

A

Dataset	Organism	Sample source	Data Type	Sample-to-sample heterogeneity	AMI	Number of batches	Number of samples (excluding pools)	Number of pool samples	Biological factors	# Features	Dataset accession
Alzheimer	Human	CSF	Proteins	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	21	408 (839 with duplicates)	84	Disorder Sex Age	889	<a href="https://github.com/spe1100/BrainBatchEffectMMS">https://github.com/spe1100/BrainBatchEffectMMS</a>
Aging Mice	Mouse	Liver tissue	Peptides precursors	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	High: samples come from cancer patients	1	3	642	74	Disease	6461	<a href="https://github.com/dengkuistat/WaveICA">https://github.com/dengkuistat/WaveICA</a>
Mixed tissues	Human & yeast	Human: ovary & prostate cancer tissue, HEK293T cell line. Yeast: BY4741 strain	Peptides	Very low: All samples are present in duplicates or triplicates. All samples are a mix of the same 3 tissue samples in different concentrations (except for the HEK293T control cell line, which is not mixed)	0.82	78	8 (1553 with replicates)	0	Proportions of mixed tissues	18114	PRIDE PXD015912
Benchmark	Bio-molecules and E. Coli	Not Applicable	Metabolites	Very low: high number of replicates. All samples from the same class are the same but at different concentrations	0.95	7	72 (1027 with replicates)	0	Biomolecules Organism	1235	<a href="https://www.research-collection.ethz.ch/handle/20.500.11850/545373">https://www.research-collection.ethz.ch/handle/20.500.11850/545373</a>

B

