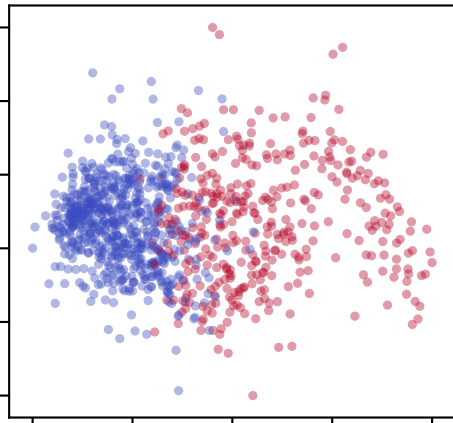
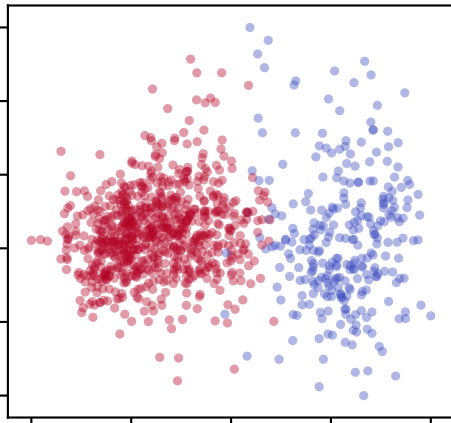
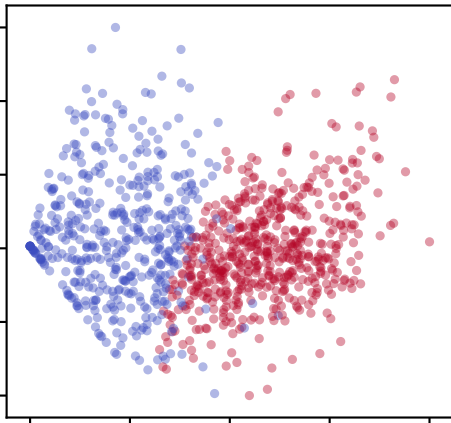
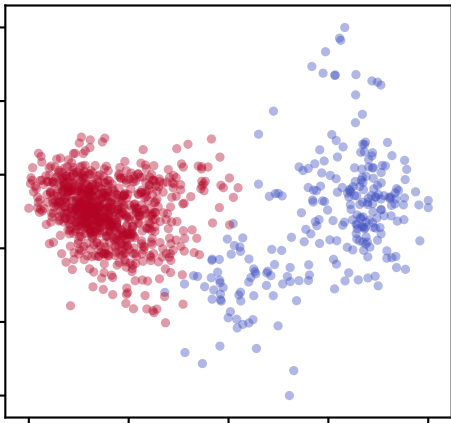
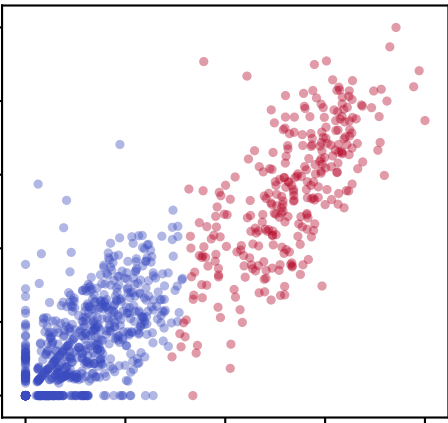


a)

BRCA
Dataset



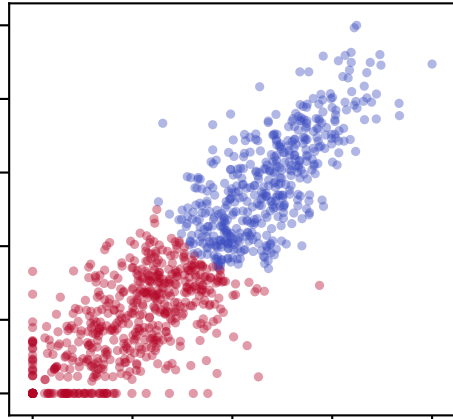
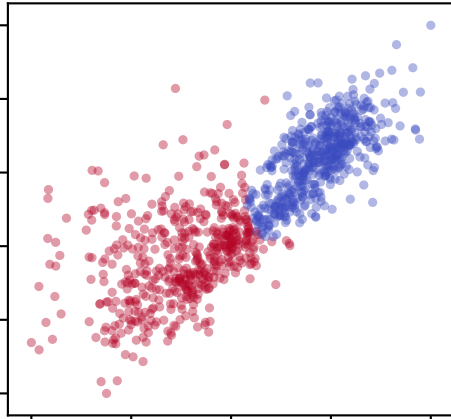
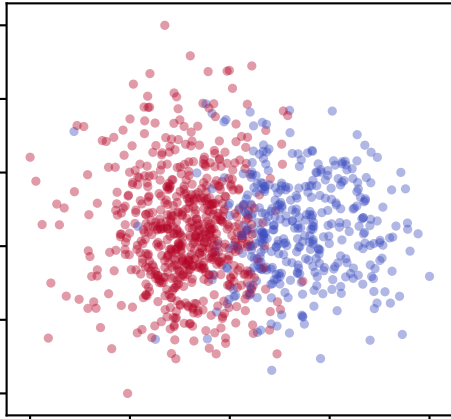
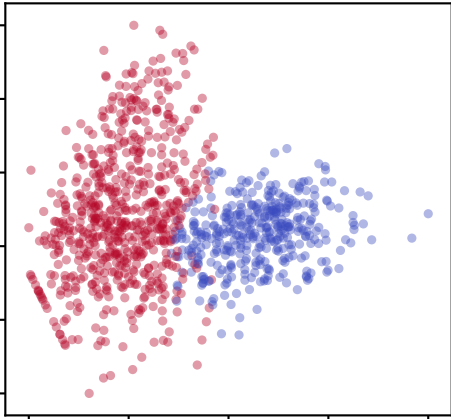
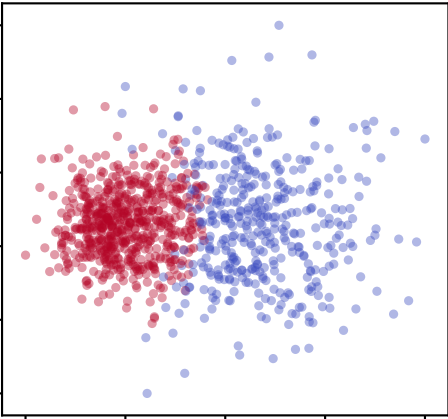
6) BCR Canonical Reason 2
ARI 0.4 (5 feaures)
Silhouette 0.46

7) Metastatic site 2
ARI 0.16 (3 feaures)
Silhouette 0.48

8) Canonical reason
ARI 0.27 (6 feaures)
Silhouette 0.41

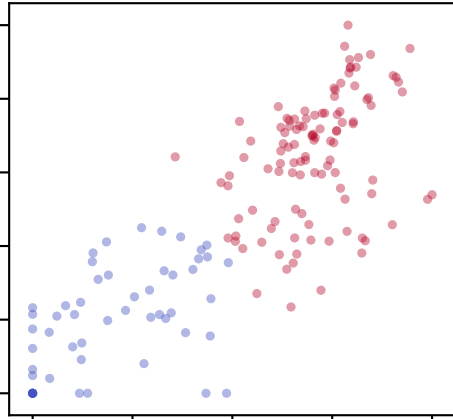
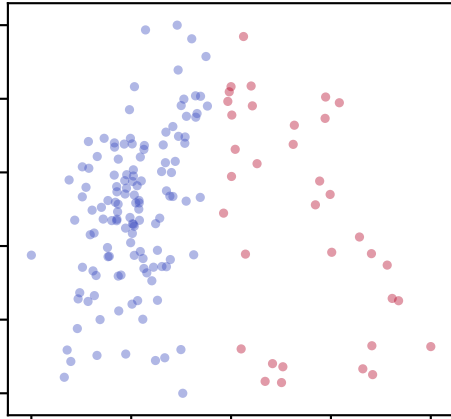
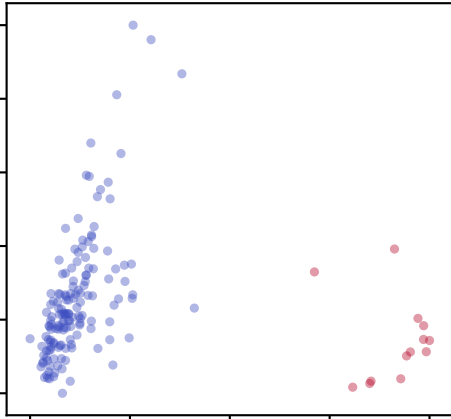
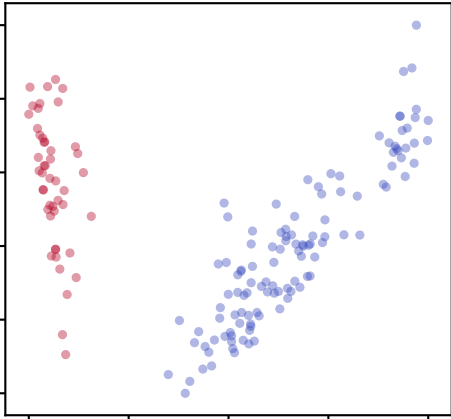
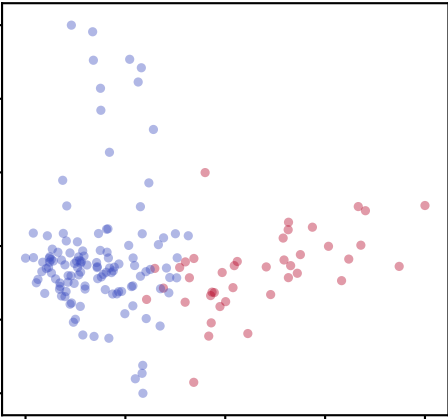
9) Metastatic site
ARI 0.36 (2 feaures)
Silhouette 0.56

10) Additional Locoregional Procedure
ARI 0.13 (2 feaures)
Silhouette 0.53



b)

KIRP
Dataset



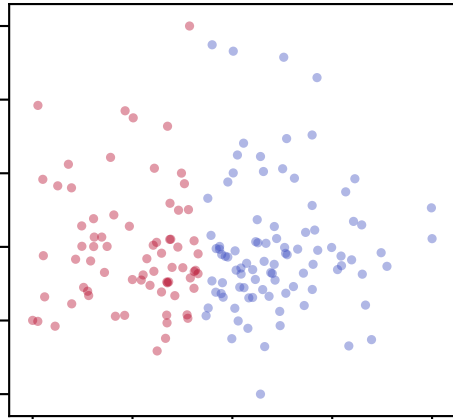
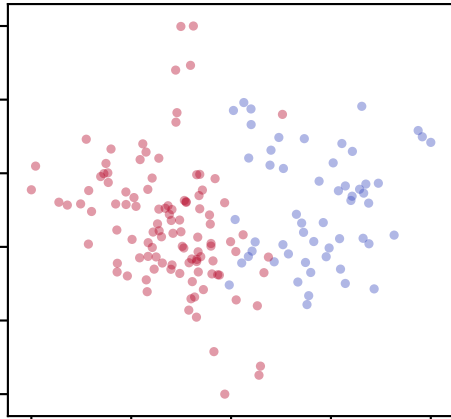
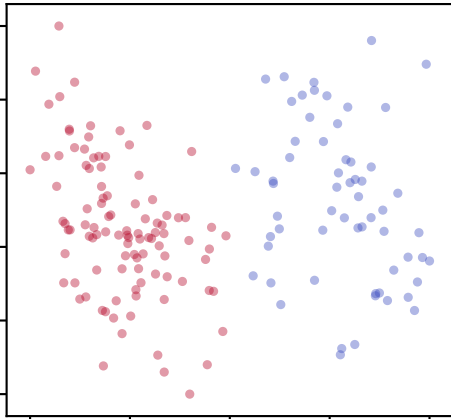
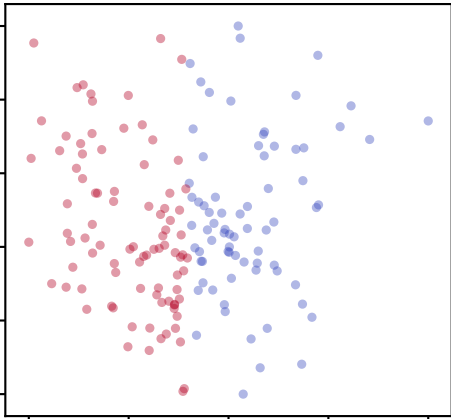
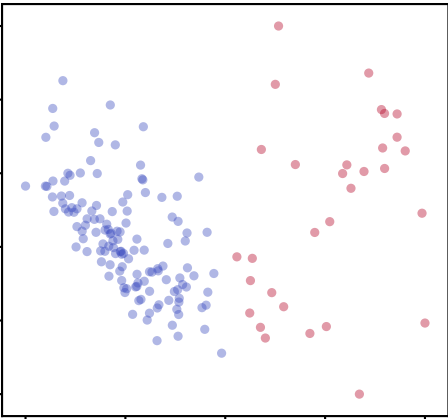
6) Clinical Stage
ARI 0.4 (157 feaures)
Silhouette 0.47

7) Lactate dehydrogenase
ARI 0.08 (107 feaures)
Silhouette 0.31

8) Ground truth
ARI 0.17 (152 feaures)
Silhouette 0.37

9) Lactate dehydrogenase
ARI 0.32 (4 feaures)
Silhouette 0.43

10) BCR Canonical Reason 2
ARI 0.67 (152 feaures)
Silhouette 0.31



c) ARI scores on bulk RNA-seq datasets

	BRCA ARI	KIRP ARI
Discover + GMM *	0.68	0.17
Discover + HDBSCAN	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*	0.71	0.38
Xgboost+HDBSCAN	0.3	0.13
MI+GMM*	0.7	0.3
MI+HDBSCAN	0.53	0.17