Description of Additional Supplementary Files

Supplementary Data 1 – Clinical data and 10X run information Anonymised clinical data of all patients used in the breast cancer dataset.

Supplementary Data 2 – Differential gene expression analysis of breast cancer clusters as seen at resolution 0.4 Table showing results of differential gene expression analysis of all clusters as seen at resolution 0.4 before merging clusters to phenotypes. The differential gene expression (MAST) analysis was done using two-sided and p values were adjusted by Bonferroni (p_val_adj).

Supplementary Data 3 – Differential gene expression analysis of all CAF and pericyte phenotypes Table showing results of differential gene expression analysis of all defined CAF and pericyte phenotypes. The differential gene expression (MAST) analysis was done using two-sided and p values were adjusted by Bonferroni (p_val_adj).

Supplementary Data 4 – Differential gene expression analysis comparing FAP+ and FAPCAFs. Table showing results of differential gene expression analysis of FAP+ versus FAP- cells. The differential gene expression (MAST) analysis was done using two-sided and p values were adjusted by Bonferroni (p_val_adj).

Supplementary Data 5 – Differential gene expression analysis of all clusters in the integrated dataset. Table showing results of differential gene expression analysis of all defined CAF and pericyte clusters in all datasets after dataset integration at the resolution of 0.7. The differential gene expression (MAST) analysis was done using two-sided and p values were adjusted by Bonferroni (p_val_adj).

Supplementary Data 6 – Differential gene expression analysis of all phenotypes in the integrated dataset, after merging of clusters. Table showing results of differential gene expression analysis of all defined CAF and pericyte phenotypes in all datasets after dataset integration. The differential gene expression (MAST) analysis was done using two-sided and p values were adjusted by Bonferroni (p_val_adj).

Supplementary Data 7 – Comparison of gene set enrichment analysis in breast cancer dataset and integrated dataset Table showing the top 20 hallmark pathways for each CAF type, for the breast cancer and integrated validation datasets. Overlapping pathways are highlighted in green, and defining pathways are highlighted in bold in the breast cancer data.