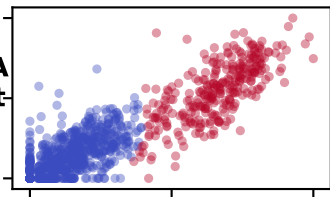
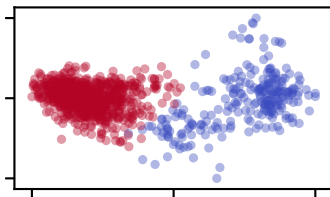


a) BRCA Dataset

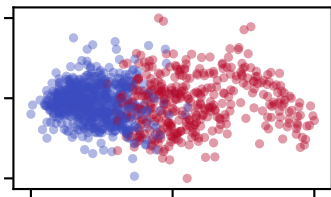
Estrogen Receptor  
ARI 0.32, Silhouette 0.65



Ground truth  
ARI 0.68, Silhouette 0.5

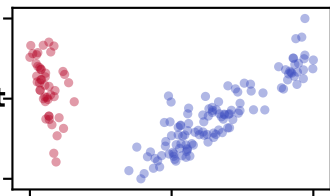


BCR Canonical Reason 2  
ARI 0.46, Silhouette 0.39

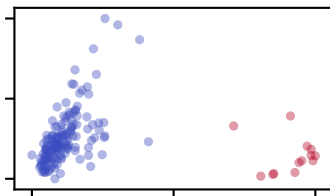


b) KIRP Dataset

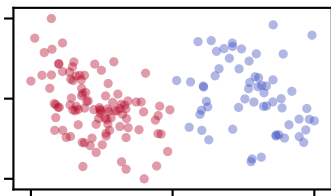
Gender  
ARI 0.97, Silhouette 0.63



BCR Canonical Status  
ARI 0.36, Silhouette 0.71



Ground truth  
ARI 0.17, Silhouette 0.37



c) ARI scores on bulk RNA-seq datasets

Dataset name	BRCA ARI	KIRP ARI
<b>Discover + GMM *</b>	0.68	0.17
<b>Discover + HDBSCAN</b>	0.57	0.0
Affinity Propagation	0.02	0.02
Kmeans*	0.39	0.0
GMM*	0.65	0.01
Spectral Clustering*	0.0	0.0
HDBSCAN	0.11	0.07
PCA50+ Kmeans*	0.35	0.0
PCA50 + GMM*	0.36	0.01
PCA50 + HDBSCAN	0.0	0.09
Xgboost+GMM*	<b>0.71</b>	<b>0.38</b>
Xgboost+HDBSCAN	0.3	0.13
MI+GMM*	0.7	0.3
MI+HDBSCAN	0.53	0.17