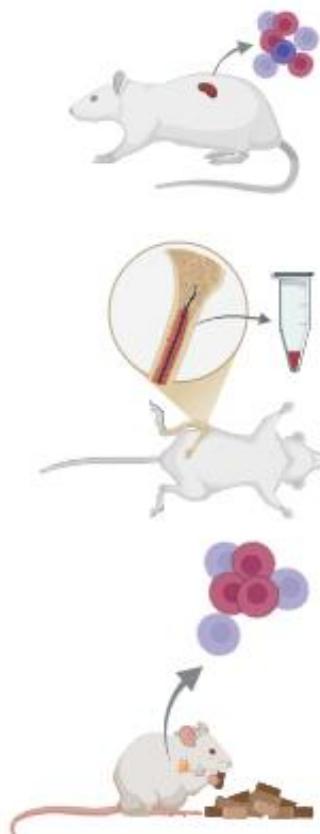


# Experimental design

13-14-mo female Balb/c



	Pooled sample	Sample name	Sample details	Hashtag used	Target recovery	Cells loaded
Bone marrow CD11b- Control vs Stress	BM_1	C-1	Control	H1	24000	30000
		C-2	Control	H2		
		CVS-1	Stressed	H3		
		CVS-2	Stressed	H4		
	BM_2	C-3	Control	H1	24000	30000
		C-4	Control	H2		
		CVS-3	Stressed	H3		
		CVS-4	Stressed	H4		
Thymus Control vs Ghrelin	GT_1	C-1	Control	H1	12000	10000
		G-1	Ghrelin Tx	H4		10000
	GT_2	C-2	Control	H2	12000	10000
		G-2	Ghrelin Tx	H5		10000
	GT_3	C-3	Control	H3	12000	10000
		G-3	Ghrelin Tx	H6		10000

# Experimental design

	Pooled sample	Sample name	Sample details	Hashtag used	Target recovery	Cells loaded	Hashtag antibodies specific against:	
 Bone marrow CD11b- Control vs Stress	BM_1	C-1	Control	H1	24000	30000	CD45 and MHC class I (Different barcodes/seq)	
		C-2	Control	H2				
		CVS-1	Stressed	H3				
		CVS-2	Stressed	H4				
	BM_2	C-3	Control	H1	24000	30000		
		C-4	Control	H2				
		CVS-3	Stressed	H3				
		CVS-4	Stressed	H4				

 Thymus Control vs Ghrelin	GT_1	C-1	Control	H1	12000	10000	CD45 and MHC class I (Different barcodes/seq)
		G-1	Ghrelin Tx	H4		10000	
	GT_2	C-2	Control	H2	12000	10000	
		G-2	Ghrelin Tx	H5		10000	
	GT_3	C-3	Control	H3	12000	10000	
		G-3	Ghrelin Tx	H6		10000	

# CD45 and MHC I

CD antigen	Cellular expression	Molecular weight (kDa)	Functions	Other names	Family relationships
CD45	All hematopoietic cells	180–240 (multiple isoforms)	Tyrosine phosphatase, augments signaling through antigen receptor of B and T cells, multiple isoforms result from alternative splicing (see below)	Leukocyte common antigen (LCA), T200, B220	Protein tyrosine phosphatase (PTP); fibronectin type III

- Mixture of 2 antibodies (CD45 + MHC I) conjugated to same oligo for multiplex single-cell sequencing.
- CD45: pan-leukocyte marker, labels all hematopoietic/immune cells (except erythrocytes & platelets), important for TCR/BCR signaling.
- MHC class I (H-2 a,b,d,j,k,s,u): labels all nucleated cells (non-hematopoietic + immune), presents antigens to CD8+ T cells.
- Use both to barcode all cells in most common mouse strains.



Why MHC I? Why non-hematopoietic? Both antibodies doesn't have to be attached to each cell. So why we introduce noise? Arent we aiming only for immune cell populations?

# So far findings

# Further potential research questions/ideas

# BM1



Pooled sample	Sample name	Sample details	Hashtag used	Target recovery	Cells loaded
BM_1	C-1	Control	H1	24000	30000
	C-2	Control	H2		
	CVS-1	Stressed	H3		
	CVS-2	Stressed	H4		

## Cell Calling Quality

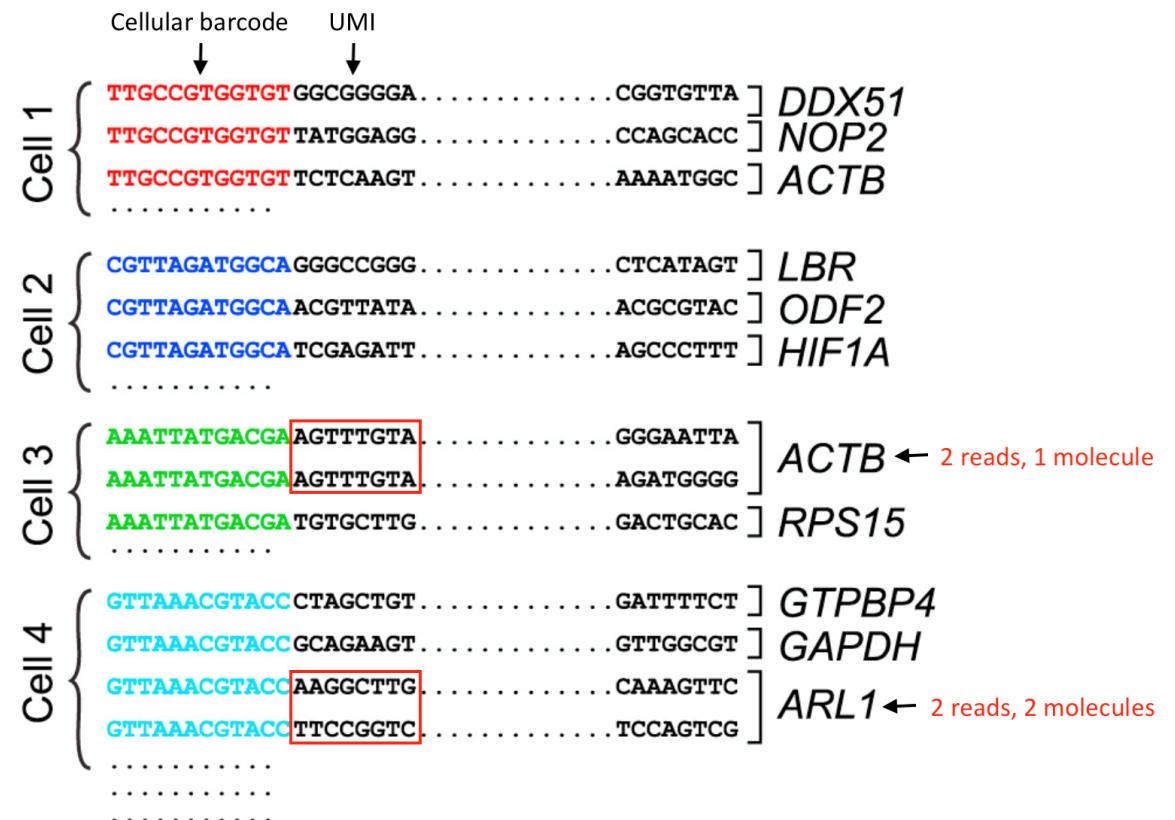
Cells	Confidently mapped reads in cells	Median genes per cell	Median UMI counts per cell	Total genes detected
6,218	88.9%	3,089	10,822	23,637

```
... reading from cache file cache/gpfs-helios-home-rostamne-CVS_SplVacc_BM_Ghr_Thy_Jul_24_scRNA-code-BM  
1-GEX_BM1_multi-outs-multi-count-raw_feature_bc_matrix-matrix.h5ad  
AnnData object with n_obs × n_vars = 2523666 × 33696  
var: 'gene_ids', 'feature_types'
```

# Why so few cells were retained in some of the experiments?

- What cells are these cells? What population of cells?

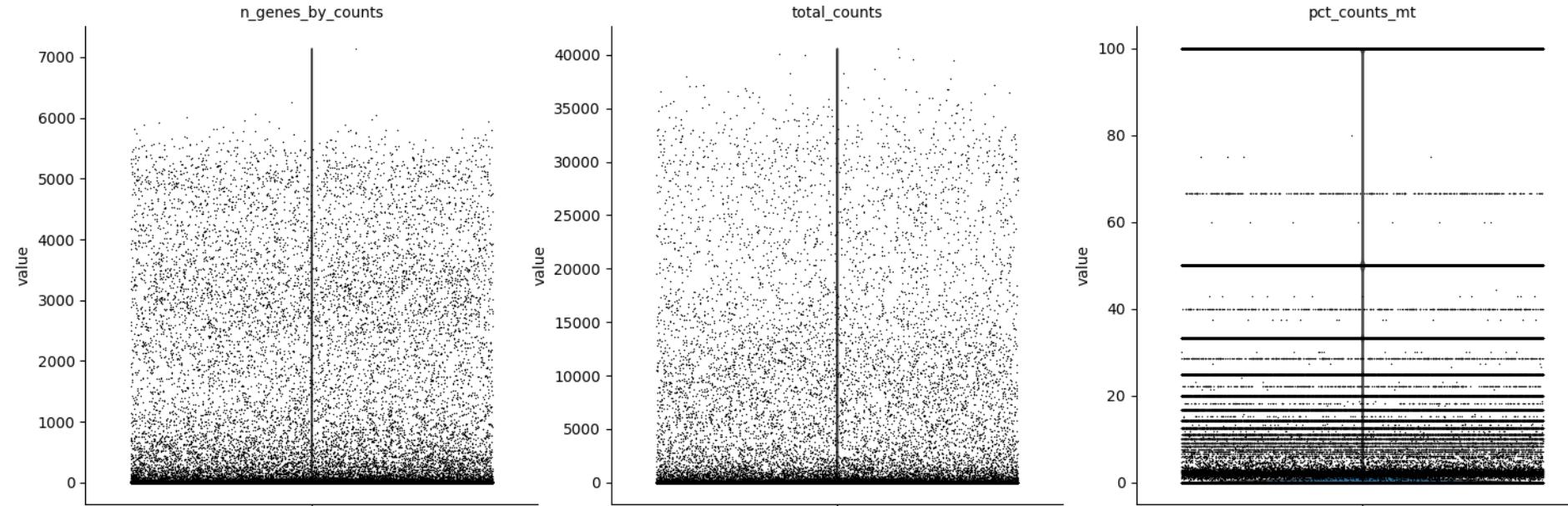
	Cell1	Cell2	...	CellN
Gene1	3	2	.	13
Gene2	2	3	.	1
Gene3	1	14	.	18
...	.	.	.	.
...	.	.	.	.
...	.	.	.	.
GeneM	25	0	.	0



(Thousands of cells)



# BM1 – raw data (all ~2 million barcodes)

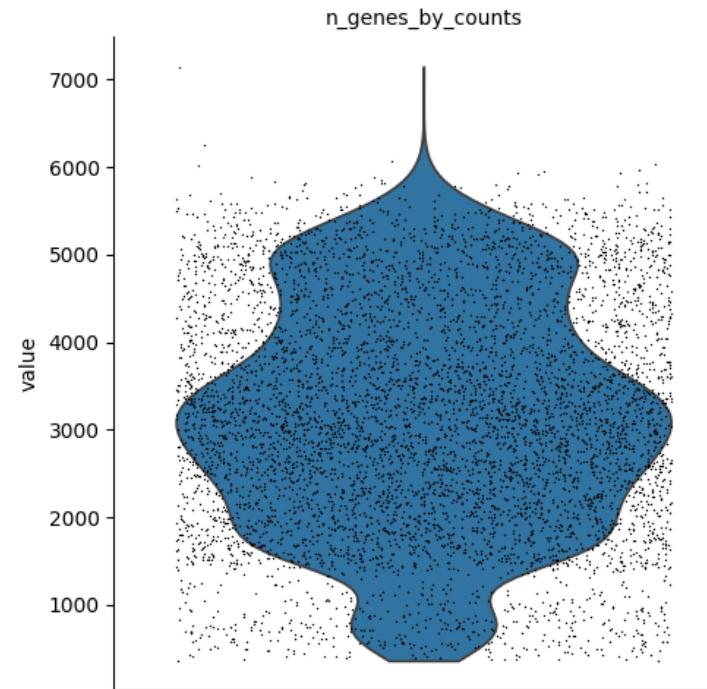


- Why so many barcodes have exact percentage?
- Why violin displayed as a line?

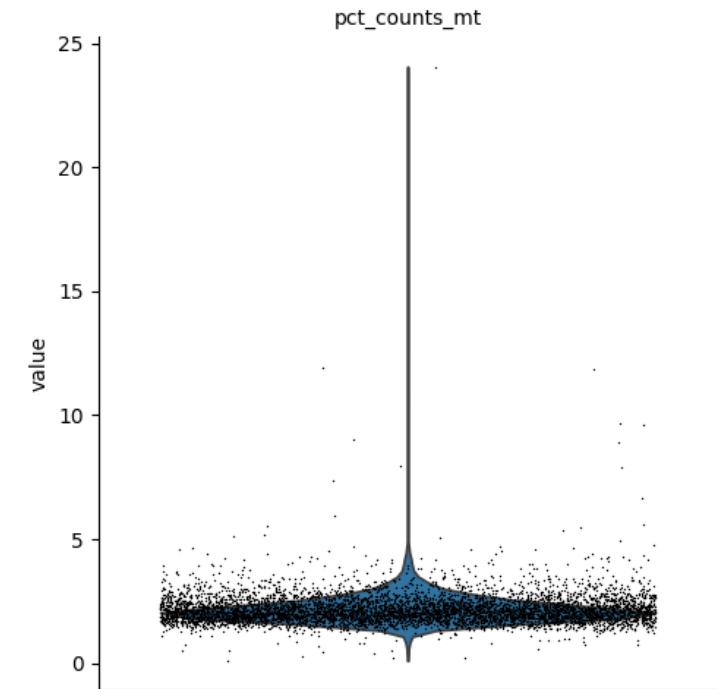
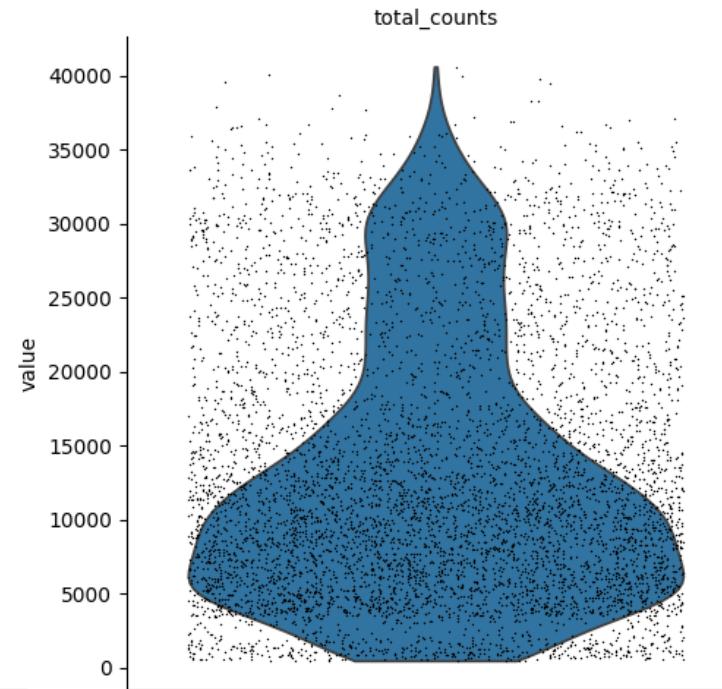
- the number of genes expressed in the count matrix
- the total counts per cell
- the percentage of counts in mitochondrial genes



# BM1-filtered



Many cells have about 3k genes detected



Very low percentage of mt genes in  
most most /all of cells

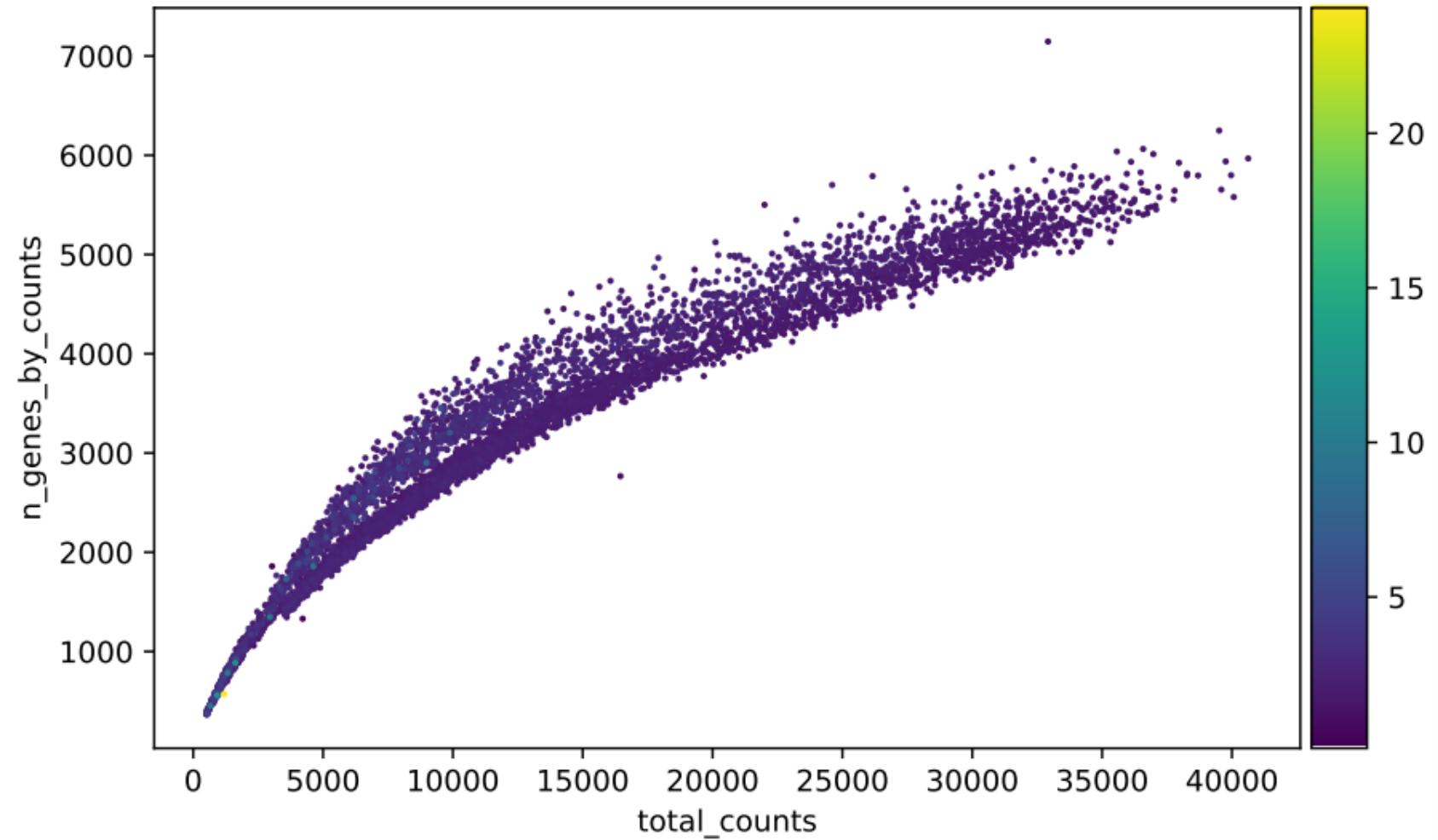
- the number of genes expressed in the count matrix
- the total counts per cell. the total number of UMIs detected
- the percentage of counts in mitochondrial genes



# BM1

sample\_filtered\_feature\_bc\_matrix

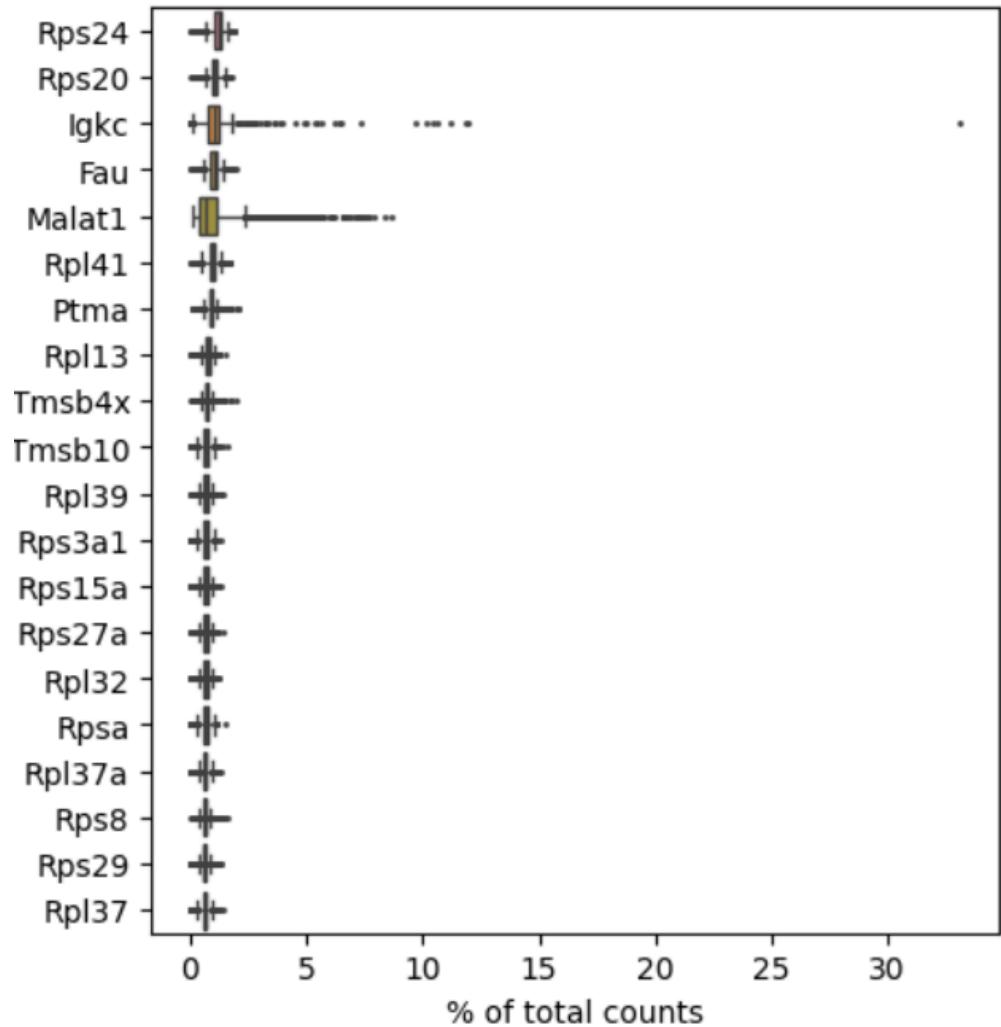
pct counts mt





# BM1 - Most dominant genes (top 20)

genes that yield the highest fraction of counts in each single cell, across all cells.



```
sc.pl.highest_expr_genes(adata_filtered, n_top=20, save="_filtered.png")
```

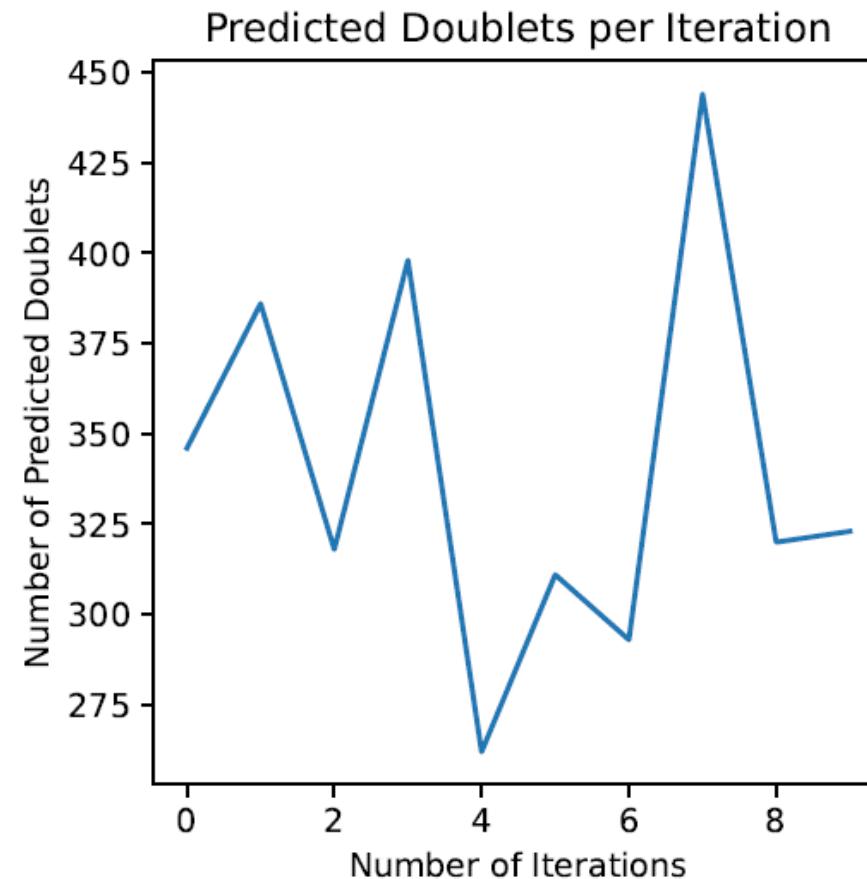


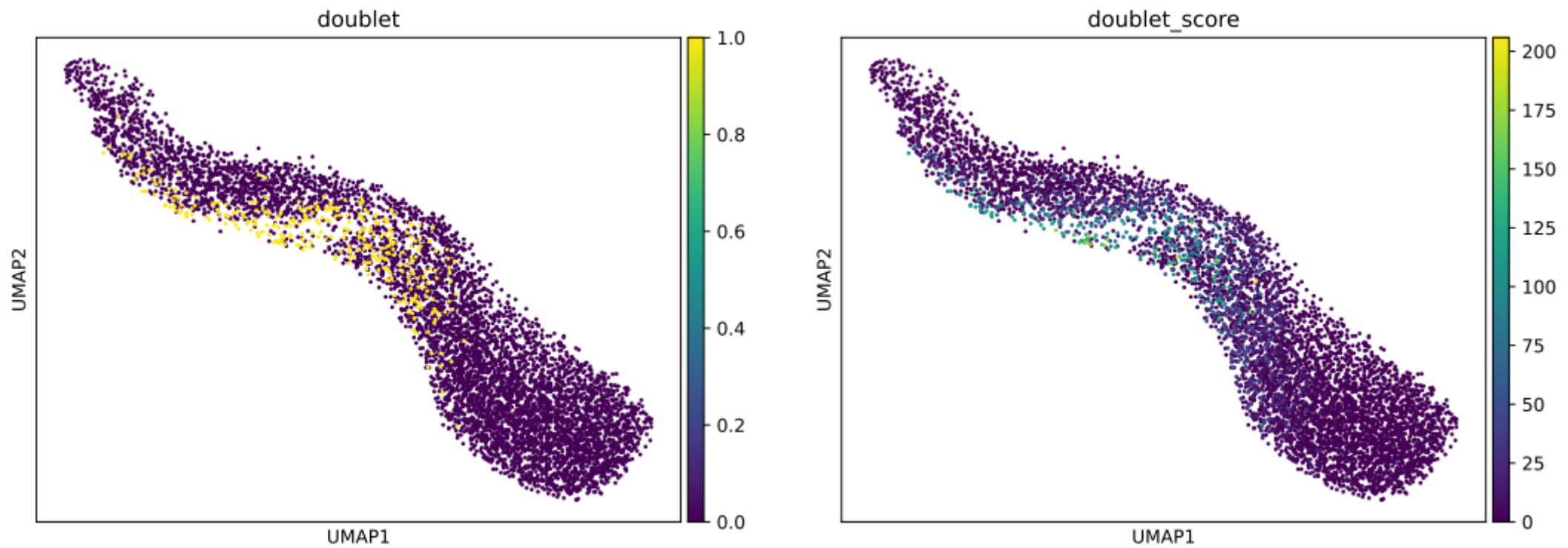
# Convergence test

Could this be the reason why so few cells are retained?

Because why so much fluctuation and even increase? In the predictions? Could it be false positive doublets?

```
doubletdetection.BoostClassifier  
clustering_algorithm="leiden"
```







# BM2

Pooled sample	Sample name	Sample details		Hashtag used	Target recovery	Cells loaded
BM_2	C-3	Control	H1	24000	30000	
	C-4	Control	H2			
	CVS-3	Stressed	H3			
	CVS-4	Stressed	H4			

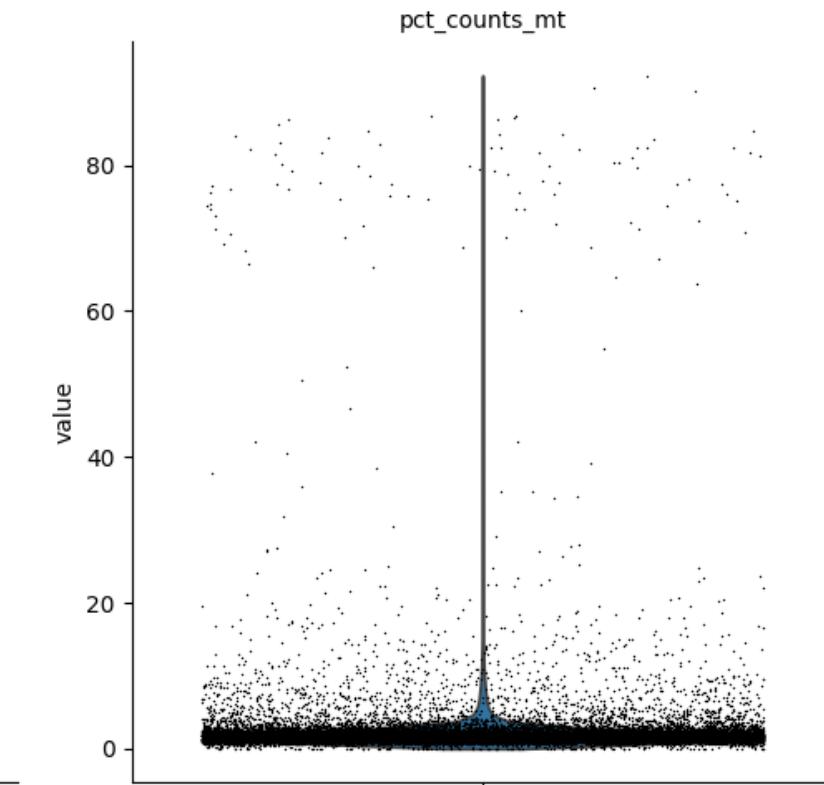
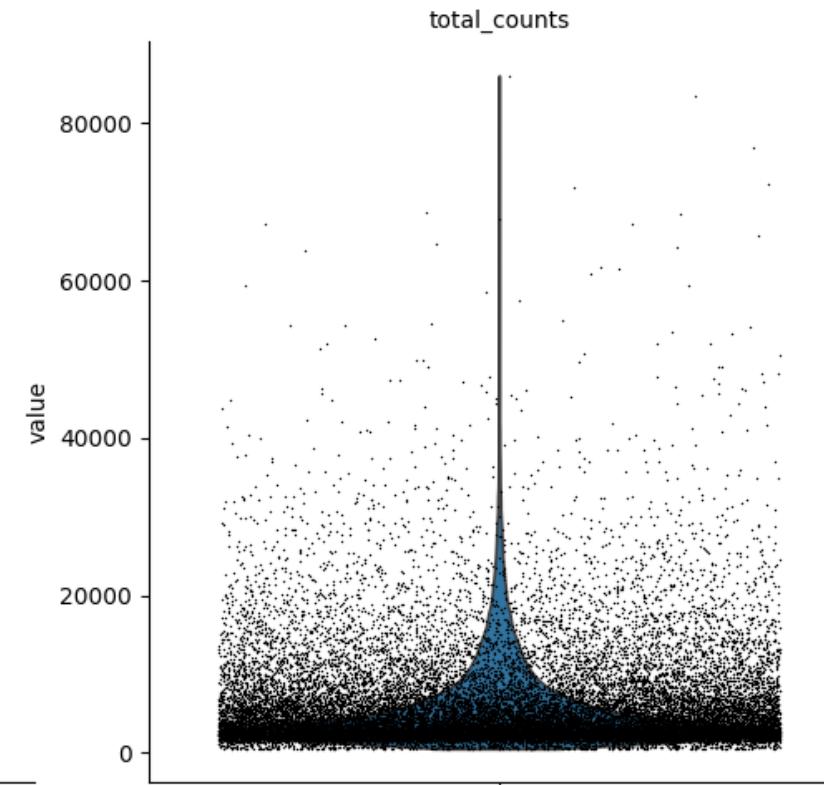
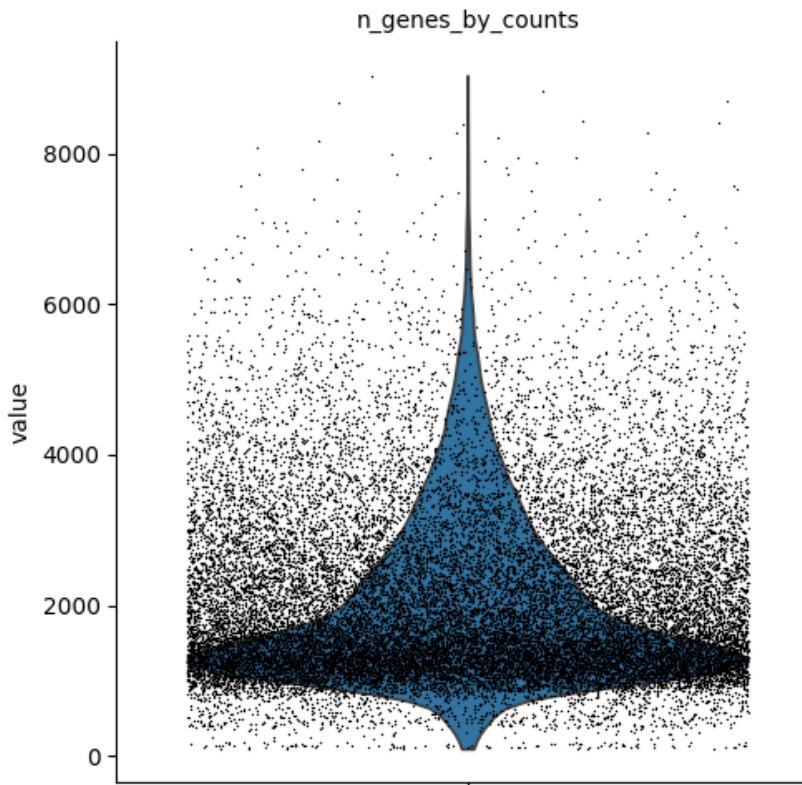
## Cell Calling Quality

Cells	Confidently mapped reads in cells	Median genes per cell	Median UMI counts per cell	Total genes detected
27,243	93.5%	1,687	3,640	24,999





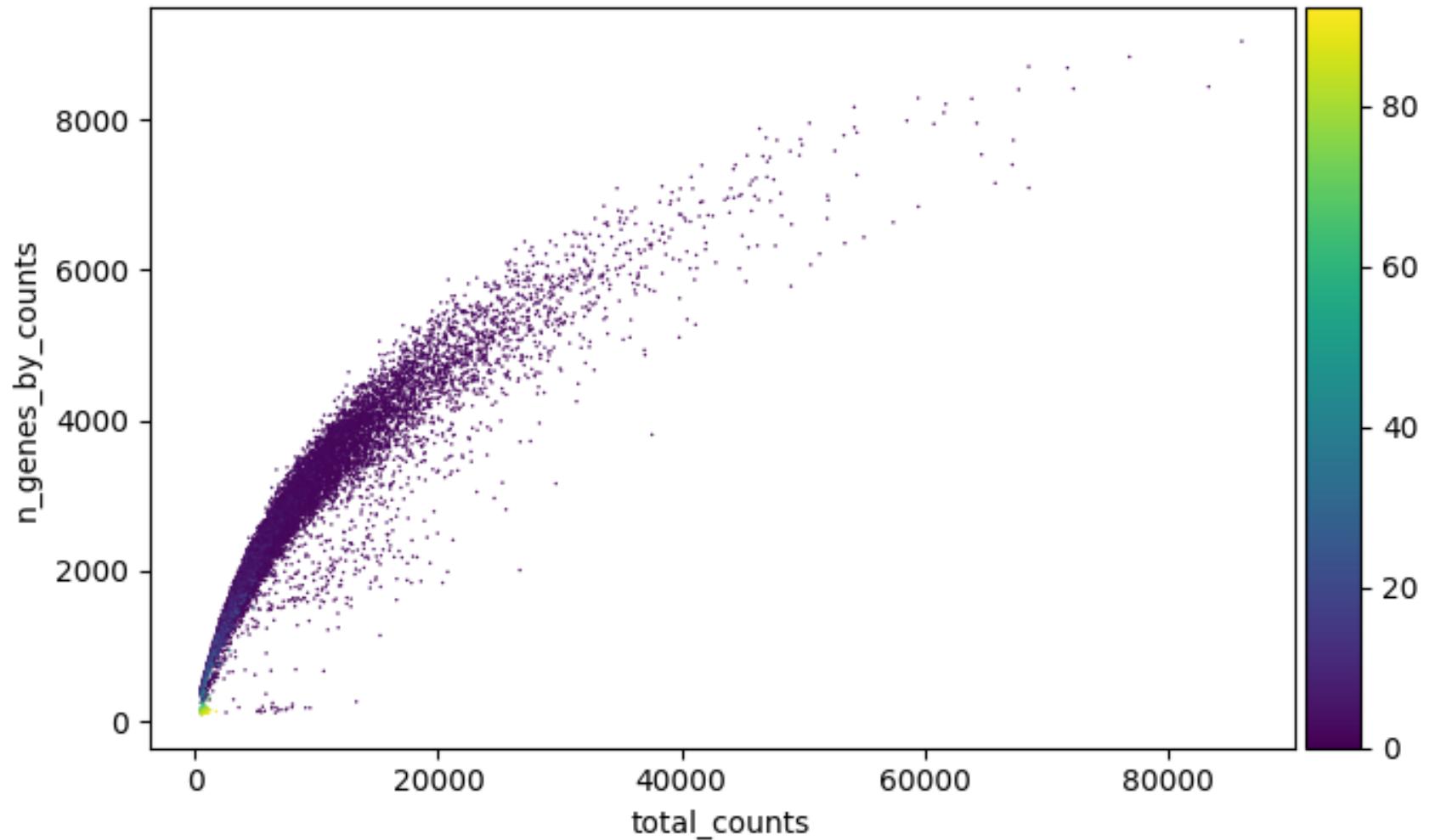
# BM2





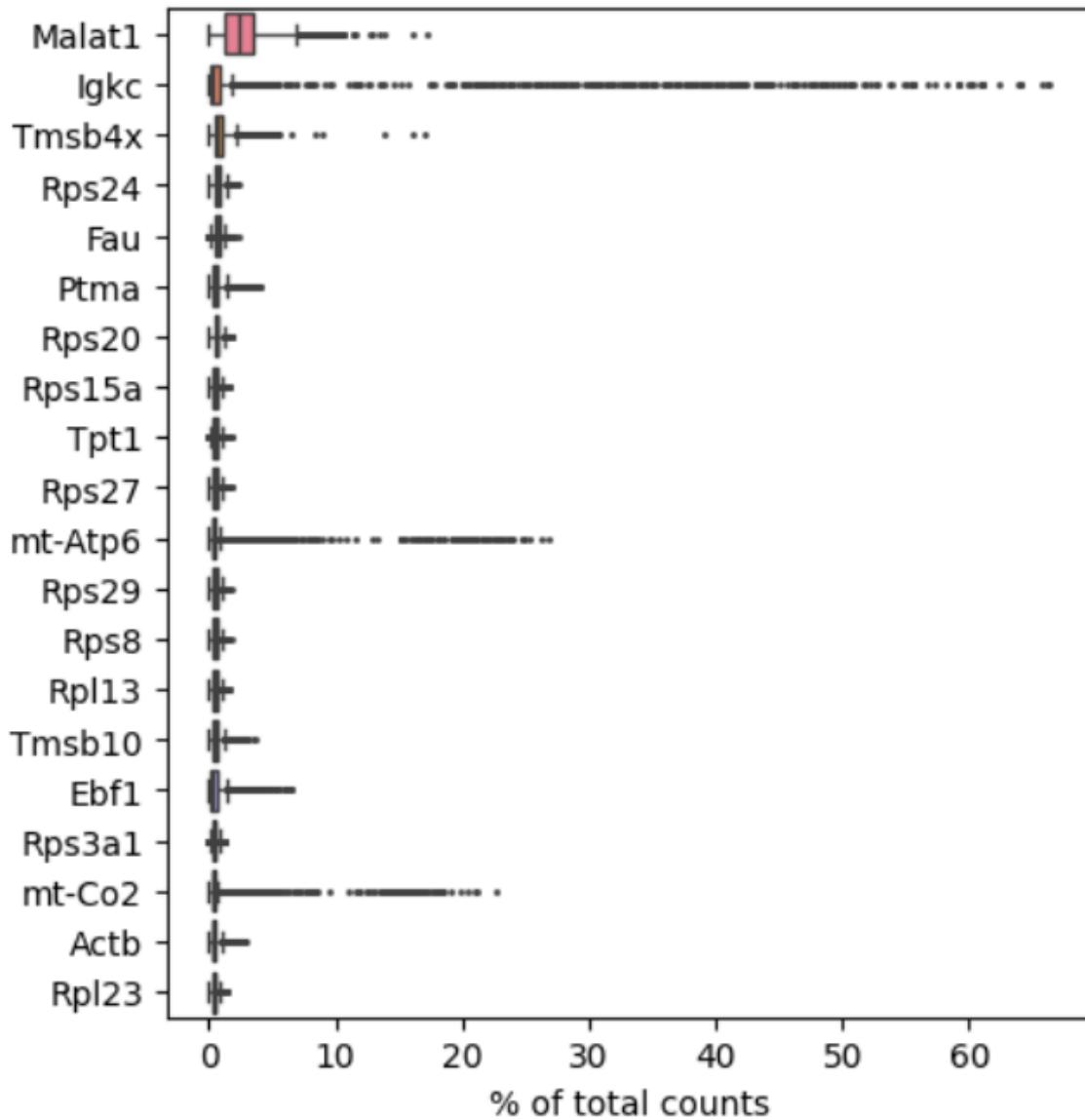
BM2

pct counts mt





# BM2





# GT1

GT_1	C-1	Control	H1	12000	10000
	G-1	Ghrelin Tx	H4		10000

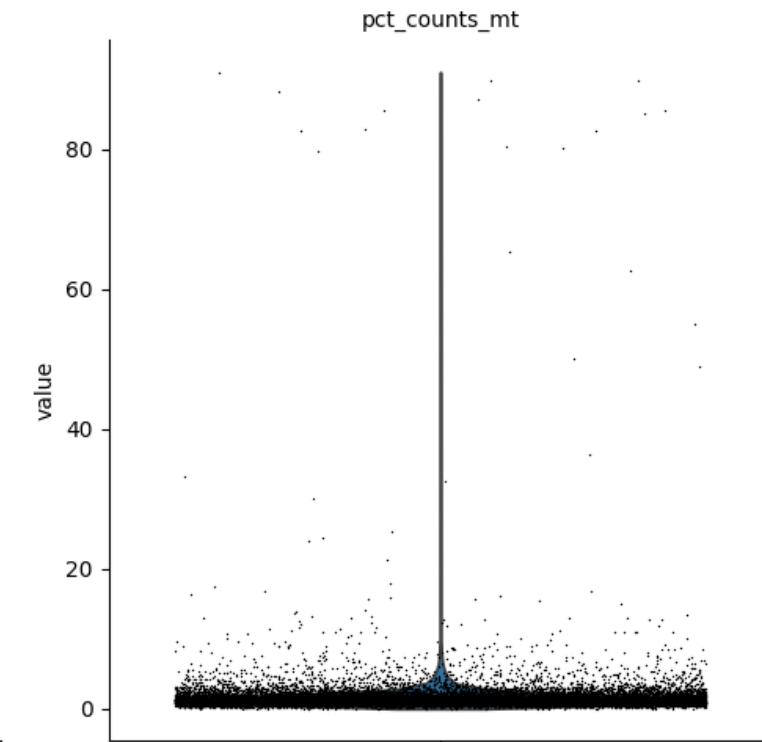
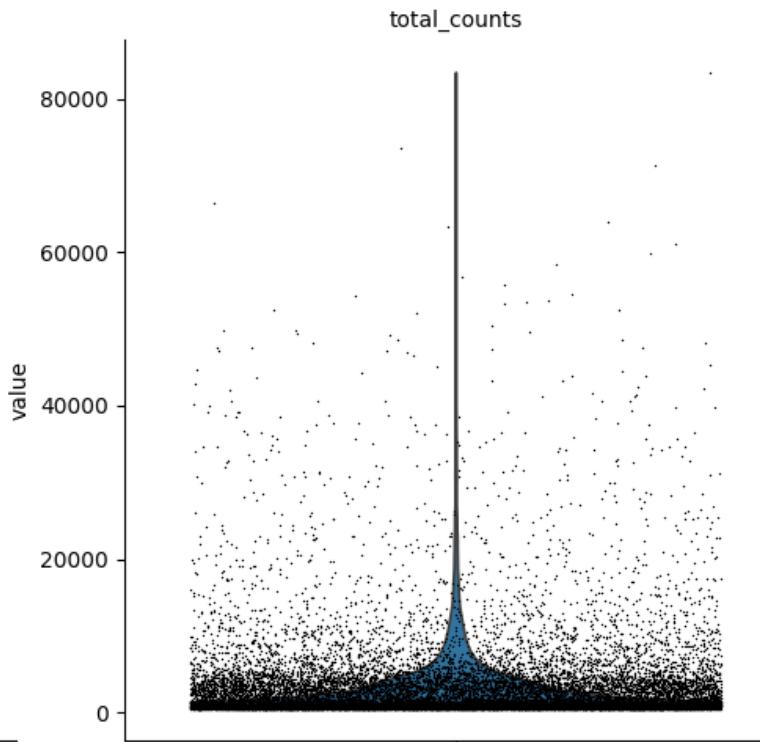
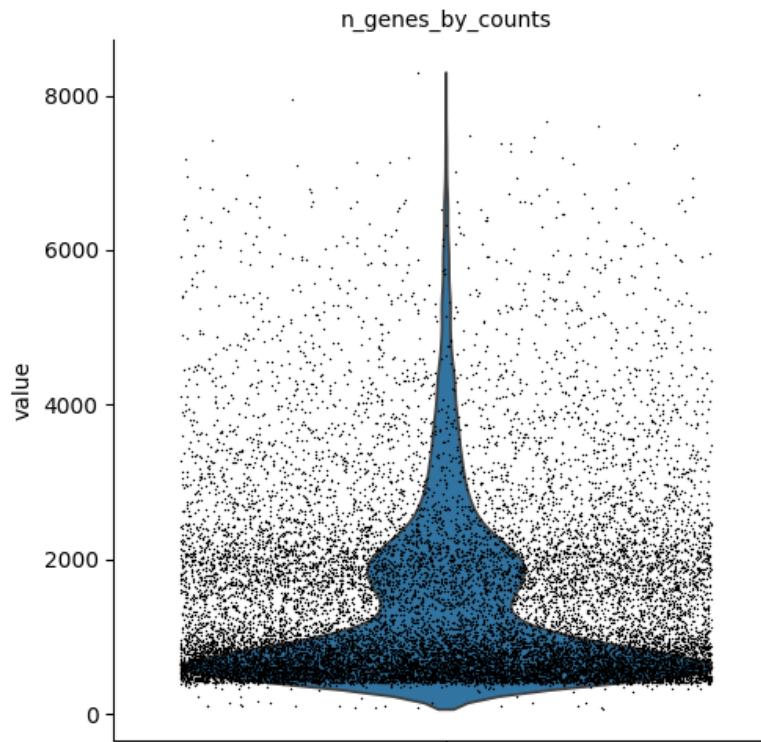
## Cell Calling Quality

Cells	Confidently mapped reads in cells	Median genes per cell	Median UMI counts per cell	Total genes detected
18,006	72.9%	937	1,448	23,193



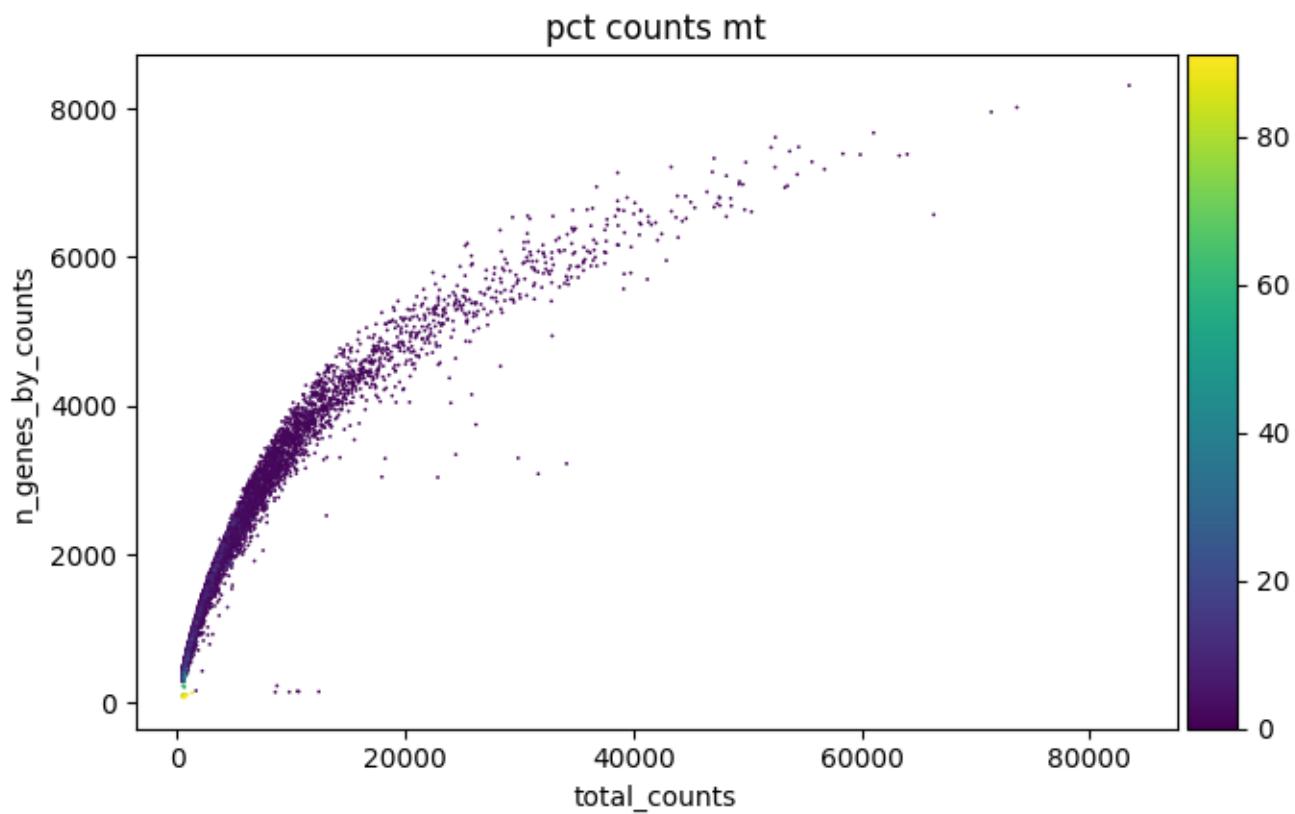


# GT1





GT1





# GT2

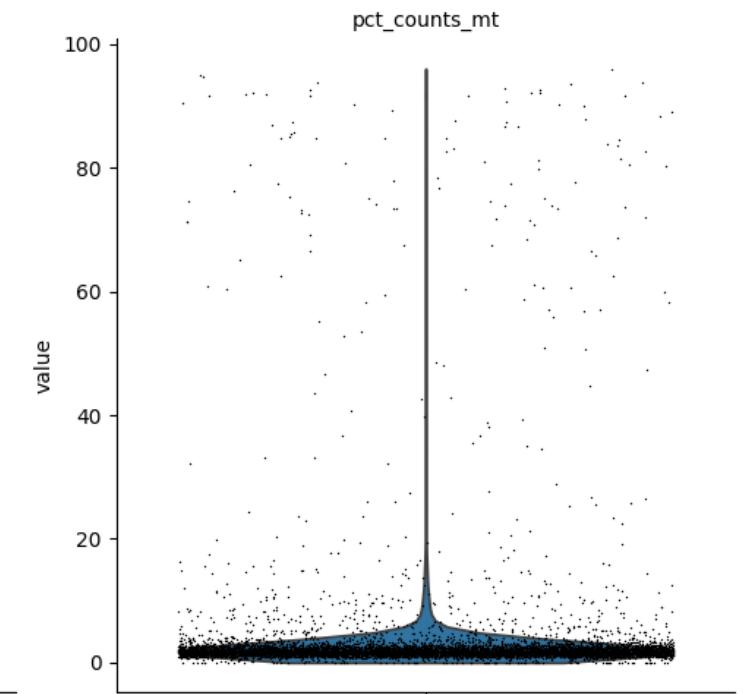
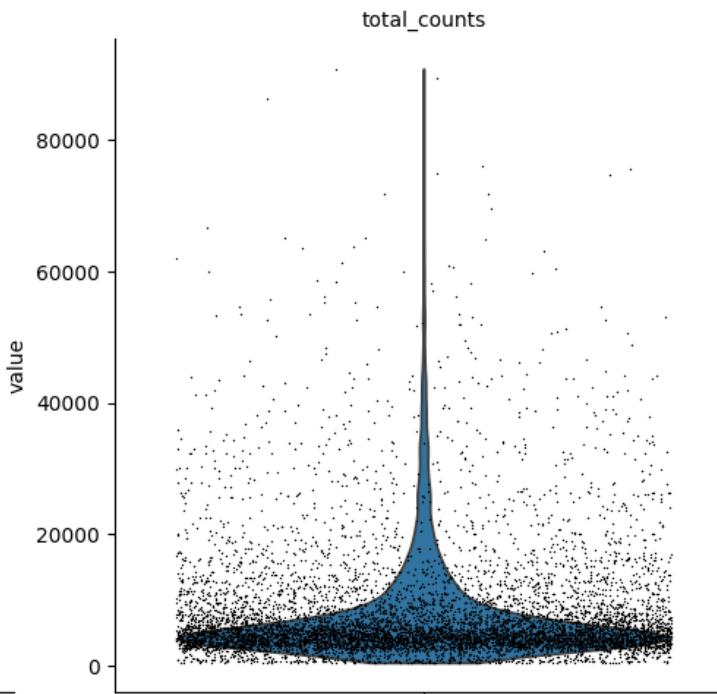
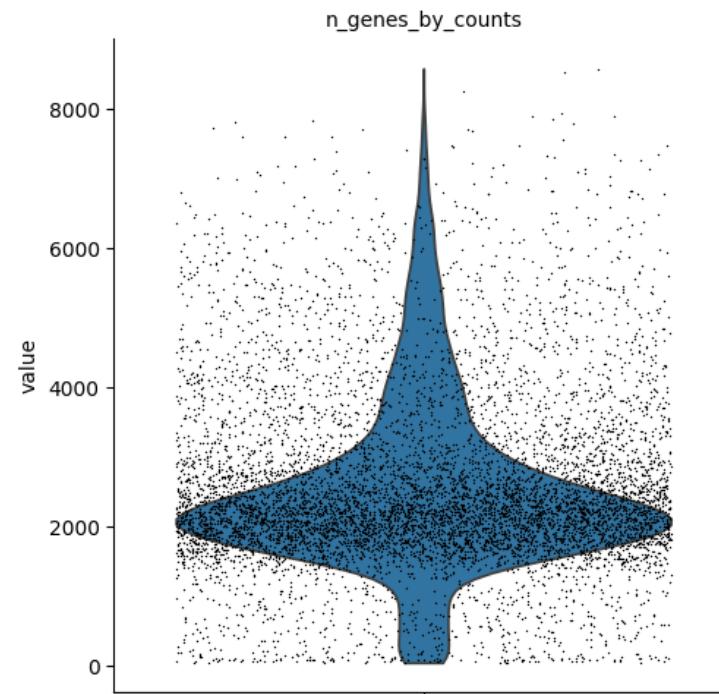
GT_2	C-2	Control	H2	12000	10000
	G-2	Ghrelin Tx	H5		10000

## Cell Calling Quality

Cells	Confidently mapped reads in cells	Median genes per cell	Median UMI counts per cell	Total genes detected
18,006	72.9%	937	1,448	23,193

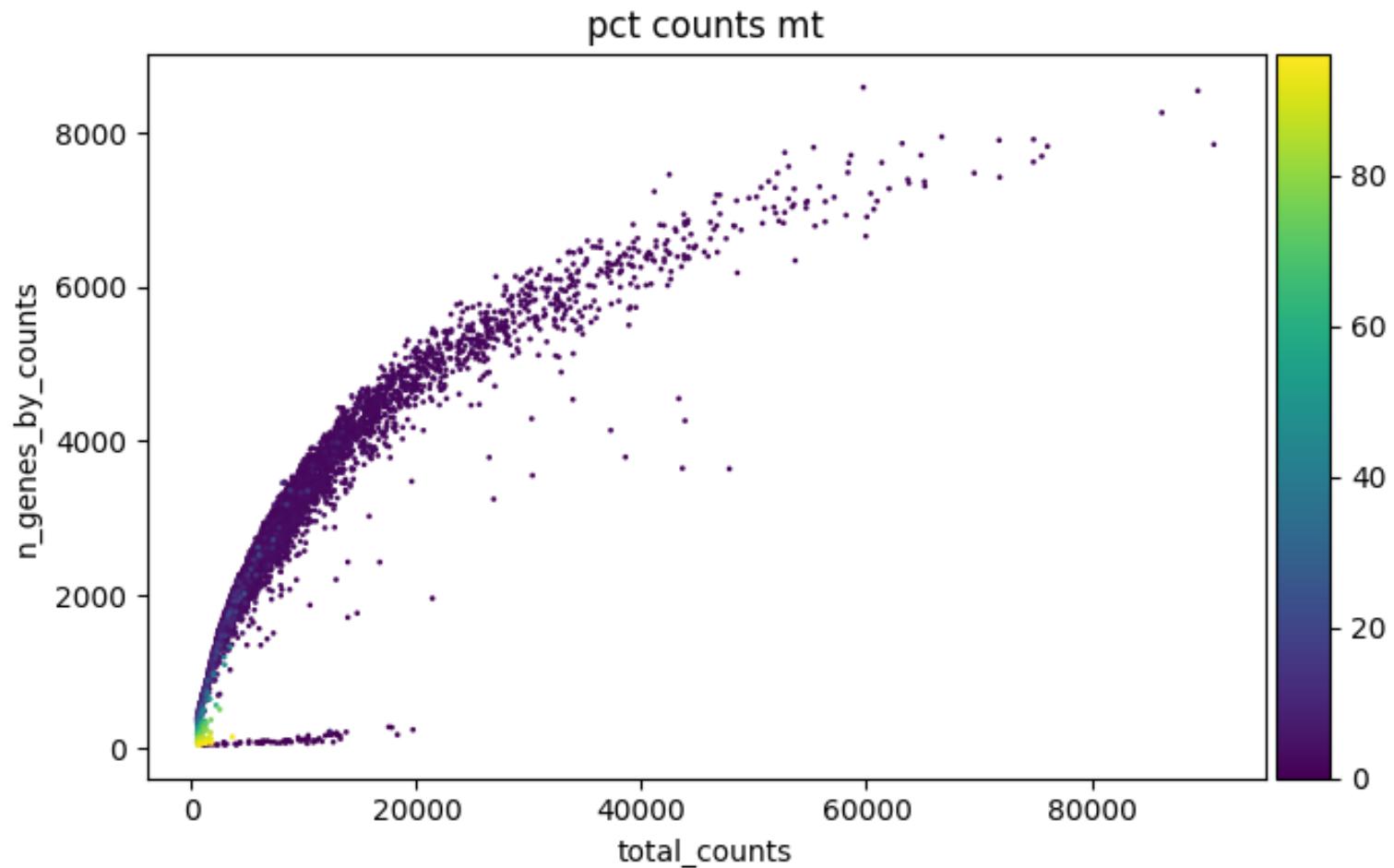


# GT2





GT2





# GT3

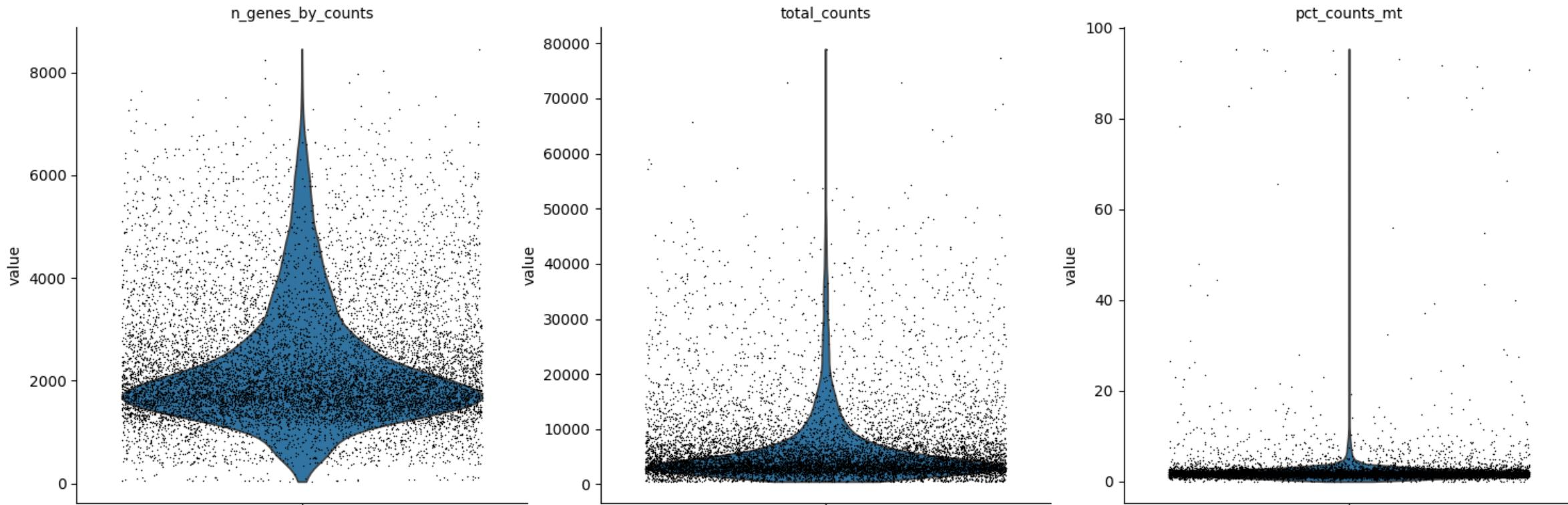
GT_3	C-3	Control	H3	12000	10000
	G-3	Ghrelin Tx	H6		10000

## Cell Calling Quality

Cells	Confidently mapped reads in cells	Median genes per cell	Median UMI counts per cell	Total genes detected
11,413	92.7%	1,942	3,940	23,443

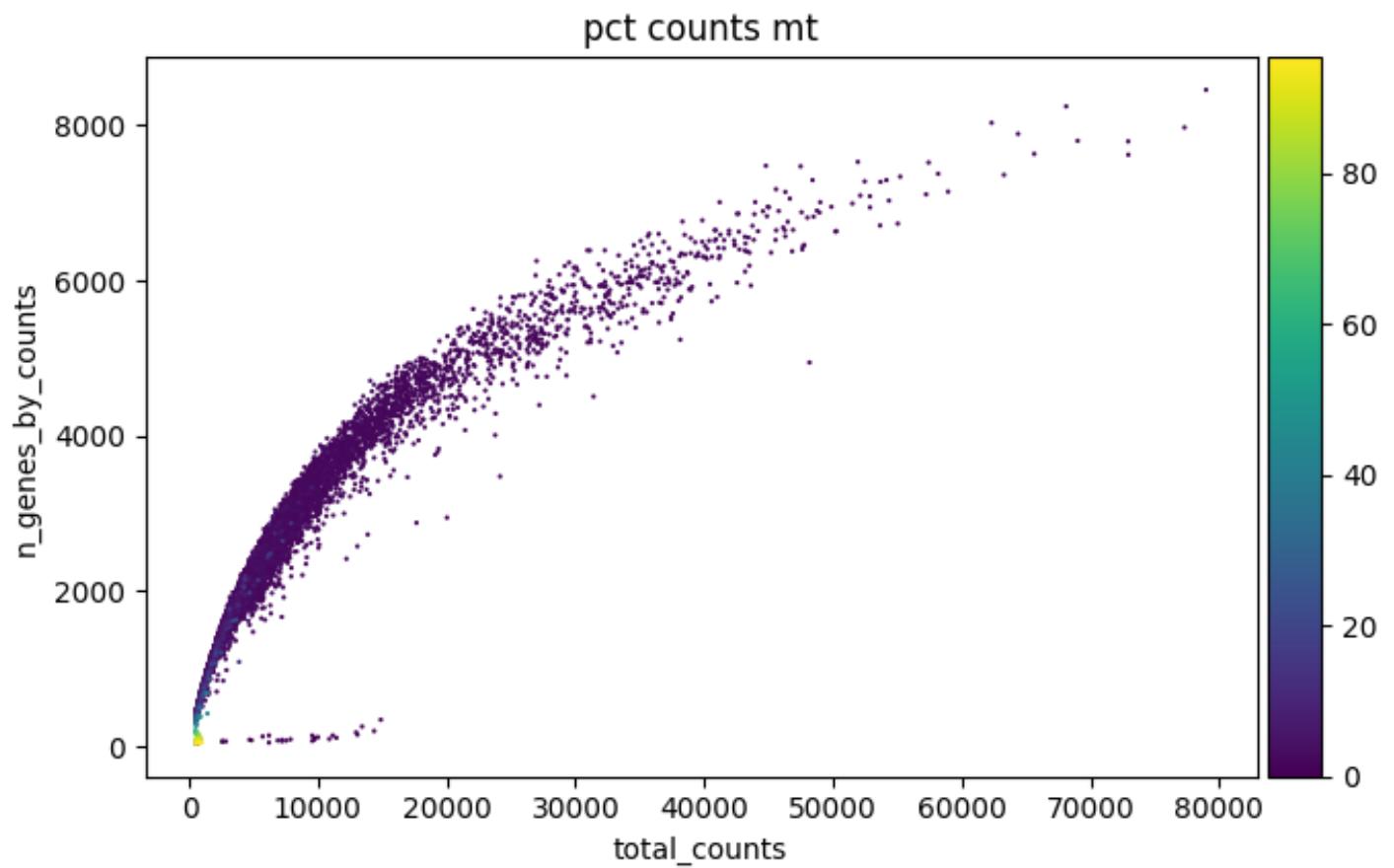


# GT3





GT3



```
mt_genes = adata_filtered.var_names[adata_filtered.var_names.str.lower().str.startswith("mt-")]
print(mt_genes)
```

```
Index(['mt-Nd1', 'mt-Nd2', 'mt-Co1', 'mt-Co2', 'mt-Atp8', 'mt-Atp6', 'mt-Co3',
       'mt-Nd3', 'mt-Nd4l', 'mt-Nd4', 'mt-Nd5', 'mt-Nd6', 'mt-Cytb'],
      dtype='object')
```

# Inputs for “multi” are:

- **FASTQ files**: *Required*. Described in the [count](#) section.
- **Reference transcriptome**: *Required*. Described in the [count](#) section. A transcriptome reference is optional for:
  - Flex Gene Expression and Antibody Capture libraries
  - Antibody Capture-only analysis (applicable to 3', 5', and Flex)
- **multi config CSV**: *Required*. A configuration file in CSV format that specifies all the parameters required to analyze Single Cell Gene Expression, V(D)J, Feature Barcode, and Flex libraries using the cellranger multi pipeline. The multi config CSV is required only for the cellranger multi pipeline and is not necessary for other Cell Ranger pipelines.
- **Feature Reference CSV**: *Optional*. Only required if a Feature Barcode library is included in your analysis. Described in the [count](#) section.

# Multi config CSV

[reference] Required. Absolute path to folder containing 10x Genomics-compatible genome reference. Optional for Flex Gene Expression and Antibody Capture libraries .

[create-bam] required

Multi config CSV for BM1:

[gene-expression]

reference,/gpfs/helios/home/rostamne/CSV\_SplVacc\_BM\_\_Ghr\_Thy\_Jul\_24\_scRNA/references/refdata-gex-GRCm39-2024-A  
create-bam,true

[feature]

reference,/gpfs/helios/home/rostamne/CSV\_SplVacc\_BM\_\_Ghr\_Thy\_Jul\_24\_scRNA/code/BM1/feature\_ref.csv

[libraries]

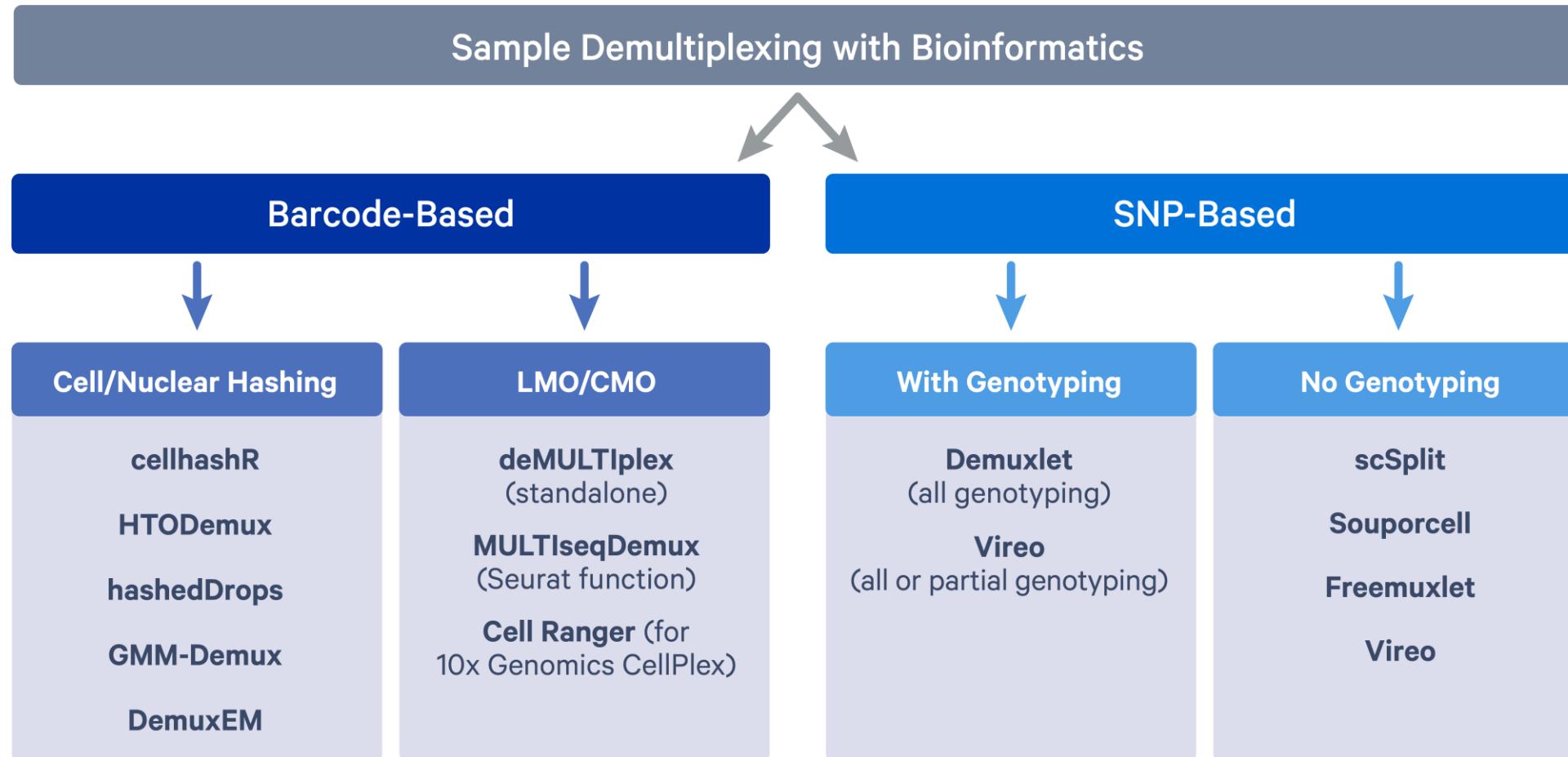
fastq\_id,fastqs,feature\_types

Antibody capture library? Optional?

# How the cellranger job was run? The parameters:

- `#!/bin/bash`
- `#SBATCH -J CR_BM1 #job name`
- `#SBATCH --partition=amd #Partition/queue to run job #GPU`
- `#SBATCH -t 6:00:00 #max time 3 hours`
- `#SBATCH --cpus-per-task=32 #nr os cpu requested`
- `#SBATCH --mem=256GB`
- `CONFIG_CSV=/gpfs/helios/home/rostamne/CSV_SplVacc_BM__Ghr_Thy_Jul_24_scRNA/code/BM1/config.csv`
- `cellranger multi \`
- `--id=GEX_BM1_multi \`
- `--csv=$CONFIG_CSV \`
- `--localcores=32 \`
- `--localmem=256`

# Why HTODemux?



# What is SEURAT OBJECT

Cell ID	nFeature	nCount	Percent_mito	Percent_ribo	...	droplet_status	Cell_type
Cell_1	18500	1000	17	2		singlet	99
Cell_2	16070	700	12	0		singlet	95
Cell_3	17780	980	5	1		singlet	92
...	18000	1600	25	5		doublet	89
Cell_50000	17070	2400	7	10		singlet	100

GRAPH DATA

Cell ID	Cell_1	Cell_2	Cell_3	...
Cell_1	1	1	1	1
Cell_2	1	1	1	1
Cell_3	1	1	1	1
...	0	0	1	1
Cell_50000	1	1	0	0

GENE METADATA

Gene ID	Avg_expression	...	variance
TP53	5		0.89
EGFR	3.5		1.2
VEGFA	0.78		2.2
...	3		0.76
BRCA1	1.2		2.4

CELL  
METADATA

DIMENSIONALITY REDUCTION DATA

Gene ID	PCA_1	PCA_2	...	UMAP_1	UMAP_2
TP53	5	5		0.89	0.89
EGFR	3.5	3.5		1.2	1.2
VEGFA	0.78	0.78		2.2	2.2
...	3	3		0.76	0.76
BRCA1	1.2	1.2		2.4	2.4

COUNTS TABLE

Gene ID	Cell_1	Cell_2	Cell_3	Cell_4	Cell_5	Cell_6	...	Cell_50000
TP53	80	50	63	36	60	0		99
EGFR	90	70	80	50	60	0		99
VEGFA	100	TP53	2.0	-2.3	-0.4	-4.3	-0.9	-9.4
APOE	95	EGFR	3.0	1.2	0.4	0.1	0.0	4.7
IL6	65	VEGFA	6.0	-2.3	-0.4	-4.3	-0.9	-9.4
TGFBI	82	APOE	4.0	1.2	0.4	0.1	0.0	4.7
AKT1	71	IL6	0.0	1.2	0.4	0.1	0.0	4.7
MTHFR	70	TGFBI	2.0	1.2	0.4	0.1	0.0	4.7
...	...	AKT1	0.0	1.2	0.4	0.1	0.0	4.7
BRCA1	75	MTHFR	0.0	1.2	0.4	0.1	0.0	4.7
		BRCA1	1.0	1.2	0.4	0.1	0.0	4.7
Gene ID	Cell_1	Cell_2	Cell_3	Cell_4	Cell_5	Cell_6	...	Cell_50000
TP53	2.0	-2.3	-0.4	-4.3	-0.9	-9.4		4.7
EGFR	3.4	13.4	4.9	-4.9	6.9	-9.4		4.7
VEGFA	6.1	-5.3	4.1	-5.3	-4.3	-9.3		4.7
APOE	4.1	8.1	2.6	-5.4	-2.6	-8.7		4.7
IL6	0.1	4.1	1.4	-6.1	0.1	-8.0		4.7
TGFBI	2.3	3.7	1.7	-5.9	-4.3	-8.0		4.7
AKT1	0.7	8.3	2.1	-5.7	-4.1	-9.4		4.7
MTHFR	0.6	4.4	3.4	-6.0	3.6	-9.4		4.7
...	...	BRCA1	1.0	1.2	0.4	0.1	0.0	4.7
BRCA1	1.3	3.4	1.7	-6.4	10.9	-9.4		4.7

Raw counts

Normalised counts

Scaled counts

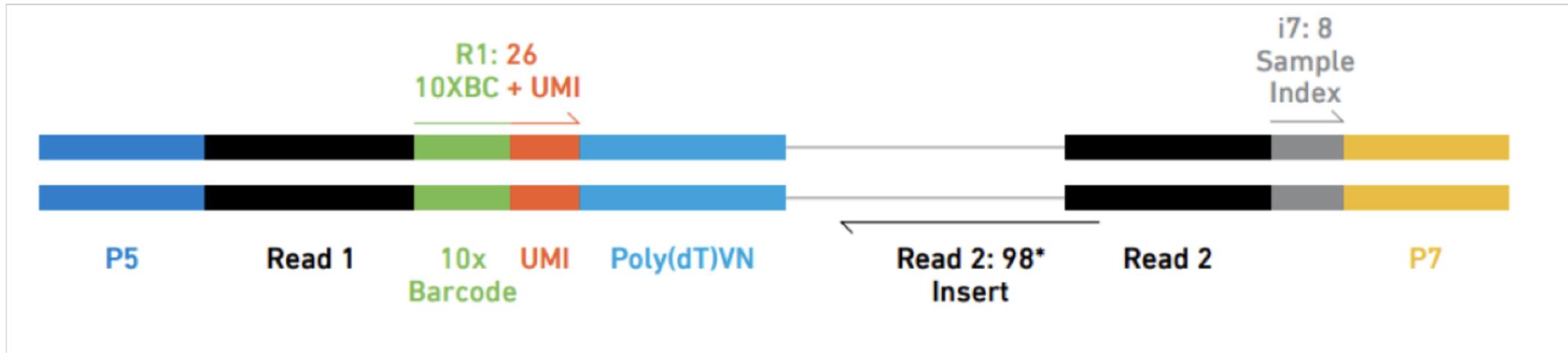
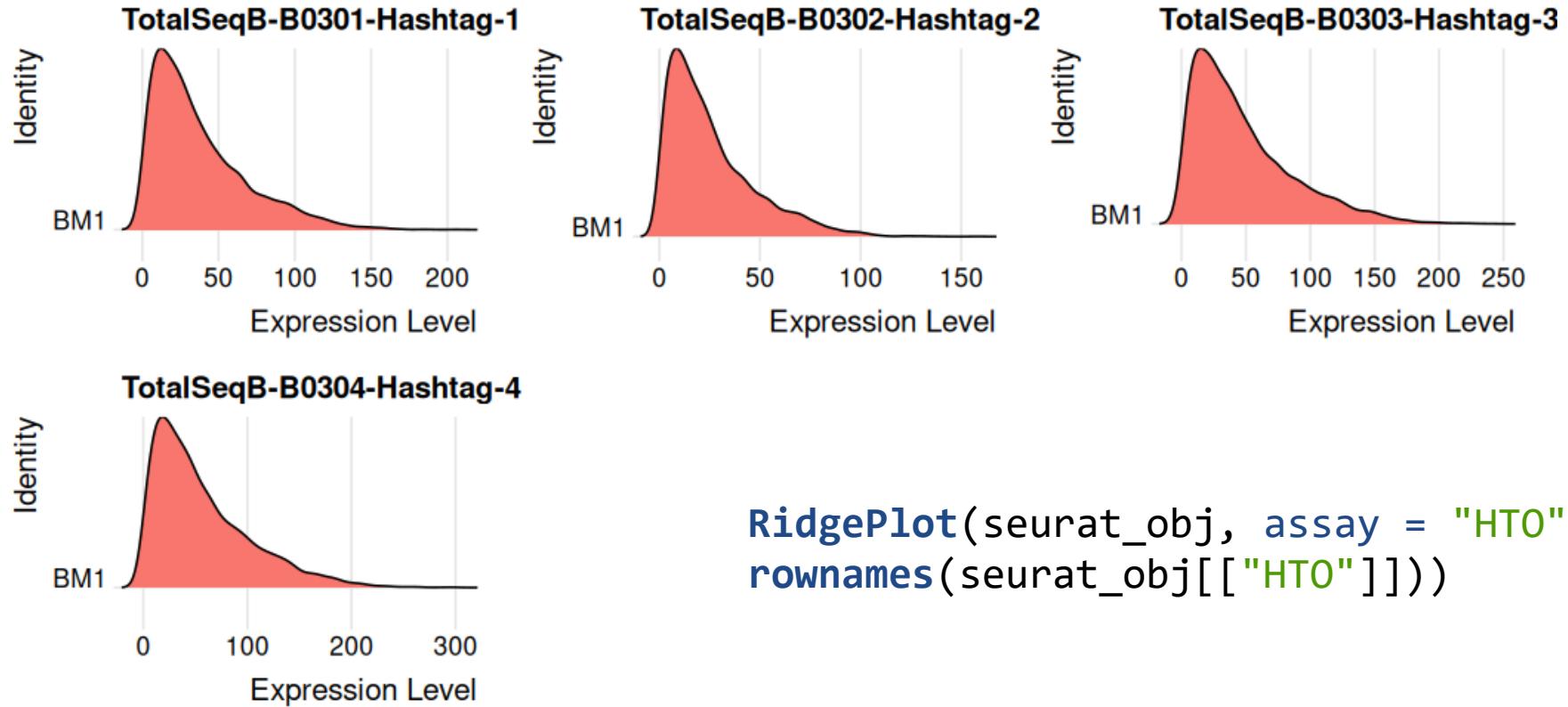


Fig. 2. Schematic of a fragment from a final Chromium™ Single Cell 3' v2 library. \*Can be adjusted.

barcodes.tsv.gz	← cell barcodes
features.tsv.gz	← features (genes or antibodies)
matrix.mtx.gz	← count matrix (UMI counts)

# Visualize demultiplexing results

## BM1



```
RidgePlot(seurat_obj, assay = "HTO", features =  
rownames(seurat_obj[["HTO"]]))
```

**distribution of HTO UMI counts per hashtag**

# BM1

```
DefaultAssay(seurat_obj) <- "HTO"
# Now fetch HTO counts
hto_values <- FetchData(
  seurat_obj,
  vars = c(
    "TotalSeqB-B0301-Hashtag-1",
    "TotalSeqB-B0302-Hashtag-2",
    "TotalSeqB-B0303-Hashtag-3",
    "TotalSeqB-B0304-Hashtag-4"
  )
)
head(hto_values)
```

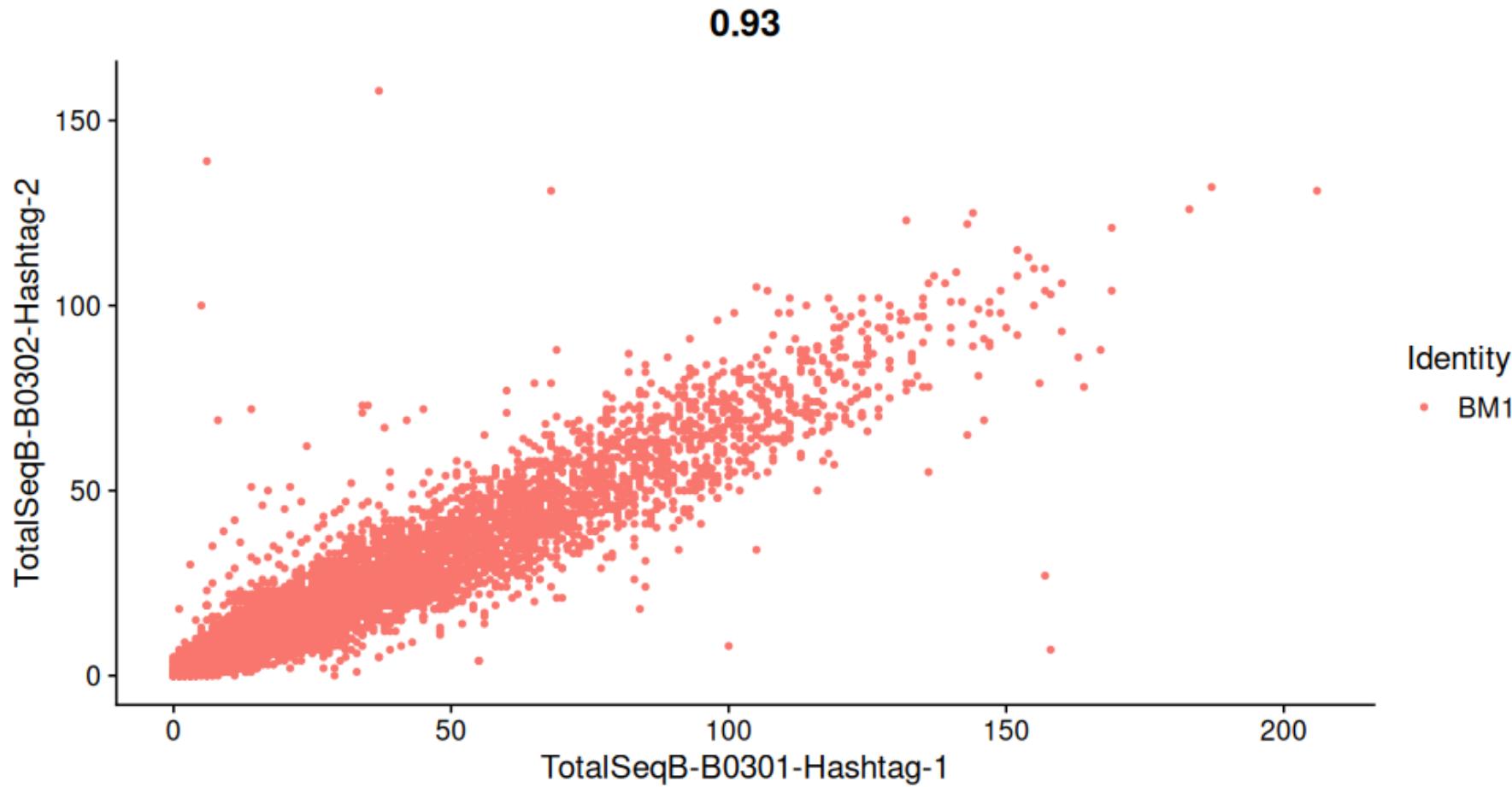
Description: df [6 x 4]

	TotalSeqB-B0301-Hashtag-1 <dbl>	TotalSeqB-B0302-Hashtag-2 <dbl>	TotalSeqB-B0303-Hashtag-3 <dbl>
AAACCCAAGCGGCTCT-1	0.2910151	0.4675008	0.3710491
AAACCCAAGCTGCCAC-1	0.7889174	0.7341150	0.8786568
AAACCCAAGGCCACCT-1	1.5185292	1.3948041	1.5377131
AAACCCACTAGCTGTT-1	0.8058257	0.5008858	0.4312012
AAACCCAGTCGTTCC-1	1.2243708	1.2190590	1.4145975
AAACGAACACCACTGG-1	1.2675534	1.2505793	1.2825090

6 rows | 1-4 of 4 columns

# Visualize demultiplexing results

## BM1



# Goals and events

- ✓ 4<sup>th</sup> Oct, Seurat was successfully installed (milestone)
- ✓ 10<sup>th</sup> Oct, Previous meeting (unsatisfactory progress)
- ✓ Literature review (Seurat/satja lab description, multiplexing, bone-healing... )
- ✓ Learning how to run and interpret the HTODemux workflow
- ✓ ~~Reading Bioconductor (not done)~~
- ✓ Running HTODemux
  
- ✓ (participating in Comp Bio conference – only as audience)

# Does chronic stress affect B cell development like aging? As in the bone-healing paper:

Checking for B cell markers in BM data

3 different mice (biological replicates)  
Young, Non-immune experienced (NE),  
Immune experienced (IE)

Pronounced impairment of B cell differentiation during bone regeneration in adult immune experienced mice

Mireille Ngokingha Tchouto<sup>1,2,3</sup>, Christian H. Bucher<sup>1,4</sup>,  
Ann-Kathrin Mess<sup>1,3</sup>, Simon Haas<sup>5,6</sup>, Katharina Schmidt-Bleek<sup>1,4</sup>,  
Georg N. Duda<sup>1,4</sup>, Dieter Beule<sup>2,7</sup> and Miha Milek<sup>2\*</sup>

---

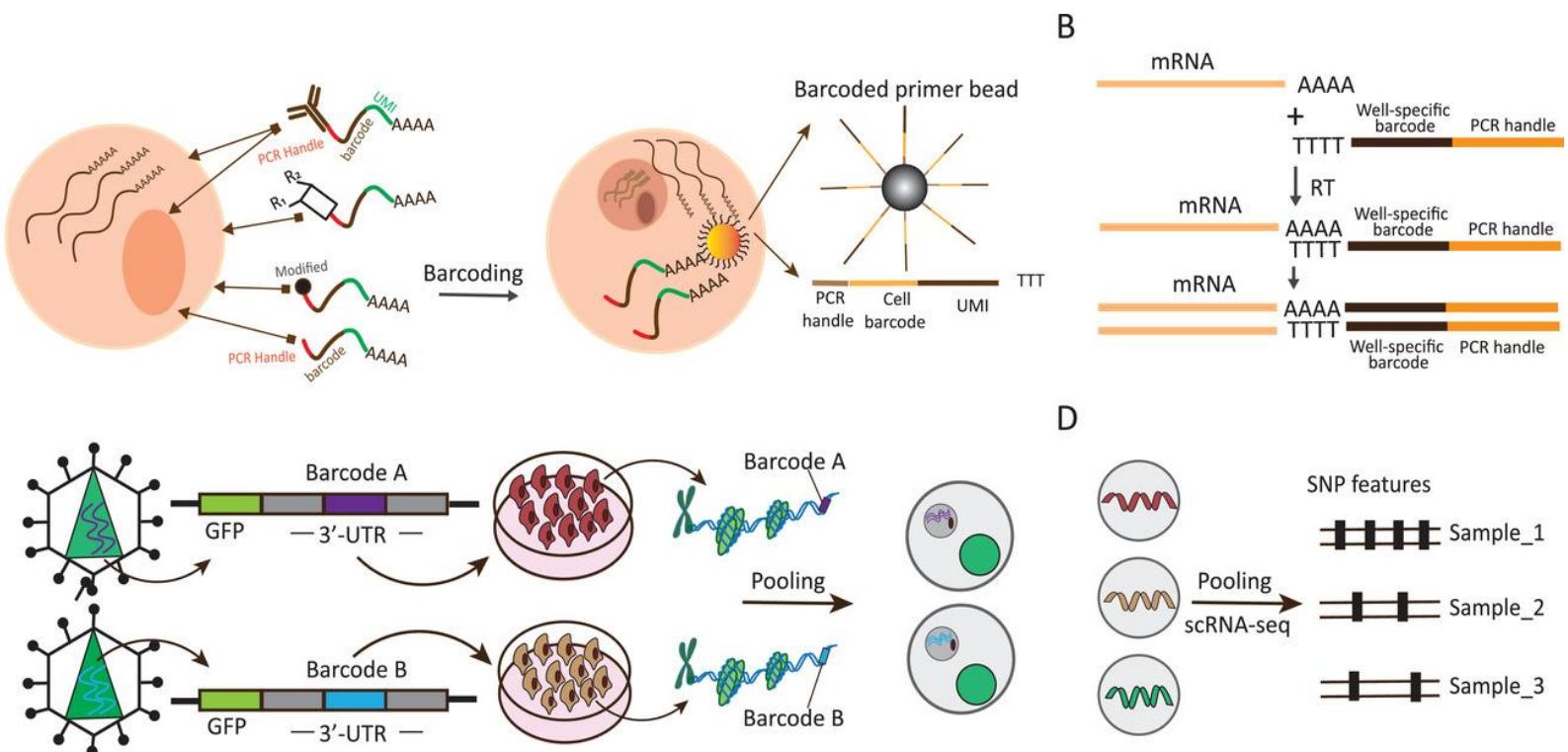
Chromium Next GEM Single Cell 3' Reagent Kits v3.1  
Feature Barcode technology for Cell Surface Protein  
Cellranger count v7.1.0 (10X Genomics)  
HTODemux function

# Brief literature review

Understanding different methods of multiplexing

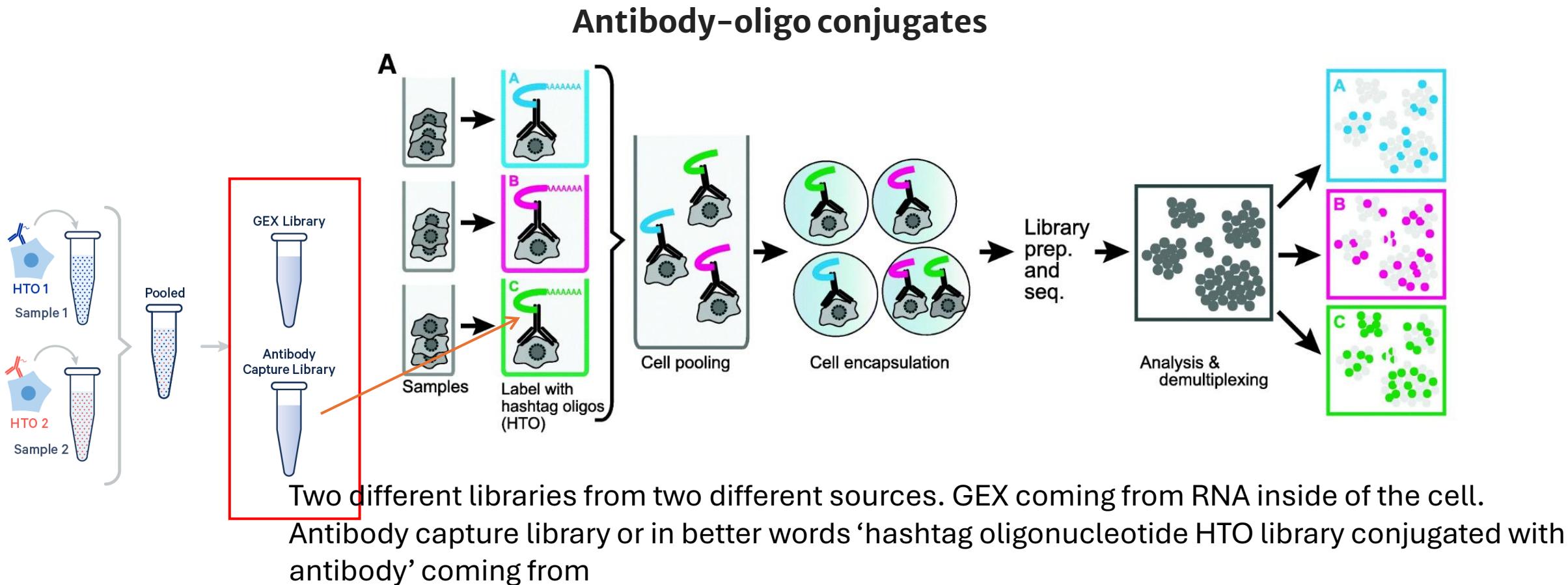
Multiplexing Methods for Simultaneous Large-Scale Transcriptomic Profiling of Samples at Single-Cell Resolution

8-plex (cell hashing)

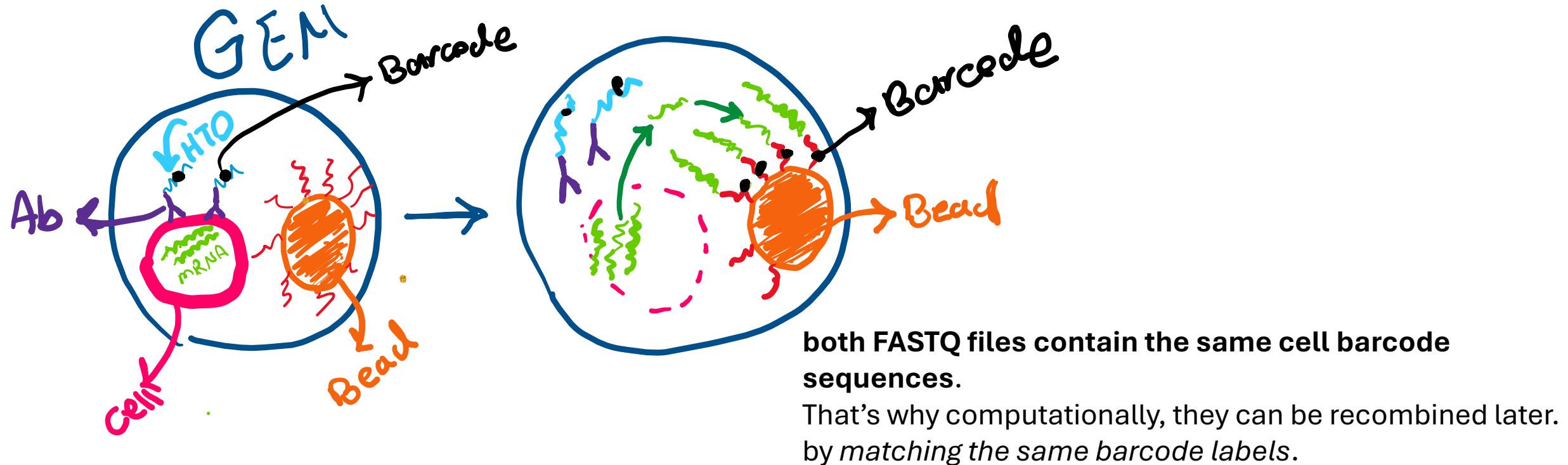


Advanced Science, Volume: 8, Issue: 17, First published: 08 July 2021, DOI: (10.1002/advs.202101229)

# Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics



# How can HTO lib and GEX lib get linked together? What do they have in common connecting them to one cell?



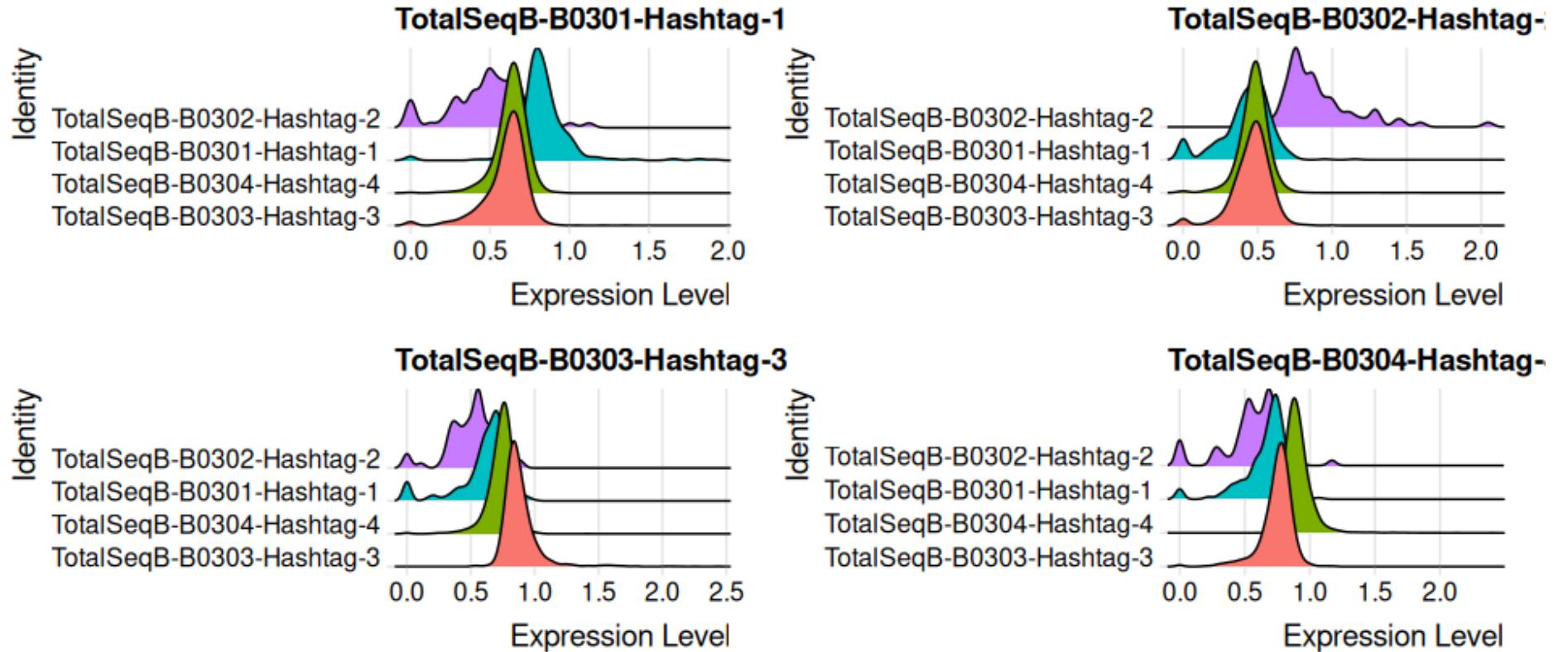
# Classification of barcodes based on HTO levels

- Three things to understand :
- ✓ Normalization with the CLR (central log ratio) transformation

$$x_i' = \log \frac{x_i}{(\prod_{i=1}^n x_i)^{\frac{1}{n}}}$$

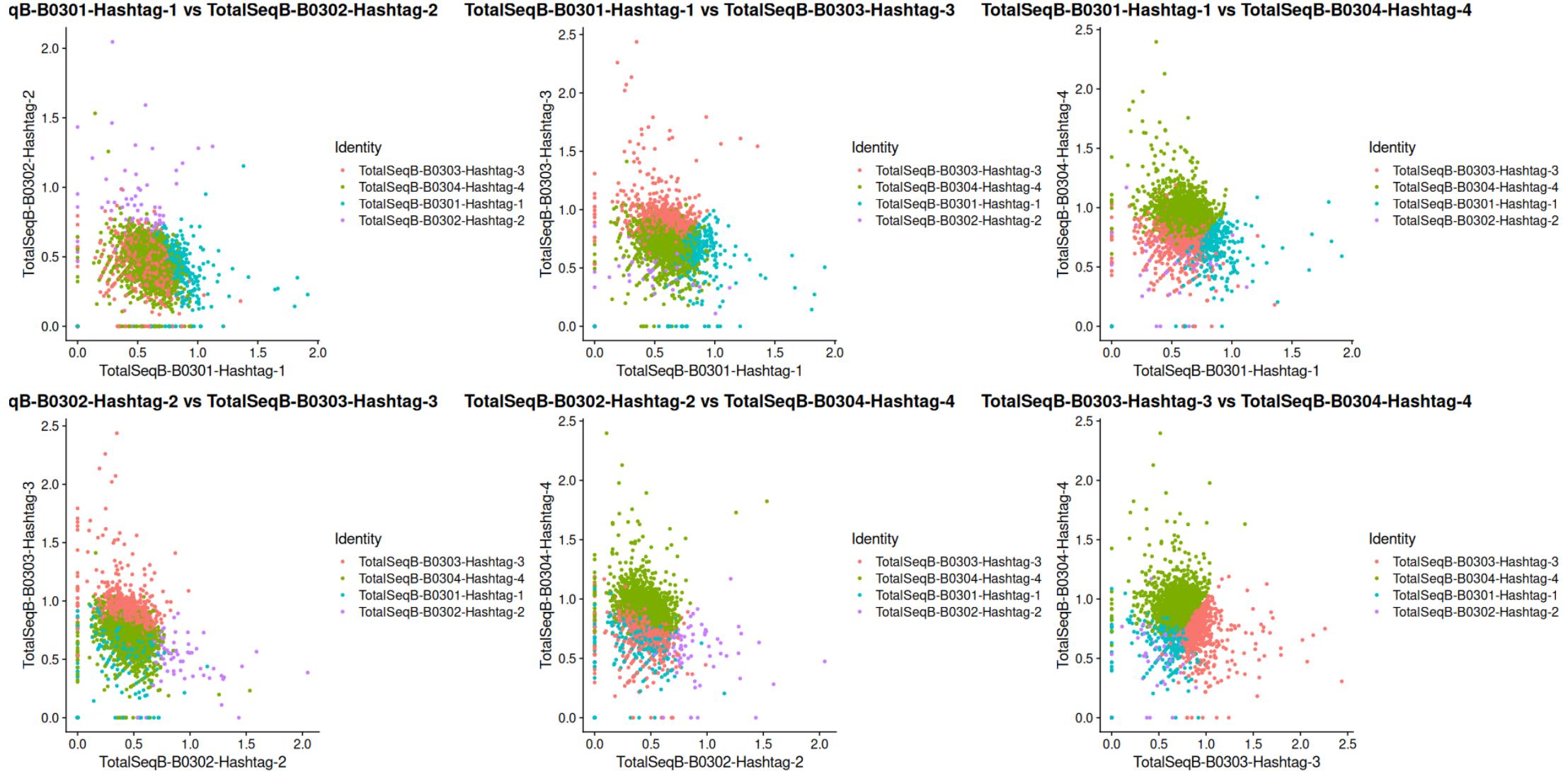
- ❑ Initial unsupervised clustering to find “likely negatives”  
**k-medoids clustering** (similar to k-means but more robust to outliers)
- ❑ Modeling background distribution for each HTO

# BM1, Ridge plot showing Barcode contaminations?

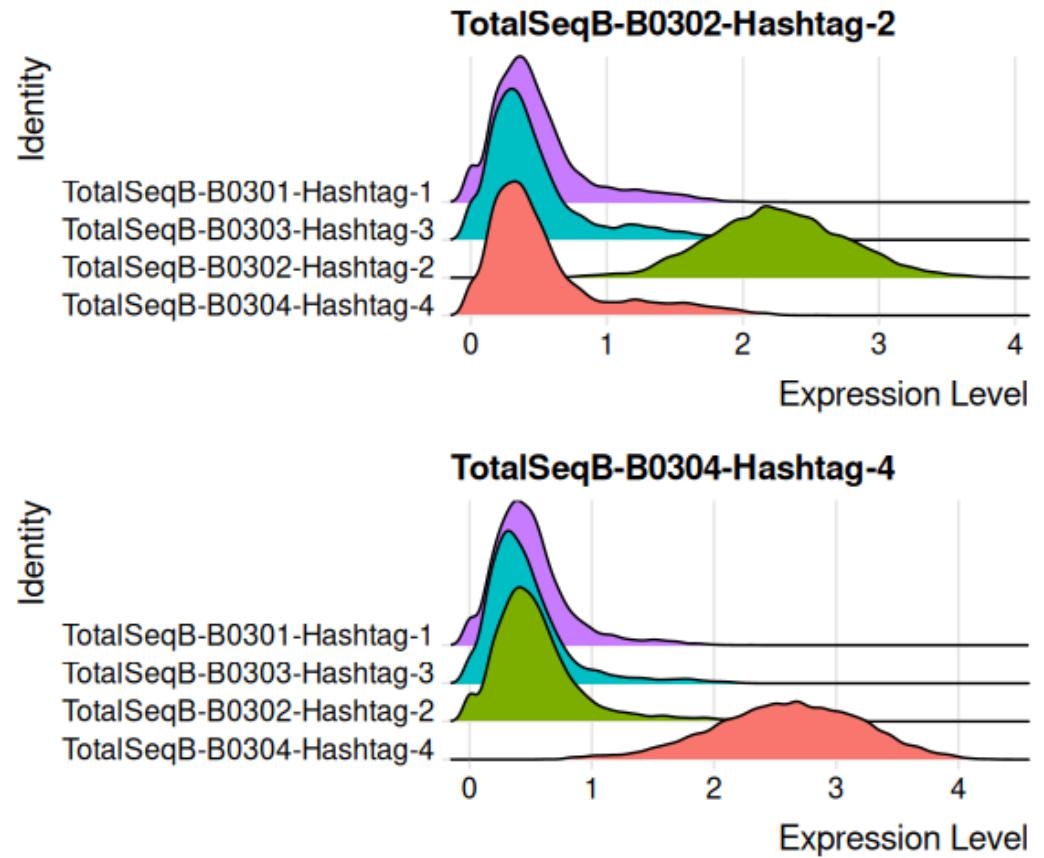
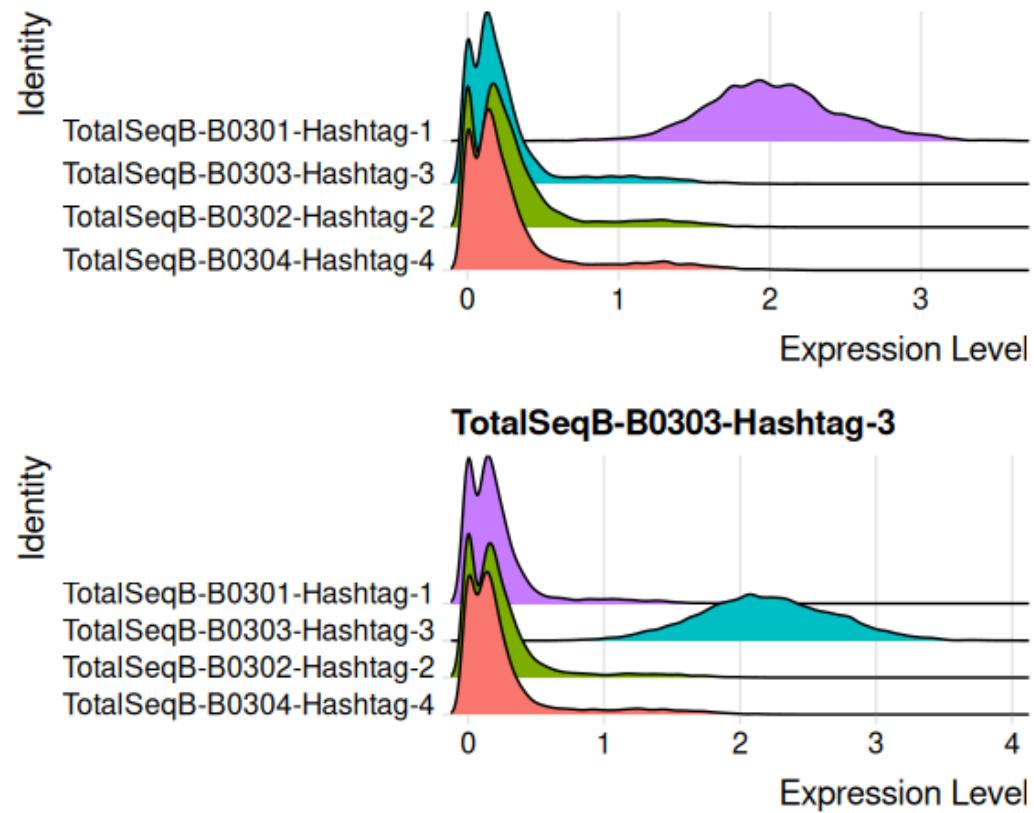


I would have expected to see peaks near 0 for all other hashtags. To show that on the cell population in which a specific HTO has been expressed the most other hashtags doesn't exist as much. But it doesn't look like it (???)

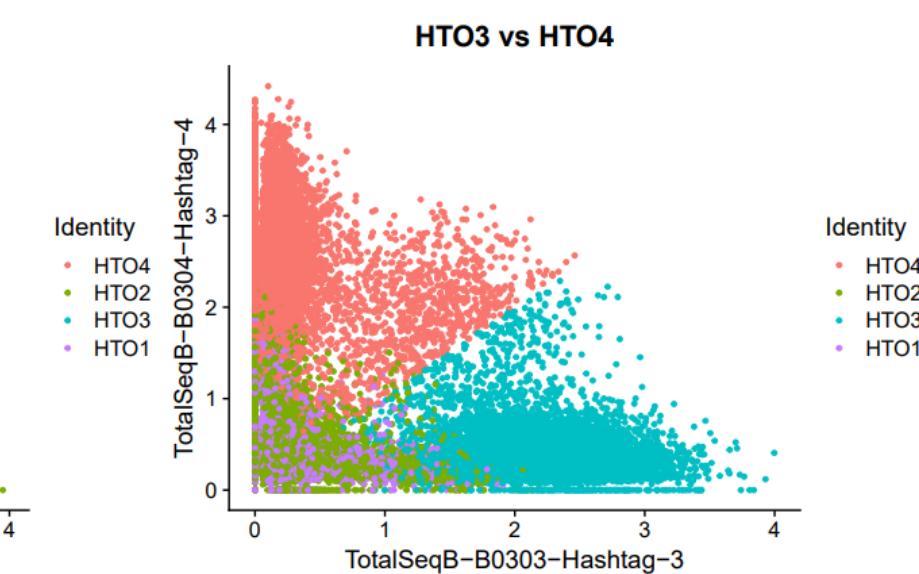
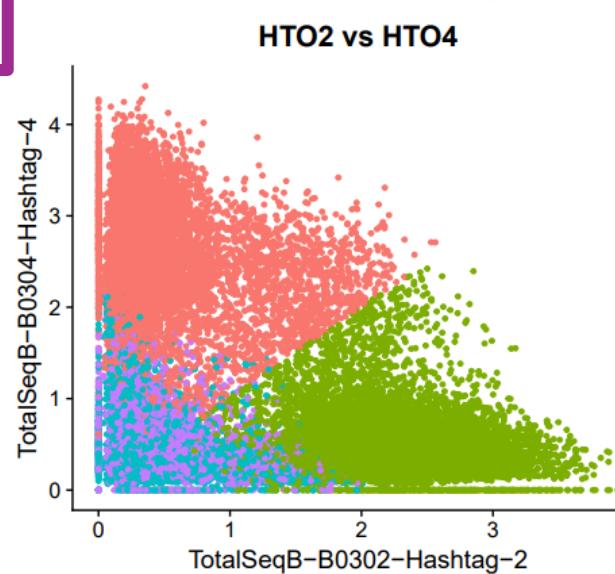
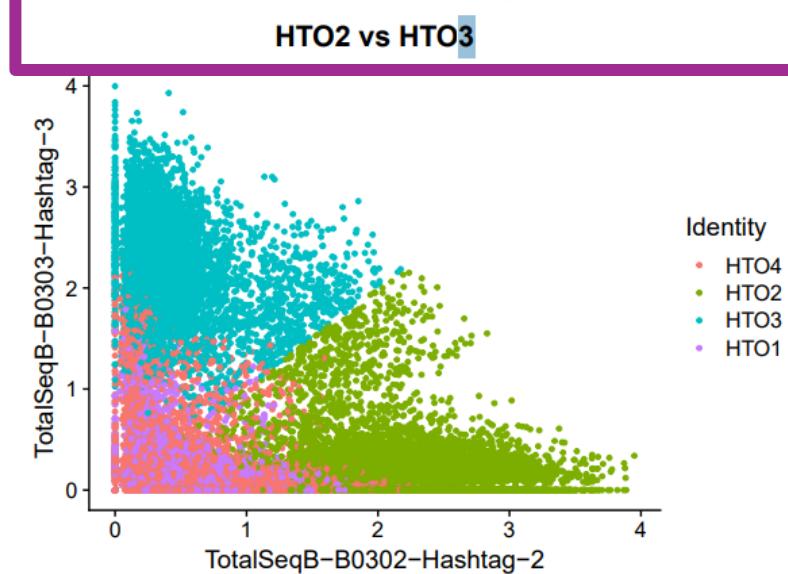
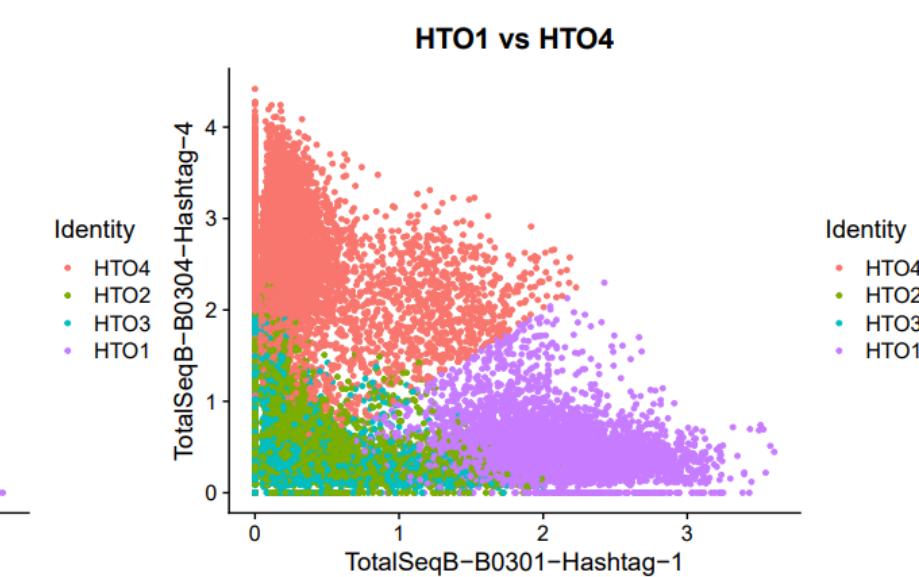
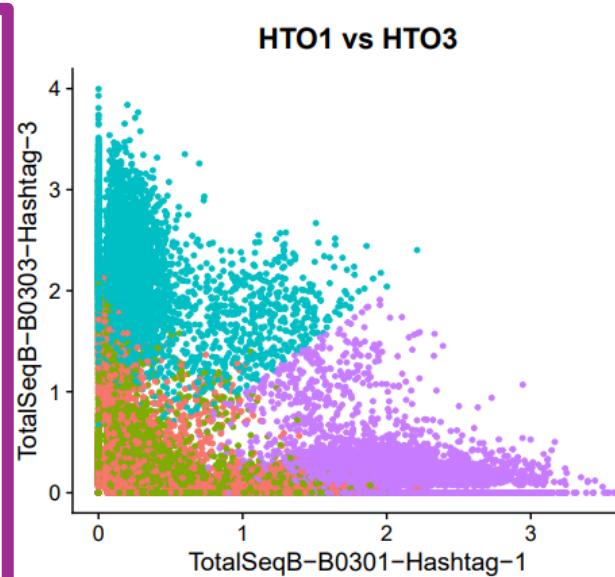
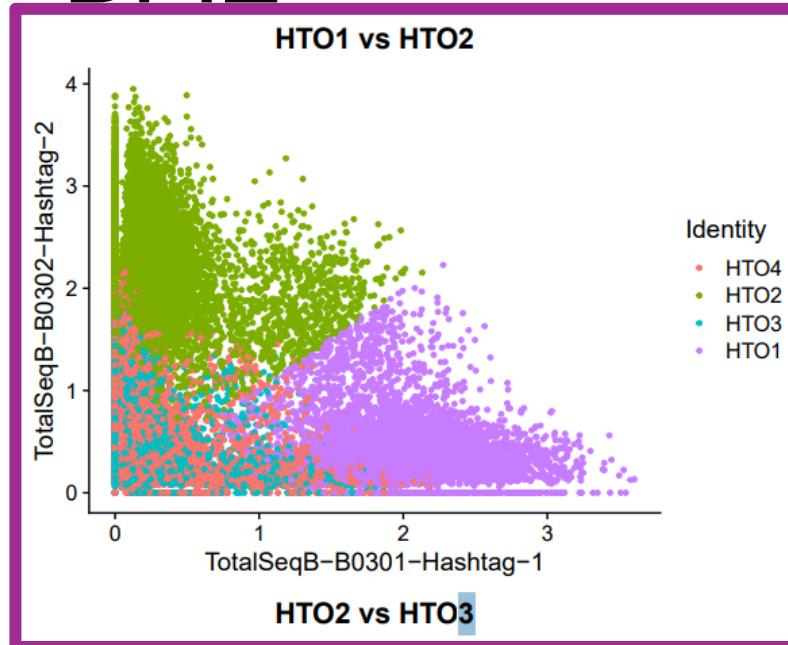
# BM1

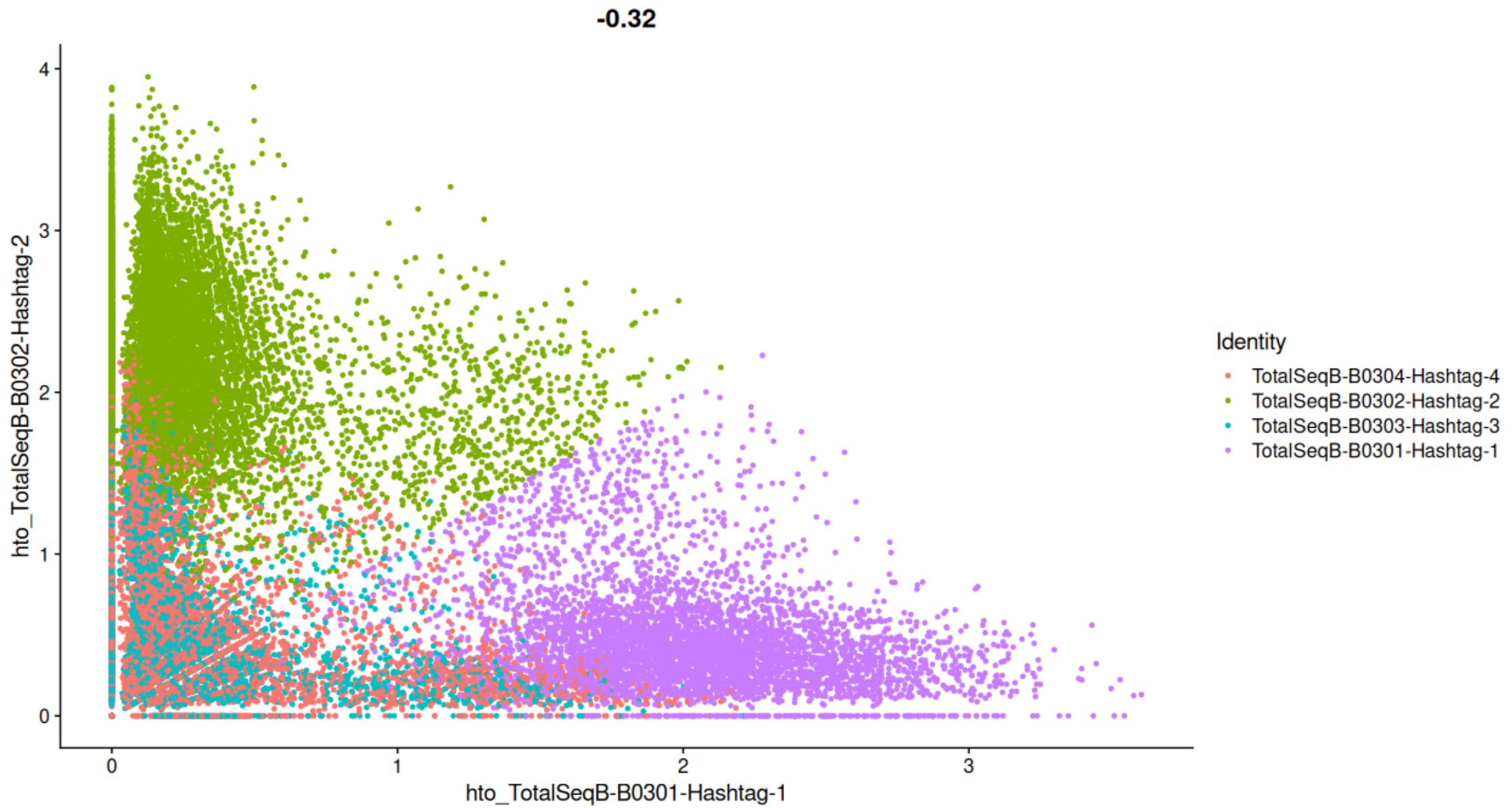


# BM2



# BM2

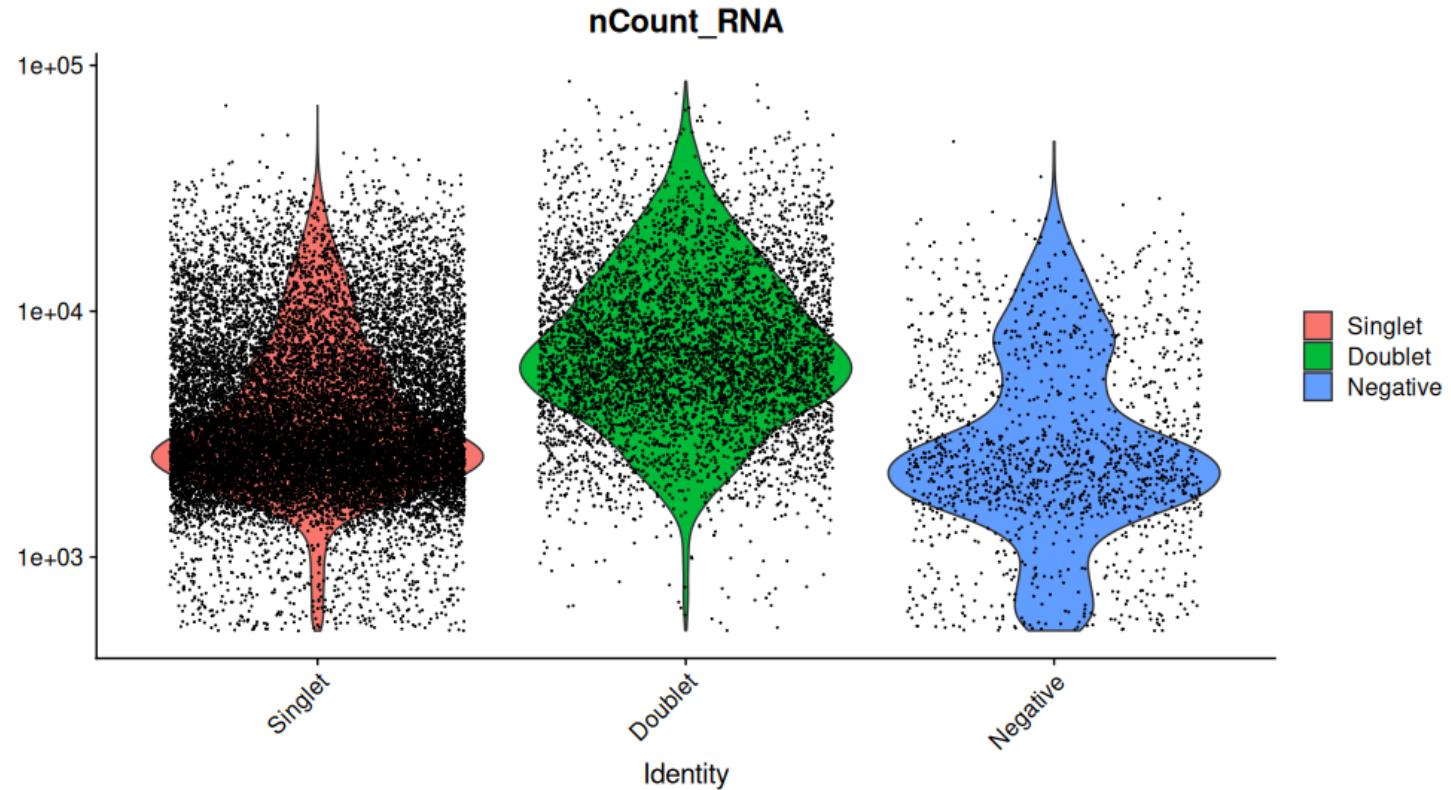




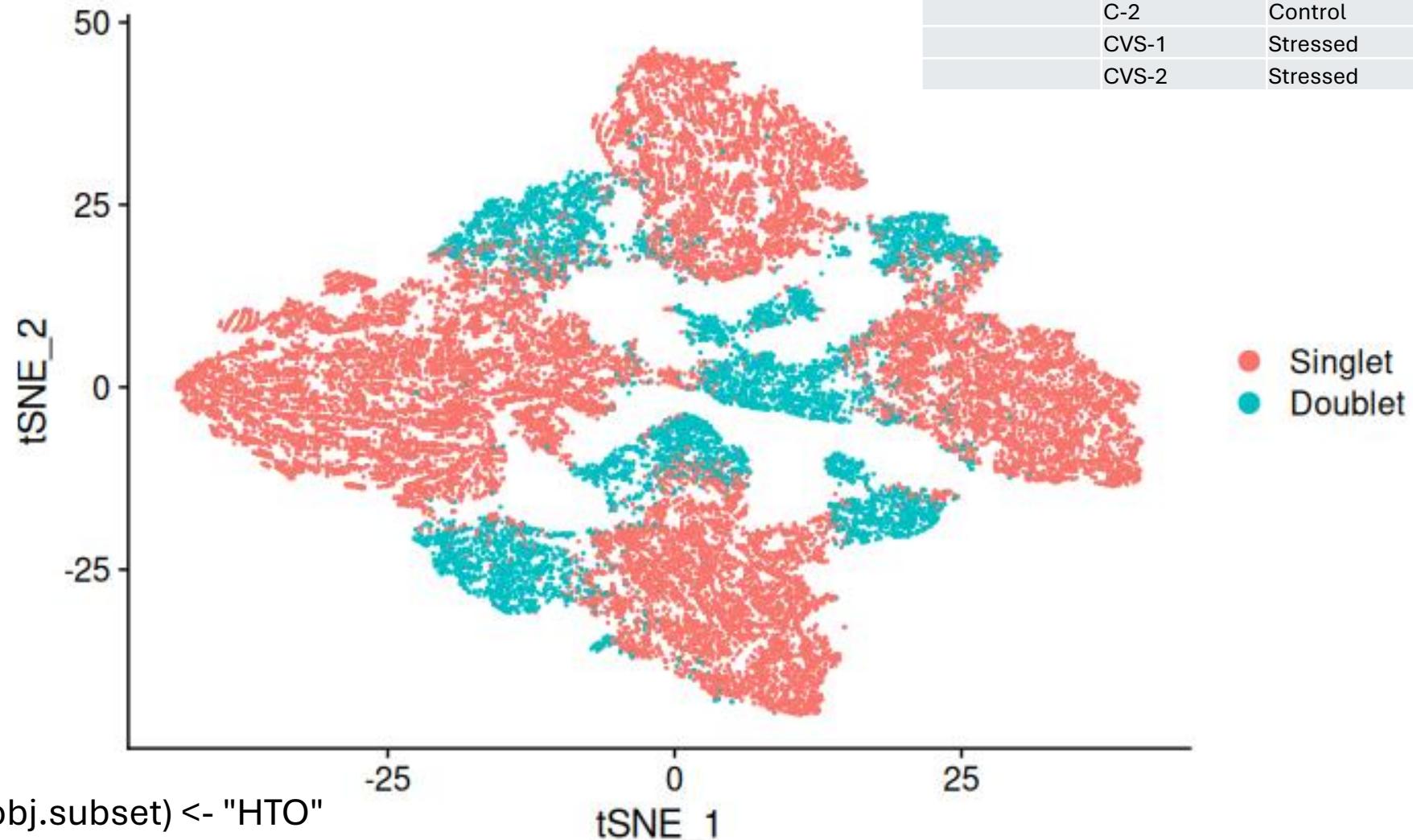
# BM2

```
table(seurat_obj$HT0_classification.global)
```

Doublet	Negative	Singlet
5875	1522	19846



# BM2



Clusters = groups of cells with similar HTO signal

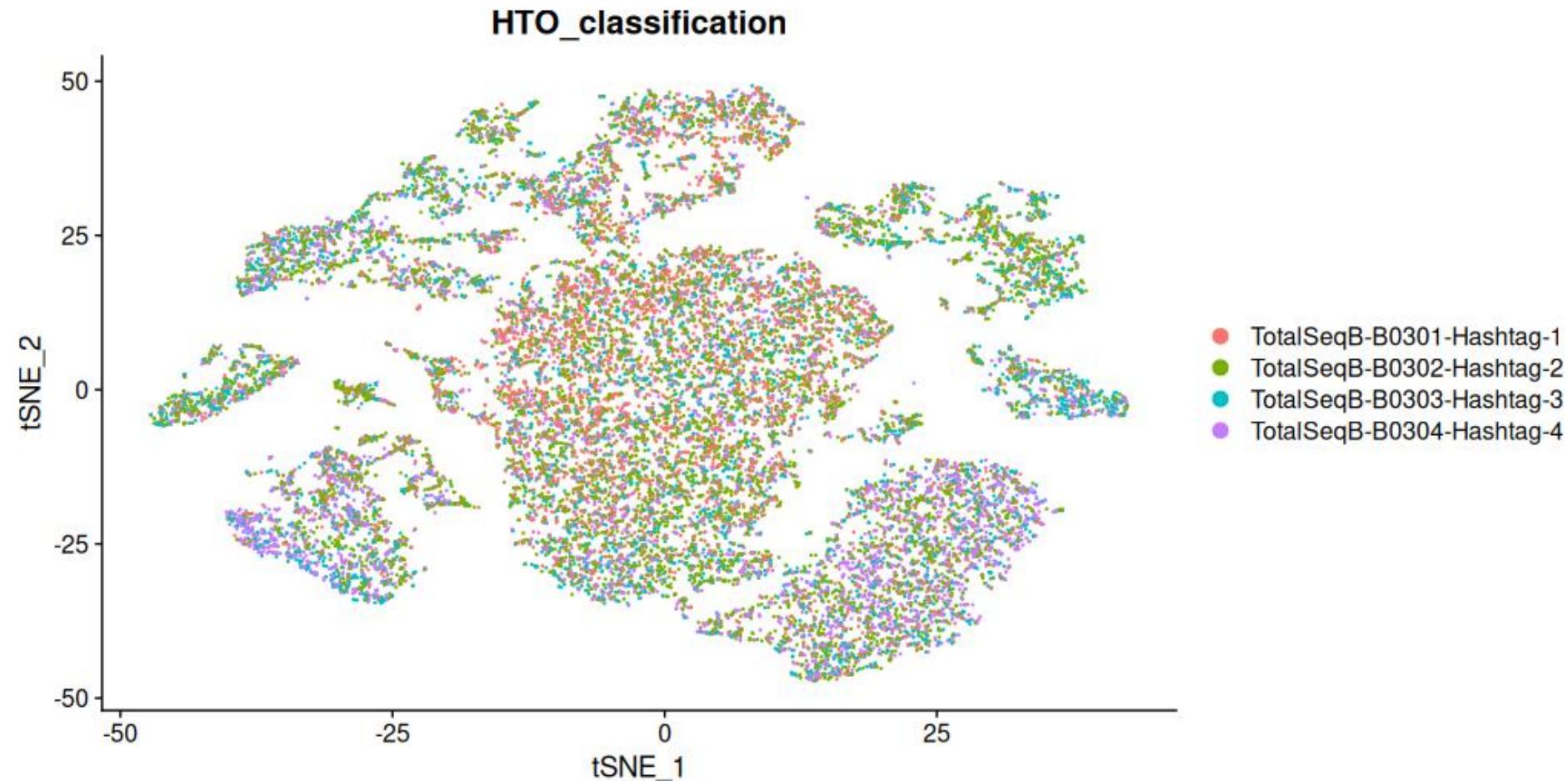
Distance between clusters = how different their expression/barcodes are

# BM2

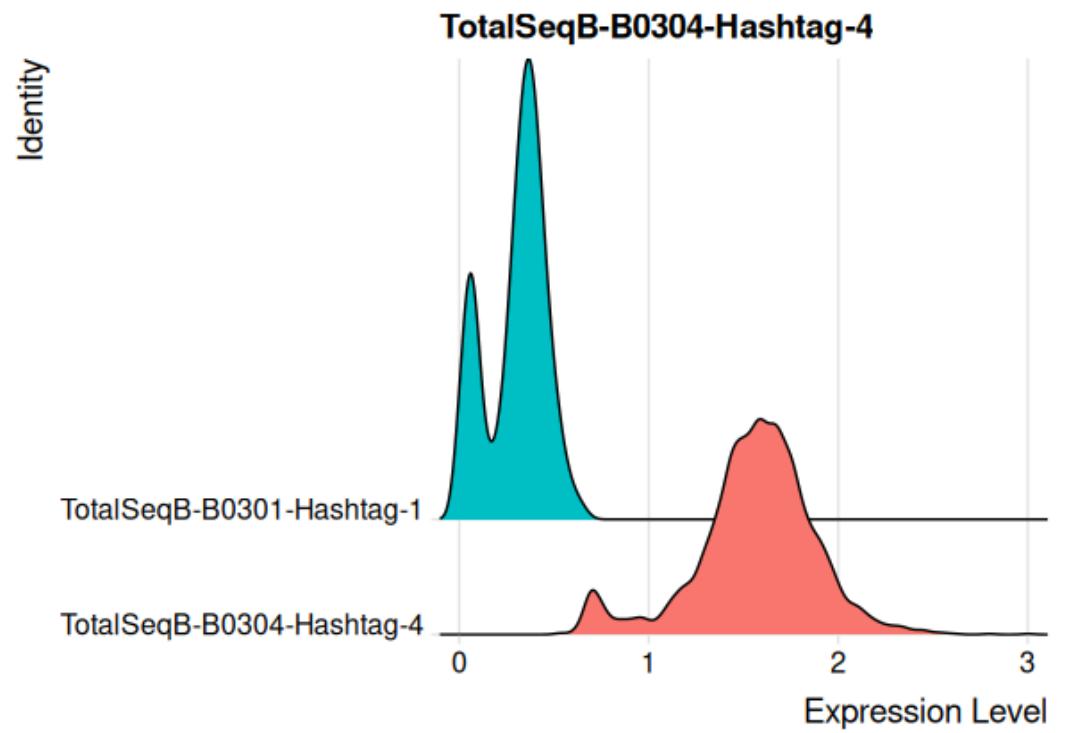
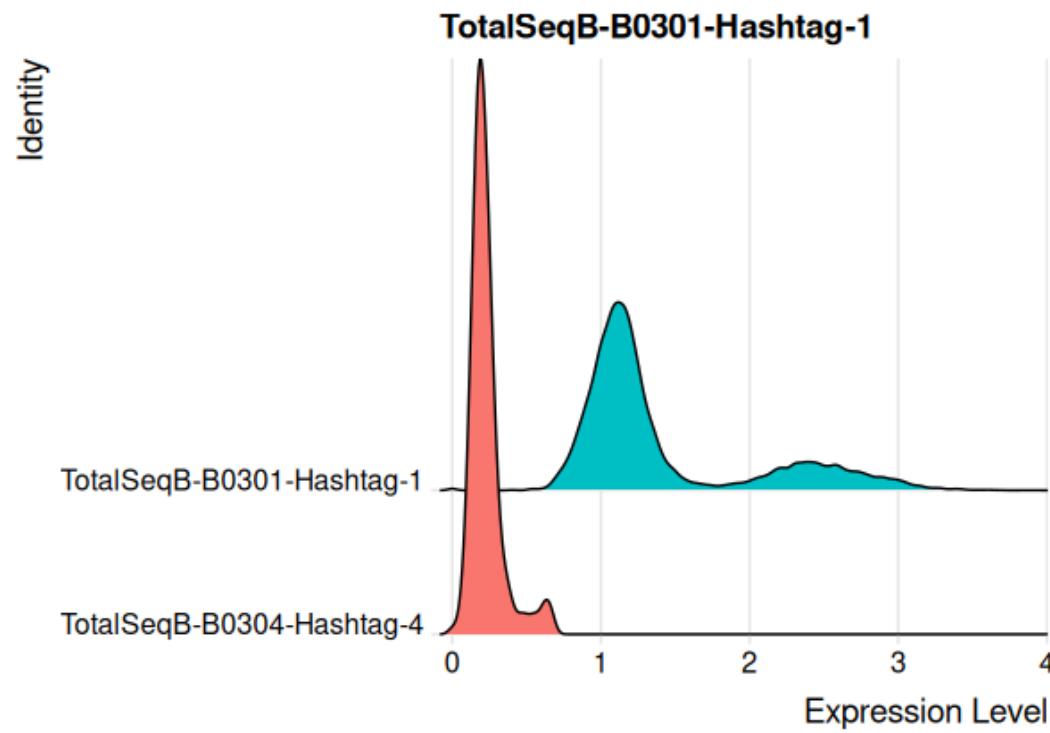


# BM2

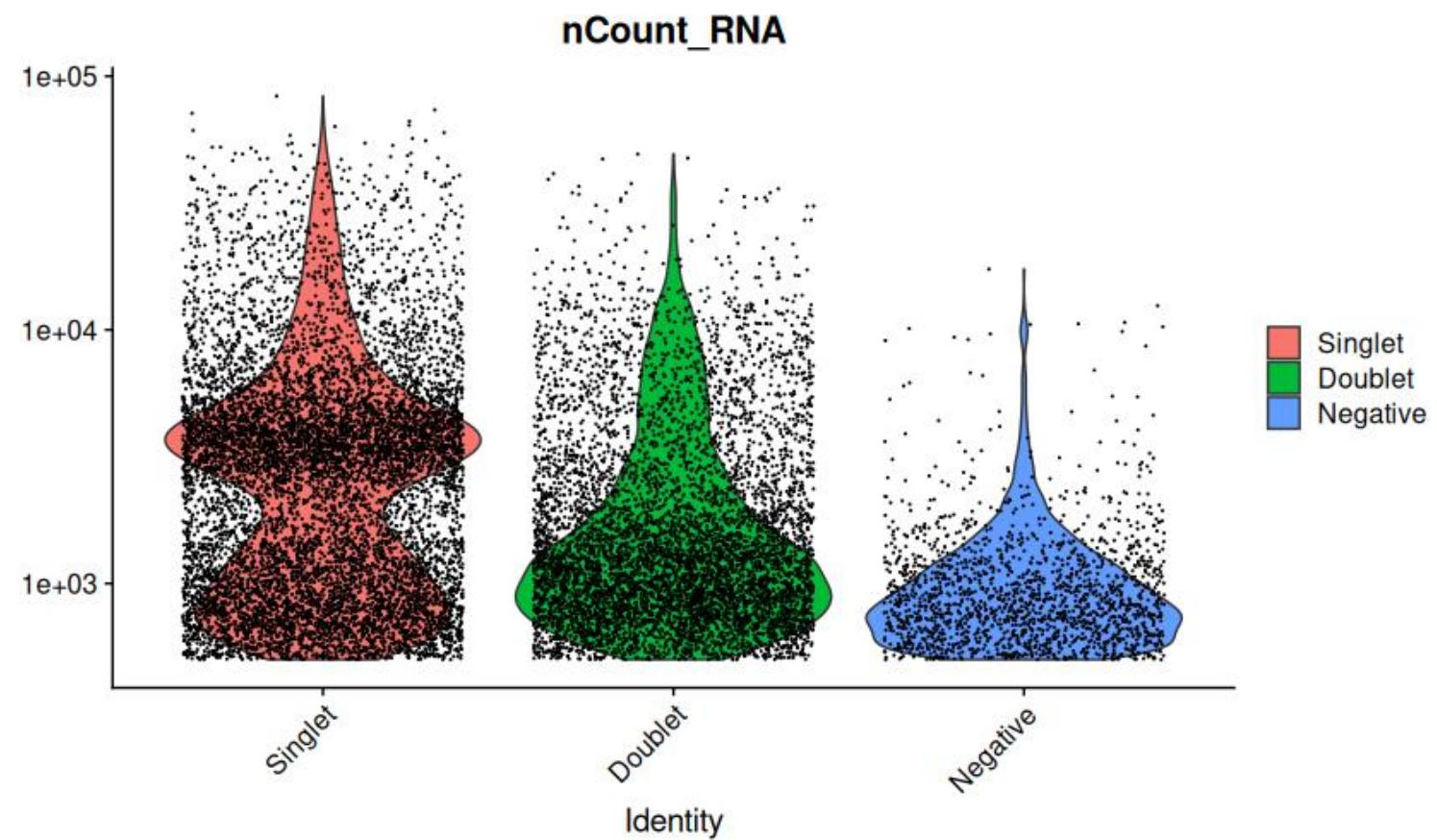
Pooled sample	Sample name	Sample details	Hashtag used
BM_1	C-1	Control	H1
	C-2	Control	H2
CVS-1	CVS-1	Stressed	H3
	CVS-2	Stressed	H4



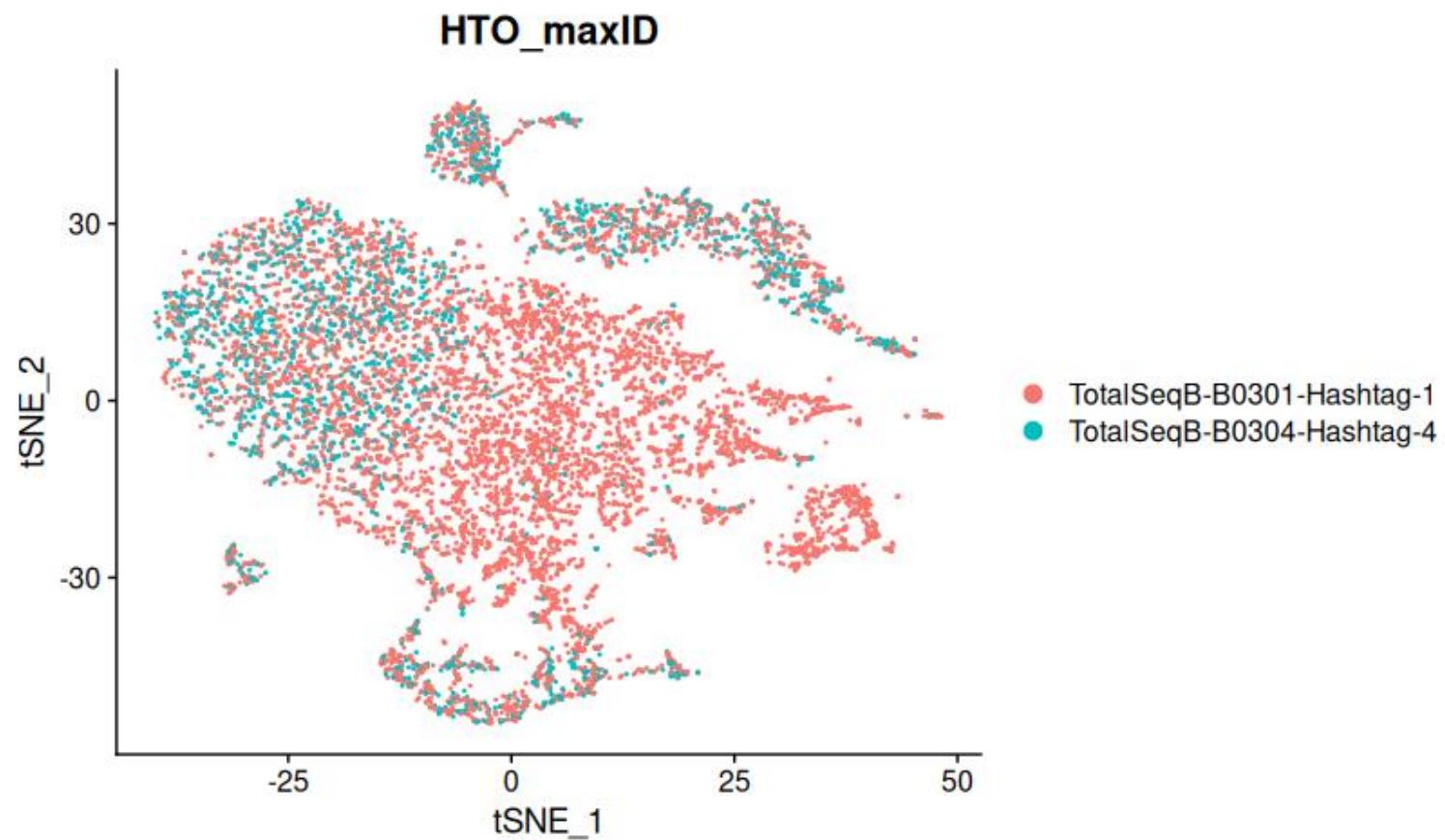
# GT1



# GT1

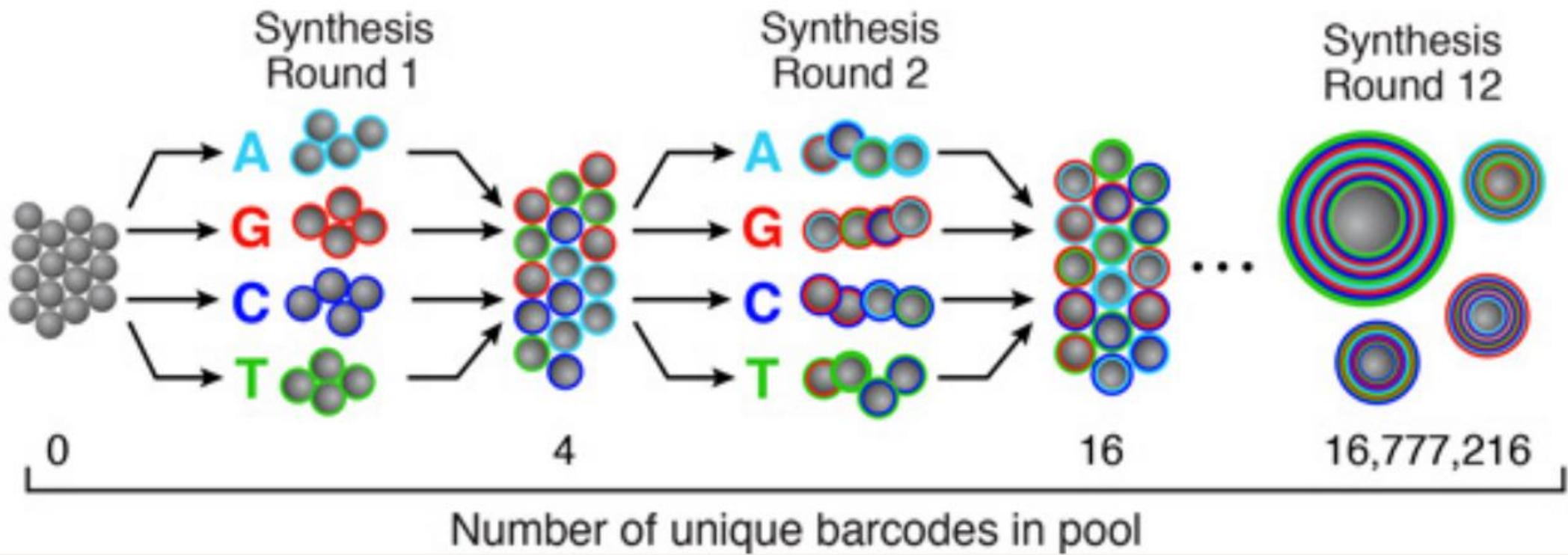


# GT1



# How thousands or millions of cell barcodes are made: using split and pool method

## C Synthesis of cell barcode (12 bases)



```
HTODemux(seurat_obj, assay = "HTO", positive.quantile = 0.99)
```

	Doublet	Negative	Singlet	Cellranger <b>web</b> <b>summary</b> - “cells”	Target recovery	Cells loaded
BM1	3953	1560	705	6,218	24000	30000
BM2	5875	1522	19846	27,243	24000	30000
GT1	6745	1763	9498	18,006	12000	20000
GT2	19	3436	4780	8,235	12000	20000
GT3	1131	48	10234	11,413	12000	20000

# BM1 was multiplexed with 1 tag?

---

Category	num_cells	pct_cells	median_umis	stddev_umis
No tag molecules	5	0.1	None	None
No tag assigned	322	5.2	None	None
1 tag assigned	5896	94.8	None	None
More than 1 tag assigned	0	0.0	None	None
B0304	5896	94.8	47.0	45.8
(END)				

GEX\_BM1\_new\_multi/outs/multi/multiplexing\_analysis/tag\_calls\_summary.csv

# GT1

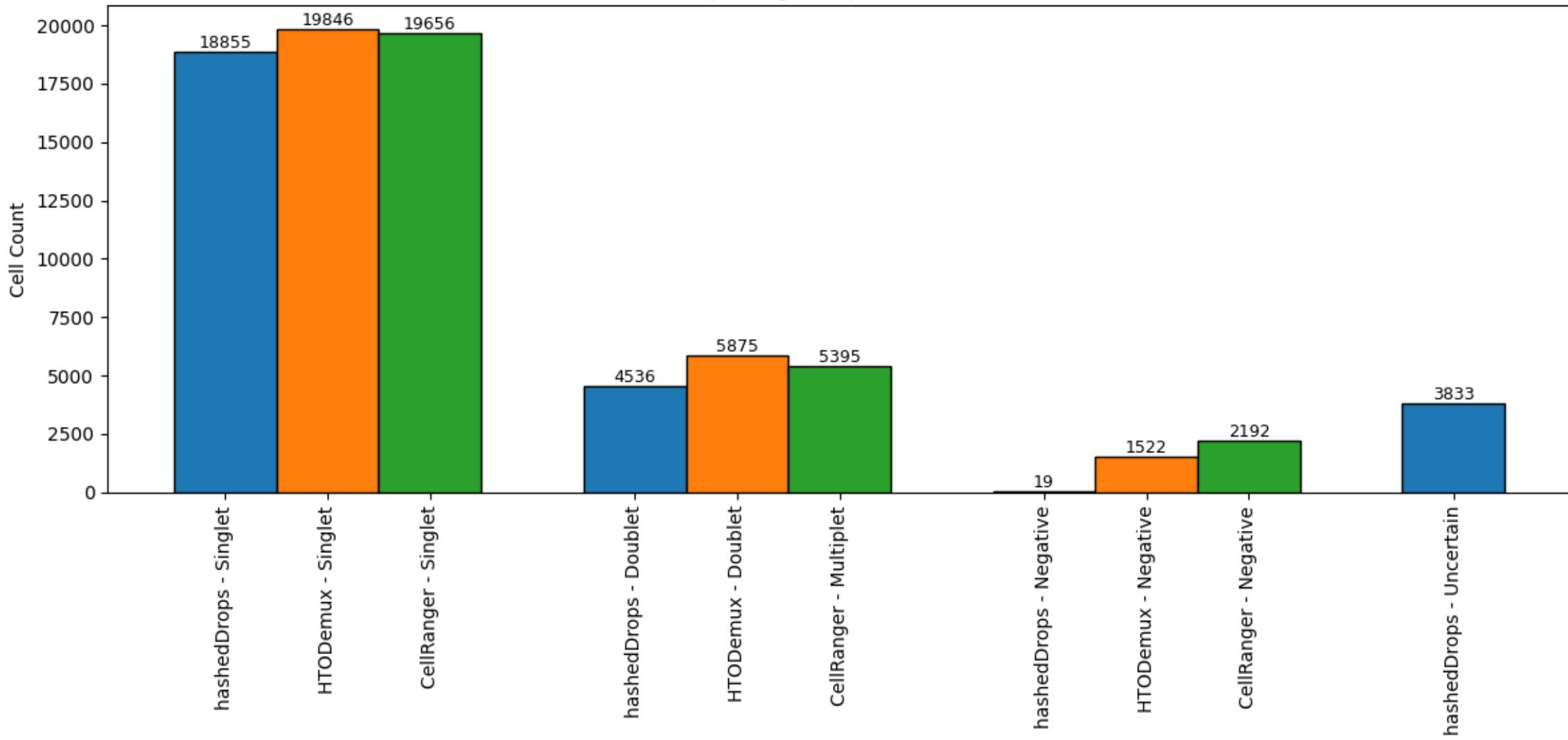
```
[rostamne@login2 multiplexing_analysis]$  
cat: less: No such file or directory  
Category,num_cells,pct_cells,median_umis  
No tag molecules,18,0.1,None,None  
No tag assigned,13933,77.4,None,None  
1 tag assigned,4073,22.6,None,None  
More than 1 tag assigned,0,0.0,None,None  
B0301,2079,11.5,406.0,342.9  
B0304,1994,11.1,145.5,94.8
```

# Cell ranger multi pipeline

- SV1:
  - Category,num\_cells,pct\_cells,median\_umis,stddev\_umis
  - No tag molecules,18,0.1,None,None
  - No tag assigned,4247,14.8,None,None
  - 1 tag assigned,19249,67.2,None,None
  - More than 1 tag assigned,5148,18.0,None,None
- SV2:
  - Category,num\_cells,pct\_cells,median\_umis,stddev\_umis
  - No tag molecules,15,0.1,None,None
  - No tag assigned,2315,8.1,None,None
  - 1 tag assigned,19461,68.1,None,None
  - More than 1 tag assigned,6786,23.8,None,None

	Seurat HTDemux			Cell Ranger multi pipeline			Cell Ranger count		
	Doublet	Negative	Singlet	Multiplet	Negative	Singlet	Cellrange r web summar y - “cells”	Target recovery	Cells loaded
BM1	3953	1560	705	0	322	5896	6,218	24000	30000
BM2	5875	1522	19846	5395	2192	19656	27,243	24000	30000
GT1	6745	1763	9498	0	13933	4073	18,006	12000	20000
GT2	19	3436	4780	415	188	7632	8,235	12000	20000
GT3	1131	48	10234	796	199	10418	11,413	12000	20000

### BM2 Demultiplexing Comparison Across Methods



# BM2

