

```
import scanpy as sc
import pandas as pd
```

Load AnnData

adata = sc.read\_h5ad("preprocessed.h5ad")

adata

AnnData object with n\_obs × n\_vars = 4579 × 178946  
obs: 'Sample', 'nMultiFragments', 'nMonoFragments', 'nFragments', 'nDIFragments', 'TSSEnrichment', 'ReadsInTSS', 'ReadsInPromoter', 'ReadsInBlacklist', 'PromoterRatio', 'PassQC', 'NucleosomeRatio', 'BlacklistRatio', 'Clusters'  
var: 'seqnames', 'start', 'end', 'width', 'strand', 'distToGeneStart', 'nearestGene', 'peakType', 'distToTSS', 'nearestTSS', 'replicateScoreQuantile'  
uns: 'GeneScoresColumns'  
obsm: 'GeneScores', 'X\_svd', 'umap'

Explore AnnData

obs

adata.obs													
	Sample	nMultiFragments	nMonoFragments	nFragments	nDIFragments	TSSEnrichment	ReadsInTSS	ReadsInPromoter	ReadsInBlacklist	PromoterRatio	PassQC	Nu	
DACE657_mKate2#GGAGTCTGTTTGGGCG-1	DACE657_mKate2	127326	380482	956946	449138	2.936	11325	66220	37633	0.034600	1	1.5	
DACE657_mKate2#ATTCAACCAGGTATTT-1	DACE657_mKate2	57136	633366	861039	170537	1.119	3236	32952	36300	0.019135	1	0.3	
DACE657_mKate2#ACATTGCAGGGATGAC-1	DACE657_mKate2	57182	600534	823762	166046	1.938	6229	47176	33188	0.028634	1	0.3	
DACE657_mKate2#CTCTTGATCTTGTC-1	DACE657_mKate2	22834	697688	819298	98776	2.809	9524	57701	32171	0.035214	1	0.1	
DACE657_mKate2#GTCATCACAATGCATAT-1	DACE657_mKate2	105585	332298	642036	204153	10.578	51937	225568	19204	0.175666	1	0.9	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
DACE657_mKate2#CGCATTTGTTGCGCT-1	DACE657_mKate2	66	575	1002	361	2.871	29	125	32	0.062375	1	0.7	
DACE657_mKate2#TGTCATAAGGGACTAA-1	DACE657_mKate2	59	526	1002	417	4.059	41	186	36	0.092814	1	0.9	
DACE657_mKate2#CGCAAATTCGTATTT-1	DACE657_mKate2	87	625	1001	289	15.239	177	626	41	0.312687	1	0.6	
DACE657_mKate2#CATCGCTTCGGTATG-1	DACE657_mKate2	62	548	1000	390	4.356	44	169	24	0.084500	1	0.8	
DACE657_mKate2#CTATGGCCACGTTACA-1	DACE657_mKate2	109	382	1000	509	2.871	29	111	46	0.055500	1	1.6	

4579 rows × 14 columns

obs\_names

adata.obs\_names

Index(['DACE657\_mKate2#GGAGTCTGTTTGGGCG-1',  
 'DACE657\_mKate2#ATTCAACCAGGTATTT-1',  
 'DACE657\_mKate2#ACATTGCAGGGATGAC-1',  
 'DACE657\_mKate2#CTCTTGATCTTGTC-1',  
 'DACE657\_mKate2#GTCATCACAATGCATAT-1',  
 'DACE657\_mKate2#AGCTACGTCCAAATCA-1',  
 'DACE657\_mKate2#AGGATTGAGCTACTGG-1',  
 'DACE657\_mKate2#TGATATCGACCCACAG-1',  
 'DACE657\_mKate2#AAACCGCGTAACCTACG-1',  
 'DACE657\_mKate2#AATGCAACGTAATT-1',  
 ...  
 'DACE657\_mKate2#GCGTGCTAGGCTATGT-1',  
 'DACE657\_mKate2#GGTAAACAGGCTAAG-1',  
 'DACE657\_mKate2#GTGCTTTCCACCTTA-1',  
 'DACE657\_mKate2#TGTGGCTCAGAAACGT-1',  
 'DACE657\_mKate2#TTTCCACGACGAG-1',  
 'DACE657\_mKate2#CGCATTGTTGCGCT-1',  
 'DACE657\_mKate2#TGTCATAAGGGACTAA-1',  
 'DACE657\_mKate2#CGCAAATTCGTATTT-1',  
 'DACE657\_mKate2#CATCGCTTCGGTATG-1',  
 'DACE657\_mKate2#CTATGGCCACGTTACA-1'],  
 dtype='object', length=4579)

var

adata.var

	seqnames	start	end	width	strand	distToGeneStart	nearestGene	peakType	distToTSS	nearestTSS	replicateScoreQuantile
Peak1	chr1	3481922	3482422	501	*	267690	Xkr4	Intronic	176731	uc007afg.1	0.114
Peak2	chr1	3493610	3494110	501	*	279378	Xkr4	Intronic	165043	uc007afg.1	0.327
Peak3	chr1	3514773	3515273	501	*	300541	Xkr4	Intronic	143880	uc007afg.1	0.965
Peak4	chr1	3516762	3517262	501	*	302530	Xkr4	Intronic	141891	uc007afg.1	0.676
Peak5	chr1	3534649	3535149	501	*	320417	Xkr4	Intronic	124004	uc007afg.1	0.835
...	...	...	...	...	...	...	...	...	...	...	...
Peak178942	chrX	169936383	169936883	501	*	251434	Mid1	Intronic	9678	uc009rcn.1	0.113
Peak178943	chrX	169937107	169937607	501	*	252158	Mid1	Intronic	10402	uc009rcn.1	0.775

	seqnames	start	end	width	strand	distToGeneStart	nearestGene	peakType	distToTSS	nearestTSS	replicateScoreQuantile
Peak178944	chrX	169937886	169938386	501	*	252937	Mid1	Intronic	11181	uc009rcn.1	0.586
Peak178945	chrX	169950318	169950818	501	*	265369	Mid1	Intronic	22316	uc009rcn.1	0.724
Peak178946	chrX	169961044	169961544	501	*	276095	Mid1	Intronic	11590	uc009rcn.1	0.668

178946 rows × 11 columns

obsm

adata.obsm
AxisArrays with keys: GeneScores, X_svd, umap

uns

adata.uns
OverloadedDict, wrapping: {'GeneScoresColumns': array(['Xkr4', 'Rp1', 'Sox17', ..., 'Mid1', '4933400A11Rik', 'Asmt'], dtype=object)} With overloaded keys: ['neighbors'].

Gene Scores

adata.obsm["GeneScores"].shape
(4579, 24333)
adata.uns["GeneScoresColumns"]
array(['Xkr4', 'Rp1', 'Sox17', ..., 'Mid1', '4933400A11Rik', 'Asmt'], dtype=object)
adata.uns["GeneScoresColumns"].shape
(24333,)
df_gene_scores = pd.DataFrame( adata.obsm["GeneScores"], columns=adata.uns["GeneScoresColumns"], index=adata.obs_names ) df_gene_scores

	Xkr4	Rp1	Sox17	Mrpl15	Lypla1	Tcea1	Rgs20	Atp6v1h	Oprk1	Npbwr1	...	Prps2	Frmpd4	Msl3	Arhgap6	Amelx	Hccs	Gm15246	Mid1
DACE657_mKate2#GGAGTCTGTTGGGCG-1	0.262	0.0	0.0	0.000	0.183	0.000	0.0	0.130	0.0	0.0	...	0.0	0.0	0.000	0.0	0.0	0.638	0.693	0.0
DACE657_mKate2#ATTCAACCAGGTATTT-1	0.000	0.0	0.0	1.181	1.605	0.795	0.0	1.931	0.0	0.0	...	0.0	0.0	0.000	0.0	0.0	2.799	3.039	0.0
DACE657_mKate2#ACATTGCAGGGATGAC-1	0.000	0.0	0.0	0.983	0.000	1.323	0.0	0.237	0.0	0.0	...	0.0	0.0	0.814	0.0	0.0	1.747	1.897	0.0
DACE657_mKate2#CTCTTGATCTTGTTCCA-1	0.000	0.0	0.0	0.322	0.814	0.806	0.0	0.807	0.0	0.0	...	0.0	0.0	0.000	0.0	0.0	0.000	0.000	0.0
DACE657_mKate2#GTCATCACATGCATAT-1	0.434	0.0	0.0	0.807	0.303	0.301	0.0	0.258	0.0	0.0	...	0.0	0.0	0.370	0.0	0.0	0.529	0.575	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
DACE657_mKate2#CGCATTGTGTCGCGCT-1	0.000	0.0	0.0	0.000	0.000	0.000	0.0	0.000	0.0	0.0	...	0.0	0.0	0.000	0.0	0.0	97.786	106.150	0.0
DACE657_mKate2#TGTCATAAGGGACTAA-1	0.000	0.0	0.0	0.000	0.000	0.000	0.0	0.000	0.0	0.0	...	0.0	0.0	0.000	0.0	0.0	0.000	0.000	0.0
DACE657_mKate2#CGCAAATTCGTATTT-1	0.000	0.0	0.0	0.000	0.000	0.000	0.0	0.000	0.0	0.0	...	0.0	0.0	0.000	0.0	0.0	0.000	0.000	0.0
DACE657_mKate2#CATCGCTTCCGGTATG-1	0.000	0.0	0.0	0.000	0.000	0.000	0.0	0.000	0.0	0.0	...	0.0	0.0	0.000	0.0	0.0	0.000	0.000	0.0
DACE657_mKate2#CTATGGCCACGTTACA-1	0.000	0.0	0.0	0.000	0.000	0.000	0.0	0.000	0.0	0.0	...	0.0	0.0	0.000	0.0	0.0	0.000	0.000	0.0

4579 rows × 24333 columns

df_gene_scores.sum(axis=1)
DACE657_mKate2#GGAGTCTGTTGGGCG-1 10000.047 DACE657_mKate2#ATTCAACCAGGTATTT-1 9999.985 DACE657_mKate2#ACATTGCAGGGATGAC-1 10000.017 DACE657_mKate2#CTCTTGATCTTGTTCCA-1 10000.018 DACE657_mKate2#GTCATCACATGCATAT-1 9999.981 ... DACE657_mKate2#CGCATTGTGTCGCGCT-1 10000.002 DACE657_mKate2#TGTCATAAGGGACTAA-1 10000.000 DACE657_mKate2#CGCAAATTCGTATTT-1 9999.996 DACE657_mKate2#CATCGCTTCCGGTATG-1 9999.999 DACE657_mKate2#CTATGGCCACGTTACA-1 10000.000 Length: 4579, dtype: float64