Biomolecules and Biomedicine
ISSN: 2831-0896 (Print) | ISSN: 2831-090X
Journal Impact Factor® (2023): 3.1
CiteScore®(2023):7.4
www.biomolbiomed.com | www.blog.bjbms.org

Machine learning integration of single-cell and bulk transcriptomics identifies fibroblast-driven prognostic markers in colorectal cancer

Ning Zhang[1,2], Ruiyan Liu[1,2], Siya Wu[1,2], Chenxi Feng[1], Boxiang Wang[1,2], Qiaoqiao Zheng[1,2], Linru Jie[1,2], Ruihua Kang[1], Xiaoli Guo[1], Xiaoyang Wang[1*], Shaokai Zhang[1*], and Jiangong Zhang[1*]

[1]Department of Cancer Epidemiology, The Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital, Zhengzhou, China

[2]College of Public Health, Zhengzhou University, Zhengzhou, China

[*]Correspondence to **Xiaoyang Wang**: xywang233@126.com; **Shaokai Zhang:** shaokaizhang@126.com and **Jiangong Zhang:** zhangjg@zzu.edu.cn

DOI: https://doi.org/10.17305/bb.2025.12038

Guest Editor: Ziheng Wang

Full article is available at the following link:
https://www.bjbms.org/ojs/index.php/bjbms/article/view/12038/3845

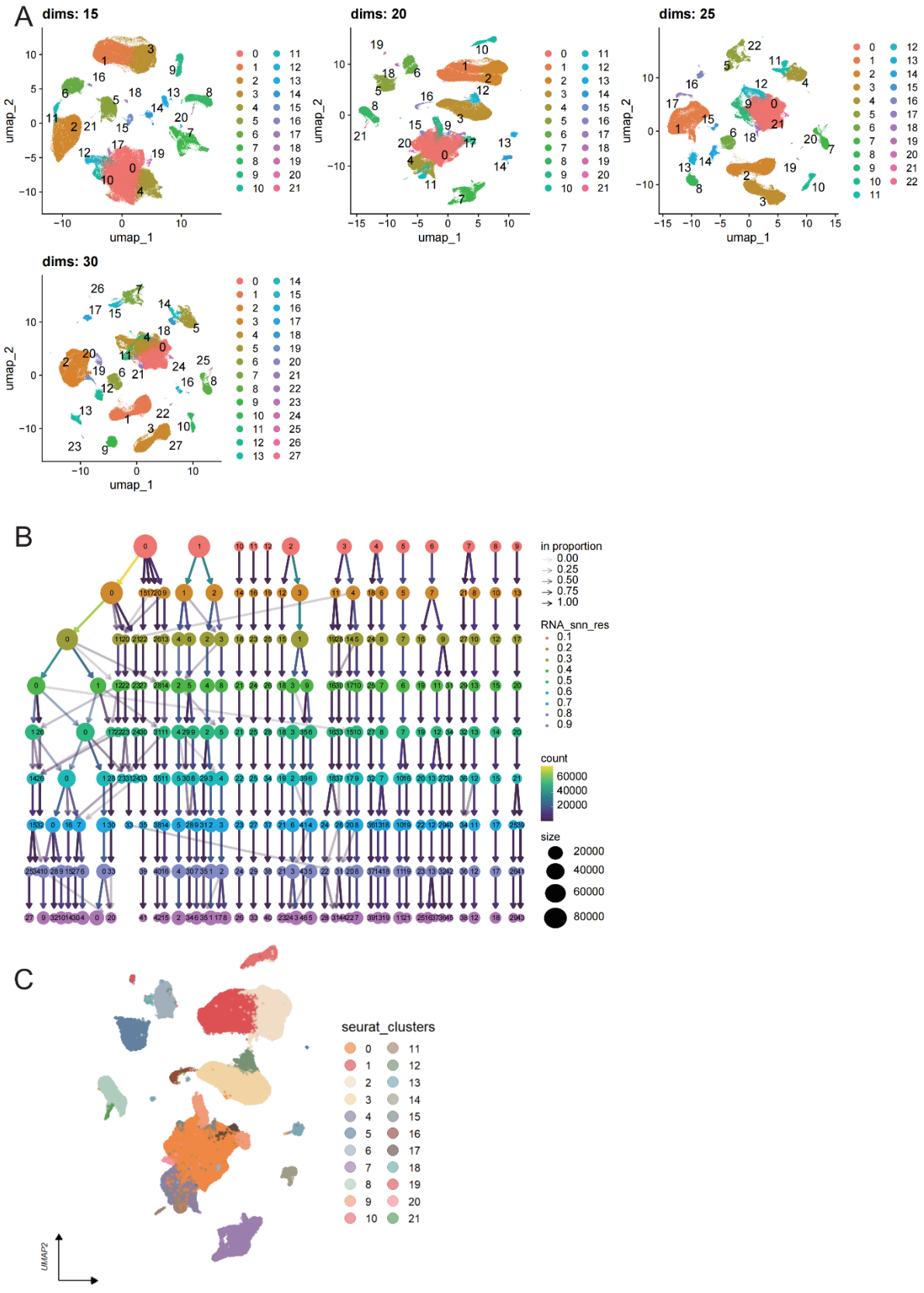

Figure S1. Single-cell sequencing data clustering process in 204 tumor center tissues.

(**A**) Cell distribution under different dimensions. Dims, which stands for dimensions, represents principal components. (**B**) The clustering tree generated using various resolution parameters for different cells has an optimal resolution of 0.2. In proportion indicates the proportion of cells within each cluster, RNA_snn_res refers to the RNA shared nearest neighbor resolution parameter used in Seurat clustering. (**C**) UMAP clustering plot at optimal resolution.

Figure S2. Supplementary information on the intercellular communication network in the CRC TME in tumor center tissues.

(**A**) Dot plot depicting cell-type specific communication patterns. Dot size indicates the count of interaction links, X and Y axes show outgoing and incoming communication probabilities, respectively. (**B**) Intuitive reflection of CellChat networks.

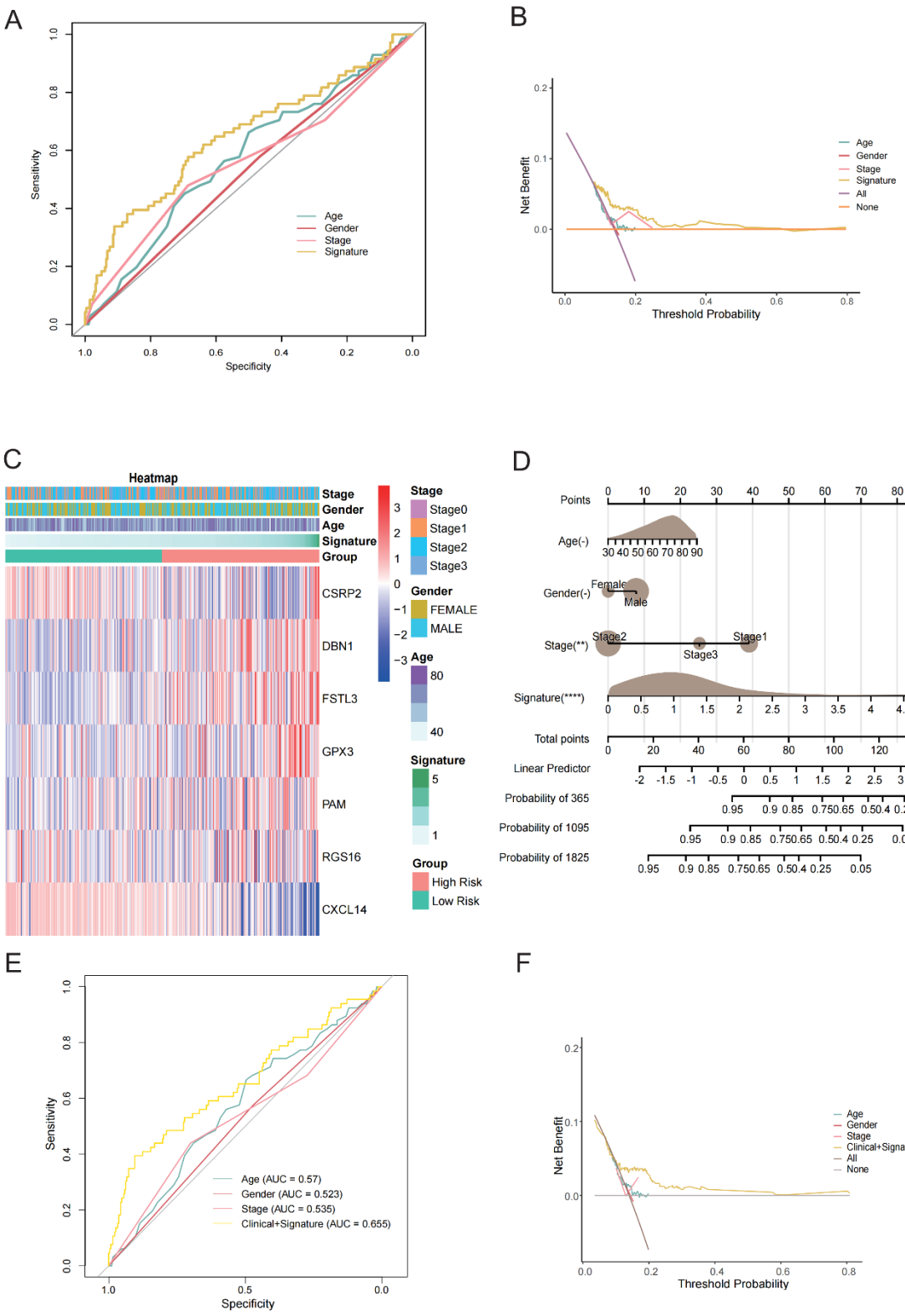

Figure S3. Validation of FRS in TCGA cohort.

(**A**) Univariate ROC curves and (**B**) DCA curves comparing FRS with clinical parameters. (**C**) Distribution of clinical characteristics and model gene expression across FRS risk groups. (**D**) Nomogram integrating FRS with clinical parameters. (**E, F**) Clinical utility assessment of the nomogram compared to individual clinical characteristics.

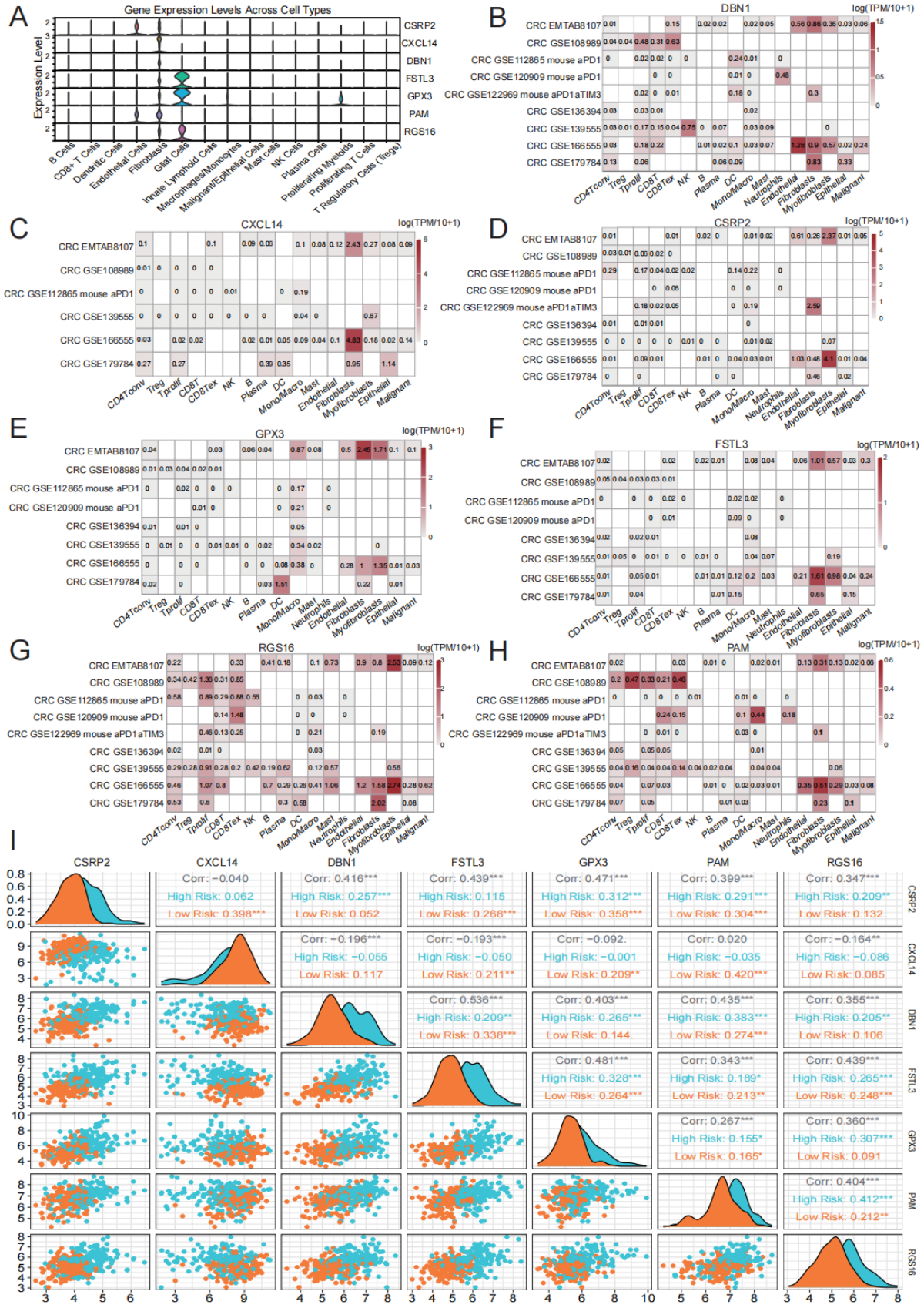

Figure S4. Seven FRS genes expression distribution in this study and other 9 CRC scRNA-seq datasets.

(**A**) The violin plot illustrates the expression distribution of 7 FRS genes in the single-cell transcriptomes of this study. (**B-H**) Expression of 7 FRS genes across various cell types from other 9 CRC scRNA-seq datasets from TISCH. (**I**) Correlation analysis of 7 genes in TCGA bulk RNA transcriptome (Corr refers to the correlation coefficient of expression levels between different genes, and it also indicates the correlation levels in different FRS subgroups, *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$).

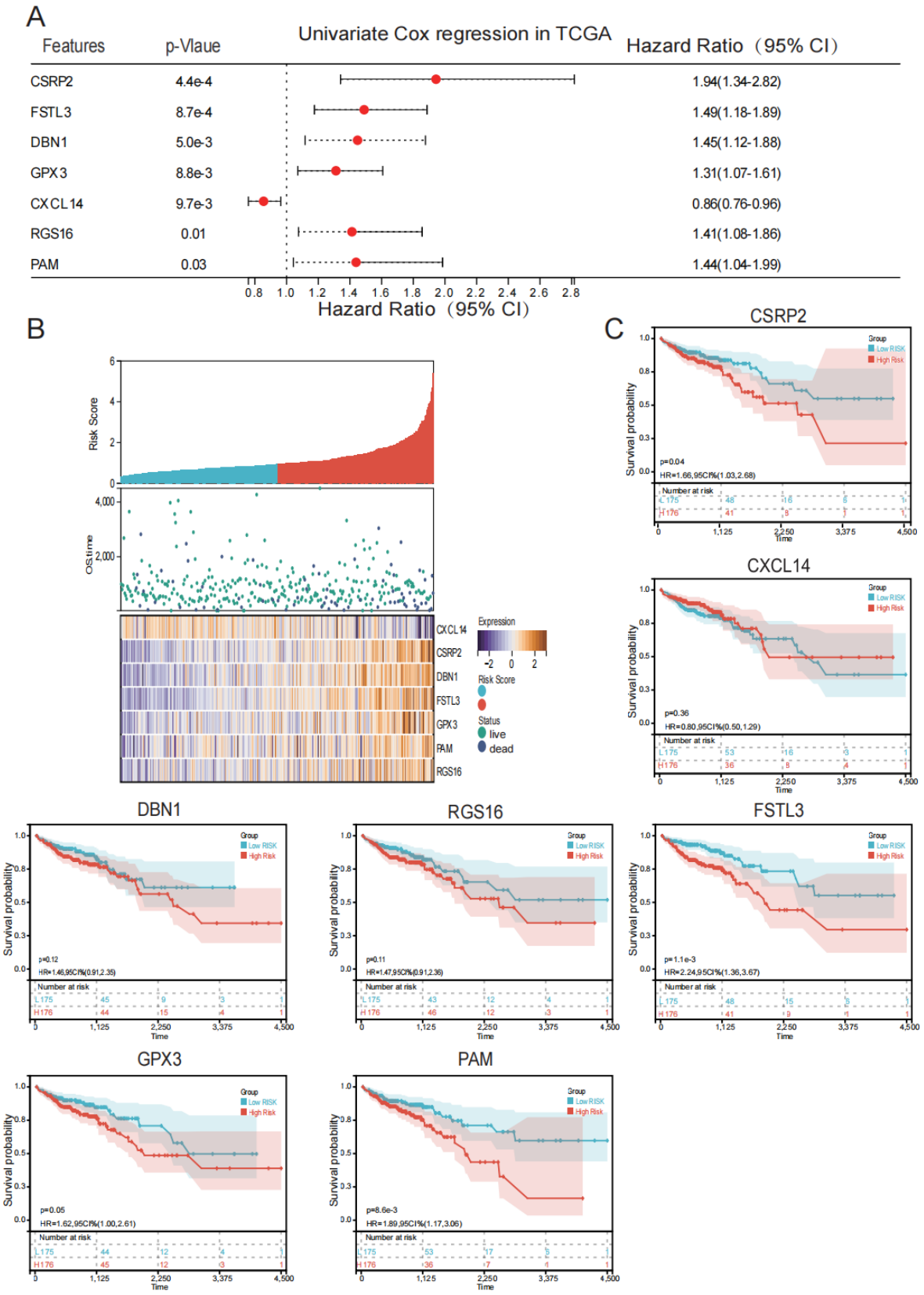

Figure S5. Single gene prognostic analysis of FRS in TCGA dataset.

(**A**)The forest plot shows the univariate cox regression results of 7 genes of FRS. (**B**) The heatmap based on risk scores displays the variation levels of gene expression in relation to

FRS risk scores and survival status. Risk Score refers to the FRS risk score. (**C**) KM curves of 7 genes of FRS.

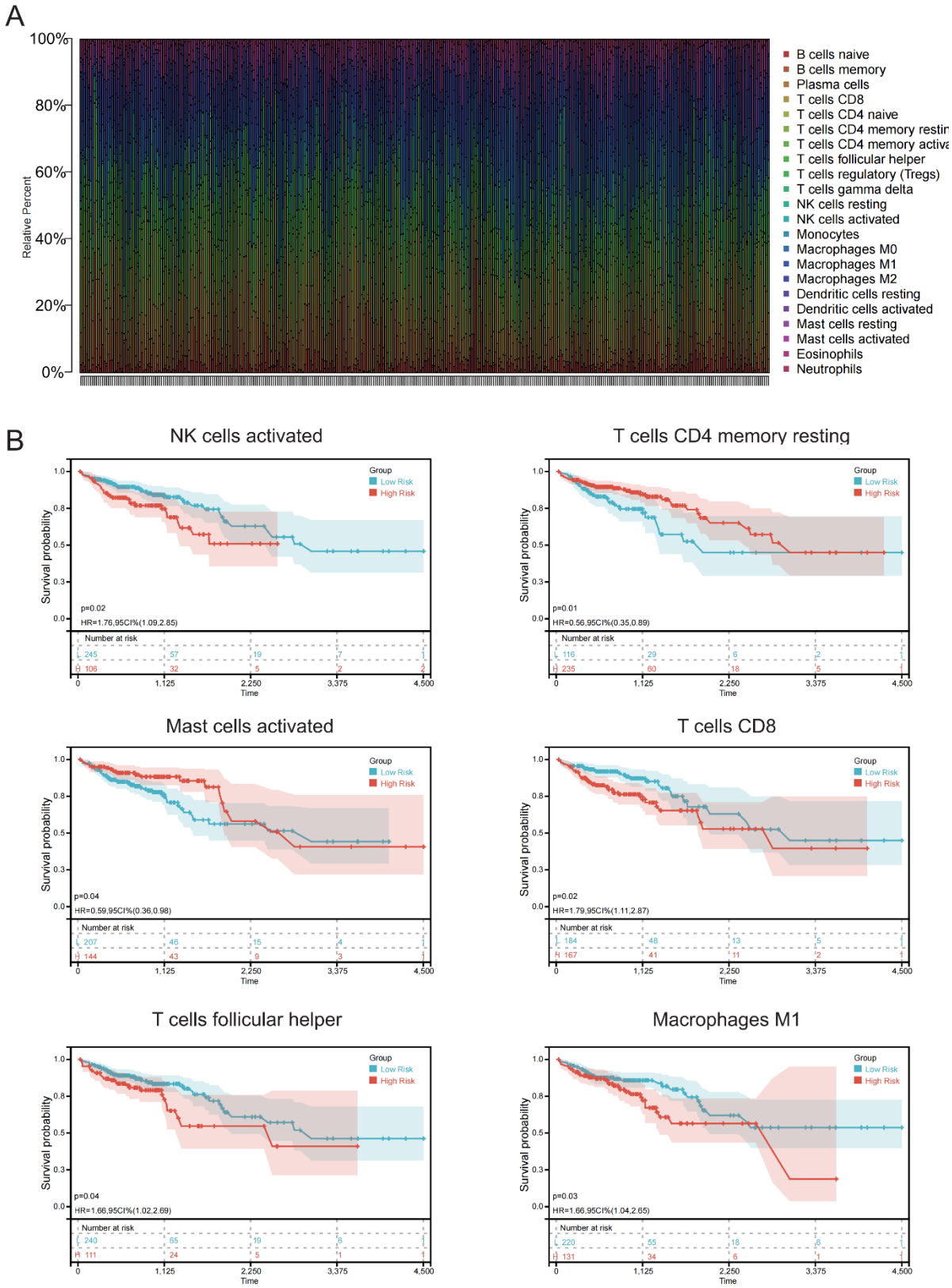

Figure S6. Supplementary explanation of immune infiltration characteristics based on TCGA bulk RNA transcriptomes.

(**A**) The relative percentage distribution of various immune cell types in each sample. (**B**) Survival analysis of 6 immune cell types with prognostic effects.