(marker 3-22)	0.88	0.76	0.88
Core-1+core-2+(marker 3-32)	0.81	0.73	0.94
Core-1+core-2+(marker 3-42)	0.81	0.74	0.87
Core-1+core-2+(marker 3-52)	0.80	0.77	0.76
Core-1+core-2+(marker 3-60)	0.80	0.77	0.76
Core-1+core-2+(marker 3-62)	0.93	0.86	0.76
Core-1+core-2+(marker 3-72)	0.94	0.83	0.82
Core-1+core-2+(marker 3-79)	0.94	0.79	0.86

[0301]

[0302] 표 4:조직학적 등급 SVM SCRIPT

```

filnamn<-"Input.txt"
# 1.1 Change FILNAME to datafile -----
---
# Läser in och logaritererar datan
rawfile <- read.delim(filnamn)
samplenames <- as.character(rawfile[,1])
Diagnosis <- rawfile[,2]
Morphology<- rawfile[,3]
Treatment<-rawfile[,4]
data <- t(rawfile[, -c(1:4)])
ProteinNames <- read.delim(filnamn,header=FALSE)
ProteinNames <- as.character(as.matrix(ProteinNames)[1,])
ProteinNames <- ProteinNames[-(1:4)]
rownames(data) <- ProteinNames
colnames(data) <- samplenames
logdata <- log(data)/log(2)

# Tar reda på vilka gruppjämförelser som ska göras
PairWiseGroups <-
as.matrix(read.delim("Comparisons_to_do.txt",header=FALSE)) # 1.2
Change filename and use criteria file -----

# Definierar Wilcoxon testet
wilcoxtest <- function(prot,subset1,subset2){
  res <- wilcox.test(prot[subset1],prot[subset2])
  res$p.value
}

# Definierar foldchange
foldchange <- function(prot,subset1,subset2){
  2^(mean(prot[subset1]) - mean(prot[subset2]))
}

```

[0303]

```
# Definierar q-värdesberäkningen
BenjaminiHochberg <- function(pvalues){
  # This function takes a vector of p-values as input and outputs
  # their q-values. No reordering of the values is performed
  NAindices <- is.na(pvalues)
  Aindices <- !NAindices
  Apvalues <- pvalues[Aindices]
  N <- length(Apvalues)
  orderedindices <- order(Apvalues)
  OrdValues <- Apvalues[orderedindices]
  CorrectedValues <- OrdValues * N / (1:N)
  MinValues <- CorrectedValues
  for (i in 1:N){MinValues[i] <- min(CorrectedValues[i:N])}
  Aqvalues <- numeric(N)
  Aqvalues[orderedindices] <- MinValues
  Qvalues <- pvalues
  Qvalues[Aindices] <- Aqvalues
  return(Qvalues)
}

# Laddar in två bibliotek
library(MASS)
library(gplots)

# Definierar färger till heatmapen
redgreen <- function(n)
{
  c(
    hsv(h=0/6, v=c( rep( seq(1,0.3,length=5) , c(13,10,8,6,4) ) ,
0 ) ) ,
    hsv(h=2/6, v=c( 0 , rep( seq(0.3,1,length=5) ,
c(3,5,7,9,11) ) ) )
  )
}
pal <- rev(redgreen(100));

#Laddar in fler bibliotek och funktioner
library(e1071)
source("NaiveBayesian")

#Definierar SVM med Leave One Out
svmLOOvalues <- function(data , fac){
  n1 <- sum(fac==levels(fac)[1])
  n2 <- sum(fac==levels(fac)[2])
  nsamples <- n1+n2
  ngenes <- nrow(data)
  SampleInformation <- paste(levels(fac)[1], " ",n1," ,
",levels(fac)[2], " ",n2,sep="")
  res <- numeric(nsamples)
  sign <- numeric(nsamples)
  for (i in 1:nsamples){
    svmtrain <- svm(t(data[, -i]) , fac[-i] , kernel="linear" )
    pred <- predict(svmtrain , t(data[,i]) , decision.values=TRUE)
    res[i] <- as.numeric(attributes(pred)$decision.values)
    facnames <- colnames(attributes(pred)$decision.values)[1]
    if (facnames ==
paste(levels(fac)[1], "/",levels(fac)[2],sep="")){sign[i] <- 1}
    if (facnames ==
```

[0304]


```

paste(levels(fac)[2], "/", levels(fac)[1], sep="")){sign[i] <- -1}
}
if (length(unique(sign)) >1){print("error")}
res <- sign * res
names <- colnames(data , do.NULL=FALSE)
orden <- order(res , decreasing=TRUE)
Samples <- data.frame(names[orden],res[orden],fac[orden])
ROCdata <- myROC(res,fac)
SenSpe <- SensitivitySpecificity(res,fac)

return(list(SampleInformation=SampleInformation,ROCarea=ROCdata[1],p
.value=ROCdata[2],SenSpe <- SenSpe,samples=Samples))
}

# Definierar hur analysen ska köras om man INTE ANVÄNDER
apriorianalyter
Analysera<- function(group1 ,group2){
  outputfiletxt <- paste(group1," versus ",group2,".txt" ,sep="")
  outputfilepdf <- paste(group1," versus ",group2,".pdf" ,sep="")
  #outputfilejpeg <- paste(group1," versus ",group2,".jpg" ,sep="")
  subset1 <- is.element(Diagnosis , strsplit(group1,"")[[1]])
  subset2 <- is.element(Diagnosis , strsplit(group2,"")[[1]])
  wilcoxpvalues <- apply(logdata , 1 , wilcoxtest , subset1 ,
subset2)
  foldchange <- apply(logdata , 1 , foldchange , subset1 , subset2)
  QvaluesAll <- BenjaminiHochberg(wilcoxpvalues)
  HugeTable <-
cbind(ProteinNames,foldchange,wilcoxpvalues,QvaluesAll)
  write.table(HugeTable, file=outputfiletxt , quote=FALSE,
sep="\t",row.names=FALSE)
  color <- rep('black' , length(subset1))
  color[subset1] <- 'red'
  color[subset2] <- 'blue'
  pdf(outputfilepdf)
  #jpeg(outputfilejpeg, quality=100, width=600, height=600)
  Sam <- sammon(dist(t(logdata[,subset1|subset2])) , k=2)
  plot(Sam$points , type="n" , xlab = NA , ylab=NA, main="All
proteins" ,asp=1)
  text(Sam$point , labels = colnames(logdata[,subset1|subset2]),
col=color[subset1|subset2])
  heatmap.2(logdata[,subset1|subset2] , labRow = row.names(logdata),
trace="none" , labCol = "" , ColSideColors=
color[subset1|subset2],col=pal , na.color= "grey", key=FALSE ,
symkey =FALSE , tracecol = "black" , main = "" , dendrogram= 'both' ,
scale = "row" ,cexRow=0.2)
  svmfac <-
factor(rep('rest',ncol(logdata)),levels=c(group1,group2,'rest'))
  svmfac[subset1] <- group1
  svmfac[subset2] <- group2
  svmResAll <- svmLOOvalues(logdata[,subset1|subset2] ,
factor(as.character(svmfac[subset1|subset2]),levels=c(group1,group2)
))
  ROCplot(svmResAll , sensspecnumber=4)

  # N <- length(ProteinNames)

```

[0305]

```
# par(mfrow=c(1,2))
# for(k in 1:N){
#   boxplot(data[k,subset1],data[k,subset2], names=c(group1, group2),
main=c(ProteinNames[k]," test"))
# }

write("", file=outputfiletxt, append=TRUE)
write("All proteins", file=outputfiletxt, append=TRUE)
write("", file=outputfiletxt, append=TRUE)
for (i in 1:5){write.table(svmResAll[[i]], file=outputfiletxt,
append=TRUE, sep="\t", quote=FALSE)
write("", file=outputfiletxt, append=TRUE)
}
dev.off()
}

Analysera("X","Y") # 1.3 Select comparisons to do-----
-----
```

[0306]

[0307]

보충 표(supplemental table) 1

Sample ID	Hist. Grade	Age	Tumor size (mm)	ER / PgR / HER2 / ki67_gt_25	Lymph_pos	Nr of Lymph pos.
6616	1	37,62	22	+/+/-/-	yes	1
6617	1	66,30	20	+/+/-/na	yes	5
7149	1	74,49	31	+/+/-/-	no	0
7454	1	47,82	22	+/+/-/-	yes	5
7940	1	53,94	30	+/+/-/-	no	0
8415	1	66,61	31	+/+/-/-	yes	4
9317	1	47,42	18	+/+/-/-	no	0
9795	1	43,26	15	+/+/-/-	yes	2
10524	1	64,34	30	+/+/-/-	no	0
4404	2	49,92	25	+/+/-/+	yes	1
5614	2	45,48	37	+/+/-/-	yes	8
5096	2	37,35	6	+/+/-/+	yes	8
5572	2	43,55	18	-/-/-/-	yes	2
6096	2	36,92	12	+/+/-/-	yes	1
6627	2	43,63	15	+/+/-/-	yes	2
7015	2	46,77	22	+/+/-/na	yes	1
7267	2	48,39	22	+/+/-/-	yes	1
7296	2	46,38	14	+/+/-/na	yes	4
8173	2	47,03	25	+/+/-/na	yes	10
9257	2	43,78	7	+/+/-/+	no	0
9340	2	52,10	29	+/+/-/-	no	0
5402	2	44,26	50	-/-/-/-	yes	5
6514	2	49,18	30	-/-/na/na	no	0
7424	2	47,98	25	+/+/-/+	yes	1
8278	2	47,54	10	+/+/-/-	yes	1
8504	2	49,66	25	+/+/-/-	yes	1
5706	3	41,19	50	-/-/-/+	yes	5
4239	3	40,66	33	-/-/-/+	no	0
5744	3	44,04	21	+/+/-/na/-	no	0
5811	3	49,75	45	-/-/-/+	yes	1
5997	3	46,37	20	-/-/-/na	no	0
6009	3	49,57	20	+/+/-/-	no	0
6029	3	42,80	25	-/-/na/+	yes	4
6158	3	55,81	20	+/+/-/+	yes	2
6191	3	45,04	25	-/-/-/-	no	0
6276	3	52,30	32	-/-/+/+	yes	6
4723	3	48,89	40	-/-/+/+	yes	3

[0308]

5198	3	46,66	32	-/+/-/na	yes	4
5203	3	33,22	30	-/-/-/-	yes	1
5634	3	44,33	25	+/-/-/na	yes	2
5996	3	50,94	22	-/-/+/-	yes	2
6013	3	41,60	50	+/-/-/na	yes	3
6176	3	50,62	35	+/-/-/+	no	0
6503	3	43,39	28	-/-/-/+	no	0
6877	3	34,39	27	+/-/-/+	yes	8
7694	3	47,66	18	-/-/-/+	yes	1
7722	3	46,61	27	+/-/na/-	no	0
8613	3	44,04	35	+/-/-/-	yes	6
9322	3	50,33	30	-/-/-/+	no	0
9460	3	49,01	17	+/-/-/+	no	0
5784	na	na	na	na/na/na/na	na	na
4917	na	na	na	na/na/na/na	na	na

na = not available

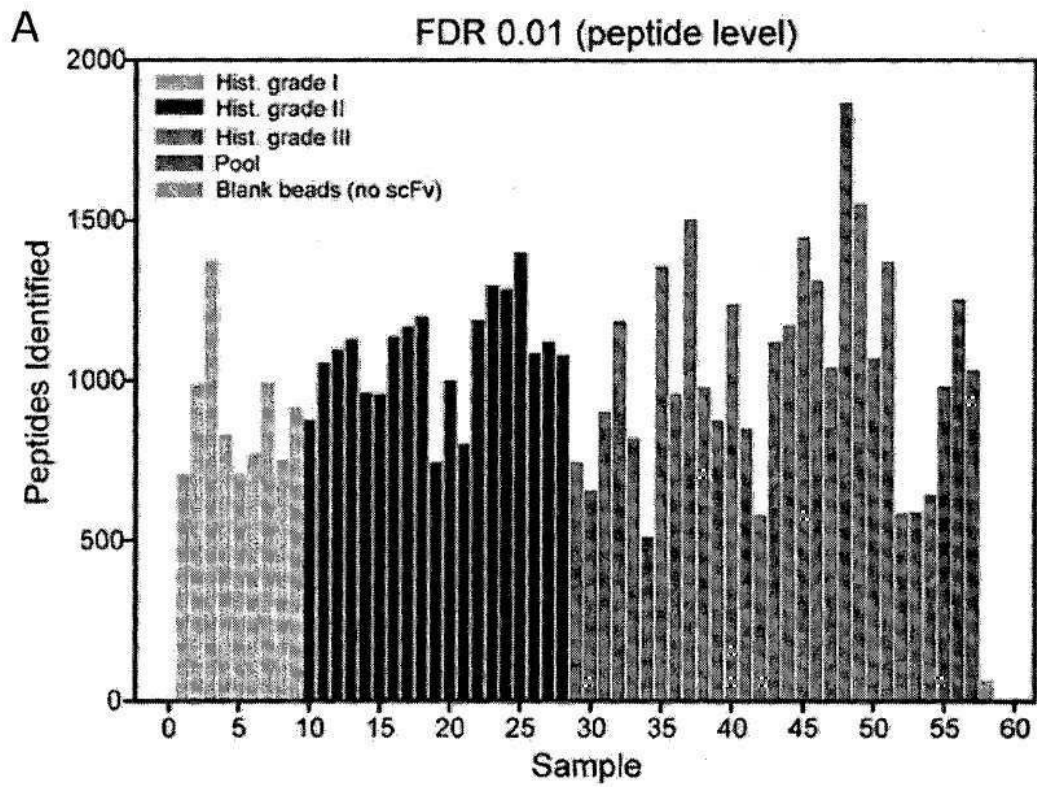
보충 표 2

CIMS antibody*	Selection peptide	Affinity (K _D) (μM)	Mix
CIMS-33-3C-A09	Biotin-SGSGLSADHR	1.6	1
CIMS-33-3D-F06	Biotin-SGSGLSADHR	5.1	1
CIMS-17-C08	Biotin-SGSGSSAYSR	0.2	2
CIMS-17-E02	Biotin-SGSGSSAYSR	0.4	2
CIMS-15-A06	Biotin-SGSGLTEFAK	2.2	3
CIMS-34-3A-D10	Biotin-SGSGSEHLR	2.5	3
CIMS-1-B03	Biotin-SGSGEDFR	3.5	4
CIMS-31-001-D01	Biotin-SGSGLNWVGK	NA	4
CIMS-32-3A-G03	Biotin-SGSGQEASFK	11.5	4

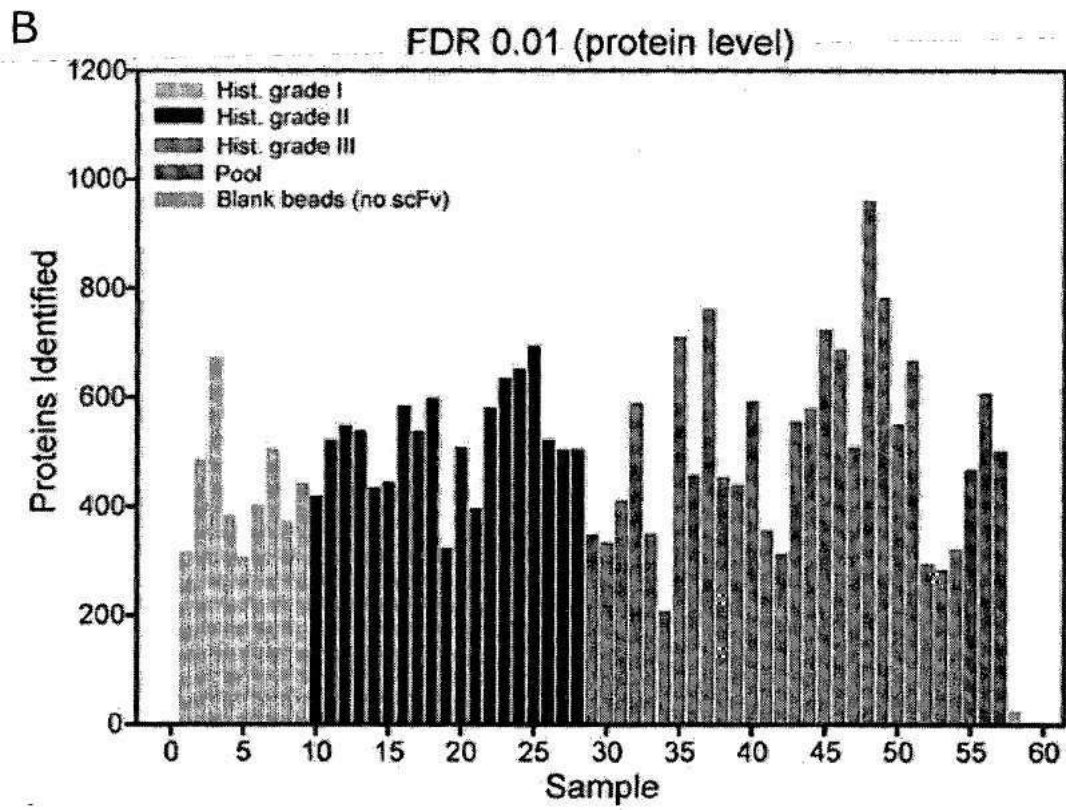
* For details regarding binder characteristics see Olsson et al. (2011) MCP M110.003962.

도면

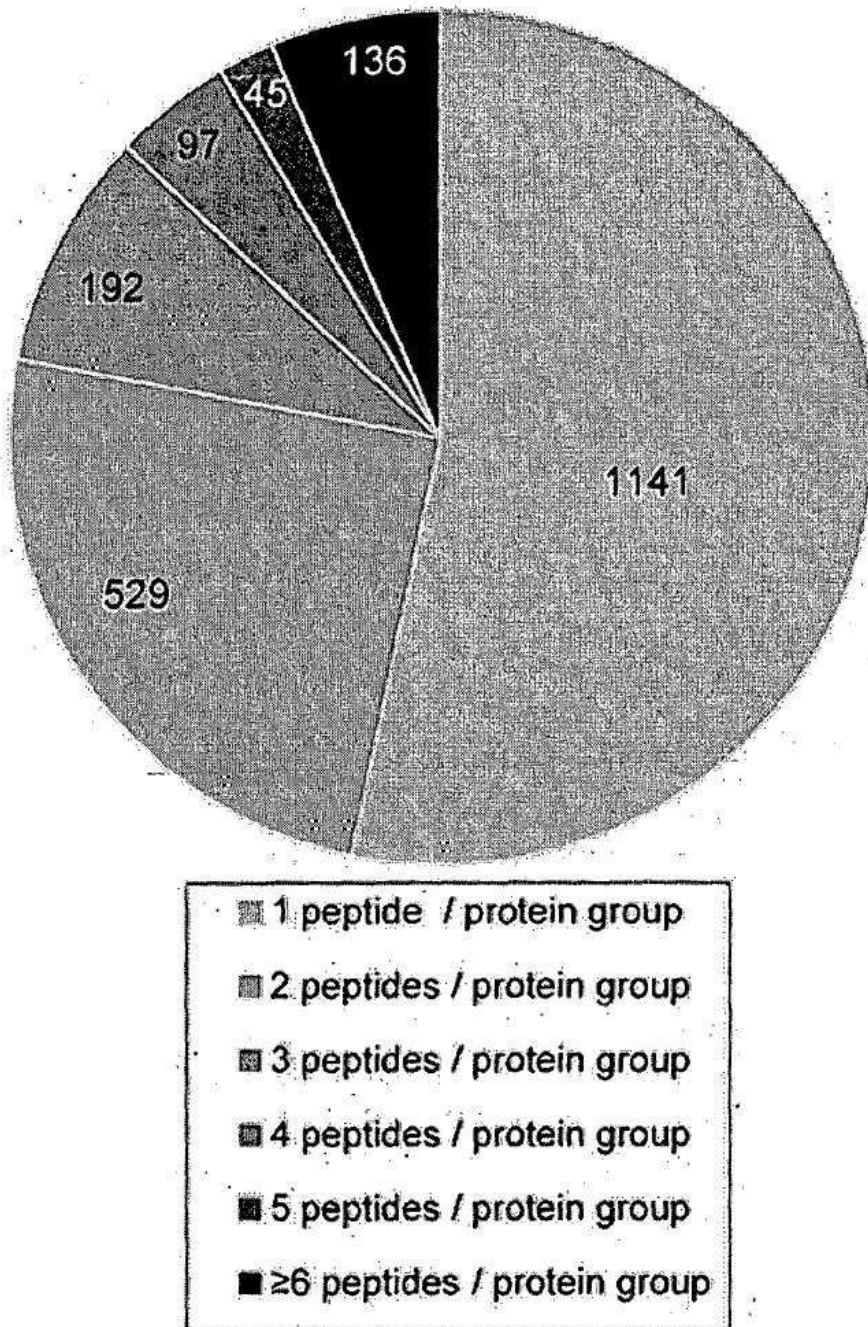
도면1a



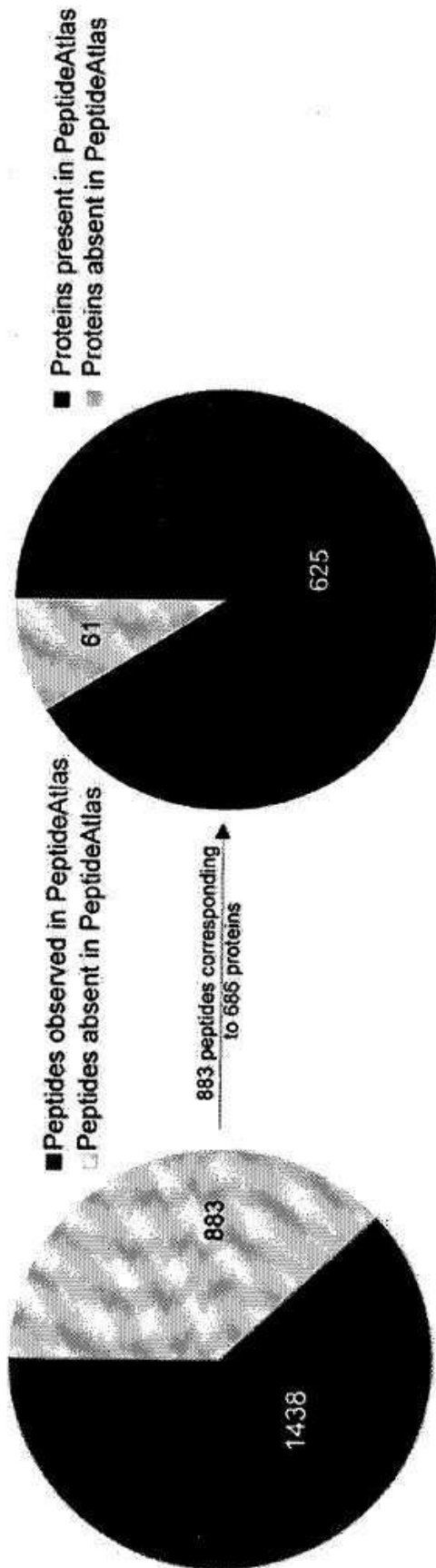
도면1b



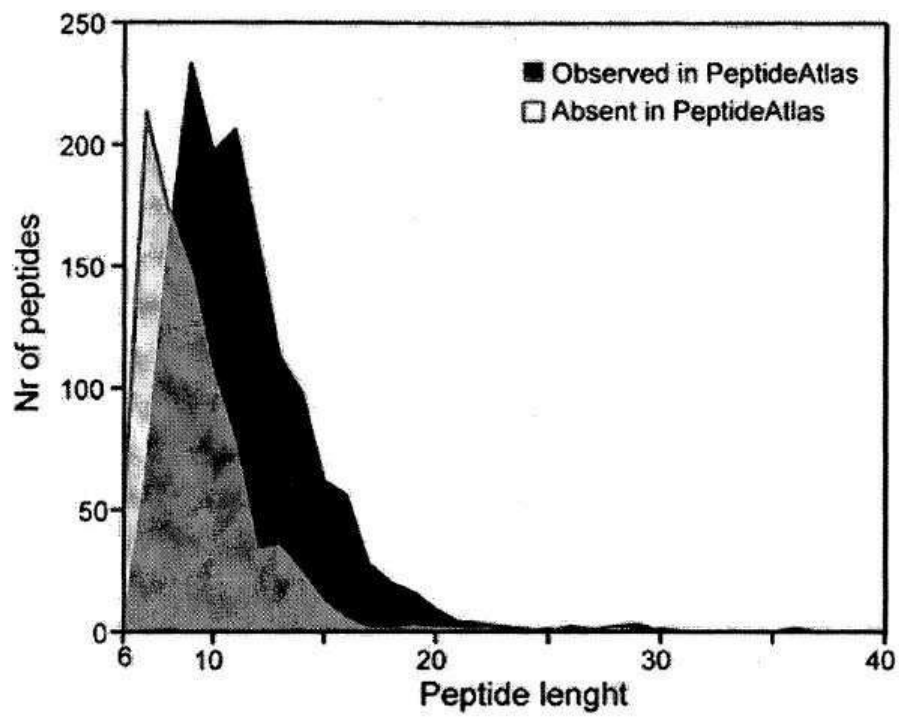
도면1c



도면1d



도면1e



도면1f

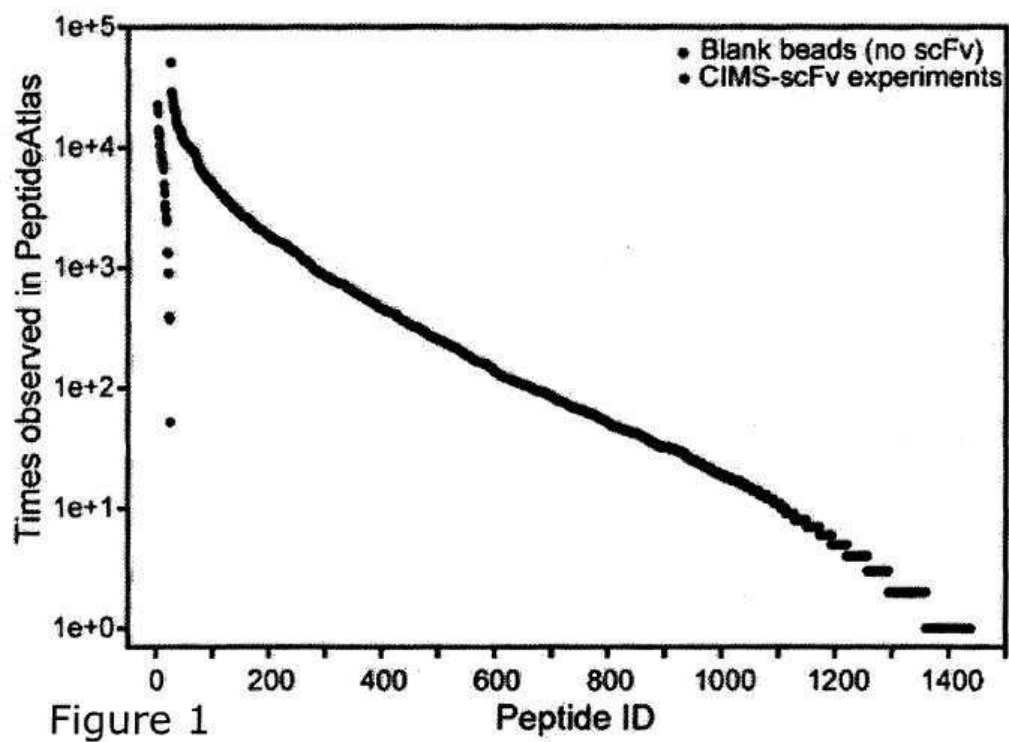
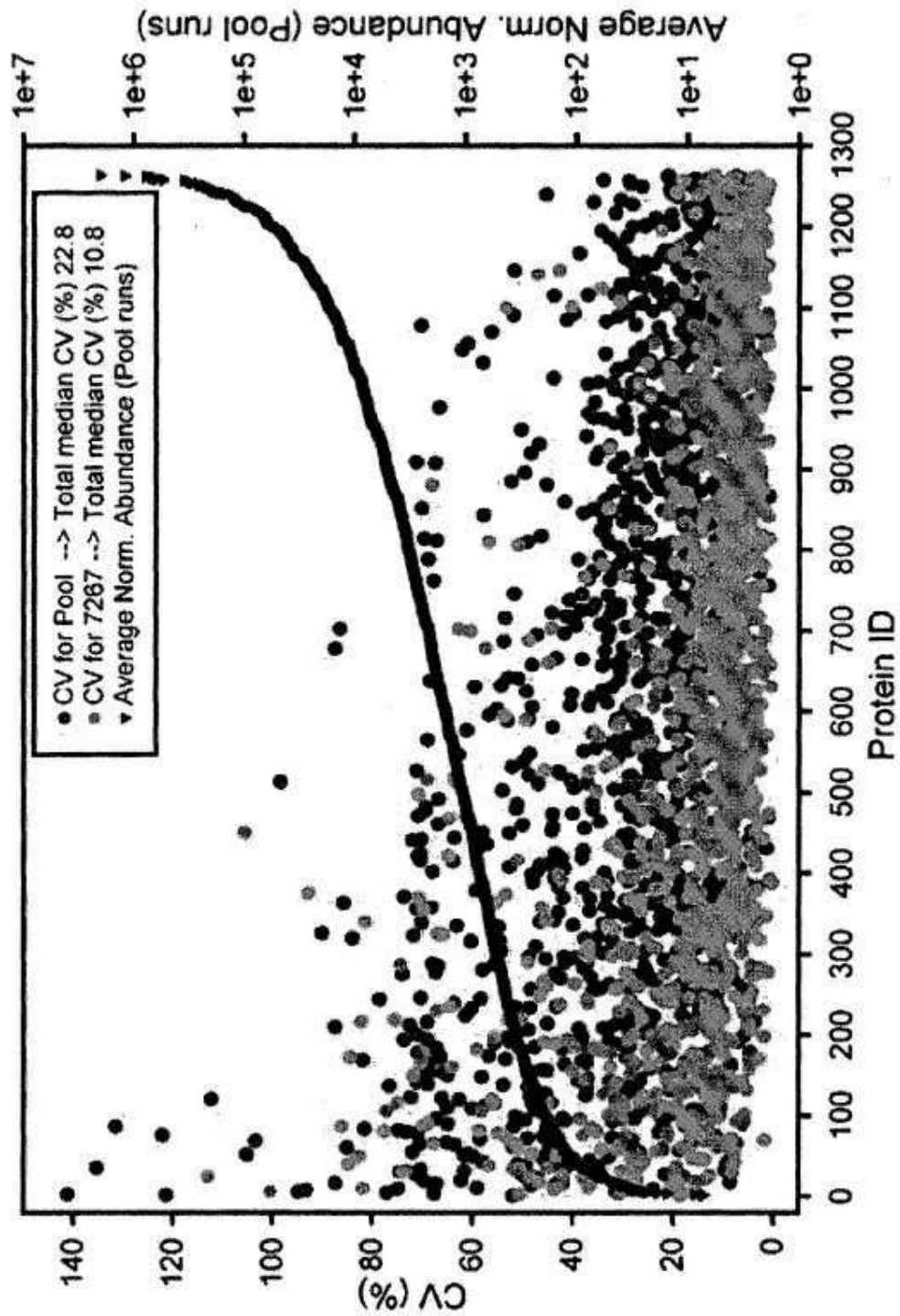
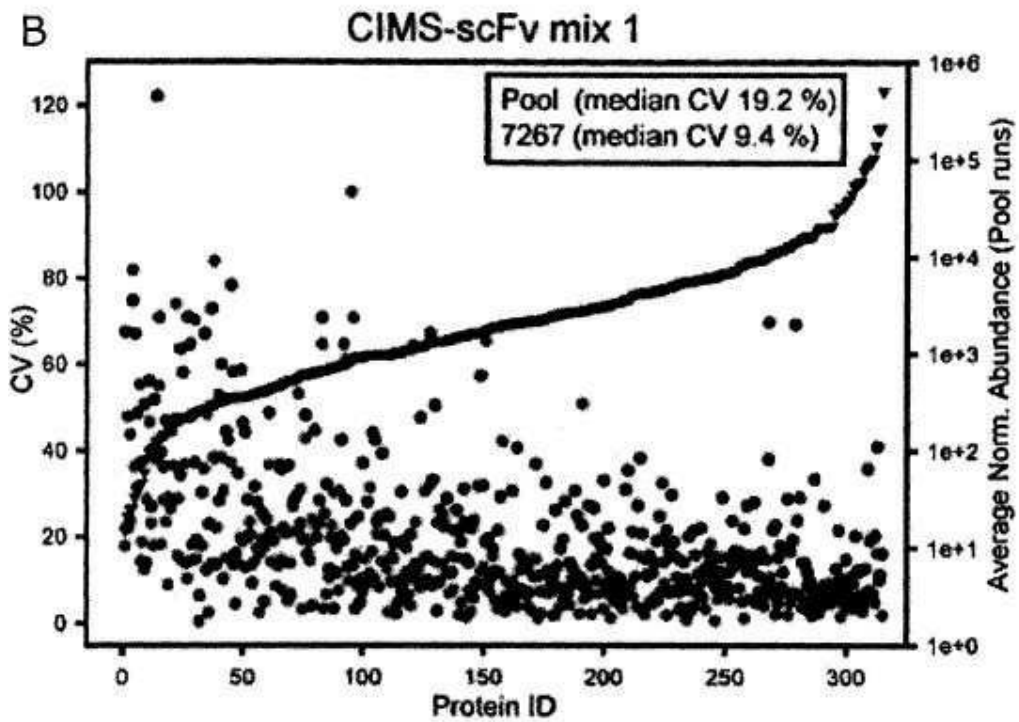


Figure 1

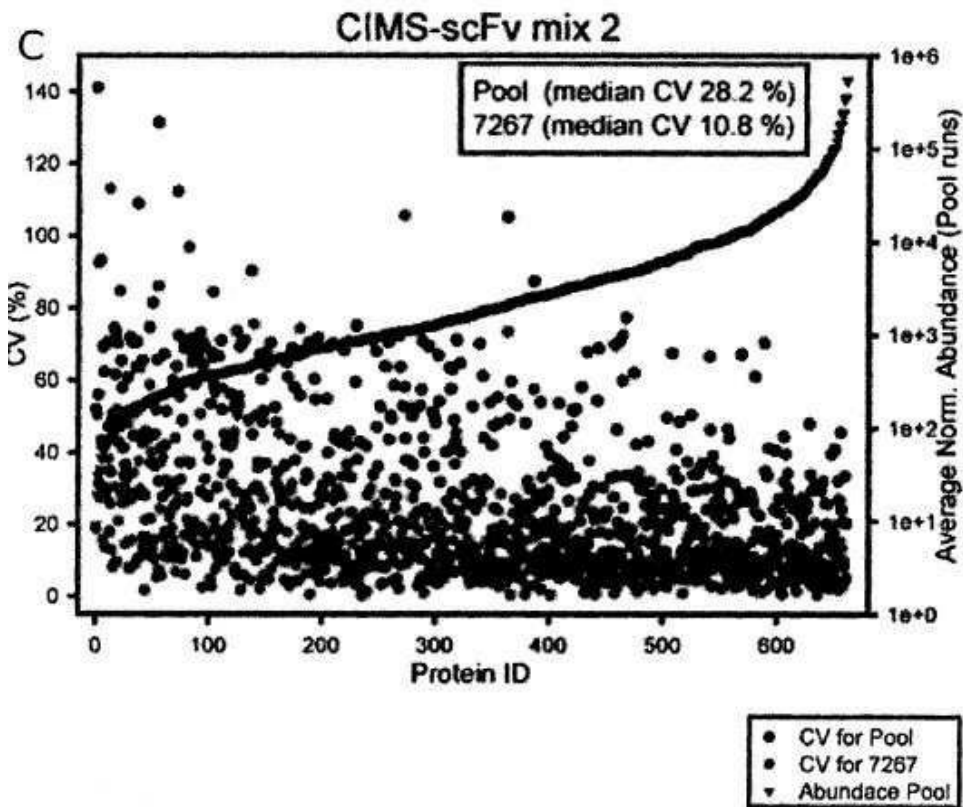
도면2a



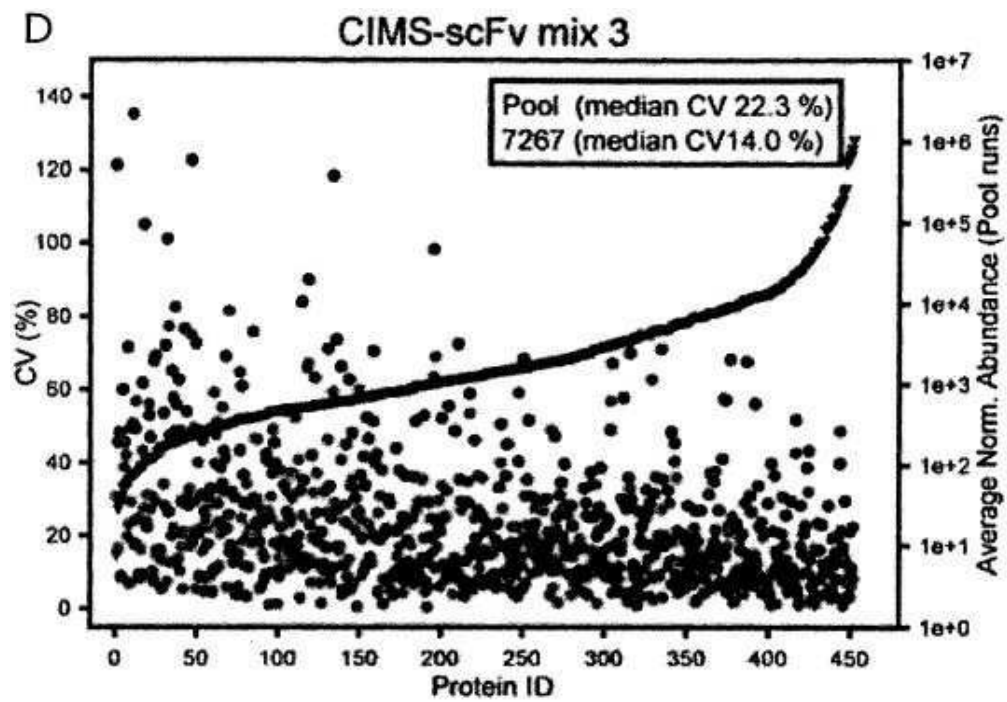
도면2b



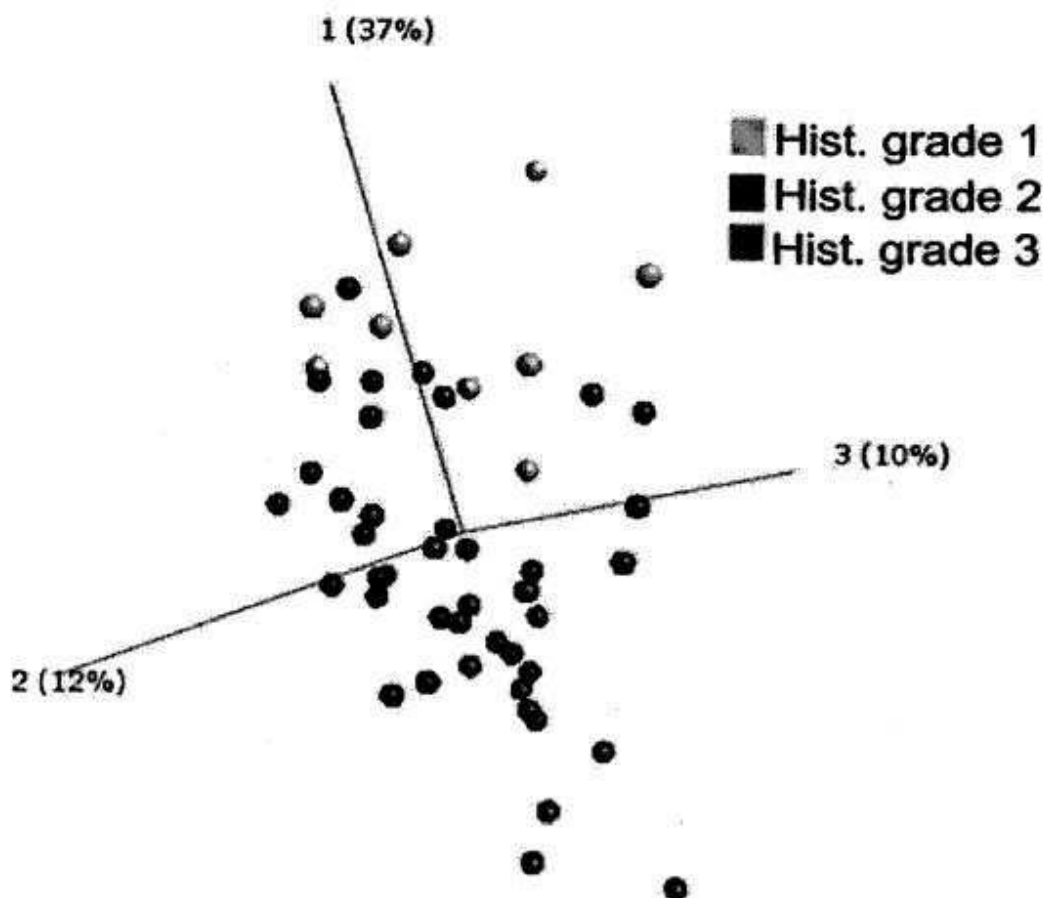
도면2c



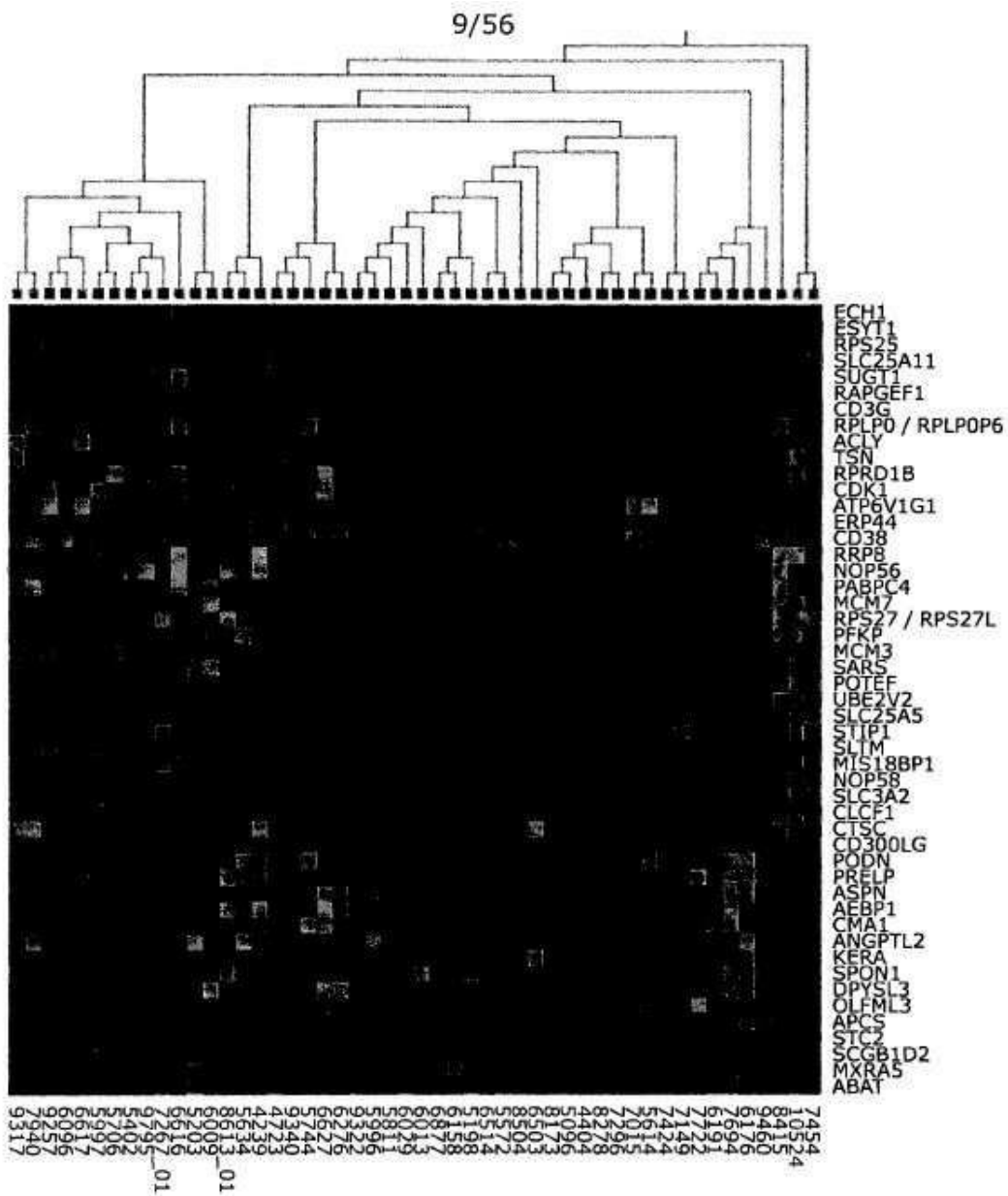
도면2d



도면3aa

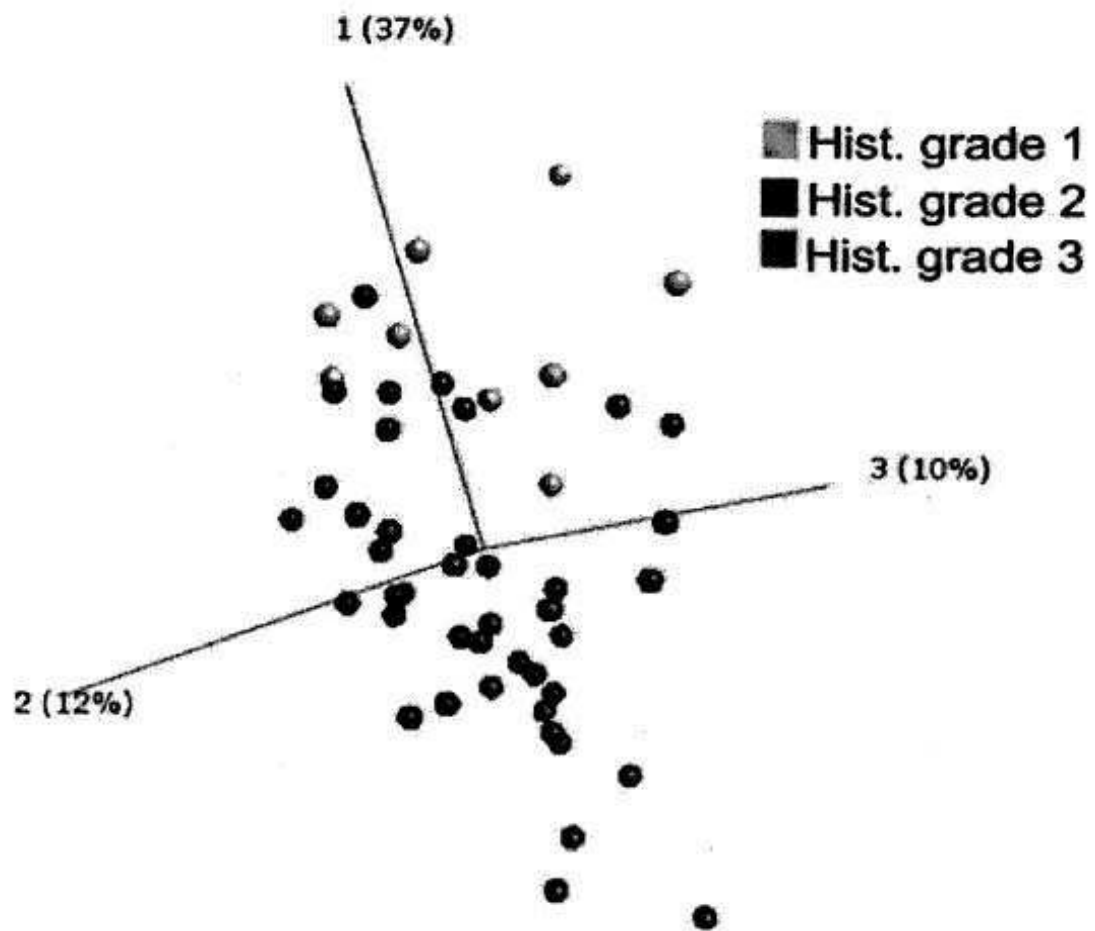


도면3ab

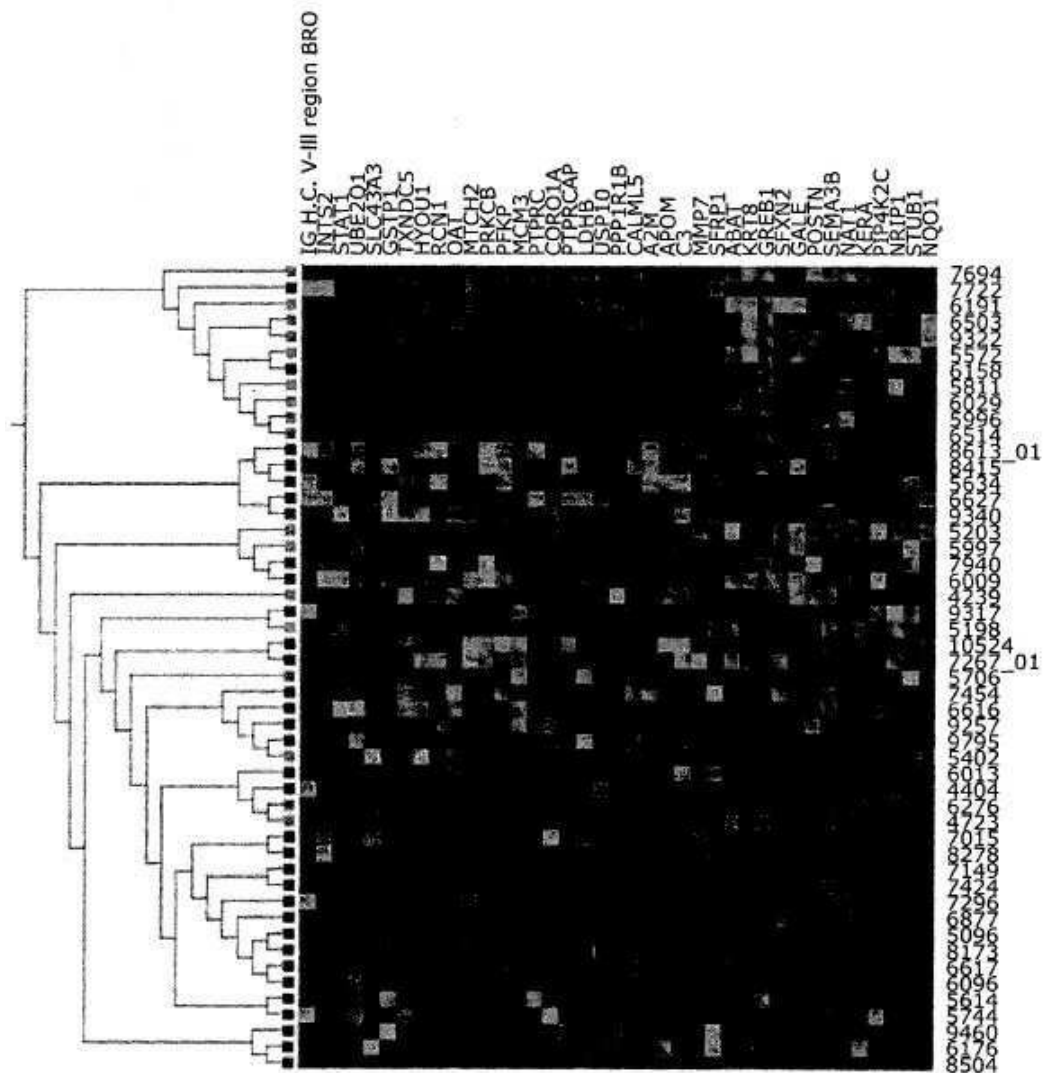


Comparison	ROC area
H1 vs. H2	0.91
H1 vs. H3	0.93
H2 vs. H3	0.75
H1, H2 vs. H3	0.77
H1 vs. H2, H3	0.88
H2 vs. H1, H3	0.7

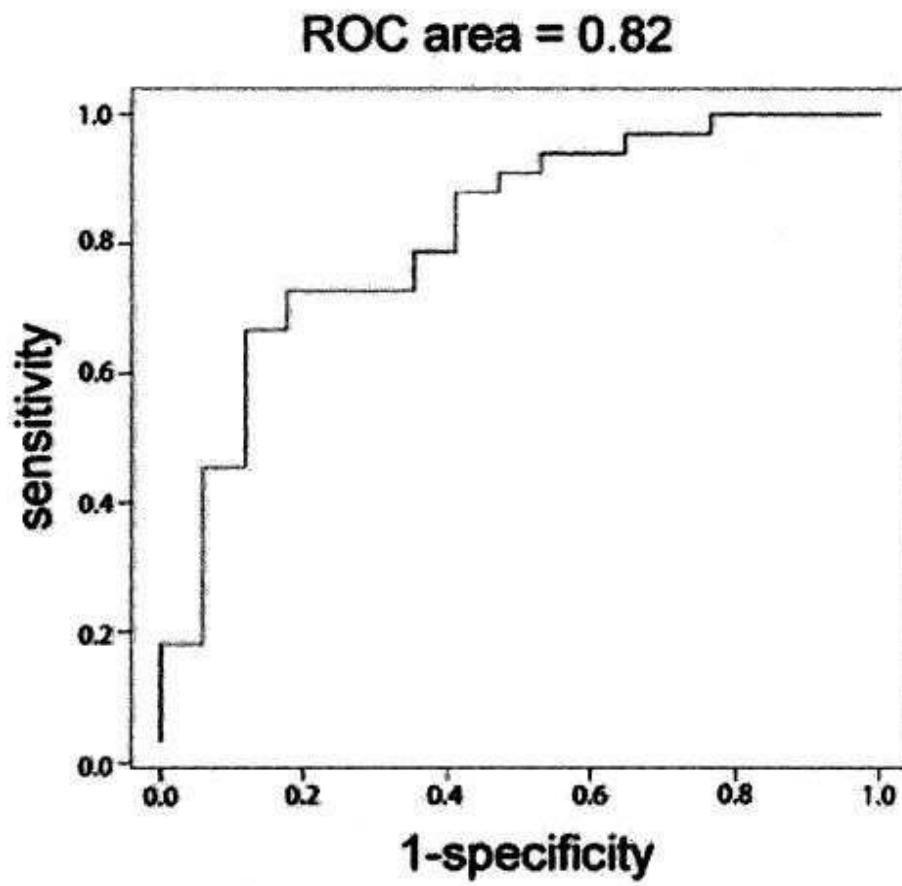
도면3ba



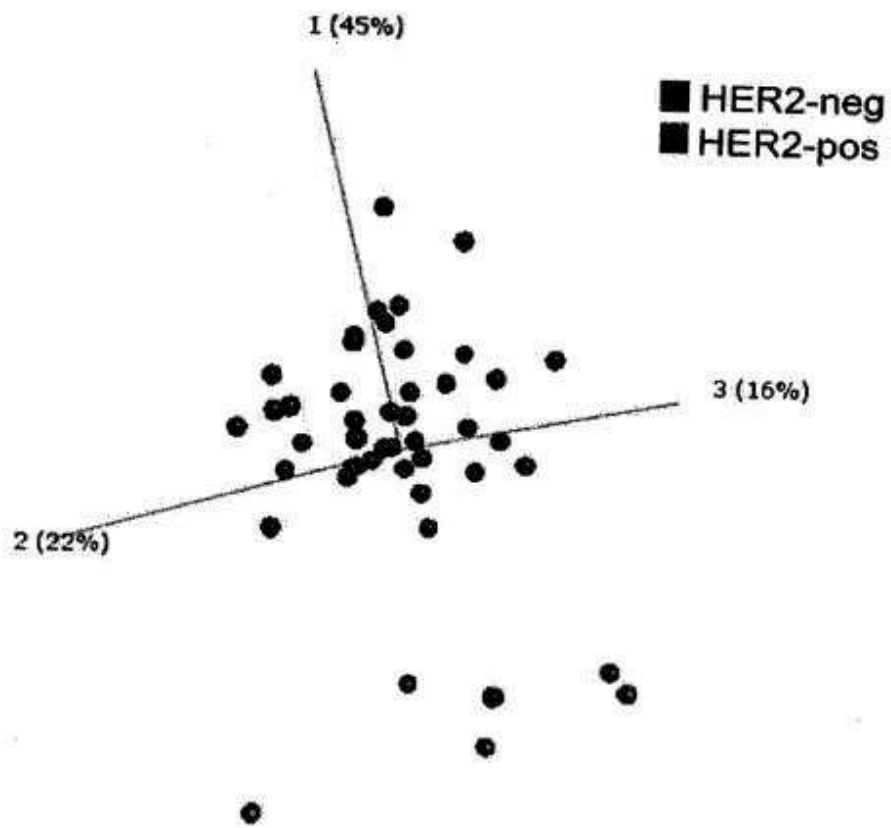
도면3bb



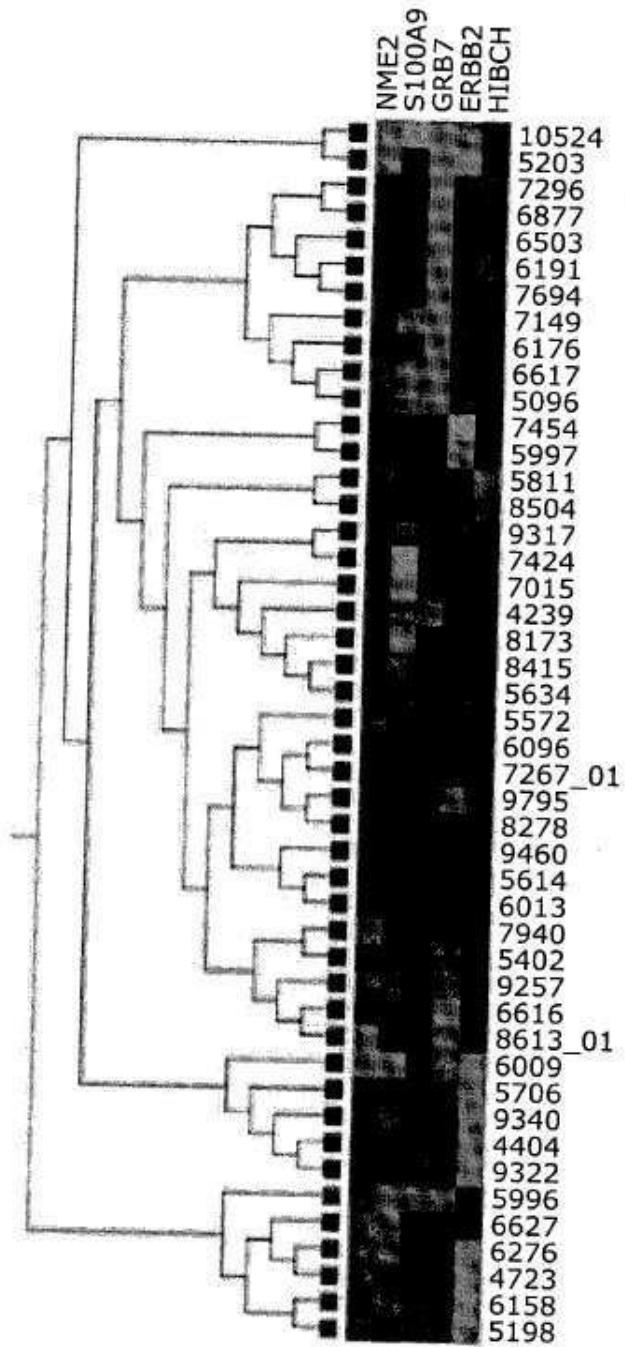
도면3bc



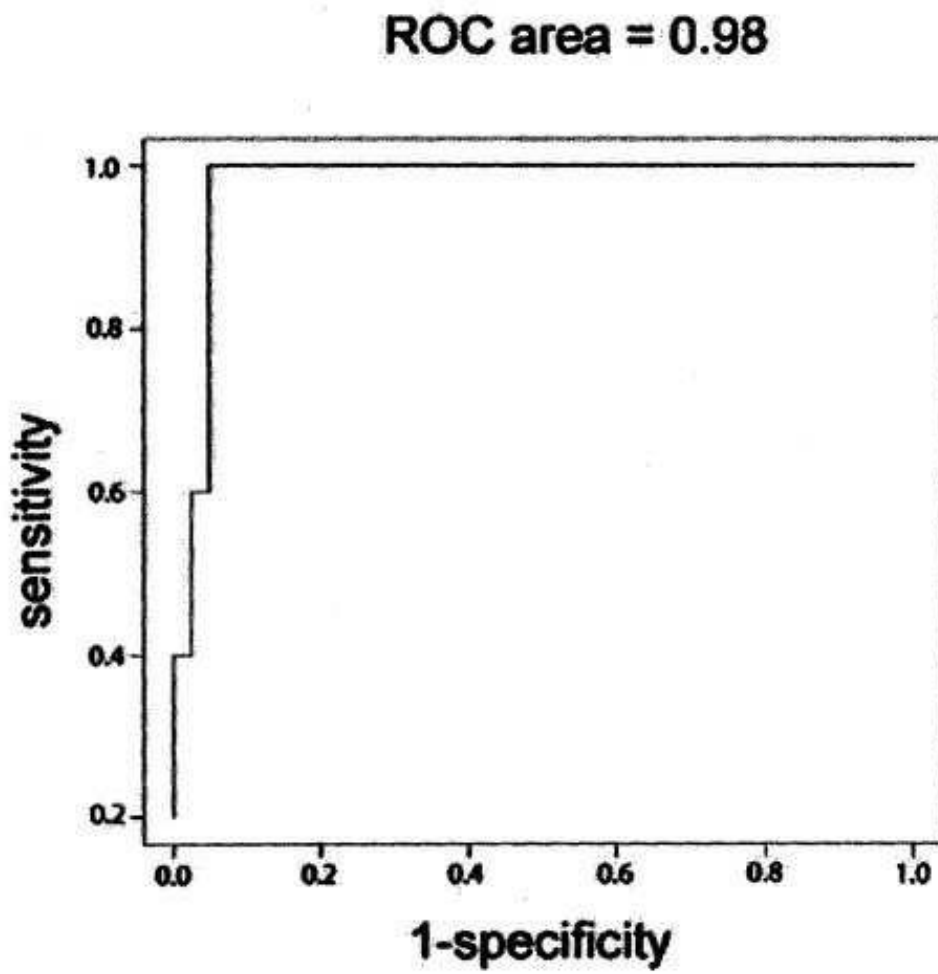
도면3ca



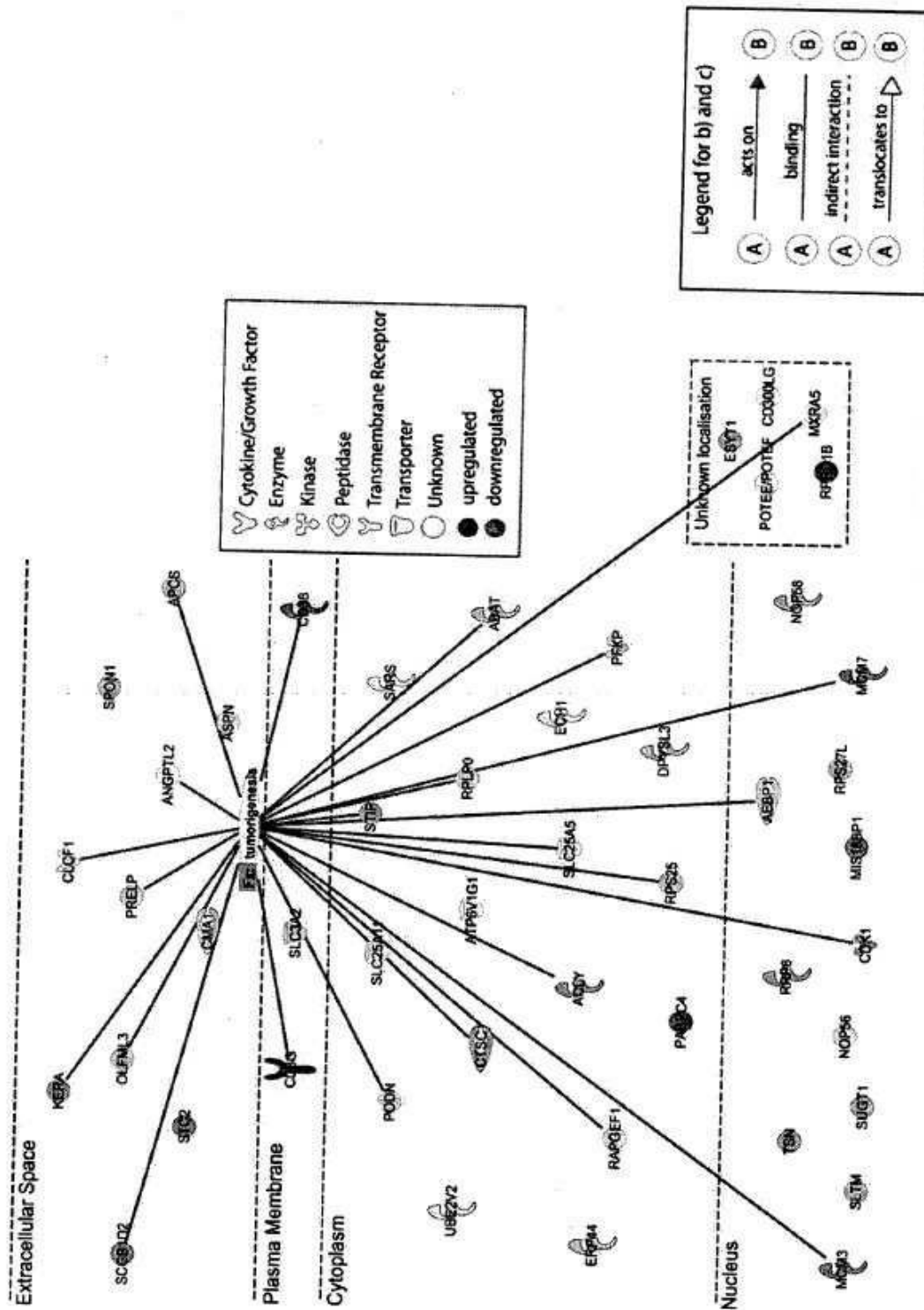
도면3cb



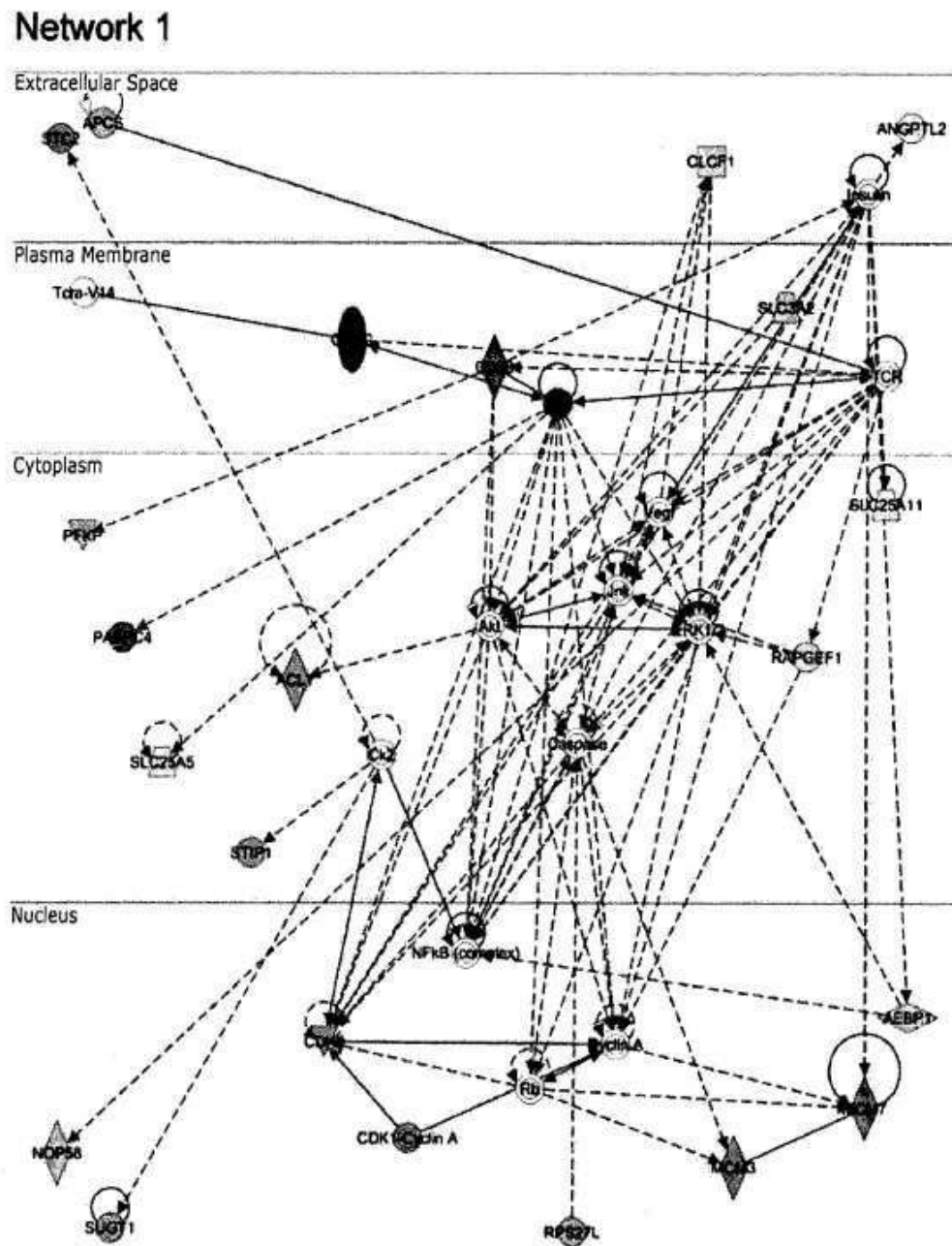
도면3cc



도면4a



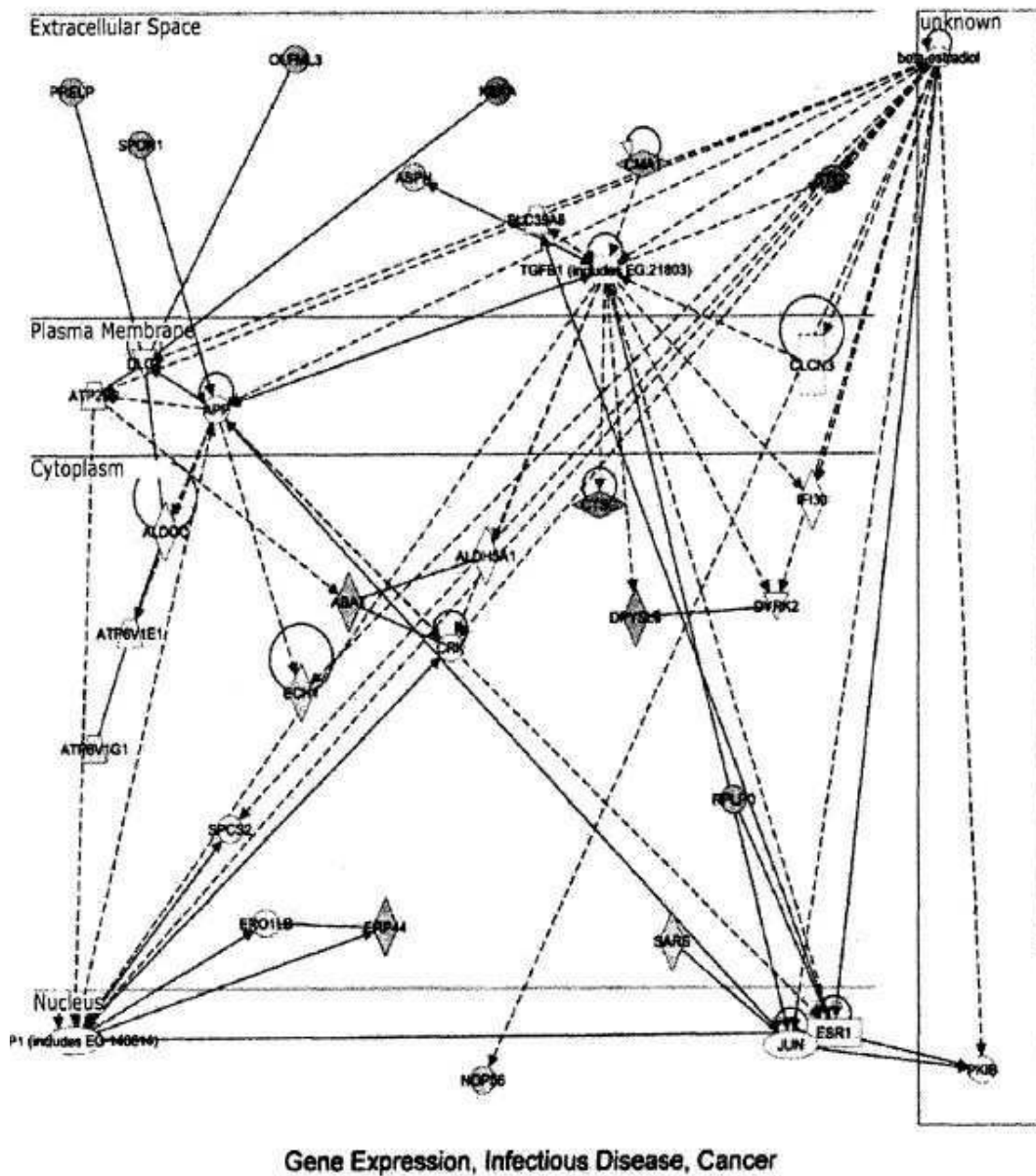
도면4b



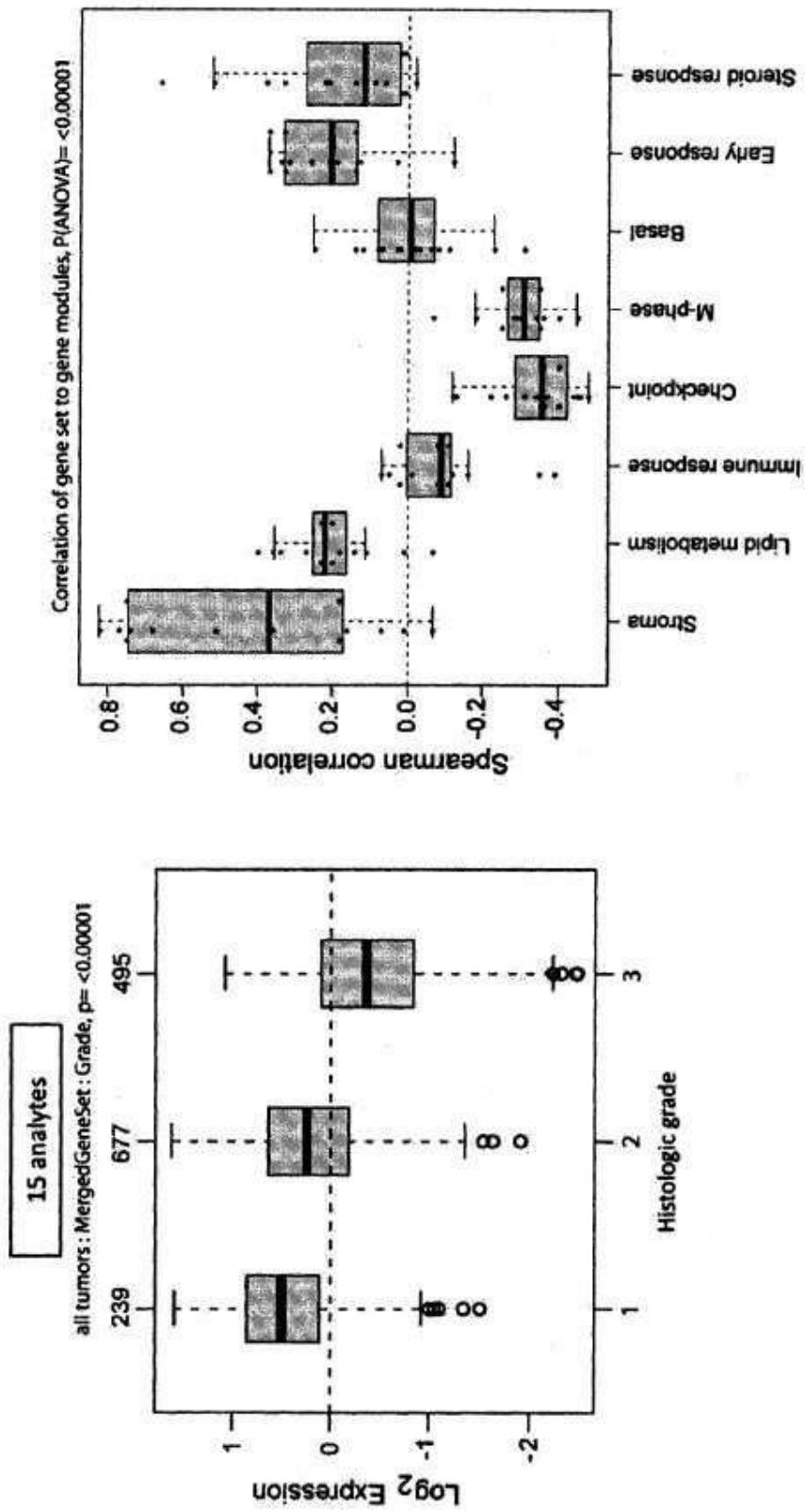
DNA Replication, Recombination, and Repair, Cell Cycle, Free Radical Scavenging

도면4c

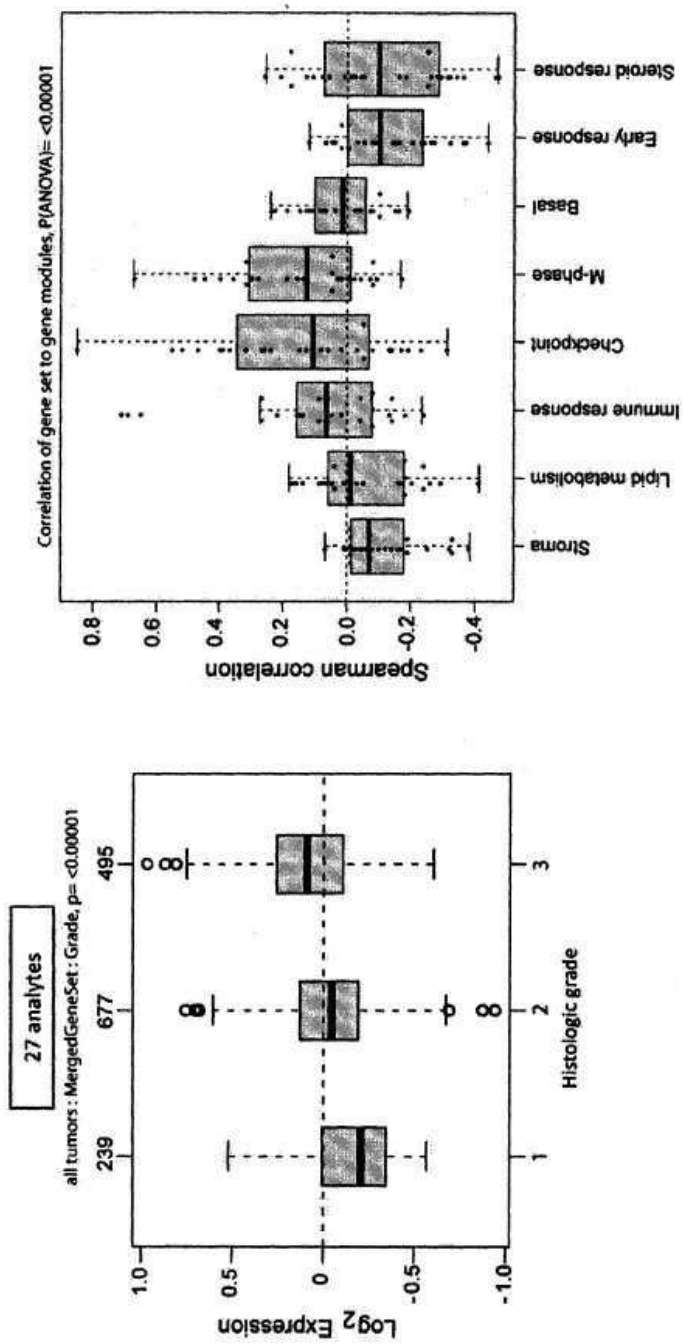
Network 2



도면5a



도면5b



도면6

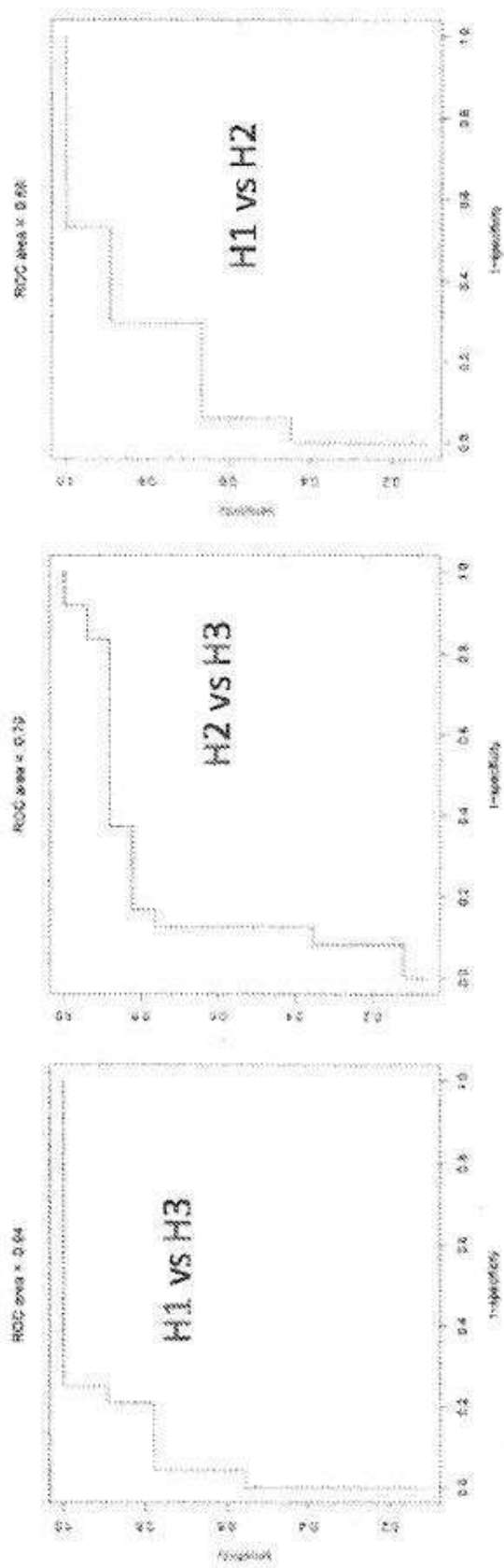
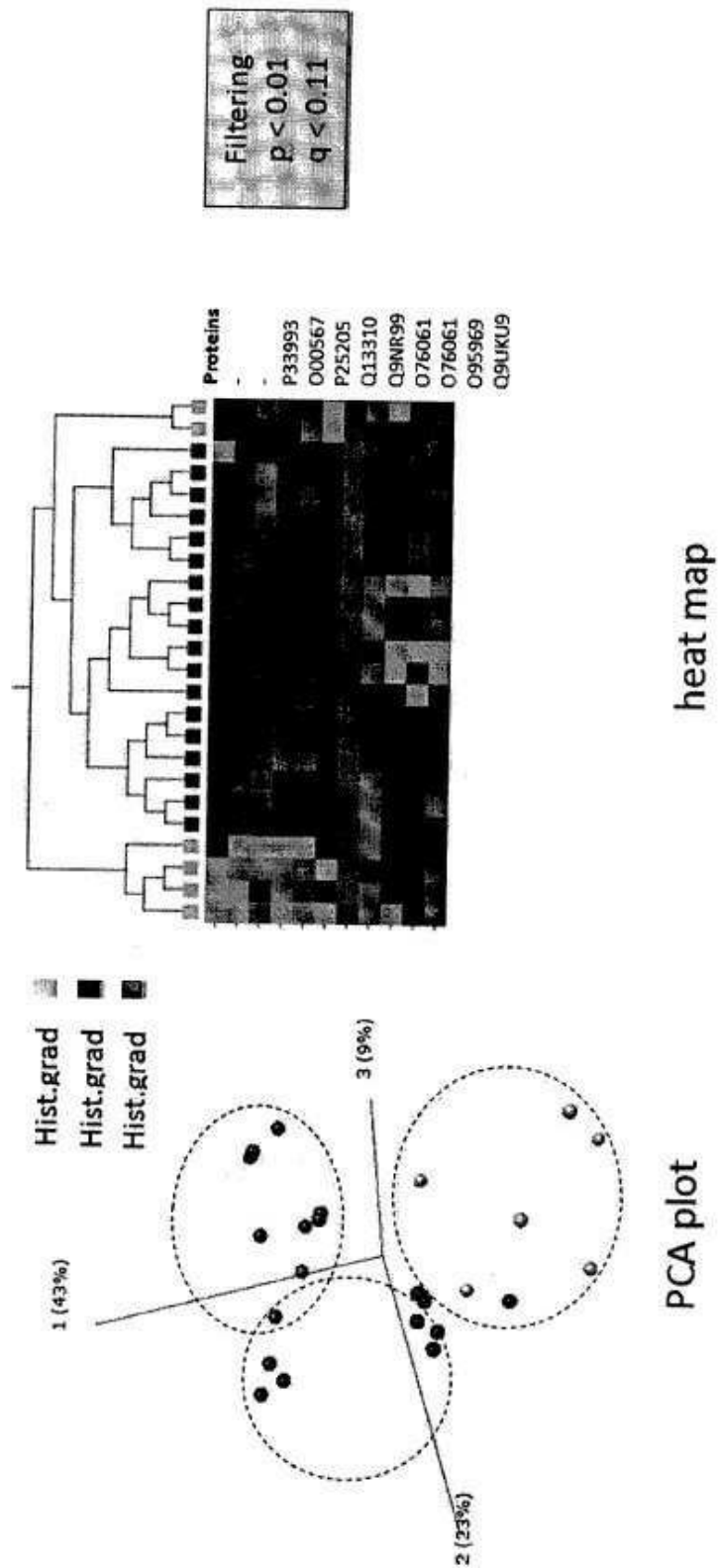
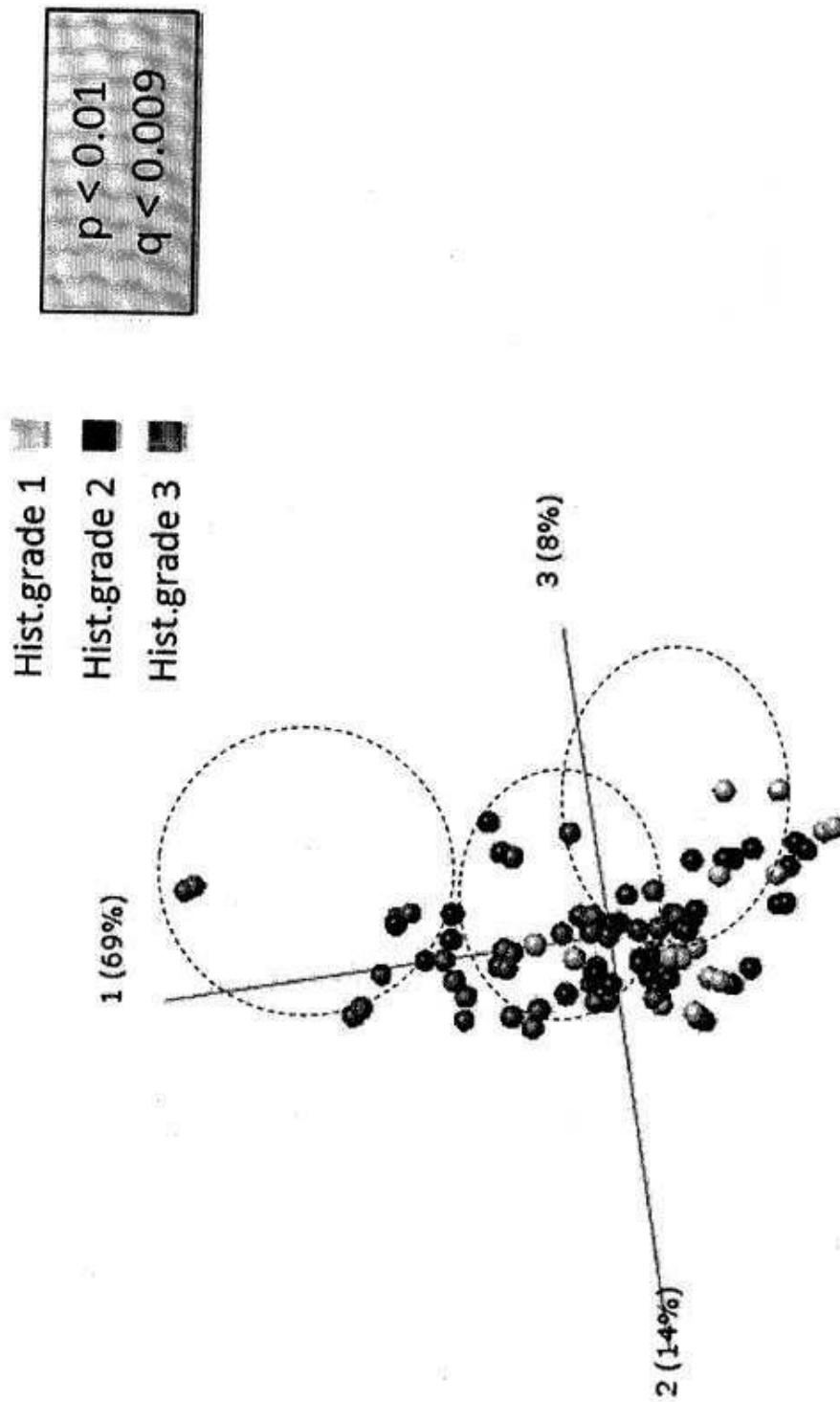


Figure 1 ROC AUC values for the 79-plex biomarker signature discriminating H1 vs H2, H2 vs H3, and H1 vs H3, respectively.

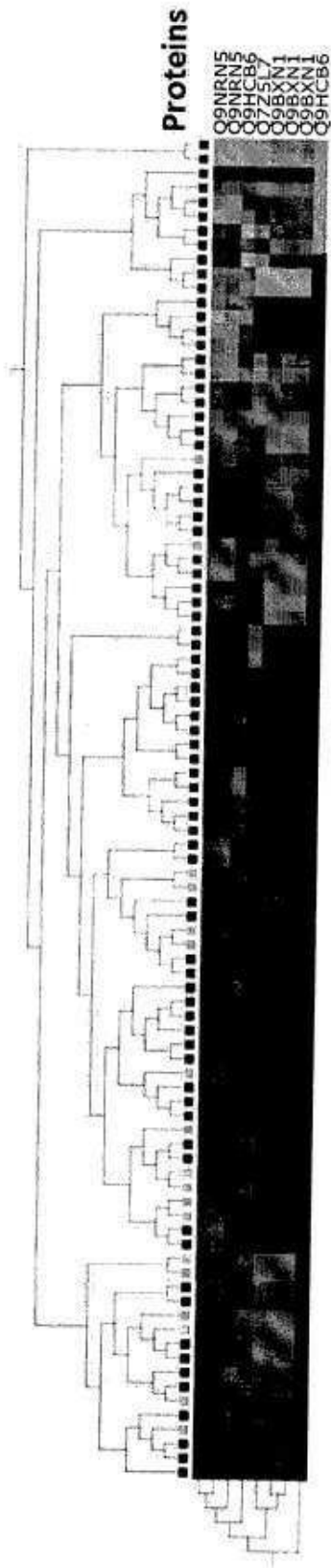
도면7



도면8a

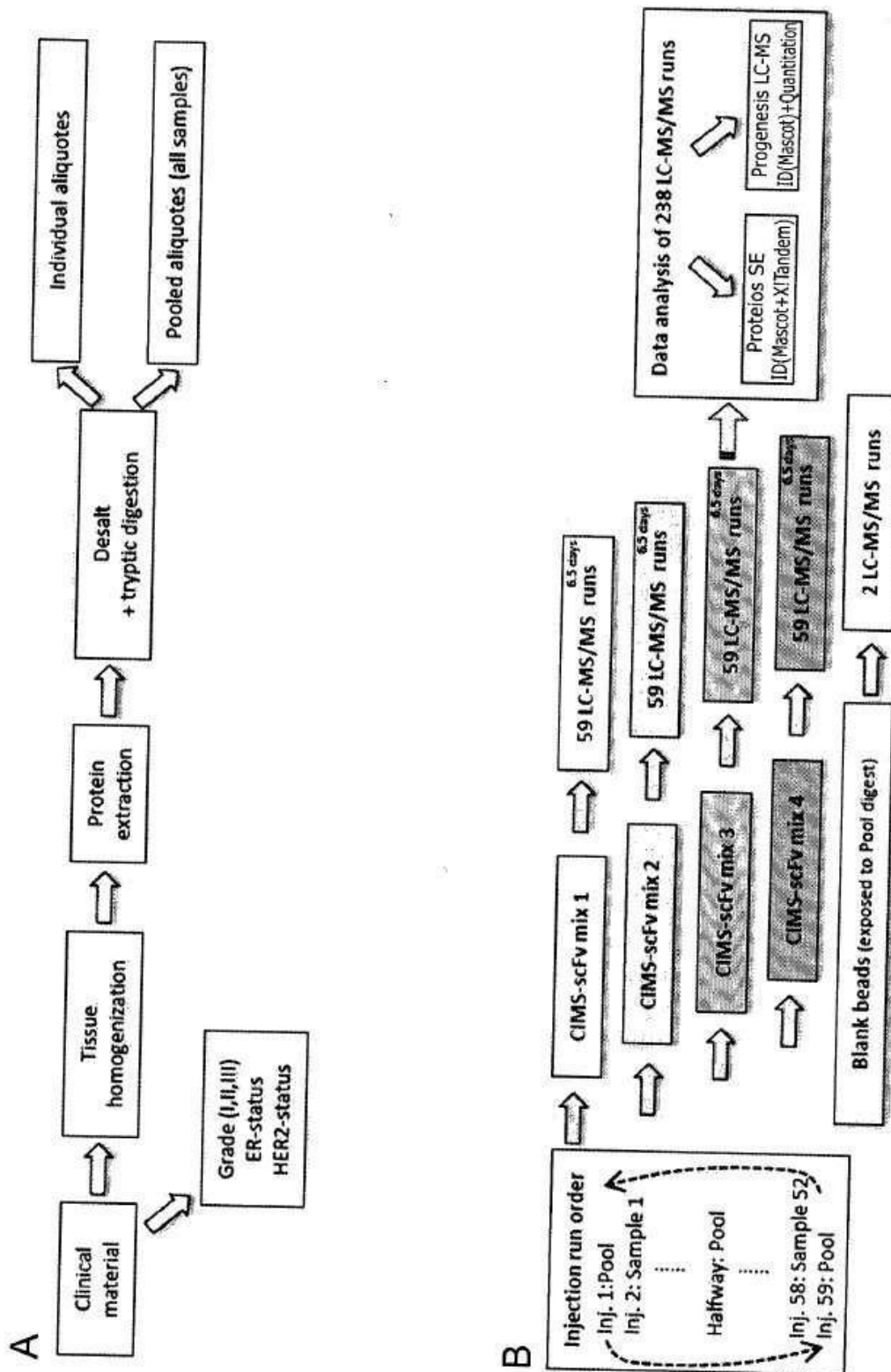


도면 8b



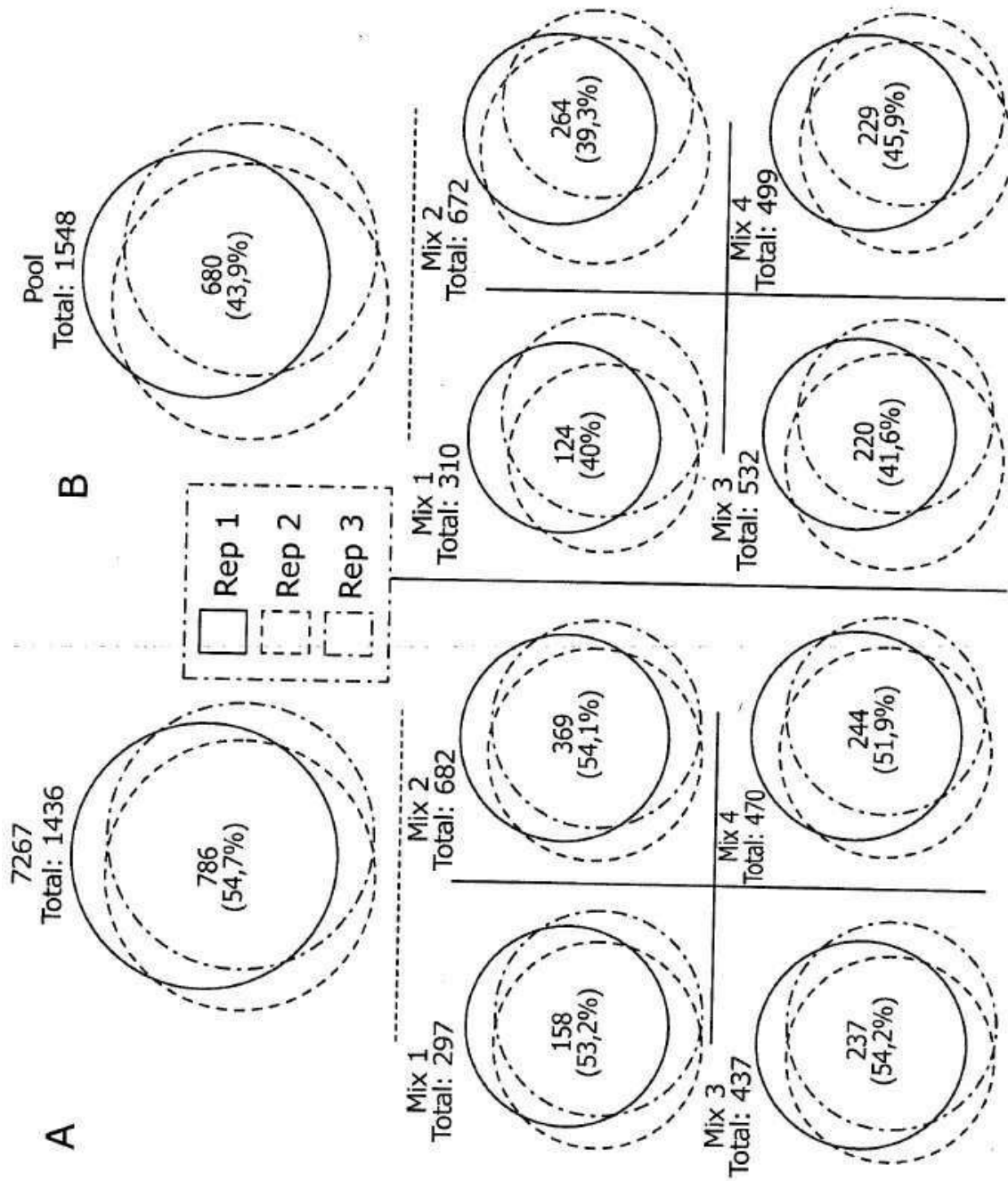
도면9

Supplementary Figure 1



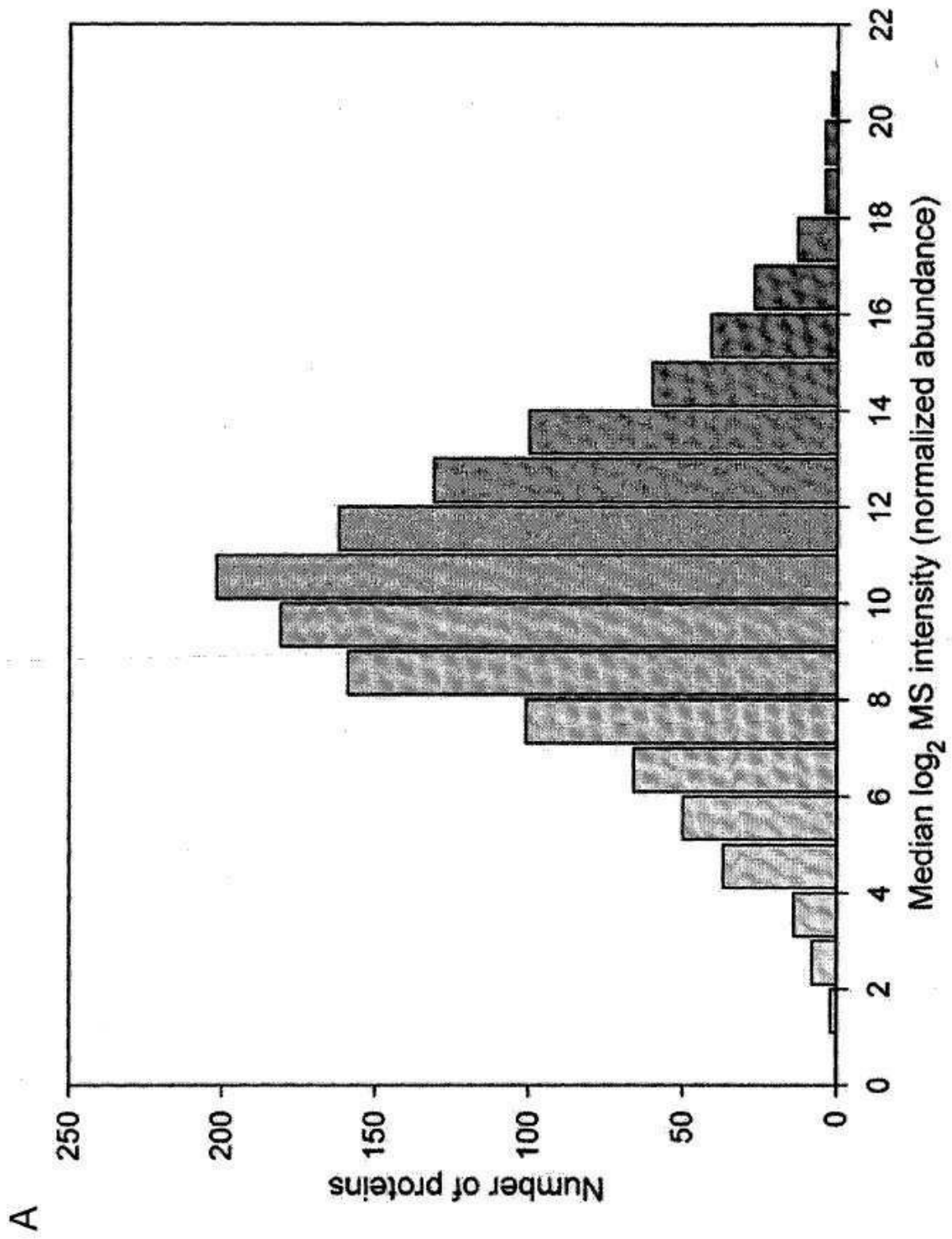
도면10

Supplementary Figure 2



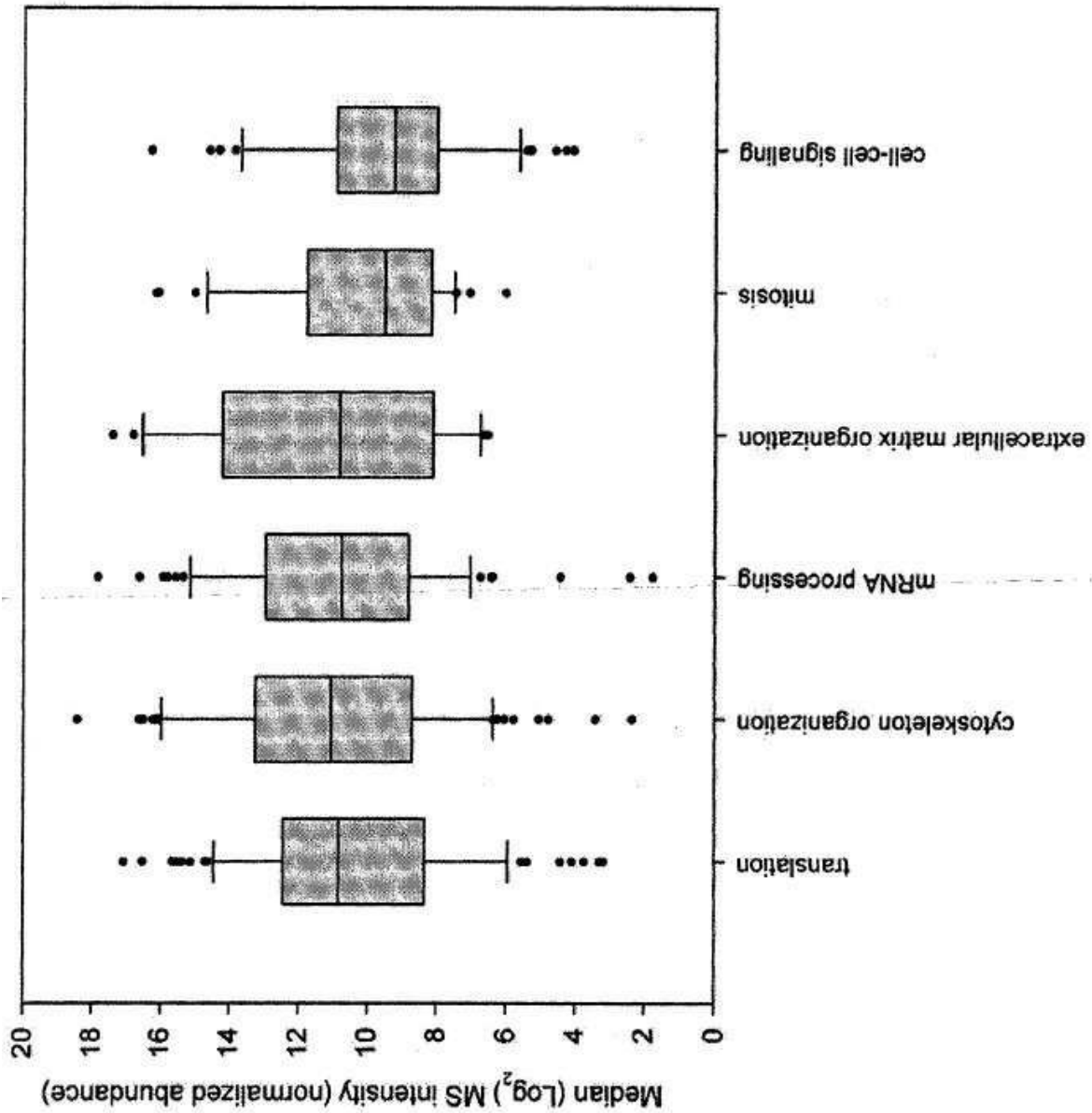
도면11

Supplementary Figure 3



도면12

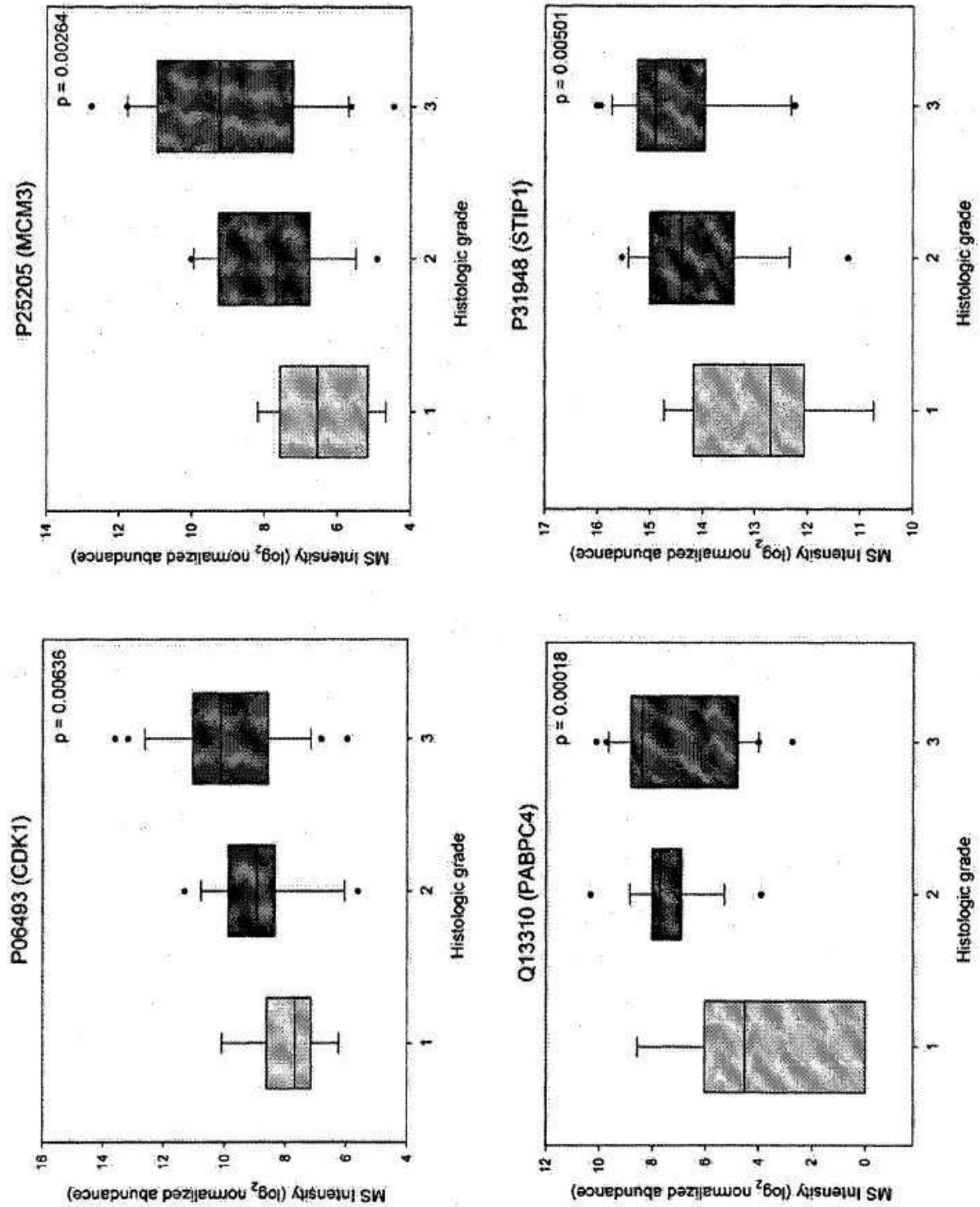
Supplementary Figure 3



B

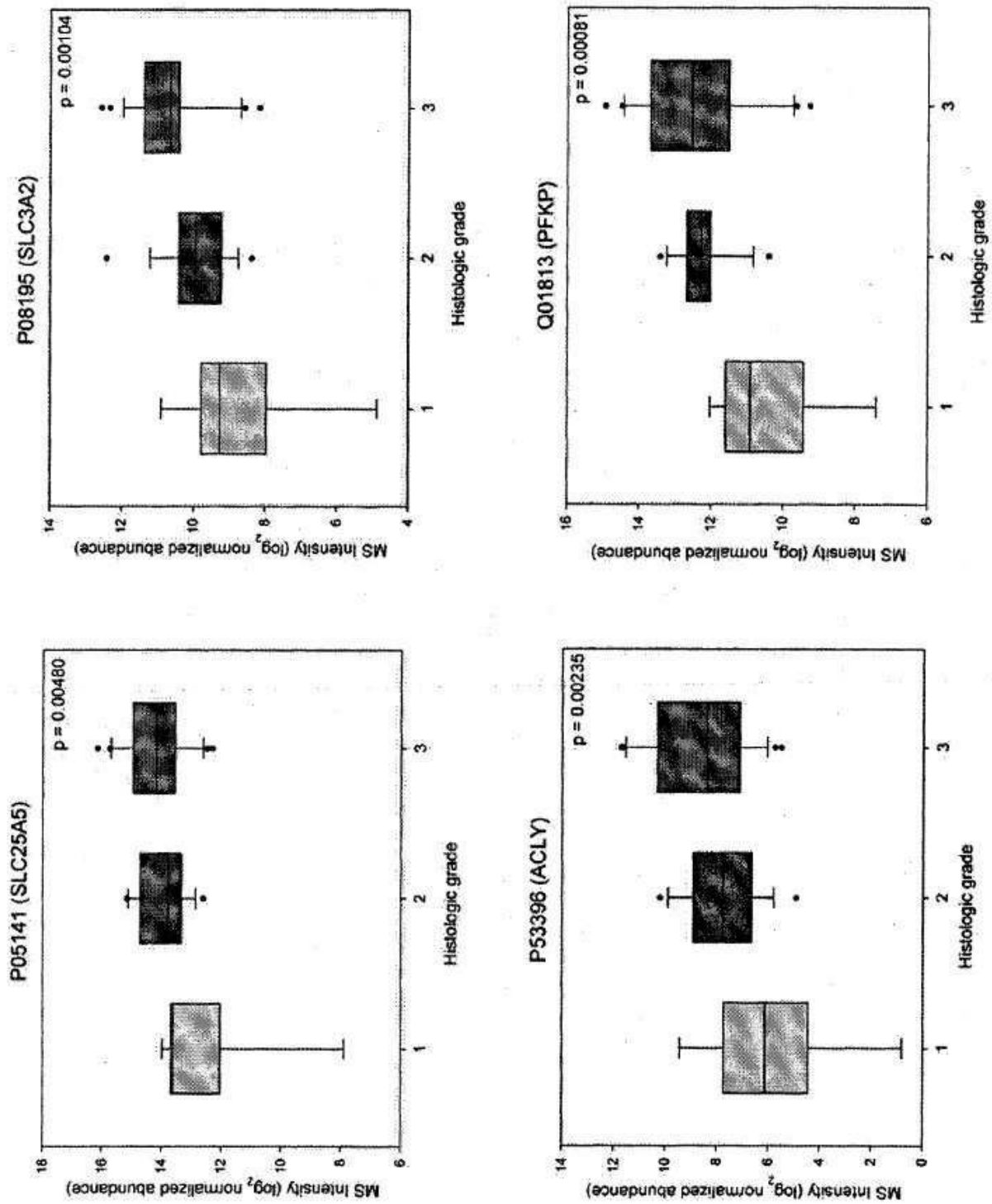
도면13

Supplementary Figure 4



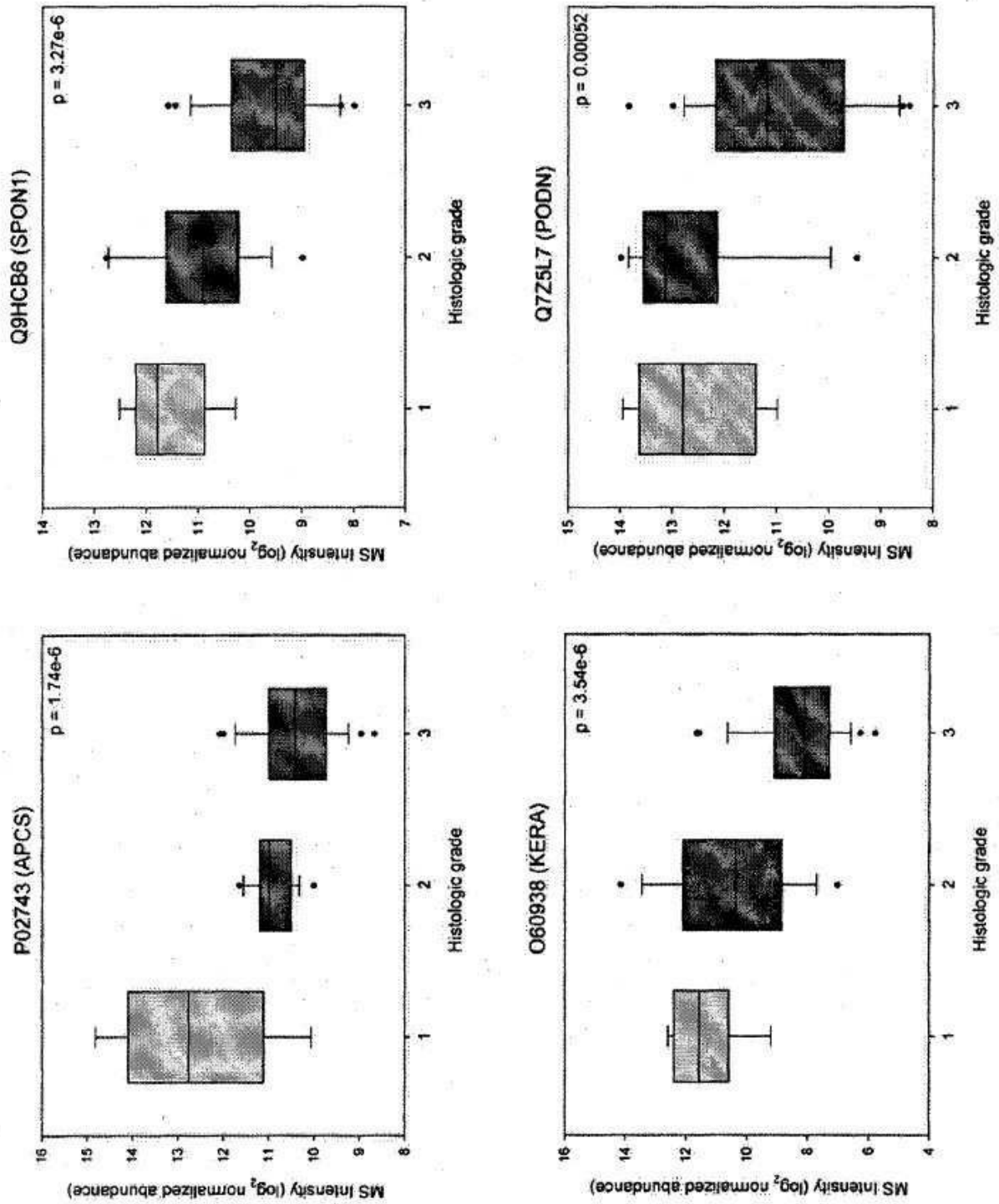
도면14

Supplementary Figure 4
(Continued)



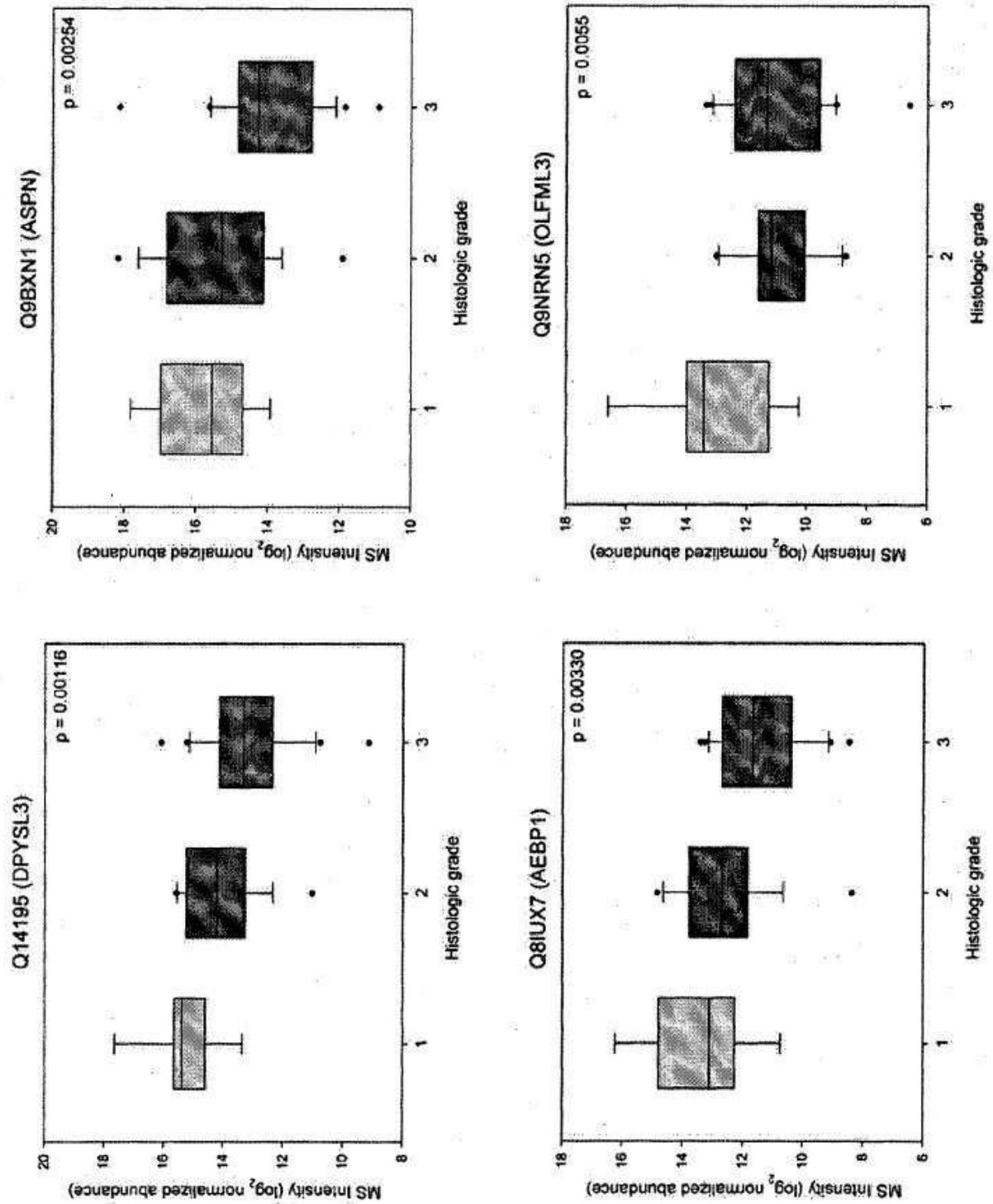
도면15

Supplementary Figure 5

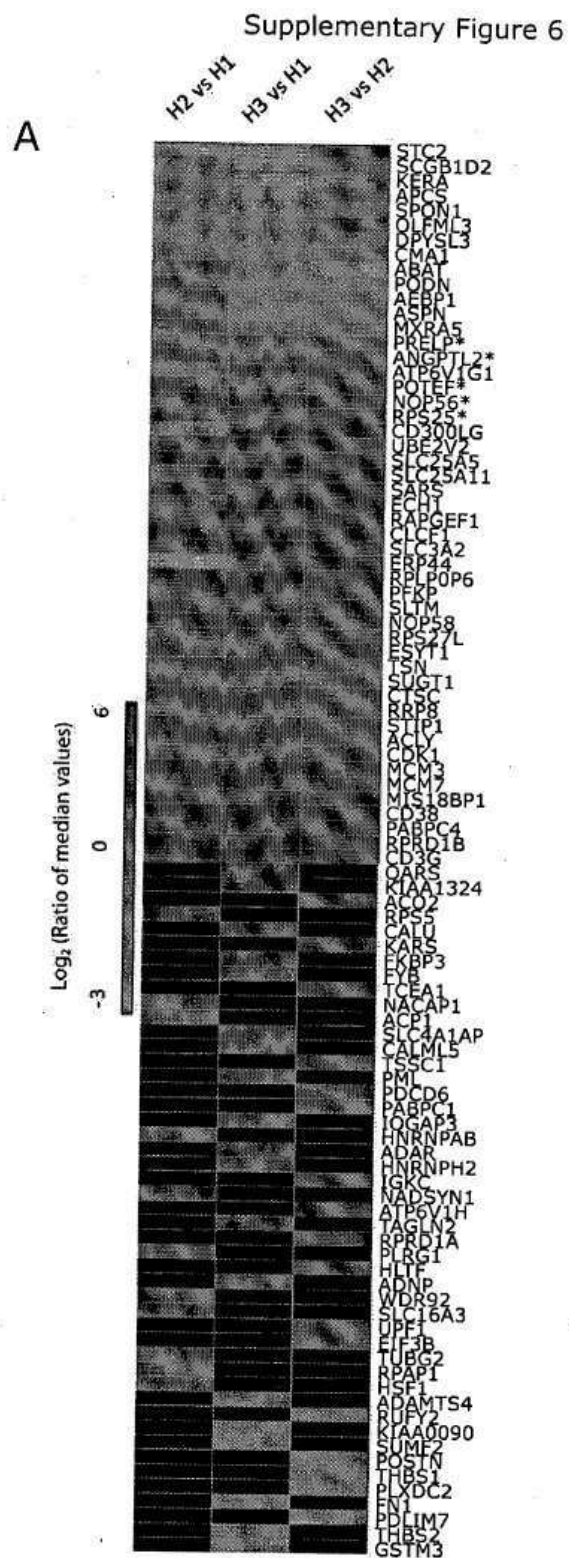


도면16

Supplementary Figure 5



도면17



도면18

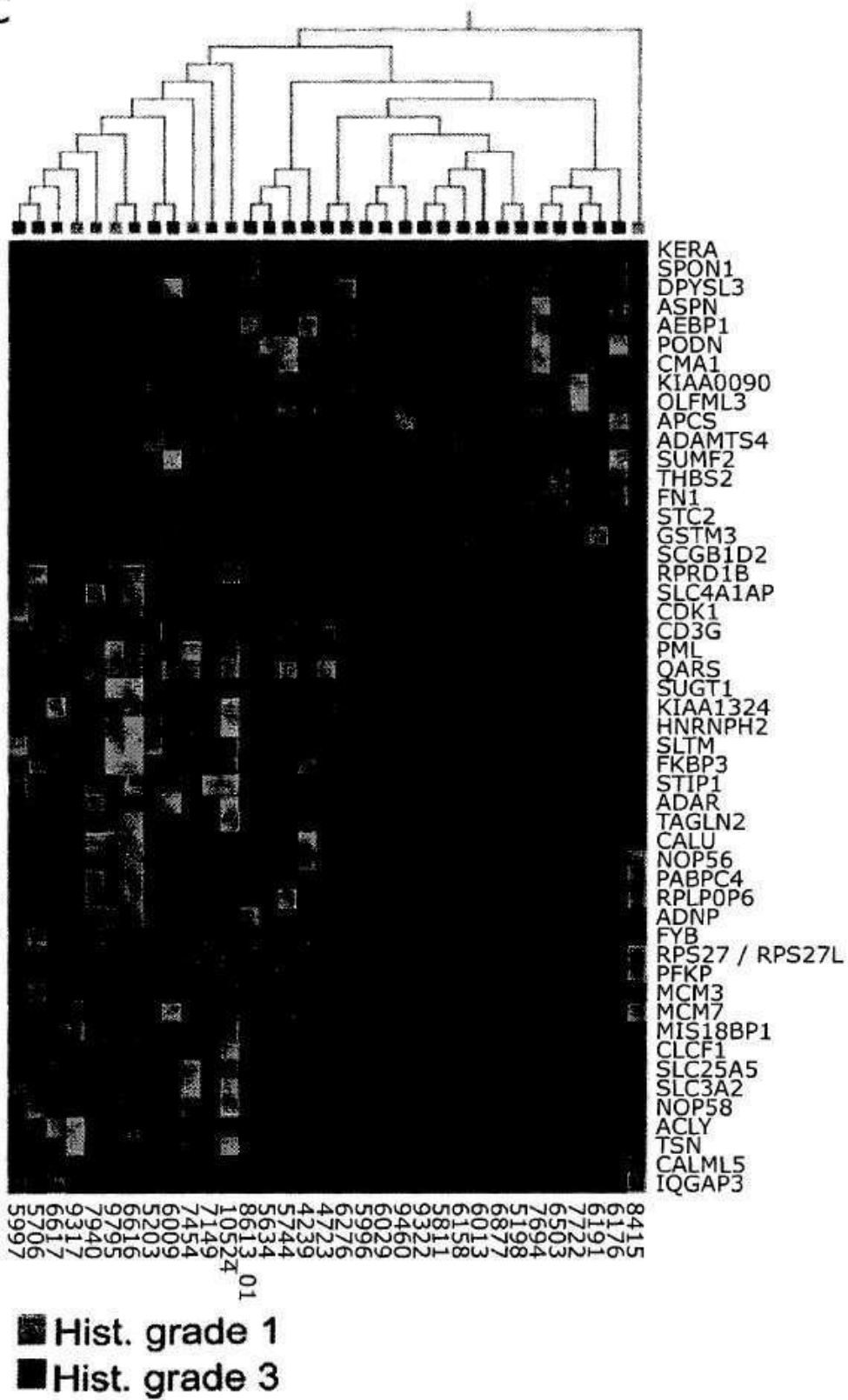
Supplementary Figure 6

B

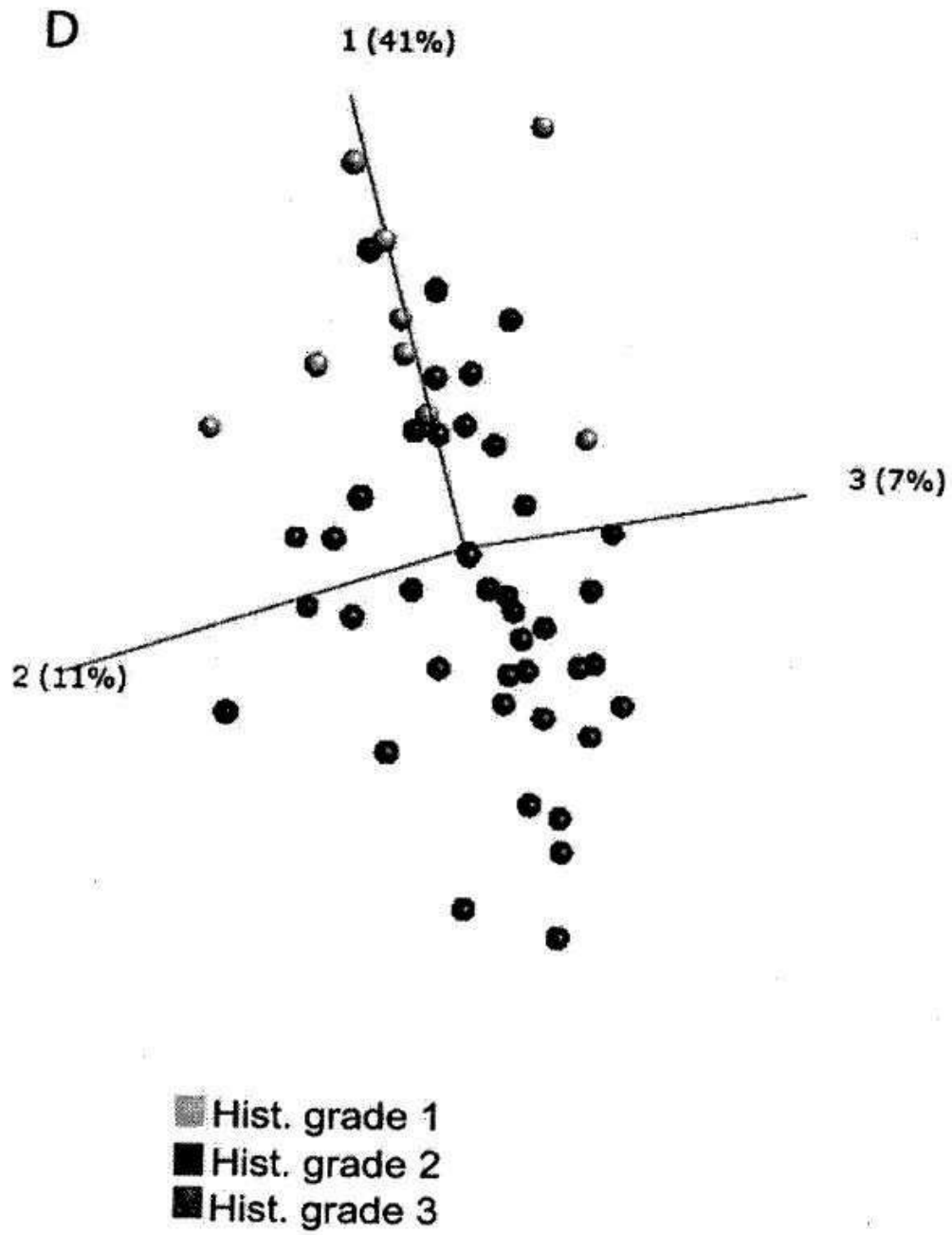
Comparison	ROC area (Unfiltered)	ROC area (Two group comparison)
H1 vs. H2	0.63	0.92
H1 vs. H3	0.71	0.92
H2 vs. H3	0.77	0.91
H1, H2 vs. H3	0.77	0.89
H1 vs. H2, H3	0.64	0.85
H2 vs. H1, H3	0.65	0.75

도면19

C

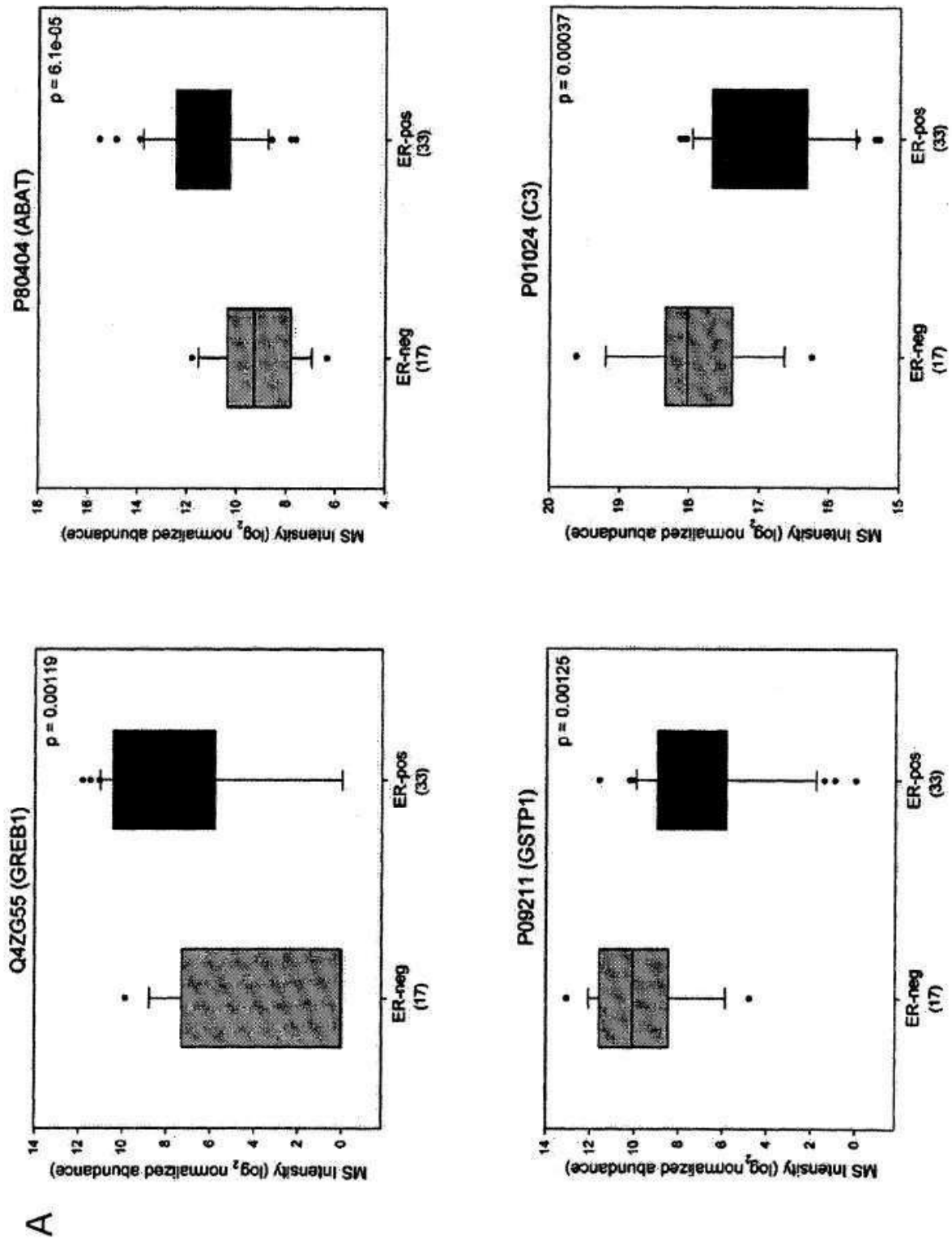


도면20



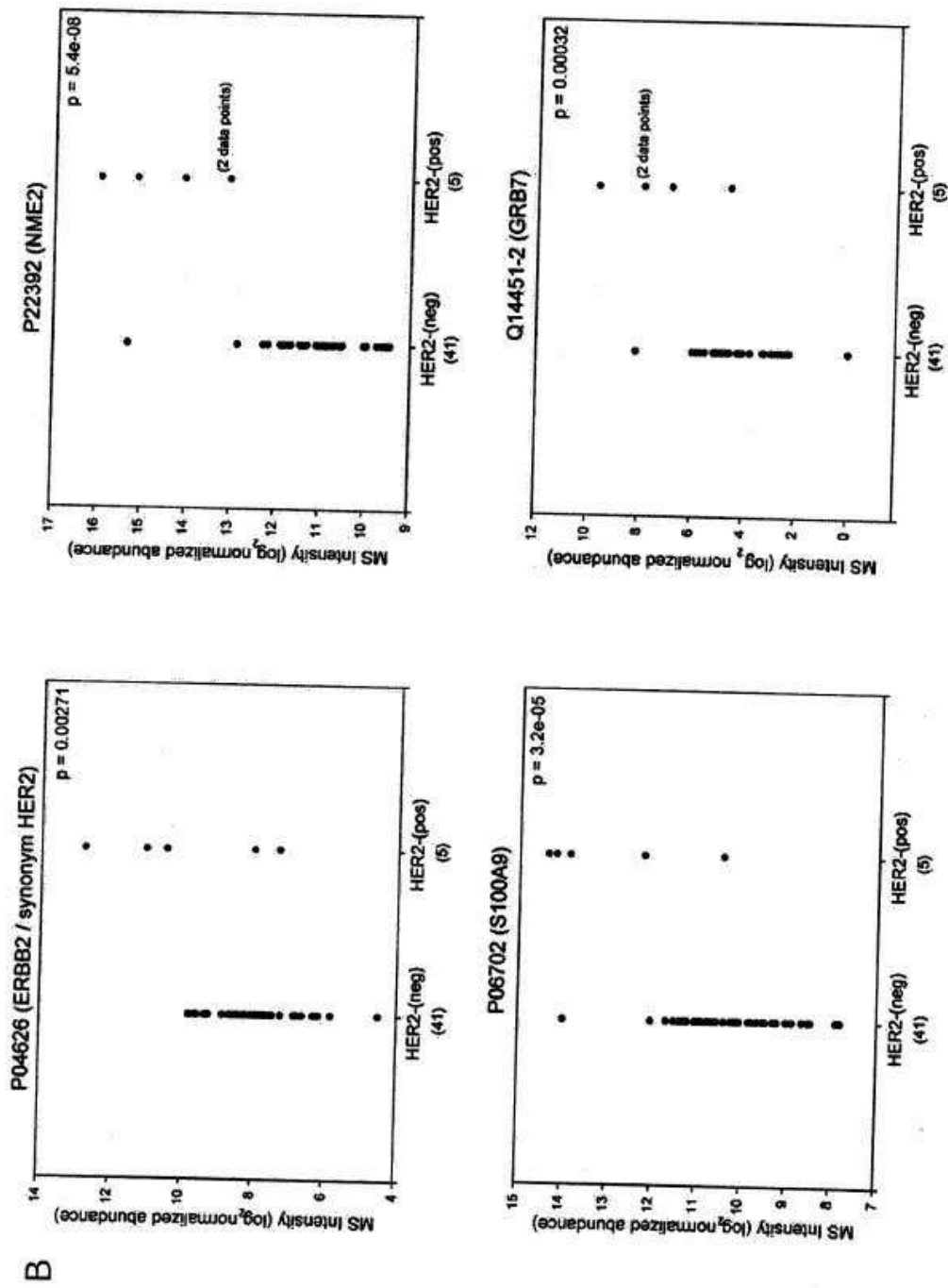
도면21

Supplementary Figure 7



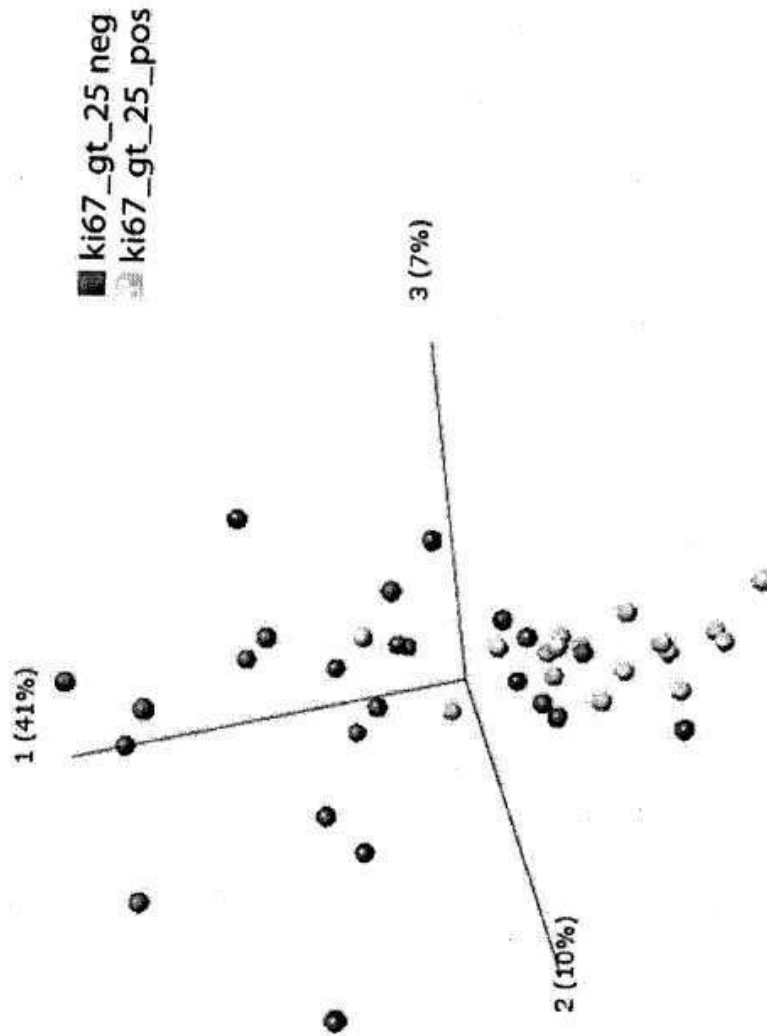
도면22

Supplementary Figure 7



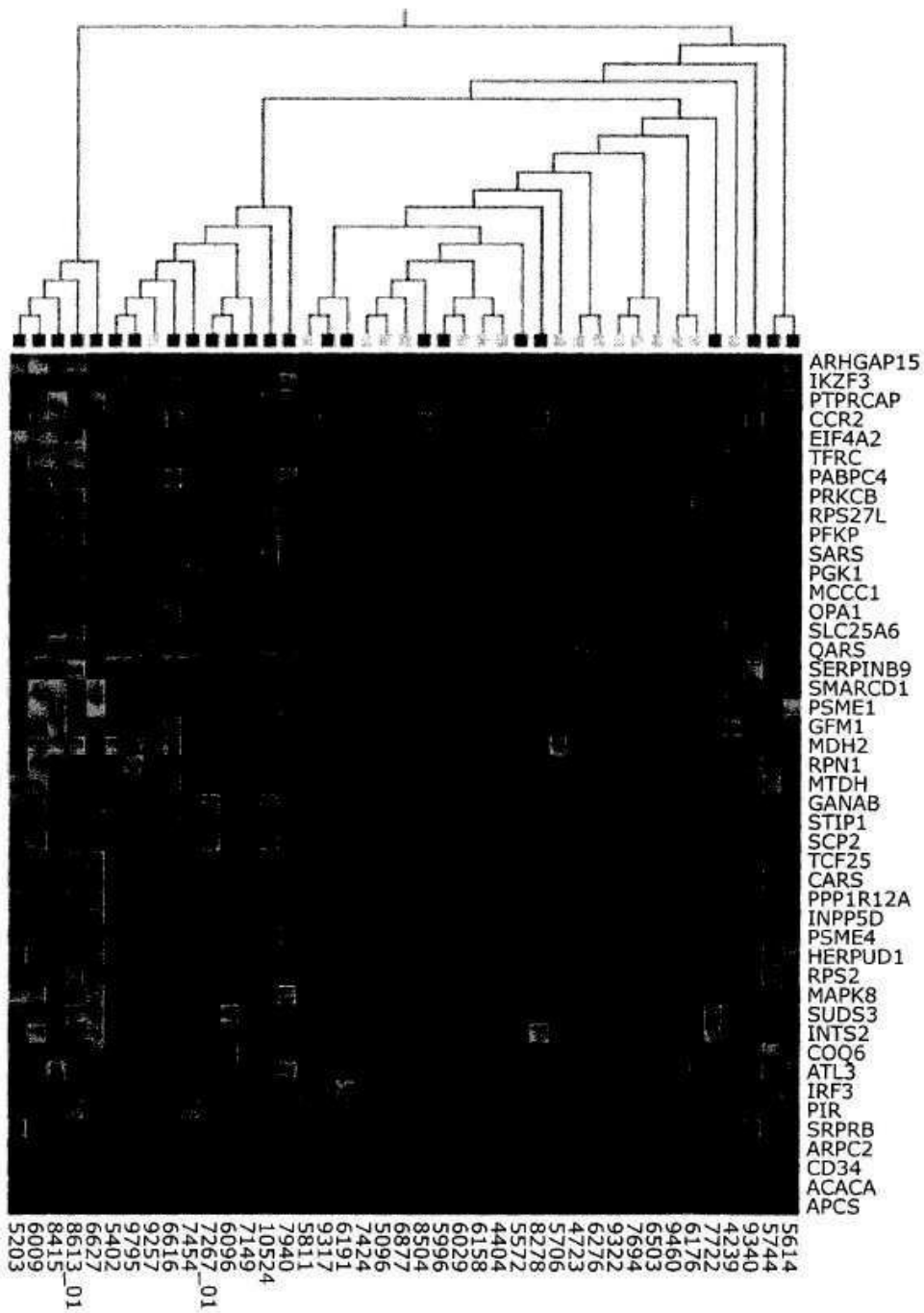
도면23

Supplementary Figure 8A



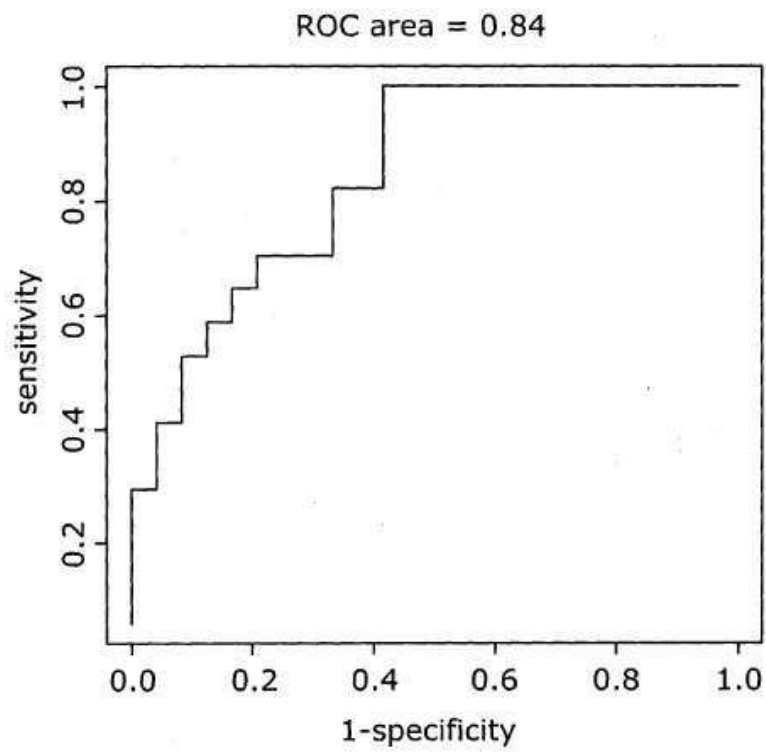
도면24

Supplementary Figure 8A
Continued



도면25

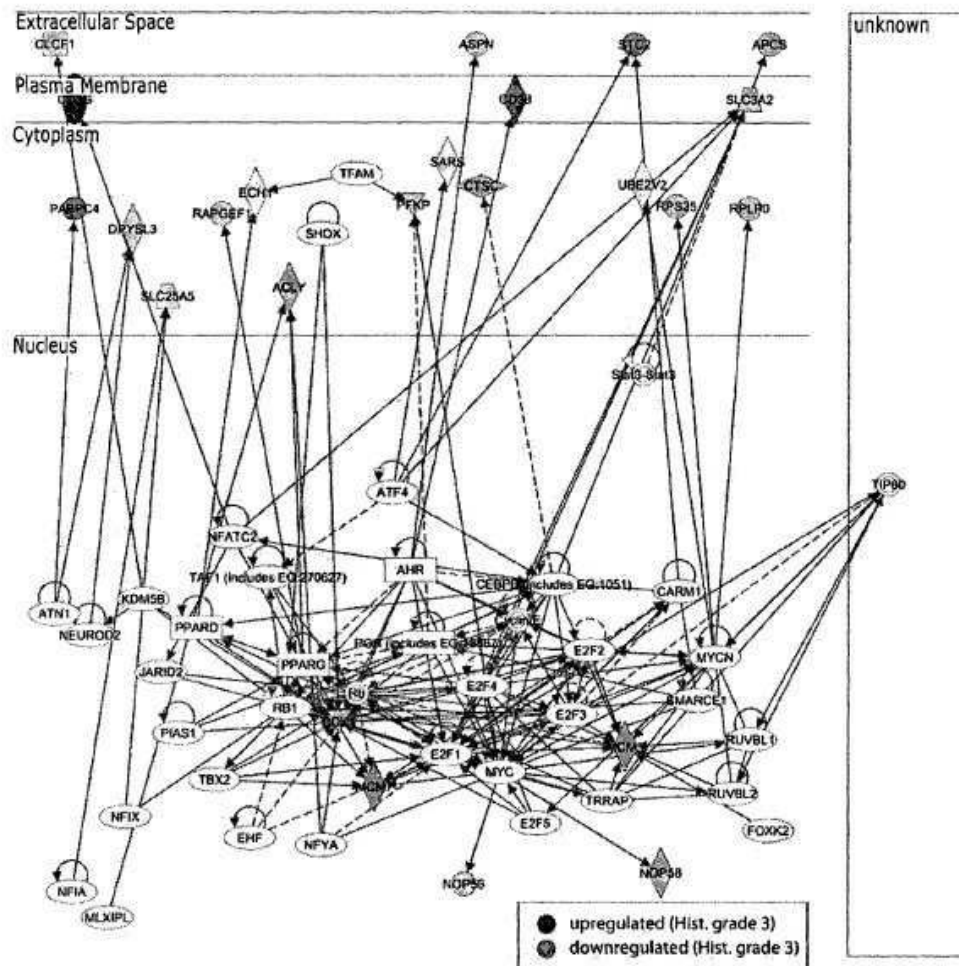
Supplementary Figure 8B



도면26

Supplementary Figure 9(a)

Histologic grade

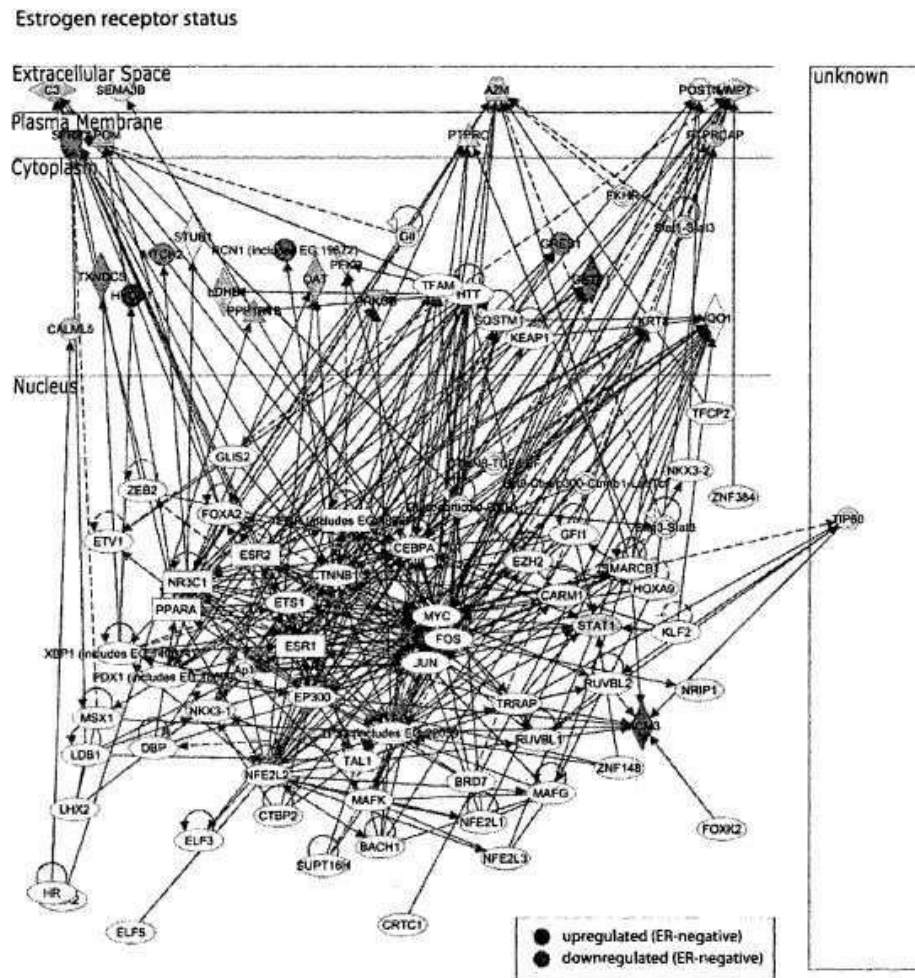


Transcription factor association network

Top Transcription Regulators
Rb
E2F2
E2F3
TFAM
ATF4

도면27

Supplementary Figure 9(b)

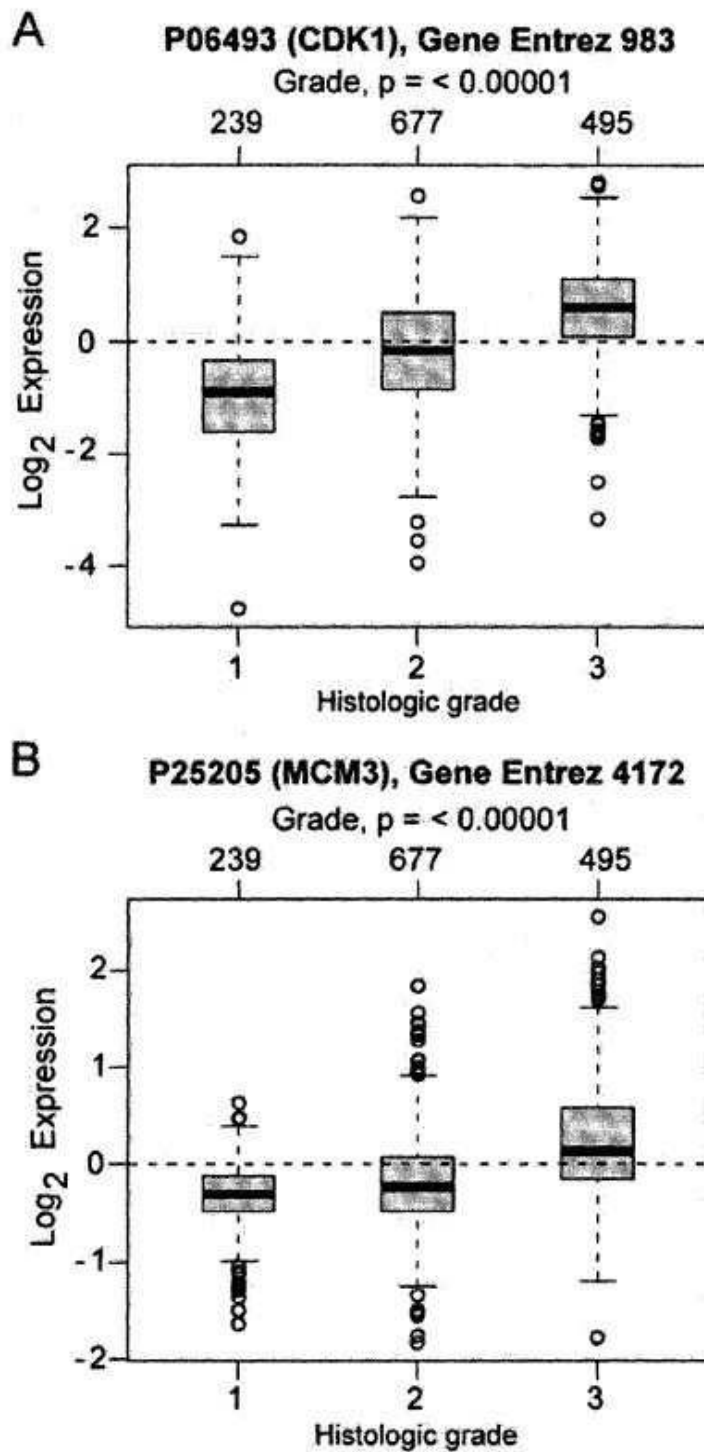


Transcription factor association network

Top Transcription Regulators
ESR2
PGR (includes EG:18667)
MAFK
FOS
JUN

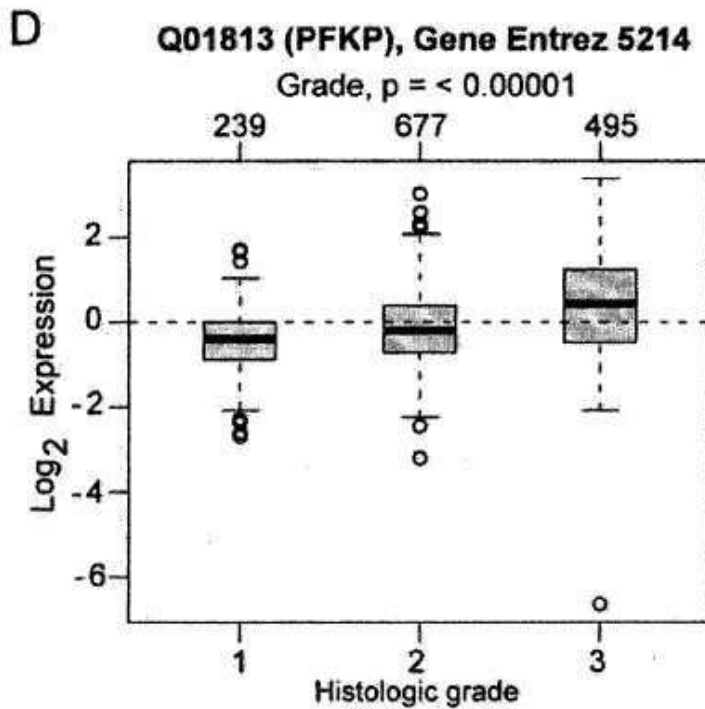
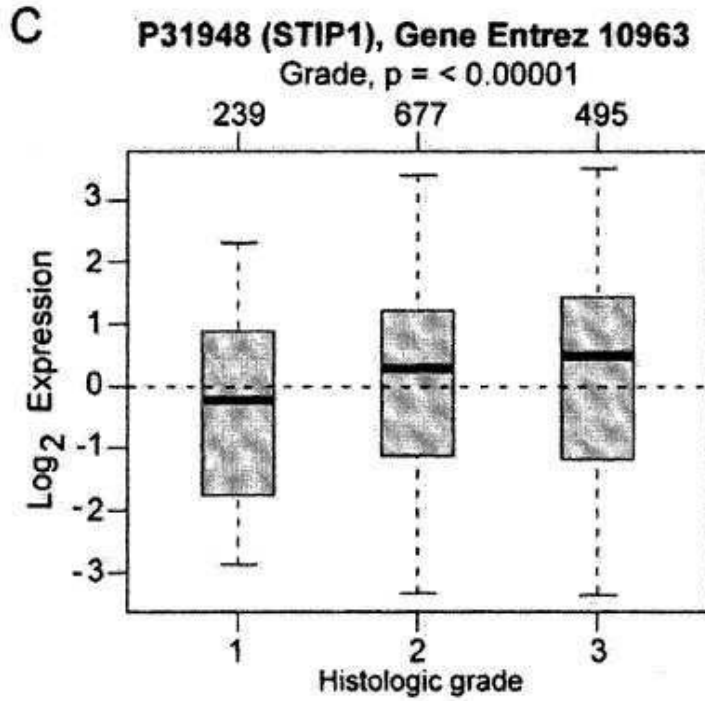
도면28

Supplementary Figure 10



도면29

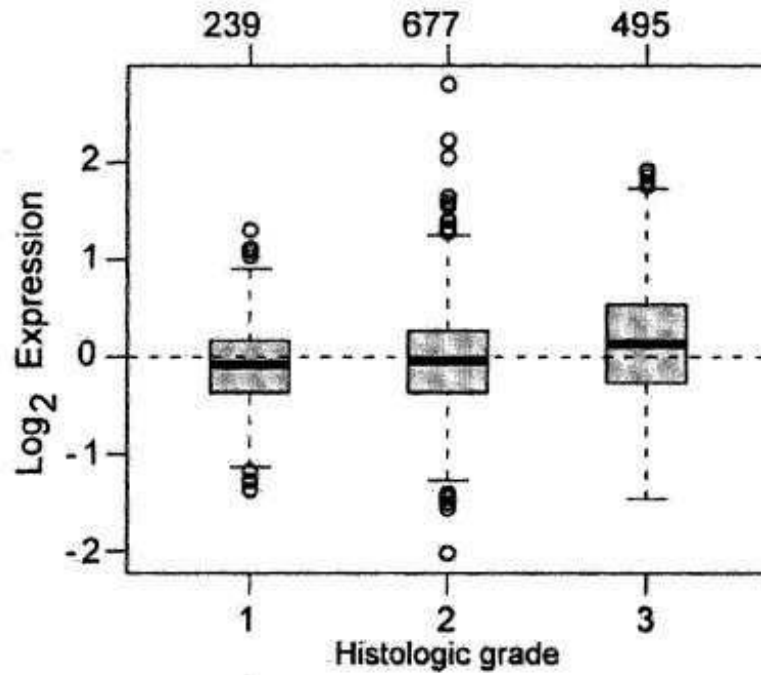
Supplementary Figure 10
Continued



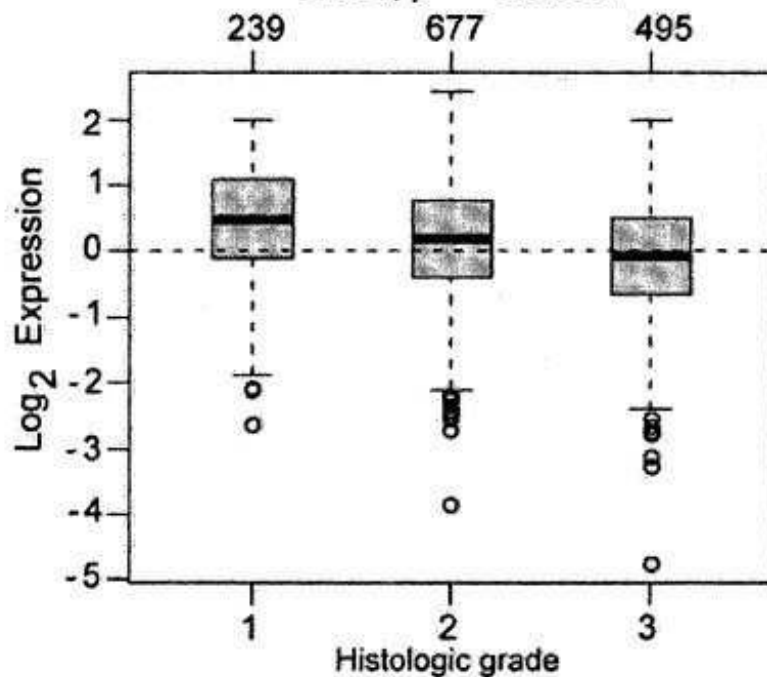
도면30

Supplementary Figure 10
Continued

E **Q13310 (PABPC4), Gene Entrez 8761**
Grade, $p = < 0.00001$

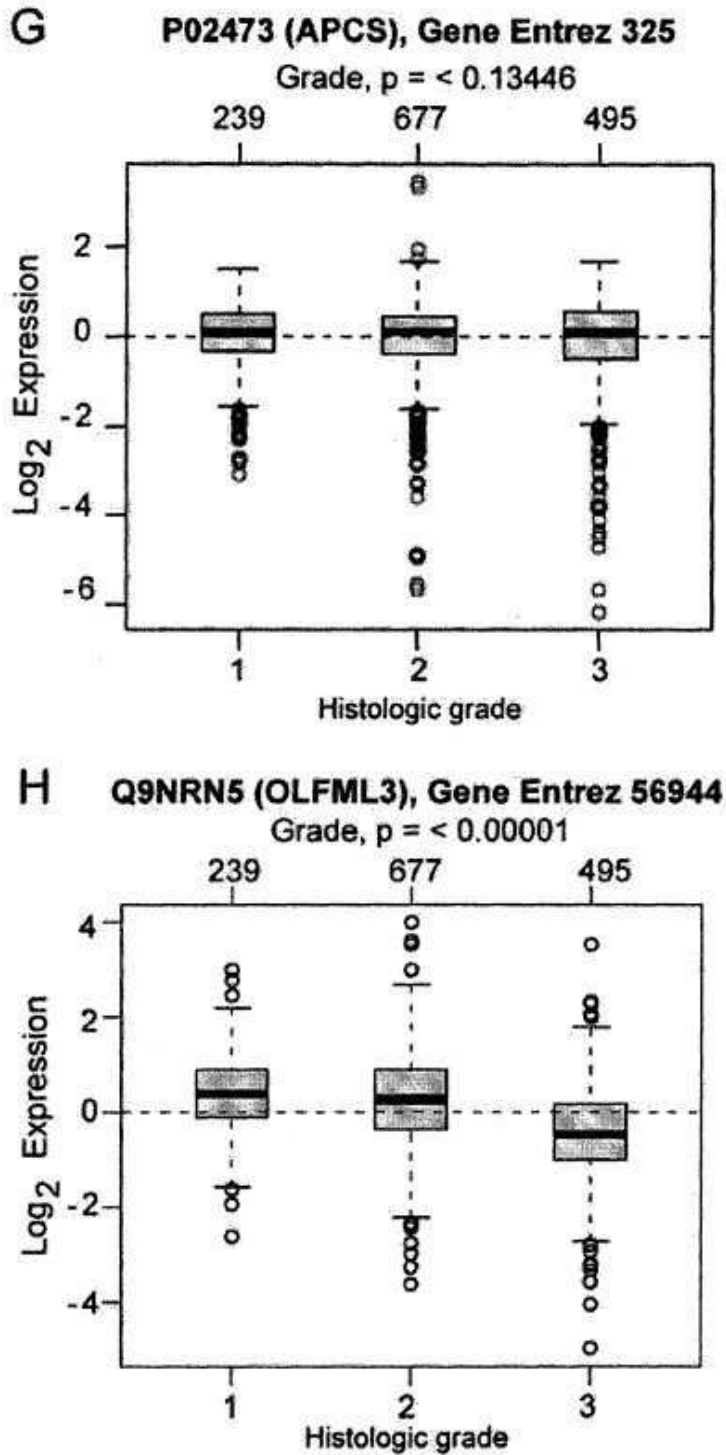


F **Q14195 (DPYSL3), Gene Entrez 1809**
Grade, $p = < 0.00001$



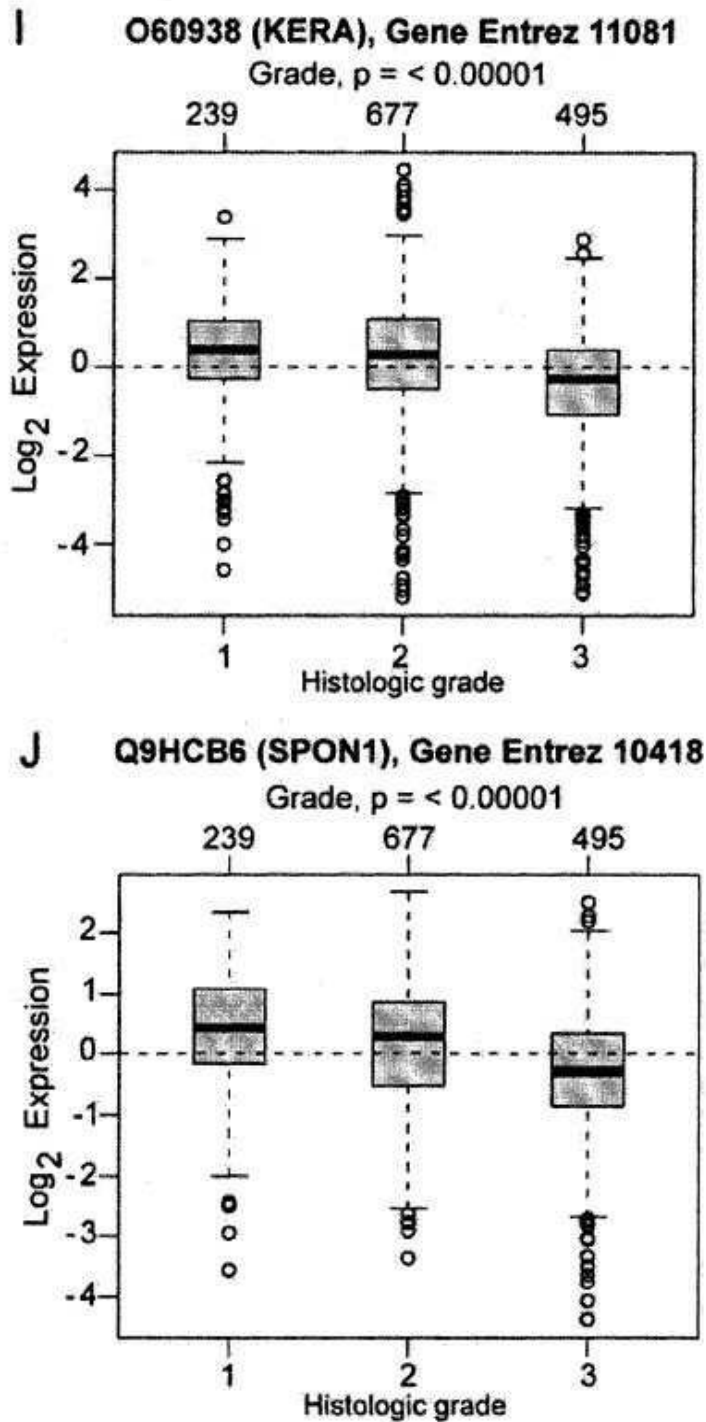
도면31

Supplementary Figure 10
Continued



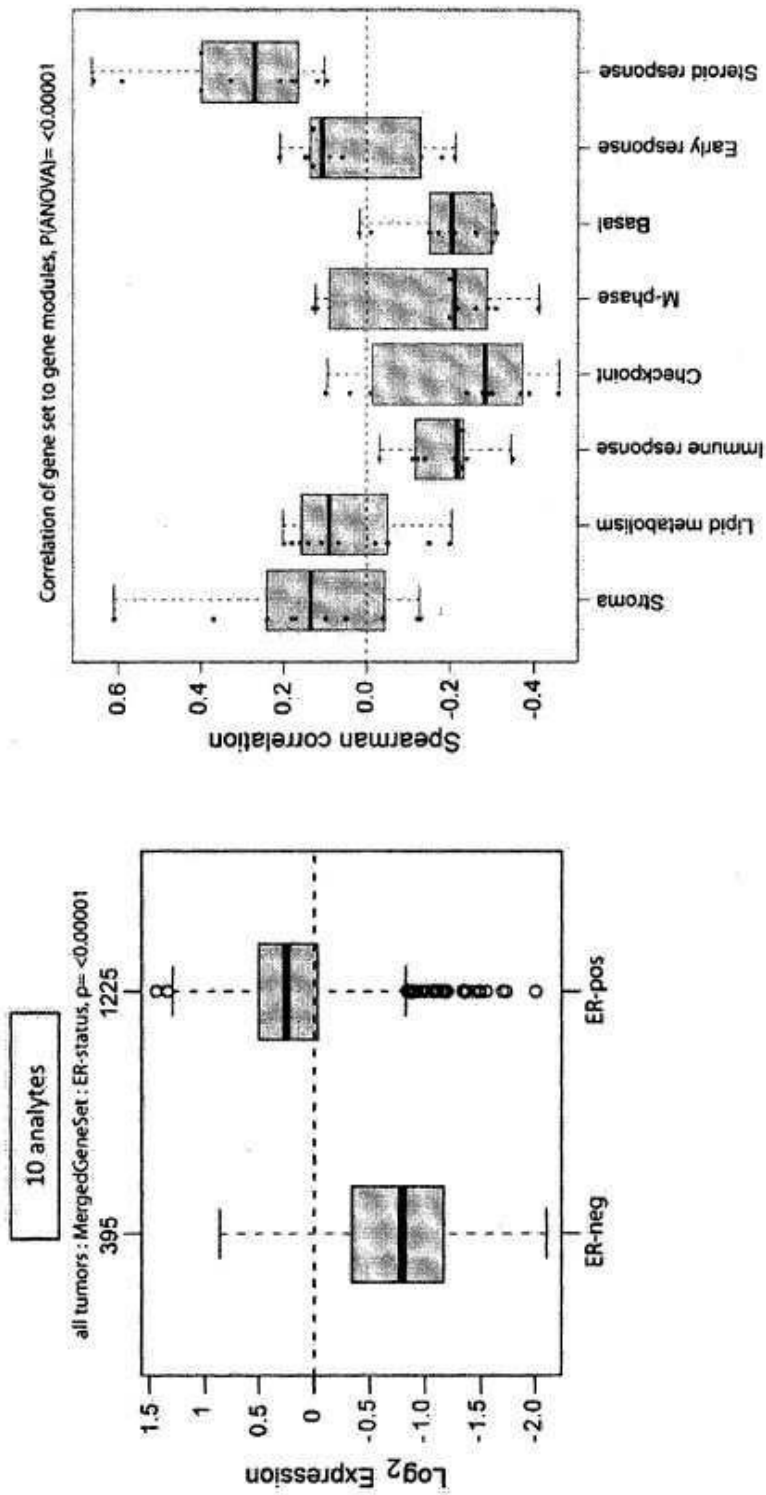
도면32

Supplementary Figure 10
Continued

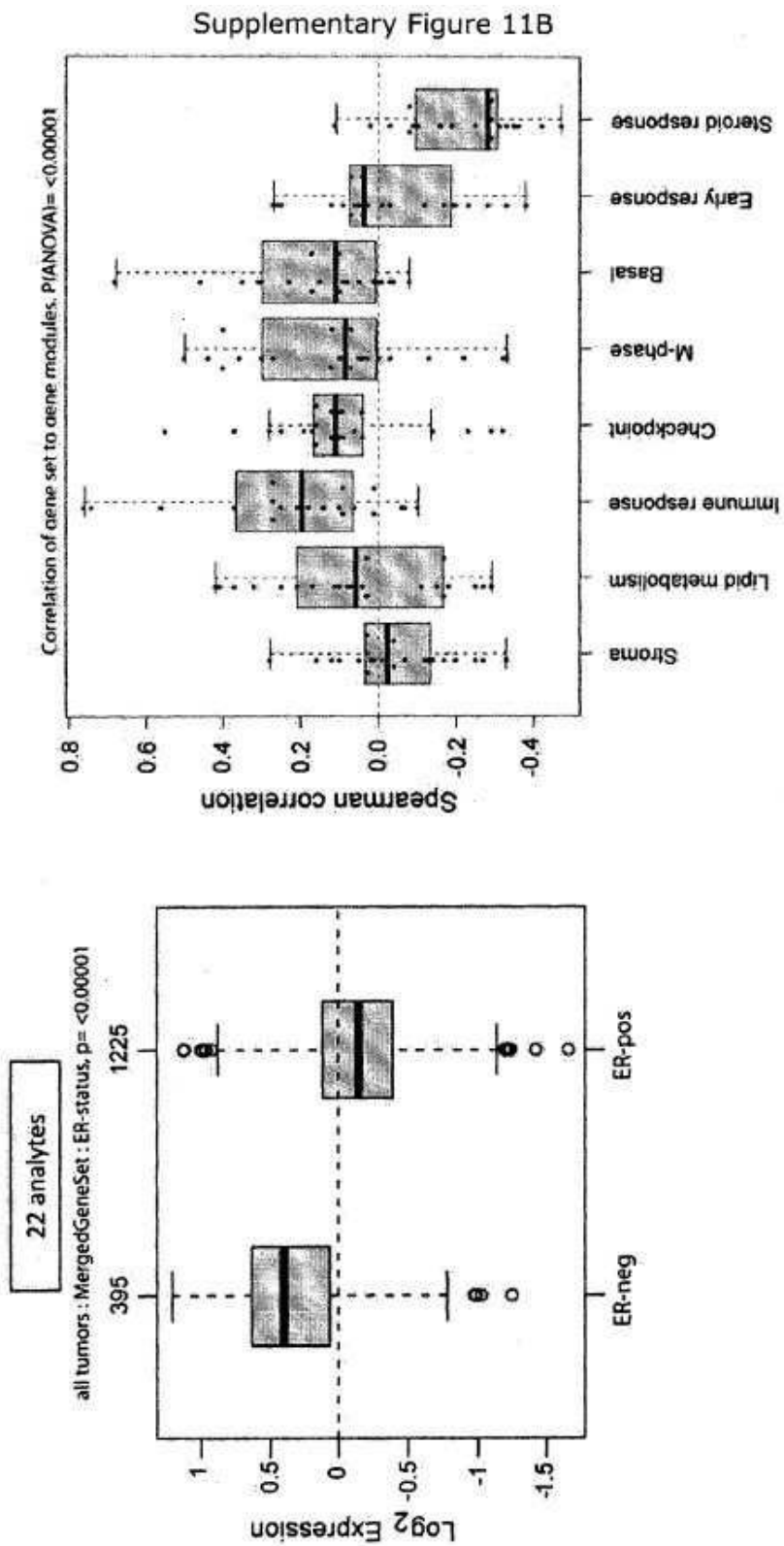


도면33

Supplementary Figure 11A

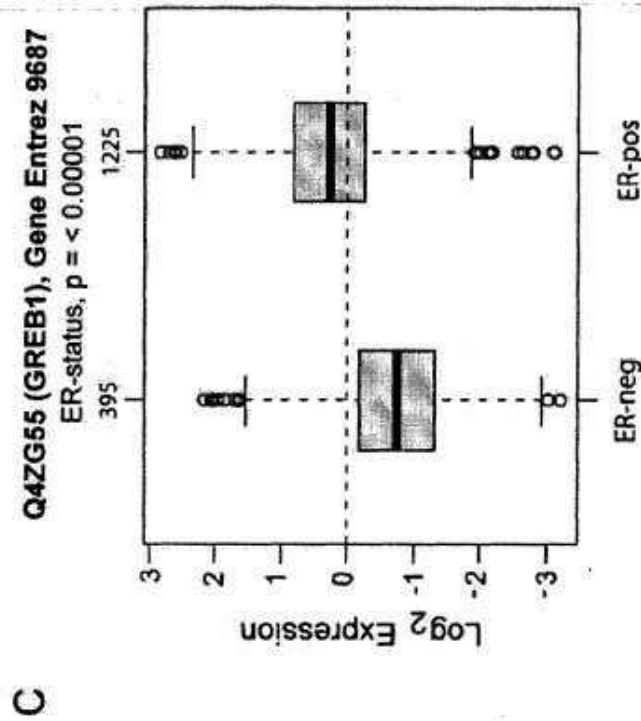
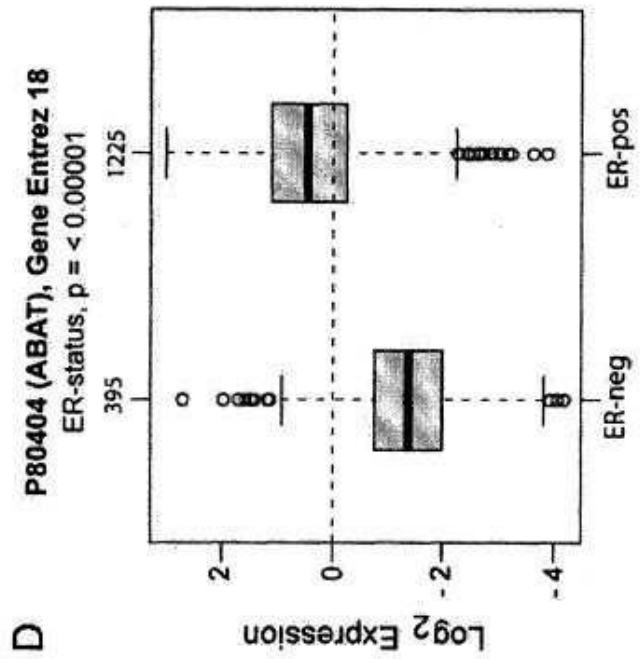


도면34

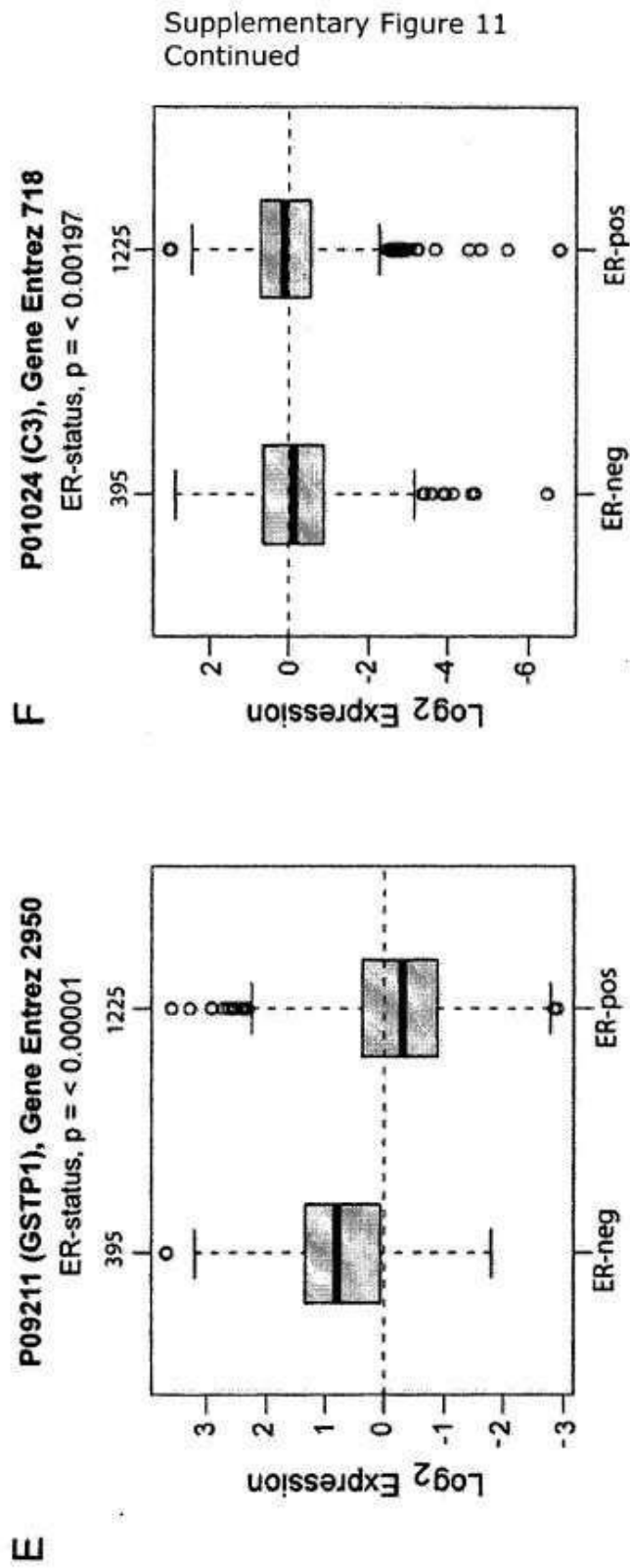


도면35

Supplementary Figure 11
Continued

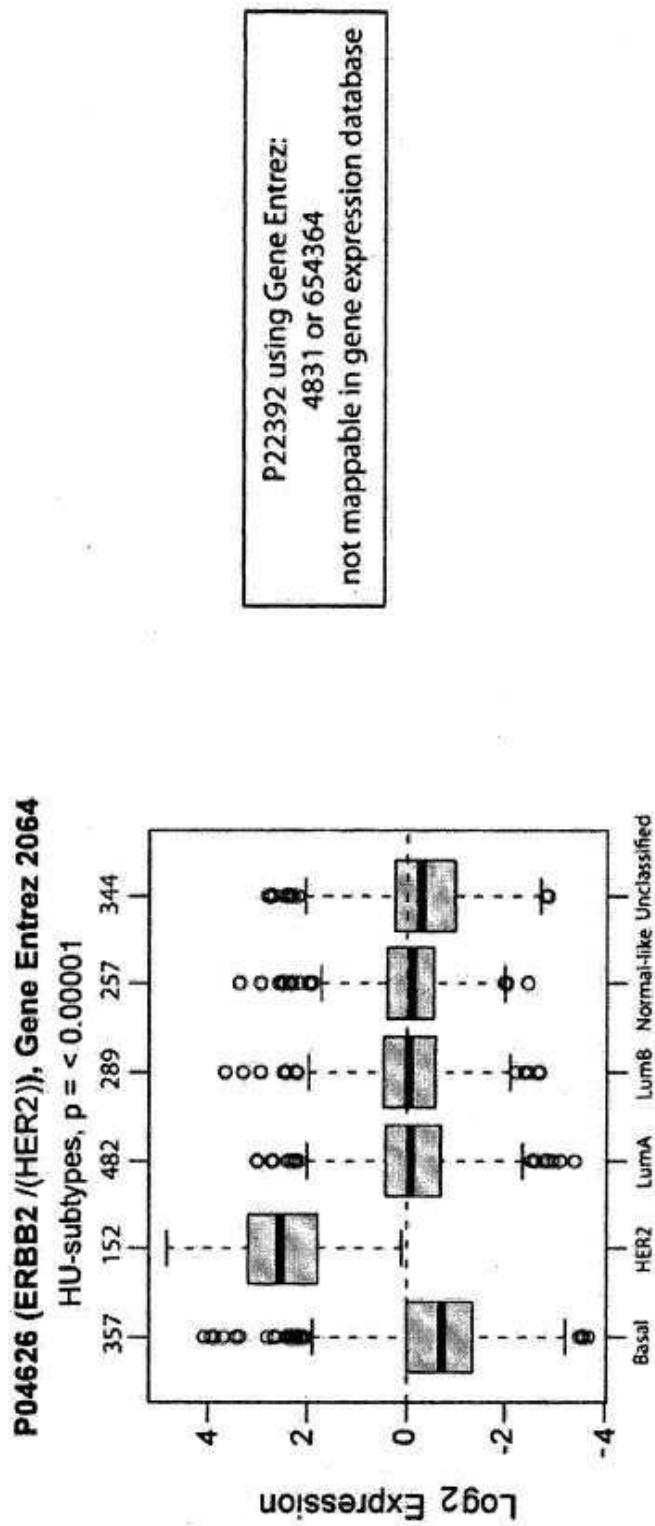


도면36



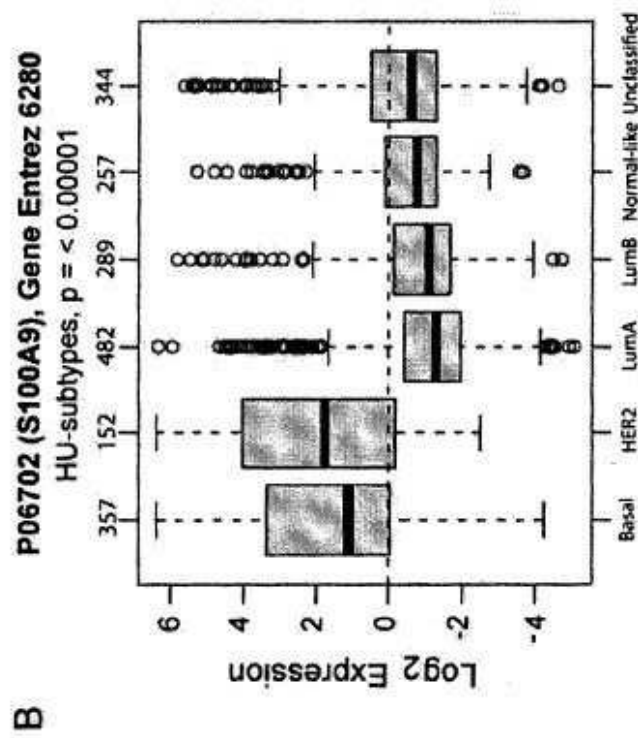
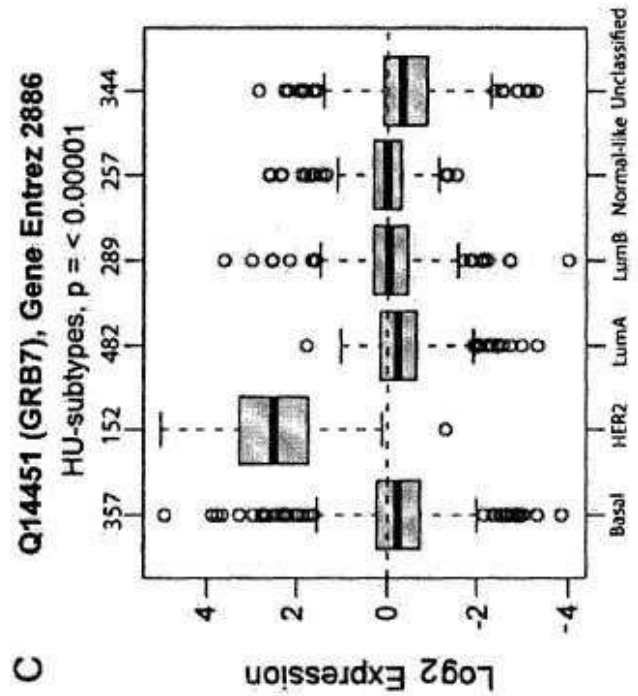
도면37

Supplementary Figure 12A



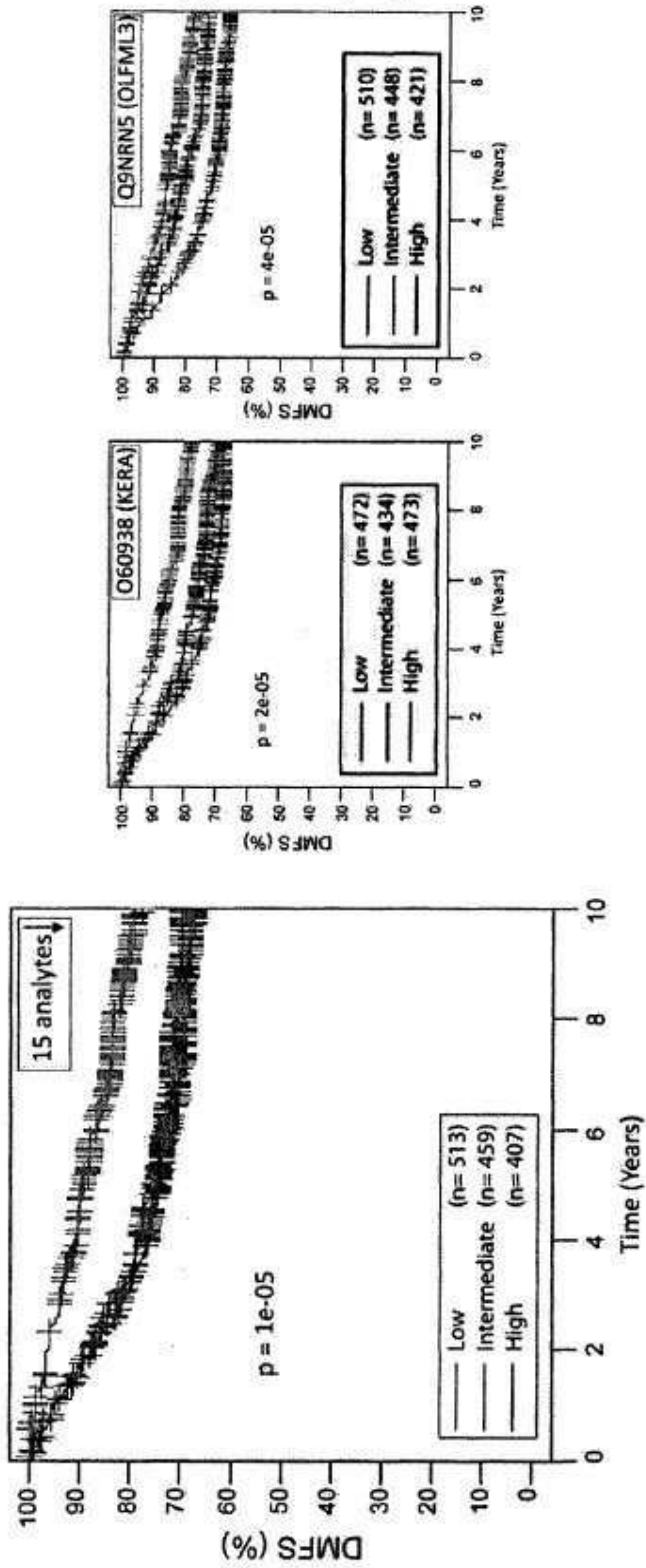
도면38

Supplementary Figure 12
Continued



도면39

Supplementary Figure 13



도면40

Supplementary Figure 13 Continued

