

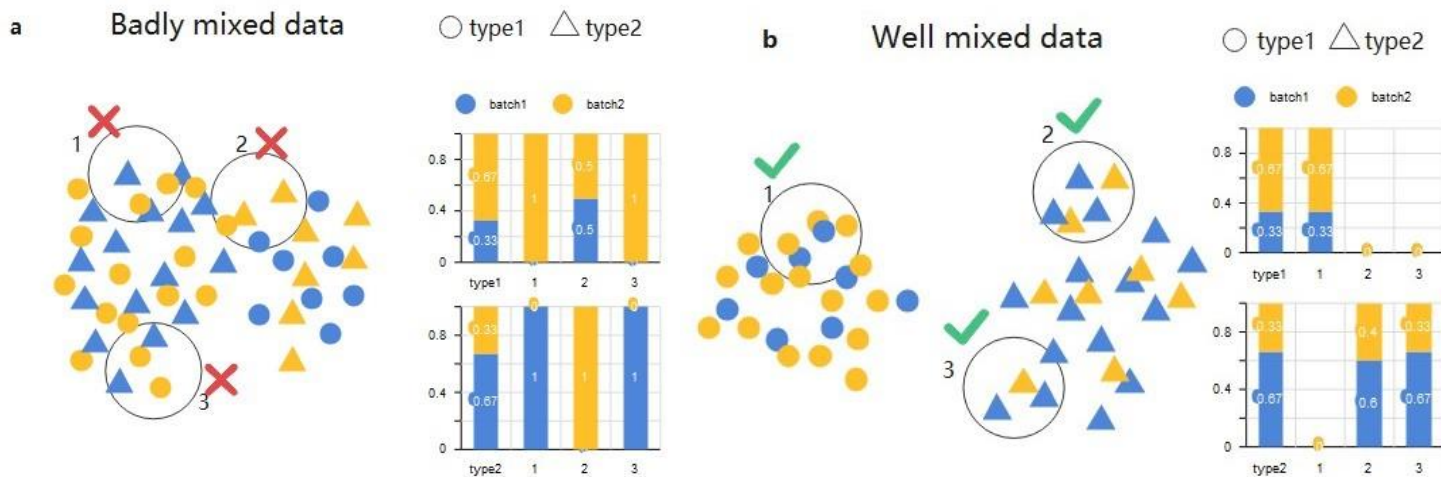
cKBET: assessing goodness of batch effect correction for single-cell RNA-seq

Yameng ZHAO, Yin GUO, Limin LI

Frontiers of Computer Science, DOI: [10.1007/s11704-022-2111-8](https://doi.org/10.1007/s11704-022-2111-8)

Problems & Ideas

- Problems of conventional stereo matching approaches:
 - Visualizing the data is subjective and inaccurate.
 - Quantification results may be biased when the proportions of cell types in different batches are quite different.
- Ideas: Use both batch and cell type information when quantifying batch effects.

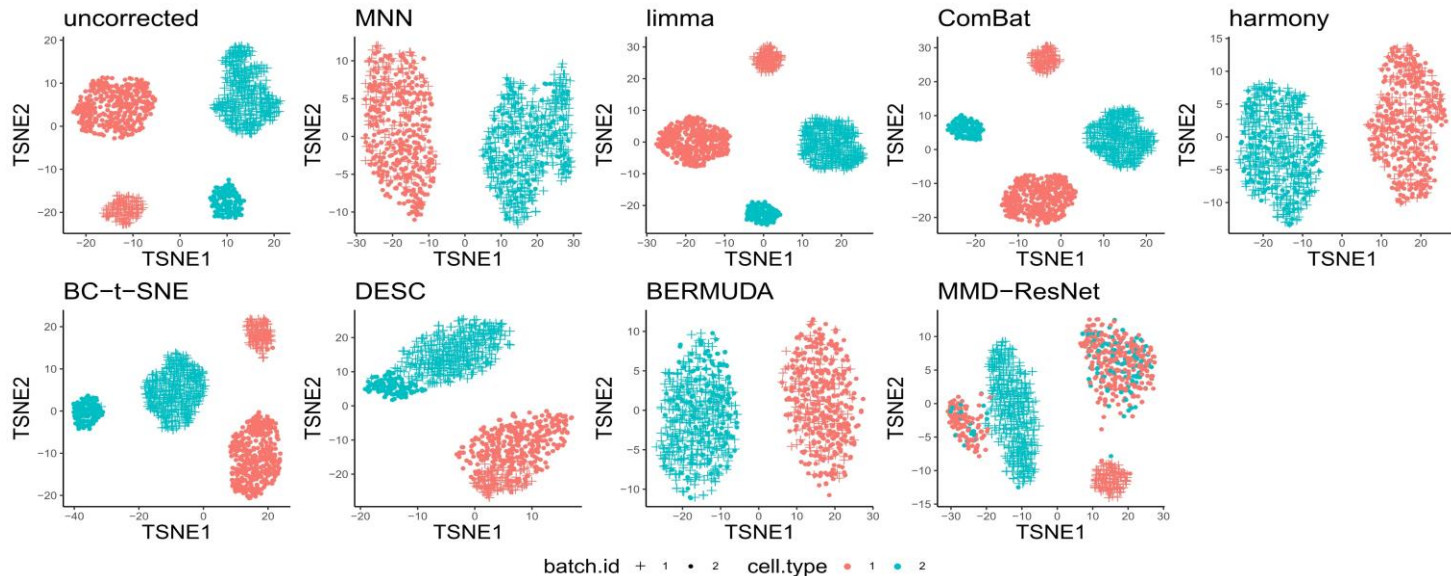


Left: The cells in different batches are well mixed, while the cells with different types are not separated, our cKBET method could correctly access the data set that has batch effect;

Right: The cells in different batches are well mixed and the cells with different types are separated, our cKBET method could correctly access the data set that has no batch effect.

Main Contributions

- Contributions:
 - A novel method to quantitatively assess whether batch effects exist in data set;
 - A method could identify biological differences in cell types when the proportions of cell types in different batches is quite different.



	Uncorrelated	MNN	limma	ComBat	harmony	BC-t-SNE	DESC	BERMUDA	MMD-ResNet
SIL	0.2424	0.2309	0.1167	0.1170	0.2259	0.3719	0.0011	0.1448	0.3681
PcR	0.9511	0.9468	0.0000	0.0000	0.9229	0.9527	0.0000	0.8901	0.8010
kBET	0.3232	0.1014	0.0212	0.0510	0.1000	0.2678	0.0328	0.2789	0.8544
cKBET	0.3174	0.1109	0.2813	0.2199	0.0620	0.2895	0.3017	0.1002	0.7157

The result of simulated data with different proportions of cell types in different batches. Above: the t-SNE visualization; Below: quantitative results of SIL, PcR, kBET and cKBET method, and the best quality dataset assessed by each method is bolded.