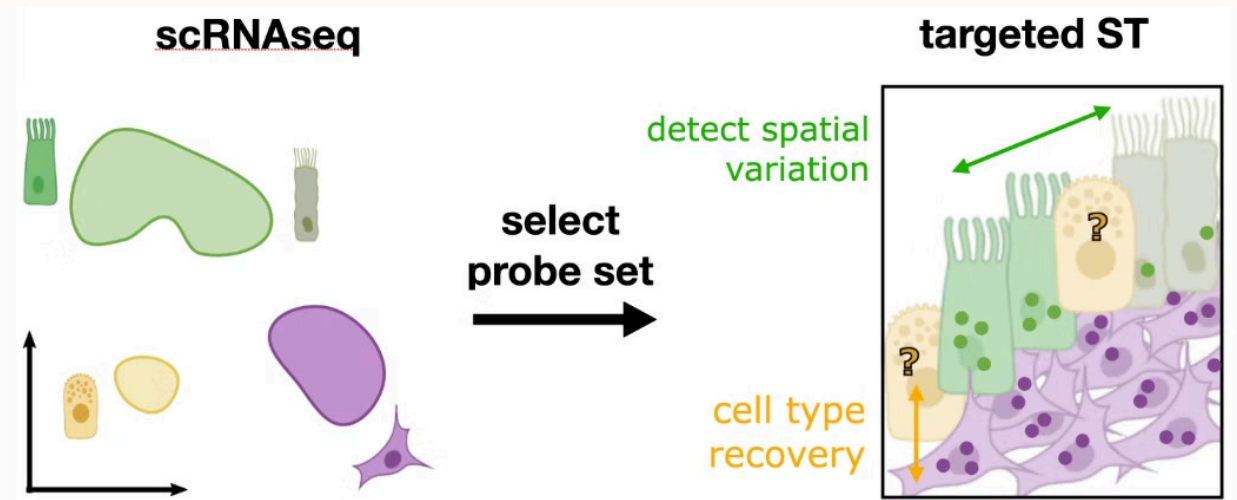
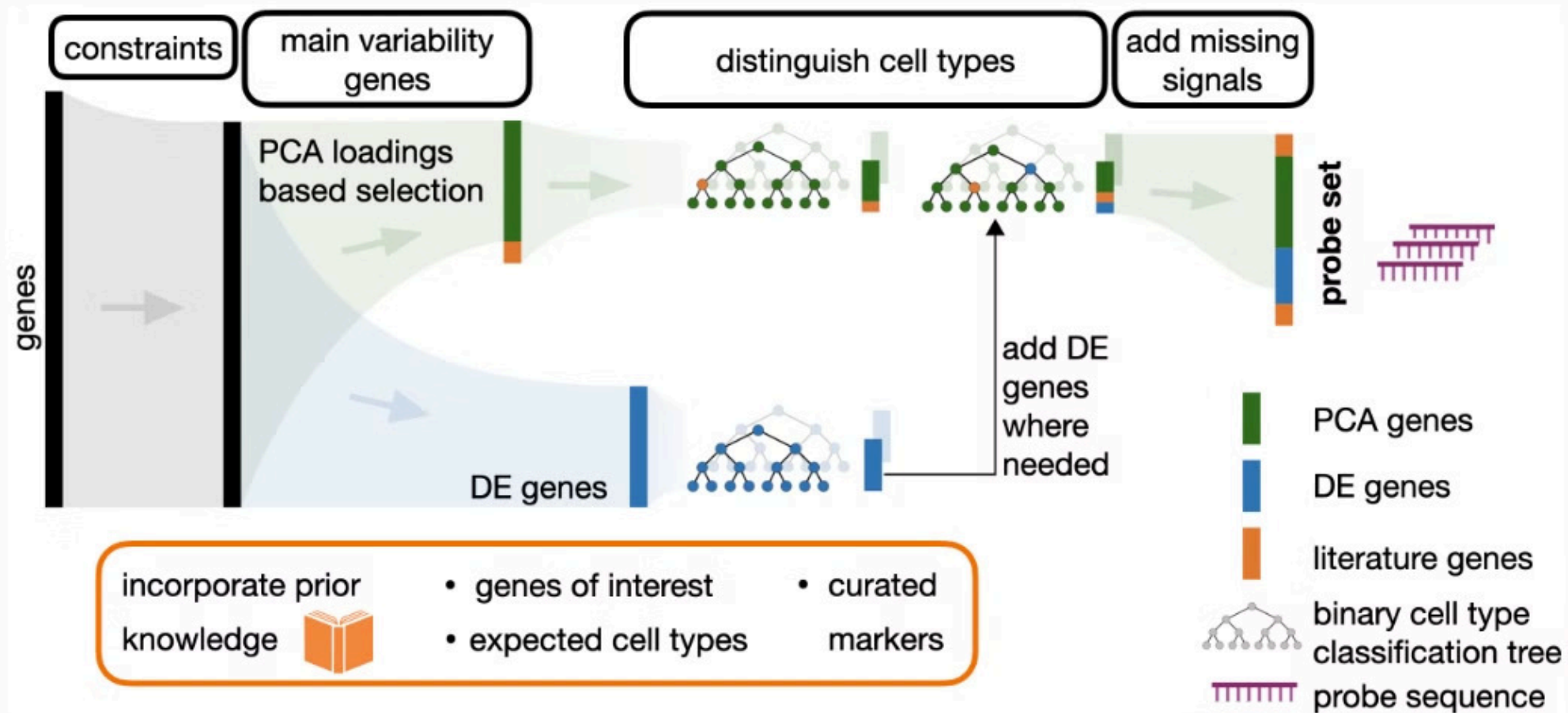


# Spapros: A Pipeline for Targeted Spatial Transcriptomics



Spapros is a probe set selection pipeline designed for targeted spatial transcriptomics

# Spapros Methodology



## Constraints

Spapros begins by defining constraints on the gene set, possibly based on prior knowledge or technical limitations. This ensures that the selected genes are suitable for the specific experiment and available resources.

## Main Variability Genes

The pipeline employs principal component analysis (PCA) to identify genes that contribute the most to data variance. This step ensures that the probe set captures the most important signals in the spatial transcriptomics data.

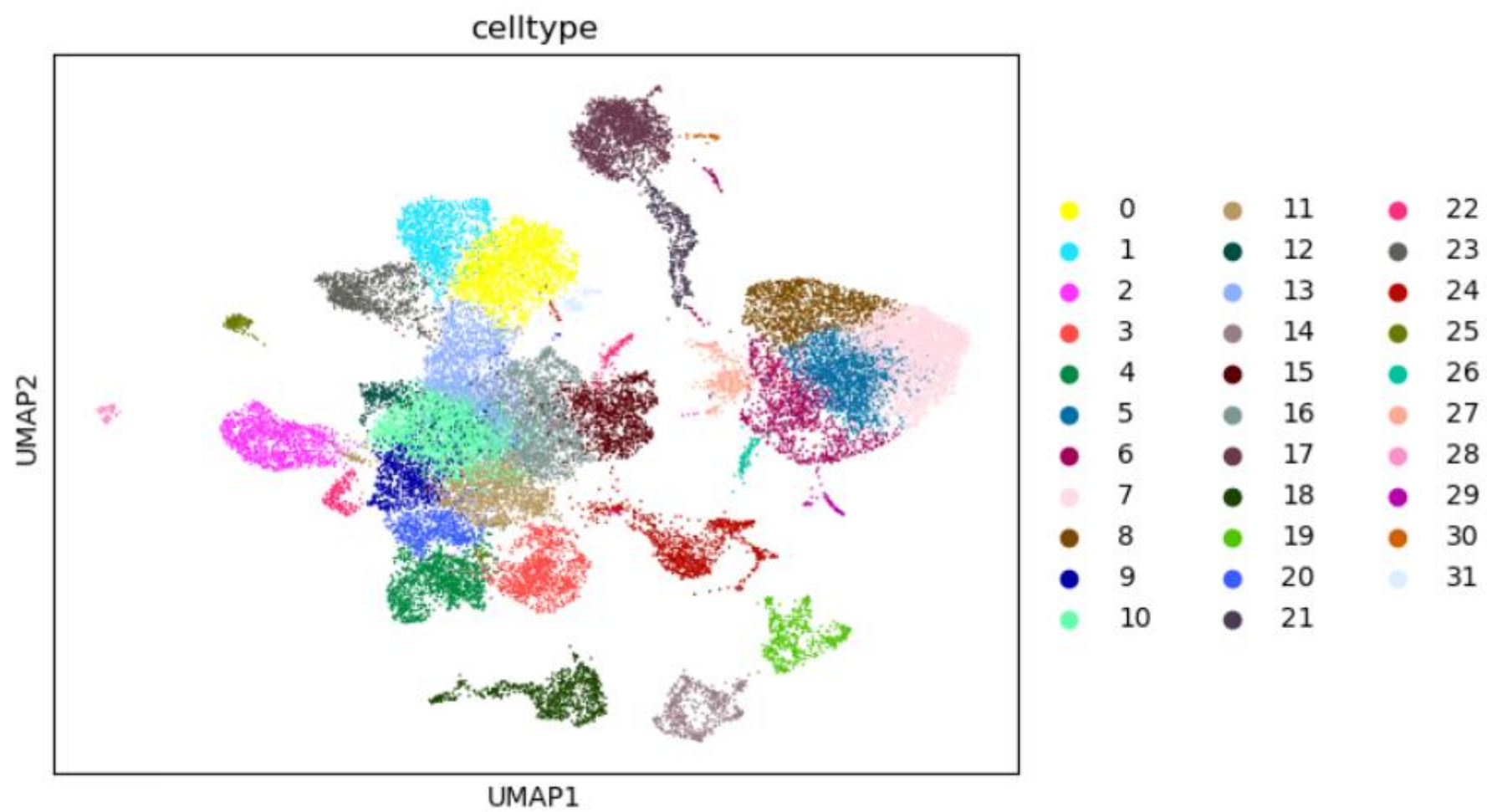
# Applications of Spapros

## **Optimal Probe Set Selection**

Spapros enables the selection of gene sets optimized for identifying specific cell types and capturing transcriptional variation beyond cell type annotations.

## **Probe Design**

The pipeline includes a probe design component that generates ready-to-order probes based on the selected gene sets, ensuring that the probes meet technical constraints and are suitable for the experimental setup.



Preprocess counts  
and get highly  
variable genes

Get info from  
processed dataset

Create first  
instance of  
probesetselector  
with n=None,  
n\_pca\_genes=0

Create second  
instance with  
n=None,  
n\_pca\_genes=0

Create third  
instance with  
n=None,  
n\_pca\_genes=0

Create fourth  
instance with  
marker genes and  
n=None,  
n\_pca\_genes=8  
000

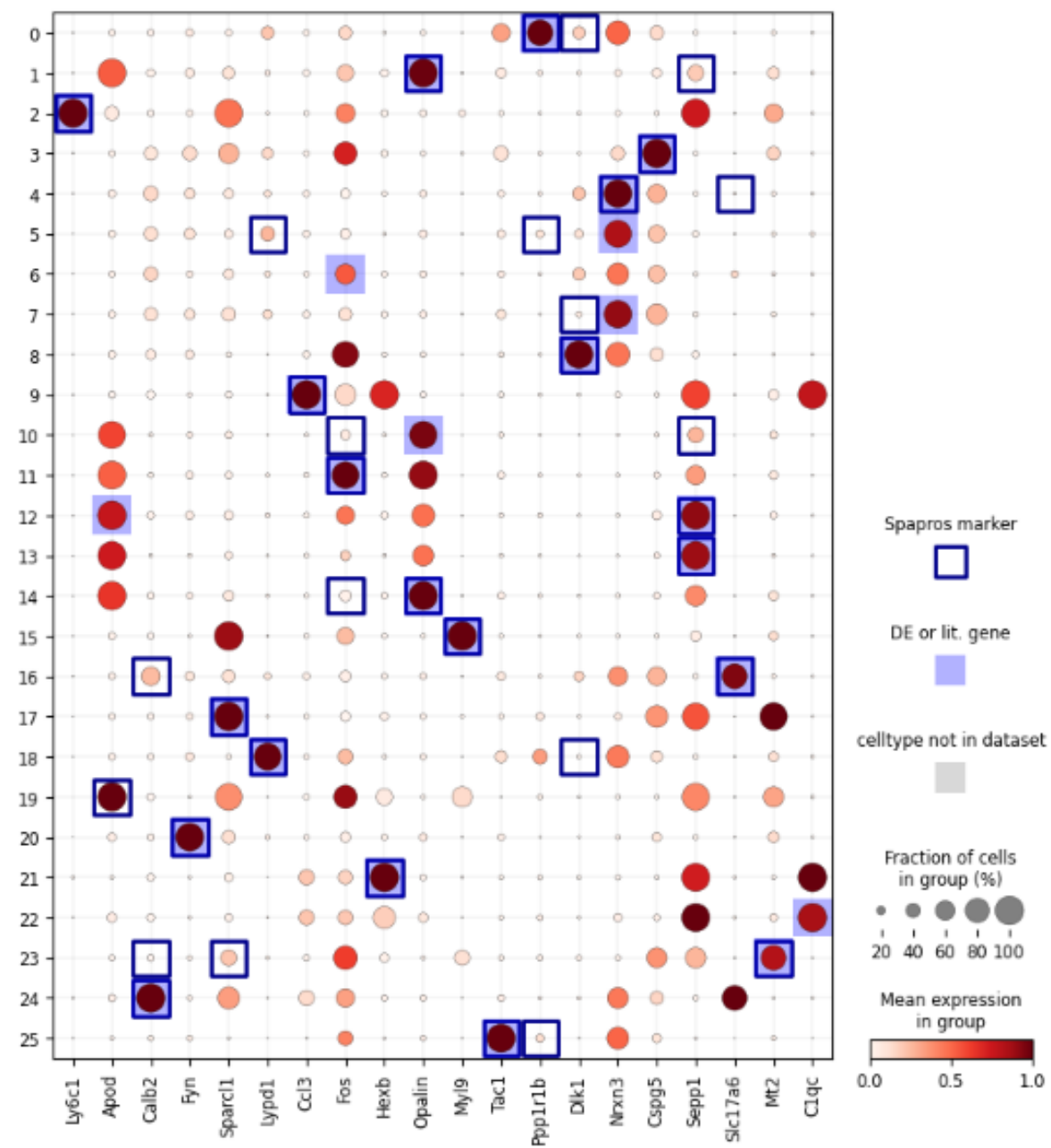
```
AnnData object with n_obs × n_vars = 29285 × 2303
obs: 'celltype'
var: 'gene_ids', 'highly_variable', 'means', 'dispersions', 'dispersions_norm'
uns: 'log1p', 'hvg'
obsm: 'X_umap'
```

128 unique genes

262 unique genes

392 unique genes

614 unique genes



	gene_nr	selection	rank	marker_rank	tree_rank	importance_score	pca_score	pre_selected	prior_selected	pca_selected	celltypes_DE_1vsall	celltypes_DE_sp
	<b>Ly6c1</b>	1	True	1.0	1.0	1.0	0.975035	0.742810	True	False	False	2
	<b>Sparcl1</b>	2	True	1.0	1.0	1.0	0.842713	1.204107	True	False	True	17
	<b>CcB</b>	3	True	1.0	1.0	1.0	0.802680	1.200838	True	False	True	9
	<b>Calb2</b>	4	True	1.0	1.0	1.0	0.780230	0.801069	True	False	False	24

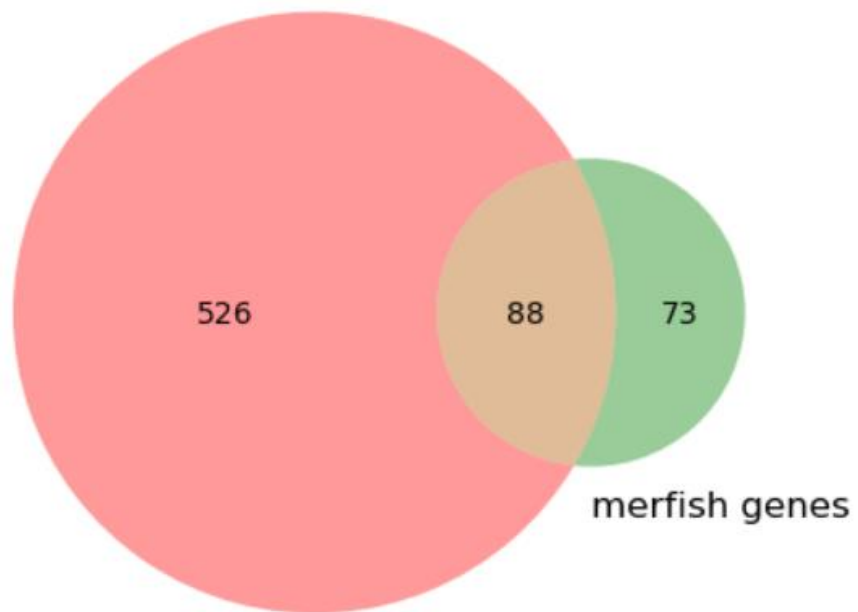
	gene_nr	selection	rank	marker_rank	tree_rank	importance_score	pca_score	pre_selected	prior_selected	pca_selected	celltypes_DE_1vsall	celltypes_DE_spe
	<b>Nrxn3</b>	19	True	1.0	1.0	1.0	0.381296	1.670207	True	False	True	4,7,5
	<b>Mal</b>	20	True	1.0	1.0	1.0	0.223368	0.989218	True	False	False	1,10,11
	<b>Mt2</b>	21	True	1.0	1.0	1.0	0.161281	0.737992	True	False	False	23
	<b>Itm2a</b>	22	True	1.0	2.0	1.0	0.492063	0.938224	True	False	False	2

	gene_nr	selection	rank	marker_rank	tree_rank	importance_score	pca_score	pre_selected	prior_selected	pca_selected	celltypes_DE_1vsall	celltypes_DE_speci
	<b>Ccl4</b>	49	True	2.0	NaN	1.0	0.052427	1.456800	False	False	True	
	<b>Anln</b>	50	True	2.0	NaN	1.0	0.044486	0.400287	False	False	False	
	<b>Junb</b>	51	True	2.0	NaN	1.0	0.032562	1.757122	False	False	True	
	<b>Xist</b>	52	True	2.0	NaN	1.0	0.024133	2.194393	False	False	True	

[34]: display(selector.probeset.iloc[266:275])

	<b>Selp1g</b>	267	True	NaN	NaN	NaN	NaN	0.366234	False	False	False	
	<b>Caln1</b>	268	True	NaN	NaN	NaN	NaN	0.366160	False	False	False	
	<b>Sox4</b>	269	False	NaN	NaN	NaN	NaN	0.364669	False	False	False	
	<b>Cxcl12</b>	270	False	NaN	NaN	NaN	NaN	0.362954	False	False	False	
	<b>Gdf15</b>	271	False	NaN	NaN	NaN	NaN	0.361086	False	False	False	
	<b>Tmem108</b>	272	False	NaN	NaN	NaN	NaN	0.361015	False	False	False	
	<b>Fxyd1</b>	273	False	NaN	NaN	NaN	NaN	0.360289	False	False	False	
	<b>Mmd2</b>	274	False	NaN	NaN	NaN	NaN	0.360022	False	False	False	
	<b>Pde10a</b>	275	False	NaN	NaN	NaN	NaN	0.359962	False	False	False	

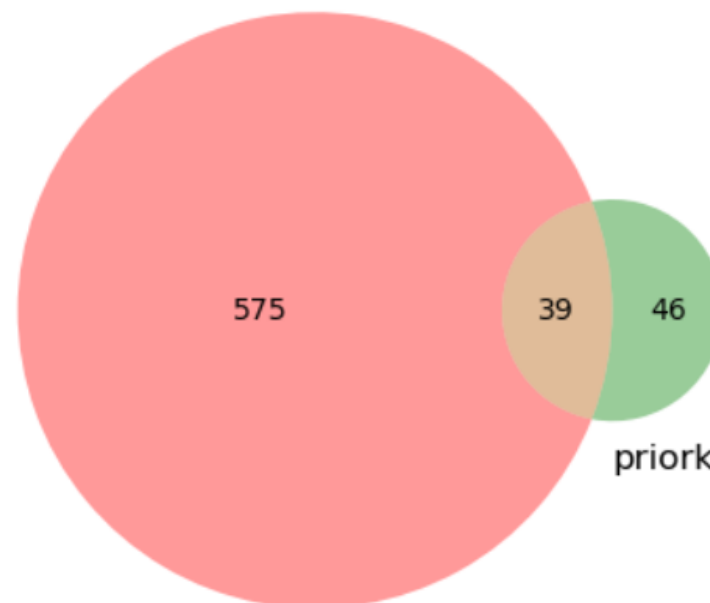
Venn Diagram of Gene Overlap  
(Overlap: 88 genes)



Selected spapros Genes

merfish genes

Venn Diagram of Gene Overlap  
(Overlap: 39 genes)



Selected spapros Genes

priorknowledge\_genes