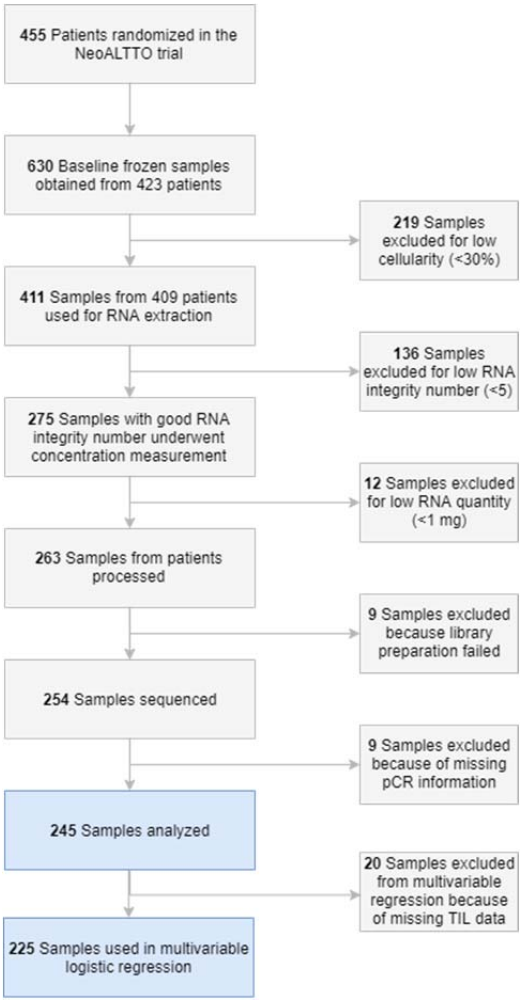
Supplementary Online Content

Powles RL, Redmond D, Sotiriou C, et al. Association of T-cell receptor repertoire use with response to combined trastuzumab-lapatinib treatment of HER2-positive breast cancer: secondary analysis of the NeoALTTO randomized clinical trial. *JAMA Oncol.* Published online June 14, 2018. doi:10.1001/jamaoncol.2018.1564

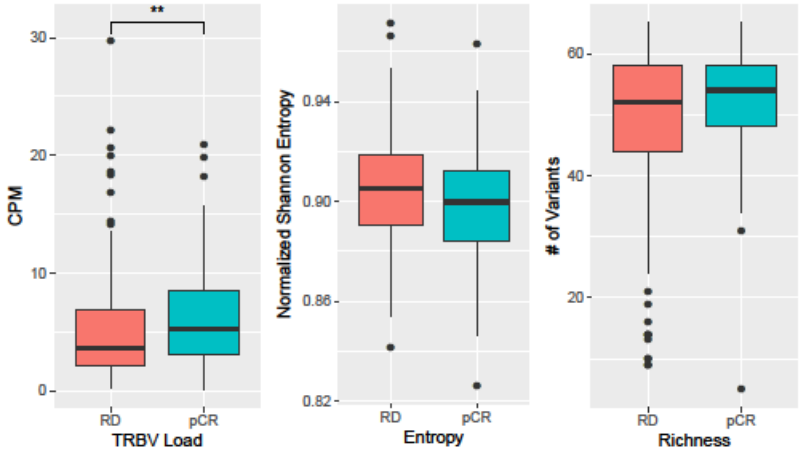
This supplementary material has been provided by the authors to give readers additional information about their work.

eFigure 1. Flow Diagram of Patients and Samples Used in Analysis



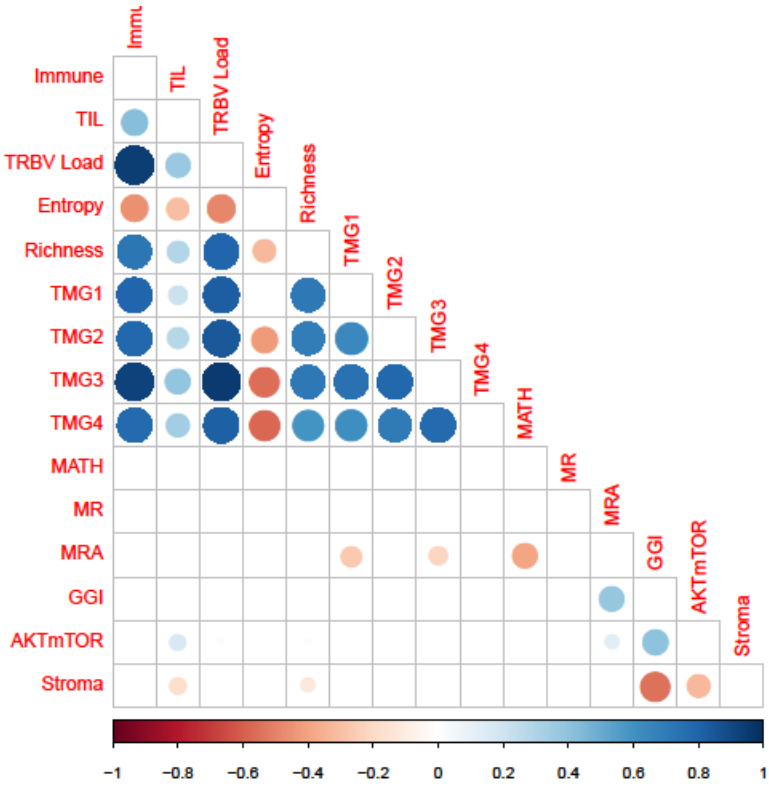
Blue boxes represent sub-cohorts used for various parts of study

eFigure 2. Comparison of Global *TRBV* Metrics



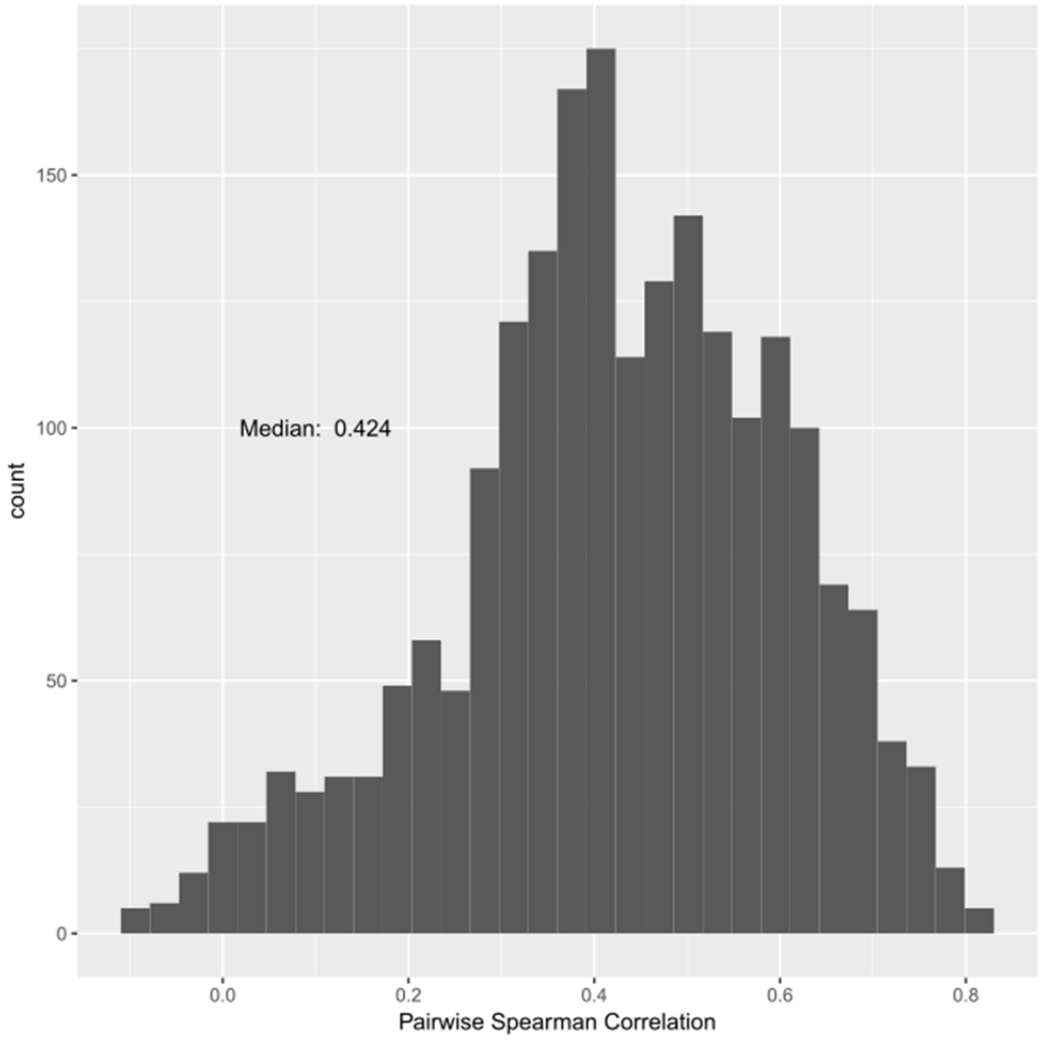
Comparison of total TRBV load in counts per million (CPM), TRBV Shannon entropy, and TRBV richness capturing the number of unique variants per sample between patients with pathologic complete response (pCR) and residual disease (RD) after neoadjuvant therapy.

eFigure 3. Correlation of *TRBV* Usage Metrics and Prognostic Features

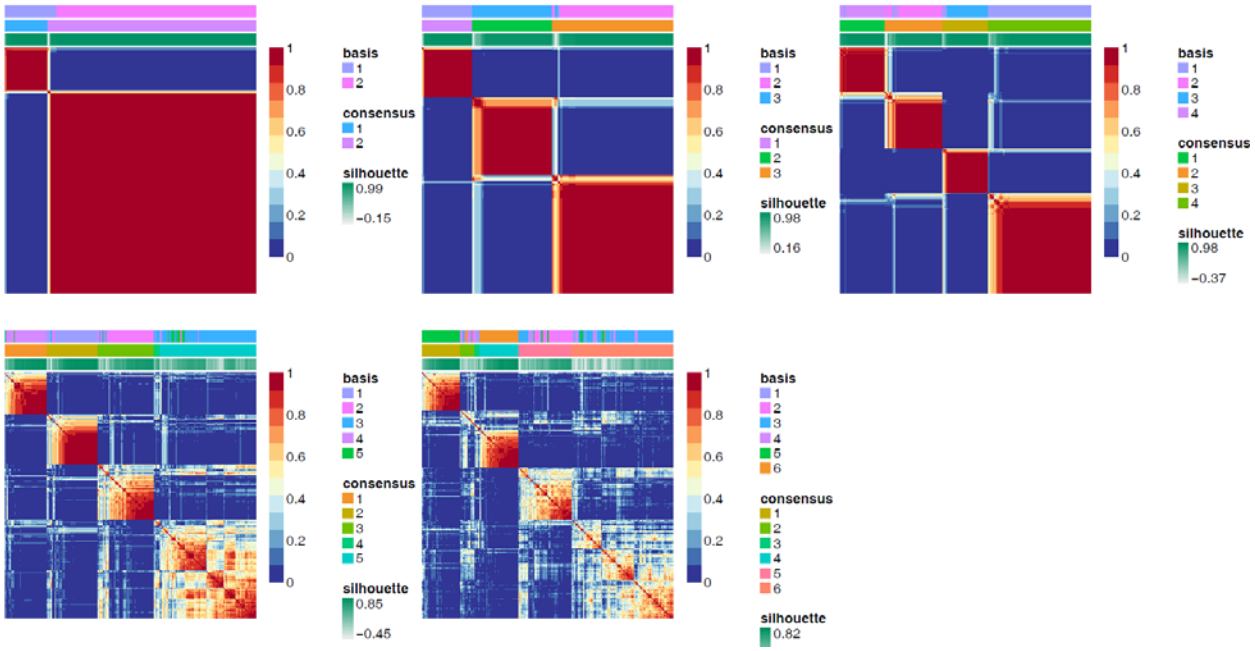


Correlation of expression of molecular signatures, T-cell receptor global metrics, T-cell receptor metagene usage, and whole exome global mutation features. Spearman correlation shown. Correlations with p > 0.05 are hidden. Tumor heterogeneity (MATH), Nonsynonymous mutation rate (MR), and overall mutation rate (MRA) correlations are assessed only in the 140 samples for which matching exome data is available. Correlation with TILs are assessed only in 224 for which TIL information is available. Genomic grade index (GGI), Stroma, and AKTmTOR metagenes are expression signatures prognostic for HER2 targeted therapies extracted from Fumagelli et al.

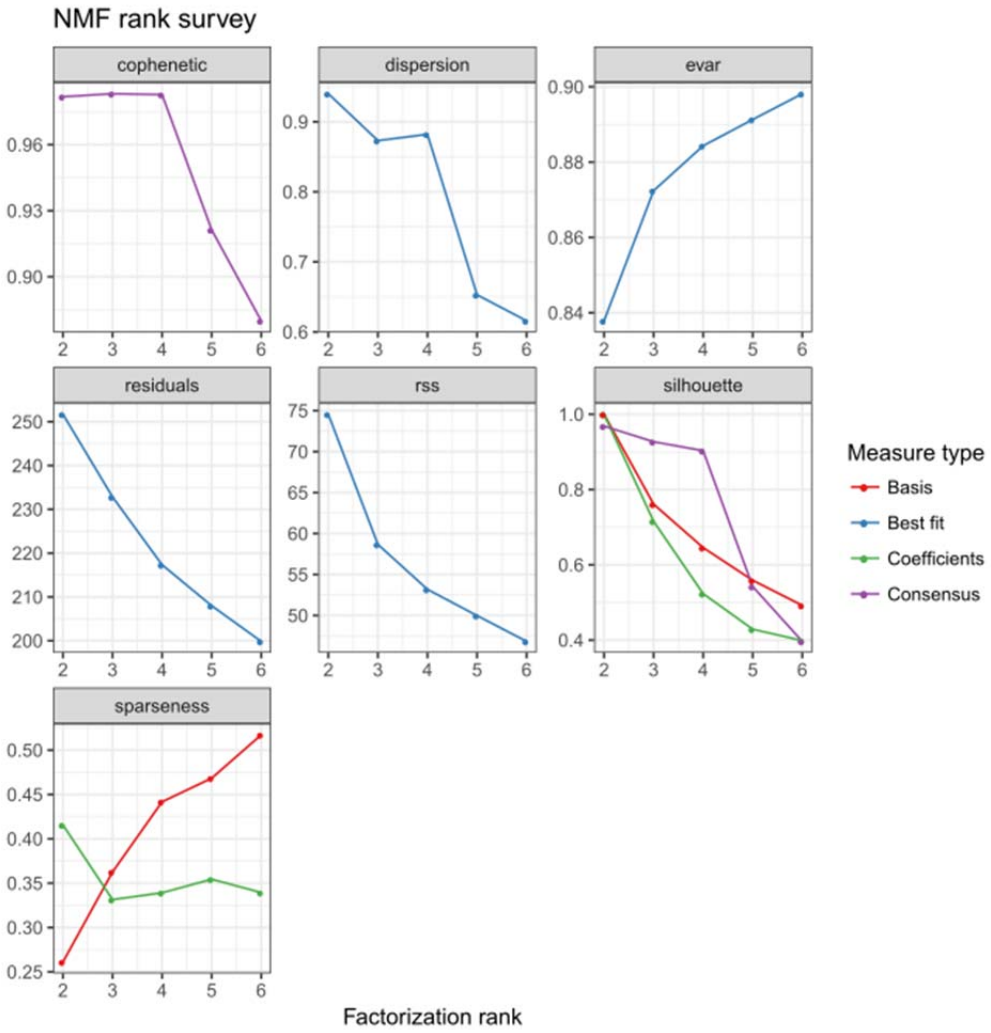
eFigure 4. Pairwise Correlation of Usage in Individual *TRBV* Genes



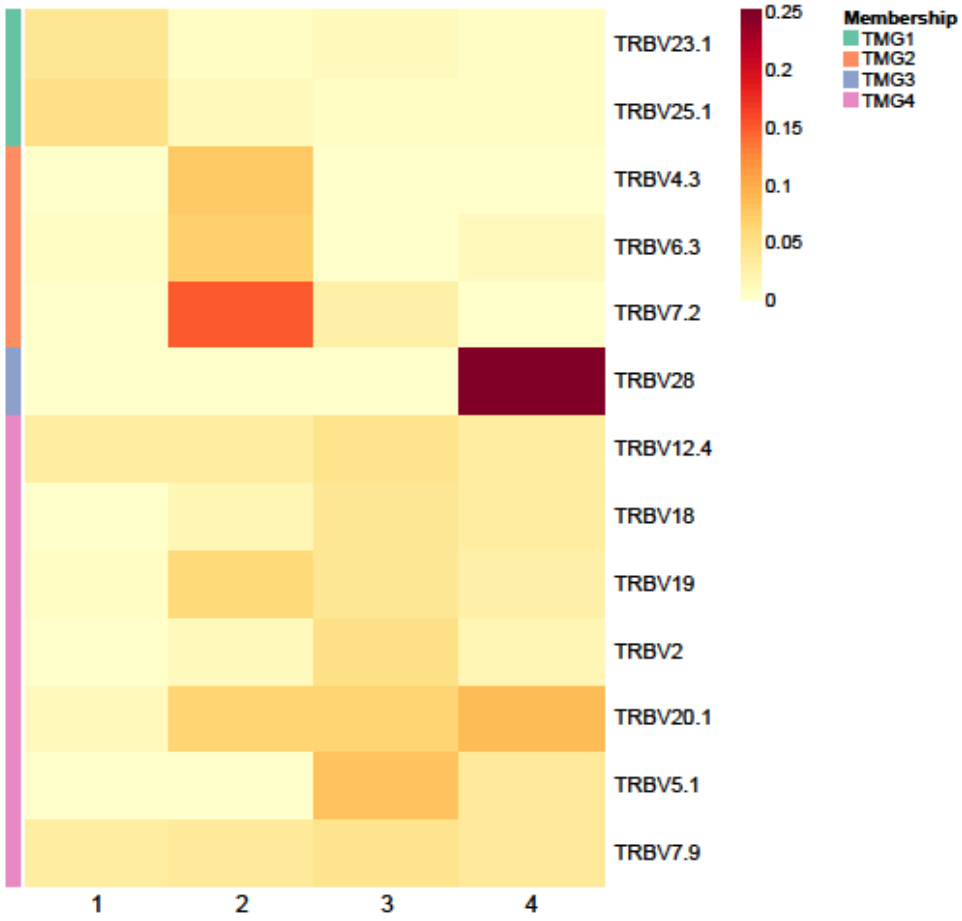
eFigure 5. Comparison of NMF-Based Clustering and Direct Clustering at Various Ranks



eFigure 6. Model Diagnostics of NMF Using Various Rank Representations

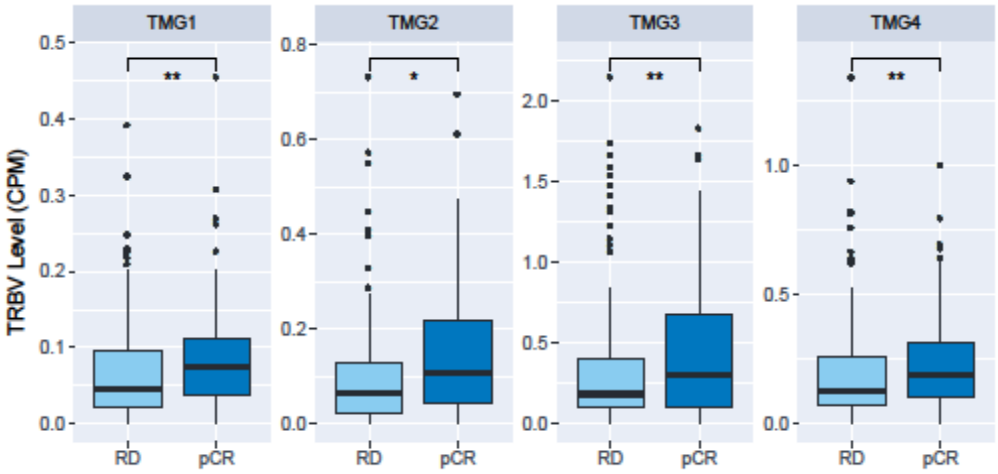


eFigure 7. Heatmap of *TRBV* Gene Use



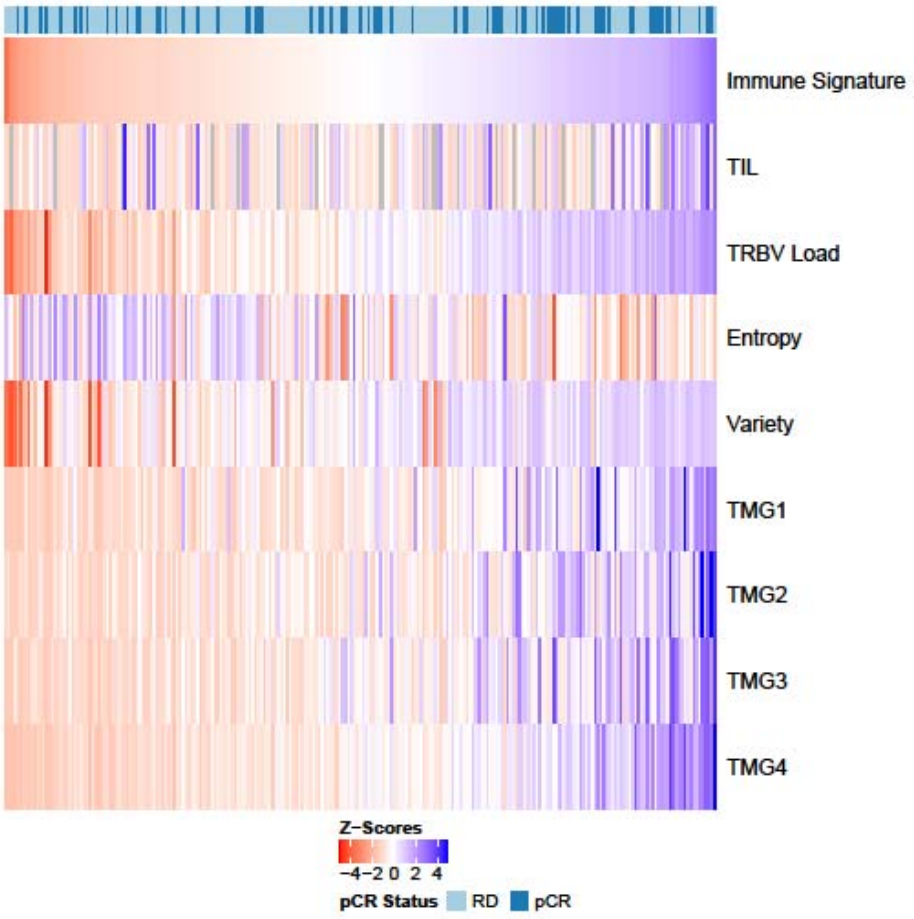
Relative contribution of individual T-cell receptor beta chain variable (TRBV) genes to the four TRBV metagenes (TMGs). TRBV contributions for each metagene are normalized to sum to 1. Only TRBV genes selected as metagene features (**eAppendix 1**) are shown. Membership on the left indicates in which metagene each TRBV gene is used.

eFigure 8. Comparison of Metagene Levels



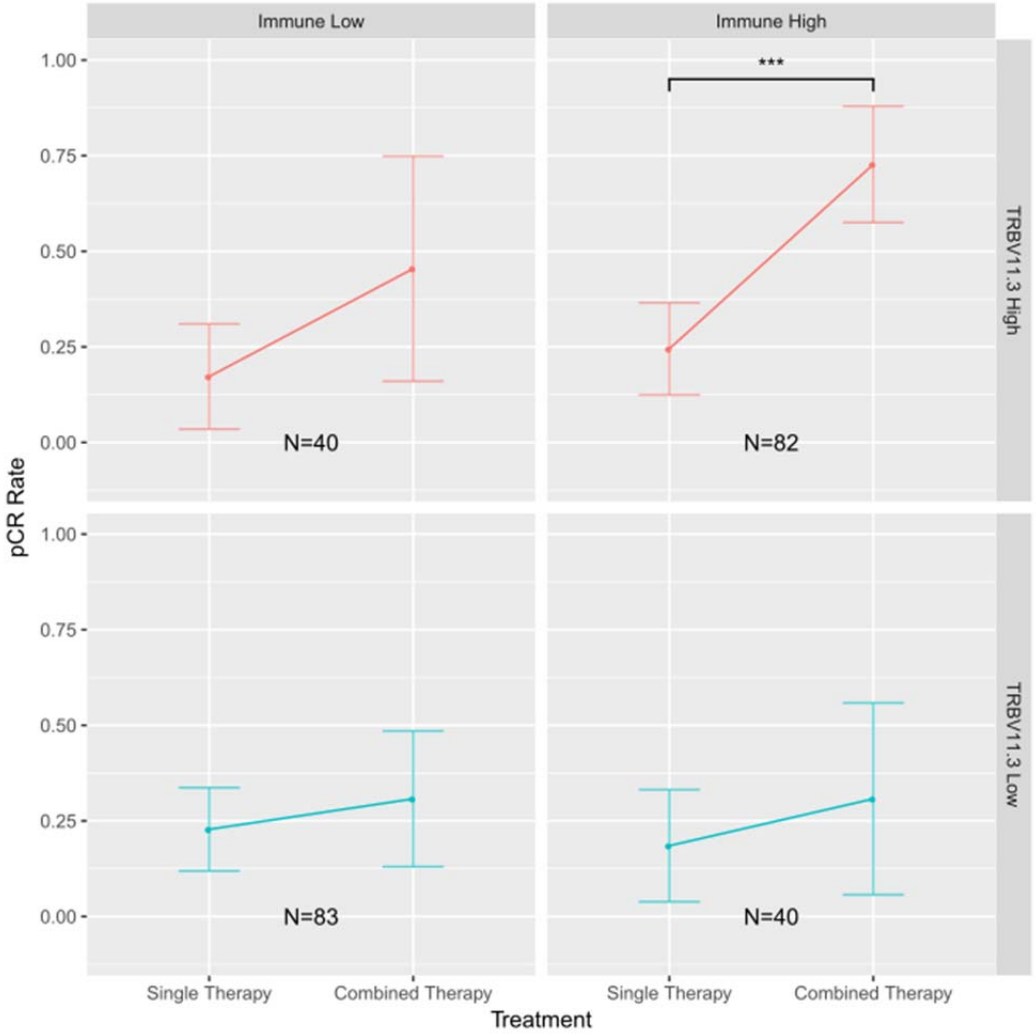
Significance assessed by Wilcoxon rank-sum test. Asterisk (*) indicates significant difference between groups (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).

eFigure 9. Direct Comparison of Immune-Related Features



Z-scores of T-cell global metrics and metagenes when compared to immune enrichment signature expression and tumor infiltrating lymphocyte percentages (TILs). Patients are sorted by increasing immune signature Z-scores. Grey heatmap color indicate missing data.

eFigure 10. Pathologic Complete Response (pCR) Rates Compared Between Treatments Dichotomized by High or Low Immune Enrichment Signature and High and Low Use of *TRBV11-3*



Asterisk (*) indicates significant difference between groups via Fisher's exact test (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$). Error bars indicate 95% confidence interval.

eTable 1. Patient Characteristics in the Sequenced and Analyzed Cohorts

Characteristics		Whole cohort (n=455)	Sequenced Cohort (n=254)	Analyzed Cohort (n=225)	*P* value (Whole vs. Analyzed)[a]	*P* value (Sequenced vs. Analyzed)[a]
Age	**Median**	50	49	49	NA[b]	0.74
	Range	23-80	23-79	26-73		
Tumor Size	**>=T3**	181 (40%)	103 (41%)	92 (41%)	0.78	0.94
	T2	274 (60%)	151 (59%)	133 (59%)		
Nodal Status	**>=N2, Nx or missing**	72 (16%)	38 (15%)	38 (17%)	0.72	0.62
	N0/1	383 (84%)	216 (85%)	187 (83%)		
Hormone Receptor Status	**negative**	223 (49%)	117 (46%)	114 (51%)	0.68	0.36
	positive	232 (51%)	137 (54%)	111 (49%)		
Treatment arm	**L**	154 (34%)	89 (35%)	79 (35%)	0.93	0.96
	T	149 (33%)	79 (31%)	71 (32%)		
	L+T	152 (33%)	86 (34%)	75 (38%)		
pCR	**No**	295 (65%)	166 (65%)	154 (86%)	0.35	0.54
	Yes	160 (35%)	88 (35%)	71 (14%)		

[a]P-values calculated with Fisher exact test for frequencies and Wilcoxon rank sum for age
[b]Raw data unavailable for calculation

eTable 2. IMGT Designation of *TRBV* Genes Included in Each Metagene

Metagene	TRBV gene Included
TMG1	TRBV25.1, TRBV23.1
TMG2	TRBV4.3, TRBV6.3, TRBV7.2
TMG3	TRBV20.1, TRBV2, TRBV5.1, TRBV12.4, TRBV7.9, TRBV18, TRBV19
TMG4	TRBV28

eTable 3. Significance of *TRBV* Features Used in ER-Adjusted Logistic Regression Models to Predict pCR

Feature	OR (95% CI)	P[a]	Adjusted P[b]
TRBV Load	1.44 (1.08-1.95)	0.015	0.12
Entropy	0.749 (0.557-0.996)	0.05	0.15
Richness	1.41 (1.04-1.98)	0.03	0.14
TMG1	1.39 (1.06-1.86)	0.02	0.13
TMG2	1.49 (1.14-1.96)	0.004	0.04
TMG3	1.38 (1.05-1.82)	0.02	0.13
TMG4	1.28 (0.981-1.67)	0.07	0.15
TRBV11-3	1.67 (1.26-2.27)	0.0006	0.04[c]

[a]P-values calculated using the Wald test.
[b]P-values adjusted using the Holm step-down procedure
[c]P-value adjusted for testing of 65 individual TRBV species
Abbreviations: TRBV, T-cell receptor beta chain variable region; TMG, T-cell receptor beta chain variable region metagene

eTable 4. Multivariable Logistic Regression of *TRBV* Features, Clinical Covariates, and Gene Expression Signatures to Predict pCR While Comparing Only Trastuzumab and Combined Therapy in Treatment Arm

Covariate Term	One-marker TMG2 Model		One-marker TRBV11-3 Model		Two-marker Model	
	OR (95% CI)	*P* value[b]	OR (95% CI)	*P* value[b]	OR (95% CI)	*P* value[b]
Age	0.99 (0.96 – 1.03)	0.74	1.00 (0.96 – 1.03)	0.84	1.00 (0.96 – 1.04)	0.97
ER Status	0.22 (0.081 – 0.53)	0.0013	0.28 (0.11 – 0.66)	0.0049	0.24 (0.088 – 0.60)	0.0032
Tumor size (>=T3 vs T2)	0.73 (0.30 – 1.70)	0.47	0.72 (0.30 – 1.66)	0.45	0.74 (0.30 – 1.75)	0.50
Arm (Combination vs Trastuzumab	4.47 (1.85 – 11.85)	0.0014	3.42 (1.49 – 8.42)	0.0049	4.28 (1.74 – 11.68)	0.0025
Grade (1-2 vs 3)	0.59 (0.25 – 1.35)	0.21	0.56 (0.23 – 1.30)	0.18	0.55 (0.22 – 1.30)	0.18
Nodal status (N0/N1 vs N3)	1.47 (0.51 – 4.18)	0.47	1.16 (0.40 – 3.33)	0.78	1.36 (0.46 – 3.94)	0.57
TIL	0.60 (0.34 – 1.00)	0.059	0.69 (0.40 – 1.15)	0.17	0.64 (0.35 – 1.08)	0.11
Immune3	1.70 (0.94 – 3.21)	0.088	1.85 (1.10 – 3.25)	0.025	1.45 (0.77 – 2.84)	0.26
GGI	1.48 (0.85 – 2.71)	0.18	1.50 (0.87 – 2.68)	0.15	1.48 (0.84 – 2.71)	0.19
Stroma1	1.17 (0.70 – 2.01)	0.56	1.12 (0.66 – 1.94)	0.67	1.17 (0.68 – 2.03)	0.58
TCR Terms						
TMG2	0.55 (0.14 – 1.90)	0.37	N/A	N/A	0.95 (0.20 – 3.83)	0.94
TMG2 x Arm[a]	8.13 (1.98 – 42.23)	0.0066	N/A	N/A	4.49 (0.89 – 27.80)	0.08
TRBV11-3	N/A	N/A	0.50 (0.17 – 1.10)	0.14	0.61 (0.19 – 1.51)	0.32
TRBV11-3 x Arm[a]	N/A	N/A	4.51 (1.64 – 16.67)	0.01	3.06 (0.95 – 12.05)	0.08
Likelihood ratio test[c]		0.00099		.0055		.001376

[a]Indicates interaction term in the model
[b]P-values calculated using the Wald test
[c]P-values calculated using a chi-squared test for nested models with and without TCR terms used in each model

eTable 5. Multivariable Logistic Regression of *TRBV* Features, Clinical Covariates, and Gene Expression Signatures to Predict pCR While Comparing Only Lapatinib and Combined Therapy in Treatment Arm

Covariate Term	One-marker TMG2 Model		One-marker TRBV11-3 Model		Two-marker Model	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age	0.98 (0.94 – 1.01)	0.21	0.98 (0.95 – 1.02)	0.41	0.98 (0.95 – 1.02)	0.42
ER Status	0.33 (0.14 – 0.77)	0.0012	0.47 (0.20 – 1.07)	0.077	0.37 (0.15 – 0.88)	0.03
Tumor size (>=T3 vs T2)	0.60 (0.24 – 1.4)	0.25	0.58 (0.23 – 1.39)	0.23	0.60 (0.24 – 1.44	0.26
Arm (Combination vs Lapatinib	5.68 (2.45 – 14.22)	.000096	4.37 (1.95 – 10.30)	0.00047	5.69 (2.36 – 14.97)	0.00020
Grade (1-2 vs 3)	1.02 (0.44 – 2.37)	0.96	0.88 (0.37 – 2.06)	0.76	0.86 (0.35 – 2.10)	0.74
Nodal status (N0/N1 vs N3)	1.10 (0.38 – 3.08)	0.85	0.88 (0.28 – 2.59)	0.82	0.97 (0.31 – 2.88)	0.95
TIL	0.69 (0.39 – 1.18)	0.19	0.82 (0.48 – 1.35)	0.44	0.75 (0.42 – 1.26)	0.30
Immune3	0.94 (0.52 – 1.69)	0.82	0.92 (0.54 – 1.60)	0.78	0.64 (0.33 – 1.23)	0.19
GGI	2.22 (1.17 – 4.42)	.018	2.23 (1.20 – 4.39)	0.015	2.32 (1.22 – 4.67)	0.013
Stroma1	1.39 (0.81 – 2.48)	0.24	1.36 (0.80 – 2.43)	0.28	1.43 (0.83 – 2.60)	0.22
TCR Terms						
TMG2	1.16 (0.59 – 2.10)	0.64	N/A	N/A	0.85 (0.35 – 1.82)	0.69
TMG2 x Arm[a]	4.74 (1.78 – 15.57)	0.0047	N/A	N/A	7.59 (2.12 – 32.96)	0.0036
TRBV11-3	N/A	N/A	1.99 (0.88 – 4.49)	0.09	3.20 (1.15 – 9.65)	0.029
TRBV11-3 x Arm[a]	N/A	N/A	1.61 (0.61 – 4.66)	0.35	0.76 (0.21 – 2.76)	0.68
Likelihood ratio test[c]		0.0027		.0011		.000026

[a]Indicates an interaction term in the model
[b]P-values calculated using the Wald test
[c]P-values calculated using a chi-squared test for nested models with and without TRBV terms used in each model

eAppendix 1. Methods

RNA sequencing

From 254 RNA samples, Illumina paired-end sequencing libraries were successfully constructed and subjected to sequencing on the Illumina HiSeq 2500 system as described previously[1]. Read pairs were trimmed using Trimmomatic[2], alignment was performed using STAR[3]. Immune, proliferation and treatment response metagene scores, calculated as the weighted sum of the log-FPKM expression of their member genes, were taken from the original publication[1].

In this analysis, we focus on the usage of TRBV genes due to their increased sequence diversity compared to alpha chain variable genes. We defined the following TRBV population-wide metrics:
(i) Total TRBV Load was calculated as the sum of all reads aligning to any TRBV variant in a sample.
(ii) TRBV richness was calculated as the number of distinct TRBV-genes per sample whose usage was non-zero. (iii) TRBV entropy was calculated as the Shannon diversity index of TRBV usage normalized by the natural logarithm of TRBV richness. All post-processing and statistical analysis was conducted in R[4].

Hierarchical clustering of TRBV usage

The Euclidean distance matrix was calculated from the natural log of TRBV CPMs across patients. To stabilize numerical computations, a pseudo-count of 1 was added to each count value before taking the logarithm. Samples were clustered using average linkage, and the stability of the cluster was evaluated using the approximately unbiased p-value from multiscale bootstrap resampling using the pvclust package[5].

Non-negative matrix factorization:

Non-negative matrix factorization (NMF) was performed using 150 runs of the Brunet et al. algorithm from the nmf package[6,7]. A factorization rank of 4 was chosen based on 50 runs for ranks 2 through 6. The highest rank at which the cophenetic coefficient decreased was chosen, as suggested in the original manuscript by Brunet et al. Individual TRBV genes from NMF were selected as significant contributors to a TRBV metagene using the feature selection method described in Carmona-Saez et al.[8]. The method considers TRBV genes by their contributing weights to each factor, and for each factor selects a sparse set of TRBV genes that contribute uniquely to that factor and no others. The median usage of these factor-specific TRBV genes were used to define the four T-cell receptor metagenes (TMG).

Logistic regression

Population-wide TRBV metrics and NMF-derived TMGs were considered in covariate-adjusted logistic regression to predict pCR. In ER-adjusted analyses, we corrected for estrogen receptor (ER) status and

tested nested models with and without each TRBV metric using the Wald approximation of the likelihood ratio test. We corrected for multiple testing using the Holm step-down procedure[9]. In multivariable models, we adjusted for age (continuous), tumor size (>=T3 vs. T2), grade (1-2 vs. 3), estrogen receptor status (negative vs. positive) and node status (N0/N1 vs. N2-N3), TIL counts, immune gene signature (Immune3), a proliferation signature (GGI), a stromal signature (Stroma1) and treatment arm in the 225 patients for which all data was available. Gene expression signatures were selected by their significance in the analysis conducted by Fumagelli et al.[1] Based on high correlation between gene expression signatures for similar phenotypes, only one immune signature and one stromal signature was chosen. All covariates were standardized. Each TRBV feature was included as a base term and as an interaction term with the treatment arm. Terms were tested for significance in their respective models using the Wald approximation of the likelihood ratio test. The significance of TCR effect overall in each model was tested using a chi-squared test of the likelihood ratio of nested models with or without TCR terms with 2 or 4 degrees of freedom depending on the number of TCR terms in the model.

eAppendix 2. Statistical Analysis Logs

```
> load("NeoALTTOData.RData")
> regressionvariables = c("Entropy", "Variety", "TTL", "TMG1", "TMG2", "TMG3",
"TMG4", "TIL", "TRBV11.3", "Immune3")
> pvalueAll = NULL
> for (v in c(regressionvariables)) {
+     print(paste(v, " ER-adjusted model", sep=""))
+     logReg = glm(as.formula(paste("factor(pCR4FL) ~ ERCat + ", v)), data =
immuneData, family = binomial)
+     print("ORs")
+     print(exp(coef(logReg)))
+     print("95% CIs")
+     print(exp(confint(logReg)))
+     print("P-values")
+     pv = coef(summary(logReg))[,4]
+     print(pv)
+     pvalueAll = c(pvalueAll, pv[3])
+ }
[1] "Entropy ER-adjusted model"
[1] "ORs"
  (Intercept) ERCatPOSITIVE       Entropy
    0.6542450     0.4154185     0.7490110
[1] "95% CIs"
                  2.5 %    97.5 %
(Intercept)   0.4539551 0.9344040
ERCatPOSITIVE 0.2326103 0.7278861
Entropy       0.5572375 0.9959316
[1] "P-values"
  (Intercept) ERCatPOSITIVE       Entropy
  0.020882390   0.002468847   0.050036478
[1] "Variety ER-adjusted model"
[1] "ORs"
  (Intercept) ERCatPOSITIVE       Variety
    0.6550552     0.4086921     1.4100306
[1] "95% CIs"
                  2.5 %    97.5 %
(Intercept)   0.4535834 0.9372531
ERCatPOSITIVE 0.2286949 0.7166964
Variety       1.0417050 1.9765512
[1] "P-values"
  (Intercept) ERCatPOSITIVE       Variety
  0.021944885   0.002076351   0.034386457
[1] "TTL ER-adjusted model"
[1] "ORs"
  (Intercept) ERCatPOSITIVE           TTL
    0.6522645     0.4112384     1.4431066
[1] "95% CIs"
                  2.5 %    97.5 %
(Intercept)   0.4510494 0.9343756
ERCatPOSITIVE 0.2298190 0.7220754
```

```
TTL           1.0824300 1.9539897
[1] "P-values"
  (Intercept) ERCatPOSITIVE           TTL
  0.021084820   0.002280884   0.014617283
[1] "TMG1 ER-adjusted model"
[1] "ORs"
  (Intercept) ERCatPOSITIVE          TMG1
    0.6555476     0.4158096     1.3939668
[1] "95% CIs"
                  2.5 %    97.5 %
(Intercept)   0.4536333 0.9387793
ERCatPOSITIVE 0.2325109 0.7297306
TMG1          1.0627697 1.8556302
[1] "P-values"
  (Intercept) ERCatPOSITIVE          TMG1
  0.022458760   0.002560404   0.018718335
[1] "TMG2 ER-adjusted model"
[1] "ORs"
  (Intercept) ERCatPOSITIVE          TMG2
    0.6706538     0.3879484     1.4854647
[1] "95% CIs"
                  2.5 %    97.5 %
(Intercept)   0.4641965 0.9608526
ERCatPOSITIVE 0.2145333 0.6860831
TMG2          1.1374688 1.9638309
[1] "P-values"
  (Intercept) ERCatPOSITIVE          TMG2
  0.030885966   0.001362001   0.004108943
[1] "TMG3 ER-adjusted model"
[1] "ORs"
  (Intercept) ERCatPOSITIVE          TMG3
    0.6500281     0.4271215     1.2787722
[1] "95% CIs"
                  2.5 %    97.5 %
(Intercept)   0.4509275 0.9283573
ERCatPOSITIVE 0.2398574 0.7471359
TMG3          0.9806118 1.6724736
[1] "P-values"
  (Intercept) ERCatPOSITIVE          TMG3
  0.019042315   0.003248887   0.067817125
[1] "TMG4 ER-adjusted model"
[1] "ORs"
  (Intercept) ERCatPOSITIVE          TMG4
    0.6679252     0.3978009     1.3802526
[1] "95% CIs"
                  2.5 %    97.5 %
(Intercept)   0.4630027 0.9554687
ERCatPOSITIVE 0.2211559 0.7005154
TMG4          1.0548370 1.8186412
[1] "P-values"
  (Intercept) ERCatPOSITIVE          TMG4
  0.028544291   0.001669548   0.019272749
[1] "TIL ER-adjusted model"
[1] "ORs"
  (Intercept) ERCatPOSITIVE           TIL
```

```
    0.6871977     0.4142081     0.8897962
[1] "95% CIs"
                  2.5 %    97.5 %
(Intercept)   0.4690444 0.9982862
ERCatPOSITIVE 0.2268940 0.7418425
TIL           0.6528438 1.1897252
[1] "P-values"
  (Intercept) ERCatPOSITIVE           TIL
  0.050856302   0.003447311   0.443054719
[1] "TRBV11.3 ER-adjusted model"
[1] "ORs"
  (Intercept) ERCatPOSITIVE      TRBV11.3
    0.6398453     0.4233211     1.6733469
[1] "95% CIs"
                  2.5 %    97.5 %
(Intercept)   0.4401136 0.9210190
ERCatPOSITIVE 0.2343655 0.7500134
TRBV11.3      1.2606328 2.2708748
[1] "P-values"
  (Intercept) ERCatPOSITIVE      TRBV11.3
  0.017445679   0.003663961   0.000577129
[1] "Immune3 ER-adjusted model"
[1] "ORs"
  (Intercept) ERCatPOSITIVE       Immune3
    0.6509423     0.4188257     1.3703523
[1] "95% CIs"
                  2.5 %    97.5 %
(Intercept)   0.4507529 0.9312305
ERCatPOSITIVE 0.2346431 0.7339409
Immune3       1.0368850 1.8270369
[1] "P-values"
  (Intercept) ERCatPOSITIVE       Immune3
  0.020019048   0.002700052   0.028665040

> adjustedPV = p.adjust(p = pvalueAll, method="holm")

> print("Holm-adjusted P-values")
[1] "Holm-adjusted P-values"

> print("Adjusted p-value for TRBV11.3 does not match eTable 2 because it is not
being tested again all 65 other individual TRBV regressions here")
[1] "Adjusted p-value for TRBV11.3 does not match eTable 2 because it is not being
tested again all 65 other individual TRBV regressions here"

> print(adjustedPV)
   Entropy    Variety        TTL       TMG1       TMG2       TMG3       TMG4
0.15010943 0.14332520 0.11693827 0.13102835 0.03698049 0.15010943 0.13102835
       TIL   TRBV11.3    Immune3
0.44305472 0.00577129 0.14332520

> print("Multivariate Regression")
[1] "Multivariate Regression"
```

```
> logReg_noTCR = glm("factor(pCR4FL) ~
Age+Hisgrade+Node+Stage+ERCat+TIL+Immune3+GGI+Stroma1+Both", data=immuneComplete,
family = binomial)

> logReg_TMG2 = glm("factor(pCR4FL) ~
Age+Hisgrade+Node+Stage+ERCat+TIL+Immune3+GGI+Stroma1+Both+TMG2+TMG2*Both", data =
immuneData, family = binomial)

> print("ORs")
[1] "ORs"

> print(exp(coef(logReg_TMG2)))
  (Intercept)           Age      Hisgrade          Node         Stage
    0.8682044     0.9893395     0.7294473     1.3062517     0.7227358
ERCatPOSITIVE           TIL       Immune3           GGI       Stroma1
    0.2980545     0.6508501     1.2457094     1.9830365     1.4404102
         Both          TMG2     Both:TMG2
    4.5082353     0.9304869     5.3829624

> print("95% CIs")
[1] "95% CIs"

> print(exp(confint(logReg_TMG2)))
                  2.5 %     97.5 %
(Intercept)   0.1726156  4.3270113
Age           0.9604253  1.0185618
Hisgrade      0.3662858  1.4372194
Node          0.5529220  2.9831080
Stage         0.3544167  1.4416791
ERCatPOSITIVE 0.1407553  0.6036092
TIL           0.4126340  0.9901476
Immune3       0.7641858  2.0585970
GGI           1.2363602  3.3054631
Stroma1       0.9511479  2.2441541
Both          2.2135036  9.5860739
TMG2          0.5048707  1.6089222
Both:TMG2     2.0495357 17.3801537

> print("P-values")
[1] "P-values"

> print(coef(summary(logReg_TMG2))[,4])
  (Intercept)           Age      Hisgrade          Node         Stage
 8.627997e-01  4.726940e-01  3.638389e-01  5.316665e-01  3.622564e-01
ERCatPOSITIVE           TIL       Immune3           GGI       Stroma1
 1.059193e-03  5.346074e-02  3.826305e-01  6.179601e-03  9.417252e-02
         Both          TMG2     Both:TMG2
 5.094433e-05  8.037474e-01  1.917145e-03

> print("Partial Chi-squared")
[1] "Partial Chi-squared"

> print(anova(logReg_TMG2, logReg_noTCR, test="Chisq"))
Analysis of Deviance Table
```

```
Model 1: factor(pCR4FL) ~ Age + Hisgrade + Node + Stage + ERCat + TIL +
    Immune3 + GGI + Stroma1 + Both + TMG2 + TMG2 * Both
Model 2: factor(pCR4FL) ~ Age + Hisgrade + Node + Stage + ERCat + TIL +
    Immune3 + GGI + Stroma1 + Both
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       212     218.01
2       214     234.61 -2  -16.594 0.0002493 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> logReg_TRBV11.3 = glm("factor(pCR4FL) ~
Age+Hisgrade+Node+Stage+ERCat+TIL+Immune3+GGI+Stroma1+Both+TRBV11.3+TRBV11.3*Both",
data = immuneData, family = binomial)

> print("ORs")
[1] "ORs"

> print(exp(coef(logReg_TRBV11.3)))
  (Intercept)           Age      Hisgrade          Node         Stage
    0.7670703     0.9923567     0.6858597     1.1321121     0.7012225
ERCatPOSITIVE           TIL       Immune3           GGI       Stroma1
    0.3690947     0.7186009     1.2822373     1.9714604     1.4086936
         Both      TRBV11.3 Both:TRBV11.3
    3.6871640     0.9922888     2.7168455

> print("95% CIs")
[1] "95% CIs"

> print(exp(confint(logReg_TRBV11.3)))
                  2.5 %    97.5 %
(Intercept)   0.1529545 3.7853120
Age           0.9636743 1.0215298
Hisgrade      0.3423703 1.3558744
Node          0.4723982 2.6111130
Stage         0.3439026 1.3964011
ERCatPOSITIVE 0.1794721 0.7337999
TIL           0.4605583 1.0844208
Immune3       0.8290502 2.0182402
GGI           1.2440927 3.2486809
Stroma1       0.9303328 2.1976530
Both          1.8641218 7.4721507
TRBV11.3      0.5598915 1.6904579
Both:TRBV11.3 1.2262508 7.0400160

> print("P-values")
[1] "P-values"

> print(coef(summary(logReg_TRBV11.3))[,4])
  (Intercept)           Age      Hisgrade          Node         Stage
 0.7448547183  0.6043545854  0.2808792341  0.7746227042  0.3185618358
ERCatPOSITIVE           TIL       Immune3           GGI       Stroma1
 0.0053238690  0.1291866847  0.2709153529  0.0054036686  0.1164623761
         Both      TRBV11.3 Both:TRBV11.3
 0.0002139193  0.9776741415  0.0241128863
```

```
> print("Partial Chi-squared")
[1] "Partial Chi-squared"

> print(anova(logReg_TRBV11.3, logReg_noTCR, test="Chisq"))
Analysis of Deviance Table

Model 1: factor(pCR4FL) ~ Age + Hisgrade + Node + Stage + ERCat + TIL +
    Immune3 + GGI + Stroma1 + Both + TRBV11.3 + TRBV11.3 * Both
Model 2: factor(pCR4FL) ~ Age + Hisgrade + Node + Stage + ERCat + TIL +
    Immune3 + GGI + Stroma1 + Both
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       212     223.71
2       214     234.61 -2  -10.904 0.004287 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> logReg_TwoMarker =
glm("factor(pCR4FL)~Age+Hisgrade+Node+Stage+ERCat+TIL+Immune3+GGI+Stroma1+TRBV11.3+TM
G2+Both+TMG2*Both+TRBV11.3*Both", data = immuneData, family = binomial)

> print("ORs")
[1] "ORs"

> print(exp(coef(logReg_TwoMarker)))
  (Intercept)           Age      Hisgrade          Node         Stage
    0.7379238     0.9931122     0.6760284     1.2738127     0.7182042
ERCatPOSITIVE           TIL       Immune3           GGI       Stroma1
    0.3231785     0.6822662     1.0469192     1.9516364     1.4558603
     TRBV11.3          TMG2          Both     TMG2:Both TRBV11.3:Both
    1.1709743     0.9586128     4.5097878     5.0314029     1.7613643

> print("95% CIs")
[1] "95% CIs"

> print(exp(confint(logReg_TwoMarker)))
                  2.5 %     97.5 %
(Intercept)   0.1404831  3.8004121
Age           0.9635590  1.0232109
Hisgrade      0.3319710  1.3540532
Node          0.5299886  2.9491304
Stage         0.3484927  1.4449191
ERCatPOSITIVE 0.1519183  0.6590409
TIL           0.4297483  1.0436203
Immune3       0.6217976  1.7780154
GGI           1.2184108  3.2470061
Stroma1       0.9571743  2.2837576
TRBV11.3      0.6180503  2.1764107
TMG2          0.4813744  1.7771499
Both          2.1616003  9.8660234
TMG2:Both     1.6164150 18.4592805
TRBV11.3:Both 0.7014197  4.8949934

> print("P-values")
[1] "P-values"
```

```
> print(coef(summary(logReg_TwoMarker))[,4])
  (Intercept)           Age      Hisgrade          Node         Stage
 7.165441e-01  6.504039e-01  2.728981e-01  5.778431e-01  3.593805e-01
ERCatPOSITIVE           TIL       Immune3           GGI       Stroma1
 2.428409e-03  9.024251e-02  8.635127e-01  7.278027e-03  8.894373e-02
     TRBV11.3          TMG2          Both     TMG2:Both TRBV11.3:Both
 6.172289e-01  8.975113e-01  9.088361e-05  8.848980e-03  2.477146e-01

> print("Partial Chi-squared")
[1] "Partial Chi-squared"

> print(anova(logReg_TwoMarker, logReg_noTCR, test="Chisq"))
Analysis of Deviance Table

Model 1: factor(pCR4FL) ~ Age + Hisgrade + Node + Stage + ERCat + TIL +
    Immune3 + GGI + Stroma1 + TRBV11.3 + TMG2 + Both + TMG2 *
    Both + TRBV11.3 * Both
Model 2: factor(pCR4FL) ~ Age + Hisgrade + Node + Stage + ERCat + TIL +
    Immune3 + GGI + Stroma1 + Both
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       210     213.72
2       214     234.61 -4   -20.89 0.0003329 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

eReferences

1. Fumagalli D, Venet D, Ignatiadis M, et al. RNA Sequencing to Predict Response to Neoadjuvant Anti-HER2 Therapy: A Secondary Analysis of the NeoALTTO Randomized Clinical Trial. *JAMA oncology.* 2016:1-8.
2. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114-2120.
3. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15-21.
4. *R: A language and environment for statistical computing* [computer program]. Vienna, Austria: R Foundation for Statistical Computing; 2008.
5. Suzuki R, Shimodaira H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics.* 2006;22:1540-1542.
6. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A.* 2004;101:4164-4169.
7. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics.* 2010;11:367.
8. Carmona-saez P, Pascual-marqui RD, Tirado F, Carazo JM, Pascual-montano A. Biclustering of gene expression data by non-smooth non-negative matrix factorization. *BMC Bioinformatics.* 2006;18:1-18.
9. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics.* 1979;6:65-70.