



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана

Кластеризация клеток периферической крови от здорового донора после single-cell RNA sequencing

Чувакова Любовь Николаевна

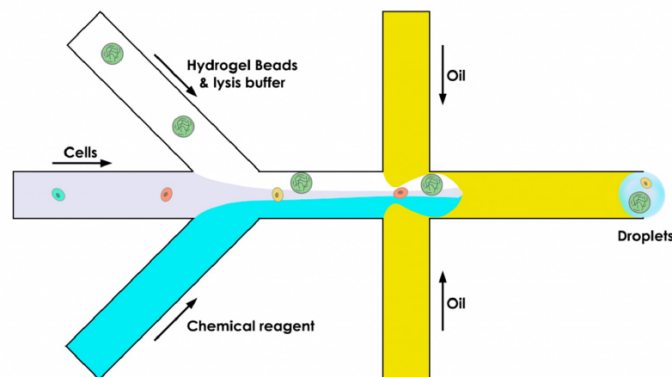


Общая схема эксперимента scRNA-Seq



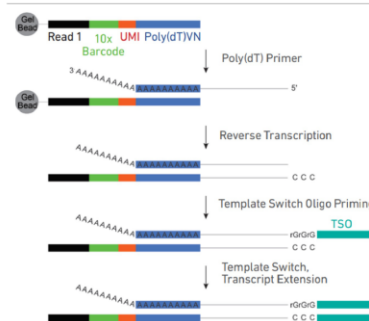
Droplet-based (капельные) методы

Капельные методы основаны на том, что клетки изолируются друг от друга, поступая по капиллярам в масляную фракцию и образуя там отдельные компартменты, содержащие необходимые реагенты и одну клетку



Источник:
<http://mccarrolllab.org/dropseq/>

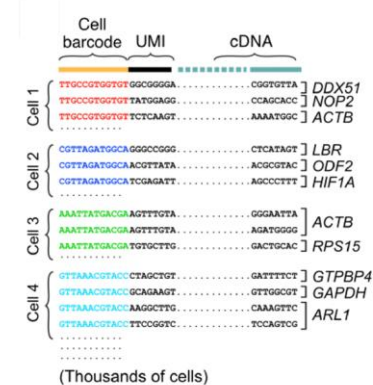
10x v3 3' Gel Beads



К каждому шарику прикреплён уникальный праймер, который состоит из

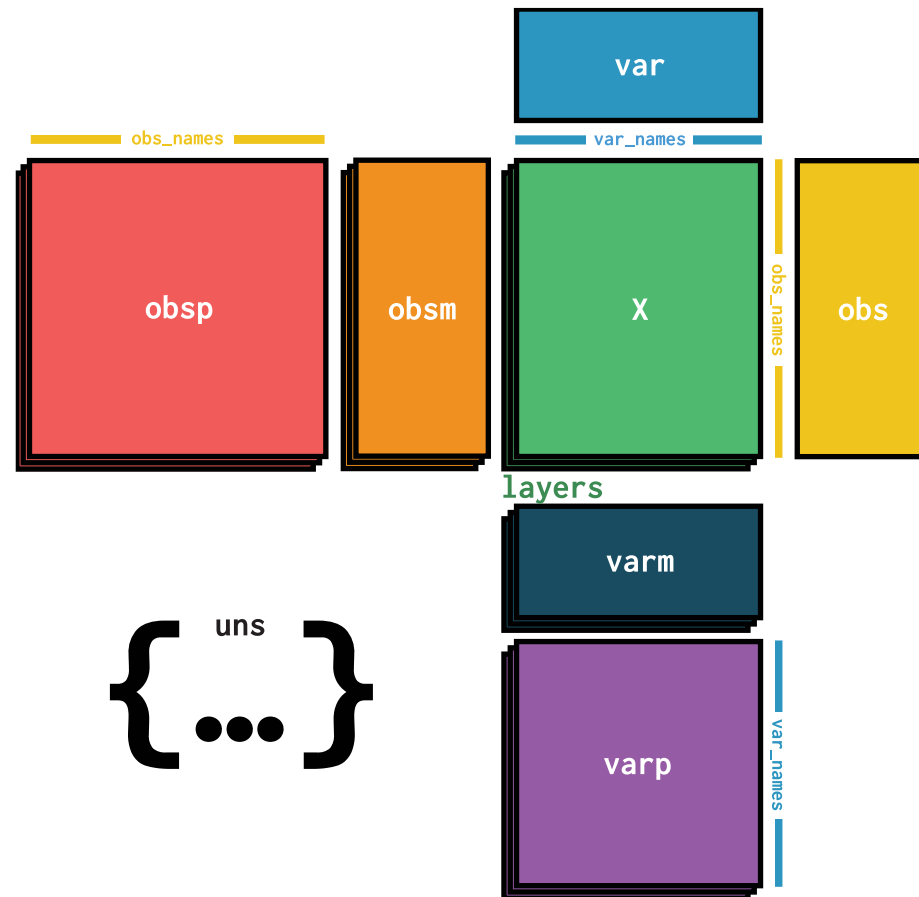
1. Праймера Illumina TruSeq Read 1,
2. **Баркода** (последовательности, которая одинакова у всех праймеров данного шарика, однако различается между всеми шариками),
3. **UMI** (последовательности, которая уникальна для всех праймеров данного шарика, но может повторяться между шариками),
4. Poly(dT)-последовательности.

Баркоды и UMI



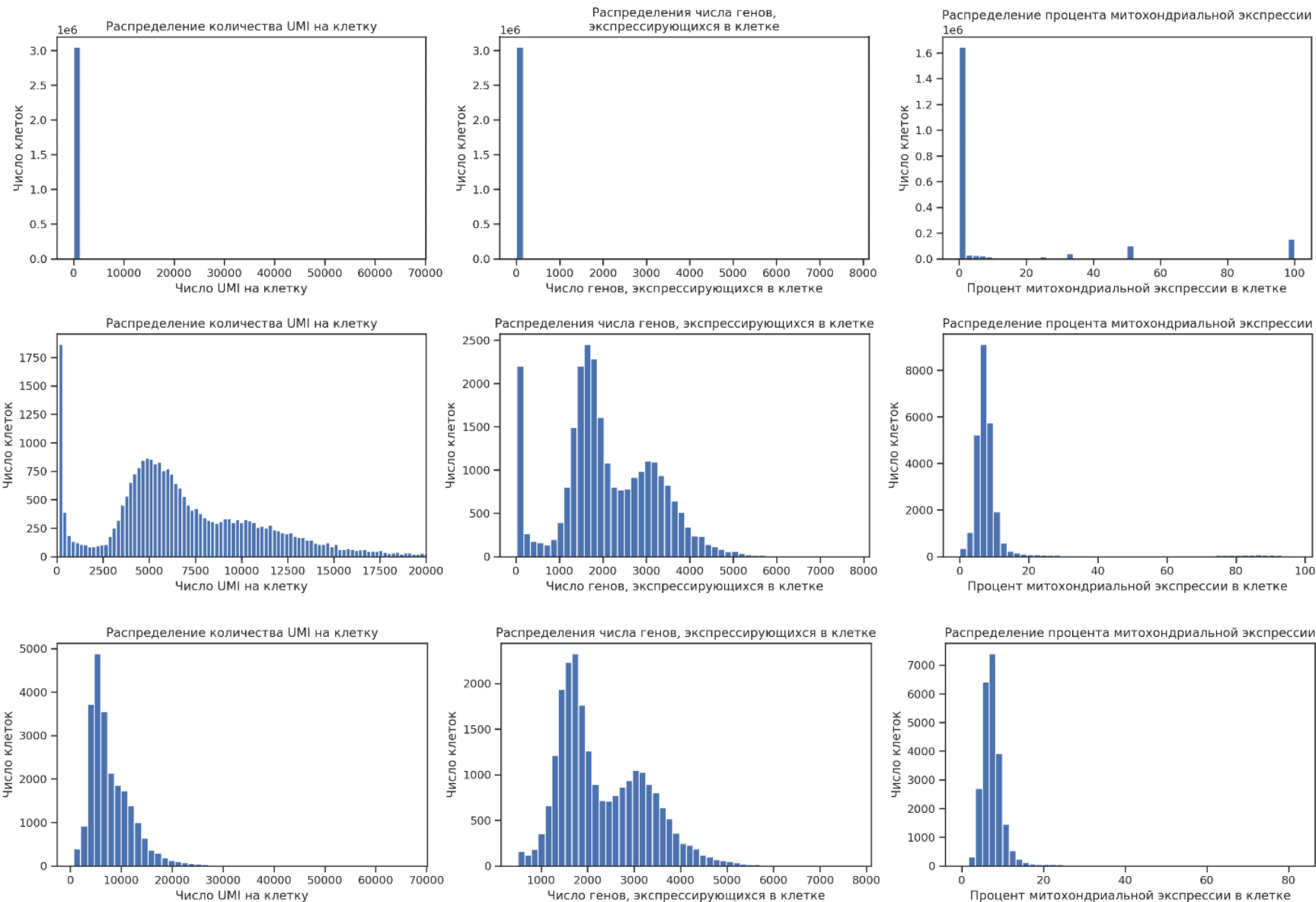


Обработка данных



Препроцессинг данных

Исходно: число клеток × число генов
3069478 × 36601.

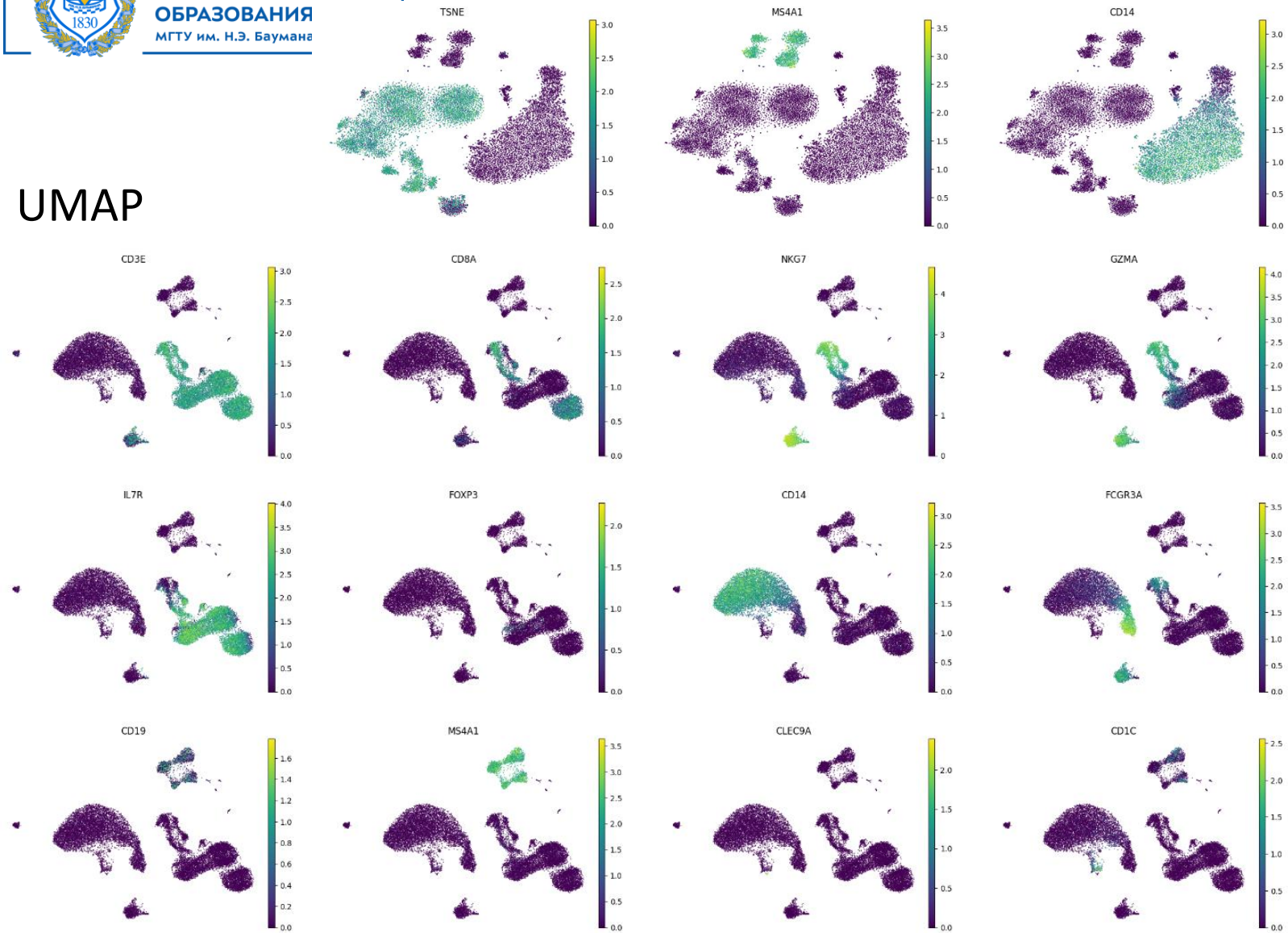


Стало: 22019 × 36601

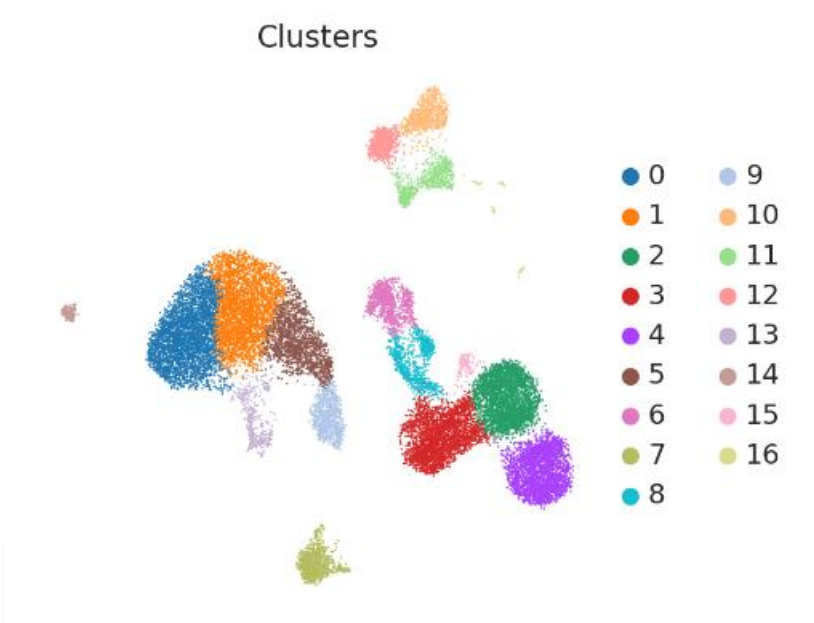
Отбор high variable genes:
22019x3000

tSNE

UMAP



Leiden

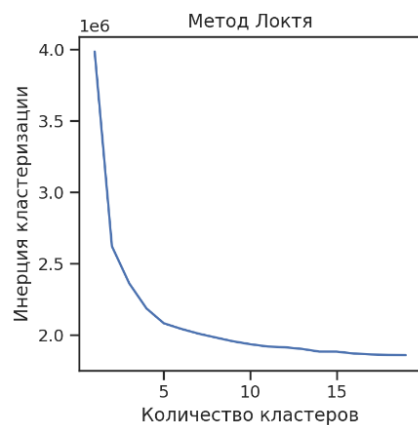


	names	scores	logfoldchanges	pvals	pvals_adj
62	IL7R	73.588806	2.593273	0.000000e+00	0.000000e+00
79	CCR7	59.787647	2.685055	0.000000e+00	0.000000e+00
95	LEF1	50.137753	2.549269	0.000000e+00	0.000000e+00
115	FHIT	36.958820	3.028283	2.112866e-245	7.851065e-244
123	CHRM3-AS2	33.802128	2.736526	5.483530e-212	1.813033e-210
...
5090	AC008581.1	0.817699	2.514848	4.135972e-01	6.747524e-01
5091	TEX101	0.817415	2.550114	4.137581e-01	6.749547e-01
5092	AC073525.1	0.817241	2.548816	4.138575e-01	6.750869e-01
5094	AC019257.2	0.816243	2.541338	4.144279e-01	6.758825e-01
5095	HPN-AS1	0.816241	2.541331	4.144286e-01	6.758825e-01

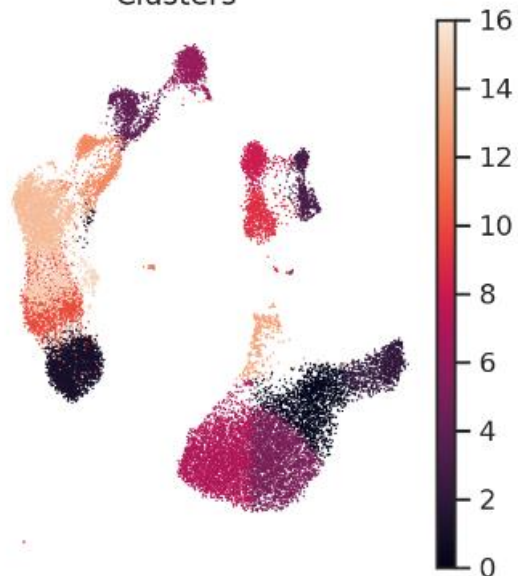
614 rows × 5 columns



KMeans



Clusters



Random forest

```
# оцениваем качество модели на тестовой выборке  
accuracy = clf.score(X_test, y_test)  
print(f'Accuracy: {accuracy:.2f}')
```

Accuracy: 0.82

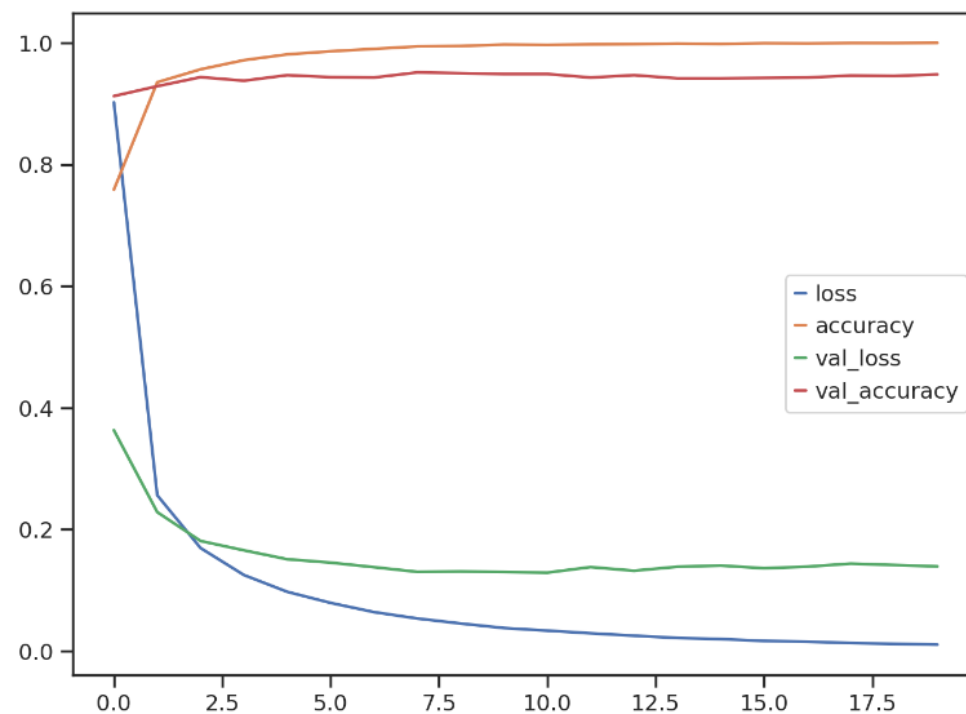
Нейронная сеть

```
model.summary() # архитектура модели
```

Model: "sequential_3"

Layer (type)	Output Shape	Param #
dense_6 (Dense)	(None, 32)	96032
dense_7 (Dense)	(None, 17)	561

Total params: 96,593
Trainable params: 96,593
Non-trainable params: 0



Выводы:

- 1) В данной работе было определено 17 кластеров популяций клеток, которые в последующем предстоит проанализировать.
- 2) Для данного анализа подходят только те методы, которые не требуют «у» (т.е. данные, которые мы и должны определить и распределить между клетками) на вход, а сами кластеризуют данные.



ЦЕНТР
ДОПОЛНИТЕЛЬНОГО
ОБРАЗОВАНИЯ
МГТУ им. Н.Э. Баумана



do.bmstu.ru