

Description of Additional Supplementary Files

Supplementary Data 1: Single cell annotations and dataset integration.

Figure 1.1. Gene signatures for the myeloid cell annotations

a) Tumour myeloid UMAP coloured by (i) patient, and (ii) broad myeloid cell type
Gene expression signatures of (b) broad myeloid cell types, (c) DC, (d) monocyte-derived macrophages (momac), (e) mast cell, (f) monocyte, and g) ILC subpopulations.
The suffix “b” indicates blood-derived populations.

Figure 1.2. Gene signatures for the T/NK cell annotations

a) UMAP coloured by sample source.
b) (i) CITE-seq and (ii) gene expression signatures of NK versus T cells.
c) (i) CITE-seq and (ii) gene expression signatures of CD4 versus CD8 T cells.
d) CITE-seq signatures of (i) CD4 T cell, (ii) CD8 T cell and (iii) NK cell populations. CITE-seq values are scaled by maximum value per cell type group.
e) Gene expression signatures of (i) CD4 T cell, (ii) CD8 T cell and (iii) NK cell populations.

Figure 1.3. Gene and VDJ signatures for the B cell annotations.

a) UMAP plots of B cells coloured by (i) source, (ii) somatic hypermutation level, (iii) isotype and (iv) VDJ expression level.
b) Correlation of IGH and IGK/L UMI counts per cell, coloured by PDAC sample type.
c) The per cell subpopulation (i) somatic hypermutation levels and ii) VDJ expression level.
d) The isotype usage percentages across cell types within each cell population.
e) Gene expression profiles of B cell subpopulations of the top differentially expressed genes.

Figure 1.4. Clone size distributions across cell types for (left) B cell and (right) T cell populations, considering only intra-tumoural immune cells.

Figure 1.5

a) Correlation of cell types between technical repeats of biopsy samples for the broad immune cell subsets, B cell subsets, T cell subsets and myeloid cell subsets.
b) Boxplots of differences between biopsy and blood immune cell proportions. Each dot represents a patient sample.

Figure 1.6. a) Schematic of SVMCellTransfer alongside the advantages of this method over established methods (left).

b) Comparison of reference annotation of T and NK cells using SVMCellTransfer (using T and NK cell PancrImmune reference), and Azimuth annotation (using either pancreas or PBMC references). High-confidence T and NK cells were subsampled from the Peng and Steele datasets (confirmed by manual checking of key T and NK genes), and applied both the SVMCellTransfer (using T and NK cell PancrImmune reference), and Azimuth annotation (using either pancreas or PBMC references). UMAP plots show the distribution of predicted cell types by each method.

Figure 1.7. UMAP distributions of the integrated PDAC150K, Peng, et al. and Steele, et al. datasets.

Figure 1.8. Gene expression signatures of (ai) T and NK cell types, (ii) DC, (d) myeloid cell types, (iii) B cell subtypes, (bi) non-immune cell types, and (ii) CAF cell types of the integrated PDAC150K, Peng, et al. and Steele, et al. datasets.

Supplementary Data 2: Summary of scRNA-seq datasets.

Supplementary Data 3: Association between B and T cells and survival.

Supplementary Data 4: Correlation p-values between cell type proportions.

Supplementary Data 5: Top 10 DGE genes between ME and AE groups (pseudobulk analysis).

Supplementary Data 6: Numbers of immunosurveilling and private B and T cells per patient.

Supplementary Data 7: Receptor-ligand list used for Figure 5.

Supplementary Data 8: Cell-cell communication statistical analyses using MANOVA.

Supplementary Data 9: CITE-seq antibody information used in the PancrImmune dataset.

Supplementary Data 10: Cell and Pathway Marker Genes.

Supplementary Data 11: Antibodies used for dual-colour assay.