Cell Reports, Volume *24*

Supplemental Information

MetaCyto: A Tool for Automated Meta-analysis of Mass and Flow Cytometry Data

Zicheng Hu, Chethan Jujjavarapu, Jacob J. Hughey, Sandra Andorf, Hao-Chih Lee, Pier Federico Gherardini, Matthew H. Spitzer, Cristel G. Thomas, John Campbell, Patrick Dunn, Jeff Wiser, Brian A. Kidd, Joel T. Dudley, Garry P. Nolan, Sanchita Bhattacharya, and Atul J. Butte

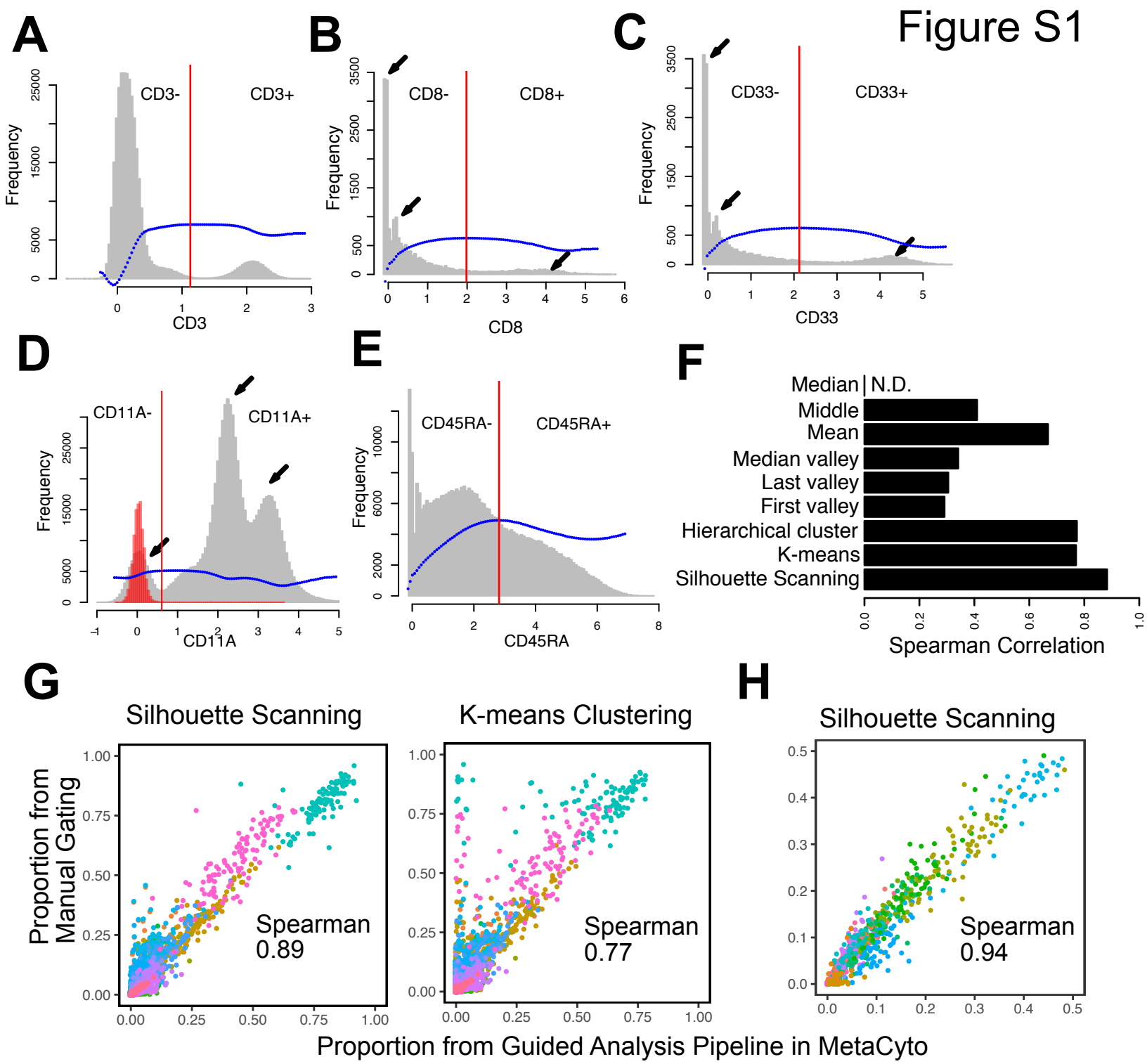


Figure S1: Silhouette scanning bisects the distribution of each marker in a biologically meaningful way. Related to Figure 1. (**A-E**) Examples illustrating how silhouette scanning bisects markers with bimodal distribution (**A**), tri-modal distribution (**B-D**) and a distribution where the positive population does not form a separate peak (**E**). The range of a marker is divided into 100 intervals using 99 breaks. The distribution is bisected at each break and the corresponding average silhouette is calculated. The break that gives rise to the largest average silhouette is used as the cutoff for bisection. Grey histogram shows the distribution of the markers. Blue dots show the average silhouette at each break. Red line shows the cutoff that maximizes the average silhouette. Black arrows show the position of 3 peaks. The red histogram in **D** represents the unstained control. (**F,G**) Using different bisection algorithms, each marker in CyTOF data from SDY420 are bisected into positive and negative regions. 24 cell types were identified using the guided analysis pipeline as described in Fig. 1c. The proportion of each cell type in each sample is calculated and compared with manual gating result. (**F**) The Spearman correlation between the estimated proportion and author's proportion are used to measure the performance of each bisection algorithm. (**G**) Scatter plots showing the result generated by using silhouette scanning and k-means clustering as the bisection algorithm on data from SDY420. Each dot represents the proportion of a cell type in a sample. Each color represents a cell type. See Supplemental Experimental Procedures for a detailed description of the 9 bisection methods tested. (**H**) A scatter plot showing the result generated by using silhouette scanning as the bisection algorithm on data from SDY820.

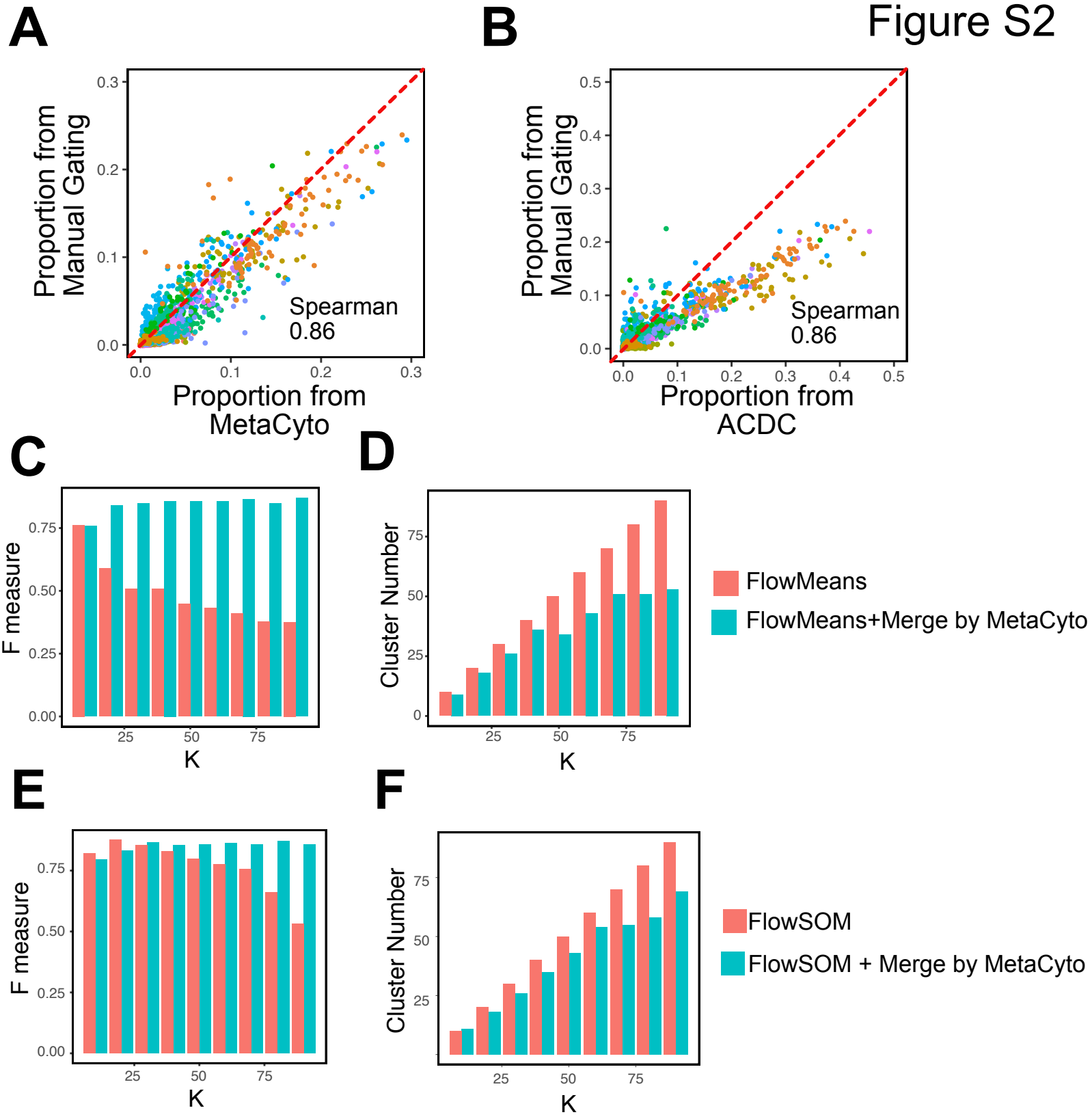


Figure S2: MetaCyto accurately identifies cell populations. Related to Figure 2. (**A-B**) Scatter plots showing the comparison between proportions of cell types estimated by the guided analysis pipeline in MetaCyto (**A**) or ACDC (**B**) and proportions provided by the authors of SDY478. Each dot represents the proportion of a cell type in a sample. Each color represents a cell type. (**C-D**) FlowMeans is used to cluster FlowCAP WNV data with K ranging from 10 to 90 with or without MetaCyto. F measure (**C**) and the number of clusters (**D**) are showed in the bar plots. (**E-F**) FlowSOM is used to cluster FlowCAP ND data with K ranging from 10 to 90 with or without MetaCyto. F measure (**E**) and the number of clusters (**F**) are showed in the bar plots.

Figure S3

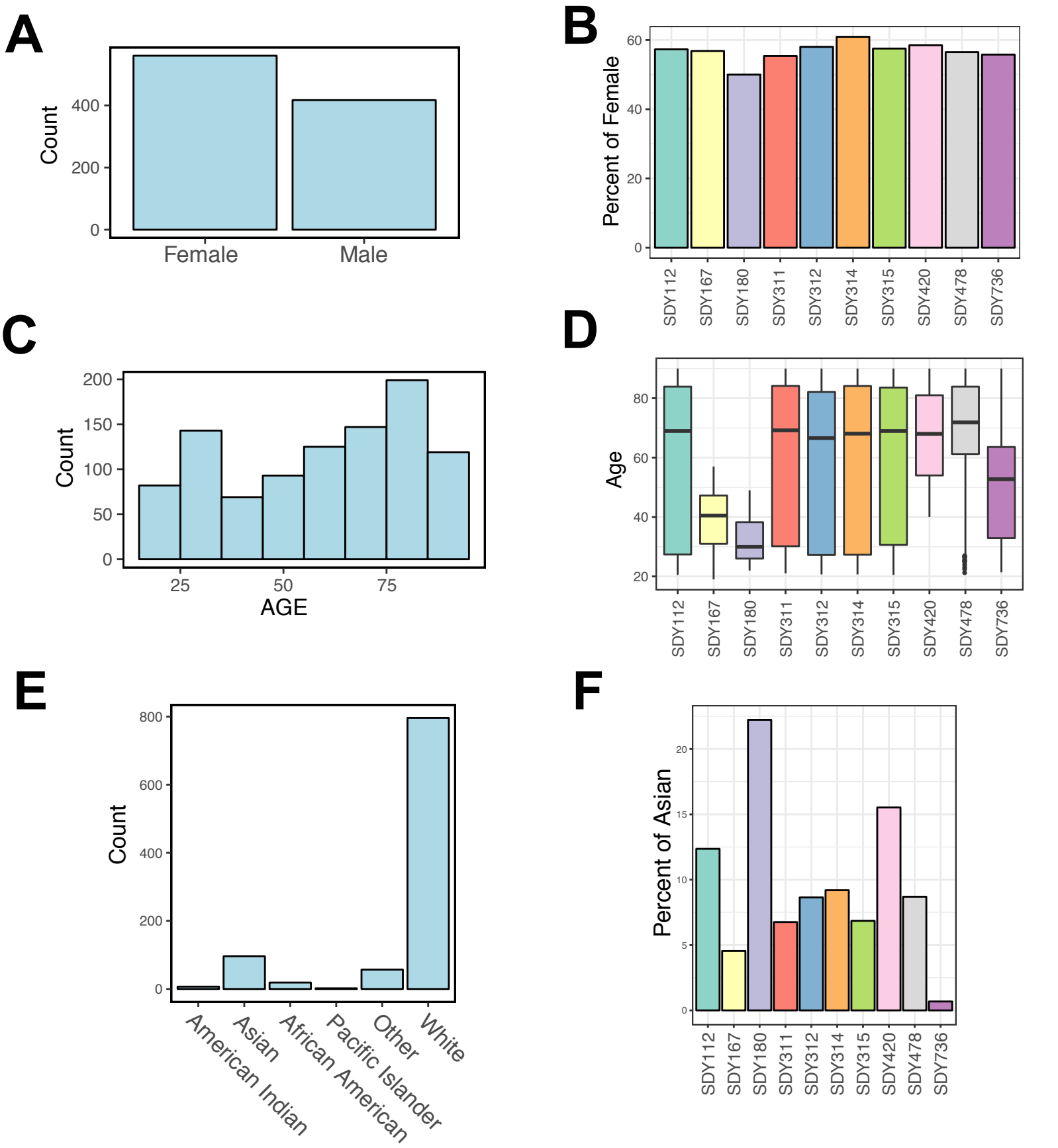


Figure S3: Demographics of subjects included in the meta-anlaysis of 10 human immunology studies. Related to Figure 3. (**A**) Distribution of gender in all studies. (**B**) Distribution of gender in individual studies. (**C**) Distribution of age in all studies. (**D**) Distribution of age in individual studies. (**E**) Distribution of ethnicity in all studies. (**F**) Percent of Asian subjects in individual studies.

Figure S4

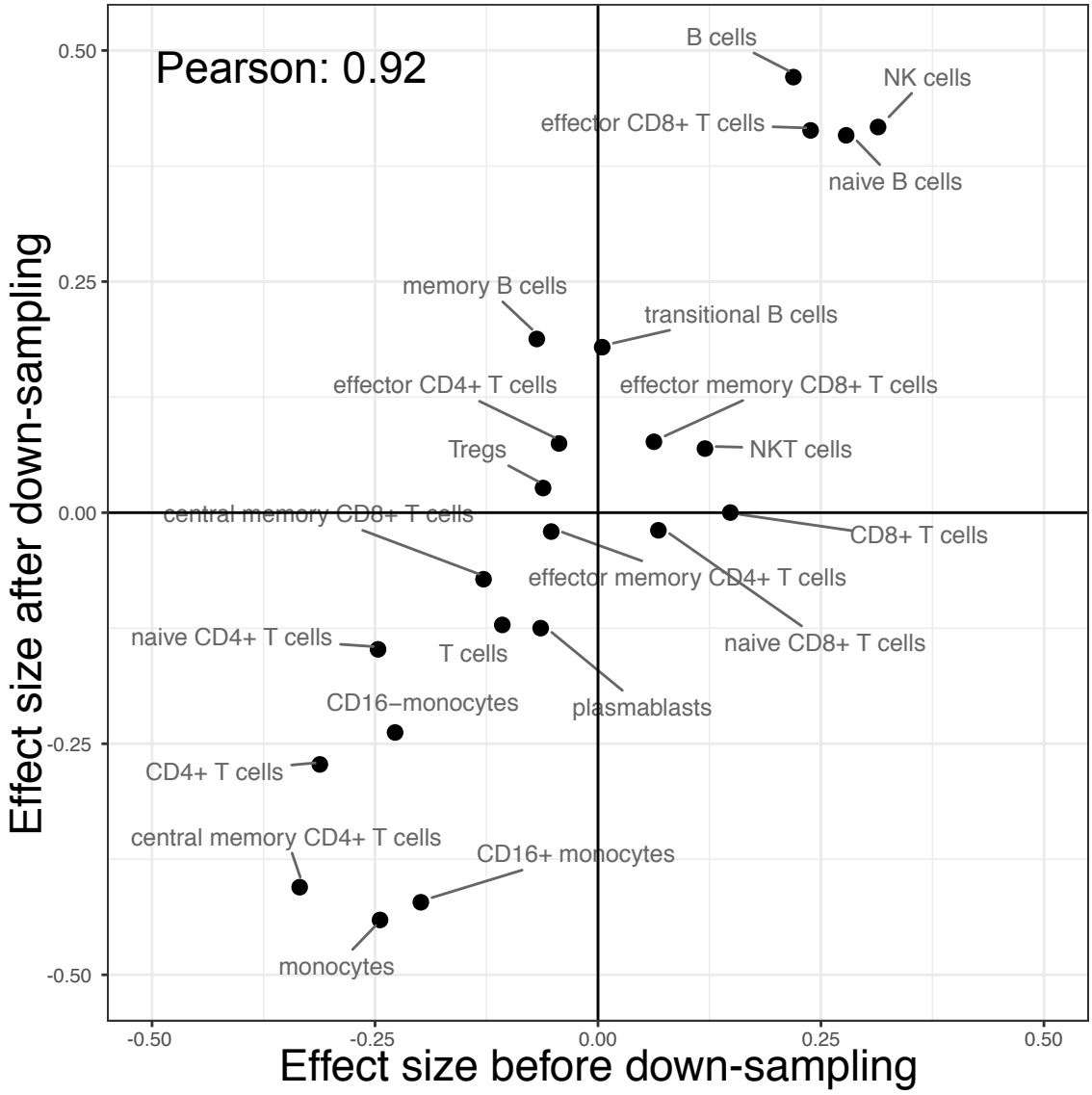
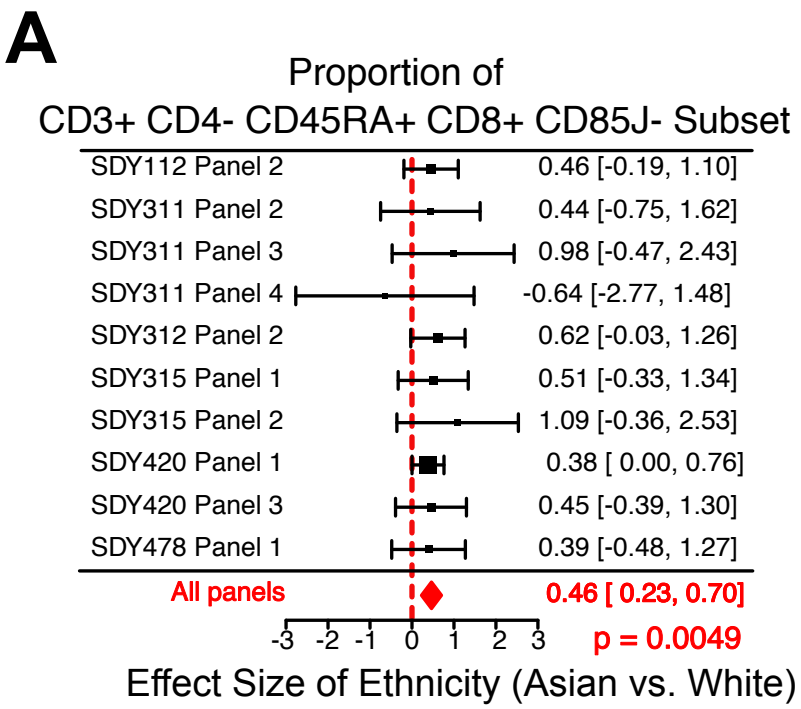


Figure S4: Meta-analysis results are not affected by data imbalance. Related to Figure 5. To test if imbalances in the makeup of ethnicity affects the meta-analysis, we down-sampled White individuals so that the number of White and Asian individuals are equal. The effect sizes of ethnicity on different immune cell types are compared before and after down-sampling.



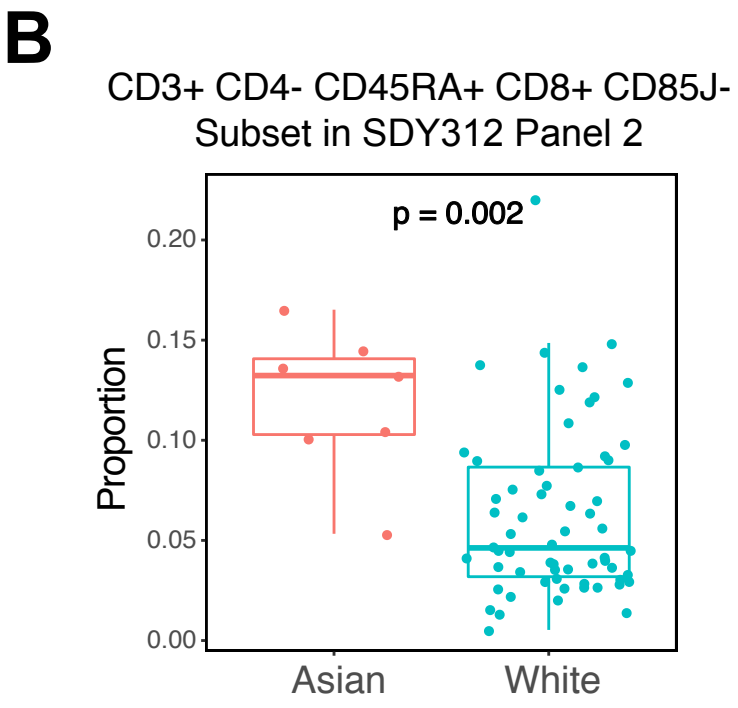


Figure S5: An example of ethnic differences identified by the unsupervised pipeline. Related to Figure 6. (**A**) Forest plot showing the effect size of ethnicity (Asian vs. White) on the proportion of a immune cell subset (CD3+CD4-CD45RA+CD8+CD85J-) identified by the unsupervised analysis pipeline. (**B**) The proportion of the CD3+CD4-CD45RA+CD8+CD85J- cell subset in PBMC from Asian and White subjects in panel 2 of SDY312. The p value in **A** is calculated using a random effect model, adjusted using Benjamini-Hochberg correction. p value in **B** is calculated from unpaired Mann-Whitney test without correction. See Table S6 for a list of differences in immune cells between ethnicities identified by the unsupervised pipeline.

Name	Cell Definitions
B cells	CD14- CD33- CD3- CD19+ CD20+
CD16- monocytes	CD14+ CD33+ CD16-
CD16+ monocytes	CD14+ CD33+ CD16+
CD4+ T cells	CD14- CD33- CD3+ CD4+
CD8+ T cells	CD14- CD33- CD3+ CD8+
central memory CD4+ T cells	CD14- CD33- CD3+ CD4+ CCR7+ CD45RA-
central memory CD8+ T cells	CD14- CD33- CD3+ CD8+ CCR7+ CD45RA-
effector CD4+ T cells	CD14- CD33- CD3+ CD4+ CCR7- CD45RA+
effector CD8+ T cells	CD14- CD33- CD3+ CD8+ CCR7- CD45RA+
effector memory CD4+ T cells	CD14- CD33- CD3+ CD4+ CCR7- CD45RA-
effector memory CD8+ T cells	CD14- CD33- CD3+ CD8+ CCR7- CD45RA-
gamma-delta T cells	CD14- CD33- TCRgd+
lymphocytes	CD14- CD33-
memory B cells	CD14- CD33- CD3- CD19+ CD20+ CD24+ CD38-
monocytes	CD14+ CD33+
naive B cells	CD14- CD33- CD3- CD19+ CD20+ CD24- CD38+
naive CD4+ T cells	CD14- CD33- CD3+ CD4+ CCR7+ CD45RA+
naive CD8+ T cells	CD14- CD33- CD3+ CD8+ CCR7+ CD45RA+
NK cells	CD14- CD33- CD3- CD16+ CD56+
NKT cells	CD14- CD33- CD3+ CD56+
plasmablasts	CD14- CD33- CD3- CD20- CD27+ CD38+
T cells	CD14- CD33- CD3+
transitional B cells	CD14- CD33- CD3- CD19+ CD20+ CD24+ CD38+
Tregs	CD14- CD33- CD3+ CD4+ CD25+ CD127-

Table S1: A list of cell definitions used to identify the 24 cell populations in cytometry data (SDY420) from ImmPort. Related to Figure 2. The cell definitions are created based on the author's gating strategy provided in SDY420.

Cell Definitions for MetaCyto	Cell Definitions for flowDensity	Cell Definitions for ACDC	Name
CD123+|HLADR-	CD123+HLADR-	CD123+|HLADR-	basophils
CD14-|CD33-	CD14-CD33-		lymphocytes
CD14-|CD33-|CD3-	CD14-CD33-/CD3-		non-T lymphocytes
CD14-|CD33-|CD3-|CD16+|CD56+	CD14-CD33-/CD3-/CD16+CD56+	CD14-|CD33-|CD3-|CD16+|CD56+	NK cells
CD14-|CD33-|CD3-|CD16+|CD56+|CD161-	CD14-CD33-/CD3-/CD16+CD56+/CD161-		CD161- NK cells
CD14-|CD33-|CD3-|CD16+|CD56+|CD161+	CD14-CD33-/CD3-/CD16+CD56+/CD161+		CD161+ NK cells
CD14-|CD33-|CD3-|CD16+|CD56+|CD57-	CD14-CD33-/CD3-/CD16+CD56+/CD57-		CD57- NK cells
CD14-|CD33-|CD3-|CD16+|CD56+|CD57+	CD14-CD33-/CD3-/CD16+CD56+/CD57+		CD57+ NK cells
CD14-|CD33-|CD3-|CD16+|CD56+|CD94-	CD14-CD33-/CD3-/CD16+CD56+/CD94-		CD94- NK cells
CD14-|CD33-|CD3-|CD16+|CD56+|CD94+	CD14-CD33-/CD3-/CD16+CD56+/CD94+		CD94+ NK cells
CD14-|CD33-|CD3-|CD16+|CD56+|HLADR-	CD14-CD33-/CD3-/CD16+CD56+/HLADR-		HLADR- NK cells
CD14-|CD33-|CD3-|CD16+|CD56+|HLADR+	CD14-CD33-/CD3-/CD16+CD56+/HLADR+		HLADR+ NK cells
CD14-|CD33-|CD3-|CD19+|CD20+	CD14-CD33-/CD3-/CD19+CD20+		B cells
CD14-|CD33-|CD3-|CD19+|CD20+|CD24-|CD38+	CD14-CD33-/CD3-/CD19+CD20+/CD24-CD38+	CD14-|CD33-|CD3-|CD19+|CD20+|CD24-|CD38+	naive B cells
CD14-|CD33-|CD3-|CD19+|CD20+|CD24+|CD38-	CD14-CD33-/CD3-/CD19+CD20+/CD24+CD38-	CD14-|CD33-|CD3-|CD19+|CD20+|CD24+|CD38-	memory B cells
CD14-|CD33-|CD3-|CD19+|CD20+|CD24+|CD38+	CD14-CD33-/CD3-/CD19+CD20+/CD24+CD38+	CD14-|CD33-|CD3-|CD19+|CD20+|CD24+|CD38+	transitional B cells
CD14-|CD33-|CD3-|CD19+|CD20+|IgD-|CD27-	CD14-CD33-/CD3-/CD19+CD20+/IgD-CD27-		IgD-CD27- B cells
CD14-|CD33-|CD3-|CD19+|CD20+|IgD-|CD27+	CD14-CD33-/CD3-/CD19+CD20+/IgD-CD27+		IgD-CD27+ B cells
CD14-|CD33-|CD3-|CD19+|CD20+|IgD+|CD27-	CD14-CD33-/CD3-/CD19+CD20+/IgD+CD27-		IgD+CD27- B cells
CD14-|CD33-|CD3-|CD19+|CD20+|IgD+|CD27+	CD14-CD33-/CD3-/CD19+CD20+/IgD+CD27+		IgD+CD27+ B cells
CD14-|CD33-|CD3-|CD20-	CD14-CD33-/CD3-/CD20-		CD20- CD3- lymphocytes
CD14-|CD33-|CD3-|CD20-|CD27+|CD38+	CD14-CD33-/CD3-/CD20-/CD27+CD38+	CD14-|CD33-|CD3-|CD20-|CD27+|CD38+	plasmablasts
CD14-|CD33-|CD3-|CD56+|CD16-	CD14-CD33-/CD3-/CD56+CD16-		CD16-CD56bright NK cells
CD14-|CD33-|CD3+	CD14-CD33-/CD3+		T cells
CD14-|CD33-|CD3+|CD4-|CD8+|CD57-	CD14-CD33-/CD3+/CD4-CD8+/CD57-		CD57-CD8+ T cells
CD14-|CD33-|CD3+|CD4-|CD8+|CD57+	CD14-CD33-/CD3+/CD4-CD8+/CD57+		CD57+CD8+ T cells
CD14-|CD33-|CD3+|CD4-|CD8+|ICOS-	CD14-CD33-/CD3+/CD4-CD8+/ICOS-		ICOS-CD8+ T cells
CD14-|CD33-|CD3+|CD4-|CD8+|ICOS+	CD14-CD33-/CD3+/CD4-CD8+/ICOS+		ICOS+CD8+ T cell
CD14-|CD33-|CD3+|CD4-|CD8+|PD1-	CD14-CD33-/CD3+/CD4-CD8+/PD1-		PD1-CD8+ T cells
CD14-|CD33-|CD3+|CD4-|CD8+|PD1+	CD14-CD33-/CD3+/CD4-CD8+/PD1+		PD1+CD8+ T cells
CD14-|CD33-|CD3+|CD4+	CD14-CD33-/CD3+/CD4+		CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CCR7-|CD45RA-	CD14-CD33-/CD3+/CD4+/CCR7-CD45RA-	CD14-|CD33-|CD3+|CD4+|CCR7-|CD45RA-	effector memory CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CCR7-|CD45RA+	CD14-CD33-/CD3+/CD4+/CCR7-CD45RA+	CD14-|CD33-|CD3+|CD4+|CCR7-|CD45RA+	effector CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CCR7+|CD45RA-	CD14-CD33-/CD3+/CD4+/CCR7+CD45RA-	CD14-|CD33-|CD3+|CD4+|CCR7+|CD45RA-	central memory CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CCR7+|CD45RA+	CD14-CD33-/CD3+/CD4+/CCR7+CD45RA+	CD14-|CD33-|CD3+|CD4+|CCR7+|CD45RA+	naive CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD161-	CD14-CD33-/CD3+/CD4+/CD161-		CD161-CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD161+	CD14-CD33-/CD3+/CD4+/CD161+		CD161+CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD25+|CD127-	CD14-CD33-/CD3+/CD4+/CD25+CD127-	CD14-|CD33-|CD3+|CD4+|CD25+|CD127-	Tregs
CD14-|CD33-|CD3+|CD4+|CD25+|CD127-|CD161-|CD45RA-	CD14-CD33-/CD3+/CD4+/CD25+CD127-/CD161-CD45RA-		CD161-CD45RA- Tregs
CD14-|CD33-|CD3+|CD4+|CD25+|CD127-|CD161-|CD45RA+	CD14-CD33-/CD3+/CD4+/CD25+CD127-/CD161-CD45RA+		CD161-CD45RA+ Tregs
CD14-|CD33-|CD3+|CD4+|CD25+|CD127-|CD161+|CD45RA-	CD14-CD33-/CD3+/CD4+/CD25+CD127-/CD161+CD45RA-		CD161+CD45RA- Tregs
CD14-|CD33-|CD3+|CD4+|CD25+|CD127-|CD161+|CD45RA+	CD14-CD33-/CD3+/CD4+/CD25+CD127-/CD161+CD45RA+		CD161+CD45RA+ Tregs
CD14-|CD33-|CD3+|CD4+|CD27-	CD14-CD33-/CD3+/CD4+/CD27-		CD4+CD27- T cells
CD14-|CD33-|CD3+|CD4+|CD27+	CD14-CD33-/CD3+/CD4+/CD27+		CD4+CD27+ T cells
CD14-|CD33-|CD3+|CD4+|CD28-	CD14-CD33-/CD3+/CD4+/CD28-		CD4+CD28- T cells
CD14-|CD33-|CD3+|CD4+|CD28+	CD14-CD33-/CD3+/CD4+/CD28+		CD4+CD28+ T cells
CD14-|CD33-|CD3+|CD4+|CD85j-	CD14-CD33-/CD3+/CD4+/CD85j-		CD85j-CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD85j+	CD14-CD33-/CD3+/CD4+/CD85j+		CD85j+CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD94-	CD14-CD33-/CD3+/CD4+/CD94-		CD94-CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD94+	CD14-CD33-/CD3+/CD4+/CD94+		CD94+CD4+ T cells
CD14-|CD33-|CD3+|CD4+|HLADR-|CD38-	CD14-CD33-/CD3+/CD4+/HLADR-CD38-		HLADR-CD38-CD4+ T cells
CD14-|CD33-|CD3+|CD4+|HLADR-|CD38+	CD14-CD33-/CD3+/CD4+/HLADR-CD38+		HLADR-CD38+CD4+ T cells
CD14-|CD33-|CD3+|CD4+|HLADR+|CD38-	CD14-CD33-/CD3+/CD4+/HLADR+CD38-		HLADR+CD38-CD4+ T cells
CD14-|CD33-|CD3+|CD4+|HLADR+|CD38+	CD14-CD33-/CD3+/CD4+/HLADR+CD38+		HLADR+CD38+CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD8-|CD57-	CD14-CD33-/CD3+/CD4+CD8-/CD57-		CD57-CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD8-|CD57+	CD14-CD33-/CD3+/CD4+CD8-/CD57+		CD57+CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD8-|ICOS-	CD14-CD33-/CD3+/CD4+CD8-/ICOS-		ICOS-CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD8-|ICOS+	CD14-CD33-/CD3+/CD4+CD8-/ICOS+		ICOS+CD4+ T cell
CD14-|CD33-|CD3+|CD4+|CD8-|PD1-	CD14-CD33-/CD3+/CD4+CD8-/PD1-		PD1-CD4+ T cells
CD14-|CD33-|CD3+|CD4+|CD8-|PD1+	CD14-CD33-/CD3+/CD4+CD8-/PD1+		PD1+CD4+ T cells
CD14-|CD33-|CD3+|CD56+	CD14-CD33-/CD3+/CD56+	CD14-|CD33-|CD3+|CD56+	NKT cells
CD14-|CD33-|CD3+|CD8+	CD14-CD33-/CD3+/CD8+		CD8+ T cells

CD14-|CD33-|CD3+|CD8+|CCR7-|CD45RA-	CD14-CD33-/CD3+/CD8+/CCR7-CD45RA-	CD14-|CD33-|CD3+|CD8+|CCR7-|CD45RA-	effector memory CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CCR7-|CD45RA+	CD14-CD33-/CD3+/CD8+/CCR7-CD45RA+	CD14-|CD33-|CD3+|CD8+|CCR7-|CD45RA+	effector CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CCR7+|CD45RA-	CD14-CD33-/CD3+/CD8+/CCR7+CD45RA-	CD14-|CD33-|CD3+|CD8+|CCR7+|CD45RA-	central memory CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CCR7+|CD45RA+	CD14-CD33-/CD3+/CD8+/CCR7+CD45RA+	CD14-|CD33-|CD3+|CD8+|CCR7+|CD45RA+	naive CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CD161-	CD14-CD33-/CD3+/CD8+/CD161-		CD161-CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CD161+	CD14-CD33-/CD3+/CD8+/CD161+		CD161+CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CD27-	CD14-CD33-/CD3+/CD8+/CD27-		CD27-CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CD27+	CD14-CD33-/CD3+/CD8+/CD27+		CD27+CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CD28-	CD14-CD33-/CD3+/CD8+/CD28-		CD28-CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CD28+	CD14-CD33-/CD3+/CD8+/CD28+		CD28+CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CD85j-	CD14-CD33-/CD3+/CD8+/CD85j-		CD85j-CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CD85j+	CD14-CD33-/CD3+/CD8+/CD85j+		CD85j+CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CD94-	CD14-CD33-/CD3+/CD8+/CD94-		CD94-CD8+ T cells
CD14-|CD33-|CD3+|CD8+|CD94+	CD14-CD33-/CD3+/CD8+/CD94+		CD94+CD8+ T cells
CD14-|CD33-|CD3+|CD8+|HLADR-|CD38-	CD14-CD33-/CD3+/CD8+/HLADR-CD38-		HLADR-CD38-CD8+ T cells
CD14-|CD33-|CD3+|CD8+|HLADR-|CD38+	CD14-CD33-/CD3+/CD8+/HLADR-CD38+		HLADR-CD38+CD8+ T cells
CD14-|CD33-|CD3+|CD8+|HLADR+|CD38-	CD14-CD33-/CD3+/CD8+/HLADR+CD38-		HLADR+CD38-CD8+ T cells
CD14-|CD33-|CD3+|CD8+|HLADR+|CD38+	CD14-CD33-/CD3+/CD8+/HLADR+CD38+		HLADR+CD38+CD8+ T cells
CD14-|CD33-|CD3+|CD56+|CD161-	CD14-CD33-/CD3+CD56+/CD161-		CD161- NKT cells
CD14-|CD33-|CD3+|CD56+|CD161+	CD14-CD33-/CD3+CD56+/CD161+		CD161+ NKT cells
CD14-|CD33-|TCRgd+	CD14-CD33-/TCRgd+	CD14-|CD33-|TCRgd+	gamma-delta T cells
CD14+|CD33+	CD14+CD33+	CD14+|CD33+	monocytes
CD14+|CD33+|CD14-|CD16+	CD14+CD33+/CD14-CD16+		CD16+CD14- monocytes
CD14+|CD33+|CD14+|CD16+	CD14+CD33+/CD14+CD16+		CD16+CD14+ monocytes
CD14+|CD33+|CD16-	CD14+CD33+/CD16-		CD16- monocytes
CD14+|CD33+|CD16+	CD14+CD33+/CD16+		CD16+ monocytes

Table S2: A list of cell definitions used to identify the 88 cell populations in cytometry data (SDY478) from ImmPort. Related to Figure 2. The cell definitions are created based on the author's gating strategy provided in SDY478.

Study	Panels	Subjects	Total Samples	CyTOF Samples	FCM Samples	Original Purpose	Institue
SDY112	3	91	188	95	93	Influenza vaccine study	Stanford
SDY167	1	44	131	0	131	Influenza vaccine study	National Institutes of Health
SDY180	30	36	510	0	510	Influenza vaccine study	Baylor Reasearch Institute
SDY311	6	75	134	79	55	Influenza vaccine study	Stanford
SDY312	11	82	616	0	616	Influenza vaccine study	Stanford
SDY314	5	87	445	0	445	Influenza vaccine study	Stanford
SDY315	5	73	125	74	51	Influenza vaccine study	Stanford
SDY420	11	278	511	284	227	Immunobiology of Aging	Stanford
SDY478	1	70	73	73	0	Influenza vaccine study	Stanford
SDY736	13	148	193	0	193	Immunobiology of aging and CMV	University of Arizona

Table S3: A summary of 10 studies included in the meta-analysis. Related to Figure 3.

Name	Cell Definitions
B cells	CD19+ CD20+
memory B cells	CD3- CD19+ CD20+ CD24+ CD38-
transitional B cells	CD3- CD19+ CD20+ CD24+ CD38+
plasmablasts	CD3- CD20- CD27+ CD38+
naive B cells	CD3- CD19+ CD20+ CD24- CD38+
T cells	CD3+
CD4+ T cells	CD3+ CD4+ CD8-
CD8+ T cells	CD3+ CD4- CD8+
naive CD4+ T cells	CD3+ CD4+ CCR7+ CD45RA+
naive CD8+ T cells	CD3+ CD8+ CCR7+ CD45RA+
central memory CD4+ T cells	CD3+ CD4+ CCR7+ CD45RA-
central memory CD8+ T cells	CD3+ CD8+ CCR7+ CD45RA-
effector CD4+ T cells	CD3+ CD4+ CCR7- CD45RA+
effector CD8+ T cells	CD3+ CD8+ CCR7- CD45RA+
effector memory CD4+ T cells	CD3+ CD4+ CCR7- CD45RA-
effector memory CD8+ T cells	CD3+ CD8+ CCR7- CD45RA-
gamma-delta T cells	TCRgd+
Tregs	CD3+ CD4+ CD25+ CD127-
monocytes	CD14+ CD33+
CD16- monocytes	CD14+ CD33+ CD16-
CD16+ monocytes	CD14+ CD33+ CD16+
NK cells	CD3- CD16+ CD56+
NKT cells	CD3+ CD56+

Table S4: A list of cell definitions used to identify the 23 cell populations in the meta-analysis. Related to Figure 4.

	Pearson Correlation									
	- SDY112	- SDY167	-SDY180	- SDY311	- SDY312	-SDY314	-SDY315	-SDY420	-SDY478	-SDY736
Age	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.98	1.00	1.00
Gender	1.00	1.00	0.99	0.99	0.99	1.00	1.00	0.92	1.00	1.00
Ethnicity	0.99	1.00	0.99	0.99	0.99	1.00	0.99	0.76	1.00	1.00

Table S5: A table reporting the Pearson correlation between the results from leave-one-out analysis and the results from the full meta-analysis . Related to Figure 4.

Cell Definitions	Parameter	Effect Size	95% CI Lower	95% CI Upper	Sample Size	p value	Adjusted p value
CD27+|CD3+|CD4+|CD8-	proportion	-0.451	-0.664	-0.239	736	3.380E-05	3.073E-03
CD127+|CD3+|CD4+|CD8-	proportion	-0.353	-0.537	-0.169	613	1.839E-04	4.940E-03
CD28-|CD3+|CD4-|CD8+	proportion	0.412	0.195	0.629	841	2.065E-04	4.940E-03
CD127-|CD3+|CD4-|CD45RA+	proportion	0.478	0.225	0.731	581	2.302E-04	4.940E-03
CD3+|CD4-|CD45RA+|CD8+|CD85J-	proportion	0.427	0.198	0.657	580	2.825E-04	4.940E-03
CD3+|CD4+|CD45RA+|CD8+	proportion	-0.249	-0.385	-0.113	871	3.407E-04	4.940E-03
CD19-|CD3+|CD33-|CD4-|CD8+	proportion	0.413	0.185	0.641	651	3.951E-04	4.940E-03
CD28-|CD3+|CD4-|CD45RA+|CD8+	proportion	0.429	0.190	0.669	623	4.598E-04	4.940E-03
CCR7+|CD3+|CD4-|CD45RA+	proportion	0.341	0.150	0.532	622	4.886E-04	4.940E-03
CD161-|CD3+|CD4+|CD8-	proportion	-0.432	-0.680	-0.183	581	6.960E-04	5.819E-03
CD27+|CD3+|CD4+|CD45RA+	proportion	-0.375	-0.592	-0.158	658	7.346E-04	5.819E-03
CD28+|CD3+|CD4+|CD8-	proportion	-0.304	-0.481	-0.127	841	7.674E-04	5.819E-03
CD27+|CD3+|CD4+|CD45RA+|CD8-	proportion	-0.372	-0.593	-0.151	658	1.011E-03	7.080E-03
CD3+|CD4-|CD45RA+|CD8+	proportion	0.286	0.107	0.465	871	1.798E-03	1.169E-02
CD19-|CD3+|CD33-|CD4+	proportion	-0.363	-0.592	-0.133	651	1.999E-03	1.213E-02
CD3+|CD4+|CD45RA+|CD8-	proportion	-0.256	-0.421	-0.091	871	2.443E-03	1.389E-02
CD27-|CD3+|CD4-|CD8+	proportion	0.338	0.108	0.569	736	4.063E-03	2.175E-02
CD19-|CD20-|CD27+|CD3+|IGD-	proportion	-0.326	-0.558	-0.095	581	5.756E-03	2.474E-02
CD27+|CD28+|CD3+|CD4+|CD85J+	proportion	-0.354	-0.606	-0.103	580	5.758E-03	2.474E-02
CCR7+|CD28+|CD3+|CD4+	proportion	-0.268	-0.458	-0.078	694	5.831E-03	2.474E-02
CD161-|CD3+|CD4+|CD45RA+|CD8-	proportion	-0.348	-0.595	-0.101	581	5.915E-03	2.474E-02
CD127+|CD3+|CD4+|CD45RA-|CD8-	proportion	-0.269	-0.461	-0.078	581	5.981E-03	2.474E-02
CD27-|CD3+|CD4-|CD45RA+	proportion	0.340	0.096	0.585	658	6.507E-03	2.575E-02
CD27-|CD3+|CD4-|CD45RA+|CD8+	proportion	0.330	0.087	0.573	658	7.877E-03	2.880E-02
CD28+|CD3+|CD4+|CD8-|CD85J-	proportion	-0.323	-0.563	-0.084	580	8.280E-03	2.880E-02
CD19-|CD27+|CD3+|IGD-	proportion	-0.303	-0.528	-0.077	581	8.679E-03	2.880E-02
CD27+|CD28+|CD3+|CD8-|CD85J-	proportion	-0.314	-0.549	-0.080	580	8.748E-03	2.880E-02
CD127+|CD3+|CD4+|CD45RA-	proportion	-0.243	-0.425	-0.061	581	8.862E-03	2.880E-02
CD3+|CD38+|CD4+|HLADR+	proportion	-0.244	-0.429	-0.059	585	9.709E-03	3.047E-02
CD27+|CD3+|CD4+|CD45RA-	proportion	-0.286	-0.510	-0.063	658	1.220E-02	3.700E-02
CD27+|CD3+|CD4+|CD45RA-|CD8-	proportion	-0.278	-0.500	-0.056	658	1.418E-02	4.164E-02
CD19+|CD20+|CD38+|IGD+	proportion	0.216	0.042	0.390	645	1.531E-02	4.354E-02
CD19+|CD20+|CD27-|CD3-	proportion	0.287	0.052	0.522	581	1.673E-02	4.613E-02

Table S6: A table summarizing the effect size of ethnicity to the proportion of cell subsets identified by the unsupervised analysis pipeline when comparing cytometry data of blood from Asian and White subjects. The effect sizes are estimated from the meta-analysis of 10 human immunology studies. Only significant findings (Adjusted p value <0.05) are listed. Related to Figure 6.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Evaluating the performance of bisection algorithm

We compared Silhouette scanning with eight other methods for bisecting the distribution of each marker.

K-means method: based on the values of a single marker, cells were clustered into 2 groups using k means clustering algorithm where k = 2. The cutoff value for bisection was the border between the 2 groups.

Hierarchical clustering method: based on the values of a single marker, cells were grouped into a Hierarchical tree. The tree was then cut into 2 groups at the top level. The cutoff value for bisection was the border between the 2 groups.

First valley method: The distribution of each marker was smoothed using the *smooth.spline* function. The peaks in the distribution were identified using the *.getPeaks* function in flowDensity package (Malek et al., 2015). The lowest points between peaks were defined as valleys. The valley with the smallest marker value was used as the cutoff for bisection.

Last valley method: The valley with the largest marker value was used as the cutoff for bisection.

Median valley method: The valley closest to the median of the marker value was used as the cutoff for bisection.

Mean method: The mean of the marker distribution was used as the cutoff.

Median method: the median of the marker value was used as the cutoff

Middle method: the mean of the max and min of the marker values were used as the cutoff.

After markers in SDY420 data were bisected, cells fulfilling the requirement of each cell definition (listed in **Table S1**) were identified. For example, for cell definition "CD3+ CD8+ CD4- ", cells falling into the CD3+ region were identified. Similarly, cells falling into CD8+ and CD4- regions were identified. The intersection of the 3 sets of cells was the cells corresponding to the cell definition "CD3+ CD8+ CD4- ". The proportion of cells corresponding to each cell definition was calculated and compared to the proportion provided by the author. The Spearman correlation was used as a measurement of the bisection algorithm.

Comparing guided analysis pipeline in MetaCyto with flowDensity and ACDC

Cell definitions were created based on the gating strategies provided by authors of SDY478. The cell definitions are available in the **Table S2**.

For MetaCyto, the proportion of each cell subset in blood was estimated by the guided analysis pipeline described above.

For flowDensity, the *flowDensity* function was used to identify the cell subsets corresponding to the cell definitions. The proportion of each cell subset in blood was calculated by dividing the number of cells in the subset by the total number of cells in the blood.

For ACDC, the cell definitions were first turned into a cell type-marker table with entries of 1, –1, and 0 representing a marker is present, absent, and ignored in a cell type respectively. The landmark points of each cell type were generated by clustering samples within each partition defined by the cell type-marker table. Landmark points were used to classify samples with the semi-supervised classification algorithm. Only the bottom level cell types (basophils, NK cells, naive B cells, memory B cells, transitional B cells, plasmablasts, effector memory CD4+ T cells, effector CD4+ T cells, central memory CD4+ T cells, naive CD4+ T cells, Tregs, NKT cells, effector memory CD8+ T cells, effector CD8+ T cells, central

memory CD8+ T cells, naive CD8+ T cells, gamma-delta T cells and monocytes) were estimated by the ACDC because it is designed to detect mutually exclusive cell types. Their parental populations, such as total T cells, were excluded.

All three methods received the same types of input: preprocessed CyTOF data and the cell definitions. No additional information or human input was provided to help identifying cell subsets. The Spearman correlation coefficient between author's result and MetaCyto or flowDensity or ACDC result was calculated to measure the performance.

Comparing unsupervised analysis pipeline in MetaCyto with FlowSOM and FlowMeans

FlowSOM (Van Gassen et al., 2015) and FlowMeans (Aghaeepour et al., 2011) used to cluster cytometry data using different K (number of clusters) values. The resulting clusters from FlowSOM and FlowMeans were labeled and merged using the same procedure described in the unsupervised pipeline.

The F measure was used to measure the performance of clustering methods with or without the merging step. The F measure was calculated as described in the FlowCAP study (Aghaeepour et al., 2013). Briefly, for each cell population in the manual gating result and each cell population in the auto-gating result, a 2 × 2 contingency table was calculated containing the false positive (FP), true positive (TP), false negative (FN) and true negative (TN). The recall (Re) was calculated as TP/(TP + FN), the precision (Pr) was calculated a TP/(TP + FP). The F measure was calculated as $F = (2 \times Pr \times Re)/(Pr + Re)$. For each population in manual gating result, the best F measure and its corresponding recall and precision were used as the F measure of the population. The overall F measure, Recall and Precision were the average of F measure, Recall and Precision of all manual gated populations, weighted by the size of each manual population.

Estimating the proportions of cell subsets using data from Carr study

The cell gating results from the Carr study, in the form of proportions of the parental gate, are available as Supplementary Data Set 1 of the Carr study publication (Carr et al., 2016). The proportion of a cell subset within blood is calculated by multiplying the proportions of parents in all gating steps together. For example, the proportion of CD4 T cell in blood is calculated by multiplying the proportion of total T cells in blood by the proportion of CD4 T cells in total T cells.

SUPPLEMENTAL REFERENCES

Aghaeepour, N., Nikolic, R., Hoos, H.H., and Brinkman, R.R. (2011). Rapid cell population identification in flow cytometry data. Cytom. Part A *79A*, 6–13.

Aghaeepour, N., Finak, G., Dougall, D., Khodabakhshi, A.H., Mah, P., Obermoser, G., Spidlen, J., Taylor, I., Wuensch, S.A., Bramson, J., et al. (2013). Critical assessment of automated flow cytometry data analysis techniques. Nat. Methods *10*, 228–238.

Carr, E.J., Dooley, J., Garcia-Perez, J.E., Lagou, V., Lee, J.C., Wouters, C., Meyts, I., Goris, A., Boeckxstaens, G., Linterman, M.A., et al. (2016). The cellular composition of the human immune system is shaped by age and cohabitation. Nat. Immunol. *17*, 461–468.

Van Gassen, S., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saeys, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytom. Part A *87*, 636–645.

Malek, M., Taghiyar, M.J., Chong, L., Finak, G., Gottardo, R., and Brinkman, R.R. (2015). flowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification. Bioinformatics *31*, 606–607.