A

Dataset	Organism	Sample source	Data Type	Inter-batch heterogeneity	Sample-to-sample heterogeneity	AMI	Number of batches	Number of samples (excluding pools)	Number of pool samples	Biological factors	Protein	Dataset accession
Alzheimer	Human	CSF	Proteins	Medium: All batches overlap with at least one other batch	Very high: samples from patients with neurological disorders. Alzheimer's disease is a heterogeneous disorder and the controls have a wide variety of of neurological disorders with no cognitive impairment	0.13	22	408 (839 with duplicates)	84	Disorder Sex Age	889	https://github.com/spell00/BrainBatchEffectMSMS
AgingMice	Mouse	Liver tissue	Peptides precursors	High: Some overlap between some batches. Some batches are very far from the others	Medium: samples come from population of inbred mice originating from two parental strains	0.82	7	372	3 (discarded)	Strain Diet Age	17887	PRIDE PXD009160
Adenocarcinoma	Human	Plasma	Metabolites	Very high: No overlap between batches. All batches are very far appart	High: samples come from cancer patients	1	3	642	74	Disease	6461	https://github.com/dengkuistat/WaveICA

B

Strength of batch effect (raw)

Strength of batch effect (corrected)

Alzheimer — Aging Mice — Adenocarcinoma

No correction / Best MCC / Over Correction