

Table 1. Real-world multi-omics datasets used in benchmark.

Domain	Dataset	Samples	Features	Classes	Note
Microbiome	Faecal	156	150	3	Faecal microbiomes (in K-tuple/30-mer feature representation)
Complex disease	GC	757	53498	4	Gastric cancer genomic and epigenomic data
	BRCA	2065	50831	4	Breast cancer genomic and epigenomic data
Single cell	NCI-scRNA	2168	1207	3	Single-cell RNA-seq in gastric cancer line NCI-N87
	Lymphoid	12067	20	5	Single-cell RNA-seq of lymphoid cells