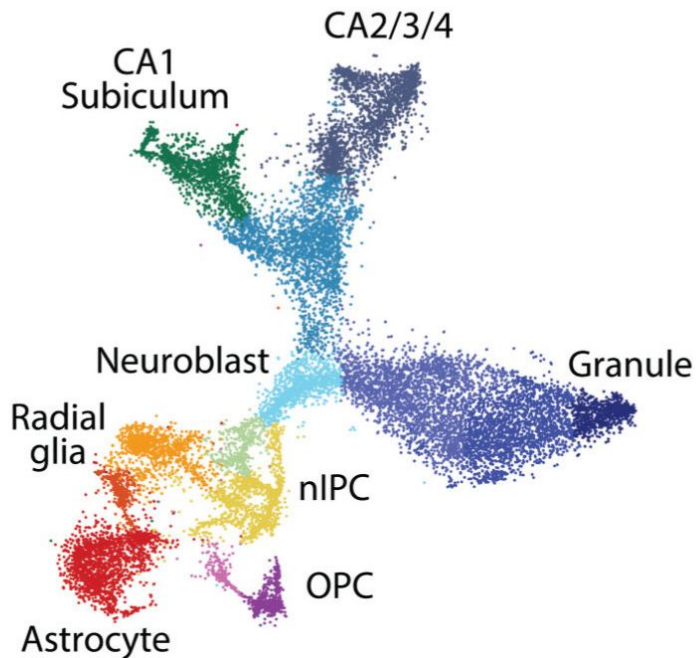


Детекция топологических сингулярностей на данных scRNA-seq

Александр Минаков

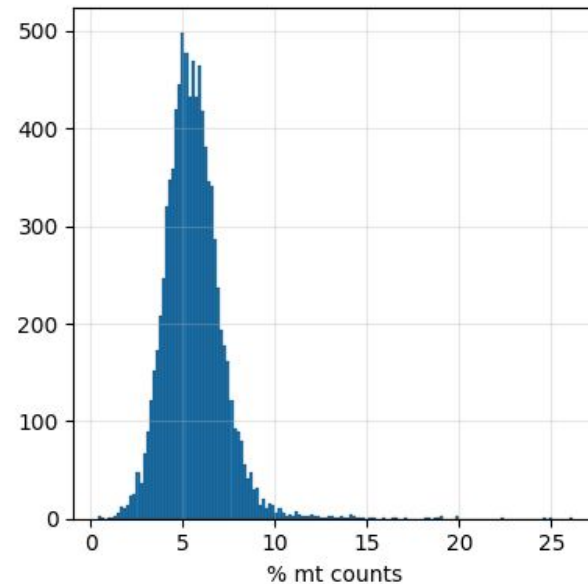
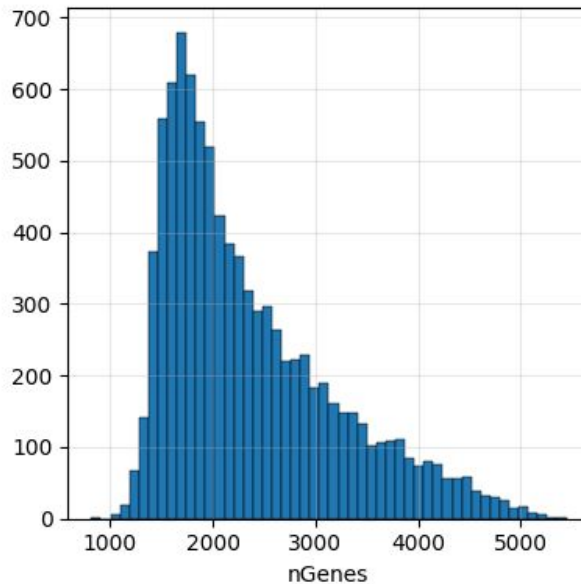
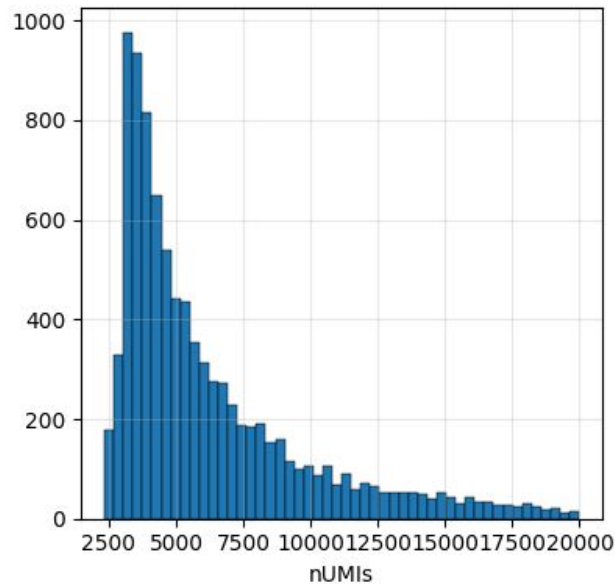
Датасет



- Развивающийся гиппокамп мыши в первый (P0) и пятый (P0) постнатальные дни
- GSE104323
- PMID: 29335606
- 9214 клеток
- 27933 генов

RNA velocity of single cells (PMID: 30089906)

Контроль качества

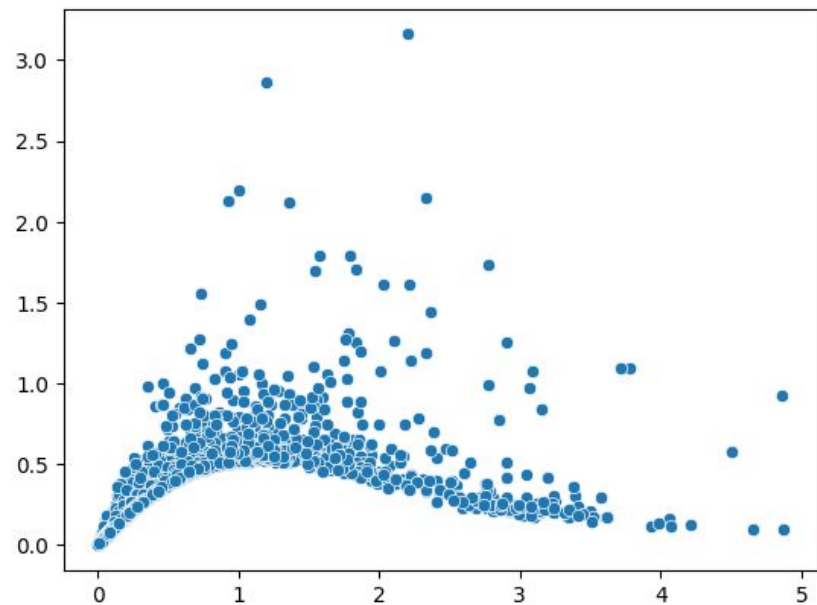
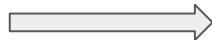
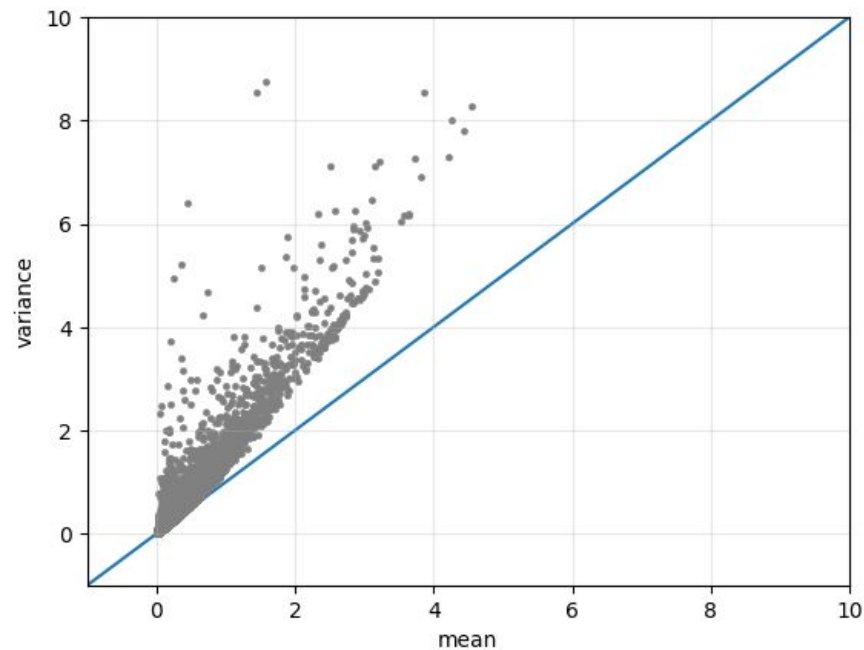


Пустых капель нет (видимо уже отфильтрованы)

18 клеток с процентом митохондриальных каунтов > 15

9 дублетов (scrublet)

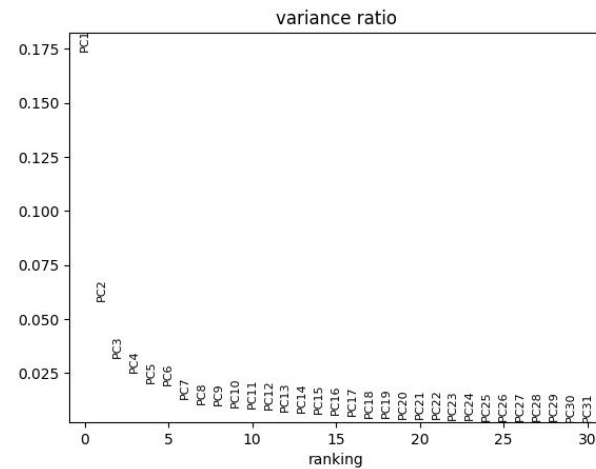
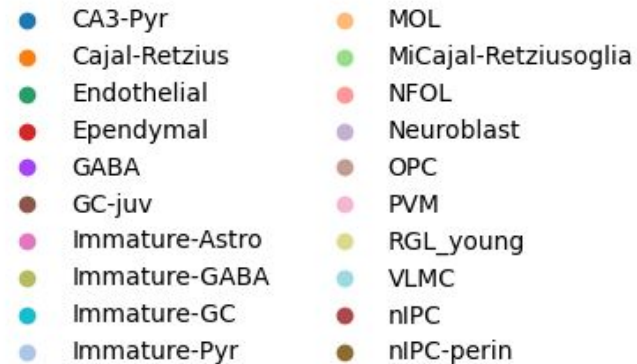
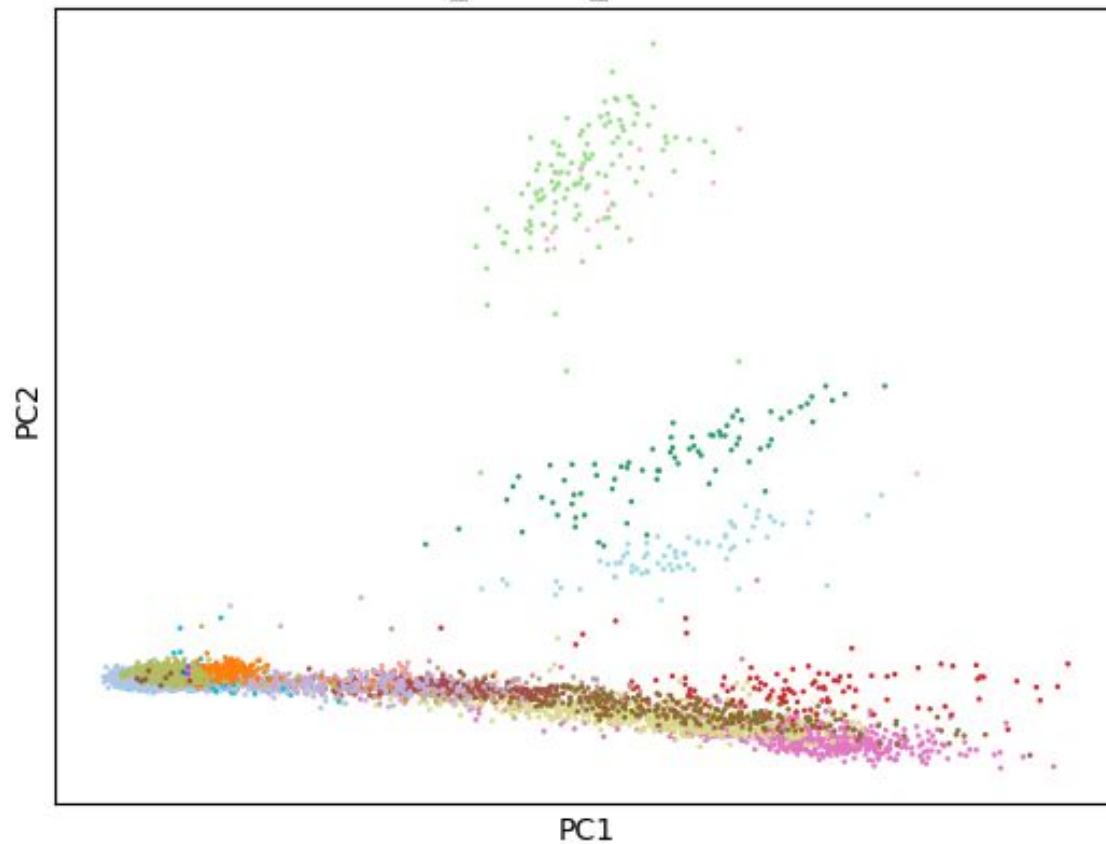
Нормализация и логарифмирование



Затем выбраны 3000 высоковариабельных генов

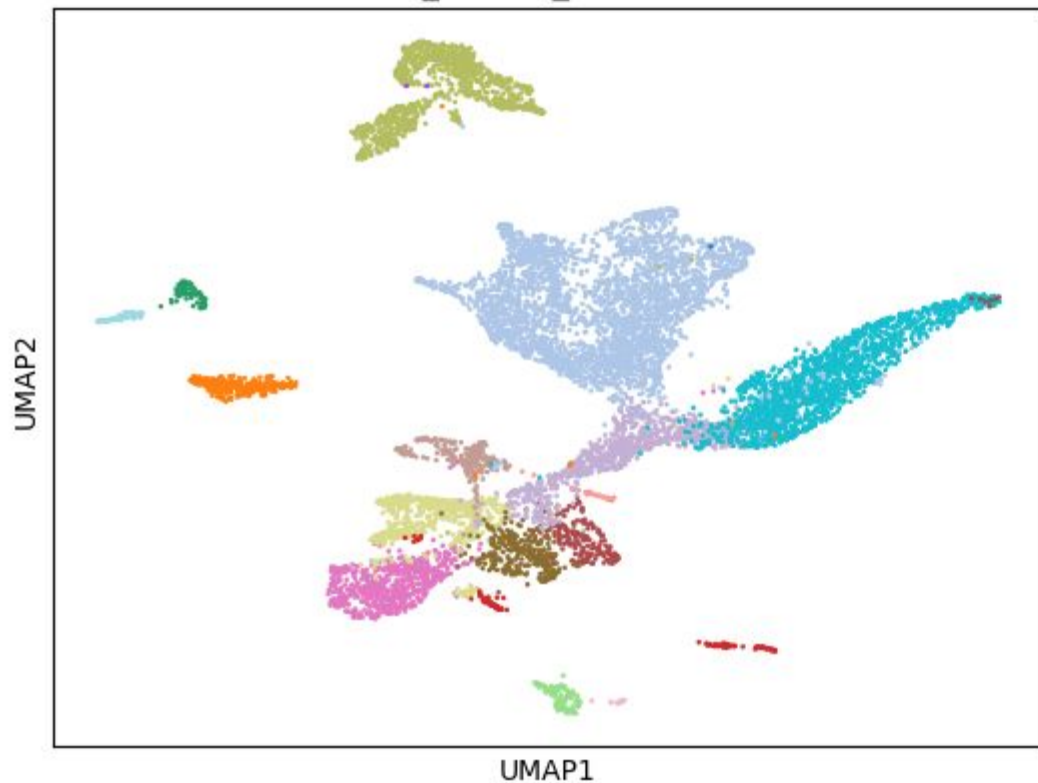
PCA

cell_cluster_metadata



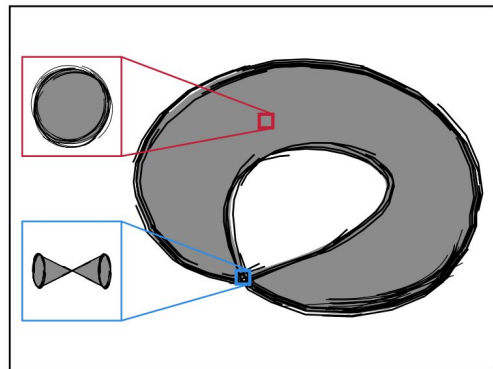
UMAP

cell_cluster_metadata

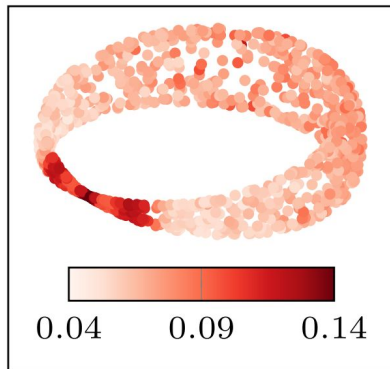


- | | |
|----------------|----------------------|
| CA3-Pyr | MOL |
| Cajal-Retzius | MiCajal-Retziusoglia |
| Endothelial | NFOL |
| Ependymal | Neuroblast |
| GABA | OPC |
| GC-juv | PVM |
| Immature-Astro | RGL_young |
| Immature-GABA | VLMC |
| Immature-GC | nIPC |
| Immature-Pyr | nIPC-perin |

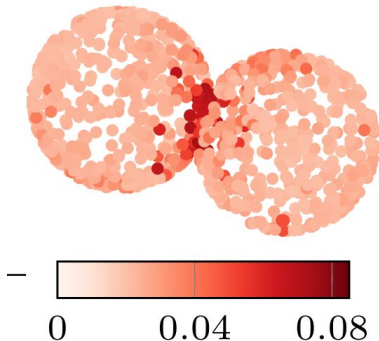
Топологические сингулярности. Euclidity



Input space with singularities



Euclidity

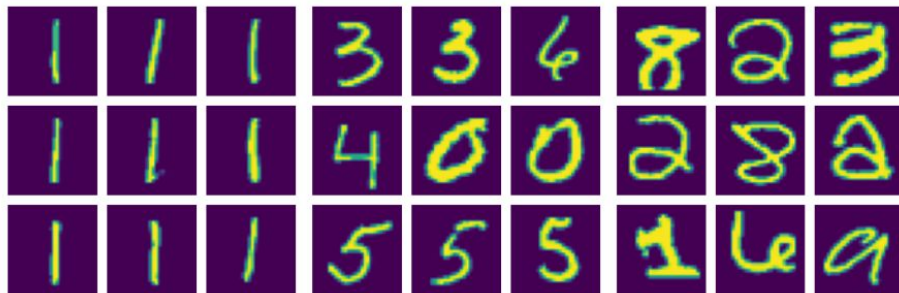


Сингулярности - non-manifold structures

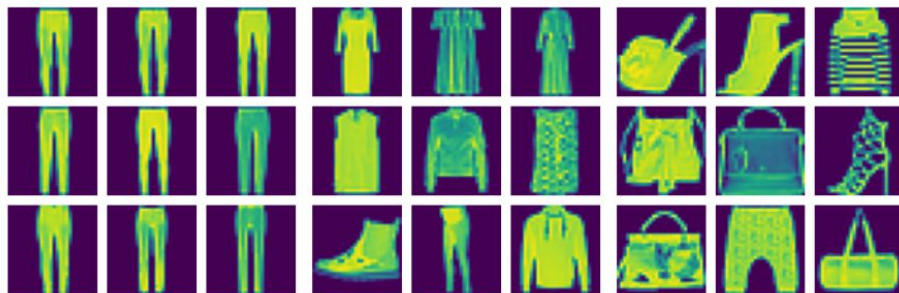
If there exists a neighbourhood U of x that is homeomorphic to \mathbb{R}^n , we say that x is a *regular point*, otherwise we call x a *singularity*.

Euclidity - low values indicate Euclidean neighbourhoods while high values indicate singular regions of a data set

Euclidicity



(a) MNIST

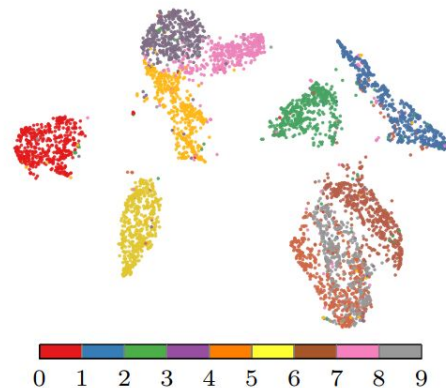


(b) FASHIONMNIST

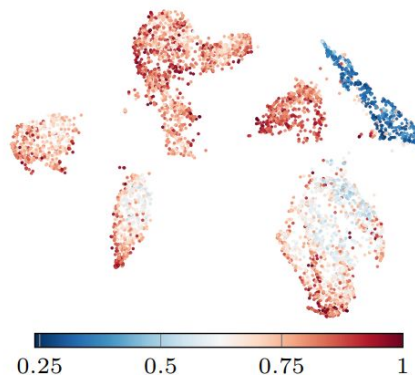
низкая
euclidicity

Средняя
euclidicity

высокая
euclidicity

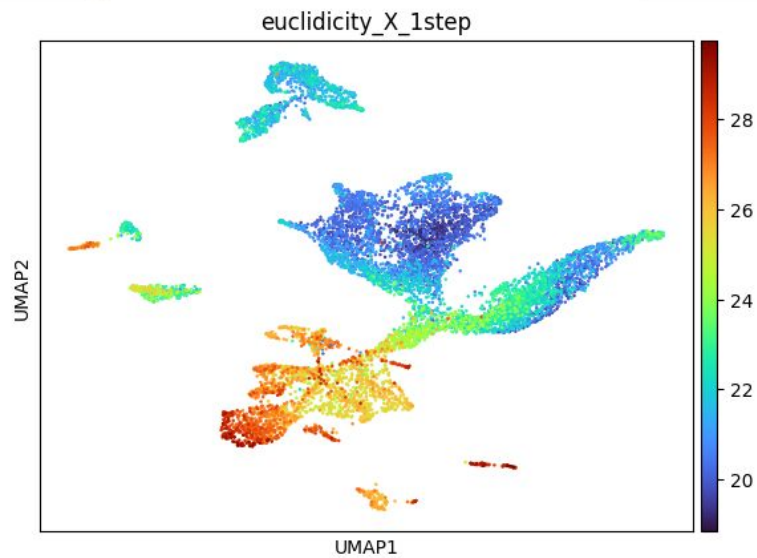
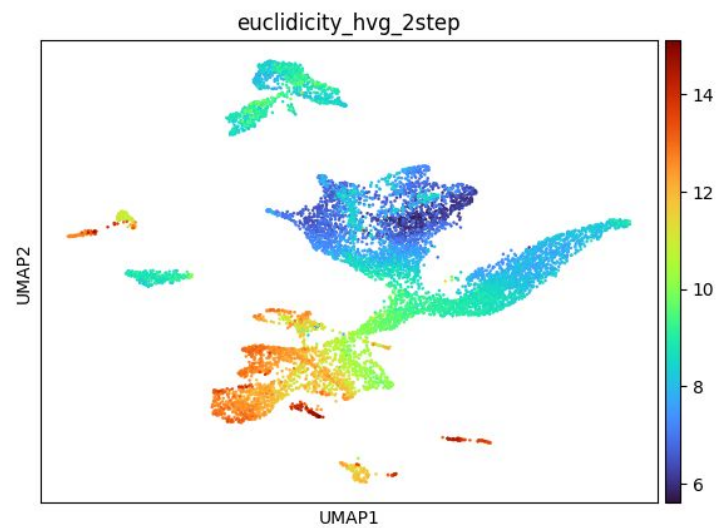
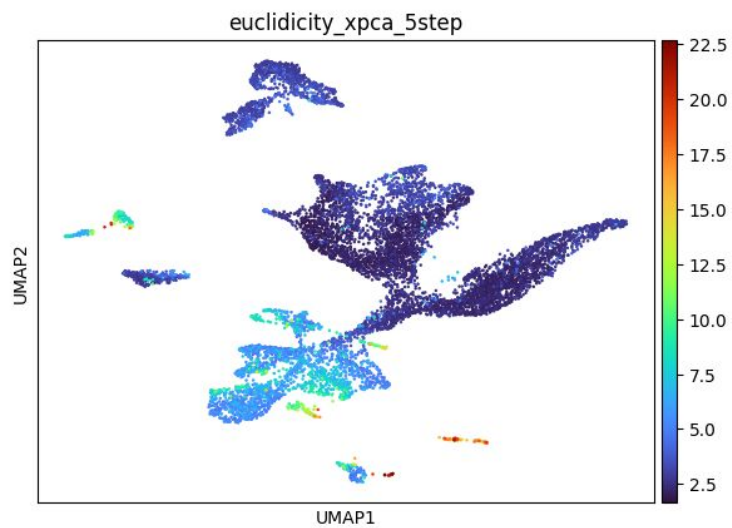


(a) Labels

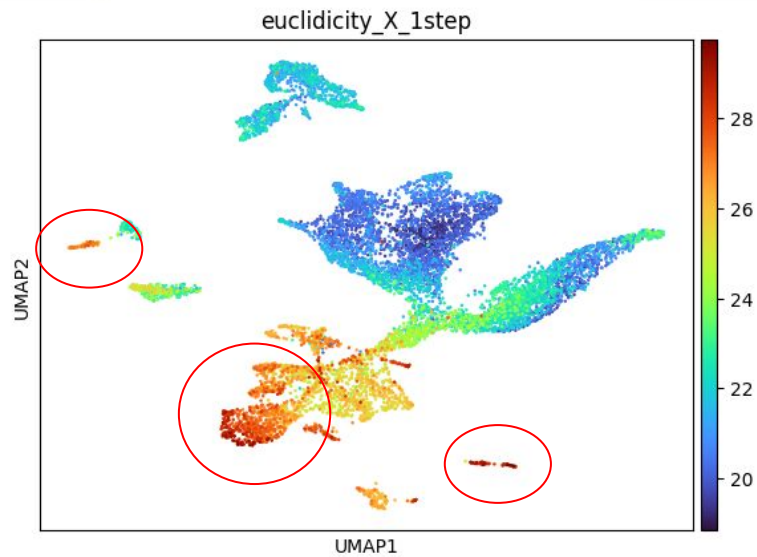
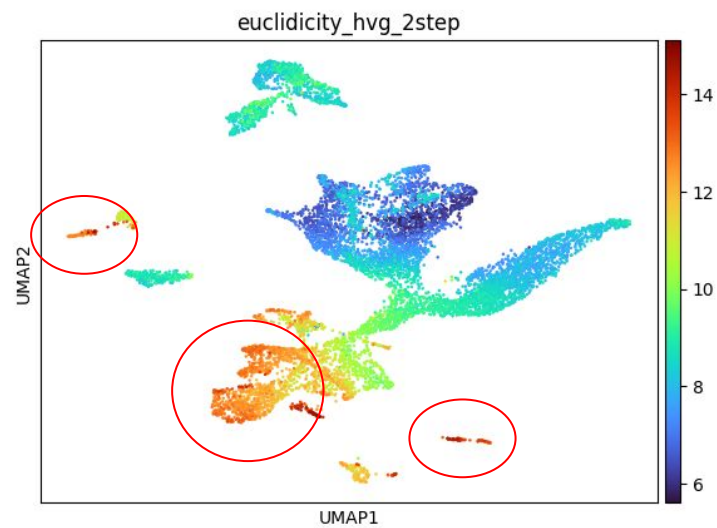
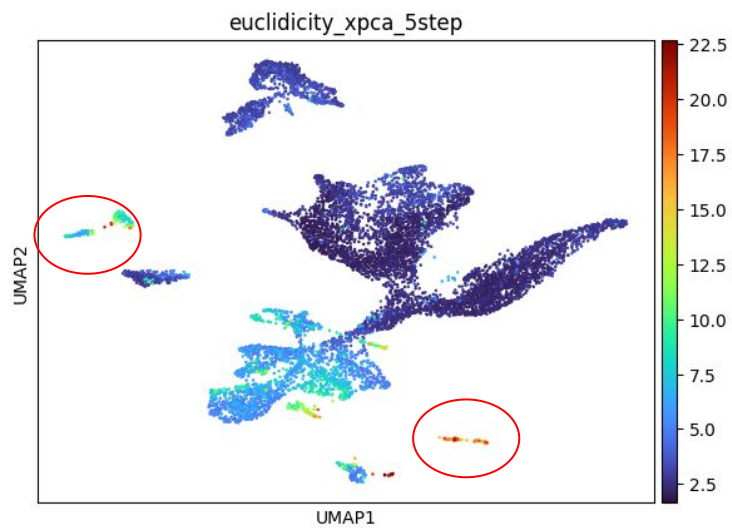


(b) Euclidicity scores

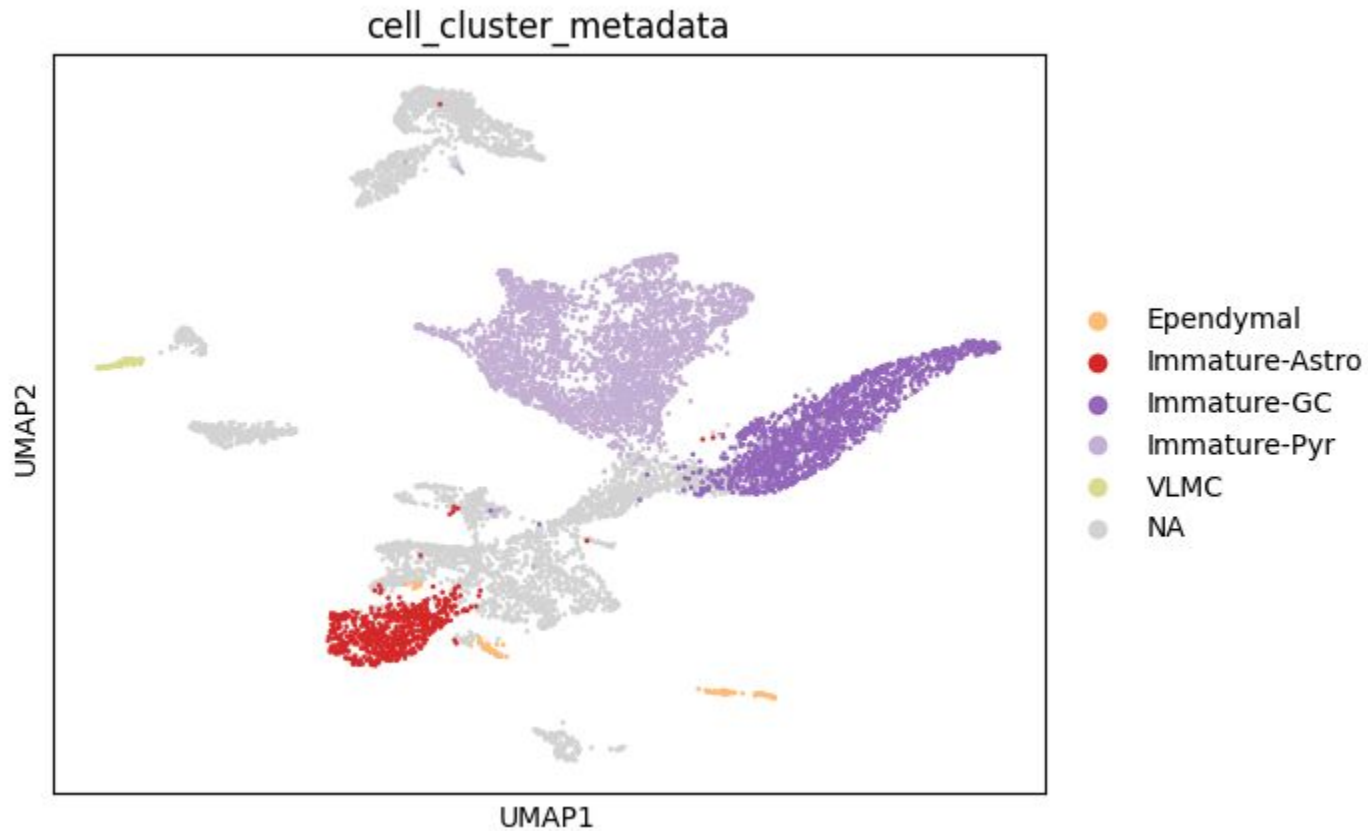
Результаты



Результаты



Результаты



	total_counts	n_genes_by_counts	total_counts_mt
cell_cluster_metadata			
CA3-Pyr	5145.000000	2368.000000	269.000000
Cajal-Retzius	6092.721191	2428.139373	310.891998
Endothelial	8728.315430	2720.657895	396.500000
Ependymal	6247.128418	2692.623932	352.068390
GABA	14659.000000	4386.000000	1547.000000
GC-juv	5801.625000	2577.375000	312.500000
Immature-Astro	4632.361816	2056.333333	254.471695
Immature-GABA	6966.404297	2582.648091	357.782257
Immature-GC	5207.594238	2148.045108	310.287048
Immature-Pyr	8285.125000	2784.988028	454.742615
MOL	12495.000000	3226.000000	1245.000000
MiCajal-Retziusoglia	5058.561523	1781.456140	306.201752
NFOL	5779.515137	2246.909091	374.090912
Neuroblast	3797.367432	1693.868481	211.944443
OPC	4518.241211	2086.780172	253.456894
PVM	4456.411621	1746.588235	216.470581
RGL_young	4104.001953	1918.381625	192.256180
VLMC	5910.157227	2213.528571	309.885712
nIPC	5981.533203	2370.044444	280.250000
nIPC-perin	6074.264648	2509.821306	265.051544