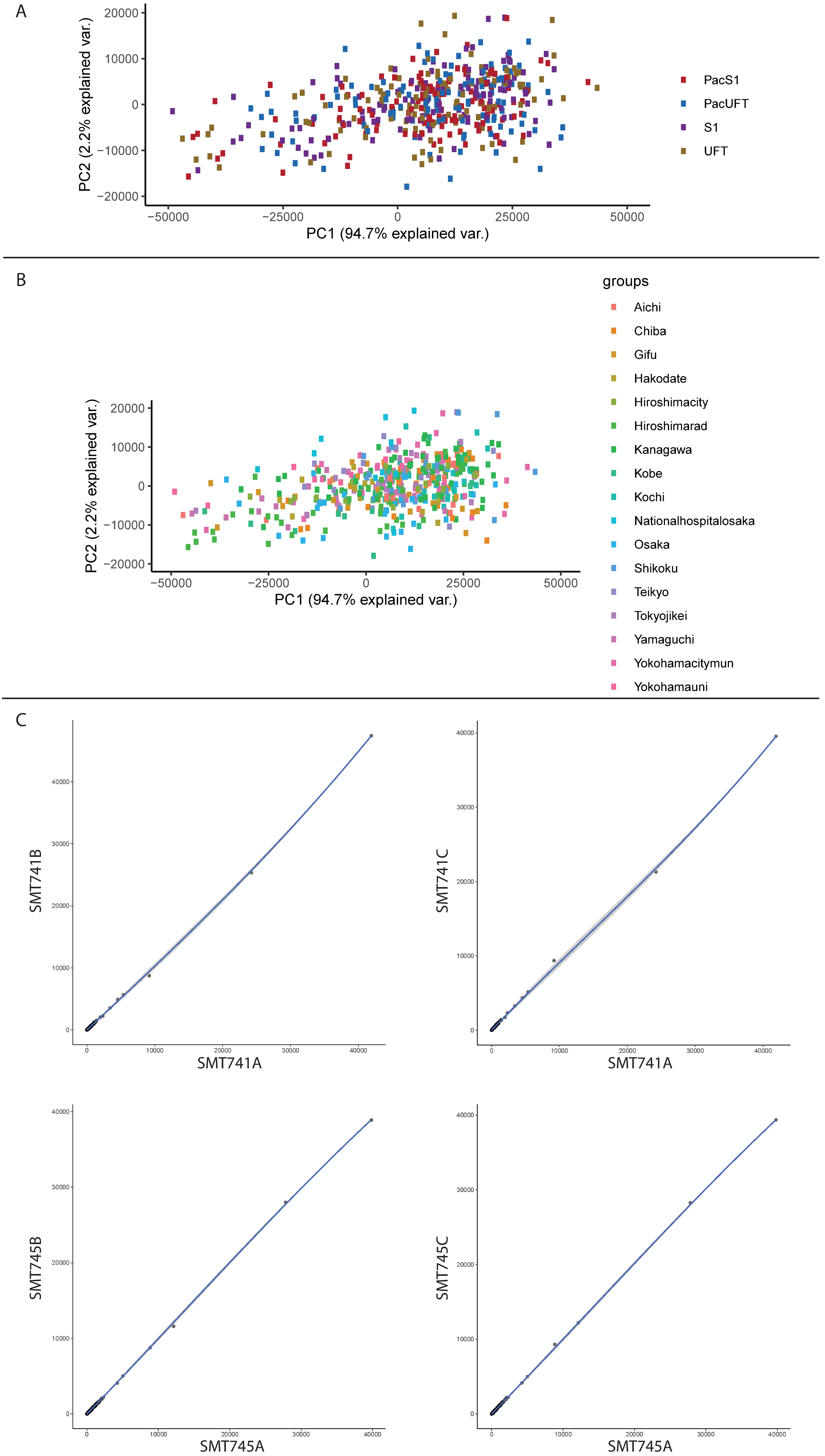
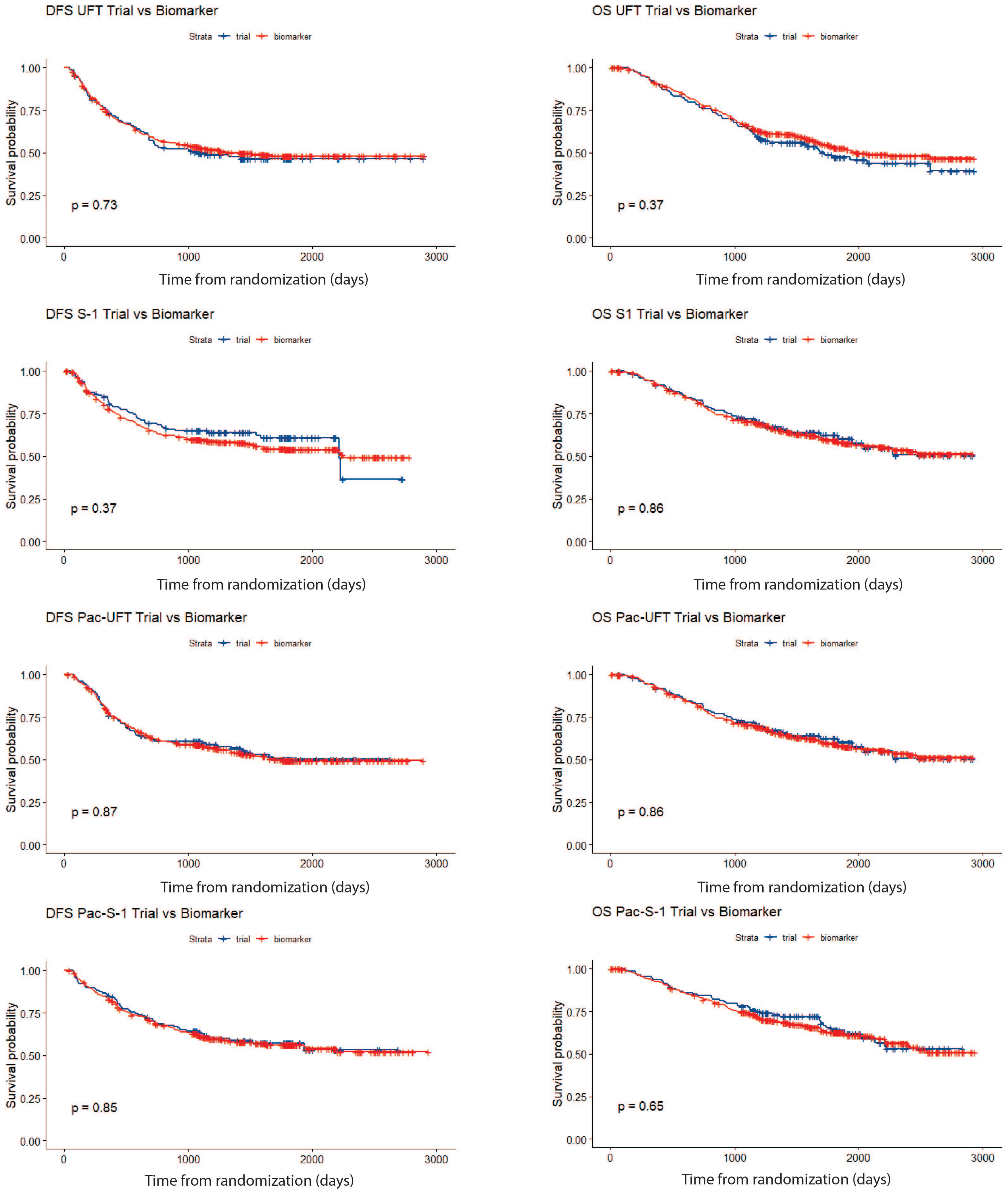
Supplementary Figure 1



Supplementary Figure 2



DFS UFT Trial vs Biomarker

OS UFT Trial vs Biomarker

DFS S-1 Trial vs Biomarker

OS S1 Trial vs Biomarker

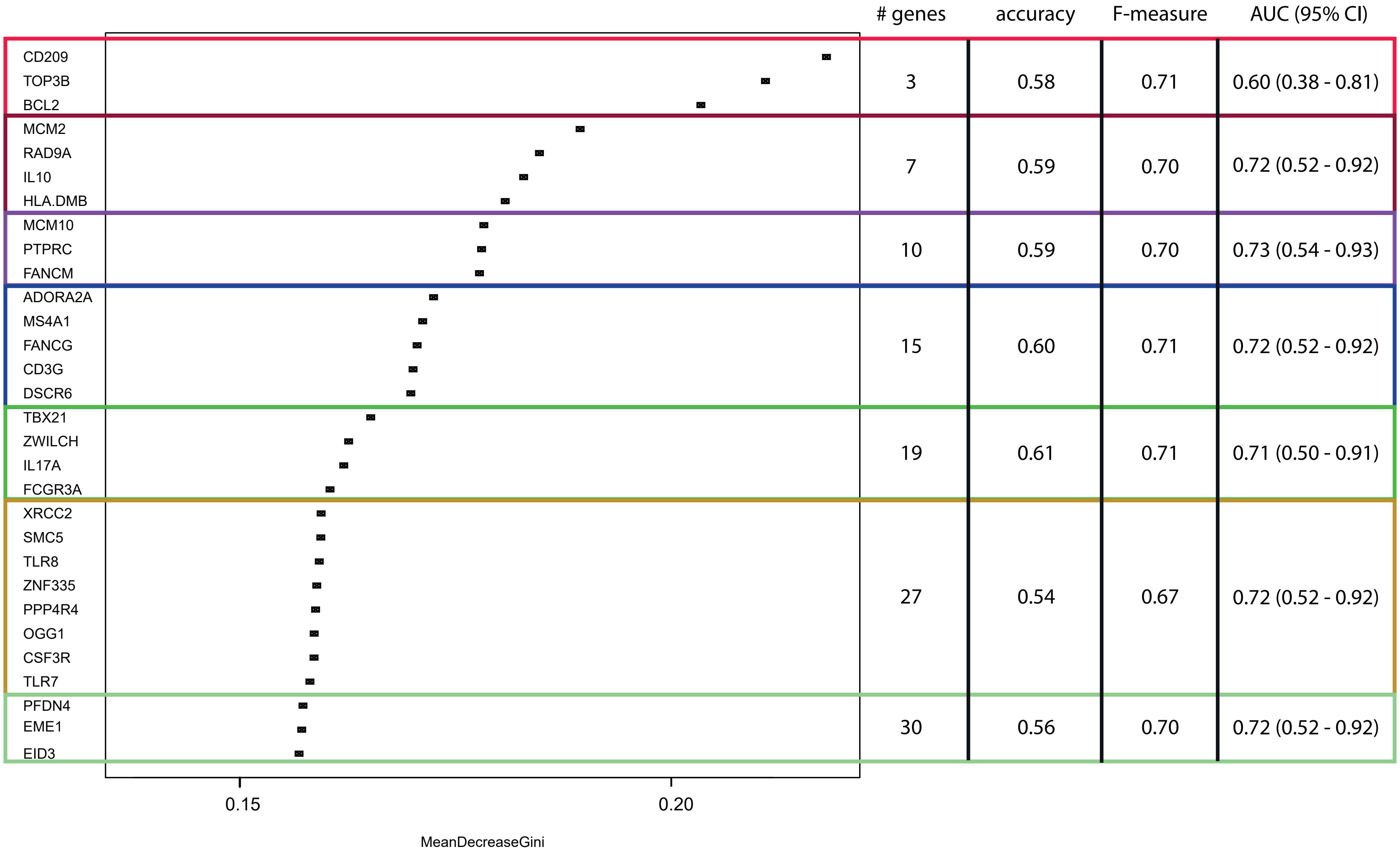
DFS Pac-UFT Trial vs Biomarker

OS Pac-UFT Trial vs Biomarker

DFS Pac-S-1 Trial vs Biomarker

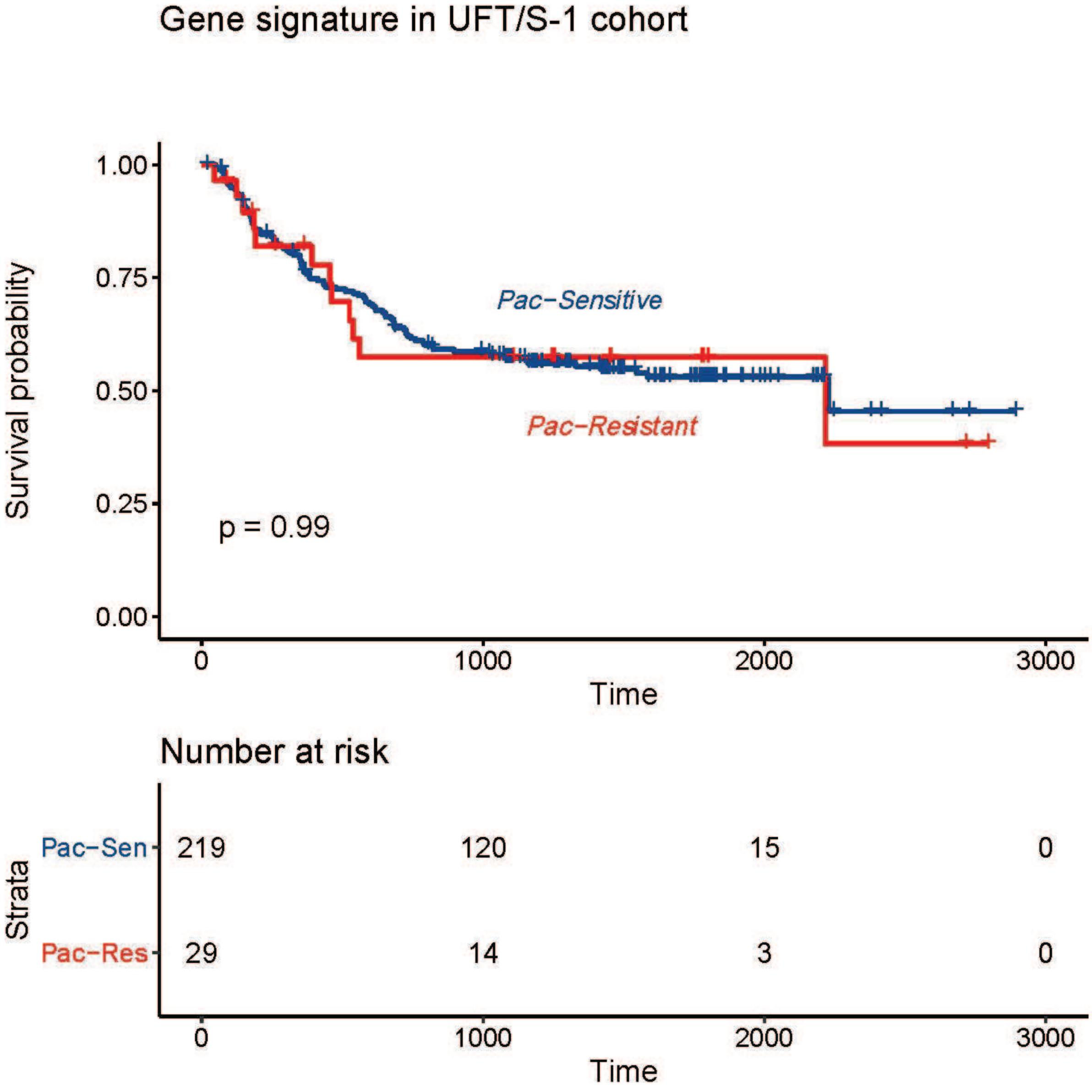
OS Pac-S-1 Trial vs Biomarker

Supplementary Figure 3



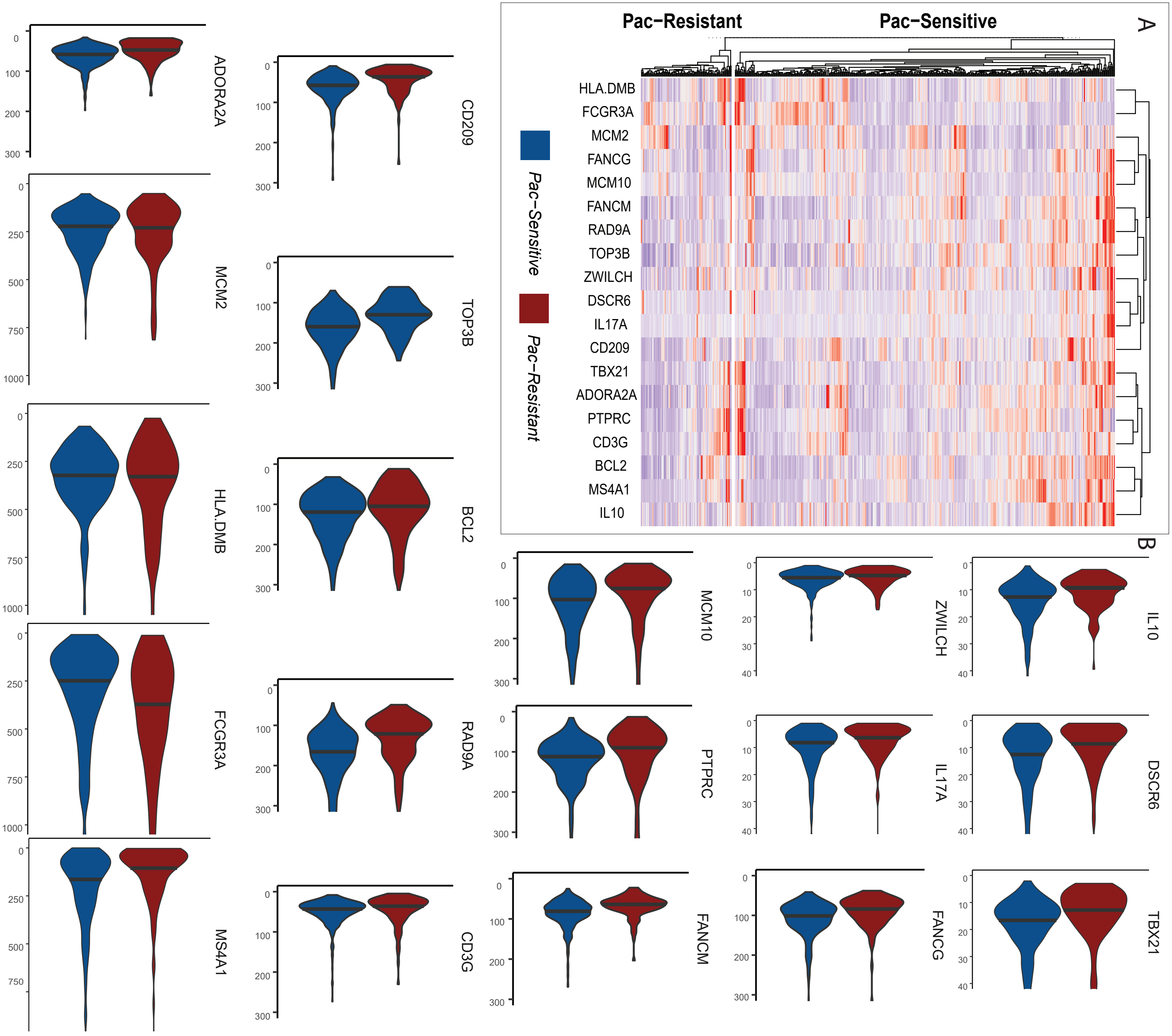
# genes	accuracy	F-measure	AUC (95% CI)
3	0.58	0.71	0.60 (0.38 - 0.81)
7	0.59	0.70	0.72 (0.52 - 0.92)
10	0.59	0.70	0.73 (0.54 - 0.93)
15	0.60	0.71	0.72 (0.52 - 0.92)
19	0.61	0.71	0.71 (0.50 - 0.91)
27	0.54	0.67	0.72 (0.52 - 0.92)
30	0.56	0.70	0.72 (0.52 - 0.92)

Supplementary Figure 4



Sundar R, *et al. Gut* 2022; 71:676–685. doi: 10.1136/gutjnl-2021-324060

Supplementary Figure 5



Supplementary Figure Legends

Supplementary Figure 1. QC of NanoString Panel

A: Principal component analysis of NanoString profiles by treatment arm: No obvious clustering seen suggesting batch effect or treatment effect

B: Principal component analysis of NanoString profiles by treatment centre: No obvious clustering seen suggesting batch effect

C: 2 samples were tested in triplicates in the NanoString profiling (SMT 741 and SMT 745). Analysis of the results between the triplicates yielded a very high correlation. Spearman's R = 0.99, p < 0.001 for all comparisons

Supplementary Figure 2. Kaplan-Meier curves of original trial and biomarker cohort

Survival curves of disease free survival (DFS) and overall survival (OS) of between the original SAMIT trial and the biomarker sub-study, by trial arms are displayed. Original trial ("Trial") is depicted by blue lines, while biomarker sub-study ("Biomarker") is depicted by red lines. There is no statistically significant difference in survival in any of the arms, between Trial and Biomarker.

Supplementary Figure 3. Variable importance plot (or Gini Index) of genes included in the NanoString panel after running random forest on the Pac-S1 arm. The higher the index, the larger the importance of the variable to the gene-signature. Sensitivity analysis was performed using accuracy and F-measure on the Pac-UFT arm at various gene number cut-points, before selecting the 19-gene signature for future analysis.

Supplementary Figure 4. Kaplan-Meier curve of DFS of patients classified by the random forest 19-gene signature into *Pac-Sensitive* (blue) and *Pac-Resistant* (red) groups in the validation S-1 and UFT arms. HR 0.99, 95% CI: 0.55 to 1.82, p = 0.99

Supplementary Figure 5. Transcriptomic characteristics of paclitaxel gene signature

A: Heatmap of NanoString profile of the 19 selected genes in signature, by *Pac-Sensitive* and *Pac-Resistant* grouping. NanoString gene expression is represented in columns, scaled. Blue to red denotes gene expression, with blue implying low gene expression and red implying high gene expression.

B: Violin plot of gene expression levels of 19 selected genes by pairwise between *Pac-Sensitive* and *Pac-Resistant* grouping.