Supplementary Materials

Supplementary Tables

Table S1. A list of public scRNA-Seq datasets used as reference datasets by ImmClassifier.

Dataset	Platform	Number of immune cells/samples	Number of immune cell types	Tissue,Cancer type/Normal	Ref
hca-bm	10X, 3' sequencing	101,935	32	Bone marrow, Normal	(Hay, Ferchen, Chetal, Grimes, & Salomonis, 2018)
pbmc	10X, 3' sequencing	69,745	31	PBMC, GI cancer	(Griffiths et al.)
liver-immune	10X, 3' sequencing	4,906	6	Liver, Normal	(MacParland et al., 2018)
jci-bm	10X, 3' sequencing	73,246	18	Bone marrow, Normal	(Oetjen et al., 2018)
nsclc-zilionis-tii-minor	inDrop, full-length	34,558	34	Tumor infiltrating immune cells, NSCLC	(Zilionis et al., 2019)
brcatil	10X, 3' sequencing	6,311	9	Tumor infiltrating immune cells, TNBC	(Savas et al., 2018)
nsclc-guo	Smart-seq2 Full-length	12,346	16	Tumor infiltrating immune cells, NSCLC	(Guo et al., 2018)

Table S2. Full name of cell types in cell type hierarchy.

Full Name	Short Name
CD34	CD34-positive stem cell
L	lymphocyte
M	myeloid cell
HSC	hematopoietic stem cell
pDC	plasmacytoid dendritic cell
mDC	monocyte-derived dendritic cell
cDC	conventional dendritic cell
Mast	mast cell
Mega	megakaryocyte
Ery	erythrocyte
Neu	neutrophil
Mono	monocyte
Mac	macrophage
Eos	eosinophil
T	T cell
NK	natural killer cell
B	B cell
unconvT	unconventional T cell
CD4+	CD4-positive T cell
CD8+	CD8-positive T cell
CD4+Tn	CD4-positive naïve T cell
CD4+Tcm	CD4-positive central memory T cell
CD4+Tem	CD4-positive effector memory T cell
CD4+Trm	CD4-positive tissue-resident memory T cell
CD4+Tex	CD4-positive exhausted T cell

Treg	regulatory T cell
Tfh	helper T cell
CD8+Tn	CD8-positive naïve T cell
CD8+Tcm	CD8-positive central memory T cell
CD8+Tem	CD8-positive effector memory T cell
CD8+Tex	CD8-positive exhausted T cell
CD8+Trm	CD8-positive tissue-resident memory T cell
MAIT	mucosal invariant T cell
gdT	Gamma-delta T cell
PC	plasma cell

Table S3. The number of cells sampled to train deep neural network.

Cell number (N)	Sampled cell number per cell type
N<1000	1000
1000 <= N <2000	1200
2000 <= N <5000	1400
N>=5000	1600

Table S4. Mapping IDs of cell types between different annotation methods.

Dataset	Cell type	Original	ImmClassifier	SingleR	Garnett	SCINA	SciBet
Brca3p	Treg	T:reg	L:T:CD4:Treg	Treg*		Treg.cells	
	CD4+T	T:CD4.*	L:T:CD4*	T_cells:CD4*	CD4 T cells		
	CD8+T	T:CD8*	L:T:CD8*	T_cells:CD8*	CD8 T cells	CD8.T.cells	
	B	B	L:B	B_cell*	B cells	B.cells	B cell, Plasma B
	NK	NK*(NKT excluded)	L:NK	NK*	NK cells	CD56dim.NK.cells	NK
	mDC	mDC	M:cDC	DC:m*	Dendritic cells	Dendritic.cells	DC
	pDC	pDC	L:pDC,M:pDC				
	MONOCYTE	MONOCYTE	M:Mono	Monocyte*	Monocytes		Monocyte
	MACROPHAGE	MACROPHAGE	M:Mac	Macrophage*		Macrophages	Macrophage
	Neutrophil	NEUTROPHIL	M:Neu	Neutrophl*		Neutrophils	
	Mast	MAST	M:Mast			Mast.cells	Mast
	Unassigned				Unknown	unknown	
	T				T cells		T cell, ILC
	NKT	NKT					

	Other		CD34*, L:unconvT, M:Eos, M:Ery, M:Mega,M:Platelet,L:PC	Chondrocytes*, CMP, Embryonic_stem_cells, Endothelial_cells*, Erythroblast, Fibroblast*, GMP, HSC*, iPS_cells*,pre-B*, pro-B*, Smooth_muscle_cells*, Tissue_stem_cells*			Epithelial, Microglia, Endothelial
Brcap5p	T:CD4+Naive	T:CD4+Naive	L:T:CD4:Naïve	T_cell:CD4+_Naive			
	T:CD4+CM	T:CD4+CM	L:T:CD4:CM	T_cell:CD4+_central_memory			
	T:CD4+EM	T:CD4+EM	L:T:CD4:EM	T_cell:CD4+_effector_memory			
	T:CD8+Naive	T:CD8+Naive	L:T:CD8:Naïve	T_cell:CD8+_naive			
	T:CD8+CM	T:CD8+CM	L:T:CD8:CM	T_cell:CD8+_Central_memory			
	T:CD8+EM	T:CD8+EM	L:T:CD8:EM	T_cell:CD8+_effector_memory*			
	Treg	Treg	L:T:CD4:Treg	T_cell:Treg:Naive		Treg.cells	
	CD4+T			T_cell:CD4+	CD4 T cells		
	CD8+T			T_cell:CD8+	CD8 T cells	CD8.T.cells	
	T				T cells	T cell	
	Exhausted T		L:T:CD4:Ex,L:T:CD8:Ex	T_cell:gamma-delta			
	Tissue-resident memory		L:T:CD4:TRM,L:T:CD8:TRM				
	Other		CD34*,L:unconvT*,M*,L:B*,L:T:CD4:Tfh	Tissue_stem_cells:*,B_cell*,Endothelial_cells*,Epithelial_cells*,Keratinocytes*,Macrophage*,Monocyte*,NK*,T_cell:gamma-delta	Tfh.cells, Tm.cells, Neutrophil, Th*.cells, Macrophage, Dendritic.cells, B.cells, CD56*NK.cells		B cell, NK, Macrophage, Plasma B, Mast, HSCs, Skeletal Muscle Myoblasts

hcc	cytotoxic CD4 cells	cytotoxic CD4 cells					
	effector memory CD8+T cells	effector memory CD8+T cells	L:T:CD8:EM				
	exhausted CD4+ T cells	exhausted CD4+ T cells	L:T:CD8:Ex				
	intermediate of EM and Exhausted T	intermediate of EM and Exhausted T	L:T:CD4:EM, L:T:CD8:EM, L:T:CD4:TRM, L:T:CD8:TRM, L:T:CD4:Ex, L:T:CD8:Ex	T_cell:effector_memory*			
	exhausted CD8+ T cells	exhausted CD8+ T cells	L:T:CD4:Ex				
	MAIT cells	MAIT cells	L:unconvT:MAIT				
	naive CD4+ T cells	naive CD4+ T cells	L:T:CD4:Naïve	T_cell:CD4+_Naive			
	Tregs	Tregs*	Treg	Treg*		Treg.cells	
	naive CD8+ T cells	naive CD8+ T cells	L:T:CD8:Naïve	T_cell:CD8+_Naive			
	CD4+T			T_cell:CD4+			
	CD8+T			T_cell:CD8+		CD8.T.cells	
	T helper	T helper	L:T:CD4:Tfh			Th*cells	
	T				T cells	T.cells	T cell, ILC

	unknown	unknown					
	central memory T		L:T:CD4:CM,L:T:CD8:CM	T_cell:Central_memory*			
	Other		CD34, L:B*, L:NK, M*	Endothelial_cells*, Epitheial_cells, NK*,T_cell:gamma-delta	B cells, CD34+, Dendritic cells, Monocytes, NK cells	Tm.cells Treg.cells, Mast.cells, *NK*, Neutrophils, Dendritic.cells, Macrophages, Tfh*	NK, Macrophage, B cell, HEK and 3T3 mix, epithelial cells, Mast, HSCs, Plasma B, foreskin fibroblast
	Unassigned				Unknown	unknown	
pbmc68k	CD14+ Monocyte	CD14+ Monocyte	M:Mono	Monocyte	Monocytes		Monocyte
	CD19+ B	CD19+ B	L:B*	B_cell*, Plasma_cell* ,pre-B*	B cells	B.cells	B cell, Plasma B
	CD34+	CD34+	CD34*	CMP,GMP,MEP,CD34+*	CD34		HSCs
	CD4+ T Helper2	CD4+ T Helper2	L:T:CD4:Tfh			Th*, Tfh*	
	CD4+/CD45RA+/CD24-Naive T	CD4+/CD45RA+/CD24-Naive T	L:T:CD4:Naïve	T_cell:CD4+_Naive			

CD4+/CD45RO+ Memory	CD4+/CD45RO+ Memory	L:T:CD4:CM,L:T:CD4:EM	T_cell:CD4+_central_memory,T_cell:CD4+_effector_memory			
CD56+ NK	CD56+ NK	L:T:NK	NK*	NK cells	*NK*	NK
CD8+ Cytotoxic T	CD8+ Cytotoxic T					
CD8+/CD45RA+ Naive Cytotoxic	CD8+/CD45RA+ Naive Cytotoxic	L:T:CD8:Naïve	T_cell:CD8+_Naive			
Dendritic	Dendritic	*DC*	DC*	Dendritic cells	Dendritic.cells	DC
CD4+T			T_cell:CD4+	T_cell:CD4+		
CD8+T			T_cell:CD8+	T_cell:CD8+	CD8.T.cells	
CD8+ memory		L:T:CD8:CM,L:T:CD8:EM	T_cell:CD8+_Central_memory,T_cell:CD8+_effector_memory*			
CD4+/CD25 T Reg		L:T:CD4:Treg	Treg*		Treg.cells	
T				T cells		T cell
Other		L:T:CD4:TRM,L:T:CD8:TRM,M:Eos,M:Ery,M:Mac,M:Mega*,M:Neu,L:unconvT* , L:T:CD4:Ex, L:T:CD8:Ex	Platelets, pro-B*, T_cell:gamma-delta			Mast,Epithelial, Pancreatic islets, Endothelial, hepatocytes
unassigned				Unknown		

Table S5. A list of public scRNA-Seq and bulk-sequencing datasets used for evaluation datasets by ImmClassifier.

Dataset	Platform	Number of immune cells/samples	Number of immune cell types	Tissue,Cancer type/Normal	Ref
brca3p	inDrop,3' sequencing	47,016	19	PBMC, tumor, adjacent normal and lymph nodes, BRCA	(Azizi et al., 2018)
brca5p	10X, 5' sequencing	28,013	8	PBMC, tumor, adjacent normal and lymph nodes, BRCA	(Azizi et al., 2018)
hcc	Smart-seq2 Full-length	5,603	12	PBMC, tumor and adjacent normal, HCC	(C. Zheng et al., 2017)
hnscc	Smart-seq2 Full-length	5,902	5	Tumor and lymph nodes, HNSCC	(Puram et al., 2017)
pbmc68k	10X, 3' sequencing	68,579	11	PBMC, Healthy	(G. X. Zheng et al., 2017)
skcm	Smart-seq2 Full-length	16,291	11	Tumor, SKCM	(Sade-Feldman et al., 2018)
bulk	Microarray	211	38	PBMC, BM, Healthy	(Novershtern et al., 2011)

Supplementary Figures

Figure S1
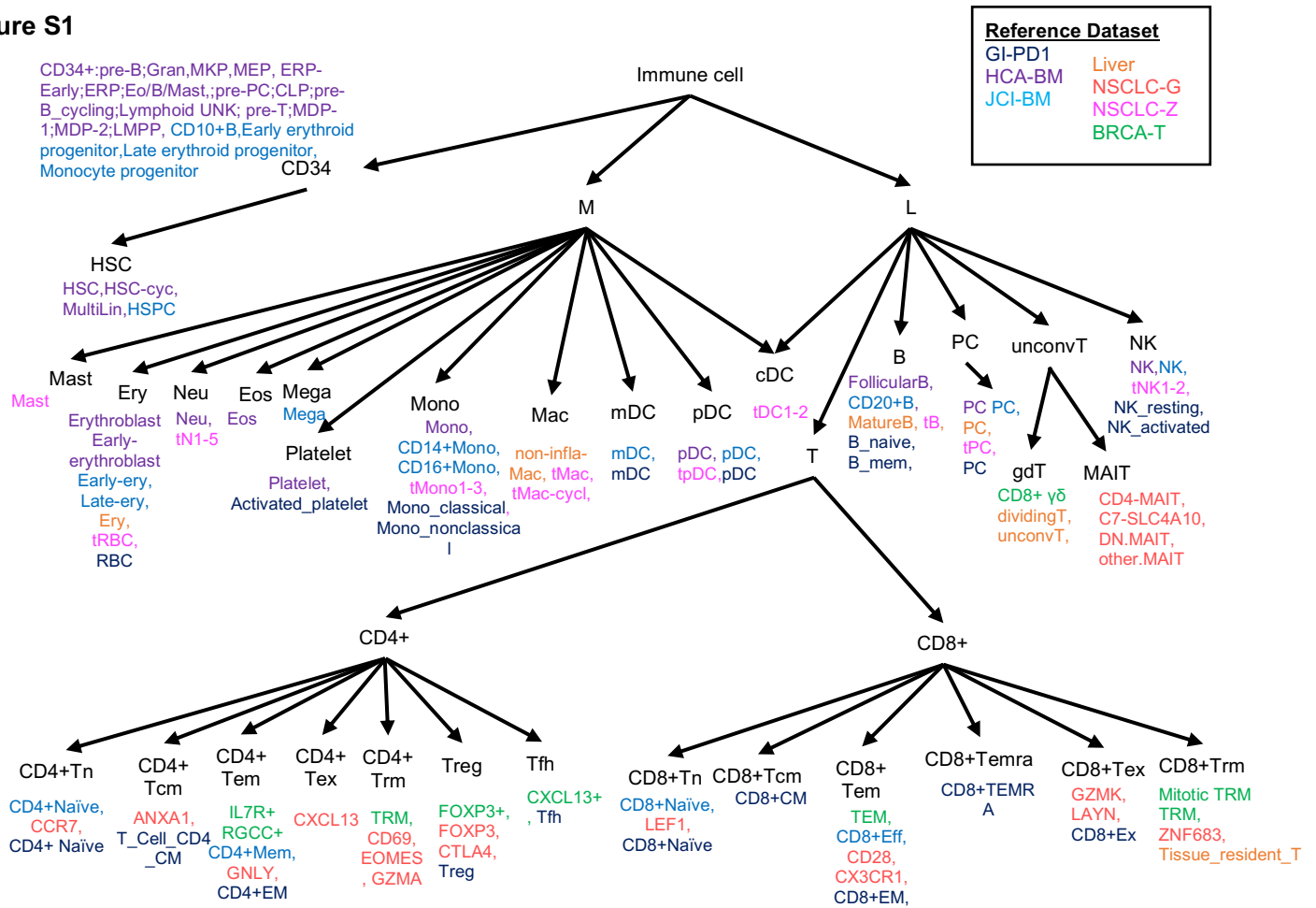
Figure S1. Cell ontology hierarchy used by ImmClassifier. The hierarchy is revised from cell differentiation hierarchy to accommodate for prediction probability. The hierarchy represents cell annotation granularity. Each node on the tree is a flat cell type. For each flat cell type (tree node), the matched cell types from training datasets are listed.
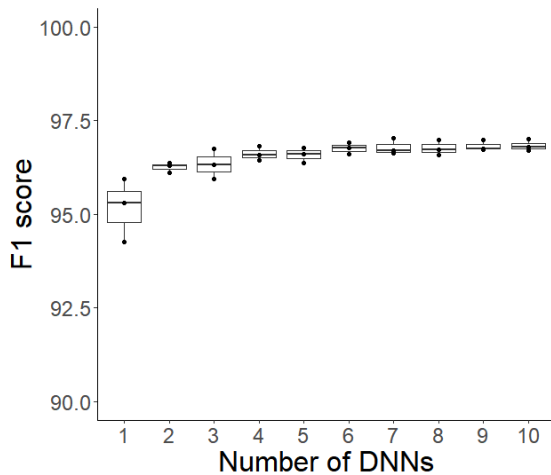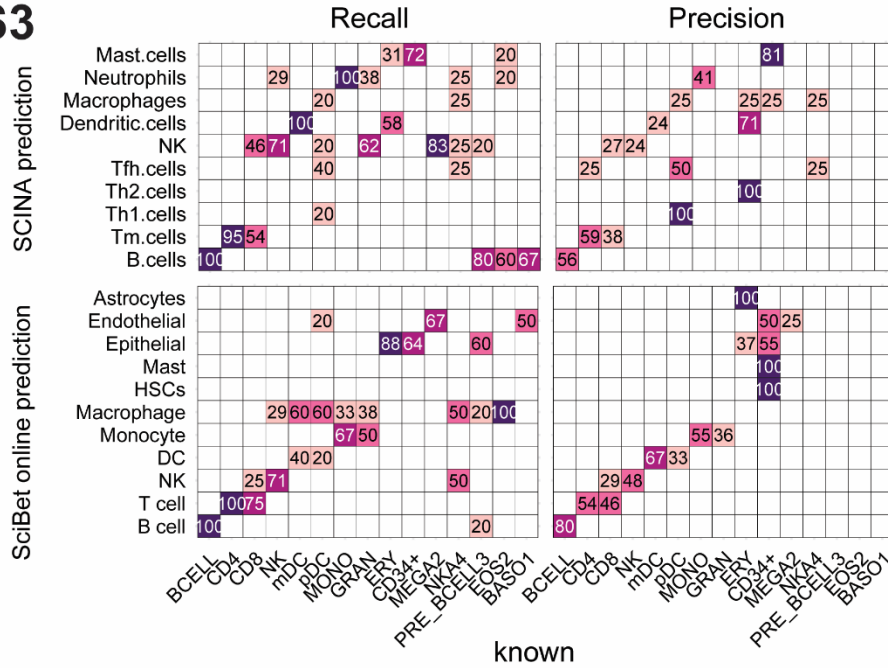
Figure S2

a
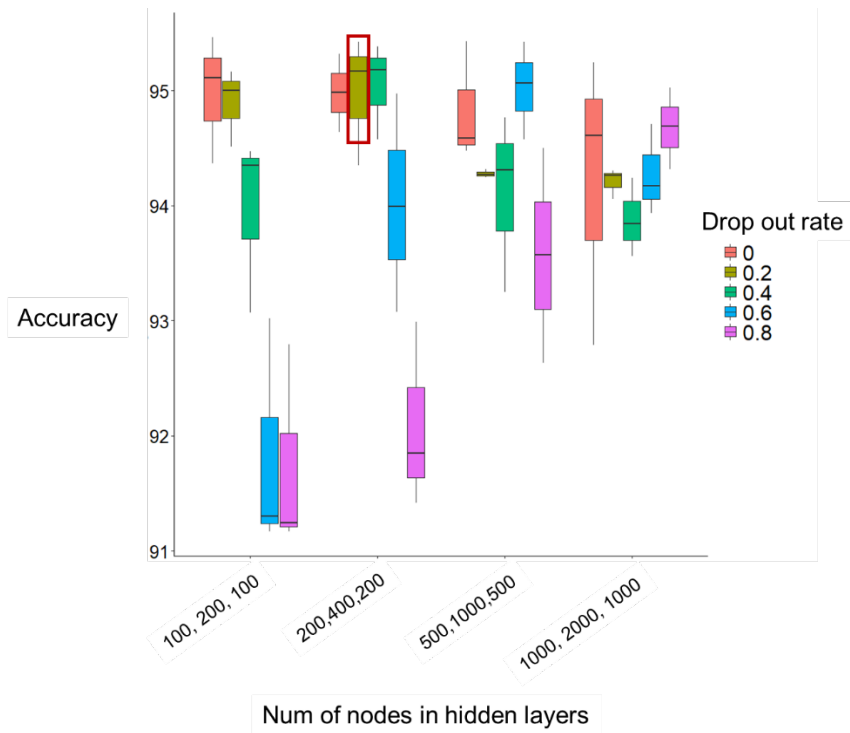
b

Figure S2. Evaluation of hyper-parameters for DNN using 3-fold cross-validation. (a) The prediction accuracy (as a percent) was shown under different combinations of the number of nodes in hidden layers and drop-out rate. The hyper-parameter with the best performance is in the red square. (b) The F1 score (as a percent) is shown as the number of DNNs ranges from 1 to 10.
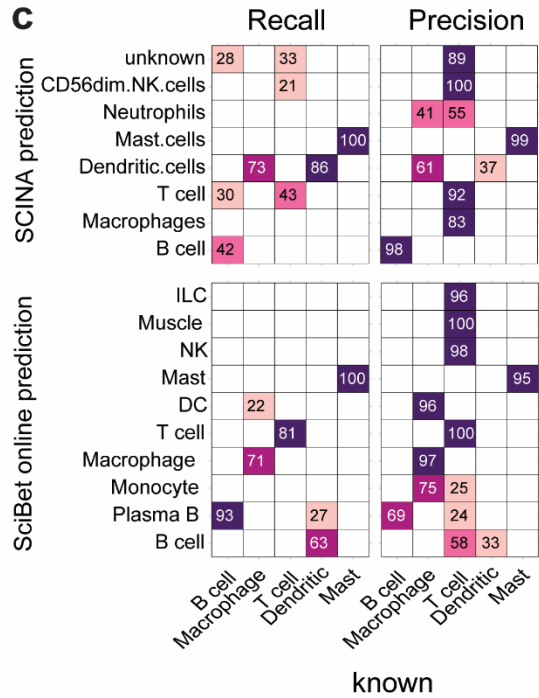
Figure S3
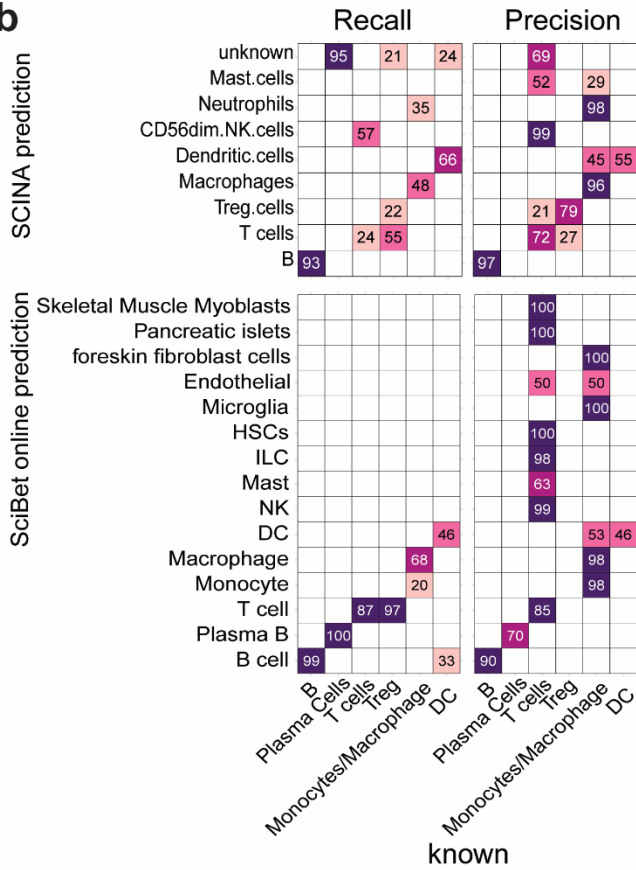
Figure S3. The performance of SCINA and SciBet to predict immune cell types. These heatmaps compare the cell types from the original publication (rows) to those inferred by SCINA and SciBet (columns). The color represents the recall and precision score (as a percent) of each original cell type predicted by ImmClassifier. Recall and precision scores no less than 20 are labeled. Three datasets tested are (a) purified immune populations sequenced by microarray platform. (b) SKCM. The *Other* row includes a small number of cells predicted to cell types not included the original annotated ones. (c) HNSCC. The *Other* row includes a small number of cells predicted to cell types not included the original annotated ones. Since ImmClassifier has finer annotation granularity than the original annotation, for the purpose of comparison, the annotation terms of equivalent granularity to the original annotation were used.
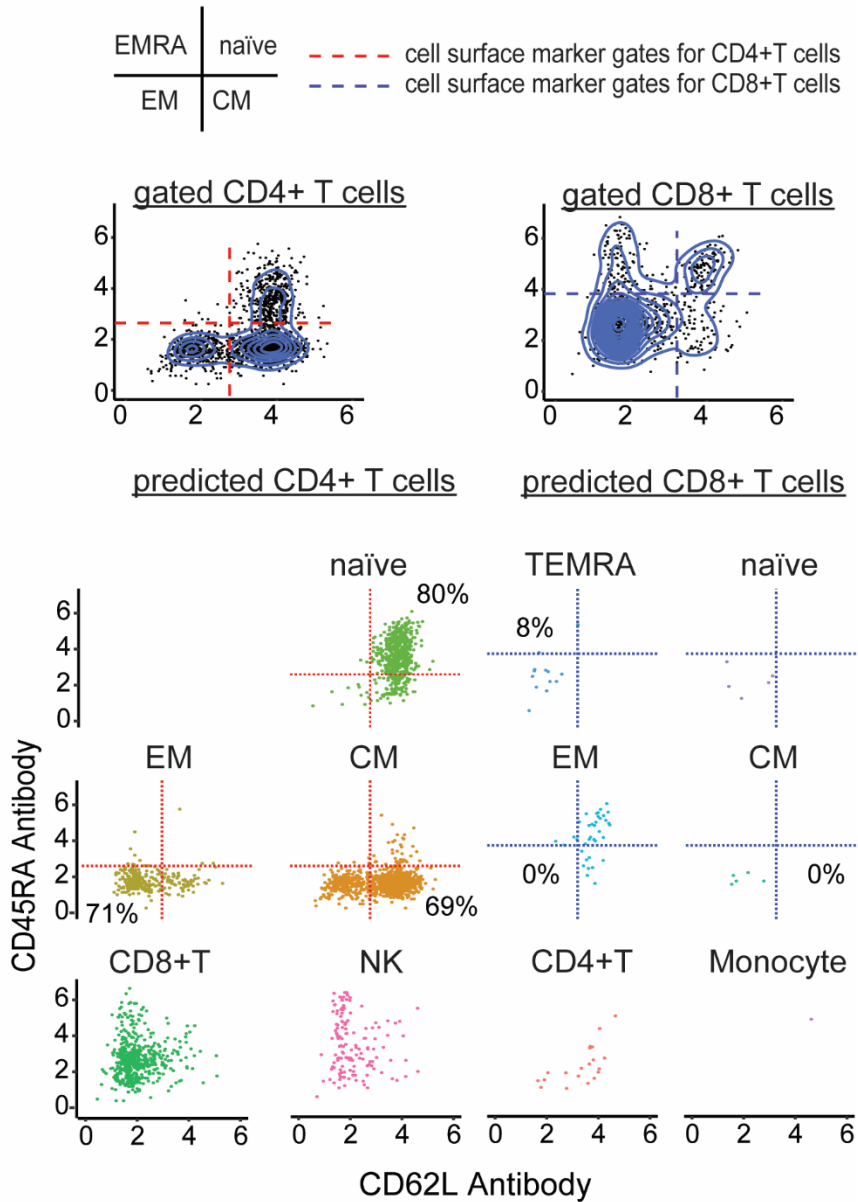
Figure S4. Visualization of the prediction of SingleR on fine-grained T cells. T cells were gated into EMRA, naïve, EM and CM immunophenotypes in the original publication. (Top panel) density plots of the gated CD4+T and CD8+T cells in the original publication. (Bottom panel) scatterplots of the cells in each predicted cell type by SingleR (fine subtype mode). The percentage of cells in each predicted cell type that fall into the concordant gated quadrant is labeled.