

Figure legends

Figure 1. An overview of the strategy to identify modules of effector T cell cytolytic activity (ECA). a, Work flow of our strategy to develop ECA modules. b, t-SNE plot of the two groups, ECA-High (N = 503) and ECA-Low (N = 487), which were identified based on the average enrichment scores of T cell cytolytic activity (as defined by GZMA and PRF1 genes). Samples in red are in ECA-High group while samples in blue are in ECA-Low group. Feature plots demonstrating expression of some key immune-related genes. Dark blue points represent a relative over-expression of that specific gene while grey points represent a relative under-expression. c, Spectral decomposition followed by Jackstraw analyses helped identify ten functional modules. Each module was partitioned into training and test datasets. Random Forest (RF) prediction models were then developed from each module-derived training set, and validated on the test data followed by an independent dataset.

Figure 2. Best RF prediction model is derived from ECA module 5 (MOD5) in TNBC. a, Diverging barplot indicating relative gene expression levels of top 60 genes in the RF model. The length of each bar denotes log₂ fold-change values. Red indicates up-regulation in ECA-High group while blue indicates down-regulation. b, c, Barplots (with standard error) of log₂ gene expression of the top genes in the Random Forest (RF) prediction model in each subtype of breast cancer within TCGA, and METABRIC datasets, respectively. d, Barplots (with standard error) of log₂ gene expression of the top genes in the Random Forest (RF) prediction model in GSE26304 (N = 114) composed of patients with ductal carcinoma in situ (DCIS; N = 31), mixed DCIS and invasive breast adenocarcinoma (IDC; N = 77). Some normal samples were also identified in this dataset. Statistical analysis was performed by one-way analysis of variance.

Figure 3. RF prediction model derived from ECA module 5 (MOD5) is prognostic in TNBC but not in hormone receptor positive tumors within TCGA dataset. a, b, c, Kaplan-Meier survival curves for progression-free survival in TCGA ER+ (N=537), HER2+ (177) and TNBC (N=150) samples, respectively.

Figure 4. RF prediction model derived from ECA module 5 (MOD5) is prognostic in TNBC but not in hormone receptor positive tumors within

METABRIC dataset. a, b, c, Kaplan-Meier survival curves for progression-free survival in METABRIC ER+ (N=1,398), HER2+ (247) and TNBC (N=320) samples, respectively.

Figure 5. RF prediction model derived from ECA module 5 (MOD5) is prognostic and predictive of cancer immunotherapy in patients with advanced melanoma. a, Diverging barplot indicating relative gene expression levels of top 60 genes in the RF model generated from ECA MOD5 within predicted ECA-High and -Low groups of patients from GSE91066 dataset. The length of each bar denotes log2 fold-change values. b, ROC curve for all the patients (N = 51) using their predicted ECA values. Non-responders are expected fall under ECA-Low group while responders in ECA-High. Area under the curve (AUC) = 0.64 (95% CI - 0.45-0.83). c, Progression-free survival for cases that were stratified into ECA-High and -Low with a log-rank test p-value of 0.029. d, Barplots (with standard error) of log2 gene expression of the top genes in the RF prediction model in responders and non-responders matched by treatment cohort (N=43). Statistical analysis was done using t-test. e, Heatmap showing average expression levels of the top genes from the RF model in patients before and after treatment with Nivolumab. Two cohorts of patients are shown separately - one set of patients are ipilimumab naive patients (NIV3-NAIVE; N = 25) and the second set of patients were previously treated with ipilimumab (NIV3-PROG; N = 26). Red indicates high expression and blue indicates low expression of genes.

Figure 6. ECA module characterization and their prognostic value in multiple solid tumors. a, Plot showing prognostic significance of each module in various solid tumors. Each module was used to develop a Random Forest prediction model and each such model was then prognosticated in several tumor types. b, ECA profile in several tumor types by genes shared amongst them. Several epithelial tumor types share ECA genes however, many of them have a distinct ECA profile.

Figure 1

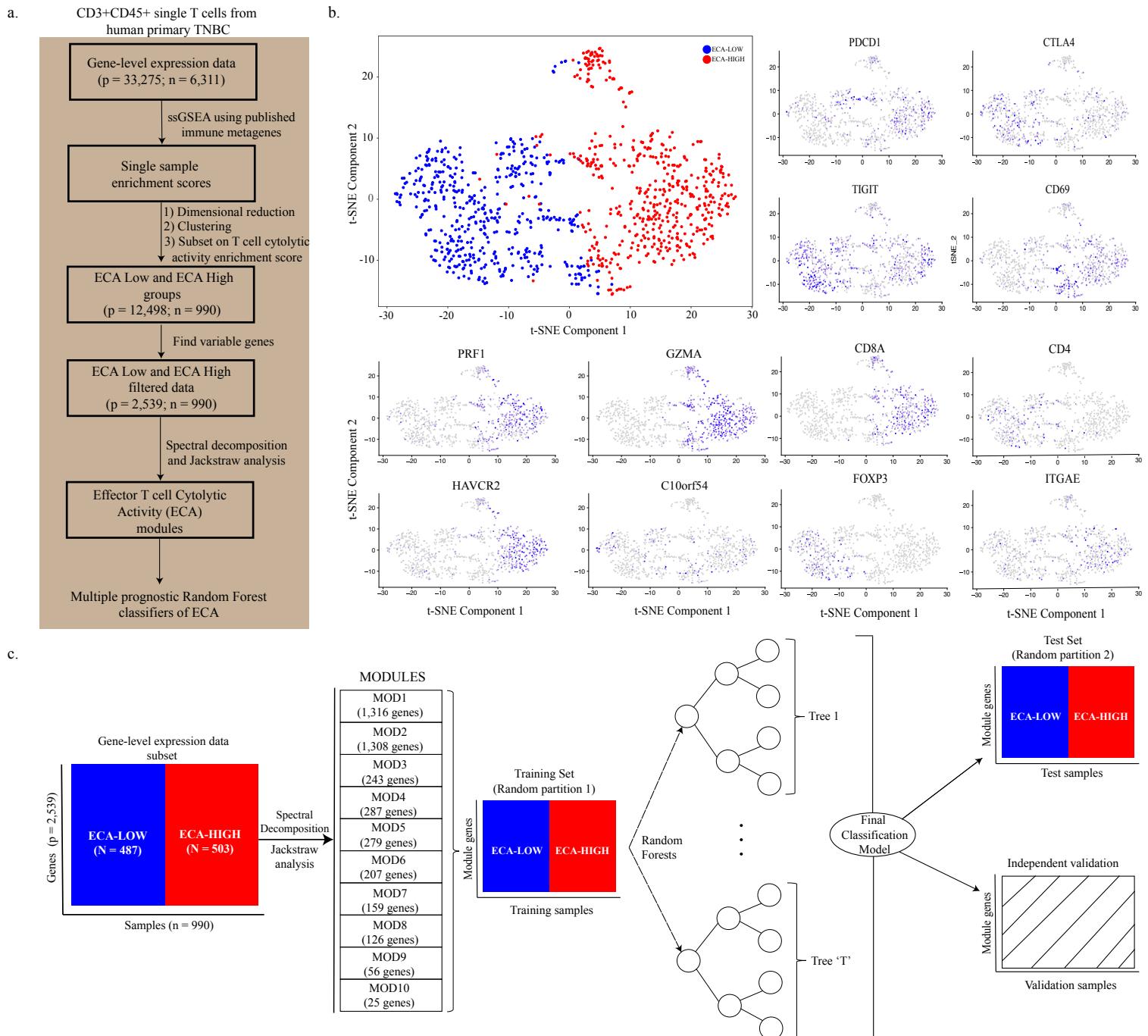
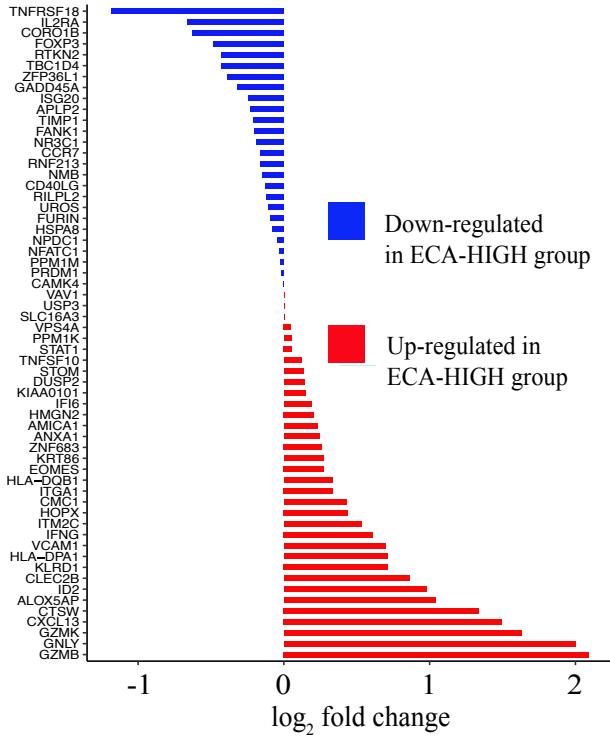


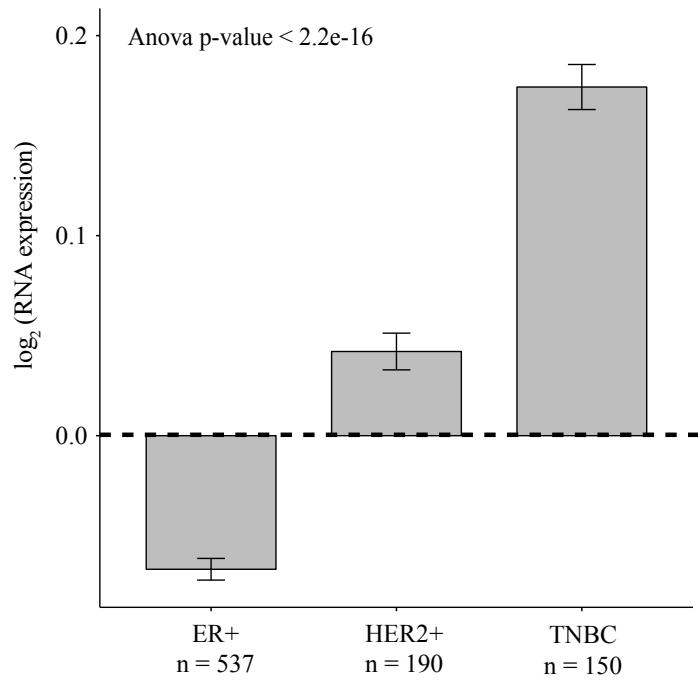
Figure 2

a. 60 genes from the Random Forest model from MOD5



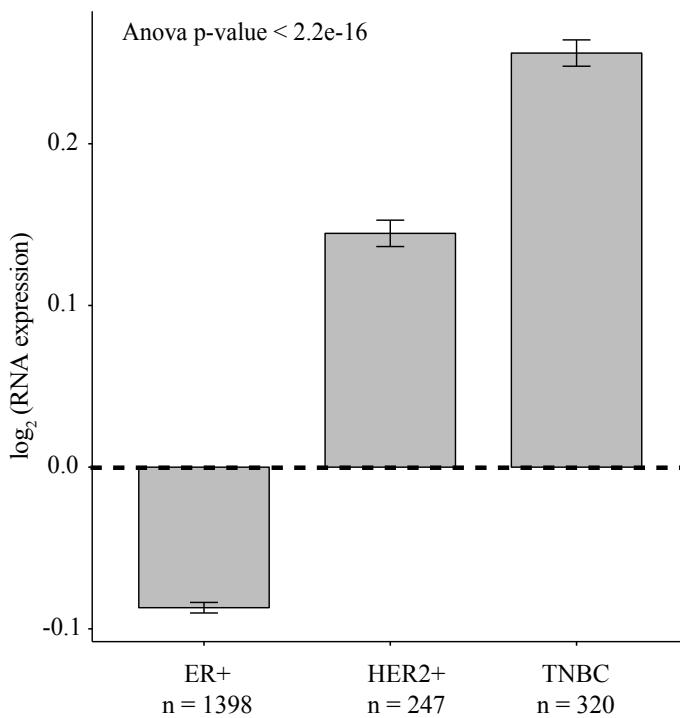
b.

TCGA breast dataset (N = 877)



c.

METABRIC dataset (N = 1965)



d.

GSE26304 (N = 115)

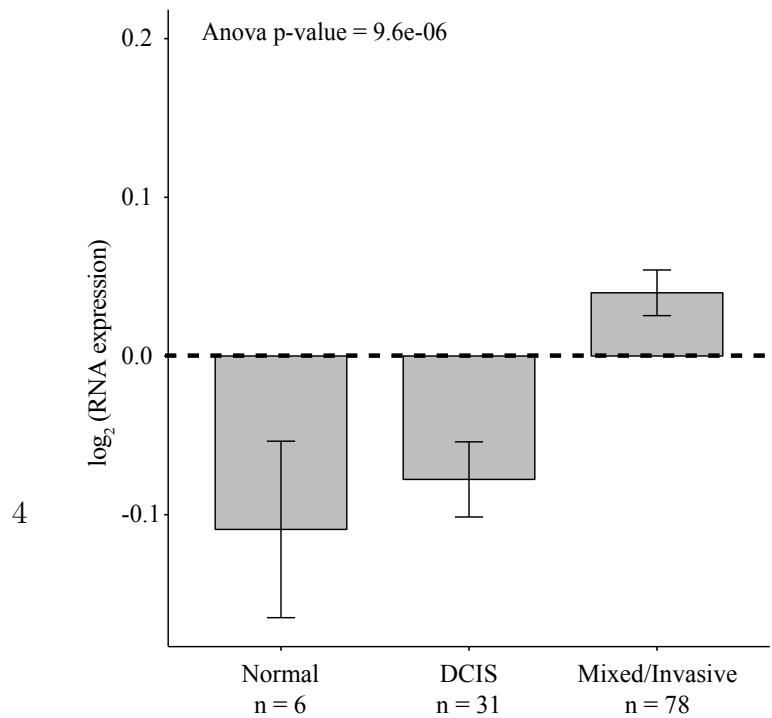


Figure 3

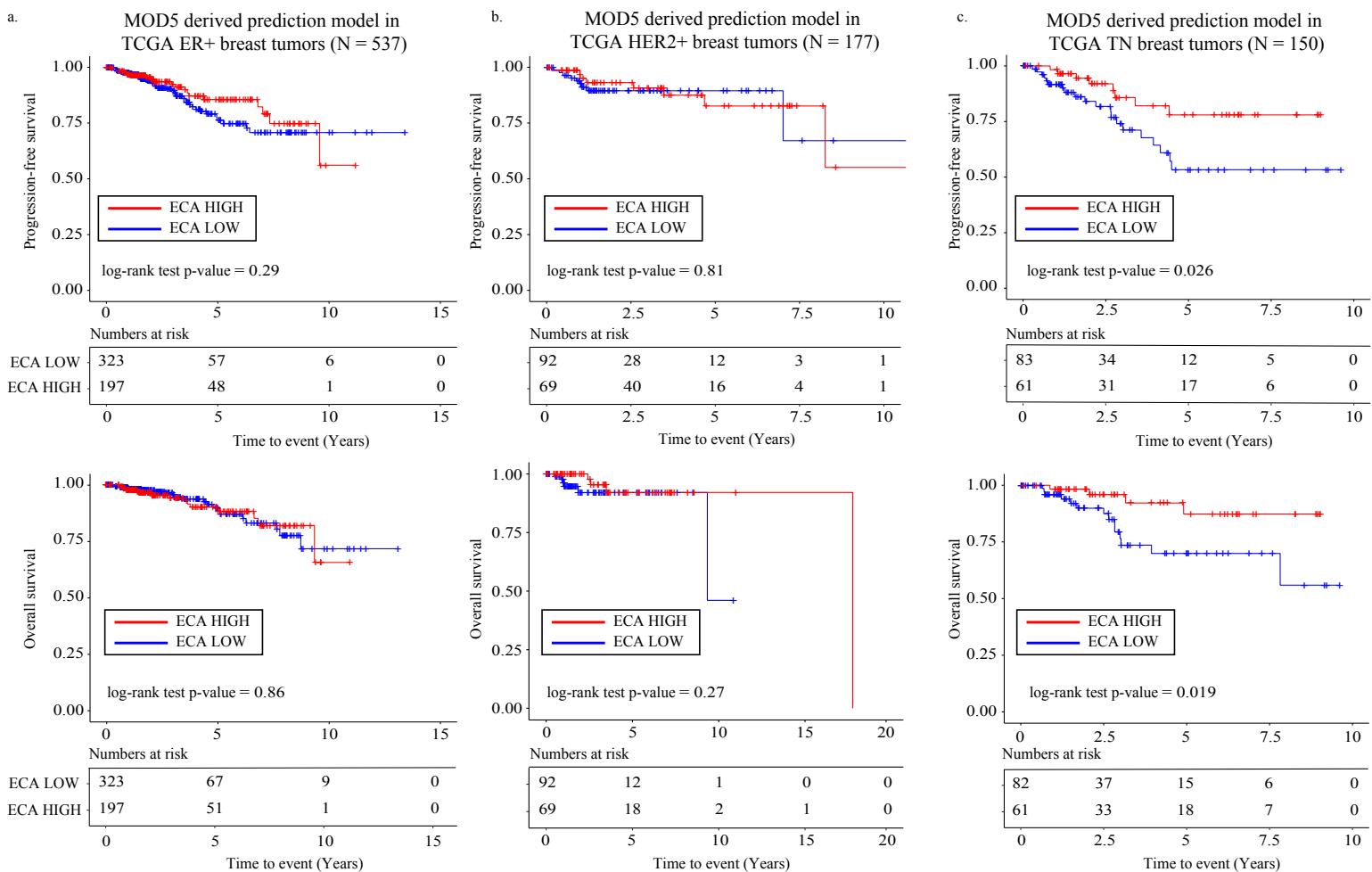


Figure 4

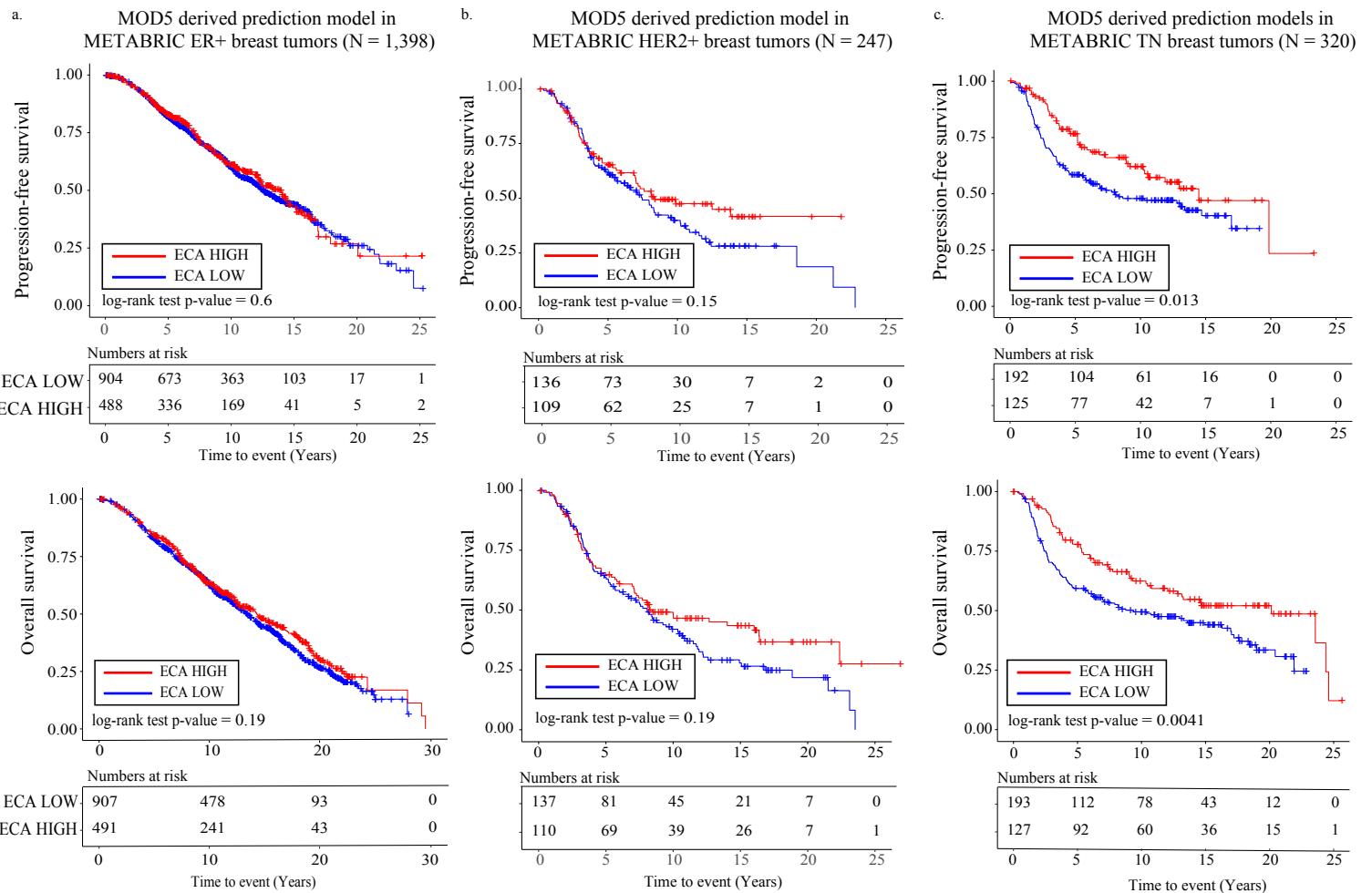


Figure 5

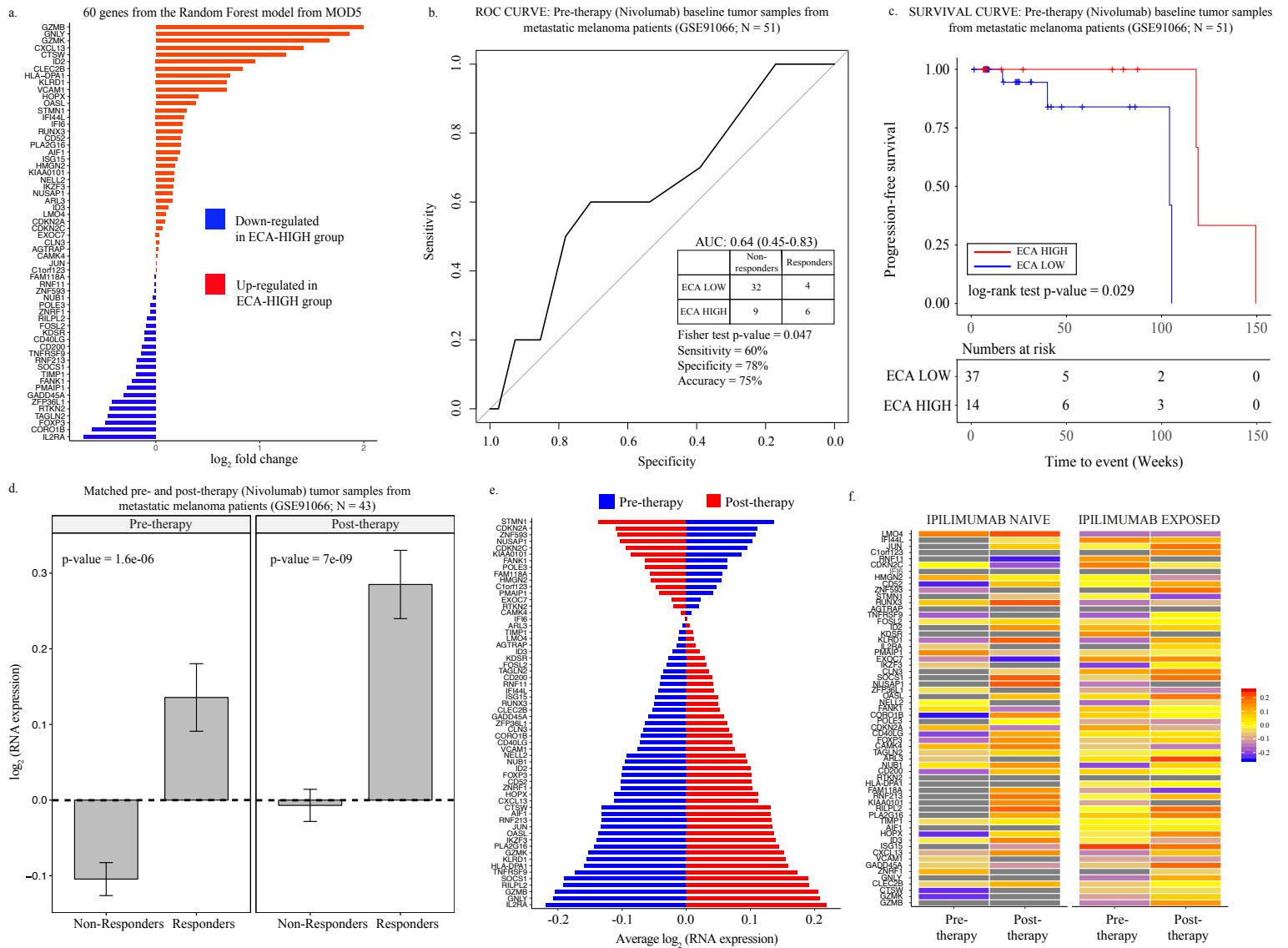


Figure 6

