

Supplementary figures

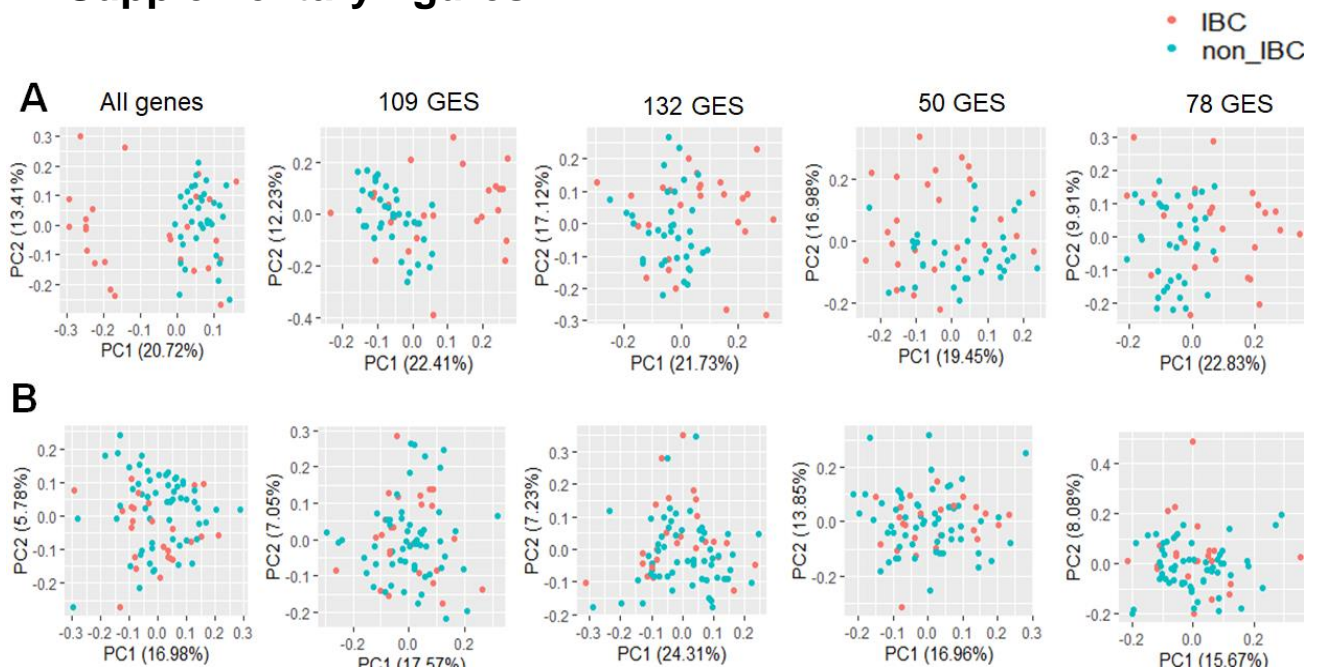


Fig S1: IBC gene signature expression using dimension reduction method. Linear dimension reduction using PCA. Here, x-axis corresponds to the first principle component and y-axis corresponds to second principle component. The variance explained by each of the component has been included in the plot. **(A) GSE22597 (B) GSE5847**

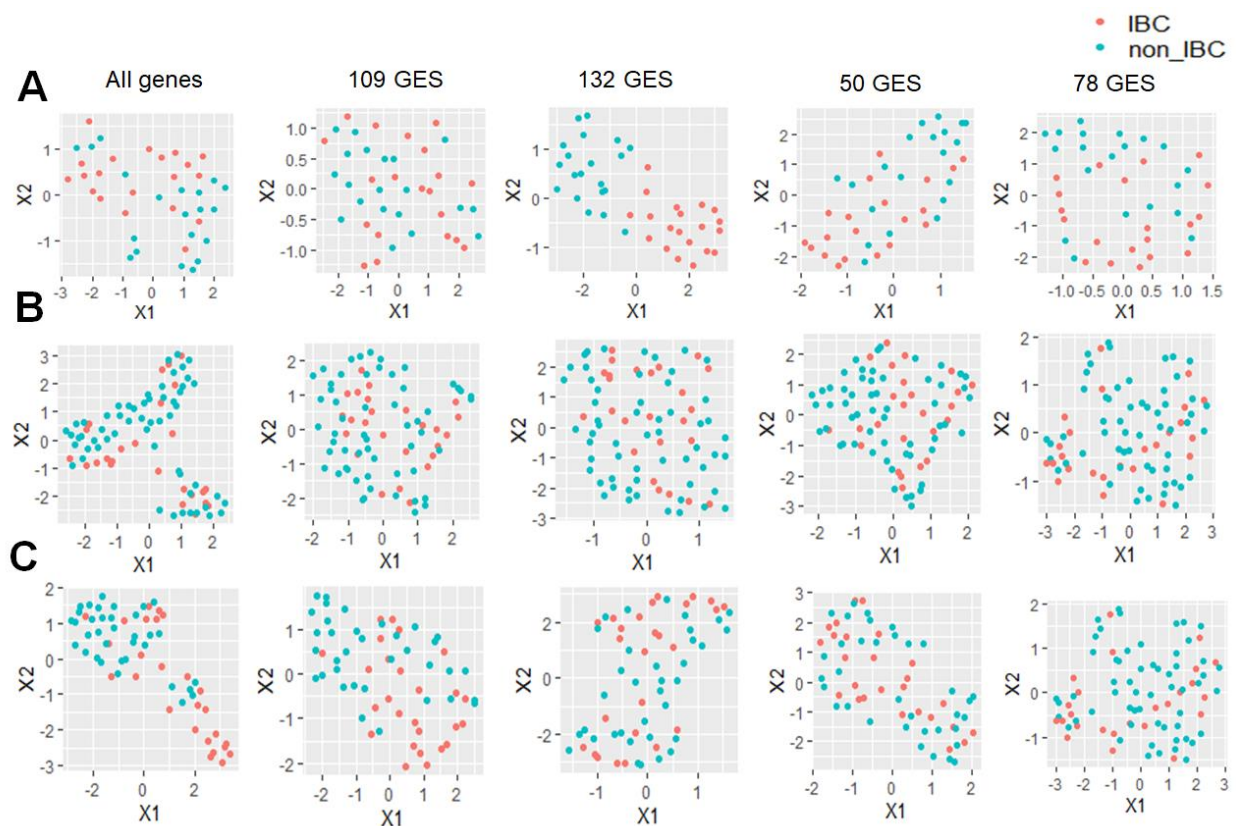


Fig S2: IBC gene signature expression using non-linear dimension reduction method uMAP. The data have been projected in two dimensions X1 and X2. **(A) GSE45581 (B) GSE22597 (C) GSE5847**

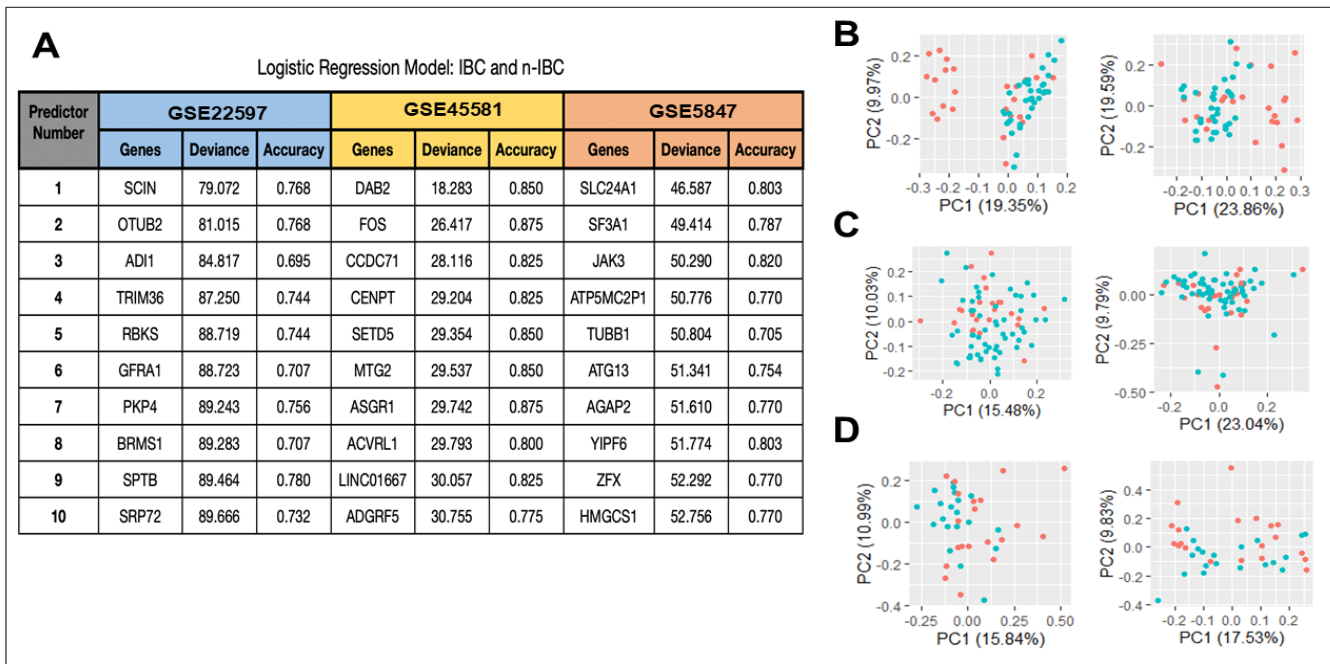


Figure S3: Features of the top LR predictors. (A) The top 10 predictors for the GSE 22597 (blue), GSE 45581 (yellow), and GSE 5847 (red) datasets were identified based on their minimal deviance values. Leave-one out predictive accuracy is reported for each top transcript. (B) GSE5847 PCA using top 100 LR predictors based on GSE22597 and GSE45581. (C) GSE22597 PCA using top 100 LR predictors based on GSE5847 and GSE45581. (D) GSE45581 PCA using top 100 LR predictors based on GSE22597 and GSE5847.

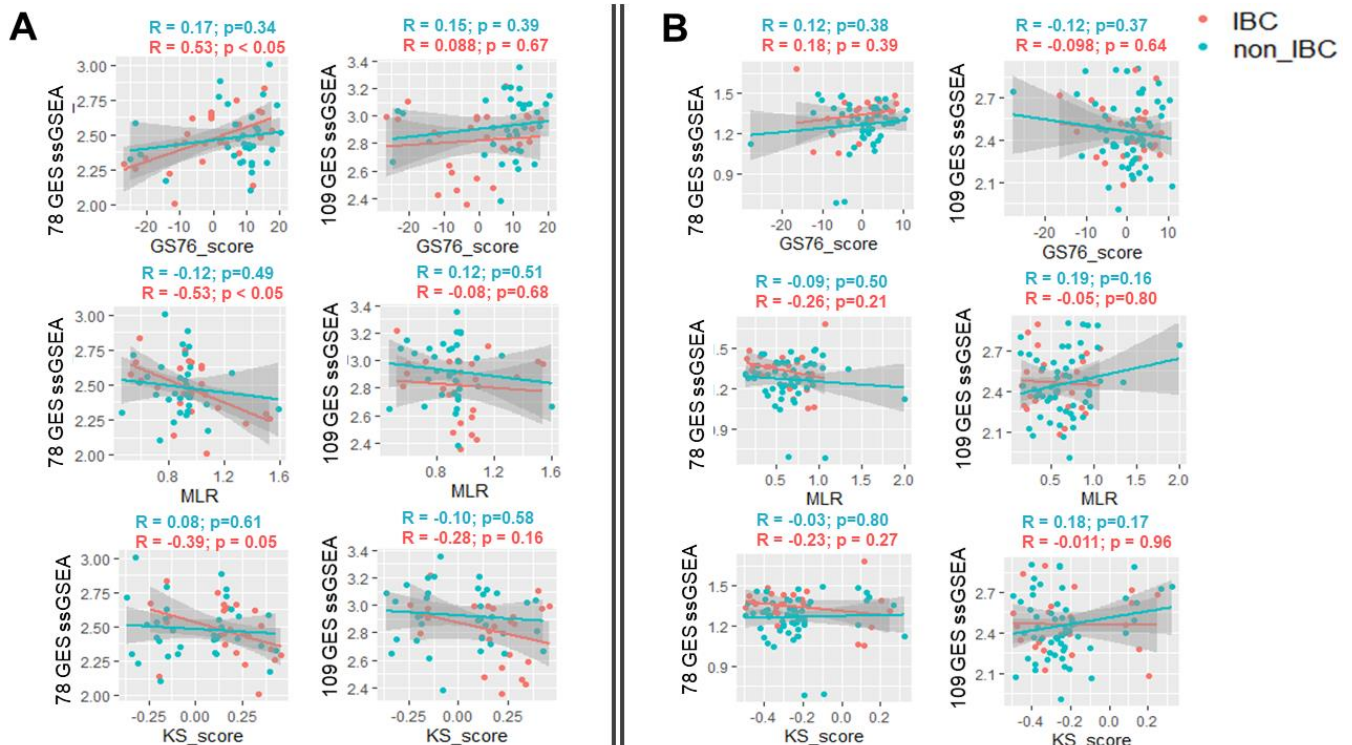


Fig S4: Correlation between ssGSEA score (78 GES and 109 GES) and EMT scoring methods. A) GSE5847 B) GSE22597. Pearson's correlation R and p-values high-lighted above each scatter plot (blue – nIBC, red – IBC).

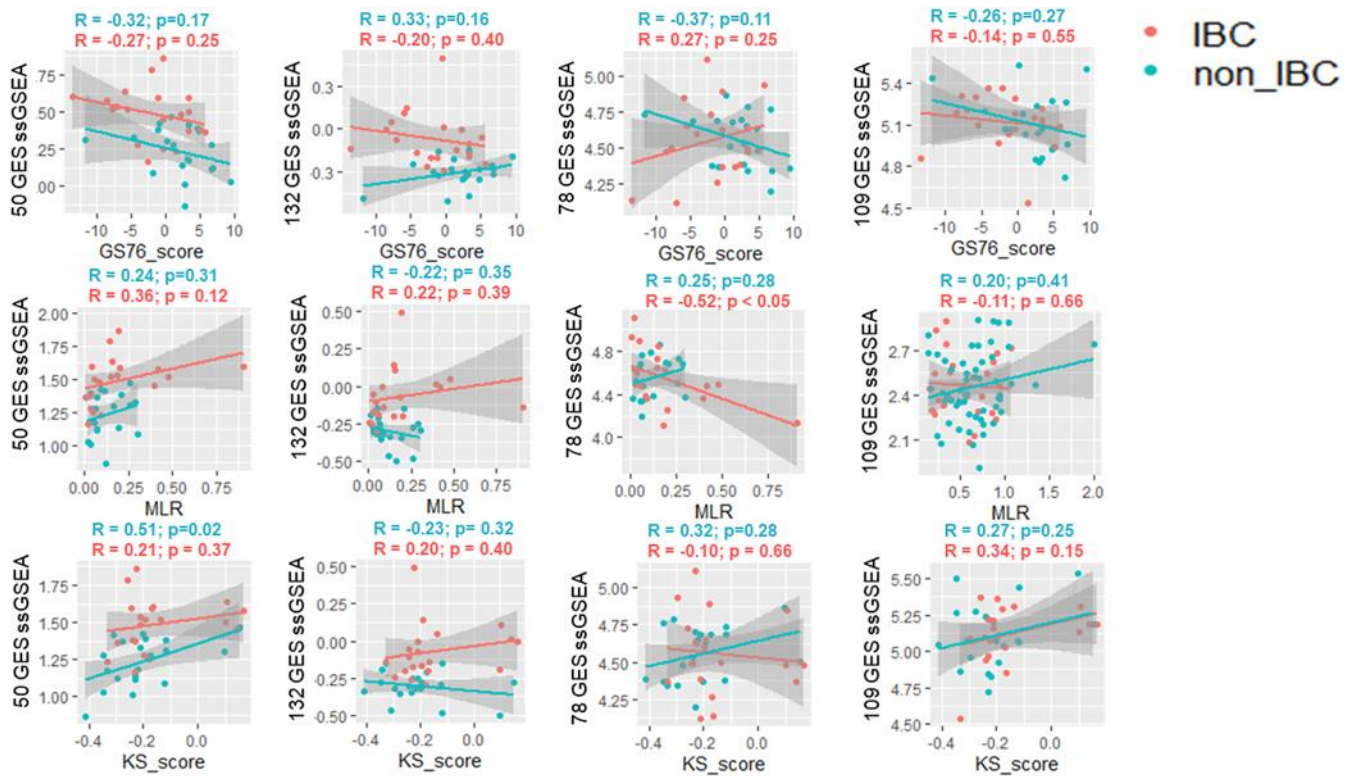


Fig S5: Correlation between ssGSEA score and EMT scoring methods in GSE45581. Pearson's correlation R and p-values high-lighted above each scatter plot (blue – nIBC , red – IBC).

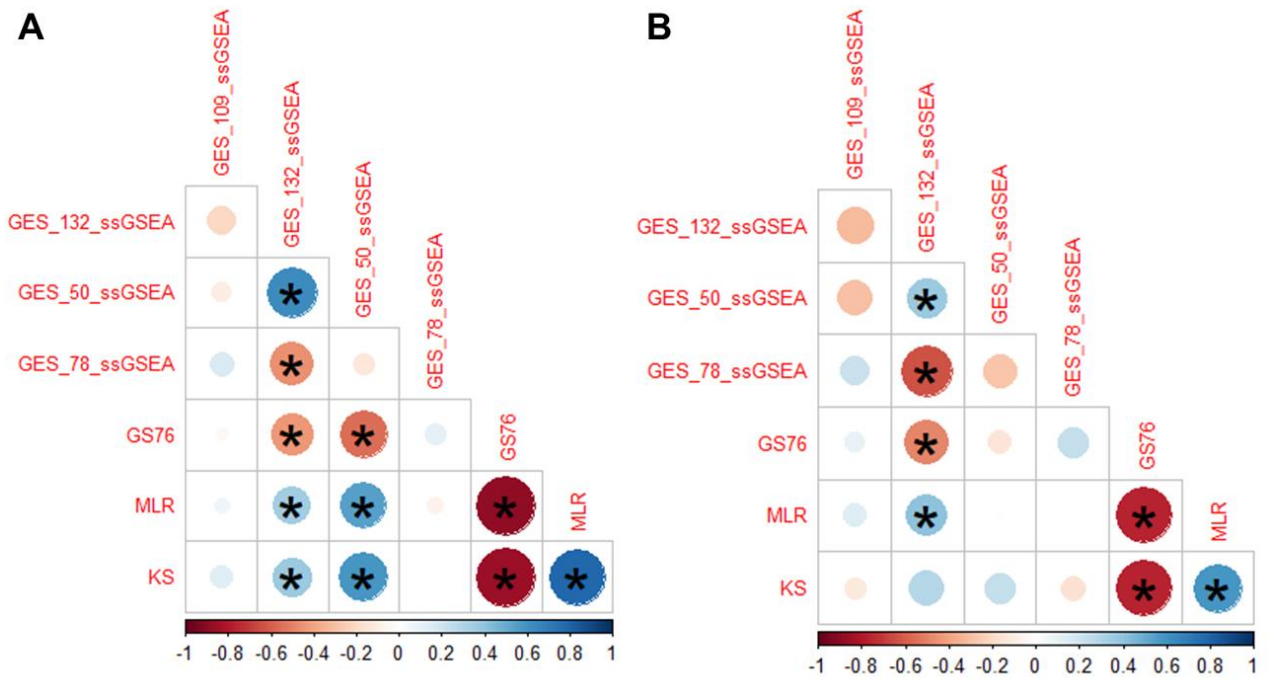


Fig S6: ssGSEA scores and EMT score spearman's correlation in GSE22597. (A) IBC (B) nIBC

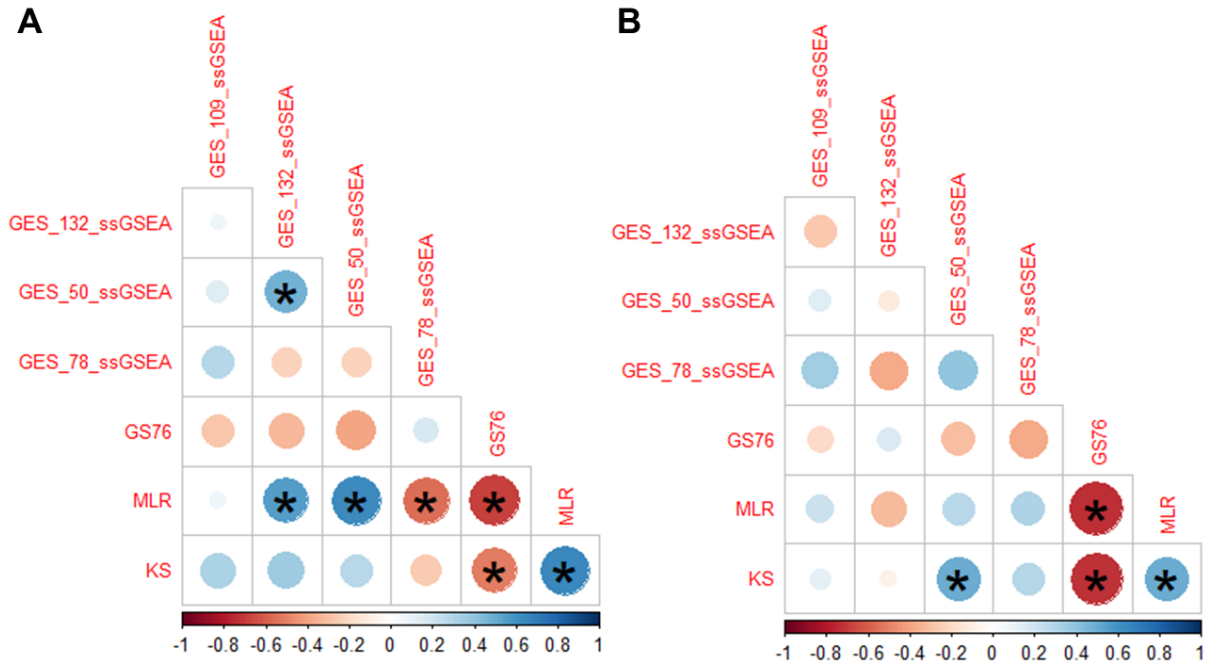


Fig S7: ssGSEA scores and EMT score spearman's correlation in GSE45581.(A) IBC (C) nIBC

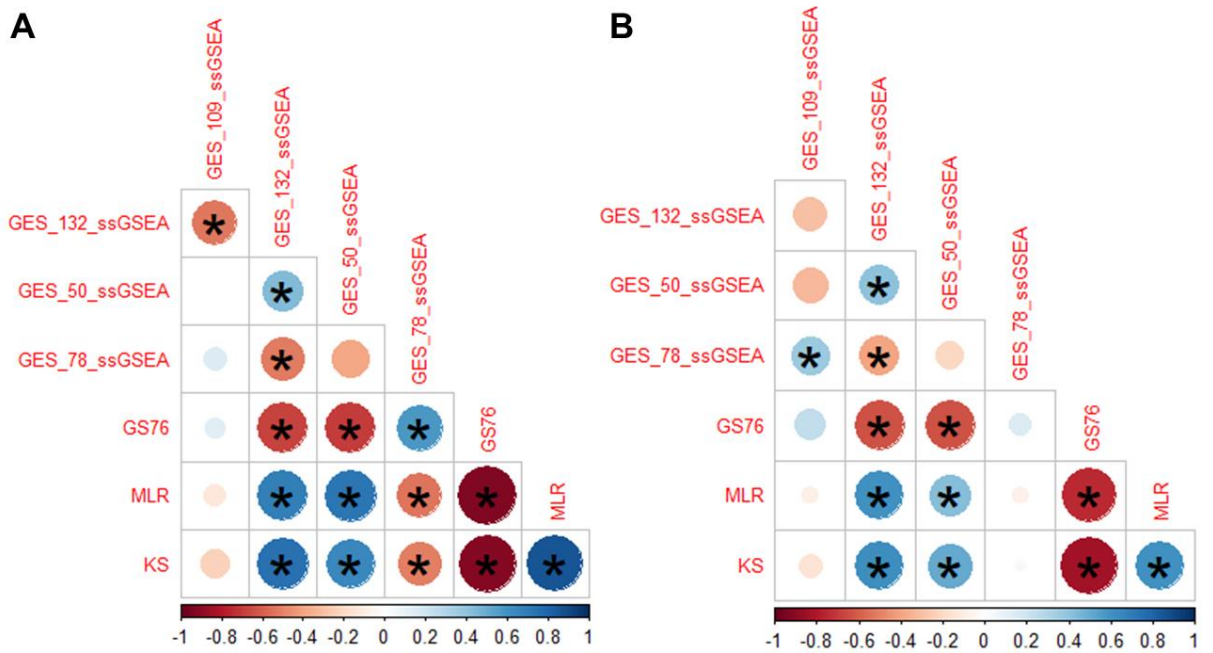


Fig S8: ssGSEA scores and EMT score spearman's correlation in GSE5847. (A) IBC (B) nIBC

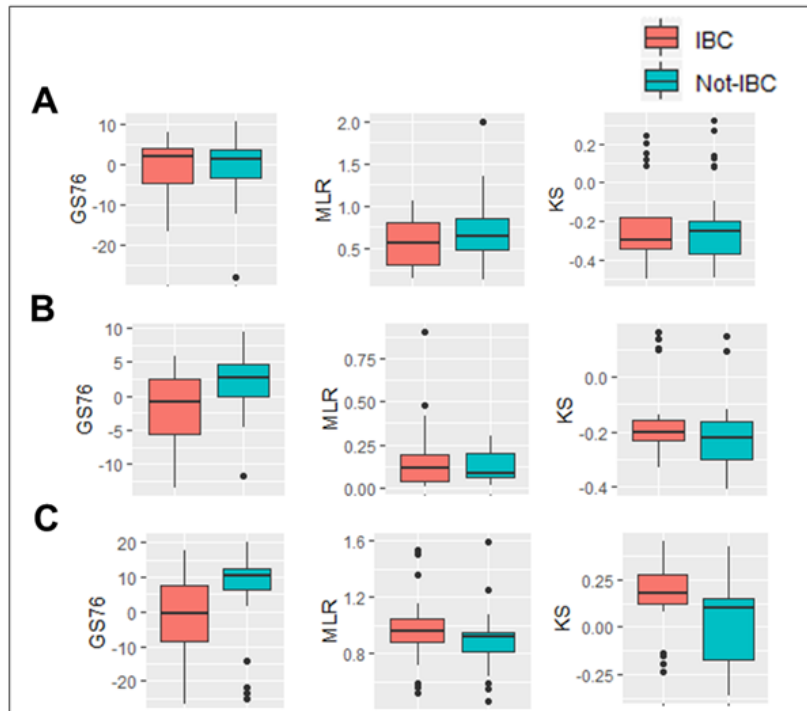


Fig S9: EMT scores across IBC and nIBC groups. (A) GSE22597 (B) GSE45581 (C) GSE5847

Supplementary table legends

S1: LR top 2000 predictors

S2: EMT scores

S3: Clustering accuracy of different combination of variables