Breast Cancer Research and Treatment (2019) 178:185–197
https://doi.org/10.1007/s10549-019-05366-x

EPIDEMIOLOGY

Risk stratification of triple-negative breast cancer with core gene signatures associated with chemoresponse and prognosis

Eun-Kyu Kim[1] · Ae Kyung Park[2] · Eunyoung Ko[3] · Woong-Yang Park[4] · Kyung-Min Lee[5] · Dong-Young Noh[5,6] · Wonshik Han[5,6]

Received: 8 May 2019 / Accepted: 16 July 2019 / Published online: 24 July 2019

Abstract

Purpose Neoadjuvant chemotherapy studies have consistently reported a strong correlation between pathologic response and long-term outcome in triple-negative breast cancer (TNBC). We aimed to define minimal gene signatures for predicting chemoresponse by a three-step approach and to further develop a risk-stratification method of TNBC.

Methods The first step involved the detection of genes associated with resistance to docetaxel in eight TNBC cell lines, leading to identification of thousands of candidate genes. Through subsequent second and third step analyses with gene set enrichment analysis and survival analysis using public expression profiles, the candidate gene list was reduced to prognostic core gene signatures comprising ten or four genes.

Results The prognostic core gene signatures include three up-regulated (*CEBPD, MMP20*, and *WLS*) and seven down-regulated genes (*ASF1A, ASPSCR1, CHAF1B, DNMT1, GINS2, GOLGA2P5,* and *SKA1*). We further develop a simple risk-stratification method based on expression profiles of the core genes. Relative expression values of the up-regulated and down-regulated core genes were averaged into two scores, Up and Down scores, respectively; then samples were stratified by a diagonal line in a xy plot of the Up and Down scores. Based on this method, the patients were successfully divided into subgroups with distinct chemoresponse and prognosis. The prognostic power of the method was validated in three independent public datasets containing 230, 141, and 117 TNBC patients with chemotherapy. In multivariable Cox regression analysis, the core gene signatures were significantly associated with prognosis independent of tumor stage and age at diagnosis. In meta-analysis, we found that five core genes (*CEBPD, WLS, CHAF1B, GINS2*, and *SKA1*) play opposing roles, either tumor promoter or suppressor, in TNBC and non-TNBC tumors respectively, depending on estrogen receptor status.

Conclusions The results may provide a promising prognostic tool for predicting chemotherapy responders among TNBC patients prior to initiation of chemotherapeutic treatment.

Keywords Triple-negative breast cancer · Chemoresponse · Core gene · Risk stratification · Prognosis

Background

Triple-negative breast cancer (TNBC) is defined by lack of expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor-2 (HER2) [1, 2]. TNBC patients constitute approximately 20% of breast cancer cases and derive no benefit from molecular targeted treatments [1, 3]. Conventional chemotherapy is currently the only modality of systemic therapy for TNBC [4]. While patients with TNBC are associated with poorer clinical outcomes, they are more sensitive to chemotherapy and exhibit a higher likelihood of achieving a pathologic complete response (pCR) to neoadjuvant chemotherapy (NAC) [5–7]. Importantly, response to NAC can be

Eun-Kyu Kim and Ae Kyung Park have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10549-019-05366-x) contains supplementary material, which is available to authorized users.

✉ Wonshik Han
hanw@snu.ac.kr

Extended author information available on the last page of the article

used to predict patient outcomes. TNBC patients achieving pCR after NAC have excellent survival, comparable with that of patients with non-TNBC achieving pCR. However, TNBC patients with residual disease (RD) after NAC have significantly poorer survival rates than non-TNBC patients with RD [8]. These observations support the existence of subgroups of TNBCs with differential sensitivity to chemotherapy and many investigations are highlighting the clinical implications of the heterogeneity among TNBC [9–15]. Prediction of the responsiveness to chemotherapy is a fundamental step toward precision oncology, and determination of effective predictive factors of chemoresistance remains a high-research priority.

Taxanes including docetaxel are a class of chemotherapeutic agents that bind to and stabilize microtubules, causing cell cycle arrest and apoptosis. They are widely used for the treatment of different breast cancer types in neoadjuvant, adjuvant, and metastatic settings [16–18]. The addition of taxane to anthracycline-based chemotherapy has been shown to be of greater benefit for TNBC than ER-positive and HER2-negative cancers [19, 20].

In this study, we develop a new risk-stratification method in TNBC based on chemoresponsive and prognostic gene signatures identified by a three-step approach starting with detection of genes correlated with the resistance to docetaxel in eight TNBC cell lines and subsequent two more steps with comprehensive public datasets. The prognostic power of the method is further validated in three independent public datasets.

Methods

Cytotoxicity assays and microarray chip processing for eight TNBC cell lines

MDA-MB-231, DU4475, MDA-MB-436, and MDA-MB-157 cell lines were obtained from the American Type Culture Collection, and MDA-MB-468, HCC1937, HS578T, and HCC38 cell lines from Korea Cell Bank. MDA-MB-157 cells were grown in Leibovitz's L-15 medium (Gibco-BRL) with 10% fetal bovine serum (FBS; HyClone Labs) and MDA-MB-436 cells in Leibovitz's L-15 medium with 10% FBS and 0.026 units/ml insulin (Sigma Chemical). MDA-MB-231, MDA-MB-468, and HS578T cell lines were cultured in DMEM (Gibco) with 10% FBS in 5% CO_2. All other cell lines were grown in RPMI-1640 with 10% FBS in 5% CO_2. Docetaxel (Sigma-Aldrich) was dissolved in dimethyl sulfoxide. Cells were harvested via trypsinization using 0.25% trypsin–EDTA (Gibco), counted in a hemacytometer, and seeded at a density of 3000 cells/well in 96-well plates for 14 h. The medium was replaced with fresh medium containing different concentrations (0, 0.01, 0.1, 1, 10, and 100 nM) of docetaxel. After 48 h of exposure, cell viability experiments were performed using the CellTiter-Glo® Luminescent Cell Viability Assay (Promega) and ATP production measured via Victor 3 luminometer (Perkin-Elmer). Experiments were repeated twice and the 50% inhibitory concentration (IC_{50}) value of each cell line was evaluated via non-linear regression analysis (SigmaPlot 10.0). Total RNA for microarray was extracted from each cell line using the Qiagen RNeasy Mini kit, amplified and labeled according to the Affymetrix GeneChip Whole Transcript Sense Target Labeling protocol. The labeled cDNA was hybridized to the Affymetrix Human Gene 1.0 ST array. Scanned raw expression values were preprocessed using the Robust Multi-array Average (RMA) method [21] and $\log_2$-transformed.

Identification of candidate genes associated with chemoresistance and prognosis with three-step analysis

To identify the genes associated with docetaxel responsiveness in the eight TNBC cell lines, we applied two significance criteria: (1) range of changes in gene expression ($|\Delta Ex|$) across the cell lines and (2) absolute value of correlation coefficient ($|r|$) between expression and rank sum of chemoresistance. The rank sum of responsiveness to docetaxel for each cell line was calculated by summation of two ranks determined by the rank of survival fractions after docetaxel treatment at concentrations of 50 and 100 nM in a cell viability assay (Table 1). Docetaxel resistance-related genes were defined in cases where $|\Delta Ex|$ was > 1, and either one of two values of $|r|$, Pearson's or Spearman's correlation coefficient, was > 0.5. The genes obtained from the analysis of eight TNBC cell lines were designated the 1st candidate genes for taxane resistance.

To narrow down the candidate genes obtained from the first step, we collected four public datasets (GSE25055, GSE25065, GSE32646, and GSE41998) that included expression profiles and information on responses (pCR and RD) to taxane-containing NAC from the Gene Expression Omnibus (GEO) database. The chemotherapeutic regimens used in individual datasets were as follows: taxane/anthracycline (GSE25055 and GSE25065, collectively designated GSE25066), paclitaxel followed by 5-fluorouracil/epirubicin/cyclophosphamide (GSE32646), and doxorubicin/cyclophosphamide, followed by paclitaxel (GSE41998). Each microarray dataset was preprocessed separately using the RMA algorithm [13]. For two series, GSE25055 and GSE25065, preprocessed and rescaled expression profiles were downloaded directly from the GEO. For separation of TNBC from non-TNBC, expression values of ER, PR, and HER2 were modeled with two normal distributions [22]. TNBC samples were extracted using cutoff points where the two

Table 1 Characteristics and chemoresistance of eight triple-negative breast cancer (TNBC) cell lines

Cell line	Histology[a]	IC_{50} (nM)	Docetaxel (50 nM)		Docetaxel (100 nM)		Rank sum	Subtype[a] [9]	Mutations [9]
			Survival fraction	Rank	Survival fraction	Rank			
MDA-MB-468	DC	1.81	0.27	1	0.26	1	2	BL1	PTEN, RB1, SMAD4, TP53
HS578T	CS	11.08	0.32	2	0.29	3	5	MSL	CDKN2A, HRAS, TP53
DU4475	DC	11.57	0.37	3	0.27	2	5	IM	APC, BRAF, MAP2K4, RB1
MDA-MB-231	IDC	29.05	0.47	4	0.35	4	8	MSL	BRAF, CDKN2A, KRAS, NF2, TP53, PDGFRA
HCC38	DC	> 100	0.57	5	0.53	5	10	BL1	CDKN2A, TP53
MDA-MB-157	MBC	> 100	0.65	6	0.65	6	12	MSL	NF1, TP53
HCC1937	DC	> 100	0.76	7	0.70	7	14	BL1	BRCA1, TP53, MAPK13, MDC1
MDA-MB-436	IDC	> 100	0.81	8	0.75	8	16	MSL	BRCA1, TP53

[a]*DC* ductal carcinoma, *CS* carcinosarcoma, *IDC* invasive ductal carcinoma, *MBC* medullary breast cancer, *BL1* basal-like 1, *MSL* mesenchymal stem-like, *IM* immunomodulatory

normal distribution curves overlapped (Online Resource 1: Fig. S1). The final four TNBC datasets designated 'exploratory datasets' included a total of 271 expression profiles (119 patients from GSE25055, 63 from GSE25065, 64 from GSE41998, and 25 from GSE32646) with clinical information on 265 chemoresponse (pCR vs. RD) and 182 distant relapse-free survival (DRFS) outcome. Subsequently, Gene Set Enrichment Analysis (GSEA) was performed with the exploratory datasets [23]. In each dataset, genes were ranked by fold change ($\log_2$) of differential expression between the RD and pCR groups, and GSEAPreranked program implemented in GSEA applied to the pre-ranked gene list using the 1st candidate genes (positively and negatively correlated gene sets, respectively) as predefined gene sets. Finally, core genes in the leading-edge subset generated by GSEA were extracted from each dataset, and then commonly detected core genes from all four analyses were defined as the 2nd candidate genes. To further select only prognosis-related genes from the 2nd candidate genes, non-parametric log-rank test and parametric Cox regression analysis were performed using the GSE25066 (GSE25055 and GSE25065) dataset that included DRFS information of 182 TNBC patients. In the log-rank test, comparisons for each gene were made between survival distributions of patients divided into high and low expression groups based on three cutoff points (0.25, 0.5, and 0.75 quantile). If statistical significance (p value < 0.05) was obtained in at least one of the three comparisons, the gene was defined as a prognostic core gene and included in the 3rd candidate genes. In the Cox regression analysis, if statistical significance (p value < 0.05) was obtained, the gene was included in the 3rd candidate genes.

Validation of prognosis-related core genes based on survival analysis

For further validation analysis, we used another 15 breast cancer datasets, including two datasets containing overall survival (OS) outcomes, The Cancer Genome Atlas (TCGA) [24] and the METABRIC [25] datasets, and 13 GEO datasets (GSE11121, GSE12093, GSE16446, GSE17705, GSE19615, GSE2034, GSE20685, GSE22219, GSE2603, GSE26971, GSE6532, GSE7390, and GSE9195) containing metastasis-related event-free survival outcomes, such as DRFS, distant metastasis-free survival (DMFS), or metastasis-free survival (MFS). To investigate the prognostic power of the 3rd candidate genes, we examined five datasets (GSE16446, GSE19615, GSE20685, GSE22219, and GSE2603), which included a total of 246 TNBCs with 230 available clinical outcomes (DRFS, DMFS, or MFS). Details of the chemotherapeutic regimens in the five datasets are as follows: neoadjuvant (GSE16446) or adjuvant (GSE19615, GSE20685, GSE22219, and GSE2603), anthracycline (GSE16446), doxorubicin/cyclophosphamide followed by paclitaxel (AC-T), ACx4 or cyclophosphamide/doxorubicin/5-fluorouracil (CAF) (GSE19615), CAF, cyclophosphamide/methotrexate/5-fluorouracil (CMF), and other agents (GSE20685), CMF (GSE22219), and mostly adjuvant with no detailed regimen (GSE2603). In the TCGA RNA-sequencing dataset ($n = 1090$), 217 samples were identified as TNBC based on two normal distributions of ER, PR, and HER2 expressions [22], among which 141 were obtained from patients who had received chemotherapy. In the METABRIC dataset ($n = 1904$), 303 samples were identified as TNBC. For validation, we further selected the samples from the patients who had received

chemotherapy but not hormone therapy and did not die of other causes. The final validation METABRIC dataset included 117 TNBC samples. Finally, meta-analysis was performed with 13 GEO datasets containing metastasis-related event-free survival outcomes (DRFS, DMFS, or MFS) by applying the fixed effects model.

In all analyses with microarray expression profiles, only one probeset with the largest inter-quartile range across the samples was selected as the representative probeset when multiple probesets existed for a given gene (Entrez Gene ID) in each dataset. To combine dataset across multiple microarray platforms, we calculated the relative expression values for each gene across the TNBC samples in each dataset (ranging from 0 to 1), then the relative expression values were combined according to Entrez Gene ID. R statistical software and Bioconductor packages (http://www.R-project.org/, http://bioconductor.org/) were used for all data analyses.

Results

To identify robust candidate genes associated with response to chemotherapy in TNBC, a three-step sequential approach was employed. An overview of the analysis is presented in Online Resource 1 (Fig. S2), incorporating detection of taxane resistance-related genes in eight TNBC cell lines (1st candidate genes), extraction of core genes based on GSEA (2nd candidate genes), and final selection of prognosis-associated core gene signatures (3rd gene signatures).

Identification of candidate genes associated with chemoresistance to docetaxel in eight TNBC cell lines

The triple-negative status of eight cell lines was confirmed in microarray expression profiles based on low expressions of ER, PR, and HER2 relative to those of ER- and PR-positive MCF7 and HER2-amplified HCC1954 cell lines (Fig. 1a). The IC_{50} values and survival fractions after docetaxel treatment at concentrations of 50 nM and 100 nM are presented in Table 1. To detect the genes positively or negatively correlated with taxane resistance, correlation coefficients were calculated between gene expression and rank sum of survival fractions following treatment with 50 nM or 100 nM docetaxel (Table 1). The large rank-sum values indicate the high resistance to docetaxel. A gene was defined as significantly correlated with resistance when two criteria were met: the range of gene expression, $|\Delta Ex|$, was > 1 and absolute value of either Pearson's or Spearman's correlation coefficient ($|r|$) was > 0.5. Consequently, the final list of taxane resistance-related candidate genes included 2113 positively correlated genes ($r > 0.5$) and 1976 negatively correlated genes ($r < -0.5$), which were designated 1st POS and 1st NEG candidate genes, respectively (Fig. 1b).

Extraction of prognostic core genes based on GSEA and survival analysis using four public TNBC datasets

To further concentrate the candidate genes obtained from the first step, GSEA was applied to four public TNBC datasets (designated 'exploratory datasets') containing information on response to chemotherapy regimens

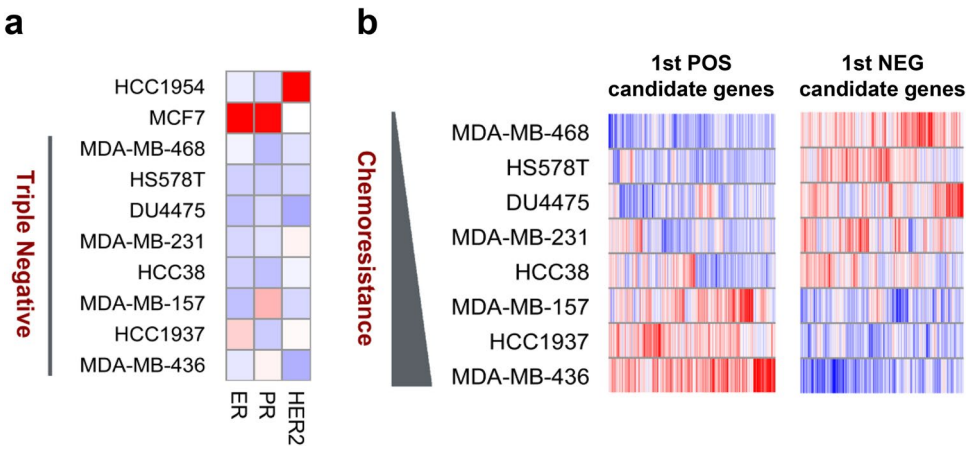

Fig. 1 Triple-negative status and 1st candidate genes correlated with chemoresistance to docetaxel in eight TNBC cell lines. **a** Heatmap of expression of ER (transcript cluster id: 8122843), PR (transcript cluster id: 7951165), and HER2 (transcript cluster id: 8006906) in ten cell lines, including eight TNBC, ER/PR-positive MCF7, and HER2-amplified HCC1954 cell lines. **b** Heatmap of 1st candidate genes including 2113 positively correlated genes (1st POS candidate genes) and 1976 negatively correlated genes (1st NEG candidate genes) with rank sums of chemoresistance to docetaxel

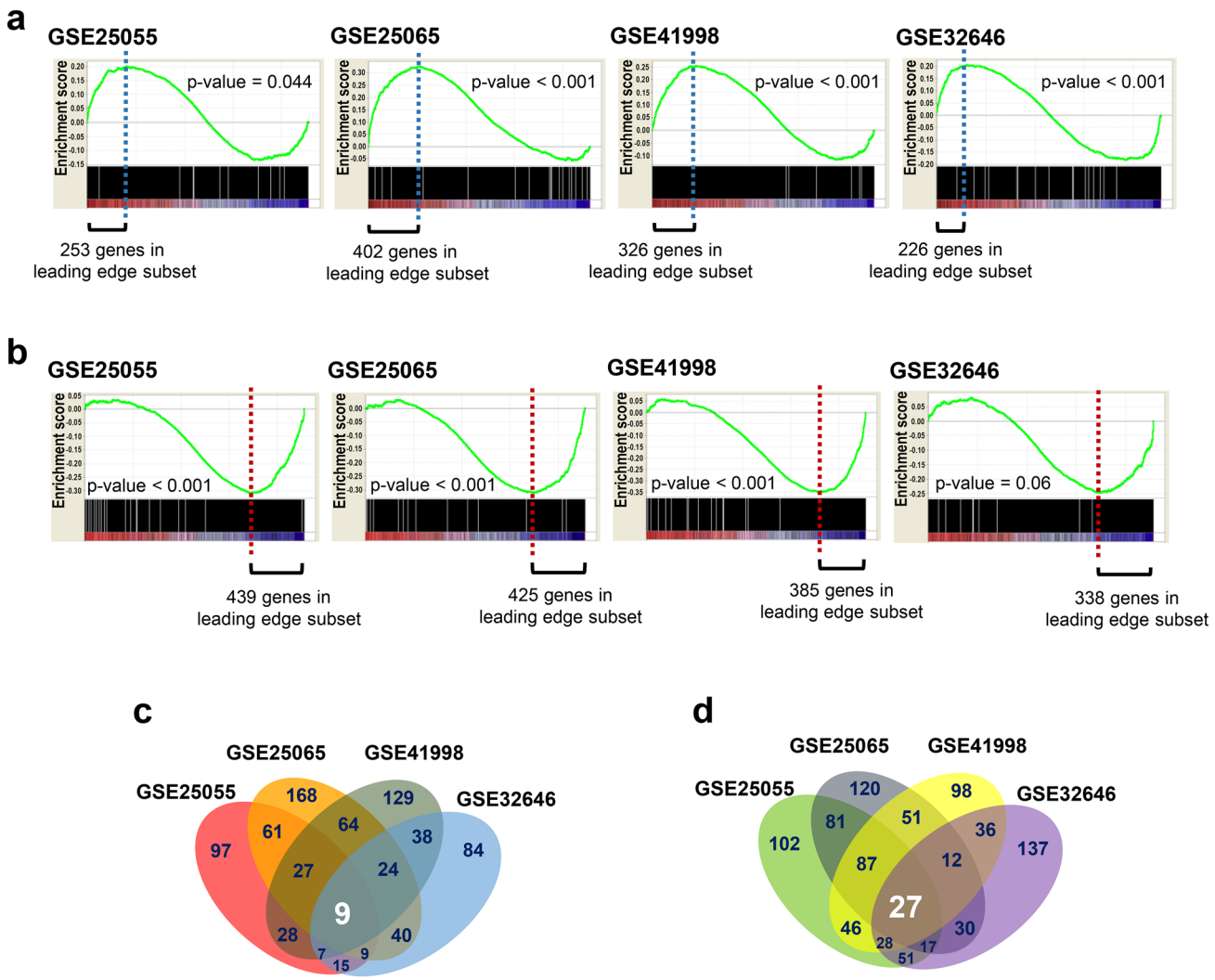

Fig. 2 Gene Set Enrichment Analysis (GSEA) using four public datasets (exploratory datasets) to extract chemoresistance-related core genes in leading-edge subsets. **a** Plots of running enrichment scores for the positively correlated gene set comprised of the 1st POS candidate genes. **b** Plots of running enrichment scores for the negatively correlated gene set comprised of the 1st NEG candidate genes. **c** Venn diagram of core genes detected in leading-edge subsets of four public datasets for the positively correlated gene set. **d** Venn diagram of core genes detected in leading-edge subsets of four public datasets for the negatively correlated gene set

including taxane (Fig. 2). Then, core genes in the leading-edge subset were extracted from the 1st POS and NEG candidate genes by applying GSEA based on differential expression between RD and pCR groups (Fig. 2a, b). The core genes in the intersection set of the genes commonly detected in all four exploratory datasets were defined as the 2nd candidate genes, which included 9 and 27 genes that were up-regulated and down-regulated, respectively, in RD groups compared to pCR groups (Fig. 2c, d). In the next step, we further narrowed down the gene list by selecting only the prognostic genes that were significantly associated with DRFS using either non-parametric log-rank test or parametric Cox regression analysis. Based on the log-rank test, we detected ten prognostic core genes, specifically, three up-regulated (*CEBPD*, *MMP20*, and *WLS*) and seven down-regulated (*ASF1A*, *ASPSCR1*, *CHAF1B*, *DNMT1*, *GINS2*, *GOLGA2P5*, and *SKA1*) genes, which showed significant association with DRFS at one of three cutoff points in log-rank tests (Online Resource 2: Table S1). Based on Cox regression analysis, we identified four prognostic core genes including one up-regulated (*CEBPD*) and three down-regulated genes (*ASPSCR1*, *CHAF1B*, and *SKA1*) (Online Resource 2: Table S1). The ten and four prognostic core genes comprise the final 3rd candidate genes (Table 2).

Table 2 Final prognostic core genes identified by three-step analysis

Class	Gene symbol	1st Step		2nd Step				3rd Step (GSE25066)							
		Correlation coefficient		Differential expression between RD versus pCR				TNBC ($n = 182$)							
		GSE25055 ($n = 119$)		GSE25065 ($n = 63$)	GSE41998 ($n = 64$)	GSE32646 ($n = 25$)		Log-rank test						Cox regression	
								0.25 quantile		0.5 quantile		0.75 quantile			
		Pearson	Spearman	Log_2FC	Log_2FC	Log_2FC	Log_2FC	Hazard ratio	*p* Value	Hazard ratio	*p* Value	Hazard ratio	*p* Value	Hazard ratio	*p* Value
Up	CEBPD	0.79	0.93	0.28	0.67	0.25	0.25	1.57	0.18	1.15	0.59	1.86	0.022	1.39	0.029
	MMP20	0.18	0.68	0.35	0.41	0.30	0.79	0.88	0.64	1.05	0.87	2.01	0.0097	1.13	0.13
	WLS	0.44	0.55	0.31	0.93	0.35	0.48	1.13	0.69	1.13	0.63	1.75	0.038	1.12	0.17
Down	ASF1A	−0.60	−0.59	−0.23	−0.52	−0.48	−0.55	0.61	0.069	0.53	0.020	0.72	0.32	0.83	0.17
	ASPSCR1	−0.47	−0.57	−0.42	−0.39	−0.13	−0.21	0.49	0.0086	0.68	0.15	0.66	0.20	0.77	0.0084
	CHAF1B	−0.77	−0.74	−1.07	−1.36	−0.27	−0.25	0.48	0.0059	0.37	0.0005	0.46	0.030	0.82	0.0019
	DNMT1	−0.88	−0.90	−0.16	−0.24	−0.21	−0.33	0.90	0.71	0.84	0.50	0.44	0.023	0.73	0.11
	GINS2	−0.57	−0.41	−0.54	−0.87	−0.23	−0.22	0.61	0.072	0.54	0.021	0.94	0.84	0.91	0.19
	GOLGA2P5	−0.57	−0.50	−0.22	−0.45	−0.14	−0.28	0.80	0.45	0.81	0.41	0.46	0.031	0.89	0.26
	SKA1	−0.42	−0.51	−0.58	−0.82	−0.23	−0.19	0.37	0.0002	0.57	0.034	0.56	0.096	0.78	0.0046

Predictive power of prognostic core genes in a combined exploratory dataset

To comprehensively determine the predictive power of the final ten and four prognostic core genes in the combined dataset including all four exploratory datasets, two scores, "Up score" and "Down score," were calculated by averaging the relative expression values of up- and down-regulated signature genes, respectively (Fig. 3a). In each exploratory dataset, the relative expression values of a given gene across TNBC patients (ranging from 0 to 1) were estimated. The relative values of the up-regulated and down-regulated genes were averaged, generating Up and Down scores for each patient (Fig. 3a). Under the assumption that chemoresponse and prognosis of TNBC patients can be distinguished by these two scores, patients were divided into two subgroups by a diagonal line in a scatter plot (Fig. 3b). The slope of the diagonal line was determined to be the ratio of the relative expression range of the Down scores to that of the Up scores. Namely, the slope was defined as the Down score range divided by the Up score range. The intercept was determined for the diagonal line to pass through the point of median values of the respective Up scores and Down scores. The two TNBC groups divided by the diagonal line showed significantly different survival curves with both ten and four prognostic core genes (Fig. 3b, upper panel). The patients were further divided into three groups (upper, middle, and lower including 25%, 50%, and 25% patients, respectively) by two lines with the same slope as the diagonal line in the upper panel of Fig. 3b. The proportions of RD in the upper, middle, and lower groups were 37.3%, 65.6%, and 85.1%, respectively, with ten prognostic core genes (Fig. 3b, lower panel). In addition, the three groups of TNBC patients showed distinct survival curves with statistical significance with both ten and four prognostic core genes although the curves of the upper and the middle groups were not distinguished with four prognostic core genes.

Validation of the prognostic core genes in independent datasets

We further investigated the prognostic power of the prognostic core genes in several independent datasets (Fig. 4). The first validation dataset was a combined dataset collected from five public datasets (GSE16446, GSE19615, GSE20685, GSE22219, and GSE2603) composed of 246 expression profiles of TNBC patients who had received chemotherapy with available clinical information on metastasis-related event-free survival outcomes in 230 cases. Due to the lack of appropriate validation datasets similar to exploratory datasets, considerably heterogeneous datasets were used for validation, as described in Methods. However, although the independent validation dataset included heterogeneous

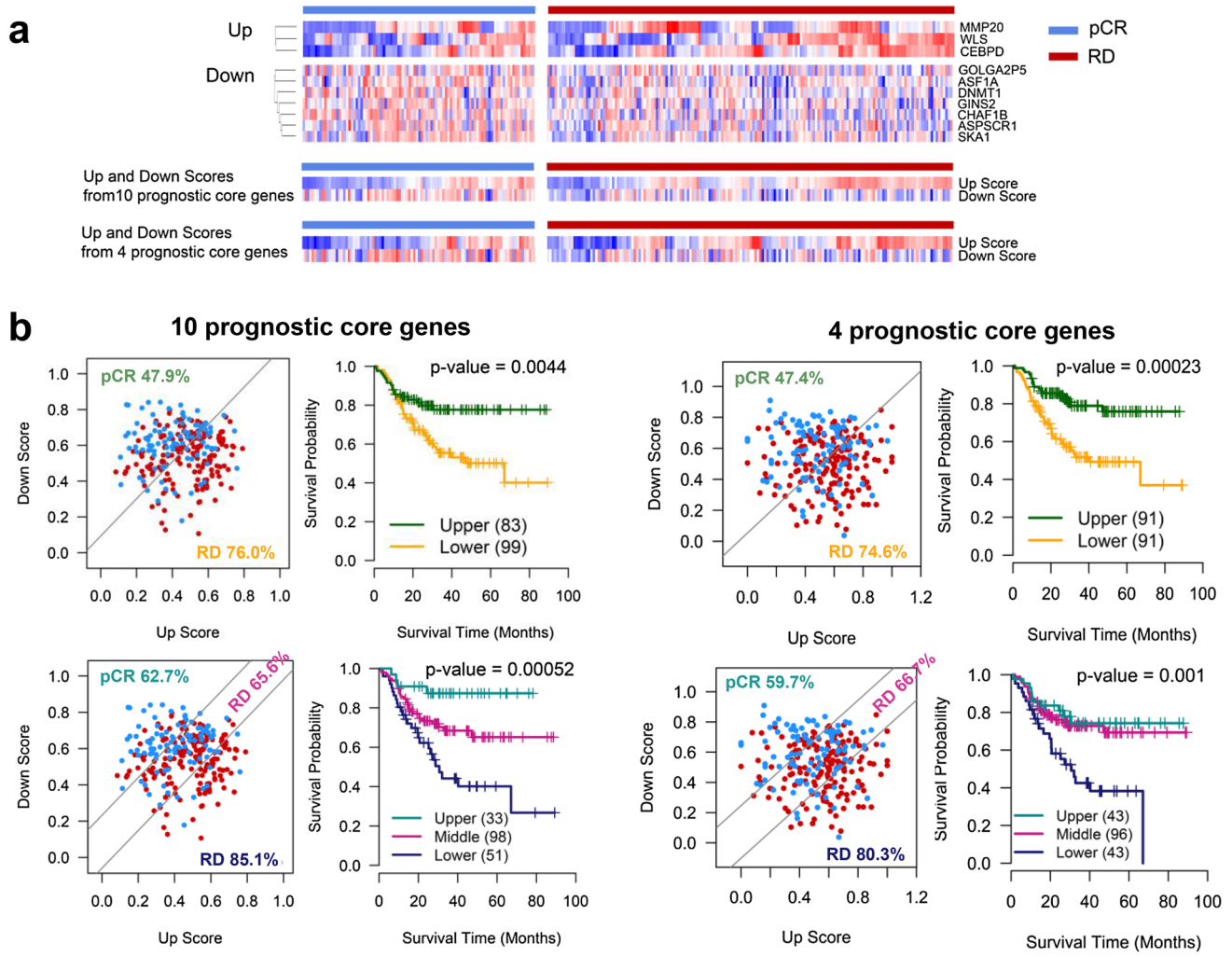

Fig. 3 Predictive power of ten and four prognostic core genes in exploratory datasets. **a** Heatmaps of relative expression values of ten and four prognostic core genes and Up and Down scores calculated by averaging the relative expression of up- and down-regulated prognostic core genes, respectively. **b** Scatter plots with chemoresponse and Kaplan–Meier plots with log-rank tests in two (upper panel) or three (lower panel) subgroups divided by diagonal lines based on Up and Down scores obtained from ten or four prognostic core genes in the combined exploratory dataset including 265 chemoresponse (pCR vs. RD) and 182 distant relapse-free survival (DRFS) cases. Blue and red dots indicate patients with pCR and RD, repectively

(or even unknown) chemotherapeutic regimens (i.e., not all neoadjuvant or taxane-containing regimens), the ten and four core genes remained significantly predictive of prognosis, showing different survival curves between the two or three subgroups divided by the same diagonal lines determined in the exploratory dataset (Fig. 4a). Similar patterns were observed for the TCGA dataset composed of 141 TNBC patients subjected to chemotherapy (Fig. 4b), although the Down score was calculated by averaging six genes instead of seven genes in the analysis of ten core genes because the expression profile of *GOLGA2P5* was not available in the TCGA dataset. The statistical significance was more obvious in the result with four prognostic core genes in the TCGA dataset. In the third validation dataset from METABRIC including 117 TNBC patients who received chemotherapy but not hormone therapy, the poor prognosis of the lower group in three-group analyses was obvious (Fig. 4c) although the expression profile of *SKA1* was not available in the METABRIC dataset. Furthermore, the prediction power was more prominent in 87 METABRIC TNBC patients with age at diagnosis ≥ 40 years (Fig. 4d).

Validation of the prognostic core genes as an independent prognostic predictor

To further investigate whether the final prognostic core genes are a significant predictor of prognosis independent of other potential prognostic variables such as tumor stage and

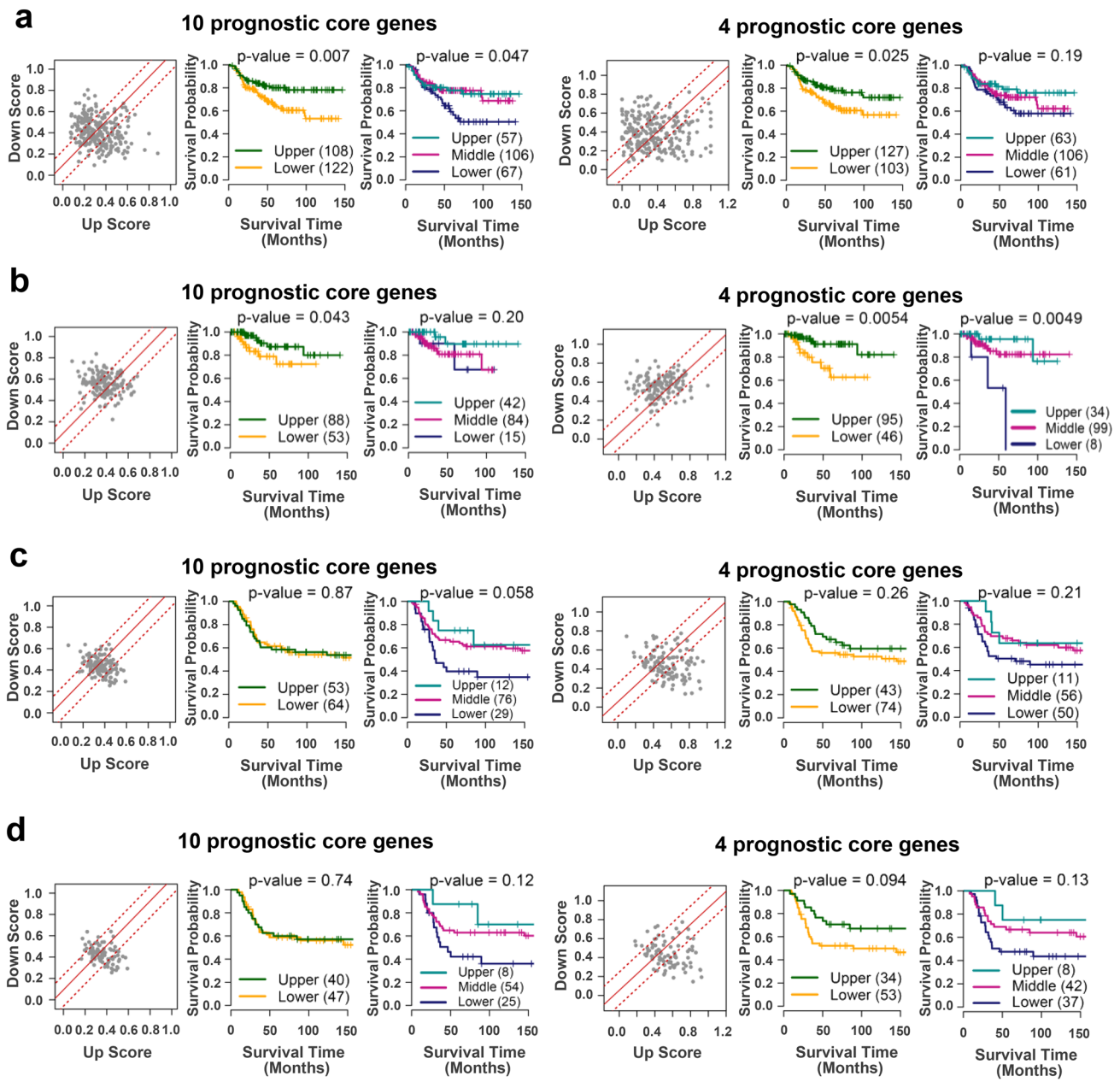

Fig. 4 Prognostic power of the prognostic core genes in validation datasets. Scatter and Kaplan–Meier plots with log-rank tests in two or three subgroups divided by the same diagonal lines in Fig. 3, **a** in the first validation dataset from GSE16446, GSE19615, GSE20685, GSE22219, and GSE2603 including 246 expression profiles and 230 metastasis-related survival outcomes from TNBC patients who received chemotherapy, **b** in the second validation dataset from TCGA including 141 expression profiles with overall survival outcomes of TNBC patients who received chemotherapy, **c** in the third validation dataset from METABRIC including 117 expression profiles with overall survival outcomes of TNBC patients who received chemotherapy without hormone therapy and did not die of other causes, **d** in the 87 TNBC patients with age at diagnosis ≥ 40 among the METABRIC 117 patients. Solid red lines in scatter plots indicate the line for dividing the patients into two subgroups and dotted ones indicate the lines for dividing the patients into three subgroups

age at diagnosis, we applied multivariable Cox regression analysis in two datasets, GSE25066 and TCGA, for which both tumor stage and age information of patients were available. The result revealed that the prognostic core genes were independent predictor of prognosis (Table 3). Kaplan–Meier plots with log-rank tests also showed that the patients in the upper group had favorable prognosis in both DRFS (GSE25066) (Fig. 5a) and OS (TCGA) (Fig. 5b) outcomes, when divided by the four prognostic core genes within the same categories of tumor stage and age at diagnosis.

Table 3 Hazard ratios and 95% confidence intervals (CIs) for age, tumor stage, and prognostic core genes

Prognostic core genes	Dataset	Covariates	Hazard ratio (95% CI)	*p* Value	Likelihood ratio test
Ten prognostic core genes	GSE25066 (Exploratory) ($n=175$)	Age (≥40 vs. <40)	1.02 (0.53, 1.97)	0.96	4.5×10^{-5}
		Stage (III vs. II)	3.08 (1.73, 5.49)	0.00013	
		Ten prognostic core genes (upper vs. lower)	0.43 (0.24, 0.77)	0.0041	
	TCGA (Validation) ($n=139$)	Age (≥40 vs. <40)	0.24 (0.06, 0.94)	0.04	1.1×10^{-7}
		Stage (III/IV vs. I/II)	23.5 (7.48, 73.8)	6.4×10^{-8}	
		Ten prognostic core genes (upper vs. lower)	0.29 (0.10, 0.83)	0.021	
Four prognostic core genes	GSE25066 (Exploratory) ($n=175$)	Age (≥40 vs. <40)	1.05 (0.54, 2.04)	0.88	4.7×10^{-5}
		Stage (III vs. II)			
		Four prognostic core genes	2.56 (1.44, 4.56)	0.0014	
		(upper vs. lower)	0.43 (0.25, 0.77)	0.0040	
	TCGA (Validation) ($n=139$)	Age (≥40 vs. <40)	0.40 (0.10, 1.50)	0.17	5.0×10^{-8}
		Stage (III/IV vs. I/II)	20.4 (6.93, 60.2)	4.5×10^{-8}	
		Four prognostic core genes (upper vs. lower)	0.24 (0.08, 0.71)	0.0099	

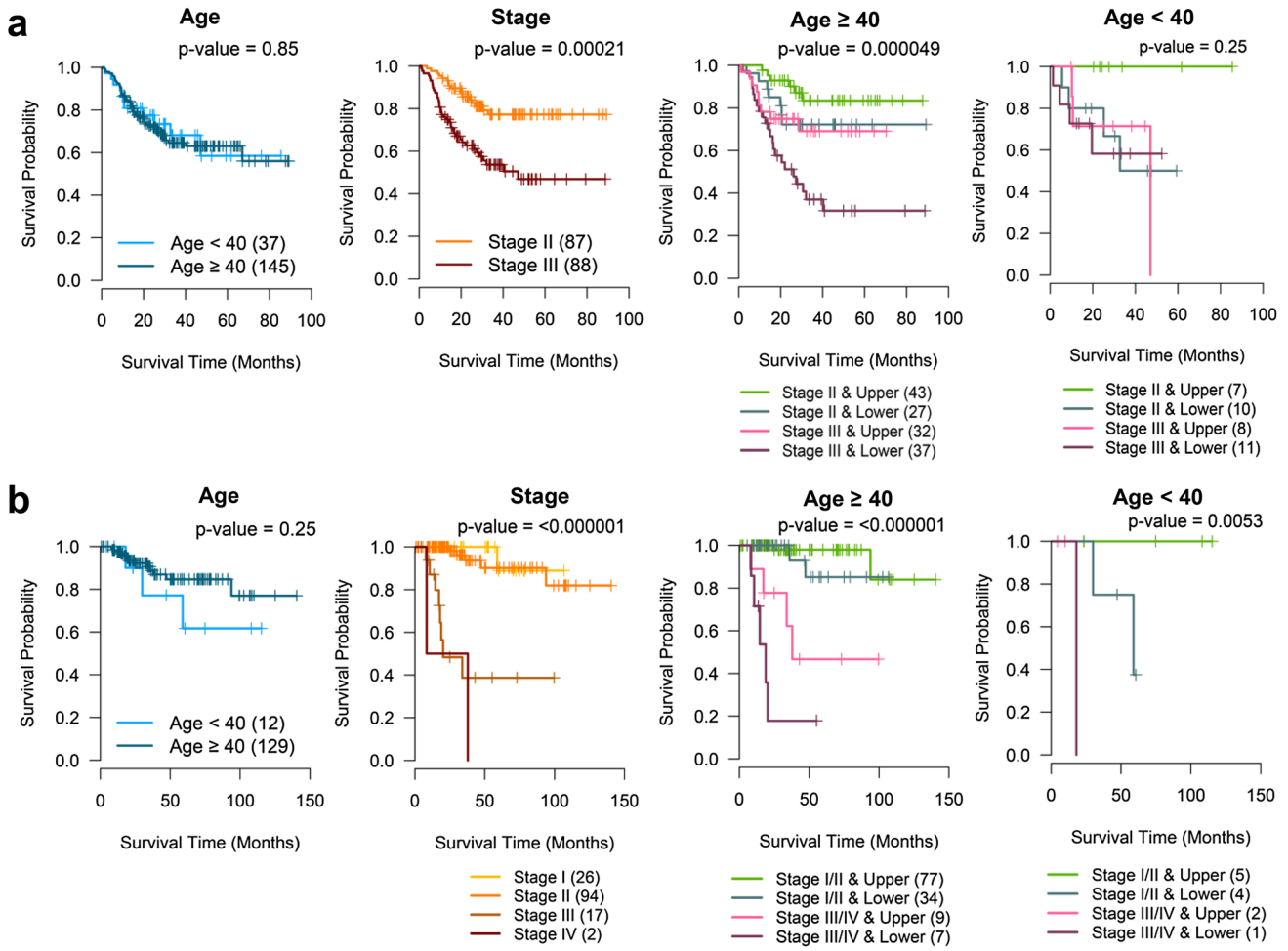

Fig. 5 Kaplan–Meier plots with log-rank tests with various TNBC patient subgroups stratified by age, tumor stage, and the four prognostic core genes. **a** GSE25066 dataset with distant relapse-free survival (DRFS) outcome **b** TCGA dataset with overall survival (OS) outcome

Discussion

Isolation of molecular predictors that can effectively facilitate identification of patients benefiting from specific agents is an important step toward precision medicine. Here, we established core gene signatures and a simple but novel method to determine chemoresponse and prognosis in TNBC patients. The core gene signature is comprised of ten prognostic core genes including three up-regulated and seven down-regulated genes, which can be further minimized into four prognostic gene signature.

One of the up-regulated prognostic core genes, *CEBPD* encoding C/EBPδ protein, is a transcription factor involved in critical regulation of immune and inflammatory responses [26]. To date, *CEBPD* has mainly been characterized as a tumor suppressor, as low expression or CpG methylation of the gene is correlated with metastasis or relapse in breast cancer [27, 28]. However, we found that up-regulation of *CEBPD* is related to chemoresistance and poor prognosis in TNBC patients (Table 2). C/EBPδ appears to play dual roles as a promoter or an inhibitor, depending on the cellular origin or physiological conditions [26]. Congruently, we observed the opposing roles of *CEBPD* by meta-analysis, a positive relationship between up-regulation of *CEBPD* and metastasis in TNBC and a negative relationship in non-TNBC (Online Resource 1: Fig. S3a, left two panels). The dual role of C/EBPδ is likely to depend on ER status (Online Resource 1: Fig. S3a, right two panels). Such subtype-dependent converse roles were additionally observed in four other genes, *WLS*, *CHAF1B*, *GINS2*, and *SKA1* (Online Resource 1: Fig. S3b–e). Another up-regulated gene was *WLS*, encoding Wntless (Wls) protein involved in Wnt signaling. The association between up-regulation of *WLS* and poor prognosis is also specific in TNBC (Online Resource 1: Fig. S3b). Wnt signaling is up-regulated in TNBC and proposed as a potential therapeutic target [29–32] and inhibition of Wnt signaling increases drug sensitivity in taxane-resistant advanced lung adenocarcinoma cells [33]. Furthermore, overexpression of *WLS* was observed in colon, gastric, ovarian, and breast cancers [34, 35], in relation to promotion of tumorigenesis in glioma [36], proliferation in breast cancer [37], and poor prognosis in B-cell precursor acute lymphoblastic leukemia [38]. Another up-regulated core gene, *MMP20*, was characterized as a broad activator of pro-kallikrein-related peptidases (KLKs). To date, several KLKs are identified as cancer biomarkers [39].

Among the seven down-regulated core prognostic genes, several appeared to be related to cell division (mitosis), proliferation, or cell cycle. Specifically, *SKA1* is involved in microtubule–kinetochore interactions for proper chromosomal segregation during mitosis [40]. *CHAF1B* and *GINS2* are involved in DNA replication and cell proliferation [41–44]. According to a previous investigation on TNBC, higher pCR rates in taxane-based NAC were observed in basal-like subtypes showing high expression of *MKI67* [9], implying that microtubule inhibitory activity of taxanes are effective on actively proliferating cells [9]. Consistently, expressions of *CHAF1B*, *GINS2*, *DNMT1*, and *SKA1* were significantly correlated with that of *MKI67* in the four exploratory datasets in our study (Online Resource 1: Fig. S4).

On the other hand, our core gene signatures are prognostic in the patients with chemotherapy independent of tumor stage and age at diagnosis (Table 3, Fig. 5). Traditional prognostic parameters of breast cancer include clinical, pathological, and immunochemistry markers such as age at occurrence, comorbidity, tumor size, histological grade, number of involved axillary lymph nodes, hormone receptor/HER2 status, and Ki-67 expression [45–47]. As a classic pathological variable, axillary lymph node metastasis is considered to be the most significant prognostic factor [48–50]. Multigene assays also provide promising molecular tools for prediction of prognosis, subtyping, and therapeutic decision making [45]. In addition, new applications of various techniques such as diffusion-weighted magnetic resonance imaging allow more precise diagnosis, monitoring of post-treatment changes, and prediction of the response to chemotherapy in breast cancer [51–54].

A limitation of this study is that the core genes were identified starting with the analysis of cell lines. However, recent research has shown that cell line collection mirrors the majority of important genomic and transcriptional abnormalities found in primary breast tumors [55, 56]. In the analysis with TNBC cell lines, the two top-ranked cell lines showing the highest level of chemoresistance were HCC1937 and MDA-MB-436 that harbor the *BRCA1* mutation (Table 1). This finding is consistent with previous reports that mutation of *BRCA1* is associated with chemoresistance to spindle poisons, such as taxanes and vinca alkaloids, which may be attributable to loss of the mitotic checkpoint function of BRCA1 that normally induces apoptotic death followed by microtubule damage by spindle poisons [57, 58].

Conclusion

We developed a new risk-stratification strategy for prediction of chemoresponse and prognosis of TNBC, which has a high potential for use in identifying chemoresponsive or chemoresistant TNBC patients prior to initiation of chemotherapy and ultimately elucidating the biological mechanisms underlying chemoresistance.

Acknowledgements We thank all the individuals who took part in the Translational Research Organization in Cancer (TROICA) project and all the researchers who have enabled this work to be carried out.

Funding This research was funded by grants of the National Research Foundation (NRF) of Korea funded by the Korea government (MSIP), 2015R1A2A2A01008264 (to WH) and 2015R1A4A1041219 (to AKP).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Not applicable.

References

1. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V (2007) Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer 109(9):1721–1728. https://doi.org/10.1002/cncr.22618
2. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA (2007) Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 13(15 Pt 1):4429–4434. https://doi.org/10.1158/1078-0432.CCR-06-3045
3. Haffty BG, Yang Q, Reiss M, Kearney T, Higgins SA, Weidhaas J, Harris L, Hait W, Toppmeyer D (2006) Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. J Clin Oncol 24(36):5652–5657. https://doi.org/10.1200/JCO.2006.06.5664
4. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, Senn HJ, Panel M (2013) Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol 24(9):2206–2223. https://doi.org/10.1093/annonc/mdt303
5. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P, Morandi P, Fan C, Rabiul I, Ross JS, Hortobagyi GN, Pusztai L (2005) Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 11(16):5678–5685. https://doi.org/10.1158/1078-0432.CCR-04-2421
6. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM (2007) The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res 13(8):2329–2334. https://doi.org/10.1158/1078-0432.CCR-06-1109
7. Bhargava R, Beriwal S, Dabbs DJ, Ozbek U, Soran A, Johnson RR, Brufsky AM, Lembersky BC, Ahrendt GM (2010) Immunohistochemical surrogate markers of breast cancer molecular classes predicts response to neoadjuvant chemotherapy: a single institutional experience with 359 cases. Cancer 116(6):1431–1439. https://doi.org/10.1002/cncr.24876
8. Liedtke C, Mazouni C, Hess KR, Andre F, Tordai A, Mejia JA, Symmans WF, Gonzalez-Angulo AM, Hennessy B, Green M, Cristofanilli M, Hortobagyi GN, Pusztai L (2008) Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol 26(8):1275–1281. https://doi.org/10.1200/JCO.2007.14.4147
9. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA (2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 121(7):2750–2767. https://doi.org/10.1172/JCI45014
10. Masuda H, Baggerly KA, Wang Y, Zhang Y, Gonzalez-Angulo AM, Meric-Bernstam F, Valero V, Lehmann BD, Pietenpol JA, Hortobagyi GN, Symmans WF, Ueno NT (2013) Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. Clin Cancer Res 19(19):5533–5540. https://doi.org/10.1158/1078-0432.CCR-13-0799
11. Lehmann BD, Jovanovic B, Chen X, Estrada MV, Johnson KN, Shyr Y, Moses HL, Sanders ME, Pietenpol JA (2016) Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. PLoS ONE 11(6):e0157368. https://doi.org/10.1371/journal.pone.0157368
12. Rody A, Karn T, Liedtke C, Pusztai L, Ruckhaeberle E, Hanker L, Gaetje R, Solbach C, Ahr A, Metzler D, Schmidt M, Muller V, Holtrich U, Kaufmann M (2011) A clinically relevant gene signature in triple negative and basal-like breast cancer. Breast Cancer Res 13(5):R97. https://doi.org/10.1186/bcr3035
13. Prat A, Lluch A, Albanell J, Barry WT, Fan C, Chacon JI, Parker JS, Calvo L, Plazaola A, Arcusa A, Segui-Palmer MA, Burgues O, Ribelles N, Rodriguez-Lescure A, Guerrero A, Ruiz-Borrego M, Munarriz B, Lopez JA, Adamo B, Cheang MC, Li Y, Hu Z, Gulley ML, Vidal MJ, Pitcher BN, Liu MC, Citron ML, Ellis MJ, Mardis E, Vickery T, Hudis CA, Winer EP, Carey LA, Caballero R, Carrasco E, Martin M, Perou CM, Alba E (2014) Predicting response and survival in chemotherapy-treated triple-negative breast cancer. Br J Cancer 111(8):1532–1541. https://doi.org/10.1038/bjc.2014.444
14. Yam C, Mani SA, Moulder SL (2017) Targeting the molecular subtypes of triple negative breast cancer: understanding the diversity to progress the field. Oncologist 22(9):1086–1093. https://doi.org/10.1634/theoncologist.2017-0095
15. Ring BZ, Hout DR, Morris SW, Lawrence K, Schweitzer BL, Bailey DB, Lehmann BD, Pietenpol JA, Seitz RS (2016) Generation of an algorithm based on minimal gene sets to clinically subtype triple negative breast cancer patients. BMC Cancer 16:143. https://doi.org/10.1186/s12885-016-2198-0
16. Noguchi S (2006) Predictive factors for response to docetaxel in human breast cancers. Cancer Sci 97(9):813–820. https://doi.org/10.1111/j.1349-7006.2006.00265.x
17. Cortes J, Baselga J (2007) Targeting the microtubules in breast cancer beyond taxanes: the epothilones. Oncologist 12(3):271–280. https://doi.org/10.1634/theoncologist.12-3-271
18. Bedard PL, Di Leo A, Piccart-Gebhart MJ (2010) Taxanes: optimizing adjuvant chemotherapy for early-stage breast cancer. Nat Rev Clin Oncol 7(1):22–36. https://doi.org/10.1038/nrclinonc.2009.186
19. Hayes DF, Thor AD, Dressler LG, Weaver D, Edgerton S, Cowan D, Broadwater G, Goldstein LJ, Martino S, Ingle JN, Henderson IC, Norton L, Winer EP, Hudis CA, Ellis MJ, Berry DA, Cancer, Leukemia Group BI (2007) HER2 and response to paclitaxel in node-positive breast cancer. N Engl J Med 357(15):1496–1506. https://doi.org/10.1056/NEJMoa071167
20. Ellis P, Barrett-Lee P, Johnson L, Cameron D, Wardley A, O'Reilly S, Verrill M, Smith I, Yarnold J, Coleman R, Earl H, Canney P, Twelves C, Poole C, Bloomfield D, Hopwood P, Johnston S, Dowsett M, Bartlett JM, Ellis I, Peckitt C, Hall E, Bliss JM, Group TTM, Trialists T (2009) Sequential docetaxel as adjuvant chemotherapy for early breast cancer (TACT): an open-label, phase III, randomised controlled trial.

Lancet 373(9676):1681–1692. https://doi.org/10.1016/S0140-6736(09)60740-6
21. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4(2):249–264
22. Karn T, Metzler D, Ruckhaberle E, Hanker L, Gatje R, Solbach C, Ahr A, Schmidt M, Holtrich U, Kaufmann M, Rody A (2010) Data-driven derivation of cutoffs from a pool of 3030 Affymetrix arrays to stratify distinct clinical types of breast cancer. Breast Cancer Res Treat 120(3):567–579. https://doi.org/10.1007/s10549-009-0416-z
23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102(43):15545–15550. https://doi.org/10.1073/pnas.0506580102
24. Cancer Genome Atlas N (2012) Comprehensive molecular portraits of human breast tumours. Nature 490(7418):61–70. https://doi.org/10.1038/nature11412
25. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Group M, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham I, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C, Aparicio S (2012) The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. Nature 486(7403):346–352. https://doi.org/10.1038/nature10983
26. Balamurugan K, Sterneck E (2013) The many faces of C/EBPdelta and their relevance for inflammation and cancer. Int J Biol Sci 9(9):917–933. https://doi.org/10.7150/ijbs.7224
27. Zucchi I, Mento E, Kuznetsov VA, Scotti M, Valsecchi V, Simionati B, Vicinanza E, Valle G, Pilotti S, Reinbold R, Vezzoni P, Albertini A, Dulbecco R (2004) Gene expression profiles of epithelial cells microscopically isolated from a breast-invasive ductal carcinoma and a nodal metastasis. Proc Natl Acad Sci USA 101(52):18147–18152. https://doi.org/10.1073/pnas.0408260101
28. Palmieri C, Monteverde M, Lattanzio L, Gojis O, Rudraraju B, Fortunato M, Syed N, Thompson A, Garrone O, Merlano M, Lo Nigro C, Crook T (2012) Site-specific CpG methylation in the CCAAT/enhancer binding protein delta (CEBPdelta) CpG island in breast cancer is associated with metastatic relapse. Br J Cancer 107(4):732–738. https://doi.org/10.1038/bjc.2012.308
29. King TD, Suto MJ, Li Y (2012) The Wnt/beta-catenin signaling pathway: a potential therapeutic target in the treatment of triple negative breast cancer. J Cell Biochem 113(1):13–18. https://doi.org/10.1002/jcb.23350
30. Dey N, Barwick BG, Moreno CS, Ordanic-Kodani M, Chen Z, Oprea-Ilies G, Tang W, Catzavelos C, Kerstann KF, Sledge GW Jr, Abramovitz M, Bouzyk M, De P, Leyland-Jones BR (2013) Wnt signaling in triple negative breast cancer is associated with metastasis. BMC Cancer 13:537. https://doi.org/10.1186/1471-2407-13-537
31. Bilir B, Kucuk O, Moreno CS (2013) Wnt signaling blockage inhibits cell proliferation and migration, and induces apoptosis in triple-negative breast cancer cells. J Transl Med 11:280. https://doi.org/10.1186/1479-5876-11-280
32. Koval AV, Vlasov P, Shichkova P, Khunderyakova S, Markov Y, Panchenko J, Volodina A, Kondrashov FA, Katanaev VL (2014) Anti-leprosy drug clofazimine inhibits growth of triple-negative breast cancer cells via inhibition of canonical Wnt signaling. Biochem Pharmacol 87(4):571–578. https://doi.org/10.1016/j.bcp.2013.12.007
33. Ren J, Wang R, Song H, Huang G, Chen L (2014) Secreted frizzled related protein 1 modulates taxane resistance of human lung adenocarcinoma. Mol Med 20:164–178. https://doi.org/10.2119/molmed.2013.00149
34. Voloshanenko O, Erdmann G, Dubash TD, Augustin I, Metzig M, Moffa G, Hundsrucker C, Kerr G, Sandmann T, Anchang B, Demir K, Boehm C, Leible S, Ball CR, Glimm H, Spang R, Boutros M (2013) Wnt secretion is required to maintain high levels of Wnt activity in colon cancer cells. Nat Commun 4:2610. https://doi.org/10.1038/ncomms3610
35. Stewart J, James J, McCluggage GW, McQuaid S, Arthur K, Boyle D, Mullan P, McArt D, Yan B, Irwin G, Harkin DP, Zhengdeng L, Ong CW, Yu J, Virshup DM, Salto-Tellez M (2015) Analysis of wntless (WLS) expression in gastric, ovarian, and breast cancers reveals a strong association with HER2 overexpression. Mod Pathol 28(3):428–436. https://doi.org/10.1038/modpathol.2014.114
36. Augustin I, Goidts V, Bongers A, Kerr G, Vollert G, Radlwimmer B, Hartmann C, Herold-Mende C, Reifenberger G, von Deimling A, Boutros M (2012) The Wnt secretion protein Evi/Gpr177 promotes glioma tumourigenesis. EMBO Mol Med 4(1):38–51. https://doi.org/10.1002/emmm.201100186
37. Lu D, Li Y, Liu QR, Wu Q, Zhang H, Xie P, Wang Q (2015) Wls promotes the proliferation of breast cancer cells via Wnt signaling. Med Oncol 32(5):140. https://doi.org/10.1007/s12032-015-0585-z
38. Chiou SS, Wang LT, Huang SB, Chai CY, Wang SN, Liao YM, Lin PC, Liu KY, Hsu SH (2014) Wntless (GPR177) expression correlates with poor prognosis in B-cell precursor acute lymphoblastic leukemia via Wnt signaling. Carcinogenesis 35(10):2357–2364. https://doi.org/10.1093/carcin/bgu166
39. Yoon H, Blaber SI, Li W, Scarisbrick IA, Blaber M (2013) Activation profiles of human kallikrein-related peptidases by matrix metalloproteinases. Biol Chem 394(1):137–147. https://doi.org/10.1515/hsz-2012-0249
40. Jeyaprakash AA, Santamaria A, Jayachandran U, Chan YW, Benda C, Nigg EA, Conti E (2012) Structural and functional organization of the Ska complex, a key component of the kinetochore-microtubule interface. Mol Cell 46(3):274–286. https://doi.org/10.1016/j.molcel.2012.03.005
41. Chang YP, Wang G, Bermudez V, Hurwitz J, Chen XS (2007) Crystal structure of the GINS complex and functional insights into its role in DNA replication. Proc Natl Acad Sci USA 104(31):12685–12690. https://doi.org/10.1073/pnas.0705558104
42. Zhang X, Zhong L, Liu BZ, Gao YJ, Gao YM, Hu XX (2013) Effect of GINS2 on proliferation and apoptosis in leukemic cell line. Int J Med Sci 10(12):1795–1804. https://doi.org/10.7150/ijms.7025
43. Kaufman PD, Kobayashi R, Kessler N, Stillman B (1995) The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. Cell 81(7):1105–1114. https://doi.org/10.1016/s0092-8674(05)80015-7
44. Mascolo M, Ilardi G, Merolla F, Russo D, Vecchione ML, de Rosa G, Staibano S (2012) Tissue microarray-based evaluation of Chromatin Assembly Factor-1 (CAF-1)/p60 as tumour prognostic marker. Int J Mol Sci 13(9):11044–11062. https://doi.org/10.3390/ijms130911044
45. Vieira AF, Schmitt F (2018) An update on breast cancer multigene prognostic tests-emergent clinical biomarkers. Front Med (Lausanne) 5:248. https://doi.org/10.3389/fmed.2018.00248
46. Razek AA, Gaballa G, Denewer A, Nada N (2010) Invasive ductal carcinoma: correlation of apparent diffusion coefficient value with pathological prognostic factors. NMR Biomed 23(6):619–623. https://doi.org/10.1002/nbm.1503
47. Inwald EC, Klinkhammer-Schalke M, Hofstadter F, Zeman F, Koller M, Gerstenhauer M, Ortmann O (2013) Ki-67 is a

prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. Breast Cancer Res Treat 139(2):539–552. https://doi.org/10.1007/s10549-013-2560-8
48. Jatoi I, Hilsenbeck SG, Clark GM, Osborne CK (1999) Significance of axillary lymph node metastasis in primary breast cancer. J Clin Oncol 17(8):2334–2340. https://doi.org/10.1200/JCO.1999.17.8.2334
49. Razek AA, Lattif MA, Denewer A, Farouk O, Nada N (2016) Assessment of axillary lymph nodes in patients with breast cancer with diffusion-weighted MR imaging in combination with routine and dynamic contrast MR imaging. Breast Cancer 23(3):525–532. https://doi.org/10.1007/s12282-015-0598-7
50. Lee JH, Kim SH, Suh YJ, Shim BY, Kim HK (2010) Predictors of axillary lymph node metastases (ALNM) in a Korean population with T1-2 breast carcinoma: triple negative breast cancer has a high incidence of ALNM irrespective of the tumor size. Cancer Res Treat 42(1):30–36. https://doi.org/10.4143/crt.2010.42.1.30
51. Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, Sun K, Tang Z, Jiang H, Li H, Xiong Q, Ding Y, Zhao X, Wang K, Liu Z, Tian J (2019) Radiomics of Multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a Multicenter Study. Clin Cancer Res 25(12):3538–3547. https://doi.org/10.1158/1078-0432.CCR-18-3190
52. Zhang M, Horvat JV, Bernard-Davila B, Marino MA, Leithner D, Ochoa-Albiztegui RE, Helbich TH, Morris EA, Thakur S, Pinker K (2019) Multiparametric MRI model with dynamic contrast-enhanced and diffusion-weighted imaging enables breast cancer diagnosis with high accuracy. J Magn Reson Imaging 49(3):864–874. https://doi.org/10.1002/jmri.26285
53. Abdel Razek AA, Gaballa G, Denewer A, Tawakol I (2010) Diffusion weighted MR imaging of the breast. Acad Radiol 17(3):382–386. https://doi.org/10.1016/j.acra.2009.10.014
54. Abdel Razek AAK, Zaky M, Bayoumi D, Taman S, Abdelwahab K, Alghandour R (2019) Diffusion tensor imaging parameters in differentiation recurrent breast cancer from post-operative changes in patients with breast-conserving surgery. Eur J Radiol 111:76–80. https://doi.org/10.1016/j.ejrad.2018.12.022
55. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM, McCormick F, Dickson RB, Johnson MD, Lippman M, Ethier S, Gazdar A, Gray JW (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10(6):515–527. https://doi.org/10.1016/j.ccr.2006.10.008
56. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10(6):529–541. https://doi.org/10.1016/j.ccr.2006.10.009
57. Wu K, Yang Q, Liu Y, Wu A, Yang Z (2014) Meta-analysis on the association between pathologic complete response and triple-negative breast cancer after neoadjuvant chemotherapy. World J Surg Oncol 12:95. https://doi.org/10.1186/1477-7819-12-95
58. Kennedy RD, Quinn JE, Mullan PB, Johnston PG, Harkin DP (2004) The role of BRCA1 in the cellular response to chemotherapy. J Natl Cancer Inst 96(22):1659–1668. https://doi.org/10.1093/jnci/djh312

Affiliations

Eun-Kyu Kim[1] · Ae Kyung Park[2] · Eunyoung Ko[3] · Woong-Yang Park[4] · Kyung-Min Lee[5] · Dong-Young Noh[5,6] · Wonshik Han[5,6]

1 Department of Surgery, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea

2 College of Pharmacy and Research Institute of Life and Pharmaceutical Sciences, Sunchon National University, Suncheon, Republic of Korea

3 Department of Surgical Oncology, Oncology Center, Sheikh Khalifa Specialty Hospital, Ras Al-Khaimah, United Arab Emirates

4 Department of Molecular Cell Biology, Samsung Medical Center and Samsung Genome Institute, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

5 Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea

6 Department of Surgery, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea