

Figure S3. Kaplan-Meier survival curves for the Stage_MSI and Stage_CMS systems. (a). Kaplan-Meier survival curve comparing OS of the four groups (Early_MSI-H, Early_non-MSI-H, Advanced_MSI-H, and Advanced_non-MSI-H) in the TCGA cohort. The indicated p-value was calculated with the log-rank test. (b). Summary table of the pairwise comparisons of survival curves between the four groups. Survival difference was tested using the log-rank test, with FDR adjusted p-values less than 0.05 were considered statistically significant and displayed in red color. Kaplan-Meier survival curves comparing DFS of the four CMS subtypes (CMS1-4) in the TCGA_early (c), TCGA_advanced (d), GEO_early (e), and GEO_advanced (f) cases. The indicated p-values were calculated with the log-rank tests.

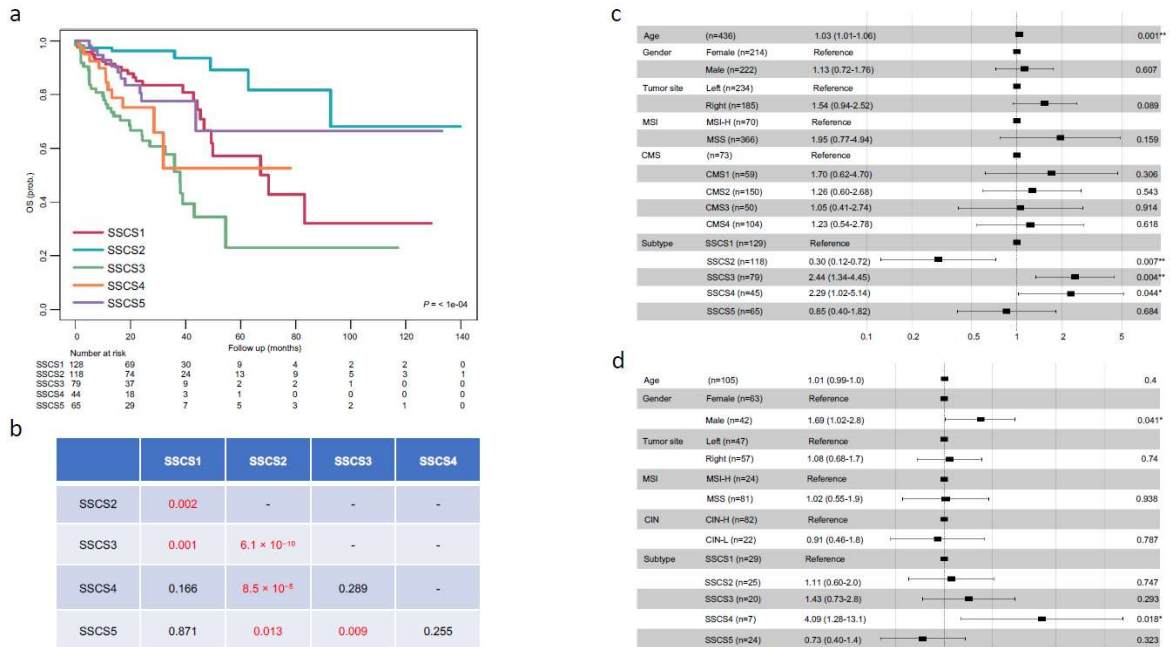


Figure S4. Kaplan-Meier survival curves and multivariate COX PH forest plots for OS. (a). Kaplan-Meier survival curve comparing OS of the five SSCS subtypes in the TCGA cohort. The indicated p-value was calculated with the log-rank test. (b). Summary table of the pairwise comparisons of survival curves between the five SSCS subtypes in the TCGA cohort. Survival differences were tested using the log-rank tests, with FDR adjusted p-values less than 0.05 were considered statistically

significant and displayed in red colors. Forest plots of Cox PH regression models illustrated the HRs, 95% CIs and log-rank p-values for different subtyping systems (SSCS, MSI, CMS, etc.) and confounder factors (age, gender, etc.) for OS in the TCGA (c) and CPTAC (d) cohorts. Log-rank p-value significance levels were given by stars: * < 0.05, and ** < 0.01.

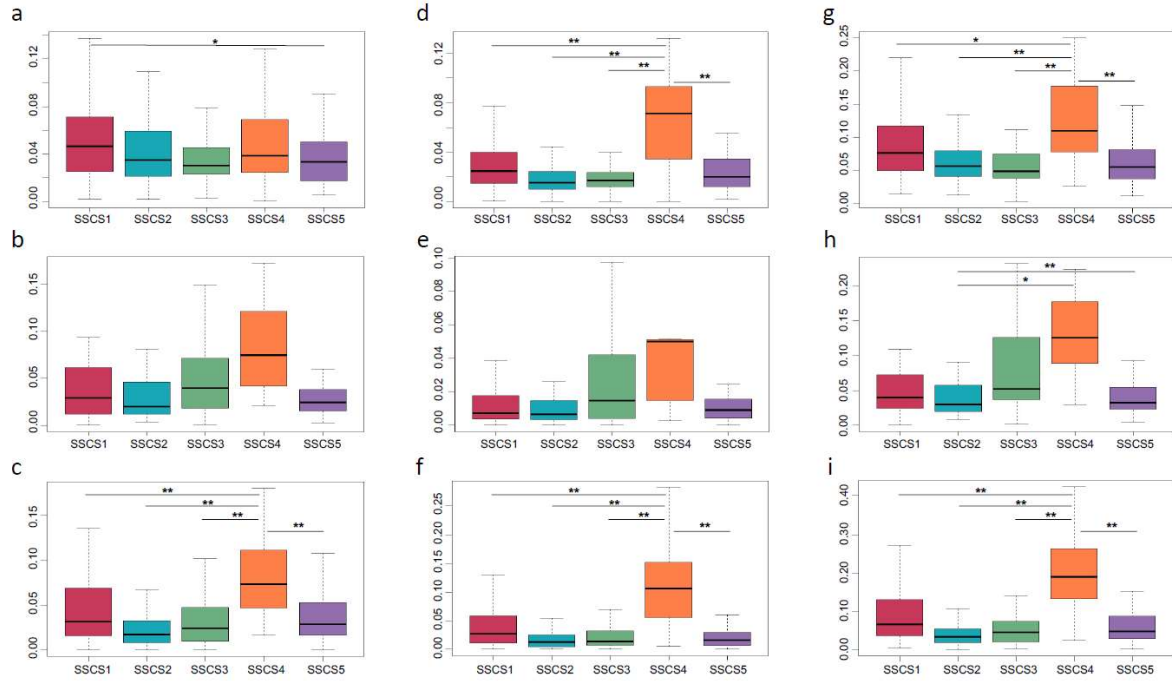


Figure S5. Boxplots of the immune, stroma, and microenvironment scores across the five SSCS subtypes. Boxplots of the immune (a-c), stroma (d-f), and microenvironment (g-i) scores across the five SSCS subtypes. Patient cohorts include TCGA (a, d, g), CPTAC (b, e, h), and GEO (c, f, i). Pairwise Wilcoxon rank sum tests with FDR adjusted p-value < 0.05 were considered statistically significant, and marked with a star. Two stars represent FDR adjusted p-value < 0.01.