CANCER
COMMUNICATIONS Open Access

Association of an anaplastic lymphoma kinase pathway signature with cell de-differentiation, neoadjuvant chemotherapy response, and recurrence risk in breast cancer

Dingxie Liu[1] | Yong Wu[2,3]

[1]Bluewater Biotech LLC, New Providence, Mandaluyong, NJ 07974, USA

[2]Division of Cancer Research and Training, Department of Internal Medicine, Charles R. Drew University of Medicine and Science, Los Angeles, CA 90059, USA

[3]David Geffen UCLA School of Medicine, and UCLA Jonsson Comprehensive Cancer Center, Los Angeles, CA 90059, USA

Correspondence
Dingxie Liu, Ph.D. Bluewater Biotech LLC, PO Box 1010, New Providence, NJ 07974, USA.
Email: dliu@bluewater-biotech.com
Yong Wu, Ph.D. Division of Cancer Research and Training, Department of Internal Medicine, Charles R. Drew University of Medicine and Science, David Geffen UCLA School of Medicine, and UCLA Jonsson Comprehensive Cancer Center, 1748 E. 118th Street, Los Angeles, CA, 90059, USA.
Email: yongwu@cdrewu.edu

Abstract

Background: Aberrant activation of anaplastic lymphoma kinase (ALK) signaling has been found to be involved in the tumorigenesis of multiple types of cancer. The aim of this study was to determine the role of this pathway in the pathogenesis of breast cancer.

Methods: An ALK pathway signature that we generated previously was used to compute the ALK pathway activity in 6381 breast cancer samples from 42 microarray datasets, and the associations between ALK pathway signature score and clinical variables were examined using logistic regression and survival analyses.

Results: Our results indicated that high ALK pathway activity was a significant risk factor for hormone receptor-negative, high-grade breast cancer in the 42 datasets. ALK pathway activity was positively associated with pathological complete response (pCR) in 15 datasets annotated with patient's neoadjuvant chemotherapy response information (overall odds ratio = 1.67, $P < 0.01$), and this association was more significant in HER2-negative and grade 1&2 tumors than in HER2-positive and grade 3 tumors. ALK pathway activity was also positively associated with recurrence risk in breast cancer patients from 30 datasets annotated with survival information (overall hazard ratio = 1.21, $P < 0.01$), particularly in patients with age > 50 years old, with positive lymph nodes, or with residual disease after neoadjuvant chemotherapy.

Conclusions: ALK may be involved in breast cancer tumorigenesis, and ALK pathway signature score may serve as a prognostic biomarker for breast cancer.

KEYWORDS

Anaplastic Lymphoma Kinase, Breast cancer, Differentiation, Disease free survival, Gene signature, Neoadjuvant chemotherapy response, Prognosis prediction

1 | BACKGROUND

Breast cancer is the most common cancer among women, accounting for approximately a quarter of all new cancer cases diagnosed in women worldwide [1]. Recent genomic studies demonstrated that breast cancers were highly heterogeneous in their molecular biology [2, 3]. In addition to those well-known genetically altered oncogenes and

tumor suppressor genes, including breast cancer susceptibility gene 1 (*BRCA1*), epidermal growth factor receptor (*EGFR*), phosphatidylinositol-3 kinase (*PI3K*), and tumor protein 53 (*TP53*), many lesser-known or lesser-characterized genes in the case of breast cancer, such as anaplastic lymphoma kinase (*ALK*), were found to be mutated or amplified in breast cancer [2, 3].

ALK is a transmembrane tyrosine kinase receptor belonging to the insulin receptor superfamily. Aberrant activation of ALK is involved in the tumorigenesis of a subset of haematopoietic, epithelial, and mesenchymal neoplasms such as anaplastic large cell lymphoma [4, 5], inflammatory myofibroblastic tumors [5], neuroblastoma [5], and lung cancer [6, 7]. The role of ALK signaling in the pathogenesis of breast cancer is not clear.

Conventionally, pathway activation is assessed by methods including immunohistochemistry that detect the levels of pathway-related protein expression or fluorescence *in situ* hybridization and quantitative PCR that detect amplification/overexpression of related oncogenes. The disadvantage of these methods is that they may sometimes be unreliable because most pathways can be activated at multiple points. The latest advances in high-throughput genomic technologies provide alternative strategies for semi-quantifying pathway activity through analyzing the expression profile of a pathway-specific gene signature by using approaches such as Bayesian binary regression (BinReg) that was developed by Nevins' group [8, 9]. We previously generated a gene signature for the ALK pathway based on the difference of gene expression profiles between tumor cells with activated and inactive ALK pathway [10]. In the current study, the ALK pathway activity in breast cancer samples from 42 microarray datasets was computed, and its associations with the clinical outcome were examined to determine the role of this pathway in the pathogenesis of breast cancer.

2 | DATA SOURCES AND METHODS

2.1 | Microarray datasets and patient cohorts

Only microarray datasets from Affymetrix U133 GeneChip microarray platforms (HU133A, HU133A 2.0, and HU133 Plus 2.0) were selected in this study since the ALK pathway signature used in this study was generated from datasets of these platforms. Datasets without patient's neoadjuvant chemotherapy response or cancer recurrence information were excluded. To avoid selection bias, only datasets with more than 50 samples were chosen.

The raw CEL files of datasets were downloaded from Gene Expression Omnibus (GEO), except for dataset MDA133 that was downloaded from the website of Department of Bioinformatics and Computational Biology at MD Anderson Cancer Center (https://bioinformatics.mdanderson.org/public-datasets/). The CEL files were normalized using Robust Multi-array Average (RMA) approaches in R environment, and array quality was assessed by R simpleaffy package [11]. Batch effects of arrays from different hybridization dates were estimated using principal component analysis. The R ComBat function was used to eliminate batch effects when multiple datasets were merged or apparent batch effects were observed in a single dataset [12].

For datasets GSE3494, GSE2990, GSE6532, and GSE7390, the patient survival data were updated, and duplicate samples were removed according to the curated clinical data made by Dr. Jonas Bergh that are available in the GEO database (accession number: GSE83232).

Since different clinical features are partially or fully missing in the 42 individual datasets, we merged these datasets into four cohorts to perform multivariate Cox regression analysis and compare the prognostic power of gene signatures in different subgroups of breast cancer. Cohort 1 was merged from the 15 datasets (MDA133, GSE16446, GSE18728, GSE18864, GSE20194, GSE20271, GSE25055, GSE25065, GSE26639, GSE32646, GSE37946, GSE41998, GSE42822, GSE50948, and GSE66305) that contain patient neoadjuvant chemotherapy response information. Cohort 2 was merged from the 16 datasets (GSE11121, GSE12276, GSE2034, GSE17705, GSE2603, GSE20685, GSE2990, GSE26971, GSE3494, GSE45255, GSE58812, GSE6532, GSE65194, GSE7390, GSE9195, and GSE88770) in which distant metastasis information is available. Cohort 3 was merged from the 8 datasets (GSE12093, GSE20711, GSE21653, GSE31519, GSE42568, GSE1456, GSE7378, and GSE71258) in which (local and distant) cancer recurrence information is available. The remaining 6 datasets (GSE16391, GSE16446, GSE17907, GSE19615, GSE25055, and GSE25065) were merged as Cohort 4 since these datasets contain short follow-up (< 5 years for most patients) information. The Combat program was used to remove the batch effects in all the 4 cohorts. It is worth noting that 3 datasets (GSE16446, GSE25055, and GSE25065) were present in both Cohort 1 and Cohort 4. Therefore, when these 2 cohorts were used for the same statistical analysis, these 3 datasets were removed from Cohort 1.

2.2 | Pathway activity prediction by using BinReg

Using the BinReg approach to generate pathway signatures and predict pathway activities of individual samples has been described in literature [9, 13]. Briefly, the gene expression patterns of two sets of samples (with one pathway being "on" and "off" respectively) were analyzed, and the pathway-specific informative genes (signature genes) were identified. Principal

components were then used to compute the weights for each signature gene, such that the weighted average of expression levels showed a clear ability to distinguish the pathway "on" and "off" groups. By applying binary regression on the principal components to the gene expression dataset of an unknown sample, a probability score of pathway activity for that sample was produced.

2.3 | ALK pathway signature

Two microarray datasets were used to generate ALK pathway signature as we described previously [10]. Briefly, the gene expression data of anaplastic large cell lymphoma cell line TS treated with or without ALK inhibitors A2 or A3 (GSE6184) was used as training set to generate a signature. This signature was then validated using the gene expression data of TS cells with or without knock-down of ALK (GSE6184) and the expression data of lung cancer cell line NCI-H2228 treated with or without ALK inhibitor CH5424802 (GSE2511817). In the present study, to enhance the reliability of this ALK pathway signature, we further validated the signature using lymphoma samples (positive or negative for ALK re-arrangement), obtained from three independent microarray datasets.

2.4 | Statistical analyses

Statistical analyses were performed using R packages including Metafor [14], Survival [15], and Survminer (www.sthda. com/english/rpkgs/survminer). Odds ratio (OR) and 95% confidence interval (CI) for the associations of pathological complete response (pCR) to neoadjuvant chemotherapy with risk score was calculated using univariate and multivariate logistic regression. Kaplan-Meier survival curves with log-rank test and cox proportional hazards regression were used to analyze the association between disease-free survival (DFS) and risk score. DFS was defined as the duration from surgery to the first confirmed recurrence or metastasis. In this study distant metastasis-free survival (DMFS) was preferred to be used in survival analysis. However, when the DMFS data were not available, recurrence-free survival (RFS) data was used. The overall hazard ratio (HR) of a variable of interest was calculated using a random-effects model. The significance of the overall effects across multiple datasets was estimated using the Z test. In analysis of different subtypes of breast cancer, only datasets with event cases > 5, non-event cases > 5, and total cases > 30 were included for regression analysis. In Kaplan-Meier survival analysis, the patients were stratified into three groups based on ALK pathway signature scores ($\geq$ 2/3 percentile as high ALK pathway activity; $\leq$ 1/3 percentile as low activity; < 2/3 percentile and > 1/3

percentile as intermediate activity). For the patient cohorts that were merged from multiple individual datasets, DFS or DMFS was censored at 8 years for Cohort 2 and Cohort 3, and at 5 years for Cohort 4 since the datasets in Cohort 4 contained shorter follow-up data. Cohort 1 was not used in survival analysis as it is not annotated with patient's survival data.

The datasets used in this study were generated from different research groups and contain different clinical variables. The sample size would be too small if all the variables were put together for adjustment in multivariate regression analysis. Therefore, in this study, ALK pathway signature score was tested with adjustment for one variable each time in multivariate regression analysis.

All statistical analyses were two-sided and considered significant when $P < 0.05$.

3 | RESULTS

3.1 | Association of ALK pathway signature score with de-differentiation of breast cancer

Forty-two publicly available datasets were analyzed. Among the 42 microarray datasets, 15 datasets contain patient's neoadjuvant chemotherapy response information and 30 datasets contain cancer recurrence information (3 datasets contain both types of information). The baseline features of breast cancer in these datasets are described in detail in our previous work [16].

We previously generated an ALK pathway signature by using gene expression data of anaplastic large-cell lymphoma cells treated with or without ALK inhibitors as training sets and using expression data of additional cell lines treated with ALK shRNA or ALK inhibitor as validating sets [10]. In the present study, to enhance the reliability of this ALK pathway signature, we further validated the signature using gene expression data from lymphoma samples with or without ALK re-arrangement (Supplementary Fig. 1).

The associations between ALK pathway signature score and seven clinical characteristics of breast cancer were analyzed in 42 datasets individually by using univariate regression analysis. Figure 1 shows that in most of the datasets, ALK pathway signature score was a significant factor of unfavorable prognosis for patients with ER-positive (overall OR = 0.30, 95% CI = 0.26-0.35, $P < 0.01$) and PR-positive tumors (overall OR = 0.45, 95% CI = 0.39-0.51, $P < 0.01$), but was a significant factor of favorable prognosis for patients with grade 3 tumors (overall OR = 2.35, 95% CI = 2.01-2.75, $P < 0.01$). By contrast, no apparent associations of ALK pathway signature score with age, HER2 status, lymph node status, and tumor size were observed in most of the datasets (Figure 2).
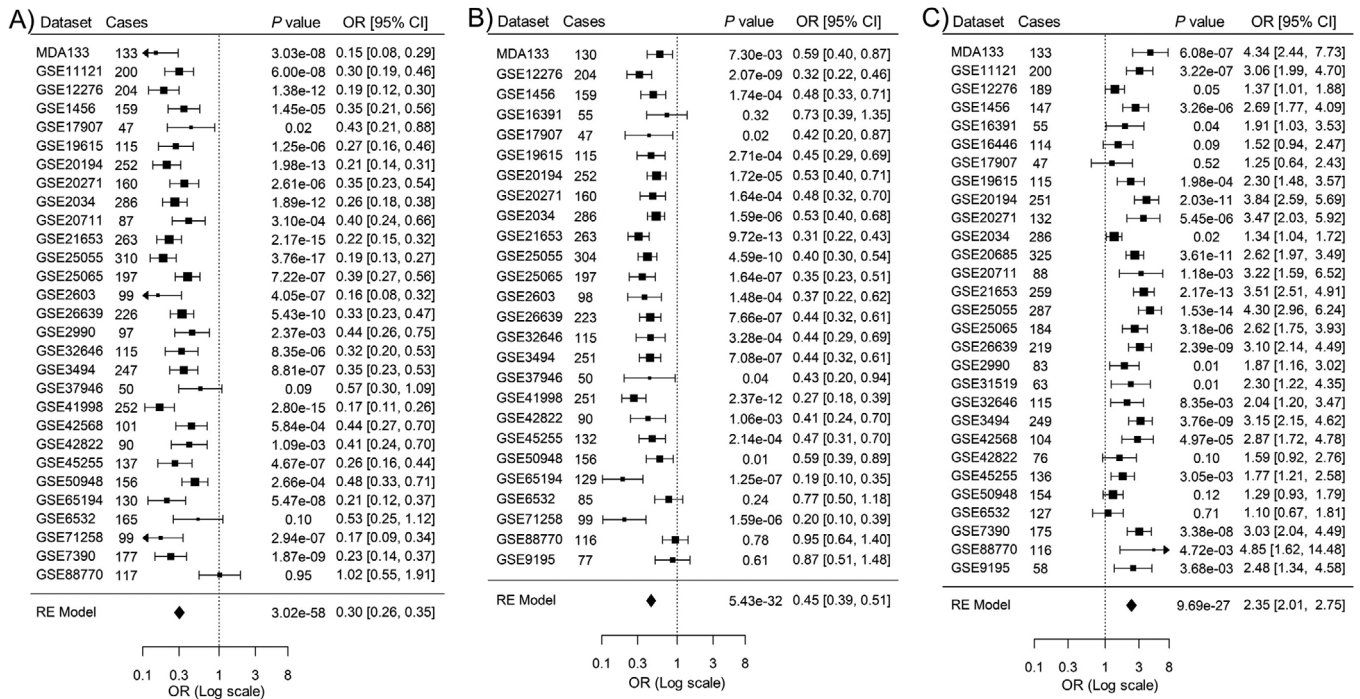
FIGURE 1 Associations of anaplastic lymphoma kinase (ALK) pathway signature score with ER/PR statuses and tumor grade in 42 individual breast cancer datasets. A) Estrogen receptor (ER) status (ER positive as event and ER negative as non-event); B) Progesterone receptor (PR) status (PR positive as event and PR negative as non-event); C) Tumor grade (grade 3 as event and grade 1&2 as non-event). The odds ratio (OR) (per one standard deviation [SD] increment) with the increase of ALK pathway signature score (used as continuous variable) was analyzed using univariate logistic regression in 42 individual breast cancer datasets. The overall effect was calculated through a random-effects (RE) model. The ORs were shown in forest plots, the squares and horizontal lines represent the OR and 95% Confidence Interval (CI) for individual cohorts, while the diamonds represent the OR and 95% CI for the overall estimate. CI was clipped to arrow in forest plot when it exceeded specified limits. Only datasets with event cases > 5, non-event cases > 5, and total cases > 30 were included for regression analysis, therefore different numbers of datasets may be analyzed for different variables

The 42 datasets were further merged into four cohorts, and similar associations of ALK pathway signature score with the seven clinical characteristics were observed in the four merged cohorts (Supplementary Fig. 2). Multivariate logistic regression was also performed on these merged cohorts to assess whether the associations of ALK pathway signature score with ER status, PR status, and tumor grade remained significant after adjustment for other covariates. Overall, the ORs for the association of ALK pathway signature score with ER status stayed constant when the pathway signature score was tested separately with adjustment for age, tumor grade, and tumor size, and statuses of PR, HER2, or lymph node (Figure 3A). Similar trend was also observed for the associations of the ALK pathway signature score with PR status (Figure 3B) and tumor grade (Figure 3C), except that the association with PR status was almost abolished when adjusted for ER status. These results indicate that the ALK pathway signature score was an independent risk factor for ER-negative and high-grade breast cancer, signifying that the ALK pathway signature score was an independent factor associated with the de-differentiation of breast cancer since the loss of ER expression and high patholog-ical grade were features of de-differentiated breast cancer [17].

3.2 | Association of ALK pathway signature score with pCR of breast cancer

Logistic regression analysis showed that ALK pathway signature score was positively associated with pCR rate of breast cancer in Cohort 1 which contains 15 individual datasets with patient's neoadjuvant chemotherapy response information (overall OR = 1.67, 95% CI = 1.45-1.93, $P < 0.01$) (Figure 4A). This association reached statistical significance ($P < 0.05$) in 8 of the 15 datasets, and reached a statistical level at $P = 0.08$ in 2 of the datasets (Figure 4A). As to the remaining 5 datasets with $P > 0.08$, two (GSE16446 and GSE18864) contain only ER-negative breast cancer samples, two (GSE37946 and GSE50948) are mainly composed of HER2-positive samples, and one (GSE25065) contains higher percentage of triple-negative breast cancer (33%) than the other datasets. These data suggest that ER and HER2 statuses may affect the association between ALK pathway
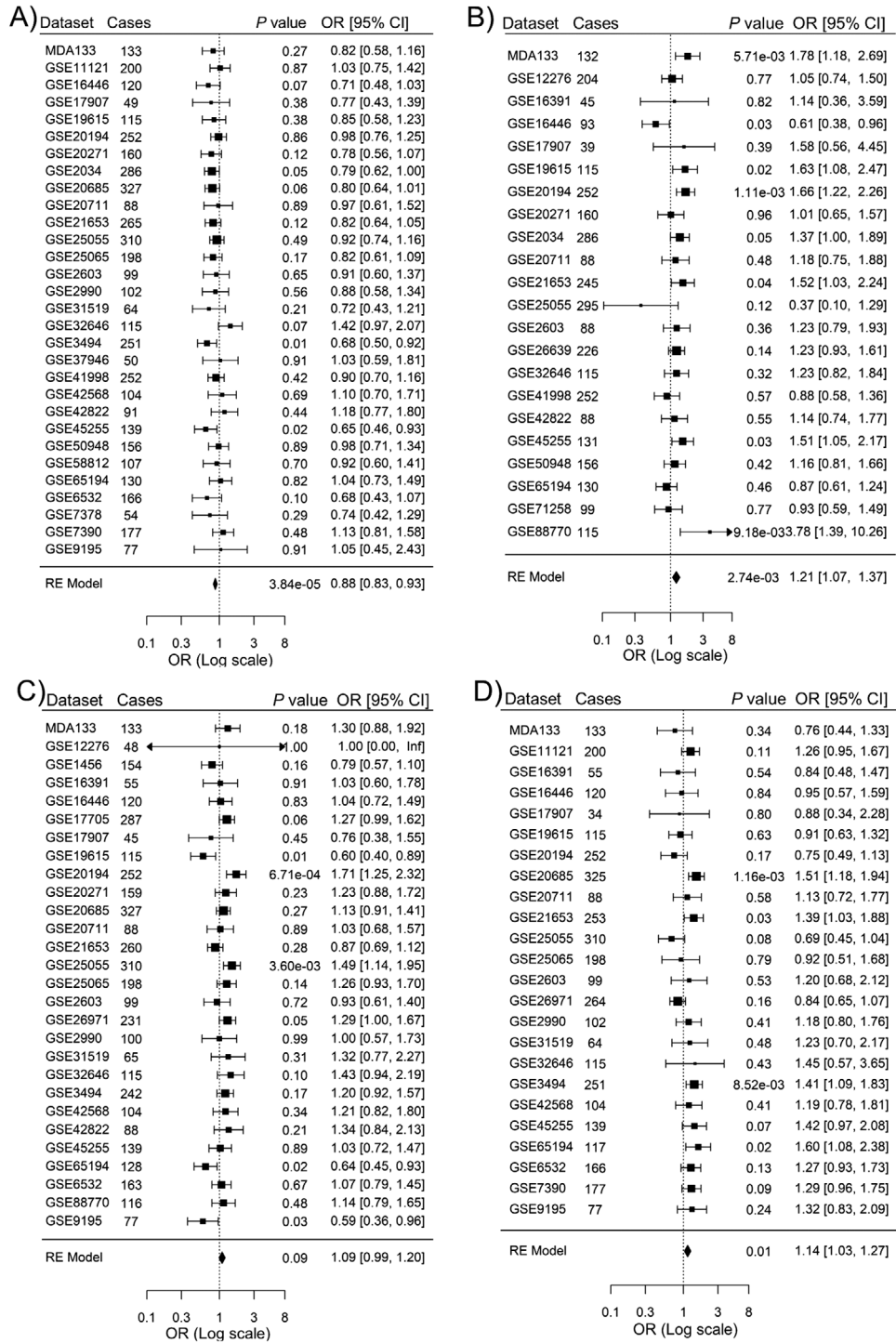
FIGURE 2 Associations of ALK pathway signature score with age, tumor size, and statuses of human epidermal growth factor receptor 2 (HER2) and node in 42 individual breast cancer datasets. A) Patient age (> 50 years as event and ≤50 years as non-event); B) HER2 status (HER2 positive as event and HER2 negative as non-event); C) Lymph node status (node positive as event and node negative as non-event); D) Tumor size (> 2 cm as event and ≤20 cm as non-event). The OR (per one standard deviation increment) with the increase of ALK pathway signature score (used as continuous variable) was analyzed using univariate logistic regression in 42 individual breast cancer datasets. Only datasets with event cases > 5, non-event cases > 5, and total cases > 30 were included for regression analysis
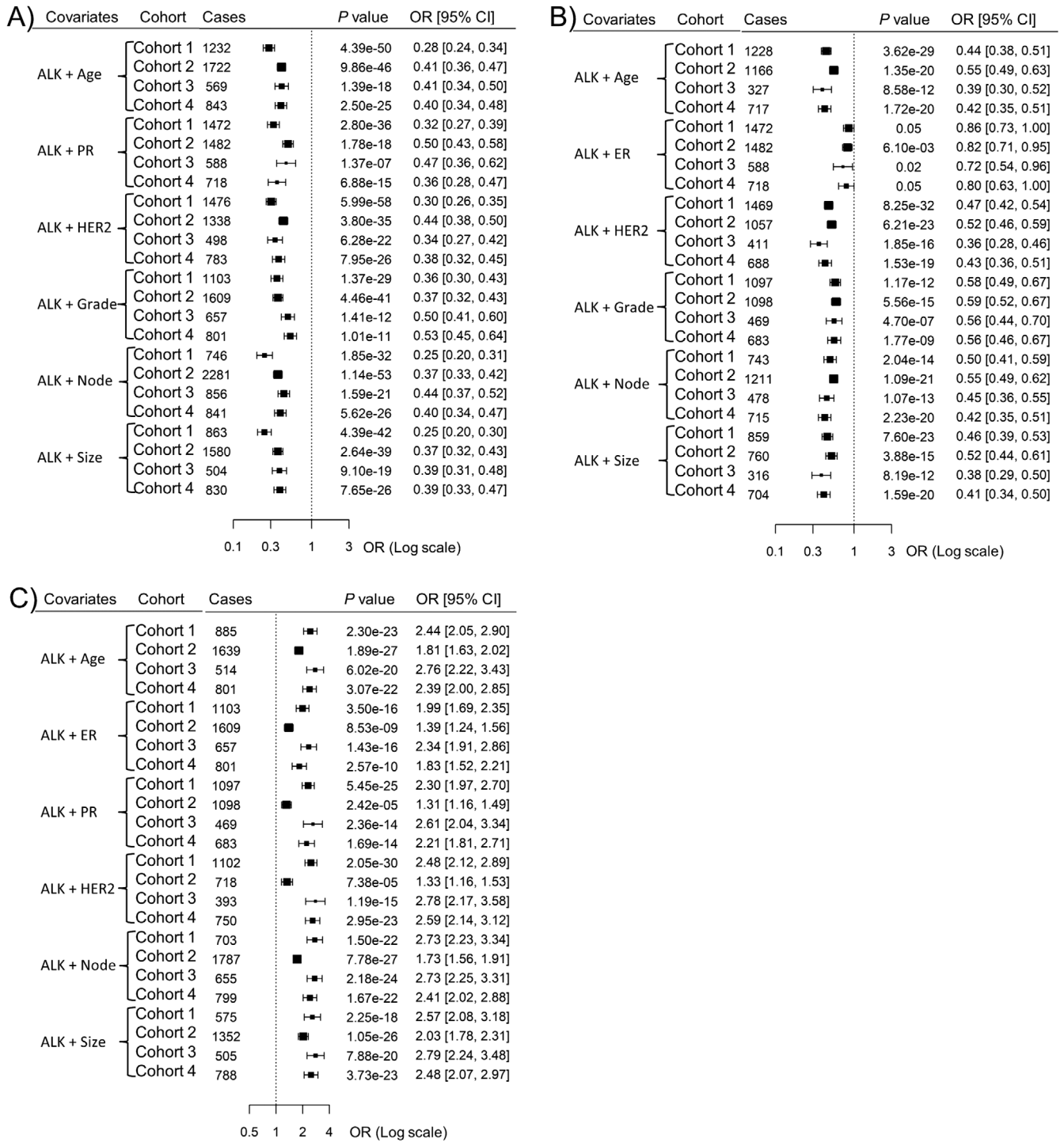
FIGURE 3 Associations of ALK pathway signature score with tumor grade and statuses of ER and PR in four merged breast cancer cohorts. A) ER status (ER positive as event and ER negative as non-event); B) PR status (PR positive as event and PR negative as non-event); C) Tumor grade (grade 3 as event and grade 1&2 as non-event). ALK pathway signature score was used as continuous variable and tested with adjustment for one variable each time in multivariate regression analysis. The OR [95% CI] is presented per one-SD increment. The four cohorts were merged from the 42 individual breast cancer datasets as described in Material and Methods. The three datasets (GSE16446, GSE25055, and GSE25065) that were present in both Cohort 1 and Cohort 4 were removed from Cohort 1 in multivariate regression analysis
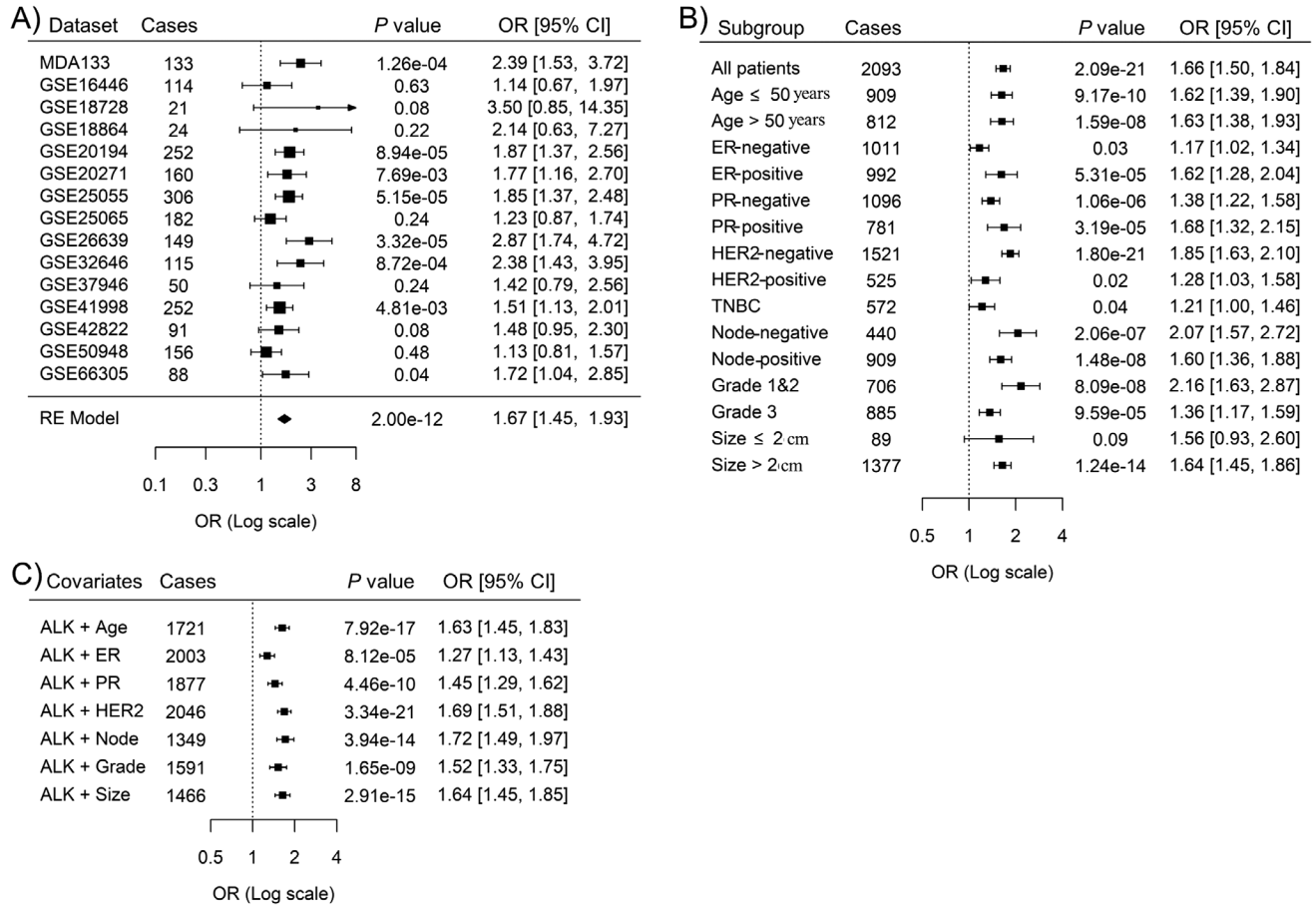
FIGURE 4 Association of ALK pathway signature score with pathological complete response (pCR) in Cohort 1 of breast cancer patients. Fifteen individual breast cancer datasets with patient's neoadjuvant chemotherapy response information were merged into Cohort 1. ALK pathway signature score was used as continuous variable, and the OR [95% CI] is presented per one-SD increment. A) Univariate logistic regression analysis of the association of ALK pathway signature score with pCR in 15 individual breast cancer datasets. B) Comparison of the association of ALK pathway signature score with pCR in different subgroups of breast cancer patients. The Combat program was used to remove the batch effects in the cohort. Logistic regression analysis was performed in different subgroups of breast cancer patients as indicated. C) Multivariate logistic regression analysis of the association of ALK pathway signature score with pCR in Cohort 1. ALK pathway signature score was tested with adjustment for one variable each time in multivariate regression analysis

signature score and pCR rate. This effect was also observed in Cohort 1 which was merged from the 15 datasets. As shown in Figure 4B, an increase of one standard deviation (SD) of ALK pathway signature score was associated with 62% and 85% increase of pCR rate in ER-positive (OR = 1.62, 95% CI = 1.28-2.04, $P < 0.01$) and HER2-negative (OR = 1.85, 95% CI = 1.63-2.10, $P < 0.01$) breast cancer respectively, but only with 17% and 28% increase in ER-negative (OR = 1.17, 95% CI = 1.02-1.34, $P = 0.03$) and HER-positive (OR = 1.28, 95% CI = 1.03-1.58, $P = 0.02$) breast cancer respectively. In addition, ALK pathway signature score was also more strongly associated with pCR rate in grade 1&2 tumors (OR = 2.16, 95% CI = 1.63-2.87, $P < 0.01$) than in grade 3 tumors (OR = 1.36, 95% CI = 1.17-1.59, $P < 0.01$).

Multivariate logistic regression analysis on Cohort 1 showed that the association of ALK pathway signature score

with pCR was not affected after adjustment for other clinical variables (Figure 4C).

3.3 | Association of ALK pathway signature score with recurrence risk of breast cancer

Logistic regression analysis showed that ALK pathway signature score was positively associated with recurrence risk of breast cancer in 30 individual datasets in which patients' survival information is available (overall HR = 1.21, 95% CI = 1.13-1.29, $P < 0.01$) (Figure 5A). Among the 30 datasets, GSE16446, GSE31519, and GSE58812 contain only ER-negative samples, and GSE17907 is mainly composed of Her2-positive samples (Supplementary Table 1). ALK pathway signature score failed to achieve significant positive association with recurrence risk in these 4 datasets. Four
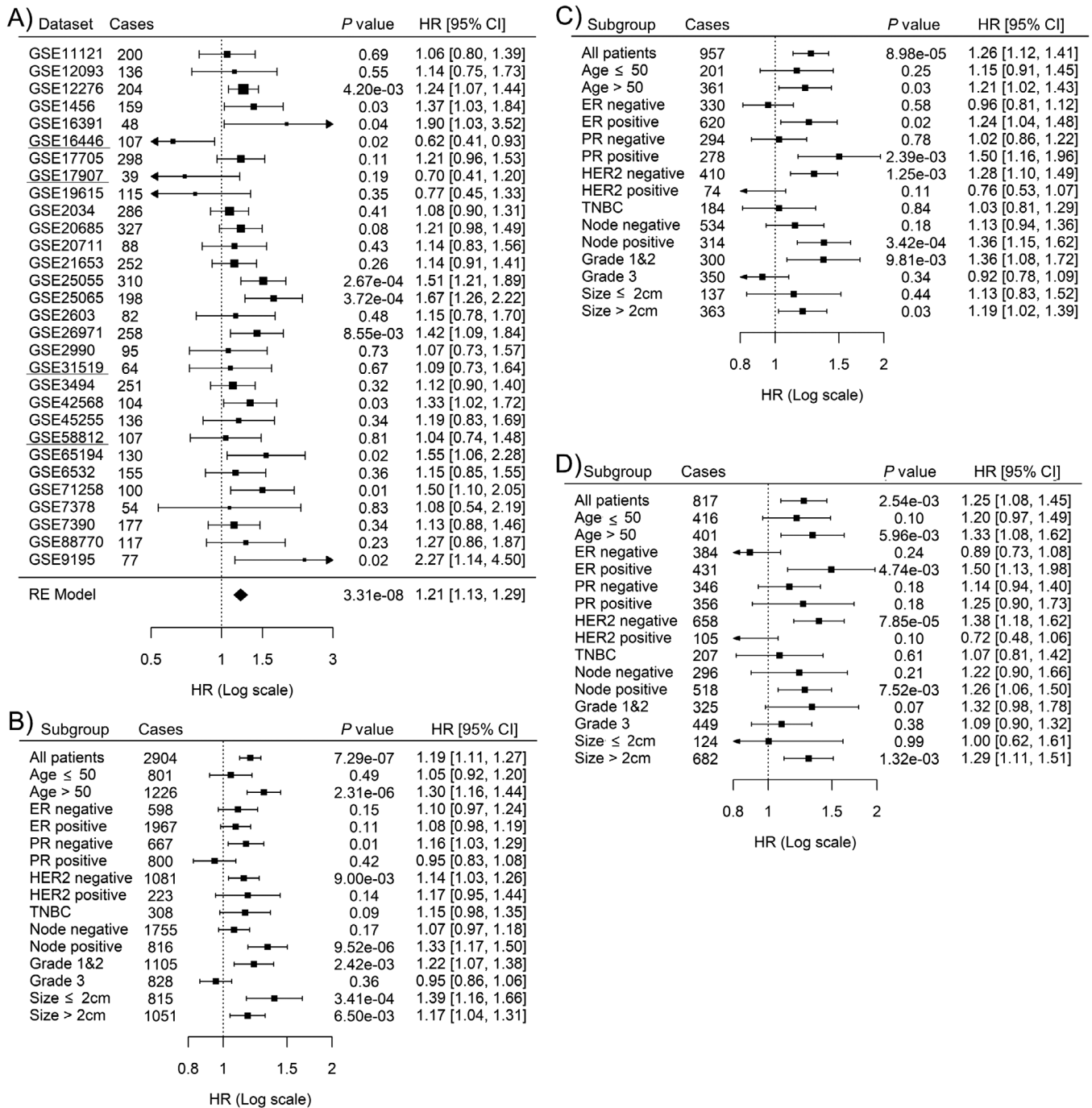
FIGURE 5 Association of ALK pathway signature score with recurrence risk in breast cancer patients. The association of ALK pathway signature score with recurrence risk was analyzed by using univariate logistic regression in 30 individual breast cancer datasets with patient's survival information (A), and were also compared among different subgroups of breast cancer patients in Cohort 2 (B), Cohort 3 (C) and Cohort 4 (D) that were merged from the 30 datasets. The details of these 3 breast cancer cohorts are described in Material and Methods. The four datasets that mainly consist of ER-negative or HER2-positive breast cancers were underlined. ALK pathway signature score was used as continuous variable, and recurrence risk is indicated by hazard ratio (HR) per one-SD increment

additional datasets (GSE20711, GSE2603, GSE2990, and GSE7378) have sample size < 100. Among the remaining 22 datasets, the association of ALK pathway signature score with recurrence risk reached statistical significance in 11 datasets (Figure 5A).

The positive association of ALK pathway signature score with recurrence risk remained significant in Cohorts 2-4 that were merged from the 30 datasets in Cox regression analysis (overall $P < 0.01$) (Figure 5B-D) and Kaplan-Meier survival curve analysis (overall $P < 0.01$) (Figure 6A-C).
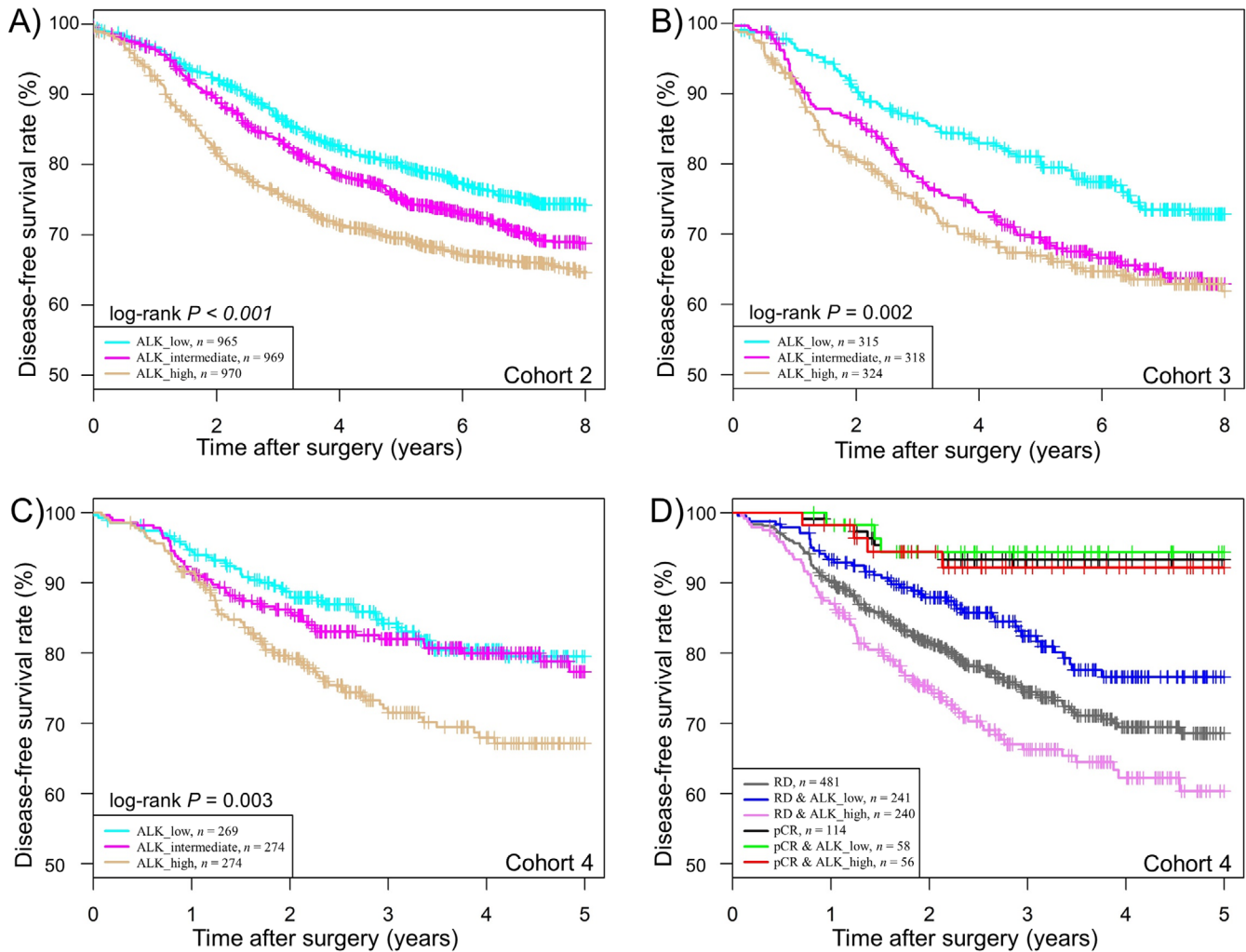
FIGURE 6 Kaplan-Meier survival analysis of the association of ALK pathway signature score with recurrence risk of breast cancer. To compare the association of ALK pathway signature score with recurrence risk in breast cancer patients with pCR or residual disease (RD) after neoadjuvant chemotherapy, the samples in Cohort 4 with both pCR and survival information were analyzed. The patients were stratified into three groups based on ALK pathway signature score ($\geq$ 2/3 percentile as high ALK pathway activity;$\leq$ 1/3 percentile as low activity; < 2/3 percentile and > 1/3 percentile as intermediate activity) in Cohort 2 (A), Cohort 3 (B), and Cohort 4 (C). (D), The patients with pCR and RD were further stratified into two subgroups based on the medium value of ALK pathway signature score, respectively

Like the association with pCR, the association of ALK pathway signature score with recurrence risk was significant in patients with age > 50 years, HER2-negative tumors, grade 1&2 tumors, or positive lymph nodes in the three cohorts (all $P < 0.05$; Figure 5B-D, Supplementary Fig. 3).

3.4 | Association of ALK pathway signature score with recurrence risk only in patients with residual disease (RD) following neoadjuvant chemotherapy

Our data showed that ALK pathway signature score was positively associated with both pCR rate and recurrence risk of

breast cancer. These results seem to be controversial since breast cancer patients with pCR generally have better prognosis than patients without pCR (i.e., patients with RD) [18]. A potential explanation is that ALK pathway signature score was associated with recurrence risk only in RD but not in pCR patients. To test this hypothesis, samples in Cohort 4 with both neoadjuvant chemotherapy response and survival information were tested. Kaplan-Meier analysis showed that patients with pCR had significantly lower recurrence rate than patients with RD (5-year DFS rate: 93% vs. 69%, $P < 0.01$) (Figure 6D). When RD patients were further stratified into two subgroups based on ALK pathway signature scores, the subgroup with low ALK pathway signature score had higher 5-year DFS rate than that with high ALK pathway signature score (77% vs.

60%, $P < 0.01$) (Figure 6D). However, the 5-year DFS rates of pCR patients with low and high ALK pathway signature scores were similar ($P = 0.67$) (Figure 6D).

4 | DISCUSSION

The role of the ALK pathway in the pathogenesis of breast cancer is not clear. In the present study, we found that high ALK pathway activity was associated with high pathological grade and loss of ER and PR expression in breast cancer patients. We also found that ALK pathway activity was positively associated with neoadjuvant chemotherapy response and recurrence of breast cancer.

It is controversial whether alterations of *ALK* gene exist in breast cancer. Although Perez-Pinera *et al*. [19] reported that activated ALK was strongly expressed in different histological subtypes of breast cancer, Fukuyoshi *et al*. [20] and Lerebours *et al*. [21] failed to observe such abnormality in breast cancer patients. Recently, Kim *et al*. [22] demonstrated that ALK copy number was significantly increased in inflammatory breast cancer (IBC) and was associated with a high recurrence risk in IBC patients. In addition, The Cancer Genome Atlas Network (TCGA) genomic analysis showed amplification of *ALK* gene in 43 of 476 breast cancer samples [3]. These results, together with our findings that higher ALK pathway signature score was associated with cell de-differentiation and higher recurrence risk in breast cancer, imply that active ALK signaling has a functional role in the pathogenesis of breast cancer.

ALK activates multiple pathways, including the phospholipase C γ (PLCγ), Janus kinase/signal transducer and activator of transcription 3 (JAK/STAT3), phosphoinositide 3-kinase/AKT (PI3K/AKT), mitogen-activated protein kinase (MAPK), sonic hedgehog, Jun B Proto-Oncogene (JUNB), and Rap guanine nucleotide exchange factor 1/Ras-related protein 1 (RAPGEF1/RAP1) signaling cascades [5]. Aberrant activation of some of these pathways, such as JAK/STAT3 [23], PI3K/AKT [24], and sonic hedgehog [25], has been found to promote tumor metastasis and be associated with cell de-differentiation in breast cancer. The downstream pathways mediating the oncogenic effects of ALK signaling in breast cancer needs further clarification.

An interesting finding in the present study is that the association between high ALK pathway signature score and tumor recurrence was only observed in patients of > 50 years old, not in those of ≤50 years old. This characteristic of ALK has not been reported in other cancer studies. The reason for this is unclear. It might be due to the complex hormonal environment of young women. A retrospective analysis performed on 260 elderly and 294 middle-aged patients with primary breast cancer showed that negative lymph node status, small tumor size, and positive ER status were favorable indicators of survival in both the elderly and the middle-aged patients [26]. In addition, another study showed that ER/PR status and HER2 gene amplification or overexpression were prognostic factors in elderly patients with breast cancer [27]. ALK, ER/PR, and HER2 share common downstream activation pathways, such as Ras/MAPK and PI3K/Akt, which ultimately lead to increased transcription, cell proliferation, growth, and survival [28–32]. This also may explain that they have similar prognostic effects on patients of these age groups.

Another interesting finding in the present study is that the relationship between ALK and tumor recurrence existed in patients with lymph node metastasis, but not in node-negative patients. A potential explanation is that the growth of breast cancer with a high metastatic capability (such as breast cancer with positive lymph nodes) will benefit more from active ALK signaling than that with a low metastatic capability (such as breast cancer with negative lymph nodes) [1]. To establish a distant tumor, a metastatic cancer cell needs to survive and adapt to a new environment, rebuild cell-matrix interactions, form a micrometastasis, and finally restart an unlimited growth process. Metabolic rewiring is a key factor for tumor cells to complete this metastatic cascade, and several pathways including transforming growth factor beta (TGF-β) and hypoxic signaling are involved in this rewiring process [33]. Recent studies revealed that the up-regulation of hypoxia-inducible factors under hypoxic conditions and the induction of vascular endothelial growth factor (VEGF) secretion induced by TGF-β were both in an ALK-dependent manner in different tumor cells [34, 35], suggesting a potential connection between ALK signaling and metabolic rewiring. Probably through this connection, ALK signaling plays more active roles in the molecular pathogenesis of metastatic cells than in that of primary tumor cells which may not benefit from metabolic rewiring.

It is well-known that de-differentiated (or high pathological grade) cancer cells are more sensitive to chemotherapy than well-differentiated cells [36]. In addition, our data showed that association of the ALK pathway with cancer recurrence was only observed in grade 1&2, but not in grade 3 tumors (Figure 5). Therefore, it is not surprising to find in the present study that ALK pathway signature score was associated with recurrence risk only in patients with RD, but not in pCR patients (Figure 6D). Based on these data, we suggest that neoadjuvant chemotherapy could be beneficial for grade 1&2 breast cancers with high ALK pathway activity and for grade 3 tumors, but not for grade 1&2 tumors with low ALK pathway activity.

A number of prognostic gene signatures have been developed for prediction of neoadjuvant chemotherapy response or recurrence risk in breast cancer [37–43]. In the present study, the ALK pathway gene signature was found to well predict both the neoadjuvant chemotherapy response and recurrence risk in multiple datasets encompassing > 5000 breast can-

cer cases, suggesting it as a potentially promising biomarker for breast cancer prognosis and management. Compared with those gene signatures reported previously, the unique feature of ALK pathway signature is that it was specifically associated with recurrence in breast cancer patients with age > 50 years, with lymph node metastasis, or with RD after neoadjuvant chemotherapy, indicating that this signature may be particularly used in risk estimation for these patients. Further validation of the ALK pathway signature in additional independent and clinically relevant analyses will be necessary before entering clinical trials.

The major limitation of the present study was that the ALK pathway data obtained by BinReg approach cannot clarify how the ALK pathway is activated in breast cancer. The ALK pathway can be over-activated through multiple mechanisms in cancer cells, such as re-arrangement [6], mutation [44], and overexpression [45] of ALK and constitute activation of PTN/RPTPβ/ζ signaling [46, 47]. Further studies are needed to clarify which mechanism mediated the over-activation of the ALK pathway in breast cancer.

5 | CONCLUSIONS

In summary, the present study highlights that ALK pathway gene signature represents a potentially promising biomarker for guiding clinical management of breast cancer. Our results also support the notion that ALK signaling may have an oncogenic role in the pathogenesis of breast cancer and therefore may be a potential molecular target for breast cancer therapy.

DECLARATIONS
ACKNOWLEDGMENTS

AUTHORS' CONTRIBUTIONS

DXL and YW contributed to the conception and design of the study and drafted the manuscript; DXL contributed to the analysis and interpretation of data, and revised the manuscript. YW participated in language expression and polishing. All authors had roles in the manuscript drafting. All authors read and approved the final manuscript.

FUNDING

AVAILABILITY OF DATA AND MATERIALS

The Affymetrix microarray datasets supporting this bioinformatics analysis were downloaded from online databases that have been cited in our manuscript. The processed data are available from the corresponding author upon request.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

COMPETING INTERESTS

Liu has equity interest in Bluewater Biotech LLC.

ORCID

Yong Wu https://orcid.org/0000-0002-1698-0897

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69:7-34.
2. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012;486:400-4.
3. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61-70.
4. Morris SW, Kirstein MN, Valentine MB, Dittmer KG, Shapiro DN, Saltman DL, et al. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. Science. 1994;263:1281-4.
5. Della Corte CM, Viscardi G, Di Liello R, Fasano M, Martinelli E, Troiani T, et al. Role and targeting of anaplastic lymphoma kinase in cancer. Mol Cancer. 2018;17:30.
6. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448:561-6.
7. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell. 2007;131:1190-203.
8. 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;439:353-7.
9. Gatza ML, Lucas JE, Barry WT, Kim JW, Wang Q, Crawford MD, et al. A pathway-based classification of human breast cancer. Proc Natl Acad Sci USA. 2010;107:6994-9.
10. Liu D, Liu X, Xing M. Activities of multiple cancer-related pathways are associated with BRAF mutation and predict the resistance to BRAF/MEK inhibitors in melanoma cells. Cell Cycle. 2014;13:208-19.
11. Wilson CL, Miller CJ. Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. Bioinformatics. 2005;21:3683-5.
12. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118-27.

13. Liu D. Concomitant dysregulation of the estrogen receptor and BRAF/MEK signaling pathways is common in colorectal cancer and predicts a worse prognosis. Cell Oncol. 2019;42:197-209.

14. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software. 2010;36:1-48.

15. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag; 2000. https://www.springer.com/us/book/9780387987842. Accessed 17 Mar 2019.

16. Liu D. AR pathway activity correlates with AR expression in a HER2-dependent manner and serves as a better prognostic factor in breast cancer. Cell Oncol. 2020;43:321-33.

17. Fisher ER, Redmond CK, Liu H, Rockette H, Fisher B. Correlation of estrogen receptor and pathologic characteristics of invasive breast cancer. Cancer. 1980;45:349-53.

18. Bear HD, Anderson S, Brown A, Smith R, Mamounas EP, Fisher B, et al. The effect on tumor response of adding sequential preoperative docetaxel to preoperative doxorubicin and cyclophosphamide: preliminary results from National Surgical Adjuvant Breast and Bowel Project Protocol B-27. J Clin Oncol. 2003;21:4165-74.

19. Perez-Pinera P, Chang Y, Astudillo A, Mortimer J, Deuel TF. Anaplastic lymphoma kinase is expressed in different subtypes of human breast cancer. Biochem Biophys Res Commun. 2007;358:399-403.

20. Fukuyoshi Y, Inoue H, Kita Y, Utsunomiya T, Ishida T, Mori M. EML4-ALK fusion transcript is not found in gastrointestinal and breast cancers. Br J Cancer. 2008;98:1536-9.

21. Lerebours F, Callens C, Vacher S, Hatem R, Guinebretière J-M, Bièche I. Rare overexpression of anaplastic lymphoma kinase gene in inflammatory and non-inflammatory breast cancer. Eur J Cancer. 2013;49:2774-6.

22. Kim MH, Lee S, Koo JS, Jung KH, Park IH, Jeong J, et al. Anaplastic lymphoma kinase gene copy number gain in inflammatory breast cancer (IBC): prevalence, clinicopathologic features and prognostic implication. PLoS One. 2015;10:e0120320.

23. Banerjee K, Resat H. Constitutive activation of STAT3 in breast cancer cells: A review. Int J Cancer. 2016;138:2570-8.

24. Kenna MM, McGarrigle S, Pidgeon GP. The next generation of PI3K-Akt-mTOR pathway inhibitors in breast cancer cohorts. Biochim Biophys Acta Rev Cancer. 2018;1870:185-97.

25. Riobo-Del Galdo NA, Lara Montero Á, Wertheimer EV. Role of Hedgehog Signaling in Breast Cancer: Pathogenesis and Therapeutics. Cells. 2019;8.

26. Alberts AS, Falkson G, van der Merwe R. Factors influencing prognosis in elderly patients with primary breast cancer. S Afr J Surg. 1991;29:8-10.

27. Cappellani A, Vita MD, Zanghì A, Cavallaro A, Piccolo G, Majorana M, et al. Prognostic factors in elderly patients with breast cancer. BMC Surg. 2013;13 (Suppl 2):S2.

28. Hrustanovic G, Bivona TG. RAS-MAPK in ALK targeted therapy resistance. Cell Cycle. 2015;14:3661-2.

29. Moore NF, Azarova AM, Bhatnagar N, Ross KN, Drake LE, Frumm S, et al. Molecular rationale for the use of PI3K/AKT/mTOR pathway inhibitors in combination with crizotinib in ALK-mutated neuroblastoma. Oncotarget. 2014;5:8737-49.

30. Ciruelos Gil EM. Targeting the PI3K/AKT/mTOR pathway in estrogen receptor-positive breast cancer. Cancer Treat Rev. 2014;40:862-71.

31. Díaz-Serrano A, Angulo B, Dominguez C, Pazo-Cid R, Salud A, Jiménez-Fonseca P, et al. Genomic Profiling of HER2-Positive Gastric Cancer: PI3K/Akt/mTOR Pathway as Predictor of Outcomes in HER2-Positive Advanced Gastric Cancer Treated with Trastuzumab. Oncologist. 2018;23:1092-102.

32. Paplomata E, O'Regan R. The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. Ther Adv Med Oncol. 2014;6:154-66.

33. Elia I, Doglioni G, Fendt S-M. Metabolic Hallmarks of Metastasis Formation. Trends Cell Biol. 2018;28:673-84.

34. Martinengo C, Poggio T, Menotti M, Scalzo MS, Mastini C, Ambrogio C, et al. ALK-dependent control of hypoxia-inducible factors mediates tumor growth and metastasis. Cancer Res. 2014;74:6094-106.

35. Seystahl K, Tritschler I, Szabo E, Tabatabai G, Weller M. Differential regulation of TGF-β-induced, ALK-5-mediated VEGF release by SMAD2/3 versus SMAD1/5/8 signaling in glioblastoma. Neuro-oncology. 2015;17:254-65.

36. Greco FA, Hainsworth JD. Poorly differentiated carcinoma or adenocarcinoma of unknown primary site: long-term results with cisplatin-based chemotherapy. Semin Oncol. 1994;21(5 Suppl) 12:77-82.

37. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415:530-6.

38. 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. New England Journal of Medicine. 2004;351:2817-26.

39. 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. J Natl Cancer Inst. 2006;98:262-72.

40. Ma X-J, 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;14:2601-8.

41. Bild AH, Parker JS, Gustafson AM, Acharya CR, Hoadley KA, Anders C, et al. An integration of complementary strategies for gene-expression analysis to reveal novel therapeutic opportunities for breast cancer. Breast Cancer Res. 2009;11:R55.

42. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27:1160-7.

43. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res. 2011;17:6012-20.

44. Mossé YP, Laudenslager M, Longo L, Cole KA, Wood A, Attiyeh EF, et al. Identification of ALK as a major familial neuroblastoma predisposition gene. Nature. 2008;455:930-5.

45. Carén H, Abel F, Kogner P, Martinsson T. High incidence of DNA mutations and gene amplifications of the ALK gene in advanced sporadic neuroblastoma tumours. Biochem J. 2008;416:153-9.

46. Perez-Pinera P, Zhang W, Chang Y, Vega JA, Deuel TF. Anaplastic lymphoma kinase is activated through the pleiotrophin/receptor protein-tyrosine phosphatase beta/zeta signaling pathway: an alternative mechanism of receptor tyrosine kinase activation. J Biol Chem. 2007;282:28683-90.

47. Perez-Pinera P, Chang Y, Astudillo A, Mortimer J, Deuel TF. Anaplastic Lymphoma Kinase is Expressed in Different Subtypes of Human Breast Cancer. Biochem Biophys Res Commun. 2007;358:399.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.