

Prognostic Predictors Do Not Equate to Predictors of Chemotherapy Benefit in HR-Positive, HER2-Negative Breast Cancer: A Population-Based Study

Jianfei Fu

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

Haiping Lin

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

Hongjuan Zheng

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

Chenyang Ge

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

LiXian Ling

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

RuiHua Ying

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

Qinghua Wang

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

Wanfen Tang

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

Xia Zhang

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

Shishi Zhou

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

Xiayun Jin

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

Xifeng Xu

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital

Shian Yu (✉ zjyhysa@163.com)

Jinhua Hospital of Zhejiang University: Jinhua Municipal Central Hospital <https://orcid.org/0000-0002-6326-7410>

Keywords: Breast cancer, Hormone receptor positive, HER2-negative, Adjuvant chemotherapy, Gene signature

Posted Date: October 6th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-78025/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

Background: Patients with hormone receptor positive (HR+), human epidermal growth factor receptor 2 negative (HER2-), early breast cancer had a favorable prognosis. We conducted a study to identify patients benefiting from adjuvant chemotherapy in this cohort.

Methods: Patients with HR+, HER2-, early breast cancer were identified from the SEER database and were classified into the chemotherapy and non-chemotherapy groups. A propensity score matching (PSM) was performed and subgroup analyses were conducted in the after-matched patients to explore which subgroups benefited from chemotherapy. Gene expression RNA-seq data and phenotypic data of HR+, HER2-, early breast cancer were identified from the TCGA database, differentially expressed genes (DEGs) from paired tumor and normal samples were identified, and DGEs were filtered through interactions between chemotherapy and gene expression level to obtain genes associated with chemotherapy benefits. A chemotherapy predictive clinical score (CPCS) and a chemotherapy predictive genetic score (CPGS) were established respectively based on variables or genes identified.

Results: In total, 86158 patients with HR+, HER2-, early breast cancer were identified from the SEER database and there were 15259 patients in each after-matched group. Race, T stage and N stage were associated with chemotherapy benefits and entered into the establishment of CPCS. Amongst the 29 normal samples and 254 tumor samples of HR+, HER2-, early breast cancer in the TCGA database, 29 pairs of samples were identified to filter 1709 DEGs, of which 84 DGEs were associated with chemotherapy benefits. SOWAHA\RP11-205M3.3\IRX6\PPBP and EMX1 were chosen to establish the CPGS according to their P-values for interaction.

Conclusions: Nodal status was an important indicator for chemotherapy decisions in HR+, HER2-, early breast cancer and most patients without nodes involved could be spared chemotherapy. In addition, a CPCS and a CPGS were established respectively using a completely new method to identify patients benefiting from chemotherapy.

Background

Breast cancer remains the most common cancer in female, while prognosis of certain patients with breast cancer is favorable due to adjuvant therapy, such as adjuvant chemotherapy[1]. With deeper understanding of biological and pathological characteristics of breast cancer, there has been a trend towards de-escalating of adjuvant chemotherapy for breast cancer in recent years[2]. For example, the PLANB trial suggested that the TC6 regimen was comparable with the EC-T regimen in disease free survival (DFS) for patients with high-risk HER2- breast cancer, introducing the "non-anthracycline regimen" of adjuvant chemotherapy in breast cancer[3].

Studies regarding endocrine therapy, such as the SOFT study and the TEXT study, emphasized the effectiveness of endocrine therapy in early breast cancer, which also has lowered the position of adjuvant chemotherapy[4–7]. Studies about neoadjuvant therapy have proved that hormone receptor positive

(HR+), HER2- breast cancer, even early breast cancer, was insensitive to neoadjuvant chemotherapy, and the pathologic complete response (pCR) rate was lower than 20%, which was much lower than the rates of HER2 + breast cancer or triple negative breast cancer[8].

When choosing patients who can be safely spared the use of adjuvant chemotherapy, clinicopathological characteristics and genetic biomarkers are often taken into consideration. From the clinicopathological perspective, risk factors, such as the tumor size and nodal statuses, were indicators for chemotherapy in previous studies. On the other hand, clinicians often face dilemma when making clinical decisions about chemotherapy, especially for patients with 1–3 positive nodes.

Genetic biomarkers play a role in this situation. According to the American Society of Clinical Oncology (ASCO) Clinical Practice Guideline, many gene assays such as the 21-gene recurrence score (Oncotype DX), the 12-gene risk score (EndoPredict), and the PAM50 risk of recurrence (ROR) score were recommended for guiding decisions on adjuvant chemotherapy, while in the National Comprehensive Cancer Network (NCCN) guidelines, only the Oncotype DX was suggested to be capable of predicting chemotherapy benefits[9–11]. However, the 21-gene assay is initially designed for predicting prognosis of estrogen receptor-positive (ER+), N0 breast cancer, and evidence for its effectiveness in predicting chemotherapy benefits in HR+, HER2-, early breast cancer is lacked because no prospective study has validated that patients with high 21-gene scores benefit from the addition of adjuvant chemotherapy (all patients with HR+, HER2-, early breast cancer with high 21-gene scores were assigned to receive adjuvant chemotherapy except for those who were not adherent to trial designs in previous studies)[12, 13].

With these considerations, we performed a population-based study to identify patients with HR+, HER2-, early breast cancer who would benefit from adjuvant chemotherapy. In addition, we aimed to identify genes associated with chemotherapy benefits in this type of cohort. We also established two chemotherapy predictive scores (one score based on clinicopathological characteristics and another based on gene expression levels) for guiding decisions on adjuvant chemotherapy.

Methods

1. Data sources from the SEER database

Patients with HR+, HER2-, early (AJCC 7th stage I-III) breast cancer were identified from the Surveillance, Epidemiology, and End Results (SEER) database. The SEER database collects information on cancer from registries covering around 34.6% of the US patients, and we obtained permission to access the database with the reference number 12296-Nov2018. Our study was approved by the review board of the Affiliated Jinhua Hospital, Zhejiang University School of Medicine.

Criteria for including the subjects were as follows: patients diagnosed during 2010–2014 (because HR and HER2 statuses were not recorded in the database until 2010); patients with age of 20–80 years old; patients undergoing cancer-directed surgery; patients with tumor site ICD-O-3 codes C-50.1-50.5, C-50.8 and histology ICD-O-3 codes 8500, 8520, 8521, 8522, 8523, 8524 (Infiltrating duct carcinoma, Lobular

carcinoma, Infiltrating ductular carcinoma, Infiltrating duct and lobular carcinoma, Infiltrating duct mixed with other types of carcinoma, and Infiltrating lobular mixed with other types of carcinoma).

Criteria for excluding the subjects were as follows: patients with incomplete recordings, such as the AJCC stage; patients with multiple primary tumors; patients with survival time less than 1 month (because these patients were at risk of death from perioperative complications).

The age and year at diagnosis, marital status, race, histological type, differentiated grade, AJCC 7th T stage and N stage, chemotherapy information, causes of death and survival months were retrieved from the SEER database. Patients was classified into the chemotherapy group and the non-chemotherapy group according to whether chemotherapy is performed or not. The chemotherapy group included the patients receiving chemotherapy (irrespective of whether receiving hormone therapy or not) and the non-chemotherapy group included the patients receiving hormone therapy or the patients who did not receive any therapy after surgery.

2. Propensity score matching

To reduce biases from confounders and achieve balance between the chemotherapy and non-chemotherapy groups, a propensity score matching (PSM) was performed. Based on demographic and clinicopathological characteristics (i.e., age, marital status, race, histological type, differentiated grade, AJCC 7th T stage and N stage), patients were matched with a 1:1 ratio using the nearest neighbor method. Further analyses were conducted on the after-matched patients[14].

3. Survival analysis and subgroup analysis

Considering that the prognosis of HR+, HER2-, early breast cancer is favorable and the risk of non-cancer specific death (non-CSD) cannot be ignored, we chose competing risk analyses to calculate the cumulative incidences of cancer specific death (CSD) using the Gray tests[15]. The time to CSD was calculated from the date of diagnosis to the date of death of cancer; the time to non-CSD was calculated from the date of diagnosis to the date of death of other causes. CSD was regarded as the outcome event, and non-CSD was regarded as the competing event. In the construction of prognostic model, subdistribution hazard ratios (SHRs) and 95% confidential intervals (CIs) were calculated.

To explore whether chemotherapy improved prognosis in a certain group, subgroup analyses were conducted in patients stratified by age, race, marital status, histological type, differentiated grade, T stage and N stage respectively. A “forest plot” was drawn to present the results. The interaction between chemotherapy and age, race, marital status, histological type, differentiated grade, T stage or N stage were calculated respectively using the generic inverse variance method, which has been reported in a previous study, and P-values for interaction were calculated[16].

4. Data sources from the TCGA database

Gene expression RNA-seq data and phenotypic data of breast cancer were downloaded from The Cancer Genome Atlas (TCGA) database (<https://genome-cancer.ucsc.edu/>). Samples from patients with HR+, HER2-, early breast cancer having both complete RNA-seq data and complete phenotypic data entered into further analyses. In total, 29 normal samples and 254 tumor samples were identified.

5. Differentially expressed genes

Amongst the 29 normal samples and 254 tumor samples, tumor samples and normal samples were paired up according to the TCGA barcode, and 29 pairs of samples were identified. Differentially expressed genes (DEGs) between the paired samples were filtered using the R package "DESeq2". We set $|Log_2 \text{ fold change (FC)}| > 2.0$ and false discovery rate (FDR) < 0.05 as the thresholds to obtain the DEGs.

6. Chemotherapy predictive genes

For each one of the identified DEGs, patients were classified into the high and the low expression groups according to the median of the expression count. Because the TCGA database did not record causes of death, we could not perform competing risk analyses and chose Cox proportional hazard analyses instead. Death or tumor recurrence were regarded as the DFS event, and the time to DFS was calculated.

Univariate Cox proportional hazard analyses for chemotherapy were performed in the high expression group and the low expression group respectively, and hazard ratios (HRs) and 95% CIs were calculated. The interaction between chemotherapy and gene expression level as a categorical variable was examined, and the P-value for interaction and the I^2 value were calculated. We set the P-value for interaction < 0.05 as the threshold to identify genes associated with chemotherapy benefits, which were called "Chemotherapy predictive genes" (CPGs) in our study. With the I^2 value increasing, the statistical strength of the association between the gene expression level and magnitude of chemotherapy effect increases.

7. Chemotherapy predictive score

In the establishment of the chemotherapy predictive clinical score (CPCS), a variable was regarded to be associated with chemotherapy benefits and entered into the establishment when the variable's P-value for interaction was less than 0.05. In the establishment of the chemotherapy predictive genetic score (CPGS), we chose 5 most statistically significant genes according to their P-values for interaction to establish the CPGS.

The processes of establishing the chemotherapy predictive score were as follows: First, coefficient values of chemotherapy in each subgroups of a variable (demographic and clinicopathological characteristics in the CPCS and gene expression levels in the CPGS) were calculated respectively; Second, for coefficient values from the same variable, the values were transformed as the highest value was equal to 0 and the other values were equal to the absolute difference between their values and the highest; then, the transformed coefficient values were put together, and the highest transformed values was set as 100 in

scale and the lowest transformed values was set as 0 in scale, with other transformed values were set as scales between 0 and 100 proportionally (**Appendix** Fig. 1).

After establishing the CPCS, a subpopulation treatment effect pattern plot (STEPP) analysis was further conducted to explore the chemotherapy-CPCS interaction, and a chow test was conducted to identify the cut-off value of the CPCS[17]. After establishing the CPGS, each patient was scored using our established CPGS, and the patients were classified into a low score group and a high score group according to the median of the calculated scores. The chemotherapy-CPGS interaction was tested to validate performance of the CPGS in predicting chemotherapy benefits.

8. Genes from previous established multiple gene assays

To test the predictive performance of genes from previous established multiple gene assays, we extracted genes in the Oncotype Dx, 70-gene (MammaPrint), ROR, 28-gene signature, EndoPredict and Breast Cancer Index (BCI), and then drew a heatmap of their expression count in the 254 tumor samples[12, 18–22]. The samples were hierarchical clustered to see whether samples with same pathological stages could be clustered together.

Results

1. Patient characteristics from the SEER database

In total, 86158 patients with HR+, HER2-, early breast cancer were identified from the SEER database, and around 30% of the patients received chemotherapy (**Appendix** Table 1). Compared to the patients not receiving chemotherapy, the patients receiving chemotherapy had poorer differentiated grade, higher T stage and N stage (P all < 0.001). 38.2% of the tumors were poorly differentiated in the chemotherapy group, in contrast to the 11.1% in the non-chemotherapy group. Most tumors were T1 stage (80.7%) and N0 stage (86.8%) in the non-chemotherapy group, while only 46.9% of the tumors were T1 stage and 41.0% of the tumors were N0 stage in the chemotherapy group.

Table 1
Demographic and clinicopathological characteristics of after-matched patients in chemotherapy and non-chemotherapy groups.

	N (%)	Non-chemotherapy	Chemotherapy	P value ^b
		n (%)	n (%)	
Total	30518	15259	15259	
Age				0.948
<=40	1783 (5.8)	882 (5.8)	901 (5.9)	
41–60	17179 (56.3)	8585 (56.3)	8594 (56.3)	
61–70	8635 (28.3)	4321 (28.3)	4314 (28.3)	
71–80	2921 (9.6)	1471 (9.6)	1450 (9.5)	
Marital status				0.913
Married	19466 (63.8)	9745 (63.9)	9721 (63.7)	
Single	4803 (15.7)	2388 (15.6)	2415 (15.8)	
Divorce	6249 (20.5)	3126 (20.5)	3123 (20.5)	
Race				0.393
White	24669 (80.8)	12295 (80.6)	12374 (81.1)	
Black	2838 (9.3)	1424 (9.3)	1414 (9.3)	
Other	3011 (9.9)	1540 (10.1)	1471 (9.6)	
Location				0.885
Central portion	1668 (5.5)	850 (5.6)	818 (5.4)	
Upper-inner	4284 (14.0)	2133 (14.0)	2151 (14.1)	
Lower-inner	1860 (6.1)	928 (6.1)	932 (6.1)	
Upper-outer	12003 (39.3)	6037 (39.6)	5966 (39.1)	
Lower-outer	2803 (9.2)	1391 (9.1)	1412 (9.3)	
Overlapping lesion	7900 (25.9)	3920 (25.7)	3980 (26.1)	
Histology				0.515
IDC	25119 (82.3)	12535 (82.1)	12584 (82.5)	

^a T stage and N stage according to 7th AJCC staging system.

^b P values obtained from the χ^2 test. All statistical tests were two-sided.

	N (%)	Non-chemotherapy n (%)	Chemotherapy n (%)	P value ^b
ILC	3191 (10.5)	1626 (10.7)	1565 (10.3)	
IDC + ILC	2208 (7.2)	1098 (7.2)	1110 (7.3)	
Grade				0.299
Well differentiated	5054 (16.6)	2489 (16.3)	2565 (16.8)	
Moderately differentiated	15788 (51.7)	7878 (51.6)	7910 (51.8)	
Poorly differentiated	9676 (31.7)	4892 (32.1)	4784 (31.4)	
T stage ^a				0.739
T1	18125 (59.4)	9032 (59.2)	9093 (59.6)	
T2	11082 (36.3)	5583 (36.6)	5499 (36.0)	
T3	1180 (3.9)	579 (3.8)	601 (3.9)	
T4	131 (0.4)	65 (0.4)	66 (0.4)	
N stage ^a				0.050
N0	18125 (59.4)	9032 (59.2)	9093 (59.6)	
N1	11082 (36.3)	5583 (36.6)	5499 (36.0)	
N2	1180 (3.9)	579 (3.8)	601 (3.9)	
N3	131 (0.4)	65 (0.4)	66 (0.4)	

^a T stage and N stage according to 7th AJCC staging system.

^b P values obtained from the χ^2 test. All statistical tests were two-sided.

Patients in the chemotherapy and non-chemotherapy groups were matched 1:1, and there were 15259 patients in each after-matched groups. The demographic and clinicopathological characteristics of the after-matched groups were shown in Table 1. The P-values of every variables were ≥ 0.05 , indicating that the demographic and clinicopathological characteristics were well balanced after matching.

2. Survival analysis for patients from the SEER database

The median follow-up period was 49 months (IQR: 34–65 months) in the after-matched patients. During the follow-up period, 873 patients suffered CSD and 630 patients suffered non-CSD.

In the entire after-matched cohort, chemotherapy was not associated with CSD in both univariable (HR: 0.97; 95% CI: 0.85–1.11; P = 0.66) and multivariable analyses (HR: 0.98; 95% CI: 0.85–1.12; P = 0.73). In

the multivariable analyses, age, marital status, race, histological type, differentiated grade, T stage and N stage were all associated with CSD (**Appendix** Table 2).

Table 2

The coefficient values, transformed coefficient values and corresponding scores of the variables of race, T stage and N stage in the chemotherapy predictive clinical score (CPCS).

Variable	Coefficient value	Transformed coefficient value	Score
Race			
White	-0.04	0.67	67.36
Black	-0.25	0.88	88.67
Other	0.63	0	0
T stage			
T1	0.35	0	0
T2	-0.17	0.52	52.19
T3	-0.34	0.69	69.65
T4	-0.33	0.68	68.45
N stage			
N0	0.29	0	0
N1	-0.07	0.35	35.76
N2	-0.53	0.81	81.72
N3	-0.71	0.99	100.00

Figure 1 showed results of the subgroup analyses. Chemotherapy did not render a reduction in CSD for patients with well, moderate or high differentiated grade tumors (HR [chemotherapy group versus non-chemotherapy group]: 1.12 in the well differentiated grade group versus 1.11 in the moderate differentiated grade group versus 0.90 in the poor differentiated grade group, P-value for interaction = 0.31), while patients with higher T stage or N stage benefited more from the chemotherapy (HR [chemotherapy group versus non-chemotherapy group]: 1.42 in the T1 group versus 0.85 in the T2 group versus 0.71 in the T3 group versus 0.72 in the T4 group, P-value for interaction < 0.01; HR [chemotherapy group versus non-chemotherapy group]: 1.33 in the N0 group versus 0.93 in the N1 group versus 0.59 in the N2 group versus 0.49 in the N3 group, P-value for interaction < 0.01).

3. Chemotherapy predictive clinical score

According to the P-values for interaction, race, T stage and N stage were associated with chemotherapy benefits and entered into the establishment of CPCS. The coefficient values, transformed coefficient

values and corresponding scores of the three variables were shown in Table 2.

The chemotherapy-CPCS interactions were shown in the STEPP (Fig. 2), from which we observed that with the CPCS increasing, the CSD of patients receiving chemotherapy was increasingly lower than that of patients receiving no chemotherapy. Results of the chow test showed that the 103.12 was the optimal cut-off value.

4. Differentially expressed genes and chemotherapy benefit genes

Amongst the 29 pairs of tumor samples and normal samples, 1709 DEGs were filtered. The DEGs were displayed in a heatmap and a volcanic map respectively (**Appendix Fig. 2**, **Appendix Fig. 3**).

For each one of the identified DEGs, the tests of statistical interactions between chemotherapy and the gene expression level were performed, and we finally identified 84 genes associated with chemotherapy benefits, which were summarized in **Appendix Table 3**.

Table 3

The coefficient values, transformed coefficient values and corresponding scores of the variables of race, T stage and N stage in the chemotherapy predictive genetic score (CPGS).

Gene	Coefficient value	Transformed coefficient value	Score
SOWAHA			
Low expression	1.11	3.92	100.00
High expression	-2.81	0.00	0.00
RP11-205M3.3			
Low expression	1.06	3.65	93.13
High expression	-2.59	0.00	0.00
IRX6			
Low expression	1.28	3.20	81.54
High expression	-1.91	0.00	0.00
PPBP			
Low expression	-2.05	0.00	0.00
High expression	0.96	3.01	76.90
EMX1			
Low expression	1.00	3.55	90.55
High expression	-2.55	0.00	0.00

5. Chemotherapy predictive genetic score

Amongst the 84 identified CPGs, we chose SOWAHA\RP11-205M3.3\IRX6\PPBP and EMX1 (their P-values for interaction were the least) to establish the CPGS. The coefficient values, transformed coefficient values and corresponding scores of the 5 genes were shown in Table 3.

According to the median of the CPGS, 125 patients were classified into the low expression group (62 patients receiving chemotherapy and 63 patients not) and 129 patients were classified into the high expression group (57 patients receiving chemotherapy and 73 patients not). The HR of chemotherapy was 8.95 (95% CI: 1.12–71.56, $P < 0.01$) in the low score group, which was 0.05 (95% CI: 0.01–0.36, $P < 0.01$) in the high score group, and the P-value for interaction between chemotherapy and CPGS as a categorical variable was less than 0.01 (Fig. 3).

6. Genes from previous established multiple gene signatures

Figure 4 showed the expression count of 130 genes from previous gene assays (values were scaled in the row direction), and we failed to observe that patients with similar age, AJCC stage, T stage or N stage to be clustered together using the hierarchical clustering.

Discussion

HR and HER2 statuses are two well-established prognostic factors in breast cancer, and the prognosis of patients with HR+, HER2-, early breast cancer is favorable[23, 24]. Our population-based study found that the non-CSD accounted for 41.92% of all causes of death, indicating that non-CSD should be emphasized and clinicians should be cautious when performing adjuvant chemotherapy for this cohort considering the side effects brought by chemotherapy.

In our survival analyses, we found that the entire cohort failed to benefit from the addition of chemotherapy, indicating some patients may be spared the use of chemotherapy. In our further subgroup analyses, N stage was associated with chemotherapy benefits, which was consistent with guidelines that lymph nodal status was an indicator for choosing adjuvant chemotherapy. We concluded from our analyses that, in patients with HR+, HER2-, early breast cancer, most patients with N0 breast cancer might be spared the use of adjuvant chemotherapy and patients with N2 and N3 cancer should be arranged to receive the adjuvant chemotherapy, while the chemotherapy benefits were uncertain for patients with N1 cancer just according to the clinical characteristics.

There has been a tendency to equate factors associated with prognosis with those related to chemotherapy benefits. However, in our study, we found these two types of factors were not entirely same. In the multivariable analyses, age, marital status, race, histological type, differentiated grade, T stage and N stage were all associated with CSD, while only race, T stage and N stage were significant when considering the interactions between adjuvant chemotherapy and demographic and clinicopathological characteristics. Therefore, we concluded that only race, T stage and N stage were

associated with chemotherapy benefits, and factors like histological type and differentiated grade were not (even though histological type and differentiated grade were associated with prognosis). In addition, we established a score for predicting chemotherapy benefits with the 3 factors (race, T stage and N stage). Different from previous predictive scores, which were transformed from models for predicting prognosis, we established our score using a totally new method.

The 21-gene recurrence score assay was the only genetic assay recommended for predicting chemotherapy benefits in the NCCN guidelines, while its predictive ability in HR+, HER2-, early breast cancer is doubtful[9]. The 21-gene assay is initially established for predicting prognosis of estrogen receptor-positive (ER+), N0 breast cancer rather than chemotherapy benefits of HR+, HER2-, early breast cancer.

Evidences supporting the 21-gene assay for predicting chemotherapy benefits were mainly from the NSABP B20 trial, the TAILORx trial and the SWOG-8814 trial. The NSABP B20 trial enrolled 2363 patients with ER+, N0 breast cancer, who were randomly assigned to chemotherapy or not[25]. Among the 2363 patients, a total of 651 patients were assessed using the 21-gene recurrence score assay and were classified into three groups according to the score: low-RS ($RS < 18$), intermediate-RS ($18 < RS < 30$) and high-RS ($RS \geq 31$). Results of the study suggested that patients with low RSs failed to receive chemotherapy benefits and patients with high RSs benefited from chemotherapy, and chemotherapy benefits were uncertain for patients with intermediate RSs. Note that the study included patients with both HER2- and HER2 + breast cancer, therefore it was doubtful to conclude that patients with HR+, HER2-, N0 breast cancer benefited from chemotherapy because the HER2 score accounts for a certain proportion of the 21-gene RS[12].

In a secondary analysis of the NSABP B20 trial, gene expression results of the 21-gene assay were obtained in the 651 patients, and 569 of them were identified as HER2- according to the HER2-gene expression from the RT-PCR assay[26]. Then chemotherapy benefits for the 569 patients were assessed using the Kaplan-Meier plots, and patients were reported to benefit from the addition of chemotherapy in the high-RS group (whether high-RS was set as 31 or more or as greater than 25), while no significant benefits in the low- and intermediate-RS groups. Considering the retrospective trait and the relatively low sample size in the high-RS group (97 patients in the $RS \geq 31$ group or 121 patients in the $RS > 25$ group), results of this secondary analysis need further validation. In addition, the NSABP B20 trial was conducted in an early era when the effectiveness of endocrine therapy is limited, and this effectiveness has increasingly improved nowadays.

The TAILORx trial considered the impact of HER2 status on survival and chemotherapy benefits, and enrolled patients with HR+, HER2-, N0 breast cancer[13]. The 6711 patients with a midrange RS of 11 to 25 by the 21-gene assay were randomly assigned to receive endocrine therapy alone or endocrine therapy plus chemotherapy. Although some patients aged 50 or younger seemed to benefit from the chemotherapy, the entire cohort failed to benefit from the chemotherapy (endocrine therapy alone versus endocrine therapy plus chemotherapy: invasive disease recurrence or death, $P = 0.26$; recurrence at a

distant site, $P = 0.48$). The 1389 patients with a RS of 26 or higher were all assigned to receive chemoendocrine therapy, thus the study did not directly demonstrate that patients with HR+, HER2-, N0 breast cancer who had a high RS benefited from chemotherapy.

In a secondary analysis of the TAILORx trial, 1389 patients with HR+, HER2-, N0 breast cancer who had a high RS of 26 or more were identified and 89 of the patients had no chemotherapy[27]. The study reported that patients without chemotherapy had a worse prognosis in invasive disease-free survival (IDFS) than patients with chemotherapy (HR with 95% CI: 0.48, 0.29–0.80), rather than freedom from recurrence at a distant site (HR with 95% CI: 0.74, 0.32–1.69). However, in the non-chemotherapy group, 57 of the 89 patients underwent breast conservation surgery while fewer patients (29 patients) received postoperative radiation therapy, indicating these non-chemotherapy patients might be of poor adherence. The radiation therapy has been validated to reduce local recurrence after breast conservation and been suggested for patients treated with conservation surgery. Therefore, results of this secondary analysis that chemotherapy rendered benefits in IDFS for patients with HR+, HER2-, N0 breast cancer who had a high RS need to be further validated.

According to the ASCO guidelines, the 70-gene assay is also recommended for treatment decision making, and the MINDACT trial is the highest level of evidence supporting its predictive values[28, 29]. In this randomized, phase 3 study, patients with early breast cancer were classified into four groups according to their clinical risk (using a modified version of Adjuvant! Online) and genomic risk (using the 70-gene signature), and the groups with discordant risk results (high clinical and low genomic risk or low clinical and high genomic risk) were randomly assigned to receive chemotherapy or not. Results of this study showed that there were no statistically significant chemotherapy benefits with respect to survival without distant metastasis, disease-free survival or overall survival, indicating that high genomic risk may not be an accurate indicator for chemotherapy benefits. Additionally, subgroup analyses of the MINDACT trial of patients with ER+, HER2-, N- showed this subgroup failed to benefit from the addition of chemotherapy (DMFS for chemotherapy versus no chemotherapy in the two discordant risk groups, in the HR+, HER2-, N- subgroup, $P = 0.456$ in high clinical risk/low genomic risk; $P = 0.333$ in low clinical risk/high genomic risk), which was consistent with our results.

The other multiple gene assays are also not suitable for predicting chemotherapy benefits in HR+, HER2-, early breast cancer because they all did not take the interaction between chemotherapy and gene expression level into consideration. Results that genes from previous multiple gene assays (including genes in the Oncotype Dx, MammaPrint, PAM 50, 28-gene signature, EndoPredict and BCI) failed to cluster patients with similar pathological stages also proved that these previous genes were not suitable for the prediction of HER2-, HR+, early breast cancer[12, 18–22]. Therefore, we adopted a new method, which considered the interaction, and identified that SOWAHA\RP11-205M3.3\IRX6\PPBP and EMX1 were genes most closely related to chemotherapy benefits in the HER2-, HR+, early breast cancer, and we established a chemotherapy predictive genetic score based on the 5 genes.

In colorectal cancer, the Iroquois Homeobox 6 (IRX6) expression has been reported to be lower in tumor samples and was a prognostic predictor[30]. Pro-platelet basic protein (PPBP) was also known as C-X-C motif chemokine ligand 7 (CXCL7), the overexpression of which has been reported to promote cell proliferation in vivo and in vitro, and it has been proposed as a novel biomarker for the diagnosis of lung cancer and renal cell carcinoma[31–33]. Zhou et al. reported that Empty spiracles homeobox 1 (EMX1) was a potential target of miR-497, which was recognized as a tumor suppressor in many cancers, and the high expression of EMX1 was related to advanced clinicopathologic characteristics in endometrial cancer[34]. Our study confirmed that the high expression of EMX1 was associated with chemotherapy benefits in breast cancer, indicating the inconsistency between prognostic and predictive values of EMX1. The role of SOWAHA \square RP11-205M3.3 in HER2-, HR+, early breast cancer needs further investigation.

There are several limitations in our study. Firstly, our study was a retrospective study and biases such as selection bias and treatment bias were inevitable in our study. In addition, our established CPCS and CPGS need further external validations to prove their effectiveness in predicting chemotherapy benefits. Finally, some important information such as regimens and courses of chemotherapy are not available in the SEER and TCGA database, which limits our further analyses.

Conclusion

For patients with HR+, HER2-, early breast cancers, our study found that adjuvant chemotherapy failed to render survival benefits in the entire cohort. Nodal status helped to identify patients benefiting from chemotherapy, and most N0 patients and some N1 patients might be spared the adjuvant chemotherapy. In addition, we adopted a new method to establish a chemotherapy predictive clinical score and a chemotherapy predictive genetic score for predicting chemotherapy benefits in the HR+, HER2-, early breast cancer.

Declarations

Ethics approval and consent to participate

The study was approved by the review board of the Affiliated Jinhua Hospital, Zhejiang University School of Medicine.

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the key program of the Jinhua Municipal Science & Technology Bureau [grant numbers 2014-3-008]; and the key program of the Jinhua Municipal Science & Technology Bureau [grant numbers 2010-3-004]. The funding plays a role in the design of the study.

Author's Contribution

JF F, HP L: Data curation, Software, Methodology, Writing - original draft; HJ Z, CY G, LX L, RH Y, QH W, WF T, X Z, SS Z, XY J, XF X: Methodology, Writing - original draft; SA Y: Funding acquisition, Methodology, Project administration, Writing - review & editing. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Ferlay J, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*. 2015;136(5):E359-86.
2. Curigliano G, et al., *De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer* 2017. *Annals of oncology: official journal of the European Society for Medical Oncology*, 2017. **28**(8): p. 1700–1712.
3. Nitz U, et al. West German Study PlanB Trial: Adjuvant Four Cycles of Epirubicin and Cyclophosphamide Plus Docetaxel Versus Six Cycles of Docetaxel and Cyclophosphamide in HER2-Negative Early Breast Cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2019;37(10):799–808.
4. Pagani O, et al. Adjuvant exemestane with ovarian suppression in premenopausal breast cancer. *N Engl J Med*. 2014;371(2):107–18.
5. Francis P, et al. Adjuvant ovarian suppression in premenopausal breast cancer. *N Engl J Med*. 2015;372(5):436–46.
6. Saha P, et al. Treatment Efficacy, Adherence, and Quality of Life Among Women Younger Than 35 Years in the International Breast Cancer Study Group TEXT and SOFT Adjuvant Endocrine Therapy Trials. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2017;35(27):3113–22.
7. Francis P, et al. Tailoring Adjuvant Endocrine Therapy for Premenopausal Breast Cancer. *N Engl J Med*. 2018;379(2):122–37.

8. von Minckwitz G, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2012;30(15):1796–804.
9. *NCCN Clinical Practice Guidelines in Oncology: Breast Cancer. Version 5.2020 - July 15, 2020.*
10. Harris L, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2016;34(10):1134–50.
11. Krop I, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Focused Update. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2017;35(24):2838–47.
12. Paik S, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27):2817–26.
13. Sparano J, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*. 2018;379(2):111–21.
14. Ho DE, et al. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political analysis*. 2007;15(3):199–236.
15. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc*. 1999;94(446):496–509.
16. Lacey B, et al. Age-specific association between blood pressure and vascular and non-vascular chronic diseases in 0·5 million adults in China: a prospective cohort study. *The Lancet Global health*. 2018;6(6):e641–9.
17. Bonetti M, *Patterns of treatment effects in subsets of patients in clinical trials*. Biostatistics (Oxford, England), 2004. 5(3): p. 465–481.
18. van 't Veer L, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415(6871):530–6.
19. Perou C, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–52.
20. Ma Y, et al. Population-based molecular prognosis of breast cancer by transcriptional profiling. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2007;13(7):2014–22.
21. Filipits M, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2011;17(18):6012–20.
22. Ma X, et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2008;14(9):2601–8.

23. Goldhirsch A, et al. 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. *Annals of oncology: official journal of the European Society for Medical Oncology*. 2013;24(9):2206–23.
24. Coates A, et al. Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Annals of oncology: official journal of the European Society for Medical Oncology*. 2015;26(8):1533–46.
25. Paik S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2006;24(23):3726–34.
26. Geyer C, et al. 21-Gene assay as predictor of chemotherapy benefit in HER2-negative breast cancer. *NPJ breast cancer*. 2018;4:37.
27. Sparano J, et al., *Clinical Outcomes in Early Breast Cancer With a High 21-Gene Recurrence Score of 26 to 100 Assigned to Adjuvant Chemotherapy Plus Endocrine Therapy: A Secondary Analysis of the TAILORx Randomized Clinical Trial*. *JAMA oncology*, 2019.
28. Henry N, et al. Role of Patient and Disease Factors in Adjuvant Systemic Therapy Decision Making for Early-Stage, Operable Breast Cancer: Update of the ASCO Endorsement of the Cancer Care Ontario Guideline. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2019;37(22):1965–77.
29. Cardoso F, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med*. 2016;375(8):717–29.
30. Zuo S, Dai G, Ren X. Identification of a 6-gene signature predicting prognosis for colorectal cancer. *Cancer cell international*. 2019;19:6.
31. Grépin R, et al. The CXCL7/CXCR1/2 axis is a key driver in the growth of clear cell renal cell carcinoma. *Cancer research*. 2014;74(3):873–83.
32. Kinouchi T, et al. Expression level of CXCL7 in peripheral blood cells is a potential biomarker for the diagnosis of renal cell carcinoma. *Cancer Sci*. 2017;108(12):2495–502.
33. Du Q, et al. CTAPIII/CXCL7: a novel biomarker for early diagnosis of lung cancer. *Cancer medicine*. 2018;7(2):325–35.
34. Xu X, et al. miRNA-mRNA Associated With Survival in Endometrial Cancer. *Frontiers in genetics*. 2019;10:743.

Figures

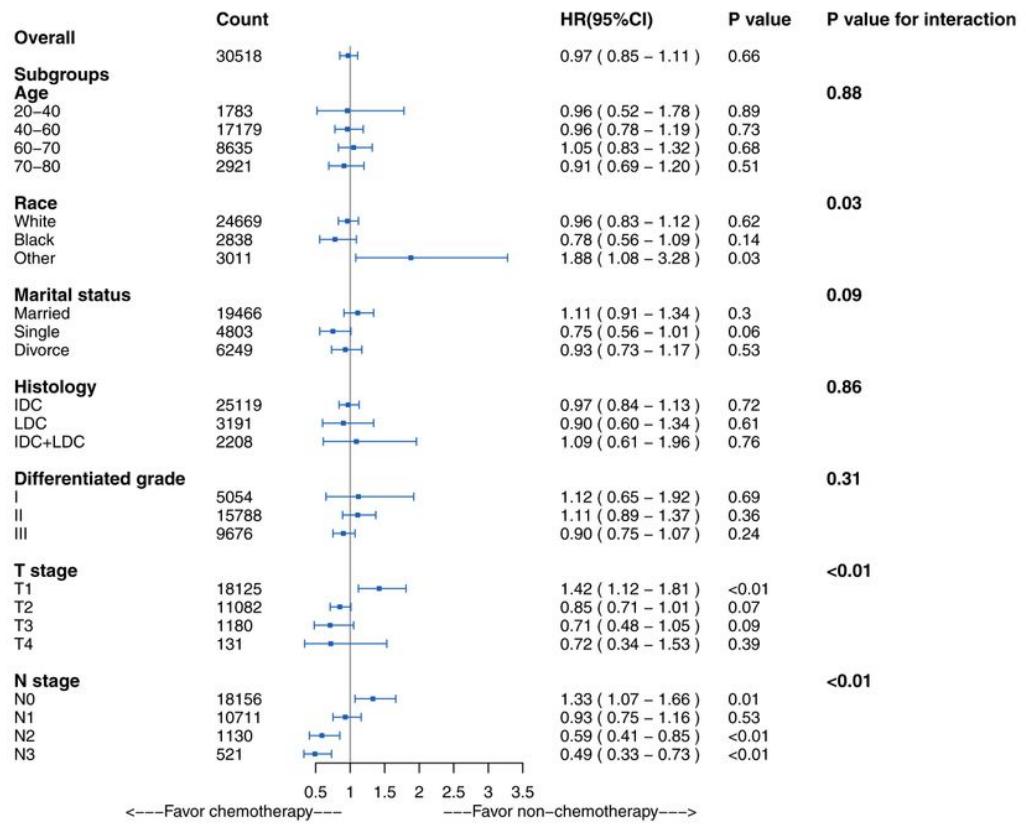


Figure 1

Chemotherapy benefits in after-matched patients stratified by age, race, marital status, histological type, differentiated grade, T stage and N stage.

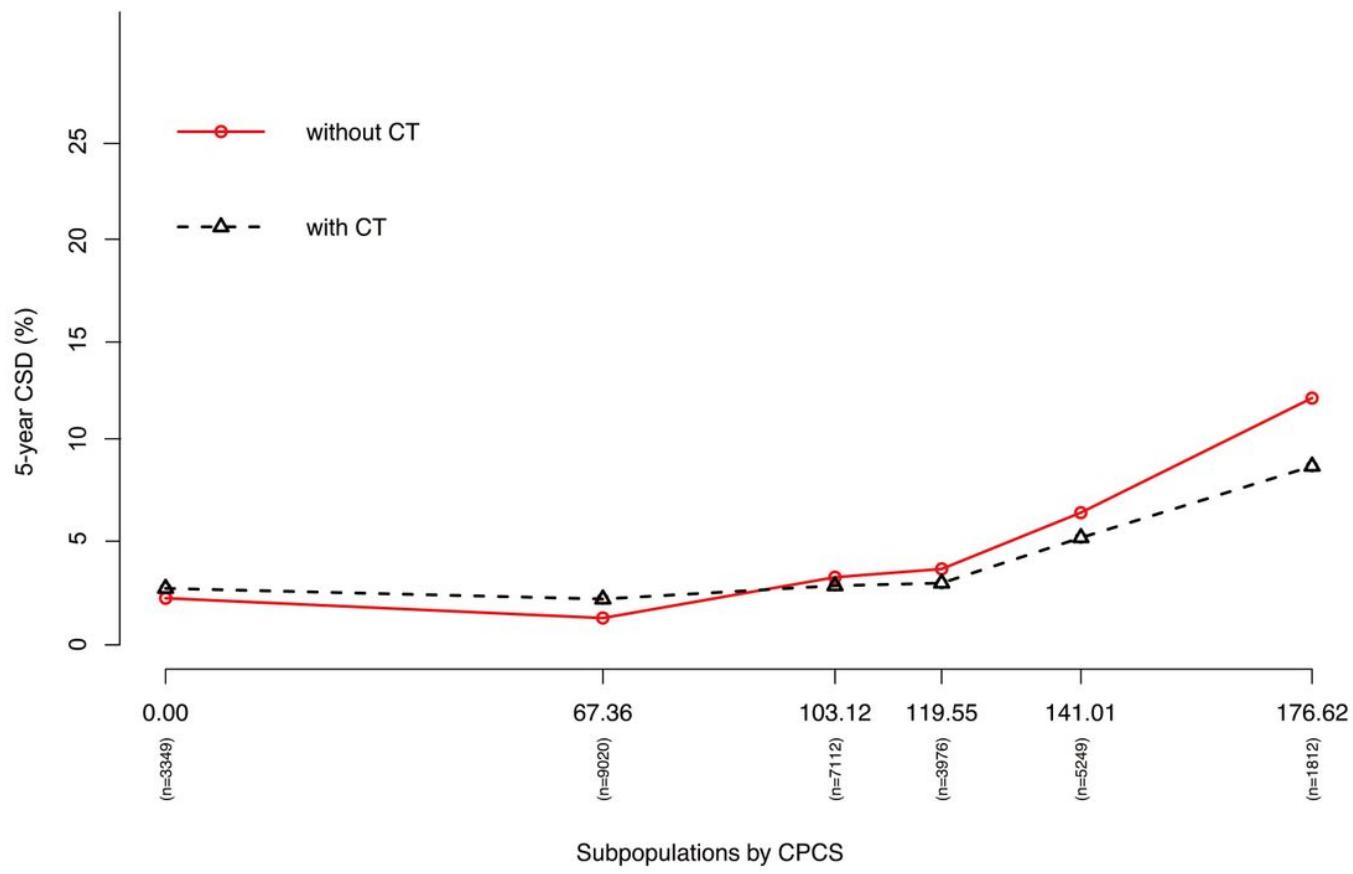


Figure 2

Subpopulation treatment effect pattern plot (STEPP) analysis of the interaction between chemotherapy and clinical predictive clinical score (CPCS).

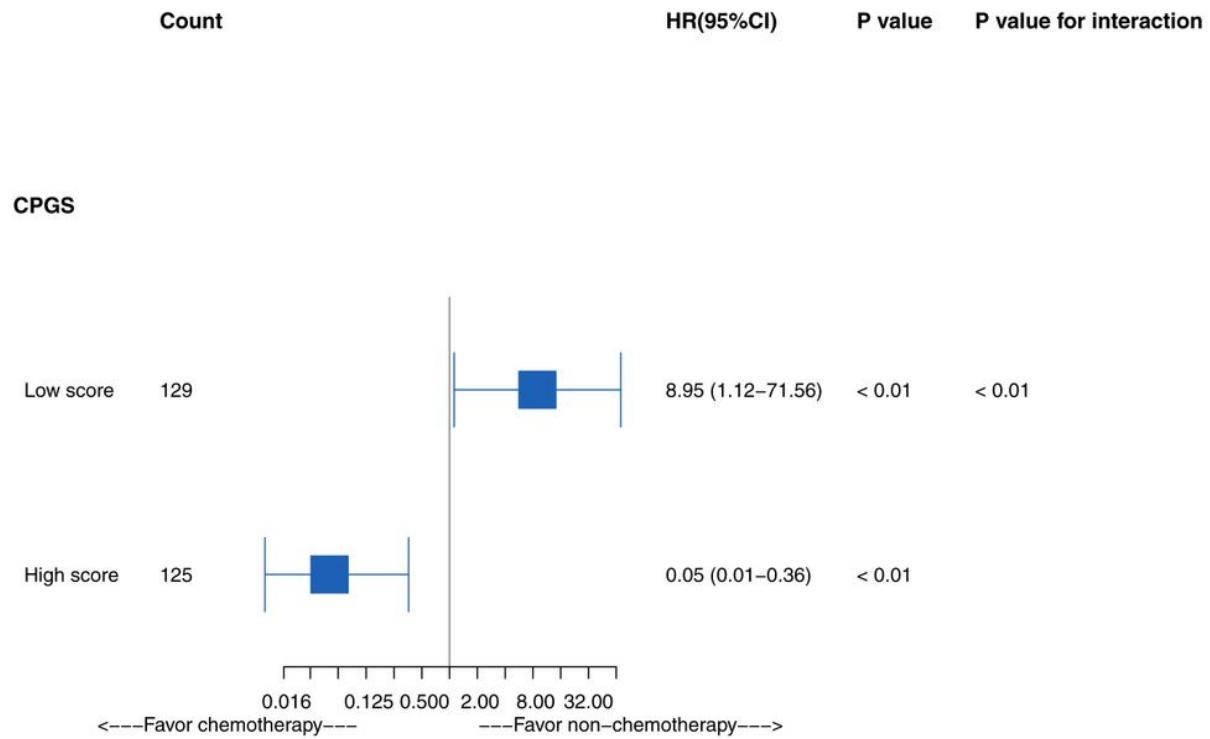


Figure 3

Chemotherapy benefits in low clinical predictive genetic score (CPGS) and high CPGS groups (stratified by the median of the CPGS).

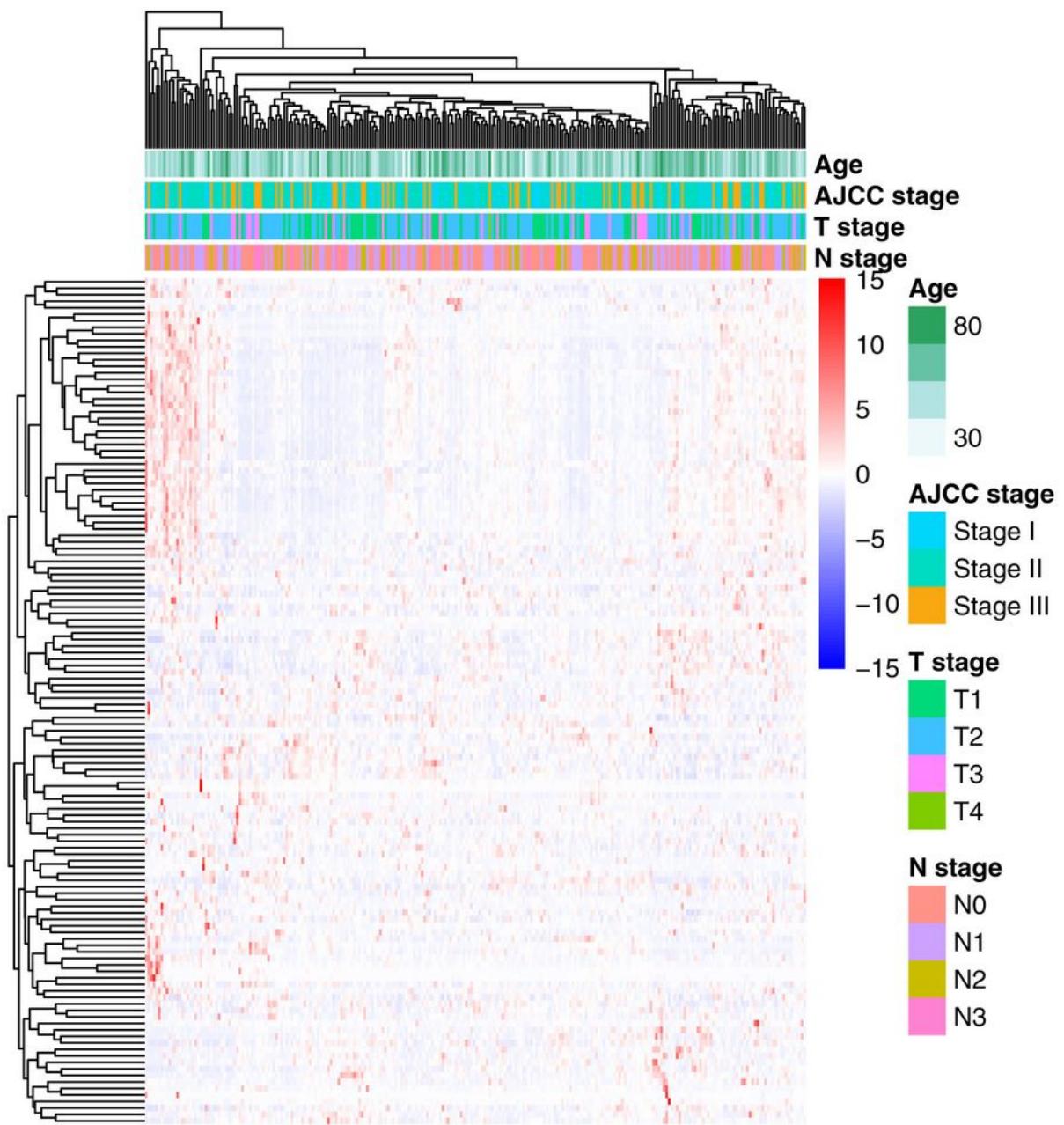


Figure 4

Heatmap showing the expression count of 130 genes from previous gene assays (values were scaled in the row direction). Each row represents a tumor sample and each column a single gene.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixFigure2.eps](#)
- [AppendixFigure3.eps](#)
- [AppendixTable1.docx](#)
- [AppendixTable2.docx](#)
- [AppendixFigure1.eps](#)