



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0099928
(43) 공개일자 2019년08월28일

- (51) 국제특허분류(Int. Cl.)
C12Q 1/6886 (2018.01) G01N 33/574 (2006.01)
G16B 20/00 (2019.01)
- (52) CPC특허분류
C12Q 1/6886 (2018.05)
G01N 33/57407 (2013.01)
- (21) 출원번호 10-2018-0020009
(22) 출원일자 2018년02월20일
심사청구일자 2018년02월20일

- (71) 출원인
서울대학교병원
서울특별시 종로구 대학로 101(연건동)
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)
- (72) 발명자
서윤석
서울특별시 종로구 대학로 101, 13619호(연건동)
나득채
서울특별시 강남구 현릉로590길 11, 306동 702호
(세곡동, 세곡리엔파크3단지)
(뒷면에 계속)
- (74) 대리인
리앤목특허법인

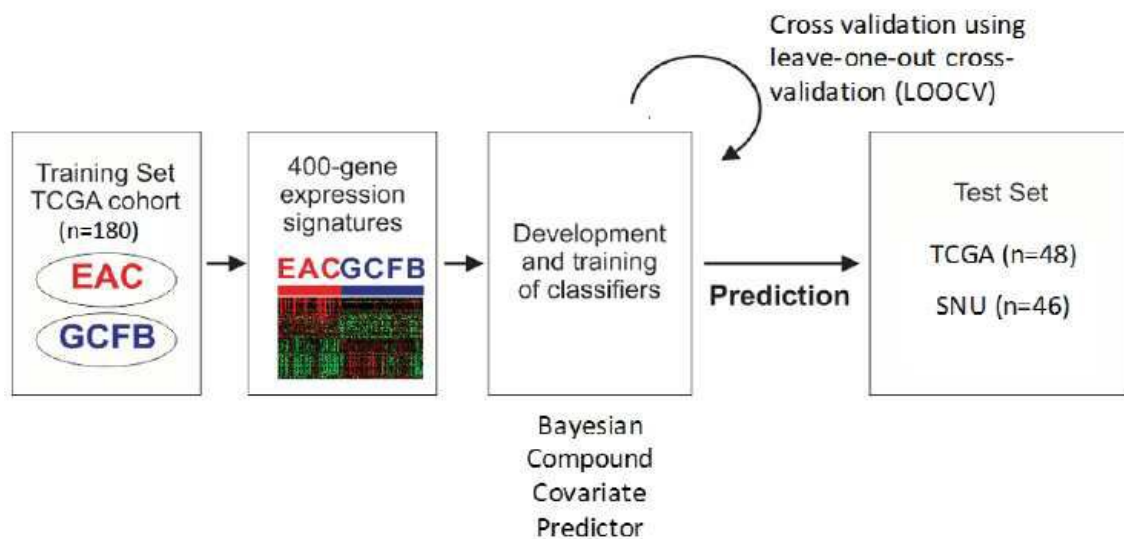
전체 청구항 수 : 총 27 항

(54) 발명의 명칭 위식도경계부선암의 진단 및 표적 치료를 위한 마커

(57) 요약

위식도경계부암을 식도암 또는 위저부/체부암으로 진단하기 위한 알고리즘, 마커 및 치료 표적을 제공한다. 상기 마커는 ERBB2 또는 EGFR을 포함한다.

대표도 - 도3



(52) CPC특허분류

G01N 33/57446 (2019.01)

G16B 20/00 (2019.02)

C12Q 2600/158 (2013.01)

G01N 2333/71 (2013.01)

G01N 2333/912 (2013.01)

(72) 발명자

이장철

미국 코네티컷주 06085 유니온빌 헤드릭슨 Ln 17

양한광

서울특별시 종로구 대학로 101, 5302호 (연건동)

이 발명을 지원한 국가연구개발사업

과제고유번호 HI14C1277

부처명 보건복지부

연구관리전문기관 한국보건산업진흥원

연구사업명 연구중심병원 육성 R&D

연구과제명 맞춤형 암-만성염증 극복을 위한 개방형 연구 비즈니스 플랫폼 구축

기 여 율 68/100

주관기관 서울대학교병원

연구기간 2014.10.01 ~ 2023.03.31

이 발명을 지원한 국가연구개발사업

과제고유번호 HI13C2148

부처명 보건복지부

연구관리전문기관 한국보건산업진흥원

연구사업명 보건의료기술연구개발사업-글로벌화장품신소재사업

연구과제명 환자맞춤형PDX동물모델확립

기 여 율 32/100

주관기관 이화여자대학교 산학협력단

연구기간 2017.12.01 ~ 2018.11.30

명세서

청구범위

청구항 1

위암 또는 식도암의 진단에 필요한 정보를 제공하기 위하여,
개체로부터 분리된 위 또는 식도 조직의 EGFR 또는 ERBB2의 발현량을 측정하는 단계를 포함하는 방법.

청구항 2

청구항 1에 있어서, 상기 위암 또는 식도암은 위식도경계부선암(adenocarcinoma of gastroesophageal junction), 분문부암(cardia cancer), 위식도경계부/분문부암, 상부위암(upperthird gastric cancer), 위저부암(gastric cancer located at fundus), 위체부암(gastric cancer located at body), 위저부/체부암(gastric adenocarcinoma located at fundus or body of the stomach: GCFB), 식도 선암(esophageal adenocarcinoma: EAC), 또는 이들의 조합인 방법.

청구항 3

청구항 1에 있어서, 상기 EGFR 또는 ERBB2의 발현량을 측정하는 단계는 EGFR 또는 ERBB2의 복제 수(copy number)를 확인하는 단계를 포함하는 것인 방법.

청구항 4

청구항 1에 있어서, 상기 EGFR 또는 ERBB2의 발현량을 측정하는 단계는 HER1 또는 HER2의 단백질의 양을 측정하는 단계를 포함하는 것인 방법.

청구항 5

청구항 1에 있어서, 상기 EGFR 또는 ERBB2의 발현량을 측정하는 단계는 역상 단백질 분석(reverse phase protein analysis), 조직 마이크로어레이(tissue microarray), 면역조직화학(immunohistochemistry: IHC), 은 제자리 부합법(silver in situ hybridization: SISH), 또는 이들의 조합을 이용하여 수행하는 것인 방법.

청구항 6

청구항 1에 있어서, 상기 EGFR 또는 ERBB2의 발현량이 증가한 경우, 식도암으로 판정하는 단계를 더 포함하는 방법.

청구항 7

청구항 6에 있어서, 상기 EGFR 또는 ERBB2의 발현량이 증가한 경우는 ERBB2 유전자의 복제 수가 증가한 경우인 것인 방법.

청구항 8

청구항 1에 있어서, 상기 EGFR 또는 ERBB2의 발현량이 증가하지 않은 경우, 위저부/체부암으로 판정하는 단계를 더 포함하는 방법.

청구항 9

청구항 1에 있어서, 상기 개체로부터 분리된 위 또는 식도 조직의 BCCP (Bayesian Compound Covariate Predictor) 점수를 측정하는 단계를 더 포함하는 방법.

청구항 10

청구항 9에 있어서, 상기 BCCP 점수가 0.4535 이상인 경우 식도암으로 판정하는 단계를 더 포함하는 방법.

청구항 11

청구항 9에 있어서, 상기 BCCP 점수가 0.4535 미만인 경우 위저부/체부암으로 판정하는 단계를 더 포함하는 방법.

청구항 12

청구항 1에 있어서, 상기 EGFR 또는 ERBB2의 발현량이 증가한 경우, EGFR 또는 ERBB2의 저해제를 투여하기로 결정하는 단계를 더 포함하는 방법.

청구항 13

청구항 12에 있어서, 상기 EGFR 또는 ERBB2의 저해제는 라파티닙(lapatinib), 트라스투주맙(Trastuzumab), 세특시맙(cetuximab), 제피티닙(gefitinib), 에피티닙(efitinib) 또는 이들의 조합을 포함하는 것인 방법.

청구항 14

청구항 1에 있어서, 상기 위암 또는 식도암은 선암(adenocarcinoma)인 방법.

청구항 15

청구항 1에 있어서, 상기 EGFR 또는 ERBB2의 발현량을 측정하는 단계는 RT-PCR, RNase 보호 분석법(RNase protection assay: RPA), 노던 블롯팅, 및 DNA 칩으로부터 선택된 하나 이상의 방법으로 수행하는 것인 방법.

청구항 16

청구항 1에 있어서, 상기 EGFR 또는 ERBB2의 발현량을 측정하는 단계는 웨스턴 블롯팅, ELISA, 방사선면역분석, 방사 면역 확산법, 오우크테로니 면역 확산법, 로케트면역전기영동, 면역침전 분석법, 보체 고정 분석법, FACS, 또는 단백질 칩으로부터 선택된 하나 이상의 방법으로 수행하는 것인 방법.

청구항 17

EGFR 또는 ERBB2 단백질, 또는 이들 각각의 단편에 특이적으로 결합하는 항체, 항원 결합 단편, 또는 폴리펩티드, 또는 EGFR 또는 ERBB2 단백질을 코딩하는 폴리뉴클레오티드 서열에 특이적으로 결합하는 프로브, 프라이머 세트, 또는 뉴클레오티드를 포함하는 개체의 위암 또는 식도암을 진단하기 위한 조성물.

청구항 18

청구항 17에 있어서, 상기 위암 또는 식도암은 위식도경계부선암(adenocarcinoma of gastroesophageal junction), 분문부암(cardia cancer), 위식도경계부/분문부암, 상부위암(upperthird gastric cancer), 위저부암(gastric cancer located at fundus), 위체부암(gastric cancer located at body), 위저부/체부암(gastric adenocarcinoma located at fundus or body of the stomach: GCFB), 식도 선암(esophageal adenocarcinoma: EAC), 또는 이들의 조합인 조성물.

청구항 19

청구항 18에 있어서, 상기 조성물은 개체의 식도암 또는 위저부/체부암을 진단하기 위한 것인 조성물.

청구항 20

청구항 17의 조성물, 시약 및 포장 단위를 포함하는 위암 또는 식도암을 진단하기 위한 키트.

청구항 21

개체의 위암 또는 식도암을 진단하기 위한 정보를 제공하기 위하여,

개체로부터 분리된 조직의 유전자의 발현 프로필을 측정하는 단계; 및

상기 유전자 발현 프로필로부터 BCCP (Bayesian Compound Covariate Predictor) 알고리즘을 실행하는 컴퓨터 시스템을 이용하여 BCCP 점수를 얻는 단계를 포함하는 유전자 발현 프로필 분석 방법으로서, 상기 유전자는 아래의 유전자로 구성된 것인 방법: KRT5, KRT13, KRT6A, KRT14, RHCG, KRT4, S100A7, SPRR3, KRT6C, SPRR2A,

MUC21, KRT6B, PKP1, SERPINB3, CLCA2, SPRR1B, MUC4, TBX5, CRNN, SPRR2D, DUOX2, GJB6, SPRR2E, S100A2, KRT16, KRT17, SCEL, DSC3, CALML3, HEPHL1, A2ML1, NKX6-1, DUOX2, SBSN, TMPRSS11D, CLCA4, ANXA8, FGFBP1, TGM1, IVL, DUOX1, DSG3, EREG, SERPINB4, KRT78, HOXC10, DUOX1, ZNF750, SPRR1A, KRT7, TMPRSS11A, CXCL6, KLK13, CRCT1, PAX9, HORMAD1, FAM83A, S100A7A, GPR110, CAPN14, ABCA12, WFDC2, TMPRSS11B, CEACAM5, MUC12, SERPINB13, PADI1, FAT2, IL1RL2, STK31, SPINK5, DDX3Y, PI3, FAM83C, LYPD2, KLK5, IL1F5, KRT15, BBOX1, GPR87, ACTL8, SLC34A2, USP9Y, TGM3, LASS3, TMPRSS13, GUCY1B2, SPRR2F, KDM5D, KRT23, LY6D, VTCN1, SLC5A12, ABCA13, S100A8, GBP6, CEACAM7, IL1F9, MUC15, PPP1R14C, KLK7, DEFB1, MMP3, TM4SF20, SYCP2, TMEM40, CEACAM6, VWA2, CYorf15B, TDRD5, DCDC2, KLK12, C10orf99, CARD14, HOXC13, LOC642587, GABRR1, DNAH2, SLC15A1, KLK8, LGALS7B, ALDH1L1, FOXN1, RAET1L, TP63, ATP13A4, UTY, DSG1, ALOX12B, AMY2B, SAA2, SAA1, GPR115, LCN2, KRT24, PRKY, PADI3, SCNN1A, PGLYRP3, PRSS27, ANXA8L2, WISP3, IL20RB, FOXE1, ZFY, ALG1L, LY6K, NCCRP1, SLC44A5, NMU, CDH17, RNASE7, FER1L4, C4BPB, CALML5, CXorf61, SPACA4, ULBP2, KRTDAP, MYO7B, , C7orf54, SLC6A20, LOC221442, GOLGA8B, GPR81, ALDH3B2, TRIM29, C21orf125, BNC1, IL1F6, TRPA1, C12orf27, LOC440905, POU5F1, LOC100131726, KCNH8, SPRR2C, IL13RA2, C3orf67, SAA4, LYPD3, GJB5, ULBP1, PHACTR3, S100A9, TRIM54, KLK11, ITIH5L, BNIPL, LOC284233, LOC387646, MRPL42P5, FLJ42393, PCDHAC2, IFNE, BCL2L10, FLJ45445, COL7A1, GDA, ARL14, VIP, CLC, PPP2R2B, FABP4, SFRP5, CCL8, CD79A, HIST1H1E, CCL4L2, SYNPO2, SLC47A1, CXCR2P1, C1QTNF7, CDH2, SIGLEC7, GZMA, NKX6-3, PGM5, LILRB4, NMUR1, XCL2, SOX10, FCGR1A, ST6GALNAC5, CCL4, ACSM5, FOLR2, KCNA1, CCL26, FCGR1C, VSIG4, C1QA, NCF1B, NXPH3, C1QC, TAGLN, PTPRN, TYROBP, ISL1, HCST, RGMA, APOBEC3H, PRND, ADIPOQ, CCL11, GLP2R, PNMA6A, CD8B, FCGR1B, PCBP3, CD8A, GPR27, DPP6, F13A1, EPYC, CXCL9, GREM2, SNORA12, DNAJC5B, HLA-DQA2, PLP1, CHGB, APOC2, FAT3, TLR7, CHRDL1, LY86, AGTR1, ISLR, CR1, FCGR3A, DPEP1, PPAPDC1A, ODZ3, MAPK4, C17orf87, KCNJ5, MSR1, RELN, APOE, C1QB, PHYHIPL, CCDC80, NCF1C, SLITRK5, TREM2, FGF14, PGA5, NKX2-3, PRELP, FCRLA, CCL5, GNAO1, PCDH9, WSCD2, MS4A1, RNF150, KLRD1, KCNK2, MATK, HUNK, IGF1, CDO1, COLEC12, FIBIN, CARTPT, VPREB3, PPP1R1A, GDF6, NCAM2, CLEC10A, KIAA0408, BHLHE22, CD52, SULT4A1, SIT1, FNDC1, GRIK3, SUCNR1, TMEM90B, CYP1B1, MKX, SV2B, BMP3, TCEAL2, CADM3, SPOCK1, GREM1, SIGLEC8, ADCY2, CDC26, OGN, FCRL6, PSD, VIPR2, GZMM, ADH1B, CCL19, PLD4, TMEM189-UBE2V1, NKG7, EOMES, STMN2, GZMH, NPTX1, CNTFR, TMEM130, CTNND2, ITGBL1, ECEL1, ACTG2, COL10A1, ST6GAL2, NRK, PI16, NALCN, GZMK, NCAM1, RMRP, CHRM2, KCNMA1, HSPB7, SLIT2, PDZRN4, CNN1, CHRDL2, OMD, PTGIS, RSPO3, PLA2G2D, SMYD1, ZNF683, NRXN3, PCSK1N, HSPB6, C16orf89, PGA3, COMP, C2orf40, C7, GKN2, DES, CILP, COL11A1, LIPF, GATA5, MFAP5, SCRG1, SFRP2, GKN1, PGA4, CCL21, SIX2, NKX3-2, SFRP4, NBLA00301, THBS4, HAND2, 및 BARX1.

청구항 22

청구항 21에 있어서, 상기 위암 또는 식도암은 위식도경계부선암(adenocarcinoma of gastroesophageal junction), 분문부암(cardia cancer), 위식도경계부/분문부암, 상부위암(upperthird gastric cancer), 위저부암(gastric cancer located at fundus), 위체부암(gastric cancer located at body), 위저부/체부암(gastric adenocarcinoma located at fundus or body of the stomach: GCFB), 식도 선암(esophageal adenocarcinoma: EAC), 또는 이들의 조합인 방법.

청구항 23

청구항 21에 있어서, 상기 BCCP 점수가 0.4535 이상인 경우 식도암으로 판정하는 단계를 더 포함하는 방법.

청구항 24

청구항 21에 있어서, 상기 BCCP 점수가 0.4535 미만인 경우 위저부/체부암으로 판정하는 단계를 더 포함하는 방법.

청구항 25

청구항 21에 있어서, 상기 유전자 발현 프로필을 측정하는 단계는 RT-PCR, RNase 보호 분석법(RNase protection assay: RPA), 노던 블롯팅, 및 DNA 칩으로부터 선택된 하나 이상의 방법으로 수행하는 것인 방법.

청구항 26

개체의 위암 또는 식도암을 진단하기 위한 정보를 제공하기 위하여, 개체로부터 분리된 조직의 유전자의 발현 프로필을 측정하기 위한 물질을 포함하는 개체의 위암 또는 식도암을 진단하기 위한 조성물로서, 상기 유전자는 아래의 유전자를 포함하는 것인 조성물: KRT5, KRT13, KRT6A, KRT14, RHCG, KRT4, S100A7, SPRR3, KRT6C, SPRR2A, MUC21, KRT6B, PKP1, SERPINB3, CLCA2, SPRR1B, MUC4, TBX5, CRNN, SPRR2D, DUOX2, GJB6, SPRR2E, S100A2, KRT16, KRT17, SCEL, DSC3, CALML3, HEPHL1, A2ML1, NKX6-1, DUOX2, SBSN, TMPRSS11D, CLCA4, ANXA8, FGFBP1, TGM1, IVL, DUOX1, DSG3, EREG, SERPINB4, KRT78, HOXC10, DUOX1, ZNF750, SPRR1A, KRT7, TMPRSS11A, CXCL6, KLK13, CRCT1, PAX9, HORMAD1, FAM83A, S100A7A, GPR110, CAPN14, ABCA12, WFDC2, TMPRSS11B, CEACAM5, MUC12, SERPINB13, PADI1, FAT2, IL1RL2, STK31, SPINK5, DDX3Y, PI3, FAM83C, LYPD2, KLK5, IL1F5, KRT15, BBOX1, GPR87, ACTL8, SLC34A2, USP9Y, TGM3, LASS3, TMPRSS13, GUCY1B2, SPRR2F, KDM5D, KRT23, LY6D, VTCN1, SLC5A12, ABCA13, S100A8, GBP6, CEACAM7, IL1F9, MUC15, PPP1R14C, KLK7, DEFB1, MMP3, TM4SF20, SYCP2, TMEM40, CEACAM6, VWA2, CYorf15B, TDRD5, DCDC2, KLK12, C10orf99, CARD14, HOXC13, LOC642587, GABRR1, DNAH2, SLC15A1, KLK8, LGALS7B, ALDH1L1, FOXN1, RAET1L, TP63, ATP13A4, UTY, DSG1, ALOX12B, AMY2B, SAA2, SAA1, GPR115, LCN2, KRT24, PRKY, PADI3, SCNN1A, PGLYRP3, PRSS27, ANXA8L2, WISP3, IL20RB, FOXE1, ZFY, ALG1L, LY6K, NCCRP1, SLC44A5, NMU, CDH17, RNASE7, FER1L4, C4BPB, CALML5, CXorf61, SPACA4, ULBP2, KRTDAP, MYO7B, , C7orf54, SLC6A20, LOC221442, GOLGA8B, GPR81, ALDH3B2, TRIM29, C21orf125, BNC1, IL1F6, TRPA1, C12orf27, LOC440905, POU5F1, LOC100131726, KCNH8, SPRR2C, IL13RA2, C3orf67, SAA4, LYPD3, GJB5, ULBP1, PHACTR3, S100A9, TRIM54, KLK11, ITIH5L, BNIPL, LOC284233, LOC387646, MRPL42P5, FLJ42393, PCDHAC2, IFNE, BCL2L10, FLJ45445, COL7A1, GDA, ARL14, VIP, CLC, PPP2R2B, FABP4, SFRP5, CCL8, CD79A, HIST1H1E, CCL4L2, SYNPO2, SLC47A1, CXCR2P1, C1QTNF7, CDH2, SIGLEC7, GZMA, NKX6-3, PGM5, LILRB4, NMUR1, XCL2, SOX10, FCGR1A, ST6GALNAC5, CCL4, ACSM5, FOLR2, KCNA1, CCL26, FCGR1C, VSIG4, C1QA, NCF1B, NXPH3, C1QC, TAGLN, PTPRN, TYROBP, ISL1, HCST, RGMA, APOBEC3H, PRND, ADIPOQ, CCL11, GLP2R, PNMA6A, CD8B, FCGR1B, PCBP3, CD8A, GPR27, DPP6, F13A1, EPYC, CXCL9, GREM2, SNORA12, DNAJC5B, HLA-DQA2, PLP1, CHGB, APOC2, FAT3, TLR7, CHRDL1, LY86, AGTR1, ISLR, CR1, FCGR3A, DPEP1, PPAPDC1A, ODZ3, MAPK4, C17orf87, KCNJ5, MSR1, RELN, APOE, C1QB, PHYHIPL, CCDC80, NCF1C, SLITRK5, TREM2, FGF14, PGA5, NKX2-3, PRELP, FCRLA, CCL5, GNAO1, PCDH9, WSCD2, MS4A1, RNF150, KLRD1, KCNK2, MATK, HUNK, IGF1, CD01, COLEC12, FIBIN, CARTPT, VPREB3, PPP1R1A, GDF6, NCAM2, CLEC10A, KIAA0408, BHLHE22, CD52, SULT4A1, SIT1, FNDC1, GRIK3, SUCNR1, TMEM90B, CYP1B1, MKX, SV2B, BMP3, TCEAL2, CADM3, SPOCK1, GREM1, SIGLEC8, ADCY2, CDC26, OGN, FCRL6, PSD, VIPR2, GZMM, ADH1B, CCL19, PLD4, TMEM189-UBE2V1, NKX7, EOMES, STMN2, GZMH, NPTX1, CNTFR, TMEM130, CTNND2, ITGBL1, ECEL1, ACTG2, COL10A1, ST6GAL2, NRK, PI16, NALCN, GZMK, NCAM1, RMRP, CHRM2, KCNMA1, HSPB7, SLIT2, PDZRN4, CNN1, CHRDL2, OMD, PTGIS, RSPO3, PLA2G2D, SMYD1, ZNF683, NRXN3, PCSK1N, HSPB6, C16orf89, PGA3, COMP, C2orf40, C7, GKN2, DES, CILP, COL11A1, LIPF, GATA5, MFAP5, SCRG1, SFRP2, GKN1, PGA4, CCL21, SIX2, NKX3-2, SFRP4, NBLA00301, THBS4, HAND2, 및 BARX1.

청구항 27

청구항 26의 조성물, 시약 및 포장 단위를 포함하는 위암 또는 식도암을 진단하기 위한 키트.

발명의 설명

기술 분야

위식도경계부선암의 진단 및 표적 치료에 관한 것이다.

배경 기술

미국의 SEER 데이터베이스의 보고에 따르면, 식도선암은 최근 30년간 600%의 증가율을 보이며 급격히 증가하였다. 특히 이러한 증가율의 대부분은 하부 1/3 식도에 발생하는 하부식도선암이며, 위식도경계부암 또한 2배에 가까운 증가율을 보였다 (Pohl H and Welch HG, 2005). 일본 국립 암 센터 (National Cancer Center)의 보고에 따르면 위식도경계부선암은 1962-1965년 2.3%에서 2001-2005년 10.0%로 5배 가까이 증가하였으며, 특히 분문부암(cardia cancer)(Siewert type II)이 차지하는 비율은 28.5%에서 57.3%로 2배 이상 증가하였다.

최근 발표된 국내 위암학회 보고에 따르면, 국내 또한 위식도경계부선암이 포함된 상부위암이 1995년 11.2%에서

2014년 16.0%로 뚜렷이 증가하고 있다 (Information Committee of Korean Gastric Cancer Association. 2016).

[0004] 위식도경계부선암은, 우리나라에서 지금까지 흔하게 관찰되었던 중하부 위선암과는 달리, 아직까지 종양의 발암 기전에 대한 가설뿐 아니라 종양의 분류법과 그 생물학적 특징, 수술 방법 등에 대한 아직 학문적 의견 일치가 부족하며 최근까지도 여러 가지 다양한 논의가 이루어지고 있다. 그러나 식도선암이나 상부위암 대비 위식도경계부 선암을 구분지을 수 있는 생물학적 특성의 차이에 대한 근본적인 연구는 아직 부족한 실정으로, 실제 이 세 가지 인접한 암종이 어떠한 분자생물학적 차이를 가지고 발현되며 어떠한 유전자가 가장 효율적인 치료 타겟이 될 수 있을지에 대해서는 아직 뚜렷한 결과가 없다.

[0005] 특히 이 위치에 있는 암종은 대부분 위전절제와 함께 매우 긴 시간의 수술법이 요구되며, 이러한 위전절제시 수술 후 전신 영양상태 및 항암치료에 대한 순응도 (compliance)가 크게 감소하여 재발율을 높이는 불량한 예후 인자로 작용한다. 따라서 위식도경계부선암에서, 수술 전 항암화학요법이나 표적치료제 적용에 대한 새로운 진단 방법 또는 치료제 선택에 관한 연구의 필요성은 계속 강조되고 있다.

[0006] 최근 암 유전체 아틀라스 그룹 (the Cancer Genome Atlas group: TCGA)에서 위암 및 식도암에 대한 대규모 분자생물학적 분석 결과가 발표되었으나, 생존율 등 실제 임상에서의 적용에는 제한이 있으며, 이 연구 역시 위식도선암과 분문부암에 대한 정의가 모호하고, 치료 타겟 측면에서 식도선암과의 유사성 또는 차이점에 대한 직접적인 비교 분석이 없다.

발명의 내용

해결하려는 과제

[0007] 일 양상은 위암 또는 식도암의 진단에 필요한 정보를 제공하기 위하여, 개체로부터 분리된 위 또는 식도 조직의 EGFR 또는 ERBB2의 발현량을 측정하는 단계를 포함하는 방법을 제공한다.

[0008] 다른 양상은 EGFR 또는 ERBB2 단백질, 또는 이들 각각의 단편에 특이적으로 결합하는 항체, 항원 결합 단편, 또는 폴리펩티드, 또는 EGFR 또는 ERBB2 단백질을 코딩하는 폴리뉴클레오티드 서열에 특이적으로 결합하는 프로브, 프라이머 세트, 또는 뉴클레오티드를 포함하는 개체의 위암 또는 식도암을 진단하기 위한 조성물 및 이를 포함하는 키트를 제공한다.

과제의 해결 수단

[0009] 일 양상은 위암 또는 식도암의 진단에 필요한 정보를 제공하기 위하여, 개체로부터 분리된 위 또는 식도 조직의 EGFR 또는 ERBB2의 발현량을 측정하는 단계를 포함하는 방법을 제공한다.

[0010] 용어 "EGFR(epidermal growth factor receptor)"(ErbB1 또는 HER1 이라고도 지칭됨)은 세포외 단백질 리간드의 표피 성장 인자 패밀리에(epidermal growth factor family)의 막에 대한 수용체로서 작용하는 막 관통 단백질 또는 이를 코딩하는 유전자를 가리킨다. 이의 서열은 당해 분야에 공지되어 있으며, 예를 들면 Genbank Accession No. AAI18666.1 (<https://www.ncbi.nlm.nih.gov/genbank/>) 등이 있다.

[0011] 용어 "ErbB2"(HER2(human epidermal growth factor receptor 2), 또는 Neu 라고도 지칭됨)는 ErbB 패밀리에 속하는 수용체 타이로신 카이네이즈(receptor tyrosine kinase) 단백질 또는 이를 코딩하는 유전자를 가리킨다. 이의 서열은 당해 분야에 공지되어 있으며, 예를 들면 Genbank Accession No. NP_001005862 (<https://www.ncbi.nlm.nih.gov/genbank/>) 등이 있다.

[0012] 상기 방법은 위암 또는 식도암의 진단 및/또는 표적 치료에 관한 것이다. 상기 위암 또는 식도암은 위식도경계부선암(adenocarcinoma of gastroesophageal junction), 분문부암(cardia cancer), 위식도경계부/분문부암, 상부위암(upperthird gastric cancer), 위저부암(gastric cancer located at fundus), 위체부암(gastric cancer located at body), 위저부/체부암(gastric adenocarcinoma located at fundus or body of the stomach: GCFB), 식도 선암(esophageal adenocarcinoma: EAC), 또는 이들의 조합일 수 있다. 상기 위암 또는 식도암은 선암(adenocarcinoma)으로 분류되는 것일 수 있다. 위와 식도는 해부학적으로 연결되어 있고, 위와 식도의 경계부가 모호하여 위식도경계부선암의 생물학적인 이해 및 위선암 혹은 식도선암과의 분자생물학적 특징의 비교에 대해 끊임 없는 논란이 지속되어 왔다. 그러나 본 발명의 일 실시예는 이러한 암들을 분자생물학적 마커를 이용하여 정확히 진단하고, 또한 상기 마커를 표적으로 하는 치료에 유용하게 이용될 수 있다.

[0013] 상기 EGFR 또는 ERBB2의 발현량을 측정하는 단계는, EGFR 또는 ERBB2의 mRNA의 양을 확인하는 단계, 단백질의

양을 확인하는 단계, 또는 유전자 복제 수 변이를 확인하는 단계일 수 있다.

- [0014] 상기 발현량을 측정하는 단계는 EGFR 또는 ERBB2 단백질 또는 그를 코딩하는 mRNA에 특이적으로 결합하는 물질을 조직 시료에 접촉시켜 복합체를 형성시키고, 형성된 복합체의 수준을 측정하는 단계를 포함할 수 있다. 이후, 상기 복합체의 수준을 측정하는 것은, 예를 들면 상기 특이적으로 결합하는 물질에 검출 가능한 표지를 부착시키고 이로부터 나오는 신호를 검출함으로써 이루어지는 것일 수 있다. 상기 측정은 RT-PCR, 경쟁적 RT-PCR(competitive RT-PCR), 실시간 RT-PCR(Real-time RT-PCR), RNase 보호 분석법(RNase protection assay: RPA), 노던 블롯팅 (Northern blotting), DNA를 포함한 핵산 마이크로어레이, 웨스턴블롯팅, ELISA(enzyme linked immunosorbent assay), 방사선 면역분석(Radioimmunoassay, RIA), 방사 면역 확산법(radioimmunoassay), 오우크테로니(Ouchterlony) 면역 확산법, 로켓트(rocket) 면역전기영동, 조직면역 염색, 면역침전 분석법 (Immunoprecipitation Assay), 보체 고정 분석법(Complement Fixation Assay), FACS, 질량분석, 자석비드-항체 면역 침강법, 단백질 칩(protein chip), 또는 그 조합으로 수행할 수 있다.
- [0015] 또한, 상기 발현량을 측정하는 단계는 역상 단백질 분석(reverse phase protein analysis), 조직 마이크로어레이(tissue microarray), 면역조직화학(immunohistochemistry: IHC), 은 제자리 부합법(silver in situ hybridization: SISH), 또는 이들의 조합으로 수행할 수 있다.
- [0016] 상기 방법은 상기 측정된 EGFR 또는 ERBB2의 발현량을 대조군에서 측정된 EGFR 또는 ERBB2의 발현량과 비교하는 단계를 포함할 수 있다. 상기 대조군은 위암 또는 식도암에 걸리지 않은 개체의 조직, 또는 위암 또는 식도암 조직일 수 있다.
- [0017] 상기 방법은 또한 상기 EGFR 또는 ERBB2의 발현량이 대조군의 발현량에 비교하여 변화한 경우 개체를 위암 또는 식도암에 걸린 것으로 결정하는 단계를 포함할 수 있다. 상기 변화는 대조군에 비하여 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900% 및 1000% 이상 증가하거나, 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% 및 100% 이상 감소하는 것을 포함할 수 있다. 예를 들면, 개체로부터 채취한 조직의 ERBB2 유전자의 복제 수가 정상 조직에 비하여 100%(2배) 이상 증가(증폭)한 경우, 상기 개체는 식도암에 걸린 것으로 판정할 수 있다.
- [0018] 상기 방법은 상기 EGFR 또는 ERBB2의 발현량이 증가한 경우, EGFR 또는 ERBB2의 저해제를 투여하기로 결정하는 단계를 더 포함할 수 있다. 상기 EGFR 또는 ERBB2의 저해제는 당해 분야에서 통상적으로 이용되는 것, 상업적으로 판매되는 것, 또는 임상 시험 중인 것이면 어느 것이든지 이용 가능하며, 예를 들면, 라파티닙(lapatinib), 트라스투주맙(Trastuzumab), 세특시맙(cetuximab), 제피티닙(gefitinib), 에피티닙(efitinib) 또는 이들의 조합일 수 있다.
- [0020] 다른 양상은 EGFR 또는 ERBB2 단백질, 또는 이들 각각의 단편에 특이적으로 결합하는 항체, 항원 결합 단편, 또는 폴리펩티드, 또는 EGFR 또는 ERBB2 단백질을 코딩하는 폴리뉴클레오티드 서열에 특이적으로 결합하는 프로브, 프라이머 세트, 또는 뉴클레오티드를 포함하는 개체의 위암 또는 식도암을 진단하기 위한 조성물 및 키트를 제공한다.
- [0021] 상기 EGFR 또는 ERBB2 단백질, 또는 이들 각각의 단편에 특이적으로 결합하는 항체, 항원 결합 단편, 또는 폴리펩티드, 또는 EGFR 또는 ERBB2 단백질을 코딩하는 폴리뉴클레오티드 서열에 특이적으로 결합하는 프로브, 프라이머 세트, 또는 뉴클레오티드는 고속 대량 스크리닝(high throughput screening: HTS) 등 당업계에 알려진 방법에 따라 실시자가 직접 제조 및 선별할 수 있고, 또는 상업적으로 구매 가능한 것일 수 있다.
- [0022] 또한 상기 키트는 하나 이상의 반응 시약을 갖는 포장단위를 포함하는 키트 형태를 제공할 수 있다. 또한 상기 키트는 하나 이상의 하기의 품목을 포함할 수 있다: 검출 대상을 검출하기 위한 물질, 완충액, 사용설명서, 및 양성 또는 음성 대조군. 상기 키트는 본 명세서에 기술된 방법을 수행하기 위해 적절한 비율로 함께 혼합된 반응 시약들의 용기들을 포함할 수 있다. 반응 시약 용기들은 바람직하게는 상기 방법을 수행할 때 측정하는 단계를 생략할 수 있도록 단위 수량의 반응 시약을 포함할 수 있다.
- [0023] 본 발명의 일 실시예에서, 식도양 선암은 분화암 및 장형암과의 연관성이 유의하게 높았으며, ERBB2의 유전자 복제수변이가 유의하게 증폭되어 있고, ERBB2 및 EGFR의 단백질 발현이 유의하게 증가되어 있음을 확인하였으므로, ERBB2 또는 EGFR 발현량을 측정함으로써, 식도암 또는 위저부/체부암을 구별하여 진단할 수 있고, 이들 암의 치료를 위한 유용한 표적까지 제공할 수 있다.

[0025] 다른 양상은 개체의 위암 또는 식도암을 진단하기 위한 정보를 제공하기 위하여, 개체로부터 분리된 조직의 유전자의 발현 프로필을 측정하는 단계; 및 상기 유전자 발현 프로필로부터 BCCP (Bayesian Compound Covariate Predictor) 알고리즘을 실행하는 컴퓨터 시스템을 이용하여 BCCP 점수를 얻는 단계를 포함하는 유전자 발현 프로필 분석 방법을 제공한다.

[0026] 상기 유전자는 아래의 400개 유전자를 포함하는 것일 수 있다: KRT5, KRT13, KRT6A, KRT14, RHCG, KRT4, S100A7, SPRR3, KRT6C, SPRR2A, MUC21, KRT6B, PKP1, SERPINB3, CLCA2, SPRR1B, MUC4, TBX5, CRNN, SPRR2D, DUOX2, GJB6, SPRR2E, S100A2, KRT16, KRT17, SCEL, DSC3, CALML3, HEPHL1, A2ML1, NKX6-1, DUOX2, SBSN, TMPRSS11D, CLCA4, ANXA8, FGFBP1, TGM1, IVL, DUOX1, DSG3, EREG, SERPINB4, KRT78, HOXC10, DUOX1, ZNF750, SPRR1A, KRT7, TMPRSS11A, CXCL6, KLK13, CRCT1, PAX9, HORMAD1, FAM83A, S100A7A, GPR110, CAPN14, ABCA12, WFDC2, TMPRSS11B, CEACAM5, MUC12, SERPINB13, PADI1, FAT2, IL1RL2, STK31, SPINK5, DDX3Y, PI3, FAM83C, LYPD2, KLK5, IL1F5, KRT15, BBOX1, GPR87, ACTL8, SLC34A2, USP9Y, TGM3, LASS3, TMPRSS13, GUCY1B2, SPRR2F, KDM5D, KRT23, LY6D, VTCN1, SLC5A12, ABCA13, S100A8, GBP6, CEACAM7, IL1F9, MUC15, PPP1R14C, KLK7, DEFB1, MMP3, TM4SF20, SYCP2, TMEM40, CEACAM6, VWA2, CYorf15B, TDRD5, DCDC2, KLK12, C10orf99, CARD14, HOXC13, LOC642587, GABRR1, DNAH2, SLC15A1, KLK8, LGALS7B, ALDH1L1, FOXN1, RAET1L, TP63, ATP13A4, UTY, DSG1, ALOX12B, AMY2B, SAA2, SAA1, GPR115, LCN2, KRT24, PRKY, PADI3, SCNN1A, PGLYRP3, PRSS27, ANXA8L2, WISP3, IL20RB, FOXE1, ZFY, ALG1L, LY6K, NCCRP1, SLC44A5, NMU, CDH17, RNASE7, FER1L4, C4BPB, CALML5, CXorf61, SPACA4, ULBP2, KRTDAP, MYO7B, , C7orf54, SLC6A20, LOC221442, GOLGA8B, GPR81, ALDH3B2, TRIM29, C21orf125, BNC1, IL1F6, TRPA1, C12orf27, LOC440905, POU5F1, LOC100131726, KCNH8, SPRR2C, IL13RA2, C3orf67, SAA4, LYPD3, GJB5, ULBP1, PHACTR3, S100A9, TRIM54, KLK11, ITIH5L, BNIPL, LOC284233, LOC387646, MRPL42P5, FLJ42393, PCDHAC2, IFNE, BCL2L10, FLJ45445, COL7A1, GDA, ARL14, VIP, CLC, PPP2R2B, FABP4, SFRP5, CCL8, CD79A, HIST1H1E, CCL4L2, SYNPO2, SLC47A1, CXCR2P1, C1QTNF7, CDH2, SIGLEC7, GZMA, NKX6-3, PGM5, LILRB4, NMUR1, XCL2, SOX10, FCGR1A, ST6GALNAC5, CCL4, ACSM5, FOLR2, KCNA1, CCL26, FCGR1C, VSIG4, C1QA, NCF1B, NXPH3, C1QC, TAGLN, PTPRN, TYROBP, ISL1, HCST, RGMA, APOBEC3H, PRND, ADIPOQ, CCL11, GLP2R, PNMA6A, CD8B, FCGR1B, PCBP3, CD8A, GPR27, DPP6, F13A1, EPYC, CXCL9, GREM2, SNORA12, DNAJC5B, HLA-DQA2, PLP1, CHGB, APOC2, FAT3, TLR7, CHRDL1, LY86, AGTR1, ISLR, CR1, FCGR3A, DPEP1, PPAPDC1A, ODZ3, MAPK4, C17orf87, KCNJ5, MSR1, RELN, APOE, C1QB, PHYHIPL, CCDC80, NCF1C, SLITRK5, TREM2, FGF14, PGA5, NKX2-3, PRELP, FCRLA, CCL5, GNAO1, PCDH9, WSCD2, MS4A1, RNF150, KLRD1, KCNK2, MATK, HUNK, IGF1, CD01, COLEC12, FIBIN, CARTPT, VPREB3, PPP1R1A, GDF6, NCAM2, CLEC10A, KIAA0408, BHLHE22, CD52, SULT4A1, SIT1, FNDC1, GRIK3, SUCNR1, TMEM90B, CYP1B1, MKX, SV2B, BMP3, TCEAL2, CADM3, SPOCK1, GREM1, SIGLEC8, ADCY2, CDC26, OGN, FCRL6, PSD, VIPR2, GZMM, ADH1B, CCL19, PLD4, TMEM189-UBE2V1, NKX7, EOMES, STMN2, GZMH, NPTX1, CNTFR, TMEM130, CTNND2, ITGBL1, ECEL1, ACTG2, COL10A1, ST6GAL2, NRK, PI16, NALCN, GZMK, NCAM1, RMRP, CHRM2, KCNA1, HSPB7, SLIT2, PDZRN4, CNN1, CHRDL2, OMD, PTGIS, RSP03, PLA2G2D, SMYD1, ZNF683, NRXN3, PCSK1N, HSPB6, C16orf89, PGA3, COMP, C2orf40, C7, GKN2, DES, CILP, COL11A1, LIPF, GATA5, MFAP5, SCRG1, SFRP2, GKN1, PGA4, CCL21, SIX2, NKX3-2, SFRP4, NBLA00301, THBS4, HAND2, 또는 BARX1.

[0027] 상기 400개 유전자에 관한, 조성물, 키트, 및 측정 방법 등은 특별히 다른 언급이 없는 한 상기 EGFR 또는 ERBB2에 관한 설명에서 언급된 내용들이 적용될 수 있다.

발명의 효과

[0028] 위식도경계부암을 진단하기 위한 마커 및 표적으로서 ERBB2 및 EGFR을 제공한다.

도면의 간단한 설명

[0029] 도 1은 TCGA 코호트의 연구군의 해부학적 분포를 나타낸다.

도 2는 SNU 코호트의 연구군의 해부학적 분포를 나타낸다.

도 3은 BCCP를 이용한 클래스 예측 모델을 나타낸다.

도 4는 분석 개요에 따른 상세한 연구군을 나타낸다.

도 5는 TCGA 코호트의 EAC 및 GCFB 사이의 5,520개 유전자의 감독되지 않은 클러스터링을 나타낸다.

도 6은 400개 시그니처 유전자 분류를 이용하여 TCGA 코호트 내의 EAC 및 GCFB의 감독되지 않은 계층적 클러스터링을 수행한 결과를 나타낸다.

도 7은 LOOCV를 이용한 교차 검증 후 ROC 곡선을 나타낸다.

도 8은 BCCP를 이용한 TCGA 코호트의 GEJ/분문부의 계층적 클러스터링을 나타낸다.

도 9는 BCCP를 이용한 SNU 코호트의 AGEJ 또는 상부위암(upperthird gastric cancer)의 계층적 클러스터링을 나타낸다.

도 10은 TCGA 코호트의 EAC-예측 및 GCFB-예측 그룹 사이의 복제 수 변화를 나타낸다.

도 11은 SNU 코호트의 EAC-예측 및 GCFB-예측 그룹 사이의 복제 수 변화를 나타낸다.

도 12는 TCGA 코호트의 역상 단백질 분석을 이용한 히트맵을 나타낸다.

도 13은 조직 마이크로어레이의 IHC 염색을 이용한 단백질 발현을 나타낸다(200x).

도 14는 SNU 코호트의 EAC-예측 그룹 및 GCFB-예측 그룹 사이의 조직 마이크로어레이의 복합 H 점수를 나타낸다.

도 15는 CCLE 데이터베이스를 이용한 예측 모델의 외부 검증을 나타낸다.

도 16은 EAC-예측 그룹 및 GCFB-예측 그룹 사이의 CCLE 데이터베이스의 계층적 클러스터링을 나타낸다.

도 17은 EAC-예측 그룹 및 GCFB-예측 그룹 사이의 CCLE 데이터베이스의 IC50 데이터를 이용한 라파티닙의 약물 반응을 나타낸다.

발명을 실시하기 위한 구체적인 내용

이하 본 발명을 실시예를 통하여 보다 상세하게 설명한다. 그러나, 이들 실시예는 본 발명을 예시적으로 설명하기 위한 것으로 본 발명의 범위가 이들 실시예에 한정되는 것은 아니다.

1. 재료 및 방법

(1) TCGA 코호트

암 유전체 지도 (The Cancer Genome Atlas: TCGA)(<https://tcga-data.nci.nih.gov/tcga/>) 및 mRNA 발현, 체세포 돌연변이, 삼입/결실, 복제 수 변화, 순수 식도 선암(pure esophageal adenocarcinoma), 위식도 경계부(gastroesophageal junction) 또는 분문부(cardia)의 선암 및 위 저부(fundus) 또는 체부(body)에 위치한 순수 위 선암 (pure gastric adenocarcinoma located at fundus or body of the stomach: GCFB)의 역상 단백질 어레이(RPPA)의 검색된 정보들을 조사하였다. 상기 언급된 각 인체 부위의 해부학적 위치는 도 1에 나타난 바와 같다.

(2) 서울대학교 코호트

위암생물학 연구실, 암 연구소, 및 서울대학교에서 1999-2015년에 모은 임상 정보로서, AGEJ 및 위의 3등분의 상위부(upperthird: UT)에 대한 임상적 정보를 포함하는 신선 냉동 조직 보관 데이터베이스를 조사하였다. 상기 신선 조직의 보관은 서울대학교 병원의 조직 검토 위원회(IRB No: H-0806-072-248)로부터 승인 받았다. 초기 진단 당시에 다른 원발성 악성 종양, 재발 선암 또는 잔류 위암을 가진 환자는 제외되었다. 서울대학교 병원의 코호트에서 AGEJ 및 위 UT의 선암은 위 식도 접합부로부터의 거리 기준을 사용하여 분류하였다. 상기 언급된 각 인체 부위의 해부학적 위치는 도 2에 나타난 바와 같다.

AGEJ II는 위식도 경계부에서 구강쪽으로 1 cm 내, 및 구강 반대쪽으로 2 cm 에 위치한 종양으로 정의되었다. 이는 Siewert II 형 암의 일반적인 정의와 같다.

AGEJ III는 위식도 경계부의 포함 여부와 관계 없이, 위식도 경계부에서 구강 반대쪽으로 2-5 cm에 위치한 종양으로 정의되었다. AGEJ II 또는 AGEJ III를 제외한 나머지 위 선암 상위부 3분의 1은 UT로 정의되었다. AGEJ II

로 분류가능한 모든 종양이 조사되고, 차세대(next-generation) 시퀀싱이 수행되었다. AGEJ III 및 UT는 가장 최근의 동일한 수의 샘플을 조사하였다. 병리학적 단계는 7th AJCC TNM 분류에 따라 진단하였다. 병리학적 분석을 위해, 유두(papillary), 잘 분화되거나 중간 정도로 분화된 유형들은 분화 그룹으로 분류하였고, 잘 분화되지 않거나, 점액성, 응집성이 좋지 않은 세포 유형은 비분화 그룹으로 분류하였다. 본 연구 프로토콜은 서울대학교 병원의 기관 검토위원회의 승인을 받았다 (IRB No: H-1501-027-639).

[0041] (3) 핵산 처리

[0042] SNU 코호트의 신선 조직 보관소로부터 각각의 냉동 종양 및 이에 상응하는 정상 위 점막을 약 $2 \times 2 \times 1 \text{ mm}^3$ 의 크기로 준비하였다. 스핀-칼럼 프로토콜을 이용하는 Qiagen DNA extraction kit (Qiagen, Venlo, Netherlands)를 사용하여 DNA를 추출하였다. 추출된 DNA는 최소 $A_{260}/_{280} \geq 1.7$ 및 dsDNA의 양 $\geq 3.0 \text{ } \mu\text{g}$ 에서 QUBIT HS dsDNA assay (Life Technologies Gaithersburg, MD, USA)를 이용하여 정량화하였다. RNA의 분리는 TRIzol의 제조사가 제공하는 프로토콜 (사용자 매뉴얼 TRIzol® Reagent(www.invitrogen.com))에 따라 Eppendorf Tube5.0 mL에서 수행하였다. 용해를 위해서 각각의 신선한 조직에 1 mL TRIzol을 첨가하였다. TRIzol 프로토콜에 따라 출발 물질 1 mL를 각각의 튜브에 옮기고, 각각에 튜브에 200 μL 클로로폼을 첨가하였다. 수상으로부터 RNA를 침전시키기 위해 이소프로판올 0.5 mL와 아래의 세척 단계에서 1 mL 에탄올(75%)를 사용하였다. RNA 침전 및 세척 단계는 12,000 x g에서 5.0 mL 튜브에서 수행하였다. 생성된 RNA 펠렛은 50 μL EDPC-처리수에 다시 현탁시켰다. NanoDrop™ 1000 (Thermo Scientific)을 이용하여, 260 nm 및 280 nm에서 OD를 측정하고 샘플의 농도 및 순도를 결정하였다.

[0044] (4) SNU 코호트의 전체 전사체(whole transcriptome) 시퀀싱

[0045] SNU 코호트로부터의 모든 종양 샘플을, Illumina Truseq RNA library preparation kit (Ribo-Zero rRNA Removal Kit)를 이용하여 전체 전사체 라이브러리를 만드는데 이용하였다. 모든 라이브러리는 Illumina HiSeq2000 platform 상에서 일 레인 당 하나의 시료를 이용하여 시퀀싱하였다(a paired-end 2 x 101 bp read length). 종양 RNA 및 이에 상응하는 정상 RNA는 같은 보통 같은 플로우 셀(flow cell)에 로드하였다. 해독 정렬 및 프로세싱은 STAR aligner 및 Broad Institute의 Picard를 이용하여, GATK 최적 사례 추천 (GATK best practice recommendation)으로서 수행하였다 (<http://broadinstitute.github.io/picard>).

[0047] (5) SNU 코호트의 전체 엑솜(exome) 시퀀싱

[0048] 종양 및 이에 상응하는 정상 위 점막 샘플으로부터 최소 3 μg 의 dsDNA의 전체 엑솜 시퀀싱을 Agilent SureSelectHumanAll Exon V5 + UTR region kit를 이용하여 수행하였다. 2 x 101 bp 해독 길이의 페어드 엔드(paired-end) 쌍을 Illumina HiSeq2000 플랫폼 상에서 시퀀싱하였다. 시퀀싱 목표 범위(depth)는 종양 및 정상 조직 모두에서 적어도 100x로 계획하였다 (종양의 경우 이상적으로는 200x). 해독 정렬 및 프로세싱은 Burrows-Wheeler Aligner(BWA)-mem 및 Broad Institute의 Picard를 이용하여 GATK 최적 사례 추천으로서 수행하였다.

[0050] (6) 전사체 시퀀싱을 이용한 예측 분류 알고리즘

[0051] 유전자 발현 데이터를 분석하기 위해 BRB Array Tool을 사용하였다. TCGA 및 SNU 코호트의 RNA-시퀀싱 데이터는 다르게 발현되는 유전자(differentially expressed genes: DEG) 및 예측 모델을 구축하기 위해 함께 분석하였다. 먼저, TCGA 코호트 중 EAC 및 GCFB 사이에서 DEG들을 Student's t 검정으로 확인하고 ($P < 0.001$), 배수 변화가 상위 또는 하위에 해당하는지 여부에 따라 추가적으로 선별하였다. 예측 모델을 구축하기 위해, 이전에 확립된 BCCP (Bayesian Compound Covariate Predictor) 알고리즘과 함께 LOOVC (leave-one-out cross validation) 접근법을 사용하였다. 이 때 사용된 방법은 Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. J Comput Biol 2002;9:505-511, Wright G, Tan B, Rosenwald A et al. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci U S A 2003;100:9991-9996, YODEN WJ. Index for rating diagnostic tests. Cancer 1950;3:32-35. Robin X, Turck N, Hainard A et al. pROC: an open-source package

for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77. doi: 10.1186/1471-2105-12-77.:77-12 등의 문헌을 참조하였으며, 상기 문헌들의 기재 내용은 본 명세서에 모두 포함된다 (도 3).

[0052] 학습된 모델 (trained model)의 민감도 및 특이도는 ROC 커브로 평가하였다. EAC 및 GCFB 사이의 최적의 컷-오프 값은 Youden 인덱스를 이용하여 결정하였다. 예측 모델과 이의 컷-오프 값에 대한 외부 검증(external validation)은 CCLE(Cancer Cell Line Encyclopedia) 데이터베이스 (<http://www.broadinstitute.org/ccle>)의 위 및 식도 선암 세포주의 RNA 마이크로어레이를 이용하여 수행하였다. BCCP 모델의 가능성에 따라, TCGA 코호트의 GEJ/분문부 종양 및 SNU 코호트의 모든 종양은 게놈 아형에 따라 재분류되었다 (EAC-예측 또는 GCFB-예측). 임상 및 분자적 수준에서의 게놈 아형의 차이는 이후 임상 병리학적 데이터, 돌연변이, 복제 수 변화를 분석하여 평가하였다. 계통 분석(pathway analysis)는 KEGG (Kyoto Encyclopedia of Genes and Genomes) 분석 툴 (<http://www.kegg.jp>)를 이용하여 수행하였다. 게놈 아형과 관련된 잠재적인 대리 마커 (surrogate marker)는 TCGA 코호트에 대한 역상 단백질 분석 및 SNU 코호트에 대한 조직 마이크로어레이로 확인하였다. 이러한 대리 마커들의 표적 약물 반응성은 CCLE 데이터 베이스의 위 및 식도 선암 세포 주의 IC50을 이용하여 비교하였다.

[0054] (7) 체세포 돌연변이 및 삽입/결실의 확인

[0055] TCGA 코호트는, 삽입/결실을 포함하는 체세포 돌연변이를 당해 분야에 기존에 보고된 방법들을 이용하여 분석하였다.

[0056] SNU 코호트는, 체세포 돌연변이 콜링(somatic mutation calling)을 위해 Mutect 및 IndelGenotyper를 이용하여 전체 엑솔 시퀀싱에 대한 BAM 파일을 이용하였다. 1) 대표 서열(reference sequence)을 기반으로 엑손 및 스플라이싱 변형, 또는 2) 8개 초과 서열 및 변이형 대립유전자(alternate allele) 4개 초과인 경우의 변이를 선별하였다.

[0057] 각각의 TCGA 및 SNU 코호트 내에서 EAC-예측 그룹 및 GCFB-예측 그룹간의 체세포 돌연변이 및 삽입/결실을 비교하였다.

[0059] (8) 체세포 복제 수 분석

[0060] TCGA 코호트는, 단일 염기 다형성(SNP) 분석으로 복제 수 변화(copy number alteration: CNA)를 당해 분야에 기존에 보고된 방법들을 이용하여 분석하였다.

[0061] SNU 코호트는, CNA를 CONIFER로부터의 RPKM(Read Per Kilobase per Million mapped reads) 값에 기반한 전체 엑솔 데이터를 이용하여 분석하였다.

[0062] TCGA 및 SNU 코호트 모두에 대해서 체세포 CNA 분석은 GISTIC 2.0 알고리즘을 이용하여 수행하였다. GISTIC 알고리즘을 이용하여 국소적 복제 수 증폭을 갖는 것으로 확인된 유전자들 중, 적어도 한 쌍의 샘플에서, 종양의 log2 복제 수 비가 그에 상응하는 정상 위 점막의 값보다 ≥ 1 인 후보 유전자들을 선별하였다.

[0063] EAC-예측 및 GCFB-예측 하위군에서 상기 후보 유전자들의 복제 수를 Student's t 검정으로 비교하였다.

[0065] (9) TCGA 코호트의 역상 단백질 어레이

[0066] 180 사례 중 132 사례(44 EAC 및 88 GCFB 포함)의 역상 단백질 어레이(RPPA) 데이터를 TCGA 데이터베이스로부터 얻었다. 클러스터링(clustering) 분석은 재중심화 정규화(recentered normalization) 후 수행하였다.

[0068] (10) SNU 코호트의 조직 마이크로어레이(TMA)

[0069] 코어 조직 생검(직경 2 mm)을 각각의 파라핀-고정 위 종양으로부터 얻고(기증 블록), 트레핀(trephine) 장치(Superbiochips Laboratories, Seoul, Korea)를 이용하여 이를 새로운 수용 파라핀 블록에 나열하였다(조직 어레이 블록). 진행된 암의 조직학적 구성 요소 또는 분자적 이상의 다양성을 고려하여, 각각의 샘플마다 3 세트의 조직 마이크로어레이를 제조하였다. 표적 약물 반응은 위암 및 식도 선암 세포주에 대한 CCLE 데이터베이스의 IC50 데이터를 이용하여 평가하였다. 조직 어레이 블록은 총 138 사례에 대하여 3개 어레이 상에 최대 46 코

어를 포함하였고, 면역조직화학(IHC) 염색을 수행하였다. 10% 초과 코어 영역을 갖는 종양을 적절한 것으로 간주 하였다. 각각의 파라핀 블록은 내부 대조군을 포함하였다. IHC는 자동화 면역염색기(BenchMark XT, Ventana Medical Systems, Tucson, AZ, USA)를 이용하여 수행하였다. 각각의 코어로부터 샘플을 채취한 후, 염색 강도를 0(막 염색 없음, 음성), 1+(약함/거의 염색되지 않음, 약한 양성), 2+(약한-중간 정도의 염색, 중간 양성), 및 3+(강한 염색, 강한 양성)로 점수화하였다. 각 TMA 코어에 대한 모든 IHC 염색 및 SISH (silver in situ hybridization)는 모든 임상적 정보에 대해 알지 못하는 한 명의 전문 병리학자가 평가 및 점수화하였다. ERBB2를 제외한 모든 단백질에 대한 염색 상태는 염색된 세포의 비율에 의한 염색 강도와 각각의 강도 수준에 대한 개별적 H-스코어를 곱하여 복합 H-스코어(complex H-score)를 이용하여 분석하였다. ERBB2의 염색 상태는 10% 이상의 세포가 적어도 하나의 TMA 코어에서 염색된 때의, 가장 강한 염색 강도 점수에서 양성으로 간주되었다. IHC 및 SISH 결과를 이용한 최종 해석은 이전에 보고된 방법들에 따라 수행하였다. ERBB2의 발현은 IHC 3+ 또는 IHC 2+ 및 SISH 흑/적 비율 ≥ 2.0 일 때 양성, IHC < 2+, 또는 IHC 2+ 및 SISH 흑/적 비율 < 2.0일 때 음성으로 결정되었다. 표적 약물 반응은 위암 및 식도 선암 세포주에 대한 CCLE 데이터베이스의 IC50 데이터를 이용하여 평가하였다.

(11) 통계학적 분석

스튜던트 t 검정 및 카이 제곱(chi-squared) 검정으로 비교 분석하였다. 생존 분석은 Kaplan-Meier 방법 및 로그 랭크 검정으로 수행하였다. 단백질 발현의 위험 인자를 확인하기 위한 다변량 분석(multivariate analysis)은 이항 로지스틱 회귀 또는 선형 회귀 분석을 이용하여 수행하였다. 모든 검정은 2-방향(sided) 검정으로 수행하였고, SPSS version 21.0(SPSS, Inc., Chicago, IL, USA)을 이용한 유의 수준 5%에서 수행하였다.

2. 결과

도 4는 분석 계획에 따른 자세한 연구군을 나타낸다. TCGA 코호트로부터의 순수 식도 선암, 순수 위선암(저부/체부), GEJ/분문부 선암 228 종양, 및 46쌍(92 샘플)의 종양에 상응하는, SNU 코호트로부터의 AGEJ II, AGEJ III, 및 UT의 정상 점막을 분석하였다.

SNU 코호트에 대해서, 신선한 냉동 조직 및 라이브러리로부터의 핵산 추출을 반복한 후, 로우 시퀀싱 데이터를 얻었다.

(1) 임상 병리학적 특성

TCGA 코호트에서, 엑스 및 전사체 데이터에 사용할 수 있는 78 EAC, 48 GEJ/분문부 및 102 GCFB 샘플을 확인하였다 (표 1).

표 1

TCGA 코호트의 임상병리학적 특성

		EAC (n=78)	GEJ/분문부 (n=48)	GCFB (n=102)	P Value
성(남:여)		69:9	37:11	57:45	<0.001
나이		66.8 $\pm$ 12.0	66.9 $\pm$ 9.2	66.6 $\pm$ 9.3	0.985
위치	식도 중간	2(2.6%)	0	0	<0.001
	식도 중-말	1(1.3%)	0	0	
	식도 말단	75(96.2%)	0	0	
	GEJ/분문부	0	48(100%)	0	
	위저부/체부	0	0	102(100%)	
WHO 분류	유두형	0	4(8.3%)	12(11.8%)	
	관형	0	23(47.9%)	49(48.0%)	
	저음집성	0	9(18.8%)	19(18.6%)	
	점액성	0	3(6.3%)	4(3.9%)	

	혼합형	0	6(12.5%)	6(5.9%)	
	이용불가	78(100%)	3(6.3%)	12(11.8%)	
Lauren 분류	장	0	32(66.7%)	70(68.6%)	
	확산형	0	9(18.8%)	19(18.6%)	
	혼합형	0	6(12.5%)	6(5.9%)	
	이용불가	78(100%)	1(2.1%)	7(6.9%)	

[0083] SNU 코호트의 임상병리학적 특성은 아래 표 2와 같다

표 2

[0085] SNU 코호트의 임상병리학적 특성

		AGEJ II (n=16)	AGEJ III (n=16)	UT (n=14)	P Value
성(남:여)		13:3	12:4	11:3	0.912
나이		58.5±10.4	66.5±9.4	63.5±8.1	0.062
WHO 분류	분화	7(43.8%)	7(43.8%)	7(50.0%)	0.919
	비분화	7(43.8%)	8(50.0%)	5(35.7%)	
	결정x	2(12.5%)	1(6.3%)	2(14.3%)	
Lauren 분류	장	5(31.3%)	6(37.5%)	7(50.0%)	0.526
	확산형	7(43.8%)	5(31.3%)	6(41.9%)	
	혼합형	4(25.0%)	5(31.3%)	1(7.1%)	

[0087] (2) 예측적 분류 모델의 개발

[0088] TCGA 코호트 내의 EAC 및 CGFB의 감독되지 않은 계층적 클러스터링(unsupervised hierarchical clustering)으로 5,520 유전자를 밝혀내었다 (도 5)(P<0.001).

[0089] 배수 변화의 등급에 따라 상위 200개 및 하위 200개 유전자를 400개의 시그니처 유전자로 선정하였다. 선정된 유전자의 목록은 다음과 같다: KRT5, KRT13, KRT6A, KRT14, RHCg, KRT4, S100A7, SPRR3, KRT6C, SPRR2A, MUC21, KRT6B, PKP1, SERPINB3, CLCA2, SPRR1B, MUC4, TBX5, CRNN, SPRR2D, DUOX2, GJB6, SPRR2E, S100A2, KRT16, KRT17, SCEL, DSC3, CALML3, HEPHL1, A2ML1, NKX6-1, DUOX2, SBSN, TMPRSS11D, CLCA4, ANXA8, FGFBP1, TGM1, IVL, DUOX1, DSG3, EREG, SERPINB4, KRT78, HOXC10, DUOX1, ZNF750, SPRR1A, KRT7, TMPRSS11A, CXCL6, KLK13, CRCT1, PAX9, HORMAD1, FAM83A, S100A7A, GPR110, CAPN14, ABCA12, WFDC2, TMPRSS11B, CEACAM5, MUC12, SERPINB13, PADI1, FAT2, IL1RL2, STK31, SPINK5, DDX3Y, PI3, FAM83C, LYPD2, KLK5, IL1F5, KRT15, BBOX1, GPR87, ACTL8, SLC34A2, USP9Y, TGM3, LASS3, TMPRSS13, GUCY1B2, SPRR2F, KDM5D, KRT23, LY6D, VTCN1, SLC5A12, ABCA13, S100A8, GBP6, CEACAM7, IL1F9, MUC15, PPP1R14C, KLK7, DEFB1, MMP3, TM4SF20, SYCP2, TMEM40, CEACAM6, VWA2, CYorf15B, TDRD5, DCDC2, KLK12, C10orf99, CARD14, HOXC13, LOC642587, GABRR1, DNAH2, SLC15A1, KLK8, LGALS7B, ALDH1L1, FOXN1, RAET1L, TP63, ATP13A4, UTY, DSG1, ALOX12B, AMY2B, SAA2, SAA1, GPR115, LCN2, KRT24, PRKY, PADI3, SCNN1A, PGLYRP3, PRSS27, ANXA8L2, WISP3, IL20RB, FOXE1, ZFY, ALG1L, LY6K, NCCRP1, SLC44A5, NMU, CDH17, RNASE7, FER1L4, C4BPB, CALML5, CXorf61, SPACA4, ULBP2, KRTDAP, MYO7B, , C7orf54, SLC6A20, LOC221442, GOLGA8B, GPR81, ALDH3B2, TRIM29, C21orf125, BNC1, IL1F6, TRPA1, C12orf27, LOC440905, POU5F1, LOC100131726, KCNH8, SPRR2C, IL13RA2, C3orf67, SAA4, LYPD3, GJB5, ULBP1, PHACTR3, S100A9, TRIM54, KLK11, ITIH5L, BNIPL, LOC284233, LOC387646, MRPL42P5, FLJ42393, PCDHAC2, IFNE, BCL2L10, FLJ45445, COL7A1, GDA, ARL14, VIP, CLC, PPP2R2B, FABP4, SFRP5, CCL8, CD79A, HIST1H1E, CCL4L2, SYNPO2, SLC47A1, CXCR2P1, C1QTNF7, CDH2, SIGLEC7, GZMA, NKX6-3, PGM5, LILRB4, NMUR1, XCL2, SOX10, FCGR1A, ST6GALNAC5, CCL4, ACSM5, FOLR2, KCNA1, CCL26, FCGR1C, VSIG4, C1QA, NCF1B, NXPH3, C1QC, TAGLN, PTPRN, TYROBP, ISL1, HCST, RGMA, APOBEC3H, PRND, ADIPOQ, CCL11, GLP2R, PNMA6A, CD8B, FCGR1B, PCBP3, CD8A, GPR27, DPP6, F13A1, EPYC,

CXCL9, GREM2, SNORA12, DNAJC5B, HLA-DQA2, PLP1, CHGB, APOC2, FAT3, TLR7, CHRDL1, LY86, AGTR1, ISLR, CR1, FCGR3A, DPEP1, PPAPDC1A, ODZ3, MAPK4, C17orf87, KCNJ5, MSR1, RELN, APOE, C1QB, PHYHIPL, CCDC80, NCF1C, SLITRK5, TREM2, FGF14, PGA5, NKX2-3, PRELP, FCRLA, CCL5, GNAO1, PCDH9, WSCD2, MS4A1, RNF150, KLRD1, KCNK2, MATK, HUNK, IGFBPL1, IGF1, CD01, COLEC12, FIBIN, CARTPT, VPREB3, PPP1R1A, GDF6, NCAM2, CLEC10A, KIAA0408, BHLHE22, CD52, SULT4A1, SIT1, FNDC1, GRIK3, SUCNR1, TMEM90B, CYP1B1, MKX, SV2B, BMP3, TCEAL2, CADM3, SPOCK1, GREM1, SIGLEC8, ADCY2, CDC26, OGN, FCRL6, PSD, VIPR2, GZMM, ADH1B, CCL19, PLD4, TMEM189-UBE2V1, NKG7, EOMES, STMN2, GZMH, NPTX1, CNTFR, TMEM130, CTNND2, ITGBL1, ECEL1, ACTG2, COL10A1, ST6GAL2, NRK, PI16, NALCN, GZMK, NCAM1, RMRP, CHRM2, KCNMA1, HSPB7, SLIT2, PDZRN4, CNN1, CHRDL2, OMD, PTGIS, RSPO3, PLA2G2D, SMYD1, ZNF683, NRXN3, PCSK1N, HSPB6, C16orf89, PGA3, COMP, C2orf40, C7, GKN2, DES, CILP, COL11A1, LIPF, GATA5, MFAP5, SCRG1, SFRP2, GKN1, PGA4, CCL21, SIX2, NKX3-2, SFRP4, NBLA00301, THBS4, HAND2, 및 BARX1.

- [0090] 상기 400개 시그니처 유전자 분류를 이용하여 TCGA 코호트 내의 EAC 및 GCFB의 감독되지 않은 계층적 클러스터링을 수행하였고, EAC 및 GCFB 사이의 확연한 클러스터 분류를 확인하였다 (도 6).
- [0091] 예측적 분류 모델은 400개 유전자 분류와 함께 BCCP에 기반하여 개발하였고, LOOCV로 학습되었다(trained). BCCP 점수를 이용한 ROC 커브는 0.957의 선 아래 영역(area under curve)과 (95% 신뢰구간=0.93-0.98), EAC 및 GCFB를 구별하는 컷-오프 값으로서 0.4535 Youden 인덱스를 밝혀내었다 (도 7).
- [0092] 상기 컷-오프 값은 EAC를 예측하는데 90.2%의 민감도 및 89.7%의 특이도를 나타내었다. 400 시그니처 유전자에 대해서, KEGG 경로 분석을 수행하였다. 5개 이상의 유전자가 포함되는 여러 암-연관 경로 중 GCFB에 관련된 PI3K-AKT 신호 경로를 확인하였으며, 이에 GCFB에서 과발현되는 200개 유전자 중 CHRM2, COMP, FGF14, IGF1, PPP2R2B, RELN, 및 THBS4가 포함되어 있었다. 결과적으로 PI3K 및 AKT가 TCGA 코호트의 RPPA 및 SNU 코호트의 조직 마이크로레이를 이용한 단백질 검증에 고려되었다.
- [0094] (3) 체세포 돌연변이 분석을 이용한 예측적 분류 모델의 검증
- [0095] 0.4535의 컷-오프 값을 갖는 BCCP 스코어를 이용하여, TCGA 코호트의 GEJ/분문부의 클러스터링을 테스트하였다 (도 8).
- [0096] TCGA 코호트의 GEJ/분문부의 계층적 클러스터링은 EAC-예측 및 GCFB-예측 그룹간의 클러스터의 스펙트럼 전이(spectral transition)를 나타냈고, 완전히 구분가능한 클러스터는 없었다. EAC로 예측된 TCGA 코호트의 GEJ/분문부는 15/48(31.2%)였고, GCFB로 예측된 것은 33/48(68.8%)였다. 체세포 돌연변이 측면에서, TCGA 코호트의 EAC-예측 그룹 및 GCFB-예측 그룹 사이에서 TP53, PIK3CA, RHOA, KRAS, 및 ARID1A의 유의한 차이는 없었다. SNU 코호트의 AGEJ II, AGEJ III, 및 UT의 클러스터링을 시험한 결과, EAC-예측 그룹 및 GCFB-예측 그룹 사이에서, TCGA 코호트와 유사한 클러스터의 스펙트럼 전이를 보였다 (도 9).
- [0097] SNU 코호트의 AGEJ II는 5/16(31.2%)가 EAC-예측 그룹으로 분류되었고, 11/16(68.8%)가 GCFB-예측 그룹으로 분류되었다. 특히, AGEJ III의 15/16(93.7%)가 GCFB-예측 그룹으로 분류되었다. SNU 코호트의 AGEJ II 및 III을 함께 고려할 때, EAC-예측 그룹과 GCFB-예측 그룹은 각각 6/32(18.8%) 및 26/32(81.2%)였다. SNU 코호트의 EAC-예측 그룹 및 GCFB-예측 그룹 사이에서 TP53, PIK3CA, RHOA, KRAS, 및 ARID1A의 유의한 차이는 없었다. 특히, TCGA 및 SNU 코호트 모두의 EAC-예측 그룹에서 RHOA, KRAS 및 PIK3CA의 체세포 돌연변이는 발견되지 않았다.
- [0099] (4) EAC-예측 및 GCFB-예측 그룹의 병리학적 분석
- [0100] SNU 코호트의 병리학적 특성 분석 결과, GEJ를 포함하는 모든 AGEJ III 및 GEJ를 포함하지 않는 AGEJ III의 80%가 GCFB-예측 그룹으로 분류됨을 확인하였다 (표 3).

[0102]

SNU 코호트의 EAC-예측 및 GCFB-예측 그룹 사이의 병리학적 특성

		EAC-예측 (n=10)	GCFB-예측 (n=36)	P Value
위치	AGEJ II	5(31.3%)	11(68.8%)	0.231
	GEJ를 포함하는 AGEJ III	0	11(100%)	
	GEJ를 포함하지 않는 AGEJ III	1(20.0%)	4(80.0%)	
	UT	4(28.6%)	10(71.4%)	
WHO	분화	8(80.0%)	13(36.1%)	0.043
	비분화	2(20.0%)	18(50.0%)	
	결정x	0	5(13.9%)	
Lauren	장	8(80.0%)	10(27.8%)	0.009
	확산형	2(20.0%)	16(44.4%)	
	혼합형	0	10(27.8%)	

[0103]

EAC-예측 그룹 및 GCFB-예측 그룹의 분포는 AGEJ II, AGEJ III 및 UT간에 유의한 차이가 없었다. 그러나, EAC-예측 그룹은 분화(differentiated) 및 장(intestinal) 유형의 비율이 유의하게 높은 반면, GCFB-예측 그룹은 비분화 및 확산(diffuse) 유형의 비율이 유의하게 높았다.

[0105]

(5) EAC-예측 및 GCFB-예측 그룹 사이의 복제 수 분석

[0106]

TCGA 코호트에서 게놈 수준의 복제 수 분석을 수행하였다. 그 결과, GISTIC 알고리즘에 의한 EAC-예측 및 GCFB-예측 그룹 사이에서 유의한 복제 수 차이가 있는 435개의 증폭된 유전자를 확인하였다(≥ 2 배 변화 및 $P < 0.05$).

[0107]

이들 435개의 유전자를 human Cancer Gene Census (<http://www.sanger.ac.uk/science/data/cancer-gene-census>)로 필터링한 결과, 6개의 암-관련 유전자를 밝혀내었고 이들은 다음과 같다 (도 10): COX6C (8q22.2, 전좌), HNRNPA2B1 (7p15.2, 전좌), NDRG1 (8q24.22, 전좌), RECQL4 (8q24.3, 넌센스, 프레임시프트/스플라이스), TCEA1 (8q11.23, 전좌), 및 TFEB (6p21.1, 전좌).

[0108]

SNU 코호트에서, EAC-예측 그룹과 GCFB-예측 그룹 사이의 GISTIC 알고리즘을 이용하여 국소적 증폭(focal amplification)을 가질 것으로 추정되는 유전자들을 비교한 결과, 유의한 복제 수 차이를 갖는 37개 유전자를 확인하였다 ($P < 0.05$)(표 4).

표 4

[0110]

SNU 코호트 중 EAC 및 GCFB 사이에서 유의한 복제 수 차이를 나타내는 유전자 ($P < 0.05$)

	EAC-유사 그룹의 평균 Log2복제 수	GCFB-유사 그룹의 평균 Log2복제 수
BOP1	0.332	0.026
C19orf12	0.279	0.039
DUSP8	0.267	0.002
EGFR	0.460	0.119
ERBB2	1.185	0.227
FOXP4	0.303	0.052
GRB7	0.826	0.219
GSTA1	0.252	0.001
GSTA2	0.320	-0.021
GSTA3	0.273	0.030
GSTA5	0.275	0.012
HIST1H1B	0.228	-0.051
HIST1H2AI	0.245	0.009
HIST1H2AK	0.234	-0.016
HIST1H2AL	0.282	-0.027
HIST1H2AM	0.281	-0.043

HIST1H2BM	0.215	-0.007
HIST1H2BN	0.230	-0.007
HIST1H2BO	0.239	-0.023
HIST1H2BN	0.230	-0.007
HIST1H2BO	0.239	-0.023
HIST1H3H	0.266	0.019
HIST1H3J	0.303	0.005
HIST1H4J	0.254	0.017
LILRA3	0.276	-0.123
LOC100287704	0.399	-0.011
LY86	0.236	-0.045
MDF1	0.381	0.047
MIEN1	1.178	0.205
OR2B2	0.209	-0.019
PI4KAP1	0.284	-0.010
PLEKHF1	0.343	0.048
POP4	0.289	0.053
SSR1	0.216	-0.002
TFEB	0.431	0.075
TMEM191B	0.381	-0.013
TRAM2	0.293	0.041
UGT2B17	0.263	-0.049
VSTM2B	0.270	0.069
ZNF439	-0.242	0.023

[0111] 37개 유전자 중, humanCancer Gene Census (<http://www.sanger.ac.uk/science/data/cancer-gene-census>)을 이용하여 필터링 한 결과 2 개의 암 연관 유전자를 밝혀내었다: ERBB2 (17q12, 증폭), 및 TFEB (6p21.1, 전좌) (도 11).

[0112] SNU 코호트에서 ERBB1(EGFR)(7p11.2)은 EAC-예측 그룹 및 GCFB-예측 그룹에서 동시에 증폭되었으나, EGFR의 복제 수는 상기 두 그룹 사이에서 유의한 차이를 나타내지 않았다. 모든 코호트의 COX6C, HNRNPA2B1, NDRG1, RECQL4, TCEA1, 및 TFEB의 주석된(annotated) 돌연변이 패턴이 복제 수 증폭에 일관되지 않았기 때문에, 이의 가능한 헤테로다이머로서 ERBB2 및 ERRB1이 TCGA 코호트의 RPPA 및 SNU 코호트의 조직 마이크로어레이를 이용하여 평가되었다.

[0114] (6) 역상 단백질 어레이(RPPA) 및 조직 마이크로 어레이(TMA)

[0115] TCGA 코호트의 44 EAC 및 88 GCFB로 구성된 RPPA 데이터의 감독된(supervised) 분석을 통해, EAC 및 GCFB 사이에서 81개 단백질의 명백히 분리되는 발현 클러스터를 확인하였다 (도 12).

[0116] 81 단백질 중, 400개 시그니처 유전자의 경로 분석에서 확인된 PIK3CA 및 AKT1과, 복제 수 분석에서 확인된 ERBB2 및 EGFR이, RPPA의 단백질 발현에서 EAC 및 GCFB 사이에서 유의한 차이를 나타내었다.

[0117] 추가적인 검증을 위해 상업적으로 판매되는 항체를 이용하여 SNU 코호트의 TMA 3세트를 이용하여 상기 4종 단백질의 발현을 분석하였다(표 5).

표 5

[0119] 조직 마이크로어레이에 사용된 항체 정보

항체	클론성 (clonality)	희석	검출키트	출처	Cat.no
EGFR	마우스 모노클론성	바로 사용	OptiView Polymer (Ventana)	Roche	790-2988

ERBB2	래빗 모노클론성	바로 사용	OptiView Polymer (Ventana)	Ventana medical systems	790-2991
PI3KinaseP110alpha	래빗 모노클론성	1:100	OptiView Polymer (Ventana)	Cell signalling	#4249
AKT1	래빗 모노클론성	1:50	OptiView Polymer (Ventana)	Abcam	ab32505

[0120] TMA에서 EGFR, ERBB, PI3Kinasep110alpha, 및 AKT1의 염색 패턴은 도 13과 같다. 3개의 TMA 세트의 발현 결과를 이용하여 EGFR, PI3Kinasep110alpha, AKT1의 Complex H 점수를 계산한 결과, EGFR의 평균 H 점수는 EAC-예측 그룹에서 GCFB-예측 그룹보다 유의하게 증가하였다 (160.7 ± 108.8 in EAC-like vs. 105.6 ± 81.6 in GCFB-like, $P=0.014$, 도 14). ERBB2에 대한 IHC(immunohistochemistry) 염색 결과, ERBB2-양성이 EAC-예측 그룹에서 GCFB-예측 그룹보다 더 높은 점수를 나타내는 경향이 확인되었다 (표 6). 그러나 PI3Kinase와 AKT1의 유의한 발현차이는 관찰되지 않았다.

표 6

[0122] ERBB2의 IHC 및 SISH

		EAC-예측 (n=10)	GCFB-예측 (n=36)	P value
IHC	0	3(30.0%)	17(47.2%)	0.081
	1+	2(20.0%)	14(38.9%)	
	2+	1(10.0%)	2(5.6%)	
	3+	4(40.0%)	3(8.3%)	
IHC 및SISH	IHC<2+, 또는 IHC2+ 및 흑/적 비율<2.0	5(50.0%)	32(88.9%)	0.015
	IHC 3+, 또는 IHC 2+, 및 흑/적 비율 ≥ 2.0	5(50.0%)	4(11.1%)	

[0123] IHC 및 SISH를 함께 비교했을 때, EAC-예측 그룹은 GCFB-예측 그룹에 비교하여 ERBB2가 유의하게 높은 양성 (IHC 3+, 또는 IHC 2+, 및 SISH ≥ 2.0 의 흑/적 비율)을 나타내었다 (50.0% EAC-예측 vs. 11.1% GCFB-예측, $P=0.015$). 표 2의 단변량 분석으로부터의 모든 유의한 변화를 다변량 분석으로 분석하여 EGFR 및 ERBB2의 발현에 대한 위험 인자를 확인하였다. EGFR의 과발현에 대한 예측 유형(EAC-예측 또는 GCFB-예측)은 조정 결정 계수 R^2 ($P=0.034$)의 0.78의 독립 위험 인자가 유일하였다 (표 7).

표 7

[0125] EGFR의 과발현에 대한 다변량 분석

변화	비표준화 계수 B±표준 편차	표준화 계수 β	t	P value	B의 95% 신뢰 구간
WHO 분류	1.822 ± 5.722	0.053	0.318	0.752	-9.733-13.378
Lauren 분류	26.389 ± 16.886	0.244	1.563	0.125	-7.665-60.443
신경주변 침 투	-38.504 ± 27.032	-0.220	-1.424	0.162	-93.057-16.046
예측 유형	62.500 ± 28.509	0.314	2.192	0.034	5.044-19.956

[0127] ERBB2 양성에 대해서는, 예측 유형 및 WHO 분류가 독립 위험 인자였다 (예측 유형에 대해 $P=0.049$ 및 분화 유형에 대해 $P=0.029$)(표 8).

표 8

[0129] ERBB2 양성에 대한 다변량 분석

변화		P value	승산비 (Odds ratio)	승산비의 95% 신뢰 구간
WHO 분류 (vs. 결정x)	분화	0.029	0.223	0.058-0.856
	비분화	0.002	0.036	0.004-0.309
Lauren 분류 (vs. 혼합형)	장	0.387	4.156	0.165-105.009
	확산	0.734	0.581	0.025-13.322
신경주변침투 (vs. 침투)	비-침투	0.576	0.532	0.058-4.870
예측 유형 (vs. GCFB-예측)	EAC-예측	0.049	6.179	1.1011-37.752

[0130] (7) CCLE 데이터베이스를 이용한 외부 검증

[0131] CCLE 데이터베이스로부터, 발현 마이크로어레이 데이터, SNP 어레이 데이터 및 라파티닙, EGFR 및 HER2 타이로신 카이네이즈 억제제 병용에 대한 IC50을 가지고 있는 식도(n=3) 및 위(n=38) 선암 세포주를 확인하였다. 각 샘플에 대해 사용 가능한 데이터는 표 9와 같다.

표 9

[0133] CCLE 데이터베이스의 식도 및 위 선암 세포주 정보

세포주	기관	BCCP 점수	예측	ERBB2 복제수*	EGFR 복제수*	IC50†
OE33	식도	0.546	EAC-예측	증폭	0	3.538
OE19	식도	0.402	GCFB-예측	증폭	0	N/A
JHESOAD1	식도	0.484	EAC-예측	N/A	N/A	N/A
FU97	위	0.109	GCFB-예측	결실	0	8.000
NUGC3	위	0.37	GCFB-예측	0	0	2.411
IM95	위	0.318	GCFB-예측	0	0	8.000
AGS	위	0.19	GCFB-예측	0	0	N/A
KATOIII	위	0.536	EAC-예측	0	0	N/A
SNU16	위	0.351	GCFB-예측	0	0	6.698
NCIN87	위	0.753	EAC-예측	증폭	0	0.066
OCUM1	위	0.347	GCFB-예측	0	0	8.000
SNU5	위	0.291	GCFB-예측	0	0	N/A
GCIY	위	0.169	GCFB-예측	0	0	7.255
SH10TC	위	0.152	GCFB-예측	0	0	8
MKN1	위	0.341	GCFB-예측	0	0	N/A
MKN74	위	0.36	GCFB-예측	0	증폭	4.69
KE39	위	0.211	GCFB-예측	증폭	0	4.056
HGC27	위	0.062	GCFB-예측	0	0	8
HUG1N	위	0.315	GCFB-예측	0	0	N/A
NUGC4	위	0.313	GCFB-예측	증폭	증폭	0.172
RERFGC1B	위	0.365	GCFB-예측	0	0	8
HS746T	위	0.143	GCFB-예측	0	0	8
NUGC2	위	0.531	EAC-예측	0	0	N/A
SNU1	위	0.176	GCFB-예측	0	0	8
MKN45	위	0.341	GCFB-예측	0	증폭	8

X2313287	위	0.509	EAC-예측	N/A	N/A	N/A
MKN7	위	0.272	GCFB-예측	증폭	0	8
SNU216	위	0.37	GCFB-예측	증폭	0	N/A
AZ521	위	0.097	GCFB-예측	0	0	1.66
LMSU	위	0.129	GCFB-예측	0	0	N/A
ECC10	위	0.153	GCFB-예측	0	0	N/A
TGBC11TKB	위	0.326	GCFB-예측	0	0	N/A
SNU520	위	0.297	GCFB-예측	0	0	N/A
GSS	위	0.223	GCFB-예측	0	증폭	N/A
SNU620	위	0.322	GCFB-예측	0	0	N/A
ECC12	위	0.074	GCFB-예측	0	0	N/A
GSU	위	0.388	GCFB-예측	0	0	N/A
SNU601	위	0.507	EAC-예측	0	0	N/A
SNU668	위	0.144	GCFB-예측	0	0	N/A
NCCSTCK140	위	0.816	EAC-예측	0	0	N/A
SNU719	위	0.305	GCFB-예측	0	증폭	N/A

[0134] *0은 복제 수가 변경되지 않았음을, N/A는 사용 가능하지 않음을 나타낸다. † IC50은 라파티닙에 대해서 최대 억제 농도의 절반을 나타낸다.

[0136] 이러한 세포주들을 이용하여 CCLE 데이터베이스의 RNA 마이크로어레이 데이터를 이용한 외부 검증으로, 식도 및 위 선암 세포주 사이에서 BCCP 점수의 유의한 차이를 확인하였다(Wilcoxon Rank Sumtest 사용, $P=0.031$)(도 15).

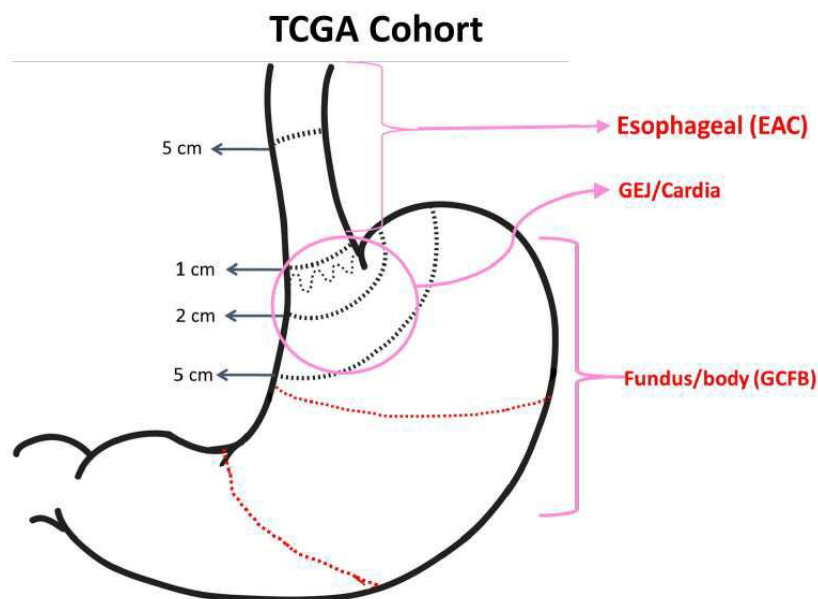
[0137] BCCP 스코어를 이용한 CCLE 데이터베이스의 계층적 클러스터링으로 EAC-예측 그룹 및 GCFB-예측 그룹 사이에서 조직 기원(식도 또는 위), ERBB2 증폭, 또는 EGFR 증폭에 따른 유의한 차이는 없는 것을 밝혀냈다(도 16).

[0138] 라파티닙, EGFR 및 HER2 타이로신 카이네이즈 억제제 병용에 대한 표적 약물 반응은 EAC-예측($n=2$) 및 GCFB-예측 그룹($n=17$) 사이에 대해서 CCLE 데이터베이스의 IC50 데이터를 이용하여 평가하였다 (도 17).

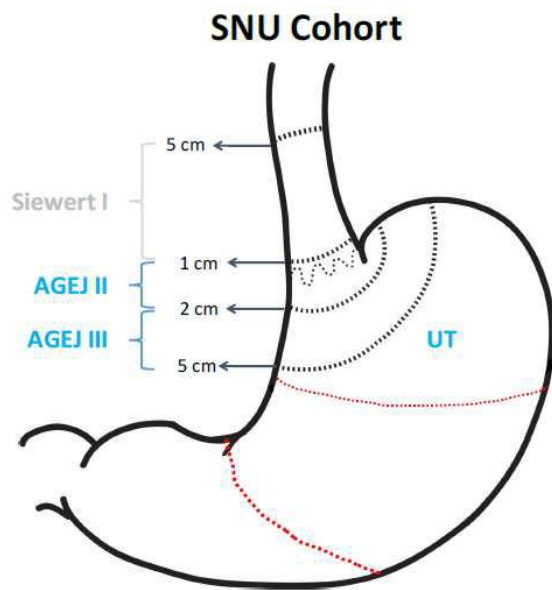
[0139] 그 결과, Wilcoxon Rank Sumtest($P=0.044$)를 이용하여, GCFB-예측 그룹보다 EAC-예측 그룹에서 IC50이 유의하게 낮게 나타남을 확인하였다.

도면

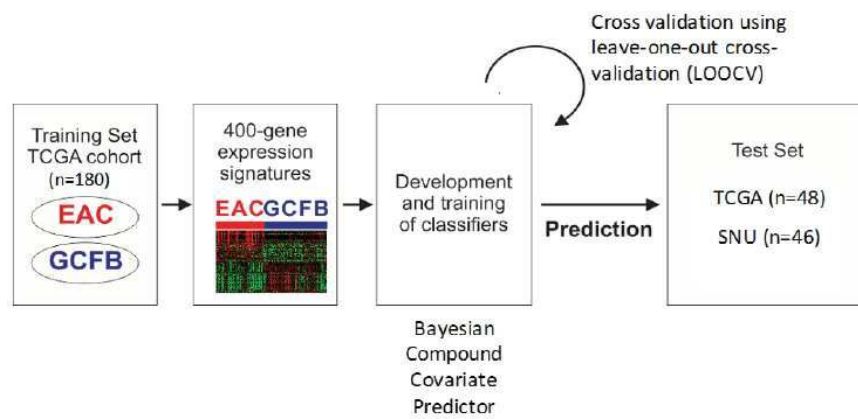
도면1



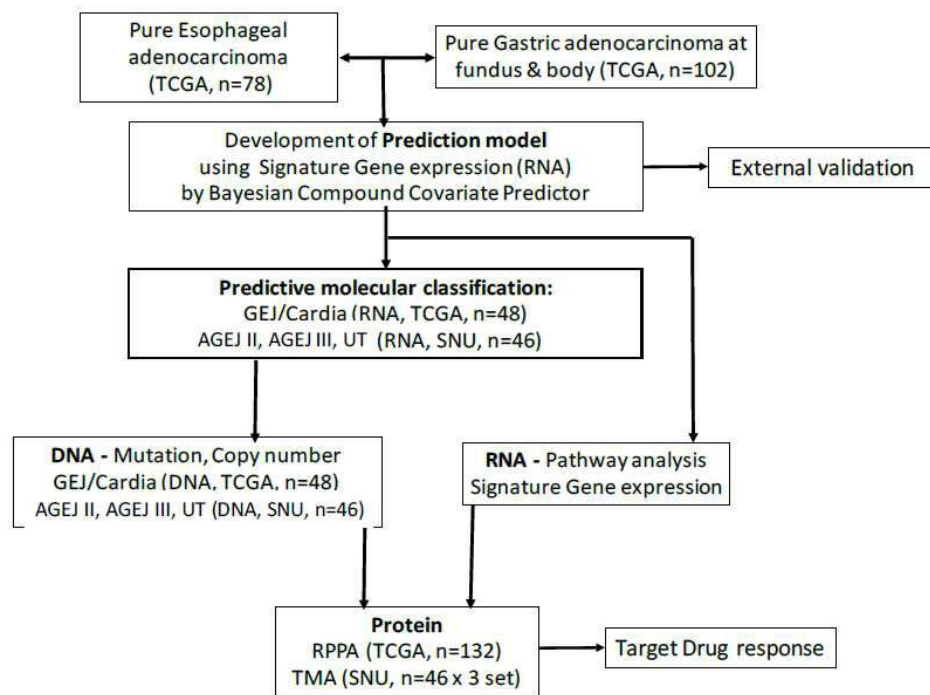
도면2



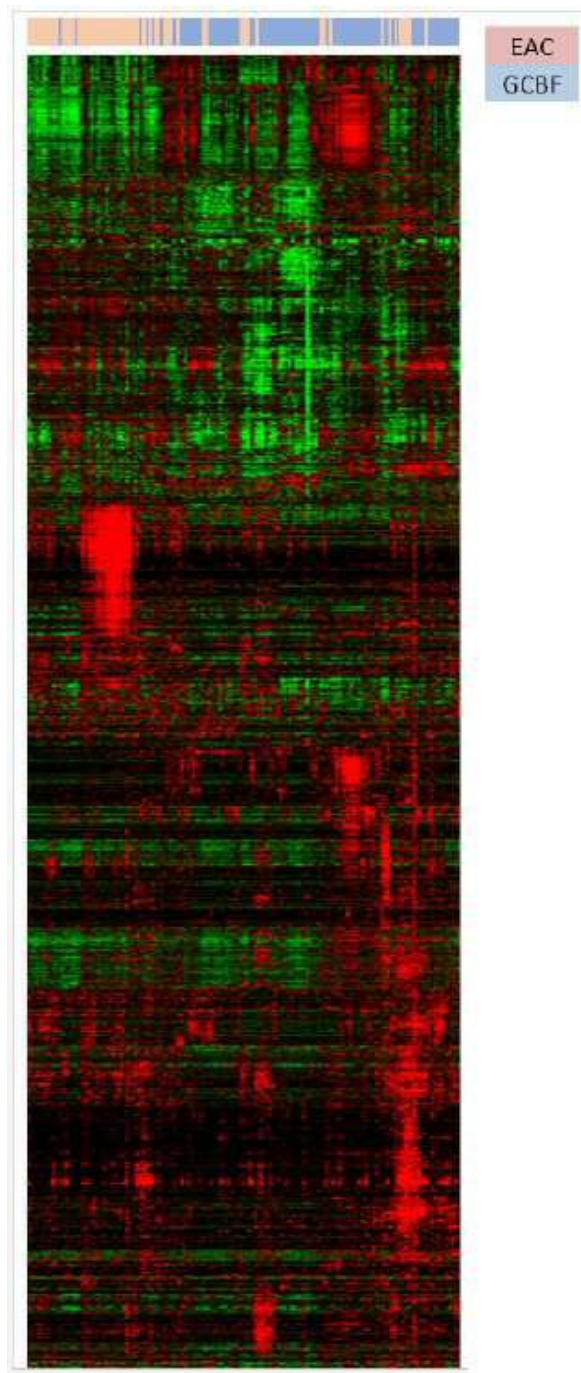
도면3



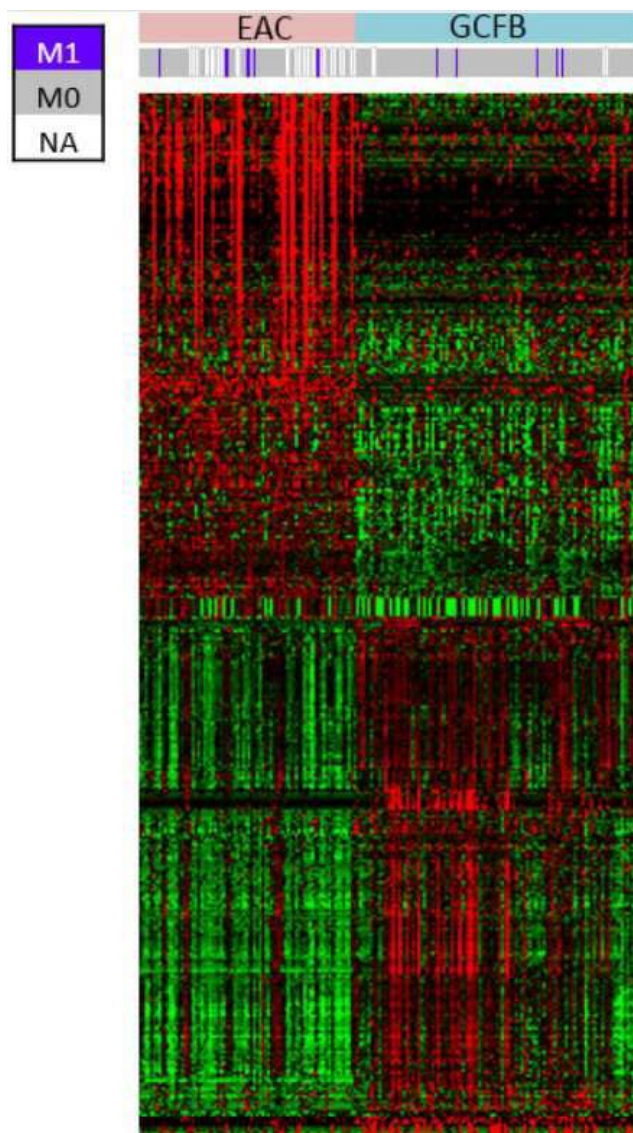
도면4



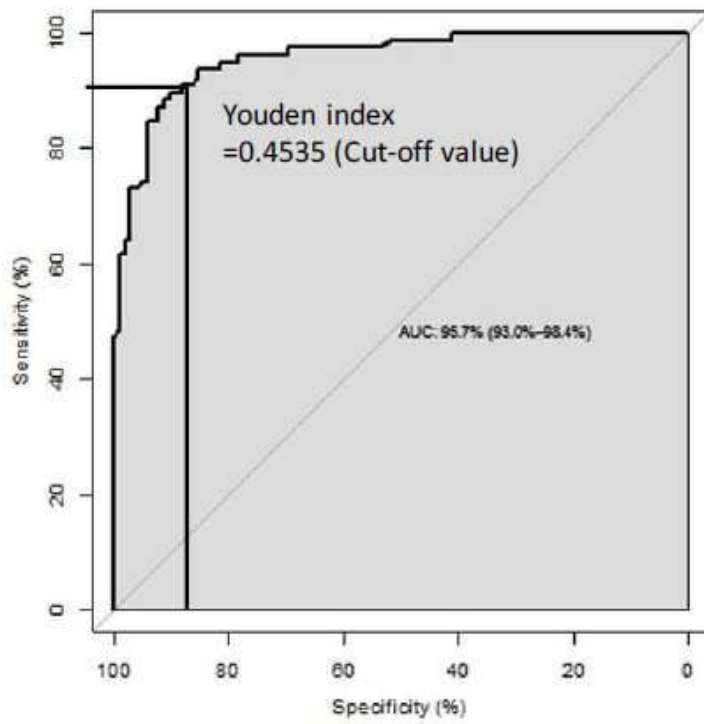
도면5



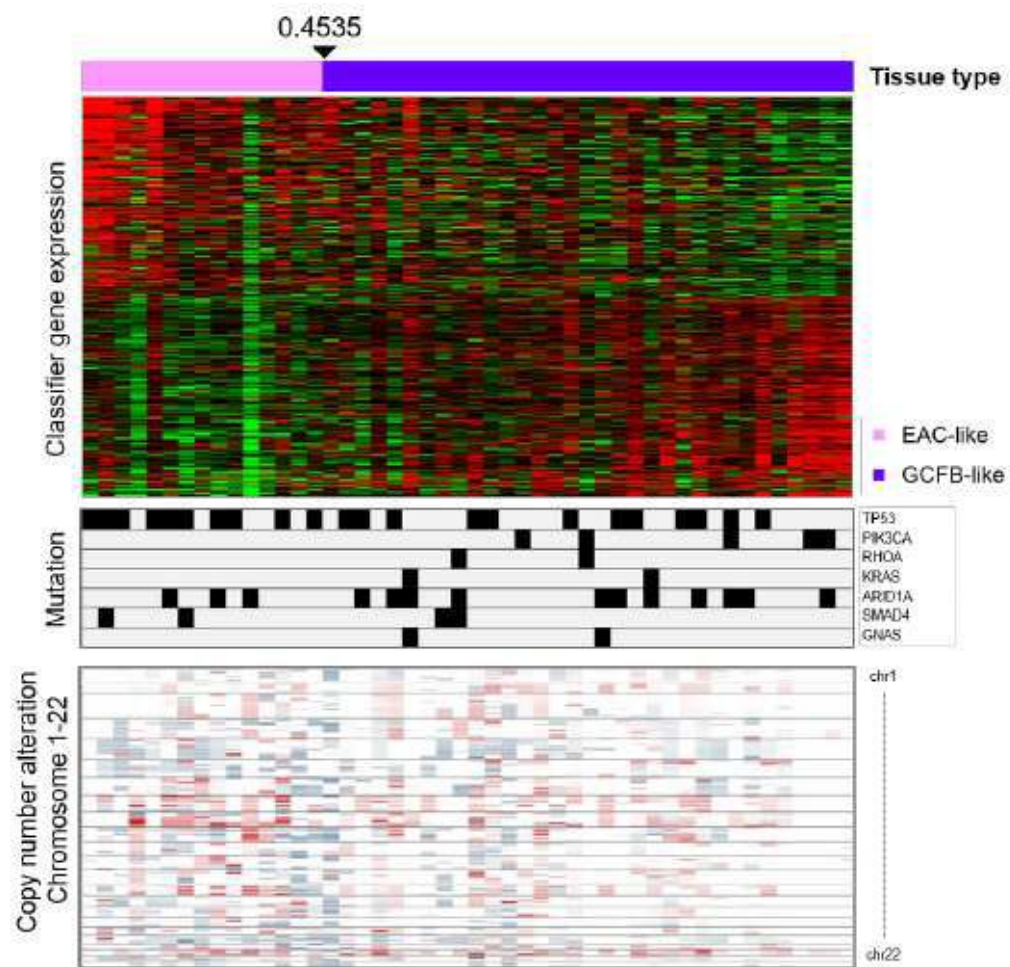
도면6



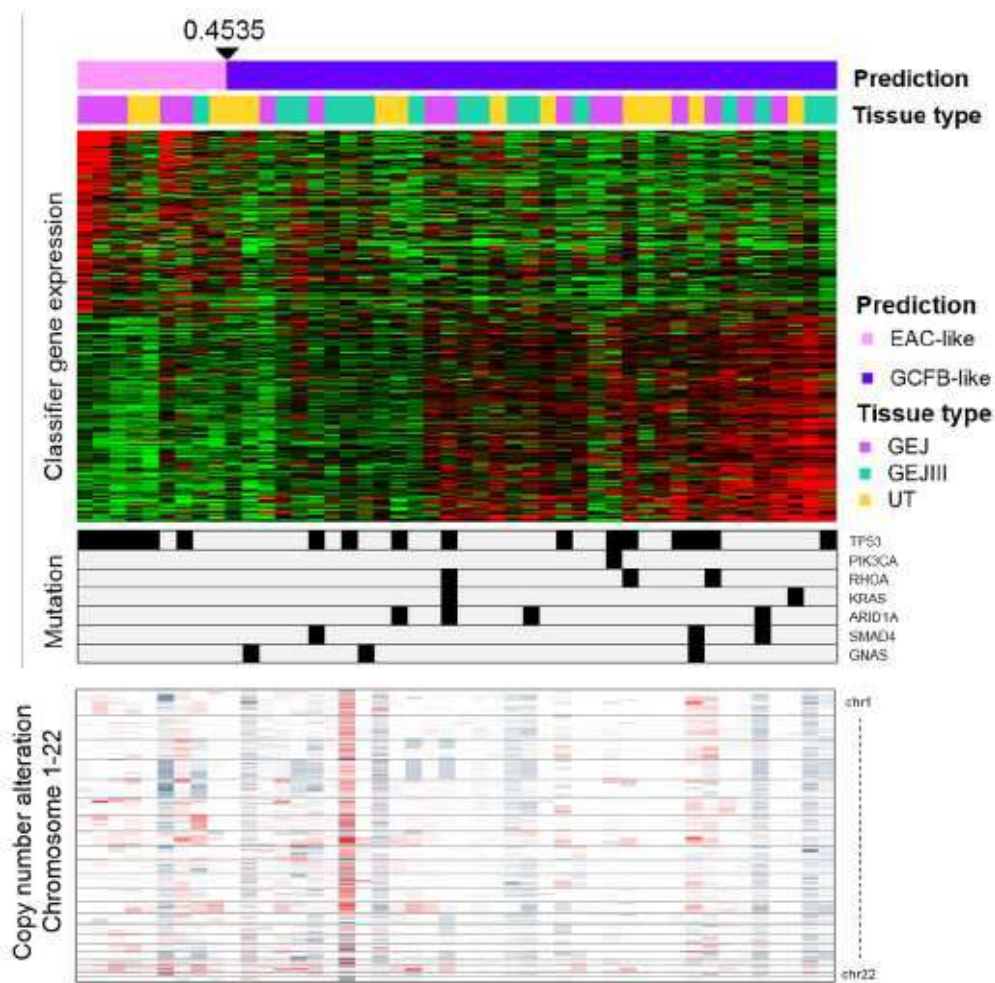
도면7



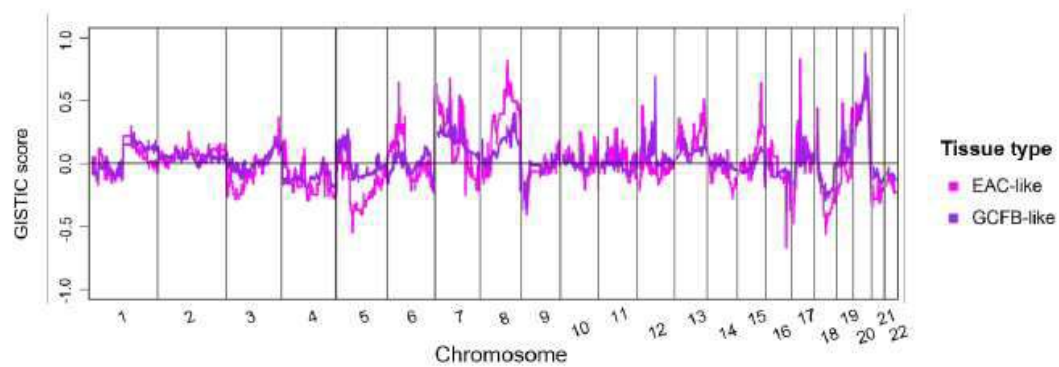
도면8



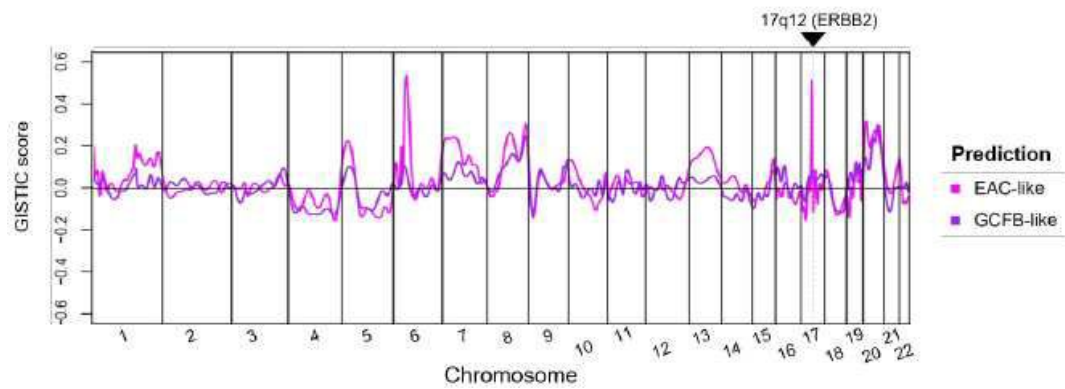
도면9



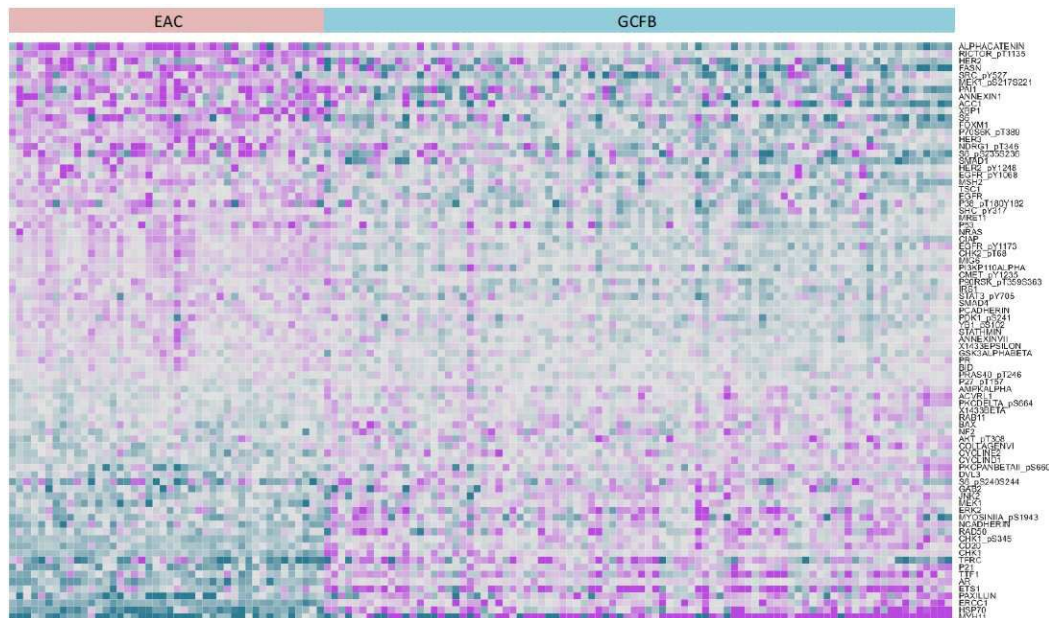
도면10



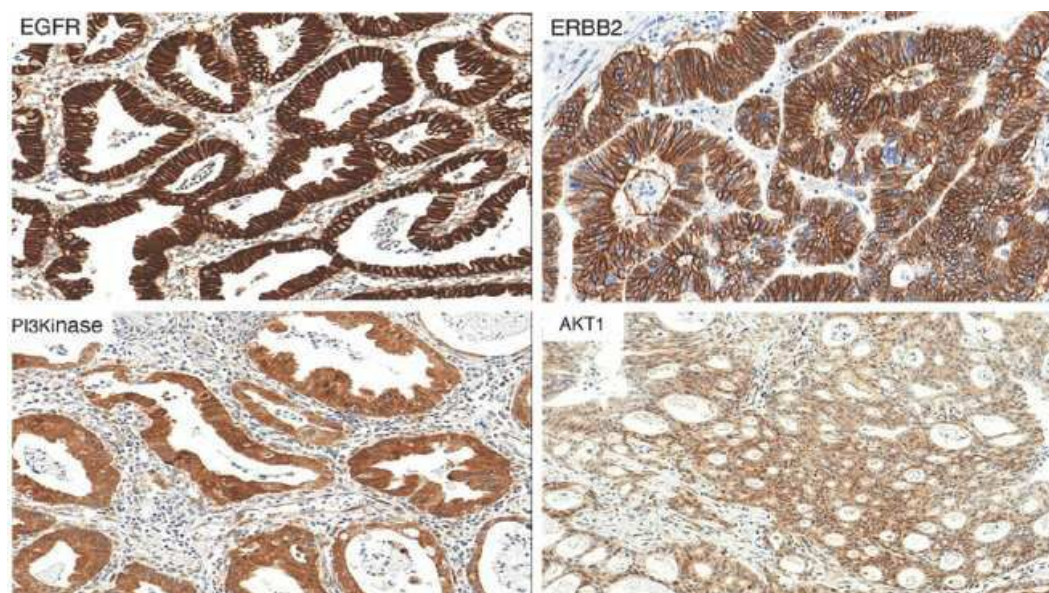
도면11



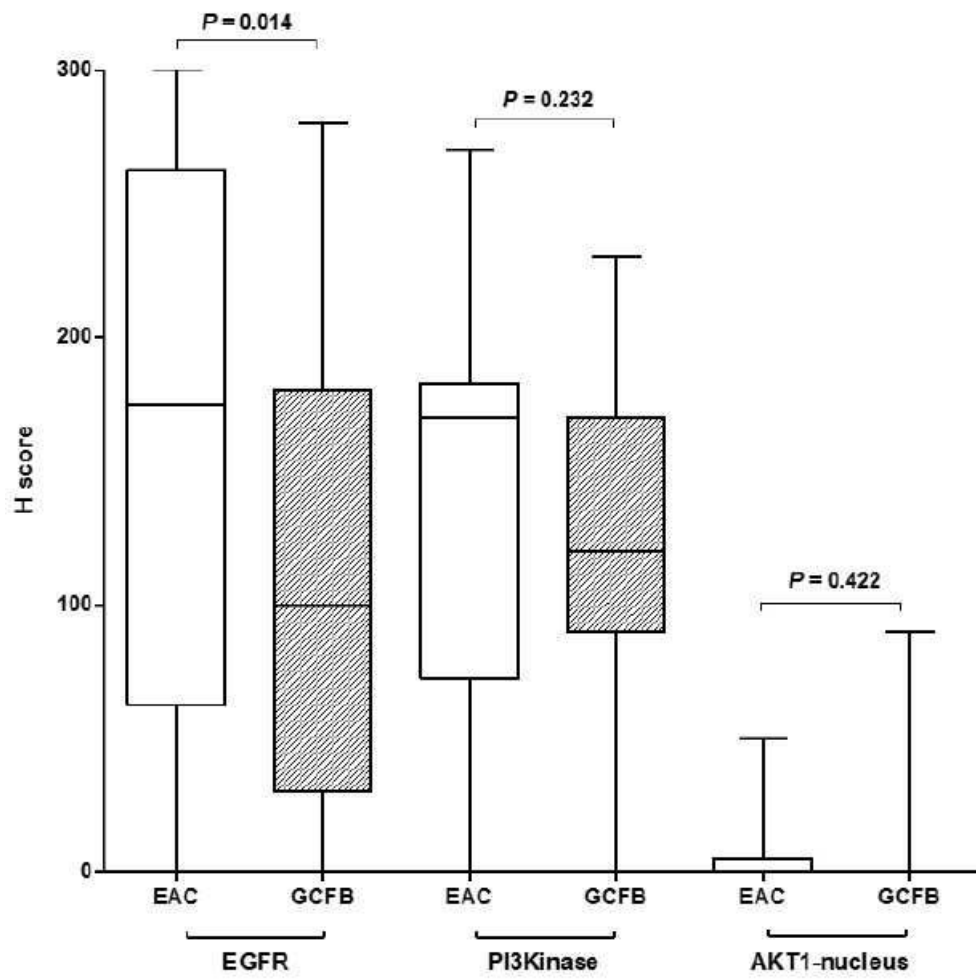
도면12



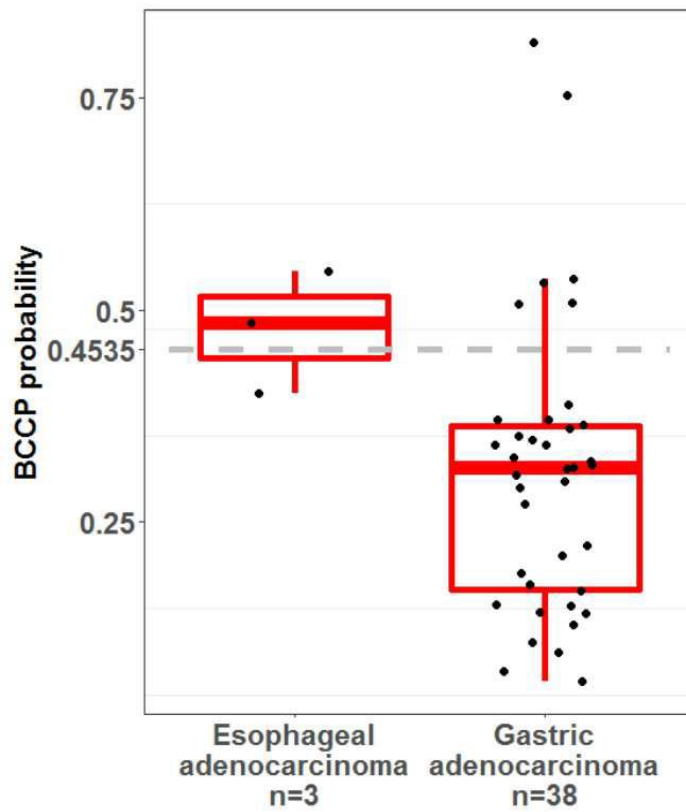
도면13



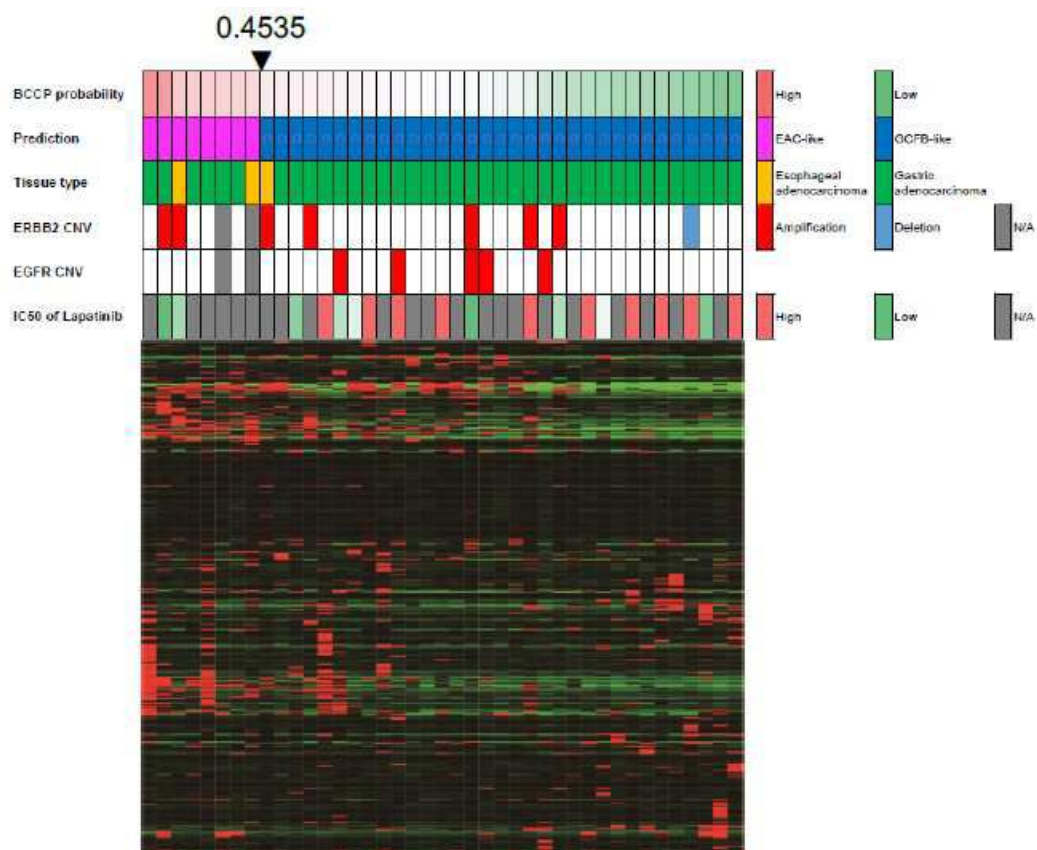
도면14



도면15



도면16



도면17

