Supplementary Table S1. Data source, data format, and normalization method of all datasets used

Data Source	Data Format[A]	Normalization Method	Cells types[B] (Number of Samples)
GSE35069	SOFT	Quantile	B cells (6), CD4+ T cells (6), CD8+ T cells (6), monocyte (6), whole blood (6), granulocyte (18)
GSE49667	SOFT	minfi::SWAN	CD8+ T cells (6). Treg cells (6)
GSE56879	BED	Max-min	Mouse E14 embryonic stem cells (32)
GSE66351	IDAT	minfi::ppNoob	Neuron (16), glia (16)
GSE83458	IDAT	minfi::ppNoob	Monocytes (3), dendritic cells (3)
GSE87196	BED	RnBeads	B cells (3), CD4+ T cells (3), CD8+ T cells (3), monocytes (3), granulocytes (3)
GSE88824	IDAT	minfi::ppNoob	B cells (8), CD4+ T cells (8), CD8+ T cells (8), monocytes (8), whole blood (14), granulocytes (8)
GSE103211	SOFT	Quantile	Dendritic cells (3)
GSE110554	IDAT	minfi::ppNoob	B cells (6), CD4+ T cells (7), CD8+ T cells (6), monocytes (6), granulocytes (6)
GSE112618	IDAT	minfi::ppNoob	Mixed immune cells (6); included B cells, CD4+ T cells, CD8+ T cells, monocytes, granulocytes
GSE121483	IDAT	minfi::ppNoob	Microglia (2), macrophage (4), infiltrating microglia (2), infiltrating monocytes (10), CMP (2), CMP (2), GMP (1)
GSE128654	IDAT	minfi::ppNoob	Glioma (54)
GSE151506[C]	BED	Quantile	Mixed immune cells (76)
Singh et al.[C]	BED	Quantile	Microglia (1)

[A]For input types, IDAT is raw Illumina array IDAT file, SOFT is data scraped from GEO SOFT files or series matrix files, BED is bedgraph-compatible file types

[B]Archetypal cell shown was used in place of samples identified by protein for B cellss (CD19+), monocytes (CD14+), natural killer (NK; CD56+), granulocytes (CD15+), Treg cells (CD25+), and endothelial cells (CD31+)

[C]Data obtained directly from authors

Abbreviations: hematopoietic stem cells (HSC; from bone and blood), multipotent progenitors (MPP; from bone and blood), common lymphoid progenitors (CMP), granulocyte-monocyte progenitors (GMP), common myeloid progenitors (CMP), common monocyte progenitors (cMOP), minfi::preprocessNood (minfi::ppNoob)