Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2941769 A1 2015/09/17

(21) 2 941 769

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) **Date de dépôt PCT/PCT Filing Date:** 2015/03/11
(87) **Date publication PCT/PCT Publication Date:** 2015/09/17
(85) **Entrée phase nationale/National Entry:** 2016/09/07
(86) **N° demande PCT/PCT Application No.:** AU 2015/050096
(87) **N° publication PCT/PCT Publication No.:** 2015/135035
(30) **Priorités/Priorities:** 2014/03/11 (AU2014900813);
2014/04/03 (AU2014901212);
2014/11/21 (AU2014904716)

(51) **Cl.Int./Int.Cl.** ***C12Q 1/68*** (2006.01)

(71) **Demandeur/Applicant:**
THE COUNCIL OF THE QUEENSLAND INSTITUTE OF MEDICAL RESEARCH, AU

(72) **Inventeur/Inventor:**
AL-EJEH, FARES, AU

(74) **Agent:** SIM & MCBURNEY

(54) **Titre : DETERMINATION DE L'AGRESSIVITE D'UN CANCER, DE SON PRONOSTIC ET DE SA SENSIBILITE A UN TRAITEMENT**
(54) **Title: DETERMINING CANCER AGGRESSIVENESS, PROGNOSIS AND RESPONSIVENESS TO TREATMENT**

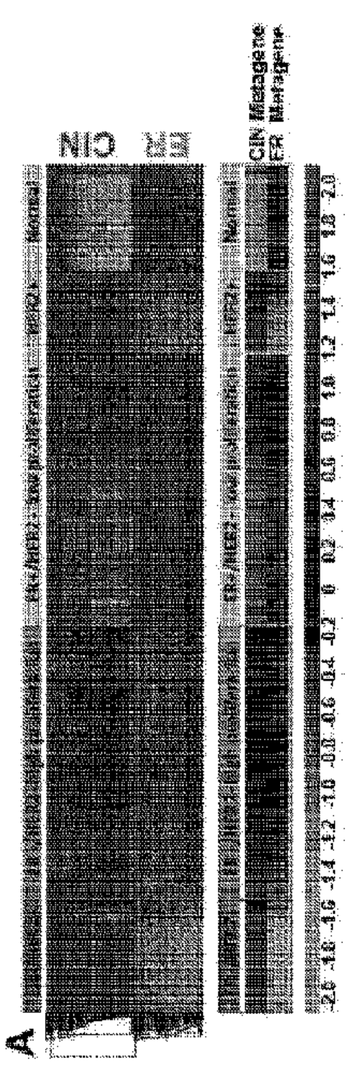
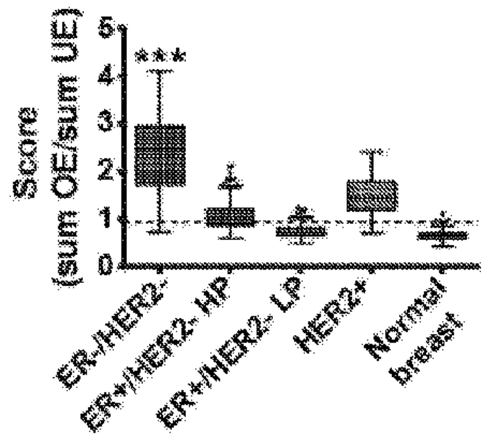


FIG. 1



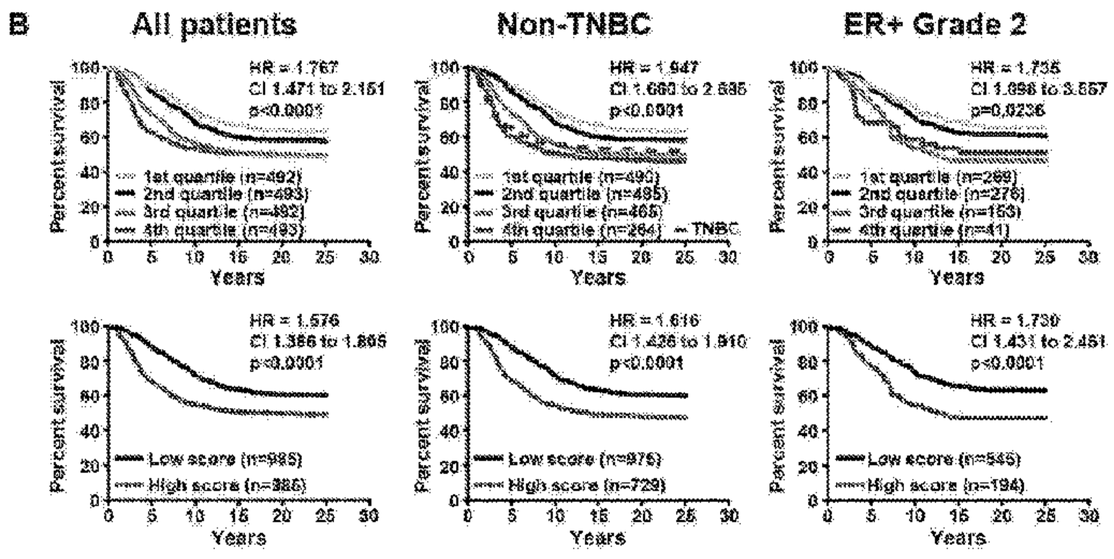


FIG. 1 cont'd

(57) **Abrégé/Abstract:**
The invention provides methods of determining the aggressiveness, prognosis and response to therapy for particular cancers, which include comparing the expression levels of one or a plurality of differentially expressed genes from one or more 5 functional metagenes, including a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune system metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein 10 Synthesis/Modification metagene and a Multiple Networks metagene. The method disclosed herein may be particularly suitable as a companion diagnostic for cancer therapies.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 September 2015 (17.09.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/135035 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/AU2015/050096

(22) International Filing Date:
11 March 2015 (11.03.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

2014900813	11 March 2014 (11.03.2014)	AU
2014901212	3 April 2014 (03.04.2014)	AU
2014904716	21 November 2014 (21.11.2014)	AU

(71) Applicant: **THE COUNCIL OF THE QUEENSLAND INSTITUTE OF MEDICAL RESEARCH** [AU/AU]; The Bancroft Centre, 300 Herston Road, Herston, Queensland 4029 (AU).

(72) Inventor: **AL-EJEH, Fares**; c/o- THE COUNCIL OF THE QUEENSLAND INSTITUTE OF MEDICAL RESEARCH, The Bancroft Centre, 300 Herston Road, Herston, Queensland 4029 (AU).

(74) Agent: **FISHER ADAMS KELLY PTY LTD**; Level 29, 12 Creek Street, Brisbane, Queensland 4000 (AU).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) **Title**: DETERMINING CANCER AGRESSIVENESS, PROGNOSIS AND RESPONSIVENESS TO TREATMENT

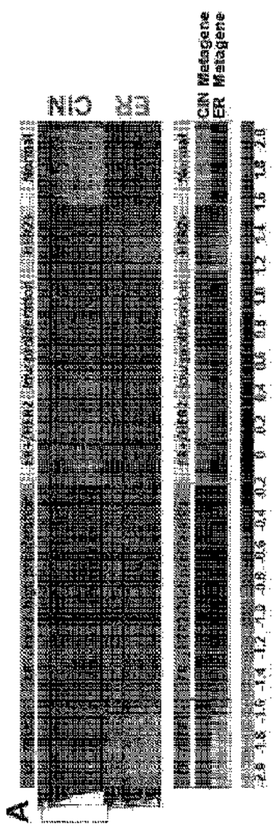


FIG. 1

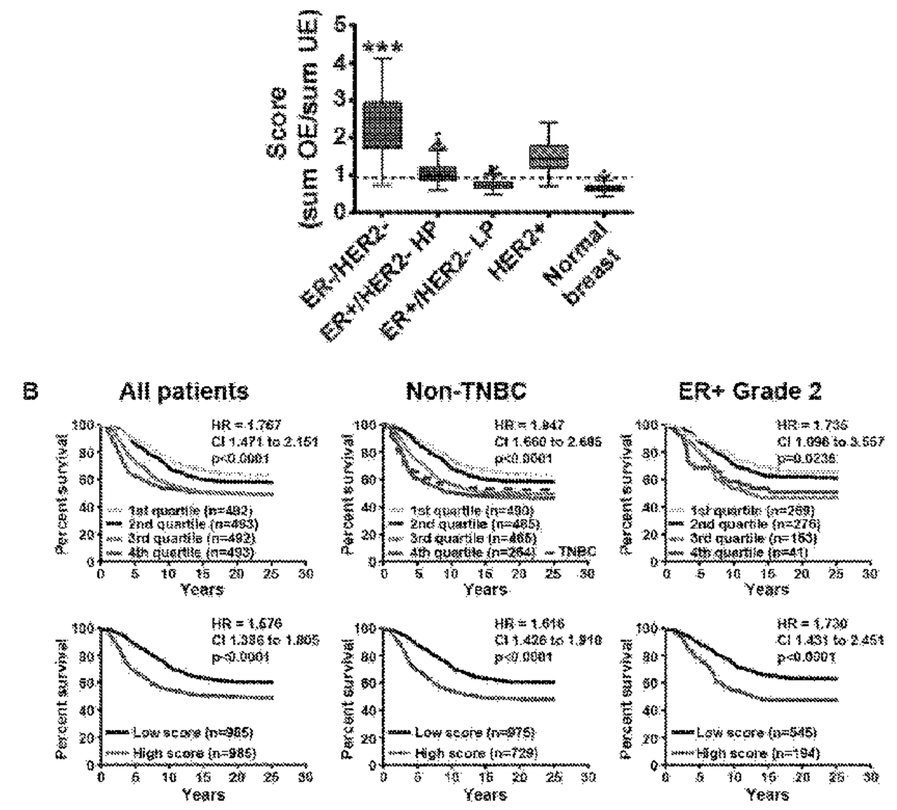


FIG. 1 cont'd

(57) **Abstract**: The invention provides methods of determining the aggressiveness, prognosis and response to therapy for particular cancers, which include comparing the expression levels of one or a plurality of differentially expressed genes from one or more 5 functional metagenes, including a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune system metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein 10 Synthesis/Modification metagene and a Multiple Networks metagene. The method disclosed herein may be particularly suitable as a companion diagnostic for cancer therapies.

WO 2015/135035 A2

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 181

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 181

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

TITLE

DETERMINING CANCER AGGRESSIVENESS, PROGNOSIS AND RESPONSIVENESS TO TREATMENT

FIELD

THIS INVENTION relates to cancer. More particularly, this invention relates to methods of determining the aggressiveness of cancers, prognosis of cancers and/or predicting responsiveness to anti-cancer therapy.

BACKGROUND

Hormone receptors (ER and PR) and HER2 are standard biomarkers used in clinical practice to aid the histopathological classification of breast cancer and management decisions. Hormone receptor (HR)- and HER2- positive tumors benefit from tamoxifen and anti-HER2 therapies, respectively. On the other hand, there are currently no targeted drug therapies for management of triple negative breast cancer (TNBC), which lacks expression of HR/HER2. TNBCs are more sensitive to chemotherapy than HR-positive tumors because they are generally more proliferative, and pathological complete responses (pCR) after chemotherapy are more likely in TNBC than in non-TNBC[1,2]. Paradoxically, TNBC is associated with poorer survival than non-TNBC, due to more frequent relapse in TNBC patients with residual disease[1,2]. Only 31% of TNBC patients experience pCR after chemotherapy[3], emphasizing the need for targeted therapies.

Transcriptome profiling has been used to dissect the heterogeneity of breast cancer into five intrinsic 'PAM50' subtypes; Luminal A, Luminal B, Basal-like, HER-2 and normal-like subtypes that relate to clinical outcomes[4-8]. Several gene signatures have been developed to predict outcome or response to treatment including: MammaPrint[9], OncotypeDx[10,11], Theros[12-15]. These commercial signatures rely on models that select genes based on clinical phenotypes such as tumor response or survival time. Notwithstanding their clinical utilities, these models fail to identify core biological mechanisms for the phenotypes of interest. Recently, an approach based on biological function-driven gene coexpression signatures, "attractor metagenes", has been applied to the prediction of survival in certain cancers. However such approaches are at an early stage and much work needs to be done to develop this attractor metagene analysis in relation to cancers in general and also for specific cancers.

SUMMARY

The present invention relates to the comparison of expression levels of a plurality of differentially expressed genes from one or a plurality of functional metagenes, including a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune system metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene; wherein the comparison of expression level of a plurality of genes in these metagenes is used to facilitate determining the aggressiveness of certain cancers. This comparison may also, or alternatively, assist in providing a cancer prognosis for a patient. The invention also relates to predicting the responsiveness of a cancer to an anti-cancer treatment by determining an expression level of one or a plurality of genes associated with one or a plurality of the aforementioned twelve functional metagenes.

The invention further relates to the comparison of expression levels of a specific signature of differentially expressed proteins to facilitate or assist in determining the aggressiveness of a particular cancer, a prognosis for a cancer patient and/or predicting responsiveness to an anti-cancer treatment. One or both of these comparisons may also be integrated with the aforementioned comparison of the expression levels of the plurality genes from one or a plurality of the aforementioned functional metagenes in determining cancer aggressiveness, prognosis and/or treatment.

In a first aspect, the invention relates to a method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein

Synthesis/Modification metagene and a Multiple Networks metagene, wherein: a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

In a second aspect, the invention relates to a method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene, wherein: a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a more favourable cancer prognosis.

In one embodiment of the above aspects, the one or plurality of overexpressed genes and/or the one or plurality of underexpressed genes are selected from one of the aforesaid metagenes. In an alternative embodiment, the one or plurality of overexpressed genes and/or one or the plurality of underexpressed genes are selected from a plurality of the aforesaid metagenes.

Suitably, for the method of the above aspects the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-

Translational Modification metagene, the Protein Synthesis/Modification metagene and/or the Multiple Networks metagene comprise one or a plurality of genes listed in Table 21.

In a third aspect, the invention relates to a method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Metabolism metagene, a Signalling metagene, a Development and Growth metagene, a Chromosome Segregation/Replication metagene, an Immune Response metagene and a Protein Synthesis/Modification metagene, wherein: a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level

In a fourth aspect, the invention relates to a method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Metabolism metagene, a Signalling metagene, a Development and Growth metagene, a Chromosome Segregation/Replication metagene, an Immune Response metagene and a Protein Synthesis/Modification metagene, wherein: a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a more favourable cancer prognosis.

In one embodiment of the third and fourth aspects, the one or plurality of overexpressed genes and/or the one or plurality of underexpressed genes are selected

from one of the aforesaid metagenes. In an alternative embodiment, the one or plurality of overexpressed genes and/or the one or plurality of underexpressed genes are selected from a plurality of the aforesaid metagenes.

Suitably, the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene comprise one or a plurality of genes listed in Table 22.

In particular embodiments of the method of the third and fourth aspects, the one or plurality of overexpressed genes and/or the one or plurality of underexpressed genes are from one or a plurality of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene.

In a fifth aspect, the invention relates to a method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes associated with chromosomal instability and/or an expression level of one or a plurality of underexpressed genes associated with estrogen receptor signalling in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed genes associated with chromosomal instability compared to the one or plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level expression level of the one or plurality of overexpressed genes associated with chromosomal instability compared to the one or plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

In a sixth aspect, the invention relates to a method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes associated with chromosomal instability and/or an expression level of one or a plurality of underexpressed genes

associated with estrogen receptor signalling in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed genes associated with chromosomal instability compared to the one or plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed genes associated with chromosomal instability compared to the one or plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with a more favourable cancer prognosis.

In certain embodiments, the genes associated with chromosomal instability are of a CIN metagene. Non-limiting examples include genes selected from the group consisting of *ATP6V1C1, RAP2A, CALM1, COG8, HELLS, KDM5A, PGK1, PLCH1, CEP55, RFC4, TAF2, SF3B3, GPI, PIR, MCM10, MELK, FOXM1, KIF2C, NUP155, TPX2, TTK, CENPA, CENPN, EXO1, MAPRE1, ACOT7, NAE1, SHMT2, TCP1, TXNRD1, ADM, CHAF1A and SYNCRIP*. Preferably, the genes are selected from the group consisting of: *MELK, MCM10, CENPA, EXO1, TTK* and *KIF2C*.

In certain embodiments, the genes associated with estrogen receptor signalling are of an ER metagene. Non-limiting examples include genes selected from the group consisting of: *BTG2, PIK3IP1, SEC14L2, FLNB, ACSF2, APOM, BIN3, GLTSCR2, ZMYND10, ABAT, BCAT2, SCUBE2, RUNX1, LRRC48, MYBPC1, BCL2, CHPT1, ITM2A, LRIG1, MAPT, PRKCB, RERE, ABHD14A, FLT3, TNN, STC2, BATF, CD1E, CFB, EVL, FBXW4, ABCB1, ACAA1, CHAD, PDCD4, RPL10, RPS28, RPS4X, RPS6, SORBS1, RPL22 and RPS4XP3*. Preferably, the genes are selected from the group consisting of: *MAPT* and *MYB*.

In certain embodiments, the method of the fifth and sixth aspects further including the step of comparing an expression level of one or a plurality of other overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and/or an expression level of one or a plurality of other underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH,*

KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the other overexpressed genes compared to the other underexpressed genes indicates or correlates with higher aggressiveness of the cancer and/or a less favourable cancer prognosis; and/or a lower relative expression level of the other overexpressed genes compared to the other underexpressed genes indicates or correlates with lower aggressiveness of the cancer and/or a more favourable cancer prognosis compared to a mammal having a higher expression level.

In one embodiment, the one or plurality of other overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*, *CD55*, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593*.

In one embodiment, the one or plurality of other underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

Suitably, the comparison of the expression level of the overexpressed genes associated with chromosomal instability and/or the expression level of the underexpressed genes associated with estrogen receptor signalling is integrated with the comparison of the expression level of the one or plurality of other overexpressed genes and/or the expression level of the one or plurality of other underexpressed genes to derive a first integrated score.

In a seventh aspect, the invention provides a method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed

genes compared to the one or plurality of underexpressed genes indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

In an eighth aspect, the invention provides a method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a more favourable cancer prognosis compared to a mammal having a higher expression level.

In one embodiment of the seventh and eighth aspects, the one or plurality of overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*, *CD55*, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593*.

In one embodiment of the seventh and eighth aspects, the one or plurality of underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

In particular embodiments, the method of the first, second, third, fourth, fifth, sixth, seventh and eighth aspects further includes the step of comparing an

expression level of one or a plurality of overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and/or an expression level of one or a plurality of underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the overexpressed proteins compared to the underexpressed proteins indicates or correlates with higher aggressiveness of the cancer and/or a less favourable cancer prognosis; and/or a lower relative expression level of the overexpressed proteins compared to the underexpressed proteins indicates or correlates with lower aggressiveness of the cancer and/or a more favourable cancer prognosis compared to a mammal having a higher expression level.

Suitably, the comparison of the expression level of the one or plurality of overexpressed proteins and/or the expression level of the one or plurality of underexpressed proteins is to thereby derive an integrated score. In one particular embodiment, the comparison of the expression level of the one or plurality of overexpressed proteins and/or the expression level of the one or plurality of underexpressed proteins is integrated with:

(i) the comparison of the expression level of the overexpressed genes associated with chromosomal instability and/or the expression level of the underexpressed genes associated with estrogen receptor signalling to derive a second integrated score; or

(ii) the first integrated score to derive a third integrated score; or

(iii) the comparison of the expression level of the overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1* and/or the expression level of the underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH,*

KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3 to derive a fourth integrated score; or

(iv) the comparison of the expression level of the overexpressed genes and/or an expression level of the underexpressed genes, wherein the genes are from one or a plurality of the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and/or the Multiple Networks metagene, to derive a fifth integrated score; or

(v) the comparison of the expression level of the overexpressed genes and/or the expression level of the underexpressed genes, wherein the genes are from one or a plurality of the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene, to derive a sixth integrated score.

wherein the second, third, fourth, fifth and/or sixth integrated score is indicative of, or correlates with, the aggressiveness and/or prognosis of the cancer in the mammal.

In particular embodiments, the second, third, fourth, fifth and/or sixth integrated score are derived, at least in part, by addition, subtraction, multiplication, division and/or exponentiation.

In a preferred embodiment, the first, second and/or third integrated scores are derived, at least in part, by exponentiation wherein the comparison of the expression level of the other overexpressed genes and the expression level of the other underexpressed genes is raised to the power of

(i) the comparison of the expression level of the overexpressed genes associated with chromosomal instability and/or the expression level of the underexpressed genes associated with estrogen receptor signalling; and/or

(ii) the comparison of the expression level of the overexpressed proteins and/or the expression level of the underexpressed proteins.

In a ninth aspect, the invention provides a method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and/or an expression level of one or a plurality of underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

In a tenth aspect, the invention provides a method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and/or an expression level of one or a plurality of underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with a more favourable cancer prognosis compared to a mammal having a higher expression level.

In an eleventh aspect, the invention provides method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method

including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene, wherein an altered or modulated relative expression level of the overexpressed genes compared to the underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

Suitably, for the present aspect the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and/or the Multiple Networks metagene comprise one or a plurality of genes listed in Table 21.

In a twelfth aspect, the invention provides a method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Metabolism metagene, a Signalling metagene, a Development and Growth metagene, a Chromosome Segregation/Replication metagene, an Immune Response metagene and a Protein Synthesis/Modification metagene, wherein an altered or modulated relative expression level of the overexpressed genes compared to the underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

In one embodiment of the eleventh and twelfth aspects, the one or plurality of overexpressed genes and/or the one or plurality of underexpressed genes are selected from one of the metagenes. In an alternative embodiment, the one or plurality of overexpressed genes and/or the one or plurality of underexpressed genes are selected from a plurality of the metagenes.

Suitably, the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene comprise one or a plurality of genes listed in Table 22.

In particular embodiments, the one or plurality of overexpressed genes and the one or plurality of underexpressed genes are from one or a plurality of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene.

According to the method of the eleventh and twelfth aspects, the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes includes comparing an average expression level of the one or plurality of overexpressed genes and/or an average expression level of the one or plurality of underexpressed genes. This may include calculating a ratio of the average expression level of the one or plurality of overexpressed genes and the average expression level of the one or plurality of underexpressed genes. Suitably, the ratio provides an aggressiveness score which is indicative of, or correlates with, cancer aggressiveness and a less favourable prognosis. Alternatively, the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes includes comparing the sum of expression levels of the one or plurality of overexpressed genes and/or the sum of expression levels of the one or plurality of underexpressed genes. This may include calculating a ratio of the sum of expression levels of the one or plurality of overexpressed genes and/or the sum of expression levels of the one or plurality of underexpressed genes.

In a thirteenth aspect, the invention provides a method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of determining an expression level of one or a plurality of genes associated with chromosomal instability in one or a plurality of non-mitotic cancer cells of the mammal, wherein a higher expression level indicates or correlates with relatively increased responsiveness of the cancer to the anti-cancer treatment

Suitably, the one or plurality of genes associated with chromosomal instability are selected from the group consisting of: *TTK, CEP55, FOXM1* and *SKIP2* and/or any CIN genes listed in Table 4.

In a fourteenth aspect, the invention provides a method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes associated with chromosomal instability and/or an expression level of one or a plurality of underexpressed genes associated with estrogen receptor signalling in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed genes associated with chromosomal instability compared to the one or plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

In certain embodiments, the genes associated with chromosomal instability are of a CIN metagene. Non-limiting examples include genes selected from the group consisting of: *ATP6V1C1, RAP2A, CALM1, COG8, HELLS, KDM5A, PGK1, PLCH1, CEP55, RFC4, TAF2, SF3B3, GPI, PIR, MCM10, MELK, FOXM1, KIF2C, NUP155, TPX2, TTK, CENPA, CENPN, EXO1, MAPRE1, ACOT7, NAE1, SHMT2, TCP1, TXNRD1, ADM, CHAF1A and SYNCRIP*. Preferably, the genes are selected from the group consisting of: *MELK, MCM10, CENPA, EXO1, TTK* and *KIF2C*.

In certain embodiments, the genes associated with estrogen receptor signalling are of an ER metagene. Non-limiting examples include genes selected from the group consisting of: *BTG2, PIK3IP1, SEC14L2, FLNB, ACSF2, APOM, BIN3, GLTSCR2, ZMYND10, ABAT, BCAT2, SCUBE2, RUNX1, LRRC48, MYBPC1, BCL2, CHPT1, ITM2A, LRIG1, MAPT, PRKCB, RERE, ABHD14A, FLT3, TNN, STC2, BATF, CD1E, CFB, EVL, FBXW4, ABCB1, ACAA1, CHAD, PDCD4, RPL10,*

RPS28, RPS4X, RPS6, SORBS1, RPL22 and RPS4XP3. Preferably, the genes are selected from the group consisting of: *MAPT* and *MYB*.

Suitably, the method of this aspect further includes the step of comparing an expression level of one or a plurality of other overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and/or an expression level of one or a plurality of other underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3* in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of other overexpressed genes compared to the one or plurality of other underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

In one embodiment, the one or plurality of other overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*, *CD55*, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593*.

In one embodiment, the one or plurality of other underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

In certain embodiments, the comparison of the expression level of the one or plurality of other overexpressed genes and/or the expression level of the one or plurality of other underexpressed genes is integrated with the comparison of the expression level of the one or plurality of overexpressed genes associated with chromosomal instability and/or the expression level of the one or plurality of underexpressed genes associated with estrogen receptor signalling to derive a first integrated score, which is indicative of, or correlates with, responsiveness of the cancer to the anti-cancer treatment. By way of example, the first integrated score may be derived, at least in part, by addition, subtraction, multiplication, division and/or exponentiation. Preferably, the integrated score is derived by exponentiation,

wherein the comparison of the expression level of the one or plurality of other overexpressed genes and the expression level of the one or plurality of other underexpressed genes is raised to the power of the comparison of the expression level of the one or plurality of overexpressed genes associated with chromosomal instability and the expression level of the one or plurality of underexpressed genes associated with estrogen receptor signalling.

In a fifteenth aspect, the invention provides a method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

In one embodiment, the one or plurality of overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*, *CD55*, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593*.

In one embodiment, the one or plurality of underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

Suitably, the method of the eleventh, twelfth, thirteenth, fourteenth and fifteenth aspects further includes the step of comparing an expression level of a one or a plurality of overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and/or an expression level of

one or a plurality of underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

Suitably, the comparison of the expression level of the one or plurality of overexpressed proteins and/or the expression level of the one or plurality of underexpressed proteins is to thereby derive an integrated score. In one particular embodiment, the comparison of the expression level of the one or plurality of overexpressed proteins and/or the expression level of the one or plurality of underexpressed proteins is integrated with:

(i) the comparison of the expression level of the overexpressed genes associated with chromosomal instability and/or the expression level of the underexpressed genes associated with estrogen receptor signalling to derive a second integrated score; or

(ii) the first integrated score to derive a third integrated score; or

(iii) the comparison of the expression level of the overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1* and/or the expression level of the underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3* to derive a fourth integrated score; or

(iv) the comparison of the expression level of the overexpressed genes and an expression level of the underexpressed genes, wherein the genes are from one or a plurality of the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development

metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and/or the Multiple Networks metagene, to derive a fifth integrated score; or

(v) the comparison of the expression level of the overexpressed genes and an expression level of the underexpressed genes, wherein the genes are from one or a plurality of the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene, to derive a sixth integrated score.

wherein the second, third, fourth, fifth and/or sixth integrated score is indicative of, or correlates with, responsiveness of the cancer to the anti-cancer treatment.

In particular embodiments the first, second, third, fourth, fifth and/or sixth integrated score are derived, at least in part, by addition, subtraction, multiplication, division and/or exponentiation.

In a preferred embodiment, the first, second and/or third integrated scores are derived, at least in part, by exponentiation wherein the comparison of the expression level of the other overexpressed genes and/or the expression level of the other underexpressed genes is raised to the power of

(i) the comparison of the expression level of the overexpressed genes associated with chromosomal instability and/or the expression level of the underexpressed genes associated with estrogen receptor signalling; and/or

(ii) the comparison of the expression level of the overexpressed proteins and/or the expression level of the underexpressed proteins.

In a sixteenth aspect, the invention provides method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2,

AKT1, NFKB1, HER2, ASNS and COL6A1, and/or an expression level of one or a plurality of underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

Suitably, the anticancer treatment of the eleventh, twelfth, thirteenth, fourteenth, fifteenth and sixteenth aspects is selected from the group consisting of endocrine therapy, chemotherapy, immunotherapy and a molecularly targeted therapy. In certain embodiments, the anticancer treatment comprises an anaplastic lymphoma kinase (ALK) inhibitor, a BCR-ABL inhibitor, a heat shock protein 90 (HSP90) inhibitor, an epidermal growth factor receptor (EGFR) inhibitor, a poly (ADP-ribose) polymerase (PARP) inhibitor, retinoic acid, a B-cell lymphoma 2 (Bcl2) inhibitor, a gluconeogenesis inhibitor, a p38 mitogen-activated protein kinase (MAPK) inhibitor, a mitogen-activated protein kinase kinase 1/2 (MEK1/2) inhibitor, a mammalian target of rapamycin (mTOR) inhibitor, a phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K) inhibitor, an insulin-like growth factor 1 receptor (IGF1R) inhibitor, a phospholipase C-γ (PLCγ) inhibitor, a c-Jun N-terminal kinase (JNK) inhibitor, a p21-activated kinase-1 (PAK1) inhibitor, a spleen tyrosine kinase (SYK) inhibitor, a histone deacetylase (HDAC) inhibitor, a fibroblast growth factor receptor (FGFR) inhibitor, an X-linked inhibitor of apoptosis (XIAP) inhibitor, a polo-like kinase 1 (PLK1) inhibitor, an extracellular-signal-regulated kinase 5 (ERK5) inhibitor and combinations thereof.

Suitably, the method of the eleventh, twelfth, thirteenth, fourteenth, fifteenth and sixteenth aspects further includes the step of administering to the mammal a therapeutically effective amount of the anticancer treatment. Preferably, the anticancer treatment is administered when the altered or modulated relative expression level indicates or correlates with relatively increased responsiveness of the cancer to the anti-cancer treatment.

In a seventeenth aspect, the invention provides a method of predicting the responsiveness of a cancer to an immunotherapeutic agent in a mammal, said method including the step of comparing an expression level of one or a plurality of

overexpressed genes selected from the group consisting of *ADORA2B, CD36, CETN3, CFDP1, KCNG1, LAMA3, NAE1, MAP2K5, PGK1, SF3B3, STAU1 and TXN* and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of*APOBEC3A, BTN2A2, BCL2, CAMK4, FBXW4, CAMSAP1, CARHSP1, GSK3B, HCFC1R1, PSEN2, MYB and ZNF593,* , in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the immunotherapeutic agent.

Suitably, the immunotherapeutic agent is an immune checkpoint inhibitor. Preferably, the immune checkpoint inhibitor is or comprises an anti-PD1 antibody or an anti-PDL1 antibody.

In an eighteenth aspect is provided a method of predicting the responsiveness of a cancer to an epidermal; growth factor receptor (EGFR) inhibitor in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *NAE1*, *GSK3B*, *TAF2*, *MAPRE1*, *BRD4*, *STAU1*, *TAF2*, *PDCD4*, *KCNG1*, *ZNRD1-AS1*, *EIF4B*, *HELLS*, *RPL22*, *ABAT*, *BTN2A2*, *CD1B*, *ITM2A*, *BCL2*, *CXCR4*, and *ARNT2* and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of *CD1C*, *CD1E*, *CD1B*, *KDM5A*, *BATF*, *EVL*, *PRKCB*, *HCFC1R1*, *CARHSP1*, *CHAD*, *KIR2DL4*, *ABHD5*, *ABHD14A*, *ACAA1*, *SRPK3*, *CFB*, *ARNT2*, *NDUFC1*, *BCL2*, *EVL*, *ULBP2*, *BIN3*, *SF3B3*, *CETN3*, *SYNCRIP*, *TAF2*, *CENPN*, *ATP6V1C1*, *CD55* and *ADORA2B* in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the EGFR inhibitor.

In a nineteenth aspect is provided a method of predicting the responsiveness of a cancer to a multikinase inhibitor in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *SCUBE*, *CHPT1*, *CDC1*, *BTG2*, *ADORA2B* and *BCL2*, and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of *NOP2*, *CALR*, *MAPRE1*, *KCNG1*, *PGK1*, *SRPK3*,

RERE, *ADM*, *LAMA3*, *KIR2DL4*, *ULBP2*, *LAMA4*, *CA9*, and *BCAP31*, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the multikinase inhibitor.

Suitably, for the method of the seventeenth, eighteenth and nineteenth aspects, a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a relatively increased responsiveness of the cancer to the immunotherapeutic agent, EGFR inhibitor or multikinase inhibitor; and/or a lower relative expression level of the one or aplurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a relatively decreased responsiveness of the cancer to the immunotherapeutic agent, EGFR inhibitor and/or multikinase inhibitor.

In some embodiments, the method of the seventeenth, eighteenth and nineteenth aspects further includes the step of administering to the mammal a therapeutically effective amount of the immunotherapeutic agent, the EGFR inhibitor or the multikinase inhibitor respectively. Preferably, the immunotherapeutic agent, the EGFR inhibitor or the multikinase inhibitor is administered when the altered or modulated relative expression level indicates or correlates with relatively increased responsiveness of the cancer to the immunotherapeutic agent, the EGFR inhibitor or the multikinase inhibitor respectively.

Suitably, for the methods of the aforementioned aspects, the step of comparing an expression level of one or a plurality ofoverexpressed genes or proteins and an expression level of one or a plurality of underexpressed genes or proteins, includes comparing an average expression level of the one or plurality of overexpressed genes or proteins and an average expression level of the one or plurality of underexpressed genes or proteins. This may include calculating a ratio of the average expression level of the one or plurality of overexpressed genes or proteins and the average expression level of the one or plurality of underexpressed genes or proteins. Suitably, the ratio provides an aggressiveness score which is indicative of, or correlates with, cancer aggressiveness and a less favourable prognosis. Alternatively, the step of comparing an expression level of one or a

plurality of overexpressed genes and an expression level of one or a plurality of underexpressed genes or proteins , includes comparing the sum of expression levels of the one or plurality of overexpressed genes or proteins and the sum of expression levels of the one or plurality of underexpressed genes or proteins. This may include calculating a ratio of the sum of expression levels of the one or plurality of overexpressed genes or protein and the sum of expression levels of the one or plurality of underexpressed genes or proteins.

In certain embodiments of the aforementioned methods, the mammal is subsequently treated for cancer.

In a twentieth aspect, the invention provides a method for identifying an agent for use in the treatment of cancer including the steps of:

(i) contacting a protein product of *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7*, *COG8, CFDP1* and/or *KCNG1* with a test agent; and

(ii) determining whether the test agent, at least partly, reduces, eliminates, suppresses or inhibits the expression and/or an activity of the protein product.

Suitably, the agent possesses or displays little or no significant off-target and/or nonspecific effects.

Preferably, the agent is an antibody or a small organic molecule.

In a twenty first aspect, the invention provides an agent for use in the treatment of cancer identified by the method of the eighteenth aspect.

In a twenty second aspect, the invention provides a method of treating a cancer in a mammal, including the step of administering to the mammal a therapeutically effective amount of an agent identified by the method of the eighteenth aspect.

Preferably, for the invention of the twentieth, twenty first and twenty second aspects, the cancer has an overexpressed gene selected from the group consisting of *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7*, *COG8, CFDP1*, *KCNG1* and any combination thereof.

Suitably, the method of the aformentioned aspects further includes the step of determining, assessing or measuring the expression level of one or plurality of the overexpressed genes, the underexpressed genes, the overexpressed proteins and/or the underexpressed proteins described herein.

Suitably, the mammal referred to in the aforementioned aspects and embodiments is a human.

In certain embodiments of the invention of the aforementioned aspects, the cancer includes breast cancer, lung cancer inclusive of lung adenocarcinoma and lung squamous cell carcinoma, cancers of the reproductive system inclusive of ovarian cancer, cervical cancer, uterine cancer and prostate cancer, cancers of the brain and nervous system, head and neck cancers, gastrointestinal cancers inclusive of colon cancer, colorectal cancer and gastric cancer, liver cancer inclusive of hepatocellular carcinoma, kidney cancer inclusive of renal clear cell carcinoma and renal papillary cell carcinoma, skin cancers such as melanoma and skin carcinomas, blood cell cancers inclusive of lymphoid cancers and myelomonocytic cancers, cancers of the endocrine system such as pancreatic cancer and pituitary cancers, musculoskeletal cancers inclusive of bone and soft tissue cancers, although without limitation thereto. By way of example, breast cancer includes aggressive breast cancers and cancer subtypes such as triple negative breast cancer, grade 2 breast cancer, grade 3 breast cancer, lymph node positive (LN^+) breast cancer, HER2 positive ($HER2^+$) breast cancer and ER positive (ER^+) breast cancer, although without limitation thereto.

Unless the context requires otherwise, the terms "*comprise*", "*comprises*" and "*comprising*", or similar terms are intended to mean a non-exclusive inclusion, such that a recited list of elements or features does not include those stated or listed elements solely, but may include other elements or features that are not listed or stated.

The indefinite articles '*a*' and '*an*' are used here to refer to or encompass singular or plural elements or features and should not be taken as meaning or defining "one" or a "single" element or feature.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Correlation of breast cancer subtypes and the aggressiveness gene list. The METABRIC dataset was visualized according to the expression of the 206 genes (Table 4) in the aggressiveness gene list. The aggressiveness score for each tumor was calculated as the ratio of the CIN metagene (average value for CIN genes expression) to the ER metagene (average value for ER genes expression). (**A**) The expression of the aggressiveness gene list according to the GENIUS histological classification. Box plot shows the aggressiveness score of the histological subtypes. (**B**) The overall survival of patients in the METABRIC dataset was analyzed

according to the aggressiveness score (upper row: by quartiles; lower row: by median) in all patients, non-TNBC patients and in patients with ER+ Grade 2 tumors. The hazard ratio (HR) and confidence interval (CI) and p-value for comparisons of upper quartile vs. lower quartiles (upper row) and at the dichotomy across the median (high vs. low) are shown (Log-rank Test, GraphPad® Prism). The number of patients (n) in each group is shown in brackets.

Figure 2: Network analysis of the aggressiveness gene list. (**A**) Ingenuity pathway analysis was performed using direct interactions on the 206 genes in the aggressiveness gene list (red is overexpressed and green is underexpressed). One network of high direct interactions was identified. (**B**) The genes in the network in A were investigated for their correlation with the aggressiveness score and overall survival (Table 5) and eight genes (MAPT, MYB, MELK, MCM10, CENPA, EXO1, TTK and KIF2C) with the highest correlation were still connected in a direct interaction network. (**C**) The overall survival of patients in the METABRIC dataset was analyzed according to score from the 8 genes in C (upper row: by quartiles; lower row: by median) in all patients, non-TNBC patients and in patients with ER+ Grade 2 tumors.

Figure 3: Survival of patients stratified by the 8-genes score in the METABRIC dataset. The overall survival of patients in the METABRIC dataset was analyzed according to the 8-genes score in selected settings in all patients (**A**) or in ER-positive patients only (**B**). (**A**) TP53 mutation was compared in high vs. low 8-genes score (split by the median). The expression of the proliferation marker Ki67 was divided by dichotomy across the median and patients in each of these groups were then stratified according to their 8-genes score (split by quartiles). Disease stages (Stage I – Stage III) were stratified by the median 8-genes score. (**B**) ER^+ Grade 3, ER+ lymph node negative (LN-) and ER+ LN+ tumors were stratified by the quartiles.

Figure 4: The 8-genes score associates with survival of breast cancer patients. Four published datasets were used to validate the 8-genes score as a predictor of survival. The 8-genes score was calculated for tumors in each of the datasets and the survival of patients was stratified according to the median 8-genes score; (**A**) GSE2990[15], (**B**) GSE3494[65], (**C**) GSE2034[66] and (**D**) GSE25066[53]. The hazard ratio (HR) and confidence interval (CI) and p-value for comparisons high vs. low 8-genes score are shown in the Kaplan-Meier survival curves (Log-rank Test, GraphPad®

Prism). The number of patients (n) is shown in brackets. The table in each panel show multivariate survival analysis in the using Cox-proportional hazard model including all available conventional indicators.

Figure 5: Therapeutic targets in the aggressiveness gene list. (**A**) The TNBC cell lines, MDA-MB-231, SUM159PT and Hs578T were treated with control siRNA (Scrambled, Sc CTRL) or siRNA targeting the specified genes and the survival of these cells was compared on day 6. Data shown is the average from the three cell lines where each cell line was treated in triplicate. * $p< 0.05$, ** $p<0.01$ and *** <0.001 from One-Way ANOVA analysis performed using GraphPad® Prism. Data for individual cell lines is shown in Table 5. (**B**) A panel of breast cancer cell lines was used to prepare lysates for immunoblotting of TTK. Tubulin was used as the loading control. (**C**) Dose response curves for the treatment of breast cancer cell lines in the absence or presence of escalating doses of the TTK inhibitor (TTKi) AZ3146. The survival of cells was measured using the CellTitre® MTS/MTA assay carried out 6 days after treatment. Percentage survival (n = 3 per dose) was calculated as the percentage of the signal from treated cells to that from control cells. (**D**) The concentration of TTK required to affect the survival of 50% of the cells (IC50) was measured by GraphPad® Prism from the dose response curves in C for each cell line.

Figure 6: TTK protein expression associates with breast cancer survival. The overall survival of patients in a large cohort of breast cancer patients (n=409) was stratified according to TTK staining by IHC (scores 0-3). Kaplan-Meier survival curves are shown for all patients (**A**) with four TTK staining (categories 0-3) and (**B**) two categories (0-2 vs. 3). Log-rank Test and p-value were used for survival curves. (**C**) The distribution of high TTK staining (category 3) across histological subgroups and mitotic indices. Data shown is the mitotic index (median + range) measured as the number of mitotic cells in 10 high power fields (hpf). The number of tumors with high TTK staining to the total number of tumors in the cohort is shown on the right. High TTK expression distributed across subtypes and did not associate with mitotic index.

Figure 7: TTK associates with aggressive subtypes and is a therapeutic target. (**A**) Kaplan-Meier survival curves are shown for Grade 3 tumors, lymph node positive patients (LN^+) and LN^+ patients with grade 3 tumors. Log-rank Test and p-value were used for these survival curves. For patients with TNBC, and HER2, survival was statistically significant using the Gehan-Breslow-Wilcoxon test (p-

values marked by asterisks) which gives more weight to deaths at early time points. The poorer survival of patients with high Ki67 tumors and high TTK staining was a trend but did not reach significance. Survival curves and statistical analyses were performed using GraphPad® Prism. (**B**) TNBC and non-TNBC cell lines were treated for 6 days with the specified concentrations of docetaxel (doc) alone, TTK inhibitor (TTKi) alone of the combinations. The survival of cells was measured using the MTS/MTA assay as described in Methods. *** $p < 0.001$ comparing the combination to single agents and to non-TNBC cell lines from Two-Way Anova in GraphPad® Prism. (**C**) MDA-MB-231 cells were treated with docetaxel or TTKi alone or in combination and collected at 96 hours to perform apoptosis assays by flow cytometry. Early apoptotic cells were defined as annexin V+/7-AAD-.

Figure 8: Global gene expression meta-analysis of genes deregulated in TNBC, metastatic events and death at 5 years in Oncomine™. (**A**) TNBC in 8 datasets were compared to non-TNBC, (**B**) tumors with metastatic events at 5 years were compared to those with no metastatic events at 5 years in 7 datasets and (**C**) tumors leading to death at 5 years were compared to those that did not lead to death at 5 years were compared in 7 datasets. The datasets used in the comparisons are stated in the legends and the key for the heatmap coloring is also included. The heatmap key denotes the top or bottom x % placement of a gene according to gene rank which is based on the p-value.

Figure 9: The derivation of the 206 aggressiveness gene list. (**A** and **B**) are Venn diagrams for the top overexpressed genes and bottom underexpressed genes shared between TNBC and/or metastasis and death at 5 years analyses in Oncomine™. (**C** and **D**) The Venn diagrams from A and B were crossed with genes which were deregulated in TNBC in comparison to adjacent normal breast tissue from the METABRIC dataset. The genes marked in bold in panels C and D are the 206 genes which constitute the unfiltered aggressiveness gene list.

Figure 10: Common genes between the 206 aggressiveness gene list and metagene attractors. Venn diagrams show common genes (in bold) between the 206 aggressiveness gene list and the chromosomal instability (CIN), lymphocyte-specific and ER attractors (Cheng et al 2013a, Cheng et al 2013b). The table below lists the shared genes. The 6 overexpressed genes (marked in red) and 2 underexpressed genes (marked in green) which constitute the 8-genes signature in this study are shown. Gene set enrichment analysis of the remaining 140 genes which

were only present in the 206 gene signature reveal that these genes function in cell cycle.

Figure 11: Correlation of breast cancer subtypes and the aggressiveness gene list. The METABRIC dataset was visualized according to the expression of the 206 genes in the aggressiveness gene list. The aggressiveness score for each tumor was calculated as the sum of normalized z-score expression values of overexpressed genes divided by that of underexpressed genes. (**A** and **B**) The expression of the aggressiveness gene list was visualized according to PAM50 intrinsic subtypes and the integrative clusters classification. Box plots show the aggressiveness score of these subtypes. The shaded lines in box plots mark the median value for the aggressiveness score. *** $p < 0.001$ One-Way ANOVA using GraphPad® Prism. Kaplan-Meier curves are of overall survival of patients in the METABRIC dataset stratified according to the quartiles (left plot) or the median (middle plot) of the aggressiveness score in ER+ patients with Grade 3 tumors. Tumors of the five PAM50 intrinsic subtypes which show high aggressiveness score (higher than the median) did not show statistical difference in overall survival (right plot). The hazard ratio (HR) and the 95% confidence interval (CI) and the p-value are reported using the Log-rank Test.

Figure 12: Survival of the PAM50 breast cancer subtypes in the METABRIC dataset according to the aggressiveness score. The survival of patients in the METABRIC dataset annotated based on the PAM50 subtypes was analyzed by dichotomy across the median aggressiveness score from the 206 gene list (**A**) and the reduced 8 gene list (**B**). The p-value are reported using the Log-rank Test in GraphPad® Prism and show that all tumors with the different PAM50 subtypes but high aggressiveness score did not show a difference in patient survival (left graphs), whereas the PAM50 subtypes showed significantly different survival only in low aggressiveness score setting.

Figure 13: TTK staining association with patient survival. The overall survival of patients in a large cohort of breast cancer patients (n=409) was stratified according to TTK staining by IHC (scores 0-3). Kaplan-Meier survival curves are shown for all patients (with four TTK staining categories 0-3 and two categories (0-2 vs. 3) with 10 and 20 years follow up. Log-rank Test and p-value were used for survival curves of all patients. There were no statistical differences in the survival of patients with

Grade 1, Grade 2 or hormone positive tumors when stratified by TTK expression. Survival curves and statistical analyses were performed using GraphPad® Prism.

Figure 14: Criteria used for assigning 'prognostic subgroups' in this study.

Figure 15: Panel 1: Overall survival curves of lung cancer patients split by ten (10) CIN and two (2) ER genes as a signature; patients are low or high according to the median of the signature; Panel 2: Survival curves for lung adenocarconima split by ten (10) CIN genes and two (2) ER genes as a signature; patients are low or high according to the median of the signature; Panel 3: Survival curves for lung adenocarconima (10 years) split by ten (10) CIN genes and two (2) ER genes as a signature; patients are low or high according to the median of the signature; Panel 4: Survival curves for lung adenocarconima split by six (6) CIN genes and two (2) ER genes as a signature; patients are low or high according to the median of the signature; and Panel 5: Survival curves for lung adenocarconima (10 years) split by six (6) CIN genes and two (2) ER genes as a signature; patients are low or high according to the median of the signature.

Figure 16: (**A**) RNA-Seq data from the breast cancer cohort of The Cancer Genome Atlas (TCGA) data. (**B**) Recurrence-free survival of breast cancer patients in the TCGA stratified by the Aggressiveness score compared to the OncotypeDx recurrence score. (**C**) Comparison of copy number variations (CNVs) of breast tumours with high aggressiveness score to those with low aggressiveness score.

Figure 17: (**A**) RNA-Seq data from all cancers of The Cancer Genome Atlas (TCGA) data. (**B**) Recurrence-free survival of all cancer patients in the TCGA stratified by the Aggressiveness score compared to the OncotypeDx recurrence score.

Figure 18: Recurrence-free survival or overall survival of cancer patients with different cancer types in the TCGA data patients stratified by the 8-genes aggressiveness score.

Figure 19: Outline of Example 2. Meta-analysis was performed in Oncomine™ using breast cancer datasets irrespective of subtypes or gene expression array platforms used. The global gene expression profiles of breast tumors that led to metastatic or death event within 5 years were compared to those that did not and the top overexpressed (OE) and underexpressed genes (UE) in these comparisons were selected. The commonly deregulated genes in the primary tumors that led to metastatic and death events (depending on the annotation of each dataset) were then interrogated using the online tool KM-Plotter™ (n>4000 patients with some overlap

with the datasets in Oncomine™). Only genes which associated with relapse-free survival (RFS), distant metastasis-free survival (DMFS) or overall survival (OS) of basal-like breast cancer (BLBC) or ER-negative (ER^-) breast cancer were selected. The 96 genes from this training were then shortlisted to 28 genes by selecting the most significant and persistent across the different outcomes (RFS, DMFS and OS). The 28-gene signature was then validated in large cohorts of breast cancer gene expression studies including The Cancer Genome Atlas (TCGA) dataset the Research Online Cancer Knowledgebase (ROCK) dataset and the homogenous TNBC dataset for prognostication of ER-, TNBC and BLBC subtypes. Finally, the TN signature was then investigated for association with pathological complete response (pCR) after neoadjuvant chemotherapy in studies which performed gene expression profiling prior to therapy.

Figure 20: The 28-gene TN signature associates with RFS, DMFS and OS of BLBC and ER- breast cancer. The 21 overexpressed and 7 underexpressed genes were used as a signature in the online tool KM-Plotter. The signature (the average expression of the 21 overexpressed genes and the inverted expression of the 7 underexpressed genes) stratified the RFS, DMFS and OS; low: under the median of the expression of the signature and high: over the median of the expression of the signature. The hazard ratio (HR) and log-rank p-value (p) for the univariate survival analyses were generated by KM-Plotter. n = number of patients.

Figure 21: The prognostication by the TN score outperforms standard clinicothapological indicators in TNCBC, BLBC and ER- breast cancer subtypes. Two datasets, (**A**) the TNBC dataset and (B&C) the ROCK dataset, were analyzed for the TN signature and the TN score was calculated as the ratio of the average expression of the 21 overexpressed genes to that of the 7 underexpressed genes. This score was calculated for each tumor and the median TN score over the entire dataset was used to classify tumors as high (above the median) or low (below the median) for the TN score. (A) RFR of TNBC patients in the TNBC cohort stratified by dichotomy across the median TN score in the cohort. Table under the survival curve shows univariate and multivariate survival analysis for the TN score and other available clinical indicators recorded in the dataset. The TN score outperformed all the clinical indicators in the multivariate analysis. (**B**) RFS and DMFS of BLBC in the ROCK dataset stratified by dichotomy across the median TN score in the dataset. The table under the survival curves shows multivariate survival

analysis for the TN score against other available clinical indicators recorded in the dataset. The TN score outperformed all the clinical indicators in the multivariate analysis of BLBC cases. (**C**) The RFS and DMFS of ER- negative breast cancer were stratified by the TN score (data not shown) and the table shows the multivariate survival analysis that the TN score outperforms clinical indicators in ER^- breast cancer cases.

Figure 22: The TN score stratifies the overall survival of ER- breast cancer patients in the TCGA dataset. The gene expression data using the Illumina HiSeq RNA-seq arrays from the TCGA breast cancer data (n = 1106) were used to calculate the TN score for all tumors. Tumors were classified as high or low for the TN score by dichotomy across the median TN score. The overall survival (OS) of ER- breast cancer cases with high TN score were compared to those with low TN score. The table below the survival curve shows that the TN score is more significant than other clinical indicators in univariate survival analysis and it is the only significant prognostic indicator in multivariate survival analysis.

Figure 23: The TN score associates with pCR after chemotherapy in ER^-HER2^- breast cancer. Gene expression datasets which profiled tumors prior to neoadjuvant chemotherapy and recorded pathological complete responses (pCR) vs. no pCR or residual disease (RD) were analyzed for the TN signature and the TN score was calculated for each tumor. Tumors were classified as high or low TN score by dichotomy across the median TN score in each dataset. Only ER-HER2- cases were used in the data shown in the Figure. (**A**) Graphs showing the percentage of cases achieving (red bars) or not achieving (black bars) pCR in low and high TN score subgroups. Fisher's exact test was used to analyze the 2x2 contingency tables and the p-value from this test was reported when statistical significance was observed. The dotted line marks the 31% pCR rate reported in literature for TNBC. Each dataset is labeled with the accession number and the chemotherapy regimen used, namely: GSE18728, GSE50948, GSE20271, GSE20194, GSE22226, GSE42822 and GSE23988. Chemotherapy abbreviation: 5-**F**U, **A**driamycin, **C**yclophosphamide, **T**axane, X: **X**eloda, **M**ethotrexate, **E**pirubicin. (**B**) The dataset GSE22226 from the ISPY-1 trial was used to compare the TN score and pCR in the prediction of ER^- patient survival after neoadjuvant chemotherapy as this dataset also recorded RFS. pCR strongly associated with RFS (first panel) as previously reported. the TN score (next three panel) was not only predictive of survival in the these patients but could

also stratified the survival of patients achieving or not achieving pCR, indicated the TN score as an independent prognostic factor for pCR after neoadjuvant chemotherapy.

Figure 24: Drug sensitivity of cancer cell lines according to the TN score. The large published study by Garnett et al. was investigated where the TN score was calculated for each cell line in the study as described in Methods. The cell lines were classified as high or low TN score according to the median TN score to compare the sensitivity of low TN score cell lines (white boxes) and high TN score cell lines (red boxes). Graphs were prepared using GraphPad® Prism showing sensitivity as -log10[IC50] in boxes (with median marked by a line) and whiskers (marking the 1st and 3rd quartiles and outliers as dots according to Tukey method for plotting the whiskers and outliers). Unpaired two-tailed t test was used for statistical analysis.

Figure 25: The iBCR score stratifies the survival of all breast cancer patients irrespective of ER status in the ROCK dataset. The TN and Agro scores were calculated for each tumor in the ROCK dataset (n=1570, Affymetrix) and then the iBCR score was calculated as the TN score to the power of the Agro score. The RFS of all patients and the RFS of ER- or ER+ patients only was compared between high score and low score by dichotomy across the median score for each of the scores. The iBCR score was prognostic in all patients as well as ER- and ER+ subsets with better separation between low score and high score tumors (increased hazard ratio [HR] and limits of the 95% confidence intervals and decreased log rank p-value). Graphs and the univariate survival analysis using the log rank test were performed using GraphPad® Prism.

Figure 26: The iBCR score stratifies the survival of all breast cancer patients irrespective of ER status in the TCGA dataset. The TN. Agro and the iBCR scores were calculated for each tumor in the TCGA dataset (n=1106, Illumina RNA-Seq). The RFS of all patients and the RFS of ER- or ER+ patients only was compared between high score and low score. As in the results in the ROCK dataset in Figure 7, The iBCR score was prognostic in all patients as well as ER- and ER+ subsets with better separation between low score and high score tumors.

Figure 27: The iBCR score associates with RFS and pCR after chemotherapy in the ISPY-1 trial. The dataset GSE22226 from the ISPY-1 trial was used to compare the Agro, TN and the integrated iBCR score in the prognosis and association with pCR after chemotherapy (**A**driamycin, **C**yclophosphamide and **T**axane) in ER$^-$HER2$^-$

and ER^+ breast cancer subtypes. Tumors were classified as high or low score by dichotomy across the median of each score in the entire dataset. High iBCR score ER^-HER2^- tumors were less likely to achieve pCR and these patients had poor survival. High iBCR ER^+ patients were more likely to achieve pCR but since a small number of ER^+ patients achieved (10/62 [16%]), the survival of high iBCR ER+ patients remained poor. Note that the Agro score identifies all but two ER-HER2- tumors as high score, thus the data from this group should not be interpreted. Also note that the Agro score is highly prognostic of survival and association with pCR in ER^+ whereas the TN score is not in these patients. The integration of these two scores in the iBCR score has overcame the limitation of each of these subtype-specific scores.

Figure 28: The iBCR score associates with pCR after chemotherapy in breast cancer. Gene expression datasets with pCR annotation after chemotherapy were used as described in Figure 5 to calculate the Agro and TN scores and the integrated iBCR score. Tumors were classified as high or low score by dichotomy across the median of each score in each dataset. (**A**) ER^-HER2^- cases with graphs showing the percentage of cases achieving (red bars) or not achieving (black bars) pCR in low and high score subgroups. (**B**) ER^+ cases were analyzed as in A. Fisher's exact test was used to analyze the 2x2 contingency tables and the p-value from this test was reported when statistical significance was observed. Each dataset is labeled with the accession number and the chemotherapy regimen used, namely: GSE18728, GSE50948, GSE20271, GSE20194, GSE22226, GSE42822 and GSE23988. Chemotherapy abbreviation: 5-**F**U, **A**driamycin, **C**yclophosphamide, **T**axane, X: **X**eloda, **M**ethotrexate, **E**pirubicin.

Figure 29: The iBCR score stratifies the survival of tamoxifen-treated ER+ patients. The Agro and TN scores and the iBCR score were calculated in two datasets of gene expression profiling prior to tamoxifen therapy: **A&B**. GSE6532 with 327 patients. 137 untreated and 190 tamoxifen-treated; **C**: GSE17705 with 298 patients treated with tamoxifen for 5 years. (**A**) ER+ N0 patients with high iBCR score have poor RFS compared low iBCR score counterparts. (**B**) RFS of all ER+ patients and N0 and N1 subsets stratified by the Agro and iBCR scores. (**C**) DMFS survival of all ER+ and N0 and N1 subsets stratified by the Agro and iBCR scores. The hazard ratios and log-rank p-values are more significant for the iBCR score than the Agro score although the Agro score was significantly prognostic.

Figure 30: Drug sensitivity of cancer cell lines according to the iBCR score. The large published study by Garnett et al. was investigated where the iBCR score was calculated for each cell line from the Agro and TN scores. The cell lines were classified as high or low iBCR score according to the median iBCR score to compare the sensitivity of low iBCR score cell lines (white boxes) and high TN score cell lines (red boxes). Results according to low and high Agro score were also included. Graphs were prepared using GraphPad® Prism and unpaired two-tailed t test was used for statistical analysis (n.s. not significant).

Figure 31: Global gene expression meta-analysis of genes deregulated in primary breast tumors with metastatic events or death at 5 years in Oncomine™. (**A**) tumors with metastatic events at 5 years were compared to those with no metastatic events at 5 years in 7 datasets and (**B**) tumors leading to death at 5 years were compared to those that did not lead to death at 5 years were compared in 7 datasets. The datasets used in the comparisons are stated in the legends and the key for the heatmap coloring is also included. The heatmap key denotes the top or bottom x % placement of a gene according to gene rank which is based on the p-value.

Figure 32: The TN signature outperforms all published signatures for TNBC/BLBC. Relapse-free survival of basal-like breast cancer patients (BLBC) was investigated in the online database KM-Plotter (Affymetrix platform) according to the TN signature in comparison to published TNBC signatures. Hazard ratios (HR) and logrank p-values were generated by KM-Plotter. (**A**) the TN score vs. signatures (**B**) from Karn et al. (PLoS One, 2011); from Rody et al. (Breast Cancer Res, 2011) (**C**) IL8, (**D**) VEGF, and (**E**) B-cell metagenes; (**F**) from Yau et al. (Breast Cancer Res, 2010); (**G**) from Yu et al. (Clin Cancer Res, 2013); (**H**) from Lee et al. (PLoS One, 2013 and (**I**) from Hallet et al. (Sci Rep, 2012).

Figure 33: The TN score stratified the survival of ER$^-$ patients in the Agilent TCGA data. The original TCGA dataset using the Agilent microarrays (n=597) were analyzed for the TN score where patients were assigned as low, intermediate or high for the TN score according to tertiles. The RFS of ER- patients only were then compared according to these tertiles. The stratification was significant according to a log-rank survival test ($P<0.0001$). High TN score group vs. low TN score group had a hazard ratio (95% confidence interval) of 3.484 (1.035 to 11.23) with a log rank p-value of 0.0179.

Figure 34: The prognostication by the TN score in ER- and BLBC is not affected by systemic treatment. The online KM-Plotter tool was used to investigate the stratification of RFS, DMFS and OS of ER- breast cancer (top two rows) and BLBC (bottom two rows) in systemically untreated patients (untreated) or in patients who were treated systemically (treated). The HR, the 95% confidence intervals and the log-rank p values were provided by KM-Plotter as well as the number of patients at risk.

Figure 35: Sensitivity of cancer cell lines to anticancer drugs according to the TN score in the Cancer Cell Line Encyclopedia (CCLE) study. The gene expression data of the cancer cell lines in the study were analyzed to calculate the TN score for each cell line and were assigned to low or high TN score by dichotomy across the median. The IC_{50} for each of the 24 drugs used in the CCLE study was compared between high and low TN score cell lines and the data shown are those with statistical differences based on unpaired two-tailed *t*-test performed using GraphPad® Prism.

Figure 36: Integration of the TN and Agro scores by addition or subtraction. The ROCK dataset was used to study the integration of the TN and Agro score with the aim to develop a test that is breast cancer subtype independent. (**A**) The raw Agro and TN scores for ER+ (black dots) and ER- (red dots) in the ROCK dataset (each dot represent one patient, n = 1570 in total). The two scores are scattered and a method of integration that can retain the information from each score in the relevant breast cancer subtype is necessary. Such methods are tested in this Figure and Figure 38. (**B**) Addition method. First column shows the TN score in ER+ tumors with low (white boxes) and high (red boxes) Agro score subgroups (top panel). In the bottom panel, the Agro score in ER- tumors with low (white boxes) and high (red boxes) TN score subgroups. This data shows that the TN score is similar for ER+ tumors with low and high Agro scores and that the Agro score is similar for ER- tumors with low and high TN scores. The lack of statistical differences (independence) suggested that integration is possible. The second column shows the linear correlation between the TN score and Agro score when they were added in each patient for ER+ (top panel) and ER- (bottom panel) patients. In the third column, the TN and Agro scores were plotted against the produced summed score showing that the information from each score is retained in the final summed score for both ER+ (top panel) and ER- (bottom panel) patients.. The last column shows the overlap of data from ER+ and ER-

patients shown separately in the second and third columns. (**C**) Identical analysis as that done in B but the integration was tested by subtraction of the TN and Agro score. The linearity of the relationship between the summed score and each of the single scores (TN and Agro score) indicated that information from each score is represented in the final score. The performance of these two methods (addition or subtraction) was tested for association with survival as shown in Figure 37.

Figure 37: Comparison of different integration methods of the TN and Agro scores for prognostication in ER- and ER+ RFS in the ROCK dataset. The methods of integration by addition or subtraction (from Figure 36) or multiplication or division (Figure 38) were tested for the association of the produced integrated score in the ROCK dataset in ER- or ER+ breast cancer. As shown in the figure, only the addition or multiplication methods were prognostic in ER- breast cancer and the multiplication was more significant in ER+ breast cancer compared to the addition. These two methods are reasonable as subtraction or division methods would reduce the value of one of the scores. Two additional methods were tested, raising one score to the power of the second score since the relationships observed when multiplication and division methods showed exponential or power curves. As shown in the last column (shaded and marked in red box), raising the TN score to the power of the Agro score should superior prognostication in both ER- and ER+ breast cancer subtypes. In fact, the prognostication of this integrated score was better than each of the score in their respective subtypes. The method was therefore used to calculate the **i**ntegrated **B**reast **C**ancer **R**ecurrence (iBCR) score.

Figure 38: Integration of the TN and Agro scores by division or multiplication. The ROCK dataset was used to study the integration of the TN and Agro as these scores were scattered when plotted against each other (panel A in Figure 36). (**A**) The box plots in the first column are identical to those in Figure 36. The shaded boxes in panel A describe integration by division (top row) or multiplication (bottom row) of the TN and Agro scores. The division produced a power curve and the multiplication produced an exponential curve for the relationship between the TN and Agro scores after dividing them or multiplying them by each other in both ER+ (black dots) and ER- (red dots). The overlay in the last column shows that the differences between ER+ and ER- patients for the scores is retained. These two methods were tested for survival association in Figure 37 and the multiplication method was suitable. (**B**) As power and exponential curves were observed in the

division and multiplication methods in A, it was reasonable to test integration by raising one score to the power of the second score. As shown in the top row in the overlay or individual plots, the integration by raising the TN score to the power of the Agro score produced a linear relationship in both ER- (red dots) and ER+ (black dots) patients. This method of integration outperformed all other methods when tested for survival association as shown in Figure 37.

Figure 39: The iBCR score is prognostic in TNBC patients. In addition to the validation of the iBCR score in the ROCK dataset (Affymetrix) and the TCGA dataset (Illumina dataset) of mixed subtypes of breast cancer, the iBCR score was investigated in the homogenous TNBC dataset. As shown in the right panel, the iBCR was as prognostic (with slight improvement) compared to the TN score. This further validates the development of the integrated score to be a prognostic test in breast cancer irrespective of ER status, unlike previous limited signatures.

Figure 40: Survival of tamoxifen-treated ER+ patients according to the Agro score vs. Oncotype Dx. **(A)** RFS and DMFS of node negative (top) and node positive (bottom) ER+ patients treated with tamoxifen in the published study (Loi *et al.*, *Clin Oncol*, 2007) stratified by the Agro Score (high vs. intermediate vs. low by tertiles). **(B)** DMFS of node negative or positive ER+ patients treated with tamoxifen for 5 years from the published study (Symmans *et al.*, *J Clin Oncol*, 2010) was stratified by the tertiles of the Agro Score. (**C**) RFS and DMFS of node negative (top) and node positive (bottom) ER+ patients treated with tamoxifen in the published study (Loi *et al.*, *Clin Oncol*, 2007) stratified by the risk groups of the OncotypeDx Recurrence Score. (**D**) DMFS of node negative or positive ER+ patients treated with tamoxifen for 5 years from the published study (Symmans *et al.*, *J Clin Oncol*, 2010) was stratified by the risk groups of the OncotypeDx Recurrence Score.

Figure 41: Comparison of the Agro Score and MammaPrint in the KM-Plotter tool. Distant metastasis-free survival according to the Agro Score (high vs. low) or according to MammaPrint (high vs. low) in all breast cancer patients, ER+, ER+ lymph node negative (LN-) or ER+ lymph node positive (LN+) patients. The KM-Plotter online tool (n = 4142 patients). The Agro score outperformed the MammaPrint signature in all patient subsets particularly for ER+ node positive patients.

Figure 42: Sensitivity of cancer cell lines to anticancer drugs according to the iBCR score in the Cancer Cell Line Encyclopedia (CCLE) study. The gene

expression data of the cancer cell lines in the study were analyzed to calculate the TN score for each cell line and were assigned to low or high iBCR score by dichotomy across the median. The IC_{50} for each of the 24 drugs used in the CCLE study was compared between high and low iBCR score cell lines and the data shown are those with statistical differences based on unpaired two-tailed *t*-test performed using GraphPad® Prism. As this analysis was also done for the TN score (Figure 35), results from analysis of the Agro score are also shown in the top row.

Figure 43: High copy number variations (CNVs) in high Agro score tumors compared to low Agro score tumors. The breast cancer tumors in the TCGA dataset were classified as high or low for the Agro score based on the gene expression data (Illumina HiSeq RNA-seq). (**A**) The TCGA copy number variations (segmented and after deletion of germline CNV) were visualized using the UCSC Genome Browser to compare patients who were classified from gene expression data as high Agro score patients (top panel) to those classified as low Agro score patients (bottom panel). (**B**) Presentation of the distribution of clinical indicators such as ER, PR and HER2 status and others. (**C**) The difference in the CNVs profile of high Agro score patients to the low Agro score patients showing gains (red) and losses (green) of whole chromosome arms in the high Agro score patients, suggesting aneuploidy.

Figure 44: High Agro and iBCR score cell lines are more sensitive to Aurora kinase inhibitors. Two studies which treated breast cancer cell lines with Aurora kinase inhibitors were analyzed based on the Agro, TN and the iBCR score for these cell lines. As shown in Figure, high Agro score and particularly high iBCR score cell lines were more sensitive to Aurora kinase inhibitors (ENMD-2076: IC50 1.4 μM vs. 5.9 μM for high vs. low iBCR Score cell lines, p=0.0125 *t*-test; AMG 900: IC50 0.3 nM vs. 0.7 nM for high vs. low iBCR score cell lines, p=0.0308 *t*-test).

Figure 45: The iBCR is prognostic in the pan-cancer TCGA data for overall and relapse-free survival. The pan-cancer TCGA data were analyzed for the iBCR gene signature using the UCSC Genome Browser and the data for this signature, survival data and cancer types were downloaded from the browser. Tumors, irrespective of cancer types, were classified into quartiles based on the iBCR signature expression and the overall and relapse free survival were compared across these quartiles. As shown in the top row, overall and relapse-free survival was stratified by the iBCR signature in this pan-cancer dataset. In the far right panel in the top row, the distribution of tumors in each cancer type across the iBCR signature quartile is

shown. Cervical cancer for example displays high iBCR signature in the majority of cases whereas on the opposite side, thyroid cancer displays low iBCR signature in all the cases. The lower panels show the stratification of overall survival according to the iBCR score from the pan-cancer dataset where the stratification was statistically significant in log-rank univariate survival analysis. In addition to the breast cancer data shown in paper, the iBCR signature was prognostic in adrenocortical cancer, endometrioid cancer, kidney clear cell cancer, bladder cancer, lower grade glioma and melanoma. The iBCR was also prognostic in lung adenocarcinoma as shown in Figure 46.

Figure 46: The iBCR signature is prognostic in lung adenocarcinoma (LUAD). The iBCR signature was tested for prognostication in lung cancer in two large datasets. (A&B) KM-Plotter (Affymetrix data) was used to investigate overall survival of lung adenocarcinoma (**A**) and squamous cell carcinoma (**B**). The iBCR signature shows a strong prognostic value in lung adenocarcinoma (LUAD). (**C**) Multivariate survival analysis was performed in KM-Plotter for the iBCR signature in lung cancer in comparison to available clinical indicators; histological type (lung adenocarcinoma vs. small cell lung cancer) and stage of disease. The iBCR signature outperformed these standard clinical indicators. (**D&E**) The TCGA data for LUAD (Illumina HiSeq RNA-seq data) were stratified by quartiles or tertiles for the iBCR signature expression to test the association of the iBCR signature with overall survival (D) and relapse-free survival (E), respectively. LUAD patients with high iBCR signature had poorest survival and suffered earlier recurrence and death compared to patients with lower iBCR signature expression. It should be noted that the TCGA data for squamous cell lung carcinoma were also investigated and there was no statistical significance for the association of the iBCR signature and survival, in agreement with the very weak association seen from the KM-Plotter data.

Figure 47: The sensitivity of breast cancer cell lines treated with 24 drugs according to the iBCR score. Breast cancer cell lines (10 cell lines) were cultured in the absence or presence of escalating doses of 24 small molecular anti-cancer drugs. This published study was re-analyzed to compare the sensitivity (calculated as the –logIC50) between high iBCR score cell lines (5 cell lines: BT-549, MDA-MB-231, MDA-MB-436, MDA-MB-468 and BT-20) to low iBCR score cell lines (5 cell lines: Hs.578T, BT-474, MCF-7, T-47D, and ZR-75-1). The iBCR scores were calculated from the Agro and TN scores using the published gene expression dataset for 51

breast cancer cell lines (Neve *et al.*, *Cancer Cell*, 2006). High iBCR score cell lines (red bars) were more sensitive than low iBCR score cell lines (white bars) to 13 drugs (shaded in grey) targeting 9 different kinases. Statistical comparison was performed in GraphPad® Prism using two tailed unpaired *t*-test.

Figure 48: Proteins and phosphoproteins associated with the iBCR mRNA gene signature. The iBCR score based on the mRNA expression of the 43 genes was used to stratify the patients in the TCGA breast cancer dataset as low, intermediate or high iBCR score. The reverse phase protein arrays (RPPA) from the TCGA breast cancer dataset (n=747 patients) were then compared between the three groups of patients according to the iBCR mRNA signature. (**A**) Overall survival of ER+ patients according to the iBCR mRNA signature. (**B**) Significantly up- or down-regulated proteins and phosphoproteins in ER+ patients in the low, intermediate and high iBCR score groups. (**C**) Overall survival of ER- according to the iBCR mRNA signature. (**D**) Significantly up- or down-regulated proteins and phosphoproteins in ER- patients in the low, intermediate and high iBCR score groups.

Figure 49: Prognostication of breast cancer patient survival by integrated mRNA and protein iBCR signature. The deregulated proteins and phosphoproteins in the three iBCR mRNA score groups were investigated for association with survival. Eight downregulated proteins and nine upregulated proteins were highly prognostic as a protein signature (iBCR protein signature). (**A**) Stratification of overall survival based on the iBCR protein signature (top row) and the integrated iBCR mRNA and protein signature (bottom row) in all breast cancer patients, ER+ and ER- cases. (**B**) Univariate and multivariate survival analysis using the Cox-proportional hazard model showing that the combined iBCR mRNA/Protein signature outperforms all clinicopathological indicators.

Figure 50: Proteins and phosphoproteins associated with the iBCR mRNA gene signature. (**A**) Stratification of lung adenocarcinoma overall survival based on the iBCR mRNA gene signature in the TCGA dataset (n=472 patients). (**B**) Comparison of proteins phosphoprotein levels between the tumors in the four quartiles of the iBCR mRNA gene signature. (**C**) Stratification of overall survival of lung adenocarcinoma patients based on six proteins deduced from panel (n=212 patients). (**D**) The combined iBCR mRNA/Protein signature stratifies the overall survival of lung adenocarcinoma patients (n=212 patients). (**E**) Multivariate Cox-proportional

hazard model for survival analysis showing that the combined iBCR mRNA/Protein score outperforms all clinicopathological indicators in lung adenocarcinoma.

Figure 51: The iBCR test is prognostic in Kidney renal clear cell carcinoma (KIRC) (left vertical panel), Skin cutaneous melanoma (SKCM) (middle vertical panel) and Uterine corpus endometrioid carcinoma (UCEC) (right vertical panel). (**A**) Stratification of overall survival based on the iBCR mRNA gene signature. (**B**) Stratification of overall survival based on iBCR protein signature. (**C**) Stratification of overall survival based on the combined iBCR mRNA/protein signature.

Figure 52: The iBCR test is prognostic in Ovarian adenocarcinoma (OVAC) (left vertical panel), Head & Neck squamous cell carcinoma (HNSC) (middle vertical panel) and Colon/Rectal Adenocarcinoma (COREAD) (right vertical panel). (**A**) Stratification of overall survival based on the iBCR mRNA gene signature. (**B**) Stratification of overall survival based on iBCR protein signature. (**C**) Stratification of overall survival based on the combined iBCR mRNA/protein signature.

Figure 53: The iBCR test is prognostic in Lower Grade Glioma (LGG) (left vertical panel), Bladder urothelial carcinoma (BLCA) (middle vertical panel) and Lung squamous cell carcinoma (LUSC) (right vertical panel). (**A**) Stratification of overall survival based on the iBCR mRNA gene signature. (**B**) Stratification of overall survival based on iBCR protein signature. (**C**) Stratification of overall survival based on the combined iBCR mRNA/protein signature.

Figure 54: The iBCR test is prognostic in (**A**) Kidney renal papillary cell carcinoma (KIRP). (**B**) Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), (**C**) Liver hepatocellular carcinoma (LIHC), (**D**) Pancreatic ductal adenocarcinoma (PDAC). For these cancer types, the TCGA datasets did not include RPPA arrays; only the iBCR mRNA gene expression test was used.

Figure 55: Protein-protein interaction of the iBCR mRNA/protein signature. The components of the iBCR test were analysed using the STRING database. The iBCR test (65 components) was significantly enriched (P=5.6E-14) for protein-protein interactions (129 interactions). The confidence of interactions is denoted by increasing thickness of the connecting blue lines. It is noteworthy that the components on the top right which do not show interactions contain several novel

genes that are not well characterised. The iBCR test is enriched for several biological functions related to the hallmarks of cancer (refer to Table 20).

Figure 56: The iBCR test as a companion diagnostic for immunotherapy. (A) Twelve genes from the iBCR test, particularly from the TN component, associated significantly with progression free survival of follicular lymphoma patients treated with pidilizumab + rituximab immunotherapy. The expression profile of the 12 genes in the tumours prior to treatment is shown (red indicates overexpression and green indicates underexpression). White and black boxes denote progression free survival or not, respectively. **(B)** A score was calculated based on the iBCR signature as the ratio of expression of the overexpressed genes to that of underexpressed genes. The survival of patients based on dichotomy across the median score was compared. The hazard ratio (HR) and the log-rank p-value for the survival comparison between low and high score tumors is shown in panel. **(C)** Eight patients were profiled pre- and post-treatment and the expression profiles of the 12 genes from the iBCR test were visualised in these patients. A trend for inversion of expression was observed and this was most evident for patient no. 9 who remained free of disease progression. **(D)** One gene was statistically significant in all patients post-treatment compared to that before treatment. This gene showed a marked different post-treatment vs. pre-treatment for patient no. 9. (E) Survival curve for the same patient group calculated from the gene signature labelled "Follicular Lymphoma" in Table 23. All conventions as per (B) above. Relapse-free survival of patients based on dichotomy across the median score is shown.

Figure 57: Network analysis of the genes from the meta-analysis of gene expression datasets.

Figure 58: Functional metagenes associate with breast cancer patient survival.

Figure 59: The iBCR test as a companion diagnostic for EGFR inhibition and multikinase inhibition. (A) Seventeen genes (see Table 23) from the iBCR test associated significantly with survival of colorectal cancer patients treated with the EGFR inhibitor cetuximab. (B) Sixteen genes (see Table 23) from the iBCR test associated significantly with overall survival of triple negative breast cancer patients treated with the EGFR inhibitor cetuximab combined with cisplatin. (C) Nineteen genes (see Table 23) from the iBCR test associated significantly with progression-free survival of lung cancer patients treated with the EGFR inhibitor erlotinib. (D) Twenty genes (see Table 23) from the iBCR test associated significantly with

progression-free survival of lung cancer patients treated with the multikinase inhibitor sorafenib.

DETAILED DESCRIPTION

The present invention is at least partly predicated on the discovery that there are genes that are associated with tumor aggressiveness and poor clinical outcome based on meta-analysis of published gene expression profiling. More particularly, the overexpression and/or underexpression of these genes (see Table 21) was found to be associated with poor survival in breast cancer. Network analysis using the Ingenuity Pathway Analysis (IPA®) software identified a number of networks or metagenes within these survival-associated genes that possess distinct biological functions as outlined in Table 21. A smaller subset of genes from each network or metagene which consistently associated with patient survival were then selected. The list of these genes and their corresponding functions are shown in Table 22. These genes were divided into six functional metagenes or networks.

The present invention is also at least partly predicated on the discovery that there are genes that are commonly de-regulated in particular subgroups that exemplify aggressive clinical behavior in triple-negative breast cancer (TNBC). More particularly, this is evident in TNBC compared to non-TNBC and normal breast, tumors associated with distant metastasis and/or death compared to their respective counterparts. Initially, a list of 206 recurrently deregulated genes was found to be particularly enriched for chromosomal instability (CIN) and estrogen receptor signaling (ER) metagenes. An aggressiveness score based on the ratio of the expression level of a CIN metagene relative to an ER metagene has been shown to identify aggressive tumors regardless of molecular subtype and clinico-pathologic indicators. Furthermore, depletion of proteins involved in kinetochore binding or chromosome segregation could be therapeutic and significantly reduced the survival of TNBC cell lines *in vitro*, particularly with regard to TTK. TTK inhibition with small molecule inhibitor affected the survival of TNBC cell lines. Also, *TTK* mRNA and protein levels were associated with aggressive tumor phenotypes. Mitosis-independent expression of TTK protein was prognostic in TNBC and other aggressive breast cancer subgroups, suggesting that protection of CIN/aneuploidy drives aggressiveness and treatment-resistance. The combination of TTK inhibition

with chemotherapy was effective *in vitro* in the treatment of cells that overexpress TTK, thus providing a therapeutic treatment for the protected CIN phenotype.

Additionally, the present invention is at least partly predicated on the discovery of a second signature of altered gene expression, including 21 overexpressed genes and 7 underexpressed genes, that is highly prognostic in patients with ER^- breast cancer, TNBC and basal-like breast cancer (BLBC). Indeed, integration of this 28 gene signature with the aforementioned aggressiveness score or gene signature produces an integrated score which is prognostic in breast cancer independent of ER status. Furthermore, the integrated score was prognostic in cancer broadly irrespective of the cancer type, as well as in specific types of cancer in addition to breast cancer, such as lung adenocarcinoma. Moreover, the 28 gene signature and the integrated score were both shown to be predictive of response to chemotherapy in breast cancer patients, as well as identify those ER^+ lymph node positive breast cancer patients who would benefit from endocrine therapy. Altered expression of the signatures described herein was also predictive of sensitivity in cancer cell lines and clinically to a range of anticancer therapeutics, and in particular, molecularly targeted inhibitors.

The inventors of the present invention have also identified a protein signature that is highly prognostic in a range of cancers, including breast cancer and lung adenocarcinoma. Furthermore, this protein signature may be integrated with the aforementioned 28 gene signature and aggressive gene signature to provide a robust prognostic indicator in cancer that was shown to outperform known clinicopathological indicators.

In one aspect, the invention relates to a method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of a plurality of overexpressed genes and an expression level of a plurality of underexpressed genes in one or more cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or more metagenes selected from the group consisting of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein

Synthesis/Modification metagene and a Multiple Networks metagene, wherein: a higher relative expression level of the plurality of the overexpressed genes compared to the plurality of the underexpressed genes indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the plurality of the overexpressed genes compared to the plurality of the underexpressed genes indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

In a futher aspect, the invention relates to a method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of a plurality of overexpressed genes and an expression level of a plurality of underexpressed genes in one or more cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or more metagenes selected from the group consisting of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene, wherein: a higher relative expression level of the plurality of overexpressed genes compared to the plurality of underexpressed genes indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the plurality of overexpressed genes compared to the plurality of underexpressed genes indicates or correlates with a more favourable cancer prognosis.

In one embodiment of the above aspects, the plurality of overexpressed genes and/or the plurality of underexpressed genes are selected from one of the metagenes. In an alternative embodiment, the plurality of overexpressed genes and/or the plurality of underexpressed genes are selected from a plurality of the metagenes.

Suitably, for the method of the above aspects the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene

and/or the Multiple Networks metagene comprise one or more genes listed in Table 21.

In another aspect, the invention relates to a method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of a plurality of overexpressed genes and an expression level of a plurality of underexpressed genes in one or more cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or more metagenes selected from the group consisting of a Metabolism metagene, a Signalling metagene, a Development and Growth metagene, a Chromosome Segregation/Replication metagene, an Immune Response metagene and a Protein Synthesis/Modification metagene, wherein: a higher relative expression level of the plurality of the overexpressed genes compared to the plurality of the underexpressed genes indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the plurality of the overexpressed genes compared to the plurality of the underexpressed genes indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level

In yet another aspect, the invention relates to a method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of a plurality of overexpressed genes and an expression level of a plurality of underexpressed genes in one or more cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or more metagenes selected from the group consisting of a Metabolism metagene, a Signalling metagene, a Development and Growth metagene, a Chromosome Segregation/Replication metagene, an Immune Response metagene and a Protein Synthesis/Modification metagene, wherein: a higher relative expression level of the plurality of overexpressed genes compared to the plurality of underexpressed genes indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the plurality of overexpressed genes compared to the plurality of underexpressed genes indicates or correlates with a more favourable cancer prognosis.

Suitably, the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication

metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene comprise one or more genes listed in Table 21.

In particular embodiments of the method of the two aforementioned aspects, the plurality of overexpressed genes and the plurality of underexpressed genes are from one or more of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene.According to the method of the above aspects, the step of comparing an expression level of a plurality of overexpressed genes and an expression level of a plurality of underexpressed genes includes comparing an average expression level of the plurality of overexpressed genes and an average expression level of the plurality of underexpressed genes. This may include calculating a ratio of the average expression level of the plurality of overexpressed genes and the average expression level of the plurality of underexpressed genes. Suitably, the ratio provides an aggressiveness score which is indicative of, or correlates with, cancer aggressiveness and a less favourable prognosis. Alternatively, the step of comparing an expression level of a plurality of overexpressed genes and an expression level of a plurality of underexpressed genes includes comparing the sum of expression levels of the plurality of overexpressed genes and the sum of expression levels of the plurality of underexpressed genes. This may include calculating a ratio of the sum of expression levels of the plurality of overexpressed genes and the sum of expression levels of the plurality of underexpressed genes.

For the purposes of this invention, by "*isolated*" is meant material that has been removed from its natural state or otherwise been subjected to human manipulation. Isolated material may be substantially or essentially free from components that normally accompany it in its natural state, or may be manipulated so as to be in an artificial state together with components that normally accompany it in its natural state. Isolated material may be in native, chemical synthetic or recombinant form.

As used herein a "*gene*" is a nucleic acid which is a structural, genetic unit of a genome that may include one or more amino acid-encoding nucleotide sequences

and one or more non-coding nucleotide sequences inclusive of promoters and other 5' untranslated sequences, introns, polyadenylation sequences and other 3' untranslated sequences, although without limitation thereto. In most cellular organisms a gene is a nucleic acid that comprises double-stranded DNA.

Non-limiting examples of genes are set forth herein, particularly in Tables 4, 21 and 22, which include Accession Numbers referencing the nucloetide sequence of the gene, or its encoded protein, as are well understood in the art.

The term "*nucleic acid*" as used herein designates single- or double-stranded DNA and RNA. DNA includes genomic DNA and cDNA. RNA includes mRNA, RNA, RNAi, siRNA, cRNA and autocatalytic RNA. Nucleic acids may also be DNA-RNA hybrids. A nucleic acid comprises a nucleotide sequence which typically includes nucleotides that comprise an A, G, C, T or U base. However, nucleotide sequences may include other bases such as inosine, methylycytosine, methylinosine, methyladenosine and/or thiouridine, although without limitation thereto.

Also included are, "*variant*" nucleic acids that include nucleic acids that comprise nucleotide sequences of naturally occurring (e.g., allelic) variants and orthologs (*e.g.*, from a different species). Preferably, nucleic acid variants share at least 70% or 75%, preferably at least 80% or 85% or more preferably at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity with a nucleotide sequence disclosed herein.

Also included are nucleic acid fragments. A "*fragment*" is a segment, domain, portion or region of a nucleic acid, which respectively constitutes less than 100% of the nucleotide sequence. A non-limilting example is an amplification product or a primer or probe. In particular embodiments, a nucleic acid fragment may comprise, for example, at least 10, 15, 20, 25, 30 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475 and 500 contiguous nucleotides of said nucleic acid.

As used herein, a "*polynucleotide*" is a nucleic acid having eighty (80) or more contiguous nucleotides, while an "*oligonucleotide*" has less than eighty (80) contiguous nucleotides. A "*probe*" may be a single or double-stranded oligonucleotide or polynucleotide, suitably labeled for the purpose of detecting complementary sequences in Northern or Southern blotting, for example. A "*primer*" is usually a single-stranded oligonucleotide, preferably having 15-50 contiguous nucleotides, which is capable of annealing to a complementary nucleic acid

"template" and being extended in a template-dependent fashion by the action of a DNA polymerase such as *Taq* polymerase, RNA-dependent DNA polymerase or Sequenase™. A "*template*" nucleic acid is a nucleic acid subjected to nucleic acid amplification.

It will be appreciated that the "*overexpressed*" genes or proteins referred to herein are genes or proteins that are expressed at a higher level in a cancer cell or tissue compared to a corresponding normal or otherwise non-cancerous cell or tissue or reference/control level or sample.

It will be appreciated that the "*underexpressed*" genes or proteins referred to herein are genes or proteins that are expressed at a lower level in a cancer cell or tissue compared to a corresponding normal or otherwise non-cancerous cell or tissue or reference/control level or sample.

In certain embodiments, the "*overexpressed*" and "*underexpressed*" genes referred to herein may form, or be components of, a metagene.

As used herein, a "*metagene*" is a grouping, cohort or network of a plurality of different genes that display a common, shared or aggregate expression profile, expression level or other expression characteristics that associate with, or are indicative of, a particular function or phenotype. Non-limiting examples include a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene. Table 21 provides non-limiting examples of genes that are components of the aforementioned twelve metagenes. Further non-limiting examples include a Metabolism metagene, a Signalling metagene, a Development and Growth metagene, a Chromosome Segregation/Replication metagene, an Immune Response metagene and a Protein Synthesis/Modification metagene. Table 22 provides non-limiting examples of genes that are components of the aforementioned six metagenes.

In particular embodiments, the plurality of overexpressed genes and/or the plurality of underexpressed genes are selected from one of the metagenes. In this regard, the plurality of overexpressed genes and/or the plurality of underexpressed genes are selected from the same metagene. By way of example, the plurality of

overexpressed genes or the plurality of underexpressed genes may be only from one of the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and the Multiple Networks metagene. In a further example, both the plurality of overexpressed genes and the plurality of underexpressed genes may be only from one of the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and the Multiple Networks metagene.

Alternatively, the plurality of overexpressed genes and/or the plurality of underexpressed genes are selected from a plurality of the metagenes described herein.

By "*aggressiveness*" and "*aggressive*" is meant a property or propensity for a cancer to have a relatively poor prognosis due to one or more of a combination of features or factors including: at least partial resistance to therapies available for cancer treatment; invasiveness; metastatic potential; recurrence after treatment; and a low probability of patient survival, although without limitation thereto.

Cancers may include any aggressive or potentially aggressive cancers, tumours or other malignancies such as listed in the NCI Cancer Index at http://www.cancer.gov/cancertopics/alphalist, including all major cancer forms such as sarcomas, carcinomas, lymphomas, leukaemias and blastomas, although without limitation thereto. These may include breast cancer, lung cancer inclusive of lung adenocarcinoma, cancers of the reproductive system inclusive of ovarian cancer, cervical cancer, uterine cancer and prostate cancer, cancers of the brain and nervous system, head and neck cancers, gastrointestinal cancers inclusive of colon cancer, colorectal cancer and gastric cancer, liver cancer, kidney cancer, skin cancers such as melanoma and skin carcinomas, blood cell cancers inclusive of lymphoid cancers and myelomonocytic cancers, cancers of the endocrine system such as pancreatic

cancer and pituitary cancers, musculoskeletal cancers inclusive of bone and soft tissue cancers, although without limitation thereto.

In certain embodiments, cancers include breast cancer, bladder cancer, colorectral cancer, glioblastoma, lower grade glioma, head & neck cancer, kidney cancer, liver cancer, lung adenocarcinoma, acute myeloid leukaemia, pancreatic cancer, adrenocortical cancer, melanoma and lung squamous cell carcinoma.

Breast cancers include all aggressive breast cancers and cancer subtypes such as triple negative breast cancer, grade 2 breast cancer, grade 3 breast cancer, lymph node positive (LN^+) breast cancer, HER2 positive ($HER2^+$) breast cancer and ER positive (ER^+) breast cancer, although without limitation thereto.

As used herein, "*triple negative breast cancer*" (TNBC) is an often aggressive breast cancer subtype lacking or having significantly reduced expression of estrogen receptor (ER) protein, progesterone receptor (PR) protein and HER2 protein. TNBC and other aggressive breast cancers are typically insensitive to some of the most effective therapies available for breast cancer treatment including HER2-directed therapy such as trastuzumab and endocrine therapies such as tamoxifen and aromatase inhibitors.

As used herein, a gene expression level may be an absolute or relative amount of an expressed gene or gene product inclusive of nucleic acids such as RNA, mRNA and cDNA and protein.

As would be appreciated by the skilled artisan, the present invention need not be limited to comparing the expression level of the overexpressed genes and/or proteins with the expression level of the underexpressed genes and/or proteins provided herein. Accordingly, in particular embodiments, the expression level of the overexpressed and/or underexpressed genes and/or proteins is compared to a control level of expression, such as the level of gene and/or protein expression of a "housekeeping" gene in one or more cancer cells, tissues or organs of the mammal.

In further embodiments, the expression level of the overexpressed and/or underexpressed genes and/or proteins is compared to a threshold level of expression, such as a level of gene and/or protein expression in non-aggressive cancerous tissue. A threshold level of expression is generally a quantified level of expression of a particular gene or set of genes, including gene products thereof. Typically, an expression level of a gene or set of genes in a sample that exceeds or falls below the threshold level of expression is predictive of a particular disease state or outcome.

The nature and numerical value (if any) of the threshold level of expression will vary based on the method chosen to determine the expression the one or more genes or proteins used in determining, for example, a prognosis, the aggressiveness and/or response to anticancer therapy, in the mammal. In light of this disclosure, any person of skill in the art would be capable of determining the threshold level of gene/protein expression in a mammal sample that may be used in determining, for example, a prognosis, the aggressiveness and/or response to anticancer therapy, using any method of measuring gene or protein expression known in the art, such as those described herein. In one embodiment, the threshold level is a mean and/or median expression level (median or absolute) of the overexpressed and/or underexpressed genes and/or proteins in a reference population, that, for example, have the same cancer type, subgroup, stage and/or grade as said mammal for which the expression level is determined. Additionally, the concept of a threshold level of expression should not be limited to a single value or result. In this regard, a threshold level of expression may encompass multiple threshold expression levels that could signify, for example, a high, medium, or low probability of, for example, progression free survival.

By "*protein*" is meant an amino acid polymer. The amino acids may be natural or non-natural amino acids, D- or L- amino acids as are well understood in the art. As would be appreciated by the skilled person, the term "protein" also includes within its scope phosphorylated forms of a protein (*i.e.*, phosphoproteins).

Also provided are protein "variants" such as natrually occurring (eg allelic variants) and orthologs. Preferably, protein variants share at least 70% or 75%, preferably at least 80% or 85% or more preferably at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity with an amino acid sequence disclosed herein.

Also provided are protein fragments, inclusive of peptide fragments thqat comprise less than 100% of an entire amino acid sequence. In particular embodiments, a protein fragment may comprise, for example, at least 10, 15, 20, 25, 30 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375 and 400 contiguous amino acids of said protein.

A "*peptide*" is a protein having no more than fifty (50) amino acids.

A "*polypeptide*" is a protein having more than fifty (50) amino acids.

It would be appreciated that in addition to comparing the expression levels of one or more genes or proteins, the methods of the present invention may further include the step of determining, assessing, evaluating, assaying or measuring the expression level of one or more of the overexpressed genes, the underexpressed genes, the overexpressed proteins and/or the underexpressed proteins described herein. The terms "*determining*", "*measuring*", "*evaluating*", "*assessing*" and "*assaying*" are used interchangeably herein and may include any form of measurement known in the art, such as those described hereinafter.

Determining, assessing, evaluating, assaying or measuring nucleic acids such as RNA, mRNA and cDNA may be performed by any technique known in the art. These may be techniques that include nucleic acid sequence amplification, nucleic acid hybridization, nucleotide sequencing, mass spectroscopy and combinations of any these.

Nucleic acid amplification techniques typically include repeated cycles of annealing one or more primers to a "template" nucleotide sequence under appropriate conditions and using a polymerase to synthesize a nucleotide sequence complementary to the target, thereby "amplifying" the target nucleotide sequence. Nucleic acid amplification techniques are well known to the skilled addressee, and include but are not limited to polymerase chain reaction (PCR); strand displacement amplification (SDA); rolling circle replication (RCR); nucleic acid sequence-based amplification (NASBA), Q-β replicase amplification; helicase-dependent amplification (HAD); loop-mediated isothermal amplification (LAMP); nicking enzyme amplification reaction (NEAR) and recombinase polymerase amplification (RPA), although without limitation thereto. As generally used herein, an *"amplification product"* refers to a nucleic acid product generated by a nucleic acid amplification technique.

PCR includes quantitative and semi-quantitative PCR, real-time PCR, allele-specific PCR, methylation-specific PCR, asymmetric PCR, nested PCR, multiplex PCR, touch-down PCR and other variations and modifications to "basic" PCR amplification.

Nucleic acid amplification techniques may be performed using DNA or RNA extracted, isolated or otherwise obtained from a cell or tissue source. In other embodiments, nucleic acid amplification may be performed directly on appropriately treated cell or tissue samples.

Nucleic acid hybridization typically includes hybridizing a nucleotide sequence (typically in the form of a probe) to a target nucleotide sequence under appropriate conditions, whereby the hybridized probe-target nucleotide sequence is subsequently detected. Non-limiting examples include Northern blotting, slot-blotting, *in situ* hybridization and fluorescence resonance energy transfer (FRET) detection, although without limitation thereto. Nucleic acid hybridization may be performed using DNA or RNA extracted, isolated, amplified or otherwise obtained from a cell or tissue source or directly on appropriately treated cell or tissue samples.

It will also be appreciated that a combination of nucleic acid amplification and nucleic acid hybridization may be utilized.

Determining, assessing, evaluating, assaying or measuring protein levels may be performed by any technique known in the art that is capable of detecting cell- or tissue-expressed proteins whether on the cell surface or intracellularly expressed, or proteins that are isolated, extracted or otherwise obtained from the cell of tissue source. These techniques include antibody-based detection that uses one or more antibodies which bind the protein, electrophoresis, isoelectric focussing, protein sequencing, chromatographic techniques and mass spectroscopy and combinations of these, although without limitation thereto. Antibody-based detection may include flow cytometry using fluorescently-labelled antibodies that bind the protein, ELISA, immunoblotting, immunoprecipitation, *in situ* hybridization, immunohistochemistry and immuncytochemistry, although without limitation thereto. Suitable techniques may be adapted for high throughput and/or rapid analysis such as using protein arrays such as a TissueMicroArray™ (TMA), MSD MultiArrays™ and multiwell ELISA, although without limitation thereto.

In certain embodiments, a gene expression level may be assessed indirectly by the measurement of a non-coding RNA, such as miRNA, that regulate gene expression. MicroRNAs (miRNAs or miRs) are post-transcriptional regulators that bind to complementary sequences in the 3' untranslated regions (3' UTRs) of target mRNA transcripts, usually resulting in gene silencing. miRNAs are short RNA molecules, on average only 22 nucleotides long. The human genome may encode over 1000 miRNAs, which may target about 60% of mammalian genes and are abundant in many human cell types. Each miRNA may alter the expression of hundreds of individual mRNAs. In particular, miRNAs may have multiple roles in negative regulation (*e.g.*, transcript degradation and sequestering, translational

suppression) and/or positive regulation (*e.g.*, transcriptional and translational activation). Additionally, aberrant miRNA expression has been implicated in various types of cancer.

In this regard, an average expression level, or alternatively a sum of the expression levels, may be calculated for the plurality of overexpressed genes and for the plurality of underexpressed genes, to thereby produce or calculate a ratio.

Accordingly, determining cancer aggressiveness and/or a prognosis for a cancer patient in certain embodiments of the present invention further includes determining the ratio of the expression level (*e.g.* an average or sum of the expression level) of the plurality of overexpressed genes to the expression level (*e.g.* an average or sum of the expression level) of the plurality of underexpressed genes.

In another aspect of the invention relates to a method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of a plurality of overexpressed genes associated with chromosomal instability and an expression level of a plurality of underexpressed genes associated with estrogen receptor signalling in one or more cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the plurality of overexpressed genes associated with chromosomal instability compared to the plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level expression level of the plurality of overexpressed genes associated with chromosomal instability compared to the plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

In yet another aspect of the invention relates to a method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of a plurality of overexpressed genes associated with chromosomal instability and an expression level of a plurality of underexpressed genes associated with estrogen receptor signalling in the mammal, wherein: a higher relative expression level of the plurality of overexpressed genes associated with chromosomal instability compared to the plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the plurality

of overexpressed genes associated with chromosomal instability compared to the plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with a more favourable cancer prognosis.

Non-limiting examples of genes in a chromosomal instability (CIN) metagene include *ATP6V1C1, RAP2A, CALM1, COG8, HELLS, KDM5A, PGK1, PLCH1, CEP55, RFC4, TAF2, SF3B3, GPI, PIR, MCM10, MELK, FOXM1, KIF2C, NUP155, TPX2, TTK, CENPA, CENPN, EXO1, MAPRE1, ACOT7, NAE1, SHMT2, TCP1, TXNRD1, ADM, CHAF1A and SYNCRIP* genes, although without limitation thereto; and an estrogen receptor signalling (ER) metagene may comprise *BTG2, PIK3IP1, SEC14L2, FLNB, ACSF2, APOM, BIN3, GLTSCR2, ZMYND10, ABAT, BCAT2, SCUBE2, RUNX1, LRRC48, MYBPC1, BCL2, CHPT1, ITM2A, LRIG1, MAPT, PRKCB, RERE, ABHD14A, FLT3, TNN, STC2, BATF, CD1E, CFB, EVL, FBXW4, ABCB1, ACAA1, CHAD, PDCD4, RPL10, RPS28, RPS4X, RPS6, SORBS1, RPL22 and RPS4XP3* genes, although without limitation thereto. Table 4 provides further examples of genes that are components of a CIN metagene or that are components of an ER metagene.

An average expression level may be calculated for the CIN metagene and for the ER metagene, to thereby produce or calculate a ratio.

Alternatively, a sum of expression levels may be calculated for the CIN metagene and for the ER metagene, to thereby produce or calculate a ratio.

In certain embodiments, a higher or increased ratio of the average or sum of expression levels of a CIN metagene relative to an ER metagene is associated with, correlates with or is indicative of, higher or increased cancer aggressiveness.

Thus, some embodiments of the invention provide an "*aggressiveness score*" which is the ratio of CIN metagene expression level (*e.g.* average or sum of expression of CIN genes) to an ER metagene expression level (*e.g* average or sum of expression of ER genes).

Accordingly, embodiments of the aforementioned aspects of the invention include determining, assessing or measuring an expression level of a plurality of overexpressed genes associated with chromosomal instability and determining, assessing or measuring an expression level of a plurality of underexpressed genes associated with estrogen receptor signalling. In this regard, reference is made to Table 4 which provides a listing of 206 genes that include genes associated with chromosomal instability and genes associated with estrogen receptor signalling.

Preferably, the chromosomal instability genes are of a CIN metagene, comprising genes such as *ATP6V1C1, RAP2A, CALM1, COG8, HELLS, KDM5A, PGK1, PLCH1, CEP55, RFC4, TAF2, SF3B3, GPI, PIR, MCM10, MELK, FOXM1, KIF2C, NUP155, TPX2, TTK, CENPA, CENPN, EXO1, MAPRE1, ACOT7, NAE1, SHMT2, TCP1, TXNRD1, ADM, CHAF1A and* SYNCRIP, although without limitation thereto. In one preferred embodiment, the chromosomal instability genes are selected from the group consisting of MELK, *MCM10*, *CENPA*, *EXO1*, *TTK* and *KIF2C*.

Preferably, the estrogen receptor signalling genes are of an ER metagene comprising genes such as *BTG2, PIK3IP1, SEC14L2, FLNB, ACSF2, APOM, BIN3, GLTSCR2, ZMYND10, ABAT, BCAT2, SCUBE2, RUNX1, LRRC48, MYBPC1, BCL2, CHPT1, ITM2A, LRIG1, MAPT, PRKCB, RERE, ABHD14A, FLT3, TNN, STC2, BATF, CD1E, CFB, EVL, FBXW4, ABCB1, ACAA1, CHAD, PDCD4, RPL10, RPS28, RPS4X, RPS6, SORBS1, RPL22 and RPS4XP3*, although without limitation thereto. In one preferred embodiment, the estrogen receptor signalling genes are selected from the group consisting of *MAPT* and *MYB*.

In certain embodiments, the method of the aforementioned two aspects further includes the step of comparing an expression level of one or more other overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and an expression level of one or more other underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*, in one or more cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or more other overexpressed genes compared to the one or more other underexpressed genes indicates or correlates with higher aggressiveness of the cancer and/or a less favourable cancer prognosis; and/or a lower relative expression level of the one or more other overexpressed genes compared to the one or more other underexpressed genes indicates or correlates with lower aggressiveness of the cancer and/or a more favourable cancer prognosis compared to a mammal having a higher expression level.

In one embodiment, the one or more other overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*, *CD55*, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593*.

In one embodiment, the one or more other underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

In this regard, an average expression level, or alternatively a sum of the expression levels, may be calculated for the one or more other overexpressed genes and for the one or more other underexpressed genes, to thereby produce or calculate a ratio.

Accordingly, determining cancer aggressiveness and/or a prognosis for a cancer patient in certain embodiments of the present invention further includes determining the ratio of the expression level (*e.g.* an average or sum of the expression level) of the one or more other overexpressed genes to the expression level (*e.g.* an average or sum of the expression level) of the one or more other underexpressed genes.

Detection and/or measurement of expression of the one or more other overexpressed genes and the one or more other underexpressed genes may be performed by any of those methods or combinations thereof described herein (*e.g* measuring mRNA levels or an amplified cDNA copy thereof and/or by measuring a protein product thereof), albeit without limitation thereto.

Suitably, the comparison of the expression level of the plurality of overexpressed genes associated with chromosomal instability and the expression level of the plurality of underexpressed genes associated with estrogen receptor signalling is integrated with the comparison of the expression level of the one or more other overexpressed genes and the expression level of the one or more other underexpressed genes to derive a first integrated score. In particular embodiments, this may include deriving the first integrated score, at least in part, by addition, subtraction, multiplication, division and/or exponentiation.

By way of example, the comparison of the expression level of the plurality of overexpressed genes associated with chromosomal instability and the expression level of the plurality of underexpressed genes associated with estrogen receptor signalling may be added to, subtracted from, multiplied by, divided by and/or raised

to the power of the comparison of the expression level of the one or more other overexpressed genes and the expression level of the one or more other underexpressed genes to derive the first integrated score. Alternatively, the comparison of the expression level of the one or more other overexpressed genes and the expression level of the one or more other underexpressed genes may be added to, subtracted from, multiplied by, divided by and/or raised to the power of the comparison of the expression level of the plurality of overexpressed genes associated with chromosomal instability and the expression level of the plurality of underexpressed genes associated with estrogen receptor signalling to derive the first integrated score.

In a particular preferred embodiment, the first integrated score is derived by exponentiation, wherein the comparison of the expression level of the one or more other overexpressed genes and the expression level of the one or more other underexpressed genes is raised to the power of the comparison of the expression level of the plurality of overexpressed genes associated with chromosomal instability and the expression level of the plurality of underexpressed genes associated with estrogen receptor signalling.

As would be appreciated by the skilled person, the other overexpressed and underexpressed genes described herein may not necessarily be associated with chromosomal instability and estrogen receptor signalling respectively.

In a further aspect, the invention provides a method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or more overexpressed genes, wherein the one or more overexpressed genes are selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and an expression level of one or more underexpressed genes, wherein the one or more underexpressed genes are selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*, in one or more cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or more overexpressed genes compared to

the one or more underexpressed genes indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or more overexpressed genes compared to the one or more underexpressed genes indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

In one embodiment, the one or more overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*, *CD55*, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593*.

In one embodiment, the one or more underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

In yet another aspect, the invention provides a method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or more overexpressed genes, wherein the one or more overexpressed genes are selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and an expression level of one or more underexpressed genes, wherein the one or more underexpressed genes are selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*, in one or more cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or more overexpressed genes compared to the one or more underexpressed genes indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or more overexpressed genes compared to the one or more underexpressed genes indicates or correlates with a more favourable cancer prognosis compared to a mammal having a higher expression level.

In one embodiment, the one or more overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*,

CD55, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593*.

In one embodiment, the one or more underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

In particular embodiments, the method of the aforementioned aspects further includes the step of comparing an expression level of one or more overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and an expression level of one or more underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or more cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or more overexpressed proteins compared to the one or more underexpressed proteins indicates or correlates with higher aggressiveness of the cancer and/or a less favourable cancer prognosis; and/or a lower relative expression level of the one or more overexpressed proteins compared to the one or more underexpressed proteins indicates or correlates with lower aggressiveness of the cancer and/or a more favourable cancer prognosis compared to a mammal having a higher expression level.

As would be appreciated by the skilled artisan, the expression level of one or more of the overexpressed proteins and/or one or more of the underexpressed proteins described herein may include one or more phosphorylated forms of said proteins (*i.e.*, a phosphoprotein). In one embodiment, EIF4EBP1 is or comprises one or more phosphoproteins selected from the group consisting of pEIF4EBP1^{S65}, pEIF4EBP1^{T37}, pEIF4EBP1^{T46} and pEIF4EBP1^{T70}. In one embodiment, EGFR is or comprises one or more phosphoproteins selected from the group consisting of pEGFR^{Y1068} and pEGFR^{Y1173}. In one embodiment, HER3 is or comprises pHER3^{Y1289}. In one embodiment, AKT1 is or comprises one or more phosphoproteins selected from the group consisting of pAKT1^{S473} and pAKT1^{T308}. In one embodiment, NFKB1 is or comprises pNFKB1^{S536}. In one embodiment, HER2 is or comprises pHER2^{Y1248}. In one embodiment, ESR1 is or comprises pESR1^{S118}. In one embodiment, PEA15 is or comprises pPEA15^{S116}. In one embodiment, RPS6 is

or comprises one or more phosphoproteins selected from the group consisting of $pRPS6^{S235}$, $pRPS6^{S236}$, $pRPS6^{S240}$ and $pRPS6^{S244}$.

An average or sum of the expression levels may be calculated for the overexpressed genes, the underexpressed genes, the overexpressed proteins and/or the underexpressed proteins, to thereby produce or calculate a ratio.

Thus, in certain embodiments of the present invention determining cancer aggressiveness and/or a prognosis for a cancer patient includes determining (i) the ratio of the expression level (*e.g.* an average or sum of the expression level) of the one or more overexpressed genes to the expression level (*e.g.* an average or sum of the expression level) of the one or more underexpressed genes; and/or (ii) the ratio of the expression level (*e.g.* an average or sum of the expression level) of the one or more overexpressed proteins to the expression level (*e.g.* an average or sum of the expression level) of the one or more underexpressed proteins.

Detection and/or measurement of expression of the overexpressed proteins and the underexpressed proteins may be performed by any of those methods or combinations thereof hereinbefore described, albeit without limitation thereto.

Suitably, the comparison of the expression level of the one or more overexpressed proteins and the expression level of the one or more underexpressed proteins is to thereby derive an integrated score. In one particular embodiment, the comparison of the expression level of the one or more overexpressed proteins and the expression level of the one or more underexpressed proteins is integrated with:

(i) the comparison of the expression level of the overexpressed genes associated with chromosomal instability and the expression level of the underexpressed genes associated with estrogen receptor signalling to derive a second integrated score; or

(ii) the first integrated score to derive a third integrated score; or

(iii) the comparison of the expression level of the overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1* and the expression level of the underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C,*

NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3 to derive a fourth integrated score; or

(iv) the comparison of the expression level of the overexpressed genes and an expression level of the underexpressed genes, wherein the genes are from one or more of the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and/or the Multiple Networks metagene, to derive a fifth integrated score; or

(v) the comparison of the expression level of the overexpressed genes and an expression level of the underexpressed genes, wherein the genes are from one or more of the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene, to derive a sixth integrated score.

In particular embodiments, the second, third, fourth, fifth and/or sixth integrated scores are derived, at least in part, by addition, subtraction, multiplication, division and/or exponentiation. By way of example, the comparison of the expression level of the one or more overexpressed proteins and the expression level of the one or more underexpressed proteins may be added to, subtracted from, multiplied by, divided by and/or raised to the power of (i) the comparison of the expression level of the plurality of overexpressed genes associated with chromosomal instability and the expression level of the plurality of underexpressed genes associated with estrogen receptor signalling; or (ii) the first integrated score. Alternatively, the comparison of the expression level of the plurality of overexpressed genes associated with chromosomal instability and the expression level of the plurality of underexpressed genes associated with estrogen receptor signalling or the first integrated score may be

added to, subtracted from, multiplied by, divided by and/or raised to the power of the comparison of the expression level of the one or more overexpressed proteins and the expression level of the one or more underexpressed proteins.

In a further aspect, the invention provides a method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or more overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and an expression level of one or more underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or more cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or more overexpressed proteins compared to the one or more underexpressed proteins indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or more overexpressed proteins compared to the one or more underexpressed proteins indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

In a related aspect, the invention provides a method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or more overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and an expression level of one or more underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or more cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or more overexpressed proteins compared to the one or more underexpressed proteins indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or more overexpressed proteins compared to the one or more underexpressed proteins indicates or correlates with a more favourable cancer prognosis compared to a mammal having a higher expression level.

In particular embodiments of the two aforementioned aspects, one or more of the overexpressed proteins and/or one or more of the underexpressed proteins are or comprise a phosphoprotein hereinbefore described.

An average or sum of the expression levels may be calculated for the one or more overexpressed proteins and the one or more underexpressed proteins, to thereby produce or calculate a ratio as hereinbefore described.

This information with respect to the aggressiveness and/or prognosis of a patient's cancer may prove useful to a physician and/or clinician in determining the most effective course of treatment. A determination of the likelihood for a cancer relapse or of the likelihood of metastasis can assist the physician and/or clinician in determining whether a more conservative or a more radical approach to therapy should be taken. As such, a prognosis may provide for the selection and classification of patients who are predicted to benefit from a given therapeutic regimen.

Accordingly, another aspect of the invention provides a method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of a plurality of overexpressed genes and an expression level of a plurality of underexpressed genes in one or more cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or more metagenes selected from the group consisting of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene, wherein an altered or modulated relative expression level of the overexpressed genes compared to the underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

As would be understood by the skilled person, the relative expression level of a gene or protein may be deemed to be "*altered*" or "*modulated*" when the expression level is higher/increased or lower/decreased when compared to a control or reference sample or expression level, such as a threshold level. In one embodiment, a relative expression level may be classified as high if it is greater than a mean and/or median relative expression level of a reference population and a relative expression

level may be classified as low if it is less than the mean and/or median relative expression level of the reference population. In this regard, a reference population may be a group of subjects who have the same cancer type, subgroup, stage and/or grade as said mammal for which the relative expression level is determined.

Suitably, for the present aspect the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and/or the Multiple Networks metagene comprise one or more genes listed in Table 21.

In a related aspect, the invention provides a method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of a plurality of overexpressed genes and an expression level of a plurality of underexpressed genes in one or more cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or more metagenes selected from the group consisting of a Metabolism metagene, a Signalling metagene, a Development and Growth metagene, a Chromosome Segregation/Replication metagene, an Immune Response metagene and a Protein Synthesis/Modification metagene, wherein an altered or modulated relative expression level of the overexpressed genes compared to the underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

In one embodiment of the two aforementioned aspects, the plurality of overexpressed genes and/or the plurality of underexpressed genes are selected from one of the metagenes. In an alternative embodiment, the plurality of overexpressed genes and/or the plurality of underexpressed genes are selected from a plurality of the metagenes.

Suitably, the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene comprise one or more genes listed in Table 22.

In particular embodiments, the plurality of overexpressed genes and the plurality of underexpressed genes are from one or more of a Carbohydrate/Lipid

Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene.

In a related aspect, the invention provides a method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of determining an expression level of one or more genes associated with chromosomal instability (CIN) in one or more cancer cells of the mammal, wherein a higher expression level indicates or correlates with relatively increased responsiveness of the cancer to the anti-cancer treatment.

As will be described in more detail, overexpression of some CIN genes may be predictive of the responsiveness of a cancer to an anti-cancer treatment, particularly although not exclusively when overexpressed by non-mitotic cancer cells. In this context, by "non-mitotic" means that the cancer cell is not in the mitotic or "M phase" of the cell cycle. Preferably, the non-mitotic cancer cells are in interphase. Broadly, any overexpressed CIN gene set forth Table 4 may be predictive of the responsiveness of a cancer to an anti-cancer treatment. In particular embodiments, the CIN gene is selected from the group consisting of: *TTK*, *CEP55*, *FOXM1* and *SKIP2*. In a particularly preferred embodiment, the CIN gene is selected from the group consisting of: *TTK*, *CEP55*, *FOXM1* and *SKIP2* and the cancer is breast cancer. In this regard, the inventors have shown that "bulk" measurements of extracted CIN gene mRNA or encoded protein do not provide a useful indication of whether overexpression of the CIN gene may be predictive of the responsiveness of a cancer to an anti-cancer treatment. More particularly, detection of CIN gene expression by individual cancer cells, particularly non-mitotic or interphase cancer cells, provides a more powerful indication of the responsiveness of a cancer to an anti-cancer treatment.

As previously described, detection and/or measurement of expression of the CIN gene may be performed by measuring RNA (*e.g* mRNA or an amplified cDNA copy thereof) or by measuring a protein product of a CIN gene. In a particularly preferred embodiment, a protein product of a CIN gene is detected or measured by immunohistochemistry. Typically, although not exclusively, a preferred

immunohistochemistry method includes binding an antibody to the protein product of a CIN gene expressed by a cell or tissue and subsequent detection of the bound antibody. By way of example only, the antibody may be unlabelled, directly labelled with an enzyme such as horseradish peroxidase, alkaline phosphatase or glucose oxidase or directly labelled with biotin or digoxigenin. In embodiments where the antibody is unlabelled, a secondary antibody (labelled such as described above) may be used to detect the bound antibody. Biotinylated antibodies may be detected using avidin complexed with an enzyme such as horseradish peroxidase, alkaline phosphatase or glucose oxidase. Suitable enzyme substrates include diaminobanzidine (DAB), permanent red, 3-ethylbenzthiazoline sulfonic acid (ABTS), 5-bromo-4-chloro-3-indolyl phosphate (BCIP), nitro blue tetrazolium (NBT), 3,3′,5,5′-tetramethyl benzidine (TNB) and 4-chloro-1-naphthol (4-CN), although without limitation thereto.

In a further aspect, the invention provides a method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of a plurality of overexpressed genes associated with chromosomal instability and an expression level of a plurality of underexpressed genes associated with estrogen receptor signalling in one or more cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the overexpressed genes associated with chromosomal instability compared to the underexpressed genes associated with estrogen receptor signalling indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

In certain embodiments, the genes associated with chromosomal instability are of a CIN metagene. Non-limiting examples include genes selected from the group consisting of: *ATP6V1C1, RAP2A, CALM1, COG8, HELLS, KDM5A, PGK1, PLCH1, CEP55, RFC4, TAF2, SF3B3, GPI, PIR, MCM10, MELK, FOXM1, KIF2C, NUP155, TPX2, TTK, CENPA, CENPN, EXO1, MAPRE1, ACOT7, NAE1, SHMT2, TCP1, TXNRD1, ADM, CHAF1A and SYNCRIP*. In one preferred embodiment, the chromosomal instability genes are selected from the group consisting of *MELK*, *MCM10*, *CENPA*, *EXO1*, *TTK* and *KIF2C*.

In certain embodiments, the genes associated with estrogen receptor signalling are of an ER metagene. Non-limiting examples include genes selected from the group consisting of: *BTG2, PIK3IP1, SEC14L2, FLNB, ACSF2, APOM,*

BIN3, GLTSCR2, ZMYND10, ABAT, BCAT2, SCUBE2, RUNX1, LRRC48, MYBPC1, BCL2, CHPT1, ITM2A, LRIG1, MAPT, PRKCB, RERE, ABHD14A, FLT3, TNN, STC2, BATF, CD1E, CFB, EVL, FBXW4, ABCB1, ACAA1, CHAD, PDCD4, RPL10, RPS28, RPS4X, RPS6, SORBS1, RPL22 and RPS4XP3. In one preferred embodiment, the estrogen receptor signalling genes are selected from the group consisting of *MAPT* and *MYB*.

Suitably, the method of this aspect further includes the step of comparing an expression level of one or more other overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and an expression level of one or more other underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3* in one or more cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or more other overexpressed genes compared to the one or more other underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

In one embodiment, the one or more other overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*, *CD55*, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593*.

In one embodiment, the one or more other underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

In certain embodiments, the comparison of the expression level of the one or more other overexpressed genes and the expression level of the one or more other underexpressed genes is integrated with the comparison of the expression level of the plurality of overexpressed genes associated with chromosomal instability and the expression level of the plurality of underexpressed genes associated with estrogen receptor signalling to derive a first integrated score as described herein, which is

indicative of, or correlates with, responsiveness of the cancer to the anti-cancer treatment.

In another related aspect, the invention provides a method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of one or more overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and an expression level of one or more underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*, in one or more cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or more overexpressed genes compared to the one or more underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

In one embodiment, the one or more overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*, *CD55*, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593*.

In one embodiment, the one or more underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

In particular embodiments, the method of the five aforementioned aspects further includes the step of comparing an expression level of one or more overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and an expression level of one or more underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or more cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or more overexpressed proteins compared to the one or

more underexpressed proteins indicates or correlates with higher aggressiveness of the cancer and/or a less favourable cancer prognosis; and/or a lower relative expression level of the one or more overexpressed proteins compared to the one or more underexpressed proteins indicates or correlates with lower aggressiveness of the cancer and/or a more favourable cancer prognosis compared to a mammal having a higher expression level.

In particular embodiments, one or more of the overexpressed proteins and/or one or more of the underexpressed proteins are or comprise a phosphoprotein hereinbefore described.

An average or sum of the expression levels may be calculated for the overexpressed genes, the underexpressed genes, the overexpressed proteins and/or the underexpressed proteins, to thereby produce or calculate a ratio, as hereinbefore described.

Detection and/or measurement of expression of the overexpressed proteins and the underexpressed proteins may be performed by any of those methods or combinations thereof hereinbefore described, albeit without limitation thereto.

Suitably, the comparison of the expression level of the one or more overexpressed proteins and the expression level of the one or more underexpressed proteins is to thereby derive an integrated score. In one particular embodiment, the comparison of the expression level of the one or more overexpressed proteins and the expression level of the one or more underexpressed proteins is integrated with:

(i) the comparison of the expression level of the overexpressed genes associated with chromosomal instability and the expression level of the underexpressed genes associated with estrogen receptor signalling to derive a second integrated score; or

(ii) the first integrated score to derive a third integrated score; or

(iii) the comparison of the expression level of the overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1* and the expression level of the underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C,*

NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3 to derive a fourth integrated score; or

(iv) the comparison of the expression level of the overexpressed genes and an expression level of the underexpressed genes, wherein the genes are from one or more of the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and/or the Multiple Networks metagene, to derive a fifth integrated score; or

(v) the comparison of the expression level of the overexpressed genes and an expression level of the underexpressed genes, wherein the genes are from one or more of the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene, to derive a sixth integrated score,

wherein the second, third, fourth, fifth and/or sixth integrated score is indicative of, or correlates with, responsiveness of the cancer to the anti-cancer treatment.

In particular embodiments, the second, third, fourth, fifth and/or sixth integrated scores are derived, at least in part, by addition, subtraction, multiplication, division and/or exponentiation, as hereinbefore described.

In a further related aspect, the invention provides a method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of one or more overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and an expression level of one or more underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1,

YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or more cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or more overexpressed proteins compared to the one or more underexpressed proteins indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

In particular embodiments, one or more of the overexpressed proteins and/or one or more of the underexpressed proteins are or comprise a phosphoprotein hereinbefore described.

It will be appreciated from the foregoing that the invention provides methods that determine the aggressiveness of a cancer, facilitate providing a cancer prognosis for a patient and/or predict the responsiveness of a cancer to an anti-cancer treatment. Particular, broad embodiments of the invention include the step of treating the patient following determining the aggressiveness of the cancer, providing a cancer prognosis and/or predicting the responsiveness of the cancer to anti-cancer treatment. Accordingly, these embodiments relate to using information obtained about the aggressiveness of the cancer, the cancer prognosis and/or the predicted responsiveness of the cancer to anti-cancer treatment to thereby construct and implement an anti-cancer treatment regime for the patient. In a preferred embodiment, this is personalized to a particular patient so that the treatment regime is optimized for that particular patient.

Cancer treatments may include drug therapy, chemotherapy, antibody, nucleic acid and other biomolecular therapies, radiation therapy, surgery, nutritional therapy, relaxation or meditational therapy and other natural or holistic therapies, although without limitation thereto. In particular embodiments, the cancer therapy may target aneuploidy or aneuploid tumours and/or chromosomal instability.

Generally, drugs, biomolecules (*e.g* antibodies, inhibitory nucleic acids such as siRNA) or chemotherapeutic agents are referred to herein as “anti-cancer therapeutic agents”. In some embodiments relating to breast cancer, the anti-cancer treatment may include HER2-directed therapy such as trastuzumab and endocrine therapies such as tamoxifen and aromatase inhibitors. In other or alternative embodiments, the therapy may include administration of inhibitors of CIN genes or CIN gene products, such as one or more of those listed in Table 4. It will be appreciated that inhibition of the CIN gene product TTK using the specific inhibitor AZ3146 was effective against TNBC cell lines. Furthermore, siRNA-mediated

knockdown of the CIN genes *TTK, TPX2, NDC80* and *PBK* was effective against TNBC cell lines.

In certain embodiments, the cancer treatment may be directed at genes or gene products other than those listed in Tables 4, 10, 21 and/or 22. By way of example, the cancer treatment may target genes or gene products such as PLK1[71,72] or others[73-76] to thereby target aneuploid tumours or tumour cells.

Suitably, when considering (i) the relative expression of one or more of the overexpressed genes of the 29 gene signature (*i.e., CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*) when compared to one or more of the underexpressed genes of the 30 gene signature (i.e., *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*); (ii) the relative expression of one or more of the overexpressed proteins (i.e., DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1) when compared to one or more of the underexpressed proteins (i.e., VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6); and/or (iii) the first, second, third and/or fourth integrated score, the anticancer therapeutic agent is selected from the group consisting of a chemotherapy, an endocrine therapy, immunotherapy and a molecularly targeted therapy. In certain embodiments, the anticancer treatment comprises an ALK inhibitor (*e.g.*, TAE684), an Aurora kinase inhibitor (e.g., Alisertib, AMG-900, BI-847325, GSK-1070916A, ilorasertib, MK-8745, danusertib), a BCR-ABL inhibitor (*e.g.*, Nilotinib, Dasatinib, Ponatinib), a HSP90 inhibitor (*e.g.*, Tanespimycin (17-AAG), PF0429113, AUY922, Luminespib, ganetespib, Debio-0932), an EGFR inhibitor (*e.g.*, Afatinib, Erlotinib, Lapatinib, cetuximab), a PARP inhibitor (*e.g.*, ABT-888, AZD-2281), retinoic acid (*e.g.*, all-trans retinoic acid or ATRA), a Bcl2 inhibitor (*e.g.*, ABT-263), a gluconeogenesis inhibitor (*e.g.*, metformin), a p38 MAPK inhibitor (*e.g.*, BIRB0796, LY2228820), a MEK1/2 inhibitor (*e.g.*, trametinib, cobimetinib, binimetinib, selumetinib, pimasertib, refametinib, TAK-733), a mTOR inhibitor (*e.g.*, BEZ235, JW-7-25-1), a PI3K inhibitor (*e.g.*, Idelalisib, buparlisib/apelisib, copanlisib, GSK-2636771,

pictilisib, AMG-319, AZD-8186), an IGF1R inhibitor (*e.g.*, BMS-754807, dalotuzumab, ganitumab, linsitinib), a PLCγ inhibitor (*e.g.*, U73122), a JNK inhibitor (*e.g.*, SP600125), a PAK1 inhibitor (*e.g.*, IPA3), a SYK inhibitor (*e.g.*, BAY613606), a HDAC inhibitor (*e.g.*, Vorinostat), an FGFR inhibitor (*e.g.*, Dovitinib), a XIAP inhibitor (*e.g.*, Embelin), a PLK1 inhibitor (*e.g.*, Volasertib, P-937), an ERK5 inhibitor (*e.g.*, XMD8-92), a MPS1/TTK inhibitor (*e.g.*, BAY-1161909) and any combination thereof.

By way of example, patients with a high relative expression level of one or more overexpressed genes, such as those of the 21 gene signature, when compared to one or more underexpressed genes, such as those of the 7 gene signature, a high relative expression level of one or more overexpressed proteins when compared to one or more underexpressed proteins and/or a high integrated score described herein are more likely to respond favourably, such as a pathological complete response, when treated with chemotherapy. In this regard, non-limiting examples of chemotherapy include a pyrimidine analogue (*e.g.*, 5-fluorouracil, capecitabine), a taxane (*e.g.*, paclitaxel), an anthracycline (*e.g.*, doxorubicin, epirubicin), an anti-folate drug (*e.g.*, the dihydrofolate reductase inhibitor methotrexate), an alkylating agent (*e.g.*, cyclophosphamide) or any combination thereof. It would be appreciated that the chemotherapy may be administered as adjuvant, neoadjuvant and/or as standard therapy, alone or in combination with other anticancer therapeutics.

Additionally, in certain embodiments, patients with a high relative expression level of one or more overexpressed genes, such as those of the 29 gene signature, when compared to one or more underexpressed genes, such as those of the 30 gene signature, a high relative expression level of one or more overexpressed proteins when compared to one or more underexpressed proteins and/or a high integrated score described herein may be more likely to respond favourably to (*i.e.*, be more sensitive to) inhibition of HSP90, EGFR, IGF1R, mTOR, PI3K, p38 MAPK, PLCγ, JNK, PAK1, ERK5, XIAP, PLK1 and/or MEK1/2 and may be less likely to respond favourably to (*i.e.*, be less sensitive to) anticancer treatment with an ALK inhibitor, a BCR-ABL inhibitor, a PARP inhibitor, retinoic acid, a Bcl2 inhibitor, a gluconeogenesis inhibitor, a p38 MAPK inhibitor, an FGFR inhibitor, a SYK inhibitor, a HDAC inhibitor and/or an IGF1R inhibitor.

It will also be understood that the gene and protein signatures described herein may be used to identify those poorer prognosis patients, such as those with

larger and/or higher grade tumours, who may benefit from one or more additional anticancer therapeutic agents to the typical or standard anti-cancer treatment regime for that particular patient group. By way of example, ER^+ breast cancer patients with or without lymph node involvement with a high integrated score, and hence a relatively poor prognosis, are more likely to respond favourably to or benefit from chemotherapy and/or endocrine therapy. This may include an improved survival and/or reduced likelihood of tumour recurrence and/or metastasis for these patients.

In certain embodiments, for patients with a high relative expression level of the overexpressed genes of the 21 gene signature when compared to the underexpressed genes of the 7 gene signature and/or a high integrated score, the cancer treatment may be directed at those genes or gene products listed in Tables 13, 15, 16 and 17.

Additionally, for patients with a high relative expression level of the overexpressed proteins when compared to the underexpressed proteins and/or a high integrated score the cancer treatment may be directed at one or more of those proteins listed in Table 19.

It would be appreciated that those methods described herein for predicting the responsiveness of a cancer to an anti-cancer treatment, such as an immunotherapeutic agent, may further include the step of administering to the mammal a therapeutically effective amount of the anticancer treatment. In a preferred embodiment, the anticancer treatment is administered when the altered or modulated relative expression level indicates or correlates with relatively increased responsiveness of the cancer to the anti-cancer treatment.

Methods of treating cancer may be prophylactic, preventative or therapeutic and suitable for treatment of cancer in mammals, particularly humans. As used herein, "*treating*", "*treat*" or "*treatment*" refers to a therapeutic intervention, course of action or protocol that at least ameliorates a symptom of cancer after the cancer and/or its symptoms have at least started to develop. As used herein, "*preventing*", "*prevent*" or "*prevention*" refers to therapeutic intervention, course of action or protocol initiated prior to the onset of cancer and/or a symptom of cancer so as to prevent, inhibit or delay or development or progression of the cancer or the symptom.

The term *"therapeutically effective amount"* describes a quantity of a specified agent sufficient to achieve a desired effect in a subject being treated with that agent. For example, this can be the amount of a composition comprising one or

more agents that binds one or more of the overexpressed and/or underexpressed genes or gene products thereof described herein, necessary to reduce, alleviate and/or prevent a cancer or cancer associated disease, disorder or condition. In some embodiments, a *"therapeutically effective amount"* is sufficient to reduce or eliminate a symptom of a cancer. In other embodiments, a *"therapeutically effective amount"* is an amount sufficient to achieve a desired biological effect, for example an amount that is effective to decrease or prevent cancer growth and/or metastasis.

Ideally, a therapeutically effective amount of an agent is an amount sufficient to induce the desired result without causing a substantial cytotoxic effect in the subject. The effective amount of an agent useful for reducing, alleviating and/or preventing a cancer will be dependent on the subject being treated, the type and severity of any associated disease, disorder and/or condition (*e.g.*, the number and location of any associated metastases), and the manner of administration of the therapeutic composition.

Suitably, the anti-cancer therapeutic agent is administered to a mammal as a pharmaceutical composition comprising a pharmaceutically-acceptable carrier, diluent or excipient.

By *"pharmaceutically-acceptable carrier, diluent or excipient"* is meant a solid or liquid filler, diluent or encapsulating substance that may be safely used in systemic administration. Depending upon the particular route of administration, a variety of carriers, well known in the art may be used. These carriers may be selected from a group including sugars, starches, cellulose and its derivatives, malt, gelatine, talc, calcium sulfate, liposomes and other lipid-based carriers, vegetable oils, synthetic oils, polyols, alginic acid, phosphate buffered solutions, emulsifiers, isotonic saline and salts such as mineral acid salts including hydrochlorides, bromides and sulfates, organic acids such as acetates, propionates and malonates and pyrogen-free water.

A useful reference describing pharmaceutically acceptable carriers, diluents and excipients is Remington's Pharmaceutical Sciences (Mack Publishing Co. N.J. USA, 1991), which is incorporated herein by reference.

Any safe route of administration may be employed for providing a patient with the composition of the invention. For example, oral, rectal, parenteral, sublingual, buccal, intravenous, intra-articular, intra-muscular, intra-dermal, subcutaneous, inhalational, intraocular, intraperitoneal, intracerebroventricular,

transdermal and the like may be employed. Intra-muscular and subcutaneous injection is appropriate, for example, for administration of immunotherapeutic compositions, proteinaceous vaccines and nucleic acid vaccines.

Dosage forms include tablets, dispersions, suspensions, injections, solutions, syrups, troches, capsules, suppositories, aerosols, transdermal patches and the like. These dosage forms may also include injecting or implanting controlled releasing devices designed specifically for this purpose or other forms of implants modified to act additionally in this fashion. Controlled release of the therapeutic agent may be effected by coating the same, for example, with hydrophobic polymers including acrylic resins, waxes, higher aliphatic alcohols, polylactic and polyglycolic acids and certain cellulose derivatives such as hydroxypropylmethyl cellulose. In addition, the controlled release may be effected by using other polymer matrices, liposomes and/or microspheres.

Compositions of the present invention suitable for oral or parenteral administration may be presented as discrete units such as capsules, sachets or tablets each containing a pre-determined amount of one or more therapeutic agents of the invention, as a powder or granules or as a solution or a suspension in an aqueous liquid, a non-aqueous liquid, an oil-in-water emulsion or a water-in-oil liquid emulsion. Such compositions may be prepared by any of the methods of pharmacy but all methods include the step of bringing into association one or more agents as described above with the carrier which constitutes one or more necessary ingredients. In general, the compositions are prepared by uniformly and intimately admixing the agents of the invention with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product into the desired presentation.

The above compositions may be administered in a manner compatible with the dosage formulation, and in such amount as is pharmaceutically-effective. The dose administered to a patient, in the context of the present invention, should be sufficient to effect a beneficial response in a patient over an appropriate period of time. The quantity of agent(s) to be administered may depend on the subject to be treated inclusive of the age, sex, weight and general health condition thereof, factors that will depend on the judgement of the practitioner.

In particular embodiments of the hereinbefore described methods, the cancer is breast cancer and the one or more overexpressed proteins are selected from the group consisting of DVL3, VEGFR2, INPP4B, EIF4EBP1, EGFR, HER3, SMAD1,

NFKB1 and HER2 and the one or more underexpressed proteins are selected from the group consisting of ASNS, MAPK9, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6.

In particular embodiments of the hereinbefore described methods, the cancer is lung cancer, such as lung adenocarcinoma, wherein:

(i) the one or more overexpressed genes are selected from the group consisting of *GNB2L1*, *TXN*, *KCNG1*, *BCAP31*, *GSK3B*, *FOXM1*, *ZNF593*, *EXO1*, *KIF2C*, *TTK*, *MELK*, *CENPA*, *TPX2*, *CA9*, *GRHPR*, *HCFC1R1*,*CEP55*, *MCM10*, *CENPN* and *CARHSP1*, and the one or more underexpressed genes are selected from the group consisting of *BTN2A2*, *MTMR7*, *ZNRD1-AS1*, *MAPT* and *BTG2*; and/or

(ii) the one or more overexpressed proteins are selected from the group consisting of DVL3, PAI-1, Ku80, GATA3, ITGA2 and AKT1, and the one or more underexpressed proteins are selected from the group consisting of ESR1.

In particular embodiments of the hereinbefore described methods, the cancer is kidney cancer, such as renal clear cell carcinoma, wherein:

(i) the one or more overexpressed genes are selected from the group consisting of *EIF3K*, *ADORA2B*, *KCNG1*, *BCAP31*, *EXOSC7*, *FOXM1*, *CD55*, *ZNF593*, *KIF2C*, *TTK*, *MELK*, *CENPA*, *TPX2*, *CEP55*, *PML*, *CENPN* and *CARHSP1*, and the one or more underexpressed genes are selected from the group consisting of *BCL2* and *MAPT*; and/or

(ii) the one or more overexpressed proteins are selected from the group consisting of DVL3, PAI-1 and EIF4EBP1, and the one or more underexpressed proteins are selected from the group consisting of HER3, MAPK9, ESR1 and RAD50.

In particular embodiments of the hereinbefore described methods, the cancer is melanoma, such as skin cutaneous melanoma, and wherein:

(i) the one or more overexpressed genes are selected from the group consisting of *EIF3K*, *ADORA2B*, *GSK3B*, *EXOSC7*, *FOXM1*, *EXO1*, *KIF2C*, *CENPA*, *TPX2*, *CAMSAP1*, *MCM10* and *ABHD5* and the one or more underexpressed genes are selected from the group consisting of *BCAP31*, *BTN2A2*, *SMPDL3B*, *MTMR7*, *ME1* and *BTG2*; and/or

(ii) the one or more overexpressed proteins are selected from the group consisting of PAI-1, EIF4EBP1, EGFR, HER3 and Ku80 and the one or more

underexpressed proteins are selected from the group consisting of ASNS, MAPK9 and ESR1.

In particular embodiments of the hereinbefore described methods, the cancer is endometrial cancer, such as uterine corpus endometrioid carcinoma, and wherein:

(i) the one or more overexpressed genes are selected from the group consisting of *GNB2L1*, *EIF3K*, *KCNG1*, *BCAP31*, *GSK3B*, *EXOSC7*, *FOXM1*, *ZNF593*, *EXO1*, *KIF2C*, *MAP2K5*, *TTK*, *MELK*, *GRHPR*, and *PML*, and the one or more underexpressed genes is *MYB*; and/or

(ii) the one or more overexpressed proteins are selected from the group consisting of DVL3, INPP4B, EIF4EBP1 and ASNS and the one or more underexpressed proteins are selected from the group consisting of MAPK9, ESR1 and YWHAE.

In particular embodiments of the hereinbefore described methods, the cancer is ovarian adenocarcinoma and wherein:

(i) the one or more overexpressed genes are selected from the group consisting of *GNB2L1*, *EIF3K*, *TXN*, *ADORA2B*, *KCNG1*, *GSK3B*, *STAU1*, *MAP2K5*, and *HCFC1R1*, and the one or more underexpressed genes are selected from the group consisting of *BTN2A2*, and *ZNRD1-AS1*; and/or

(ii) the one or more overexpressed proteins are selected from the group consisting of PAI-1 and VEGFR2 and the one or more underexpressed proteins are selected from the group consisting of ASNS, MAPK9, ESR1, YWHAE and PGR.

In particular embodiments of the hereinbefore described methods, the cancer is head and neck cancer, such as head and neck squamous cell carcinoma, and wherein:

(i) the one or more overexpressed genes are selected from the group consisting of *GNB2L1*, *TXN*, *ADORA2B*, *KCNG1*, *CD55*, *ZNF593*, *NDUFC1*, and *HCFC1R1*, and the one or more underexpressed genes are selected from the group consisting of *BTN2A2*, and *MTMR7*; and/or

(ii) the one or more overexpressed proteins are selected from the group consisting of PAI-1, INPP4B, EGFR, HER3, SMAD1, GATA3, ITGA2 and COL6A1 and the one or more underexpressed proteins are selected from the group consisting of VEGFR2 and ASNS.

In particular embodiments of the hereinbefore described methods, the cancer is colorectal cancer, such as colorectal adenocarcinoma, and wherein:

(i) the one or more overexpressed genes are selected from the group consisting of *EIF3K*, *TXN*, *CD55*, *NDUFC1*, *HCFC1R1*, and *PML*, and the one or more underexpressed genes are selected from the group consisting of *BTN2A2*, *SMPDL3B*, and *ME1*; and/or

(ii) the one or more overexpressed proteins are selected from the group consisting of DVL3, PAI-1, INPP4B, EIF4EBP1, EGFR and HER3 and the one or more underexpressed proteins are selected from the group consisting of ASNS, MAPK9, YWHAE, RAD50 and PEA15.

In particular embodiments of the hereinbefore described methods, the cancer is glioma, such as lower grade glioma, and wherein:

(i) the one or more overexpressed genes are selected from the group consisting of *TXN*, *BCAP31*, *STAU1*, *PML*, *CARHSP1*, and *BTN2A2*; and/or

(ii) the one or more overexpressed proteins are selected from the group consisting of DVL3, PAI-1, VEGFR2, Ku80, SMAD1 and NFKB1 and the one or more underexpressed proteins are selected from the group consisting of ESR1, YWHAE and PGR.

In particular embodiments of the hereinbefore described methods, the cancer is bladder cancer, such as urothelial carcinoma, and wherein:

(i) the one or more overexpressed genes are selected from the group consisting of *ADORA2B*, *KCNG1*, *STAU1*, *MAP2K5*, and *CAMSAP1*, and the one or more underexpressed genes are selected from the group consisting of *GNB2L1*, *EIF3K*, *TXN*, *BCAP31*, *EXOSC7*, *CD55*, *NDUFC1*, *GRHPR*, *CETN3*, *BTN2A2*, *SMPDL3B*, and *ERC2*,; and/or

(ii) the one or more overexpressed proteins are selected from the group consisting of DVL3, VEGFR2, Ku80, SMAD1 and AKT1 and the one or more underexpressed proteins is ASNS.

In particular embodiments of the hereinbefore described methods, the cancer is lung cancer, such as lung squamous cell carcinoma, and wherein:

(i) the one or more overexpressed genes are selected from the group consisting of *GNB2L1*, *ZNF593*, and *SMPDL3B*, and the one or more underexpressed genes are selected from the group consisting of *GSK3B*, *MAP2K5*, *NDUFC1*, *CAMSAP1*, *ABHD5*, and *ME1*; and/or

(ii) the one or more overexpressed proteins are selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EGFR and GATA3 and the one or more underexpressed proteins is ASNS.

In particular embodiments of the hereinbefore described methods, the cancer is adrenocortical carcinoma, and wherein:

the one or more overexpressed genes are selected from the group consisting of *GNB2L1*, *EIF3K*, *TXN*, *ADORA2B*, *KCNG1*, *BCAP31*, *FOXM1*, *ZNF593*, *EXO1*, *KIF2C*, *MAP2K5*, *TTK*, *MELK*, *CENPA*, *TPX2*, *GRHPR*, *CEP55*, *MCM10*, and *CENPN*, and the one or more underexpressed genes are selected from the group consisting of *MTMR7*, *BCL2*, *MAPT*, *MYB*, and *STC2*.

In particular embodiments of the hereinbefore described methods, the cancer is kidney renal papillary cell carcinoma and wherein:

the one or more overexpressed genes are selected from the group consisting of *GNB2L1*, *ADORA2B*, *KCNG1*, *GSK3B*, *FOXM1*, *CD55*, *EXO1*, *KIF2C*, *STAU1*, *TTK*, *MELK*, *CENPA*, *TPX2*, *CA9*, *CEP55*, and *MCM10*, and the one or more underexpressed genes are selected from the group consisting of *SMPDL3B*, and *BCL2*.

In particular embodiments of the hereinbefore described methods, the cancer is pancreatic ductal adenocarcinoma and wherein:

the one or more overexpressed genes are selected from the group consisting of *EIF3K*, *ADORA2B*, *GSK3B*, *EXOSC7*, *FOXM1*, *CD55*, *EXO1*, *STAU1*, *CAMSAP1*, and *CETN3* and the one or more underexpressed genes are selected from the group consisting of *BTN2A2*, *SMPDL3B*, *MTMR7*, *ME1*, *BCL2*, and *ERC2*.

In particular embodiments of the hereinbefore described methods, the cancer is liver hepatocellular carcinoma and wherein:

the one or more overexpressed genes are selected from the group consisting of *GNB2L1*, *TXN*, *EXOSC7*, and *CA9*, and the one or more underexpressed genes is *MTMR7*.

In particular embodiments of the hereinbefore described methods, the cancer is cervical squamous cell carcinoma and/or endocervical adenocarcinoma and wherein:

the one or more overexpressed genes are selected from the group consisting of *STAU1*, *CA9*, and *ME1* and the one or more underexpressed genes are selected from the group consisting of *EIF3K*, *TXN*, *BCAP31*, *EXOSC7*, and *ZNRD1-AS1*.

Furthermore, in certain embodiments, patients with a high relative expression level of one or more overexpressed genes, such as those of the 29 gene signature, when compared to one or more underexpressed genes, such as those of the 30 gene signature, a high relative expression level of one or more overexpressed proteins when compared to one or more underexpressed proteins and/or a high integrated score as described herein may be more likely to respond favourably to immunotherapy.

Accordingly, one aspect provides a method of predicting the responsiveness of a cancer to an immunotherapeutic agent in a mammal, said method including the step of comparing an expression level of one or more overexpressed genes selected from the group consisting of *ADORA2B*, *CD36*, *CETN3*, *KCNG1*, *LAMA3*, *MAP2K5*, *NAE1*, *PGK1*, *STAU1*, *CFDP1*, *SF3B3* and *TXN*, and an expression level of one or more underexpressed genes selected from the group consisting of *APOBEC3A*, *BCL2*, *BTN2A2*, *CAMSAP1*, *CAMK4*, *CARHSP1*, *FBXW4*, *GSK3B*, *HCFC1R1*, *MYB*, *PSEN2* and *ZNF593*, in one or more cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or more overexpressed genes compared to the one or more underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the immunotherapeutic agent.

In one embodiment the one or more overexpressed genes are selected from the group consisting of *ADORA2B*, *CETN3*, *KCNG1*, *MAP2K5*, *STAU1* and *TXN*, and/or an expression level of one or more underexpressed genes are selected from the group consisting of *BTN2A2*, *CAMSAP1*, *CARHSP1*, *GSK3B*, *HCFC1R1*, and *ZNF593*.

In one embodiment, the one or more overexpressed genes are selected from the group consisting of *ADORA2B*, *CD36*, *KCNG1*, *LAMA3*, *MAP2K5*, *NAE1*, *PGK1*, *STAU1*, *CFDP1*, and *SF3B3* and/or an expression level of one or more underexpressed genes are selected from the group consisting of *APOBEC3A*, *BCL2*, *BTN2A2*, *CAMK4*, *FBXW4*, *PSEN2* and, *MYB*.

It would be understood for particular embodiments of the present aspect that one or more other overexpressed genes and/or one or more other underexpressed genes from one or more of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene,

an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene. such as those listed in Table 21, may be included in the step of comparing an expression level of one or more overexpressed genes and an expression level of one or more underexpressed genes.

Insofar as they relate to cancer, immunotherapy or immunotherapeutic agents use or modify the immune mechanisms of a subject so as to promote or facilitate treatment of a cancer. In this regard, immunotherapy or immunotherapeutic agents used to treat cancer include cell-based therapies, antibody therapies (*e.g.*, anti-PD1 or anti-PDL1 antibodies) and cytokine therapies. These therapies all exploit the phenomenon that cancer cells often have subtly different molecules termed cancer antigens on their surface that can be detected by the immune system of the cancer subject. Accordingly, immunotherapy is used to provoke the immune system of a cancer patient into attacking the cancer's cells by using these cancer antigens as targets.

Non-limiting examples of immunotherapy or immunotherapeutic agents include adalimumab, alemtuzumab, basiliximab, belimumab, bevacizumab, BMS-936559, brentuximab, certolizumab, cituximab, daclizumab, eculizumab, ibritumomab, infliximab, ipilimumab, lambrolkizumab, mepolizumab, MPDL3280A muromonab, natalizumab, nivolumab, ofatumumab, omalizumab, pembrolizumab, pexelizumab, pidilizumab, rituximab, tocilizumab, tositumomab, trastuzumab, ustekinumab, abatacept, alefacept and denileukin diftitox. In particular preferred embodiments, the immunotherapeutic agent is an immune checkpoint inhibitor, such as an anti-PD1 antibody (*e.g.*, pidilizumab, nivolumab, lambrolkizumab, pembrolizumab), an anti-PDL1 antibody (*e.g.*, BMS-936559, MPDL3280A) and/or an anti-CTLA4 antibody (*e.g.*, ipilimumab).

As would be appreciated by the skilled artisan, immune checkpoints refer to a variety of inhibitory pathways of the immune system that are crucial for maintaining self-tolerance and for modulating the duration and/or amplitude of an immune response in a subject. Cancers can use particular immune checkpoint pathways as a major mechanism of immune resistance, particularly against T cells that are specific for tumour antigens. Accordingly, immune checkpoint inhibitors include any agent that blocks or inhibits the inhibitory pathways of the immune system. Such inhibitors

may include small molecule inhibitors or may include antibodies, or antigen binding fragments thereof, that bind to and block or inhibit immune checkpoint receptors or antibodies that bind to and block or inhibit immune checkpoint receptor ligands. By way of example, immune checkpoint receptors or receptor ligands that may be targeted for blocking or inhibition include, but are not limited to, CTLA-4, 4-1BB (CD137), 4-1BBL (CD137L), PDL1, PDL2, PD1, B7-H3, B7-H4, BTLA, HVEM, TIM3, GAL9, LAG3, TIM3, B7H3, B7H4, VISTA, KIR, 2B4, CD160 and CGEN-15049. Illustrative immune checkpoint inhibitors include tremelimumab (CTLA-4 blocking antibody), anti-OX40, PD-L1 monoclonal Antibody (Anti-B7-H1; MEDI4736), MK-3475 (PD-1 blocker), nivolumab (anti-PD1 antibody), pidilizamab (CT-011; anti-PD1 antibody), BY55 monoclonal antibody, AMP224 (anti-PDL1 antibody), BMS-936559 (anti-PDL1 antibody), MPLDL3280A (anti-PDL1 antibody), MSB0010718C (anti-PDL1 antibody) and yervoy/ipilimumab (anti-CTLA-4 checkpoint inhibitor), albeit without limitation thereto.

In one embodiment, the method of predicting the responsiveness of a cancer to an immunotherapeutic agent, may further include the step of administering to the mammal a therapeutically effective amount of the immunotherapeutic agent.

In a related aspect is provided a method of predicting the responsiveness of a cancer to an EGFR inhibitor in a mammal, said method including the step of comparing an expression level of one or more overexpressed genes selected from the group consisting of *NAE1*, *GSK3B*, *TAF2*, *MAPRE1*, *BRD4*, *STAU1*, *TAF2*, *PDCD4*, *KCNG1*, *ZNRD1-AS1*, *EIF4B*, *HELLS*, *RPL22*, *ABAT*, *BTN2A2*, *CD1B*, *ITM2A*, *BCL2*, *CXCR4*, and *ARNT2*and an expression level of one or more underexpressed genes selected from the group consisting of *CD1C*, *CD1E*, *CD1B*, *KDM5A*, *BATF*, *EVL*, *PRKCB*, *HCFC1R1*, *CARHSP1*, *CHAD*, *KIR2DL4*, *ABHD5*, *ABHD14A*, *ACAA1*, *SRPK3*, *CFB*, *ARNT2*, *NDUFC1*, *BCL2*, *EVL*, *ULBP2*, *BIN3*, *SF3B3*, *CETN3*, *SYNCRIP*, *TAF2*, *CENPN*, *ATP6V1C1*, *CD55* and *ADORA2B* in one or more cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or more overexpressed genes compared to the one or more underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the EGFR inhibitor.

It would be appreciated that the EGFR inhibitor may be any known in the art, including monoclonal antibody and small molecule inhibitors thereof, such as those

hereinbefore described. In particular embodiments, the EGFR inhibitor is or comprises erlotinib and/or cetuximab.

In certain embodiments, the cancer is or comprises lung cancer, colorectal cancer or breast cancer.

In one embodiment, the one or more overexpressed genes are selected from the group consisting of *NAE1*, *GSK3B*, and *TAF2* and/or the one or more underexpressed genes are selected from the group consisting of *CD1C*, *CD1E*, *CD1B*, *KDM5A*, *BATF*, *EVL*, *PRKCB*, *HCFC1R1*, *CARHSP1*, *CHAD*, *KIR2DL4*, *ABHD5*, *ABHD14A*, *ACAA1*, *SRPK3*, and *CFB*.

In one embodiment, the one or more overexpressed genes are selected from the group consisting of *MAPRE1*, *BRD4*, *STAU1*, *TAF2*, *GSK3B*, *PDCD4*, *KCNG1*, *ZNRD1-AS1*, *EIF4B* and *HELLS* and/or the one or more underexpressed genes are selected from the group consisting of *ARNT2*, *NDUFC1*, *BCL2*, *ABHD14A*, *EVL*, *ULBP2*, and *BIN3*.

In one embodiment, the one or more overexpressed genes are selected from the group consisting of *RPL22*, *ABAT*, *BTN2A2*, *CD1B*, *ITM2A*, *BCL2*, *CXCR4*, and *ARNT2* and/or the one or more underexpressed genes are selected from the group consisting of *SF3B3*, *CETN3*, *SYNCRIP*, *TAF2*, *CENPN*, *ATP6V1C1*, *CD55* and *ADORA2B*.

In a related aspect is provided a method of predicting the responsiveness of a cancer to a multikinase inhibitor in a mammal, said method including the step of comparing an expression level of one or more overexpressed genes selected from the group consisting of *SCUBE*, *CHPT1*, *CDC1*, *BTG2*, *ADORA2B* and *BCL2*, and an expression level of one or more underexpressed genes selected from the group consisting of *NOP2*, *CALR*, *MAPRE1*, *KCNG1*, *PGK1*, *SRPK3*, *RERE*, *ADM*, *LAMA3*, *KIR2DL4*, *ULBP2*, *LAMA4*, *CA9*, and *BCAP31*, in one or more cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or more overexpressed genes compared to the one or more underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the EGFR inhibitor.

Multikinase inhibitors typically work by inhibiting multiple intracellular and/or cell surface kinases, some of which may be implicated in tumor growth and metastatic progression of a cancer, thus decreasing tumor growth and replication. It would be appreciated that the multikinase inhibitor may be any known in the art,

including small molecule inhibitors, such as those hereinbefore described. Non-limiting examples of multikinase inhibitors include sorafenib, trametinib, dabrafenib, vemurafenib, crizotinib, sunitinib, axitinib, ponatinib, ruxolitinib, vandetanib, cabozantinib, afatinib, ibrutinib and regorafenib. In a particular embodiment, the multikinase inhibitor is or comprises sorafenib.

In one embodiment, the cancer is or comprises lung cancer.

Suitably, with regard to predicting the responsiveness of a cancer to an immunotherapeutic agent, an EGFR inhibitor or a multikinase inhibitor, a higher relative expression level of the one or more overexpressed genes compared to the one or more underexpressed genes indicates or correlates with a relatively increased responsiveness of the cancer to the agent or inhibitor; and/or a lower relative expression level of the one or more overexpressed genes compared to the one or more underexpressed genes indicates or correlates with a relatively decreased responsiveness of the cancer to the agent or inhibitor.

In a further aspect, the invention provides a method for identifying an agent for use in the treatment of cancer including the steps of:

(i) contacting a protein product of *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7*, *COG8*, *CFDP1* and/or *KCNG1* with a test agent; and

(ii) determining whether the test agent, at least partly, reduces, eliminates, suppresses or inhibits the expression and/or an activity of the protein product.

Suitably, the cancer is of a type hereinbefore described, albeit without limitation thereto. Preferably, the cancer has an overexpressed gene selected from the group consisting of *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7*, *COG8*, *CFDP1* and *KCNG1* and any combination thereof,

Suitably, the agent possesses or displays little or no significant off-target and/or nonspecific effects.

Preferably, the agent is an antibody or a small organic molecule.

In embodiments relating to antibody inhibitors, the antibody may be polyclonal or monoclonal, native or recombinant. Well-known protocols applicable to antibody production, purification and use may be found, for example, in Chapter 2 of Coligan *et al.*, CURRENT PROTOCOLS IN IMMUNOLOGY (John Wiley & Sons NY, 1991-1994) and Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor, Cold Spring Harbor Laboratory, 1988, which are both herein incorporated by reference.

Generally, antibodies of the invention bind to or conjugate with an isolated protein, fragment, variant, or derivative of the protein product of one or more of *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7*, *COG8*, *CFDP1* and *KCNG1*. For example, the antibodies may be polyclonal antibodies. Such antibodies may be prepared for example by injecting an isolated protein, fragment, variant or derivative of the protein product into a production species, which may include mice or rabbits, to obtain polyclonal antisera. Methods of producing polyclonal antibodies are well known to those skilled in the art. Exemplary protocols which may be used are described for example in Coligan *et al.*, CURRENT PROTOCOLS IN IMMUNOLOGY, *supra*, and in Harlow & Lane, 1988, *supra*.

Monoclonal antibodies may be produced using the standard method as for example, described in an article by Köhler & Milstein, 1975, Nature **256**, 495, which is herein incorporated by reference, or by more recent modifications thereof as for example, described in Coligan *et al.*, CURRENT PROTOCOLS IN IMMUNOLOGY, *supra* by immortalizing spleen or other antibody producing cells derived from a production species which has been inoculated with one or more of the isolated protein products and/or fragments, variants and/or derivatives thereof.

Typically, the inhibitory activity of candidate inhibitor antibodies may be assessed by *in vitro* and/or *in vivo* assays that detect or measure the expression levels and/or activity of the protein products of one or more of *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7*, *COG8*, *CFDP1* and *KCNG1* in the presence of the antibody.

In embodiments relating to small organic molecule inhibitors, this may involve screening of large compound libraries, numbering hundreds of thousands to millions of candidate inhibitors (chemical compounds including synthetic, small organic molecules or natural products, for example) which may be screened or tested for biological activity at any one of hundreds of molecular targets in order to find potential new drugs, or lead compounds. Screening methods may include, but are not limited to, computer-based ("in silico") screening and high throughput screening based on in vitro assays.

Typically, the active compounds, or “hits”, from this initial screening process are then tested sequentially through a series of other in vitro and/or in vivo tests to further characterize the active compounds. A progressively smaller number of the “successful” compounds at each stage are selected for subsequent testing, eventually

leading to one or more drug candidates being selected to proceed to being tested in human clinical trials.

At the clinical level, screening a test agent may include obtaining samples from test subjects before and after the subjects have been exposed to a test compound. The levels in the samples of the protein product of the overexpressed genes may then be measured and analysed to determine whether the levels and/or activity of the protein products change after exposure to a test agent. By way of example, protein product levels in the samples may be determined by mass spectrometry, western blot, ELISA and/or by any other appropriate means known to one of skill in the art. Additionally, the activity of the protein products, such as their enzymatic activity, may be determined by any method known in the art. This may include, for example, enzymatic assays, such as spectrophotometric, fluorometric, calorimetric, chemiluminescent, light scattering, microscale thermophoresis, radiometric and chromatographic assays.

It would be appreciated that subjects who have been treated with test agents may be routinely examined for any physiological effects which may result from the treatment. In particular, the test agents will be evaluated for their ability to decrease cancer likelihood or occurrence in a subject. Alternatively, if the test agents are administered to subjects who have previously been diagnosed with cancer, they will be screened for their ability to slow or stop the progression of the cancer as well as induce disease remission.

In a particular embodiment, the invention may provide a "*companion diagnostic*" whereby the one or more genes that are detected as having elevated expression are the same genes that are targeted by the anti-cancer treatment.

In a related aspect, the invention provides an agent for use in the treatment of cancer identified by the method hereinbefore described.

Suitably, the cancer is of a type hereinbefore described, albeit without limitation thereto. Preferably, the cancer has an overexpressed gene selected from the group consisting of *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7*, *COG8*, *CFDP1*, *KCNG1* and any combination thereof.

In another related aspect, the invention provides a method of treating a cancer in a mammal, including the step of administering to the mammal a therapeutically effective amount of an agent hereinbefore described.

In this regard, test agents that are identified of being capable of reducing, eliminating, suppressing or inhibiting the expression level and/or activity of a protein product of *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7*, *COG8*, *CFDP1* and/or *KCNG1* may then be administered to patients who are suffering from or are at risk of developing cancer,. For example, the administration of a test agent which inhibits or decreases the activity and/or expression of the protein product of one or more of the aforementioned genes may treat the cancer and/or decrease the risk cancer, if the increased activity of the biomarker is responsible, at least in part, for the progression and/or onset of the cancer.

Suitably, the cancer is of a type hereinbefore described, albeit without limitation thereto. Preferably, the cancer has an overexpressed gene selected from the group consisting of *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7*, *COG8*, *CFDP1*, *KCNG1* and any combination thereof.

All computer programs, algorithms, patent and scientific literature referred to herein is incorporated herein by reference.

For the present invention, the database accession number or unique identifier provided herein for a gene or a protein, such as those presented in Tables 4, 5, 10, 15, 16, 17 and 18, as well as the gene and/or protein sequence or sequences associated therewith, are incorporated by reference herein.

So that preferred embodiments of the invention may be fully understood and put into practical effect, reference is made to the following non-limiting examples.

EXAMPLE 1

Materials and Methods

Meta-analysis of global gene expression in TNBC

We performed a meta-analysis of global gene expression data in the Oncomine™ database[19] (Compendia Bioscience, MI) using a primary filter for breast cancer (130 datasets), sample filter to use clinical specimens and dataset filters to use mRNA datasets with more than 151 patients (22 datasets). Patients of all ages, gender, disease stages or treatments were included. Three additional filters were applied to perform three independent differential analyses: (1) triple negative (TNBC cases vs. non-TNBC cases, 8 datasets[49-56]; (2) metastatic event analysis at 5 years

(metastatic events vs. no metastatic events, 7 datasets[53,54,57-61]) and (3) survival at 5 years (patients who died vs. patients who survived, 7 datasets[49,54,56,58,61-63]). Deregulated genes were selected based on the median p-value of the median gene rank in overexpression or underexpression patterns across the datasets (Figure 8). The union of these three deregulated gene lists resulted in a gene list of deregulated genes in aggressive breast cancers (Figure 9). The METBRIC dataset[21] was used as the validation set for further analysis. The normalized z-score expression data of the METABRIC dataset was extracted from Oncomine™ and imported into BRB-ArrayTools[64] (V4.2, Biometric Research Branch, NCI, Maryland, USA) with built in R Bioconductor packages. Survival curves for the METABRIC dataset were constructed using GraphPad® Prism v6.0 (GraphPad Software, CA, USA) and the Log-rank (Mantel-Cox) Test was used for statistical comparisons of survival curves.

Ingenuity Pathway Analysis and derivation of the eight gene list

Pathway analysis was performed using the Ingenuity Pathway Analysis® (Ingenuity Systems®, CA). For pathway analysis in IPA®, we used only direct relationships. After pathway analysis, we set to identify the minimum gene list that recapitulates the aggressiveness 206 gene list. We used the METABRIC dataset to perform statistical filtering in the BRB-ArrayTools software to derive the minimum gene list as follows: (1) the correlation of each gene in the CIN metagene and the ER metagene to the metagene itself was determined by quantitative trait analysis using the Pearson's correlation coefficient (univariate p-value threshold of 0.001); (2) the association of each gene with overall survival using univariate Cox proportional hazards model (univariate test p-value < 0.001); and (3) the fold-change of gene expression between high aggressiveness score tumors and low aggressiveness score tumors was calculated for each gene. We selected genes with Pearson's correlation coefficient > 0.7 to the metagenes, strongest survival association and more than 2-fold deregulation between high and low agressiveness score tumors. The METABRIC dataset and four publically available datasets were used to validate the 8-genes score. The four datasets (GSE25066[53], GSE3494[65], GSE2990[15], GSE2034[66]) were analyzed as described previously[67].

Cell culture and drug treatments

Breast cancer cell lines were obtained from ATCC™ (VA, USA) and cultured as per ATCC™ instructions. All cell lines were regularly tested for mycoplasma and authenticated using STR profiling. For the siRNA screen, siRNA

solutions (Shanghai Gene Pharma, China) were used to transfect cells (MDA-MB-231, SUM159PT and Hs578T) with 10 nM of respective siRNA using Lipofectamine® RNAiMAX (Life Technologies, CA, USA). For drug treatments, docetaxel and the TTK inhibitor AZ3146 were purchased from Selleck Chemicals LLC (TX, USA) and diluted in DMSO. Six days after siRNA knockdown or after drug treatments the survival of cells in comparison to control was determined using the CellTiter 96® Assay as per manufacturer instructions (Promega Corporation, WI, USA). For immunoblotting, standard protocols were used and membranes were probed with antibodies against TTK (anti-MPS1 mouse monoclonal antibody [N1] ab11108 (Abcam, Cambridge), and γ-tubulin (Sigma-Aldrich®) then developed using chemiluminescence reagent plus (Milipore, MA, USA). Flow cytometry to quantify apoptosis was performed using Annexin V-$Alexa_{488}$ and 7-AAD (Life Technologies) as per manufacturer instruction using BD FACSCanto II™ flow cytometer (BD Biosciences, CA, USA).

Breast cancer tissue microarrays, immunohistochemical and survival analysis

The Brisbane Breast Bank collected fresh breast tumor samples from consenting patients; the study was approved by the local ethics committees. Tissue microarrays (TMAs) were constructed from duplicate cores of formalin-fixed, paraffin-embedded (FFPE) breast tumor samples from patients undergoing resection at the Royal Brisbane and Women's Hospital between 1987 and 1994. For biomarker analysis, whole tumor sections or TMAs (depending on the marker) were stained with antibodies against ER, PR, Ki67, HER2, CK5/6, CK14, EGFR and TTK (Table 8), and scored by trained Pathologists. The Vectastain® Universal ABC kit (Vector laboratories, CA) was used for signal detection according to the manufacturer's instructions. Stained sections were scanned at high resolution (ScanScope Aperio, Leica Microsystems, Wetzlar, Germany), and then images were segmented into individual cores for analysis using Spectrum software (Aperio). Survival and other clinical data were collected from the Queensland Cancer Registry and original diagnostic Pathology reports, and in addition we performed an internal histopathological review (SRL) of representative tumor sections from each case, stained with H&E. For analysis of HER2-amplification TMAs were analyzed using HER2 CISH. Criteria for assigning prognostic subgroups in this study are summarized in Figure 14.

Other statistical analysis

Statistical analyses were prepared using GraphPad® Prism v6.0. The types of tests used are stated in Figure Legends. Univariate and multivariate Cox proportional hazards regression analyses were performed using MedCalc for Windows, version 12.7 (MedCalc Software, Ostend, Belgium).

Results

Meta-analysis of gene expression profiles in TNBC

We performed a meta-analysis of published gene expression data, irrespective of platform, using the Oncomine™ database[19] (version 4.5). We compared the expression profiles of 492 TNBC cases vs. 1382 non-TNBC cases in 8 datasets and found 1600 overexpressed and 1580 underexpressed genes in the TNBC cases (cutoff median p-value across the 8 datasets < $1x10^{-5}$ from a Student's *t*-test, Figure 8). We also compared the expression profiles of primary breast cancers from 512 patients who developed metastases vs. 732 patients who did not develop metastases at 5 years (7 datasets in total) to identify 500 overexpressed and 480 underexpressed genes in the metastasis cases (cutoff median p-value across the 7 datasets < 0.05 from a Student's *t*-test, Figure 8). Finally, we compared the expression profiles of 232 primary breast tumors from patients who died within 5 years vs. 879 patients who survived in 7 datasets and found 500 overexpressed and 500 underexpressed genes in the poor survivors (cutoff median p-value across the 7 datasets < 0.05 from a Student's *t*-test, Figure 8). The union of these analyses – genes deregulated in TNBC and in tumors that metastasized or resulted in death within 5 years - generated a gene list of 305 overexpressed and 341 underexpressed genes (Figure 9A&B). The deregulated genes from our analyses did not consider deregulation in comparison to normal breast tissue. To identify cancer-related genes, we used the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) dataset[21] as a validation dataset. Of the 305 overexpressed and 341 underexpressed genes identified in the meta-analysis, 117 overexpressed and 89 underexpressed genes (206 genes) were deregulated in TNBC (250 cases) vs. 144 adjacent normal tissue (1.5 fold-change cutoff; Figure 9C&D).

Clinicopathological features of the aggressiveness gene list

We compared the 206 genes from the above analysis, we called the "aggressiveness gene list" (Table 4), to the recently described metagene attractors[16,17] and found that 45 of the overexpressed genes were in the CIN metagene, whereas 19

of the underexpressed genes were in the ER metagene (Figure 10). The expression of the aggressiveness gene list was visualized in the METABRIC dataset, stratified according to the histological subtypes by the GENUIS classification[22]. As shown in Figure 1A, $ER^-/HER2^-$ (TNBC), in comparison to adjacent normal breast tissue, showed the highest upregulation of CIN genes (red in the heat map) and downregulation of ER signaling genes (green in the heat map). Tumors of other subtypes showed a range of deregulation of these genes. To quantify these trends, we calculated the "aggressiveness score" as the ratio of the CIN metagene (average of expression of CIN genes) to the ER metagene (average of expression of ER genes). The aggressiveness score was highest for $ER^-/HER2^-$ (TNBC), followed by $HER2^+$ then ER^+ tumors (box plot in Figure 1). We also analyzed the aggressiveness score in the five intrinsic breast cancer subtypes predefined by the PAM50 classification[8] and the ten integrative clustering (intClust) subtypes defined by combined clustering of gene expression and copy number data subtypes[21] (Figure 11). The aggressiveness score was highest in the basal-like and the intClust 10 subtypes which are enriched for TNBC and have poor prognosis.

Interestingly, tumors of various subtypes scored higher than the median aggressiveness score (line in box plots in Figure 1 and Figure 11). To this end, we examined the overall survival of patients in the METABRIC dataset stratified by quartiles and also dichotomized by the median of the aggressiveness score. Tumors with high aggressiveness score had worse survival than those with low aggressiveness score. The survival of patients with non-TNBC tumors with high aggressiveness score had poor survival that was similar to TNBC patients (Figure 1B). Among ER^+ tumors we found that high aggressiveness score predicted poor survival in both Grade 2 (Figure 1B) and Grade 3 (Figure 11) tumors. Tumors with high aggressiveness score showed poor survival regardless of the PAM50 intrinsic breast cancer subtypes (Figure 11). The PAM50 classifier was prognostic only in low aggressiveness score tumors (Figure 12).

One network of direct interactions in the aggressiveness gene list associates with patient survival

We performed network analysis on the aggressiveness gene list using the Ingenuity Pathway Analysis (IPA®) and found a network with direct interactions between 97 of the 206 deregulated genes (Figure 2A). To find the minimal genes that represent the aggressiveness genes and this network, the 97 genes in this network

were analyzed for their correlation with the CIN or ER metagenes and overall survival in the METABRIC dataset (Table 5). We selected genes according to the following criteria: (1) highest correlation with the metagenes (Pearson's correlation coefficient > 0.7); (2) association with overall survival (Cox proportional hazards model, p <0.001), and (3) more than 2-fold deregulation with least standard deviation of expression between high and low aggressiveness score tumors. These analyses identified two genes from the ER metagene (*MAPT* and *MYB*) and six genes from the CIN metagene (*MELK*, *MCM10*, *CENPA*, *EXO1*, *TTK* and *KIF2C*). These 8 genes were maintained in a directly connected network (Figure 2B). The classification of tumors (high vs. low across the median) from these eight genes, again representing the ratio of CIN and ER metagenes, predicted the classification from the 206 genes with 95% sensitivity and 97% specificity by prediction of microarray (PAM) analysis (data not shown). Importantly, a high score from these eight genes identified poor survival in all patients, non-TNBC patients and ER^+ Grade 2 (Figure 2C).

Next, we explored the 8-genes score for prognosis in several molecular and histological settings in the METABRIC dataset. The survival of patients with tumors with wild-type TP53 were stratified by the 8-genes score (Figure 3A). Patients with mutant TP53, which were mainly of high score, showed worse survival than those with wild-type TP53, suggesting that TP53 mutation is an independent prognostic factor. Patients with tumors with low or high expression of the proliferation marker Ki67 were stratified by the 8-genes score suggesting that the 8-genes score is independent of proliferation (Figure 3A). We also found that the 8-genes score stratified the survival of patients from all stages of disease (Stage I – Stage III, Figure 3A). We focused on ER^+ and found that, as in the case of ER^+ Grade 2 tumors (Figure 2C); the 8-genes score stratified the survival of patients with ER^+ Grade 3 tumors (Figure 3B). Importantly, the 8-genes score identified ER^+LN^- and ER^+LN^+ patients who had poor survival similar to ER^-LN^- and ER^-LN^+ patients, respectively (Figure 3B). High 8-genes score identified poor survival of patients with tumors of all PAM50 subtypes and the prognostication by PAM50 classification was only evident in low 8-genes score tumors (Figure 12).

The 8-genes aggressiveness score in multivariate survival analysis

To exclude the possibility that the aggressiveness score - calculated using the 206 genes or the 8 genes – was redundant; we performed multivariate Cox-proportional hazards model analysis in the METABRIC dataset (with Illumina

platform) in comparison to conventional clinical variables and current gene signatures. As detailed in Table 1, the aggressiveness scores significantly associated with patient survival when compared with conventional variables and outperformed MammaPrint[9], OncotypeDx[10,11], proliferation/cell cycle[16,20] and CIN[20] signatures. Moreover, our aggressiveness scores outperformed the CIN4 classier[23] which was recently developed from the CIN signature.

We validated the six CIN and two ER genes in univariate survival association using the online tool Kaplan-Meier (KM)-plotter[24] (Tables 6 & 7) which has the gene expression and survival data of more than 2000 patients (but are not part of the METABRIC dataset). We found that the collective expression of the six overexpressed genes (*MELK, MCM10, CENPA, EXO1, TTK* and *KIF2C*) significantly associated with relapse free survival (RFS) and distant metastasis free survival (DMFS) in all patients, ER^+ patients, lymph node negative (LN^-) or positive (LN^+) patients (Table 6). The two underexpressed genes (*MAPT* and *MYB*) also significantly associated with RFS and DMFS in these patient groups (Table 7).

More importantly, we performed multivariate survival analysis of the 8-genes score in four datasets (with Affymetrix platform from the Gene Expression Omnibus [GEO]; GSE2990, GSE3494, GSE2034 and GSE25066). Again, the score was significantly associated with survival in a multivariate Cox-proportional hazards model in every dataset tested (Figure 4). Altogether, we found that in multiple datasets that used different platforms, the 8-genes score identified patients with poor survival independently of other clinico-pathologic indicators and outperforming current signatures.

Therapeutic targets in the aggressiveness gene list

The overexpressed genes in the CIN metagene are involved in or regulate mitosis, spindle assembly and checkpoint, kinetochore attachment, chromosome segregation and mitotic exit. Thus it is not surprising that several of the overexpressed genes are targets for molecular inhibitors, such as CDK1[25,26] and AURKA/AURKB[27] and have been trialed pre-clinically and clinically[28]. To this end, we performed siRNA depletion against 25 genes of the CIN metagene in three TNBC cell lines, MDA-MB-231, SUM159PT and Hs578T. We found that knockdown of four genes (*TTK, TPX2, NDC80* and *PBK*) consistently affected the survival of these cells (Figure 5A and Table 5). The knockdown of *TTK* showed the worst survival and since it was in the 8-genes score we selected TTK for further studies. We found

that TTK protein was higher in TNBC cell lines compared to the near-normal MCF10A cell line, and luminal/HER2 cell lines (Figure 5B). Next, we used the specific TTK inhibitor (TTKi), AZ3146, against a panel of breast cancer cell lines and found that TNBC cell lines were more sensitive to the TTKi (Figure 5C).

TTK expression in aggressive tumors and potential for combination therapy

To further study the potential of TTK as therapeutic target, we investigated TTK expression at the mRNA and protein levels in breast cancer patients. We analyzed the correlation of *TTK* mRNA expression, dichotomized at the median, with clinicopathological indicators in the METABRIC dataset of 2000 patients (Table 2). High *TTK* mRNA expression associated with younger age of tumor diagnosis, larger tumor size, higher tumor grade, higher Ki67 expression, TP53 mutations, an ER/PR negative tumor phenotype, HER2 positivity and TNBC. Based on PAM50 subtyping, high *TTK* mRNA was associated with luminal B, HER2-enriched and basal-like tumors.

We also analyzed TTK expression in a cohort of breast cancer patients (406 patients) by IHC. TTK and its activity is detected at all stages of the cell cycle, however, it is upregulated during mitosis[29]. Thus, we observed TTK staining in non-mitotic cells to define high TTK levels (score of 3) in order to exclude the bias of elevated TTK level during mitosis. Similar to *TTK* mRNA, high TTK protein level (Table 3) associated with high tumor grade, high Ki67 expression and TNBC status (particularly basal TNBC). Moreover, in agreement with the *TTK* mRNA associations with the PAM50 intrinsic subtypes, high TTK protein was observed in HER2-positive and proliferative $ER^+/HER2^-$ tumors (most related to luminal B) but low TTK protein in non-proliferative $ER^+/HER2^-$ tumors (most related to luminal A). In addition to these associations with aggressive phenotypes, we also found that high TTK protein significantly associated with aggressive histological features including ductal histology, pushing tumor border, lymph node involvement, nuclear pleomorphism, lymphocytic infiltration and higher mitotic scores (Table 3). Altogether, like the high aggressiveness score from the 206 or 8 genes, high level of TTK mRNA and protein span across breast cancer subtypes marking aggressive behavior.

We examined the association of TTK protein level with patient survival and found that breast tumors with high TTK staining (category 3) had worse survival than other staining groups at 5 years (Figure 6A&B) and 10 and 20 years (Figure 13).

Importantly, high TTK staining (category 3) was not restricted to a particular histological subgroup or to tumors with high mitotic index (Figure 6C). Next, we focused on prognostication of aggressive subgroups (Grade 3, lymph node positive, TNBC, HER2 or high Ki67) and found that high TTK protein level identified exceptionally aggressive tumors that lead to poor survival of less than 2 years (Figure 7A). Finally, to exploit our finding that TTK, as a part of the aggressiveness score, was associated with aggressive breast tumors and that TTK inhibition was effective in TNBC cell lines that overexpress this protein (Figure 5), we investigated the therapeutic potential of combining TTK inhibition with chemotherapy. We found that TTKi synergized with docetaxel at very low (sub-lethal doses) in the treatment of TNBC cell lines which overexpress TTK in comparison to cell lines which do not (Figure 7B) and that this combination induced apoptotic cell death (Figure 7C).

CIN metagene and ER metagenes in lung adenocarcinoma

There is also reason to believe that the metagene signature may work for other cancers, such as lung cancer. FIG. 15 provides overall survival curves of lung cancer patients split by ten (10) CIN genes that include the aforementioned six (6) (genes as well as *CENPN*, *CEP55*, *FOXM1* and *TPX2*; and the two (2) ER genes *MAPT and MYB* as a signature; patients are low or high according to the median of the signature. The signature outperformed tumour grade and disease stage and remained significant when adjusted for AJCC T (size) and N (lymph node) stages (tumour size (T stage) and lymph node status (N stage) in multivariate Cox regression analysis in lung cancer patients (Table 9). In particular, the signature was prognostic in lung adenocarcinoma. The prognostication of lung adenocarcinoma was significant even when including a minimal gene set of 6 CIN genes and 2 ER genes.

In Figure 16A we show the global gene expression (by RNAseq) of the breast cancer patients in the TCGA dataset. From these data the 8-genes score (Aggressiveness score) and the OncotypeDx (Recurrence score) were investigated for association with survival. The 8-genes score stratified breast cancer survival better than the OncotypeDx (Figure 16B). Further, the 8-genes score (Aggressiveness score) identified tumours with high genomic copy number variations involving whole chromosome arms deletions and duplications reflecting aneuploidy (Figure 16C).

We also find that the 8-genes score (Aggressiveness score) stratifies the survival of all cancers collectively in the TCGA data better than the OncotypeDx (Figure 17) and that the 8-genes score (Aggressiveness score)was prognostic in each of the tested cancers (Figure 18). Similarly, as in breast cancer (Figure 16C), the 8-genes score (Aggressiveness score) identified tumors of all cancer types with high genomic copy number variations involving whole chromosome arms deletions and duplications reflecting aneuploidy (data not shown). These cancer types include breast cancer, bladder cancer, colorectral cancer, glioblastoma, lower grade glioma, head & neck cancer, kidney cancer, liver cancer, lung adenocarcinoma, abute myeloid leukaemia, pancreatic cancer and lung squamous cell carcinoma.

Discussion

This meta-analysis of gene expression in the Oncomine^{TM} database identified a list of 206 was enriched with two core biological functions/metagenes; chromosomal instability (CIN) and ER signaling. We calculated the aggressiveness score, the ratio of CIN to ER metagenes, which associated with overall survival of breast cancer. A core of eight genes (six CIN genes and two ER signaling genes) was representative and recapitulated the correlations with outcome from the 206 genes. The score from the six CIN genes to the 2 ER signaling genes, 8-genes score, associated with survival in several breast cancer datasets. Our aggressiveness scores outperformed conventional variable and published signatures in multivariate survival analysis. Particularly in ER^+ tumors, some cases have survival as poor as that of the aggressive $HER2^+$ and TNBC subtypes. Our data suggest that the interplay of cancer-related biological functions, namely CIN and ER signaling, are better predictors of phenotypes than single genes or single functions. This notion is in line with recent studies showing that the interaction of biologically-driven predictors provide better prognosis[16,17,30]. Recently, all ER^- tumors were described to have a high level of CIN metagene, however, it was not clear that ER^+ tumors could be described as low CIN tumors[16]. In our study, we clarify that ER^+ disease contains a considerable fraction of tumors that have high level of CIN genes and that the relationship between CIN and ER genes is a powerful predictor of survival in these patients.

The fidelity of chromosome segregation is ensured by the proper attachment of the microtubules from the mitotic spindle to the kinetochores of chromosomes in a tightly regulated process and CIN refers to the missegregation of whole

chromosomes thus producing aneuploidy[31]. Using aneuploidy as a surrogate marker for CIN, Carter et al developed a gene signature and found that this "CIN signature" predicts clinical outcome in multiple cancers[20]. More recently, a minimal gene set that captures the CIN signature, CIN4 (*AURKA*, *FOXM1*, *TOP2A* and *TPX2*) was described as the first clinically applicable qPCR derived measure of tumor aneuploidy from FFPE tissue. Since Grade 2 tumors heterogeneous characteristics in terms of clinical outcome, the significance of the CIN4 classier is the stratification of Grade 2 tumors into good and poor prognosis groups[23]. Our aggressiveness scores were prognostic in all tumor grades and disease stages (stages I-III and lymph node negative and positive) and outperformed the CIN signature and the CIN4 classier in multivariate survival analysis in the METABRIC dataset. Strikingly, but in agreement with previous studies[32,33], the prognostication using the CIN metagene and our aggressiveness scores from gene expression levels were restricted to ER^+ disease but not in the TNBC or HER2 subtypes. This may be explained that ER^- tumors have a high level of CIN metagene as per our results and published previously[16]. However, our results with TTK protein level clearly demonstrate that TNBC, HER2, high grade, lymph node positive and proliferative tumors contain subgroups with high TTK levels exclusive of mitotic cells and have poorer survival than those with low TTK expression or TTK expression in mitotic cells. We propose that there are two types of high expression of CIN genes that may not be clearly differentiated by mRNA expression studies. One form of elevated CIN genes relates to high level of mitosis and proliferation whereas the second form that we measured by IHC exclusive of mitotic cells is driven by another aggressive phenotype; protection of aneuploidy and genomic instability. The recent study of the CIN4 classifier lends support to our proposition. In this study, using flow cytometry to measure aneuploidy by DNA content, the authors found that a substantial proportion of tumors with high CIN4 scores have a normal DNA ploidy and that a significant proportion of aneuploid cases had low CIN4 score[23].

Chromosome missegregation and aneuploidy enhance genetic recombination and defective DNA damage repair[34] to drive a "mutator phenotype" required for oncogenesis[35]. Genomic instability caused by deregulated mitotic spindle assembly checkpoint (SAC) and aneuploidy has been termed "non-oncogene addiction"[36,37]. It is tempting to suggest that CIN and aneuploidy are exploited by breast cancer stem cells which are high in TNBC[38] due to the link between cancer stem cells, aneuploidy

and therapy resistance[39,40]. This is supported by studies that implicate several genes involved in the SAC and chromosome segregation in tumor initiation, progression and cancer stem cells, e.g. AURKA in ovarian cancer[41], MELK/FOXM1 in glioblastoma[42,43], MELK[44] and MAD2[45] in breast cancer and SKP2 in several cancers[46]. The role of CIN genes to protecting aneuploidy could provide an insight to the paradox that TNBC show a better response to chemotherapy due to higher level of proliferation, yet these tumors have poorer outcome. We propose that resistance in TNBC could be attributed to the ability of aneuploid cells to adapt and drive recurrence. At least *in vivo*, chemotherapy has been shown to induce the proliferation quiescent aneuploid cells as a mechanism for therapy resistance[39]. We envisage that the high level of the CIN metagene in TNBC, particularly genes involved in chromosome segregation, is protective of this state. Indeed, one study found that a high level of TTK is protective of aneuploidy in breast cancer cells and its silencing reduces the tumorigenicity of breast cancer cell lines *in vivo*[47]. Our results from the patient cohort demonstrate that high TTK protein expression exclusive of mitosis was indeed prognostic aggressive tumors and support the concept that protection from aneuploidy and genomic instability is an aggressive phenotype that drives poor outcome.

Our results with the TTK molecular inhibitor, in agreement with published studies using siRNA depletion[47,48], supports the idea of targeting chromosomal segregation in tumors with a high CIN phenotype as a therapeutic strategy. We also suggest that while TTK is high in TNBC as previously described[47,48], a considerable proportion of non-TNBC tumors that display aggressive features also show an elevated level of CIN genes, and would benefit from such targeted therapies. To our knowledge the combination of sub-lethal doses of taxanes with TTK inhibition has not been investigated so far in breast cancer, but in other cancers[33,50-53]. Our results reveal that TTK inhibition indeed sensitizes breast cancer cells with high TTK to docetaxel.

Referring particularly in FIGS 16-18, as well as the 8-genes score (Aggressiveness score) being prognostic for the survival of cancer patients after treatment, the aggressiveness score also identifies tumors with high copy number variations involving whole chromosome arms reflecting aneuploid status. Thus, the aggressiveness score may also serve as a companion diagnostic for drugs that target aneuploidy by means of targeting genes listed in Table 4, inclusive of the 8 genes

used to produce the aggressiveness score (such as TTK[67-70]) or by other drugs that target the aneuploidy state (such as PLK1[71,72] or others[73-76]).

In conclusion, our study emphasizes that classification of breast cancer based on biological phenotypes facilitates understanding the drivers of oncogenic phenotypes and therapeutic potentials. Importantly, our studies demonstrate that IHC assessment of CIN genes, exemplified by TTK here; provide better characterization and understanding for the contribution of CIN to tumor aggressiveness and prognosis.

Throughout this specification, the aim has been to describe the preferred embodiments of the invention without limiting the invention to any one embodiment or specific collection of features. Various changes and modifications may be made to the embodiments described and illustrated herein without departing from the broad spirit and scope of the invention.

All computer programs, algorithms, patent and scientific literature referred to herein is incorporated herein by reference in their entirety.

REFERENCES

1. Liedtke C, Mazouni C, Hess KR, Andre F, Tordai A, Mejia JA *et al*. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol 2008; 26: 1275-1281.
2. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F *et al*. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res 2007; 13: 2329-2334.
3. von Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, Fasching PA *et al*. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol 2012; 30: 1796-1804.
4. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al*. 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 2001; 98: 10869-10874.
5. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA *et al*. Molecular portraits of human breast tumours. Nature 2000; 406: 747-752.
6. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG *et al*. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer research 2005; 65: 9155-9158.
7. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF *et al*. The molecular portraits of breast tumors are conserved across microarray platforms. BMC genomics 2006; 7: 96-107.
8. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T *et al*. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009; 27: 1160-1167.
9. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M *et al*. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415: 530-536.
10. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M *et al*. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. The New England journal of medicine 2004; 351: 2817-2826.

11. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM *et al.* Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. Journal of the National Cancer Institute 2006; 98: 1183-1192.

12. Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C *et al.* Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol 2007; 25: 1239-1246.

13. Ma XJ, Salunga R, Dahiya S, Wang W, Carney E, Durbecq V *et al.* A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. Clin Cancer Res 2008; 14: 2601-2608.

14. Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A *et al.* A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer cell 2004; 5: 607-616.

15. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J *et al.* Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. Journal of the National Cancer Institute 2006; 98: 262-272.

16. Cheng WY, Ou Yang TH, Anastassiou D. Biomolecular events in cancer revealed by attractor metagenes. PLoS Comput Biol 2013; 9: e1002920.

17. Cheng WY, Ou Yang TH, Anastassiou D. Development of a prognostic model for breast cancer survival in an open challenge environment. Sci Transl Med 2013; 5: 181ra150.

18. Dai H, van't Veer L, Lamb J, He YD, Mao M, Fine BM *et al.* A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. Cancer research 2005; 65: 4059-4066.

19. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia (New York, NY 2004; 6: 1-6.

20. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nature genetics 2006; 38: 1043-1048.

21. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ *et al*. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486: 346-352.

22. Haibe-Kains B, Desmedt C, Rothe F, Piccart M, Sotiriou C, Bontempi G. A fuzzy gene expression-based computational approach improves breast cancer prognostication. Genome Biol 2010; 11: R18.

23. Szasz AM, Li Q, Eklund AC, Sztupinszki Z, Rowan A, Tokes AM *et al*. The CIN4 chromosomal instability qPCR classifier defines tumor aneuploidy and stratifies outcome in grade 2 breast cancer. PLoS ONE 2013; 8: e56707.

24. Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q *et al*. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast cancer research and treatment 2010; 123: 725-731.

25. Rizzolio F, Tuccinardi T, Caligiuri I, Lucchetti C, Giordano A. CDK inhibitors: from the bench to clinical trials. Curr Drug Targets 2010; 11: 279-290.

26. Horiuchi D, Kusdra L, Huskey NE, Chandriani S, Lenburg ME, Gonzalez-Angulo AM *et al*. MYC pathway activation in triple-negative breast cancer is synthetic lethal with CDK inhibition. The Journal of experimental medicine 2012; 209: 679-696.

27. Manchado E, Guillamot M, Malumbres M. Killing cells by targeting mitosis. Cell death and differentiation 2012.

28. Janssen A, Medema RH. Mitosis as an anti-cancer target. Oncogene 2011; 30: 2799-2809.

29. Stucke VM, Sillje HH, Arnaud L, Nigg EA. Human Mps1 kinase is required for the spindle assembly checkpoint but not for centrosome duplication. The EMBO journal 2002; 21: 1723-1732.

30. Nagalla S, Chou JW, Willingham MC, Ruiz J, Vaughn JP, Dubey P *et al*. Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis. Genome Biol 2013; 14: R34.

31. Bakhoum SF, Compton DA. Chromosomal instability and cancer: a complex relationship with therapeutic potential. The Journal of clinical investigation 2012; 122: 1138-1143.

32. Roylance R, Endesfelder D, Gorman P, Burrell RA, Sander J, Tomlinson I *et al*. Relationship of extreme chromosomal instability with long-term survival in a retrospective analysis of primary breast cancer. Cancer Epidemiol Biomarkers Prev 2011; 20: 2183-2194.

33. Birkbak NJ, Eklund AC, Li Q, McClelland SE, Endesfelder D, Tan P *et al*. Paradoxical relationship between chromosomal instability and survival outcome in cancer. Cancer research 2011; 71: 3447-3452.

34. Janssen A, van der Burg M, Szuhai K, Kops GJ, Medema RH. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. Science (New York, NY 2011; 333: 1895-1898.

35. Kolodner RD, Cleveland DW, Putnam CD. Cancer. Aneuploidy drives a mutator phenotype in cancer. Science (New York, NY 2011; 333: 942-943.

36. Luo J, Solimini NL, Elledge SJ. Principles of cancer therapy: oncogene and non-oncogene addiction. Cell 2009; 136: 823-837.

37. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144: 646-674.

38. Al-Ejeh F, Smart CE, Morrison BJ, Chenevix-Trench G, Lopez JA, Lakhani SR *et al*. Breast cancer stem cells: treatment resistance and therapeutic opportunities. Carcinogenesis 2011; 32: 650-658.

39. Kusumbe AP, Bapat SA. Cancer stem cells and aneuploid populations within developing tumors are the major determinants of tumor dormancy. Cancer research 2009; 69: 9245-9253.

40. Liang Y, Zhong Z, Huang Y, Deng W, Cao J, Tsao G *et al*. Stem-like cancer cells are inducible by increasing genomic instability in cancer cells. The Journal of biological chemistry 2010; 285: 4931-4940.

41. Do TV, Xiao F, Bickel LE, Klein-Szanto AJ, Pathak HB, Hua X *et al*. Aurora kinase A mediates epithelial ovarian cancer cell migration and adhesion. Oncogene 2014; 33: 539-549.

42. Joshi K, Banasavadi-Siddegowda Y, Mo X, Kim SH, Mao P, Kig C *et al*. MELK-dependent FOXM1 phosphorylation is essential for proliferation of glioma stem cells. Stem cells (Dayton, Ohio) 2013; 31: 1051-1063.

43. Gu C, Banasavadi-Siddegowda YK, Joshi K, Nakamura Y, Kurt H, Gupta S *et al*. Tumor-specific activation of the C-JUN/MELK pathway regulates

glioma stem cell growth in a p53-dependent manner. Stem cells (Dayton, Ohio) 2013; 31: 870-881.

44. Hebbard LW, Maurer J, Miller A, Lesperance J, Hassell J, Oshima RG *et al.* Maternal embryonic leucine zipper kinase is upregulated and required in mammary tumor-initiating cells in vivo. Cancer research 2010; 70: 8863-8873.

45. Schvartzman JM, Duijf PH, Sotillo R, Coker C, Benezra R. Mad2 is a critical mediator of the chromosome instability observed upon Rb and p53 pathway inhibition. Cancer cell 2011; 19: 701-714.

46. Chan CH, Morrow JK, Li CF, Gao Y, Jin G, Moten A *et al.* Pharmacological inactivation of Skp2 SCF ubiquitin ligase restricts cancer stem cell traits and cancer progression. Cell 2013; 154: 556-568.

47. Daniel J, Coulter J, Woo JH, Wilsbach K, Gabrielson E. High levels of the Mps1 checkpoint protein are protective of aneuploidy in breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America 2011; 108: 5384-5389.

48. Maire V, Baldeyron C, Richardson M, Tesson B, Vincent-Salomon A, Gravier E *et al.* TTK/hMPS1 is an attractive therapeutic target for triple-negative breast cancer. PLoS ONE 2013; 8: e63712.

49. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 2006; 439: 353-357.

50. Bittner M. Expression Project for Oncology - Breast Samples. International Genomics Consortium, Phoeniz, AZ 85004 Oncomine. Not Published 2005/01/15

51. Bonnefoi H, Potti A, Delorenzi M, Mauriac L, Campone M, Tubiana-Hulin M *et al.* Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial. The lancet oncology 2007; 8: 1071-1078.

52. Gluck S, Ross JS, Royce M, McKenna EF, Jr., Perou CM, Avisar E *et al.* TP53 genomics predict higher clinical and pathologic tumor response in operable early-stage breast cancer treated with docetaxel-capecitabine +/- trastuzumab. Breast cancer research and treatment 2012; 132: 781-791.

53. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A *et al.* A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. JAMA 2011; 305: 1873-1881.

54. Kao KJ, Chang KM, Hsu HC, Huang AT. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. BMC cancer 2011; 11: 143.

55. Tabchy A, Valero V, Vidaurre T, Lluch A, Gomez H, Martin M *et al.* Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer. Clin Cancer Res 2010; 16: 5351-5361.

56. TCGA. Comprehensive molecular portraits of human breast tumours. Nature 2012; 490: 61-70.

57. Bos PD, Zhang XH, Nadal C, Shu W, Gomis RR, Nguyen DX *et al.* Genes that mediate breast cancer metastasis to the brain. Nature 2009; 459: 1005-1009.

58. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B *et al.* Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res 2007; 13: 3207-3214.

59. Schmidt M, Bohm D, von Torne C, Steiner E, Puhl A, Pilch H *et al.* The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer research 2008; 68: 5405-5413.

60. Symmans WF, Hatzis C, Sotiriou C, Andre F, Peintinger F, Regitnig P *et al.* Genomic index of sensitivity to endocrine therapy for breast cancer. J Clin Oncol 2010; 28: 4111-4119.

61. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW *et al.* A gene-expression signature as a predictor of survival in breast cancer. The New England journal of medicine 2002; 347: 1999-2009.

62. Pawitan Y, Bjohle J, Amler L, Borg AL, Egyhazi S, Hall P *et al.* Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res 2005; 7: R953-964.

63. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A *et al.* Repeated observation of breast tumor subtypes in independent gene

expression data sets. Proceedings of the National Academy of Sciences of the United States of America 2003; 100: 8418-8423.

64. Zhao Y, Simon R. BRB-ArrayTools Data Archive for human cancer gene expression: a unique and efficient data sharing resource. Cancer Inform 2008; 6: 9-15.

65. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A *et al.* An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proceedings of the National Academy of Sciences of the United States of America 2005; 102: 13550-13555.

66. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005; 365: 671-679.

67. Al-Ejeh F, Shi W, Miranda M, Simpson PT, Vargas AC, Song S *et al.* Treatment of triple-negative breast cancer using anti-EGFR-directed radioimmunotherapy combined with radiosensitizing chemotherapy and PARP inhibitor. J Nucl Med 2013; 54: 913-921.

67. Colombo R, Caldarelli M, Mennecozzi M, Giorgini ML, Sola F, Cappella P *et al.* Targeting the mitotic checkpoint for cancer therapy with NMS-P715, an inhibitor of MPS1 kinase. Cancer research 2010; 70: 10255-10264.

68. Janssen A, Kops GJ, Medema RH. Elevating the frequency of chromosome mis-segregation as a strategy to kill tumor cells. Proceedings of the National Academy of Sciences of the United States of America 2009; 106: 19108-19113.

69. Janssen A, Medema RH. Mitosis as an anti-cancer target. Oncogene 2011; 30: 2799-2809.

70. Janssen A, van der Burg M, Szuhai K, Kops GJ, Medema RH. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. Science (New York, NY 2011; 333: 1895-1898.

71. Degenhardt Y, Lampkin T. Targeting Polo-like kinase in cancer therapy. Clin Cancer Res 2010; 16: 384-389.

72. Strebhardt K, Ullrich A. Targeting polo-like kinase 1 for cancer therapy. Nature reviews 2006; 6: 321-330.

73. Malumbres M, Barbacid M. Cell cycle kinases in cancer. Curr Opin Genet Dev 2007; 17: 60-65.

74. Manchado E, Guillamot M, Malumbres M. Killing cells by targeting mitosis. Cell death and differentiation 2012.

75. Manchado E, Malumbres M. Targeting aneuploidy for cancer therapy. Cell 2011; 144: 465-466.

76. Colombo R, Moll J. Destabilizing aneuploidy by targeting cell cycle and mitotic checkpoint proteins in cancer cells. Curr Drug Targets 2010; 11: 1325-1335.

Table 1: Univariate and multivariate survival analysis of the aggressiveness score in the METABRIC dataset

	Univariate Cox-proportional hazards model		Multivariate Cox-proportional hazards model (stepwise)	
	HR (95 % CI)	p-value	HR (95 % CI)	p-value
206 genes score (high, low)	1.6173 (1.4174 - 1.8454)	<0.0001	1.5188 (1.3227 - 1.7440)	<0.0001
8 genes score (high, low)	1.5853 (1.2883 - 1.8103)	<0.0001	1.4760 (1.2198 - 1.6344)	<0.0001
Lymph node (+, -)	1.8594 (1.6289 - 2.1224)	<0.0001	1.6807 (1.4610 - 1.9334)	<0.0001
Tumor size (T1, T2, T3)	1.4354 (1.2813 - 1.6080)	<0.0001	1.3666 (1.1642 - 1.6041)	0.0001
HER2 status (+, -)	1.4565 (1.2537 - 1.6920)	<0.0001	1.1983 (1.0183 - 1.4101)	0.0302
Tumor grade (1, 2, 3)	1.3500 (1.2095 - 1.5067)	<0.0001	ns	ns
Ki67 (+, -)	1.4184 (1.2399 - 1.6226)	<0.0001	ns	ns
MammaPrint (high, low)	1.3320 (1.1669 -1.5204)	<0.0001	ns	ns
CIN4 (high, low)	1.5310 (1.3413 - 1.7476)	<0.0001	ns	ns
CIN75 (high, low)	1.5004 (1.3132 - 1.7143)	<0.0001	ns	ns
Cell Cycle (high, low)	1.5018 (1.3145 - 1.7158)	<0.0001	ns	ns
ER status (+, -)	1.3016 (1.1167 - 1.5170)	0.0008	ns	ns
OncotypeDx (L, I, H)	1.2672 (1.0909 - 1.4720)	0.0021	ns	ns
Treatment (yes, no)	1.1646 (0.9753 - 1.2639)	0.0939		
Age (<40, >40)	1.1235 (0.8480 - 1.4886)	0.4196		

HR: Hazard Ratio. CI: confidence interval. ns: not significant. OncoTypeDx scores are low (L, < 18), intermediate (I, 18-31), high (H >31). All variables were included in the multivariate Cox-proportional hazards model analysis and by stepwise model, only significant co-variants were included in the final analysis shown in Table.

Table 2: Correlation of *TTK* mRNA level and clinico-pathological indicators in the METABRIC dataset

Comparison		TTK Low	TTK high	X^2
Tumor size				
	<2cm	346 (18%)	280 (14%)	p<1.0E-6
	>2cm <5cm	509 (26%)	685 (35%)	p=3.2E-5
	>5cm	60 (3%)	92 (5%)	p=1.25E-2
Tumor Grade				
	Grade 1	137 (7%)	33 (2%)	p<1.0E-6
	Grade 2	479 (25%)	296 (16%)	p<1.0E-6
	Grade 3	251 (13%)	706 (37%)	p<1.0E-6
Ki67 expression				
	Low	826 (39%)	242 (11%)	
	High	237 (11%)	831 (39%)	p<1.0E-6
Immunohistochemical subtypes				
	ER negative	71 (4%)	369 (19%)	p<1.0E-6
	ER positive	827 (42%)	681 (35%)	
	PR negative	306 (15%)	637 (32%)	p<1.0E-6
	PR positive	617 (31%)	432 (22%)	
	HER2 negative	802 (40%)	744 (37%)	
	HER2 positive	118 (6%)	323 (16%)	p<1.0E-6
	non-TNBC	885 (45%)	840 (43%)	
	Triple negative (TNBC)	29 (1%)	221 (11%)	p<1.0E-6
Intrinsic subtypes				
	Luminal A	552 (28%)	169 (9%)	p<1.0E-6
	Luminal B	142 (7%)	350 (18%)	p<1.0E-6
	HER2-enriched	40 (2%)	200 (10%)	p<1.0E-6
	Normal-like	161 (8%)	41 (2%)	p<1.0E-6
	Basal-like	26 (1%)	305 (15%)	p<1.0E-6
Age (years)				
	<50	167 (8%)	259 (13%)	p=8.68E-4
	50-74	485 (24%)	549 (27%)	ns
	75-100	282 (14%)	253 (13%)	ns
TP53 mutation				
	Wildtype	390 (48%)	331 (40%)	
	Mutant	14 (2%)	85 (10%)	p<1.0E-6

X2: Chi square test performed using GraphPad® Prism. ns not significant

Table 3: Associations between TTK protein expression and clinico-pathological indicators

Parameter	TTK (0-1)	TTK (2)	TTK (3)	P value[#]
Histological type				
Ductal NOS	147(60.7 %)	67(27.7 %)	28 (11.6 %)	0.0265
Lobular	43(76.8 %)	10 (17.9 %)	3 (5.4 %)	
Mixed ducto-lobular	31(88.6 %)	4 (11.4 %)	0 (0.0 %)	
Metaplastic	9(56.3 %)	7 (43.8 %)	0 (0.0 %)	
Tubular/cribiform	8(80.0 %)	2 (20.0 %)	0 (0.0 %)	
Other special types (incl mixed)	37(66.1 %)	14 (25.0 %)	5 (8.9 %)	
Overall grade				
1	43(76.8 %)	13(23.2 %)	0 (0.0 %)	<0.0001
2	162(77.5 %)	41 (19.6 %)	6 (2.9 %)	
3	73(47.7 %)	50 (32.7 %)	30 (19.6 %)	
Mitotic score				
1	193(79.8 %)	44(18.2 %)	5 (2.1 %)	<0.0001
2	33(61.1 %)	18 (33.3 %)	3 (5.6 %)	
3	52(43.0 %)	42 (34.7 %)	27 (22.3 %)	
Nuclear pleomorphism score				
1-2	164(75.2 %)	49(22.5 %)	5 (2.3 %)	<0.0001
3	115(57.2 %)	55 (27.4 %)	31 (15.4 %)	
Tubule score				
1	10(76.9 %)	3(23.1 %)	0 (0.0 %)	ns
2	52(69.3 %)	20 (26.7 %)	3 (4.0 %)	
3	216(65.5 %)	81 (24.5 %)	33 (10.0 %)	
Lymph node status				
Positive	77(62.1 %)	41(33.1 %)	6 (4.8 %)	0.0056
Negative	81(73.0 %)	18 (16.2 %)	12 (10.8 %)	
Tumor size				
<2 cm	112(68.3 %)	40(24.4 %)	12 (7.3 %)	ns
2-5 cm	104(66.2 %)	38 (24.2 %)	15 (9.6 %)	
>5 cm	19(61.3 %)	6 (19.4 %)	6 (19.4 %)	
Lymphovascular invasion				
Absent	214(67.3 %)	77(24.2 %)	27 (8.5 %)	ns
Present	63(63.6 %)	27 (27.3 %)	9 (9.1 %)	
Lymphocytic infiltrate				
Absent	119(78.3 %)	28(18.4 %)	5 (3.3 %)	0.0007
Mild	115(63.9 %)	47 (26.1 %)	18 (10.0 %)	
Moderate	36(53.7 %)	23 (34.3 %)	8 (11.9 %)	
Severe	7(41.2 %)	6 (35.3 %)	4 (23.5 %)	
Central scarring/fibrosis				
Absent	254(67.7 %)	90(24.0 %)	31 (8.3 %)	ns
Present	25(56.8 %)	14 (31.8 %)	5 (11.4 %)	
Tumor border				
Infiltrative	250(69.1 %)	88(24.3 %)	24 (6.6 %)	0.0003
Pushing (<50%)	11(36.7 %)	11 (36.7 %)	8 (26.7 %)	
Pushing (>50%)	16(64.0 %)	5 (20.0 %)	4 (16.0 %)	
Ki67 expression (20% threshold)				
Low	240(71.6 %)	77(23.0 %)	18 (5.4 %)	<0.0001
High	14(25.9 %)	23 (42.6 %)	17 (31.5 %)	
Prognostic subgroups				
HER2+	21(51.2 %)	14(34.1 %)	6 (14.6 %)	<0.0001
HR+/HER2-neg (Ki67-high)	6(24.0 %)	13 (52.0 %)	6 (24.0 %)	
HR+/HER2-neg (Ki67-low)	196(76.0 %)	53 (20.5 %)	9 (3.5 %)	
TN (basal-like)	23(41.8 %)	20 (36.4 %)	12 (21.8 %)	
TN (non-basal)	10(71.4 %)	1 (7.1 %)	3 (21.4 %)	

TMAs were scored by two independent assessors according to the following categories: 0, negative; 1, weak and focal staining (pooled with negative cases for this analysis); 2, moderate-strong focal staining (collectively <50% of tumour cells); 3 = moderate-strong diffuse staining (>50% of tumour cells). Regarding % cells stained, we disregarded mitotic cells to assess mitosis-independent TTK expression. [#]Chi square test (GraphPad® Prism. ns: not significant)

Table 4: The aggressiveness genelist (206 genes)

Input	Approved Name	HGNC ID	Location
ADIRF	adipogenesis regulatory factor	HGNC:24043	10q23.31
AFF3	AF4/FMR2 family, member 3	HGNC:6473	2q11.2-q12
AGO2	argonaute RISC catalytic component 2	HGNC:3263	8q24.3
AGR3	anterior gradient 3 homolog (Xenopus laevis)	HGNC:24167	7p21.1
AHNAK	AHNAK nucleoprotein	HGNC:347	11q12-q13
ALDH3A2	aldehyde dehydrogenase 3 family, member A2	HGNC:403	17p11.2
ANLN	anillin, actin binding protein	HGNC:14082	7p15-p14
APOBEC3B	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B	HGNC:17352	22q13.1-q13.2
AQP9	aquaporin 9	HGNC:643	15q
ATP6V1C2	ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C2	HGNC:18264	2p25.1
AUNIP	aurora kinase A and ninein interacting protein	HGNC:28363	1p36.11
AURKA	aurora kinase A	HGNC:11393	20q13
AURKB	aurora kinase B	HGNC:11390	17p13.1
AZGP1	alpha-2-glycoprotein 1, zinc-binding	HGNC:910	7q22.1
BBS1	Bardet-Biedl syndrome 1	HGNC:966	11q13
BCL2	B-cell CLL/lymphoma 2	HGNC:990	18q21.3
BIRC5	baculoviral IAP repeat containing 5	HGNC:593	17q25.3
BLM	Bloom syndrome, RecQ helicase-like	HGNC:1058	15q26.1
BTG2	BTG family, member 2	HGNC:1131	1q32
BUB1	BUB1 mitotic checkpoint serine/threonine kinase	HGNC:1148	2q13
BYSL	bystin-like	HGNC:1157	6p21.1
C10orf32	chromosome 10 open reading frame 32	HGNC:23516	10q24.33
C18orf56	chromosome 18 open reading frame 56	HGNC:29553	18p11.32
C1orf106	chromosome 1 open reading frame 106	HGNC:25599	1q32.1
C1orf21	chromosome 1 open reading frame 21	HGNC:15494	1q25
C7orf63	chromosome 7 open reading frame 63	HGNC:26107	7q21.13
CA9	carbonic anhydrase IX	HGNC:1383	9p13.3
CARD10	caspase recruitment domain family, member 10	HGNC:16422	22q13.1
CASC1	cancer susceptibility candidate 1	HGNC:29599	12p12.1
CCDC170	coiled-coil domain containing 170	HGNC:21177	6q25.1
CCDC176	coiled-coil domain containing 176	HGNC:19855	14q24.3
CCNA2	cyclin A2	HGNC:1578	4q27
CCNB2	cyclin B2	HGNC:1580	15q21.3
CCNE1	cyclin E1	HGNC:1589	19q12
CCNG2	cyclin G2	HGNC:1593	4q21.22
CD163	CD163 molecule	HGNC:1631	12p13
CDC20	cell division cycle 20	HGNC:1723	1p34.1
CDC25A	cell division cycle 25A	HGNC:1725	3p21
CDC25B	cell division cycle 25B	HGNC:1726	20p13
CDC45	cell division cycle 45	HGNC:1739	22q11.21

CDCA3	cell division cycle associated 3	HGNC:14624	12p13.31
CDCA5	cell division cycle associated 5	HGNC:14626	11q13.1
CDCA7	cell division cycle associated 7	HGNC:14628	2q31.1
CDCA8	cell division cycle associated 8	HGNC:14629	1p34.3
CDK1	cyclin-dependent kinase 1	HGNC:1722	10q21.2
CDKN2A	cyclin-dependent kinase inhibitor 2A	HGNC:1787	9p21
CENPA	centromere protein A	HGNC:1851	2p23.3
CENPE	centromere protein E, 312kDa	HGNC:1856	4q24-q25
CENPN	centromere protein N	HGNC:30873	16q23.2
CENPW	centromere protein W	HGNC:21488	6q22.32
CEP55	centrosomal protein 55kDa	HGNC:1161	10q24.1
CHEK1	checkpoint kinase 1	HGNC:1925	11q24.2
CIRBP	cold inducible RNA binding protein	HGNC:1982	19p13.3
CKAP2L	cytoskeleton associated protein 2-like	HGNC:26877	2q13
CKS1B	CDC28 protein kinase regulatory subunit 1B	HGNC:19083	1q21.2
CKS2	CDC28 protein kinase regulatory subunit 2	HGNC:2000	9q22
CLIC6	chloride intracellular channel 6	HGNC:2065	21q22.12
CMC2	COX assembly mitochondrial protein 2 homolog (S. cerevisiae)	HGNC:24447	16q23.2
CMYA5	cardiomyopathy associated 5	HGNC:14305	5q14.1
CPEB2	cytoplasmic polyadenylation element binding protein 2	HGNC:21745	4p15.33
CST3	cystatin C	HGNC:2475	20p11.2
CSTB	cystatin B (stefin B)	HGNC:2482	21q22.3
CTSV	cathepsin V	HGNC:2538	9q22.33
CYB5D1	cytochrome b5 domain containing 1	HGNC:26516	17p13.1
CYBRD1	cytochrome b reductase 1	HGNC:20797	2q31
DACH1	dachshund homolog 1 (Drosophila)	HGNC:2663	13q22
DAPK1	death-associated protein kinase 1	HGNC:2674	9q34.1
DEPDC1	DEP domain containing 1	HGNC:22949	1p31.2
DKC1	dyskeratosis congenita 1, dyskerin	HGNC:2890	Xq28
DLGAP5	discs, large (Drosophila) homolog-associated protein 5	HGNC:16864	14q22.3
DNAJC12	DnaJ (Hsp40) homolog, subfamily C, member 12	HGNC:28908	10q21.3
DNALI1	dynein, axonemal, light intermediate chain 1	HGNC:14353	1p35.1
ECI2	enoyl-CoA delta isomerase 2	HGNC:14601	6p24.3
ELOVL5	ELOVL fatty acid elongase 5	HGNC:21308	6p21.1-p12.1
ESR1	estrogen receptor 1	HGNC:3467	6q24-q27
EXO1	exonuclease 1	HGNC:3511	1q42-q43
FAM198B	family with sequence similarity 198, member B	HGNC:25312	4q32.1
FAM214A	family with sequence similarity 214, member A	HGNC:25609	15q21.2-q21.3
FAM64A	family with sequence similarity 64, member A	HGNC:25483	17p13.2
FAM83D	family with sequence similarity 83, member D	HGNC:16122	20
FOXA1	forkhead box A1	HGNC:5021	14q12-q13
FOXM1	forkhead box M1	HGNC:3818	12p13

FPR3	formyl peptide receptor 3	HGNC:3828	19q13.3-q13.4
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	HGNC:4141	12p13.31
GFRA1	GDNF family receptor alpha 1	HGNC:4243	10q25-q26
GGH	gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)	HGNC:4248	8q12.3
GLI3	GLI family zinc finger 3	HGNC:4319	7p13
GLYATL2	glycine-N-acyltransferase-like 2	HGNC:24178	11q12.1
GPD1L	glycerol-3-phosphate dehydrogenase 1-like	HGNC:28956	3p22.3
GPSM2	G-protein signaling modulator 2	HGNC:29501	1p13.3
GSTM1	glutathione S-transferase mu 1	HGNC:4632	1p13.3
GSTM3	glutathione S-transferase mu 3 (brain)	HGNC:4635	1p13.3
GTPBP4	GTP binding protein 4	HGNC:21535	10p15-p14
GTSE1	G-2 and S-phase expressed 1	HGNC:13698	22q13.2-q13.3
HJURP	Holliday junction recognition protein	HGNC:25444	2q37.1
HRASLS	HRAS-like suppressor	HGNC:14922	3q29
HSD17B4	hydroxysteroid (17-beta) dehydrogenase 4	HGNC:5213	5q2
HSD17B8	hydroxysteroid (17-beta) dehydrogenase 8	HGNC:3554	6p21.3
IGFBP2	insulin-like growth factor binding protein 2, 36kDa	HGNC:5471	2q33-q34
IGFBP4	insulin-like growth factor binding protein 4	HGNC:5473	17q12-q21.1
IL6ST	interleukin 6 signal transducer (gp130, oncostatin M receptor)	HGNC:6021	5q11.2
IL8	interleukin 8	HGNC:6025	4q13-q21
IMPA2	inositol(myo)-1(or 4)-monophosphatase 2	HGNC:6051	18p11.2
IRAK1	interleukin-1 receptor-associated kinase 1	HGNC:6112	Xq28
KCNG1	potassium voltage-gated channel, subfamily G, member 1	HGNC:6248	20q13
KCNMA1	potassium large conductance calcium-activated channel, subfamily M, alpha member 1	HGNC:6284	10q22
KCTD3	potassium channel tetramerization domain containing 3	HGNC:21305	1q41
KIF13B	kinesin family member 13B	HGNC:14405	8p21
KIF14	kinesin family member 14	HGNC:19181	1q32.1
KIF20A	kinesin family member 20A	HGNC:9787	5q31
KIF23	kinesin family member 23	HGNC:6392	15q23
KIF2C	kinesin family member 2C	HGNC:6393	1p34.1
KIF5C	kinesin family member 5C	HGNC:6325	2q23
KRT6A	keratin 6A	HGNC:6443	12q13.13
LAD1	ladinin 1	HGNC:6472	1q25.1-q32.3
LAPTM4B	lysosomal protein transmembrane 4 beta	HGNC:13646	8q22.1
LFNG	LFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase	HGNC:6560	7p22.3
LMNB2	lamin B2	HGNC:6638	19p13.3
LOC100286909	-	-	-
LRIG1	leucine-rich repeats and immunoglobulin-like domains 1	HGNC:17360	3p14
LRP8	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor	HGNC:6700	1p32.3
LYPD6	LY6/PLAUR domain containing 6	HGNC:28751	2q23.2

MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)	HGNC:6763	4q27
MAPT	microtubule-associated protein tau	HGNC:6893	17q21
MCM10	minichromosome maintenance complex component 10	HGNC:18043	10p13
MCM2	minichromosome maintenance complex component 2	HGNC:6944	3q21
MCM4	minichromosome maintenance complex component 4	HGNC:6947	8q12-q13
MCM6	minichromosome maintenance complex component 6	HGNC:6949	2q14-q21
MCM7	minichromosome maintenance complex component 7	HGNC:6950	7q21.3-q22.1
MEIS3P1	Meis homeobox 3 pseudogene 1	HGNC:7002	17p12
MELK	maternal embryonic leucine zipper kinase	HGNC:16870	9p13.1
MLPH	melanophilin	HGNC:29643	2q37.2
MST1	macrophage stimulating 1 (hepatocyte growth factor-like)	HGNC:7380	3p21
MTHFD1L	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like	HGNC:21055	6q25.1
MX2	myxovirus (influenza virus) resistance 2 (mouse)	HGNC:7533	21q22.3
MYB	v-myb avian myeloblastosis viral oncogene homolog	HGNC:7545	6q22-q23
NCAPG	non-SMC condensin I complex, subunit G	HGNC:24304	4p15.32
NDC80	NDC80 kinetochore complex component	HGNC:16909	18p11.31
NFIA	nuclear factor I/A	HGNC:7784	1p31.3-p31.2
NME5	NME/NM23 family member 5	HGNC:7853	5q31.2
NOP2	NOP2 nucleolar protein	HGNC:7867	12p13
NOSTRIN	nitric oxide synthase trafficker	HGNC:20203	2q31.1
NOVA1	neuro-oncological ventral antigen 1	HGNC:7886	14q12
NRIP1	nuclear receptor interacting protein 1	HGNC:8001	21q11.2
NUP205	nucleoporin 205kDa	HGNC:18658	7q31.32
NUP93	nucleoporin 93kDa	HGNC:28958	16q13
NUSAP1	nucleolar and spindle associated protein 1	HGNC:18538	15q14
OGN	osteoglycin	HGNC:8126	9q22
PDCD4	programmed cell death 4 (neoplastic transformation inhibitor)	HGNC:8763	10q24
PFKP	phosphofructokinase, platelet	HGNC:8878	10p15.3-p15.2
PHYHD1	phytanoyl-CoA dioxygenase domain containing 1	HGNC:23396	9q34.13
PIP	prolactin-induced protein	HGNC:8993	7q32-qter
PLAT	plasminogen activator, tissue	HGNC:9051	8p11.21
PLCH1	phospholipase C, eta 1	HGNC:29185	3q25
PNP	purine nucleoside phosphorylase	HGNC:7892	14q11.2
PNPLA7	patatin-like phospholipase domain containing 7	HGNC:24768	9q34.3
PRC1	protein regulator of cytokinesis 1	HGNC:9341	15q26.1
PSMB2	proteasome (prosome, macropain) subunit, beta type, 2	HGNC:9539	1p34.2
PTGER3	prostaglandin E receptor 3 (subtype EP3)	HGNC:9595	1p31.2
PTPRT	protein tyrosine phosphatase, receptor type, T	HGNC:9682	20q12-q13

PTTG1	pituitary tumor-transforming 1	HGNC 9690	5q35.1
QDPR	quinoid dihydropteridine reductase	HGNC:9752	4p15.31
RAB27B	RAB27B, member RAS oncogene family	HGNC 9767	18q21.2
RABEP1	rabaptin, RAB GTPase binding effector protein 1	HGNC:17677	17p13.2
RAD51AP1	RAD51 associated protein 1	HGNC 16956	12p13.2-p13.1
RBM38	RNA binding motif protein 38	HGNC:15818	20q13.31
RERG	RAS-like, estrogen-regulated, growth inhibitor	HGNC 15980	12p13.1
RFC4	replication factor C (activator 1) 4, 37kDa	HGNC:9972	3q27
RIPK2	receptor-interacting serine-threonine kinase 2	HGNC 10020	8q21
RNASE4	ribonuclease, RNase A family, 4	HGNC:10047	14q11
RPP40	ribonuclease P/MRP 40kDa subunit	HGNC 20992	6p25.1
RPS23	ribosomal protein S23	HGNC:10410	5q14.2
S100A8	S100 calcium binding protein A8	HGNC 10498	1q12-q22
SCUBE2	signal peptide, CUB domain, EGF-like 2	HGNC:30425	11p15.3
SH3BGRL	SH3 domain binding glutamic acid-rich protein like	HGNC 10823	Xq13.3
SKP1	S-phase kinase-associated protein 1	HGNC:10899	5q31
SKP2	S-phase kinase-associated protein 2, E3 ubiquitin protein ligase	HGNC 10901	5p13
SLC16A10	solute carrier family 16 (aromatic amino acid transporter), member 10	HGNC:17027	6q21-q22
SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1	HGNC 11005	1p34.2
SLC39A6	solute carrier family 39 (zinc transporter), member 6	HGNC:18607	18q12.2
SLC40A1	solute carrier family 40 (iron-regulated transporter), member 1	HGNC 10909	2q32
SLC7A5	solute carrier family 7 (amino acid transporter light chain, L system), member 5	HGNC:11063	16q24.3
SOD2	superoxide dismutase 2, mitochondrial	HGNC 11180	6q25
SOX11	SRY (sex determining region Y)-box 11	HGNC:11191	2p25
SRD5A1	steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)	HGNC 11284	5p15.31
SRPK1	SRSF protein kinase 1	HGNC:11305	6p21.31
STC2	stanniocalcin 2	HGNC 11374	5q35.2
STIL	SCL/TAL1 interrupting locus	HGNC:10879	1p32
STK32B	serine/threonine kinase 32B	HGNC 14217	4p16
SYTL4	synaptotagmin-like 4	HGNC:15588	Xq21.33
TAT	tyrosine aminotransferase	HGNC 11573	16q22.1
TBC1D9	TBC1 domain family, member 9 (with GRAM domain)	HGNC:21710	4q31.1
TEAD4	TEA domain family member 4	HGNC 11717	12p13.3-p13.2
TFF1	trefoil factor 1	HGNC:11755	21q22.3
TFF3	trefoil factor 3 (intestinal)	HGNC 11757	21q22.3
TMEM26	transmembrane protein 26	HGNC:28550	10q21.3
TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)	HGNC 1249	20q11.2
TRIP13	thyroid hormone receptor interactor 13	HGNC:12307	5p15
TROAP	trophinin associated protein	HGNC 12327	12q13.12

TTK	TTK protein kinase	HGNC:12401	6q13-q21
TUBA4A	tubulin, alpha 4a	HGNC:12407	2q36.1
UBE2C	ubiquitin-conjugating enzyme E2C	HGNC:15937	20q13.12
USB1	U6 snRNA biogenesis 1	HGNC:25792	16q13
VGLL1	vestigial like 1 (Drosophila)	HGNC:20985	Xq26.3
XBP1	X-box binding protein 1	HGNC:12801	22q12.1
YEATS2	YEATS domain containing 2	HGNC:25489	3q27.3

Table 5: Degregulated genes from Ingenuity Pathway Analysis and correlation with aggressiveness score

			Pearson correlation coefficient with respective metagene				High vs. Low score
	Symbol	UniqueID	Correlation coefficient	Parametric p-value	FDR	Permutation p-value	Fold-change
ER Metagene	*MAPT*	*ILMN_2310814*	*0.755*	*< 1e-07*	*< 1e-07*	*< 1e-07*	*-4.55*
	MYB	*ILMN_1711894*	*0.718*	*< 1e-07*	*< 1e-07*	*< 1e-07*	*-3.33*
	BCL2	ILMN_2246956	0.682	< 1e-07	< 1e-07	< 1e-07	-2.86
	STC2	ILMN_1691884	0.65	< 1e-07	< 1e-07	< 1e-07	-4.55
	BTG2	ILMN_1770085	0.544	< 1e-07	< 1e-07	< 1e-07	-2.38
	CLIC6	ILMN_1699665	0.403	< 1e-07	< 1e-07	< 1e-07	-3.45
	ESR1	ILMN_1678535	0.842	< 1e-07	< 1e-07	< 1e-07	-12.99
	FOXA1	ILMN_1766650	0.78	< 1e-07	< 1e-07	< 1e-07	-5.56
	XBP1	ILMN_1809433	0.741	< 1e-07	< 1e-07	< 1e-07	-2.56
	TFF3	ILMN_1811387	0.73	< 1e-07	< 1e-07	< 1e-07	-10.75
	DACH1	ILMN_1755741	0.684	< 1e-07	< 1e-07	< 1e-07	-3.85
	TFF1	ILMN_1722489	0.645	< 1e-07	< 1e-07	< 1e-07	-10.00
	PTPRT	ILMN_1698885	0.637	< 1e-07	< 1e-07	< 1e-07	-2.63
	PLAT	ILMN_1738742	0.557	< 1e-07	< 1e-07	< 1e-07	-2.38
	GSTM3	ILMN_1736184	0.5	< 1e-07	< 1e-07	< 1e-07	-2.00
	RPS23	ILMN_1772459	0.467	< 1e-07	< 1e-07	< 1e-07	-2.00
	GSTM1	ILMN_2391861	0.384	< 1e-07	< 1e-07	< 1e-07	-2.00
	ITPR1	ILMN_1789505	0.546	< 1e-07	< 1e-07	< 1e-07	-1.85
	SKP1	ILMN_1711766	0.689	< 1e-07	< 1e-07	< 1e-07	-1.75
	IGFBP2	ILMN_1725193	0.38	< 1e-07	< 1e-07	< 1e-07	-1.72
	GLI3	ILMN_1771962	0.486	< 1e-07	< 1e-07	< 1e-07	-1.61
	AHNAK	ILMN_1714567	0.551	< 1e-07	< 1e-07	< 1e-07	-1.59
	NRIP1	ILMN_1718629	0.548	< 1e-07	< 1e-07	< 1e-07	-1.59
	TAT	ILMN_1791678	0.304	< 1e-07	< 1e-07	< 1e-07	-1.56
	PDCD4	ILMN_1768004	0.44	< 1e-07	< 1e-07	< 1e-07	-1.54
CIN Metagene	*MELK*	*ILMN_1731184*	*0.9*	*< 1e-07*	*< 1e-07*	*< 1e-07*	*2.29*
	MCM10	*ILMN_2413898*	*0.868*	*< 1e-07*	*< 1e-07*	*< 1e-07*	*2.22*
	CENPA	*ILMN_1801257*	*0.909*	*< 1e-07*	*< 1e-07*	*< 1e-07*	*2.2*
	EXO1	*ILMN_1673721*	*0.885*	*< 1e-07*	*< 1e-07*	*< 1e-07*	*2.15*
	TTK	*ILMN_1788166*	*0.837*	*< 1e-07*	*< 1e-07*	*< 1e-07*	*2.15*
	KIF2C	*ILMN_1685916*	*0.888*	*< 1e-07*	*< 1e-07*	*< 1e-07*	*2.13*
	CENPN	ILMN_1720526	0.814	< 1e-07	< 1e-07	< 1e-07	2.04
	CEP55	ILMN_1747016	0.891	< 1e-07	< 1e-07	< 1e-07	2.03
	FOXM1	ILMN_2344971	0.869	< 1e-07	< 1e-07	< 1e-07	2.01
	TPX2	ILMN_1796949	0.88	< 1e-07	< 1e-07	< 1e-07	2.01
	AURKB	ILMN_1684217	0.884	< 1e-07	< 1e-07	< 1e-07	2.25
	AURKA	ILMN_1680955	0.854	< 1e-07	< 1e-07	< 1e-07	2.09

	CCNE1	ILMN_2374425	0.813	< 1e-07	< 1e-07	< 1e-07		2.64
	CDC45	ILMN_1670238	0.861	< 1e-07	< 1e-07	< 1e-07		2.22
	PTTG1	ILMN_2042771	0.846	< 1e-07	< 1e-07	< 1e-07		2.15
	BIRC5	ILMN_2349459	0.822	< 1e-07	< 1e-07	< 1e-07		2.28
	CCNB2	ILMN_1801939	0.89	< 1e-07	< 1e-07	< 1e-07		2.39
	CDCA5	ILMN_1683450	0.89	< 1e-07	< 1e-07	< 1e-07		2.49
	CDC20	ILMN_1663390	0.873	< 1e-07	< 1e-07	< 1e-07		3.22
	UBE2C	ILMN_2301083	0.861	< 1e-07	< 1e-07	< 1e-07		2.75
	SLC7A5	ILMN_1720373	0.749	< 1e-07	< 1e-07	< 1e-07		3.75
	BUB1	ILMN_2202948	0.896	< 1e-07	< 1e-07	< 1e-07		1.97
	CDCA8	ILMN_1709294	0.87	< 1e-07	< 1e-07	< 1e-07		1.96
	PRC1	ILMN_1728934	0.817	< 1e-07	< 1e-07	< 1e-07		1.96
	CCNA2	ILMN_1786125	0.888	< 1e-07	< 1e-07	< 1e-07		1.95
	CDCA3	ILMN_1737728	0.859	< 1e-07	< 1e-07	< 1e-07		1.95
	HJURP	ILMN_1703906	0.861	< 1e-07	< 1e-07	< 1e-07		1.9
	KIF20A	ILMN_1695658	0.844	< 1e-07	< 1e-07	< 1e-07		1.9
	NCAPG	ILMN_1751444	0.848	< 1e-07	< 1e-07	< 1e-07		1.89
	CHEK1	ILMN_1664630	0.853	< 1e-07	< 1e-07	< 1e-07		1.85
	SKP2	ILMN_1665538	0.709	< 1e-07	< 1e-07	< 1e-07		1.83
	MCM4	ILMN_1737205	0.75	< 1e-07	< 1e-07	< 1e-07		1.79
	CDK1	ILMN_1747911	0.776	< 1e-07	< 1e-07	< 1e-07		1.75
	LMNB2	ILMN_1708101	0.799	< 1e-07	< 1e-07	< 1e-07		1.74
	RAD51AP1	ILMN_1670353	0.818	< 1e-07	< 1e-07	< 1e-07		1.74
	TROAP	ILMN_1700337	0.784	< 1e-07	< 1e-07	< 1e-07		1.73
	DLGAP5	ILMN_1749829	0.86	< 1e-07	< 1e-07	< 1e-07		1.7
	MAD2L1	ILMN_1777564	0.709	< 1e-07	< 1e-07	< 1e-07		1.7
	GTSE1	ILMN_1771039	0.858	< 1e-07	< 1e-07	< 1e-07		1.69
	CENPE	ILMN_1716279	0.858	< 1e-07	< 1e-07	< 1e-07		1.67
	MCM2	ILMN_1681503	0.764	< 1e-07	< 1e-07	< 1e-07		1.66
	MCM6	ILMN_1798654	0.792	< 1e-07	< 1e-07	< 1e-07		1.6
	BLM	ILMN_1709484	0.792	< 1e-07	< 1e-07	< 1e-07		1.57
	KIF14	ILMN_1808071	0.808	< 1e-07	< 1e-07	< 1e-07		1.55
	KIF23	ILMN_1811472	0.824	< 1e-07	< 1e-07	< 1e-07		1.55
	BYSL	ILMN_1682792	0.741	< 1e-07	< 1e-07	< 1e-07		1.52
	CDC25A	ILMN_1711005	0.773	< 1e-07	< 1e-07	< 1e-07		1.52
	GTPBP4	ILMN_1742577	0.755	< 1e-07	< 1e-07	< 1e-07		1.67
	SRPK1	ILMN_1798804	0.763	< 1e-07	< 1e-07	< 1e-07		1.62
	CKS2	ILMN_1756326	0.717	< 1e-07	< 1e-07	< 1e-07		1.6
	NDC80	ILMN_1664511	0.765	< 1e-07	< 1e-07	< 1e-07		1.46
	RFC4	ILMN_1724489	0.759	< 1e-07	< 1e-07	< 1e-07		1.45
	MCM7	ILMN_1663195	0.704	< 1e-07	< 1e-07	< 1e-07		1.64
	S100A8	ILMN_1729801	0.517	< 1e-07	< 1e-07	< 1e-07		8.16
	KRT6A	ILMN_2219002	0.474	< 1e-07	< 1e-07	< 1e-07		3.27

)

	PFKP	ILMN_1805737	0.595	< 1e-07	< 1e-07	< 1e-07	2.87
	SOD2	ILMN_2336781	0.565	< 1e-07	< 1e-07	< 1e-07	2.69
	IL8	ILMN_2184373	0.357	< 1e-07	< 1e-07	< 1e-07	2.27
	CDC25B	ILMN_2338323	0.594	< 1e-07	< 1e-07	< 1e-07	2.01
	CKS1B	ILMN_2041046	0.691	< 1e-07	< 1e-07	< 1e-07	1.75
	GAPDH	ILMN_1802252	0.685	< 1e-07	< 1e-07	< 1e-07	1.61
	SLC16A10	ILMN_1782938	0.45	< 1e-07	< 1e-07	< 1e-07	1.45
	VGLL1	ILMN_1719753	0.524	< 1e-07	< 1e-07	< 1e-07	3.16
	CDKN2A	ILMN_1717714	0.591	< 1e-07	< 1e-07	< 1e-07	1.91
	RIPK2	ILMN_1758939	0.674	< 1e-07	< 1e-07	< 1e-07	1.7
	TUBA4A	ILMN_1784300	0.399	< 1e-07	< 1e-07	< 1e-07	1.66
	PSMB2	ILMN_1764794	0.633	< 1e-07	< 1e-07	< 1e-07	1.52
	NOP2	ILMN_1723158	0.617	< 1e-07	< 1e-07	< 1e-07	1.52
	IRAK1	ILMN_2379130	0.625	< 1e-07	< 1e-07	< 1e-07	1.51
	DKC1	ILMN_1671257	0.692	< 1e-07	< 1e-07	< 1e-07	1.5
	DAPK1	ILMN_1708340	0.432	< 1e-07	< 1e-07	< 1e-07	1.43
	TEAD4	ILMN_1705301	0.554	< 1e-07	< 1e-07	< 1e-07	1.42

Cox proportional hazards model, Wald Statistic
p-value cutoff univariate test: 0.001
10000 random permutations,

Cell survival after siRNA depletion in vitro

Parametric p-value	FDR	Permutation p-value	Hazard Ratio	SD of log intensities	MDA-MB-231	Sum159PT	Hs578T
5.00E-07	*1.47E-06*	*< 1e-07*	*-1.112347052*	*1.567*			
0.0008816	*1.28E-03*	*8.00E-04*	*-1.091703057*	*1.258*			
< 1e-07	< 1e-07	< 1e-07	-1.209189843	1.043			
2.86E-05	5.55E-05	< 1e-07	-1.074113856	1.938			
2.45E-05	4.95E-05	< 1e-07	-1.154734411	1.03			
< 1e-07	< 1e-07	< 1e-07	-1.111111111	2.214			

< 1e-07	< 1e-07	< 1e-07	-1.207729469	1.076			
6.00E-07	1.62E-06	< 1e-07	-1.133786848	1.358			
2.43E-05	4.95E-05	< 1e-07	-1.204819277	0.736			
5.66E-05	9.98E-05	< 1e-07	-1.096491228	1.625			
5.50E-06	1.21E-05	< 1e-07	-1.242236025	0.723			
< 1e-07	*< 1e-07*	*< 1e-07*	*1.274*	*0.799*	91 ± 3% (n=3)	46 ± 4 (n=5)	92 ± 3 (n=3)
7.00E-07	*1.84E-06*	*< 1e-07*	*1.221*	*0.791*	98 ± 6% (n=3)	68 ± 8 (n=6)	83 ± 2 (n=3)
1.00E-07	*3.46E-07*	*< 1e-07*	*1.253*	*0.762*	155 ± 17% (n=3)	100 ± 1 (n=3)	52 ± 2 (n=3)
< 1e-07	*< 1e-07*	*< 1e-07*	*1.267*	*0.767*			
4.00E-07	*1.25E-06*	*< 1e-07*	*1.222*	*0.799*	51 ± 1% (n=3)	34 ± 6 (n=6)	19 ± 1 (n=3)
< 1e-07	*< 1e-07*	*< 1e-07*	*1.271*	*0.756*			
5.00E-06	1.13E-05	< 1e-07	1.204	0.779			
1.00E-07	3.46E-07	< 1e-07	1.26	0.75	87 ± 4% (n=3)	37 ± 1 (n=6)	37 ± 1 (n=3)
1.00E-07	3.46E-07	< 1e-07	1.247	0.744	66 ± 1% (n=3)	72 ± 2 (n=5)	62 ± 2 (n=3)
< 1e-07	< 1e-07	< 1e-07	1.326	0.782	57 ± 7% (n=3)	64 ± 1 (n=3)	24 ± 3 (n=3)
< 1e-07	< 1e-07	< 1e-07	1.256	0.836			
< 1e-07	< 1e-07	< 1e-07	1.273	0.873			
3.30E-06	7.81E-06	< 1e-07	1.182	0.878			
< 1e-07	< 1e-07	< 1e-07	1.246	0.882			
< 1e-07	< 1e-07	< 1e-07	1.259	0.958			
5.00E-07	1.47E-06	< 1e-07	1.184	0.965	100 ± 26% (n=3)	79 ± 3 (n=6)	100 ± 1 (n=3)
< 1e-07	< 1e-07	< 1e-07	1.232	0.966			
< 1e-07	< 1e-07	< 1e-07	1.271	0.998			
< 1e-07	< 1e-07	< 1e-07	1.184	1.194	98 ± 5% (n=3)	101 ± 1 (n=6)	110 ± 1 (n=3)
< 1e-07	< 1e-07	< 1e-07	1.208	1.239			
3.29E-05	6.06E-05	< 1e-07	1.109	1.299			
< 1e-07	< 1e-07	< 1e-07	1.279	0.726	91 ± 4%	76 ± 2	83 ± 3

					(n=3)	(n=6)	(n=3)
2.00E-07	6.69E-07	< 1e-07	1.267	0.692			
< 1e-07	< 1e-07	< 1e-07	1.269	0.893			
< 1e-07	< 1e-07	< 1e-07	1.3	0.709			
4.00E-07	1.25E-06	< 1e-07	1.232	0.765			
< 1e-07	< 1e-07	< 1e-07	1.301	0.693			
< 1e-07	< 1e-07	< 1e-07	1.322	0.806			
< 1e-07	< 1e-07	< 1e-07	1.271	0.784			
0.0001213	1.96E-04	2.00E-04	1.23	0.603			
0.0008164	1.20E-03	6.00E-04	1.191	0.612	76 ± 3% (n=3)	45 ± 6 (n=6)	63 ± 3 (n=3)
6.00E-07	1.62E-06	< 1e-07	1.24	0.764			
9.00E-07	2.18E-06	< 1e-07	1.214	0.84			
1.08E-05	2.28E-05	< 1e-07	1.284	0.583			
6.12E-05	1.06E-04	< 1e-07	1.194	0.728	83 ± 4% (n=3)	65 ± 4 (n=6)	91 ± 2 (n=3)
< 1e-07	< 1e-07	< 1e-07	1.378	0.717	86 ± 7% (n=3)	55 ± 7 (n=6)	68 ± 10 (n=3)
8.00E-07	2.04E-06	< 1e-07	1.311	0.595	106 ± 3% (n=3)	62 ± 5 (n=6)	54 ± 1 (n=3)
4.80E-06	1.11E-05	< 1e-07	1.188	0.884			
< 1e-07	< 1e-07	< 1e-07	1.464	0.52			
< 1e-07	< 1e-07	< 1e-07	1.374	0.601			
9.00E-07	2.18E-06	< 1e-07	1.264	0.704			
0.0007066	1.07E-03	3.00E-04	1.205	0.587			
9.41E-05	1.57E-04	1.00E-04	1.282	0.491			
6.00E-07	1.62E-06	< 1e-07	1.36	0.51			
< 1e-07	< 1e-07	< 1e-07	1.381	0.577			
8.50E-06	1.83E-05	< 1e-07	1.363	0.466			
7.55E-05	1.28E-04	1.00E-04	1.297	0.472			
					56 ± 4% (n=3)	65 ± 1 (n=3)	28 ± 2 (n=3)
					119 ± 6% (n=3)	90 ± 1 (n=3)	39 ± 5 (n=3)
2.59E-05	5.13E-05	< 1e-07	1.063	2.187			
5.27E-05	9.47E-05	< 1e-07	1.098	1.328			
0.000120	1.96E-	< 1e-07	1.133	1.029			

5	04						
0.0002994	4.76E-04	1.00E-04	1.123	0.999			
0.0003927	6.05E-04	2.00E-04	1.101	1.239			
3.25E-05	6.06E-05	< 1e-07	1.174	0.818			
0.0007626	1.14E-03	5.00E-04	1.153	0.768	105 ± 7% (n=3)	94 ± 1 (n=3)	84 ± 4 (n=3)
0.0003141	4.91E-04	1.00E-04	1.196	0.66			
3.31E-05	6.06E-05	< 1e-07	1.278	0.548			

CENPF	88 ± 4% (n=3)	86 ± 3 (n=6)	70 ± 3 (n=3)
MYBL1	73 ± 5% (n=3)	83 ± 1 (n=3)	47 ± 4 (n=3)
GPSM2	87 ± 4% (n=3)	80 ± 4 (n=5)	60 ± 6 (n=3)
ANP32E	95 ± 2% (n=3)	79 ± 8 (n=6)	94 ± 4 (n=3)
TOP2A	122 ± 8% (n=3)	73 ± 1 (n=3)	69 ± 5 (n=3)
TYMS	75 ± 8% (n=3)	71 ± 2 (n=5)	68 ± 4 (n=3)
PBK	59 ± 2% (n=3)	55 ± 7 (n=6)	46 ± 1 (n=3)
ASPM	94 ± 6% (n=3)	37 ± 1 (n=6)	65 ± 2 (n=3)

Table 6: Association of the 6 overexpressed genes in the 8-genes score with RFS and DMFS at 5 and 10 years

Patient subgroup and cut-off		RFS			
		5 years		10 years	
		HR (95 % CI)	p-value	HR (95 % CI)	p-value
All	Median	1.32 (1.17 - 1.83)	<1.00E-16	2.16 (1.90 - 2.47)	<1.00E-16
	Quartile	1.93 (2.43 - 3.78)	<1.00E-16	2.46 (2.09 - 2.89)	<1.00E-16
	Tertile	2.83 (2.35 - 3.41)	<1.00E-16	2.70 (2.24 - 3.27)	<1.00E-16
ER+	Median	2.67 (2.23 - 3.17)	<1.00E-16	2.28 (1.89 - 2.65)	<1.00E-16
	Quartile	2.96 (2.39 - 3.67)	<1.00E-16	2.64 (2.16 - 3.23)	<1.00E-16
	Tertile	2.87 (2.34 - 3.53)	<1.00E-16	2.51 (2.11 - 2.99)	<1.00E-16
LN-	Median	2.76 (2.19 - 3.48)	<1.00E-16	2.25 (1.84 - 2.74)	2.20E-16
	Quartile	2.76 (2.08 - 3.80)	1.10E-10	2.51 (1.91 - 3.29)	5.10E-12
	Tertile	2.92 (2.21 - 3.87)	4.30E-15	2.53 (2.00 - 3.19)	1.00E-15
LN+	Median	2.20 (1.57 - 3.08)	2.40E-06	1.88 (1.40 - 2.53)	1.90E-05
	Quartile	3.19 (1.87 - 5.44)	6.80E-06	2.56 (1.67 - 3.94)	8.10E-06
	Tertile	2.43 (1.61 - 3.37)	1.70E-06	2.05 (1.48 - 2.81)	4.00E-06

Patient subgroup and cut-off		DMFS			
		5 years		10 years	
		HR (95 % CI)	p-value	HR (95 % CI)	p-value
All	Median	2.87 (2.17 - 3.79)	8.80E-15	2.37 (1.87 - 3.01)	1.90E-13
	Quartile	3.64 (2.41 - 5.52)	5.80E-11	3.43 (2.41 - 4.88)	3.20E-12
	Tertile	3.53 (2.48 - 5.04)	8.70E-14	2.93 (2.18 - 3.93)	3.70E-14
ER+	Median	3.43 (2.49 - 4.74)	1.30E-13	2.63 (2.00 - 3.45)	4.20E-13
	Quartile	3.41 (2.21 - 5.27)	8.00E-08	3.27 (2.28 - 4.71)	2.20E-11
	Tertile	3.87 (2.62 - 5.74)	8.10E-13	3.07 (2.24 - 4.20)	2.30E-13
LN-	Median	4.84 (2.53 - 9.26)	1.40E-07	2.80 (2.00 - 3.93)	4.30E-10
	Quartile	4.86 (2.82 - 8.37)	2.70E-10	4.46 (2.61 - 7.60)	1.80E-09
	Tertile	3.88 (2.61 - 6.07)	4.30E-12	3.78 (2.46 - 5.74)	4.50E-11
LN+	Median	2.13 (1.19 - 3.80)	9.40E-03	2.16 (1.28 - 3.62)	3.00E-03
	Quartile	3.97 (1.18 - 7.44)	1.50E-02	3.07 (1.31 - 6.29)	6.80E-03
	Tertile	3.51 (1.50 - 8.21)	2.00E-03	2.88 (1.47 - 5.68)	1.40E-03

p-values are from Log-rank test from GE-plots

Table 7: Association of the 2 overexpressed genes in the 8-genes score with RFS and DMFS at 5 and 10 years

Patient subgroup and cut-off		RFS			
		5 years		10 years	
		HR (95 % CI)	p-value	HR (95 % CI)	p-value
All	Median	0.53 (0.46 – 0.61)	<1.00E-16	0.59 (0.52 – 0.67)	5.60E-16
	Quartile	0.53 (0.35 – 0.81)	<1.00E-16	0.59 (0.52 – 0.68)	8.20E-14
	Tertile	0.49 (0.43 – 0.57)	<1.00E-16	0.57 (0.50 – 0.65)	<1.00E-16
ER+	Median	0.60 (0.51 – 0.71)	2.10E-09	0.65 (0.56 – 0.76)	2.60E-08
	Quartile	0.62 (0.48 – 0.79)	1.30E-04	0.69 (0.54 – 0.86)	1.20E-03
	Tertile	0.54 (0.45 – 0.66)	6.00E-10	0.63 (0.52 – 0.76)	5.00E-07
LN-	Median	0.61 (0.49 – 0.76)	8.00E-06	0.71 (0.58 – 0.86)	4.20E-04
	Quartile	0.56 (0.44 – 0.72)	3.50E-06	0.63 (0.50 – 0.79)	6.00E-05
	Tertile	0.53 (0.42 – 0.66)	2.20E-08	0.62 (0.50 – 0.76)	4.60E-06
LN+	Median	0.58 (0.42 – 0.80)	9.70E-04	0.70 (0.52 – 0.93)	1.30E-02
	Quartile	0.59 (0.42 – 0.84)	3.00E-03	0.70 (0.50 – 0.98)	3.50E-02
	Tertile	0.57 (0.41 – 0.78)	4.00E-04	0.68 (0.50 – 0.91)	9.70E-03

Patient subgroup and cut-off		DMFS			
		5 years		10 years	
		HR (95 % CI)	p-value	HR (95 % CI)	p-value
All	Median	0.59 (0.47 – 0.74)	2.50E-06	0.59 (0.47 – 0.74)	2.50E-06
	Quartile	0.58 (0.46 – 0.74)	4.40E-06	0.58 (0.46 – 0.74)	4.40E-06
	Tertile	0.57 (0.46 – 0.71)	4.90E-07	0.57 (0.46 – 0.71)	4.90E-07
ER+	Median	0.62 (0.48 – 0.81)	3.00E-04	0.62 (0.48 – 0.81)	3.00E-04
	Quartile	0.56 (0.39 – 0.82)	2.40E-03	0.56 (0.39 – 0.82)	2.40E-03
	Tertile	0.57 (0.42 – 0.78)	3.00E-04	0.57 (0.42 – 0.78)	3.00E-04
LN-	Median	0.74 (0.54 – 1.00)	4.80E-02	0.74 (0.54 – 1.00)	4.80E-02
	Quartile	0.64 (0.46 – 0.89)	7.80E-03	0.64 (0.46 – 0.89)	7.80E-03
	Tertile	0.60 (0.44 – 0.81)	1.00E-03	0.60 (0.44 – 0.81)	1.00E-03
LN+	Median	0.58 (0.35 – 0.96)	3.20E-02	0.58 (0.35 – 0.96)	3.20E-02
	Quartile	0.49 (0.29 – 0.83)	6.90E-03	0.49 (0.29 – 0.83)	6.90E-03
	Tertile	0.56 (0.34 – 0.92)	2.10E-02	0.56 (0.34 – 0.92)	2.10E-02

p-values are from Log-rank test from KM-plotter

Table 8: details of antibodies and immunohistochemistry conditions used for breast cancer TMA analysis in this study

Antibody	Clone	Species	Source	Dilution	Antigen Retrieval*	Cellular Localization	Cut-off used for classification as 'positive'
ER	6F11	Mouse	Novocastra	1:100	Citrate	Nucleus	> 1%
PR	1A6	Mouse	Novocastra	1:200	Citrate	Nucleus	> 1%
HER2	CB11	Rabbit	Dako	1:200	Citrate	Cell Membrane	3+ (>30%)
CK5/6	D5/16 B4	Mouse	Chemicon	1:400	Citrate	Membrane + Cytoplasm	Any positivity
CK14	LL002	Mouse	Novocastra	1:40	Citrate	Membrane + Cytoplasm	Any positivity
EGFR	31G7	Mouse	Invitrogen	1:100	EDTA	Cell Membrane	Any positivity
Ki-67	MIB-1	Mouse	Dako	1:200	Citrate	Nucleus	Any positivity (20% cells stained classed as 'Ki67-high')
TTK	N1	Mouse	Abcam	1:100	EDTA	Cytoplasm	0 Negative 1 weak and focal staining 2 moderate-strong focal staining (collectively <50% tumor cells) 3 moderate-strong diffuse staining (>50% tumor cells) *Regarding estimating % of cells stained, we disregarded mitotic cells to assess mitosis-independent expression of TTK*

*Antigen retrieval in 0.01M citric acid buffer (pH 6.0) at 125 °C for 5 min in a pressure cooker, or in 0.001 M Tris/EDTA; pH 8.8, at 105°C for 15 min in a pressure cooker.

Table 9: Multivariate analyses

Covariants	P value	Hazard Ratio	Covariants	P value	Hazard Ratio	Covariants	P value	Hazard Ratio
Grade	0.7045	1.04 (0.86 - 1.25)	Stage	0	1.46 (1.26 - 1.68)	AJCC stage T	0.0002	1.35 (1.16 - 1.58)
10CIN 2ER signature	0.0002	1.59 (1.25 - 2.04)	10CIN 2ER signature	0	2.2 (1.73 - 2.79)	AJCC stage N	0	1.73 (1.5 - 1.99)
						10CIN 2ER signature	0.0075	1.35 (1.08 - 1.69)

EXAMPLE 2

Materials and Methods

Meta-analysis of global gene expression in TNBC

We performed a meta-analysis of global gene expression data in the Oncomine™ database [37] (Compendia Bioscience, Ann Arbor, MI) using a primary filter for breast cancer (130 datasets), sample filter to use clinical specimens and dataset filters to use mRNA datasets with more 151 patients (22 datasets). Two additional filters were applied to perform two independent differential analyses. The first differential was metastatic event analysis at 5 years (metastatic events vs. no metastatic events, 7 datasets [51, 56-61]) and the second differential analysis was survival at 5 years (patients who died vs. patients who survived, 7 datasets [39, 57, 59, 61-64]). Deregulated genes were selected based on the median p-value of the median gene rank in overexpression or underexpression patterns across the datasets for each of the two differential analyses.

Deriving the 28-signature (the TN signature)

The online tool KM-Plotter [38] which collates gene expression data from Affymterix platform for more than 4000 breast cancer patients were used for developing the 28-gene signature. From the deregulated genes in primary tumors which led to metastatic or death events within 5 years discovered in the meta-analysis in Oncomine™, 166 genes were common in both survival events. These genes were then interrogated one by one in KM-Plotter restricting the univariate survival analysis to ER^- or BLBC subtypes. Genes which significantly associated with relapse-free survival (RFS), distant metastasis-free survival (DMFS) or overall survival (OS) in either ER^- or BLBC subtypes were short selected. The 96 genes that were significant in this filtering where then sorted for their level of significance as well as the prevalence of significance across the different survival outcomes (RFS, DMFS and OS) and across ER^- and BLBC subtypes. Based on this sorting, six groups of gene lists were obtained with different levels of survival association (Table 14). Each of these groups were then used as a metagene and the average expression of genes in each group was investigated for association with survival in KM-Plotter in ER^- and BLBC subtypes. Based on these analysis, four groups were selected and two were excluded. Furthermore, for two groups, the top 4 and 3 genes were found to be more prognostic than the rest of the group and these were selected. In total, the 7

genes (which their downregulation associates with poor survival) from these two groups and 21 genes (which their upregulation associates with poor survival) in the other two groups were selected to test for association with survival in KM-Plotter. These 28 genes showed the highest association with survival as a gene signature compared to any single gene in the original list or any groups from this list. These 28 genes were selected as the triple negative (TN) signature and was subjected to validation as described below.

Validation of the TN signature in breast cancer cohorts

Three large breast cancer gene expression datasets were used for validation. The Research Online Cancer Knowledgebase (ROCK) dataset [40] (GSE47561; n=1570 patients) and the homogenous TNBC dataset [32] (GSE31519; n=579 TNBC patients) were obtained from Gene Expression Omnibus (GEO) and the data was imported into BRB-ArrayTools [65] (V4.2, Biometric Research Branch, NCI, Maryland, USA) with built in R Bioconductor packages. The Cancer Genome Atlas (TCGA) dataset [39]; using the Illumina HiSeq RNA-Seq arrays (n=1106 patients) or the Agilent custom arrays (Agilent G4502A-07-3) on 597 patients of the 1106 total patients, were obtained from the UCSC Genome Browser [66, 67]. The TN signature was investigated in each of these datasets where a score was devised to quantify the signature; the TN score = average expression of the 21 genes whose overexpression associated with poor survival ÷ average expression of the 7 genes whose underexpression associated with poor survival. The TN score for each tumor in each dataset was calculated and tumors were assigned as high or low TN score tumors by dichotomy across the median TN score in each dataset. In some cases, tertiles of the TN score in each dataset were used to classify tumors as high, intermediate or low TN score tumors and in other cases the quartiles of the TN score were used to classify tumors in the 1^{st}, 2^{nd}, 3^{rd} or 4^{th} quartiles. The survival of patients in high (over the median, last tertile of the 4th quartile) vs. low TN score groups was compared. Survival analyses were constructed using GraphPad® Prism v6.0 (GraphPad Software, CA, USA) and the Log-rank (Mantel-Cox) Test was used for statistical comparisons of survival curves.

Association of the TN score and signatures with pathological complete responses (pCR) after neoadjuvant chemotherapy and response to endocrine therapy

Datasets which performed gene expression profiling prior to neoadjuvant chemotherapy or endocrine therapy alone were obtained from GEO. The datasets

used in this study for neoadjuvant chemotherapy and recorded pathological complete response (pCR) include: GSE18728 [42], GSE50948 [43], GSE20271 [44], GSE20194 [45], GSE22226 [41, 46], GSE42822 [47] and GSE23988 [48]. For datasets which performed gene expression profiling prior to endocrine therapy (tamoxifen) and recorded patient survival include: GSE6532 [25] and GSE17705 [51]. These datasets using the Affymetrix gene expression array platforms were imported into BRB-ArrayTools and normalized as described previously [68]. Each tumor in the datasets were assigned as high or low score for our signatures as described in the previous sections. The rate of pCR after chemotherapy or the survival of patients after endocrine therapy were compared between high score tumors and low score tumors using GraphPad® Prism.

Global gene expression profiles comparison by class comparison

Global gene expression comparison was carried out to compare tumors with high TN or iBCR scores to those with low TN or iBCR scores to characterize additional differences between these tumors and identify deregulated genes which could be suitable as for drug targeting. These comparisons were carried out in the large cohort of 1570 patients in the ROCK dataset and BRB-ArrayTools was used to perform the Class Comparison test. The two classes were high vs. low score tumors and the parameters selected in this plugin in ArrayTools were as follows: Type of univariate test used = Two-sample T-test; Class variable = TN score (high or low) or iBCR score (high or low); fold-change cutoff = 1.5 fold; Permutation p-values for significant genes were computed based on 10000 random permutations and Nominal significance level of each univariate test: 0.05. The results from these analyses are shown in Tables 13 and 15-17.

Integration of the Agro and TN signatures in the integrated Breast Cancer Recurrence (iBCR) score

We previously published the Aggressiveness (Agro) signature and score also from meta-analysis and extensive validation and show that this signature is prognostic in ER+ breast cancer [36]. To test whether the Agro signatures could be integrated with the TN signature (prognostic in ER^- breast cancer) to produce an integrated test that is independent of ER status, several integration methods were investigated. The hypothesis behind the integration methods was to identify a direct relationship that can describe the relationship between the TN and Agro scores in both ER^- and ER^+ breast cancer subtypes that is also in direct relationship with the integrated score. In

other words, the integrated score would retain the information from each the Agro and TN scores relevant to their prognostic value in ER^+ and ER^- breast cancers, respectively. The ROCK dataset was used to test the different methods of integration and the performance of these methods in the stratification of survival of ER^+ and ER^- breast cancer. The addition or subtraction of the scores produced a direct relationship between the TN and Agro score and the produced integrated score (Figure 36). These two methods were then analyzed for prognostication of ER^+ and ER^- subtypes in the ROCK dataset and only the addition method retained prognostication in ER^- breast cancer (Figure 37). Similarly, multiplying and dividing the TN and Agro scores were tested and an exponential and power curve relationships described the relation between the two scores and with the integrated score (Figure 38). Again, these two methods were tested from prognostication in the ROCK dataset and only the multiplication method retained prognostication in ER^- breast cancer (Figure 37). Because the multiplication and division methods produced exponential and power curves for the relationship between the scores, integration by raising one score to the power of the other score appeared reasonable. Exponential and power curves are the result of power equations. Indeed, integration by raising the TN score to the power of the Agro score was highly prognostic in both ER^+ and ER^- breast cancers (Figures 37 and 38). This integrated score, the integrated Breast Cancer Recurrence (iBCR) score was in fact more prognostic in ER^+ and ER^- patients in the ROCK dataset than the single Agro and TN scores, respectively. The iBCR score was validated in the ROCK and homogenous TNBC datasets (Affymetrix platform), the TCGA dataset (Illumina RNA-Seq platform) and the ISPY-1 trial dataset (GSE22226 [41, 46], Agilent platform), illustrating the platform-independence of the iBCR score which is driven by the platform independence of the Agro and TN signatures as they were discovered from meta-analysis irrespective of array platforms used from independent studies.

Mining drug screen studies

Two large studies which treated large panels of cancer cell lines with large panels of anticancer drugs were investigated to determine whether cell lines with high Agro, TN or iBCR scores show different sensitivity to particular anticancer drugs in comparison to cancer cell lines with low Agro, TN or iBCR scores. Briefly, the datasets of gene expression profiling from Genentech (mRNA Cancer Cell Line Profiles GSE10843), Pfizer (Pfizer Molecular Profile Data for Cell Line GSE34211) and Broad Institute/Novartis (Cancer Cell Line Encyclopedia [CCLE] GSE3613)

were obtained from GEO and imported into ArrayTools as described earlier. The Agro, TN and iBCR scores for all the cell lines profiled were calculated and cell lines were assigned as high or low for each of the scores based on dichotomy across the median in each dataset. For cell lines which were profiled in more than one dataset, the average scores were used. Using this data, the sensitivity of cancer cell lines with high and low Agro, TN or iBCR scores was compared to those with low scores to anticancer drugs was investigated in two studies [49, 50]. Drugs which had significantly different IC50 in high score cell lines compared to low score cell lines are described herein. Statistical significance was determined from unpaired two-tailed t-test using GraphPad® Prism.

Other statistical analysis

Univariate and multivariate Cox proportional hazards regression analyses were performed using MedCalc for Windows, version 12.7 (MedCalc Software, Ostend, Belgium).

Results

Meta-analysis of gene expression profile in Oncomine™

We performed a meta-analysis of published gene expression data, irrespective of platform or breast cancer subtype, using the Oncomine™ database [37] (version 4.5). We were able to compared the expression profiles of primary breast tumors from 512 patients who developed metastases vs. 732 patients who did not develop metastases at 5 years (7 datasets in total) to identify 500 overexpressed genes and 500 underexpressed genes in the metastasis cases (cutoff median p-value across the datasets < 0.05 from a Student's *t*-test, Figure 31). We also compared the expression profiles of 232 primary breast tumors from patients who died within 5 years vs. 879 patients who survived in 7 datasets and found 500 overexpressed genes and 500 underexpressed genes in the poor survivors (cutoff median p-value across the datasets < 0.05 from a Student's *t*-test, Figure 31). Since several datasets were annotated for one of these outcomes but not both, we rationalized that the union of these analyses is more appropriate particularly that death is the most likely outcome in metastatic disease. The union of the over- and expressed genes in tumors that associated with metastasis and those that associated with death within 5 years revealed common 101 overexpressed and 65 underexpressed genes (Figure 19). These 166 deregulated genes were then subjected to training using the online tool

KM-plotter [38] to derive a 28 gene signature as described in methods below followed by validation of this signature, the TN signature, in several large cohorts of breast cancer gene expression datasets (Figure 19).

The TN signature is prognostic in TNBC, BLBC and ER- breast cancer subtypes

The 166 deregulated genes in primary breast tumors that associated with poor outcome discovered from the Oncomine™ meta-analysis were interrogated using KM-Plotter. The overexpression of 31 genes and the underexpression of 65 genes associated with RFS, DMFS or OS of BLBC or ER- breast cancer (Table 14). Based on the level of significance in univariate survival analysis and the prevalence of this significance across the different disease outcomes (RFS, DMFS and OS), a list of 21 overexpressed and 7 underexpressed genes (Table 1) were shortlisted as a signature with the strongest association with survival in both BLBC and ER- breast cancer subtypes (Figure 20).

The 28-gene signature, the TN signature, was then validated in multivariate survival analysis in two breast cancer cohorts, the homogenous TNBC dataset [32] and the Research Online Cancer Knowledgebase (ROCK) dataset [40]. We devised a score to quantify trends in the TN signature, the TN score, which is calculated as the ratio of the average expression of the 21 overexpressed genes to that of the 7 underexpressed genes. Dichotomy across the median TN score stratified the survival of TNBC (Figure 21A), BLBC (Figure 21B) and ER- (Figure 21C) patients and outperformed all standard clinicopathological indicators. These analyses indicated that the TN score is an independent prognostic factor that identified TNBC, BLBC or ER- patients with poor survival irrespective to tumor size and grade, patient age, lymph node status or treatment. The TN signature also outperformed all previously published signatures that are prognostic in ER-, TNBC or BLBC subtypes [30-35] (Figure 32).

While the discovery of the signature in Oncomine™ included datasets using the Affymterix, Illumina and Agilent platforms, the training and validation above was limited to the Affymterix platform. Thus, we validated the TN score in The Cancer Genome Atlas (TCGA) dataset [39] which used the Illumina HiSeq RNA-seq platform. As shown in Figure 22, the RFS of ER- patients in the TCGA dataset was stratified by TN score and this stratification outperformed that by standard clinicopathological indicators. The original TCGA publication used Agilent custom arrays (Agilent G4502A-07-3) on 597 patients and we analyzed the prognosis of the

TN score in this data. The TN score stratified the survival of ER$^-$ patients in the Agilent TCGA data (Figure 33). Altogether, the prognostic value of the TN signature/score was validated in large, independent cohorts of breast cancer in TNBC, BLBC and ER$^-$ breast cancer subtypes irrespective of the gene expression array platforms used.

The TN score and the likelihood of pCR after chemotherapy

Chemotherapy is a standard therapy for ER$^-$ breast cancer and the only mode of therapy for ER$^-$HER2$^-$ (TNBC) breast cancer. Although, pathological complete response (pCR) differs by receptor status, it remains highly predictive of survival within the different breast cancer subtypes [41]. Given the association of the TN score with outcome in TNBC, BLBC and ER$^-$ breast cancer, we questioned whether this score is also associated with pCR after chemotherapy. To this end, we analyzed publically available datasets of neoadjuvant chemotherapy trials which recorded pCR and performed pre-treatment gene expression profiling. As shown in Figure 23A, pCR after chemotherapy in ER$^-$/HER2$^-$ patients was less likely after TX (GSE18728), AT/CMF (GSE50948) or FAC (GSE20271) chemotherapy regimens when these patients had a high TN score. TFAC chemotherapy regimen was less likely to produce pCR in high TN score tumors in one study (GSE20194) but without a significant association in a second study (GSE20271). ER$^-$HER2$^-$ tumors with high TN score had a trend to lower response to AC/T chemotherapy (GSE22226 AC/T). In contrast, pCR was achieved in 57% and 60% of ER$^-$HER2$^-$ tumors with high TN score after treatment with the FEC/TX (GSE42822) and FAC/TX (GSE23988) regimens, respectively. Altogether, the rate of pCR stratified by the TN score was significantly different in either the low or high TN score tumor from the reported general 31% pCR rate in TNBC [9] (dotted line in Figure 23A). In one dataset, the ISPY-1 trial (GSE22226), the relapse-free survival (RFS) was also recorded. As shown in Figure 23B, pCR was a strong predictor of RFS in ER$^-$HER2$^-$ breast cancer as previously published [41]. The TN score was not only a strong predictor of RFS after chemotherapy, but also could stratify the survival of patients who achieved pCR further in addition to the stratification of patients who did not achieve pCR to good and poor prognosis groups (Figure 23B). This data indicates that the TN score is independent and has additional value to monitoring pCR after neoadjuvant chemotherapy in ER$^-$HER2$^-$ (TNBC) breast cancer patients. To further illustrate the utility of the TN score, we analyzed ER$^-$ and BLBC patient outcome in KM-plotter

for systemically untreated and treated patients separately. As summarized in Table 11 (Figure 34 for survival curves), the TN signature was prognostic in either systemically untreated or treated ER- and BLBC subtypes.

Therapeutic targets based on the TN signature

The overexpressed genes in the TN signature contains novel genes which have limited literature describing their function, particularly in cancer. These genes includes *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7* and *KCNG1*. These genes are novel candidates for future studies to investigate the effect of their knockdown on the survival of ER^- or TNBC breast cancer cell lines. In addition, we took two approaches to identify possible therapeutic strategies envisioned by the TN signature to benefit the poor survival of patients identified by this signature. First, we compared the global gene expression profile of TNBC/BLBC tumors with high TN score to those with low TN score. Secondly, we analyzed published pre-clinical studies which treated cancer cell lines with panels of molecularly targeted drugs to determine whether cell lines with high TN score display sensitive to particular drugs. In the first approach, a class comparison between the global gene expression profiles of BLBC or ER- tumors with high TN score to those with low TN score was carried out in the ROCK dataset. In comparison to low TN score BLBC tumors, high TN score BLBC tumors overexpressed 171 probes and underexpressed 251 probes (Table 15). In a similar analysis, high TN score ER^- tumors overexpressed 307 probes and underexpressed 332 probes (Table 16). Of the overexpressed probes, 87 probes (82 genes) were commonly overexpressed in high TN score BLBC and ER^- breast cancer compared to low TN score counterparts. Of the 87 probes, 39 probes were prognostic in BLBC and ER- breast cancer (marked in bold in Table 15). More importantly, the 87 probes include genes which encode several kinases, enzymes and ion channels which could be targets or current for future drug development for the treatment of the high TN score tumors that have poor outcome.

In the second approach, published studies which surveyed panels of molecular drugs against cancer cell lines were analyzed. The Cancer Cell Line Encyclopedia (CCLE) study [50] investigated the pharmacological profiles for 24 anticancer drugs across 479 cancer cell lines which were also profiled with gene expression arrays. We calculated the TN score for each cell line in this study and compared the sensitivity of these cell lines to the anticancer drugs according to the TN score. Cancer cell lines

with high TN score were less sensitive to inhibition of ALK (TAE684) and BCR-ABL (Nilotinib) but more sensitive to the inhibition of HSP90 (Tanespimycin [17-AAG]) and EGFR (Erlotinib or Lapatinib) (Figure 35). In a similar method, we also analyzed a second large study, Garnett et al. [49], which tested 130 drugs against more than 600 cancer cell lines. As shown in Figure 24, cell lines with high TN score were less sensitive to inhibition of PARP (ABT-888), retinoic acid (ATRA), Bcl2 (ABT-263), DHFR (methotrexate), glucose (metformin) and p38MAPK (BIRB 0796). Two IGF1R inhibitors showed different results; high TN score cell lines were less sensitive to the OSI-906 inhibitor but more sensitive to the BMS-536924 inhibitor. As shown in Figure 24, cell lines with high TN score were also sensitive to HSP90 inhibition (17-AAG and Elesclomol) in agreement with the findings from the CCLE study (Figure 35). High TN score cell lines were also more sensitive to mTOR/PI3K (BEZ235) and MEK (RDEA-119) inhibition.

Integration of the TN score and the aggressiveness score

We have recently published the aggressiveness gene signature/score (Agro score) [36] from a meta-analysis in Oncomine™ and validated that this score is prognostic in ER^+ breast cancer at the gene level. ER^- breast cancer, BLBC and TNBC almost consistently express high level of the Agro score thus this signature was not prognostic in these subtypes. We further showed that one of these genes, TTK/MPS1, is upregulated in TNBC cell lines and some ER- negative cell lines, and that TTK is a therapeutic target in these cell lines. Moreover, we showed that the TTK protein level by immunohistochemistry (IHC) is prognostic in very aggressive subgroups of breast cancer including high grade, proliferative tumors, lymph node positive, TNBC and $HER2^+$ subtypes [36]. The integration of the TN gene signature (prognostic in ER^-/BLBC/TNBC) and the Agro gene signature (prognostic in ER^+) would allow one integrated signature and score which will be prognostic in breast cancer irrespective of subtypes. As detailed in the methods section, the addition, subtraction, multiplication or division of the TN and Agro scores were investigated in the ROCK dataset to identify a direct relationship that would retain the information provided from each of the scores. A linear relationship was observed by the addition or subtraction of the TN and Agro scores (Figure 36), but only the integration by addition was prognostic in ER- patients (Figure 37). On the other hand, the multiplication and division of the TN and Agro score produced exponential and power curves relationships, respectively (Figure 38). Only the multiplication of the

scores was prognostic in ER- breast cancer (Figure 37). Since multiplication and division produced exponential and power curves for the relationship between the TN and Agro score, we also tested integration by one score raised to the power of the second score. Indeed, the TN score raised to the power of Agro score was highly prognostic in ER- and ER+ patients in the ROCK dataset (Figure 37). This method to integrate the TN and Agro scores, the integrated breast cancer recurrence (iBCR) score, was prognostic in all patients, ER- and ER+ patients in the ROCK dataset (Figure 25) and the TCGA dataset (Figure 26). Moreover, the iBCR score was as prognostic as the TN score in the homogenous TNBC dataset [32] (Figure 39), supporting the iBCR score as prognostic test in breast cancer.

The iBCR score and the likelihood of pCR after chemotherapy

The association of the iBCR score with patient survival and the likelihood of pCR after chemotherapy was investigated in the ISPY-1 trial (GSE22226). The RFS of $ER^-/HER2^-$ patients was stratified by iBCR score better than the TN score alone (Figure 27). High iBCR score $ER^-/HER2^-$ patients were less likely to achieve pCR (Figure 27), which could explain the poorer survival of these patients. In ER^+ breast cancer, the iBCR score stratified the RFS patients similarly to the Agro score. Although higher likelihood pCR was observed in high iBCR score ER^+ tumors (Figure 27), this subgroup had poor RFS. This can be explained by the small number of ER^+ patients who achieved pCR (10/62 [16%] vs. 10/34 [29%] in ER^-HER2^-). These results provide further validation and evidence for the value of the iBCR score as a single test which incorporates the Agro score (prognostic in ER^+) and the TN score (prognostic in ER^-). The results in Figure 25 from the ROCK dataset (Affymetrix platform), Figure 26 from the TCGA dataset (Illumina platform) and Figure 27 from the ISPY-1 trial (Agilent platform) also provide evidence for the robustness of the Agro and TN scores and the derived iBCR score across independent studies across the three major gene expression array platforms.

Next, the association of the iBCR score with pCR was investigated in other neoadjuvant chemotherapy datasets in both ER^-HER2^- and ER^+ patients. pCR was less likely in high iBCR ER^-/HER^- patients after TX (GSE18728) chemotherapy regimen and not different to low iBCR ER-/HER2- patients when treated with AT/CMF (GSE50948). In the other datasets, pCR was more likely in high iBCR score ER-/HER2- patients after treatment with FAC (GSE20271), TFAC (GSE20271

and GSE20194), FEC/TX (GSE42822) and FAC/TX (GSE23988) neoadjuvant chemotherapy regimens (Figure 28A).

As shown in the summary from these four studies in Table 12, of the total 183 ER^-HER2^- patients, 120 patients (65.6%) had high iBCR score and of these 54 patients (29.5%) achieved pCR while 66 patients (36.1%) did not achieve pCR. The larger number of patients with high iBCR score that did not achieving pCR (66/120, 55%) and that recurrence may be observed on high iBCR score patients after pCR (55/120, 45%) could explain the poorer survival of high iBCR score ER^-HER2^- patients (40-50% survival at 10 years in Figure 25 and Figure 26). Based on these studies and that chemotherapy is the mainstay in the treatment of $ER^-/HER2^-$ breast cancer, low iBCR score patients may be spared from additional treatments particularly if they achieve pCR after chemotherapy. On the other hand, high iBCR ER-HER2- patients and particularly those who do not achieve pCR should be offered additional therapy which could be based on the upregulated genes in the Agro or TN signatures or based on other overexpressed genes in these tumors (Tables 15 and 16) or from the pre-clinical analysis we performed from drug sensitivity studies (Figures 24 and 35).

High iBCR score in ER^+ was associated with higher likelihood of pCR after AT/CMF (GSE50948), TX (GSE18728), TFAC (GSE20271 and GSE20194) and FAC/TX (GSE23988) neoadjuvant chemotherapy regimens (Figure 38B). Despite this higher pCR likelihood, high iBCR ER+ patients have poorer survival (Figures 25 and 26) which could be explained by the small number of ER+ patients who achieve pCR (of the 207 ER^+ patients in the above five studies, 5 [2.5%] with low iBCR and 20 [9.7%] with high iBCR score achieved pCR). Thus, for ER^+ breast cancer where a decision about including chemotherapy with the standard endocrine therapy in the treatment planning may be informed by the iBCR score. The value of the iBCR score in the treatment planning of ER+ patients is the described next section.

The iBCR score and the treatment of ER^+ breast cancer

ER^+ breast cancer patients are treated with endocrine therapy, particularly tamoxifen. When these patients are lymph node positive (N1), adjuvant chemotherapy is also included. For lymph node negative (N0) ER^+ patients, decision to include chemotherapy is less certain as good prognosis patients (small and lower grade tumors) would be over-treated if chemotherapy is included whereas poorer prognosis patients (larger and higher grade tumors) would be under-treated if chemotherapy is not included. This clinical decision has been the motivation for the development of

Oncotype Dx® recurrence score, the MammaPrint® and more recently the PAM50 risk of recurrence score. We have previously published that the Agro score outperformed the Oncotype Dx and the MammaPrint tests in multivariate survival analysis in the METABRIC dataset of 2000 patients [36]. This finding is further supported by direct comparison of the Agro score to Oncotype Dx (Figure 40) and MammaPrint (Figure 41) in all ER^+ patients and in the N0 and N1 subsets. For the iBCR score, as shown in Figure 29A, this score was prognostic in ER^+ N0 patients who were not treated with tamoxifen indicating that high iBCR ER^+ N0 patients should be treated with tamoxifen. When ER+ N0 or N1 patients are treated with tamoxifen, the iBCR score can still identify patients who have poor RFS (Figure 29B) and DMFS (Figure 29C). Thus, ER+ N0 or N1 patients with high iBCR score may benefit from the inclusion of adjuvant chemotherapy in their treatment as these patients may experience better pCR (Figure 28B). Nonetheless, as pCR rate in ER^+ is not high, high iBCR score ER+ patients, particularly N1, should be offered additional targeted therapies. The type of targeted therapies for these patients is suggested in the next section.

The iBCR score predicts therapies for $ER^-/HER2^-$ and ER^+ and breast cancer subtypes

The overexpressed genes in the Agro and TN signature contain targetable genes which could be useful for therapeutic intervention against the high iBCR tumors which have poor survival after the standard treatments. Similar to the analysis performed for the TN signature above, we took two approached to identify additional possible targets in the high iBCR score breast tumors. In the first approach, a class comparison between the global gene expression profiles of ER^+ or ER^- tumors with high iBCR score to those with low iBCR score was carried out in the ROCK dataset. The produced gene-list (1178 probes, data not shown) was then filtered by comparison to normal breast tissue which was also profiled in this dataset. In comparison to low iBCR score tumors and normal breast tissue, high iBCR score tumors overexpressed 204 probes (181 genes) and underexpressed 124 probes (116 genes) (Table 17). Of the 181 overexpressed genes, 134 genes were specifically upregulated in high iBCR score ER^+ vs. normal breast and low iBCR ER^+ and 95 genes were specifically upregulated in high iBCR score ER^- vs. normal breast and low iBCR ER^-. As shown in Table 13, 49 genes were uniquely upregulated in high iBCR score ER- tumors compared to low score iBCR score ER^- tumors and normal

breast tissue. Similar comparison revealed that high iBCR score ER^+ tumors have unique upregulation of 86 genes. High iBCR score ER^- and ER^+ tumors commonly overexpressed 46 genes in comparison to low score iBCR counterparts and normal breast tissue. These genes encode several kinases, enzymes and ion channels which could be targets for current or future drug development for the treatment of the high iBCR score tumors with poor outcome. Of the downregulated probes, a particularly interesting hit was the micro-RNA (miRNA) hsa-mir-568 (9.3- and 2.2-fold downregulated in high iBCR score ER^- vs. normal breast and low iBCR score ER^-, respectively; 5.6- and 2.9-fold downregulated in high iBCR score ER^+ vs. normal breast and low iBCR score ER^+, respectively). This downregulated miRNA in the high iBCR score tumors targets several of the upregulated genes in these tumors, particularly those which are upregulated compared to normal breast tissue (Table 18). This miRNA could be a genomic-based treatment against high iBCR score breast cancers.

In the second approach, again similar to the above analysis for the TN score, published studies of drug screens were analyzed for the association of the iBCR score with sensitivity of cancer cell lines to anti-cancer drugs. In the CCLE study (Figure 42), cancer cell lines with high iBCR score were less sensitive to inhibition of ALK (TAE684) and BCR-ABL (Nilotinib) similar to results from the TN score. In addition, high iBCR cell lines were less sensitive to inhibition of FGFR (TKI258) and IGF1R (AEW541). High iBCR score cell lines were more sensitive to the inhibition of HSP90 (Tanespimycin [17-AAG]) (Figure 42). In the second large study by Garnett et al. [49], high iBCR score cell lines were more sensitive to low iBCR score cell lines to 8 anticancer drugs (Figure 30). These include inhibitors of HSP90 (17AAG), mTOR/PI3K (BEZ235) and IGF1R (BMS-536924) as also observed in the TN score results. Additionally, high iBCR score cell lines were more sensitive to inhibition of PI3K (GDC0941), mTOR (JW-7-25-1), XIAP (Embelin) and PLK1 (BI-2536) which also matched results from Agro score results (Figure 30). The Agro score also identified sensitivity to inhibition of RSK (CMK), MEK (PD0325901) and DNA damage (Bleomycin). Similar to results from high TN score, high iBCR score cell lines were also less sensitive to the inhibition of PARP (ABT-888 and AZD-2281), retinoic acid (ATRA), Bcl2 (ABT-263), DHFR (methotrexate) and glucose (metformin). Additionally, high iBCR score cell lines were less sensitive to inhibition of SYK (BAY613606), HDAC (Vorinostat) and BCR-ABL (Nilotinib)

and p38MAPK (BIRB 0796). High Agro score cell lines were less sensitive to an additional drug against GSK3A/B (SB216763). Altogether, the TN score (Figures 24 and 35) and the Agro score and the combined iBCR score (Figures 30 and 42) associate with sensitivity to several anticancer drugs and future experimental validation would establish these scores as companion diagnostic for these drugs and benefit breast cancer patients by directing these drugs to the high score patients with poor survival.

Sensitivity of breast cancer cell lines to targeted inhibitors according to the iBCR score

Breast cancer cell lines (10 cell lines); BT-549, MDA-MB-231, MDA-MB-436, MDA-MB-468, BT-20, Hs.578T, BT-474, MCF-7, T-47D, and ZR-75-1, were cultured in the absence or presence of escalating doses of 24 anti-cancer drugs. The survival of cells was determined six days in comparison to untreated cells using the MTS/MTA assay. The response of the cell lines to the drugs was analyzed in GraphPad® Prism using a dose response curve to calculate the $\log_{10}$ of IC50 (IC50 is the dose required to kill 50% of the cells). Sensitivity was presented as the -$\log_{10}$[IC50]. This drug screen which we published previously (Al-Ejeh *et al.*, *Oncotarget*, 2014) was re-analyzed according to the iBCR score. The gene expression datasets of 51 breast cancer cell lines by Neve et al. (*Cancer Cell*, 2006), was analyzed to calculate the Agro and TN scores for each cell line to calculate the iBCR score. Each cell line was assigned as low of high iBCR score by dichotomy across the median of all the cell lines in the Neve et al. dataset. Based on the low or high iBCR score classification, the sensitivity of the 10 cell lines used in our screen was compared between high iBCR score cell lines (5 cell lines) to low iBCR score cell lines (5 cell lines). As shown in Figure 47, high iBCR score cell lines were significantly more sensitive to the inhibition of p38MAPK (LY2228820), PLC□ (U73122), JNK (SP600125), PAK1 (IPA3), MEK (AS703026 and AZD6244), ERK5 (XMD 8-92 and BIX02188), HSP90 (17-AAG, PF0429113 and AUY922), IGF1R (GSK1904529A) and EGFR (Afatinib). The results from our screen are in agreement with the higher sensitivity of high iBCR score cancer cell lines to HSP90, IGF1R and MEK inhibitors we identified from the two previously published large cell line studies.

Discussion

Our meta-analysis of gene expression datasets in the Oncomine™ database has previously identified a signature, the Aggressiveness signature (Agro signature), which was prognostic in ER^+ breast cancer. We validated one of the genes in this signature, TTK/MPS1, by IHC and found that TTK positivity in interphase cells (exclusive of mitotic cells) was prognostic in highly aggressive breast cancers such as high grade, high grade and lymph node positive and highly proliferative (Ki67 positive) cases [36]. In this study, we used our meta-analysis approach to identify a second signature, the triple negative signature (TN signature), which was highly prognostic in ER^-, TNBC and BLBC subtypes. The TN signature outperformed all standard clincopatholical indicators in multivariate survival analysis and also outperformed published signatures in ER- breast cancer. We were also able to integrate the Agro signature (prognostic in ER^+ breast cancer) to produce the **i**ntegrated **B**reast **C**ancer **R**ecurrence (iBCR) test. The two signatures and the iBCR were validated in large independent cohorts of breast cancer studies irrespective of the gene expression arrays used indicating the experimenter/technology independence of our signatures. Importantly, both the Agro and TN signatures and the iBCR test associated with response and outcome after endocrine therapy for ER^+ and neoadjuvant chemotherapy for ER^- and ER^+ breast cancers. Moreover, by comparison of the global gene expression profiles of high iBCR score tumors to low iBCR score tumors, we were able to identify several overexpressed targets which can be used for the targeted therapy of these poor prognosis patients who are not really benefiting from the current treatment standards. In addition, mining of large preclinical studies of drug screens against cancer cell lines showed that the signatures and iBCR score predict higher sensitivity of cell lines to particular drugs. Thus, the signatures and the iBCR test could be used as a companion diagnostic to direct targeted therapies to those patients who would benefit from these treatments to increase their low survival rates. Altogether, our studies have not only extensively illustrated the potential of our signatures in personalized medicine, but may also shed light for future studies to understand the underlying mechanisms for the aggressiveness of tumors that the iBCR test identified that lead to poor survival

To date, there is an unmet medical need for the prognostication of ER- breast cancer and the development of effective therapies against these tumors particularly when lacking HER2 expression. Chemotherapy remains to be the only standard therapy in these patients and the response rate after chemotherapy in the neoadjuvant setting is

reported as 31% in ER$^-$HER2$^-$ (TNBC) patients [9]. Identifying patients who would truly benefit from chemotherapy would aid clinicians to determine patients who may require longer or additional treatment regimens including investigational clinical trial enrolment. Our signatures and the iBCR score predict higher pCR after chemotherapy in patients who have high scores compared to those with low score. The low score patients have better survival and may not require additional therapy. On the other hand, despite the higher pCR in high score patients, this patient subgroup still has poor survival and recurrences were present even after achieving pCR in high score patients when we analyzed the data from the ISPY-1 trial. Our results from comparative analysis and mining pre-clinical drug screens identified several targets and sensitivity to drugs in development. Thus, ER- and particularly TNBC patients with high scores for our signatures/iBCR test may benefit from the inclusion of therapies envisioned by these signatures to increase their survival rates. Such clinical development will depend on future prospective validation of our signatures and the iBCR test in clinical trials and pre-clinical studies.

In ER$^+$ breast cancer, three commercial tests exist for clinical decisions to spare or include adjuvant chemotherapy with the standard endocrine therapy; Oncotype Dx®, MammaPrint® and Prosigna®. These have been validated for ER$^+$ lymph node negative (N0) breast cancer patients treated with endocrine therapy whether patients with high risk according to these tests are recommended for adjuvant chemotherapy. Our signatures and the iBCR test outperformed these tests in a direct comparison in ER$^+$ N0 patient-survival after tamoxifen therapy. Moreover, our tests also predicted the response of ER$^+$ patients to chemotherapy and importantly could predict sensitivity to targeted therapies. The current commercial tests do not have this capability. Importantly, our signatures and the iBCR test was also prognostic in the subgroup with unmet need, ER$^+$ lymph node positive breast cancer (ER$^+$ N1). The survival of these patients was stratified to poor and good prognosis groups by our signatures and iBCR test which also informed whether these patients are benefiting from endocrine therapy. Clinical validation of our signatures and the iBCR test along with validation of drug sensitivity predictions would aid the development of new treatment regimens for ER$^+$ patients who are at high risk of relapse or metastatic spread after the current treatment standards.

The comparison of aggressive ER$^-$ tumors identified by our signatures to their counterparts and to normal breast tissue identified several kinases, enzymes (redox

particularly) and potassium channels which could inform new directions in developing targeted treatments against ER^- breast cancer. On the other hand, for aggressive ER^+ tumors identified by our signatures, although targets were not restricted to cell cycle and proliferation, these functions were notably enriched. This high proliferation profile could explain the higher pCR in these tumors after chemotherapy as proliferative tumors would be more responsive to chemotherapeutics. Nonetheless, we have previously clarified that the overexpressed genes in the Agro signature, thus the iBCR test, are genes that are involved in kinetochore binding and chromosome segregations and that the signature is prognostic even in proliferative tumors (high Ki67 expression) [36]. Deregulation of genes involved in chromosome segregation would produce aneuploidy and chromosomal instability (CIN) [52]. At least *in vivo*, chemotherapy has been shown to induce the proliferation quiescent aneuploid cells as a mechanism for therapy resistance [53]. In support of the notion that high Agro score is related to aneuploidy, analysis of the copy number variations (CNVs) TCGA data showed that high Agro score tumours, compared to low Agro score tumors, have high level of CNVs, particularly those involving whole chromosomes or chromosome arms (Figure 43). Thus, although proliferation may be a characteristic of high Agro/iBCR score ER^+ tumors, these tumors appear to be aneuploid. In line with this notion, the sensitivity of high Agro/iBCR score cell lines to PLK1 and HSP90 inhibition (Figure 30) and aurora kinase inhibitors (Figure 44) support that high Agro/iBCR scores predict sensitivity to anti-aneuploid therapy. PLK1 and Aurora kinases are classical targets in aneuploidy and HSP90 inhibition has been reported to selectively kill aneuploid cancer cells [54]. HSP90 sensitivity was also found for high TN score tumors and interestingly, we have previously identified HSP90 as a target in TNBC by kinome profiling of breast cancer. We showed that HSP90 inhibition in combination therapy is effective *in vitro* and *in vivo* [55]. We propose that anti-aneuploid drugs should be effective against ER^+ tumors with high Agro/iBCR scores including PLK1, Aurora kinase and HSP90 inhibitors and that HSP90 inhibition should be effective in high TN/iBCR score ER^- tumors. While other therapies envisioned by our signatures and the iBCR test should also be investigated, the above targets represent first line targets for initial validation and development.

In conclusion, our meta-analysis in Oncomine™ and extensive subsequent validation and analysis have developed novel signatures and an integrated genomic test for the

prognosis of breast cancer and prediction of response to standard treatments irrespective of ER status. The novel signatures and their integration also have the potential as companion diagnostic tests for several classes of targeted therapies in breast cancer patients who suffer poor survival. Future validation and clinical development of our signatures and the iBCR test holds a great potential and impact on personalized and precision medicine for breast cancer. Finally, it should be noted that the iBCR test has value in the prognosis of several other cancers (Figure 45) and particularly in lung adenocarcinoma (Figure 46), thus our approach and novel signatures may extend benefit to other cancer types.

REFERENCES

1. Kang, S.P., M. Martel, and L.N. Harris, *Triple negative breast cancer: current understanding of biology and treatment options.* Curr Opin Obstet Gynecol, 2008. **20**(1): p. 40-6.

2. Schneider, B.P., et al., *Triple-negative breast cancer: risk factors to potential targets.* Clin Cancer Res, 2008. **14**(24): p. 8010-8.

3. Rakha, E.A., J.S. Reis-Filho, and I.O. Ellis, *Basal-like breast cancer: a critical review.* J Clin Oncol, 2008. **26**(15): p. 2568-81.

4. Fulford, L.G., et al., *Basal-like grade III invasive ductal carcinoma of the breast: patterns of metastasis and long-term survival.* Breast Cancer Res, 2007. **9**(1): p. R4.

5. Goldstein, L.J., et al., *Concurrent doxorubicin plus docetaxel is not more effective than concurrent doxorubicin plus cyclophosphamide in operable breast cancer with 0 to 3 positive axillary nodes: North American Breast Cancer Intergroup Trial E 2197.* J Clin Oncol, 2008. **26**(25): p. 4092-9.

6. Keam, B., et al., *Prognostic impact of clinicopathologic parameters in stage II/III breast cancer treated with neoadjuvant docetaxel and doxorubicin chemotherapy: paradoxical features of the triple negative breast cancer.* BMC Cancer, 2007. **7**: p. 203.

7. Liedtke, C., et al., *Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer.* J Clin Oncol, 2008. **26**(8): p. 1275-81.

8. Carey, L.A., et al., *The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes.* Clin Cancer Res, 2007. **13**(8): p. 2329-34.

9. von Minckwitz, G., et al., *Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes.* J Clin Oncol, 2012. **30**(15): p. 1796-804.

10. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.

11. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.

12. Hu, Z., et al., *The molecular portraits of breast tumors are conserved across microarray platforms.* BMC Genomics, 2006. **7**: p. 96-107.

13. Parker, J.S., et al., *Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes.* J Clin Oncol, 2009.

14. Weigelt, B., et al., *Breast cancer molecular profiling with single sample predictors: a retrospective analysis.* Lancet Oncol. **11**(4): p. 339-49.

15. Weigelt, B., et al., *Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer.* Cancer Res, 2005. **65**(20): p. 9155-8.

16. Parker, J.S., et al., *Supervised risk predictor of breast cancer based on intrinsic subtypes.* J Clin Oncol, 2009. **27**(8): p. 1160-7.

17. Lehmann, B.D., et al., *Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies.* J Clin Invest, 2011. **121**(7): p. 2750-67.

18. Shah, S.P., et al., *The clonal and mutational evolution spectrum of primary triple-negative breast cancers.* Nature, 2012. **486**(7403): p. 395-9.

19. Irshad, S., P. Ellis, and A. Tutt, *Molecular heterogeneity of triple-negative breast cancer and its clinical implications.* Curr Opin Oncol, 2011. **23**(6): p. 566-77.

20. Criscitiello, C., et al., *Understanding the biology of triple-negative breast cancer.* Ann Oncol, 2012. **23 Suppl 6**: p. vi13-8.

21. Masuda, H., et al., *Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes.* Clin Cancer Res, 2013. **19**(19): p. 5533-40.

22. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. **415**(6871): p. 530-6.

23. Paik, S., et al., *A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.* N Engl J Med, 2004. **351**(27): p. 2817-26.

24. Buyse, M., et al., *Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer.* J Natl Cancer Inst, 2006. **98**(17): p. 1183-92.

25. Loi, S., et al., *Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade.* J Clin Oncol, 2007. **25**(10): p. 1239-46.

26. Ma, X.J., et al., *A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer.* Clin Cancer Res, 2008. **14**(9): p. 2601-8.

27. Ma, X.J., et al., *A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen.* Cancer Cell, 2004. **5**(6): p. 607-16.

28. Sotiriou, C., et al., *Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis.* Journal of the National Cancer Institute, 2006. **98**(4): p. 262-72.

29. Dowsett, M., et al., *Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy.* J Clin Oncol, 2013. **31**(22): p. 2783-90.

30. Yau, C., et al., *A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer.* Breast Cancer Res, 2010. **12**(5): p. R85.

31. Rody, A., et al., *A clinically relevant gene signature in triple negative and basal-like breast cancer.* Breast Cancer Res, 2011. **13**(5): p. R97.

32. Karn, T., et al., *Homogeneous datasets of triple negative breast cancers enable the identification of novel prognostic and predictive signatures.* PLoS One, 2011. **6**(12): p. e28403.

33. Yu, K.D., et al., *Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer.* Clin Cancer Res, 2013. **19**(10): p. 2723-33.

34. Lee, U., et al., *A prognostic gene signature for metastasis-free survival of triple negative breast cancer patients.* PLoS One, 2013. **8**(12): p. e82125.

35. Hallett, R.M., et al., *A gene signature for predicting outcome in patients with basal-like breast cancer.* Sci Rep, 2012. **2**: p. 227.

36. Al-Ejeh, F., et al., *Meta-analysis of the global gene expression profile of triple-negative breast cancer identifies genes for the prognostication and treatment of aggressive breast cancer.* Oncogenesis, 2014. **3**: p. e100.

37. Rhodes, D.R., et al., *ONCOMINE: a cancer microarray database and integrated data-mining platform.* Neoplasia, 2004. **6**(1): p. 1-6.

38. Gyorffy, B., et al., *An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients.* Breast Cancer Res Treat, 2010. **123**(3): p. 725-31.

39. TCGA, *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.

40. Ur-Rehman, S., et al., *ROCK: a resource for integrative breast cancer data analysis.* Breast Cancer Res Treat, 2013. **139**(3): p. 907-21.

41. Esserman, L.J., et al., *Pathologic complete response predicts recurrence-free survival more effectively by cancer subset: results from the I-SPY 1 TRIAL--CALGB 150007/150012, ACRIN 6657.* J Clin Oncol, 2012. **30**(26): p. 3242-9.

42. Korde, L.A., et al., *Gene expression pathway analysis to predict response to neoadjuvant docetaxel and capecitabine for breast cancer.* Breast Cancer Res Treat, 2010. **119**(3): p. 685-99.

43. Prat, A., et al., *Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study.* Clin Cancer Res, 2014. **20**(2): p. 511-21.

44. Tabchy, A., et al., *Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer.* Clin Cancer Res, 2010. **16**(21): p. 5351-61.

45. Popovici, V., et al., *Effect of training-sample size and classification difficulty on the accuracy of genomic predictors.* Breast Cancer Res, 2010. **12**(1): p. R5.

46. Esserman, L.J., et al., *Chemotherapy response and recurrence-free survival in neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657).* Breast Cancer Res Treat, 2012. **132**(3): p. 1049-62.

47. Shen, K., et al., *Cell line derived multi-gene predictor of pathologic response to neoadjuvant chemotherapy in breast cancer: a validation study on US Oncology 02-103 clinical trial.* BMC Med Genomics, 2012. **5**: p. 51.

48. Iwamoto, T., et al., *Gene pathways associated with prognosis and chemotherapy sensitivity in molecular subtypes of breast cancer.* J Natl Cancer Inst, 2011. **103**(3): p. 264-72.

49. Garnett, M.J., et al., *Systematic identification of genomic markers of drug sensitivity in cancer cells.* Nature, 2012. **483**(7391): p. 570-5.

50. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.* Nature, 2012. **483**(7391): p. 603-7.

51. Symmans, W.F., et al., *Genomic index of sensitivity to endocrine therapy for breast cancer.* J Clin Oncol, 2010. **28**(27): p. 4111-9.

52. Bakhoum, S.F. and D.A. Compton, *Chromosomal instability and cancer: a complex relationship with therapeutic potential.* J Clin Invest, 2012. **122**(4): p. 1138-43.

53. Kusumbe, A.P. and S.A. Bapat, *Cancer stem cells and aneuploid populations within developing tumors are the major determinants of tumor dormancy.* Cancer Res, 2009. **69**(24): p. 9245-53.

54. Tang, Y.C., et al., *Identification of aneuploidy-selective antiproliferation compounds.* Cell, 2011. **144**(4): p. 499-512.

55. Al-Ejeh, F., et al., *Kinome profiling reveals breast cancer heterogeneity and identifies targeted therapeutic opportunities for triple negative breast cancer.* Oncotarget, 2014.

56. Bos, P.D., et al., *Genes that mediate breast cancer metastasis to the brain.* Nature, 2009. **459**(7249): p. 1005-9.

57. Desmedt, C., et al., *Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series.* Clin Cancer Res, 2007. **13**(11): p. 3207-14.

58. Hatzis, C., et al., *A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer.* JAMA, 2011. **305**(18): p. 1873-81.

59. Kao, K.J., et al., *Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization.* BMC Cancer, 2011. **11**: p. 143.

60. Schmidt, M., et al., *The humoral immune system has a key prognostic impact in node-negative breast cancer.* Cancer Res, 2008. **68**(13): p. 5405-13.

61. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. **347**(25): p. 1999-2009.

62. Bild, A.H., et al., *Oncogenic pathway signatures in human cancers as a guide to targeted therapies.* Nature, 2006. **439**(7074): p. 353-7.

63. Pawitan, Y., et al., *Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts.* Breast Cancer Res, 2005. **7**(6): p. R953-64.

64. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.

65. Zhao, Y. and R. Simon, *BRB-ArrayTools Data Archive for human cancer gene expression: a unique and efficient data sharing resource.* Cancer Inform, 2008. **6**: p. 9-15.

66. Cline, M.S., et al., *Exploring TCGA Pan-Cancer data at the UCSC Cancer Genomics Browser.* Sci Rep, 2013. **3**: p. 2652.

67. Goldman, M., et al., *The UCSC Cancer Genomics Browser: update 2013.* Nucleic Acids Res, 2013. **41**(Database issue): p. D949-54.

68. Al-Ejeh, F., et al., *Treatment of triple-negative breast cancer using anti-EGFR-directed radioimmunotherapy combined with radiosensitizing chemotherapy and PARP inhibitor.* J Nucl Med, 2013. **54**(6): p. 913-21.

69. Diamond, J.R., et al., *Predictive biomarkers of sensitivity to the aurora and angiogenic kinase inhibitor ENMD-2076 in preclinical breast cancer models.* Clin Cancer Res, 2013. **19**(1): p. 291-303.

70. Kalous, O., et al., *AMG 900, pan-Aurora kinase inhibitor, preferentially inhibits the proliferation of breast cancer cell lines with dysfunctional p53.* Breast Cancer Res Treat, 2013. **141**(3): p. 397-408.

Table 10: The 28-gene signature discovered from a meta-analysis of gene expression data in breast cancer in Oncomine™

Gene Symbol	Affymetrix probe	Entrez	Gene name
↑ABHD5	213935_at	51099	abhydrolase domain containing 5; 1-acylglycerol-3-phosphate O-acyltransferase
↑ADORA 2B	205891_at	136	adenosine A2b receptor
↑BCAP31	200837_at	10134	B-cell receptor-associated protein 31
↑CA9	205199_at	768	carbonic anhydrase IX
↑CAMSA P1	212711_at	157922	calmodulin regulated spectrin-associated protein 1
↑CARHSP 1	218384_at	23589	calcium regulated heat stable protein 1, 24kDa
↑CD55	201926_s_at	1604	CD55 molecule, decay accelerating factor for complement (Cromer blood group)
↑CETN3	209662_at	1070	centrin, EF-hand protein, 3
↑EIF3K	221494_x_at	27335	eukaryotic translation initiation factor 3, subunit K
↑EXOSC7	212627_s_at	23016	exosome component 7
↑GNB2L1	200651_at	10399	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1
↑GRHPR	214864_s_at	9380	glyoxylate reductase/hydroxypyruvate reductase
↑GSK3B	209945_s_at	2932	glycogen synthase kinase 3 beta
↑HCFC1R 1	218537_at	54985	host cell factor C1 regulator 1 (XPO1 dependent)
↑KCNG1	214595_at	3755	potassium voltage-gated channel, subfamily G, member 1
↑MAP2K5	211370_s_at	5607	mitogen-activated protein kinase kinase 5
↑NDUFC1	203478_at	4717	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa
↑PML	206503_x_at	5371	promyelocytic leukemia
↑STAU1	208948_s_at	6780	staufen, RNA binding protein, homolog 1 (Drosophila)
↑TXN	216609_at	7295	thioredoxin
↑ZNF593	204175_at	51042	zinc finger protein 593
↓BTN2A2	205298_s_at	10385	butyrophilin, subfamily 2, member A2
↓ERC2	213938_at	26059	ELKS/RAB6-interacting/CAST family member 2
↓IGH	211649_x_at	3492	immunoglobulin heavy locus
↓ME1	211204_at	4199	malic enzyme 1, NADP(+)-dependent, cytosolic
↓MTMR7	217292_at	9108	myotubularin related protein 7
↓SMPDL3 B	205309_at	27293	sphingomyelin phosphodiesterase, acid-like 3B
↓ZNRD1-↓AS1	215985_at	80862	ZNRD1 antisense RNA 1

Table 11: The TN signature is prognostic in ER- and BLBC irrespective of systemic therapy.

		Untreated			Treated		
		HR	**CI 95%**	**p-value**	**HR**	**CI 95%**	**p-value**
ER-	RFS	2.02	1.25 – 3.26	3.20E-03	2.59	1.84 – 3.60	1.70E-08
	DMFS	4.10	1.44 – 11.7	4.20E-03	1.89	1.04 – 3.43	3.40E-02
	OS	1.77	0.65 – 4.83	*0.26*	3.82	1.43 – 10.18	3.90E-03
BLBC	RFS	2.48	1.46 – 4.21	5.10E-04	2.88	1.94 – 4.28	4.50E-08
	DMFS	5.54	1.66 – 18.48	1.70E-03	3.14	1.38 – 7.19	4.20E-03
	OS	2.42	0.79 – 7.47	*0.11*	4.89	1.65 – 14.46	1.50E-03

The 28-gene signature was used as described in Figure 2 in the online tool KM-plotter but restricting the analysis on ER- or BLBC patients who were untreated systemically or systemically treated. The survival curves for RFS, DMFS and OS are shown in Figure 34; only the hazard ratio (HR), the 95% confidence interval (CI 95%) and the log-rank p-value from these curves are reported in the Table.

Table 12: The likelihood of pCR in ER-HER2- patients according to the iBCR score

	pCR	**no pCR**	**Sum**
Low Score	12 (6.6%)	51 (27.9%)	63 (34.4%)
High Score	54 (29.5%)	66 (36.1%)	120 (65.6%)
Total	66 (36.1%)	117 (63.1%)	183 (100%)

ER-/HER2- patients stratified by low and high iBCR scores from four studies were compared for achieving or not achieving pCR after four chemotherapy regimens: FAC (GSE20271), TFAC (GSE20271 and GSE30194), FEC/TX (GSE42822) and FAC/TX (GSE23988)

Table 13: Upregulated genes in high iBCR score tumors compared to low iBCR tumors and normal breast tissue

High iBCR score ER- vs. low iBCR score ER- and normal breast			High iBCR score ER+ vs. low iBCR score ER+ and normal breast				Common in high iBCR score ER-/+ vs. low iBCR score and normal		
ACE2	HMGB3		ACP1	ENO1	MCM4		ASPM	HN1	
ADM	IL8		APOBEC3B	EPRS	MCM6	RANBP1	AURKA	KCNK1	
AR	IMPA2		ATAD2	EXOSC4	MCM7	RECQL4	BIRC5	KIF4A	
BNIP3	KYNU		AURKB	FADS1	MRPL13	RFC2	BUB1	MKI67	
C1orf106	LBP		BOP1	FANCI	MRPL15	RMDN1	BUB1B	MLF1IP	
CALML5	LRP8		CACYBP	GINS2	MSH6	RSAD2	CCNB1	MMP1	
CBS	MAGEA3	TDO2	CALU	GTSE1	MYBL2	SHMT2	CCNB2	MTFR1	
CCL18	MAGEA6	TMEM45A	CCNA2	H2AFZ	NCAPG	SMC4	CCNE2	NDC80	YKT6
CD24	ME1	TMSB15A	CCT2	HELLS	NDUFS8	SPAG5	CDC20	NEK2	ZWINT
CLIC3	MMP12	VEGFA	CDCA3	HMMR	NUDT21	SQLE	CDC6	NUSAP1	
CORO1C	PFKP	VGLL1	CEBPG	HSPH1	NUTF2	STIP1	CDK1	PDXK	
CP	PHLDA2		CKS1B	KIAA0101	OIP5	TACC3	CDKN3	PHB	
CRISP3	PTPN12		CXCL10	KIF11	PBK	TBCE	CENPF	PRC1	
DDC	QPRT		CXCL11	KIF14	PCNA	TIMM17A	CKS2	PTTG1	
ECT2	S100A7		DERL1	KIF20A	PGK1	TMPO	CNIH4	RRM2	
EZH2	S100A9		DHFR	KPNA2	PLOD2	TSN	CNTNAP2	S100A8	
FABP7	SCD		DNPH1	KPNA4	PRAME	TYMS	DDA1	S100P	
FAR2	SLC7A5		DONSON	LAPTM4B	PSMA7	UBE2S	DLGAP5	SPP1	
GABBR2	SOD2		DSCC1	LMNB1	PSMC3	UCK2	DTL	TK1	
GALNT3	SOX11		EIF4EBP1	LSM4	RACGAP1	WHSC1	ESRP1	TOP2A	
GMPS	SRD5A1		EIF5A	LY6E	RAD21	ZWILCH	GINS1	TRIP13	
GPSM2	ST14		EMC8	MAD2L1	RAD54B		HIST1H2BG	UBE2C	

Table 14. Univariate survival analysis of genes from the Oncomine metanalysis in the KM-Plotter online tool in BLBC and ER- breast cancer. Deriving the 28-gene signature.

		Basal-like breast cancer					
		RFS		DMFS		OS	
Gene Name	Affy Probe ID	HR	P	HR	P	HR	P
BTN2A2	205298_s_at	0.41	6.00E-12	0.34	7.20E-05	0.4	1.10E-03
IGHA1	211649_x_at	0.55	6.80E-06	0.39	1.50E-04	0.48	8.40E-03
MTMR7	217292_at	0.73	1.70E-02	0.50	1.10E-03	0.44	7.20E-03
ME1	211204_at	0.65	2.30E-03	0.38	2.10E-03	0.66	2.00E-01
BRI3	221004_s_at	0.59	9.00E-04	0.44	3.30E-03	0.34	1.60E-04
RBM38	212430_at	0.55	3.20E-04	0.48	4.20E-03	0.47	6.00E-03
GGA2	210658_s_at	0.65	1.10E-03	0.48	4.40E-03	0.6	7.90E-02
DDX11	208149_x_at	0.66	1.70E-03	0.46	7.50E-03	0.5	2.40E-02
ESPL1	38158_at	0.55	7.30E-06	0.50	8.60E-03	0.4	1.00E-03
CALR	214315_x_at	0.55	5.80E-06	0.50	9.20E-03	0.62	9.80E-02
CCL13	206407_s_at	0.58	4.70E-05	0.52	1.10E-02	0.87	6.50E-01
FANCA	203805_s_at	0.60	2.20E-04	0.54	1.70E-02	0.45	3.70E-03
NUP205	222382_x_at	0.52	3.00E-03	0.54	1.70E-02	0.56	3.90E-02
PDK3	221957_at	0.64	6.50E-04	0.55	2.10E-02	0.54	2.60E-02
MX2	204994_at	0.52	6.70E-07	0.56	2.80E-02	0.62	9.10E-02
SIRPA	202895_s_at	0.63	3.70E-04	0.58	4.10E-02	0.77	3.70E-01
FBXO41	221945_at	0.59	1.10E-04	0.53	4.20E-02	0.6	1.80E-01
COX5A	217451_at	0.75	3.60E-02	0.58	4.60E-02	0.73	2.70E-01
NOP2	214427_at	0.65	1.80E-03	0.61	5.90E-02	0.57	4.30E-02
C1QB	202953_at	0.60	1.30E-04	0.64	9.60E-02	0.66	1.50E-01
NCRNA00171	215985_at	0.59	2.40E-04	0.44	1.50E-03	0.75	3.50E-01
SMPDL3B	205309_at	1.04	7.80E-01	0.48	4.90E-03	0.61	8.80E-02
ERC2	213938_at	0.67	2.60E-03	0.50	6.80E-03	0.53	2.40E-02
DCLRE1C	219678_x_at	0.57	1.70E-05	0.52	1.10E-02	0.8	4.40E-01
FOXK2	203064_s_at	0.72	1.80E-02	0.69	1.50E-02	0.59	5.80E-02
PTCD3	217895_at	0.70	7.80E-03	0.53	1.70E-02	0.83	5.10E-01
GART	210005_at	0.56	2.10E-05	0.55	2.30E-02	0.64	1.30E-01
NF1	204325_s_at	0.70	1.20E-02	0.55	3.70E-02	0.9	7.30E-01
ASF1B	218115_at	0.70	8.00E-03	0.58	4.00E-02	0.66	1.40E-01
MLLT10	205408_at	0.63	5.70E-04	0.59	4.20E-02	0.92	7.60E-01
C11orf41	214772_at	0.70	7.90E-03	0.60	4.90E-02	0.82	4.80E-01
KIF18B	222039_at	0.58	6.10E-04	0.63	1.30E-01	0.5	3.70E-02
RECQL5	221686_s_at	0.59	3.30E-04	0.64	1.50E-01	0.54	5.40E-02
PKMYT1	204267_x_at	0.69	5.60E-03	0.66	1.10E-01	0.59	6.20E-02
TRAPPC10	208184_s_at	0.64	5.80E-04	0.64	9.00E-02	0.63	1.10E-01
ATAD5	220223_at	0.66	1.70E-03	0.68	1.40E-01	0.65	1.30E-01
RHBDD3	217622_at	0.60	1.20E-04	0.63	9.90E-02	0.68	1.80E-01
SLCO1A2	211481_at	0.62	4.60E-04	0.62	9.10E-02	0.69	2.10E-01
MARS	213672_at	0.73	2.00E-02	0.47	7.60E-03	0.75	3.20E-01

CD163	216233_at	0.57	2.10E-05	0.63	8.30E-02	0.76	3.20E-01
LOC100131613	210928_at	0.61	4.10E-04	0.80	4.30E-01	0.63	1.30E-01
PPIL2	206064_s_at	0.67	3.80E-03	0.69	1.90E-01	0.76	4.00E-01
FAM49B	217535_at	0.59	1.20E-04	0.68	1.60E-01	0.8	4.50E-01
MRPS12	213840_s_at	0.66	2.50E-03	1.23	4.30E-01	1.4	2.30E-01
GCH1	204224_s_at	0.66	1.70E-03	0.66	1.20E-01	1.14	6.80E-01
LAD1	203287_at	0.66	2.60E-03	0.82	4.90E-01	0.78	3.70E-01
DHTKD1	209916_at	0.62	3.00E-04	0.87	5.90E-01	0.74	2.80E-01
GTPBP2	221050_s_at	0.64	6.50E-04	0.80	3.90E-01	0.87	6.30E-01
TSKS	220544_at	0.66	2.00E-03	0.94	8.10E-01	0.68	2.20E-01
AIM1L	220290_at	0.64	9.10E-04	0.76	3.00E-01	0.95	8.40E-01
NUP62	207740_s_at	0.59	4.90E-05	0.77	3.20E-01	0.95	8.70E-01
C21orf45	219004_s_at	0.63	4.10E-04	1.30	4.10E-01	0.94	8.50E-01
TYRO3	211431_s_at	0.69	5.60E-03	0.84	5.20E-01	0.92	7.70E-01
SSX3	215881_x_at	0.63	9.20E-04	0.94	8.20E-01	0.8	4.80E-01
MTDH	212251_at	0.70	6.90E-03	1.14	6.40E-01	0.9	7.30E-01
P2RY6	208373_s_at	0.67	2.90E-03	0.91	7.30E-01	0.89	6.80E-01
C1orf38	207571_x_at	0.70	1.70E-02	0.95	8.60E-01	0.82	5.40E-01
CR1	217552_x_at	0.65	8.70E-04	0.98	9.40E-01	1.22	5.00E-01
ZNF710	39891_at	0.66	1.90E-03	0.90	6.90E-01	0.92	7.60E-01
LRP8	208433_s_at	0.60	8.50E-04	0.98	9.60E-01	0.78	5.70E-01
SLC35A2	207440_at	0.58	5.00E-05	0.97	9.20E-01	1.11	7.10E-01
SRPK3	206216_at	0.70	8.10E-03	1.02	9.50E-01	0.89	6.80E-01
PPP3R1	204506_at	0.72	1.50E-02	0.90	7.00E-01	1	9.90E-01
FCGR3B	204007_at	0.69	4.60E-03	0.97	9.20E-01	1.07	8.00E-01
CD68	203507_at	0.62	3.10E-04	1.01	9.80E-01	0.99	9.70E-01
STAU1	208948_s_at	1.46	4.90E-03	1.95	9.40E-03	1.76	4.40E-02
EXOSC7	212627_s_at	1.68	1.10E-04	1.83	2.20E-02	1.71	5.50E-02
BCAP31	200837_at	1.59	8.10E-04	1.81	3.40E-02	1.49	1.60E-01
GNB2L1	200651_at	1.75	1.80E-05	2.15	3.20E-03	2.07	8.70E-03
ADORA2B	205891_at	1.58	7.30E-03	2.17	2.90E-02	2.87	6.60E-03
ABHD5	213935_at	1.28	4.60E-02	1.92	1.20E-02	1.92	2.00E-02
MAP2K5	211370_s_at	1.37	2.20E-02	1.84	3.40E-02	1.88	3.70E-02
KCNG1	214595_at	1.88	1.20E-03	2.78	2.30E-02	1.87	1.40E-01
CAMSAP1	212711_at	1.32	3.50E-02	1.84	1.80E-02	1.11	7.20E-01
EIF3K	221494_x_at	1.45	8.00E-03	1.59	9.30E-02	1.86	3.20E-02
CD55	201926_s_at	1.35	2.80E-02	1.99	1.30E-01	1.93	2.30E-02
CARHSP1	218384_at	1.63	2.70E-03	1.16	5.80E-01	1.01	9.80E-01
GSK3B	209945_s_at	2.02	6.10E-07	1.49	1.30E-01	1.55	1.30E-01
CA9	205199_at	1.40	2.30E-02	2.29	5.30E-03	1.99	2.90E-02
PML	206503_x_at	1.48	1.30E-02	2.16	3.70E-02	2.11	6.20E-02
TXN	216609_at	1.54	1.50E-03	2.16	3.80E-03	1.32	3.30E-01
HCFC1R1	218537_at	1.51	3.60E-03	1.19	5.70E-01	1.55	1.60E-01
NDUFC1	203478_at	1.42	1.00E-02	1.35	2.70E-01	1.19	5.50E-01
CETN3	209662_at	1.43	7.00E-03	1.12	6.60E-01	1.23	4.60E-01
GRHPR	214864_s_at	1.31	4.60E-03	1.62	6.20E-02	1.59	1.00E-01

ZNF593	204175_at	1.30	4.90E-02	0.92	7.40E-01	1.19	5.50E-01
DYNC1LI2	203590_at	1.84	2.60E-04	1.68	6.70E-02	1.62	1.60E-01
CALM1	200655_s_at	1.85	5.80E-06	1.49	1.30E-01	1.47	1.80E-01
C7orf10	219655_at	1.66	2.60E-04	1.40	2.00E-01	1.32	3.20E-01
C16orf80	217957_at	1.82	9.00E-04	2.38	5.60E-02	1.07	8.50E-01
HNRNPAB	201277_s_at	1.63	4.80E-04	1.67	6.00E-02	1.02	9.50E-01
PRRG1	205618_at	1.59	6.80E-03	1.44	2.40E-01	1.79	1.10E-01
EIF3B	211501_s_at	1.57	4.00E-03	1.33	3.30E-01	1.48	2.20E-01
PGK1	217356_s_at	1.55	8.60E-03	1.63	1.70E-01	1.32	3.90E-01
HIF1A	200989_at	1.56	7.60E-03	0.87	6.60E-01	2.15	1.30E-01
ETFA	201931_at	1.65	2.10E-04	1.30	3.50E-01	1.21	5.20E-01

ER- breast cancer						Sorting	
RFS		DMFS		OS			
HR	P	HR	P	HR	P	Avg P	Count of Sig. p val
0.45	6.50E-13	0.44	1.10E-04	0.55	1.10E-01	1.85E-02	5
0.59	2.40E-06	0.51	8.50E-04	0.59	2.00E-02	4.90E-03	6
0.8	5.00E-02	0.75	1.70E-01	0.67	9.60E-02	5.69E-02	4
0.73	5.90E-03	0.65	4.10E-02	0.88	6.00E-01	1.42E-01	4
0.58	4.30E-05	0.45	2.20E-04	0.36	8.40E-06	7.72E-04	6
0.58	3.50E-06	0.65	3.50E-02	0.68	9.60E-02	2.36E-02	5
0.65	1.40E-04	0.59	1.20E-02	0.73	2.10E-01	5.11E-02	4
0.67	4.40E-04	0.67	7.10E-02	0.82	4.20E-01	8.74E-02	4
0.58	1.70E-06	0.58	8.70E-03	0.6	2.30E-02	6.88E-03	6
0.57	4.70E-07	0.66	4.20E-02	0.74	2.00E-01	5.82E-02	4
0.53	1.50E-08	0.42	1.00E-05	0.58	1.40E-02	1.13E-01	5
0.62	3.20E-05	0.69	6.80E-02	0.68	8.40E-02	2.88E-02	4
0.67	3.00E-04	0.74	1.40E-01	0.79	3.00E-01	8.32E-02	4
0.65	1.00E-04	0.6	1.40E-02	0.64	4.90E-02	1.85E-02	6
0.55	5.40E-08	0.67	5.10E-02	0.64	5.10E-02	3.68E-02	5
0.6	6.60E-06	0.64	3.20E-02	0.82	4.20E-01	1.44E-01	4
0.62	5.90E-05	0.64	5.60E-02	0.69	1.50E-01	7.14E-02	4
0.73	7.30E-03	0.58	1.10E-02	0.79	3.20E-01	1.15E-01	4
0.67	3.20E-03	0.69	7.70E-02	0.69	1.10E-01	4.90E-02	4
0.64	9.50E-05	0.58	6.70E-03	0.54	6.50E-03	4.32E-02	4
0.63	1.80E-04	0.67	1.00E-01	0.93	7.80E-01	2.05E-01	3
1.01	9.50E-01	0.56	4.90E-03	0.63	4.40E-02	3.12E-01	3
0.82	1.00E-01	0.79	2.50E-01	0.86	5.20E-01	1.51E-01	3
0.64	5.10E-05	0.81	3.00E-01	1.05	8.50E-01	2.67E-01	3
0.75	1.40E-02	0.81	3.00E-01	0.74	1.90E-01	9.92E-02	3
0.76	1.40E-02	0.76	1.90E-01	1.08	7.40E-01	2.46E-01	3
0.6	3.90E-06	0.69	6.80E-02	0.86	5.30E-01	1.25E-01	3
0.69	1.80E-03	0.7	9.40E-02	0.87	5.50E-01	2.37E-01	3
0.71	2.20E-03	0.77	2.00E-01	0.88	5.80E-01	1.62E-01	3
0.65	1.60E-04	0.74	1.40E-01	1.02	9.40E-01	3.14E-01	3

0.76	1.40E-02	0.83	3.70E-01	0.94	7.90E-01	2.85E-01	3
0.61	7.70E-05	0.84	4.60E-01	1.07	8.20E-01	2.41E-01	3
0.69	1.90E-03	1	1.00E+00	0.87	5.70E-01	2.96E-01	3
0.74	8.80E-03	0.9	6.20E-01	0.89	6.00E-01	2.34E-01	2
0.65	9.50E-05	0.8	2.70E-01	0.83	4.30E-01	1.50E-01	2
0.73	5.90E-03	0.92	6.70E-01	1.05	8.30E-01	2.96E-01	2
0.66	2.10E-04	0.88	5.30E-01	0.85	5.00E-01	2.18E-01	2
0.63	8.40E-05	0.72	1.40E-01	0.82	4.00E-01	1.40E-01	2
0.83	1.00E-01	0.87	4.90E-01	1.18	4.60E-01	2.33E-01	2
0.58	1.20E-06	0.8	2.70E-01	1.06	8.10E-01	2.47E-01	2
0.63	1.30E-04	0.75	2.10E-01	0.74	2.30E-01	1.67E-01	2
0.72	5.90E-03	0.97	8.70E-01	1.2	4.40E-01	3.18E-01	2
0.7	1.70E-03	0.78	2.50E-01	0.9	6.50E-01	2.52E-01	2
0.62	5.20E-05	0.88	5.40E-01	1.25	3.30E-01	2.55E-01	2
0.72	4.10E-03	0.71	1.20E-01	1.12	6.70E-01	2.66E-01	2
0.71	4.60E-03	1.03	8.80E-01	0.87	5.60E-01	3.85E-01	2
0.64	7.00E-05	0.91	6.40E-01	0.99	9.60E-01	4.12E-01	2
0.67	2.90E-04	1.14	5.30E-01	1.18	4.60E-01	3.35E-01	2
0.63	9.30E-05	0.92	7.20E-01	0.75	2.90E-01	3.40E-01	2
0.55	2.30E-07	0.69	7.10E-02	0.83	4.20E-01	2.72E-01	2
0.6	5.10E-06	0.88	5.50E-01	1.05	8.20E-01	4.27E-01	2
0.65	1.20E-04	1.16	5.40E-01	0.96	8.80E-01	4.47E-01	2
0.77	1.90E-02	1.24	3.10E-01	1.26	3.10E-01	3.22E-01	2
0.65	2.20E-04	1.02	9.40E-01	0.91	7.10E-01	4.92E-01	2
0.7	1.80E-03	1.08	7.20E-01	1.04	8.80E-01	4.96E-01	2
0.63	4.10E-05	0.73	1.20E-01	0.8	3.50E-01	3.14E-01	2
0.59	9.40E-06	0.73	1.50E-01	0.75	2.30E-01	3.00E-01	2
0.65	1.30E-04	0.85	4.20E-01	1.03	9.00E-01	4.60E-01	2
0.69	8.60E-04	0.99	9.70E-01	1.18	4.90E-01	4.85E-01	2
0.58	2.20E-05	1.05	8.70E-01	0.83	5.70E-01	4.95E-01	2
0.67	3.20E-04	1.08	7.20E-01	1.11	6.80E-01	5.05E-01	2
0.76	1.30E-02	1.16	4.70E-01	1.14	5.70E-01	4.49E-01	2
0.69	1.20E-03	0.83	3.70E-01	1.01	9.60E-01	5.06E-01	2
0.61	8.30E-06	0.7	7.80E-02	0.86	5.00E-01	3.84E-01	2
0.68	7.00E-04	1.1	6.30E-01	1.14	5.80E-01	5.27E-01	2
1.46	1.20E-03	1.78	5.20E-03	1.88	7.00E-03	1.20E-02	6
1.56	8.50E-05	1.4	5.20E-02	1.64	2.90E-02	2.64E-02	6
1.59	9.90E-05	1.83	6.20E-03	1.66	3.10E-02	3.87E-02	5
1.68	3.30E-06	1.42	8.70E-02	1.33	2.10E-01	5.15E-02	4
1.14	3.00E-01	1.43	1.20E-01	1.96	1.30E-02	7.93E-02	4
1.15	2.10E-01	1.2	3.80E-01	1.56	5.00E-02	1.20E-01	4
1.25	5.70E-02	1.47	9.10E-02	1.25	4.10E-01	1.09E-01	4
1.56	2.20E-03	1.83	2.60E-02	1.58	1.20E-01	5.21E-02	4
1.33	1.20E-02	1.88	1.90E-03	1.28	2.80E-01	1.78E-01	4
1.35	1.00E-02	1.22	3.40E-01	1.29	2.70E-01	1.26E-01	3
1.2	1.20E-01	1.6	3.60E-02	1.36	2.00E-01	8.95E-02	3

1.7	9.30E-06	1.59	3.30E-02	1.78	5.50E-03	2.67E-01	3
1.87	2.50E-07	1.55	3.90E-02	1.16	5.30E-01	1.38E-01	3
1.11	3.90E-01	1.44	8.90E-02	1.44	1.20E-01	1.09E-01	3
1.23	9.70E-02	1.17	5.10E-01	1.04	8.70E-01	2.65E-01	2
1.38	4.70E-03	1.43	8.80E-02	1.43	8.30E-02	8.52E-02	2
1.47	6.60E-04	1.29	2.60E-01	1.35	1.00E-02	1.67E-01	2
1.32	1.30E-02	0.94	7.80E-01	1	9.90E-01	4.36E-01	2
1.51	3.20E-04	1.06	7.70E-01	1.39	1.50E-01	3.41E-01	2
1.18	1.50E-01	1.14	5.40E-01	1.42	1.50E-01	1.68E-01	1
1.2	1.00E-01	0.81	3.10E-01	1.1	6.70E-01	4.03E-01	1
1.77	4.40E-06	1.66	3.10E-02	1.52	1.40E-01	6.64E-02	3
1.92	3.80E-08	1.49	6.00E-02	1.61	5.80E-02	7.13E-02	3
1.72	4.90E-06	1.63	1.90E-02	1.34	2.00E-01	1.23E-01	3
1.5	3.10E-03	1.35	2.50E-01	0.92	7.50E-01	3.18E-01	3
1.68	2.20E-05	1.89	5.00E-03	1.24	3.50E-01	2.28E-01	3
1.33	3.00E-02	1.3	2.60E-01	1.53	1.10E-01	1.26E-01	2
1.4	9.30E-03	1.32	2.30E-01	1.36	2.20E-01	1.69E-01	2
1.36	2.80E-02	1.4	2.10E-01	1.24	4.20E-01	2.04E-01	2
1.39	2.30E-02	0.73	2.30E-01	2.03	1.20E-01	1.95E-01	2
1.59	1.50E-04	1.18	4.80E-01	1.04	8.70E-01	3.70E-01	2

Top 4 genes in the module

	BLBC	
RFS	-2.33	4.30E-11
DMFS	**-3.57**	**6.30E-07**
OS	**-3.57**	**5.70E-06**
	ER negative	
RFS	-2.22	6.50E-13
DMFS	**-2.70**	**7.80E-07**
OS	**-2.78**	**5.10E-06**

Rest of the genes in module

	BLBC	
RFS	-2.78	1.90E-14
DMFS	-2.63	1.30E-04
OS	-2.63	4.60E-04
	ER negative	
RFS	-2.38	3.30E-15
DMFS	-2.08	2.40E-04
OS	-2.13	7.10E-04

Using top 4 genes from first module and top 3 genes from the second module

	BLBC	
RFS	**-2.86**	**1.10E-16**
DMFS	**-5.00**	**7.80E-10**
OS	**-4.76**	**1.50E-08**
	ER negative	
RFS	**-2.50**	**1.00E-16**
DMFS	**-3.13**	**2.40E-08**

BTN2A2
IGHA1
MTMR7
ME1
NCRNA00171
SMPDL3B
ERC2

*Only top 3 genes as a module had sig p value **both** DMFS and OS in **both** BLBC and ER-*

	BLBC	
RFS	-2.13	1.60E-08
DMFS	*-1.89*	*1.90E-02*
OS	*-2.63*	*9.60E-04*
	ER negative	
RFS	-1.82	1.20E-07
DMFS	*-1.54*	*4.50E-02*
OS	*-1.75*	*2.00E-02*

	BLBC	
RFS	-2.04	2.60E-07
DMFS	1.06	0.86
OS	0.71	0.25
	ER negative	
RFS	-2.00	3.40E-09
DMFS	0.95	0.82
OS	0.89	0.66

OS	**-3.45**	**1.20E-07**

	BLBC	
RFS	2.38	9.70E-10
DMFS	3.23	5.00E-05
OS	2.98	2.00E-04
	ER negative	
RFS	2.1	2.80E-10
DMFS	1.94	1.50E-03
OS	1.86	6.70E-03

	BLBC	
RFS	1.51	2.10E-03
DMFS	1.48	1.40E-02
OS	1.58	0.1
	ER negative	
RFS	1.4	3.30E-03
DMFS	1.18	0.41
OS	1.27	0.3

	BLBC	
RFS	2	3.10E-04
DMFS	1.64	0.19
OS	1.18	0.66
	ER negative	
RFS	2.06	1.60E-05
DMFS	1.93	0.07
OS	1.38	0.38

	BLBC	
RFS	2.79	2.00E-14
DMFS	4.1	6.00E-07
OS	3.81	5.00E-06
	ER negative	
RFS	2.41	2.00E-14
DMFS	2.41	2.00E-05
OS	2.37	1.50E-04

STAU1
EXOSC7
BCAP31
GNB2L1
ADORA2B
ABHD5
MAP2K5
KCNG1
CAMSAP1
EIF3K
CD55
CARHSP1
GSK3B
CA9
PML
TXN
HCFC1R1
NDUFC1
CETN3
GRHPR
ZNF593

Table 15. Class comparison of the global gene expression profiles of high TN score BLBC tumors to low TN score BLBC tumors in the ROCK dataset (highlighted probe set indicates common in high TN score BLBC and ER- breast tumours and bold probe set indicates common and prognostic in BLBC and ER- breast cancer).

Parametric p-value	FDR	Permutation p-value	Fold-change for high TN score vs. Low TN score	Probe Set	Symbol	Name	EntrezID
4.22E-05	0.0174	< 1e-07	2.9	**203108_at**	GPRC5A	G protein-coupled receptor, family C, group 5, member A	9052
4.60E-04	0.0436	0.0005	2.5	**213693_s_at**	MUC1	mucin 1, cell surface associated	4582
1.03E-04	0.0249	< 1e-07	2.4	219956_at	GALNT6	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6)	11226
1.88E-02	0.221	0.0172	2.3	204268_at	S100A2	S100 calcium binding protein A2	6273
2.35E-02	0.244	0.0239	2.3	204351_at	S100P	S100 calcium binding protein P	6286
7.09E-03	0.137	0.0054	2.1	**203021_at**	SLPI	secretory leukocyte peptidase inhibitor	6590
5.95E-03	0.126	0.0046	2.1	**207847_s_at**	MUC1	mucin 1, cell surface associated	4582
1.83E-03	0.0746	0.002	2.1	**202489_s_at**	FXYD3	FXYD domain containing ion transport regulator 3	5349
6.71E-05	0.0214	< 1e-07	2.0	**201468_s_at**	NQO1	NAD(P)H dehydrogenase, quinone 1	1728
2.47E-05	0.0167	< 1e-07	2.0	204818_at	HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2	3294
1.76E-03	0.0737	0.0019	2.0	**212444_at**	GPRC5A	G protein-coupled receptor, family C, group 5, member A	9052
2.56E-03	0.085	0.0028	1.9	218309_at	CAMK2N1	calcium/calmodulin-dependent protein kinase II inhibitor 1	55450
7.04E-04	0.0509	0.0004	1.9	209803_s_at	PHLDA2	pleckstrin homology-like domain, family A, member 2	7262
5.08E-05	0.0195	< 1e-07	1.9	**201286_at**	SDC1	syndecan 1	6382
1.65E-03	0.0716	0.0013	1.9	210519_s_at	NQO1	NAD(P)H dehydrogenase, quinone 1	1728
1.23E-02	0.18	0.012	1.9	209160_at	AKR1C3	aldo-keto reductase family 1, member C3	8644

4.27E-02	0.321	0.0403	1.8	**209800_at**	KRT16	keratin 16	3868
1.13E-03	0.0602	0.0008	1.8	201951_at	ALCAM	activated leukocyte cell adhesion molecule	214
5.40E-04	0.0475	0.0003	1.8	210239_at	IRX5	iroquois homeobox 5	10265
2.92E-03	0.09	0.003	1.8	221024_s_at	SLC2A10	solute carrier family 2 (facilitated glucose transporter), member 10	81031
1.92E-03	0.0755	0.0018	1.8	**203060_s_at**	PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	9060
2.71E-02	0.261	0.0282	1.8	218211_s_at	MLPH	melanophilin	79083
4.52E-03	0.111	0.0044	1.8	**201242_s_at**	ATP1B1	ATPase, Na+/K+ transporting, beta 1 polypeptide	481
1.19E-03	0.0614	0.0009	1.8	220230_s_at	CYB5R2	cytochrome b5 reductase 2	51700
2.38E-03	0.0823	0.0028	1.8	218959_at	HOXC10	homeobox C10	3226
5.29E-03	0.12	0.0039	1.8	219959_at	MOCOS	molybdenum cofactor sulfurase	55034
2.58E-03	0.0851	0.0022	1.8	219580_s_at	TMC5	transmembrane channel-like 5	79838
1.98E-03	0.076	0.0012	1.8	**204595_s_at**	STC1	stanniocalcin 1	6781
3.06E-03	0.0914	0.0022	1.8	200862_at	DHCR24	24-dehydrocholesterol reductase	1718
3.44E-02	0.291	0.0339	1.7	211657_at	CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	4680
8.94E-03	0.154	0.0078	1.7	210652_s_at	TTC39A	tetratricopeptide repeat domain 39A	22996
2.16E-03	0.0787	0.0021	1.7	201467_s_at	NQO1	NAD(P)H dehydrogenase, quinone 1	1728
2.45E-02	0.248	0.0218	1.7	209699_x_at	AKR1C2	aldo-keto reductase family 1, member C2	1646
1.02E-02	0.165	0.0111	1.7	203215_s_at	MYO6	myosin VI	4646
8.95E-04	0.0547	0.0008	1.7	204284_at	PPP1R3C	protein phosphatase 1, regulatory subunit 3C	5507
4.66E-02	0.332	0.0457	1.7	203453_at	SCNN1A	sodium channel, non-voltage-gated 1 alpha subunit	6337
4.38E-04	0.0435	0.0005	1.7	206326_at	GRP	gastrin-releasing peptide	2922
2.35E-03	0.0821	0.0025	1.7	219127_at	PRR15L	proline rich 15-like	79170
3.86E-02	0.306	0.0346	1.7	214580_x_at	KRT6C		
1.94E-03	0.0756	0.0018	1.7	203058_s_at	PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	9060
8.03E-03	0.146	0.0067	1.7	**209373_at**	MALL	mal, T-cell differentiation protein-like	7851
5.88E-04	0.0488	0.0006	1.7	**213285_at**	TMEM30B	transmembrane protein 30B	161291
8.97E-03	0.154	0.0098	1.7	**205110_s_at**	FGF13	fibroblast growth factor 13	2258

4.44E-04	0.0435	0.0005	1.7	221042_s_at	CLMN	calmin (calponin-like, transmembrane)	79789
1.44E-02	0.193	0.0151	1.7	214079_at	DHRS2	dehydrogenase/reductase (SDR family) member 2	10202
3.70E-02	0.302	0.0377	1.7	204913_s_at	SOX11	SRY (sex determining region Y)-box 11	6664
9.09E-04	0.0547	0.0014	1.6	209260_at	SFN	stratifin	2810
5.49E-03	0.121	0.0057	1.6	203946_s_at	ARG2	arginase, type II	384
1.01E-03	0.0571	0.0012	1.6	221589_s_at	ALDH6A1	aldehyde dehydrogenase 6 family, member A1	4329
8.30E-04	0.0537	0.0005	1.6	222258_s_at	SH3BP4	SH3-domain binding protein 4	23677
8.16E-03	0.147	0.0083	1.6	205968_at	KCNS3	potassium voltage-gated channel, delayed-rectifier, subfamily S, member 3	3790
8.44E-03	0.15	0.0094	1.6	203180_at	ALDH1A3	aldehyde dehydrogenase 1 family, member A3	220
2.13E-03	0.0786	0.003	1.6	201596_x_at	KRT18	keratin 18	3875
1.60E-03	0.0707	0.0022	1.6	219232_s_at	EGLN3	egl nine homolog 3 (C. elegans)	112399
4.28E-04	0.0435	0.0003	1.6	208710_s_at	AP3D1	adaptor-related protein complex 3, delta 1 subunit	8943
9.84E-04	0.0569	0.0009	1.6	218273_s_at	PDP1	pyruvate dehyrogenase phosphatase catalytic subunit 1	54704
1.83E-04	0.0306	0.0003	1.6	209945_s_at	GSK3B	glycogen synthase kinase 3 beta	2932
3.77E-03	0.101	0.0046	1.6	203407_at	PPL	periplakin	5493
3.03E-02	0.275	0.0263	1.6	207802_at	CRISP3	cysteine-rich secretory protein 3	10321
1.65E-02	0.205	0.0179	1.6	205594_at	ZNF652	zinc finger protein 652	22834
1.72E-02	0.21	0.0165	1.6	210372_s_at	TPD52L1	tumor protein D52-like 1	7164
2.32E-02	0.242	0.0222	1.6	203803_at	PCYOX1	prenylcysteine oxidase 1	51449
2.10E-02	0.233	0.0239	1.6	209875_s_at	SPP1	secreted phosphoprotein 1	6696
6.66E-03	0.133	0.0077	1.6	213577_at	SQLE	squalene epoxidase	6713
2.54E-05	0.0167	< 1e-07	1.6	201926_s_at	CD55	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	1604
7.46E-03	0.141	0.0065	1.6	213397_x_at	RNASE4		
2.15E-03	0.0786	0.0025	1.6	219468_s_at	CUEDC1	CUE domain containing 1	404093
3.26E-03	0.094	0.0034	1.6	36711_at	MAFF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog F (avian)	23764
3.75E-03	0.101	0.0039	1.6	205709_s_at	CDS1	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 1	1040
1.72E-	0.21	0.016	1.6	220161_	EPB41	erythrocyte membrane	54566

02		4		s_at	L4B	protein band 4.1 like 4B	
2.33E-03	0.0816	0.0026	1.5	202068_s_at	LDLR	low density lipoprotein receptor	3949
1.80E-03	0.074	0.0009	1.5	201849_at	BNIP3	BCL2/adenovirus E1B 19kDa interacting protein 3	664
3.09E-03	0.0916	0.0033	1.5	219496_at	SOWAHC	sosondowah ankyrin repeat domain family member C	65124
6.63E-05	0.0214	< 1e-07	1.5	203042_at	LAMP2	lysosomal-associated membrane protein 2	3920
2.70E-03	0.0874	0.0027	1.5	222158_s_at	DESI2	desumoylating isopeptidase 2	51029
9.04E-03	0.155	0.0091	1.5	204044_at	QPRT	quinolinate phosphoribosyltransferase	23475
3.81E-05	0.0173	< 1e-07	1.5	204175_at	ZNF593	zinc finger protein 593	51042
5.28E-04	0.0472	0.0006	1.5	221526_x_at	PARD3	par-3 partitioning defective 3 homolog (C. elegans)	56288
6.97E-03	0.136	0.0074	1.5	202984_s_at	BAG5	BCL2-associated athanogene 5	9529
6.84E-05	0.0214	0.0002	1.5	219377_at	GAREM	GRB2 associated, regulator of MAPK1	64762
1.49E-04	0.0299	0.0003	1.5	202733_at	P4HA2	prolyl 4-hydroxylase, alpha polypeptide II	8974
5.37E-03	0.12	0.005	1.5	221577_x_at	GDF15	growth differentiation factor 15	9518
5.73E-04	0.0486	0.0004	1.5	209146_at	MSMO1	methylsterol monooxygenase 1	6307
6.05E-05	0.0201	0.0001	1.5	202929_s_at	DDT	D-dopachrome tautomerase	1652
7.09E-03	0.137	0.0083	1.5	213506_at	F2RL1	coagulation factor II (thrombin) receptor-like 1	2150
4.00E-07	0.00223	< 1e-07	1.5	202562_s_at	C14orf1	chromosome 14 open reading frame 1	11161
8.53E-04	0.054	0.0008	1.5	202314_at	CYP51A1		
1.47E-03	0.0676	0.0014	2.3	212094_at	PEG10	paternally expressed 10	23089
2.70E-03	0.0874	0.0018	2.2	219225_at	PGBD5	piggyBac transposable element derived 5	79605
3.43E-02	0.29	0.0339	2.0	213680_at	KRT6B	keratin 6B	3854
5.29E-03	0.12	0.0044	1.9	202286_s_at	TACSTD2	tumor-associated calcium signal transducer 2	4070
1.41E-03	0.0667	0.0015	1.9	202669_s_at	EFNB2	ephrin-B2	1948
9.41E-03	0.159	0.0091	1.9	204750_s_at	DSC2	desmocollin 2	1824
2.18E-02	0.236	0.0207	1.9	221690_s_at	NLRP2	NLR family, pyrin domain containing 2	55655
7.15E-04	0.0512	0.0008	1.9	211538_s_at	HSPA2	heat shock 70kDa protein 2	3306
4.95E-04	0.046	0.0002	1.8	206125_s_at	KLK8	kallikrein-related peptidase 8	11202
9.06E-03	0.155	0.009	1.8	205428_s_at	CALB2	calbindin 2	794
3.04E-02	0.275	0.034	1.8	202376_at	SERPINA3	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	12

2.65E-02	0.26	0.0262	1.8	205595_at	DSG3	desmoglein 3	1830
1.58E-03	0.0704	0.0014	1.8	204614_at	SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	5055
2.67E-04	0.0353	0.0003	1.8	201324_at	EMP1	epithelial membrane protein 1	2012
3.19E-02	0.281	0.0378	1.7	203628_at	IGF1R	insulin-like growth factor 1 receptor	3480
3.61E-03	0.0993	0.0036	1.7	214595_at	KCNG1	potassium voltage-gated channel, subfamily G, member 1	3755
1.91E-02	0.223	0.0205	1.7	209602_s_at	GATA3	GATA binding protein 3	2625
1.98E-03	0.076	0.0025	1.7	204059_s_at	ME1	malic enzyme 1, NADP(+)-dependent, cytosolic	4199
1.94E-04	0.0317	0.0002	1.7	205809_s_at	WASL	Wiskott-Aldrich syndrome-like	8976
1.15E-05	0.0128	< 1e-07	1.7	204032_at	BCAR3	breast cancer anti-estrogen resistance 3	8412
3.60E-03	0.0993	0.0029	1.7	219995_s_at	ZNF750	zinc finger protein 750	79755
1.35E-03	0.0655	0.0016	1.7	212451_at	SECISBP2L	SECIS binding protein 2-like	9728
1.20E-04	0.0275	< 1e-07	1.7	203566_s_at	AGL	amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase	178
1.30E-03	0.0644	0.0019	1.7	204058_at	ME1	malic enzyme 1, NADP(+)-dependent, cytosolic	4199
1.49E-02	0.196	0.0155	1.6	218678_at	NES	nestin	10763
4.14E-02	0.315	0.0456	1.6	208900_s_at	TOP1	topoisomerase (DNA) I	7150
1.18E-03	0.0614	0.0013	1.6	208610_s_at	SRRM2	serine/arginine repetitive matrix 2	23524
1.00E-07	0.00223	< 1e-07	1.6	213526_s_at	LIN37	lin-37 homolog (C. elegans)	55957
3.34E-02	0.287	0.034	1.6	209581_at	PLA2G16	phospholipase A2, group XVI	11145
1.21E-02	0.179	0.0113	1.6	218858_at	DEPTOR	DEP domain containing MTOR-interacting protein	64798
2.63E-02	0.259	0.0258	1.6	204288_s_at	SORBS2	sorbin and SH3 domain containing 2	8470
3.97E-03	0.104	0.0038	1.6	204688_at	SGCE	sarcoglycan, epsilon	8910
1.86E-02	0.22	0.0194	1.6	217996_at	PHLDA1	pleckstrin homology-like domain, family A, member 1	22822
2.13E-04	0.0328	0.0003	1.6	209254_at	KLHDC10	kelch domain containing 10	23008
7.78E-03	0.144	0.0088	1.6	219263_at	RNF128	ring finger protein 128, E3 ubiquitin protein ligase	79589
1.05E-02	0.167	0.0083	1.6	219476_at	C1orf116	chromosome 1 open reading frame 116	79098
3.45E-02	0.291	0.0354	1.6	202600_s_at	NRIP1	nuclear receptor interacting protein 1	8204
2.73E-02	0.263	0.0249	1.6	209126_x_at	KRT6B	keratin 6B	3854
3.04E-03	0.0914	0.0031	1.6	206421_s_at	SERPINB7	serpin peptidase inhibitor, clade B (ovalbumin),	8710

						member 7	
3.35E-02	0.288	0.0341	1.6	209604_s_at	GATA3	GATA binding protein 3	2625
1.59E-03	0.0704	0.0016	1.6	212775_at	OBSL1	obscurin-like 1	23363
2.80E-02	0.266	0.026	1.6	205440_s_at	NPY1R	neuropeptide Y receptor Y1	4886
3.12E-02	0.278	0.033	1.6	204508_s_at	CA12	carbonic anhydrase XII	771
3.17E-02	0.28	0.0298	1.6	209301_at	CA2	carbonic anhydrase II	760
3.31E-02	0.286	0.0322	1.6	201860_s_at	PLAT	plasminogen activator, tissue	5327
8.35E-04	0.0537	0.0011	1.6	212294_at	GNG12	guanine nucleotide binding protein (G protein), gamma 12	55970
5.35E-04	0.0475	0.0003	1.6	201325_s_at	EMP1	epithelial membrane protein 1	2012
3.90E-02	0.307	0.0387	1.6	212992_at	AHNAK2	AHNAK nucleoprotein 2	113146
5.98E-03	0.126	0.0055	1.6	201996_s_at	SPEN	spen homolog, transcriptional regulator (Drosophila)	23013
4.43E-03	0.109	0.004	1.6	207480_s_at	MEIS2	Meis homeobox 2	4212
6.59E-03	0.133	0.0068	1.6	202454_s_at	ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	2065
4.22E-03	0.107	0.0048	1.6	212492_s_at	KDM4B	lysine (K)-specific demethylase 4B	23030
2.87E-02	0.268	0.0296	1.6	204748_at	PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	5743
1.29E-02	0.183	0.0121	1.6	206307_s_at	FOXD1	forkhead box D1	2297
2.41E-02	0.247	0.0248	1.6	213110_s_at	COL4A5	collagen, type IV, alpha 5	1287
8.36E-04	0.0537	0.001	1.6	212634_at	UFL1	UFM1-specific ligase 1	23376
9.30E-03	0.157	0.0101	1.6	219681_s_at	RAB11FIP1	RAB11 family interacting protein 1 (class I)	80223
8.13E-03	0.147	0.0103	1.6	203319_s_at	ZNF148	zinc finger protein 148	7707
6.73E-04	0.0508	0.0007	1.6	204066_s_at	AGAP1	ArfGAP with GTPase domain, ankyrin repeat and PH domain 1	116987
3.68E-03	0.0999	0.004	1.6	219298_at	ECHDC3	enoyl CoA hydratase domain containing 3	79746
5.77E-03	0.124	0.0043	1.6	209720_s_at	SERPINB3	serpin peptidase inhibitor, clade B (ovalbumin), member 3	6317
3.05E-02	0.275	0.0294	1.6	210467_x_at	MAGEA12	melanoma antigen family A, 12	4111
1.88E-03	0.0755	0.0017	1.5	204029_at	CELSR2	cadherin, EGF LAG seven-pass G-type receptor 2	1952
1.09E-02	0.17	0.0114	1.5	204779_s_at	HOXB7	homeobox B7	3217
2.63E-	0.259	0.028	1.5	204686_	IRS1	insulin receptor substrate 1	3667

02		8		at			
6.28E-04	0.0493	0.001	1.5	209123_at	QDPR	quinoid dihydropteridine reductase	5860
1.04E-03	0.058	0.0009	1.5	212417_at	SCAMP1	secretory carrier membrane protein 1	9522
2.68E-03	0.0873	0.002	1.5	209719_x_at	SERPINB3	serpin peptidase inhibitor, clade B (ovalbumin), member 3	6317
9.14E-04	0.0547	0.0002	1.5	211906_s_at	SERPINB4	serpin peptidase inhibitor, clade B (ovalbumin), member 4	6318
7.11E-04	0.0511	0.0007	1.5	219073_s_at	OSBPL10	oxysterol binding protein-like 10	114884
2.98E-02	0.273	0.0313	1.5	209488_s_at	RBPMS	RNA binding protein with multiple splicing	11030
6.36E-03	0.13	0.0073	1.5	203542_s_at	KLF9	Kruppel-like factor 9	687
3.76E-03	0.101	0.004	1.5	203780_at	MPZL2	myelin protein zero-like 2	10205
2.87E-02	0.268	0.0318	1.5	209443_at	SERPINA5	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 5	5104
2.95E-03	0.09	0.0046	1.5	210612_s_at	SYNJ2	synaptojanin 2	8871
9.59E-03	0.16	0.0103	1.5	213030_s_at	PLXNA2	plexin A2	5362
1.19E-02	0.178	0.0124	1.5	218435_at	DNAJC15	DnaJ (Hsp40) homolog, subfamily C, member 15	29103
1.24E-02	0.181	0.0136	1.5	202998_s_at	LOXL2	lysyl oxidase-like 2	4017
2.27E-04	0.0336	0.0001	1.5	218253_s_at	EIF2D	eukaryotic translation initiation factor 2D	1939
6.56E-03	0.133	0.0062	1.5	203439_s_at	STC2	stanniocalcin 2	8614
3.01E-02	0.274	0.031	1.5	203929_s_at	MAPT	microtubule-associated protein tau	4137
4.26E-03	0.108	0.0038	1.5	204256_at	ELOVL6	ELOVL fatty acid elongase 6	79071
1.60E-04	0.0299	0.0003	1.5	218407_x_at	NENF	neudesin neurotrophic factor	29937
1.46E-03	0.0676	0.0019	1.5	221588_x_at	ALDH6A1	aldehyde dehydrogenase 6 family, member A1	4329
1.23E-05	0.013	< 1e-07	0.3	211634_x_at			
3.61E-05	0.0173	0.0001	0.3	211635_x_at			
5.90E-04	0.0488	0.001	0.3	216491_x_at	IGHM	immunoglobulin heavy constant mu	3507
4.51E-04	0.0435	0.0007	0.4	205242_at	CXCL13	chemokine (C-X-C motif) ligand 13	10563
1.65E-03	0.0716	0.0019	0.4	214768_x_at	IGKC	immunoglobulin kappa constant	3514
5.56E-04	0.0479	0.0005	0.4	203915_at	CXCL9	chemokine (C-X-C motif) ligand 9	4283
2.06E-04	0.0327	0.0006	0.4	211637_x_at			
1.66E-03	0.0717	0.0021	0.4	214777_at			

1.57E-03	0.0701	0.0019	0.4	217148_x_at			
2.77E-04	0.0358	0.0006	0.4	214916_x_at			
3.35E-05	0.0173	< 1e-07	0.4	211633_x_at	IGHG1	immunoglobulin heavy constant gamma 1 (G1m marker)	3500
2.82E-04	0.0358	0.0003	0.4	205267_at	POU2AF1	POU class 2 associating factor 1	5450
1.31E-03	0.0645	0.0018	0.4	216576_x_at			
1.22E-04	0.0277	0.0001	0.4	214973_x_at	IGHD	immunoglobulin heavy constant delta	3495
1.79E-04	0.0303	< 1e-07	0.4	217179_x_at			
2.22E-04	0.0332	0.0003	0.4	217281_x_at			
1.64E-03	0.0716	0.0011	0.4	216510_x_at			
2.50E-06	0.00587	< 1e-07	0.4	207238_s_at	PTPRC	protein tyrosine phosphatase, receptor type, C	5788
7.63E-03	0.142	0.008	0.4	205890_s_at			
9.88E-05	0.0245	< 1e-07	0.4	217235_x_at	IGLL5	immunoglobulin lambda-like polypeptide 5	1E+08
7.52E-03	0.141	0.0079	0.4	211644_x_at			
1.73E-05	0.0159	< 1e-07	0.4	207339_s_at	LTB	lymphotoxin beta (TNF superfamily, member 3)	4050
5.60E-04	0.0479	0.0007	0.4	216557_x_at			
3.98E-05	0.0173	0.0002	0.4	212588_at	PTPRC	protein tyrosine phosphatase, receptor type, C	5788
1.05E-04	0.0249	0.0001	0.4	211796_s_at			
5.16E-04	0.0464	0.0008	0.4	211650_x_at			
7.81E-05	0.0223	0.0002	0.5	204563_at	SELL	selectin L	6402
1.04E-03	0.058	0.0013	0.5	211643_x_at			
6.02E-03	0.127	0.0077	0.5	211645_x_at			
1.50E-05	0.0145	< 1e-07	0.5	215949_x_at			
8.04E-05	0.0223	0.0001	0.5	211640_x_at			
1.07E-03	0.0587	0.0011	0.5	205861_at	SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	6689
8.17E-04	0.0537	0.0006	0.5	210915_x_at	TRBC1	T cell receptor beta constant 1	28639
1.10E-02	0.171	0.0109	0.5	211122_s_at	CXCL11	chemokine (C-X-C motif) ligand 11	6373
1.51E-03	0.0688	0.0019	0.5	216207_x_at			
4.11E-04	0.0435	0.0002	0.5	219014_at	PLAC8	placenta-specific 8	51316

6.57E-03	0.133	0.0077	0.5	216560_x_at	IGLC1	immunoglobulin lambda constant 1 (Mcg marker)	3537
2.74E-03	0.088	0.0025	0.5	204439_at	IFI44L	interferon-induced protein 44-like	10964
3.00E-07	0.00223	< 1e-07	0.5	211649_x_at			
3.16E-04	0.0384	0.0002	0.5	AFFX-HUMISGF3A/M97935_MA_at	STAT1	signal transducer and activator of transcription 1, 91kDa	6772
3.65E-04	0.0414	0.0005	0.5	216541_x_at			
1.85E-03	0.075	0.0022	0.5	217227_x_at	IGLV1-44	immunoglobulin lambda variable 1-44	28823
4.80E-03	0.115	0.0067	0.5	216984_x_at			
1.41E-02	0.191	0.0147	0.5	210029_at	IDO1	indoleamine 2,3-dioxygenase 1	3620
4.18E-04	0.0435	0.0003	0.5	211881_x_at	IGLJ3	immunoglobulin lambda joining 3	28831
4.54E-04	0.0436	0.0008	0.5	205831_at	CD2	CD2 molecule	914
2.84E-04	0.0358	0.0003	0.5	206666_at	GZMK	granzyme K (granzyme 3; tryptase II)	3003
2.66E-04	0.0353	0.0003	0.5	211908_x_at	IGK@	immunoglobulin kappa locus	50802
1.30E-02	0.184	0.0148	0.5	215176_x_at	IGKC	immunoglobulin kappa constant	3514
6.25E-03	0.129	0.005	0.5	206134_at	ADAMDEC1	ADAM-like, decysin 1	27299
1.28E-05	0.013	< 1e-07	0.5	209670_at	TRAC	T cell receptor alpha constant	28755
8.47E-03	0.15	0.0098	0.5	217378_x_at			
1.68E-04	0.0299	0.0004	0.5	211868_x_at			
1.31E-02	0.184	0.0142	0.5	210163_at	CXCL11	chemokine (C-X-C motif) ligand 11	6373
1.88E-03	0.0755	0.0025	0.5	211798_x_at	IGLJ3	immunoglobulin lambda joining 3	28831
2.66E-04	0.0353	0.0002	0.5	211641_x_at			
4.18E-03	0.107	0.0034	0.5	204533_at	CXCL10	chemokine (C-X-C motif) ligand 10	3627
1.76E-03	0.0737	0.0017	0.5	214657_s_at			
3.66E-03	0.0999	0.0029	0.5	205569_at	LAMP3	lysosomal-associated membrane protein 3	27074
1.56E-02	0.2	0.0171	0.5	216401_x_at			
1.93E-05	0.0159	< 1e-07	0.5	204912_at	IL10RA	interleukin 10 receptor, alpha	3587
1.10E-04	0.0258	< 1e-07	0.5	210538_s_at	BIRC3	baculoviral IAP repeat containing 3	330
6.40E-06	0.00839	0.0001	0.5	203471_s_at	PLEK	pleckstrin	5341
2.71E-03	0.0876	0.0033	0.5	217480_x_at			

6.37E-04	0.0493	0.0006	0.5	214453_s_at	IFI44	interferon-induced protein 44	10561
8.90E-04	0.0547	0.0007	0.5	AFFX-HUMISGF3A/M97935_MB_at	STAT1	signal transducer and activator of transcription 1, 91kDa	6772
2.46E-03	0.0834	0.0028	0.5	212671_s_at			
7.63E-05	0.0223	0.0002	0.5	211639_x_at			
2.12E-04	0.0328	< 1e-07	0.5	205671_s_at	HLA-DOB	major histocompatibility complex, class II, DO beta	3112
3.29E-05	0.0173	< 1e-07	0.5	212314_at	SEL1L3	sel-1 suppressor of lin-12-like 3 (C. elegans)	23231
1.38E-03	0.066	0.0011	0.5	204891_s_at	LCK	lymphocyte-specific protein tyrosine kinase	3932
5.95E-05	0.0201	0.0001	0.5	204118_at	CD48	CD48 molecule	962
3.18E-04	0.0385	0.0004	0.5	203868_s_at	VCAM1	vascular cell adhesion molecule 1	7412
6.25E-04	0.0493	0.0009	0.5	217258_x_at	IGLV1-44	immunoglobulin lambda variable 1-44	28823
1.22E-03	0.0622	0.0014	0.5	213888_s_at	TRAF3IP3	TRAF3 interacting protein 3	80342
1.91E-03	0.0755	0.002	0.5	204279_at	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional peptidase 2)	5698
4.04E-03	0.105	0.0037	0.5	205159_at	CSF2RB	colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage)	1439
2.62E-04	0.0353	0.0001	0.5	213193_x_at	TRBC1	T cell receptor beta constant 1	28639
8.93E-03	0.154	0.0087	0.5	204006_s_at			
3.28E-02	0.284	0.0353	0.5	209374_s_at	IGHM	immunoglobulin heavy constant mu	3507
8.66E-04	0.0543	0.0004	0.5	210972_x_at			
8.23E-04	0.0537	0.0006	0.5	213539_at	CD3D	CD3d molecule, delta (CD3-TCR complex)	915
9.90E-05	0.0245	< 1e-07	0.5	209671_x_at			
3.98E-04	0.0433	0.0006	0.6	217157_x_at			
2.57E-03	0.0851	0.0025	0.6	212311_at	SEL1L3	sel-1 suppressor of lin-12-like 3 (C. elegans)	23231
4.26E-02	0.321	0.0396	0.6	209116_x_at	HBB	hemoglobin, beta	3043
1.15E-03	0.0608	0.0008	0.6	206715_at	TFEC	transcription factor EC	22797
8.48E-03	0.15	0.0099	0.6	203639_s_at	FGFR2	fibroblast growth factor receptor 2	2263
2.87E-03	0.0895	0.0027	0.6	204834_at	FGL2	fibrinogen-like 2	10875
2.99E-02	0.273	0.0283	0.6	210072_at	CCL19	chemokine (C-C motif) ligand 19	6363
1.81E-	0.074	0.001	0.6	215049_	CD163	CD163 molecule	9332

03		8		x_at			
4.40E-03	0.109	0.0039	0.6	205488_at	GZMA	granzyme A (granzyme 1, cytotoxic T-lymphocyte-associated serine esterase 3)	3001
5.45E-04	0.0478	0.0006	0.6	209606_at	CYTIP	cytohesin 1 interacting protein	9595
8.00E-06	0.0099	< 1e-07	0.6	212307_s_at	OGT	O-linked N-acetylglucosamine (GlcNAc) transferase	8473
1.46E-03	0.0676	0.0012	0.6	209823_x_at	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	3119
2.10E-04	0.0328	< 1e-07	0.6	204890_s_at	LCK	lymphocyte-specific protein tyrosine kinase	3932
9.72E-04	0.0567	0.001	0.6	203922_s_at	CYBB	cytochrome b-245, beta polypeptide	1536
2.89E-04	0.0361	0.0005	0.6	216250_s_at	LPXN	leupaxin	9404
2.32E-03	0.0816	0.0018	0.6	209846_s_at	BTN3A2	butyrophilin, subfamily 3, member A2	11118
8.91E-05	0.0231	0.0001	0.6	202524_s_at	SPOCK2	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	9806
3.98E-04	0.0433	0.0004	0.6	213915_at	NKG7	natural killer cell group 7 sequence	4818
3.46E-03	0.0971	0.0028	0.6	205541_s_at	GSPT2	G1 to S phase transition 2	23708
2.70E-05	0.0167	< 1e-07	0.6	221978_at	HLA-F	major histocompatibility complex, class I, F	3134
9.54E-04	0.056	0.0003	0.6	204116_at	IL2RG	interleukin 2 receptor, gamma	3561
3.31E-05	0.0173	< 1e-07	0.6	221087_s_at	APOL3	apolipoprotein L, 3	80833
4.17E-04	0.0435	0.0002	0.6	202510_s_at	TNFAIP2	tumor necrosis factor, alpha-induced protein 2	7127
6.02E-04	0.049	0.0004	0.6	210356_x_at	MS4A1	membrane-spanning 4-domains, subfamily A, member 1	931
1.44E-04	0.0299	0.0002	0.6	218805_at			
7.51E-04	0.0523	0.0005	0.6	221973_at			
5.02E-03	0.116	0.0035	0.6	209969_s_at	STAT1	signal transducer and activator of transcription 1, 91kDa	6772
2.73E-02	0.263	0.0282	0.6	209924_at	CCL18	chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)	6362
3.89E-03	0.103	0.0033	0.6	208798_x_at	GOLGA8A	golgin A8 family, member A	23015
2.79E-02	0.266	0.0264	0.6	205681_at	BCL2A1	BCL2-related protein A1	597
1.65E-03	0.0716	0.0016	0.6	203645_s_at	CD163	CD163 molecule	9332
3.67E-03	0.0999	0.0037	0.6	205692_s_at	CD38	CD38 molecule	952
6.18E-03	0.129	0.0049	0.6	34210_at	CD52	CD52 molecule	1043
2.67E-05	0.0167	< 1e-07	0.6	203416_at	CD53	CD53 molecule	963

1.97E-03	0.076	0.0019	0.6	204057_at	IRF8	interferon regulatory factor 8	3394
4.55E-04	0.0436	0.0003	0.6	222150_s_at	PION	pigeon homolog (Drosophila)	54103
5.98E-04	0.049	0.0006	0.6	210982_s_at	HLA-DRA	major histocompatibility complex, class II, DR alpha	3122
2.20E-03	0.0794	0.0023	0.6	200796_s_at	MCL1	myeloid cell leukemia sequence 1 (BCL2-related)	4170
3.21E-04	0.0386	0.0005	0.6	202531_at	IRF1	interferon regulatory factor 1	3659
2.07E-04	0.0327	0.0002	0.6	38149_at	ARHGAP25	Rho GTPase activating protein 25	9938
2.33E-04	0.034	< 1e-07	0.6	204821_at	BTN3A3	butyrophilin, subfamily 3, member A3	10384
8.49E-05	0.0223	0.0001	0.6	212980_at	USP34	ubiquitin specific peptidase 34	9736
1.05E-02	0.167	0.01	0.6	202902_s_at	CTSS	cathepsin S	1520
7.52E-04	0.0523	0.0004	0.6	205049_s_at	CD79A	CD79a molecule, immunoglobulin-associated alpha	973
2.90E-06	0.00587	< 1e-07	0.6	207564_x_at	OGT	O-linked N-acetylglucosamine (GlcNAc) transferase	8473
4.23E-04	0.0435	0.001	0.6	212384_at			
1.35E-04	0.0291	0.0003	0.6	213160_at	DOCK2	dedicator of cytokinesis 2	1794
1.14E-03	0.0603	0.0013	0.6	221286_s_at	MZB1	marginal zone B and B1 cell-specific protein	51237
1.64E-04	0.0299	0.0001	0.6	205997_at	ADAM28	ADAM metallopeptidase domain 28	10863
5.11E-04	0.0463	0.0002	0.6	204192_at	CD37	CD37 molecule	951
1.46E-03	0.0676	0.0012	0.6	212827_at	IGHM	immunoglobulin heavy constant mu	3507
8.33E-05	0.0223	< 1e-07	0.6	207563_s_at	OGT	O-linked N-acetylglucosamine (GlcNAc) transferase	8473
1.04E-04	0.0249	0.0002	0.6	214093_s_at	FUBP1	far upstream element (FUSE) binding protein 1	8880
1.59E-02	0.201	0.0168	0.6	208228_s_at	FGFR2	fibroblast growth factor receptor 2	2263
1.62E-04	0.0299	< 1e-07	0.6	205291_at	IL2RB	interleukin 2 receptor, beta	3560
4.74E-03	0.114	0.0062	0.6	214753_at	N4BP2L2	NEDD4 binding protein 2-like 2	10443
3.50E-04	0.0407	0.0003	0.6	206337_at	CCR7	chemokine (C-C motif) receptor 7	1236
5.11E-04	0.0463	0.0003	0.6	211991_s_at	HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	3113
2.76E-03	0.0881	0.0014	0.6	204661_at	CD52	CD52 molecule	1043
6.84E-04	0.0509	0.0007	0.6	203760_s_at	SLA	Src-like-adaptor	6503
2.95E-03	0.09	0.0029	0.6	202988_s_at	RGS1	regulator of G-protein signaling 1	5996
2.75E-	0.088	0.002	0.6	203381_	APOE	apolipoprotein E	348

03	1	5		s_at			
3.65E-02	0.3	0.0351	0.6	214023_x_at	TUBB2B	tubulin, beta 2B class IIb	347733
1.92E-05	0.0159	0.0001	0.6	212577_at	SMCHD1	structural maintenance of chromosomes flexible hinge domain containing 1	23347
4.68E-02	0.333	0.0425	0.6	217414_x_at			
3.00E-07	0.00223	< 1e-07	0.6	205298_s_at	BTN2A2	butyrophilin, subfamily 2, member A2	10385
2.51E-03	0.084	0.002	0.6	202953_at	C1QB	complement component 1, q subcomponent, B chain	713
4.09E-02	0.313	0.0404	0.6	215214_at			
7.67E-04	0.0525	0.0004	0.6	210425_x_at			
1.76E-04	0.0303	0.0001	0.6	209685_s_at	PRKCB	protein kinase C, beta	5579
6.04E-04	0.049	0.0006	0.6	202643_s_at	TNFAIP3	tumor necrosis factor, alpha-induced protein 3	7128
4.20E-06	0.00624	< 1e-07	0.6	212232_at	FNBP4	formin binding protein 4	23360
4.86E-02	0.337	0.0442	0.6	209458_x_at			
2.47E-04	0.0352	0.0002	0.6	209312_x_at			
4.92E-03	0.116	0.0047	0.6	201858_s_at	SRGN	serglycin	5552
3.21E-02	0.281	0.0292	0.6	216853_x_at			
2.09E-03	0.0782	0.0021	0.6	208894_at	HLA-DRA	major histocompatibility complex, class II, DR alpha	3122
1.44E-03	0.0676	0.0012	0.6	211902_x_at	YME1L1	YME1-like 1 (S. cerevisiae)	10730
3.15E-03	0.0927	0.0033	0.6	AFFX-HUMISGF3A/M97935_5_at	STAT1	signal transducer and activator of transcription 1, 91kDa	6772
4.50E-03	0.11	0.0052	0.6	218232_at	C1QA	complement component 1, q subcomponent, A chain	712
9.04E-04	0.0547	0.0009	0.6	213502_x_at	GUSBP11	glucuronidase, beta pseudogene 11	91316
2.47E-02	0.249	0.0283	0.6	221728_x_at	XIST	X inactive specific transcript (non-protein coding)	7503
7.01E-04	0.0509	0.0006	0.6	221989_at			
7.23E-04	0.0515	0.0007	0.6	211656_x_at	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	3119
5.81E-04	0.0487	0.0005	0.6	215193_x_at			
5.71E-04	0.0485	0.0009	0.6	203470_s_at	PLEK	pleckstrin	5341
5.58E-04	0.0479	0.0004	0.6	216542_x_at			
2.43E-03	0.083	0.0018	0.6	214617_at	PRF1	perforin 1 (pore forming protein)	5551
3.23E-03	0.0933	0.0028	0.6	217143_s_at	YME1L1	YME1-like 1 (S. cerevisiae)	10730

2.76E-03	0.0881	0.0033	0.6	203923_s_at	CYBB	cytochrome b-245, beta polypeptide	1536
1.91E-03	0.0755	0.0016	0.6	213703_at	LINC00342	long intergenic non-protein coding RNA 342	150759
2.96E-03	0.09	0.0021	0.6	209083_at	CORO1A	coronin, actin binding protein, 1A	11151
5.67E-05	0.02	0.0002	0.6	203932_at	HLA-DMB	major histocompatibility complex, class II, DM beta	3109
5.46E-03	0.121	0.0047	0.6	201718_s_at	EPB41L2	erythrocyte membrane protein band 4.1-like 2	2037
3.11E-02	0.277	0.0332	0.6	210164_at	GZMB	granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1)	3002
1.78E-04	0.0303	0.0004	0.6	204670_x_at			
5.68E-03	0.123	0.006	0.6	211742_s_at	EVI2B	ecotropic viral integration site 2B	2124
1.62E-03	0.071	0.0022	0.6	221768_at			
3.55E-04	0.041	0.0001	0.6	201720_s_at	LAPTM5	lysosomal protein transmembrane 5	7805
1.80E-06	0.00573	< 1e-07	0.6	212036_s_at	PNN	pinin, desmosome associated protein	5411
8.48E-05	0.0223	< 1e-07	0.6	204674_at	LRMP	lymphoid-restricted membrane protein	4033
1.16E-03	0.0608	0.0011	0.6	213142_x_at	PION	pigeon homolog (Drosophila)	54103
5.74E-05	0.02	< 1e-07	0.6	205270_s_at	LCP2	lymphocyte cytosolic protein 2 (SH2 domain containing leukocyte protein of 76kDa)	3937
5.88E-03	0.126	0.005	0.6	211654_x_at	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	3119
3.00E-03	0.0909	0.0033	0.6	215946_x_at	IGLL3P	immunoglobulin lambda-like polypeptide 3, pseudogene	91353
1.00E-03	0.0571	0.0008	0.6	209723_at	SERPINB9	serpin peptidase inhibitor, clade B (ovalbumin), member 9	5272
8.78E-04	0.0545	0.0007	0.6	205842_s_at	JAK2	Janus kinase 2	3717
4.43E-04	0.0435	0.0004	0.6	203236_s_at	LGALS9	lectin, galactoside-binding, soluble, 9	3965
2.42E-03	0.083	0.002	0.6	217418_x_at	MS4A1	membrane-spanning 4-domains, subfamily A, member 1	931
1.14E-02	0.174	0.0105	0.6	202086_at	MX1	myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)	4599
9.48E-03	0.159	0.0085	0.6	213875_x_at	C6orf62	chromosome 6 open reading frame 62	81688
5.77E-03	0.124	0.0047	0.6	203879_at	PIK3CD	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit delta	5293
8.98E-04	0.0547	0.0003	0.6	204882_at	ARHGAP25	Rho GTPase activating protein 25	9938
6.21E-04	0.0493	0.0009	0.6	221427_s_at	CCNL2	cyclin L2	81669
1.60E-	0.202	0.013	0.6	210663_	KYNU	kynureninase	8942

02		7		s_at			
9.08E-04	0.0547	0.0004	0.6	210116_at	SH2D1A	SH2 domain containing 1A	4068
4.28E-03	0.108	0.0041	0.6	216614_at			
1.30E-03	0.0644	0.0015	0.6	213326_at	VAMP1	vesicle-associated membrane protein 1 (synaptobrevin 1)	6843
1.78E-03	0.074	0.0016	0.6	221602_s_at	FAIM3	Fas apoptotic inhibitory molecule 3	9214
3.19E-03	0.0931	0.0029	0.6	38241_at	BTN3A3	butyrophilin, subfamily 3, member A3	10384
2.58E-03	0.0851	0.0021	0.6	205758_at	CD8A	CD8a molecule	925
4.94E-03	0.116	0.0048	0.6	213359_at	HNRNPD	heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa)	3184
1.04E-02	0.166	0.0093	0.6	204959_at	MNDA	myeloid cell nuclear differentiation antigen	4332
3.44E-03	0.0969	0.0034	0.6	206133_at	XAF1	XIAP associated factor 1	54739
4.31E-02	0.322	0.043	0.6	217388_s_at	KYNU	kynureninase	8942
1.20E-03	0.0616	0.0011	0.6	213293_s_at	TRIM22	tripartite motif containing 22	10346
6.36E-03	0.13	0.0067	0.6	208018_s_at	HCK	hemopoietic cell kinase	3055
7.89E-04	0.0528	0.0018	0.6	214132_at	ATP5C1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	509
1.49E-02	0.196	0.0147	0.6	212998_x_at	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	3119
1.13E-02	0.174	0.0111	0.6	219666_at	MS4A6A	membrane-spanning 4-domains, subfamily A, member 6A	64231
2.30E-06	0.00587	< 1e-07	0.6	202380_s_at	NKTR	natural killer-tumor recognition sequence	4820
1.05E-02	0.167	0.0102	0.6	219209_at	IFIH1	interferon induced with helicase C domain 1	64135
1.76E-02	0.212	0.0158	0.6	206641_at	TNFRSF17	tumor necrosis factor receptor superfamily, member 17	608
3.90E-03	0.103	0.0036	0.7	206978_at	CCR2	chemokine (C-C motif) receptor 2	729230
4.75E-05	0.0186	0.0001	0.7	204923_at	SASH3	SAM and SH3 domain containing 3	54440
1.94E-03	0.0755	0.0013	0.7	218543_s_at	PARP12	poly (ADP-ribose) polymerase family, member 12	64761
6.79E-05	0.0214	0.0002	0.7	213269_at	ZNF248	zinc finger protein 248	57209
8.26E-05	0.0223	0.0001	0.7	204234_s_at	ZNF195	zinc finger protein 195	7748
6.48E-04	0.0498	0.0005	0.7	203761_at	SLA	Src-like-adaptor	6503
6.59E-03	0.133	0.0047	0.7	201104_x_at			

1.92E-02	0.223	0.018	0.7	211090_s_at	PRPF4B	PRP4 pre-mRNA processing factor 4 homolog B (yeast)	8899
3.28E-03	0.0942	0.0022	0.7	213603_s_at	RAC2	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	5880
2.15E-03	0.0786	0.0023	0.7	203382_s_at	APOE	apolipoprotein E	348
1.23E-02	0.18	0.0124	0.7	216920_s_at			
1.74E-02	0.212	0.0196	0.7	211996_s_at			
4.63E-03	0.112	0.0049	0.7	202803_s_at	ITGB2	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)	3689
1.73E-04	0.0301	0.0001	0.7	212613_at	BTN3A2	butyrophilin, subfamily 3, member A2	11118
1.81E-03	0.074	0.0016	0.7	210279_at	GPR18	G protein-coupled receptor 18	2841
3.02E-04	0.0374	0.0006	0.7	221971_x_at			
1.40E-02	0.191	0.0148	0.7	211317_s_at	CFLAR	CASP8 and FADD-like apoptosis regulator	8837
2.51E-03	0.084	0.0016	0.7	203185_at	RASSF2	Ras association (RalGDS/AF-6) domain family member 2	9770
1.58E-02	0.201	0.0169	0.7	214059_at	IFI44	interferon-induced protein 44	10561
6.57E-03	0.133	0.0067	0.7	208747_s_at	C1S	complement component 1, s subcomponent	716
7.43E-04	0.0523	0.0008	0.7	208885_at	LCP1	lymphocyte cytosolic protein 1 (L-plastin)	3936
1.12E-02	0.173	0.0099	0.7	216829_at			
1.57E-03	0.0701	0.001	0.7	210031_at	CD247	CD247 molecule	919
1.87E-02	0.22	0.0177	0.7	219505_at	CECR1	cat eye syndrome chromosome region, candidate 1	51816
2.67E-03	0.0869	0.0027	0.7	202957_at	HCLS1	hematopoietic cell-specific Lyn substrate 1	3059
1.67E-04	0.0299	0.0001	0.7	208306_x_at			
2.56E-04	0.0353	0.0004	0.7	221850_x_at			
5.66E-03	0.123	0.0054	0.7	212187_x_at	PTGDS	prostaglandin D2 synthase 21kDa (brain)	5730
3.30E-04	0.0394	0.0005	0.7	220046_s_at	CCNL1	cyclin L1	57018
7.28E-03	0.139	0.0073	0.7	222018_at	NACA	nascent polypeptide-associated complex alpha subunit	4666
1.32E-02	0.185	0.0133	0.7	214567_s_at			
2.61E-02	0.258	0.0261	0.7	213537_at	HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	3113
2.18E-04	0.0328	0.0003	0.7	207734_at	LAX1	lymphocyte transmembrane adaptor 1	54900

1.65E-04	0.0299	0.0003	0.7	217610_at	SPDYE2	speedy homolog E2 (Xenopus laevis)	441273
4.96E-04	0.046	0.0001	0.7	206150_at	CD27	CD27 molecule	939

Table 16. Class comparison of the global gene expression profiles of high TN score ER- tumors to low TN score ER- tumors in the ROCK dataset (highlighted probe set indicates common in high TN score BLBC and ER- breast tumours).

Parametric p-value	FDR	Permutation p-value	Fold-change for high TN score vs. Low TN score	ProbeSet	Symbol	Name	EntrezID
< 1e-07	< 1e-07	< 1e-07	3.0	209803_s_at	PHLDA2	pleckstrin homology-like domain, family A, member 2	7262
2.72E-05	0.000683	0.0002	2.9	211657_at	CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	4680
5.00E-07	0.0000338	< 1e-07	2.8	203108_at	GPRC5A	G protein-coupled receptor, family C, group 5, member A	9052
1.28E-04	0.00209	0.0001	2.7	204351_at	S100P	S100 calcium binding protein P	6286
2.00E-07	0.0000168	< 1e-07	2.4	202489_s_at	FXYD3	FXYD domain containing ion transport regulator 3	5349
< 1e-07	< 1e-07	< 1e-07	2.3	201467_s_at	NQO1	NAD(P)H dehydrogenase, quinone 1	1728
5.00E-07	0.0000338	< 1e-07	2.2	210519_s_at	NQO1	NAD(P)H dehydrogenase, quinone 1	1728
4.88E-04	0.00588	0.0005	2.1	203021_at	SLPI	secretory leukocyte peptidase inhibitor	6590
< 1e-07	< 1e-07	< 1e-07	2.1	219232_s_at	EGLN3	egl nine homolog 3 (C. elegans)	112399
1.96E-05	0.000531	< 1e-07	2.1	218309_at	CAMK2N1	calcium/calmodulin-dependent protein kinase II inhibitor 1	55450
5.00E-07	0.0000338	< 1e-07	2.1	204044_at	QPRT	quinolinate phosphoribosyltransferase	23475
2.00E-07	0.0000168	< 1e-07	2.1	201468_s_at	NQO1	NAD(P)H dehydrogenase, quinone 1	1728
3.00E-07	0.0000224	< 1e-07	2.0	205968_at	KCNS3	potassium voltage-gated channel, delayed-rectifier, subfamily S, member 3	3790
< 1e-07	< 1e-07	< 1e-07	2.0	201286_at	SDC1	syndecan 1	6382
3.60E-06	0.000144	< 1e-07	2.0	203803_at	PCYOX1	prenylcysteine oxidase 1	51449
1.00E-05	0.000317	< 1e-07	2.0	200862_at	DHCR24	24-dehydrocholesterol reductase	1718
8.43E-	0.048	0.009	2.0	204268_	S100A2	S100 calcium binding	6273

03	1	9		at		protein A2	
< 1e-07	< 1e-07	< 1e-07	2.0	209146_at	MSMO1	methylsterol monooxygenase 1	6307
1.30E-06	0.0000677	< 1e-07	1.9	203058_s_at	PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	9060
3.20E-05	0.000761	0.0001	1.9	203060_s_at	PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	9060
3.84E-05	0.00087	< 1e-07	1.9	210652_s_at	TTC39A	tetratricopeptide repeat domain 39A	22996
6.51E-04	0.00727	0.0006	1.9	213693_s_at	MUC1	mucin 1, cell surface associated	4582
1.56E-04	0.00243	0.0002	1.9	212444_at	GPRC5A	G protein-coupled receptor, family C, group 5, member A	9052
2.00E-07	0.0000168	< 1e-07	1.9	205709_s_at	CDS1	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 1	1040
< 1e-07	< 1e-07	< 1e-07	1.9	202068_s_at	LDLR	low density lipoprotein receptor	3949
< 1e-07	< 1e-07	< 1e-07	1.8	201849_at	BNIP3	BCL2/adenovirus E1B 19kDa interacting protein 3	664
1.10E-06	0.0000599	< 1e-07	1.8	218273_s_at	PDP1	pyruvate dehyrogenase phosphatase catalytic subunit 1	54704
9.00E-07	0.0000513	< 1e-07	1.8	213577_at	SQLE	squalene epoxidase	6713
2.60E-06	0.000114	< 1e-07	1.8	36711_at	MAFF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog F (avian)	23764
3.60E-03	0.0257	0.0025	1.8	207847_s_at	MUC1	mucin 1, cell surface associated	4582
6.90E-06	0.000234	< 1e-07	1.8	201596_x_at	KRT18	keratin 18	3875
3.11E-04	0.00416	0.0005	1.8	201951_at	ALCAM	activated leukocyte cell adhesion molecule	214
8.00E-07	0.0000477	< 1e-07	1.8	222258_s_at	SH3BP4	SH3-domain binding protein 4	23677
1.36E-02	0.0676	0.0127	1.8	203453_at	SCNN1A	sodium channel, non-voltage-gated 1 alpha subunit	6337
3.00E-07	0.0000224	< 1e-07	1.8	219496_at	SOWAHC	sosondowah ankyrin repeat domain family member C	65124
8.54E-05	0.00157	0.0001	1.8	218959_at	HOXC10	homeobox C10	3226
1.01E-05	0.000319	< 1e-07	1.8	203407_at	PPL	periplakin	5493
5.20E-05	0.0011	0.0001	1.8	219127_at	PRR15L	proline rich 15-like	79170
2.00E-07	0.0000168	< 1e-07	1.8	221042_s_at	CLMN	calmin (calponin-like, transmembrane)	79789
1.84E-02	0.083	0.0181	1.8	214079_at	DHRS2	dehydrogenase/reductase (SDR family) member 2	10202
1.00E-06	0.0000561	< 1e-07	1.8	209260_at	SFN	stratifin	2810
1.80E-02	0.0818	0.0183	1.8	218211_s_at	MLPH	melanophilin	79083
1.57E-03	0.0141	0.0021	1.8	221024_s_at	SLC2A10	solute carrier family 2 (facilitated glucose transporter), member 10	81031

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 181

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 181

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CLAIMS

1. A method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene, wherein: a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

2. A method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple

Networks metagene, wherein: a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a more favourable cancer prognosis.

3. The method of Claim 1 or Claim 2, wherein the one or plurality of overexpressed genes and/or the one or plurality of underexpressed genes are selected from one of the metagenes or are selected from a plurality of the metagenes.

4. The method of any one of the preceding claims, wherein the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and/or the Multiple Networks metagene comprise one or a plurality of genes listed in Table 21.

5. A method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Metabolism metagene, a Signalling metagene, a Development and Growth metagene, a Chromosome Segregation/Replication metagene, an Immune Response metagene and a Protein Synthesis/Modification metagene, wherein: a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

6. A method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Metabolism metagene, a Signalling metagene, a Development and Growth metagene, a Chromosome Segregation/Replication metagene, an Immune Response metagene and a Protein Synthesis/Modification metagene, wherein: a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a more favourable cancer prognosis.

7. The method of Claim 5 or Claim 6, wherein the one or plurality of overexpressed genes and/or the one or plurality of underexpressed genes are selected from one of the metagenes or are selected from a plurality of the metagenes.

8. The method of any one of Claims 5 to 7, wherein the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene comprise one or more genes listed in Table 22.

9. The method of any one of Claims 5 to 8, wherein the one or plurality of overexpressed genes and the one or plurality of underexpressed genes are from one or a plurality of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene.

10. The method of any one of the preceding claims, wherein the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes includes comparing an average expression level of the one or plurality of overexpressed genes and/or an average expression level of the one or plurality of underexpressed genes.

11. The method of Claim 10, which includes calculating a ratio of the average expression level of the one or plurality of overexpressed genes and the average expression level of the plurality of underexpressed genes.

12. The method of any one of Claims 1-9, wherein the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes includes comparing the sum of expression levels of the one or plurality of overexpressed genes and/or the sum of expression levels of the one or plurality of underexpressed genes.

13. The method of Claim 12, which includes calculating a ratio of the sum of expression levels of the one or plurality of overexpressed genes and the sum of expression levels of the one or plurality of underexpressed genes.

14. A method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes associated with chromosomal instability and/or an expression level of one or a plurality of underexpressed genes associated with estrogen receptor signalling in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed genes associated with chromosomal instability compared to the one or plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or plurality of overexpressed genes associated with chromosomal instability compared to the one or plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

15. A method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes associated with chromosomal instability and/or an expression level of one or a plurality of underexpressed genes associated with estrogen receptor signalling in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed genes associated with chromosomal instability compared to the one or plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed genes associated with chromosomal instability compared to the one or plurality of underexpressed genes associated with estrogen receptor signalling indicates or correlates with a more favourable cancer prognosis.

16. The method of Claim 15, wherein the cancer prognosis includes determining responsiveness to anti-cancer therapies targeting aneuploid tumours.

17. The method of Claim 15, wherein the cancer prognosis includes determining responsiveness to anti-cancer therapies targeting chromosomal instability.

18. The method of any one of Claims 15 to 17, wherein the cancer prognosis includes determining responsiveness to one or more anti-cancer therapies that comprise targeting TTK, PLK1 and/or one or more Aurora Kinases.

19. The method of any one of Claims 14 to 18, wherein the step of comparing an expression level of one or a plurality of overexpressed genes associated with chromosomal instability and/or an expression level of one or a plurality of underexpressed genes associated with estrogen receptor signalling includes comparing an average expression level of the one or plurality of overexpressed genes associated with chromosomal instability and/or an average expression level of the one or plurality of underexpressed genes associated with estrogen receptor signalling.

20. The method of Claim 19, which includes calculating a ratio of the average expression level of the one or plurality of overexpressed genes associated with chromosomal instability and the average expression level of the one or plurality of underexpressed genes associated with estrogen receptor signalling.

21. The method of any one of Claims 14-18, wherein the step of comparing an expression level of one or a plurality of overexpressed genes associated with chromosomal instability and/or an expression level of one or a plurality of underexpressed genes associated with estrogen receptor signalling includes comparing the sum of expression levels of the one or plurality of overexpressed genes associated with chromosomal instability and/or the sum of expression levels of the one or plurality of underexpressed genes associated with estrogen receptor signalling.

22. The method of Claim 21, which includes calculating a ratio of the sum of expression levels of the one or plurality of overexpressed genes associated with chromosomal instability and the sum of expression levels of the one or plurality of underexpressed genes associated with estrogen receptor signalling.

23. The method of Claim 20 or Claim 22, wherein the ratio provides an aggressiveness score which is indicative of, or correlates with, cancer aggressiveness and a less favourable prognosis.

24. The method of any preceding claim, wherein the genes associated with chromosomal instability are of a CIN metagene.

25. The method of Claim 24, wherein the CIN metagene comprises a plurality of genes listed in Table 4.

26. The method of Claim 25, wherein the genes are selected from the group consisting of: *ATP6V1C1, RAP2A, CALM1, COG8, HELLS, KDM5A, PGK1, PLCH1, CEP55, RFC4, TAF2, SF3B3, GPI, PIR, MCM10, MELK, FOXM1, KIF2C, NUP155, TPX2, TTK, CENPA, CENPN, EXO1, MAPRE1, ACOT7, NAE1, SHMT2, TCP1, TXNRD1, ADM, CHAF1A and SYNCRIP.*

27. The method of Claim 26, wherein the genes are selected from the group consisting of: *MELK, MCM10, CENPA, EXO1, TTK* and *KIF2C.*

28. The method of any one of Claims 14 to 27, wherein the genes associated with estrogen receptor signalling are of an ER metagene.

29. The method of Claim 28, wherein the genes are selected from the group consisting of: *BTG2, PIK3IP1, SEC14L2, FLNB, ACSF2, APOM, BIN3, GLTSCR2, ZMYND10, ABAT, BCAT2, SCUBE2, RUNX1, LRRC48, MYBPC1, BCL2, CHPT1, ITM2A, LRIG1, MAPT, PRKCB, RERE, ABHD14A, FLT3, TNN, STC2, BATF, CD1E, CFB, EVL, FBXW4, ABCB1,*

ACAA1, CHAD, PDCD4, RPL10, RPS28, RPS4X, RPS6, SORBS1, RPL22 and RPS4XP3.

30. The method of Claim 29, wherein the genes are selected from the group consisting of: *MAPT* and *MYB*.

31. The method of any one of Claims 14 to 30, further including the step of comparing an expression level of one or a plurality of other overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and/or an expression level of one or a plurality of other underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of other overexpressed genes compared to the one or plurality of other underexpressed genes indicates or correlates with higher aggressiveness of the cancer and/or a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of other overexpressed genes compared to the one or plurality of other underexpressed genes indicates or correlates with lower aggressiveness of the cancer and/or a more favourable cancer prognosis compared to a mammal having a higher expression level.

32. The method of Claim 31, wherein the one or plurality of other overexpressed genes are selected from the group consisting of *ABHD5, ADORA2B, BCAP31, CA9, CAMSAP1, CARHSP1, CD55, CETN3, EIF3K, EXOSC7, GNB2L1, GRHPR, GSK3B, HCFC1R1, KCNG1, MAP2K5, NDUFC1, PML, STAU1, TXN* and *ZNF593* and/or the one or plurality of other underexpressed genes are selected from the group consisting of *BTN2A2, ERC2, IGH, ME1, MTMR7, SMPDL3B* and *ZNRD1-AS1*.

33. The method of Claim 31 or Claim 32, wherein the step of comparing the expression level of the one or plurality of other overexpressed genes and/or the expression level of the one or plurality of other underexpressed genes

includes comparing an average expression level of the one or plurality of other overexpressed genes and/or an average expression level of the one or plurality of other underexpressed genes.

34. The method of Claim 33, which includes calculating a ratio of the average expression level of the other overexpressed genes and the average expression level of the other underexpressed genes.

35. The method of Claim 31 or Claim 32, wherein the step of comparing an expression level of the one or plurality of other overexpressed genes and/or an expression level of the one or plurality of other underexpressed genes includes comparing the sum of expression levels of the one or plurality of other overexpressed genes and/or the sum of expression levels of the one or plurality of other underexpressed genes.

36. The method of Claim 35, which includes calculating a ratio of the sum of expression levels of the one or plurality of other overexpressed genes and the sum of expression levels of the one or plurality of other underexpressed genes.

37. The method of any one of Claims 31 to 36, wherein the comparison of the expression level of the overexpressed genes associated with chromosomal instability and/or the expression level of the underexpressed genes associated with estrogen receptor signalling is integrated with the comparison of the expression level of the other overexpressed genes and/or the expression level of the other underexpressed genes to derive a first integrated score.

38. A method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative

expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

39. A method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1*, and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3*, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a more favourable cancer prognosis compared to a mammal having a higher expression level.

40. The method of Claim 38 or Claim 39, wherein the one or plurality of overexpressed genes are selected from the group consisting of *ABHD5, ADORA2B, BCAP31, CA9, CAMSAP1, CARHSP1, CD55, CETN3, EIF3K, EXOSC7, GNB2L1, GRHPR, GSK3B, HCFC1R1, KCNG1, MAP2K5, NDUFC1, PML, STAU1, TXN* and *ZNF593* and/or the one or plurality of underexpressed genes are selected from the group consisting of *BTN2A2, ERC2, IGH, ME1, MTMR7, SMPDL3B* and *ZNRD1-AS1*.

41. The method of any one of Claims 38 to 40, wherein the step of comparing the expression level of the one or plurality of overexpressed genes and/or the expression level of the one or plurality of underexpressed genes includes comparing an average expression level of the one or plurality of overexpressed genes and/or an average expression level of the one or plurality of underexpressed genes.

42. The method of Claim 41, which includes calculating a ratio of the average expression level of the one or plurality of overexpressed genes and the average expression level of the one or plurality of underexpressed genes.

43. The method of any one of Claims 38 to Claim 40, wherein the step of comparing an expression level of the one or plurality of overexpressed genes and/or an expression level of the one or plurality of underexpressed genes includes comparing the sum of expression levels of the one or plurality of overexpressed genes and/or the sum of expression levels of the one or plurality of underexpressed genes.

44. The method of Claim 43, which includes calculating a ratio of the sum of expression levels of the one or plurality of overexpressed genes and the sum of expression levels of the one or plurality of underexpressed genes.

45. The method of any one of Claims 1 to 44, further including the step of comparing an expression level of a one or a plurality of overexpressed proteins, and/or an expression level of one or a plurality of underexpressed proteins, in one or a plurality of cancer cells, tissues or organs of the mammal to thereby derive an integrated score.

46. The method of Claim 38, wherein the one or plurality of overexpressed proteins are selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and/or the one or plurality of underexpressed proteins are selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, wherein: a higher relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with higher aggressiveness of the cancer and/or a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed proteins compared to the one or

plurality of underexpressed proteins indicates or correlates with lower aggressiveness of the cancer and/or a more favourable cancer prognosis compared to a mammal having a higher expression level.

47. The method of Claim 45 or Claim 46, wherein the step of comparing the expression level of the one or plurality of overexpressed proteins and/or the expression level of the one or plurality of underexpressed proteins includes comparing an average expression level of the one or plurality of overexpressed proteins and/or an average expression level of the one or plurality of underexpressed proteins.

48. The method of Claim 47, which includes calculating a ratio of the average expression level of the one or plurality of overexpressed proteins and the average expression level of the one or plurality of underexpressed proteins.

49. The method of Claim 45 or Claim 46, wherein the step of comparing an expression level of the one or plurality of overexpressed proteins and/or an expression level of the one or plurality of underexpressed proteins includes comparing the sum of expression levels of the one or plurality of overexpressed proteins and/or the sum of expression levels of the one or plurality of underexpressed proteins.

50. The method of Claim 49, which includes calculating a ratio of the sum of expression levels of the one or plurality of overexpressed proteins and the sum of expression levels of the one or plurality of underexpressed proteins.

51. The method of any one of Claims 45 to 50, wherein the comparison of the expression level of the one or plurality of overexpressed proteins and the expression level of the one or plurality of underexpressed proteins is integrated with:

(i) the comparison of the expression level of the overexpressed genes associated with chromosomal instability and/or the expression level of the underexpressed genes associated with estrogen receptor signalling to derive a second integrated score; or

(ii) the first integrated score to derive a third integrated score; or

(iii) the comparison of the expression level of the overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML,*

CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1 and *STAU1* and/or the expression level of the underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3* to derive a fourth integrated score; or

(iv) the comparison of the expression level of the overexpressed genes and an expression level of the underexpressed genes, wherein the genes are from one or a plurality of the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and/or the Multiple Networks metagene, to derive a fifth integrated score; or

(v) the comparison of the expression level of the overexpressed genes and an expression level of the underexpressed genes, wherein the genes are from one or a plurality of the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene, to derive a sixth integrated score.

52. The method of Claim 51, wherein the first, second, third, fourth, fifth and/or sixth integrated scores are derived, at least in part, by addition, subtraction, multiplication, division and/or exponentiation.

53. A method of determining the aggressiveness of a cancer in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1,

and/or an expression level of one or a plurality of underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with higher aggressiveness of the cancer; and/or a lower relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with lower aggressiveness of the cancer compared to a mammal having a higher expression level.

54. A method of determining a cancer prognosis for a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and/or an expression level of one or a plurality of underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein: a higher relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with a more favourable cancer prognosis compared to a mammal having a higher expression level.

55. The method of Claim 53 or Claim 54, wherein the step of comparing the expression level of the one or plurality of overexpressed proteins and/or the expression level of the one or plurality of underexpressed proteins includes comparing an average expression level of the one or plurality of overexpressed proteins and/or an average expression level of the one or plurality of underexpressed proteins.

56. The method of Claim 55, which includes calculating a ratio of the average expression level of the one or plurality of overexpressed proteins and the average expression level of the one or plurality of underexpressed proteins.

57. The method of Claim 53 or Claim 54, wherein the step of comparing an expression level of the one or plurality of overexpressed proteins and/or an expression level of the one or plurality of underexpressed proteins includes comparing the sum of expression levels of the one or plurality of overexpressed proteins and/or the sum of expression levels of the one or plurality of underexpressed proteins.

58. The method of Claim 57, which includes calculating a ratio of the sum of expression levels of the overexpressed proteins and the sum of expression levels of the underexpressed proteins.

59. A method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of determining an expression level of one or plurality of genes associated with chromosomal instability in one or a plurality of non-mitotic cells of the mammal, wherein a higher expression level indicates or correlates with relatively increased responsiveness of the cancer to the anti-cancer treatment.

60. The method of Claim 59, wherein the one or plurality of genes associated with chromosomal instability are targeted by the anti-cancer treatment.

61. The method of Claim 59 or Claim 60, wherein the one or plurality of genes associated with chromosomal instability are listed in Table 4 and/or include one or more genes associated with aneuploidy.

62. The method of Claim 61, wherein the one or plurality of genes associated with chromosomal instability and/or aneuploidy are selected from the group consisting of: *TTK*, *CEP55*, *FOXM1*, *SKIP2*, *PLK1* and/or Aurora kinases.

63. The method of any one of Claims 59 to 62, wherein the anti-cancer treatment is a treatment targeted to aneuploid tumours.

64. The method of any one of Claims 59 to 63, wherein the anti-cancer treatment is a treatment targeted to chromosomal instability.

65. A method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a

plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene, wherein an altered or modulated relative expression level of the overexpressed genes compared to the underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

66. The method of Claim 65, wherein the one or plurality of overexpressed genes and/or the one or plurality of underexpressed genes are selected from one metagene or are selected from a plurality of metagenes.

67. The method of Claim 65 or Claim 66, wherein the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein Synthesis/Modification metagene and/or the Multiple Networks metagene comprise one or more genes listed in Table 21.

68. A method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes and/or an expression level of one or a plurality of underexpressed genes in one or a plurality of cancer cells, tissues or organs of the mammal, wherein the overexpressed genes and the underexpressed genes are from one or a plurality of metagenes selected from the group consisting of a Metabolism metagene, a Signalling metagene, a Development and Growth metagene, a Chromosome Segregation/Replication metagene, an Immune Response metagene and a Protein Synthesis/Modification metagene, wherein an altered or modulated

relative expression level of the overexpressed genes compared to the underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

69. The method of Claim 68, wherein the one or plurality of overexpressed genes and/or the one or plurality of underexpressed genes are selected from one metagene or are selected from a plurality of metagenes.

70. The method of Claim 68 or Claim 69, wherein the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene comprise one or more genes listed in Table 22.

71. The method of any one of Claims 68 to 70, wherein the one or plurality of overexpressed genes and the one or plurality of underexpressed genes are from one or a plurality of of a Carbohydrate/Lipid Metabolism metagene, a Cell Signalling metagene, a Cellular Development metagene, a Cellular Growth metagene, a Chromosome Segregation metagene, a DNA Replication/Recombination metagene, an Immune System metagene, a Metabolic Disease metagene, a Nucleic Acid Metabolism metagene, a Post-Translational Modification metagene, a Protein Synthesis/Modification metagene and a Multiple Networks metagene.

72. The method of any one of Claims 65 to 71, wherein the step of comparing an expression level of the one or plurality of overexpressed genes and/or an expression level of the one or plurality of underexpressed genes includes comparing an average expression level of the plurality of overexpressed genes and/or an average expression level of the plurality of underexpressed genes.

73. The method of Claim 72, which includes calculating a ratio of the average expression level of the one or plurality of overexpressed genes and the average expression level of the one or plurality of underexpressed genes.

74. The method of any one of Claims 65 to 71, wherein the step of comparing an expression level of the one or plurality of overexpressed genes and/or an expression level of the one or plurality of underexpressed genes includes comparing the sum of expression levels of the one or plurality of

overexpressed genes and/or the sum of expression levels of the one or plurality of underexpressed genes.

75. The method of Claim 74, which includes calculating a ratio of the sum of expression levels of the one or plurality of overexpressed genes and the sum of expression levels of the one or plurality of underexpressed genes.

76. A method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of a one or plurality of overexpressed genes associated with chromosomal instability and/or an expression level of one or a plurality of underexpressed genes associated with estrogen receptor signalling in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the overexpressed genes associated with chromosomal instability compared to the underexpressed genes associated with estrogen receptor signalling indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

77. The method of Claim 76, wherein the genes associated with chromosomal instability are of a CIN metagene.

78. The method of Claim 77, wherein the CIN metagene comprises a plurality of genes listed in Table 4.

79. The method of Claim 78, wherein the genes are selected from the group consisting of: *ATP6V1C1, RAP2A, CALM1, COG8, HELLS, KDM5A, PGK1, PLCH1, CEP55, RFC4, TAF2, SF3B3, GPI, PIR, MCM10, MELK, FOXM1, KIF2C, NUP155, TPX2, TTK, CENPA, CENPN, EXO1, MAPRE1, ACOT7, NAE1, SHMT2, TCP1, TXNRD1, ADM, CHAF1A and SYNCRIP.*

80. The method of Claim 79, wherein the genes are selected from the group consisting of: *MELK, MCM10, CENPA, EXO1, TTK* and *KIF2C.*

81. The method of any one or Claims 76 to 80, wherein the genes associated with estrogen receptor signalling are of an ER metagene.

82. The method of Claim 81, wherein the genes are selected from the group consisting of: *BTG2, PIK3IP1, SEC14L2, FLNB, ACSF2, APOM, BIN3, GLTSCR2, ZMYND10, ABAT, BCAT2, SCUBE2, RUNX1, LRRC48, MYBPC1, BCL2, CHPT1, ITM2A, LRIG1, MAPT, PRKCB, RERE, ABHD14A, FLT3, TNN, STC2, BATF, CD1E, CFB, EVL, FBXW4, ABCB1,*

ACAA1, CHAD, PDCD4, RPL10, RPS28, RPS4X, RPS6, SORBS1, RPL22 and RPS4XP3.

83. The method of Claim 82, wherein the genes are selected from the group consisting of: *MAPT* and *MYB*.

84. The method of any one of Claims 76 to 83, wherein the step of comparing an expression level of the one or plurality of overexpressed genes associated with chromosomal instability and/or an expression level of the one or plurality of underexpressed genes associated with estrogen receptor signalling includes comparing an average expression level of the one or plurality of overexpressed genes associated with chromosomal instability and/or an average expression level of the one or plurality of underexpressed genes associated with estrogen receptor signalling.

85. The method of Claim 84, which includes calculating a ratio of the average expression level of the one or plurality of overexpressed genes associated with chromosomal instability and the average expression level of the one or plurality of underexpressed genes associated with estrogen receptor signalling.

86. The method of any one of Claims 76 to 83, wherein the step of comparing an expression level of the one or plurality of overexpressed genes associated with chromosomal instability and/or an expression level of the one or plurality of underexpressed genes associated with estrogen receptor signalling includes comparing the sum of expression levels of the one or plurality of overexpressed genes associated with chromosomal instability and/or the sum of expression levels of the one or plurality of underexpressed genes associated with estrogen receptor signalling.

87. The method of Claim 86, which includes calculating a ratio of the sum of expression levels of the one or plurality of overexpressed genes associated with chromosomal instability and the sum of expression levels of the one or plurality of underexpressed genes associated with estrogen receptor signalling.

88. The method of any one or Claims 76 to 87, further including the step of comparing an expression level of one or a plurality of other overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55,*

GEMIN4, *TXN*, *ABHD5*, *EIF3K*, *EIF4B*, *EXOSC7*, *GNB2L1*, *LAMA3*, *NDUFC1* and *STAU1*, and/or an expression level of one or a plurality of other underexpressed genes selected from the group consisting of *BRD8*, *BTN2A2*. *KIR2DL4*. *ME1*, *PSEN2*, *CALR*, *CAMK4*, *ITM2C*, *NOP2*, *NSUN5*, *SF3B1*, *ZNRD1-AS1*, *ARNT2*, *ERC2*, *SLC11A1*, *BRD4*, *APOBEC3A*, *CD1A*, *CD1B*, *CD1C*, *CXCR4*, *HLA-B*, *IGH*, *KIR2DL3*, *SMPDL3B*, *MYB*, *RLN1*, *MTMR7*, *SORBS1 and SRPK3* in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of other overexpressed genes compared to the one or plurality of other underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

89. The method of Claim 88, wherein the one or plurality of other overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*, *CD55*, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593* and/or the one or plurality of other underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

90. The method of Claim 88 or Claim 89, wherein the comparison of the expression level of the one or plurality of other overexpressed genes and/or the expression level of the one or plurality of other underexpressed genes is integrated with the comparison of the expression level of the one or plurality of overexpressed genes associated with chromosomal instability and/or the expression level of the one or plurality of underexpressed genes associated with estrogen receptor signalling to derive a first integrated score, which is indicative of, or correlates with, responsiveness of the cancer to the anti-cancer treatment.

91. The method of Claim 90, wherein the first integrated score is derived, at least in part, by addition, subtraction, multiplication, division and/or exponentiation.

92. The method of Claim 91, wherein the first integrated score is derived by exponentiation, wherein the comparison of the expression level of one or a plurality of other overexpressed genes and/or the expression level of one or a

plurality of other underexpressed genes is raised to the power of the comparison of the expression level of the one or plurality of overexpressed genes associated with chromosomal instability and/or the expression level of the one or plurality of underexpressed genes associated with estrogen receptor signalling.

93. The method of any one of Claims 88 to 92, wherein the step of comparing an expression level of the one or plurality of other overexpressed genes and/or an expression level of the one or plurality of other underexpressed genes includes comparing an average expression level of the one or plurality of other overexpressed genes and/or an average expression level of the one or plurality of other underexpressed genes.

94. The method of Claim 93, which includes calculating a ratio of the average expression level of the one or plurality of other overexpressed genes and the average expression level of the one or plurality of other underexpressed genes.

95. The method of any one of Claims 88 to 92, wherein the step of comparing an expression level of the one or plurality of other overexpressed genes and/or an expression level of the one or plurality of other underexpressed genes includes comparing the sum of expression levels of the one or plurality of other overexpressed genes and/or the sum of expression levels of the one or plurality of other underexpressed genes.

96. The method of Claim 95, which includes calculating a ratio of the sum of expression levels of the one or plurality of other overexpressed genes and the sum of expression levels of the one or plurality of other underexpressed genes.

97. A method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *CAMSAP1*, *CETN3*, *GRHPR*, *ZNF593*, *CA9*, *CFDP1*, *VPS28*, *ADORA2B*, *GSK3B*, *LAMA4*, *MAP2K5*, *HCFC1R1*, *KCNG1*, *BCAP31*, *ULBP2*, *CARHSP1*, *PML*, *CD36*, *CD55*, *GEMIN4*, *TXN*, *ABHD5*, *EIF3K*, *EIF4B*, *EXOSC7*, *GNB2L1*, *LAMA3*, *NDUFC1* and *STAU1*, and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of *BRD8*, *BTN2A2*. *KIR2DL4*. *ME1*, *PSEN2*, *CALR*, *CAMK4*, *ITM2C*, *NOP2*, *NSUN5*, *SF3B1*, *ZNRD1-AS1*, *ARNT2*, *ERC2*,

SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

98. The method of Claim 97, wherein the one or plurality of overexpressed genes are selected from the group consisting of *ABHD5*, *ADORA2B*, *BCAP31*, *CA9*, *CAMSAP1*, *CARHSP1*, *CD55*, *CETN3*, *EIF3K*, *EXOSC7*, *GNB2L1*, *GRHPR*, *GSK3B*, *HCFC1R1*, *KCNG1*, *MAP2K5*, *NDUFC1*, *PML*, *STAU1*, *TXN* and *ZNF593* and/or the one or plurality of underexpressed genes are selected from the group consisting of *BTN2A2*, *ERC2*, *IGH*, *ME1*, *MTMR7*, *SMPDL3B* and *ZNRD1-AS1*.

99. The method of Claim 97 or Claim 98, wherein the step of comparing the expression level of the one or plurality of overexpressed genes and/or the expression level of the one or plurality of underexpressed genes includes comparing an average expression level of the one or plurality of overexpressed genes and/or an average expression level of the one or plurality of underexpressed genes.

100. The method of Claim 99, which includes calculating a ratio of the average expression level of the one or plurality of overexpressed genes and the average expression level of the one or plurality of underexpressed genes.

101. The method of Claim 97 or Claim 98, wherein the step of comparing an expression level of the one or plurality of overexpressed genes and/or an expression level of the one or a plurality of underexpressed genes includes comparing the sum of expression levels of the one or plurality of overexpressed genes and/or the sum of expression levels of the one or plurality of underexpressed genes.

102. The method of Claim 101, which includes calculating a ratio of the sum of expression levels of the one or plurality of overexpressed genes and the sum of expression levels of the one or plurality of underexpressed genes.

103. The method of any one of Claims 65 to 103, further including the step of comparing an expression level of a one or plurality of overexpressed

proteins, and/or an expression level of one or a plurality of underexpressed proteins, in one or a plurality of cancer cells, tissues or organs of the mammal to thereby derive an integrated score.

104. The method of Claim 103, wherein the one or plurality of overexpressed proteins are selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and/or the one or plurality of underexpressed proteins are selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, wherein: a higher relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with higher aggressiveness of the cancer and/or a less favourable cancer prognosis; and/or a lower relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with lower aggressiveness of the cancer and/or a more favourable cancer prognosis compared to a mammal having a higher expression level.

105. The method of Claim 103 or Claim 104, wherein the step of comparing the expression level of the one or plurality of overexpressed proteins and/or the expression level of the one or plurality of underexpressed proteins includes comparing an average expression level of the one or plurality of overexpressed proteins and/or an average expression level of the one or plurality of underexpressed proteins.

106. The method of Claim 105, which includes calculating a ratio of the average expression level of the one or plurality of overexpressed proteins and the average expression level of the one or plurality of underexpressed proteins.

107. The method of Claim 103 or Claim 104, wherein the step of comparing an expression level of the one or plurality of overexpressed proteins and/or an expression level of the one or plurality of underexpressed proteins includes comparing the sum of expression levels of the one or plurality of overexpressed proteins and/or the sum of expression levels of the one or a plurality of underexpressed proteins.

108. The method of Claim 107, which includes calculating a ratio of the sum of expression levels of the one or plurality of overexpressed proteins and the sum of expression levels of the one or plurality of underexpressed proteins.

109. The method of any one of Claims 103 to 108, wherein the comparison of the expression level of the one or plurality of overexpressed proteins and the expression level of the one or plurality of underexpressed proteins is integrated with:

(i) the comparison of the expression level of the overexpressed genes associated with chromosomal instability and/or the expression level of the underexpressed genes associated with estrogen receptor signalling to derive a second integrated score; or

(ii) the first integrated score to derive a third integrated score; or

(i) the comparison of the expression level of the overexpressed genes selected from the group consisting of *CAMSAP1, CETN3, GRHPR, ZNF593, CA9, CFDP1, VPS28, ADORA2B, GSK3B, LAMA4, MAP2K5, HCFC1R1, KCNG1, BCAP31, ULBP2, CARHSP1, PML, CD36, CD55, GEMIN4, TXN, ABHD5, EIF3K, EIF4B, EXOSC7, GNB2L1, LAMA3, NDUFC1* and *STAU1* and/or the expression level of the underexpressed genes selected from the group consisting of *BRD8, BTN2A2. KIR2DL4. ME1, PSEN2, CALR, CAMK4, ITM2C, NOP2, NSUN5, SF3B1, ZNRD1-AS1, ARNT2, ERC2, SLC11A1, BRD4, APOBEC3A, CD1A, CD1B, CD1C, CXCR4, HLA-B, IGH, KIR2DL3, SMPDL3B, MYB, RLN1, MTMR7, SORBS1 and SRPK3* to derive a fourth integrated score; or

(ii) the comparison of the expression level of the overexpressed genes and/or an expression level of the underexpressed genes, wherein the genes are from one or a plurality of the Carbohydrate/Lipid Metabolism metagene, the Cell Signalling metagene, the Cellular Development metagene, the Cellular Growth metagene, the Chromosome Segregation metagene, the DNA Replication/Recombination metagene, the Immune System metagene, the Metabolic Disease metagene, the Nucleic Acid Metabolism metagene, the Post-Translational Modification metagene, the Protein

Synthesis/Modification metagene and/or the Multiple Networks metagene, to derive a fifth integrated score; or

(iii) the comparison of the expression level of the overexpressed genes and/or an expression level of the underexpressed genes, wherein the genes are from one or a plurality of the Metabolism metagene, the Signalling metagene, the Development and Growth metagene, the Chromosome Segregation/Replication metagene, the Immune Response metagene and/or the Protein Synthesis/Modification metagene, to derive a sixth integrated score.

110. The method of Claim 109, wherein the first, second, third, fourth, fifth and/or sixth integrated scores are derived, at least in part, by addition, subtraction, multiplication, division and/or exponentiation.

111. A method of predicting the responsiveness of a cancer to an anti-cancer treatment in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed proteins selected from the group consisting of DVL3, PAI-1, VEGFR2, INPP4B, EIF4EBP1, EGFR, Ku80, HER3, SMAD1, GATA3, ITGA2, AKT1, NFKB1, HER2, ASNS and COL6A1, and/or an expression level of one or a plurality of underexpressed proteins selected from the group consisting of VEGFR2, HER3, ASNS, MAPK9, ESR1, YWHAE, RAD50, PGR, COL6A1, PEA15 and RPS6, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed proteins compared to the one or plurality of underexpressed proteins indicates or correlates with relatively increased or decreased responsiveness of the cancer to the anti-cancer treatment.

112. The method of Claim 111, wherein the step of comparing the expression level of the one or plurality of overexpressed proteins and/or the expression level of the one or plurality of underexpressed proteins includes comparing an average expression level of the one or plurality of overexpressed proteins and/or an average expression level of the one or plurality of underexpressed proteins.

113. The method of Claim 112, which includes calculating a ratio of the average expression level of the one or plurality of overexpressed proteins and

the average expression level of the one or plurality of underexpressed proteins.

114. The method of Claim 111, wherein the step of comparing an expression level of the one or plurality of overexpressed proteins and/or an expression level of the one or plurality of underexpressed proteins includes comparing the sum of expression levels of the one or plurality of overexpressed proteins and/or the sum of expression levels of the one or plurality of underexpressed proteins.

115. The method of Claim 114, which includes calculating a ratio of the sum of expression levels of the one or plurality of overexpressed proteins and the sum of expression levels of the one or plurality of underexpressed proteins.

116. The method of any one of Claims 59 to 115, wherein the anti-cancer treatment is selected from the group consisting of a endocrine therapy, chemotherapy, immunotherapy and a molecularly targeted therapy.

117. The method of Claim 116, wherein the treatment comprises administration of an agent selected from the group consisting of an ALK inhibitor, a BCR-ABL inhibitor, a HSP90 inhibitor, an EGFR inhibitor, a PARP inhibitor, retinoic acid, a Bcl2 inhibitor, a gluconeogenesis inhibitor, a p38 MAPK inhibitor, a MEK1/2 inhibitor, a mTOR inhibitor, a PI3K inhibitor, an IGF1R inhibitor, a PLCγ inhibitor, a JNK inhibitor, a PAK1 inhibitor, a SYK inhibitor, a HDAC inhibitor, an FGFR inhibitor, a XIAP inhibitor, a PLK1 inhibitor, an ERK5 inhibitor, a TTK inhibitor, an Aurora Kinase Inhibitor and/or any combination thereof.

118. The method of Claim 116, wherein immunotherapy is or comprises an immune checkpoint inhibitor.

119. The method of Claim 118, wherein the immune checkpoint inhibitor is or comprises an anti-PD1 antibody or an anti-PDL1 antibody.

120. A method of predicting the responsiveness of a cancer to an immunotherapeutic agent in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *ADORA2B*, *CD36*, *CETN3*, *KCNG1*, *LAMA3*, *MAP2K5*, *NAE1*, *PGK1*, *STAU1*, *CFDP1*, *SF3B3* and *TXN*, and/or an expression level of one or a plurality of underexpressed genes selected

from the group consisting of *APOBEC3A*, *BCL2*, *BTN2A2*, *CAMSAP1*, *CAMK4*, *CARHSP1*, *FBXW4*, *GSK3B*, *HCFC1R1*, *MYB*, *PSEN2* and *ZNF593*, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the immunotherapeutic agent.

121. The method of Claim 120, wherein a higher relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a relatively increased responsiveness of the cancer to the immunotherapeutic agent; and/or a lower relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with a relatively decreased responsiveness of the cancer to the immunotherapeutic agent.

122. The method of Claim 120 or Claim 121, wherein the immunotherapeutic agent is an immune checkpoint inhibitor.

123. The method of Claim 122, wherein the immune checkpoint inhibitor is or comprises an anti-PD1 antibody or an anti-PDL1 antibody.

124. A method of predicting the responsiveness of a cancer to an epidermal growth factor receptor (EGFR) inhibitor in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *NAE1*, *GSK3B*, *TAF2*, *MAPRE1*, *BRD4*, *STAU1*, *TAF2*, *PDCD4*, *KCNG1*, *ZNRD1-AS1*, *EIF4B*, *HELLS*, *RPL22*, *ABAT*, *BTN2A2*, *CD1B*, *ITM2A*, *BCL2*, *CXCR4*, and *ARNT2* and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of *CD1C*, *CD1E*, *CD1B*, *KDM5A*, *BATF*, *EVL*, *PRKCB*, *HCFC1R1*, *CARHSP1*, *CHAD*, *KIR2DL4*, *ABHD5*, *ABHD14A*, *ACAA1*, *SRPK3*, *CFB*, *ARNT2*, *NDUFC1*, *BCL2*, *EVL*, *ULBP2*, *BIN3*, *SF3B3*, *CETN3*, *SYNCRIP*, *TAF2*, *CENPN*, *ATP6V1C1*, *CD55* and *ADORA2B*, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of

underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the immunotherapeutic agent.

125. A method of predicting the responsiveness of a cancer to a multikinase inhibitor in a mammal, said method including the step of comparing an expression level of one or a plurality of overexpressed genes selected from the group consisting of *SCUBE*, *CHPT1*, *CDC1*, *BTG2*, *ADORA2B* and *BCL2*, and/or an expression level of one or a plurality of underexpressed genes selected from the group consisting of *NOP2*, *CALR*, *MAPRE1*, *KCNG1*, *PGK1*, *SRPK3*, *RERE*, *ADM*, *LAMA3*, *KIR2DL4*, *ULBP2*, *LAMA4*, *CA9*, and *BCAP31*, in one or a plurality of cancer cells, tissues or organs of the mammal, wherein an altered or modulated relative expression level of the one or plurality of overexpressed genes compared to the one or plurality of underexpressed genes indicates or correlates with relatively increased or decreased responsiveness of the cancer to the multikinase inhibitor.

126. The method of any preceding claim, which includes the further step of treating cancer in the mammal.

127. A method for identifying an agent for use in the treatment of cancer including the steps of:

(i) contacting a protein product of *GRHPR*, *NDUFC1*, *CAMSAP1*, *CETN3*, *EIF3K*, *STAU1*, *EXOSC7*, *COG8*, *CFDP1* and/or *KCNG1* with a test agent; and

(ii) determining whether the test agent, at least partly, reduces, eliminates, suppresses or inhibits the expression and/or an activity of the protein product.

128. The method of Claim 127, wherein the agent possesses or displays little or no significant off-target and/or nonspecific effects.

129. The method of Claim 127 or Claim 128, wherein the agent is an antibody or a small organic molecule.

130. A method of treating a cancer in a mammal, including the step of administering to the mammal a therapeutically effective amount of the agent identified by the method of any one of Claims 127 to 129.

131. The method of any preceding claim wherein the mammal is a human.

132. The method of any preceding claim wherein the cancer includes breast cancer, lung cancer, ovarian cancer, cervical cancer, uterine cancer,

prostate cancer, cancer of the brain and nervous system, head and neck cancer, colon cancer, colorectal cancer, gastric cancer, liver cancer, kidney cancer, bladder cancer, melanoma, lymphoid cancers, myelomonocytic cancers, pancreatic cancer, pituitary cancer, adrenal cancer or musculoskeletal cancer.

133. The method of Claim 132, wherein breast cancer includes aggressive breast cancers and cancer subtypes such as triple negative breast cancer, grade 2 breast cancer, grade 3 breast cancer, lymph node positive (LN^+) breast cancer, HER2 positive ($HER2^+$) breast cancer and ER positive (ER^+) breast cancer.

134. An agent identified by the method of any one of Claims 127 to 129 for use in the treatment of cancer.

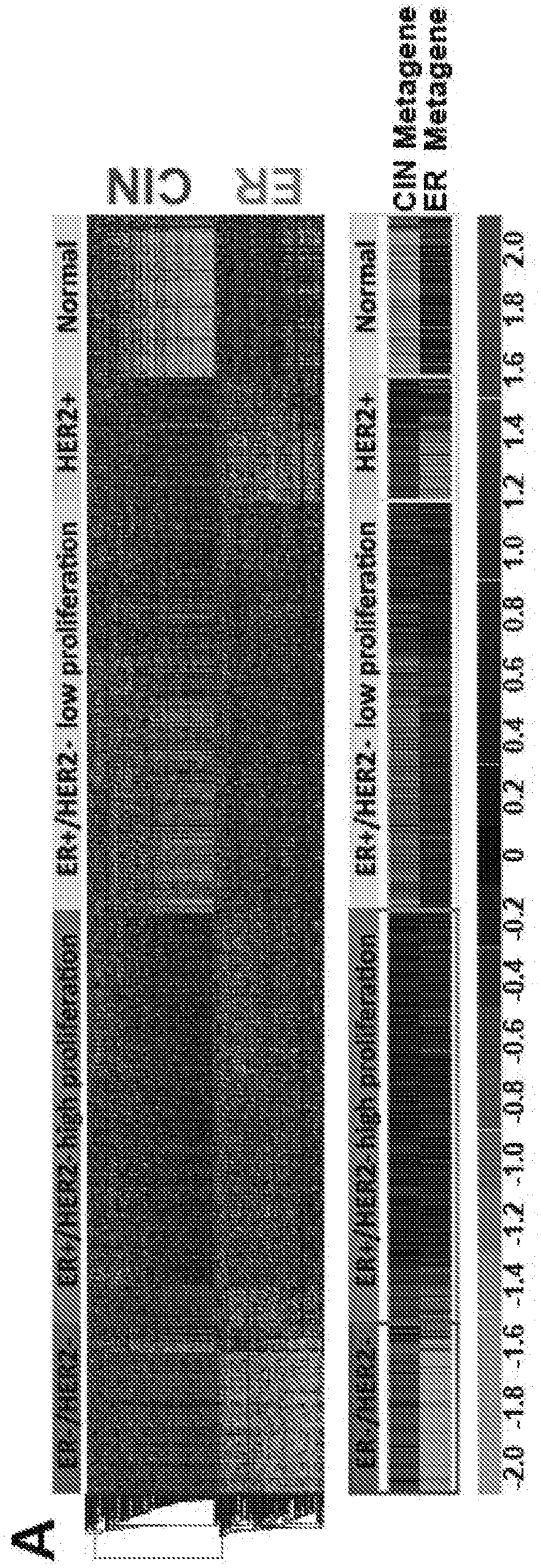
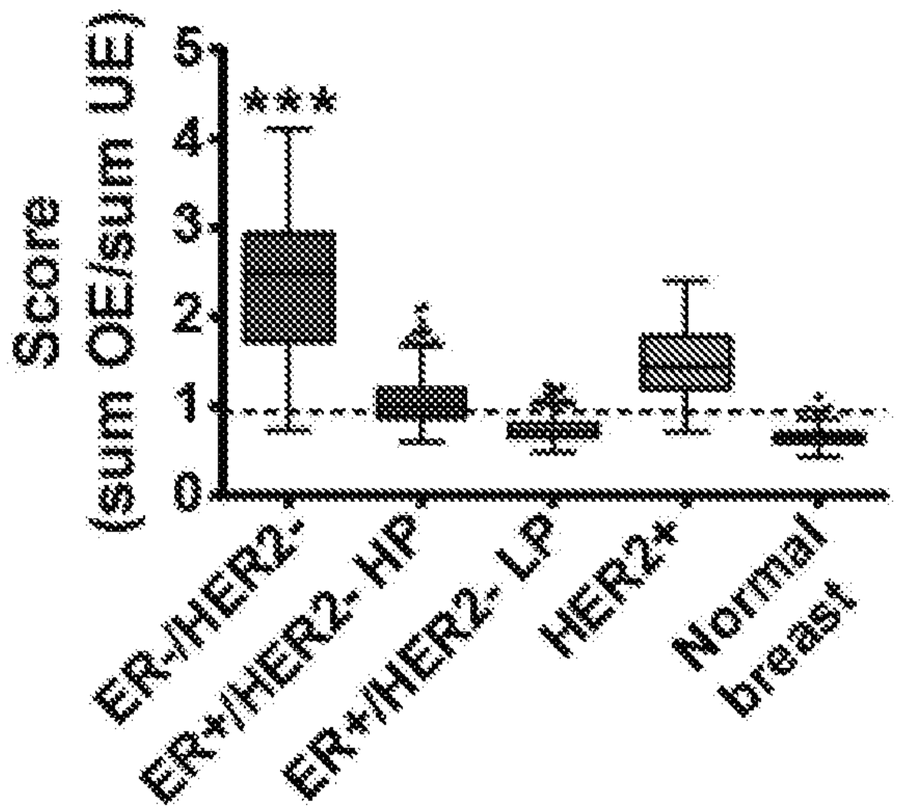

CIN
ER
CIN Metagene
ER Metagene
Normal
ER+/HER2- low proliferation
HER2+
ER+/HER2- high proliferation
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
-1.6
-1.8
-2.0
A

FIG. 1


Score (sum OE/sum UE)
0
1
2
3
4
5
***
ER-/HER2-
ER+/HER2- HP
ER+/HER2- LP
HER2+
Normal breast

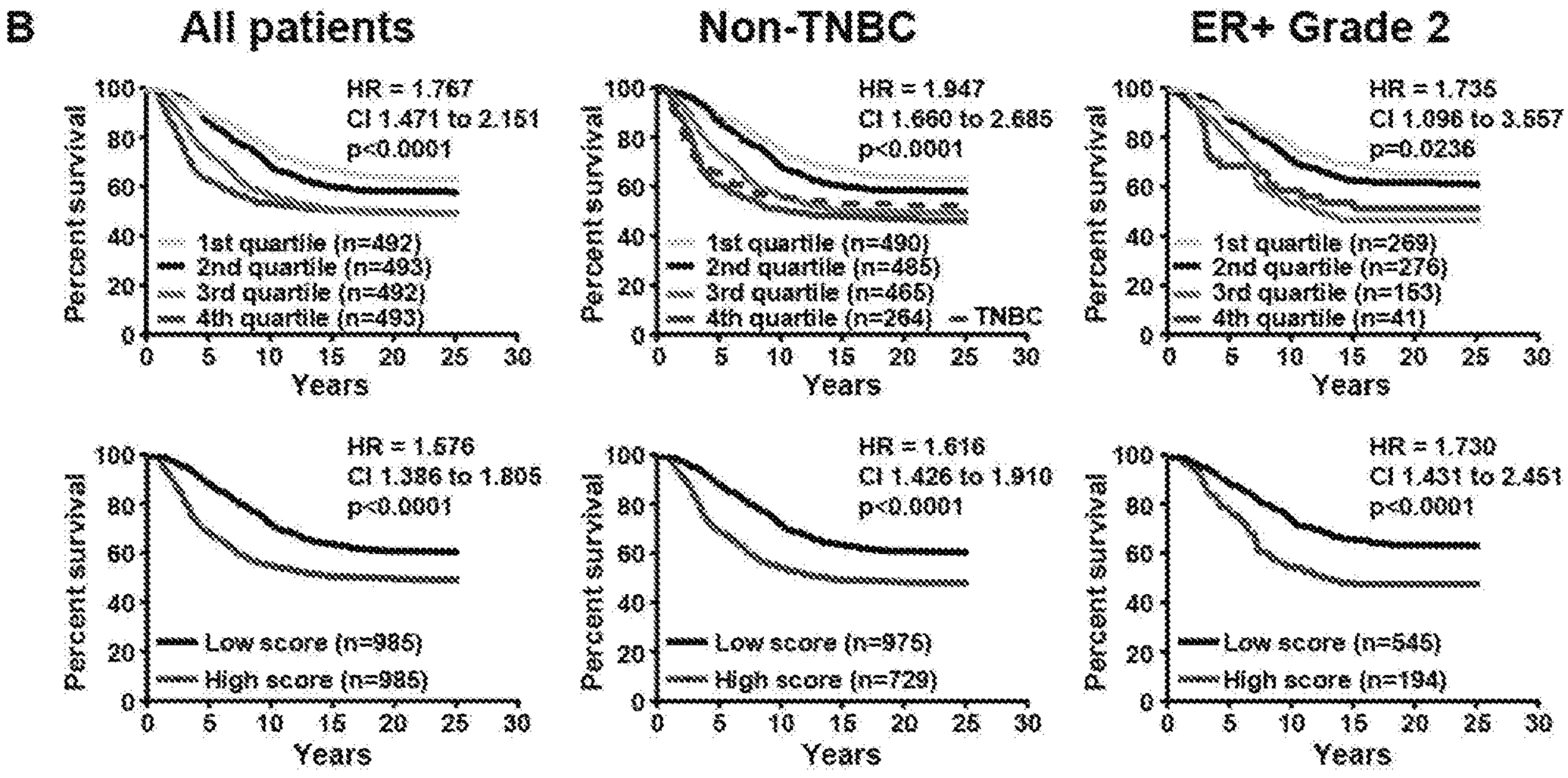

B
All patients
Non-TNBC
ER+ Grade 2
Percent survival
Years
HR = 1.767
CI 1.471 to 2.151
p<0.0001
1st quartile (n=492)
2nd quartile (n=493)
3rd quartile (n=492)
4th quartile (n=493)
HR = 1.947
CI 1.660 to 2.685
p<0.0001
1st quartile (n=490)
2nd quartile (n=485)
3rd quartile (n=465)
4th quartile (n=264)
TNBC
HR = 1.735
CI 1.096 to 3.557
p=0.0236
1st quartile (n=269)
2nd quartile (n=276)
3rd quartile (n=153)
4th quartile (n=41)
HR = 1.576
CI 1.386 to 1.805
p<0.0001
Low score (n=985)
High score (n=985)
HR = 1.616
CI 1.426 to 1.910
p<0.0001
Low score (n=975)
High score (n=729)
HR = 1.730
CI 1.431 to 2.461
p<0.0001
Low score (n=545)
High score (n=194)

FIG. 1 cont'd

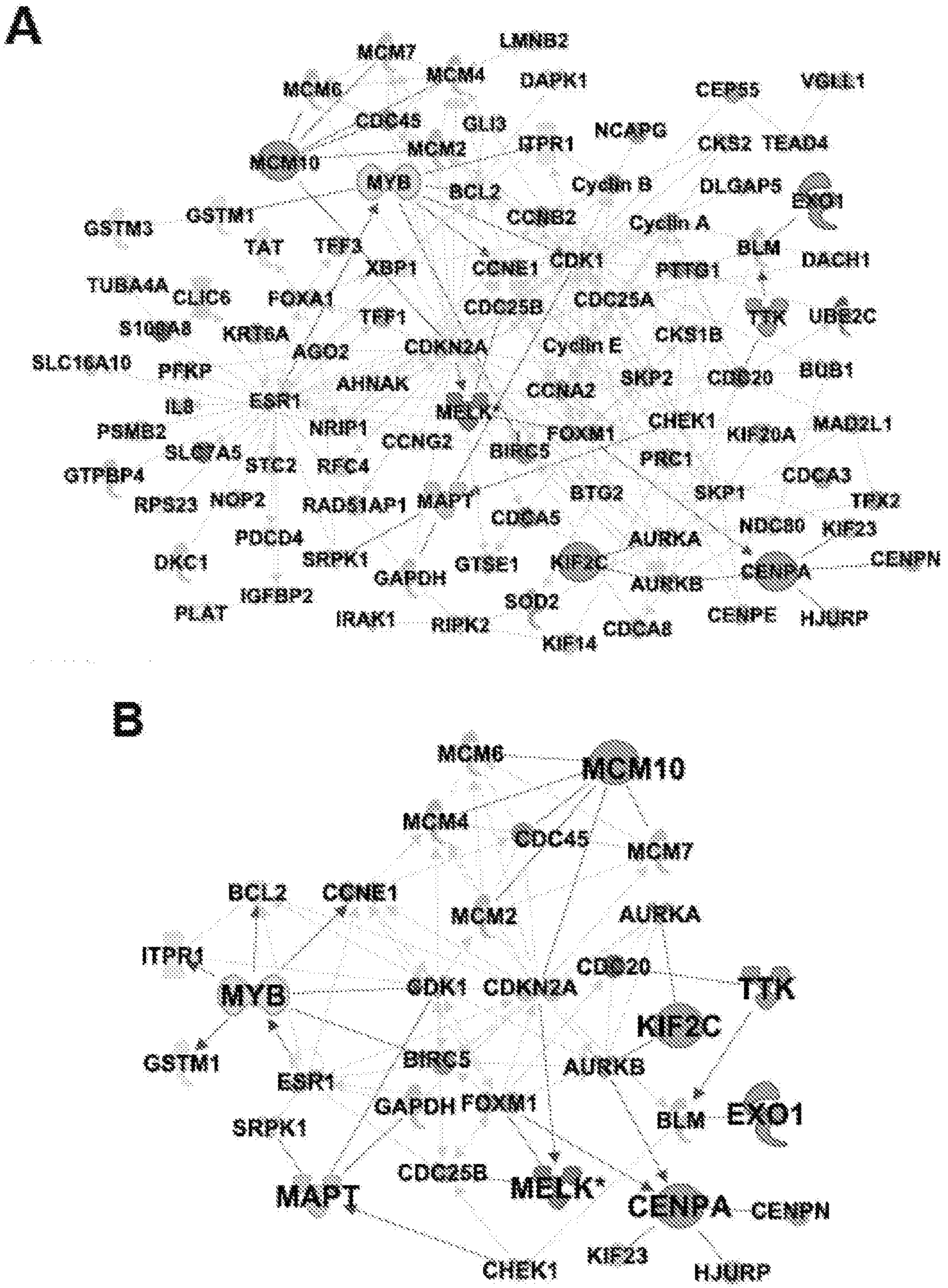

A
MCM7
LMNB2
MCM4
DAPK1
MCM6
CEP55
VGLL1
CDC45
GLI3
NCAPG
MCM2
ITPR1
CKS2
TEAD4
MCM10
MYB
BCL2
Cyclin B
DLGAP5
EXO1
GSTM1
GSTM3
CCNB2
Cyclin A
TAT
TFF3
BLM
DACH1
XBP1
CCNE1
CDK1
PTTG1
TUBA4A
CLIC6
FOXA1
CDC25B
CDC25A
TFF1
TTK
UBE2C
S100A8
KRT6A
CKS1B
AGO2
CDKN2A
Cyclin E
SLC16A10
PFKP
SKP2
CCNB2
BUB1
AHNAK
CCNA2
IL8
ESR1
MELK*
CHEK1
PSMB2
NRIP1
FOXM1
MAD2L1
CCNG2
KIF20A
SLC7A5
BIRC5
STC2
RFC4
PRC1
GTPBP4
CDCA3
BTG2
SKP1
RPS23
NOP2
RAD51AP1
MAPT
TPX2
CDCA5
NDC80
KIF23
PDCD4
AURKA
DKC1
SRPK1
GTSE1
KIF2C
CENPN
GAPDH
CENPA
AURKB
IGFBP2
SOD2
PLAT
IRAK1
RIPK2
CENPE
HJURP
CDCA8
KIF14
B
MCM6
MCM10
MCM4
CDC45
MCM7
BCL2
CCNE1
MCM2
AURKA
ITPR1
CCNB2
MYB
CDK1
CDKN2A
TTK
KIF2C
GSTM1
BIRC5
AURKB
ESR1
GAPDH
FOXM1
BLM
EXO1
SRPK1
CDC25B
MELK*
MAPT
CENPA
CENPN
KIF23
CHEK1
HJURP

FIG. 2

C

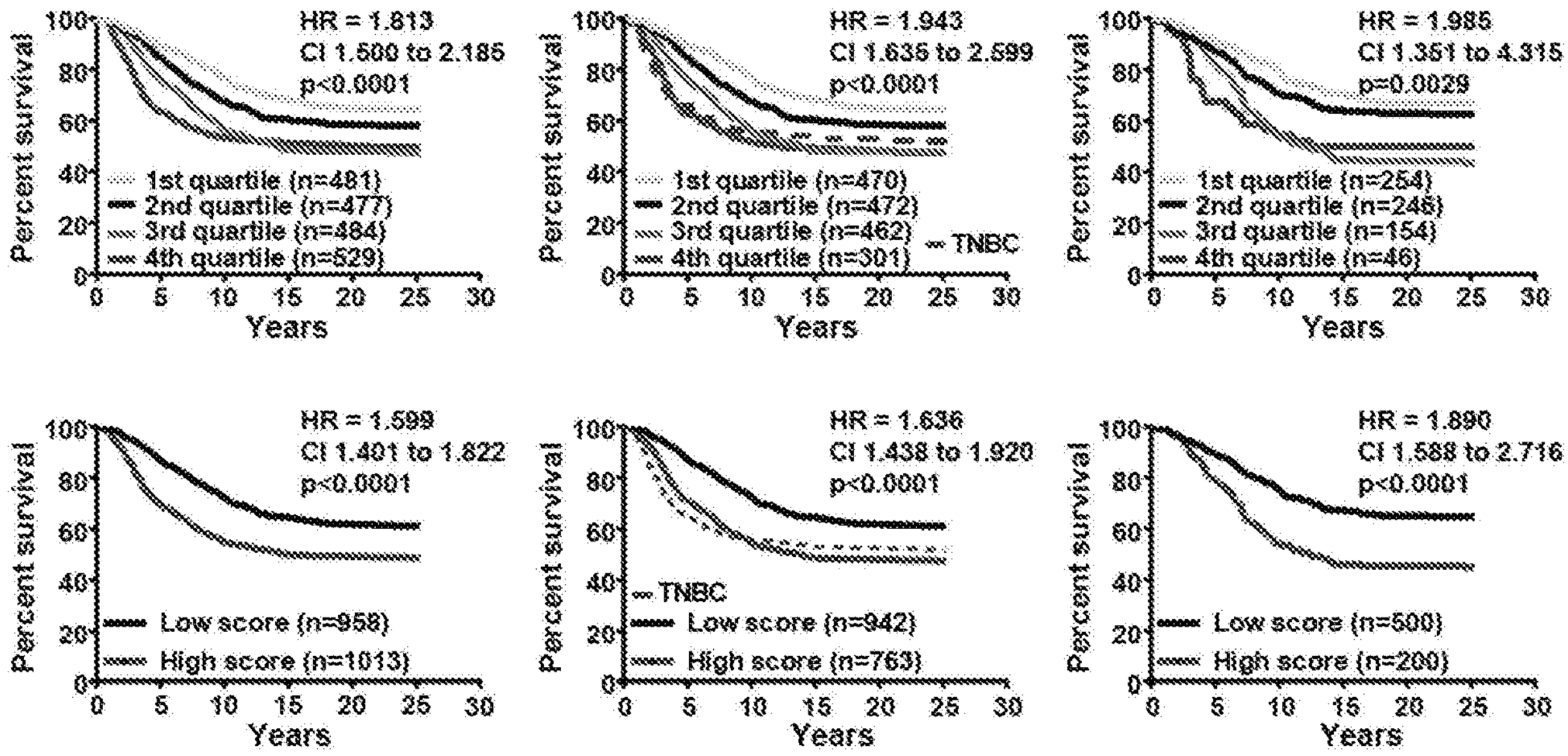

All patients
HR = 1.813
CI 1.500 to 2.185
p<0.0001
1st quartile (n=481)
2nd quartile (n=477)
3rd quartile (n=484)
4th quartile (n=529)
Non-TNBC
HR = 1.943
CI 1.635 to 2.599
p<0.0001
1st quartile (n=470)
2nd quartile (n=472)
3rd quartile (n=462)
4th quartile (n=301)
TNBC
ER+ Grade 2
HR = 1.985
CI 1.351 to 4.315
p=0.0029
1st quartile (n=254)
2nd quartile (n=248)
3rd quartile (n=154)
4th quartile (n=46)
HR = 1.599
CI 1.401 to 1.822
p<0.0001
Low score (n=958)
High score (n=1013)
HR = 1.636
CI 1.438 to 1.920
p<0.0001
TNBC
Low score (n=942)
High score (n=763)
HR = 1.890
CI 1.588 to 2.716
p<0.0001
Low score (n=500)
High score (n=200)
Percent survival
Years

FIG. 2 cont'd

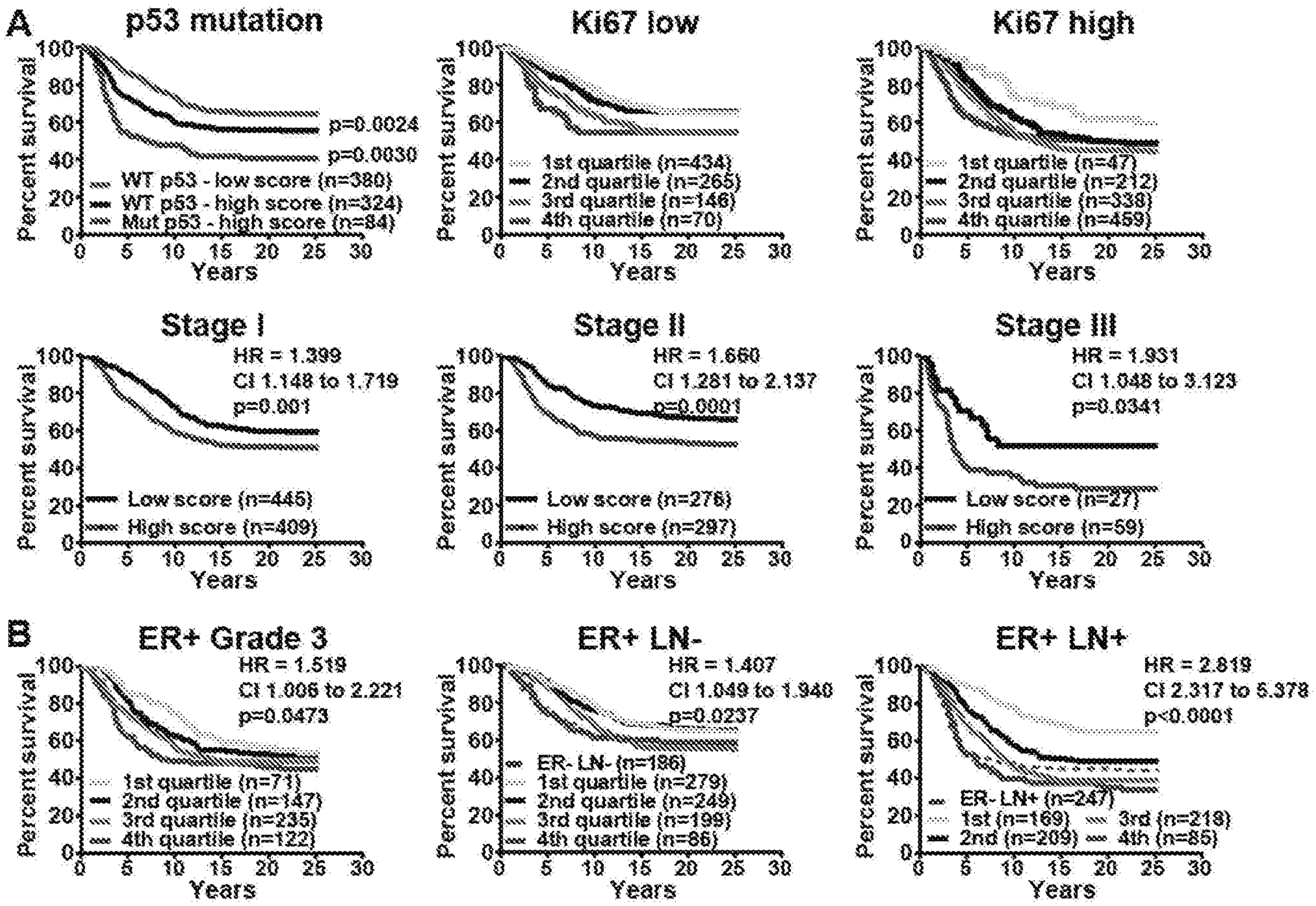
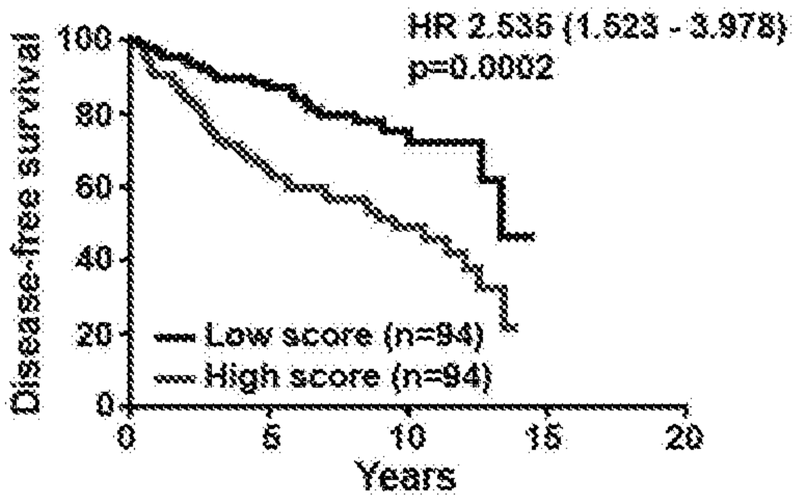

A
p53 mutation
Ki67 low
Ki67 high
Percent survival
Years
p=0.0024
p=0.0030
WT p53 - low score (n=380)
WT p53 - high score (n=324)
Mut p53 - high score (n=84)
1st quartile (n=434)
2nd quartile (n=265)
3rd quartile (n=146)
4th quartile (n=70)
1st quartile (n=47)
2nd quartile (n=212)
3rd quartile (n=338)
4th quartile (n=459)
Stage I
Stage II
Stage III
HR = 1.399
CI 1.148 to 1.719
p=0.001
Low score (n=445)
High score (n=409)
HR = 1.660
CI 1.281 to 2.137
p=0.0001
Low score (n=276)
High score (n=297)
HR = 1.931
CI 1.048 to 3.123
p=0.0341
Low score (n=27)
High score (n=59)
B
ER+ Grade 3
ER+ LN-
ER+ LN+
HR = 1.519
CI 1.006 to 2.221
p=0.0473
1st quartile (n=71)
2nd quartile (n=147)
3rd quartile (n=235)
4th quartile (n=122)
HR = 1.407
CI 1.049 to 1.940
p=0.0237
ER- LN- (n=186)
1st quartile (n=279)
2nd quartile (n=249)
3rd quartile (n=199)
4th quartile (n=86)
HR = 2.819
CI 2.317 to 5.378
p<0.0001
ER- LN+ (n=247)
1st (n=169)
2nd (n=209)
3rd (n=218)
4th (n=85)

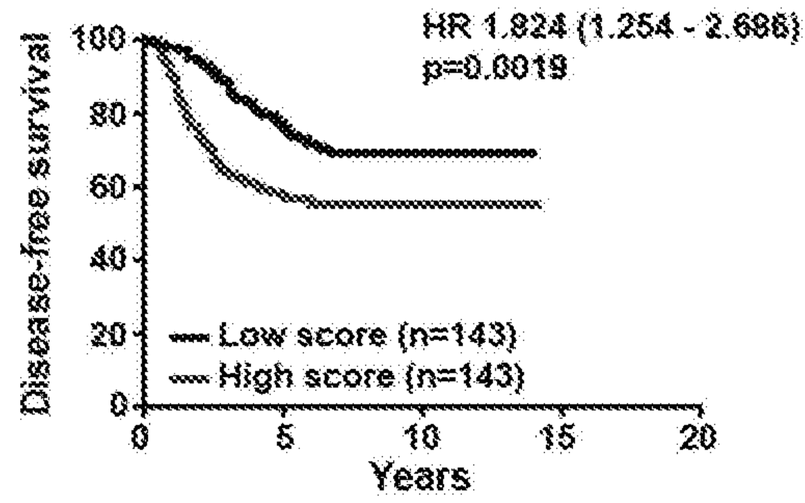
FIG. 3

A



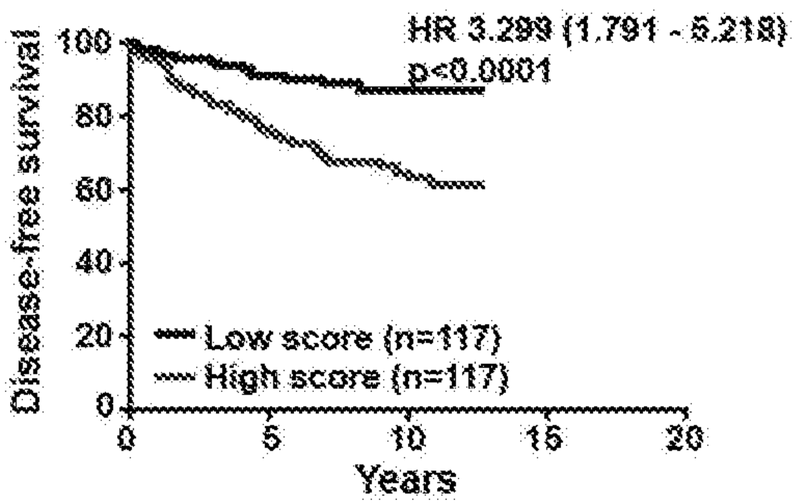
Covariate	HR (95% CI)	P-value
8 genes Score (high, low)	2.6267 (1.3685 - 5.0378)	0.0038
Lymph node (+, -)	0.9409 (0.4327 - 2.0459)	0.8785
Grade (1, 2, 3)	1.0276 (0.7178 - 1.4713)	0.8826
ER (+,-)	0.9833 (0.5223 - 1.8512)	0.9587

C



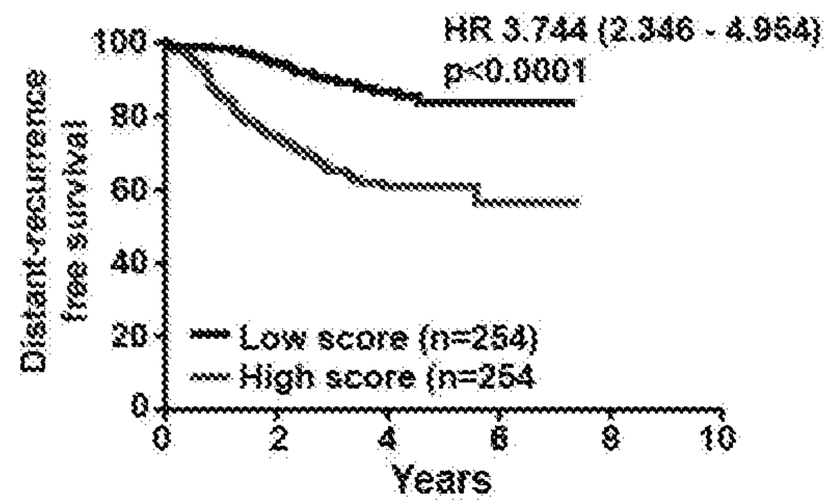
Covariate	HR (95% CI)	P-value
8 genes Score (high, low)	2.2123 (1.4438 - 3.3910)	0.0003
ER (+,-)	1.5586 (0.9834 - 2.5214)	0.072

B



Covariate	HR (95% CI)	P-value
8 genes Score (high, low)	2.1445 (1.0936 - 4.2176)	0.015
Lymph node (+, -)	2.8619 (1.5138 - 5.0751)	0.0003
ER (+,-)	0.5862 (0.2470 - 1.4394)	0.2526
Grade (1, 2, 3)	1.2947 (0.8032 - 2.0869)	0.2914
p53 (mut, WT)	1.2607 (0.6670 - 2.3825)	0.478

D



Covariate	HR (95% CI)	P-value
8 genes Score (high, low)	3.4442 (1.5914 - 7.4540)	0.0018
Lymph node (+, -)	5.4825 (1.7112 - 16.753)	0.0044
Tumor size (1, 2, 3)	1.6679 (1.0037 - 2.7718)	0.0496
ER (+,-)	1.6261 (0.7056 - 3.7473)	0.2561
Grade (1, 2, 3)	0.8233 (0.4696 - 1.4435)	0.4996
PR (+,-)	1.2716 (0.5866 - 2.7581)	0.5448

FIG. 4

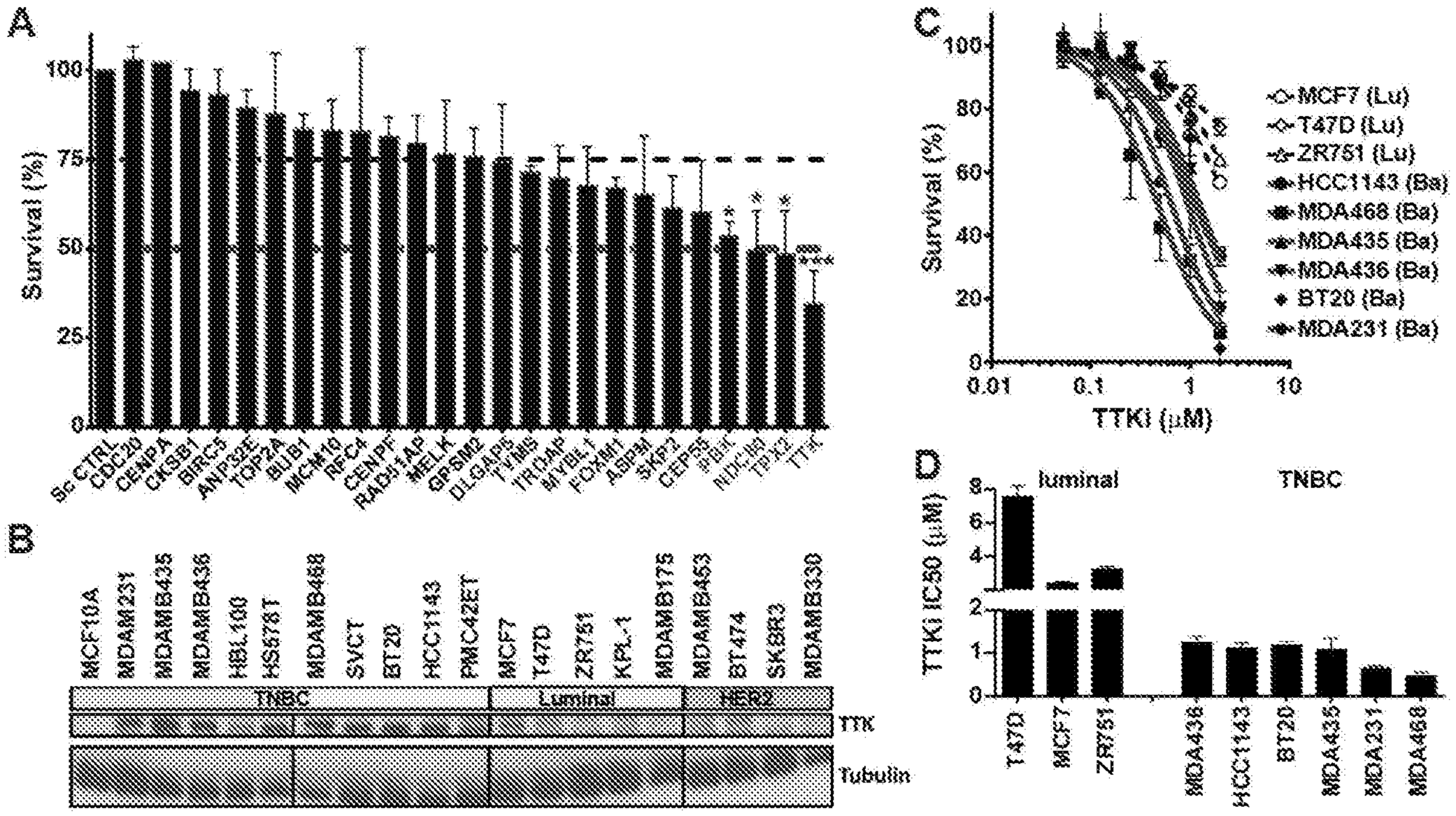

A
Survival (%)
100
75
50
25
0
Sc CTRL
CDC20
CENPA
CKSB1
BIRC5
ANP32E
TOP2A
BUB1
MCM10
RFC4
CENPF
RAD51AP
MELK
GPSM2
DLGAP5
TYMS
TROAP
MYBL1
FOXM1
ASPM
SKP2
CEP55
PBK
NDC80
TPX2
TTK
C
Survival (%)
100
80
60
40
20
0
0.01
0.1
1
10
TTKi (μM)
MCF7 (Lu)
T47D (Lu)
ZR751 (Lu)
HCC1143 (Ba)
MDA468 (Ba)
MDA435 (Ba)
MDA436 (Ba)
BT20 (Ba)
MDA231 (Ba)
D
MCF10A
MDAM231
MDAMB435
MDAMB436
HBL100
HS578T
MDAMB468
SVCT
BT20
HCC1143
PMC42ET
MCF7
T47D
ZR751
KPL-1
MDAMB175
MDAMB453
BT474
SK8R3
MDAMB330
TNBC
Luminal
HER2
TTK
Tubulin
B
TTKi IC50 (μM)
8
6
4
2
1
0
luminal
TNBC
T47D
MCF7
ZR751
MDA436
HCC1143
BT20
MDA435
MDA231
MDA468

FIG. 5

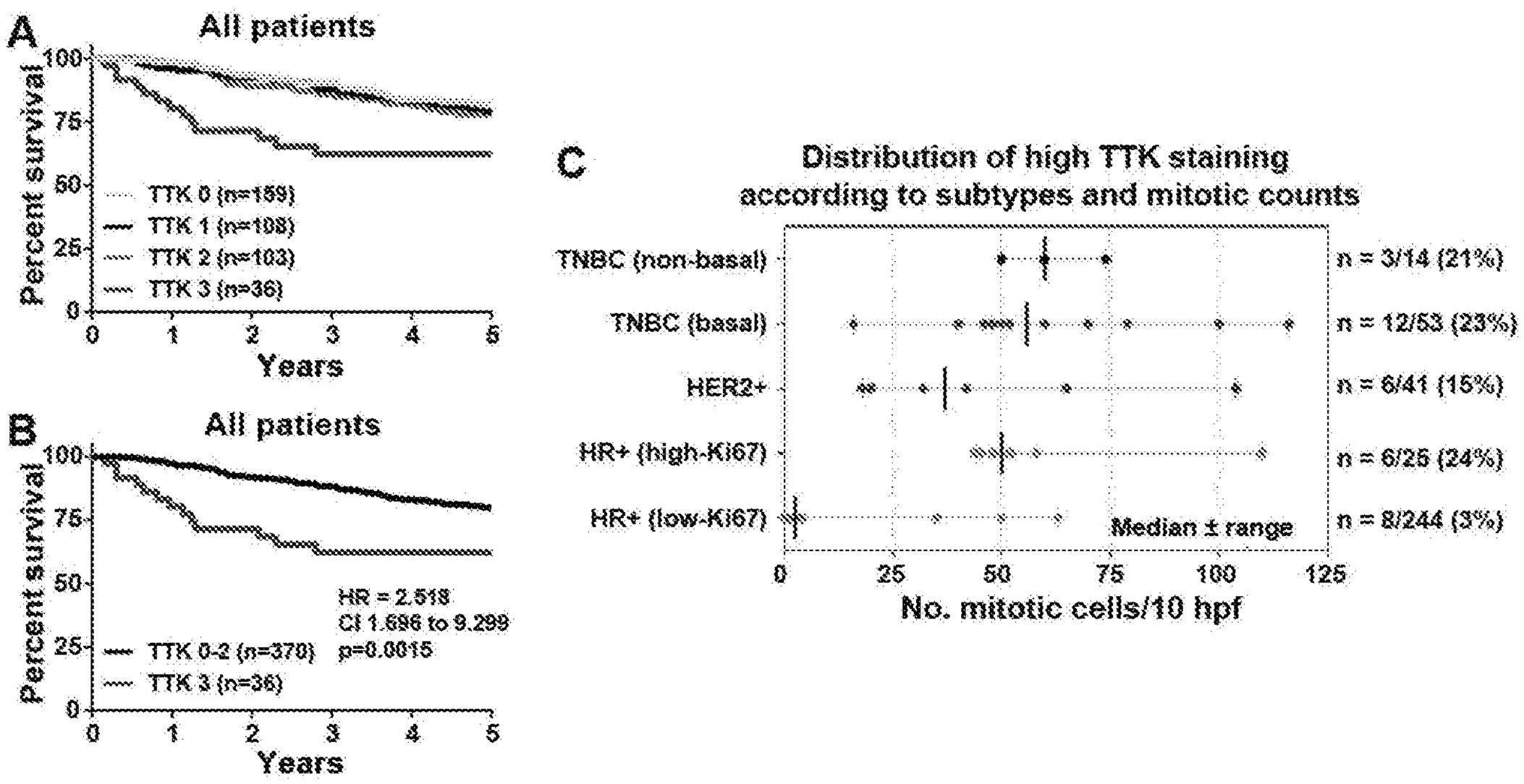

A
All patients
Percent survival
100
75
50
25
0
TTK 0 (n=159)
TTK 1 (n=108)
TTK 2 (n=103)
TTK 3 (n=36)
0
1
2
3
4
5
Years
B
All patients
Percent survival
100
75
50
25
0
HR = 2.518
CI 1.696 to 9.299
p=0.0015
TTK 0-2 (n=370)
TTK 3 (n=36)
0
1
2
3
4
5
Years
C
Distribution of high TTK staining
according to subtypes and mitotic counts
TNBC (non-basal)
TNBC (basal)
HER2+
HR+ (high-Ki67)
HR+ (low-Ki67)
n = 3/14 (21%)
n = 12/53 (23%)
n = 6/41 (15%)
n = 6/25 (24%)
n = 8/244 (3%)
Median ± range
0
25
50
75
100
125
No. mitotic cells/10 hpf

FIG. 6

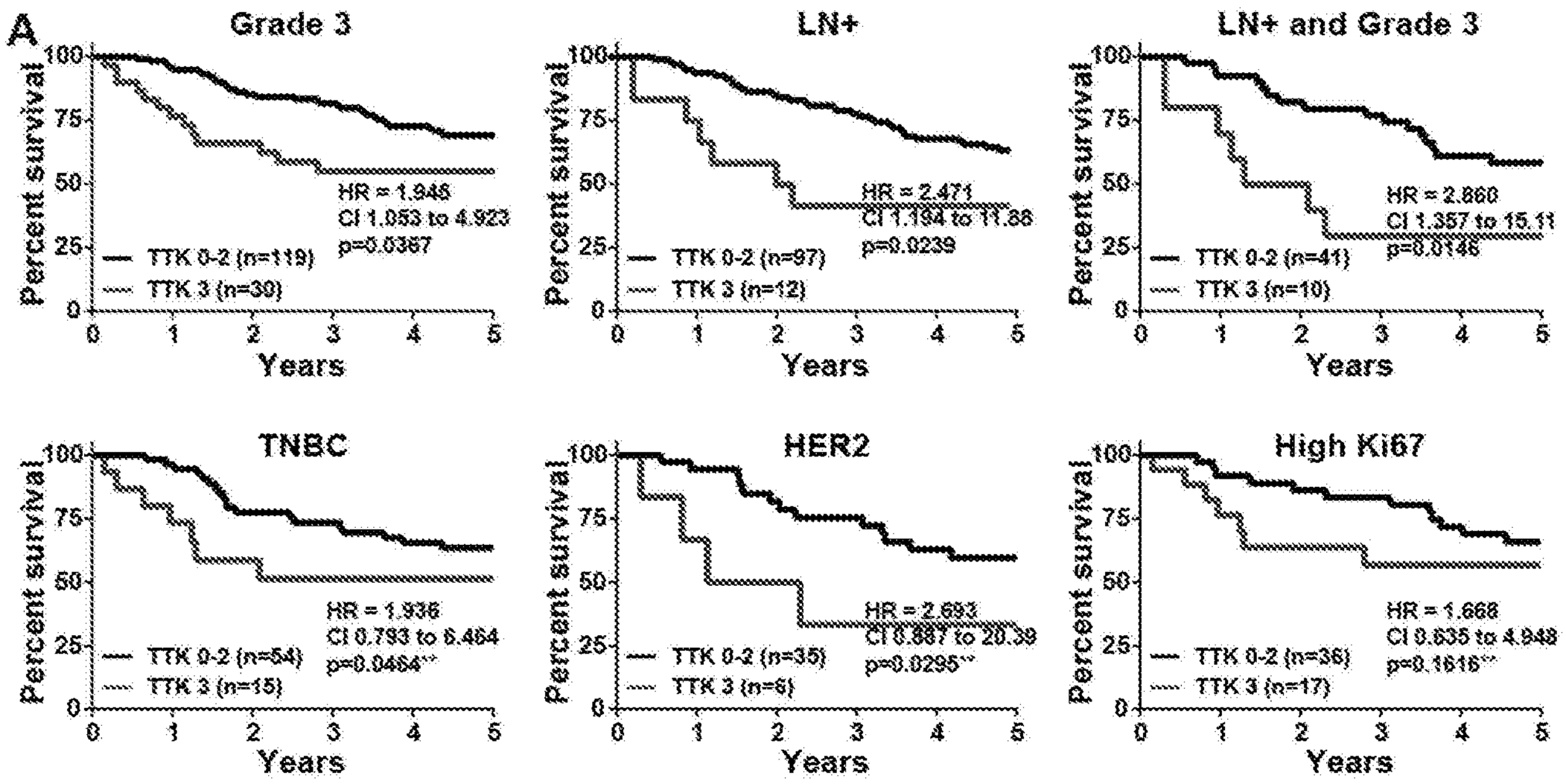
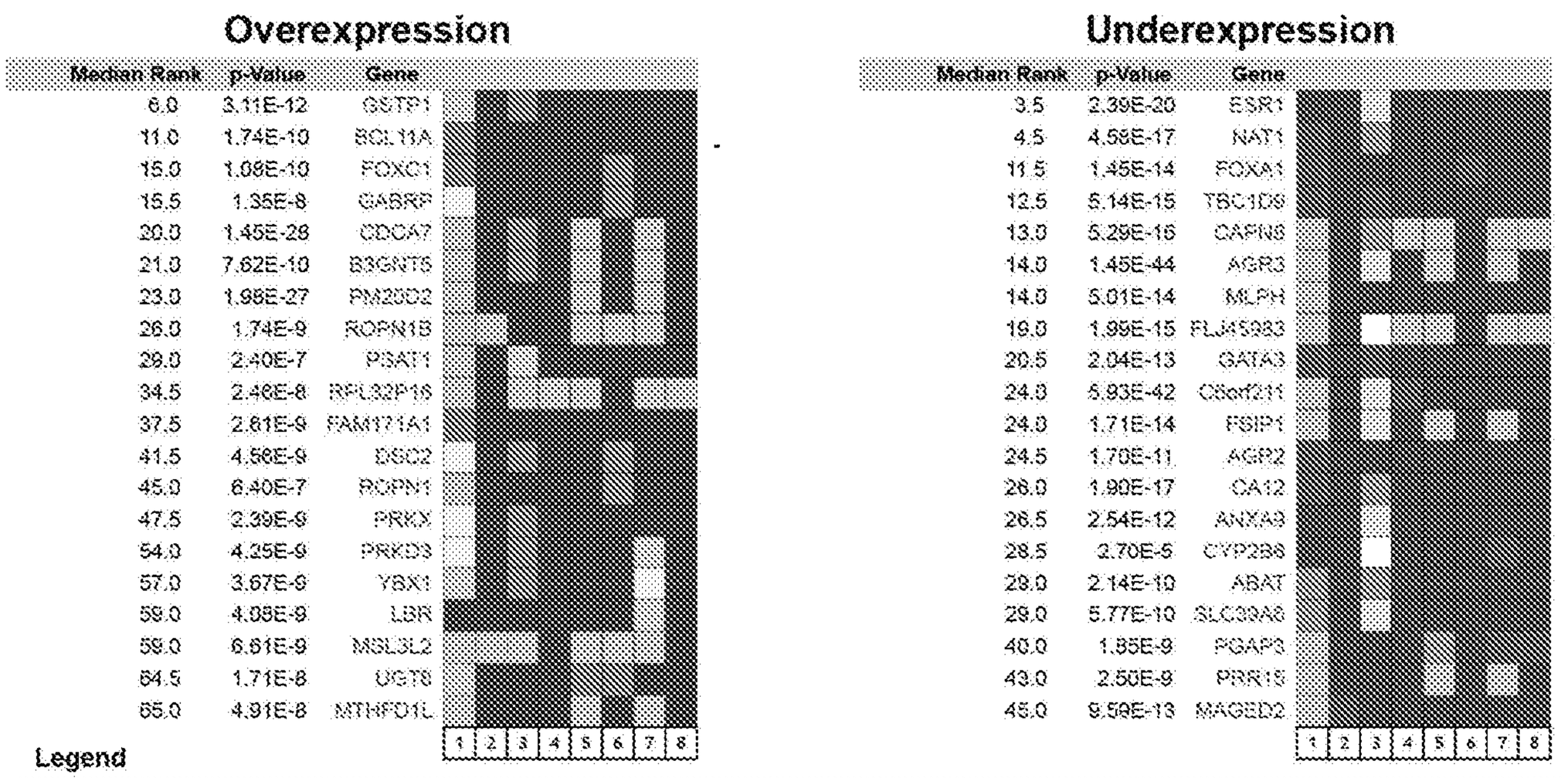

A
Grade 3
Percent survival
100
75
50
25
0
0 1 2 3 4 5
Years
HR = 1.946
CI 1.053 to 4.923
p=0.0367
TTK 0-2 (n=119)
TTK 3 (n=30)
LN+
HR = 2.471
CI 1.194 to 11.88
p=0.0239
TTK 0-2 (n=97)
TTK 3 (n=12)
LN+ and Grade 3
HR = 2.860
CI 1.357 to 15.11
p=0.0148
TTK 0-2 (n=41)
TTK 3 (n=10)
TNBC
HR = 1.936
CI 0.793 to 6.464
p=0.0464**
TTK 0-2 (n=54)
TTK 3 (n=15)
HER2
HR = 2.693
CI 0.887 to 20.39
p=0.0295**
TTK 0-2 (n=35)
TTK 3 (n=6)
High Ki67
HR = 1.668
CI 0.635 to 4.948
p=0.1616**
TTK 0-2 (n=36)
TTK 3 (n=17)
B
TNBC
non-TNBC
Survival (%)
Control
Doc (0.25 nM)
TTKi (250 nM)
Combination
MDA-MB-231
BT20
MDA-MB-468
MDA-MB-435
MDA-MB-436
MCF-7
T47D
ZR751
SKBR3
C
Apoptosis
(% Annexin V+/7-AAD-)
Control
Doc (0.25 nM)
TTKi (250 nM)
Combination

FIG. 7

A



Legend

1. Breast Carcinoma - ERBB2/ER/PR Negative
Bild Breast, Nature, 2006

2. Ductal Breast Carcinoma - ERBB2/ER/PR Negative
Bittner Breast, Not Published, 2005

3. Ductal Breast Carcinoma - ERBB2/ER/PR Negative
Bonnefoi Breast, Lancet Oncol, 2007

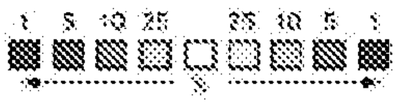
4. Invasive Breast Carcinoma - ERBB2/ER/PR Negative
Gluck Breast, Breast Cancer Res Treat, 2011

5. Invasive Breast Carcinoma - ERBB2/ER/PR Negative
Hatzis Breast, JAMA, 2011

6. Breast Carcinoma - ERBB2/ER/PR Negative
Kao Breast, BMC Cancer, 2011

7. Invasive Ductal Breast Carcinoma - ERBB2/ER/PR Negative
Tabchy Breast, Clin Cancer Res, 2010

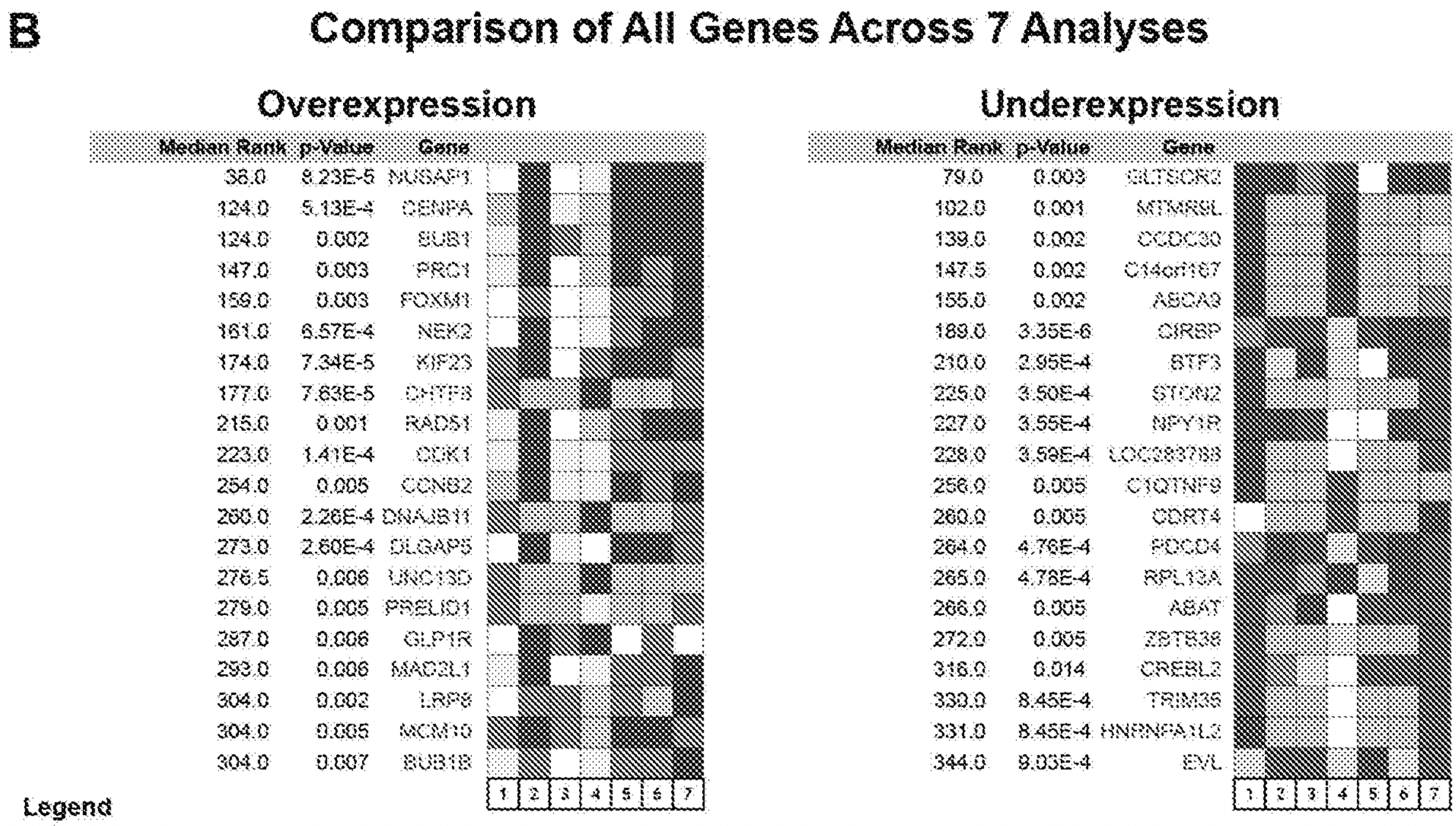
8. Invasive Ductal Breast Carcinoma - ERBB2/ER/PR Negative
TCGA Breast, No Associated Paper, 2011



The rank for a gene is the median rank for that gene across each of the analyses.
The p-Value for a gene is its p-Value for the median-ranked analysis

Copyright© 2006-11 Compendia Bioscience, Inc. Images from Oncomine™ may be used in publications with proper citation.
The citation for Oncomine is as follows: Oncomine™ (Compendia Bioscience, Ann Arbor, MI) was used for analysis and visualization.
For further information, refer to the terms described in the license agreement.

FIG. 8



Copyright© 2008-11 Compendia Bioscience, Inc. Images from Oncomine™ may be used in publications with proper citation.
The citation for Oncomine is as follows: Oncomine™ (Compendia Bioscience, Ann Arbor, MI) was used for analysis and visualization.
For further information, refer to the terms described in the license agreement.

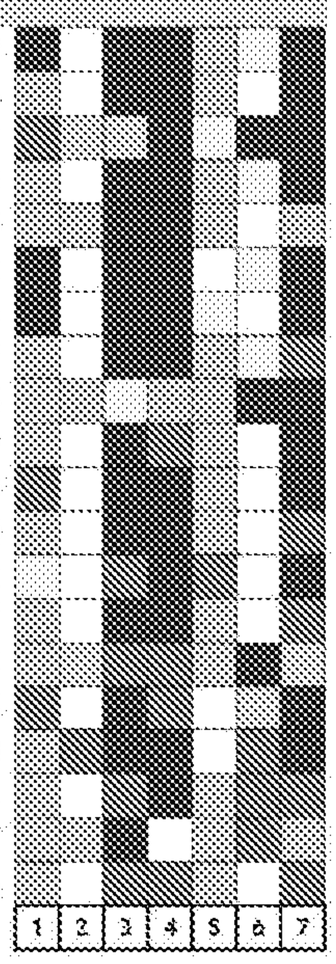
FIG. 8 cont'd

C

Comparison of All Genes Across 7 Analyses

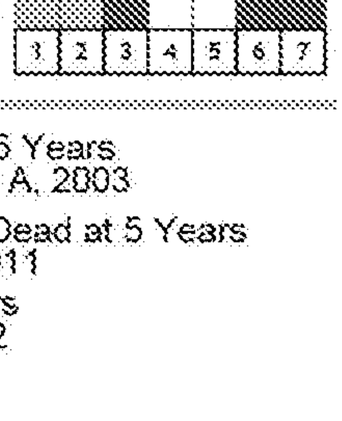
Overexpression

Median Rank	p-Value	Gene
109.5	1.48E-4	CCNB2
110.0	1.61E-4	KIF20A
126.0	0.005	PGK1
135.0	2.23E-4	RACGAP1
136.0	1.47E-4	CDCA5
146.0	2.45E-4	CCNB1
168.0	2.96E-4	FOXM1
171.0	2.96E-5	DEPDC1
171.0	0.003	PGAM5
184.0	2.36E-4	HJURP
187.0	0.004	CDC20
231.0	5.72E-5	MCM10
241.0	0.015	BUB1
242.0	6.04E-5	GTSE1
255.0	4.38E-4	ARFGAP1
257.0	0.010	BIRC5
266.0	0.008	CENPN
273.0	0.005	C16orf61
312.0	0.005	JMY
317.0	1.39E-4	KIF4A



Underexpression

Median Rank	p-Value	Gene
100.0	0.003	CBLN2
121.5	2.00E-4	DCDC5
181.5	2.78E-5	FGD3
195.0	0.002	TMEM26
214.0	6.88E-4	ZNF385B
230.0	0.003	RPL26
248.0	0.019	RERG
249.0	0.002	CCDC74B
255.0	8.65E-4	C5orf41
256.0	3.76E-4	SEC14L2
261.5	0.009	PAM105A
270.0	0.002	BEND5
271.0	0.011	NAT1
271.0	0.021	PTGER3
276.0	0.001	ERBB4
277.5	0.002	KIAA2026
280.0	0.008	MEIS3P1
290.0	0.002	PAPLN
307.5	0.004	SUSD3
311.0	0.001	ADHFE1



Legend

1. Breast Carcinoma - Dead at 5 Years
Bild Breast, Nature, 2006

2. Invasive Ductal Breast Carcinoma - Dead at 5 Years
Desmedt Breast, Clin Cancer Res, 2007

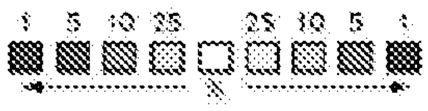
3. Breast Carcinoma - Dead at 5 Years
Kao Breast, BMC Cancer, 2011

4. Breast Carcinoma - Dead at 5 Years
Pawitan Breast, Breast Cancer Res, 2005

5. Ductal Breast Carcinoma - Dead at 5 Years
Sorlie Breast 2, Proc Natl Acad Sci U S A, 2003

6. Invasive Ductal Breast Carcinoma - Dead at 5 Years
TCGA Breast, No Associated Paper, 2011

7. Breast Carcinoma - Dead at 5 Years
vandeVijver Breast, N Engl J Med, 2002



The rank for a gene is the median rank for that gene across each of the analyses.
The p-Value for a gene is its p-Value for the median-ranked analysis.

Copyright© 2008-11 Compendia Bioscience, Inc. Images from Oncomine™ may be used in publications with proper citation.
The citation for Oncomine is as follows: Oncomine™ (Compendia Bioscience, Ann Arbor, MI) was used for analysis and visualization.
For further information, refer to the terms described in the license agreement.

FIG. 8 cont'd

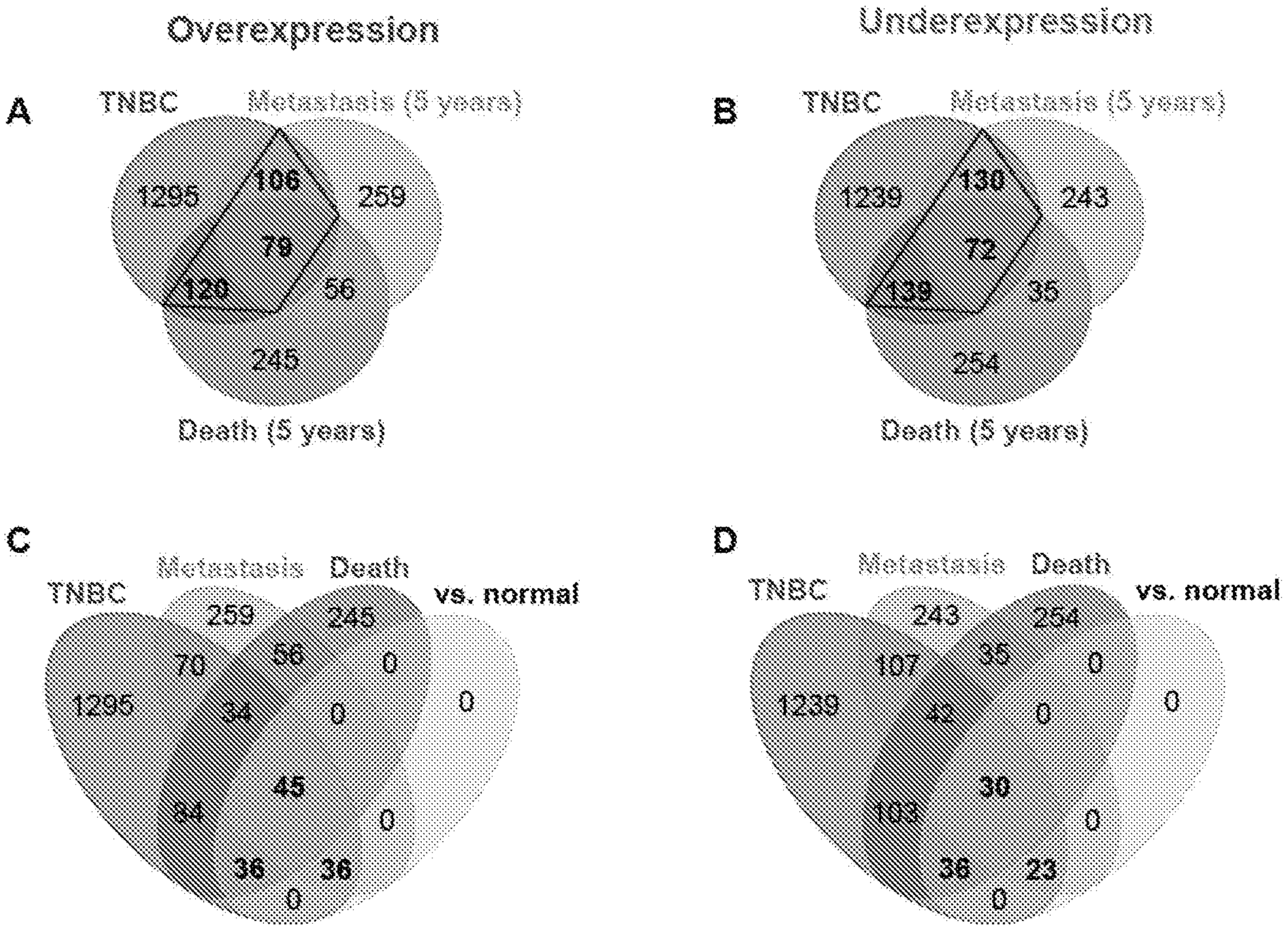
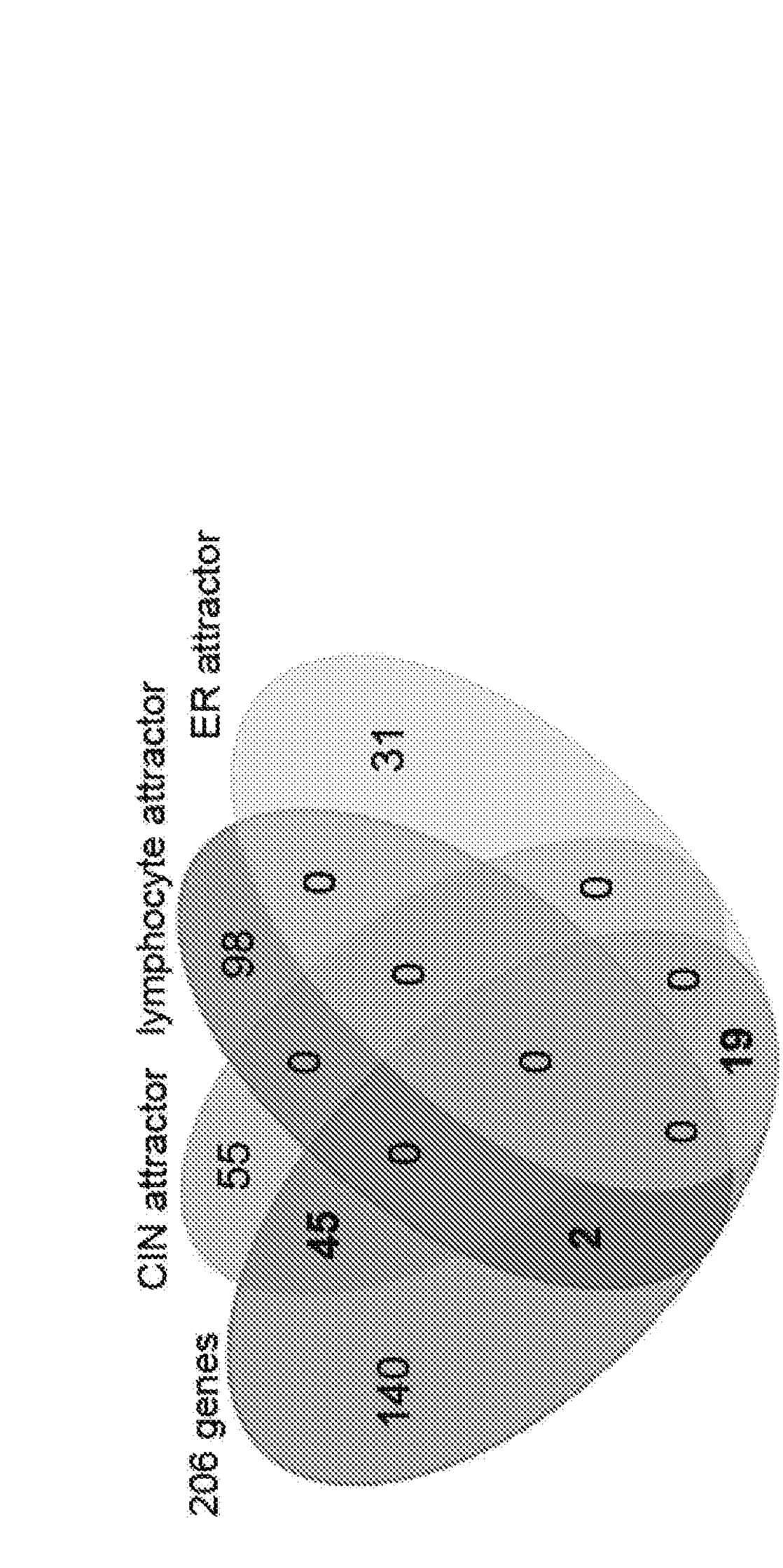

Overexpression
A
TNBC
Metastasis (5 years)
1295
106
259
79
120
56
245
Death (5 years)
Underexpression
B
TNBC
Metastasis (5 years)
1239
130
243
72
139
35
254
Death (5 years)
C
TNBC
Metastasis
Death
vs. normal
259
245
70
56
0
1295
34
0
0
45
84
0
36
36
0
D
TNBC
Metastasis
Death
vs. normal
243
254
107
35
0
1239
42
0
0
30
103
0
36
23
0

FIG. 9



Gene lists	total	Gene
206 genelist AND CIN attractor	45	AURKB FAM83D TPX2 CENPE EXO1 CDCA3 ANLN KIF2C BIRC5 KIF20A MCM10 STIL UBE2C FOXM1 CCNB2 PRC1 CDK1 CENPA CHEK1 MELK CENPW CEP55 CDC20 CDC45 BUB1 NCAPG TRIP13 PTTG1 CDCA5 AURKA FAM64A NUSAP1 KIF14 MAD2L1 DLGAP5 KIF23 DEPDC1 HJURP RAD51AP1 CDCA8
206 genelist AND lymphocyte specific attractor	2	CD163 FPR3
206 genelist AND estrogen receptor attractor	19	CCDC170 FAM214A AGR3 SCUBE2 DACH1 GFRA1 FOXA1 ESR1 DNAJC12 MLPH TBC1D9 TFF1 ILEST TFF3 MAPT XBP1 BNALI1 [illegible] SLC39A6
206 genelist Unique (cell cycle enriched based on gene set enrichment analysis, GSEA)	140	PTGER3 KIF13B KCNMA1 SH3BGRL C1orf106 BTG2 CKS1B STK32B BYSL PFKP CKAP2L KCNG1 IRAK1 GSHSYTL4 TUBA4A ATP6V1C2 RAB27B GPSM2 CMYA5 SOD2 AQP3 NOSTRIN BLM AGO2 SRPK1 PSMB2 OGN NFIA PLCH1 IL6 ELOVL5 PIP CDC25B GSTM1 CYB5D1 MCM6 FAM198B CASC1 CIRBP CSTB LMNB2 CPEB2 RPS23 NOVA1 CA9 RPP40 PNPLA7 NUP93 PDCD4 SLC7A5 HSD17B8 PLAT ALDH3A2 RERG TAT SOX11 SKP1 CARD10 NUP205 GTPBP4 CMC2 MEIS3P1 CCNE1 IGFBP2 STC2 IMPA2 MCM7 QDPR SRD5A1 CCDC176 CDCA7 YEATS2 LRP8 BCL2 SKP2 IGFBP4 SLC40A1 LRIG1 PTPRT AZGP1 KCTD3 GAPDH HSD17B4 AUNIP DKC1 LAD1 NOP2 CDKN2A GL13 LOC100286903 CYBRD1 CTSV MST1 RFC4 PHYHD1 S100A9 NRIP1 CCNG2 BBS1 MTHFD1L LFNG CST3 KRT6A GPD1L USB1 TMEM26 DAPK1 KIF5C RABEP1 MCM2 LAPTM4B HRASLS MCM4 RNASE4 CLIC6 APOBEC3B C18orf56 TRDAP NME5 ECI2 GSTM3 C18orf32 SLC2A1 RIPK2 ADIRF C1orf21 SLC16A10 AHNAK CKS2 VGLL1 MX2 TEAD4 PNP RBM36 GLYATL2 C7orf63 LYRD6 AFF3 CDC25A

FIG. 10

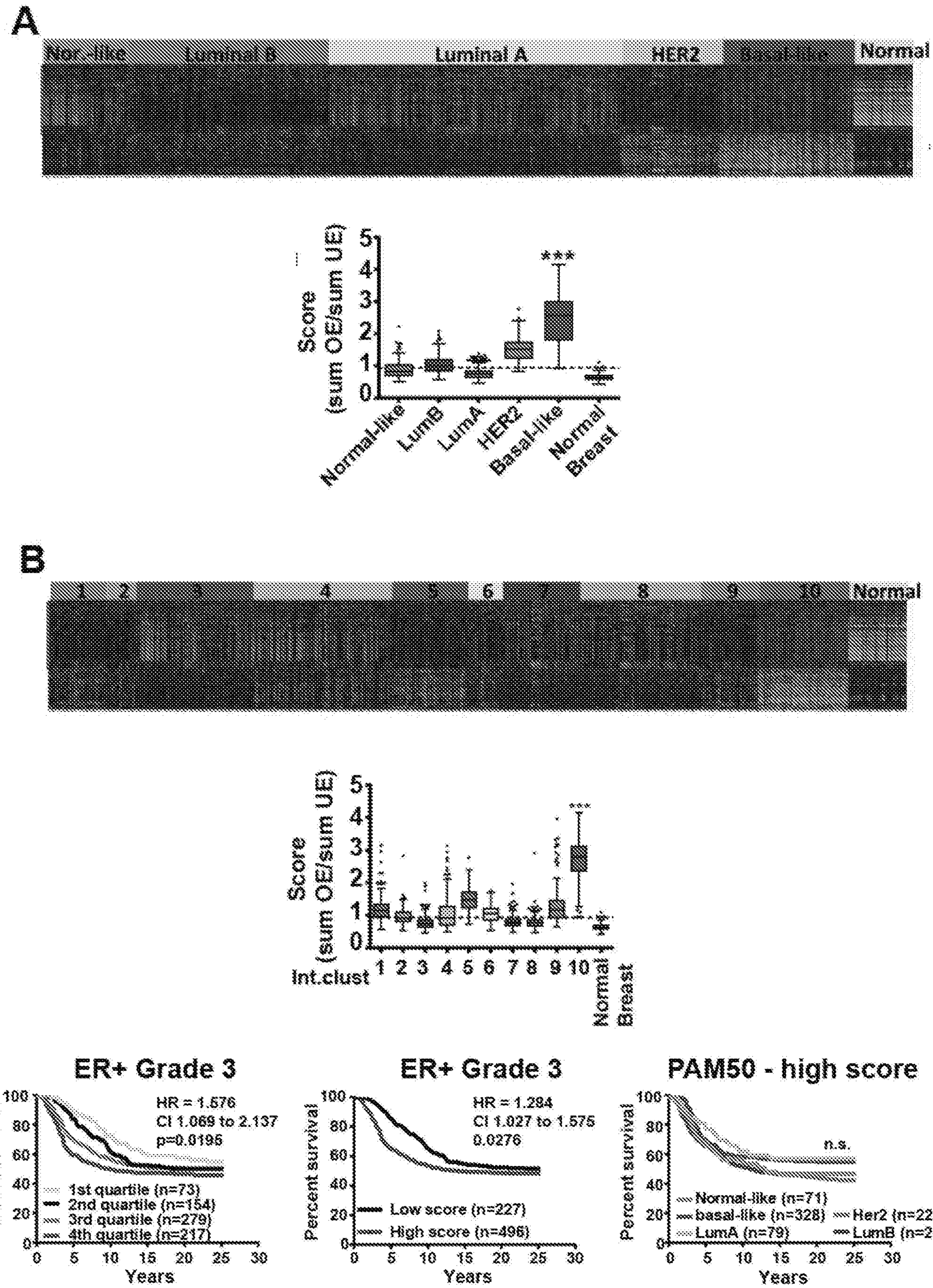

A
Normal
B
Normal
Score (sum OE/sum UE)
Normal-like
LumB
LumA
HER2
Basal-like
Normal Breast
Int.clust
ER+ Grade 3
HR = 1.576
CI 1.069 to 2.137
p=0.0196
1st quartile (n=73)
2nd quartile (n=154)
3rd quartile (n=279)
4th quartile (n=217)
Years
ER+ Grade 3
HR = 1.294
CI 1.027 to 1.575
0.0276
Low score (n=227)
High score (n=496)
Years
PAM50 - high score
n.s.
Normal-like (n=71)
basal-like (n=328)
Her2 (n=226)
LumA (n=79)
LumB (n=279)
Years

FIG. 11

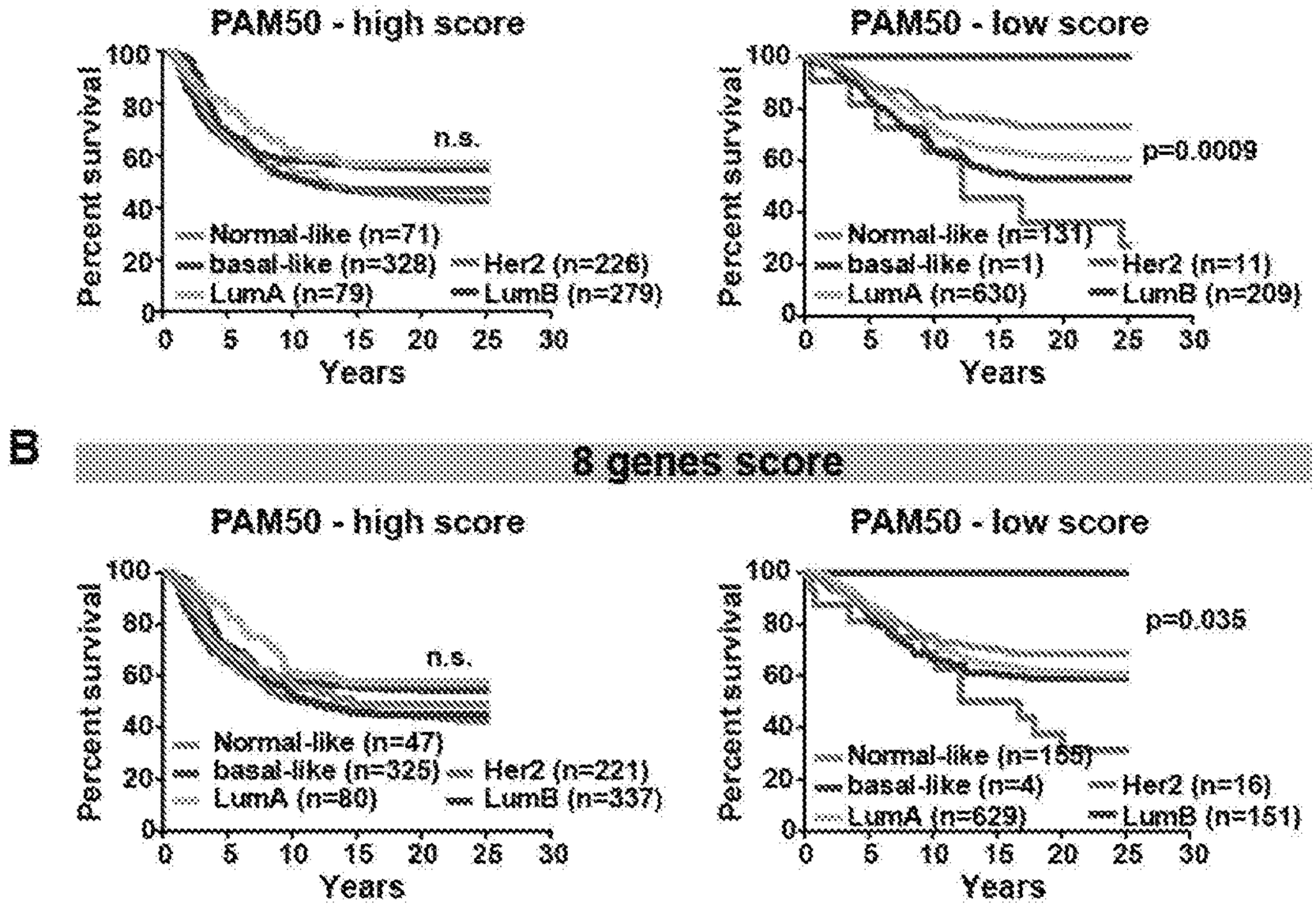
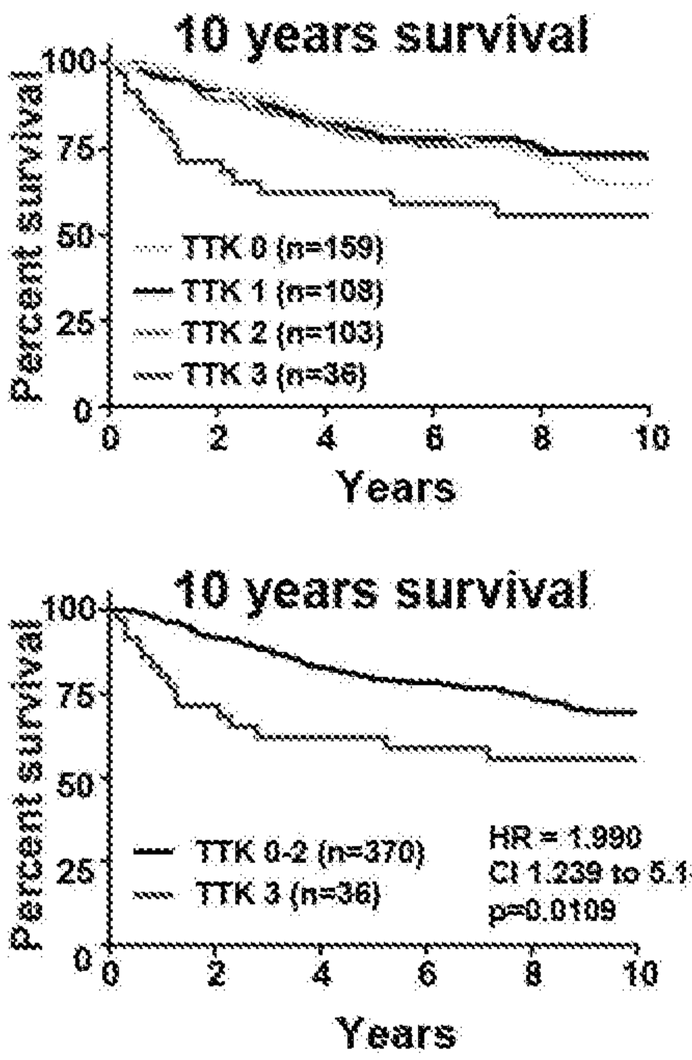
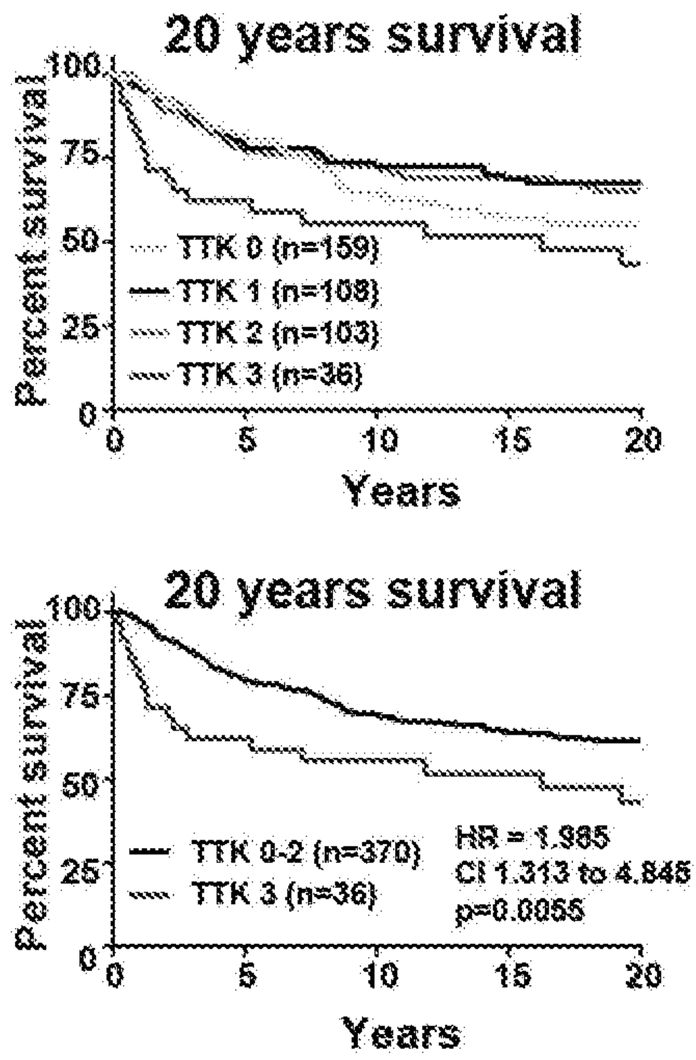

A
206 genes score
PAM50 - high score
Percent survival
100
80
60
40
20
0
n.s.
Normal-like (n=71)
basal-like (n=328)
Her2 (n=226)
LumA (n=79)
LumB (n=279)
0 5 10 15 20 25 30
Years
PAM50 - low score
p=0.0009
Normal-like (n=131)
basal-like (n=1)
Her2 (n=11)
LumA (n=630)
LumB (n=209)
B
8 genes score
PAM50 - high score
n.s.
Normal-like (n=47)
basal-like (n=325)
Her2 (n=221)
LumA (n=80)
LumB (n=337)
PAM50 - low score
p=0.035
Normal-like (n=155)
basal-like (n=4)
Her2 (n=16)
LumA (n=629)
LumB (n=151)

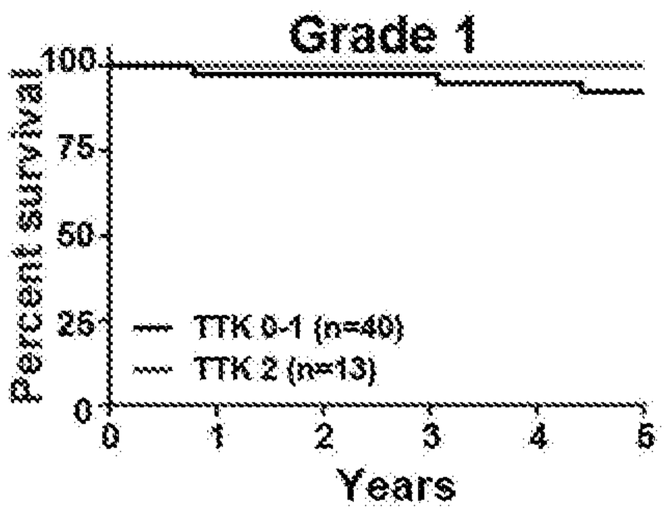
FIG. 12

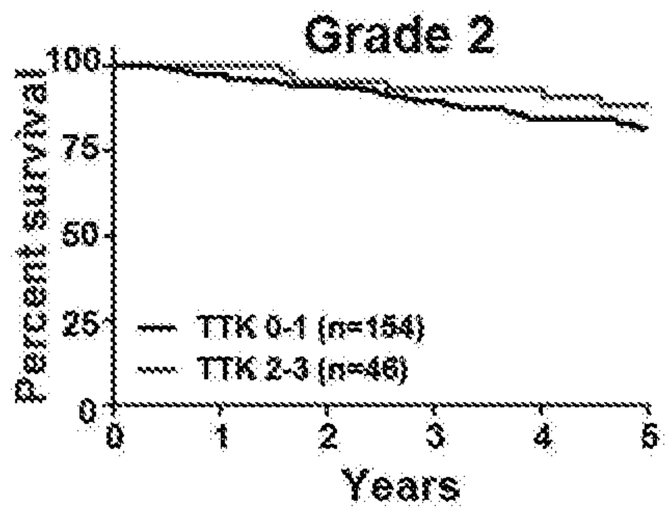

10 years survival
Percent survival
100
75
50
25
0
TTK 0 (n=159)
TTK 1 (n=108)
TTK 2 (n=103)
TTK 3 (n=36)
0
2
4
6
8
10
Years


20 years survival
Percent survival
100
75
50
25
0
TTK 0 (n=159)
TTK 1 (n=108)
TTK 2 (n=103)
TTK 3 (n=36)
0
5
10
15
20
Years

10 years survival
Percent survival
100
75
50
25
0
TTK 0-2 (n=370)
TTK 3 (n=36)
HR = 1.990
CI 1.239 to 5.184
p=0.0109
0
2
4
6
8
10
Years

20 years survival
Percent survival
100
75
50
25
0
TTK 0-2 (n=370)
TTK 3 (n=36)
HR = 1.985
CI 1.313 to 4.845
p=0.0055
0
5
10
15
20
Years


Grade 1
Percent survival
100
75
50
25
0
TTK 0-1 (n=40)
TTK 2 (n=13)
0
1
2
3
4
5
Years


Grade 2
Percent survival
100
75
50
25
0
TTK 0-1 (n=154)
TTK 2-3 (n=46)
0
1
2
3
4
5
Years

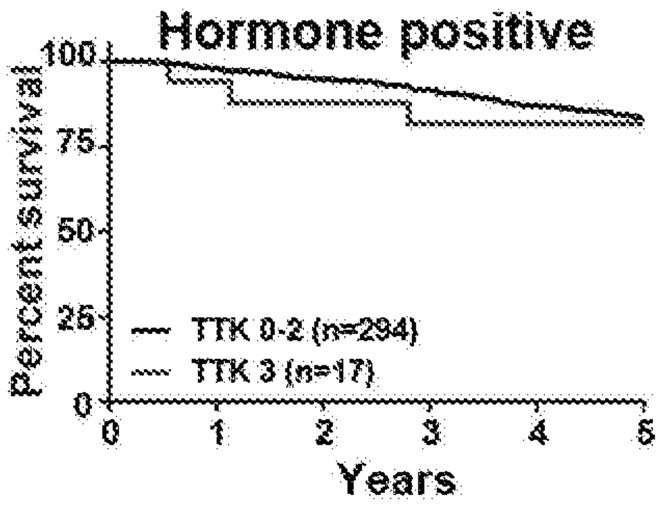

Hormone positive
Percent survival
100
75
50
25
0
TTK 0-2 (n=294)
TTK 3 (n=17)
0
1
2
3
4
5
Years

FIG. 13

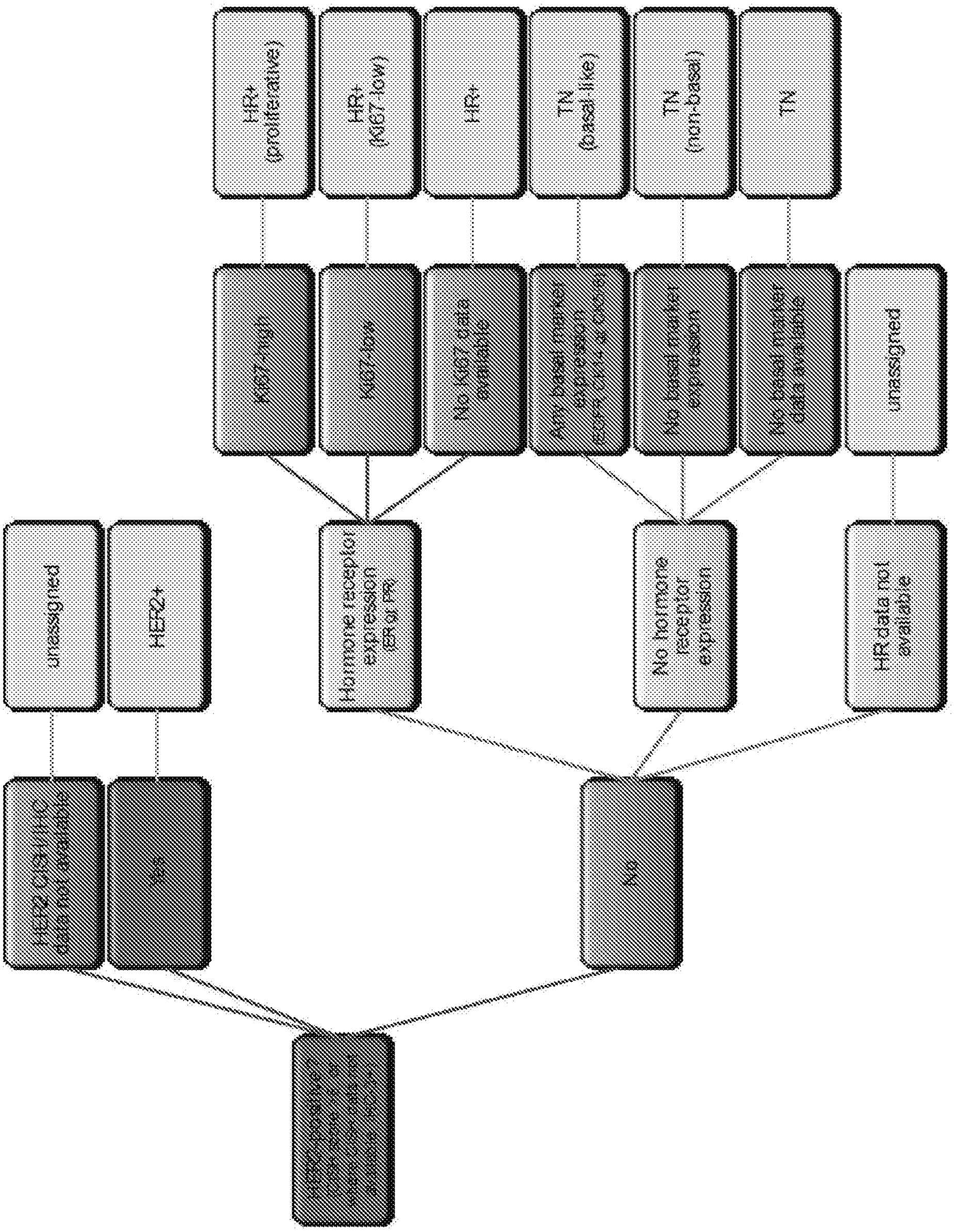
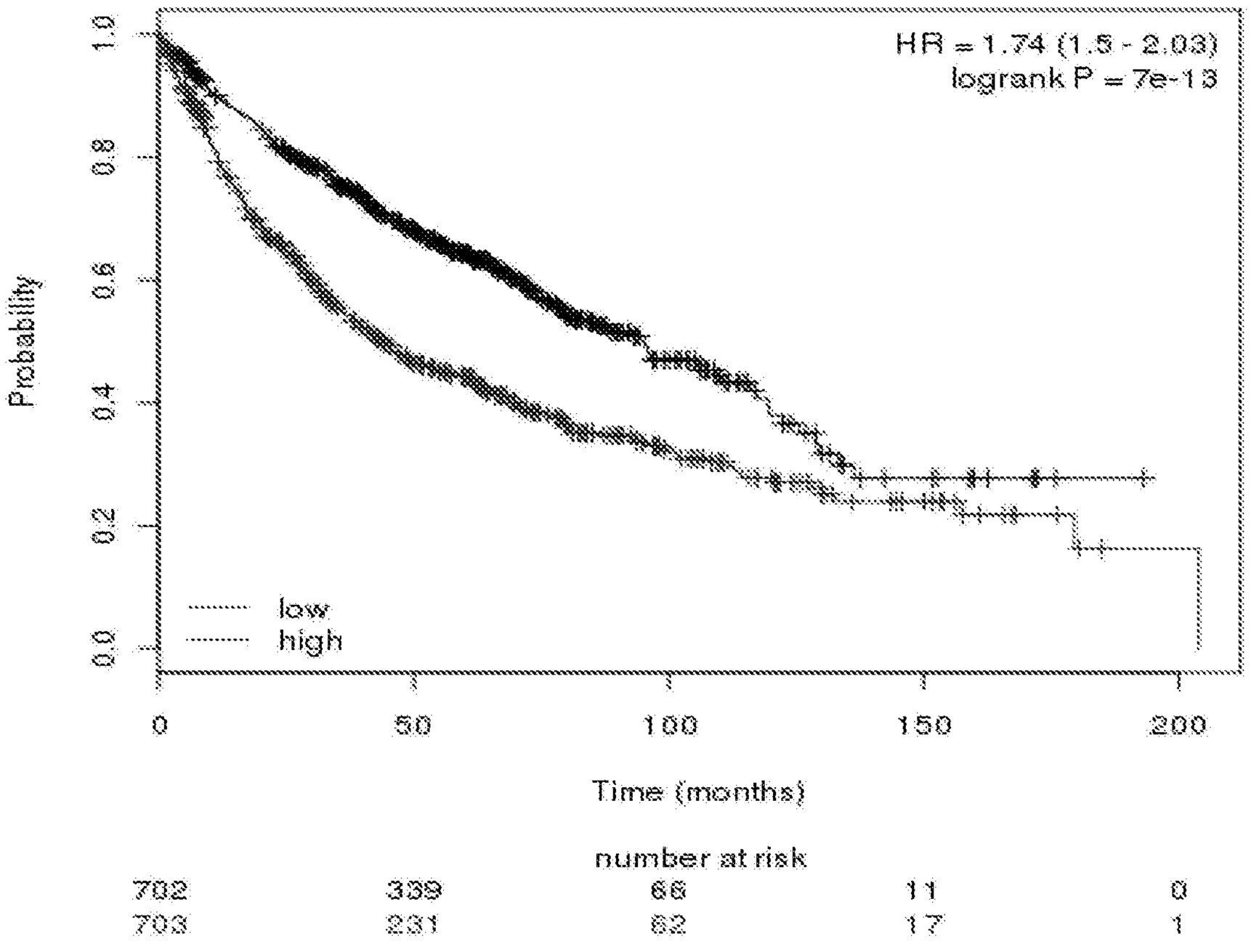

HR+ (proliferative)
HR+ (Ki67-low)
HR+
TN (basal-like)
TN (non-basal)
TN
Ki67-high
No Ki67 data available
Any basal marker expression (ECs, CK5/6 or EGFR)
No basal marker expression
Ki67 basal marker data available
unassigned
Hormone receptor expression (ER or PR)
No hormone receptor expression
HR data not available
unassigned
HER2+
HER2 GISH/IHC data not available
HER2
HER2-
HER2 GISH/IHC data not available

FIG. 14


HR = 1.74 (1.5 - 2.03)
logrank P = 7e-13
Probability
1.0
0.8
0.6
0.4
0.2
0.0
low
high
0
50
100
150
200
Time (months)
number at risk
702 339 66 11 0
703 231 62 17 1

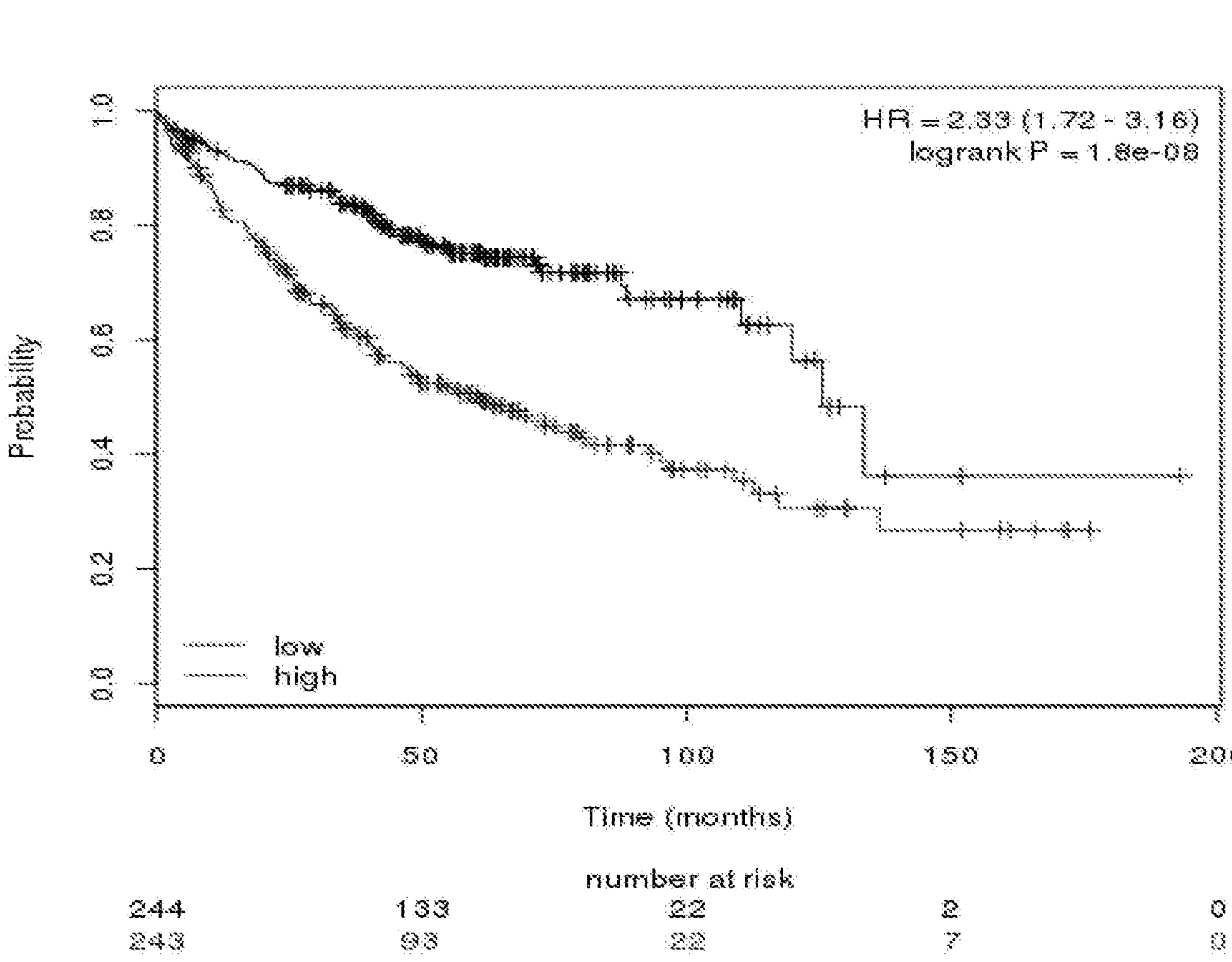

HR = 2.33 (1.72 - 3.16)
logrank P = 1.8e-08
Probability
1.0
0.8
0.6
0.4
0.2
0.0
low
high
0
50
100
150
200
Time (months)
number at risk
244 133 22 2 0
243 93 22 7 0

FIG. 15

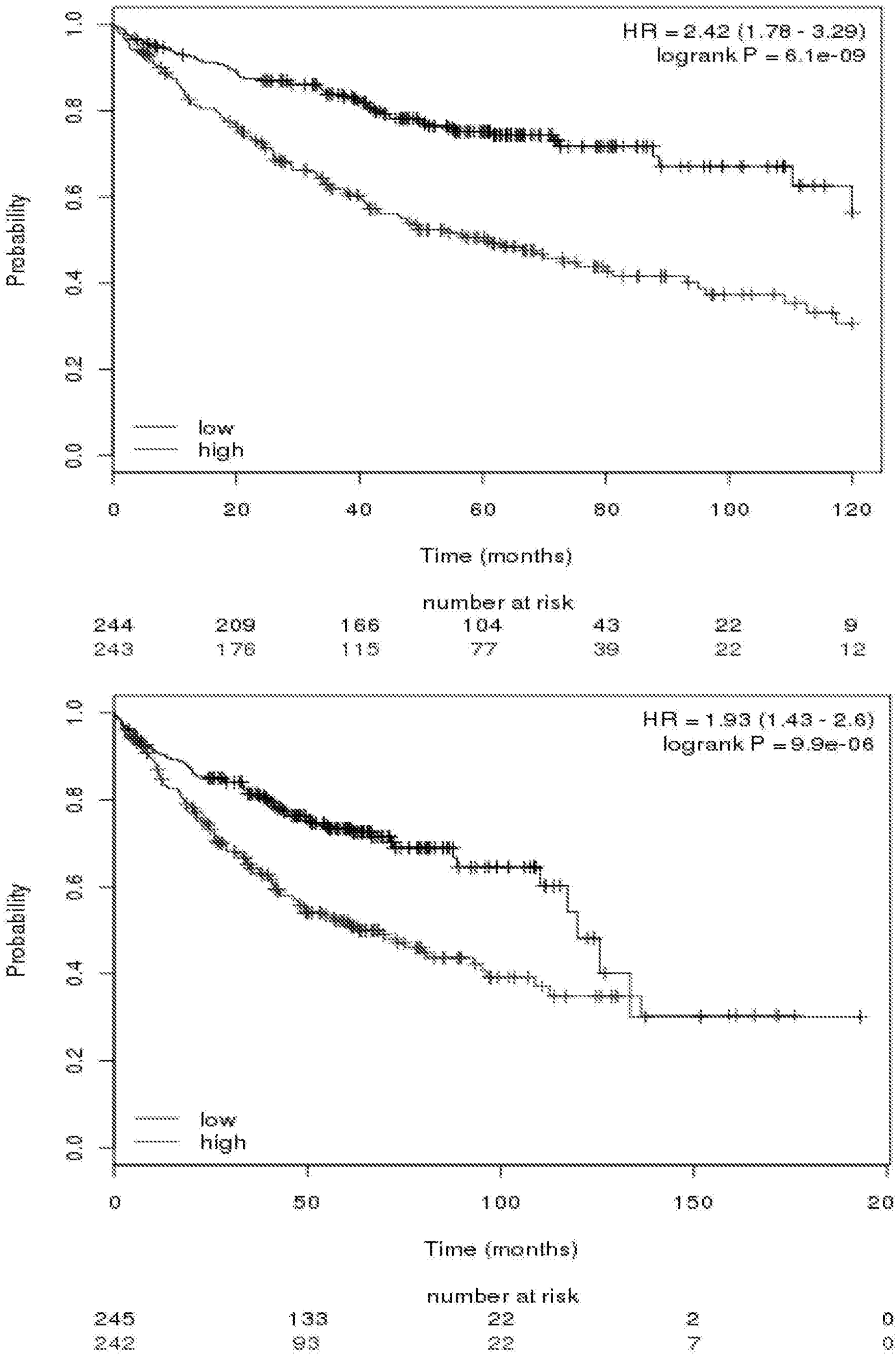

HR = 2.42 (1.78 - 3.29)
logrank P = 6.1e-09
Probability
1.0
0.8
0.6
0.4
0.2
0.0
low
high
0
20
40
60
80
100
120
Time (months)
number at risk
244 209 166 104 43 22 9
243 176 115 77 39 22 12
HR = 1.93 (1.43 - 2.6)
logrank P = 9.9e-06
Probability
1.0
0.8
0.6
0.4
0.2
0.0
low
high
0
50
100
150
200
Time (months)
number at risk
245 133 22 2 0
242 93 22 7 0

FIG. 15 cont'd

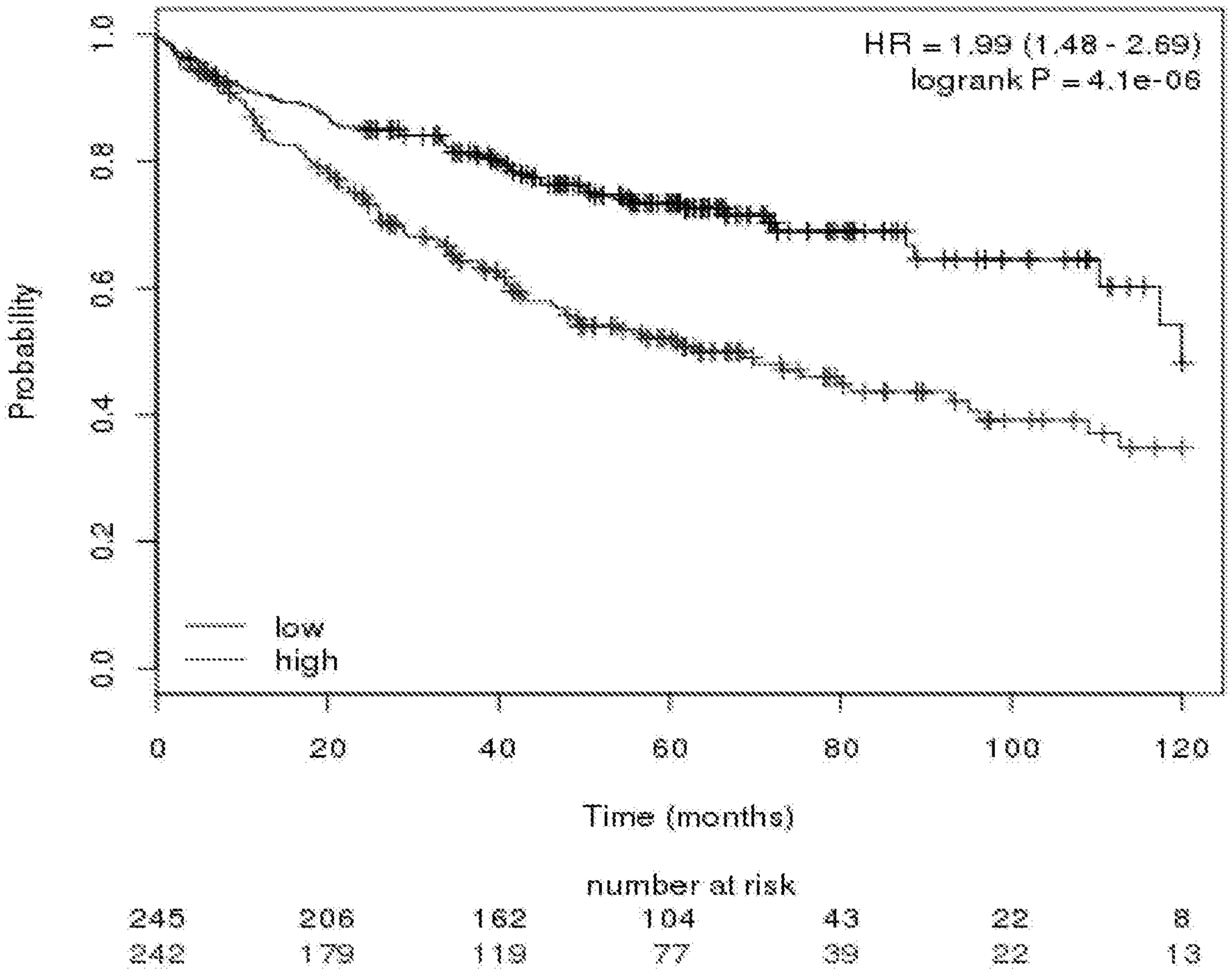

HR = 1.99 (1.48 - 2.69)
logrank P = 4.1e-06
Probability
0.0
0.2
0.4
0.6
0.8
1.0
low
high
0
20
40
60
80
100
120
Time (months)
number at risk
245 206 162 104 43 22 8
242 179 119 77 38 22 13

FIG. 15 cont'd

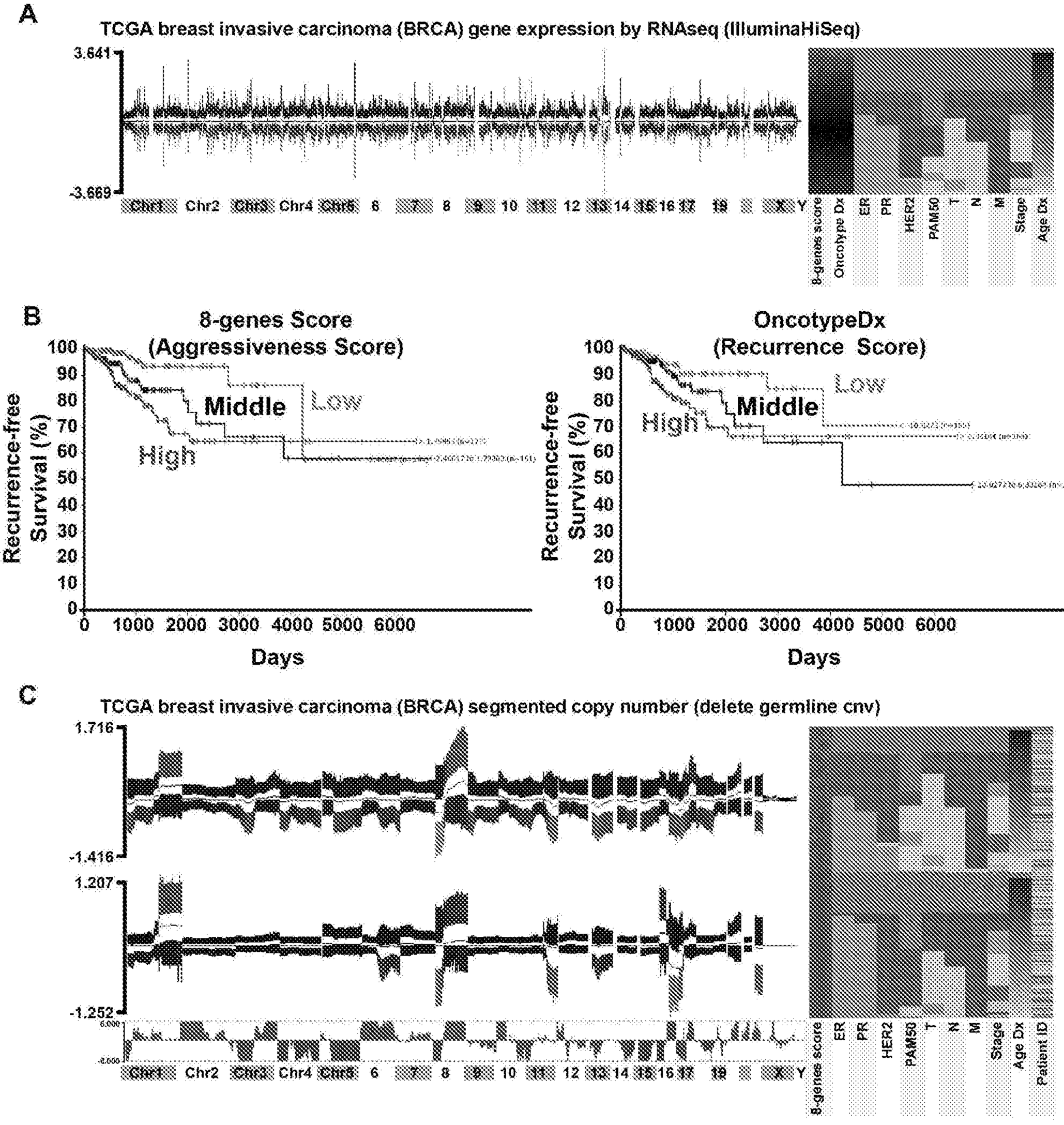

A
TCGA breast invasive carcinoma (BRCA) gene expression by RNAseq (IlluminaHiSeq)
3.641
-3.669
Chr1 Chr2 Chr3 Chr4 Chr5 6 7 8 9 10 11 12 13 14 15 16 17 18 X Y
8-genes score
Oncotype Dx
ER
PR
HER2
PAM50
T
N
M
Stage
Age Dx
B
8-genes Score
(Aggressiveness Score)
Recurrence-free Survival (%)
Middle
Low
High
Days
OncotypeDx
(Recurrence Score)
Recurrence-free Survival (%)
Low
Middle
High
Days
C
TCGA breast invasive carcinoma (BRCA) segmented copy number (delete germline cnv)
1.716
-1.416
1.207
-1.252
6.000
-6.000
Chr1 Chr2 Chr3 Chr4 Chr5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X Y
8-genes score
ER
PR
HER2
PAM50
T
N
M
Stage
Age Dx
Patient ID

FIG. 16

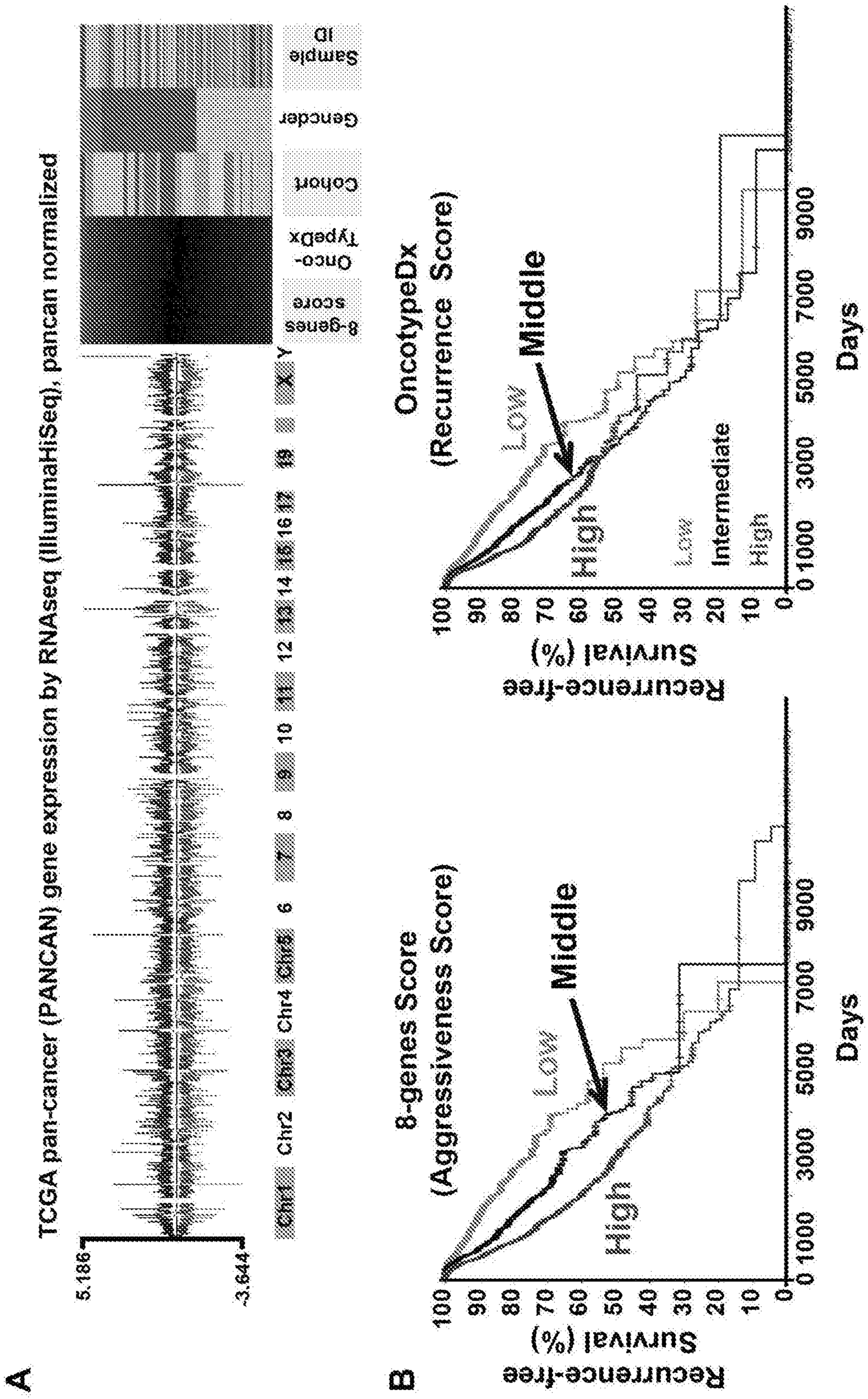

A
TCGA pan-cancer (PANCAN) gene expression by RNAseq (IlluminaHiSeq), pancan normalized
5.186
-3.644
Chr1 Chr2 Chr3 Chr4 Chr5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X Y
8-genes score
Onco-TypeDx
Cohort
Gencder
Sample ID
OncotypeDx
(Recurrence Score)
Low
Middle
High
Recurrence-free
Survival (%)
100 90 80 70 60 50 40 30 20 10 0
0 1000 3000 5000 7000 9000
Days
Intermediate
8-genes Score
(Aggressiveness Score)
B

FIG. 17

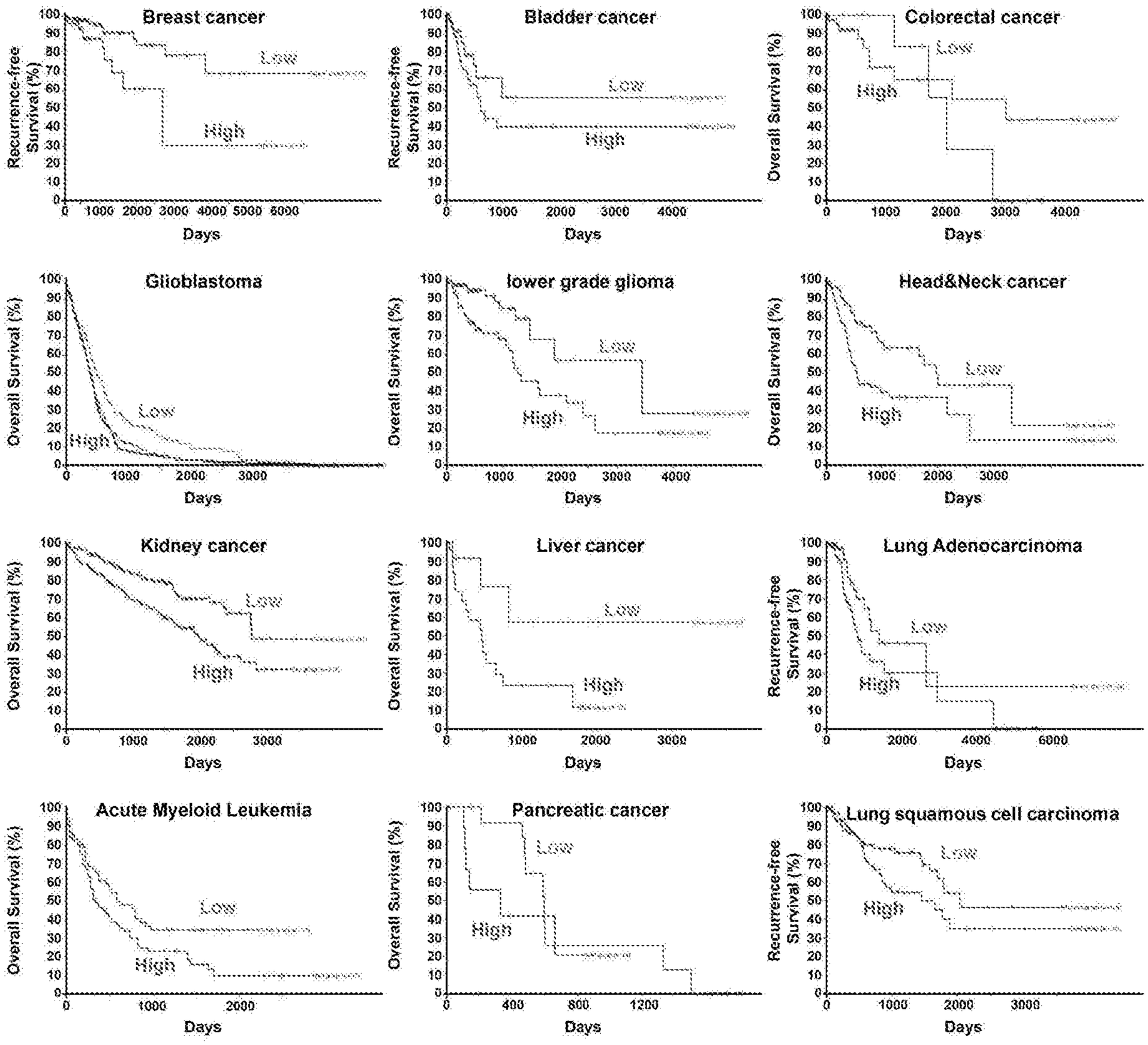

Breast cancer
Recurrence-free Survival (%)
Low
High
Days
Bladder cancer
Recurrence-free Survival (%)
Low
High
Days
Colorectal cancer
Overall Survival (%)
Low
High
Days
Glioblastoma
Overall Survival (%)
Low
High
Days
lower grade glioma
Overall Survival (%)
Low
High
Days
Head&Neck cancer
Overall Survival (%)
Low
High
Days
Kidney cancer
Overall Survival (%)
Low
High
Days
Liver cancer
Overall Survival (%)
Low
High
Days
Lung Adenocarcinoma
Recurrence-free Survival (%)
Low
High
Days
Acute Myeloid Leukemia
Overall Survival (%)
Low
High
Days
Pancreatic cancer
Overall Survival (%)
Low
High
Days
Lung squamous cell carcinoma
Recurrence-free Survival (%)
Low
High
Days

FIG. 18

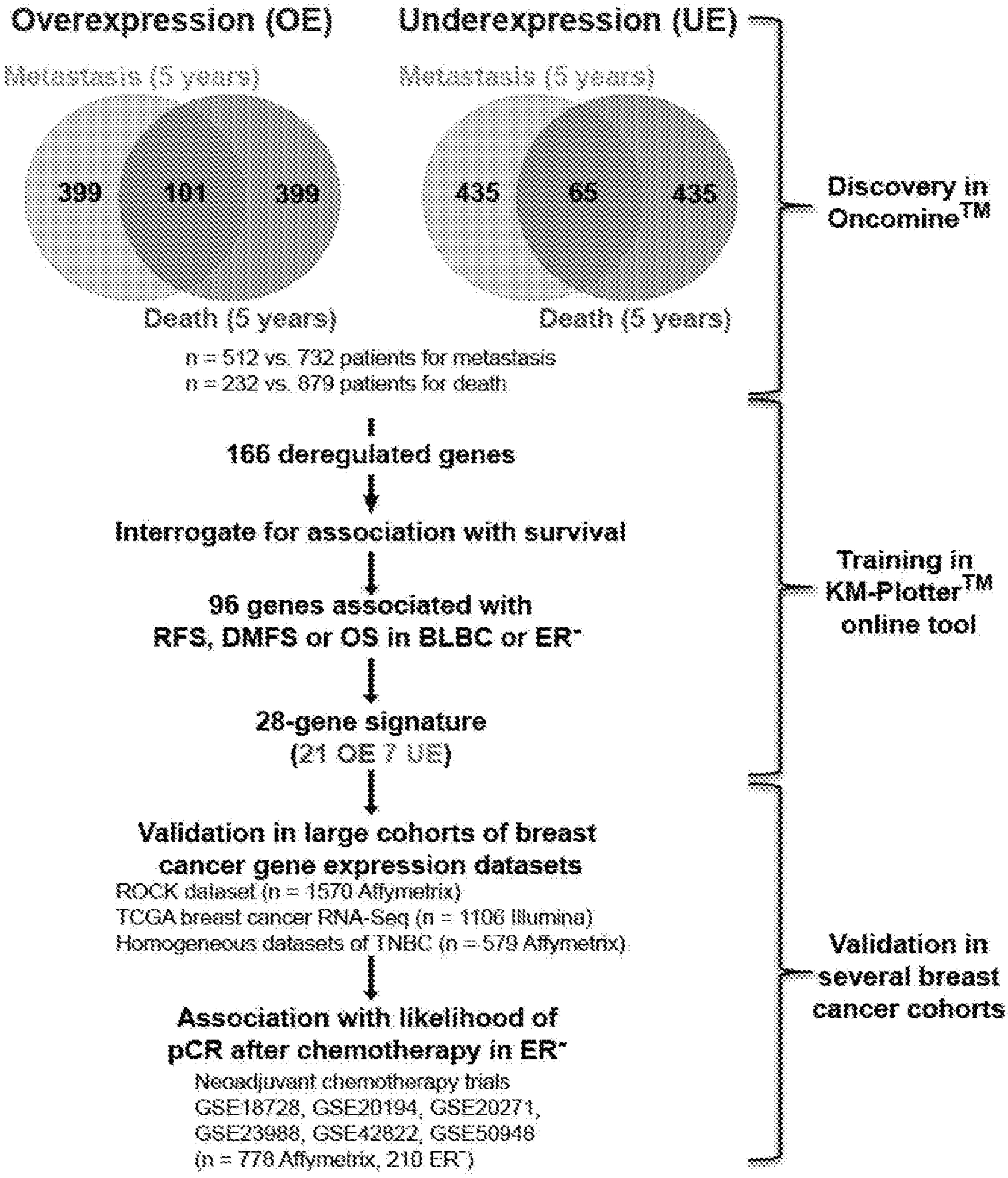

Overexpression (OE)
Underexpression (UE)
Metastasis (5 years)
Metastasis (5 years)
399
101
399
435
65
435
Death (5 years)
Death (5 years)
n = 512 vs. 732 patients for metastasis
n = 232 vs. 879 patients for death
166 deregulated genes
Interrogate for association with survival
96 genes associated with RFS, DMFS or OS in BLBC or ER-
28-gene signature (21 OE 7 UE)
Validation in large cohorts of breast cancer gene expression datasets
ROCK dataset (n = 1570 Affymetrix)
TCGA breast cancer RNA-Seq (n = 1106 Illumina)
Homogeneous datasets of TNBC (n = 579 Affymetrix)
Association with likelihood of pCR after chemotherapy in ER-
Neoadjuvant chemotherapy trials GSE18728, GSE20194, GSE20271, GSE23988, GSE42822, GSE50948 (n = 778 Affymetrix, 210 ER-)
Discovery in Oncomine™
Training in KM-Plotter™ online tool
Validation in several breast cancer cohorts

FIG. 19

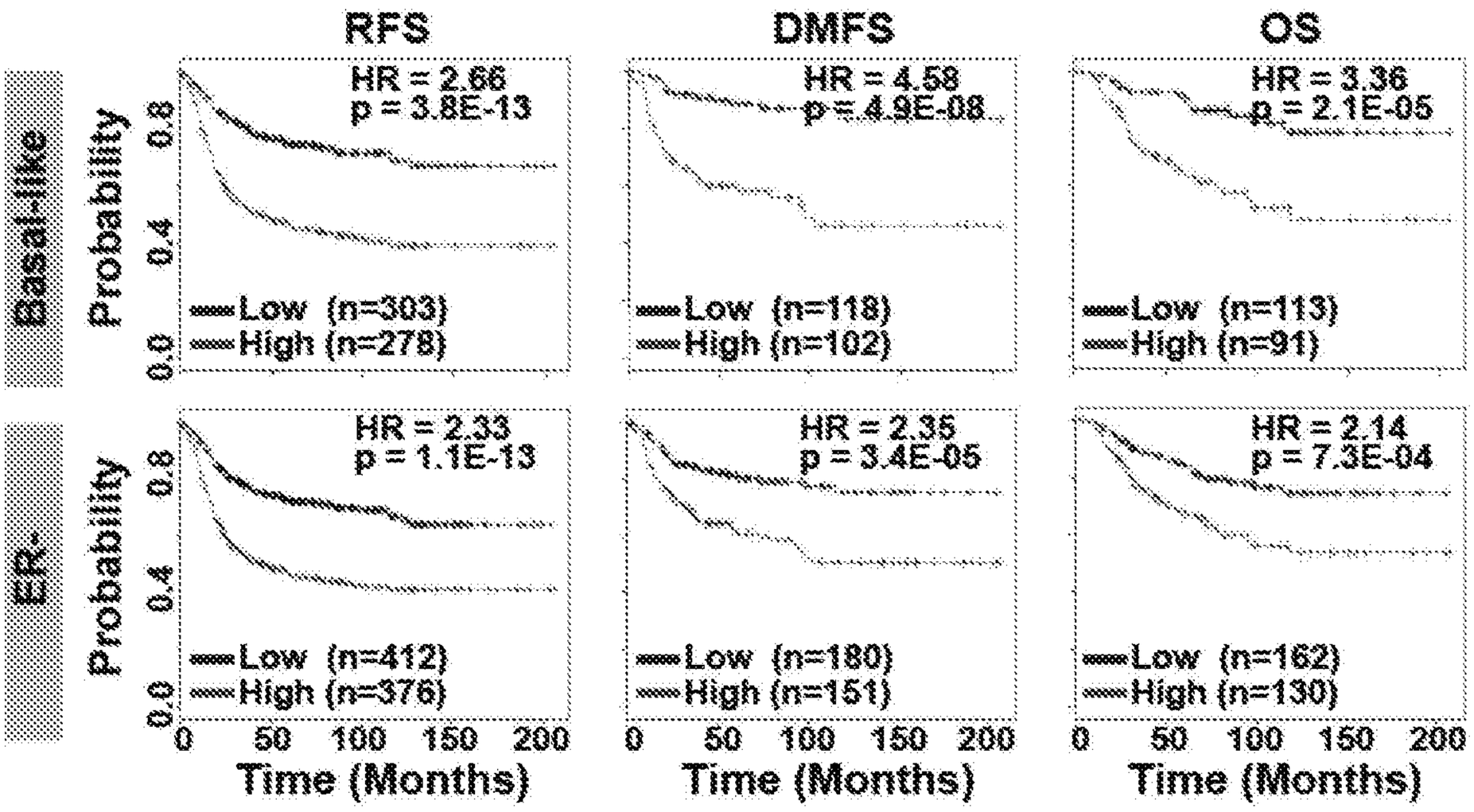
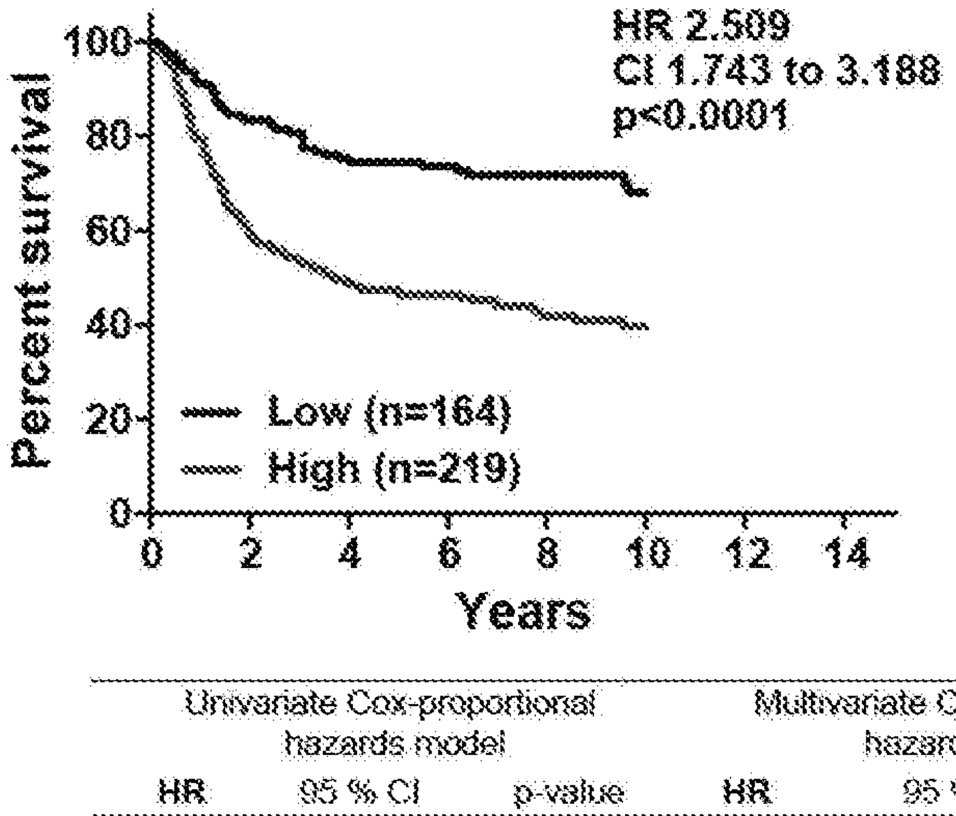

RFS
DMFS
OS
Basal-like
ER-
Probability
0.0
0.4
0.8
HR = 2.66
p = 3.8E-13
Low (n=303)
High (n=278)
HR = 4.58
p = 4.9E-08
Low (n=118)
High (n=102)
HR = 3.36
p = 2.1E-05
Low (n=113)
High (n=91)
HR = 2.33
p = 1.1E-13
Low (n=412)
High (n=376)
HR = 2.35
p = 3.4E-05
Low (n=180)
High (n=151)
HR = 2.14
p = 7.3E-04
Low (n=162)
High (n=130)
0
50
100
150
200
Time (Months)

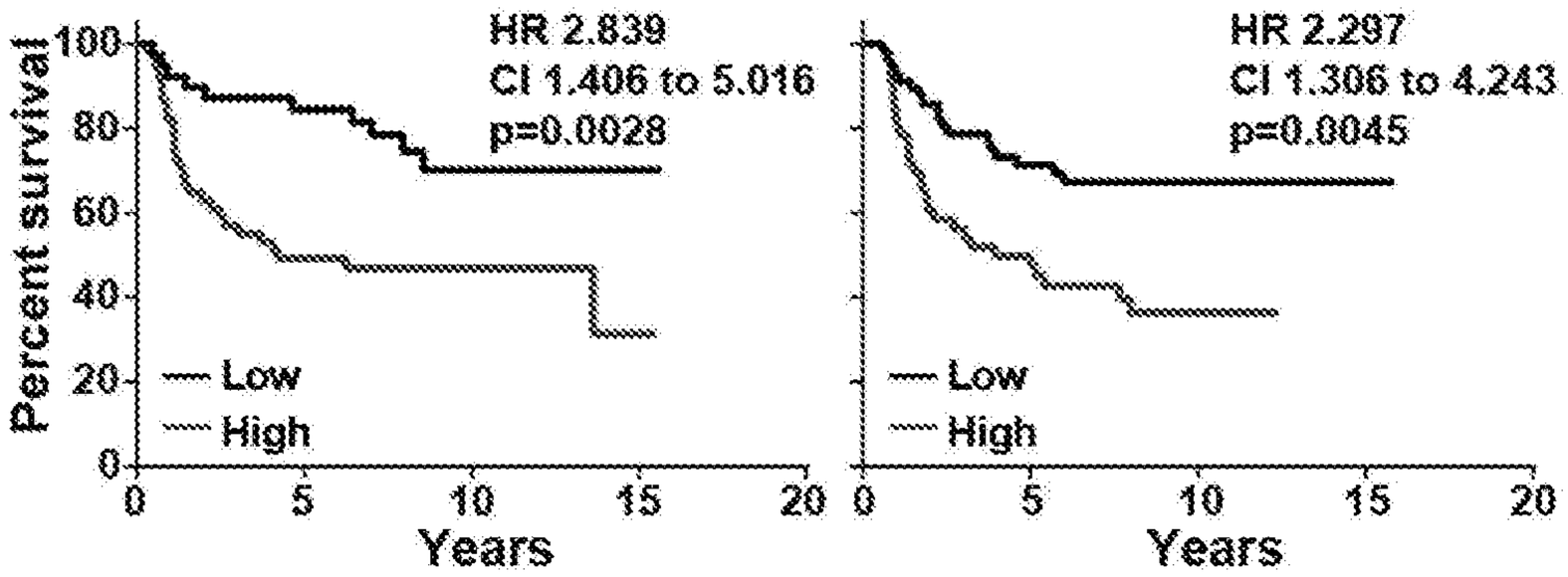
FIG. 20

A



	Univariate Cox-proportional hazards model			Multivariate Cox-proportional hazards model		
	HR	95 % CI	p-value	HR	95 % CI	p-value
TN score (high vs. low)	**2.5172**	1.7919-3.5361	**<0.0001**	**2.2882**	1.5217-3.4438	**0.0001**
Tumor size (T1, T2, T3)	**0.7097**	0.4893-1.0292	0.0720	**0.8148**	0.5522-1.2023	0.3046
Grade (G1, G2, G3)	**1.3036**	0.9102-1.8670	0.1501	**1.1784**	0.8023-1.7308	0.4052
Age (<50, >50)	**0.9009**	0.6572-1.2349	0.5186	**0.9908**	0.6985-1.4055	0.9592
Nodal status (N0, N1)	**1.0922**	0.6949-1.7165	0.7037	**1.4758**	0.6935-3.1406	0.3150
Adj. chemo (Yes, No)	1.0564	0.7320-1.5247	0.7705	1.1582	0.7464-1.7974	0.5145

B

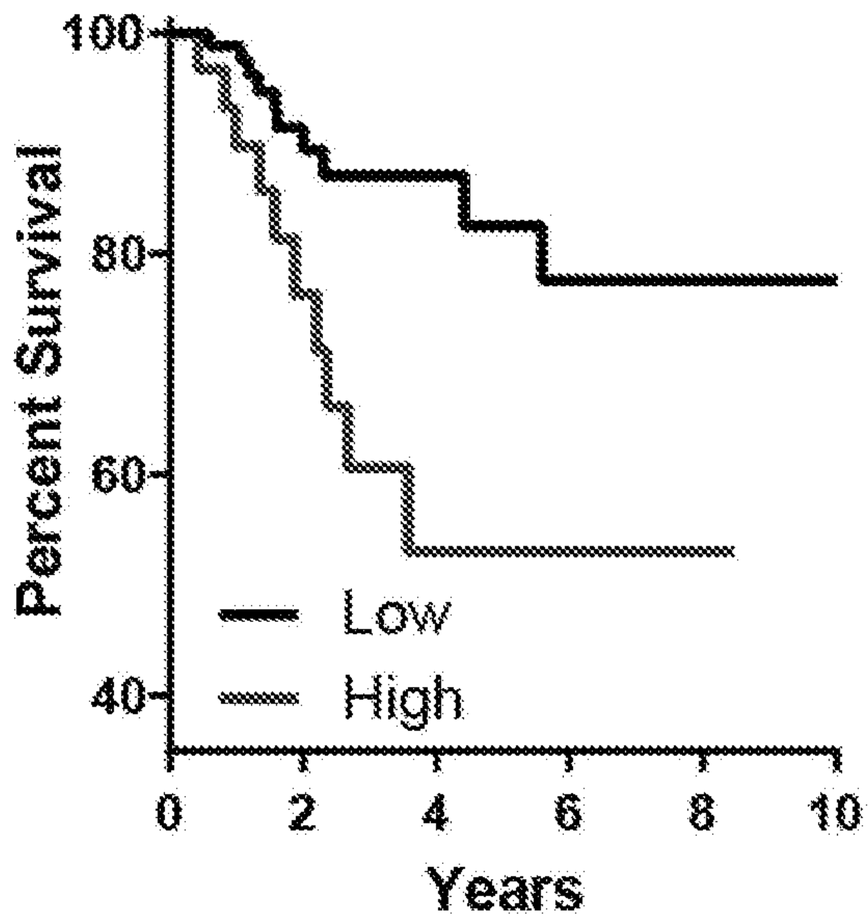


	RFS			DMFS		
	Multivariate Cox-proportional hazards model			Multivariate Cox-proportional hazards model		
	HR	95 % CI	p-value	HR	95 % CI	p-value
TN score (high vs. low)	**3.0343**	1.3703-6.7194	**0.0065**	**2.7139**	1.1723-6.2829	**0.0204**
Tumor size (T1, T2, T3)	**3.3223**	0.4135-26.694	0.2612	**2.8371**	0.4681-31.452	0.2126
Grade (G1, G2, G3)	**1.2288**	0.3377-4.4713	0.7557	**1.8560**	0.5153-6.6855	0.3467
Age (<50, >50)	**1.0481**	0.4964-2.2129	0.9024	**0.9172**	0.4225-1.9911	0.8279
Nodal status (N0, N1)	**2.5022**	0.4817-12.997	0.2777	**0.9240**	0.1052-8.1178	0.9435
ER status (ER+, ER-)	**0.5828**	0.1753-1.9374	0.3808	**0.6780**	0.1986-2.3141	0.5370

C

	RFS			DMFS		
	Multivariate Cox-proportional hazards model			Multivariate Cox-proportional hazards model		
	HR	95 % CI	p-value	HR	95 % CI	p-value
TN score (high vs. low)	**2.0323**	1.2211-3.3822	**0.0066**	**2.2830**	1.1859-4.3953	**0.0140**
Tumor size (T1, T2, T3)	**1.1703**	0.2562-5.3460	0.8400	**0.7795**	0.0908-6.6911	0.8213
Grade (G1, G2, G3)	**0.7299**	0.0983-5.4204	0.7535	**1.3221**	0.1718-10.177	0.7898
Age (<50, >50)	**1.0969**	0.6935-1.7604	0.7030	**0.9745**	0.5558-1.7087	0.9286
Nodal status (N0, N1)	**1.5486**	0.5542-4.3274	0.4066	**1.6229**	0.5052-5.2136	0.4185

FIG. 21



	Univariate Cox-proportional hazards model			Multivariate Cox-proportional hazards model		
	HR	95 % CI	p-value	HR	95 % CI	p-value
TN score (high vs. low)	**2.9993**	1.2533-7.1778	**0.0141**	**3.3316**	1.2323-9.0069	**0.0183**
Stage (I, II, III)	**3.4112**	1.2738-9.1355	**0.0151**	**3.2814**	0.8902-12.095	0.0757
Nodal status (N0, N1)	**3.4618**	1.2525-9.5680	**0.0172**	**0.9415**	0.2092-4.2375	0.9377
HER2 (Neg, Pos)	**0.3487**	0.0456-2.6680	0.3127			
Age (<50, >50)	**1.2375**	0.4502-3.4018	0.6812			
Tumor size (T2/3 vs T1)	**1.0393**	0.2916-3.7046	0.9528			

FIG. 22

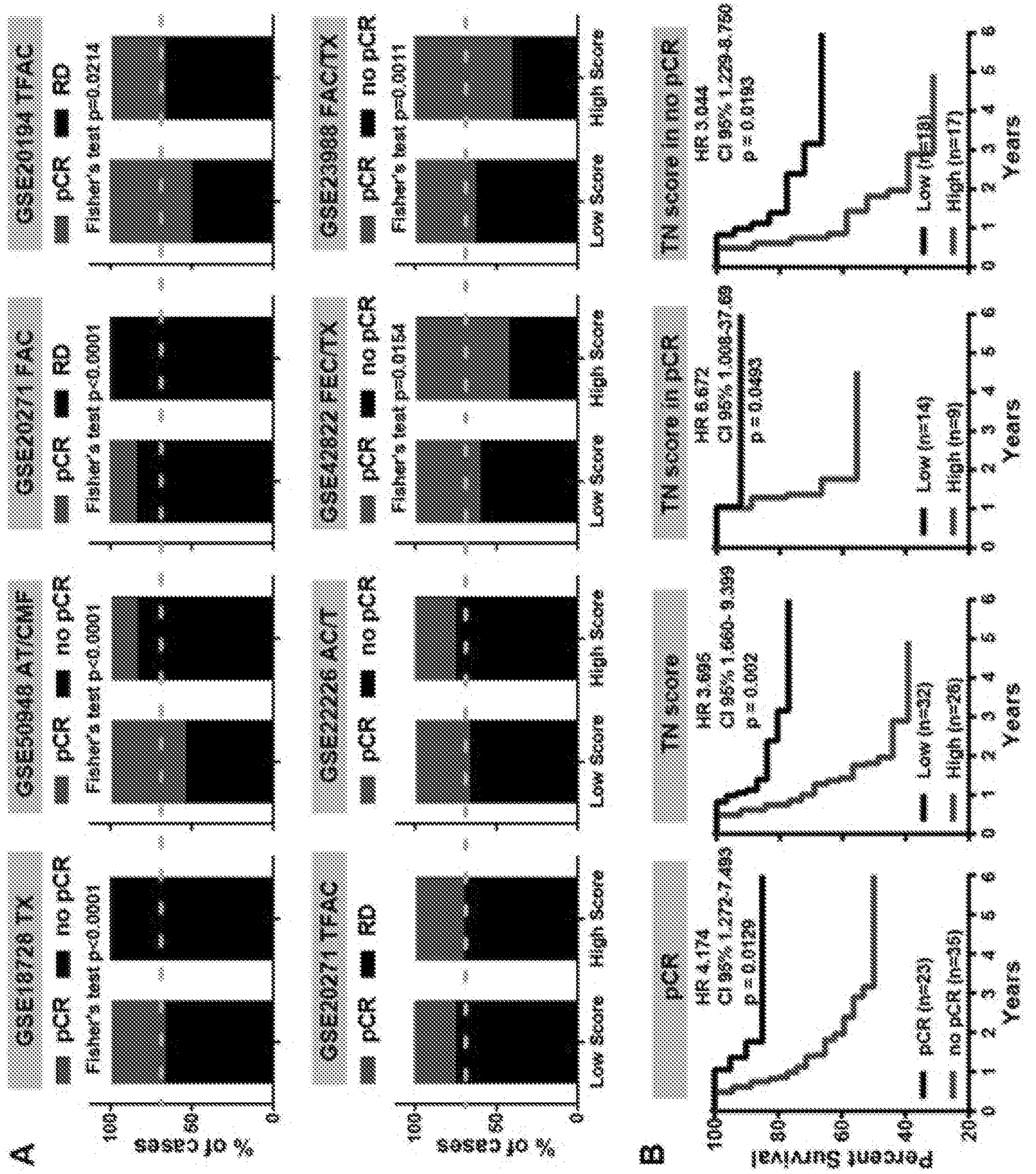

A
GSE16716 TX
pCR
no pCR
Fisher's test p<0.0001
% of cases
100
50
0
Low Score
High Score
GSE20271 TFAC
pCR
RD
GSE18728 TX
pCR
no pCR
Fisher's test p<0.0001
GSE20271 TFAC
pCR
RD
GSE50948 AT/CMF
pCR
no pCR
Fisher's test p<0.0001
GSE22226 ACT
pCR
no pCR
GSE22226 ACT
pCR
no pCR
GSE20271 FAC
pCR
RD
Fisher's test p<0.0001
GSE23988 FAC/TX
pCR
no pCR
Fisher's test p=0.0011
GSE42822 FEC/TX
pCR
no pCR
Fisher's test p=0.0154
GSE20194 TFAC
pCR
RD
Fisher's test p=0.0214
B
pCR
HR 4.174
CI 95% 1.272-7.493
p = 0.0129
pCR (n=23)
no pCR (n=36)
TN score in pCR
HR 3.685
CI 95% 1.660- 9.399
p = 0.002
Low (n=32)
High (n=26)
TN score in no pCR
HR 6.672
CI 95% 1.008-37.69
p = 0.0423
Low (n=14)
High (n=9)
HR 3.044
CI 95% 1.229-8.760
p = 0.0183
Low (n=18)
High (n=17)
Percent Survival
100
80
60
40
20
0
1
2
3
4
5
6
Years

FIG. 23

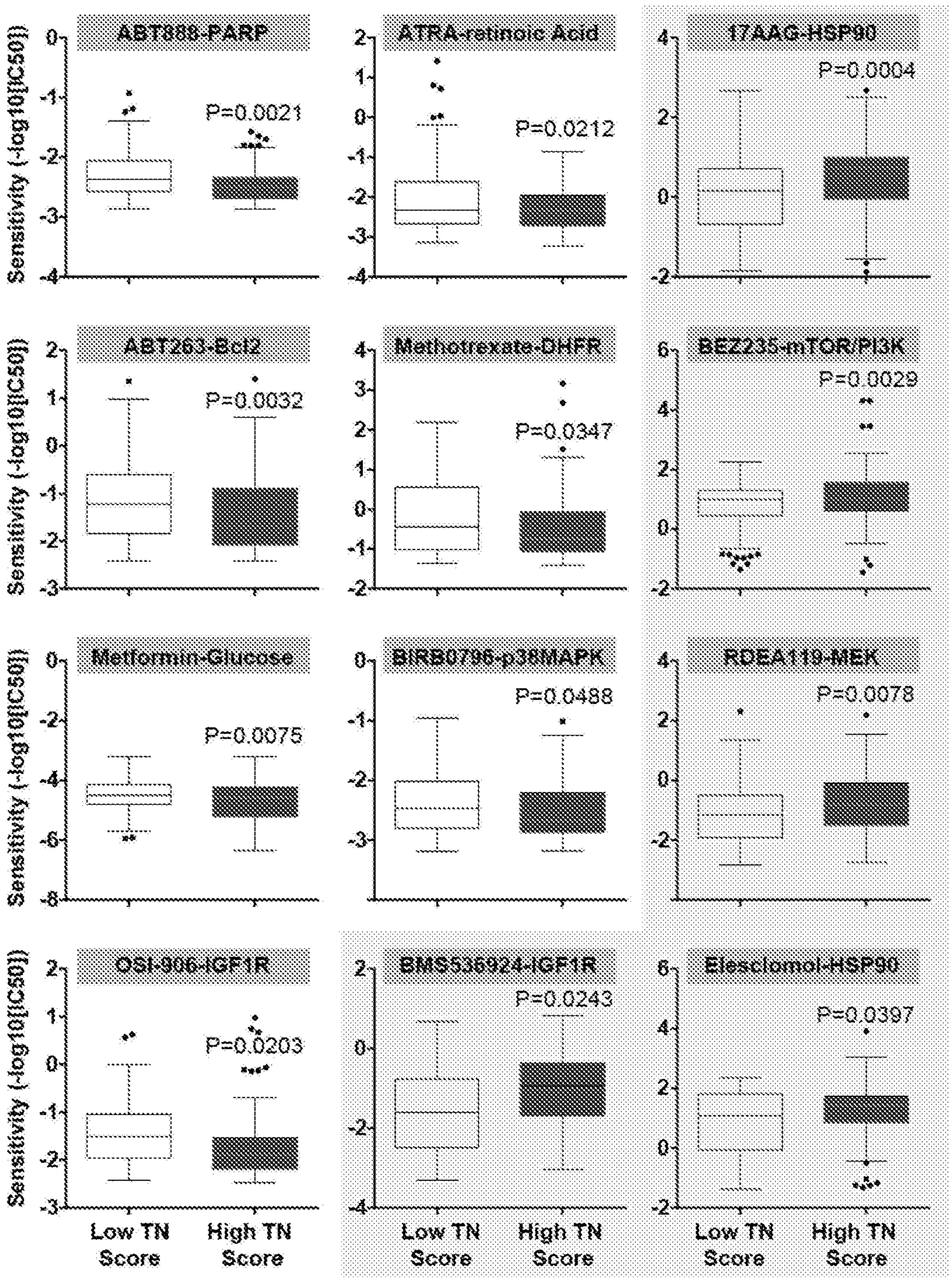

ABT888-PARP
P=0.0021
ATRA-retinoic Acid
P=0.0212
17AAG-HSP90
P=0.0004
ABT263-Bcl2
P=0.0032
Methotrexate-DHFR
P=0.0347
BEZ235-mTOR/PI3K
P=0.0029
Metformin-Glucose
P=0.0075
BIRB0796-p38MAPK
P=0.0488
RDEA119-MEK
P=0.0078
OSI-906-IGF1R
P=0.0203
BMS536924-IGF1R
P=0.0243
Elesclomol-HSP90
P=0.0397
Sensitivity (-log10[IC50])
Low TN Score
High TN Score

FIG. 24

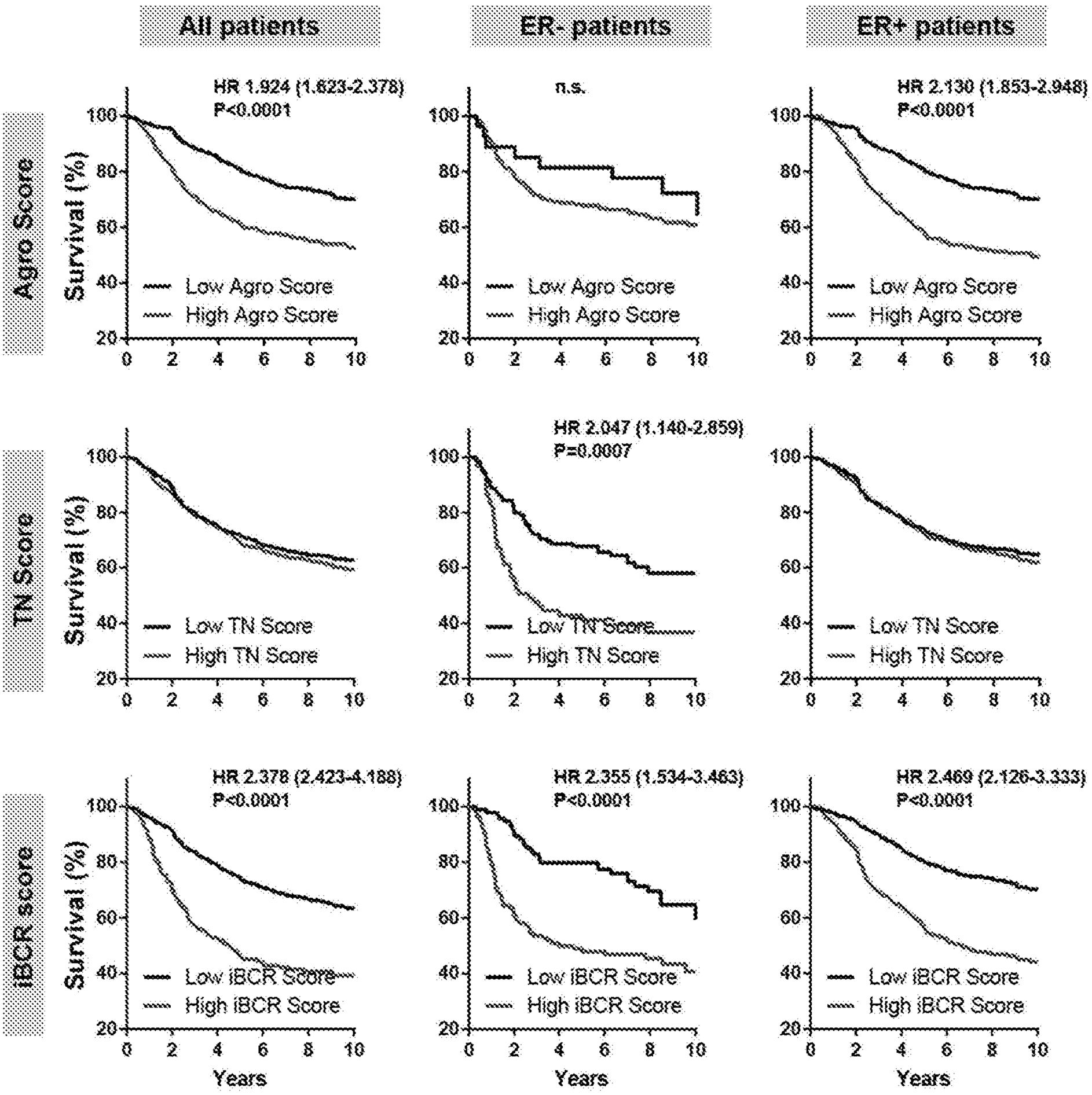

All patients
ER- patients
ER+ patients
Agro Score
TN Score
iBCR score
Survival (%)
HR 1.924 (1.623-2.378)
P<0.0001
n.s.
HR 2.130 (1.853-2.948)
P<0.0001
HR 2.047 (1.140-2.859)
P=0.0007
HR 2.378 (2.423-4.188)
P<0.0001
HR 2.355 (1.534-3.463)
P<0.0001
HR 2.469 (2.126-3.333)
P<0.0001
Low Agro Score
High Agro Score
Low TN Score
High TN Score
Low iBCR Score
High iBCR Score
Years

FIG. 25

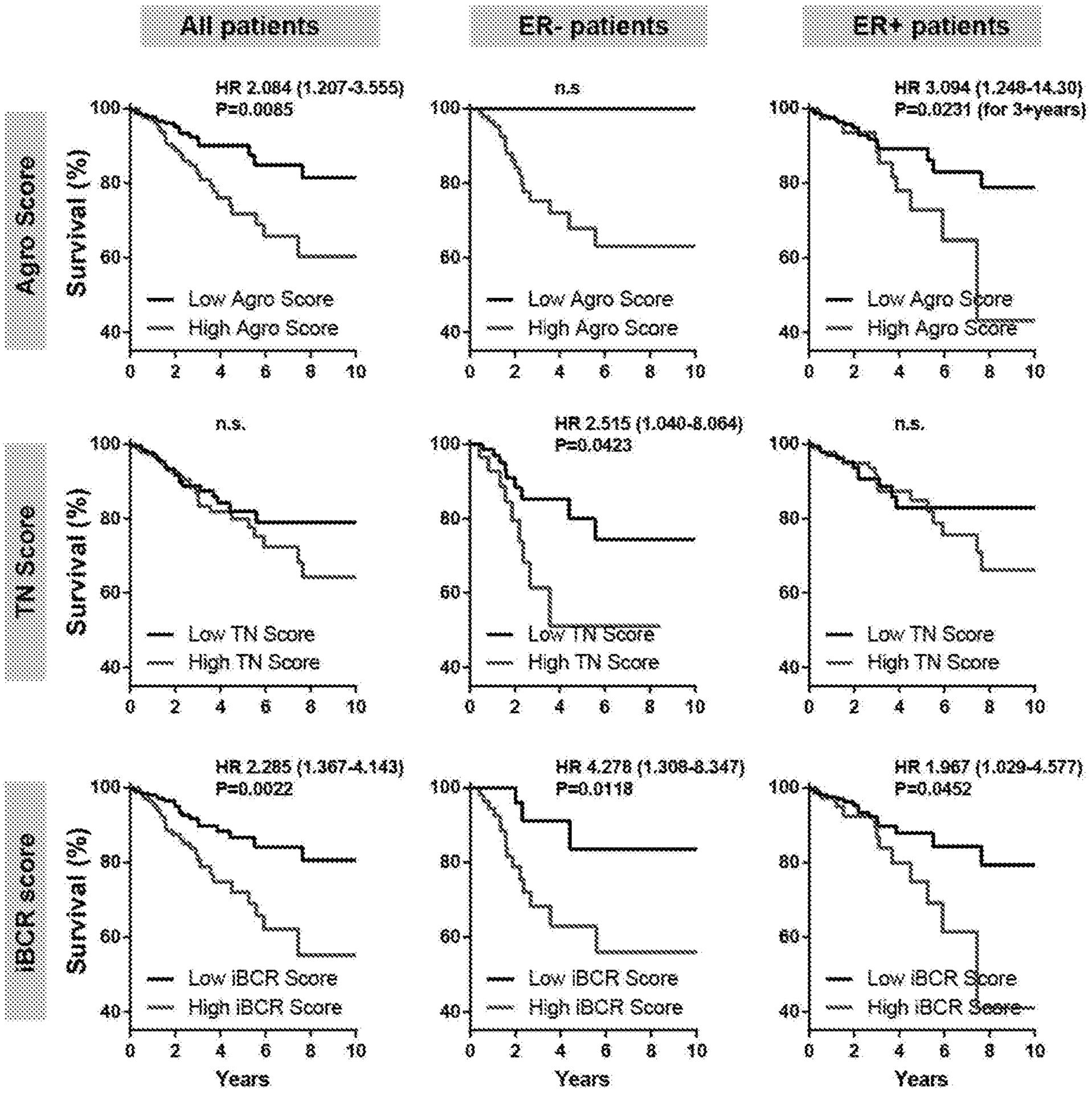

All patients
ER- patients
ER+ patients
Agro Score
TN Score
IBCR score
Survival (%)
HR 2.084 (1.207-3.555)
P=0.0085
n.s
HR 3.094 (1.248-14.30)
P=0.0231 (for 3+years)
n.s.
HR 2.515 (1.040-8.064)
P=0.0423
n.s.
HR 2.285 (1.367-4.143)
P=0.0022
HR 4.278 (1.308-8.347)
P=0.0118
HR 1.967 (1.029-4.577)
P=0.0452
Low Agro Score
High Agro Score
Low TN Score
High TN Score
Low iBCR Score
High iBCR Score
100
80
60
40
0 2 4 6 8 10
Years

FIG. 26

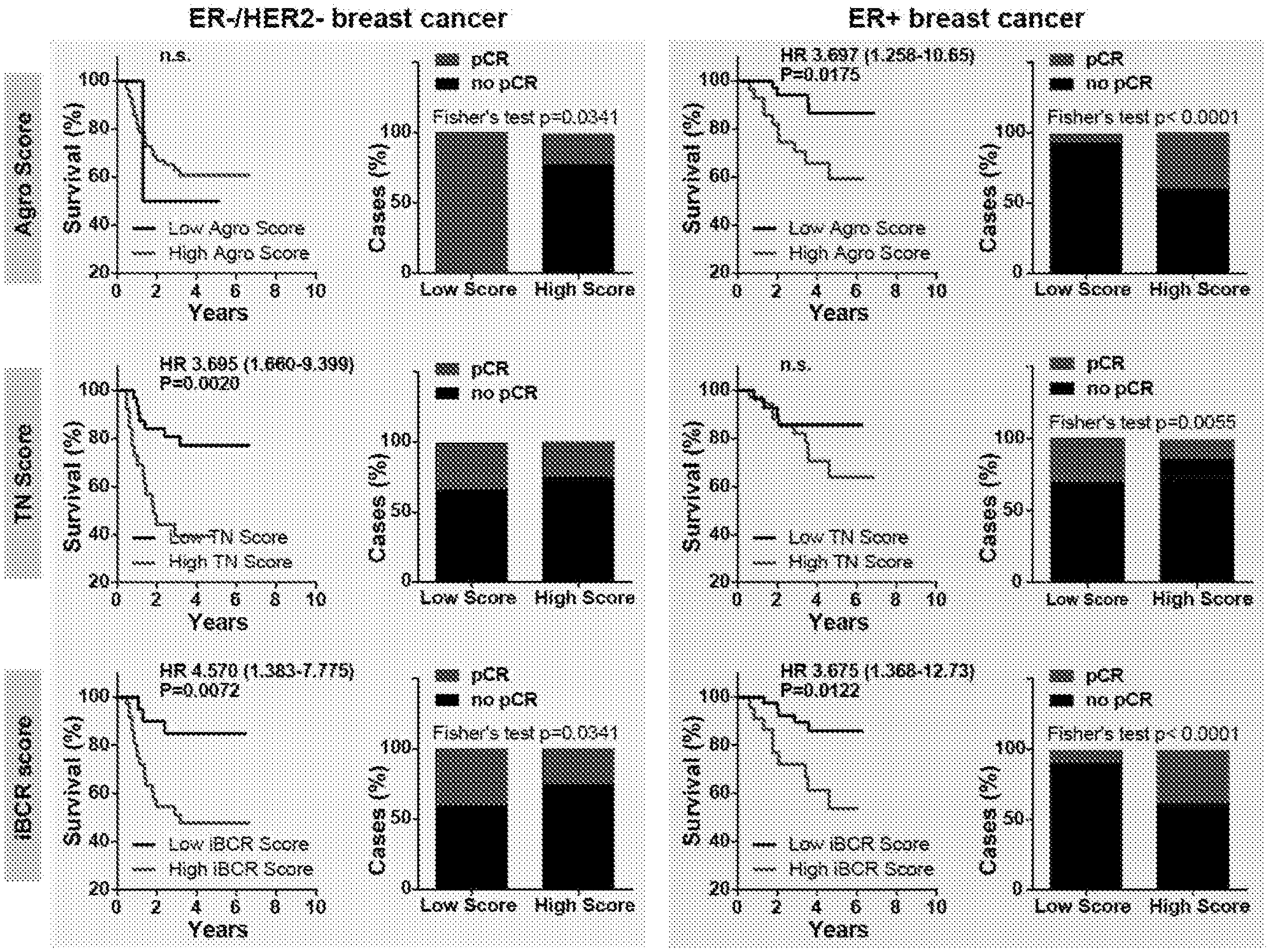

ER-/HER2- breast cancer
ER+ breast cancer
Agro Score
TN Score
iBCR score
n.s.
HR 3.697 (1.258-10.65)
P=0.0175
HR 3.695 (1.660-9.399)
P=0.0020
HR 4.570 (1.383-7.775)
P=0.0072
HR 3.675 (1.368-12.73)
P=0.0122
Survival (%)
Years
Low Agro Score
High Agro Score
Low TN Score
High TN Score
Low iBCR Score
High iBCR Score
pCR
no pCR
Cases (%)
Fisher's test p=0.0341
Fisher's test p< 0.0001
Fisher's test p=0.0055
Low Score
High Score

FIG. 27

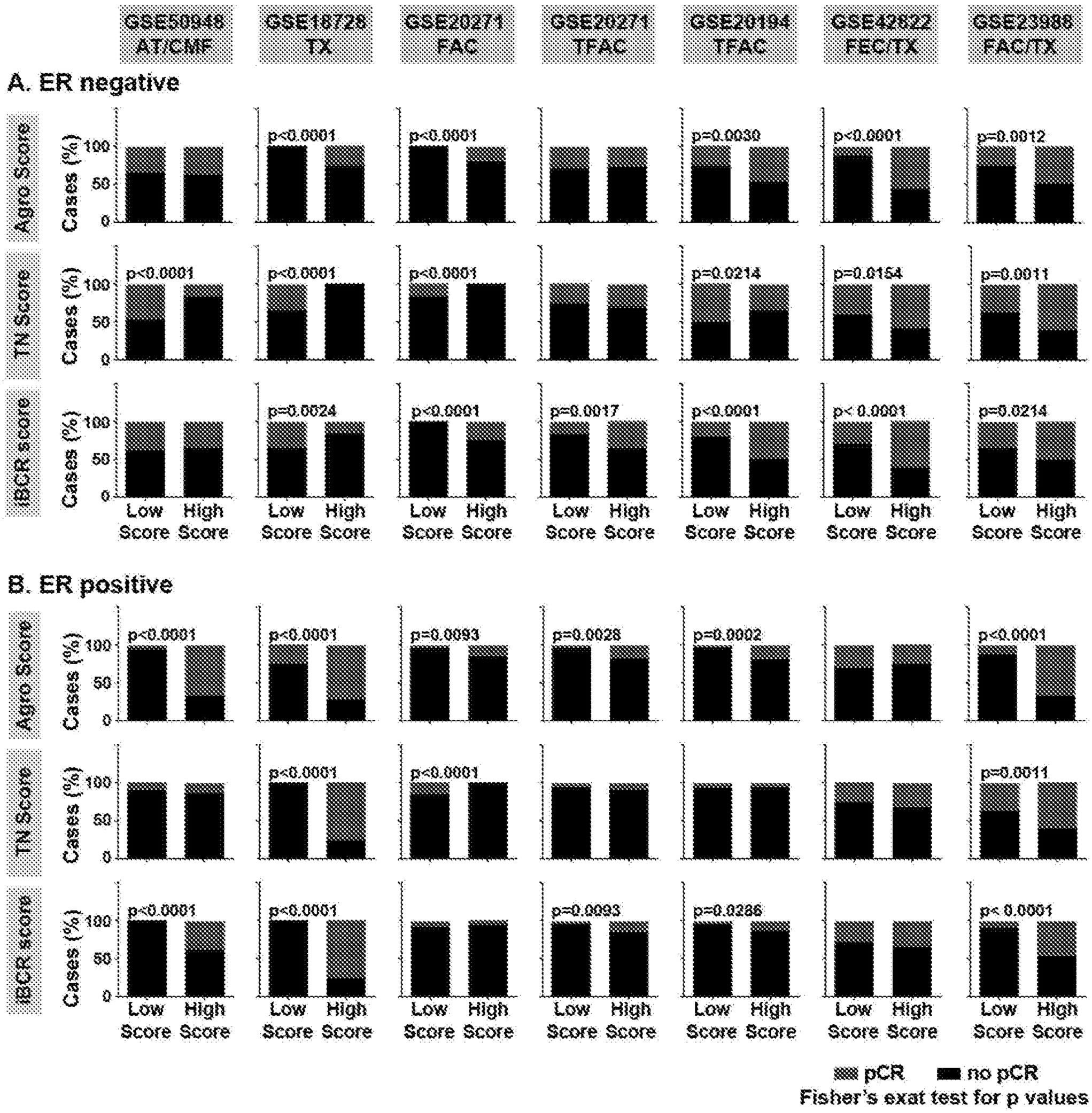

GSE50948 AT/CMF
GSE16726 TX
GSE20271 FAC
GSE20271 TFAC
GSE20194 TFAC
GSE42822 FEC/TX
GSE23988 FAC/TX
A. ER negative
Agro Score
TN Score
iBCR score
Cases (%)
100
50
0
p<0.0001
p=0.0030
p<0.0001
p=0.0012
p=0.0214
p=0.0154
p=0.0011
p=0.0024
p=0.0017
p< 0.0001
p=0.0214
Low Score
High Score
B. ER positive
p=0.0093
p=0.0028
p=0.0002
p=0.0011
p=0.0093
p=0.0286
pCR
no pCR
Fisher's exat test for p values

FIG. 28

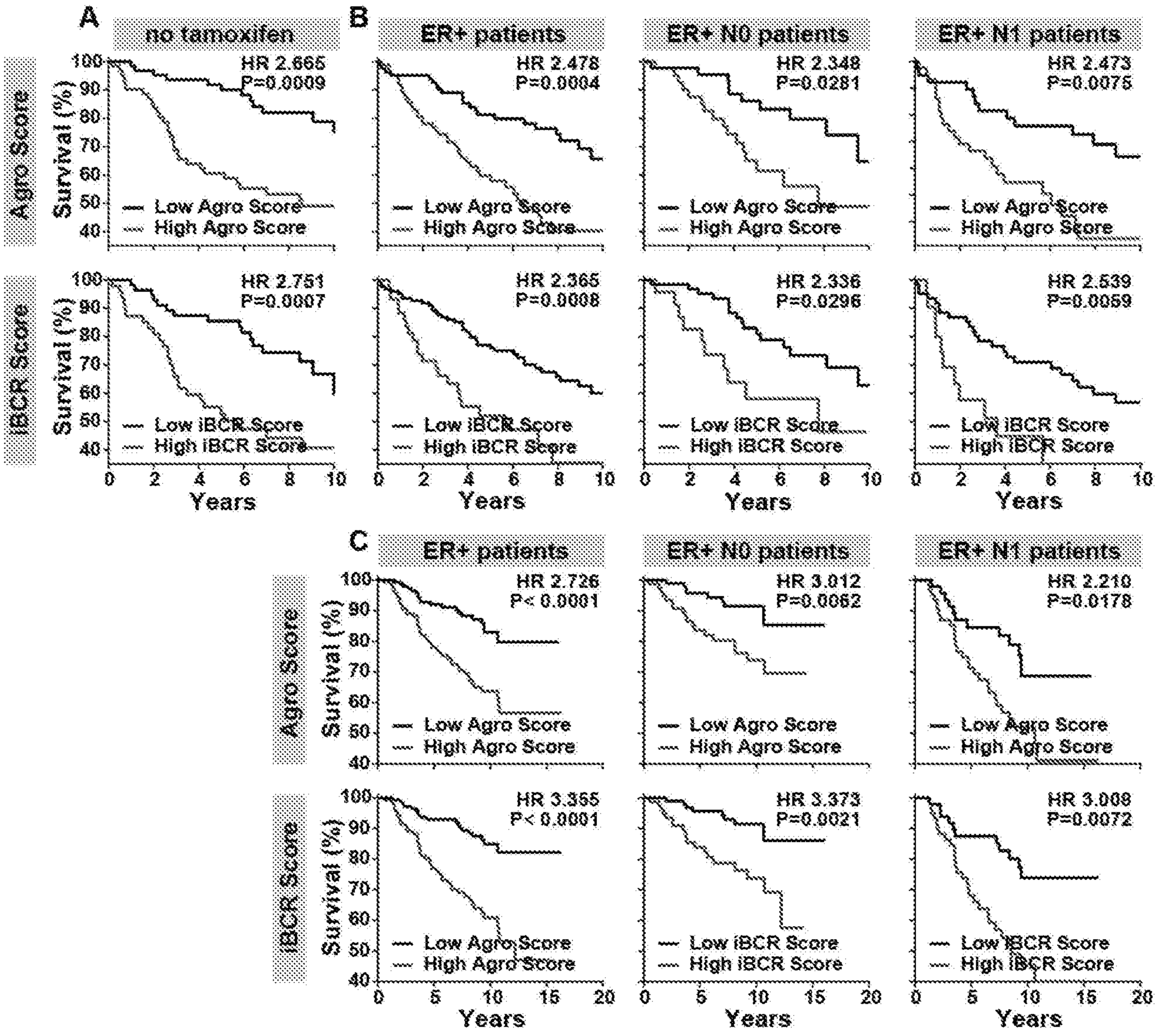

A
no tamoxifen
HR 2.665
P=0.0009
Low Agro Score
High Agro Score
Agro Score
Survival (%)
HR 2.751
P=0.0007
Low iBCR Score
High iBCR Score
iBCR Score
Years
B
ER+ patients
HR 2.478
P=0.0004
HR 2.365
P=0.0008
ER+ N0 patients
HR 2.348
P=0.0281
HR 2.336
P=0.0296
ER+ N1 patients
HR 2.473
P=0.0075
HR 2.539
P=0.0059
C
ER+ patients
HR 2.726
P< 0.0001
HR 3.355
P< 0.0001
ER+ N0 patients
HR 3.012
P=0.0062
HR 3.373
P=0.0021
ER+ N1 patients
HR 2.210
P=0.0178
HR 3.008
P=0.0072

FIG. 29

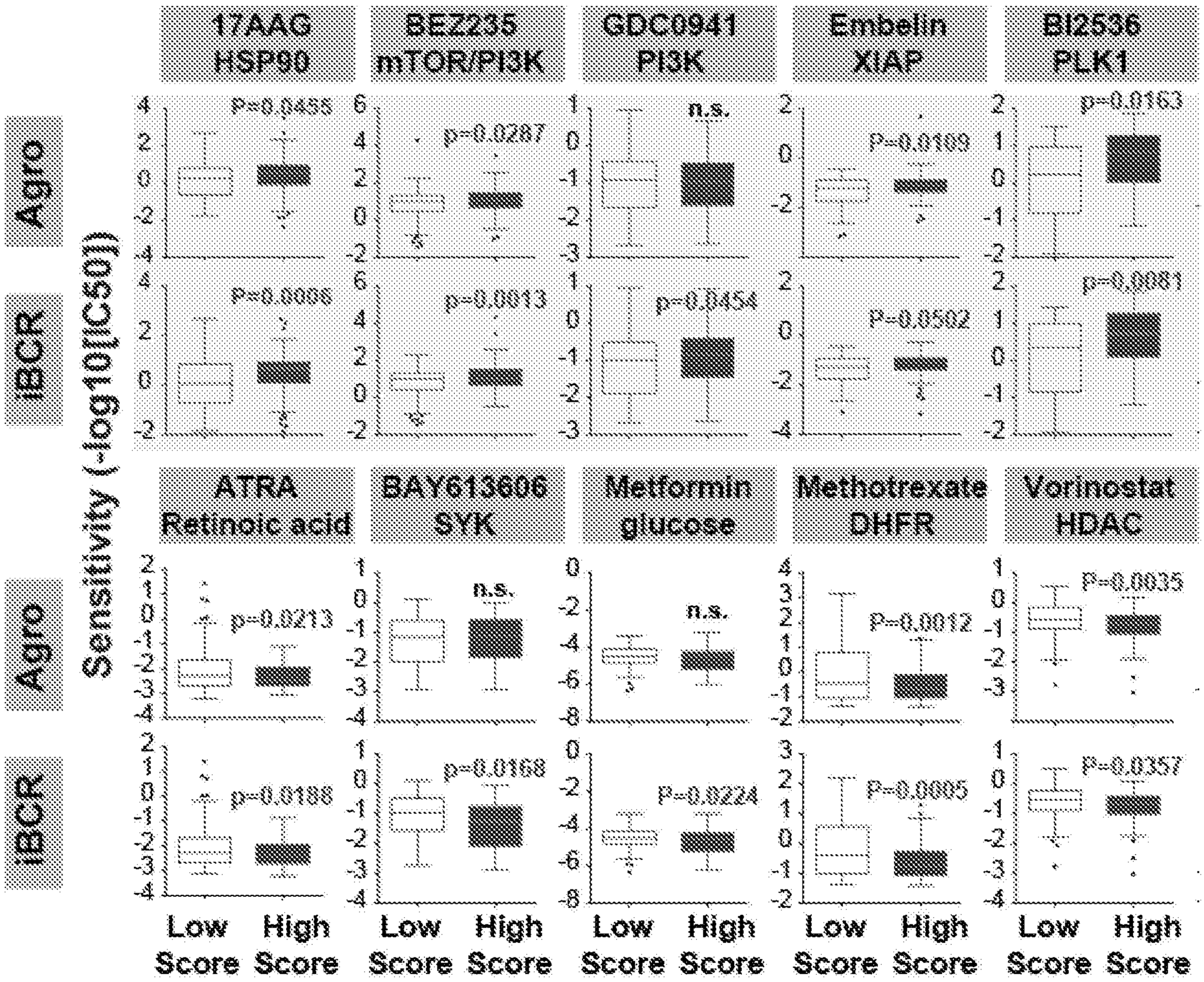

17AAG HSP90
BEZ235 mTOR/PI3K
GDC0941 PI3K
Embelin XIAP
BI2536 PLK1
Agro
IBCR
Sensitivity (-log10IC50)
P=0.0455
p=0.0287
n.s.
P=0.0109
p=0.0163
P=0.0006
p=0.0013
p=0.0454
P=0.0502
p=0.0081
ATRA Retinoic acid
BAY613606 SYK
Metformin glucose
Methotrexate DHFR
Vorinostat HDAC
p=0.0213
n.s.
n.s.
P=0.0012
P=0.0035
p=0.0188
p=0.0168
P=0.0224
P=0.0005
P=0.0357
Low Score
High Score

FIG. 30

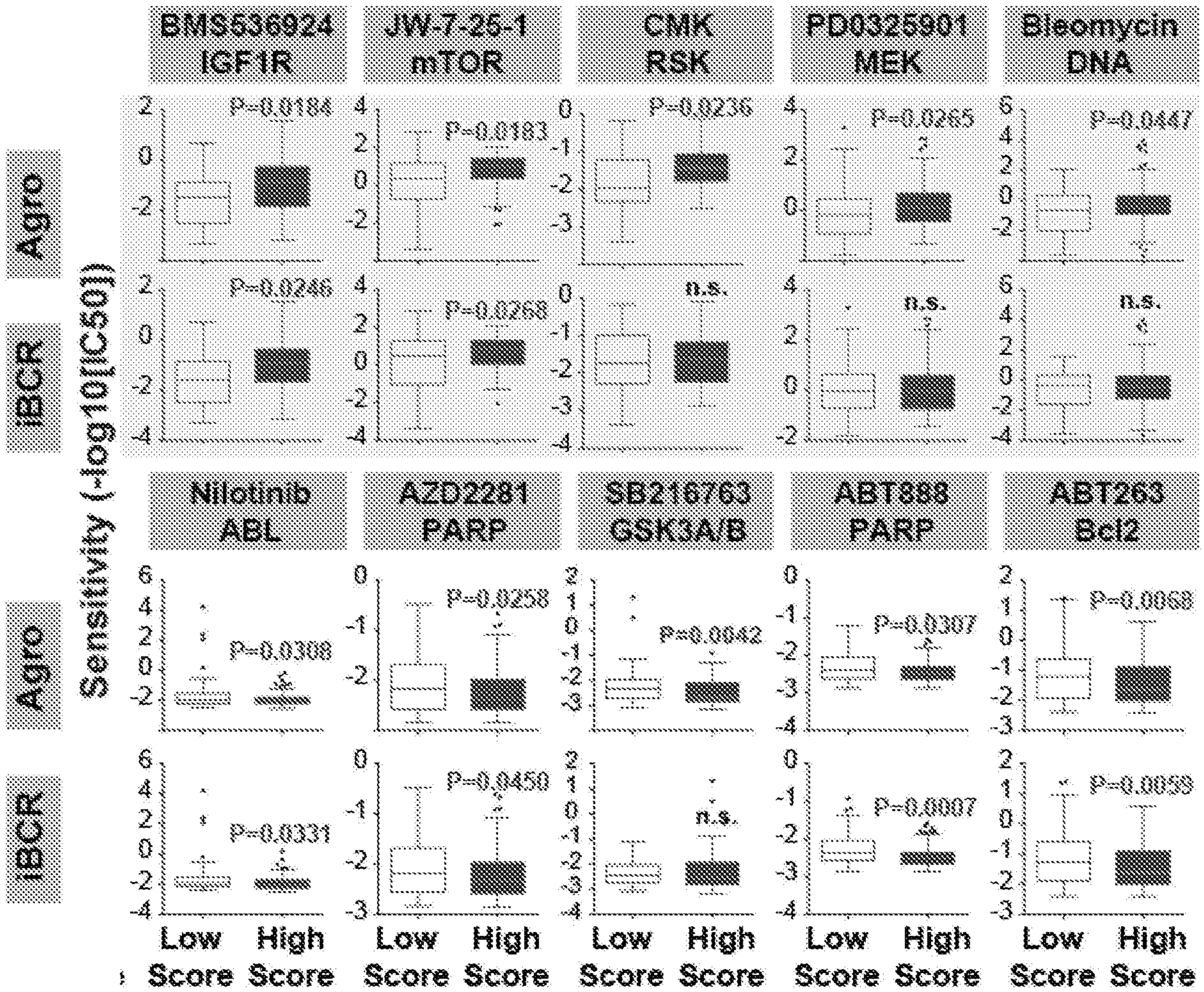

BMS536924
IGF1R
JW-7-25-1
mTOR
CMK
RSK
PD0325901
MEK
Bleomycin
DNA
Nilotinib
ABL
AZD2281
PARP
SB216763
GSK3A/B
ABT888
PARP
ABT263
Bcl2
Agro
iBCR
Sensitivity (-log10IC50)
P=0.0184
P=0.0183
P=0.0236
P=0.0265
P=0.0447
P=0.0246
P=0.0268
n.s.
n.s.
n.s.
P=0.0308
P=0.0258
P=0.0042
P=0.0307
P=0.0068
P=0.0331
P=0.0450
n.s.
P=0.0007
P=0.0059
Low Score
High Score

FIG. 30 cont'd

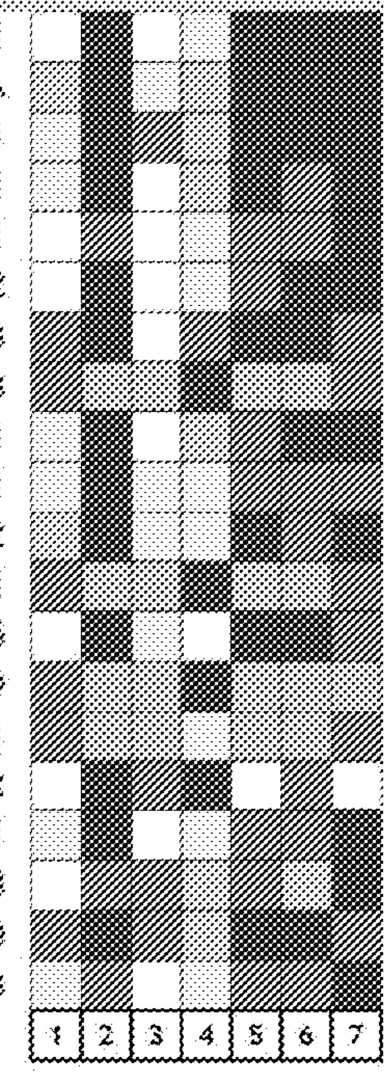
CA 02941769 2016-09-07

A

Comparison of All Genes Across 7 Analyses

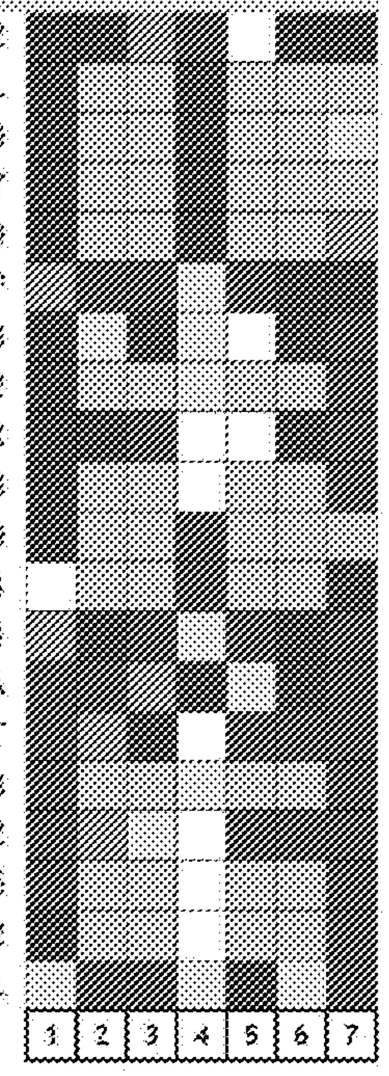
Overexpression

Median Rank	p-Value	Gene
36.0	8.23E-5	NUSAP1
124.0	5.13E-4	CENPA
124.0	0.002	BUB1
147.0	0.003	PRC1
159.0	0.003	FOXM1
161.0	6.57E-4	NEK2
174.0	7.34E-5	KIF23
177.0	7.63E-5	CHTF8
215.0	0.001	RAD51
223.0	1.41E-4	CDK1
254.0	0.005	CCNB2
260.0	3.26E-4	DNAJB11
273.0	2.60E-4	DLGAP5
276.5	0.006	UNC13D
279.0	0.005	PRELID1
287.0	0.006	GLP1R
293.0	0.006	MAD2L1
304.0	0.002	LRP8
304.0	0.005	MCM10
304.0	0.007	BUB1B



Underexpression

Median Rank	p-Value	Gene
79.0	0.003	GLTSCR2
102.0	0.001	MTMR8L
139.0	0.002	CCDC30
147.5	0.002	C14orf187
155.0	0.002	ABCA8
189.0	3.35E-6	CIRBP
210.0	2.95E-4	BTF3
225.0	3.50E-4	STON2
227.0	3.55E-4	NPY1R
228.0	3.59E-4	LOC283788
256.0	0.005	C1QTNF9
260.0	0.005	CDRT4
264.0	4.76E-4	PDCD4
265.0	4.78E-4	RPL13A
266.0	0.005	ABAT
272.0	0.005	ZBTB38
316.0	0.014	CREBL2
330.0	8.45E-4	TRIM35
331.0	8.45E-4	HNRNPA1L2
344.0	9.03E-4	EVL



Legend

1. Breast Carcinoma - Metastatic Event at 5 Years
Bos Breast, Nature, 2009

2. Invasive Ductal Breast Carcinoma - Metastatic Event at 5 Years
Desmedt Breast, Clin Cancer Res, 2007

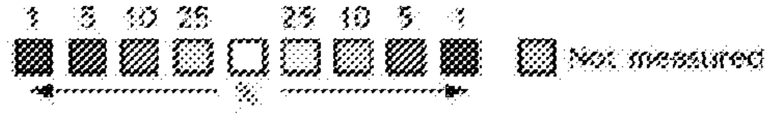
3. Invasive Breast Carcinoma - Metastatic Event at 5 Years
Hatzis Breast, JAMA, 2011

4. Breast Carcinoma - Metastatic Event at 5 Years
Kao Breast, BMC Cancer, 2011

5. Invasive Breast Carcinoma - Metastatic Event at 5 Years
Schmidt Breast, Cancer Res, 2008

6. Invasive Breast Carcinoma - Metastatic Event at 5 Years
Symmans Breast, J Clin Oncol, 2010

7. Breast Carcinoma - Metastatic Event at 5 Years
vandeVijver Breast, N Engl J Med, 2002



The rank for a gene is the median rank for that gene across each of the analyses.
The p-Value for a gene is its p-Value for the median-ranked analysis.

Copyright© 2008-11 Compendia Bioscience, Inc. Images from Oncomine™ may be used in publications with proper citation.
The citation for Oncomine is as follows: Oncomine™ (Compendia Bioscience, Ann Arbor, MI) was used for analysis and visualization.
For further information, refer to the terms described in the license agreement.

FIG. 31

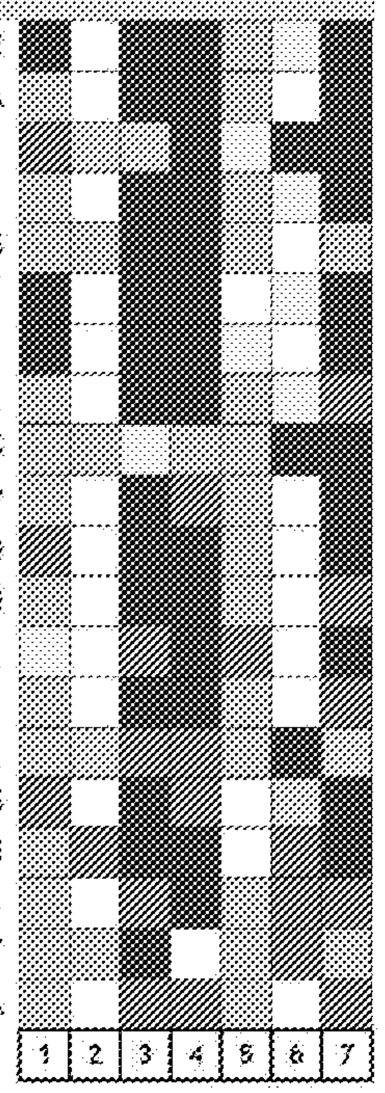
FIG. 31 cont'd

B

Comparison of All Genes Across 7 Analyses

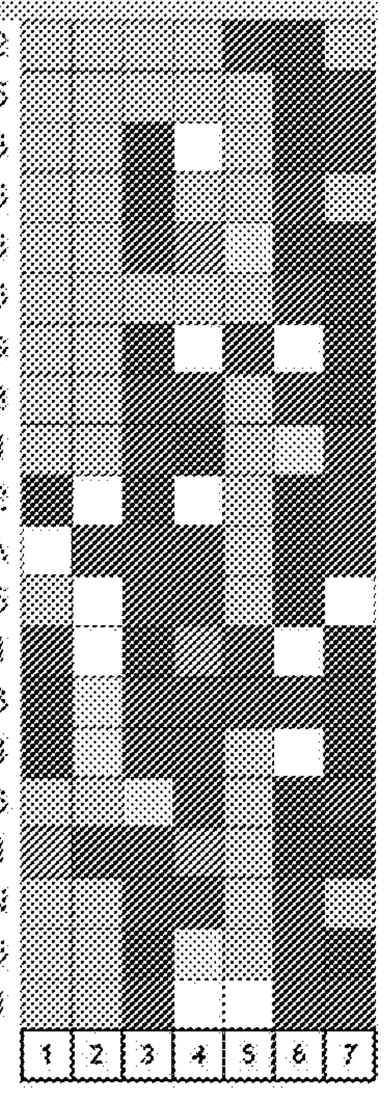
Overexpression

Median Rank	p-Value	Gene
109.5	1.48E-4	CCNB2
110.0	1.81E-4	KIF20A
126.0	0.005	PGK1
135.0	2.23E-4	RACGAP1
136.0	1.47E-4	CDCA5
146.0	2.45E-4	CCNB1
168.0	2.96E-4	FOXM1
171.0	2.96E-5	DEPDC1
171.0	0.003	PGAM5
184.0	2.36E-4	HJURP
187.0	0.004	CDC20
231.0	5.72E-5	MCM10
241.0	0.015	BUB1
242.0	6.04E-5	GTSE1
255.0	4.38E-4	ARFGAP1
257.0	0.010	BIRC5
266.0	0.008	CENPN
273.0	0.005	C18orf61
312.0	0.005	JMY
317.0	1.39E-4	KIF4A



Underexpression

Median Rank	p-Value	Gene
100.0	0.003	CBLN2
121.5	2.00E-4	DCDC5
181.5	2.78E-5	FGD3
195.0	0.002	TMEM26
214.0	6.88E-4	ZNF385B
230.0	0.003	RPL26
248.0	0.019	RERG
249.0	0.002	CCDC74B
255.0	8.65E-4	C5orf41
256.0	3.76E-4	SEC14L2
261.5	0.009	FAM105A
270.0	0.002	SEND5
271.0	0.011	NAT1
271.0	0.021	PTGER3
276.0	0.001	ERBB4
277.5	0.002	KIAA2026
280.0	0.009	MEIS3P1
290.0	0.002	PAPLN
307.5	0.004	SUSD3
311.0	0.001	ADHFE1



Legend

1. Breast Carcinoma - Dead at 5 Years
Bild Breast, Nature, 2006

2. Invasive Ductal Breast Carcinoma - Dead at 5 Years
Desmedt Breast, Clin Cancer Res, 2007

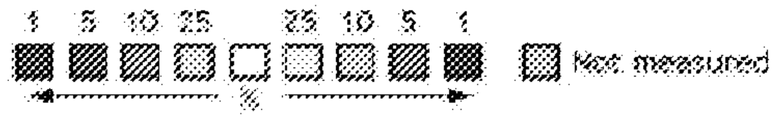
3. Breast Carcinoma - Dead at 5 Years
Kao Breast, BMC Cancer, 2011

4. Breast Carcinoma - Dead at 5 Years
Pawitan Breast, Breast Cancer Res, 2005

5. Ductal Breast Carcinoma - Dead at 5 Years
Sorlie Breast 2, Proc Natl Acad Sci U S A, 2003

6. Invasive Ductal Breast Carcinoma - Dead at 5 Years
TCGA Breast, No Associated Paper, 2011

7. Breast Carcinoma - Dead at 5 Years
vandeVijver Breast, N Engl J Med, 2002



The rank for a gene is the median rank for that gene across each of the analyses.
The p-Value for a gene is its p-Value for the median-ranked analysis.

Copyright© 2008-11 Compendia Bioscience, Inc. Images from Oncomine™ may be used in publications with proper citation.
The citation for Oncomine is as follows: Oncomine™ (Compendia Bioscience, Ann Arbor, MI) was used for analysis and visualization.
For further information, refer to the terms described in the license agreement.

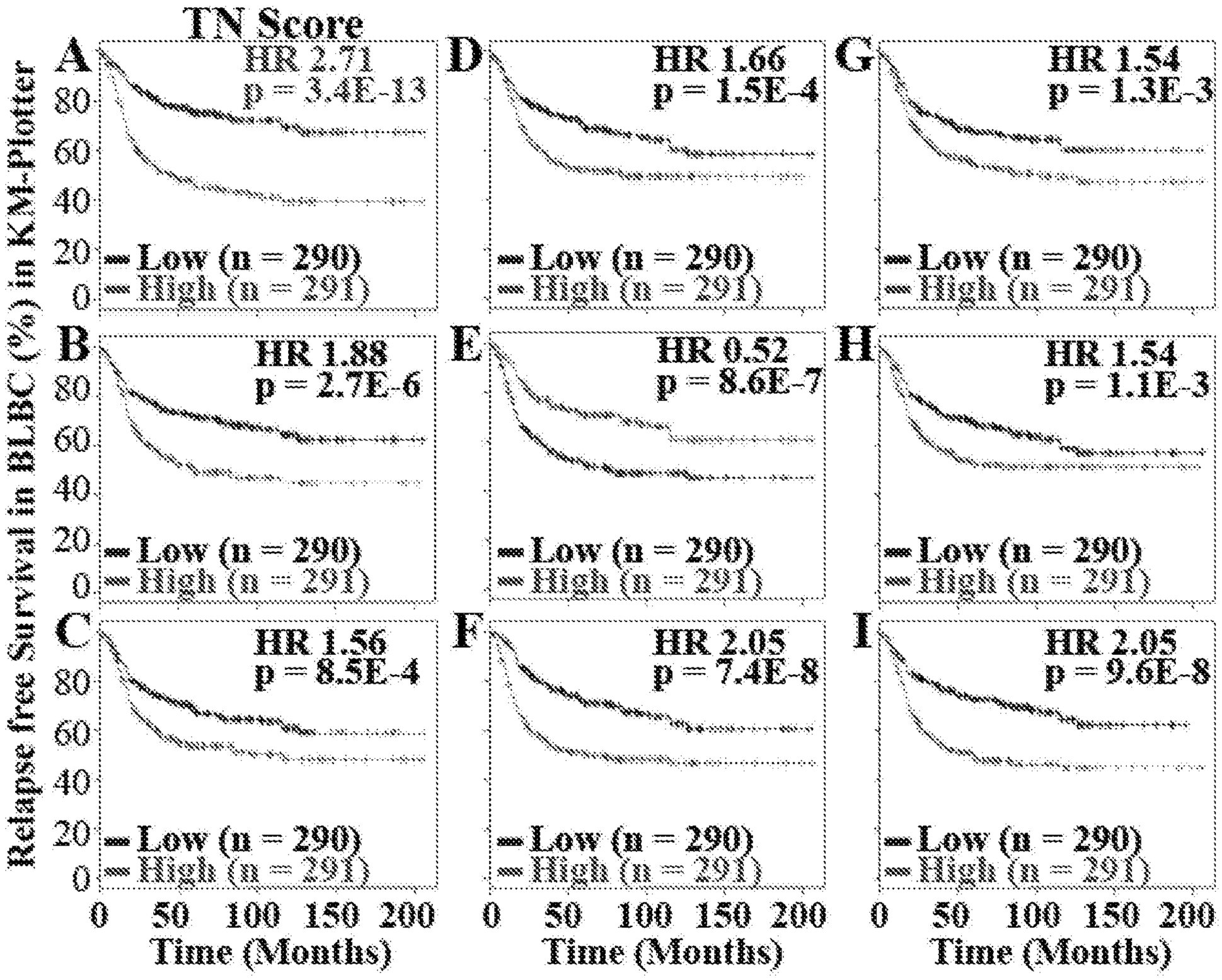

TN Score
A
HR 2.71
p = 3.4E-13
D
HR 1.66
p = 1.5E-4
G
HR 1.54
p = 1.3E-3
B
HR 1.88
p = 2.7E-6
E
HR 0.52
p = 8.6E-7
H
HR 1.54
p = 1.1E-3
C
HR 1.56
p = 8.5E-4
F
HR 2.05
p = 7.4E-8
I
HR 2.05
p = 9.6E-8
Low (n = 290)
High (n = 291)
Relapse free Survival in BLBC (%) in KM-Plotter
0
20
40
60
80
0 50 100 150 200
Time (Months)

FIG. 32

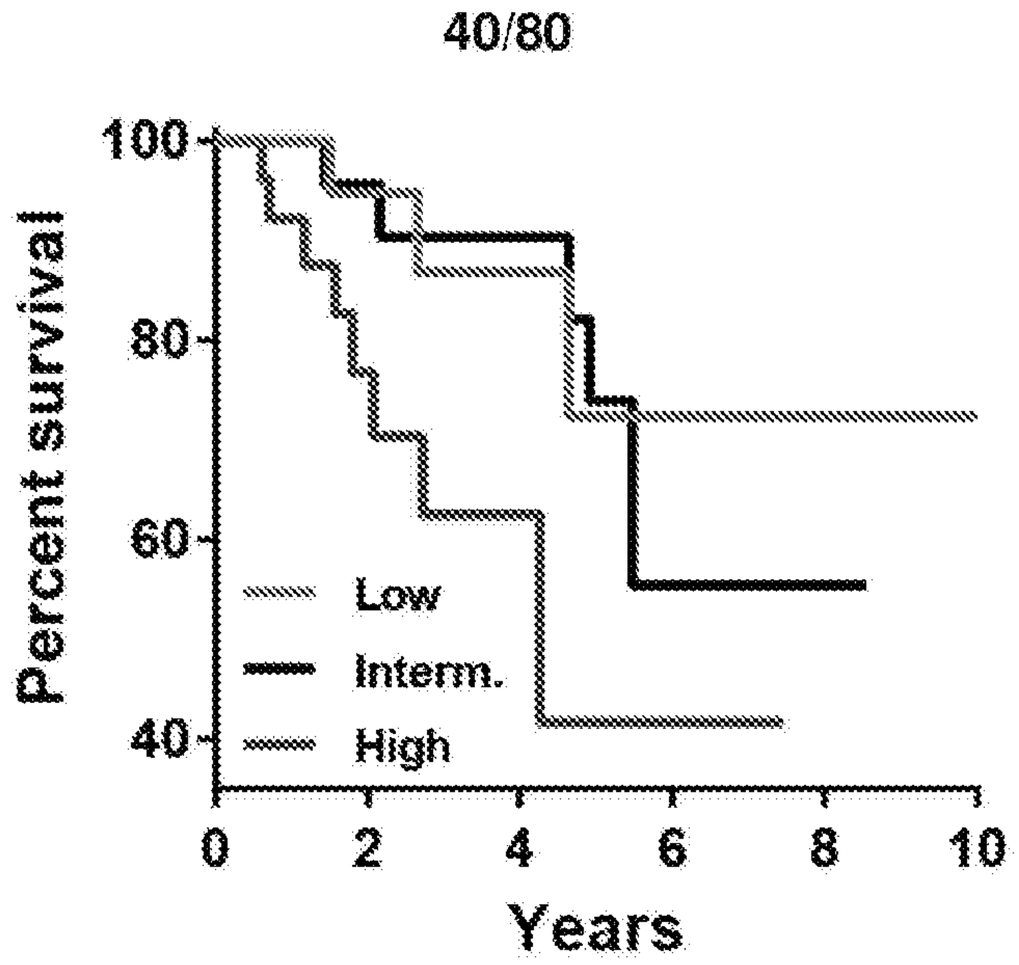

40/80
Percent survival
100
80
60
40
Low
Interm.
High
0
2
4
6
8
10
Years

FIG. 33

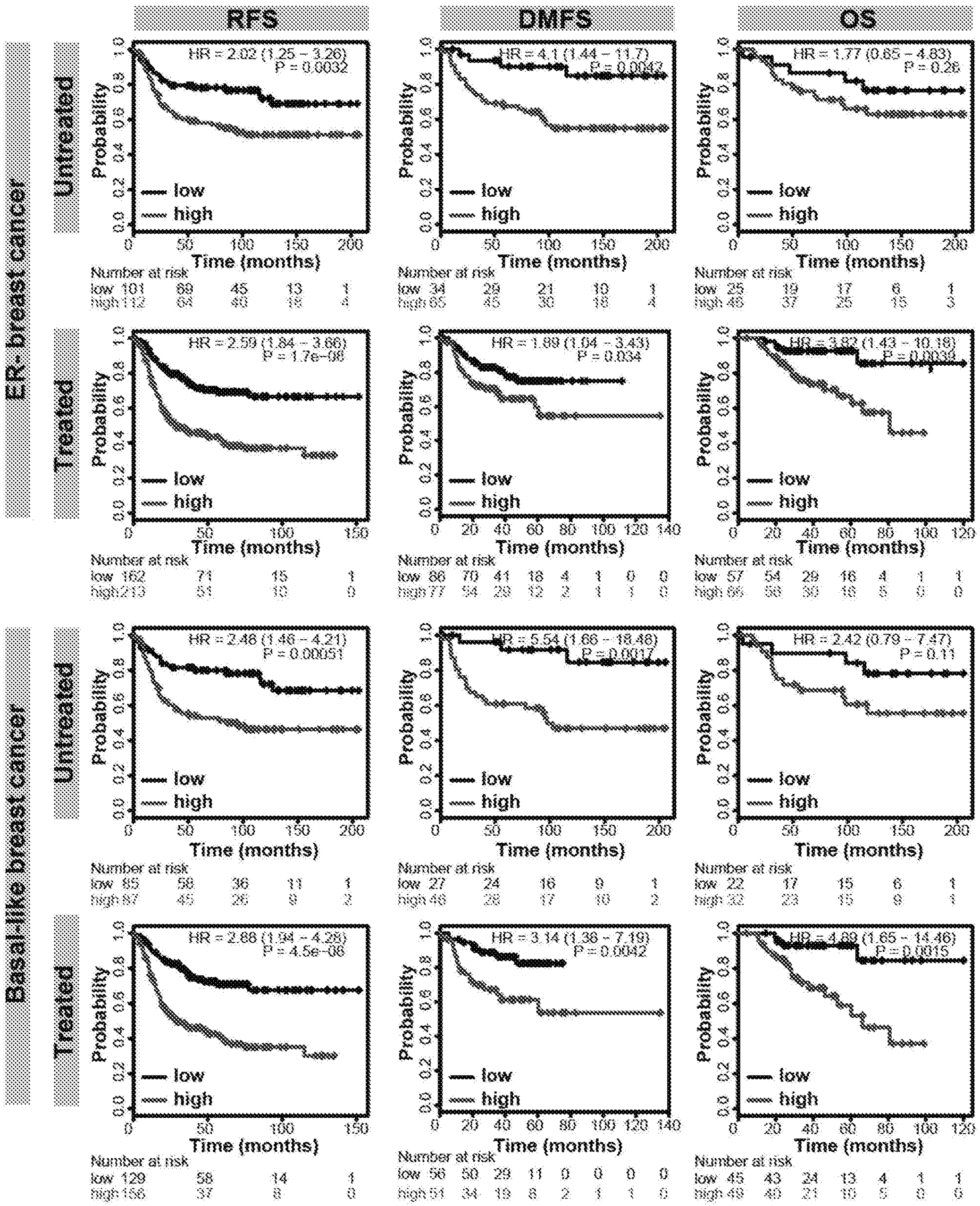

RFS
DMFS
OS
ER- breast cancer
Basal-like breast cancer
Untreated
Treated
Probability
Time (months)
low
high
HR = 2.02 (1.25 – 3.26)
P = 0.0032
HR = 4.1 (1.44 – 11.7)
P = 0.0042
HR = 1.77 (0.85 – 4.83)
P = 0.26
HR = 2.59 (1.84 – 3.66)
P = 1.7e–08
HR = 1.89 (1.04 – 3.43)
P = 0.034
HR = 3.82 (1.43 – 10.18)
P = 0.0039
HR = 2.46 (1.46 – 4.21)
P = 0.00051
HR = 5.54 (1.66 – 18.46)
P = 0.0017
HR = 2.42 (0.79 – 7.47)
P = 0.11
HR = 2.88 (1.94 – 4.28)
P = 4.5e–08
HR = 3.14 (1.36 – 7.19)
P = 0.0042
HR = 4.89 (1.65 – 14.46)
P = 0.0015
Number at risk
low 101 69 45 13 1
high 112 64 40 18 4
low 34 29 21 10 1
high 68 45 30 18 4
low 25 19 17 6 1
high 48 37 25 15 3
low 162 71 15 1
high 213 61 10 0
low 86 70 41 18 4 1 0 0
high 77 54 29 12 2 1 1 0
low 57 54 29 16 4 1 1
high 66 58 30 16 5 0 0
low 85 58 36 11 1
high 87 45 26 9 2
low 27 24 16 9 1
high 46 28 17 10 2
low 22 17 15 6 1
high 32 23 15 9 1
low 129 58 14 1
high 156 37 8 0
low 56 50 29 11 0 0 0 0
high 51 34 19 8 2 1 1 0
low 45 43 24 13 4 1 1
high 49 40 21 10 5 0 0

FIG. 34

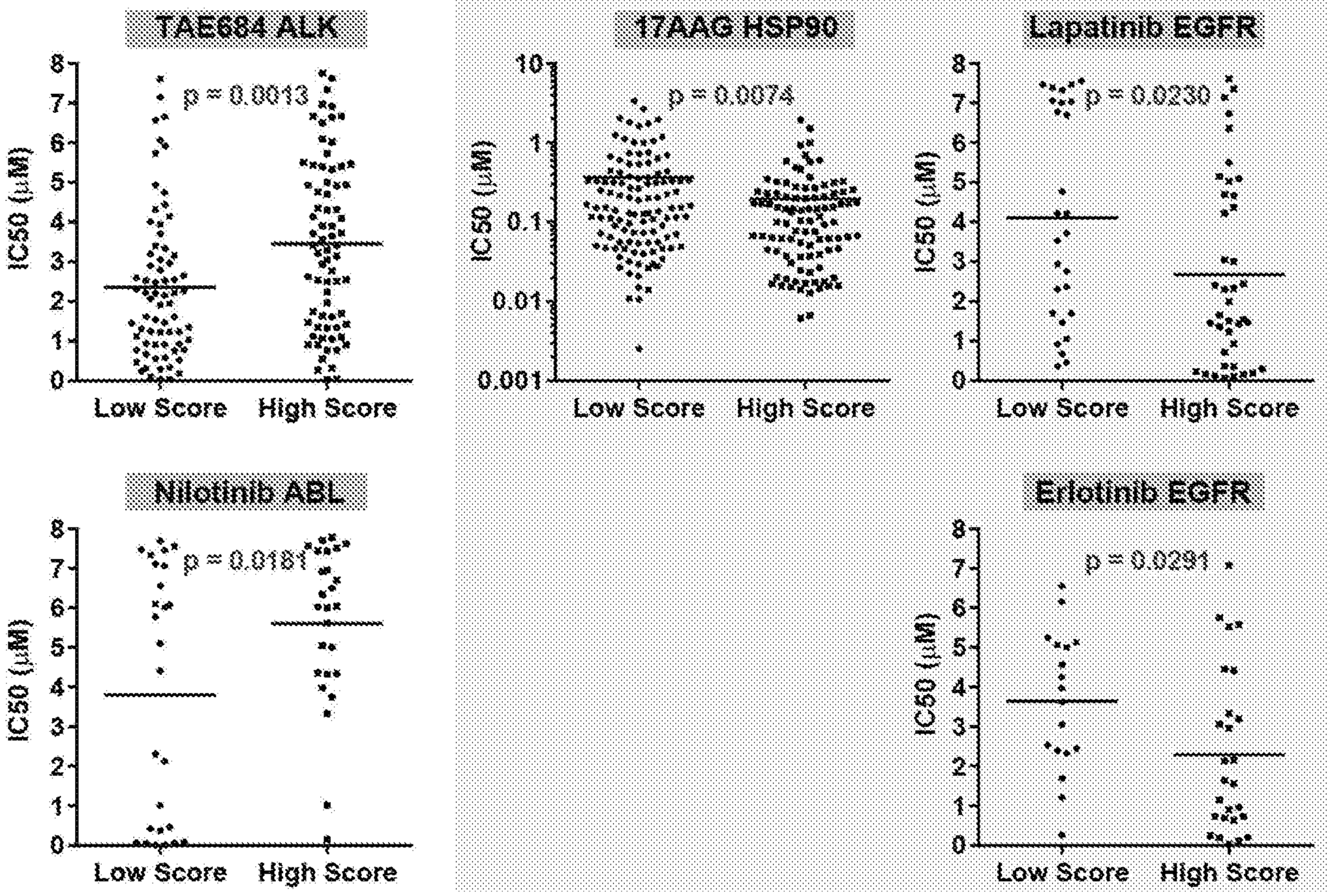
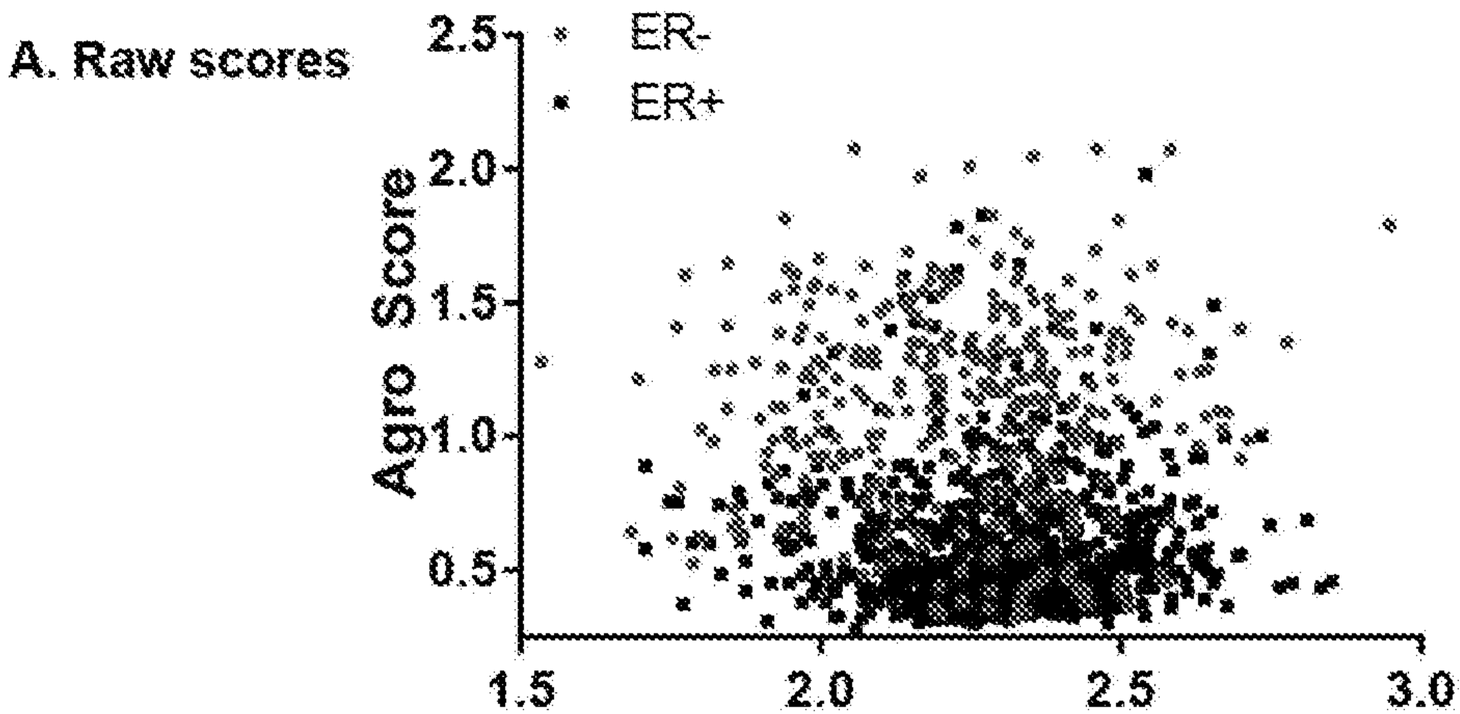

TAE684 ALK
p = 0.0013
IC50 (μM)
Low Score
High Score
17AAG HSP90
p = 0.0074
IC50 (μM)
Low Score
High Score
Lapatinib EGFR
p = 0.0230
IC50 (μM)
Low Score
High Score
Nilotinib ABL
p = 0.0181
IC50 (μM)
Low Score
High Score
Erlotinib EGFR
p = 0.0291
IC50 (μM)
Low Score
High Score

FIG. 35


A. Raw scores
ER-
ER+
Agro Score
2.5
2.0
1.5
1.0
0.5
1.5
2.0
2.5
3.0
TN Score

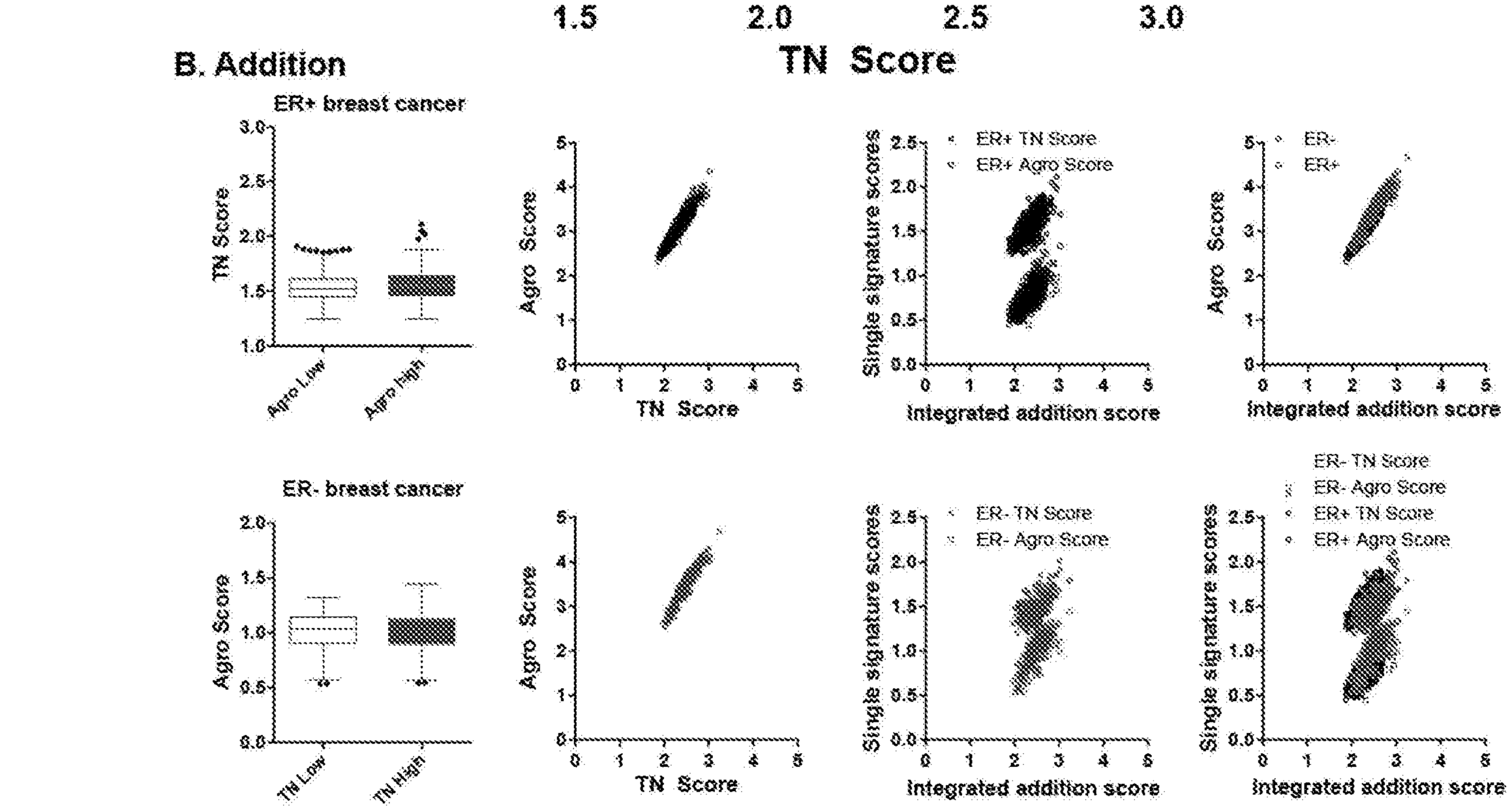

B. Addition
ER+ breast cancer
TN Score
Agro Low
Agro High
Agro Score
TN Score
ER+ TN Score
ER+ Agro Score
Single signature scores
Integrated addition score
ER-
ER+
Agro Score
Integrated addition score
ER- breast cancer
Agro Score
TN Low
TN High
TN Score
ER- TN Score
ER- Agro Score
Single signature scores
Integrated addition score
ER+ TN Score
ER+ Agro Score
Single signature scores
Integrated addition score

FIG. 36

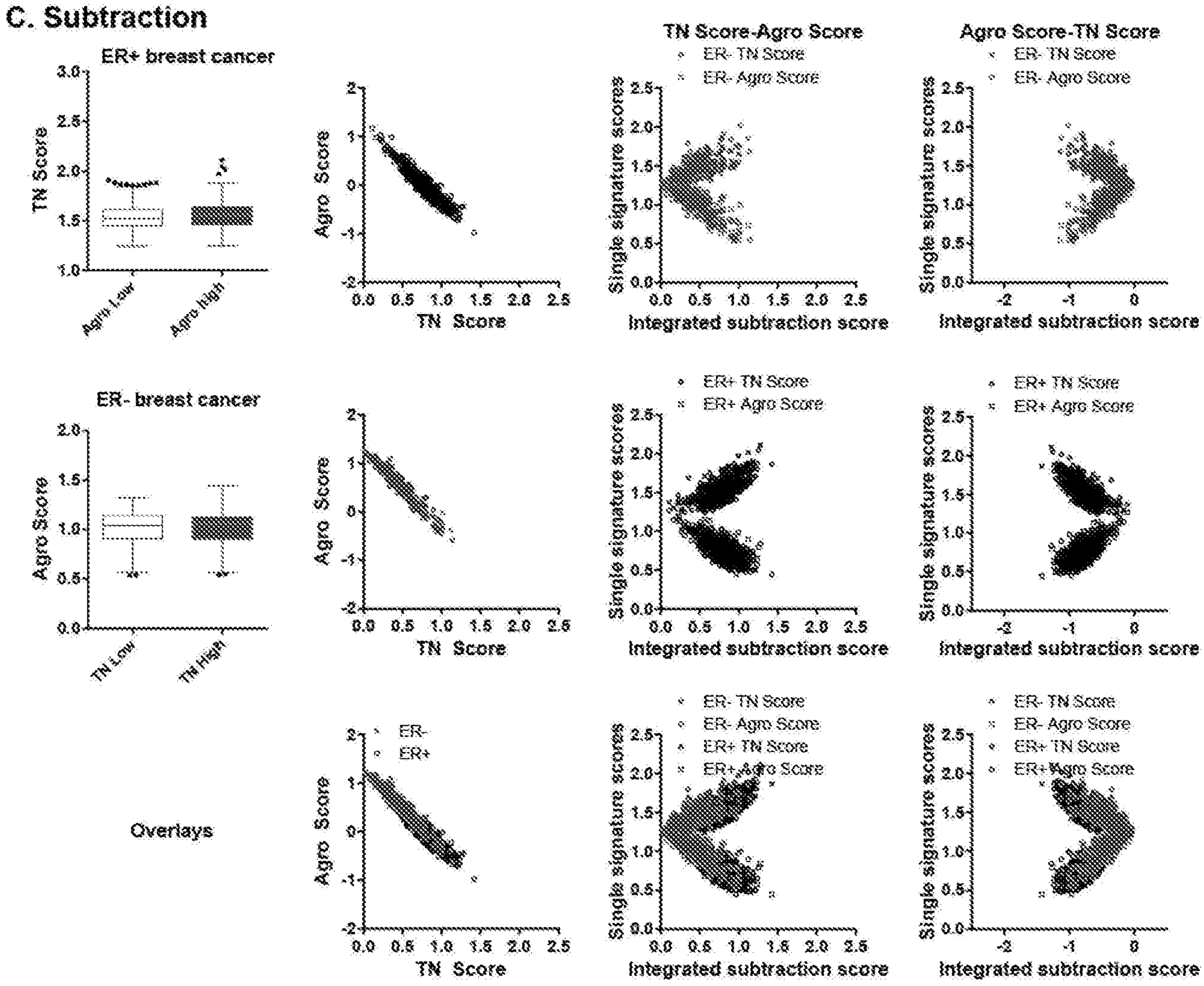

C. Subtraction
ER+ breast cancer
TN Score
Agro Low
Agro High
Agro Score
TN Score
TN Score-Agro Score
ER- TN Score
ER- Agro Score
Single signature scores
Integrated subtraction score
Agro Score-TN Score
ER- TN Score
ER- Agro Score
Single signature scores
Integrated subtraction score
ER- breast cancer
Agro Score
TN Low
TN High
ER+ TN Score
ER+ Agro Score
Overlays
ER-
ER+
ER- TN Score
ER- Agro Score
ER+ TN Score
ER+ Agro Score

FIG. 36 cont'd

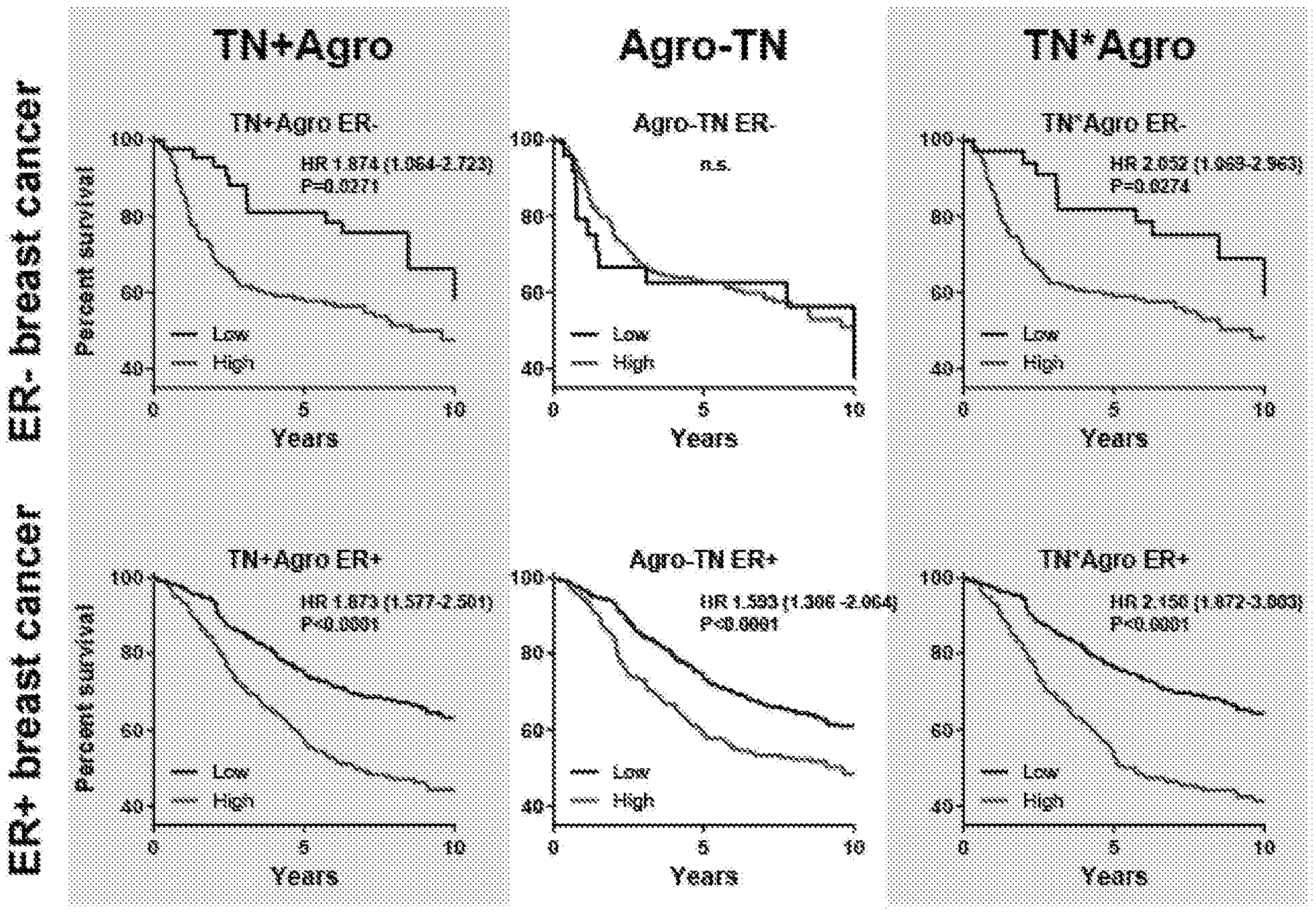

ER- breast cancer
ER+ breast cancer
TN+Agro
TN+Agro ER-
HR 1.874 (1.364-2.723)
P=0.0271
Agro-TN
Agro-TN ER-
n.s.
TN*Agro
TN*Agro ER-
HR 2.052 (1.393-2.963)
P=0.0274
TN+Agro ER+
HR 1.873 (1.577-2.501)
P<0.0001
Agro-TN ER+
HR 1.593 (1.308 -2.064)
P<0.0001
TN*Agro ER+
HR 2.158 (1.672-3.003)
P<0.0001
Agro/TN
Agro/TN ER-
n.s.
Agro^TN
Agro^TN ER-
n.s.
TN^Agro
TN^Agro ER-
HR 2.355 (1.524-3.463)
P<0.0001
Agro/TN ER+
HR 2.137 (1.654-2.973)
P<0.0001
Agro^TN ER+
HR 2.106 (1.618-2.908)
P<0.0001
TN^Agro ER+
HR 2.435 (2.065-3.270)
P<0.0001
Percent survival
100
80
60
40
0
5
10
Years
Low
High

FIG. 37

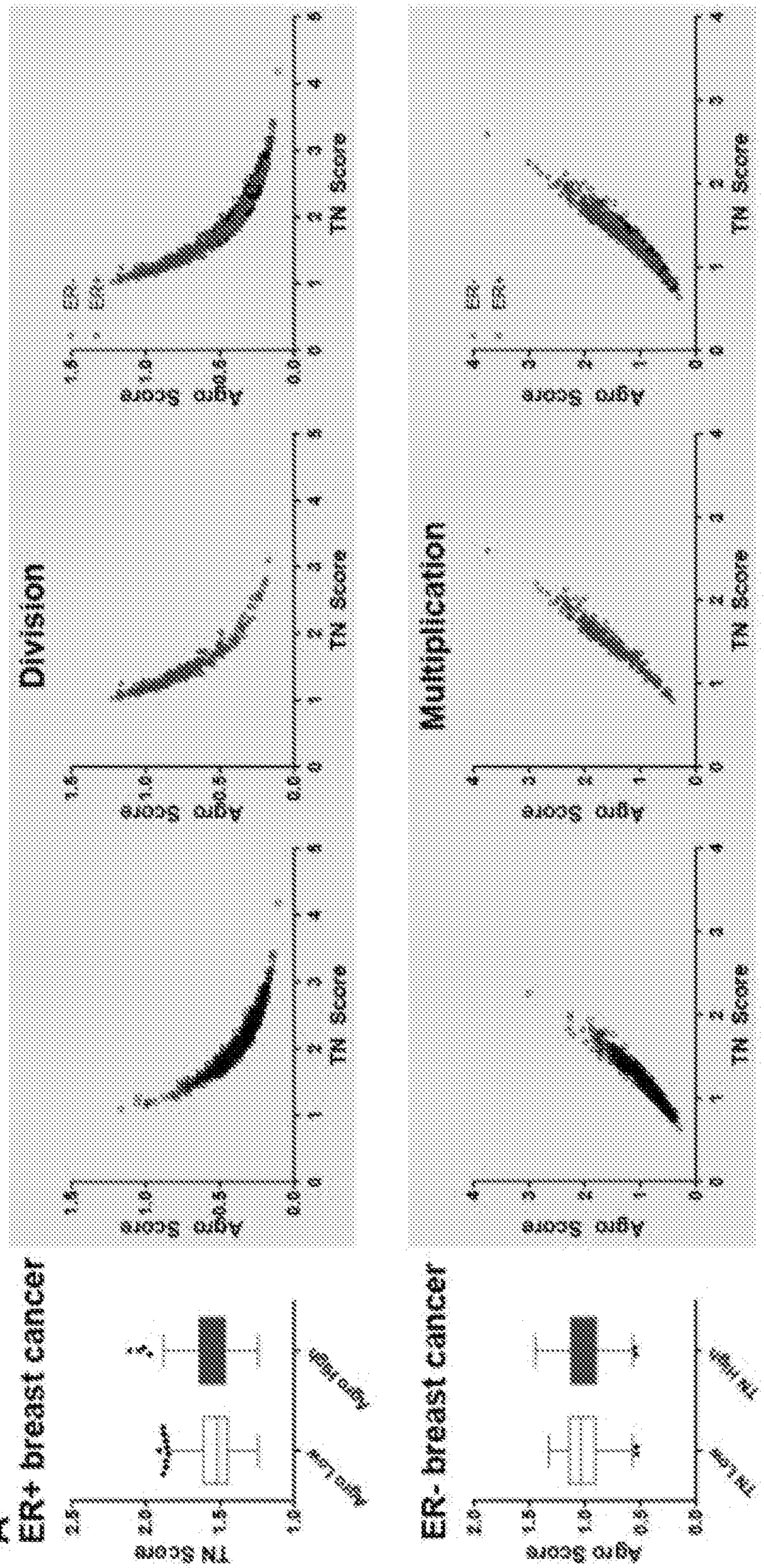

Division
Multiplication
Age score
TM score
ER+ breast cancer
ER- breast cancer
A

FIG. 38

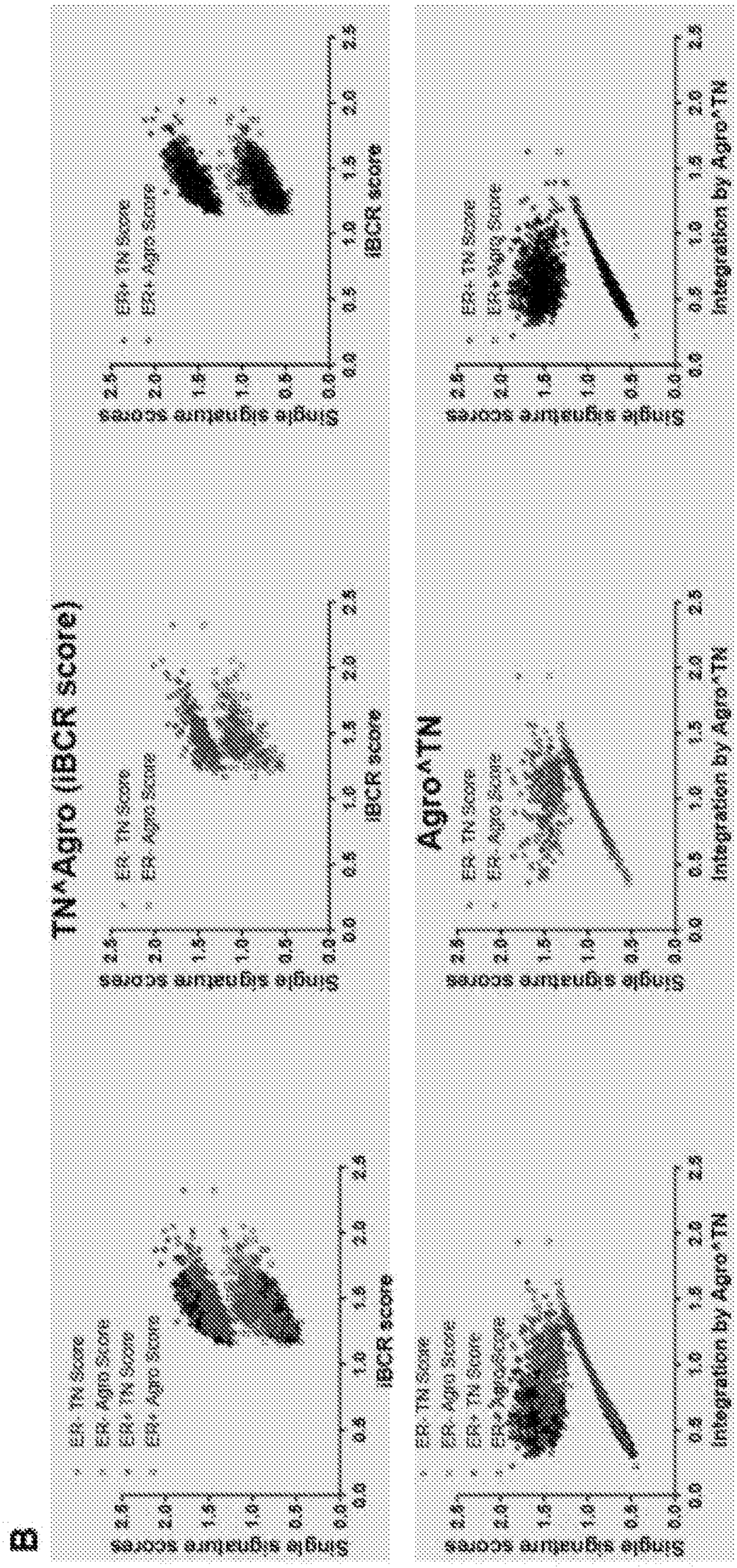

B
ER- TN+ score
ER- Agro score
ER+ TN- score
ER+ Agro score
iBCR score
Single signature scores
TN^Agro (iBCR score)
AgroTN
Integration by Agro^TN

FIG. 38 cont'd

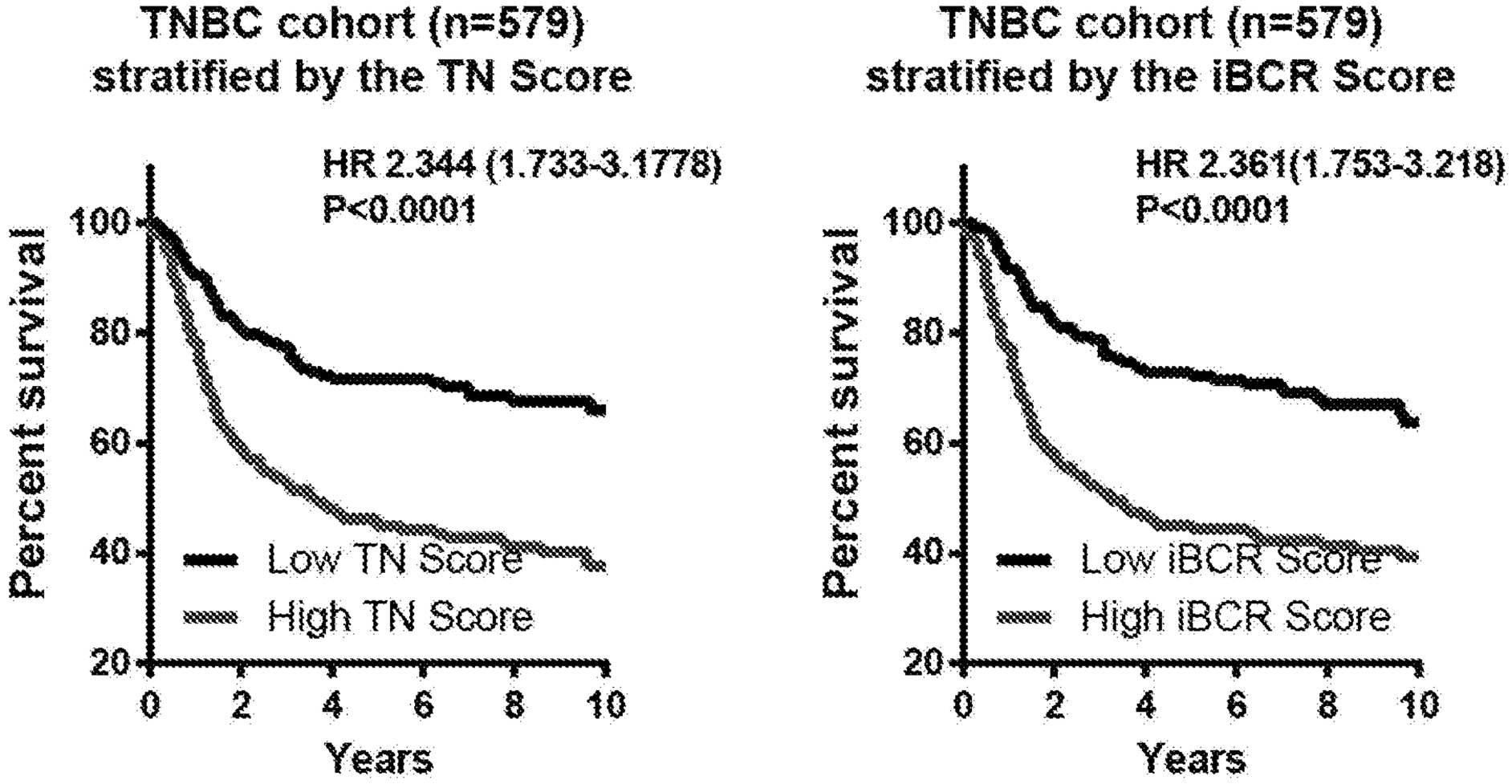

TNBC cohort (n=579)
stratified by the TN Score
HR 2.344 (1.733-3.1778)
P<0.0001
Percent survival
100
80
60
40
20
0
2
4
6
8
10
Years
Low TN Score
High TN Score
TNBC cohort (n=579)
stratified by the iBCR Score
HR 2.361(1.753-3.218)
P<0.0001
Percent survival
100
80
60
40
20
0
2
4
6
8
10
Years
Low iBCR Score
High iBCR Score

FIG. 39

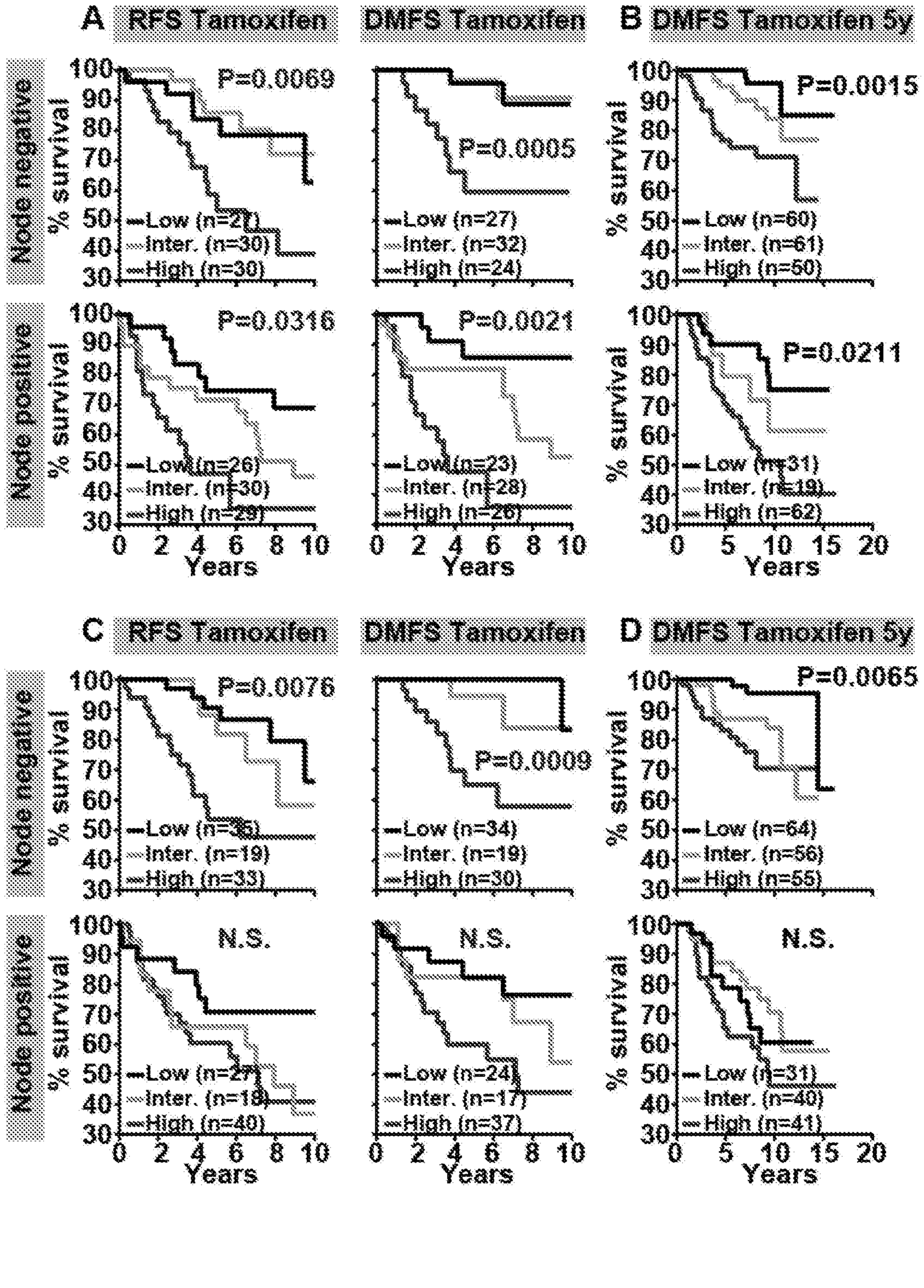

A
RFS Tamoxifen
DMFS Tamoxifen
B
DMFS Tamoxifen 5y
Node negative
% survival
P=0.0069
P=0.0005
P=0.0015
Low (n=27)
Inter. (n=32)
High (n=24)
Low (n=60)
Inter. (n=61)
High (n=50)
Node positive
P=0.0316
P=0.0021
P=0.0211
Low (n=26)
Inter. (n=30)
High (n=29)
Low (n=23)
Inter. (n=28)
High (n=26)
Low (n=31)
Inter. (n=19)
High (n=62)
0 2 4 6 8 10
0 5 10 15 20
Years
C
RFS Tamoxifen
DMFS Tamoxifen
D
DMFS Tamoxifen 5y
P=0.0076
P=0.0009
P=0.0065
Low (n=35)
Inter. (n=19)
High (n=33)
Low (n=34)
Inter. (n=19)
High (n=30)
Low (n=64)
Inter. (n=56)
High (n=55)
N.S.
Low (n=27)
Inter. (n=19)
High (n=40)
Low (n=24)
Inter. (n=17)
High (n=37)
Low (n=31)
Inter. (n=40)
High (n=41)

FIG. 40

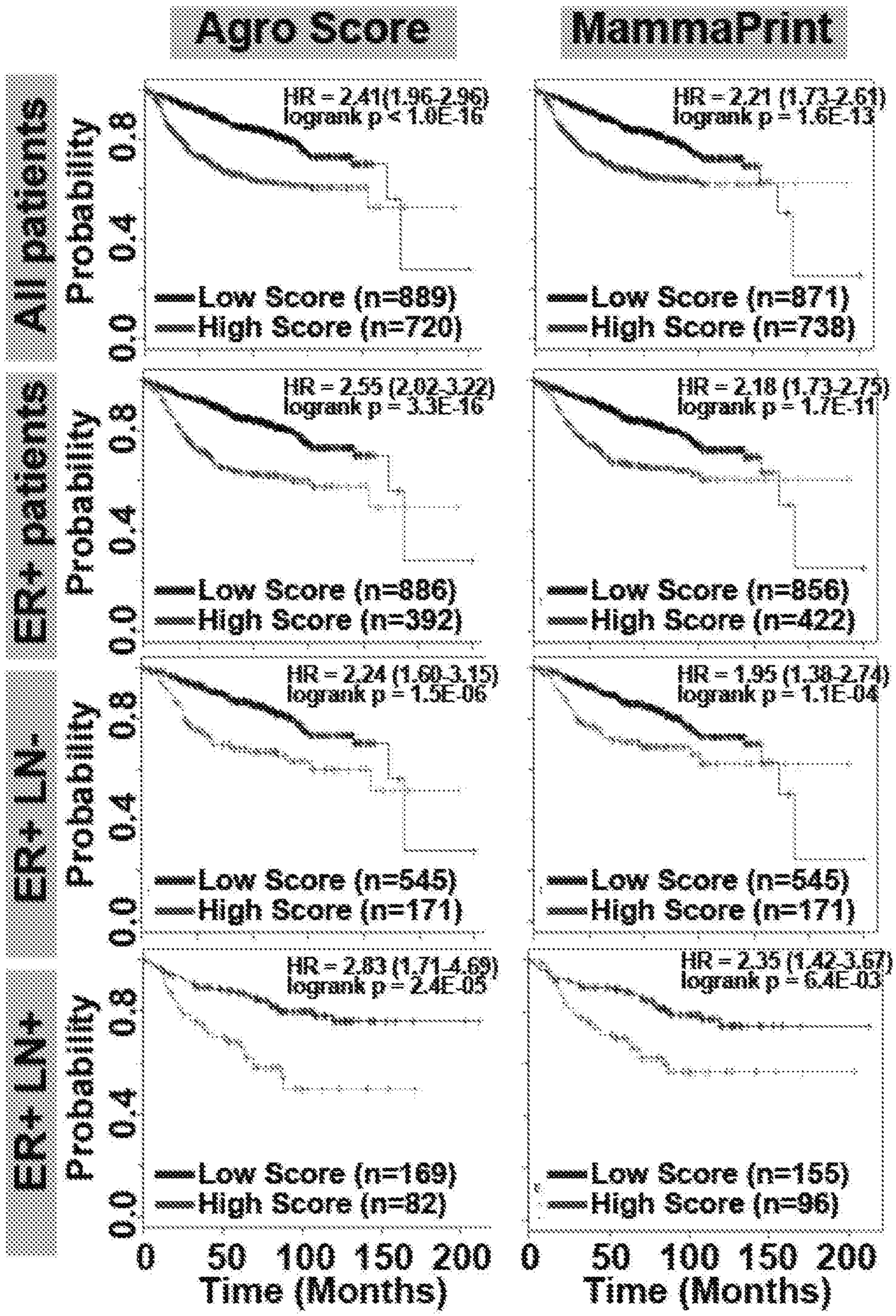

Agro Score
MammaPrint
All patients
ER+ patients
ER+ LN-
ER+ LN+
Probability
0.0
0.4
0.8
HR = 2.41(1.96-2.96)
logrank p < 1.0E-16
Low Score (n=889)
High Score (n=720)
HR = 2.21 (1.73-2.81)
logrank p = 1.6E-13
Low Score (n=871)
High Score (n=738)
HR = 2.55 (2.02-3.22)
logrank p = 3.3E-16
Low Score (n=886)
High Score (n=392)
HR = 2.18 (1.73-2.75)
logrank p = 1.7E-11
Low Score (n=856)
High Score (n=422)
HR = 2.24 (1.60-3.15)
logrank p = 1.5E-06
Low Score (n=545)
High Score (n=171)
HR = 1.95 (1.38-2.74)
logrank p = 1.1E-04
Low Score (n=545)
High Score (n=171)
HR = 2.83 (1.71-4.69)
logrank p = 2.4E-05
Low Score (n=169)
High Score (n=82)
HR = 2.35 (1.42-3.87)
logrank p = 6.4E-03
Low Score (n=155)
High Score (n=96)
0
50
100
150
200
Time (Months)

FIG. 41

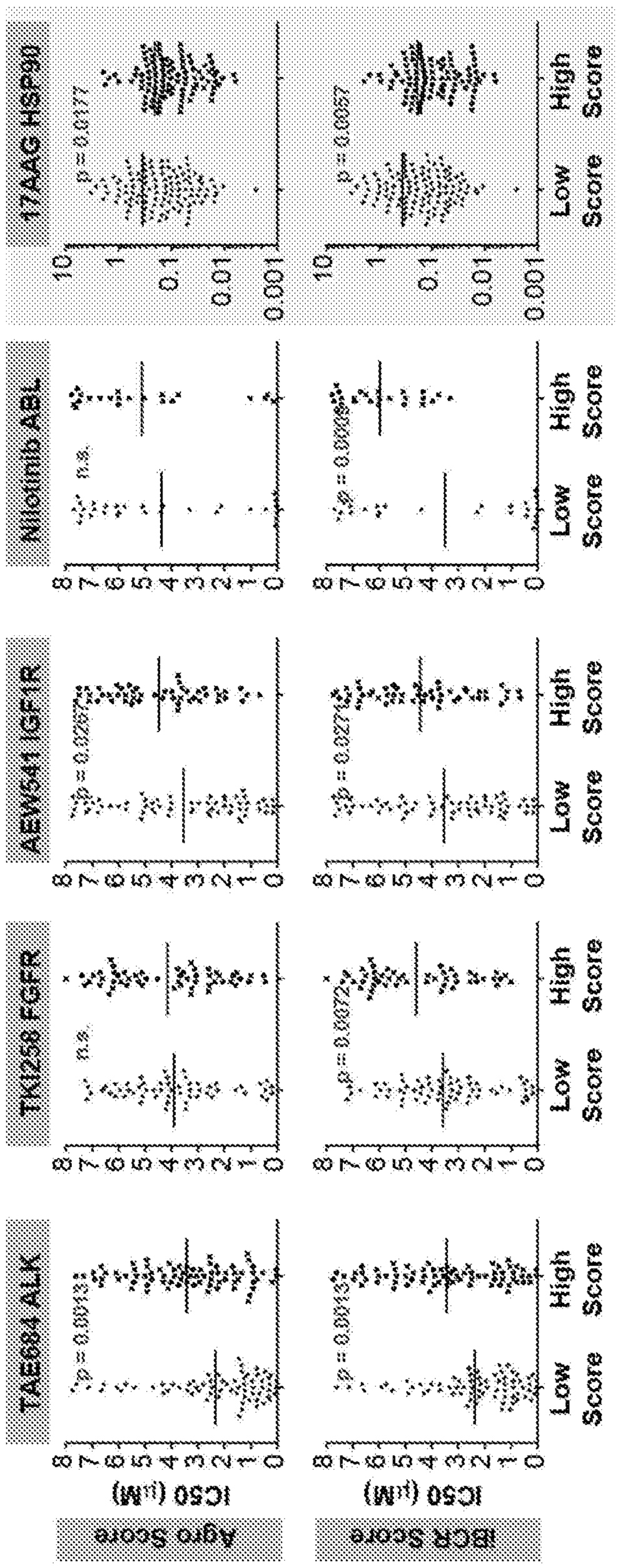

17AAG HSP90
p = 0.0177
p = 0.0057
10
1
0.1
0.01
0.001
Low Score
High Score
Nilotinib ABL
n.s.
p = 0.0008
AEW541 IGF1R
p = 0.0267
p = 0.0277
TKI258 FGFR
n.s.
p = 0.0072
TAE684 ALK
p = 0.0013
p = 0.0013
0
1
2
3
4
5
6
7
8
Low Score
High Score
IC50 (μM)
IBCR Score
Agro Score

FIG. 42

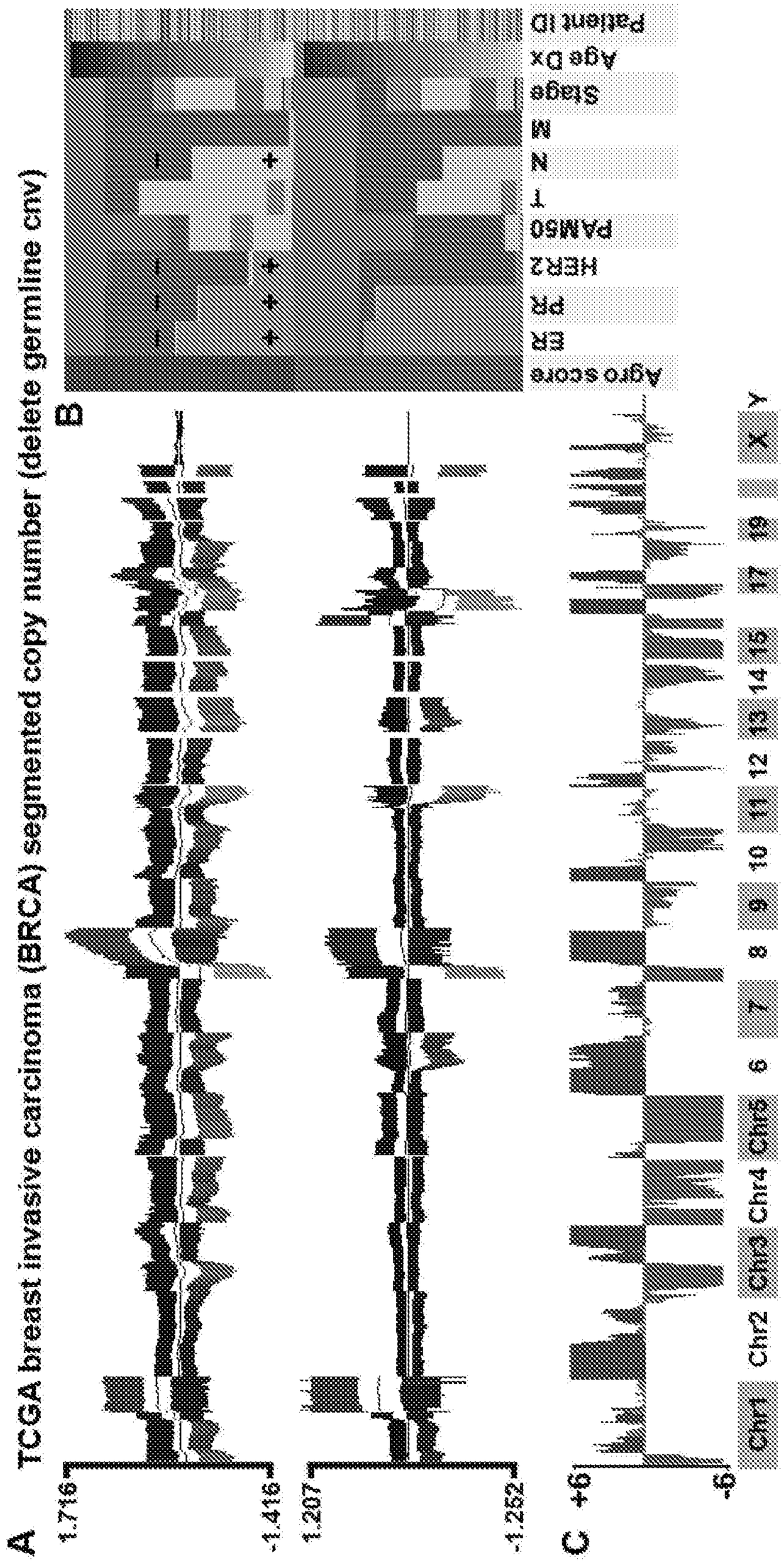

A TCGA breast invasive carcinoma (BRCA) segmented copy number (deleted germline cnv)
1.716
-1.416
1.207
-1.252
B
C +6
-6
Chr1 Chr2 Chr3 Chr4 Chr5 6 7 8 9 10 11 12 13 14 15 17 19 X Y
Agro score
ER
PR
HER2
PAM50
T
N
M
Stage
Age Dx
Patient ID

FIG. 43

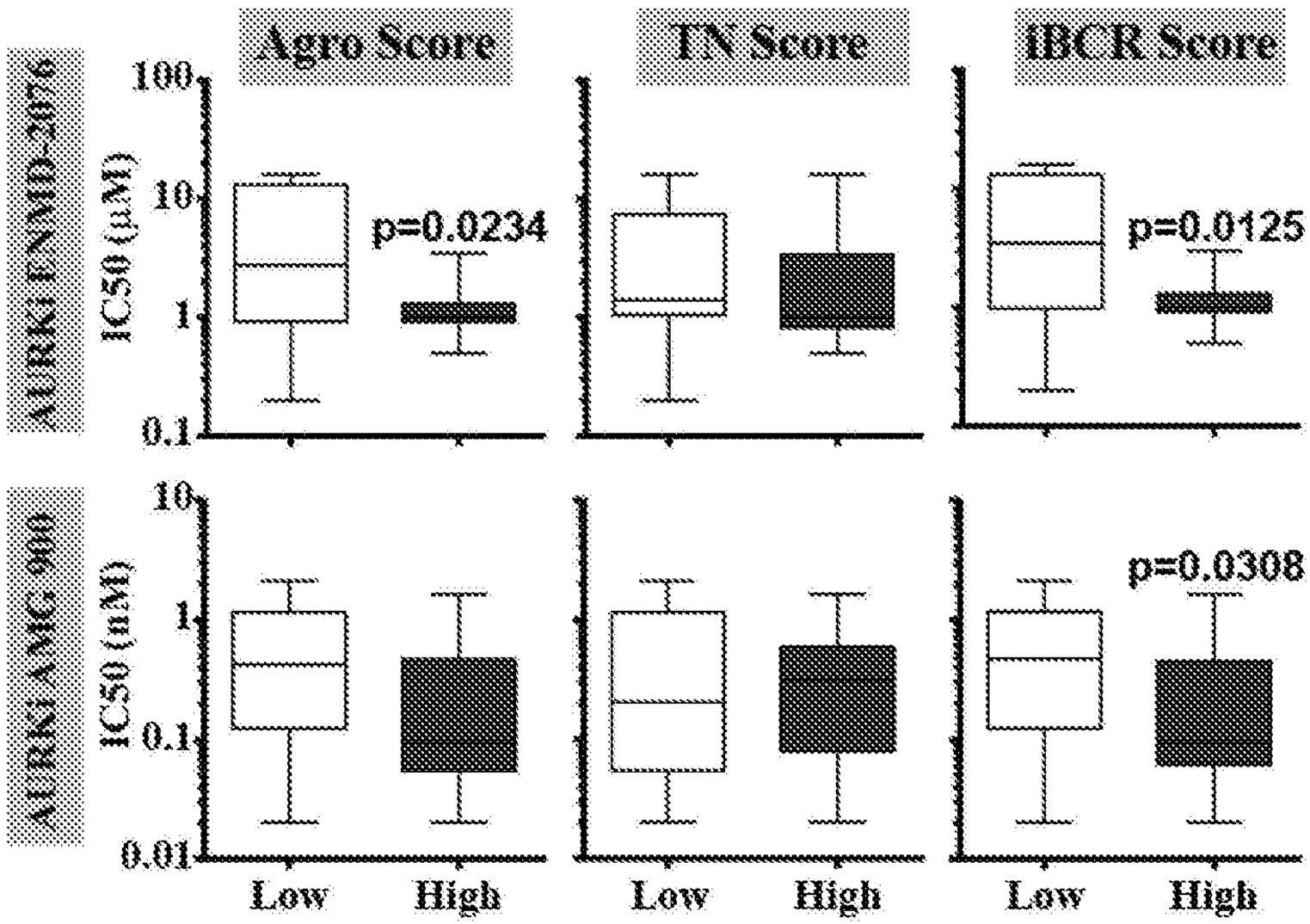
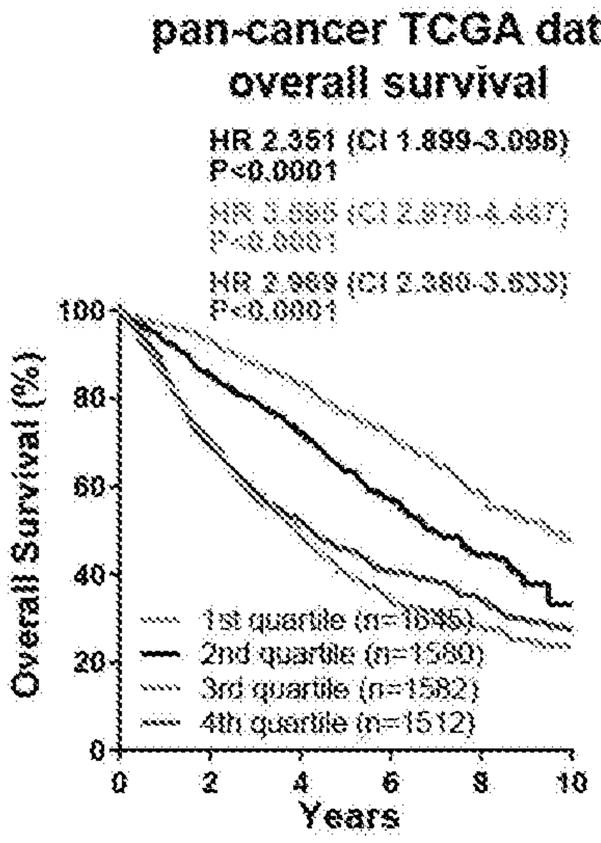
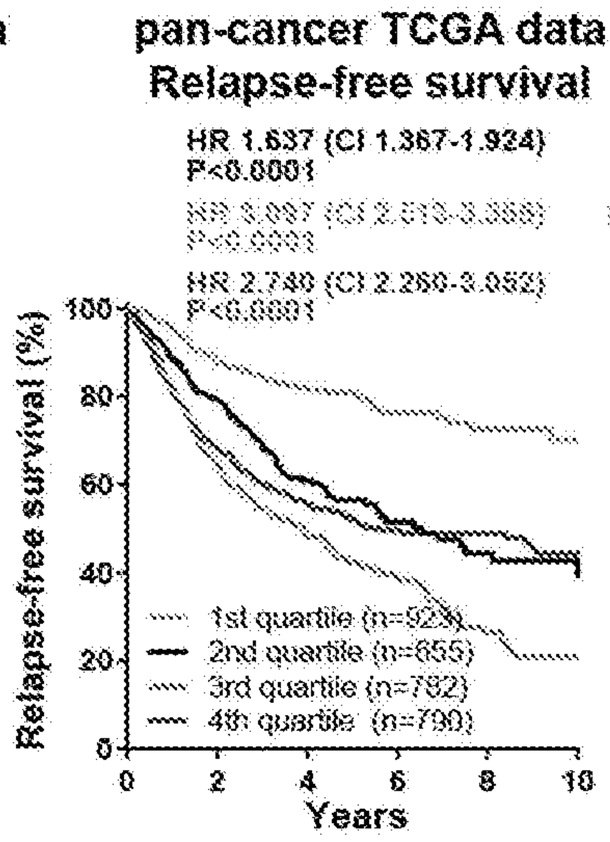
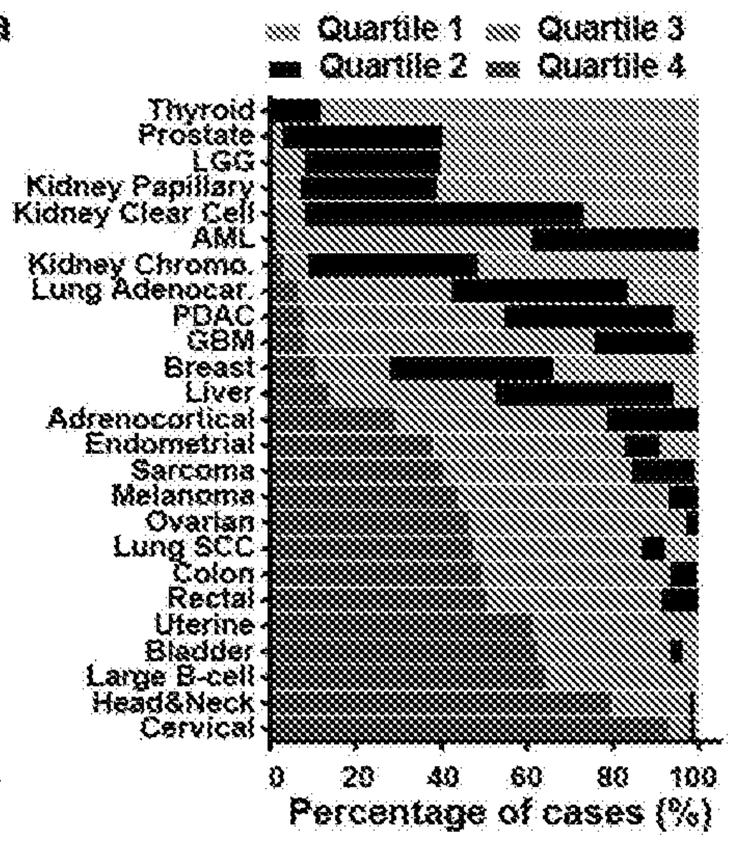
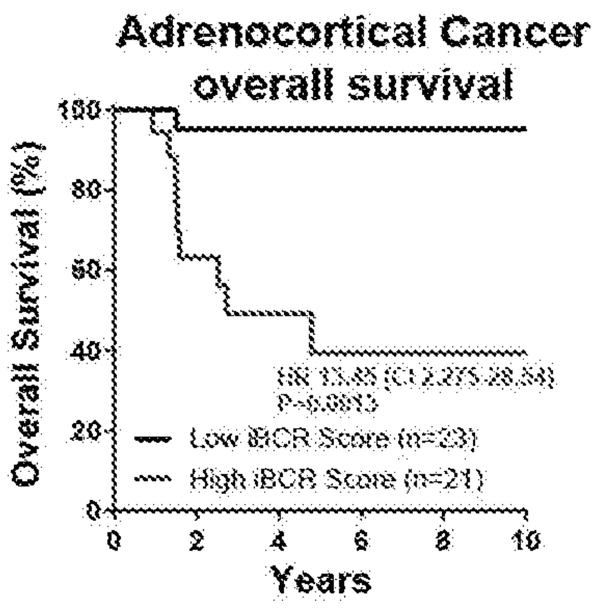

Agro Score
TN Score
IBCR Score
AURKi ENMD-2076
IC50 (μM)
100
10
1
0.1
p=0.0234
p=0.0125
AURKi AMG 900
IC50 (nM)
10
1
0.1
0.01
p=0.0308
Low
High
Low
High
Low
High

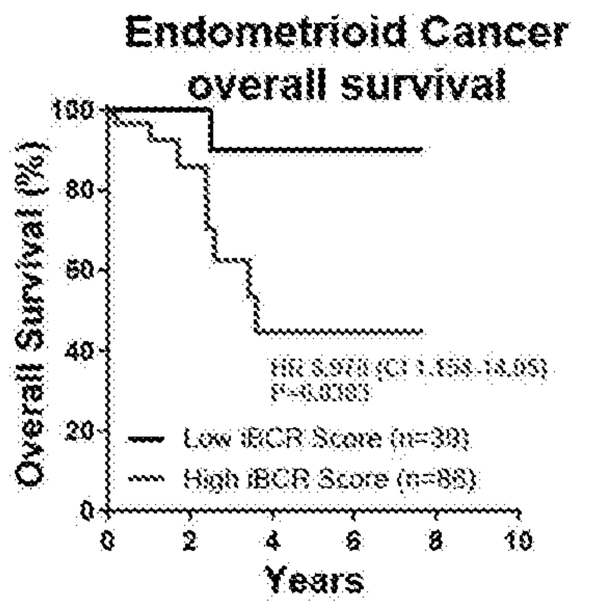
FIG. 44

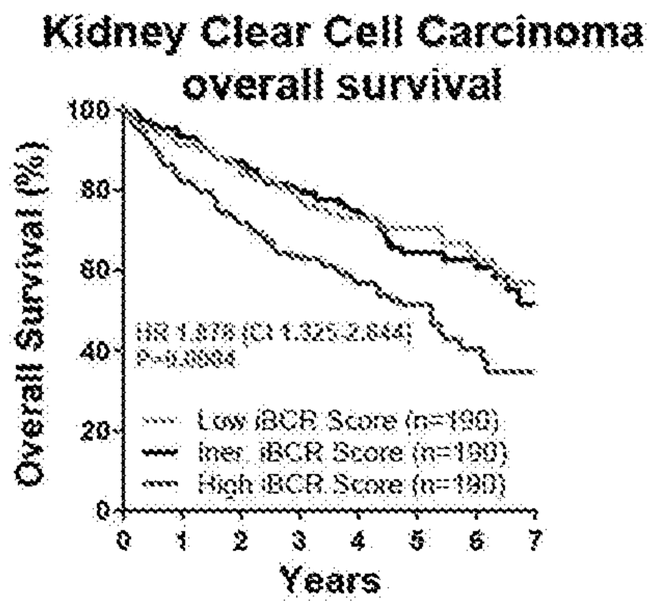

pan-cancer TCGA data
overall survival
HR 2.351 (CI 1.899-3.098)
P<0.0001
HR 3.585 (CI 2.870-4.447)
P<0.0001
HR 2.989 (CI 2.380-3.633)
P<0.0001
Overall Survival (%)
100
80
60
40
20
0
0
2
4
6
8
10
Years
1st quartile (n=1545)
2nd quartile (n=1580)
3rd quartile (n=1582)
4th quartile (n=1512)

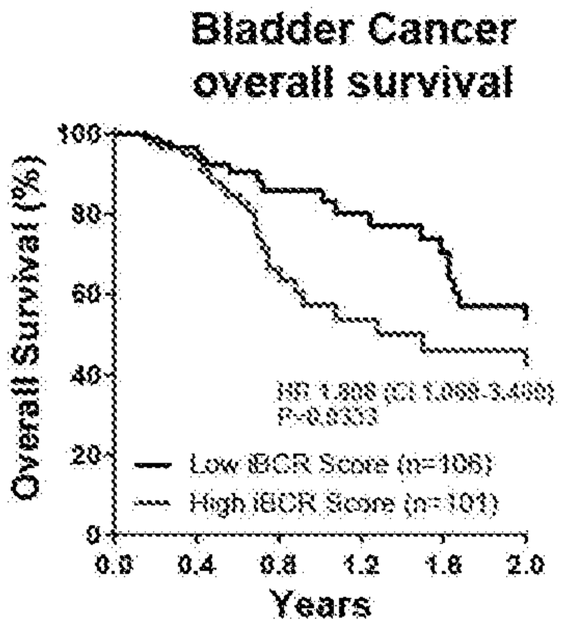

pan-cancer TCGA data
Relapse-free survival
HR 1.637 (CI 1.367-1.924)
P<0.0001
HR 3.097 (CI 2.513-3.358)
P<0.0001
HR 2.740 (CI 2.260-3.052)
P<0.0001
Relapse-free survival (%)
100
80
60
40
20
0
0
2
4
6
8
10
Years
1st quartile (n=923)
2nd quartile (n=655)
3rd quartile (n=782)
4th quartile (n=790)

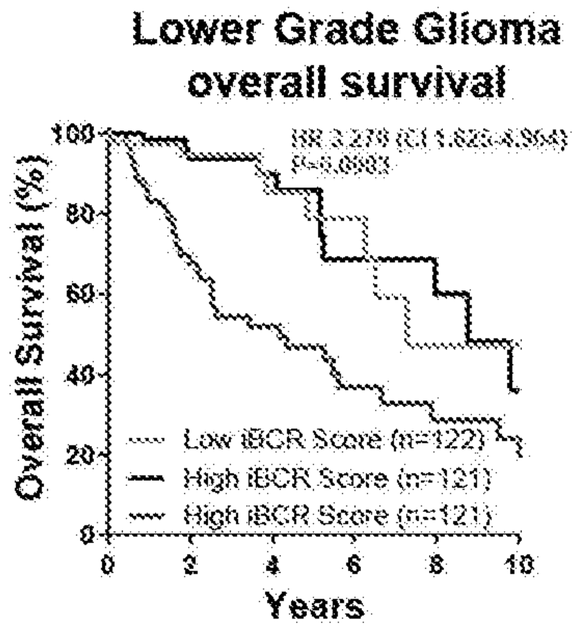

Quartile 1
Quartile 2
Quartile 3
Quartile 4
Thyroid
Prostate
LGG
Kidney Papillary
Kidney Clear Cell
AML
Kidney Chromo.
Lung Adenocar.
PDAC
GBM
Breast
Liver
Adrenocortical
Endometrial
Sarcoma
Melanoma
Ovarian
Lung SCC
Colon
Rectal
Uterine
Bladder
Large B-cell
Head&Neck
Cervical
0
20
40
60
80
100
Percentage of cases (%)


Adrenocortical Cancer
overall survival
Overall Survival (%)
100
80
60
40
20
0
HR 13.85 (CI 2.275-28.54)
P=0.0013
Low iBCR Score (n=23)
High iBCR Score (n=21)
0
2
4
6
8
10
Years


Endometrioid Cancer
overall survival
Overall Survival (%)
100
80
60
40
20
0
HR 5.978 (CI 1.658-18.05)
P=0.0383
Low iBCR Score (n=39)
High iBCR Score (n=88)
0
2
4
6
8
10
Years


Kidney Clear Cell Carcinoma
overall survival
Overall Survival (%)
100
80
60
40
20
0
HR 1.879 (CI 1.325-2.644)
P=0.0004
Low iBCR Score (n=190)
Inter. iBCR Score (n=190)
High iBCR Score (n=190)
0
1
2
3
4
5
6
7
Years


Bladder Cancer
overall survival
Overall Survival (%)
100
80
60
40
20
0
HR 1.838 (CI 1.068-3.488)
P=0.0333
Low iBCR Score (n=106)
High iBCR Score (n=101)
0.0
0.4
0.8
1.2
1.6
2.0
Years


Lower Grade Glioma
overall survival
HR 3.279 (CI 1.826-6.964)
P=0.0003
Overall Survival (%)
100
80
60
40
20
0
Low iBCR Score (n=122)
High iBCR Score (n=121)
High iBCR Score (n=121)
0
2
4
6
8
10
Years

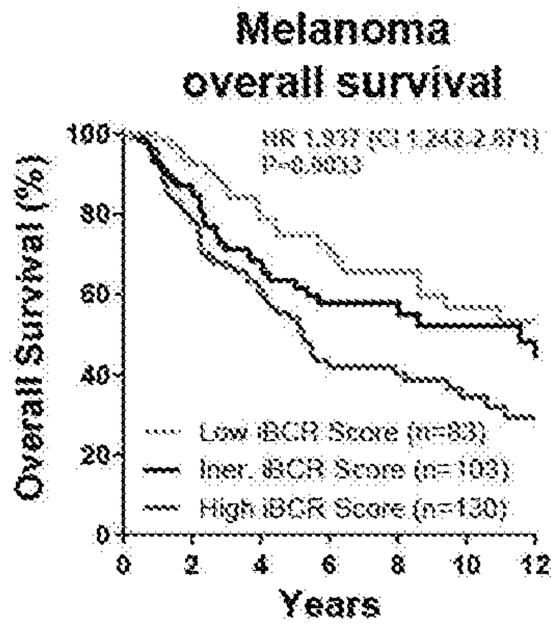
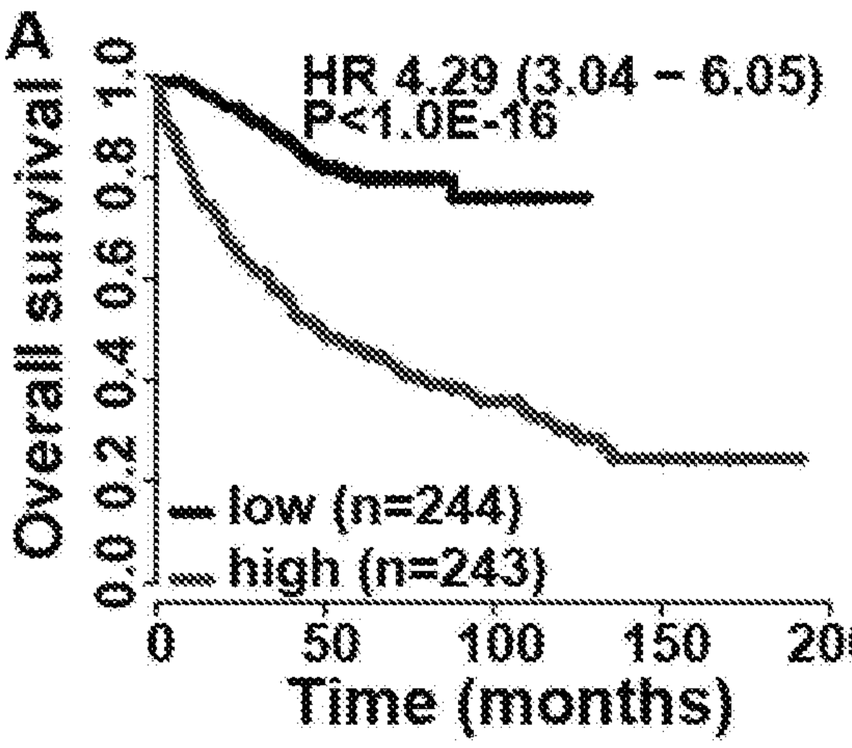
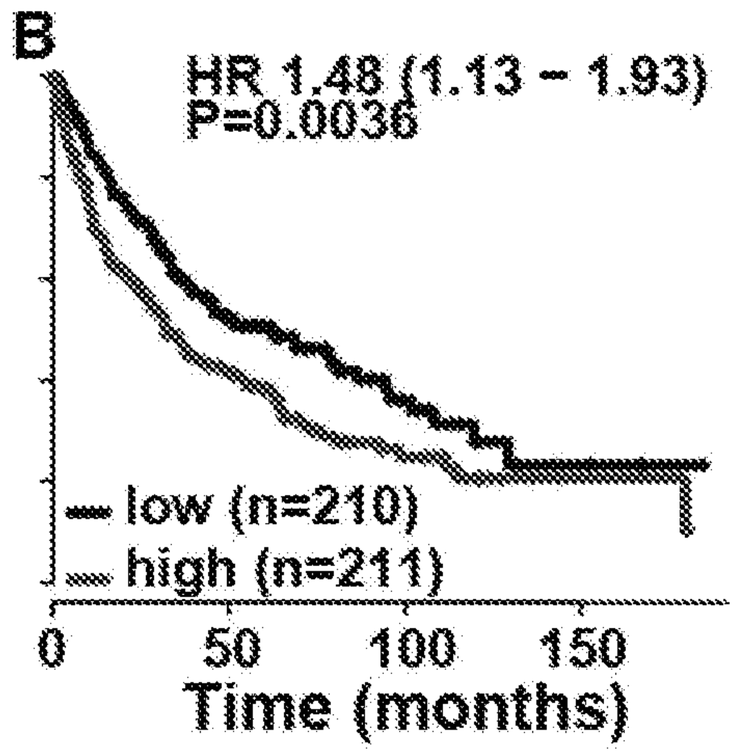

Melanoma
overall survival
HR 1.937 (CI 1.243-2.871)
P=0.0033
Overall Survival (%)
100
80
60
40
20
0
Low iBCR Score (n=83)
Inter. iBCR Score (n=103)
High iBCR Score (n=130)
0
2
4
6
8
10
12
Years

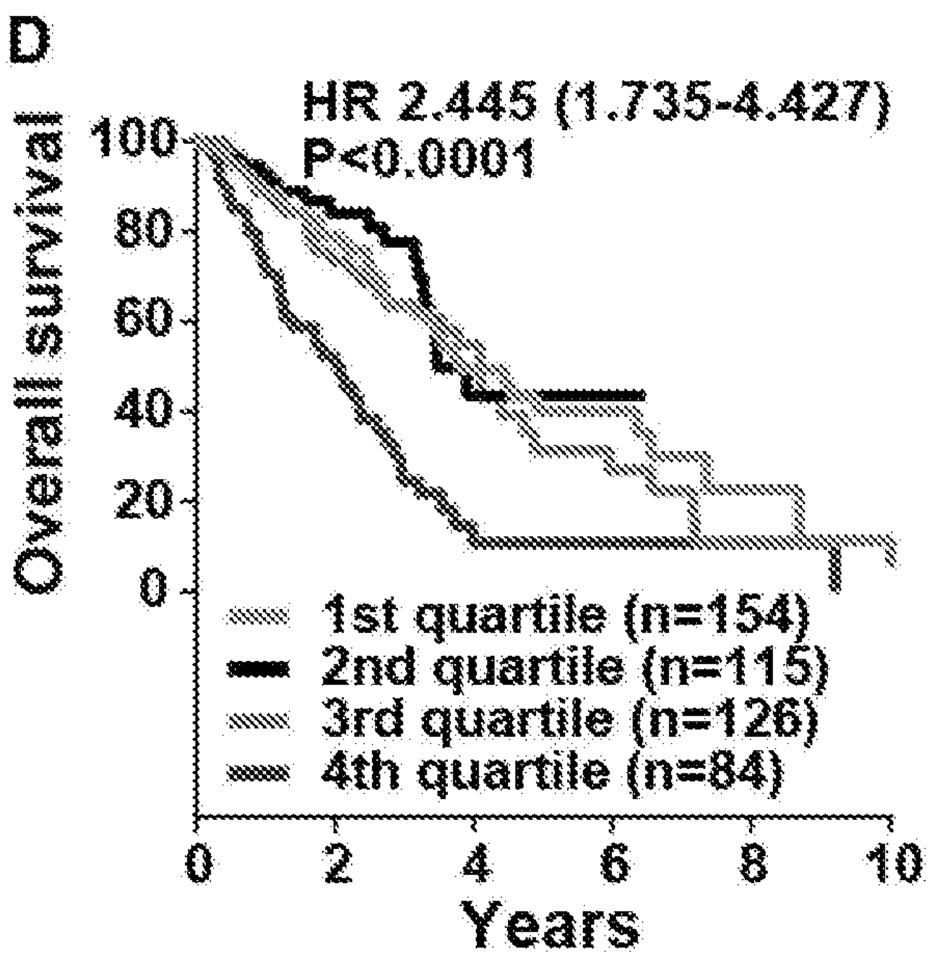
FIG. 45





C Multivariate analysis

	P value	Hazard Ratio
iBCR sig.	<1.0E-16	2.53 (1.93-3.32)
Histology	0.0249	1.25 (1.03-1.52)
Stage	<1.0E-16	1.42 (1.24-1.64)



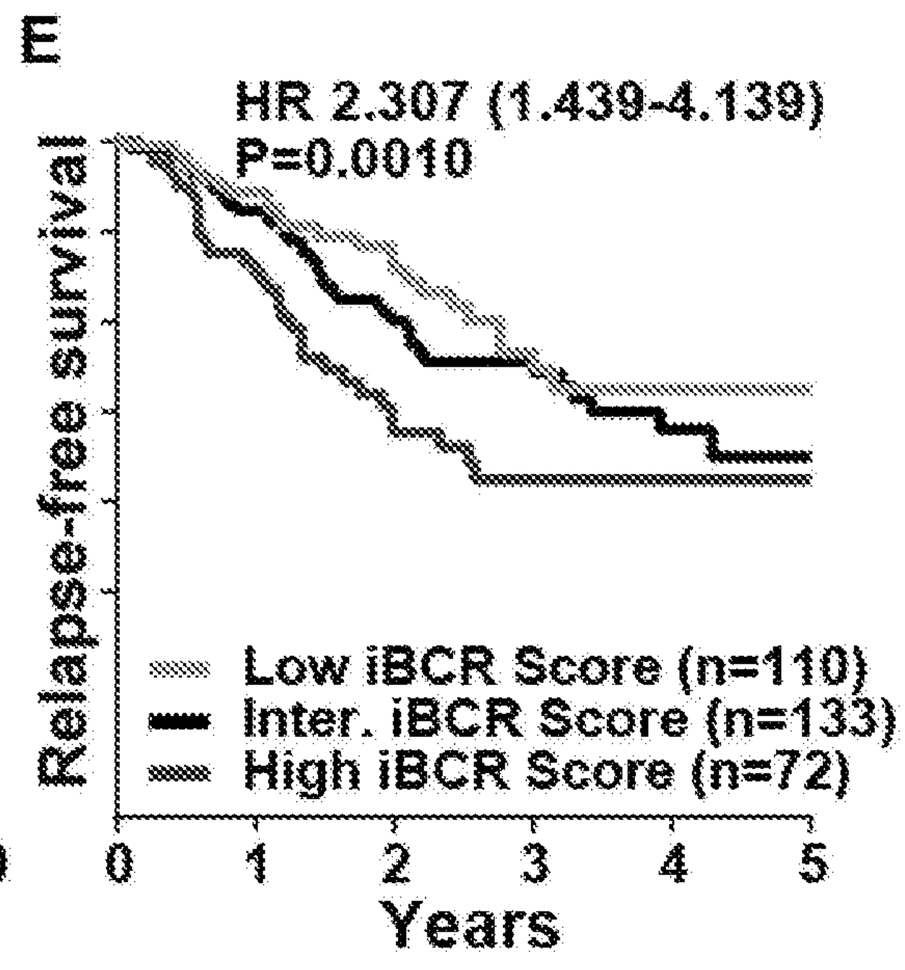


FIG. 46

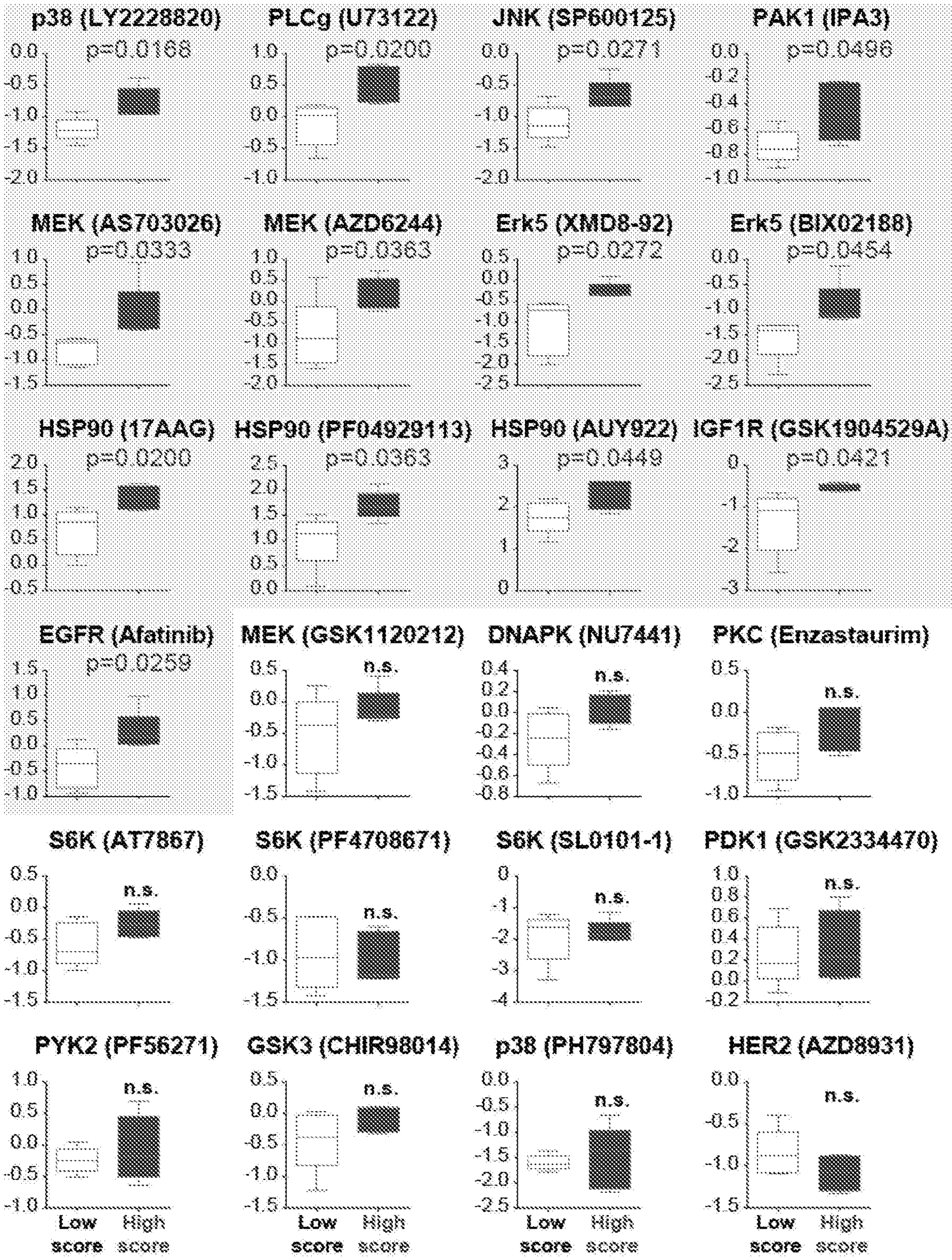

p38 (LY2228820)
p=0.0168
PLCg (U73122)
p=0.0200
JNK (SP600125)
p=0.0271
PAK1 (IPA3)
p=0.0496
MEK (AS703026)
p=0.0333
MEK (AZD6244)
p=0.0363
Erk5 (XMD8-92)
p=0.0272
Erk5 (BIX02188)
p=0.0454
HSP90 (17AAG)
p=0.0200
HSP90 (PF04929113)
p=0.0363
HSP90 (AUY922)
p=0.0449
IGF1R (GSK1904529A)
p=0.0421
EGFR (Afatinib)
p=0.0259
MEK (GSK1120212)
n.s.
DNAPK (NU7441)
n.s.
PKC (Enzastaurim)
n.s.
S6K (AT7867)
n.s.
S6K (PF4708671)
n.s.
S6K (SL0101-1)
n.s.
PDK1 (GSK2334470)
n.s.
PYK2 (PF56271)
n.s.
GSK3 (CHIR98014)
n.s.
p38 (PH797804)
n.s.
HER2 (AZD8931)
n.s.
Sensitivity (-Log10[IC50], μM)
Low score
High score

FIG. 47

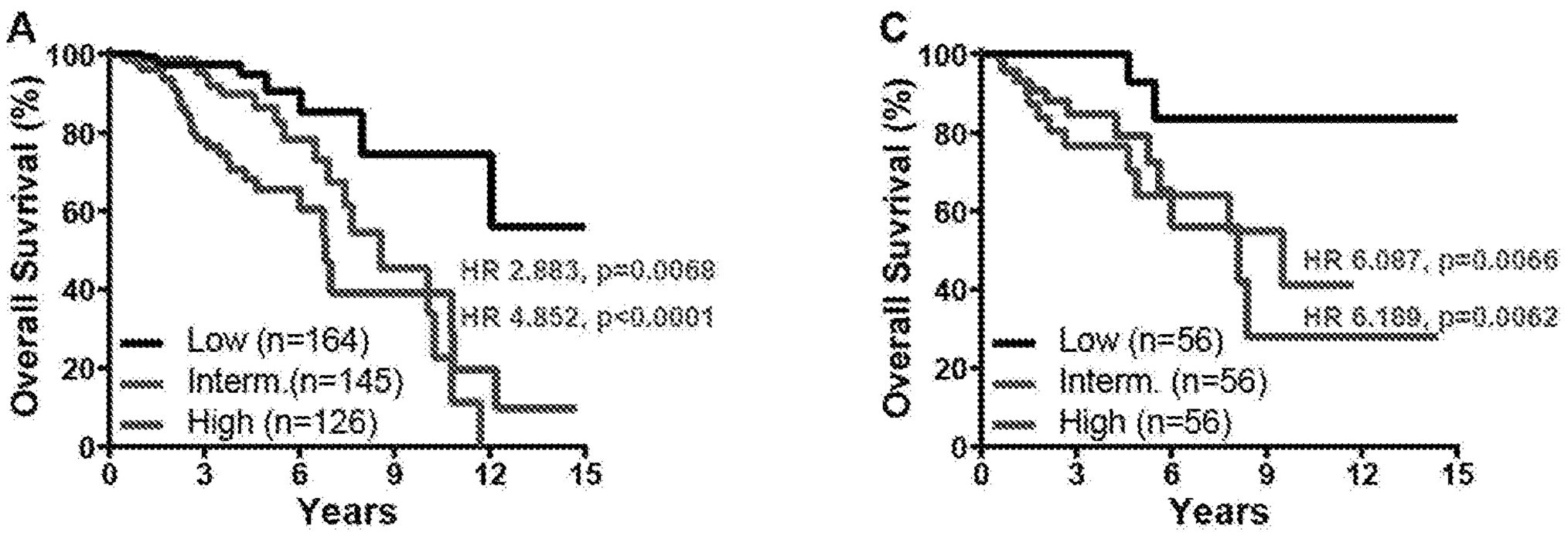

A
Overall Survival (%)
100
80
60
40
20
0
0
3
6
9
12
15
Years
Low (n=164)
Interm.(n=145)
High (n=126)
HR 2.883, p=0.0069
HR 4.852, p<0.0001
C
Overall Survival (%)
Low (n=56)
Interm. (n=56)
High (n=56)
HR 6.097, p=0.0066
HR 6.189, p=0.0062

FIG. 48

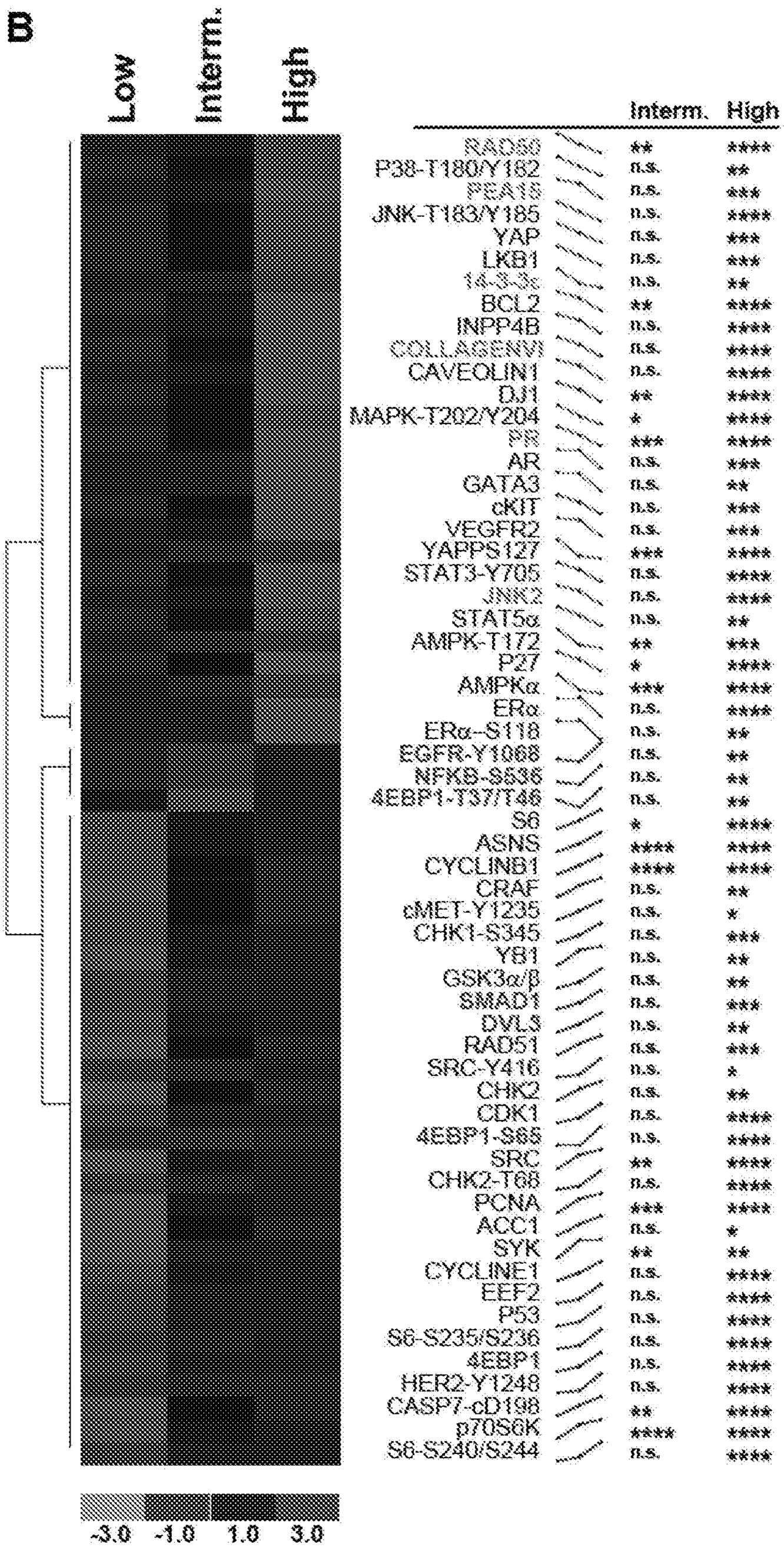

B
Low
Interm.
High
Interm. High
RAD50 ** ****
P38-T180/Y182 n.s. **
PEA15 n.s. ***
JNK-T183/Y185 n.s. ****
YAP n.s. ***
LKB1 n.s. ***
14-3-3ε n.s. **
BCL2 ** ****
INPP4B n.s. ****
COLLAGENVI n.s. ****
CAVEOLIN1 n.s. ****
DJ1 ** ****
MAPK-T202/Y204 * ****
PR *** ****
AR n.s. ***
GATA3 n.s. **
cKIT n.s. ***
VEGFR2 n.s. ***
YAPPS127 *** ****
STAT3-Y705 n.s. ****
JNK2 n.s. ****
STAT5α n.s. **
AMPK-T172 ** ***
P27 * ****
AMPKα *** ****
ERα n.s. ****
ERα-S118 n.s. **
EGFR-Y1068 n.s. **
NFKB-S536 n.s. **
4EBP1-T37/T46 n.s. **
S6 * ****
ASNS **** ****
CYCLINB1 **** ****
CRAF n.s. **
cMET-Y1235 n.s. *
CHK1-S345 n.s. ***
YB1 n.s. **
GSK3α/β n.s. **
SMAD1 n.s. ***
DVL3 n.s. **
RAD51 n.s. ***
SRC-Y416 n.s. *
CHK2 n.s. **
CDK1 n.s. ****
4EBP1-S65 n.s. ****
SRC ** ****
CHK2-T68 n.s. ****
PCNA *** ****
ACC1 n.s. *
SYK ** **
CYCLINE1 n.s. ****
EEF2 n.s. ****
P53 n.s. ****
S6-S235/S236 n.s. ****
4EBP1 n.s. ****
HER2-Y1248 n.s. ****
CASP7-cD198 ** ****
p70S6K **** ****
S6-S240/S244 n.s. ****
-3.0 -1.0 1.0 3.0

FIG. 48 cont'd

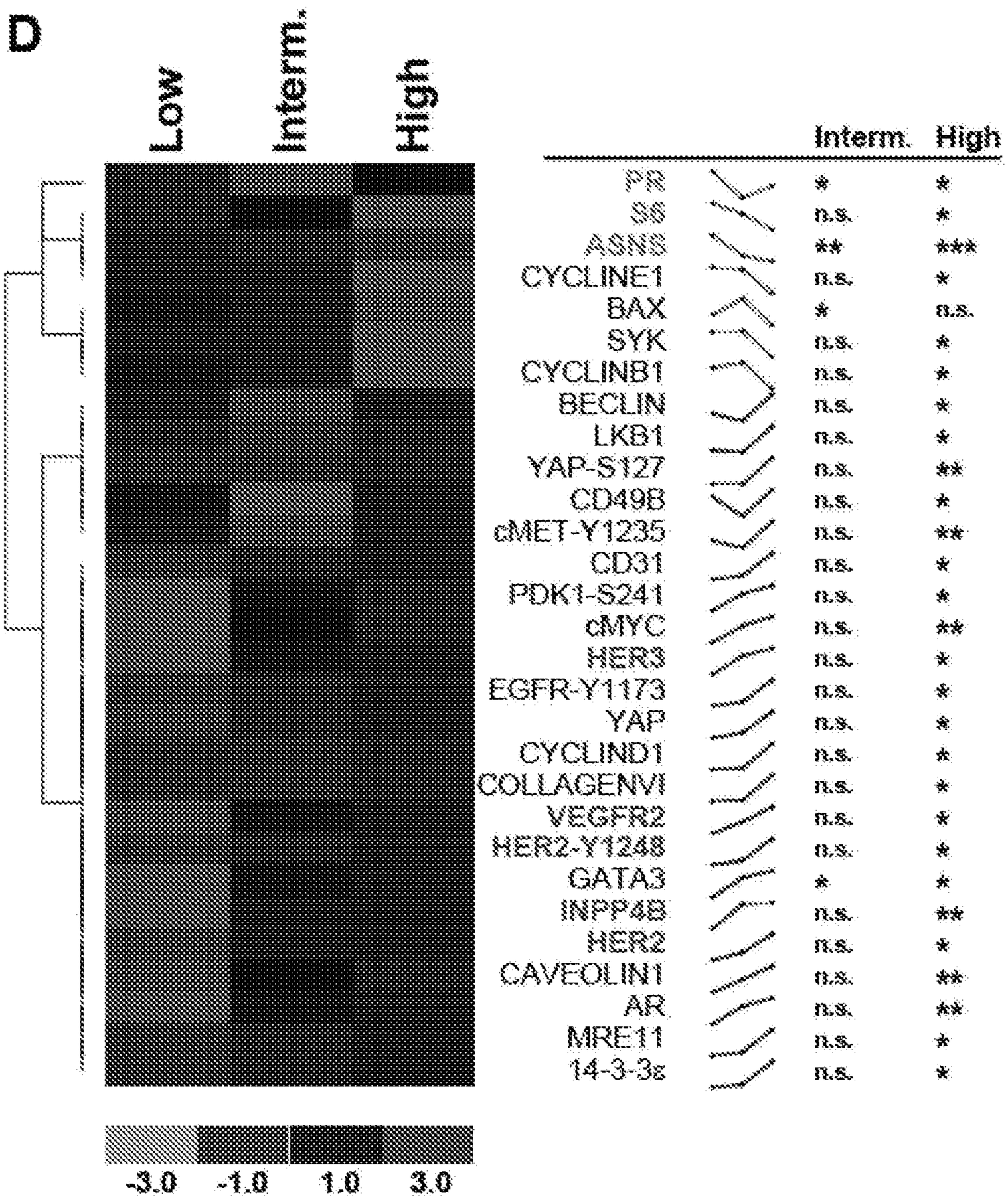

D
Low
Interm.
High
-3.0
-1.0
1.0
3.0
Interm. High
PR * *
S6 n.s. *
ASNS ** ***
CYCLINE1 n.s. *
BAX * n.s.
SYK n.s. *
CYCLINB1 n.s. *
BECLIN n.s. *
LKB1 n.s. *
YAP-S127 n.s. **
CD49B n.s. *
cMET-Y1235 n.s. **
CD31 n.s. *
PDK1-S241 n.s. *
cMYC n.s. **
HER3 n.s. *
EGFR-Y1173 n.s. *
YAP n.s. *
CYCLIND1 n.s. *
COLLAGENVI n.s. *
VEGFR2 n.s. *
HER2-Y1248 n.s. *
GATA3 * *
INPP4B n.s. **
HER2 n.s. *
CAVEOLIN1 n.s. **
AR n.s. **
MRE11 n.s. *
14-3-3ε n.s. *

FIG. 48 cont'd

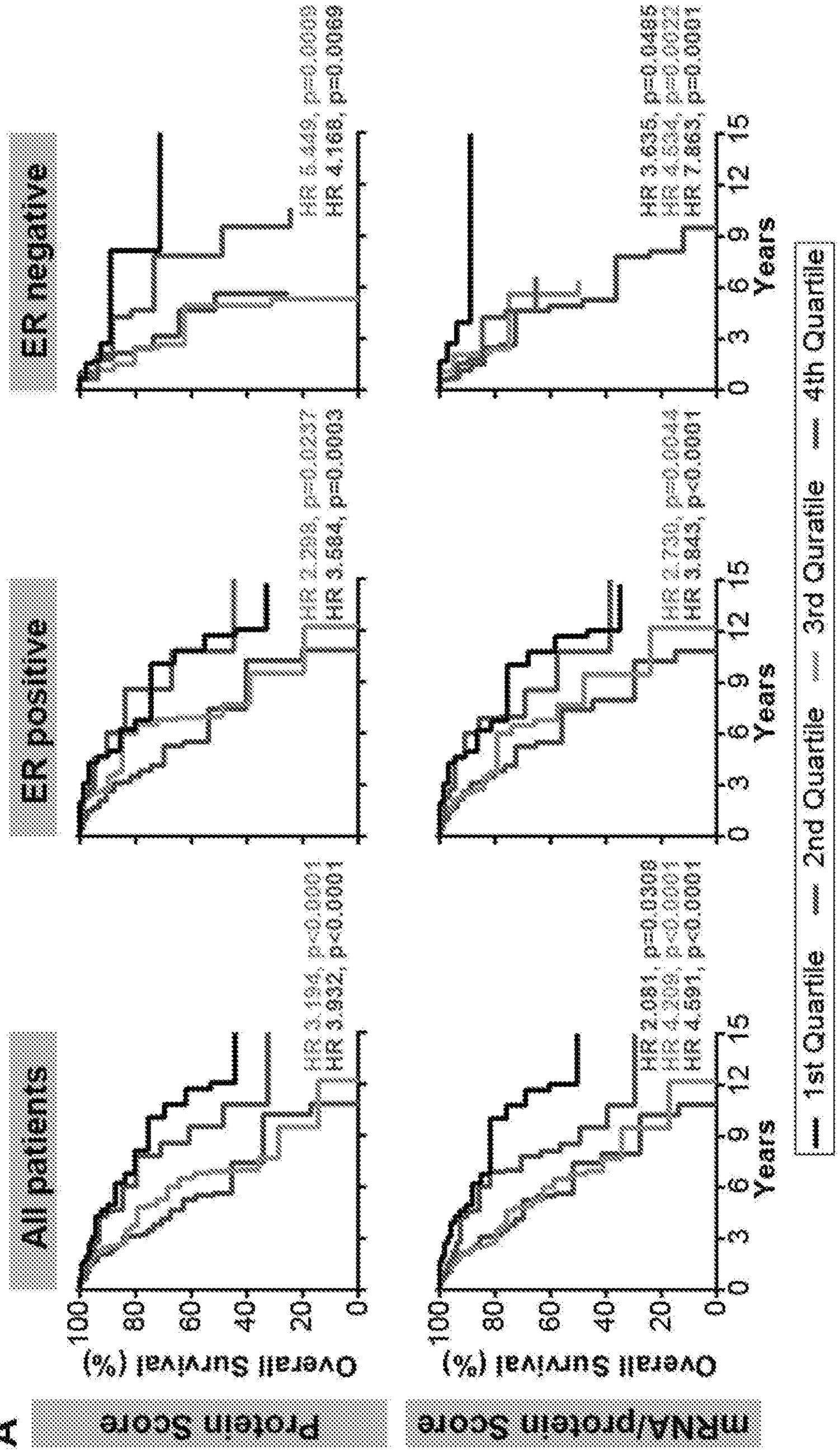

mRNA/protein Score
Protein Score
ER negative
ER positive
All patients
HR 3.635, p=0.0485
HR 4.634, p=0.0022
HR 7.863, p=0.0001
HR 2.730, p=0.0044
HR 3.843, p<0.0001
HR 2.081, p=0.0308
HR 4.208, p<0.0001
HR 4.591, p<0.0001
HR 3.449, p=0.0069
HR 2.298, p=0.0237
HR 3.584, p=0.0033
HR 3.194, p<0.0001
HR 3.932, p<0.0001
Overall Survival (%)
0
20
40
60
80
100
0 3 6 9 12 15
Years
1st Quartile
2nd Quartile
3rd Quratile
4th Quartile

FIG. 49

CA 02941769 2016-09-07

B

Parameter	Univariant Cox-proportional hazard model HR	95% CI	P	Multivariant Cox-proportional hazard model HR	95% CI	P
Combined iBCR mRNA/Protein Score (Q1, Q2, Q3, Q4)	1.75	1.45 - 2.12	<0.0001	3.46	1.51 - 7.92	0.0035
Stage (I, IIA, IIB, IIIA, IIIB, IIIC)	1.49	1.15 - 1.92	0.0023	3.81	1.30 - 11.17	0.0153
Tumor size (T1, T2, T3)	1.12	0.81 - 1.55	0.5094	3.88	0.72 - 20.53	0.1157
HER2 (pos vs. neg)	1.82	1.01 - 3.28	0.0466	6.24	0.43 - 9.34	0.1816
Radiotherapy (yes vs. no)	0.38	0.19 - 0.76	0.0063	0.27	0.04 - 1.98	0.1991
PAM50 subtypes	1.08	0.87 - 1.34	0.5048	0.69	0.36 - 1.34	0.2804
Lymph node (pos vs. neg)	2.14	1.35 - 3.39	0.0012	3.32	0.33 - 3.38	0.3127
ER (pos vs. neg)	0.76	0.48 - 1.21	0.2493	0.40	0.22 - 7.49	0.5439
Age at Dx (<50 vs. >50)	0.60	0.37 - 0.95	0.0295	0.28	0.04 - 2.32	0.5775
Surgery (lump, mastec, mod. rad. mastec., Other)	1.28	1.04 - 1.58	0.0191	0.75	0.27 - 2.12	0.5901
Molecular therapy (no vs. yes)	6.17	2.63 - 14.43	<0.0001	1.62	0.17 - 15.91	0.6805
Histology (ILC vs. IDC)	0.53	0.23 - 1.22	0.1399	0.54	0.27 - 10.69	0.6847
PR (pos vs. neg)	0.76	0.49 - 1.16	0.2035	0.72	0.414 - 1.25	0.8219
Menopause (post vs. pre)	2.04	1.03 - 4.05	0.0419	1.15	0.15 - 8.65	0.9483

FIG. 49 cont'd

SUBSTITUTE SHEET (RULE 26)

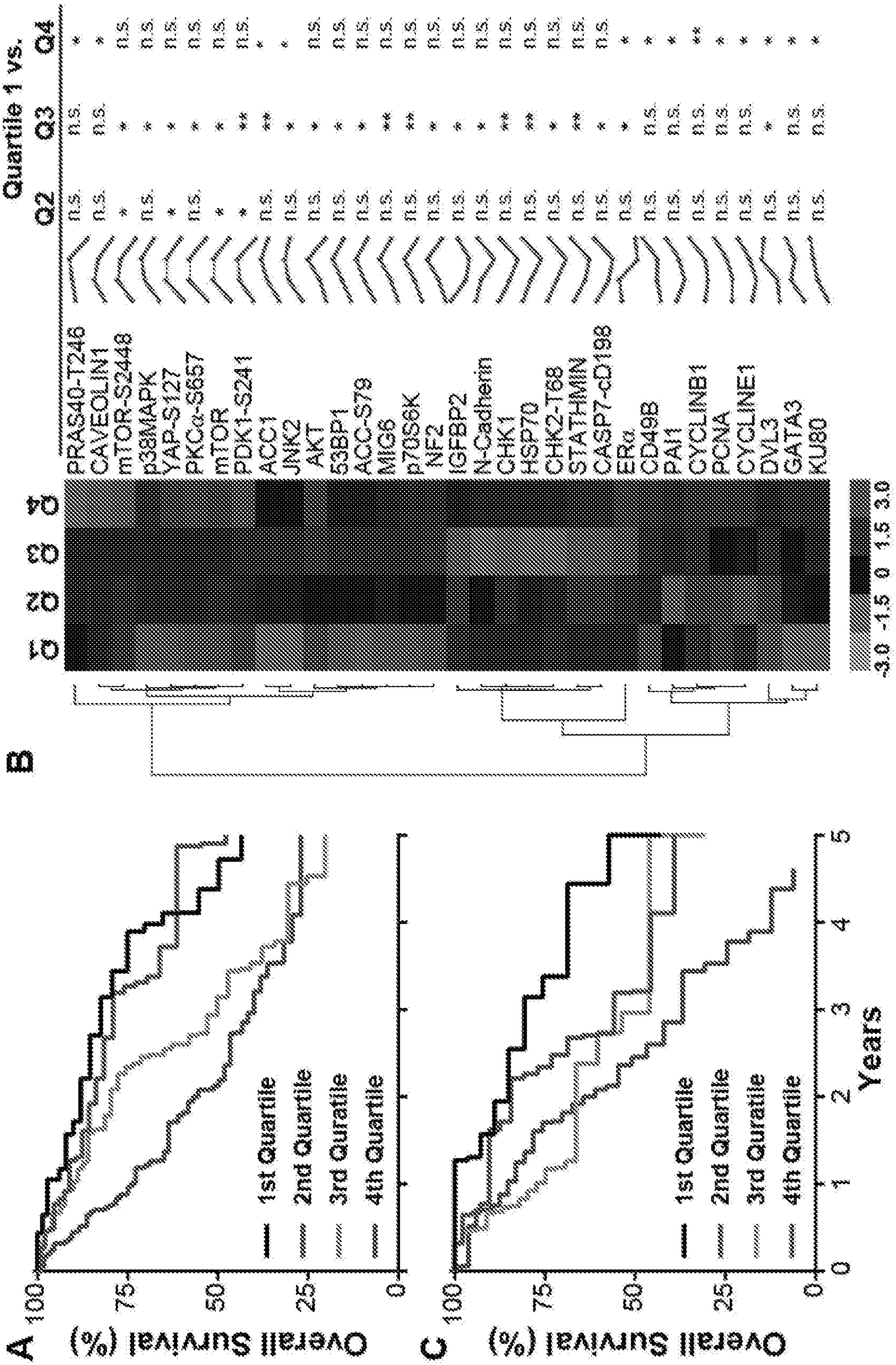

Quartile 1 vs.
Q2
Q3
Q4
n.s.
*
**
PRAS40-T246
CAVEOLIN1
mTOR-S2448
p38MAPK
YAP-S127
PKCα-S657
mTOR
PDK1-S241
ACC1
JNK2
AKT
53BP1
ACC-S79
MIG6
p70S6K
NF2
IGFBP2
N-Cadherin
CHK1
HSP70
CHK2-T68
STATHMIN
CASP7-cD198
ERα
CD49B
PAI1
CYCLINB1
PCNA
CYCLINE1
DVL3
GATA3
KU80
Q1
Q2
Q3
Q4
-3.0 -1.5 0 1.5 3.0
B
1st Quartile
2nd Quartile
3rd Quratile
4th Quartile
Overall Survival (%)
100
75
50
25
0
A
C
0
1
2
3
4
5
Years

FIG. 50

E

Covariate	HR	95% CI	p
Combined iSCR mRNA/Protein (Q1-Q4)	2.08	1.24 - 3.50	0.0056
Age (<60, 60-70, >70)	2.25	1.14 - 4.43	0.0198
Nodal status (N1 vs. N0)	1.73	1.40 - 2.10	0.0254
Tumor size (T1, T2, T3, T4)	1.63	0.83 - 3.19	0.1553
Residual tumor (yes vs. no)	3.70	0.46 - 30.06	0.2229
EGFR (mut vs. WT)	2.43	0.46 - 12.87	0.2995
Stage (Stage I, II, III)	0.58	0.11 - 3.03	0.5187
Targeted therapy (yes vs. no)	1.96	0.25 - 14.29	0.5218
Gender (F or M)	1.32	0.35 - 4.96	0.6797
KRAS (mut vs. WT)	1.33	0.25 - 7.01	0.7403
PIK3CA (mut vs. WT)	0.79	0.09 - 7.07	0.8321
ERBB4 (mut vs. WT)	0.84	0.11 - 6.72	0.8714
mRNA Subtype (Bronchioid, Magnoid, Squamoid)	0.98	0.54 - 1.78	0.9528

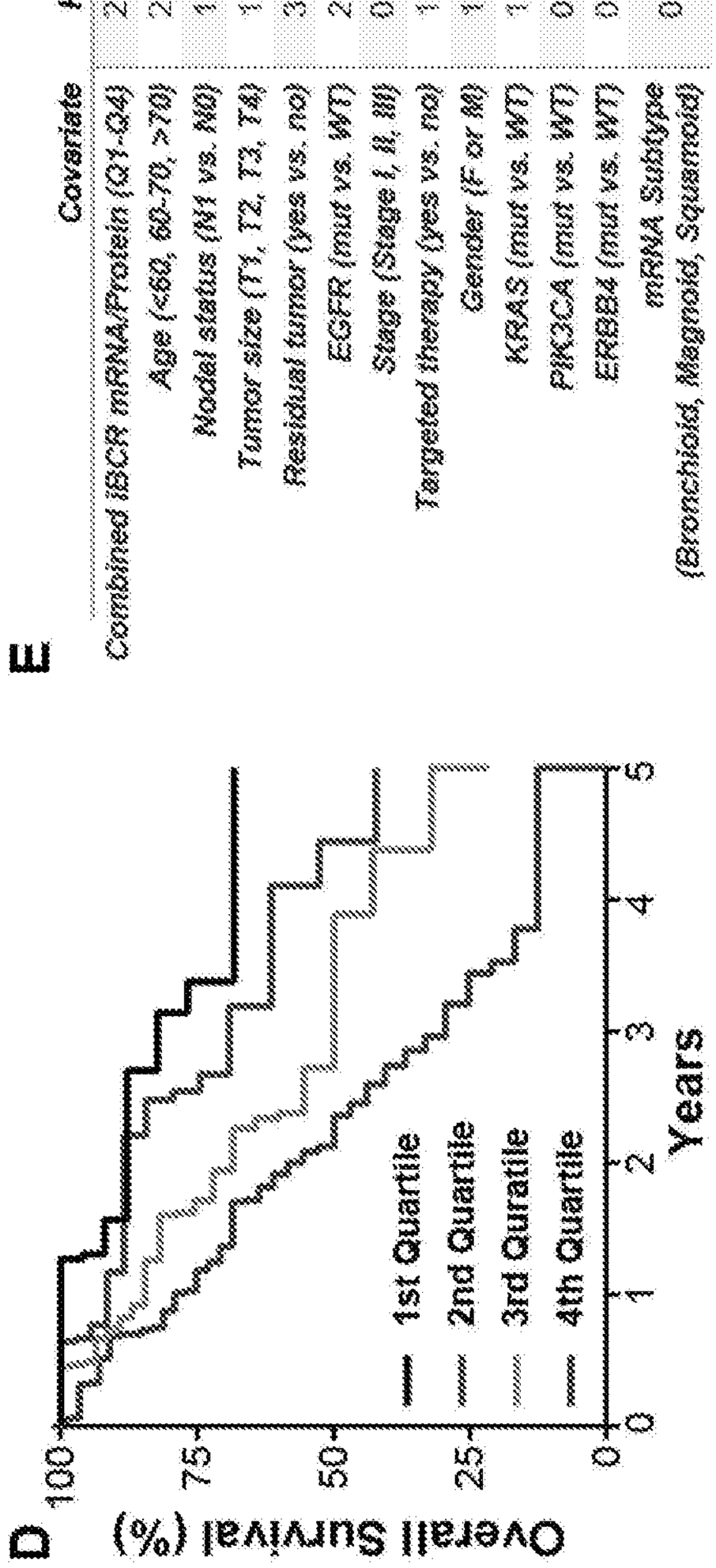


FIG. 50 cont'd

Kidney renal clear cell carcinoma (KIRC)

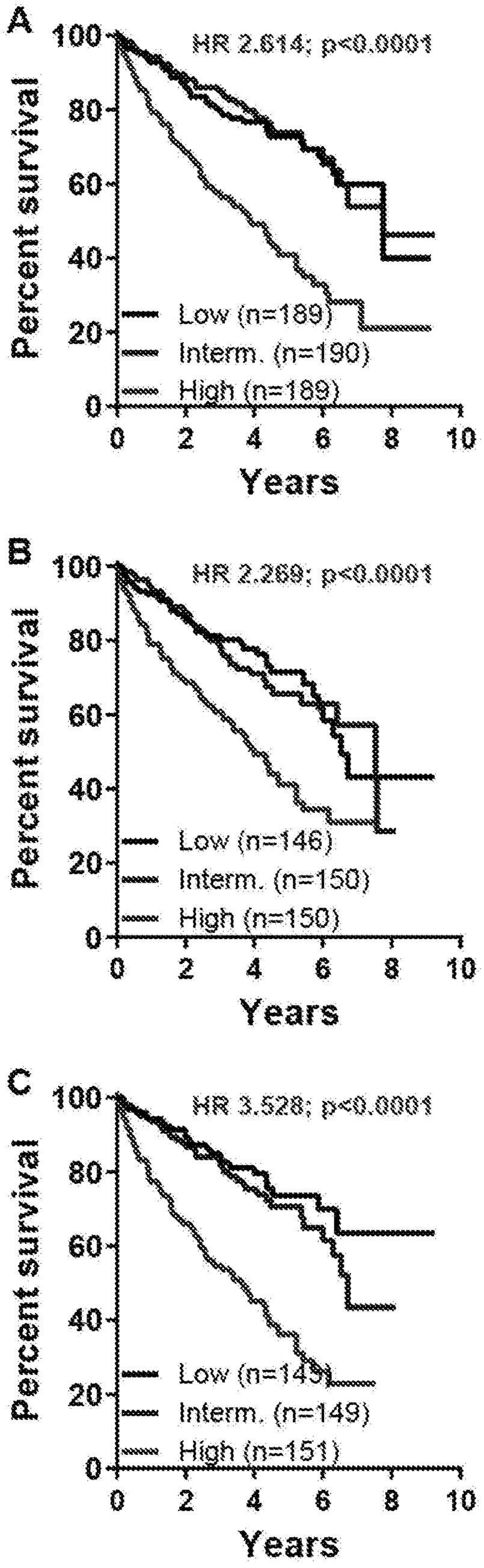


FIG. 51

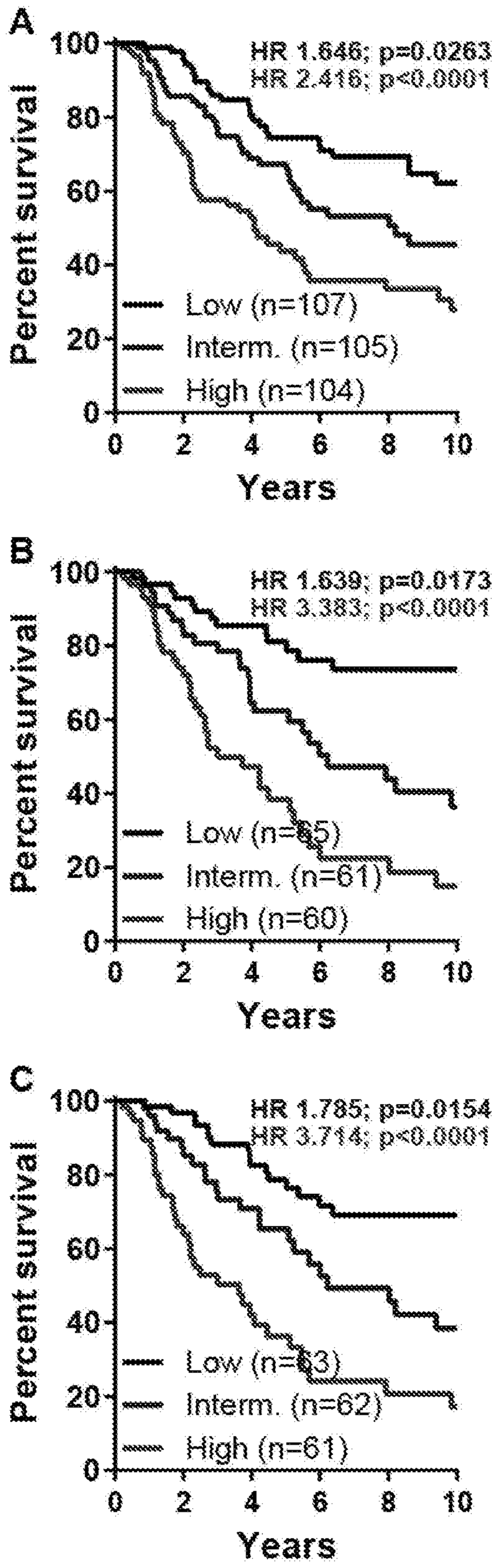

Skin cutaneous melanoma (SKCM)
A
Percent survival
100
80
60
40
20
0
HR 1.646; p=0.0263
HR 2.416; p<0.0001
Low (n=107)
Interm. (n=105)
High (n=104)
0 2 4 6 8 10
Years
B
Percent survival
HR 1.639; p=0.0173
HR 3.383; p<0.0001
Low (n=65)
Interm. (n=61)
High (n=60)
Years
C
Percent survival
HR 1.785; p=0.0154
HR 3.714; p<0.0001
Low (n=63)
Interm. (n=62)
High (n=61)
Years

FIG. 51 cont'd

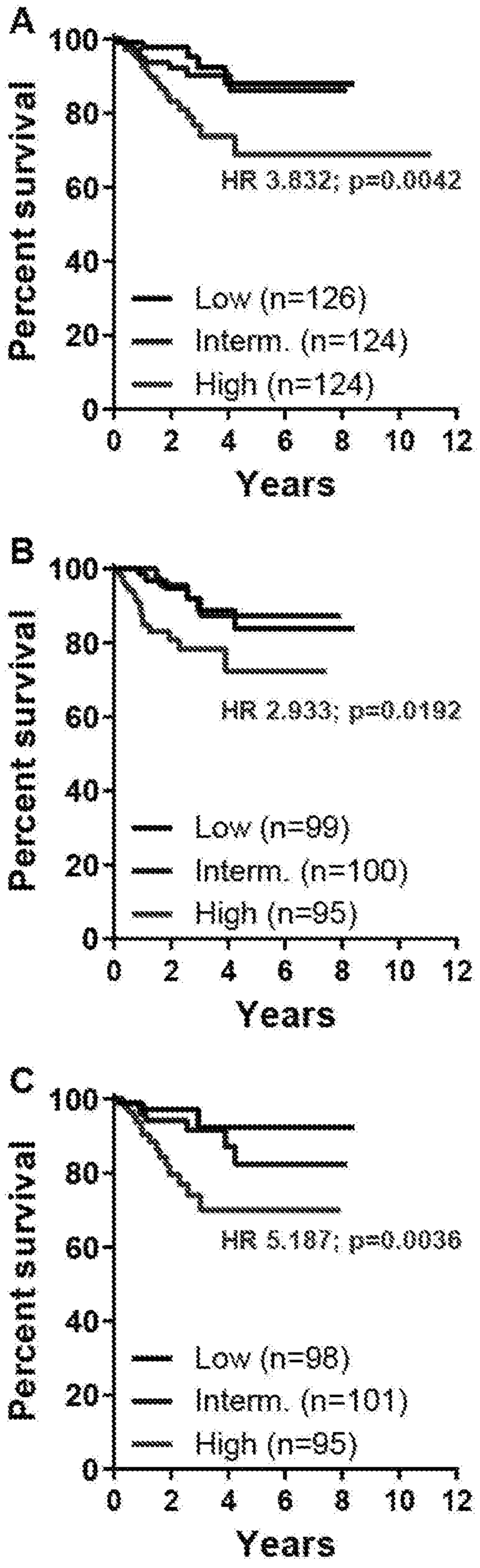

Uterine corpus endometrioid carcinoma (UCEC)
A
Percent survival
100
80
60
40
20
0
HR 3.832; p=0.0042
Low (n=126)
Interm. (n=124)
High (n=124)
0 2 4 6 8 10 12
Years
B
Percent survival
100
80
60
40
20
0
HR 2.933; p=0.0192
Low (n=99)
Interm. (n=100)
High (n=95)
0 2 4 6 8 10 12
Years
C
Percent survival
100
80
60
40
20
0
HR 5.187; p=0.0036
Low (n=98)
Interm. (n=101)
High (n=95)
0 2 4 6 8 10 12
Years

FIG. 51 cont'd

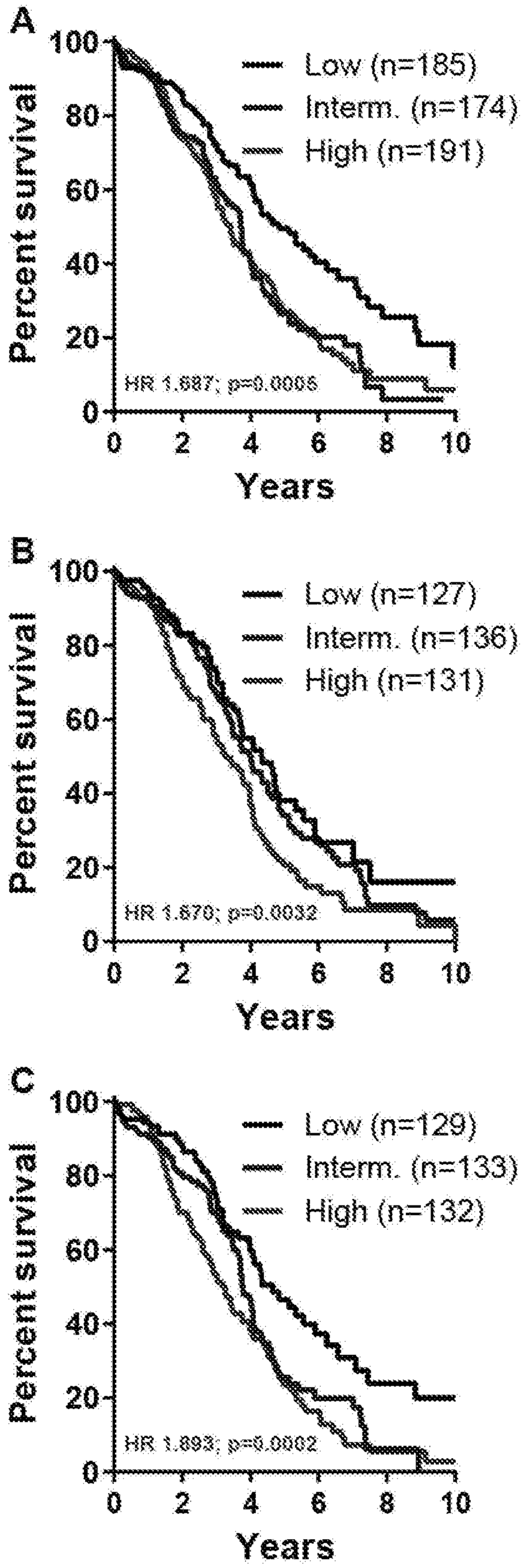

Ovarian adenocarcinoma (OVAC)
A
Percent survival
100
80
60
40
20
0
Low (n=185)
Interm. (n=174)
High (n=191)
HR 1.687; p=0.0005
0 2 4 6 8 10
Years
B
Percent survival
Low (n=127)
Interm. (n=136)
High (n=131)
HR 1.670; p=0.0032
Years
C
Percent survival
Low (n=129)
Interm. (n=133)
High (n=132)
HR 1.893; p=0.0002
Years

FIG. 52

Head & Neck squamous cell carcinoma (HNSC)

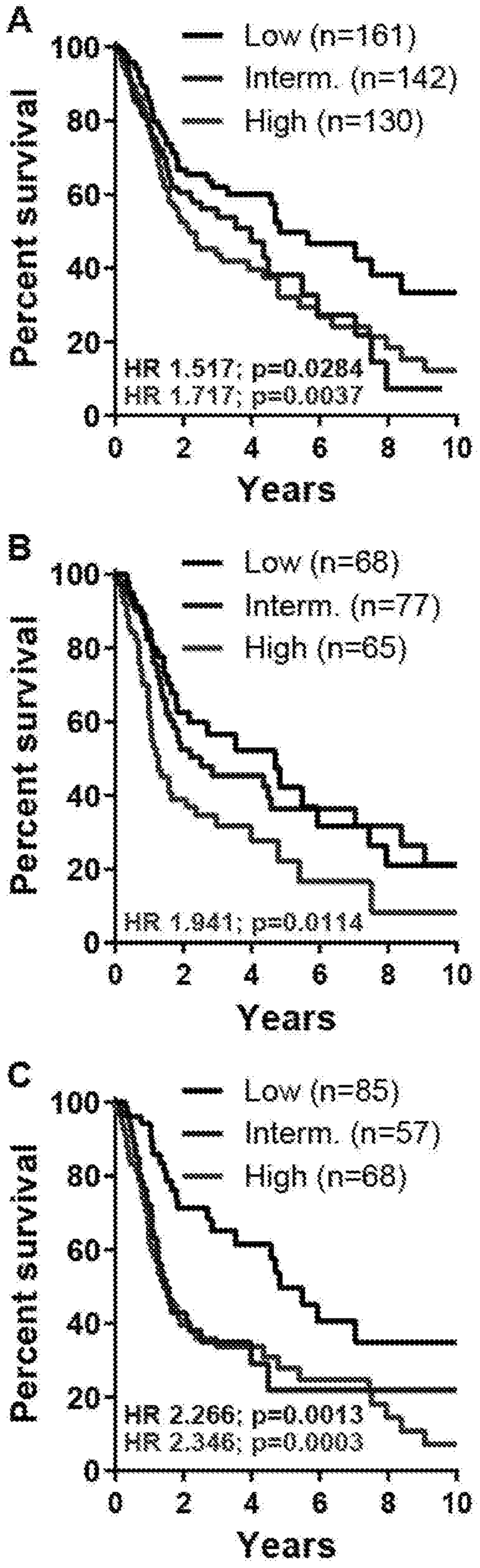


FIG. 52 cont'd

Colon/Rectal Adenocarcinoma (COREAD)

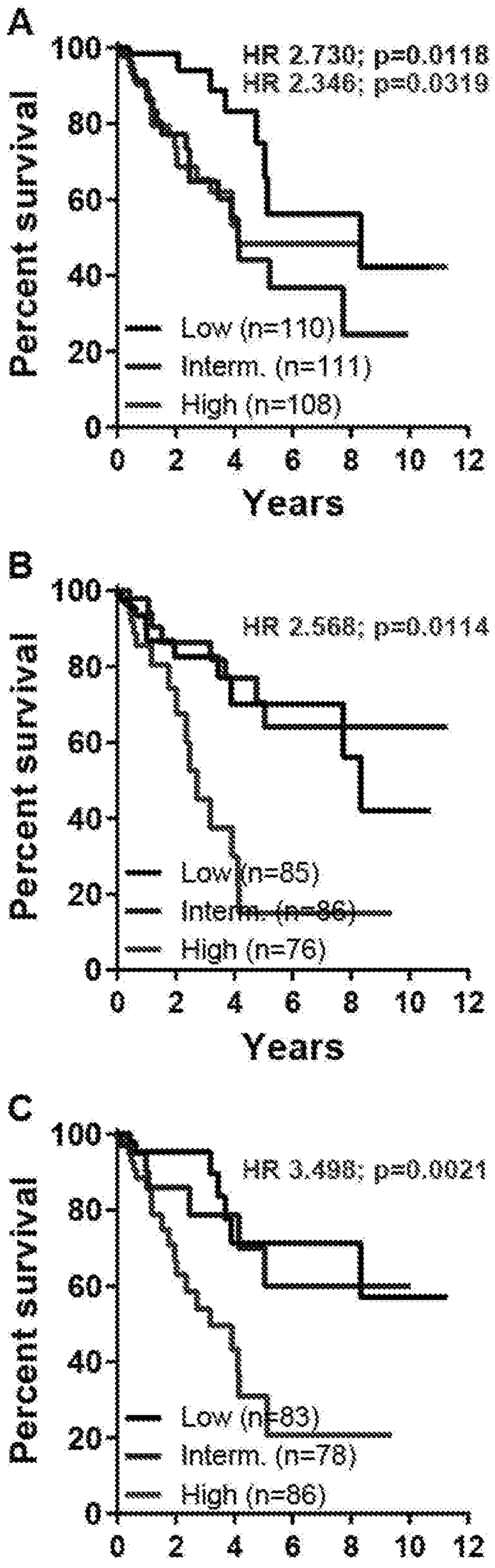


FIG. 52 cont'd

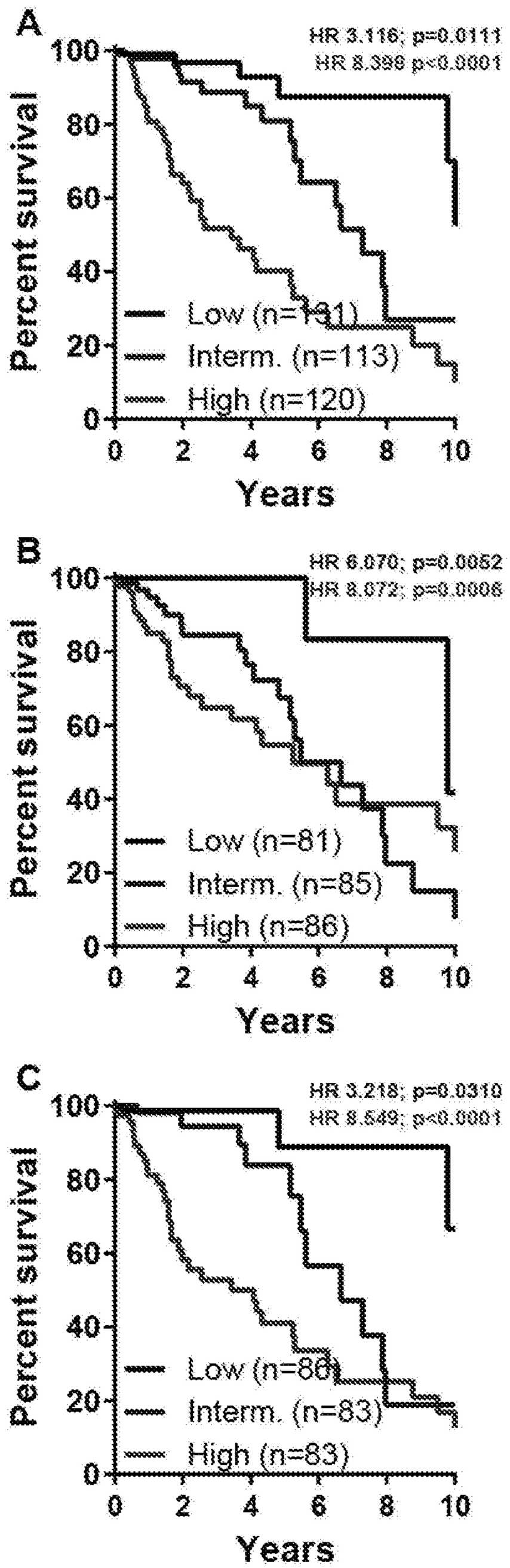

Lower Grade Glioma (LGG)
A
HR 3.116; p=0.0111
HR 8.398 p<0.0001
Percent survival
100
80
60
40
20
0
Low (n=131)
Interm. (n=113)
High (n=120)
0 2 4 6 8 10
Years
B
HR 6.070; p=0.0052
HR 8.072; p=0.0006
Percent survival
Low (n=81)
Interm. (n=85)
High (n=86)
Years
C
HR 3.218; p=0.0310
HR 8.549; p<0.0001
Percent survival
Low (n=86)
Interm. (n=83)
High (n=83)
Years

FIG. 53

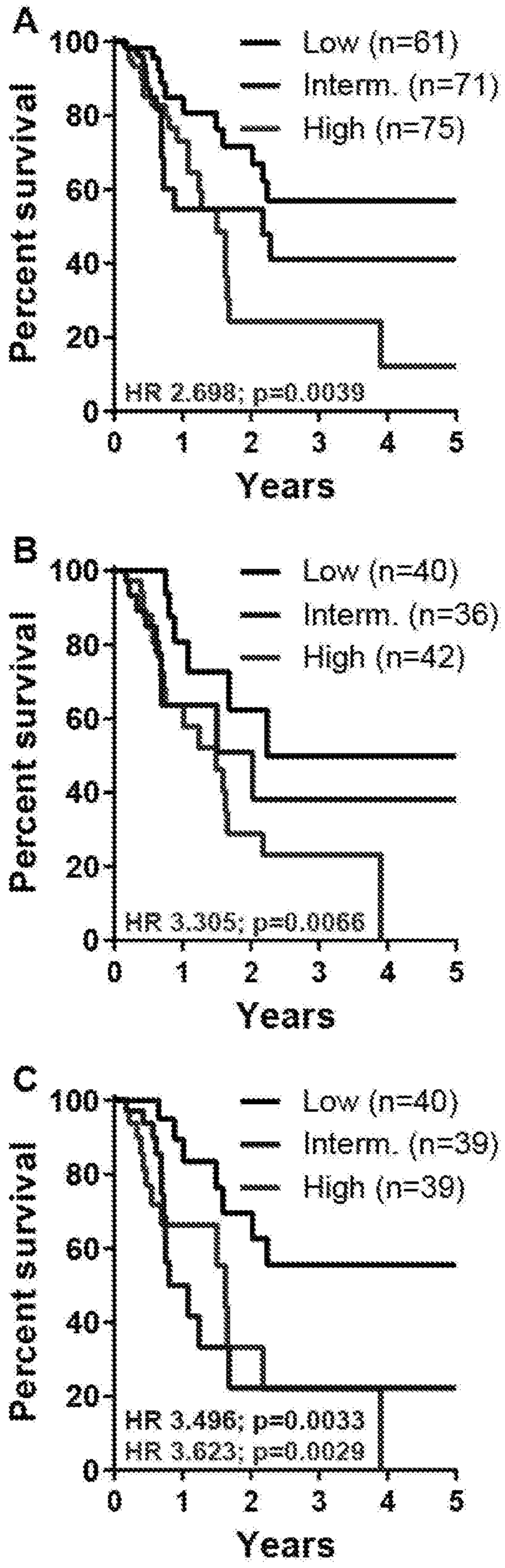

Bladder urothelial carcinoma (BLCA)
A
Percent survival
100
80
60
40
20
0
Low (n=61)
Interm. (n=71)
High (n=75)
HR 2.698; p=0.0039
0 1 2 3 4 5
Years
B
Percent survival
Low (n=40)
Interm. (n=36)
High (n=42)
HR 3.305; p=0.0066
Years
C
Percent survival
Low (n=40)
Interm. (n=39)
High (n=39)
HR 3.496; p=0.0033
HR 3.623; p=0.0029
Years

FIG. 53 cont'd

Lung squamous cell carcinoma (LUSC)

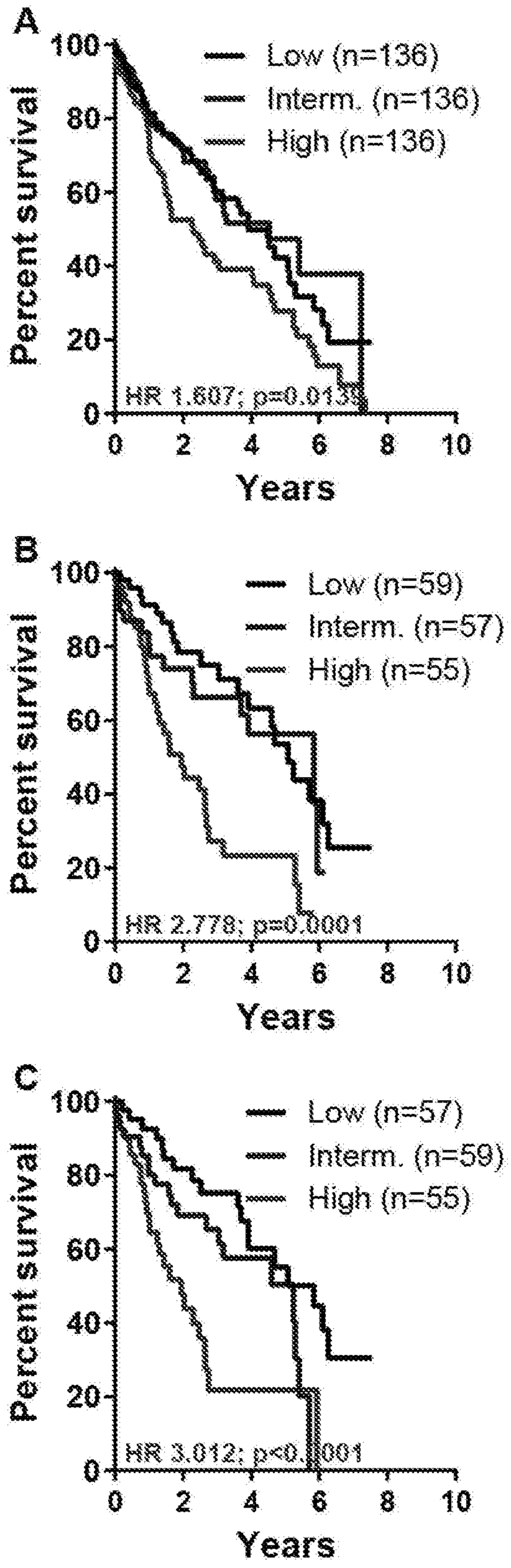


FIG. 53 cont'd

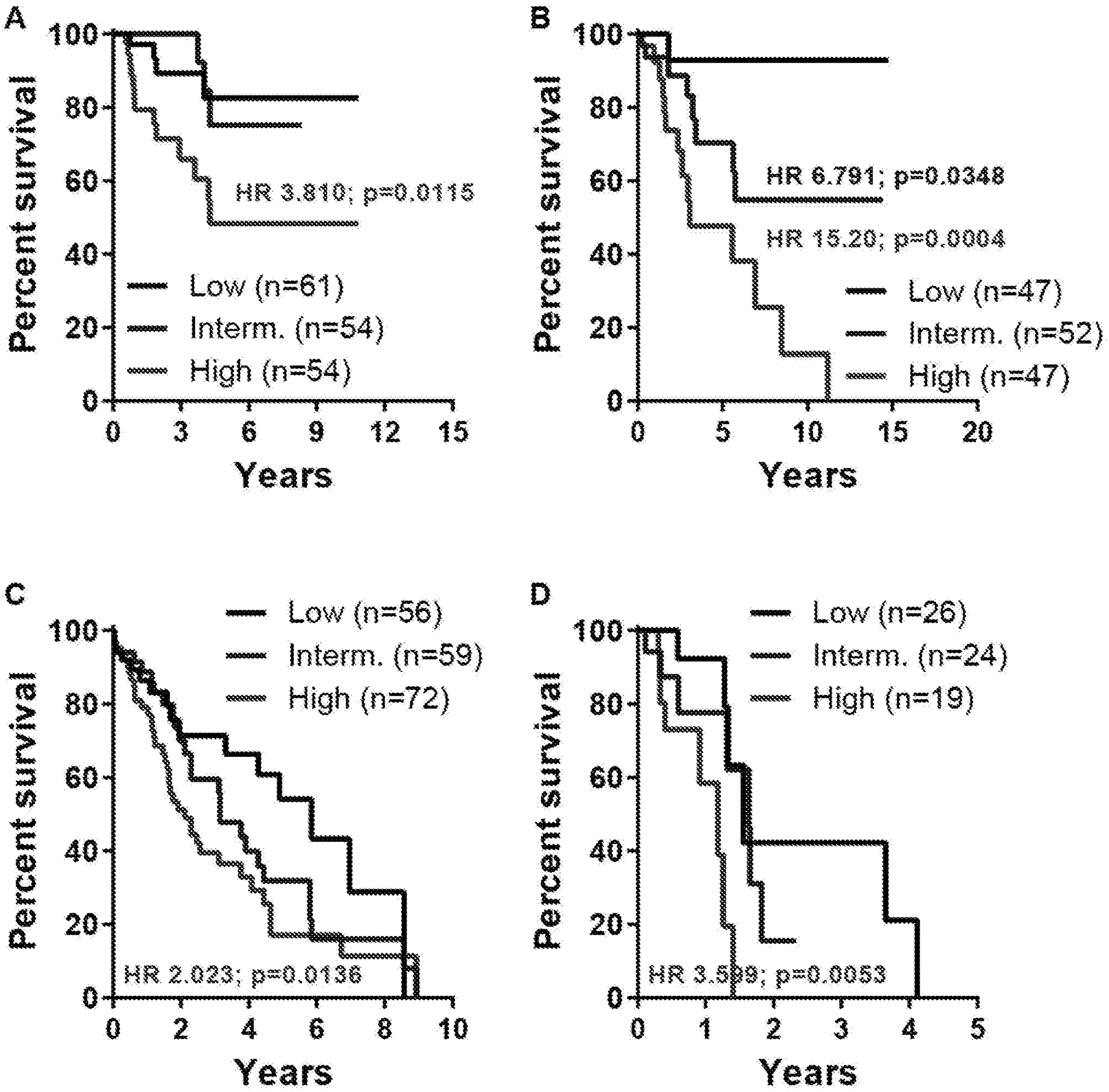

A
Percent survival
100
80
60
40
20
0
HR 3.810; p=0.0115
Low (n=61)
Interm. (n=54)
High (n=54)
0 3 6 9 12 15
Years
B
Percent survival
100
80
60
40
20
0
HR 6.791; p=0.0348
HR 15.20; p=0.0004
Low (n=47)
Interm. (n=52)
High (n=47)
0 5 10 15 20
Years
C
Percent survival
100
80
60
40
20
0
Low (n=56)
Interm. (n=59)
High (n=72)
HR 2.023; p=0.0136
0 2 4 6 8 10
Years
D
Percent survival
100
80
60
40
20
0
Low (n=26)
Interm. (n=24)
High (n=19)
HR 3.599; p=0.0053
0 1 2 3 4 5
Years

FIG. 54

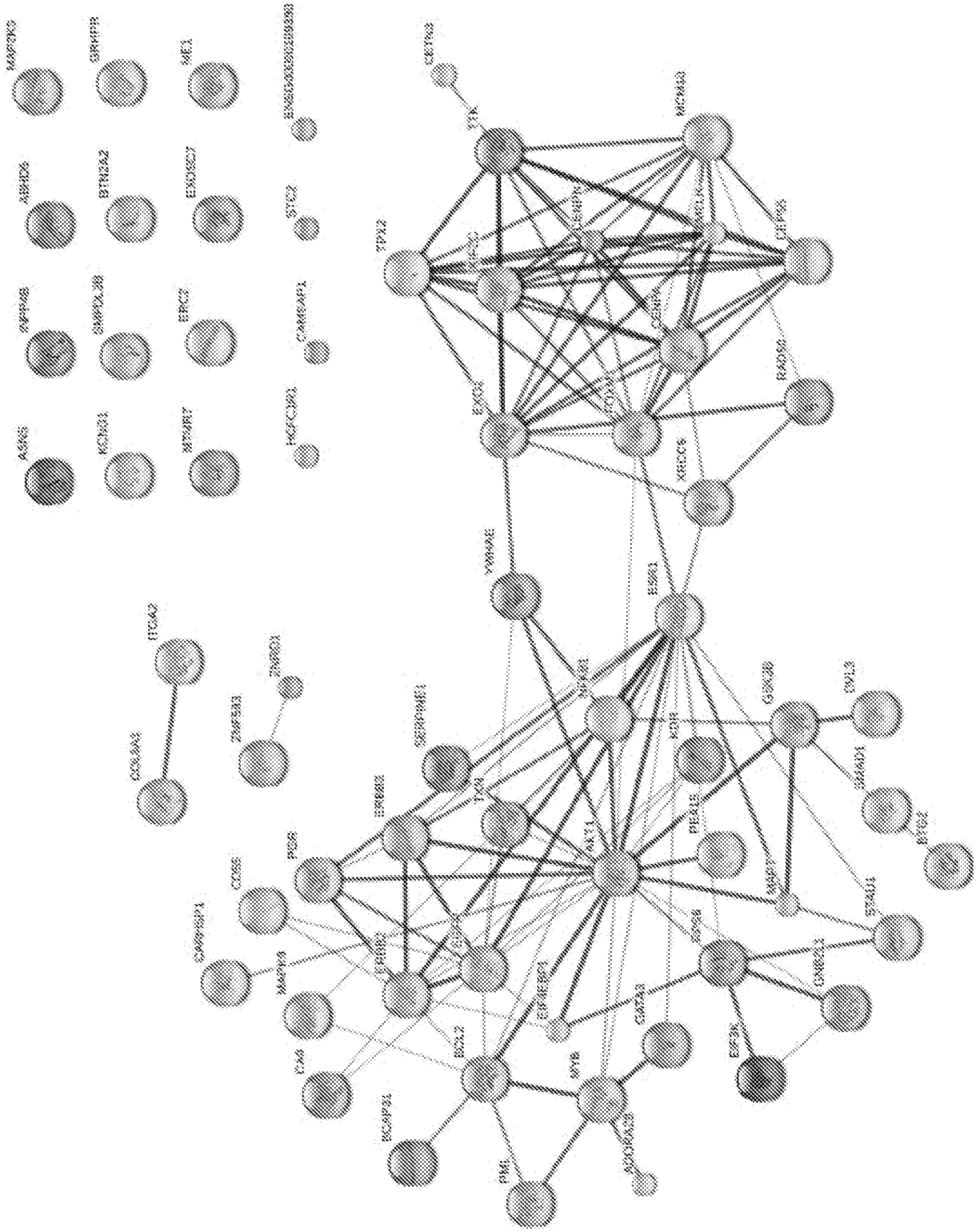


FIG. 55

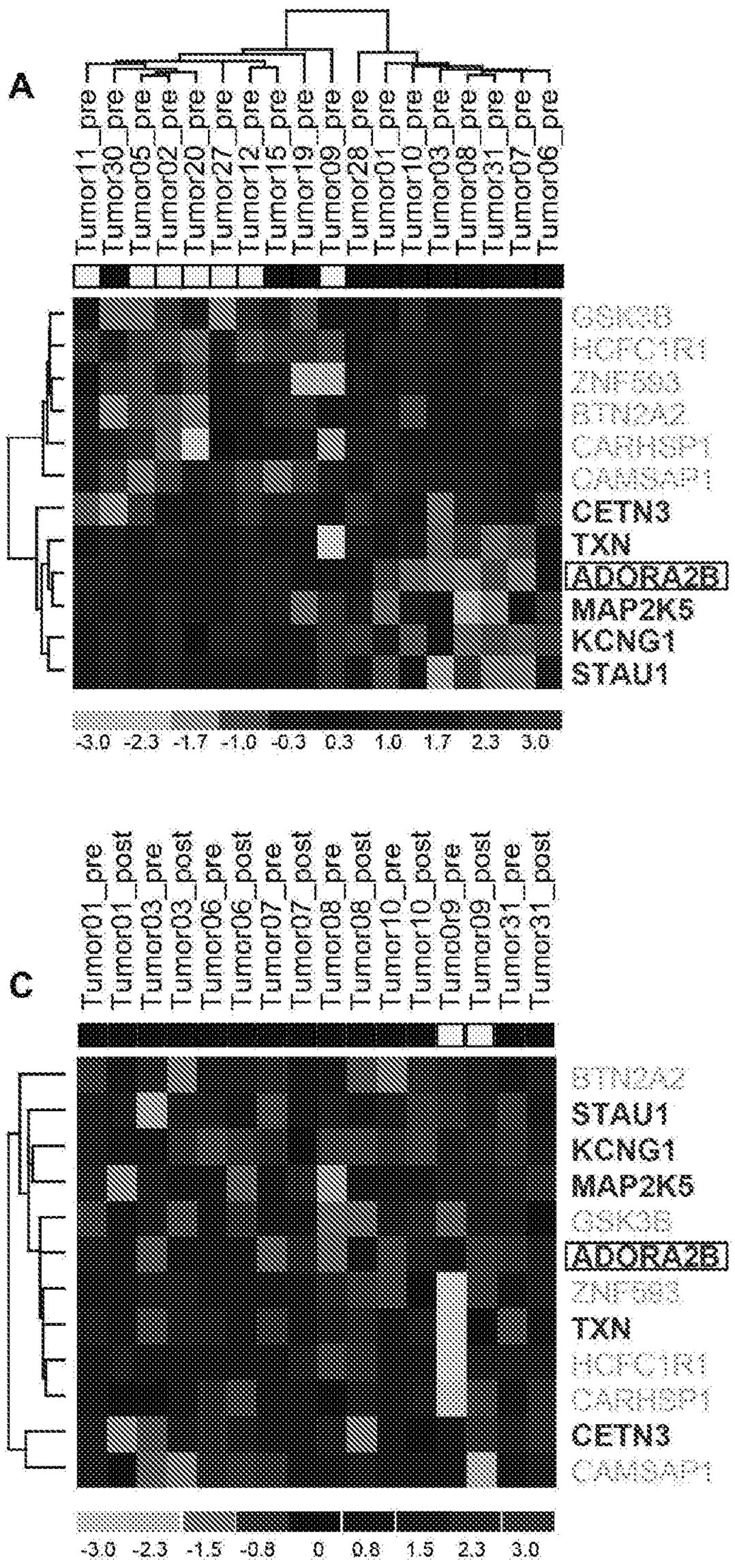

A
Tumor11_pre
Tumor30_pre
Tumor05_pre
Tumor02_pre
Tumor20_pre
Tumor27_pre
Tumor12_pre
Tumor15_pre
Tumor19_pre
Tumor09_pre
Tumor28_pre
Tumor01_pre
Tumor10_pre
Tumor03_pre
Tumor08_pre
Tumor31_pre
Tumor07_pre
Tumor06_pre
GSK3B
HCFC1R1
ZNF593
BTN2A2
CARHSP1
CAMSAP1
CETN3
TXN
ADORA2B
MAP2K5
KCNG1
STAU1
-3.0 -2.3 -1.7 -1.0 -0.3 0.3 1.0 1.7 2.3 3.0
B
Low (n=9)
High (n=9)
Percent survival
100
80
60
40
20
0
HR 8.119; p=0.0011
0 12 24 36 48
Months
C
BTN2A2
STAU1
KCNG1
MAP2K5
GSK3B
ADORA2B
ZNF593
TXN
HCFC1R1
CARHSP1
CETN3
CAMSAP1
-3.0 -2.3 -1.5 -0.8 0 0.8 1.5 2.3 3.0
D
ADORA2B
mRNA expression (log2)
7.5
7.0
6.5
6.0
5.5
t-test p-value =0.0339
Pre
Post

FIG. 56

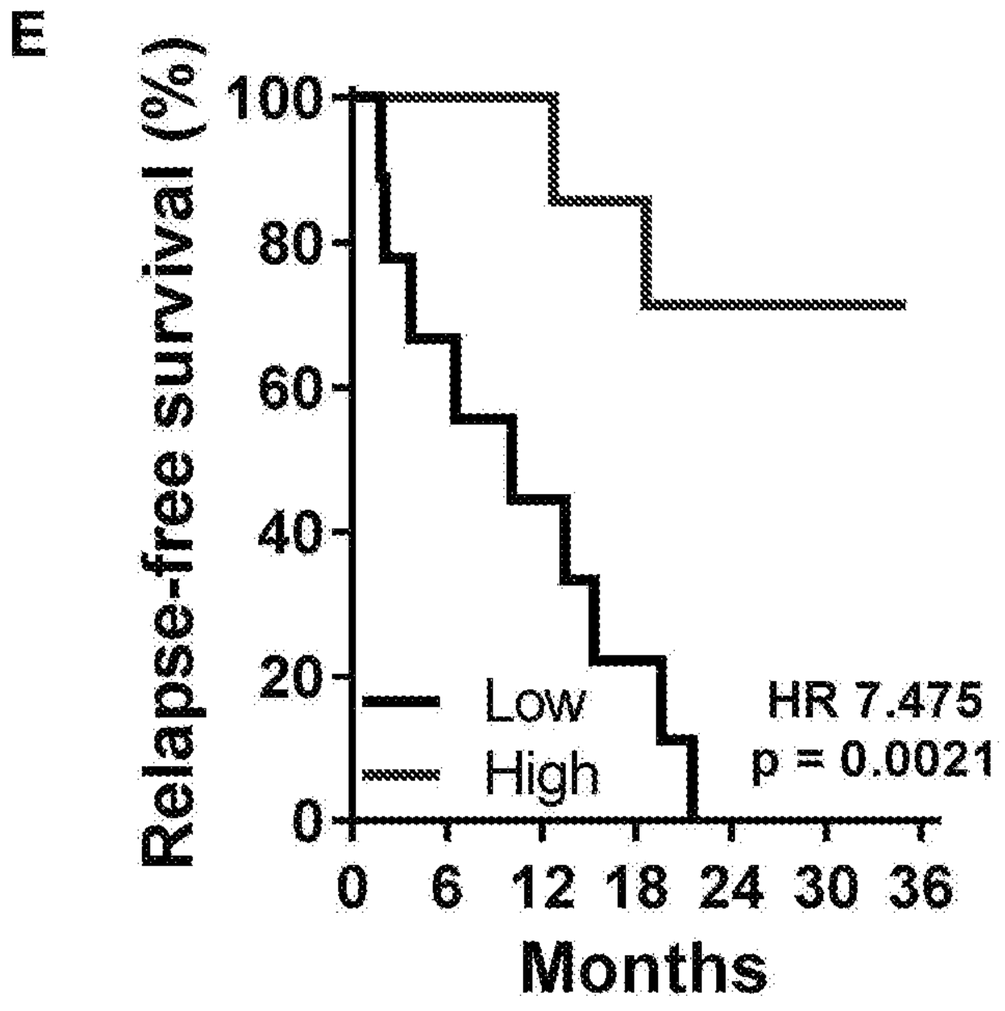

E
Relapse-free survival (%)
100
80
60
40
20
0
Low
High
HR 7.475
p = 0.0021
0 6 12 18 24 30 36
Months

FIG. 56 cont'd

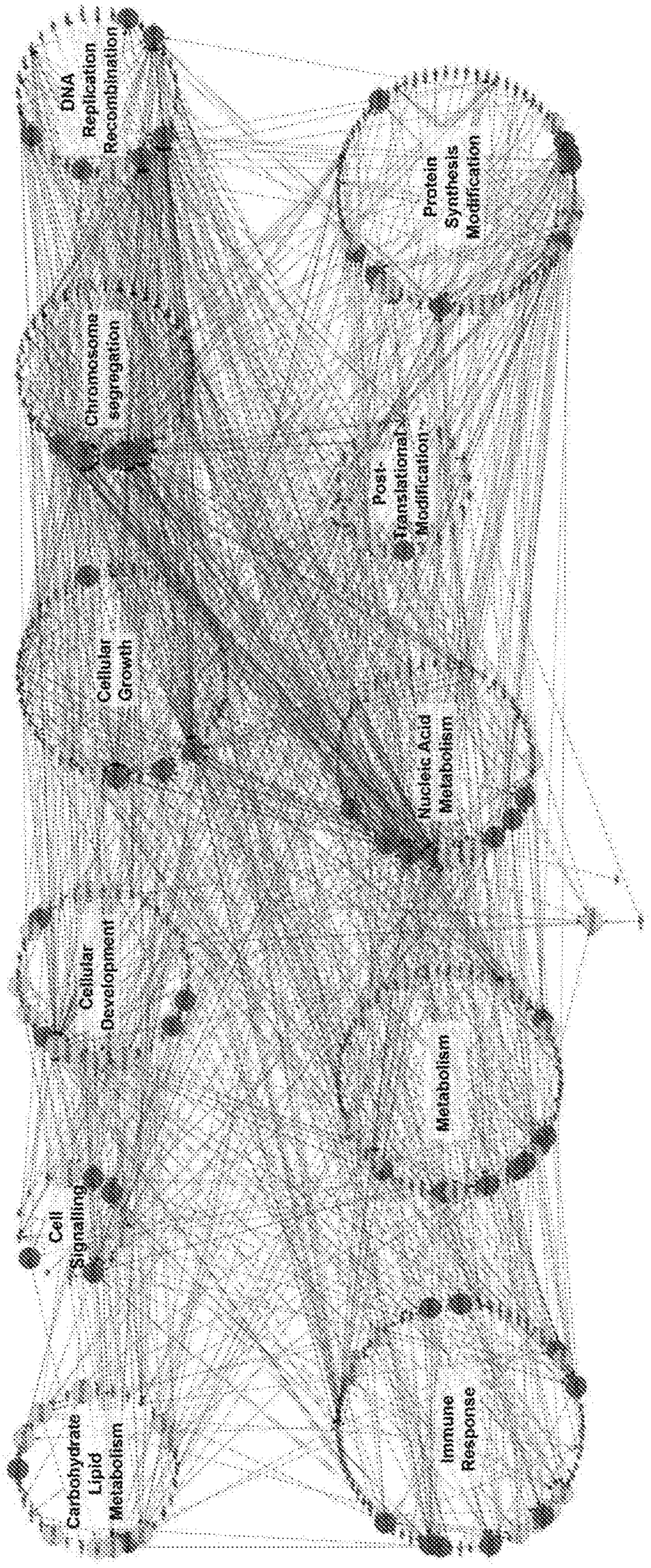

DNA
Replication
Recombination
Chromosome
segregation
Cellular
Growth
Post-
translational
Modification
Protein
Synthesis
Modification
Nucleic Acid
Metabolism
Cellular
Development
Cell
Signalling
Metabolism
Carbohydrate
Lipid
Metabolism
Immune
Response

FIG. 57

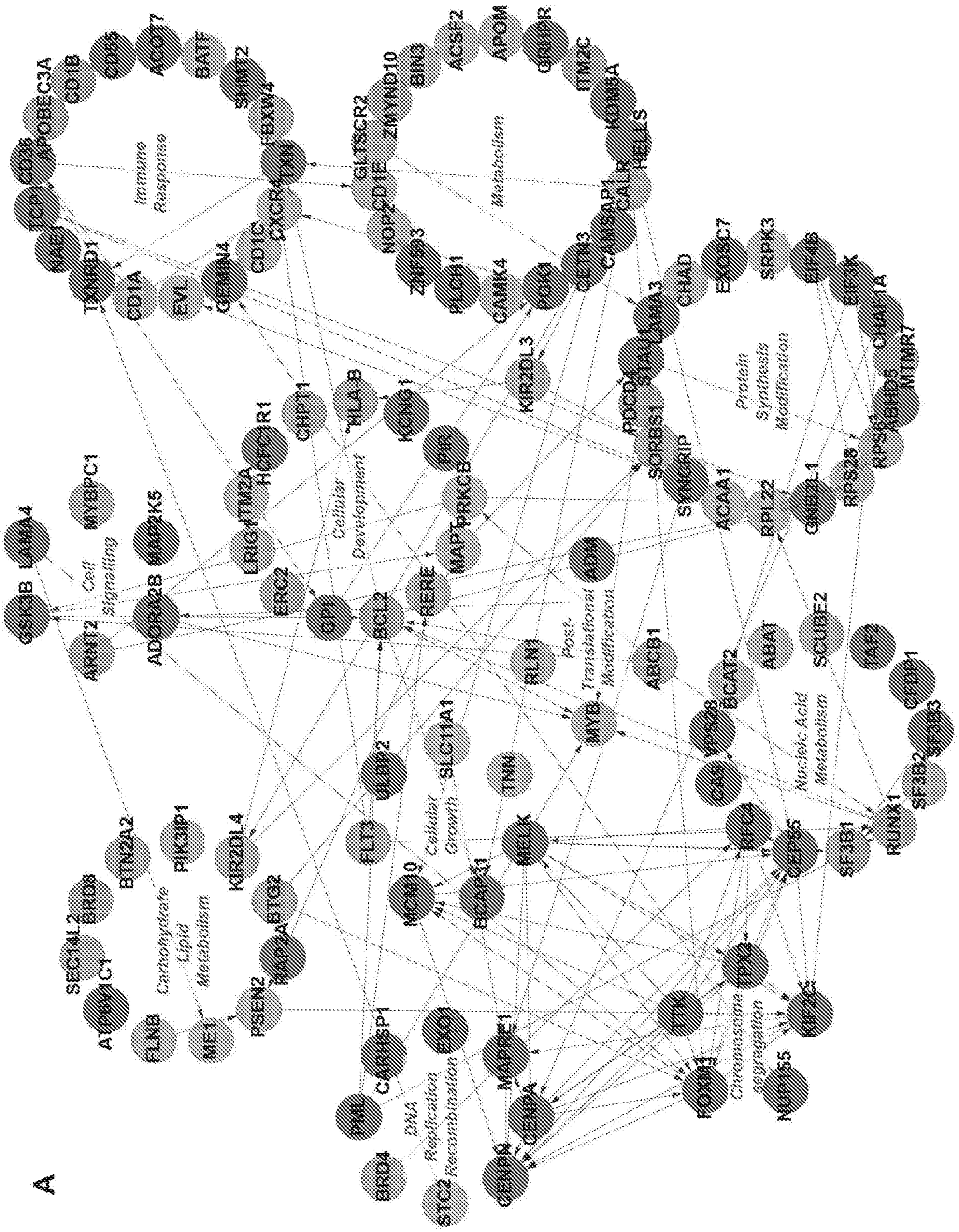

Immune Response
Metabolism
Cell signaling
Cellular Development
Post-Translational Modification
Protein Synthesis Modification
Carbohydrate Lipid Metabolism
Cellular Growth
Nucleic Acid Metabolism
DNA Replication
Chromosome segregation
A

FIG. 58

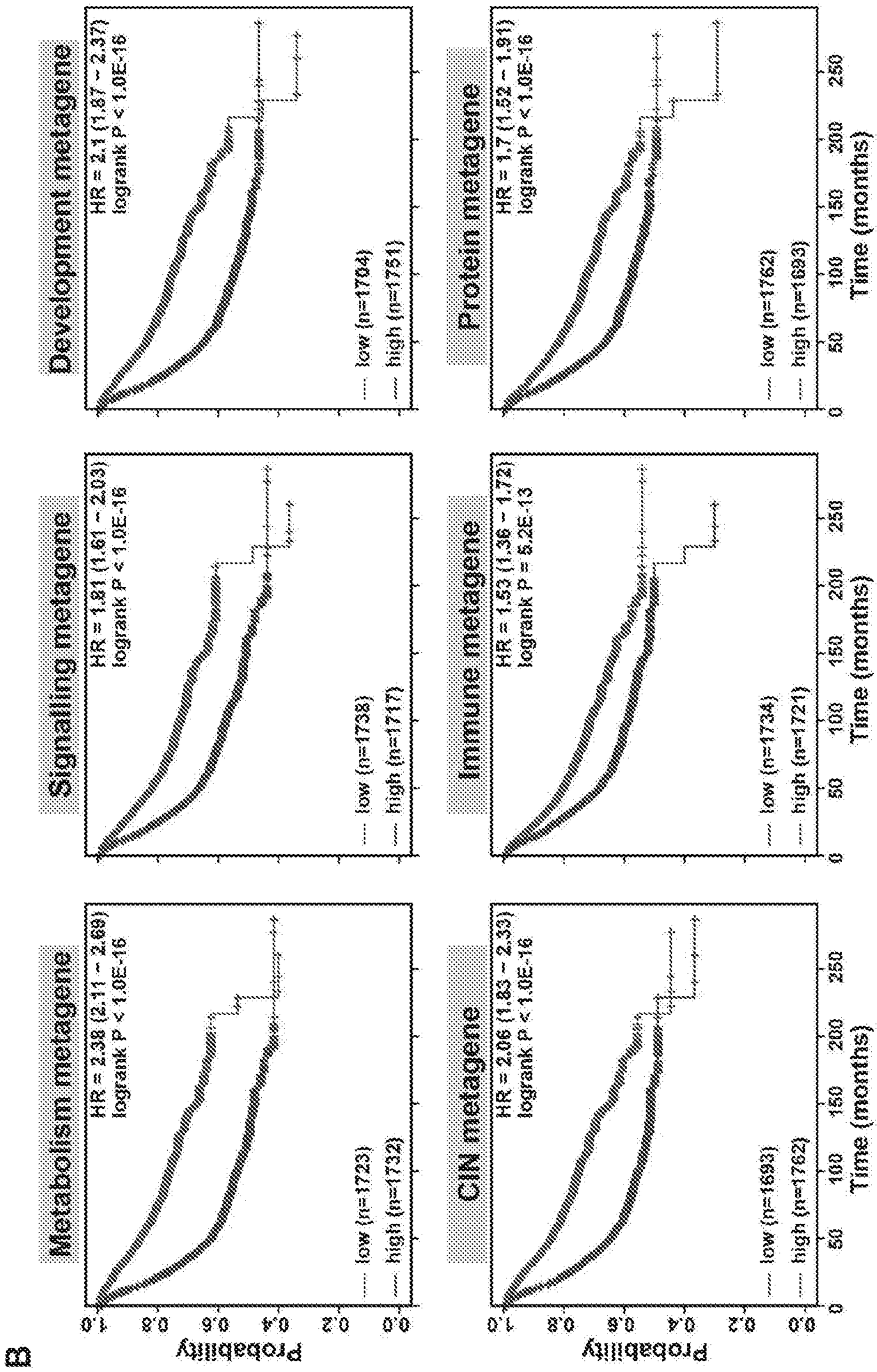

Development metagene
HR = 2.1 (1.87 ~ 2.37)
logrank P < 1.0E-16
low (n=1704)
high (n=1751)
Protein metagene
HR = 1.7 (1.52 ~ 1.91)
logrank P < 1.0E-16
low (n=1762)
high (n=1693)
Signalling metagene
HR = 1.81 (1.61 ~ 2.03)
logrank P < 1.0E-16
low (n=1738)
high (n=1717)
Immune metagene
HR = 1.53 (1.36 ~ 1.72)
logrank P = 5.2E-13
low (n=1734)
high (n=1721)
Metabolism metagene
HR = 2.38 (2.11 ~ 2.69)
logrank P < 1.0E-16
low (n=1723)
high (n=1732)
CIN metagene
HR = 2.06 (1.83 ~ 2.33)
logrank P < 1.0E-16
low (n=1693)
high (n=1762)
Time (months)
Probability

FIG. 58 cont'd

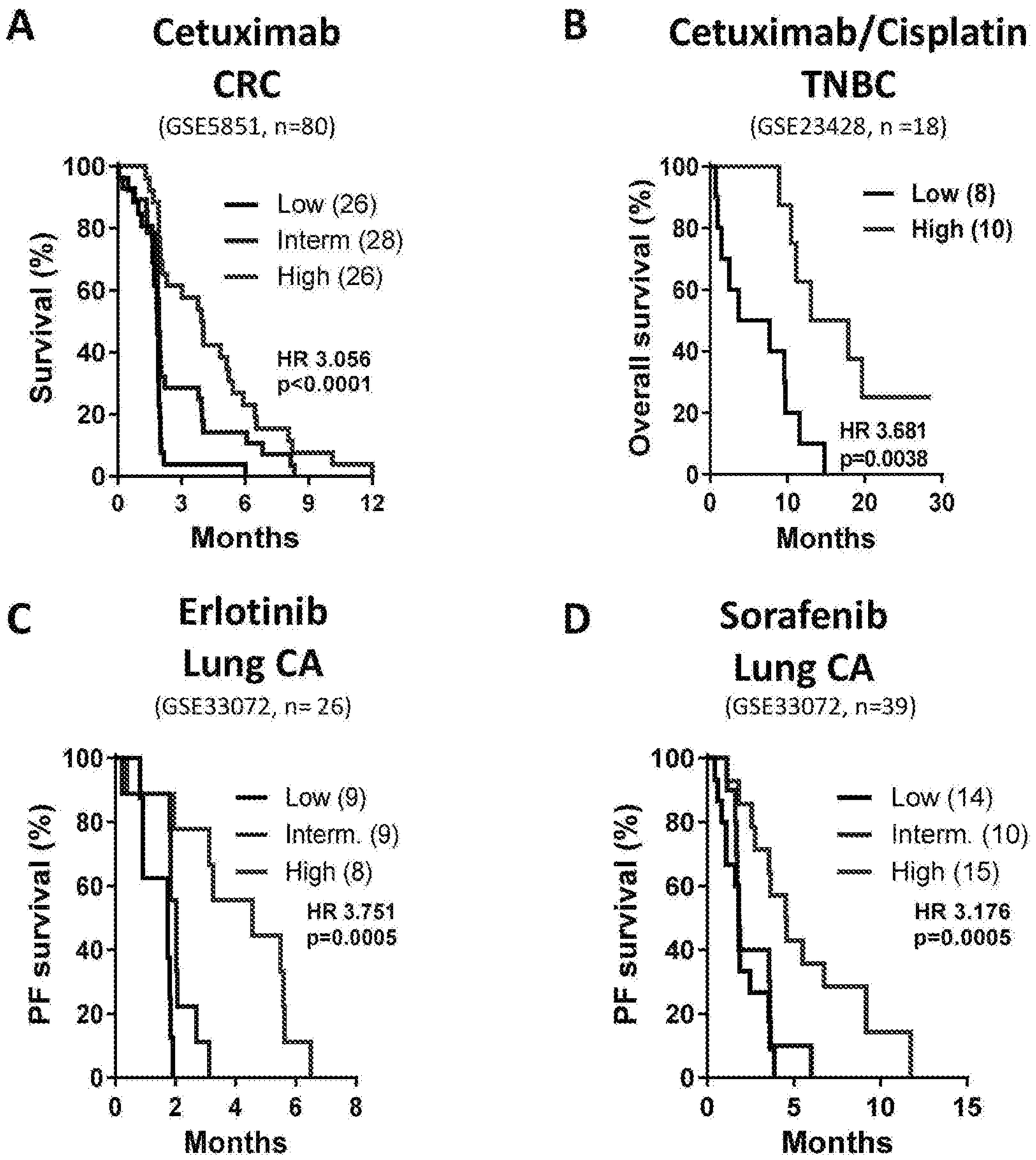

A
Cetuximab
CRC
(GSE5851, n=80)
Survival (%)
100
80
60
40
20
0
0
3
6
9
12
Months
Low (26)
Interm (28)
High (26)
HR 3.056
p<0.0001
B
Cetuximab/Cisplatin
TNBC
(GSE23428, n =18)
Overall survival (%)
100
80
60
40
20
0
0
10
20
30
Months
Low (8)
High (10)
HR 3.681
p=0.0038
C
Erlotinib
Lung CA
(GSE33072, n= 26)
PF survival (%)
100
80
60
40
20
0
0
2
4
6
8
Months
Low (9)
Interm. (9)
High (8)
HR 3.751
p=0.0005
D
Sorafenib
Lung CA
(GSE33072, n=39)
PF survival (%)
100
80
60
40
20
0
0
5
10
15
Months
Low (14)
Interm. (10)
High (15)
HR 3.176
p=0.0005

FIG. 59

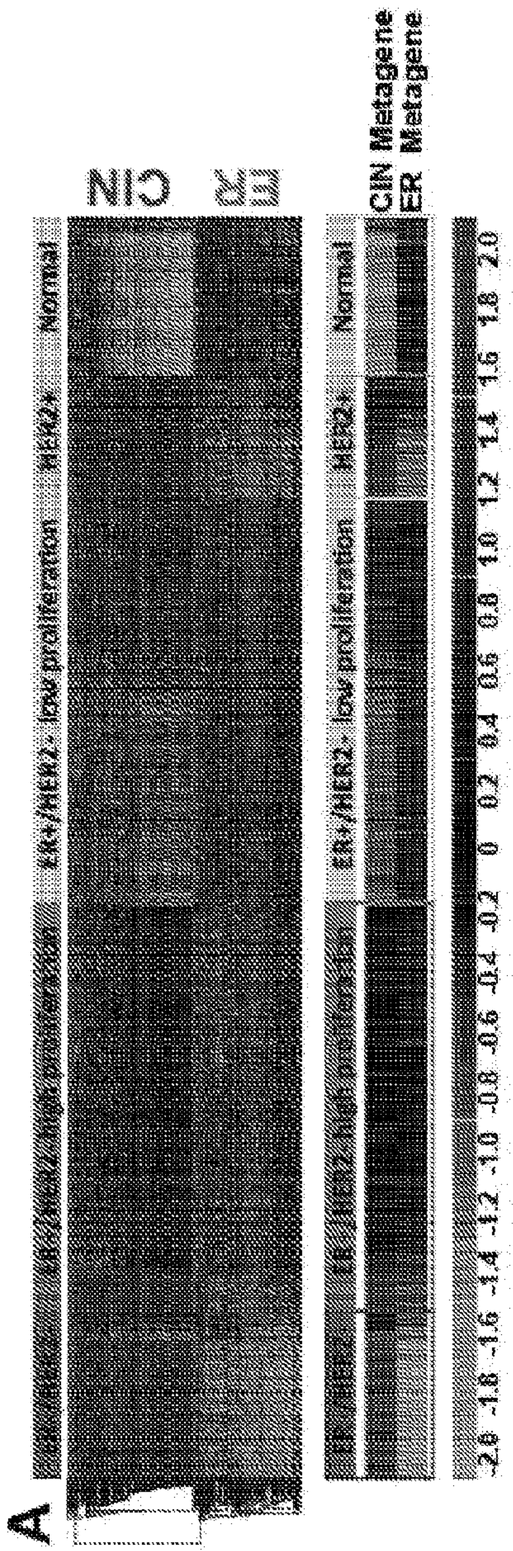
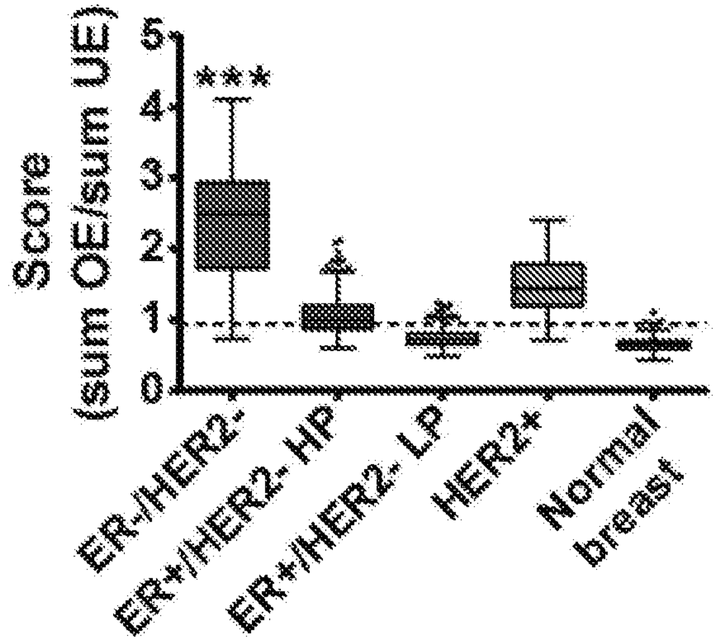


FIG. 1



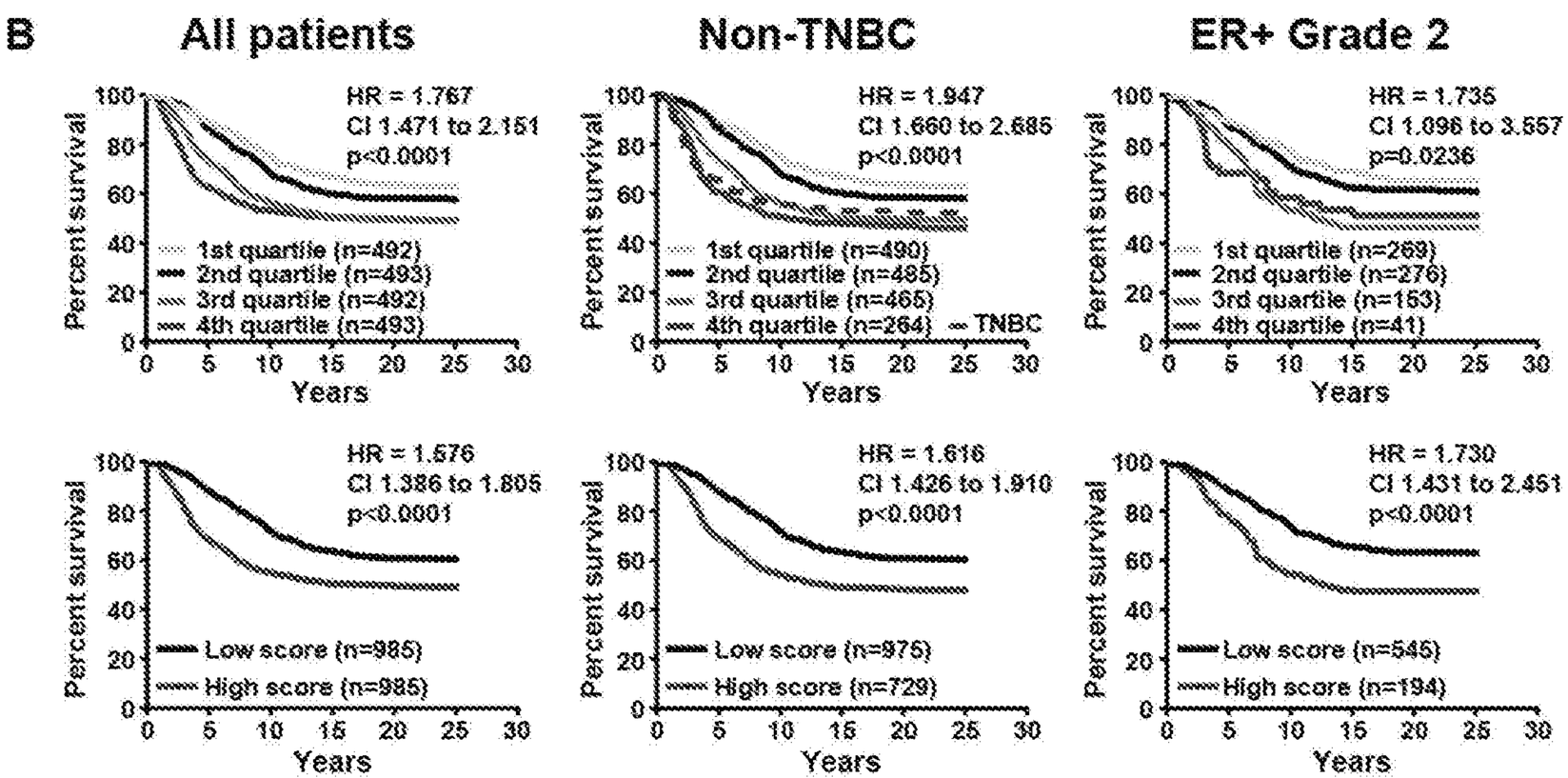


FIG. 1 cont'd