

# Precision, Recall, and F1 Score

All 24 Populations — LLM Gate vs. Spitzer Reference

Sorted by F1 score (descending). Computed from the confusion matrix comparing LLM gate labels against Spitzer reference labels.

Population	Precision	Recall	F1	LLM	Ref	TP
IgD <sup>+</sup> IgM <sup>+</sup> B cells	87.9%	85.2%	86.5%	8,496	8,768	7,470
CD8 T cells	65.2%	87.1%	74.6%	1,066	798	695
IgD <sup>−</sup> IgM <sup>+</sup> B cells	64.1%	83.5%	72.5%	4,568	3,508	2,928
CD4 T cells	59.9%	83.5%	69.8%	566	406	339
Basophils	38.0%	77.4%	51.0%	595	292	226
Eosinophils	43.3%	39.4%	41.3%	4,709	5,173	2,039
Classical Monocytes	75.4%	24.9%	37.5%	4,499	13,607	3,391
IgM <sup>−</sup> IgD <sup>−</sup> B cells	21.4%	80.0%	33.7%	4,203	1,123	898
$\gamma\delta$ T cells	14.7%	85.7%	25.1%	204	35	30
Plasma Cells	14.1%	50.0%	22.0%	418	118	59
Non-Classical Monocytes	13.6%	55.0%	21.9%	4,099	1,016	559
MPP	9.8%	36.0%	15.4%	723	197	71
Intermediate Monocytes	29.1%	10.4%	15.3%	4,296	12,045	1,249
pDCs	48.4%	5.6%	10.1%	345	2,974	167
NK cells	6.1%	27.5%	10.0%	846	189	52
MEP	9.9%	7.4%	8.5%	313	420	31
CMP	6.4%	12.2%	8.4%	375	197	24
GMP	12.1%	5.4%	7.5%	313	703	38
Pro-B (Frac A–C)	16.7%	4.1%	6.6%	66	266	11
mDCs	2.3%	15.5%	4.1%	770	116	18
NKT cells	3.1%	4.7%	3.7%	1,458	961	45
Macrophages	0.3%	2.5%	0.5%	1,980	199	5
HSC	0.2%	33.3%	0.3%	636	3	1
CLP	0.0%	0.0%	0.0%	79	59	0