Supplementary Note:

Assessing Garnett performance with missing cell types and low data quality

We further used the PBMC datasets to explore the limits of Garnett. For example, to assess how Garnett would perform if a cell type was missing from the marker file, we removed the T cell definition from the PBMC marker file, trained a classifier on the PBMC V2 data, and applied it to the V1 dataset. Nearly all T cells were labelled as "Unknown", while other cell types remained accurately labeled (**Supplementary Figure 2A-B**). A small subset of T cells were labelled as natural killer cells and on further investigation, we found that they expressed markers of NKT cells (**Supplementary Figure 2C**). Conversely, removing T cells from the training data ablated Garnett's ability to recognize them during classification, leaving them labeled "Unknown" - although once again, this potentially revealed NKT cells (**Supplementary Figure 2D-E**).

We next wanted to explore the limits of Garnett with respect to data quality. For each test, we trained a classifier from the V2 dataset and then classified cells from the V1 dataset where we had FACS based cell type assignments. Downsampling the reads on the training set to as little as 25% of the original depth per cell did not degrade classification accuracy (**Supplementary Figure 3A**). Downsampling reads on the test data reduced Garnett's accuracy, suggesting that even well-trained classifiers have difficulty labeling cells with shallow molecular profiles. However, cluster-extension was largely able to rescue classification accuracy with shallow sequencing, even at only 10% of the original depth (**Supplementary Figure 3B**). Along similar lines, we wondered how well Garnett could recognize rare cell populations, so we downsampled the number of T cells in the training dataset and found that even with only a few T cells included, classification accuracy remained high (**Supplementary Figure 3C**).

Supplementary Table 1. Consensus cell types for lung datasets.

Consensus type	Tabula muris type	MCA type	Lambrechts type
Alveolar	epithelial cell of lung	AT2 Cell AT1 Cell Alveolar bipotent progenitor	Alveolar
B cells	B cell	Ig−producing B cell B Cell	B cells
Ciliated cells	ciliated columnar cell of tracheobronchial tree	Ciliated cell Clara Cell Dividing cells	Epithelial
Endothelial	lung endothelial cell	Endothelial cell_Tmem100 high Endothelial cells_Vwf high Endothelial cell_Kdr high	Endothelial
Fibroblasts	stromal cell	Stromal cell_Inmt high Stromal cell_Dcn high Stromal cell_Acta2 high	Fibroblasts
Granulocytes		Neutrophil granulocyte Eosinophil granulocyte Basophil	
Leukocytes	leukocyte		
Macs/Monos/DCs	classical monocyte monocyte	Conventional dendritic cell_H2-M2 high Dendritic cell_Naaa high Dividing dendritic cells Conventional dendritic cell_Tubb5 high Conventional dendritic cell_Gngt2 high Plasmacytoid dendritic cell Conventional dendritic cell_Mgl2 high Monocyte progenitor cell Interstitial macrophage Alveolar macrophage_Ear2 high Alveolar macrophage_Pclaf high	
Myeloid	myeloid cell		Myeloid
Natural killer cells	natural killer cell	NK Cell	
T cells	T cell	T Cell_Cd8b1 high Dividing T cells Nuocyte	T cells
Tumor			Tumor