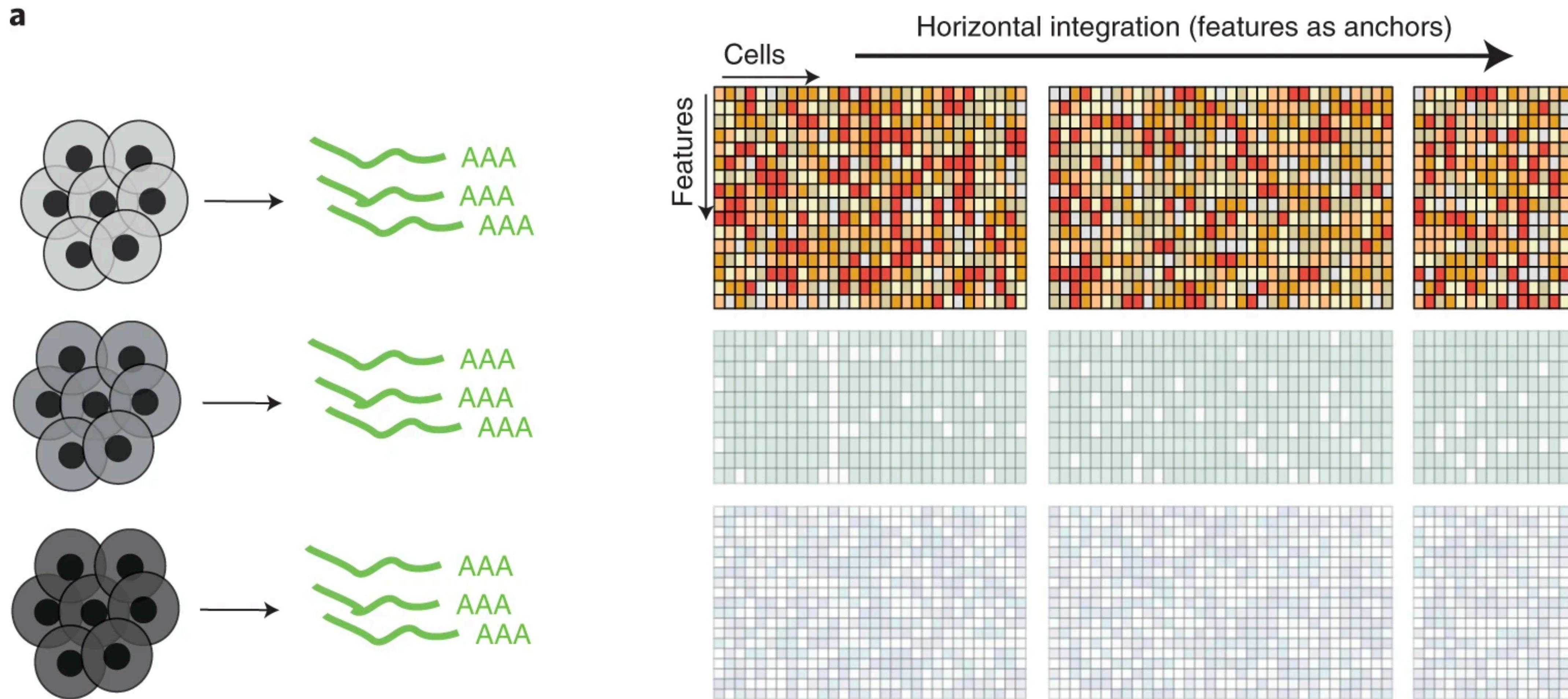
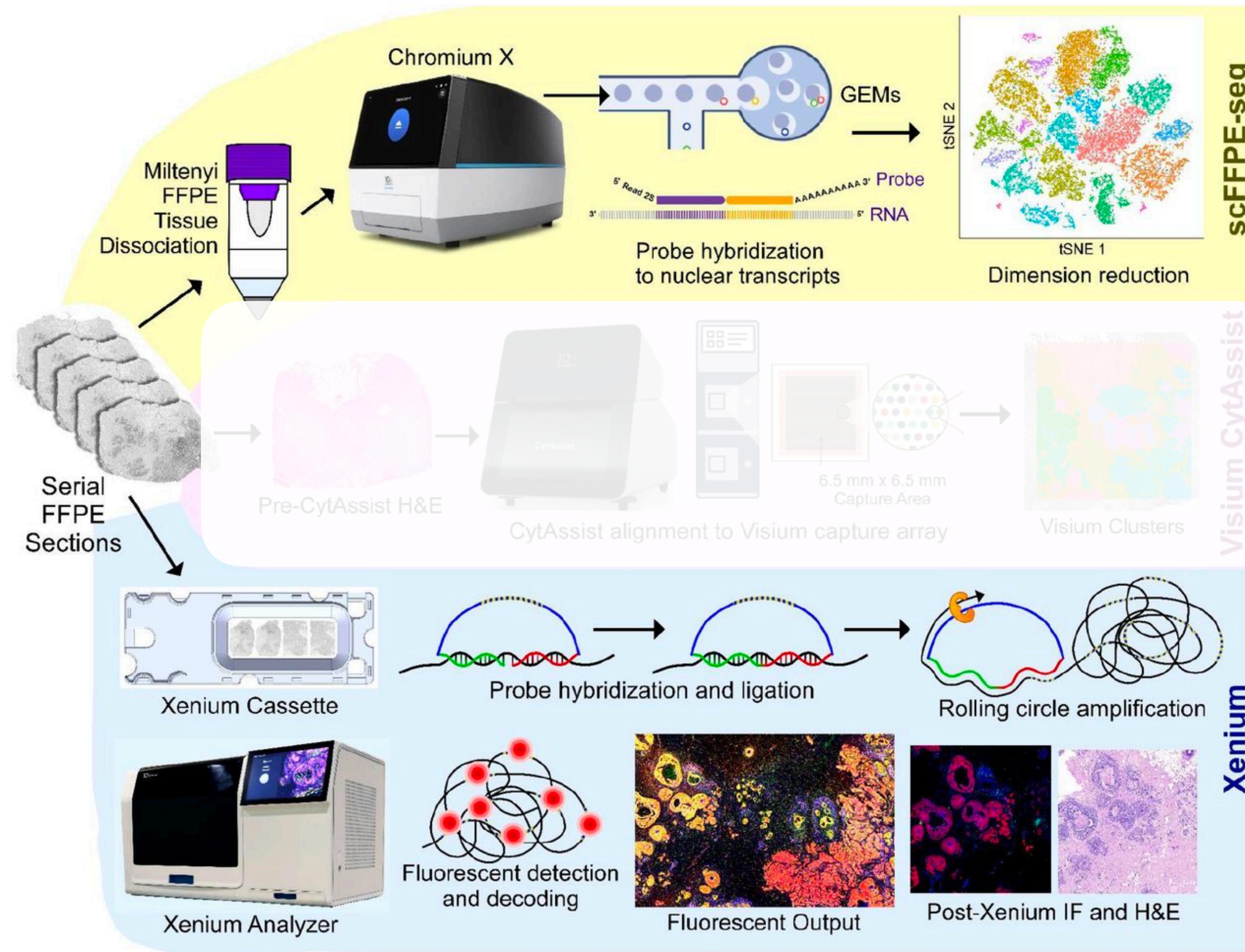


Part 2: Horizontal data integration

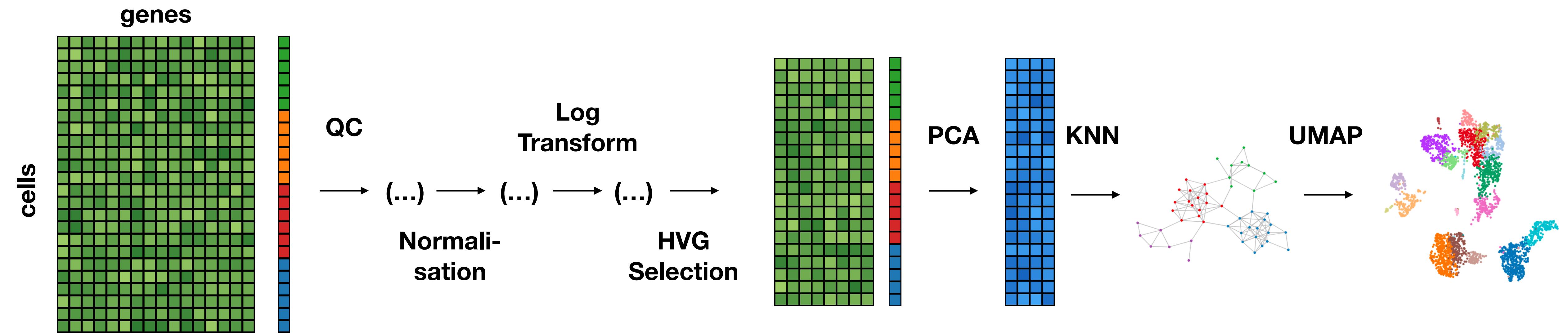


Part 2: Horizontal data integration

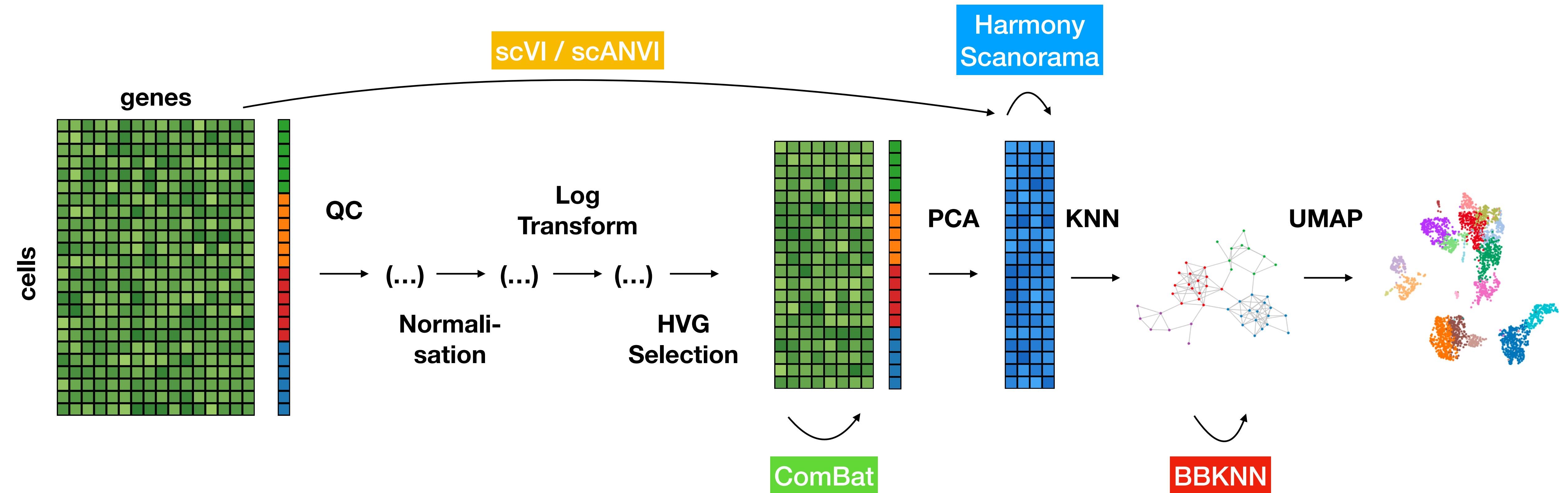


**Integrate single cells
from scFFPE-seq
with cells from Xenium**

The single analysis workflow



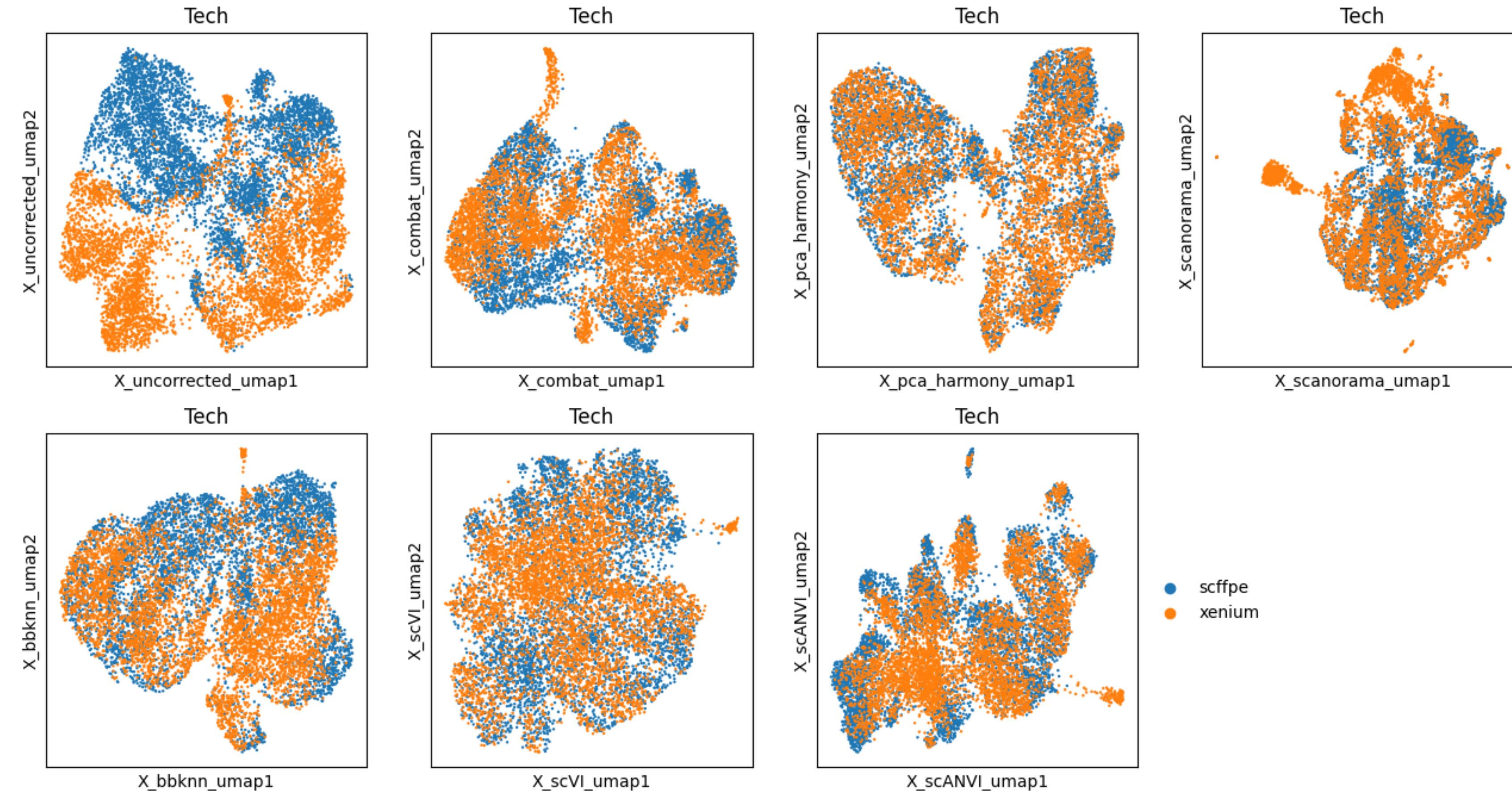
(Some) batch integration methods



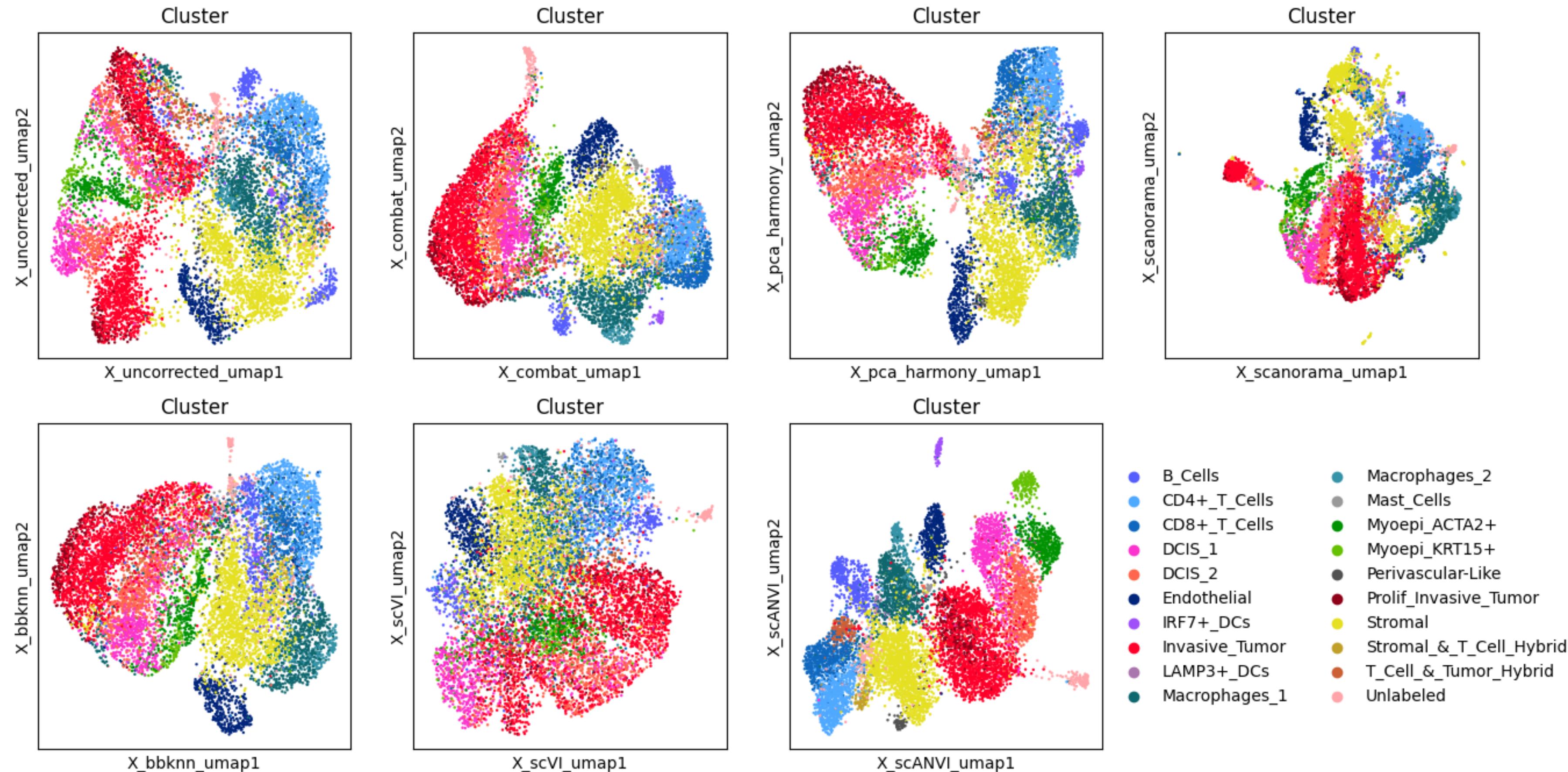
An overview of integration tools

	Combat	BBKNN	Harmony	Scanorama	scVI/scANVI
Input (+ batch covariate)	(Log1p transformed) expression matrix	KNN graph	PCA embeddings	PCA embeddings	Raw counts (+ cell type annotations for scANVI)
Output	Batch corrected expression matrix	Batch corrected KNN graph	Batch corrected PCA embeddings	Batch corrected PCA embeddings	Batch corrected cell embeddings
Methodology	Linear regression (+ Empirical bayes)	“approximate neighbour detection” (annoy package)	Soft clustering (“maximum diversity clustering”)	Mutual nearest neighbour matching	(Variational) autoencoder
Notes	Linear, Loc/scale model	Purely dependent on the graph representation			Non-linear, Complex

Results - Technology



Results - Cell types



Results - Integration metrics

Method	Bio conservation					Batch correction					Aggregate score		
	Isolated labels	KMeans NMI	KMeans ARI	Silhouette label	cLISI	Silhouette batch	iLISI	KBET	Graph connectivity comparison	PCR comparison	Batch correction	Bio conservation	Total
X_pca_harmony	0.38	0.52	0.34	0.50	0.98	0.92	0.54	0.45	0.78	0.96	0.73	0.55	0.62
X_scANVI	0.48	0.57	0.38	0.52	0.97	0.94	0.41	0.36	0.79	0.71	0.64	0.58	0.61
X_combat	0.39	0.51	0.35	0.50	0.99	0.88	0.17	0.28	0.81	1.00	0.63	0.55	0.58
X_scVI	0.47	0.37	0.21	0.50	0.94	0.94	0.39	0.31	0.58	0.69	0.59	0.50	0.53
X_scanorama	0.41	0.48	0.27	0.50	0.98	0.86	0.33	0.34	0.68	0.24	0.49	0.53	0.51
X_bbknn_umap	0.39	0.48	0.26	0.48	0.96	0.69	0.53	0.15	0.55	0.55	0.49	0.51	0.51
X_uncorrected	0.44	0.49	0.28	0.50	0.99	0.81	0.00	0.09	0.82	0.00	0.34	0.54	0.46