① DATA PREPROCESSING AND HVGs

normal hsad files

8 files

using scanpy and PCA getting HVGs list

Foundation model (scPRINT / Transcriptformer) to get gene embedding

gene embedding averaged per dataset

Use genes only in the HVG list

new matrix

gene x embedding

	e1	e2	e3	...	e2048
gene1					
gene2					

similarly 8 files

finally averaged for single hsad file

gene x embedding

	em1	em2	...	em
gene1				
gene2				

8 normal files → DATA PRC PRO → gene x embeddg matrix → 4154 x 2048

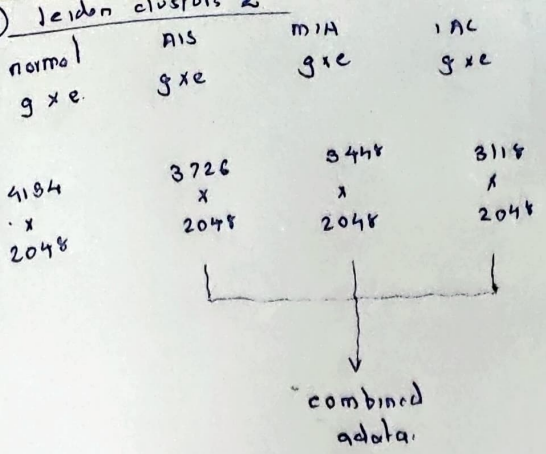
3 IAC files → DATA PRC PRO → gene x emb matrix → 3118 x 2048

3 mIA files → DATA PRC PRO → gene x emb matrix → 3446 x 2048

3 AIS files → DATA PRC PRO → gene x emb matrix → 3726 x 2048

② Leiden clusters set 1

③ leiden clusters 2



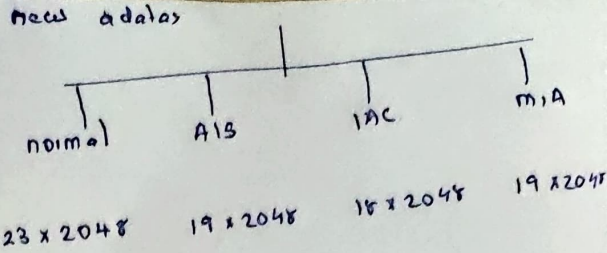
combined adata. obs

cluster 2 here.

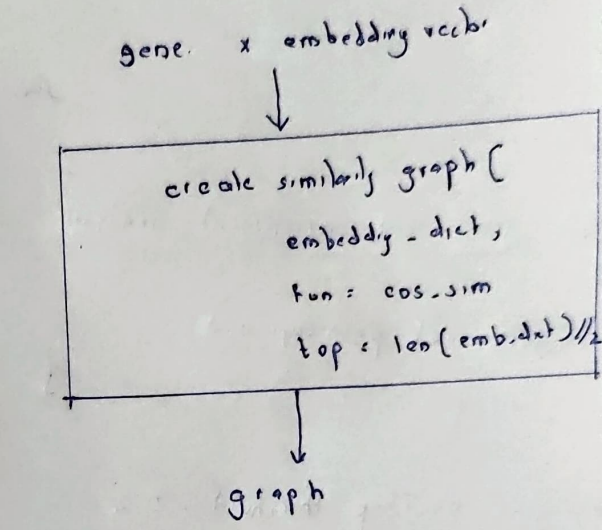
④ DAUG TARGET GENE FILTERATION

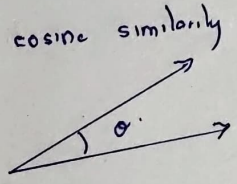
pass list here.

Embeddings labels	no of missing genes
normal	9 missing
IAC	14.
AIS	13
mIA	13



⑤ create similarity graph.





angle θ close to 0
cos sim close to 1

highest ranked gene pairs have highest cosine similarity.
means those two genes are close in that case

plot the graphs here.

AIS

⑥ gene set enrichment analysis

gene set pathways used

: {

}

gene overlap threshold = 0.2

Results

IAC

MIA

GSR

normal

Steps

1. DATA PREPROCESSING & HVG SELECTION
2. LEIDEN CLUSTERS SET1
3. LEIDEN CLUSTERS SET2
4. DRUG TARGET GENES SELECTION
5. CREATING SIMILARITY GRAPH
6. GENE SET ENRICHMENT ANALYSIS

7. gene set enrichment analysis