Data S1. Detailed information about the operation of these machine learning programs

For the CoxBoost model, the 'optimCoxBoostPenalty' function was first called to determine the optimal penalty (shrinkage) value. This was then combined with 10-fold cross-validation on the CoxBoost model to search for the best number of boosting steps. Finally, the 'CoxBoost' function was used to fit the model. The stepwise Cox analysis was performed using the survival package, and the complexity of the statistical model was evaluated based on the Akaike information criterion (AIC). All possible combinations for the direction parameter in stepwise Cox analysis were calculated, including 'both', 'backward' and 'forward'. Lasso, Ridge and Enet models were implemented using the glmnet package and the 'cv.glmnet' function. The regularization parameter lambda was determined through 10-fold cross-validation, while the tradeoff parameter alpha was set between 0 and 1 (interval=0.1). When alpha is equal to 1, Lasso is executed, while Ridge is executed when alpha is equal to 0. For other values of alpha, Enet is executed. The survival-SVM model was implemented using the 'survivalsvm' function from the survivalsvm package, which employs support vector analysis on datasets with survival outcomes. The GBM model was implemented using the gbm package. The 'gbm' function was used with 10-fold cross-validation to fit a GBM. The SuperPC model was implemented using the superpc package, which is an extension of PCA. The 'superpc.cv' function was also used with 10-fold cross-validation. For the plsRcox model, the 'cv.plsRcox' function of the plsRcox package was used. For the RSF model, the randomForestSRC package was utilized, employing the 'rfsrc' function with two important parameters: 'ntree' and 'nodesize'. The parameter 'ntree' represents the number of trees in the random forest, while 'nodesize' represents the minimum size of the terminal nodes. In this study, 'ntree' was set to 1,000 and 'nodesize' to 5.

Figure S1. Potential biomarkers among CRGs in BRCA. (A) Differentially expressed genes between BRCA and normal tissues. (B) Univariate Cox regression analysis identified potential biomarkers in BRCA. (C) Overlap between differentially expressed genes and CRGs. The genes included in the signature are indicated with red boxes. FC, fold-change; CDS, programmed cell death-based gene signature; BRCA, breast cancer; DEGs, differentially expressed genes; CRGs, programmed cell death-related genes.

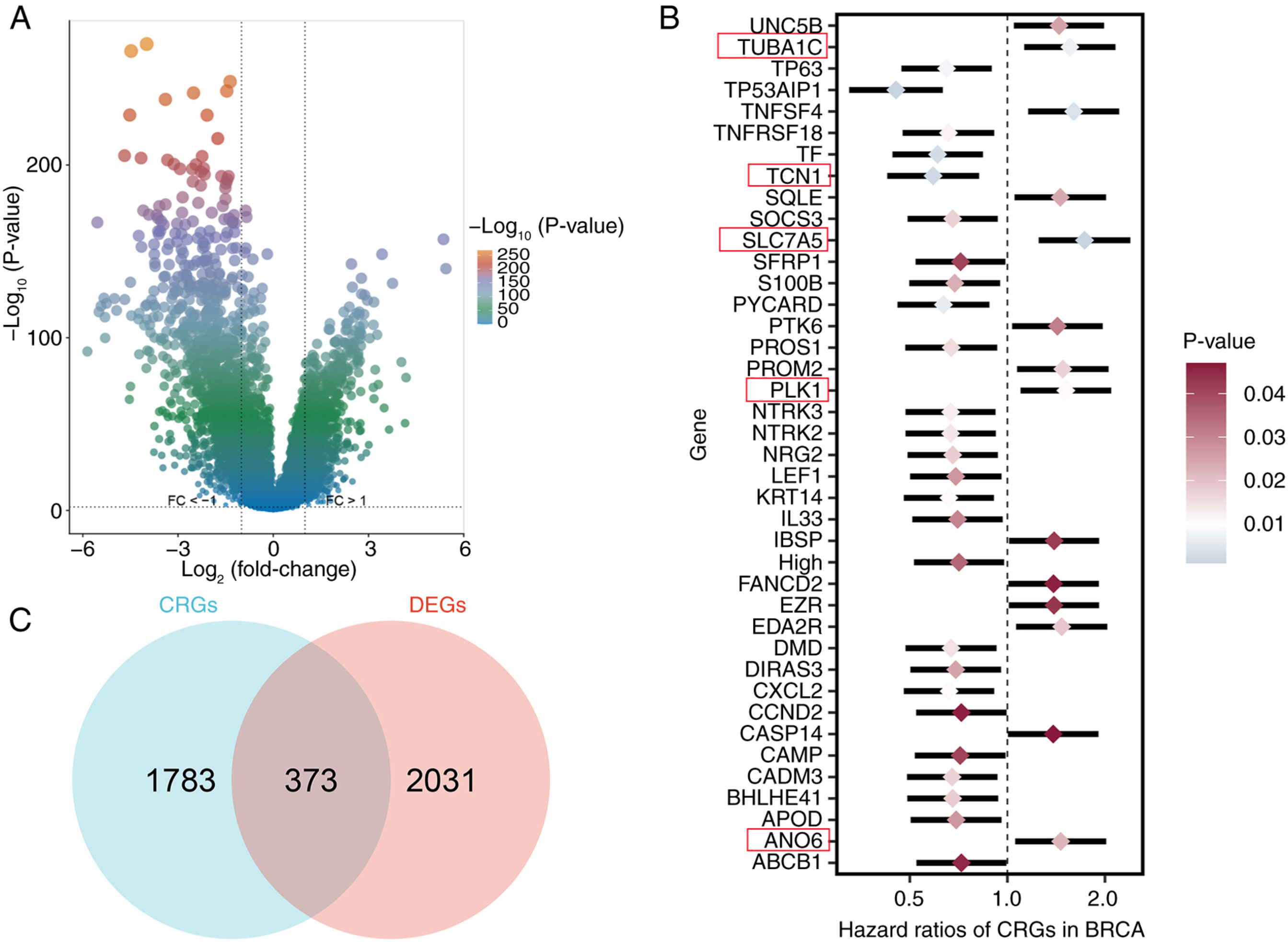

Figure S2. Correlation between the CDS score and abundance of immune cells in breast cancer based on the results of Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts method. CDS score demonstrates significant negative correlations with the abundance of (A) $CD4^+$ memory resting T cell, (B) $CD4^+$ memory-activated T cell, (C) memory B cells, (D) M1 macrophages, (E) monocyte, (F) resting myeloid dendritic cells, (G) activated mast cells and (H) resting mast cells. CDS, programmed cell death-based gene signature.

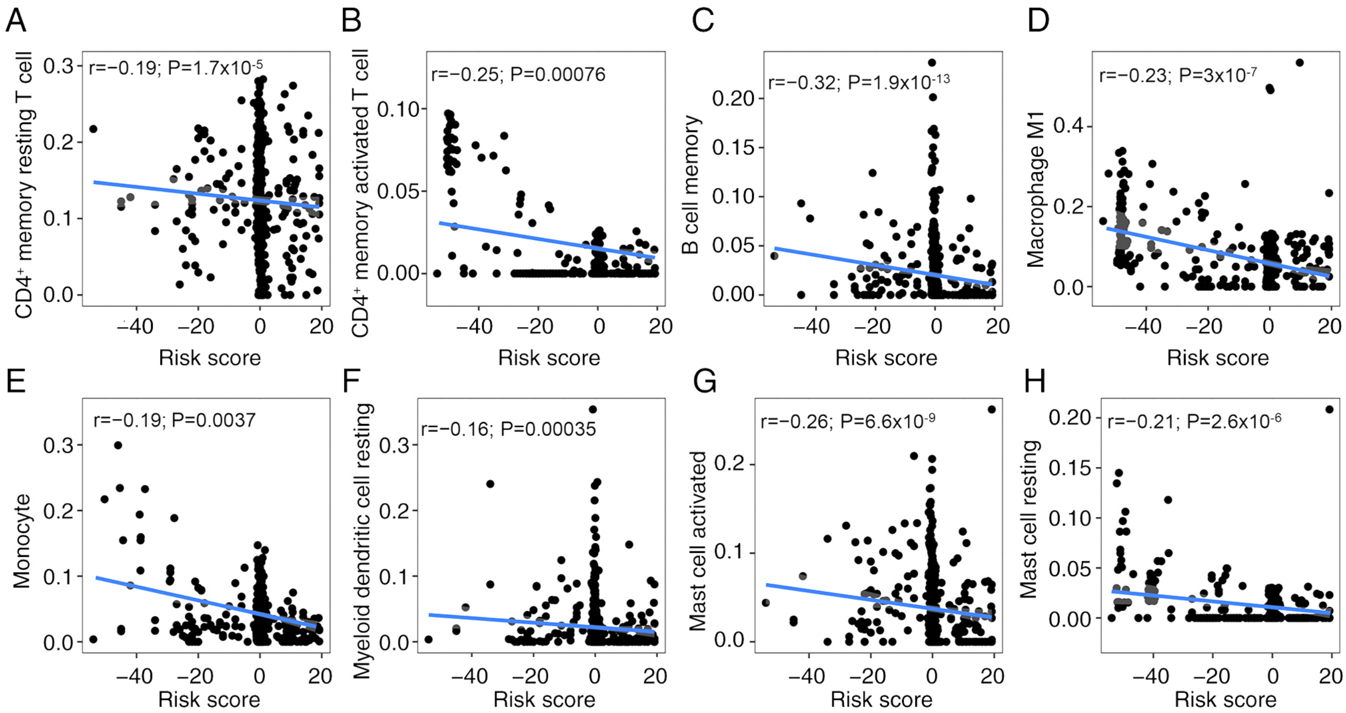

Figure S3. Difference of immune indicators in patients with breast cancer with different CDS scores. (A) Level of gene set score involved in antigen presentation, (B) inflammatory response and (C) T-cell exhaustion in different CDS score group. CDS, programmed cell death-based gene signature.