Validation of the NSABP/NRG Oncology 8-gene trastuzumab-benefit signature in Alliance/NCCTG N9831

Katherine L. Pogue-Geile, Nan Song, Daniel J. Serie, et al.

Supplementary Materials

Supplementary Methods

Method of predicting benefit groups in N9831

All N9831 patients provided written consent for research, as approved by the Mayo Clinic Institutional Review Board (PR07-005237).

1,388 N9831 samples were profiled by two nCounter code Sets, called v1 and v3. v1 has four housekeeping genes (*GAPDH, POLR2A, TBP, YWHAZ*) and v3 has 12 housekeeping genes (*ABCF1, B2M, G6PD, NRDE2, OAZ1, POLR2A, SDHA, STK11IP, TBC1D10B, TBP, UBB, YWHAZ*). In addition, v1 and v3 share 17 genes. For the 8-gene classifier, *ERBB2, IGF1R, ESR1, GATA3, GRB7,* and *NAT1* are in v1, and CA12 and C17orf37 (which is now annotated as *MIEN1*) are in v3.

After technical normalization, i.e., adjusted by positive control, median [geomean(positive Controls)]/geomean (positive Control), we normalized data using housekeeping genes in the same platform with the formula: $\Delta Cq = Cq(\text{target gene}) - Cq(\text{reference index})$.

To remove batch effects among the two code sets of N9831 (v1.N9831 and v3.N9831), and B-31, each gene was normalized as mean zero and standard deviation one. Correspondingly, the 8-gene signature method was modified slightly by changing the input from raw gene expression to normalized gene expression. Based on the pre-specified cutoff, the N9831 samples were assigned to no-, moderate-, or high-trastuzumab benefit groups. Note, for the B-31 test dataset, the modified 8-gene algorithm has very similar performance as the original method as shown in **Supplementary Table 1**.

Details for assigning patients to the large-, moderate-, and no-benefit groups are described in Pogue-Geile et al.[1] The R code is listed below. The assignment of patients to an 8-gene-trastuzumab benefit group were made by NSABP investigators blinded to clinical information.

Methods testing of outcomes in the *predicted* trastuzumab-benefit groups

The endpoints for analysis were disease-free survival (DFS) and recurrence-free survival (RFS). DFS was defined as time from random assignment to breast cancer recurrence (local, region, or distant recurrence of breast cancer), breast cancer-related death, or death from any cause. RFS

was defined as the time from random assignment to breast cancer recurrence (local, regional, or distant recurrence of breast cancer), or breast cancer-related death.

Follow-up included events recorded before April 5, 2015. For each 8-gene group, Cox models were used to assess whether DFS or RFS differed with respect to treatment by adjusting age, nodes, ER/PR, tumor size, and grade. The proportionality assumption was tested using Schoenfeld Residuals. For all tests, the proportionality assumptions were not violated. The predictive value of treatment benefit was assessed using Cox models with an interaction term between the treatment and 8-gene group. All reported P-values are 2-sided and the statistical significance level was set to <0.05.

Reference(s):

1. Pogue-Geile KL, et al., Predicting degree of benefit from adjuvant trastuzumab in NSABP trial B-31. *J Natl Cancer Inst.* 2013. 105(23): 1782-8.

R code to generate 8 gene benefit groups

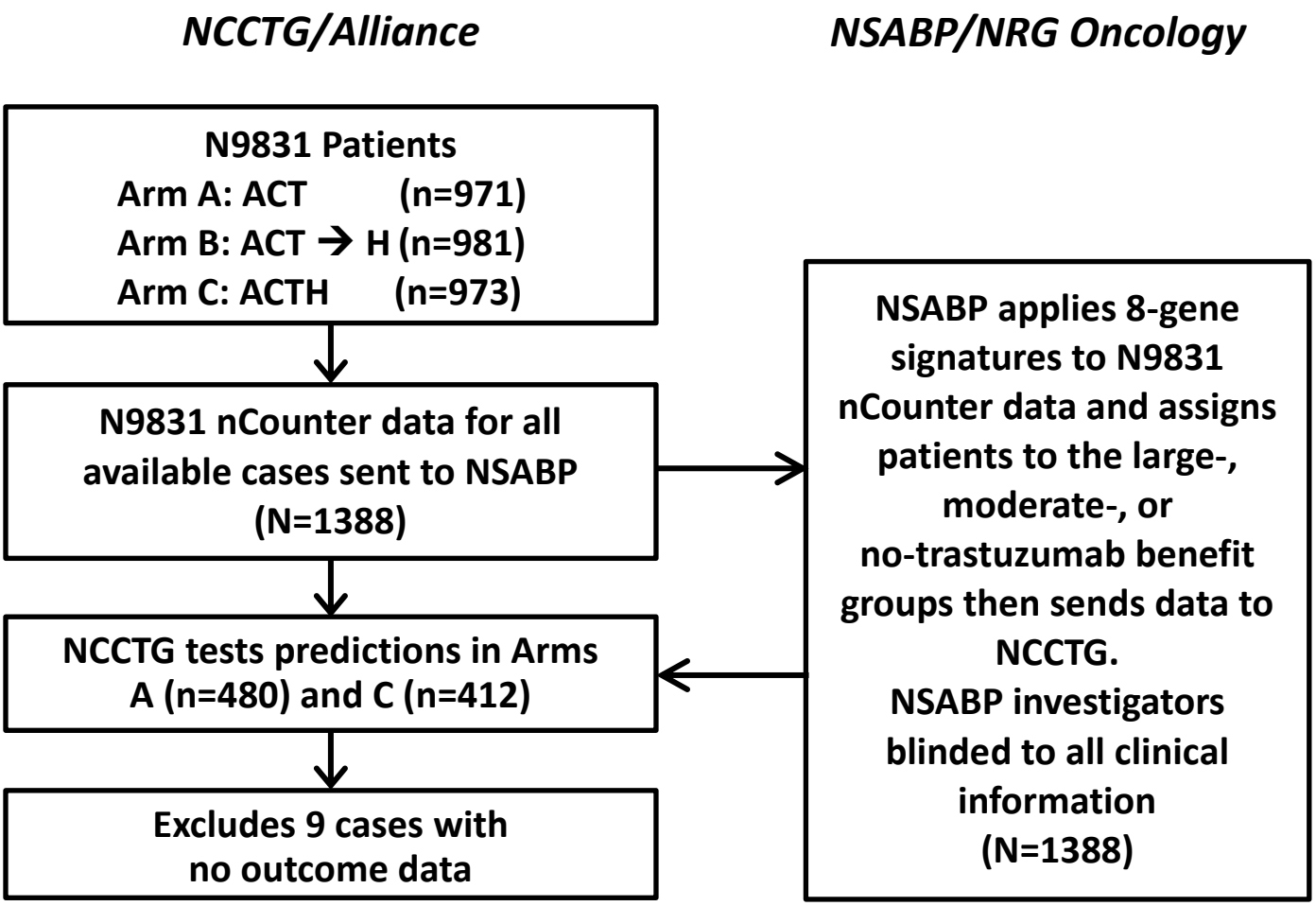
```
#########
# input_data: gene expression data
# rows represent samples, columns represent genes
##########
        predict_8gene_sig = function(input_data){
                pca.scale.rotation = matrix(c(
                        -0.3402804, 0.4547020,
                        -0.3214482, 0.4606466,
                        0.3824512, 0.1263290,
                        0.4106668, 0.2517401,
                        0.3145151, 0.3692210,
                        -0.3401935, 0.4603961,
                        0.3471010, 0.3264948,
                        0.3617094, 0.2164050),
                ncol=2, byrow=TRUE)
                rownames(pca.scale.rotation)=c('ERBB2','C17orf37','IGF1R','ESR1',
                                                'GATA3','GRB7','CA12','NAT1')

                input_data = input_data[, rownames(pca.scale.rotation)]
                pca_data = scale(input_data) %*% as.matrix(pca.scale.rotation)
                pca.d.scale_sd_1 = 2.051814
                pca.d.scale_sd_2 = 1.472679
                f1 <- scale( pca_data[,1], center=0, scale=2.051814)
                f2 <- scale( pca_data[,2], center=0, scale= 1.472679)
```

```
        ## assign groups according to cutoffs in the manuscript
        group <- ifelse( f1>0.6 & f2>0.1, 0,
                         ifelse( ( f1 > -0.6 & f1 < 0.6 ) & f2 >= 0.6, 2,
                         ifelse( f1 > -0.12 & f2 <= 0.1, 1,
                         ifelse( f1 <= -0.12 & f2 <= -0.55, 1,
                         ifelse( f1 <= -0.12 & ( f2 > -0.55 & f2 < 0.6 ), 2,
                         ifelse( f1 <= -0.6 & f2 >= 0.6, 1,
                         ifelse( ( f1 > -0.12 & f1 <= 0.6 ) & ( f2 > 0.1 & f2 <= 0.6 ) &
f2 > f1+0.22, 2,
                         ifelse( ( f1 > -0.12 & f1 <= 0.6 ) & ( f2 > 0.1 & f2 <= 0.6 ) &
f2 <= f1+0.22, 1,
                         NA ) ) ) ) ) ) ) ) )
        return (group)
    }
```

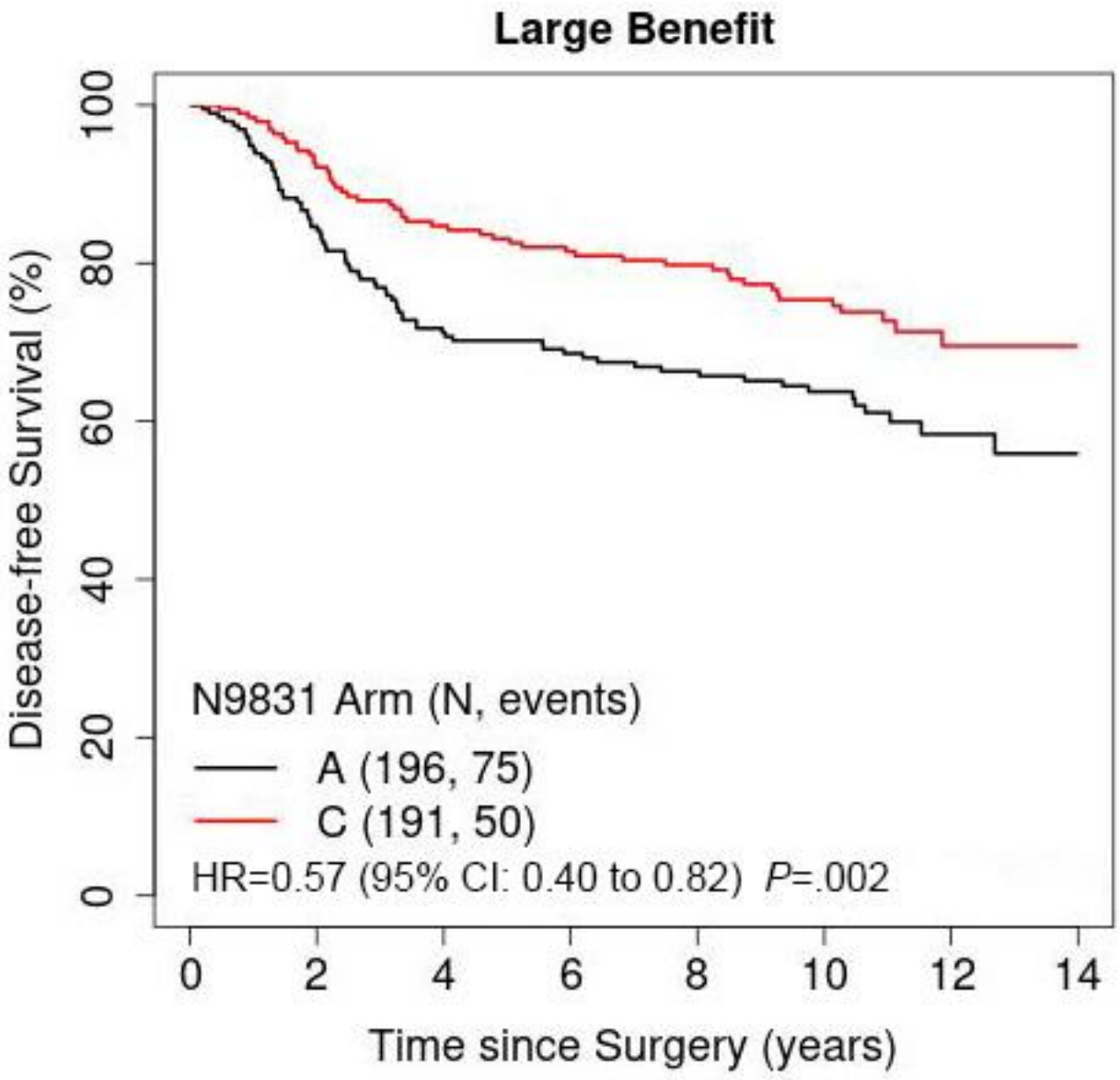
Supplementary Table 1. Comparison of prediction results using original and modified 8-gene signature algorithm on NSABP B-31 test dataset.

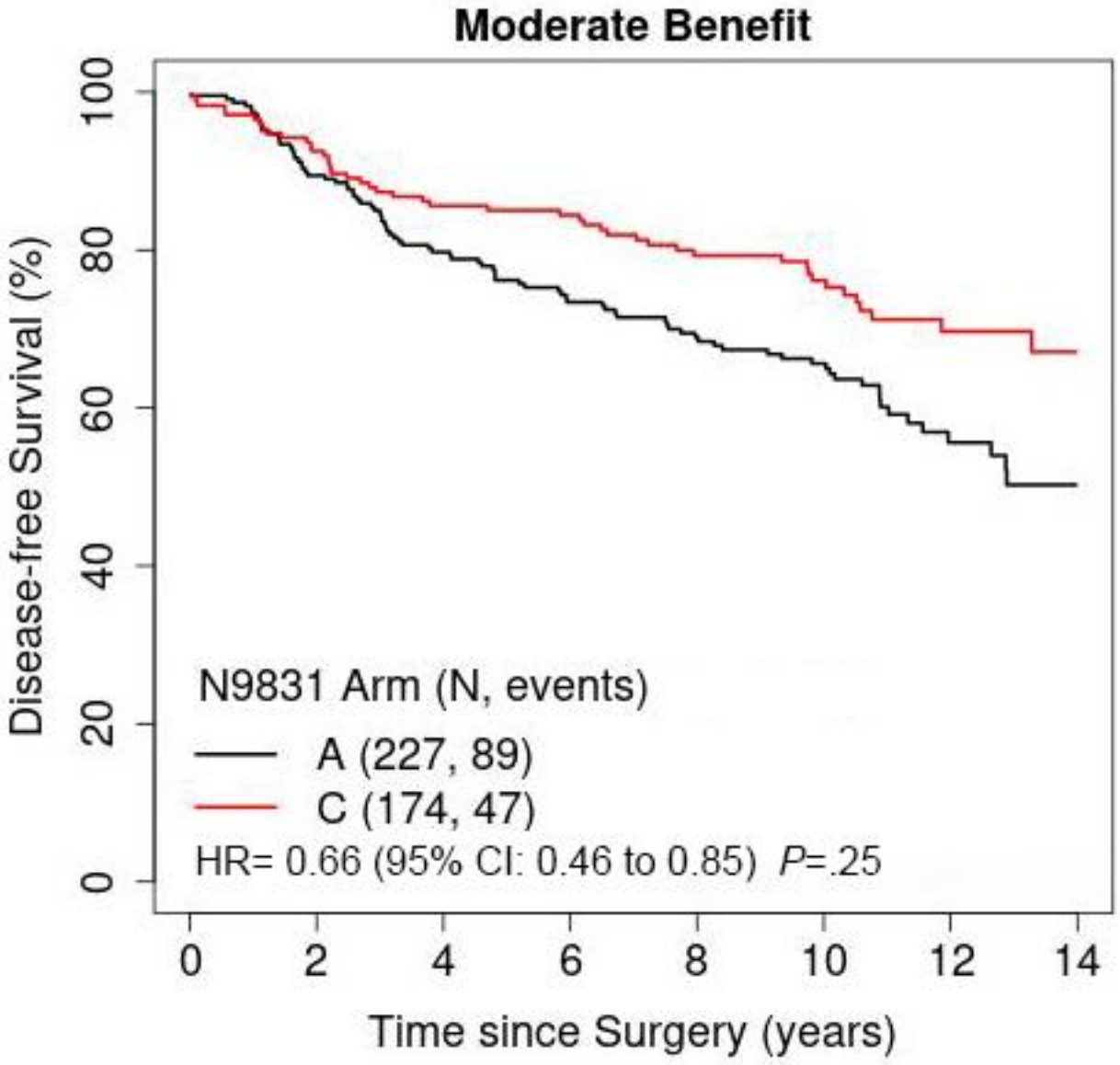
		Modified method		
	Risk	No benefit	Moderate benefit	Large benefit
Original method	No benefit	100	0	0
	Moderate benefit	2	443	4
	Large benefit	9	5	428

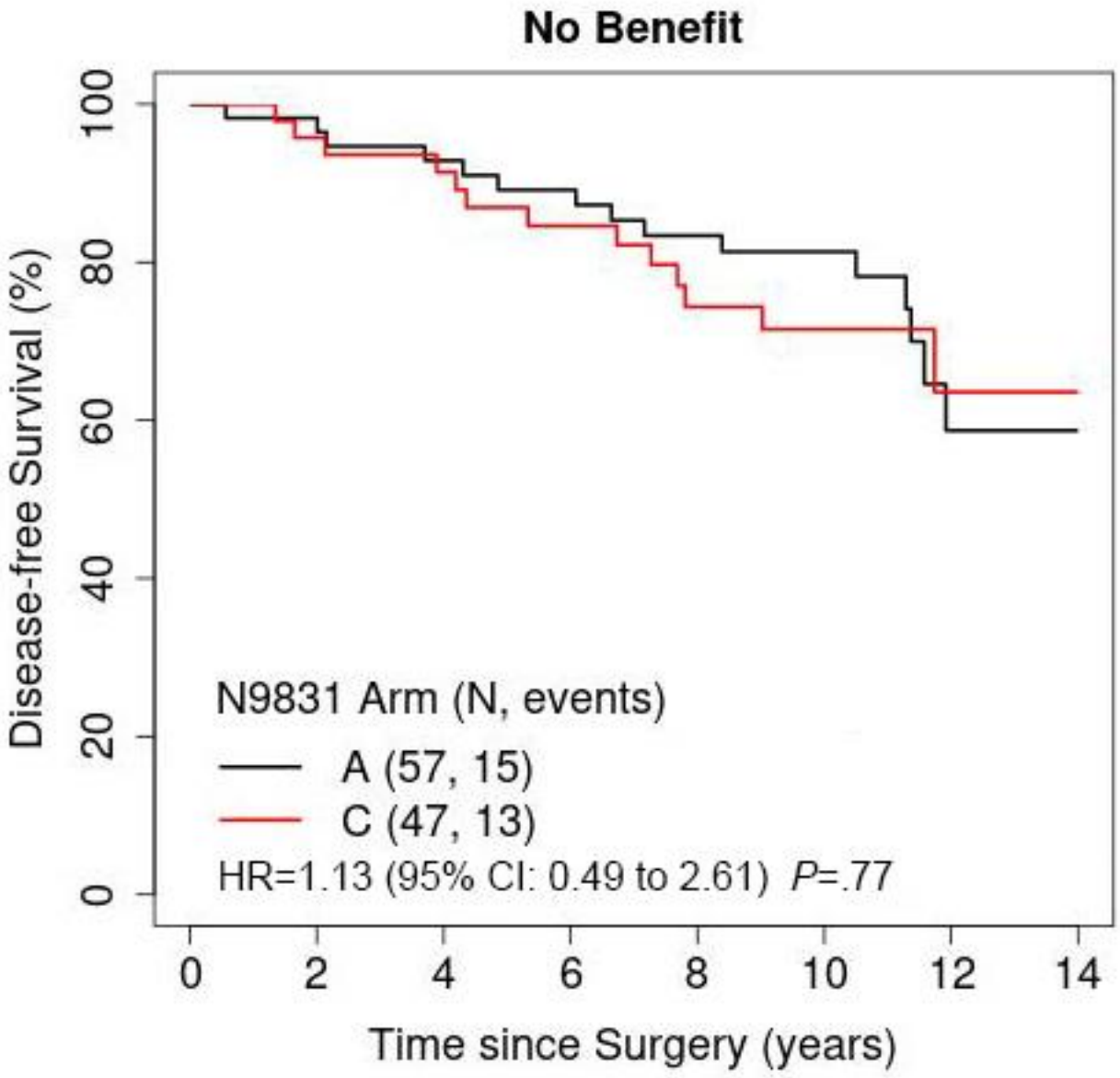
Supplementary Figure 1. Flow of Gene Expression Data and Analysis



Supplementary Figure 2.
Disease-free survival in the 8-gene trastuzumab-benefit groups of Alliance/NCCTG N9831




Moderate Benefit
Disease-free Survival (%)
0
20
40
60
80
100
0
2
4
6
8
10
12
14
Time since Surgery (years)
N9831 Arm (N, events)
A (227, 89)
C (174, 47)
HR= 0.66 (95% CI: 0.46 to 0.85) P=.25


No Benefit
Disease-free Survival (%)
0
20
40
60
80
100
0
2
4
6
8
10
12
14
Time since Surgery (years)
N9831 Arm (N, events)
A (57, 15)
C (47, 13)
HR=1.13 (95% CI: 0.49 to 2.61) P=.77