

Iteration 2: QC for BM-32 sequencing

Analysis & QC with no filters

Zinger Yang Loureiro

August 30, 2021

QC results revealed uneven target cell recovery (expected: 10000)

Source: CellRanger QC report

	BMY-32	BMR-32	BMS-32
Number of Reads	883,279,537	609,220,469	647,635,467
Estimated Number of Cells	9,691	16,505	9,376
Fraction Reads in Cells	79.2%	88.7%	91.5%
Mean Reads per Cell	91,144	36,911	69,074
Median Genes per Cell	1,510	2,615	3,780
Total Genes Detected	27,321	27,563	26,340
Median UMI Counts per Cell	5,911	8,371	22,467

Estimated Number of Cells The number of barcodes associated with at least one cell.

Fraction Reads in Cells The fraction of valid-barcode, confidently-mapped-to-transcriptome reads with cell-associated barcodes.

Mean Reads per Cell The total number of sequenced reads divided by the number of barcodes associated with cell-containing partitions.

Median Genes per Cell The median number of genes detected per cell-associated barcode. Detection is defined as the presence of at least 1 UMI count.

Total Genes Detected The number of genes with at least one UMI count in any cell.

Median UMI Counts per Cell The median number of UMI counts per %s cell-associated barcode.

How to interpret the "Fraction Reads in Cells" metric?

Question: I see a low value for the "Fraction Reads in Cells". How can I interpret this metric?

Answer: A low "Fraction Reads in Cells" value is typically explained by the following:

- 1) High ambient RNA (background) in your sample. This ambient RNA comes from lysed/dead cells in your sample. Cell Ranger is able to confidently align the reads from ambient RNA to the transcriptome but the reads are not associated with a valid cell-containing GEM.
- 2) The cell-calling heuristic did not apply. For example, there may be higher variation in RNA content than expected (more cells with lower RNA content). The current cell-calling heuristic assumes a ten-fold variation in RNA content.

<https://kb.10xgenomics.com/hc/en-us/articles/360003919491-How-to-interpret-the-Fraction-Reads-in-Cells-metric->

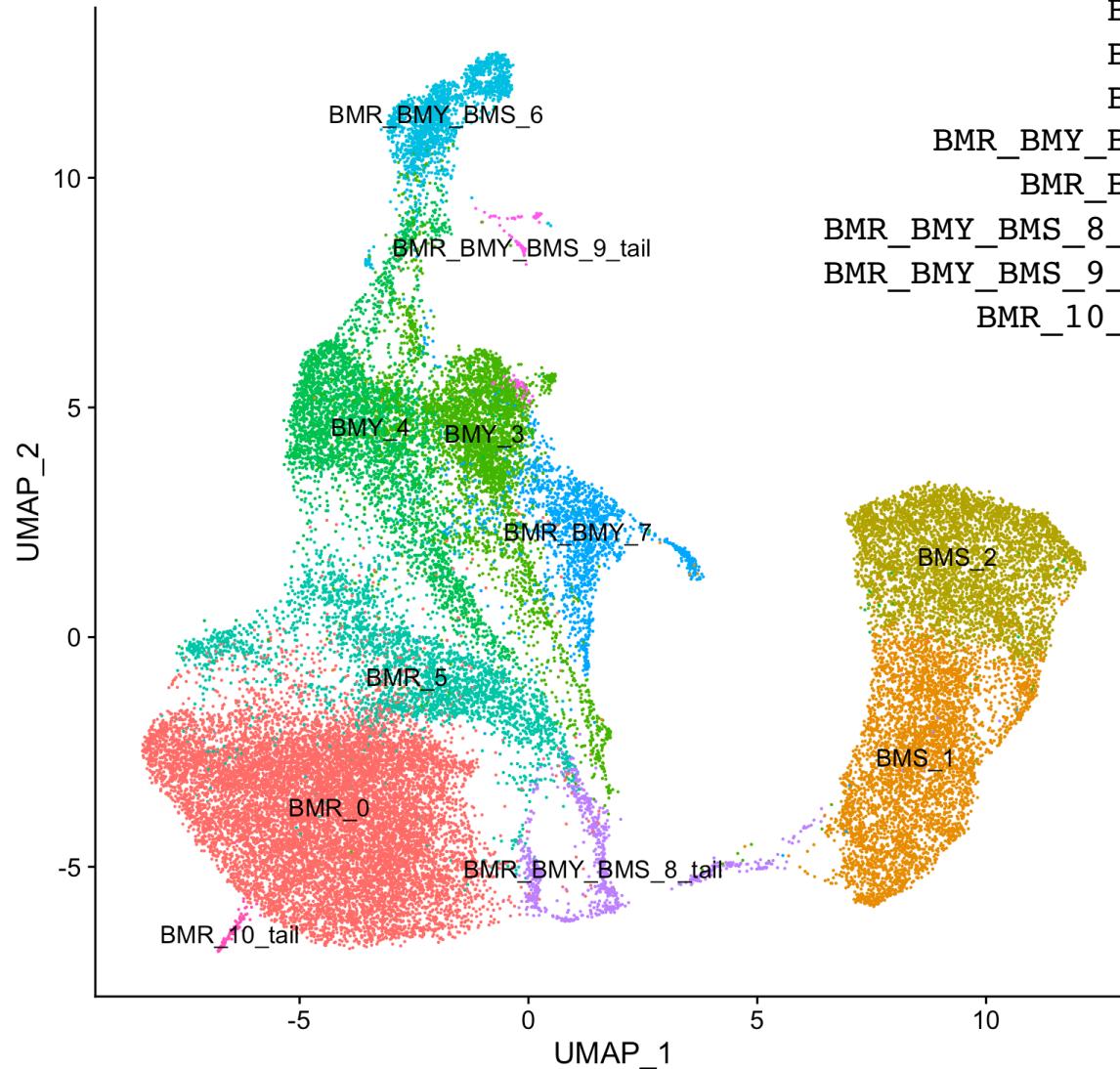
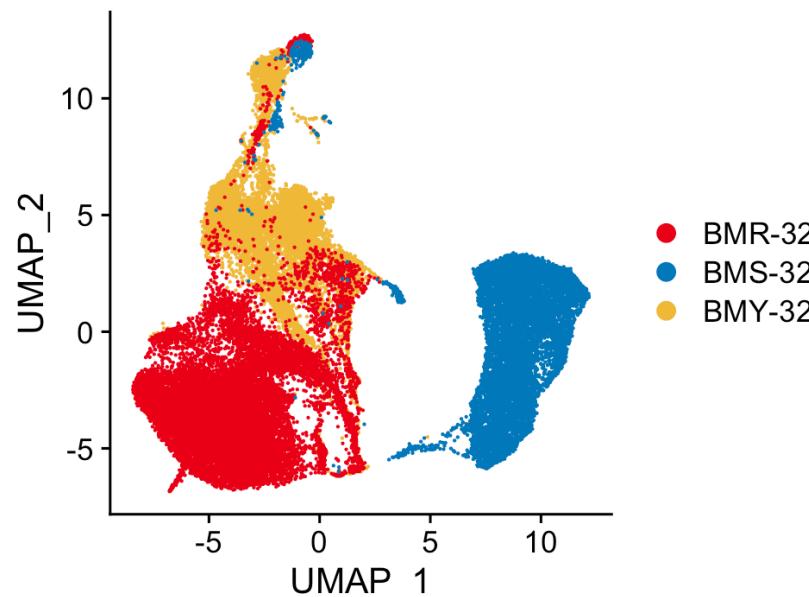
Seurat preliminary clustering for QC purpose **Iteration 2**

No filtering applied to the dataset

BMY-32 9691 cells

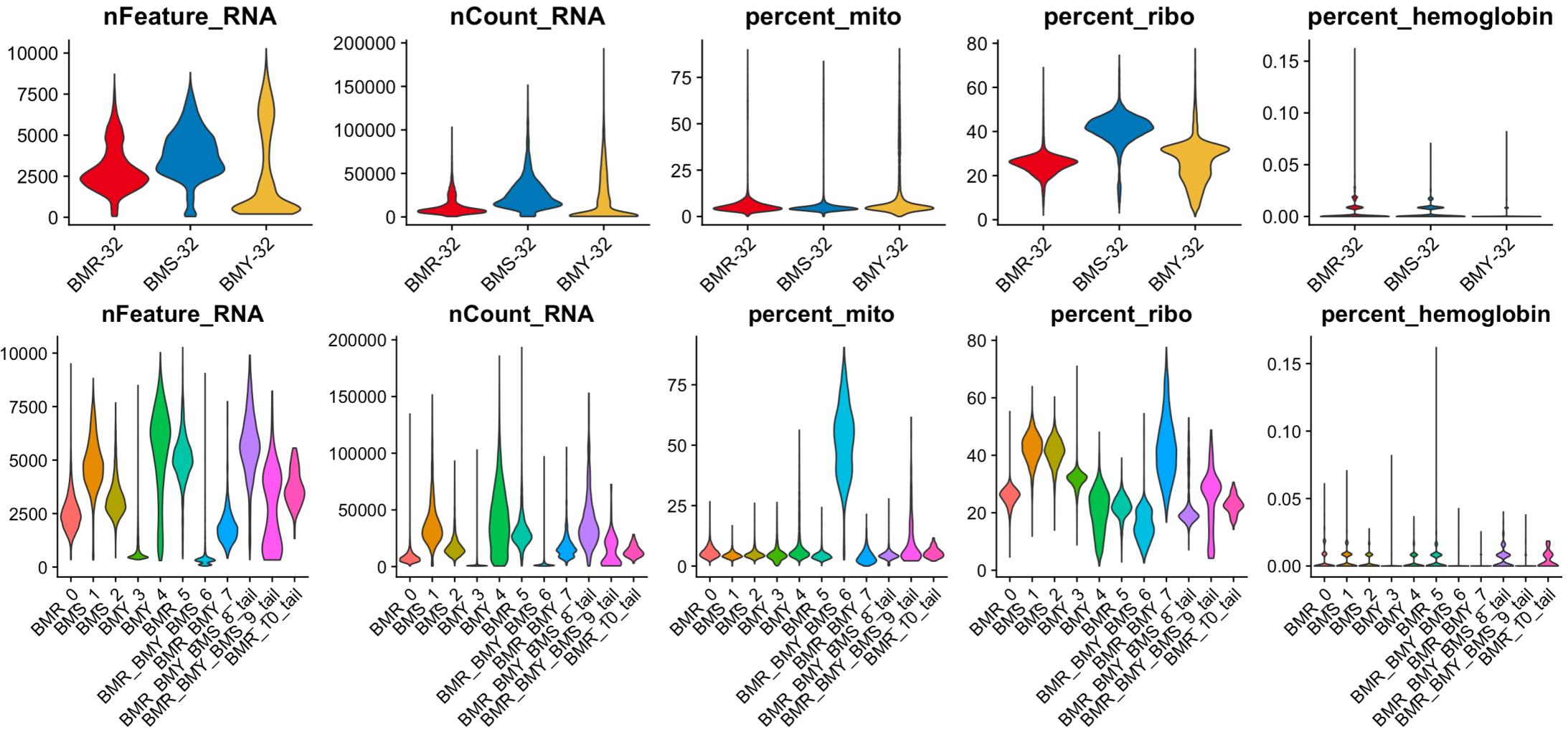
BMR-32 16505 cells

BMS-32 9376 cells



BMR_0	11894	cells
BMS_1	4393	cells
BMS_2	4349	cells
BMY_3	4046	cells
BMY_4	3755	cells
BMR_5	2938	cells
BMR_BMY_BMS_6	1670	cells
BMR_BMY_7	1504	cells
BMR_BMY_BMS_8_tail	765	cells
BMR_BMY_BMS_9_tail	147	cells
BMR_10_tail	111	cells

QC by sample / cluster

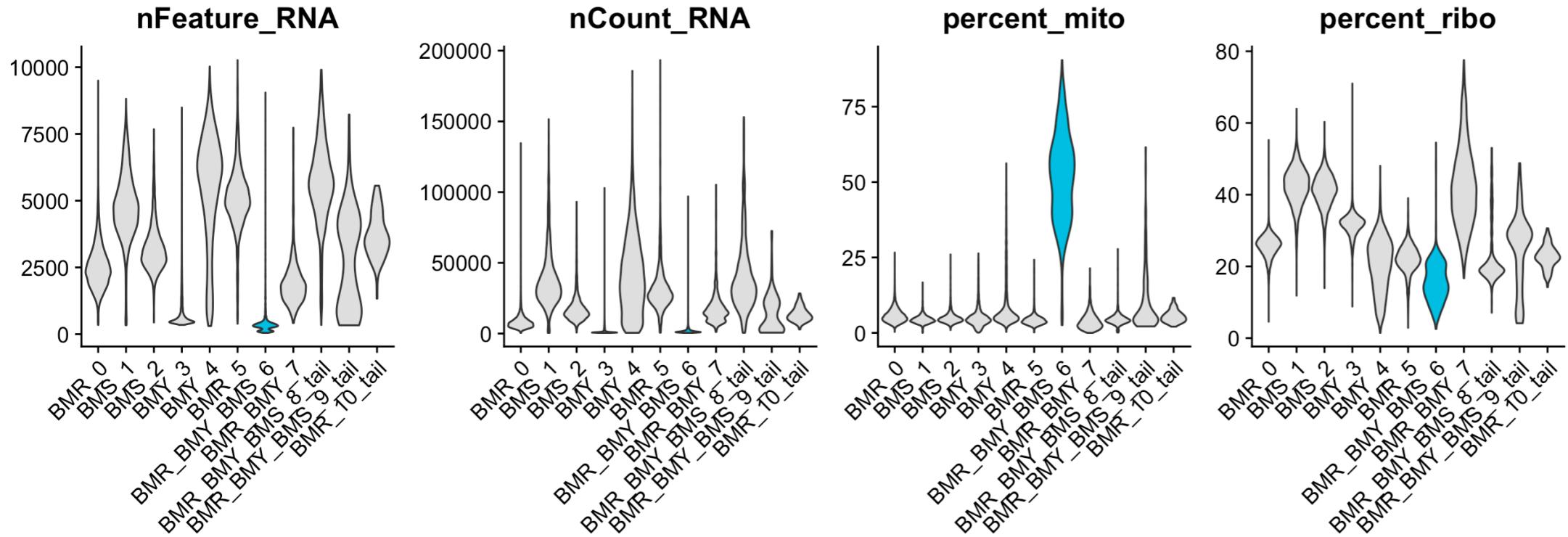


BMR-32 249 cells
 BMS-32 255 cells
 BMY-32 1166 cells



Cluster 6 is a low count, high mitochondrial gene cluster

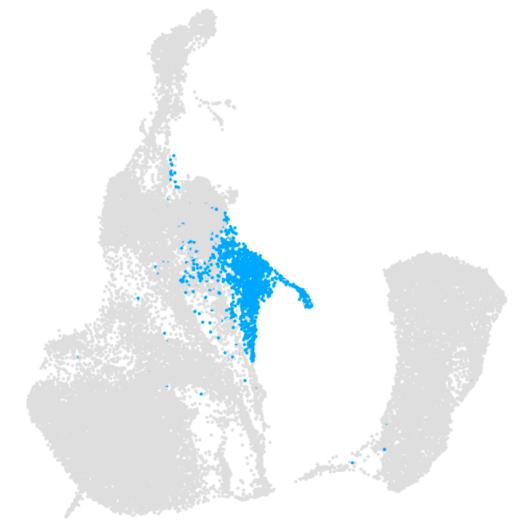
Top markers: MT-CO3, MT-ATP6, MT-ND1, MT-ND2, MT-CO1, MT-ND4, MT-ND3, MT-CYB, MT-CO2, MT-ND5



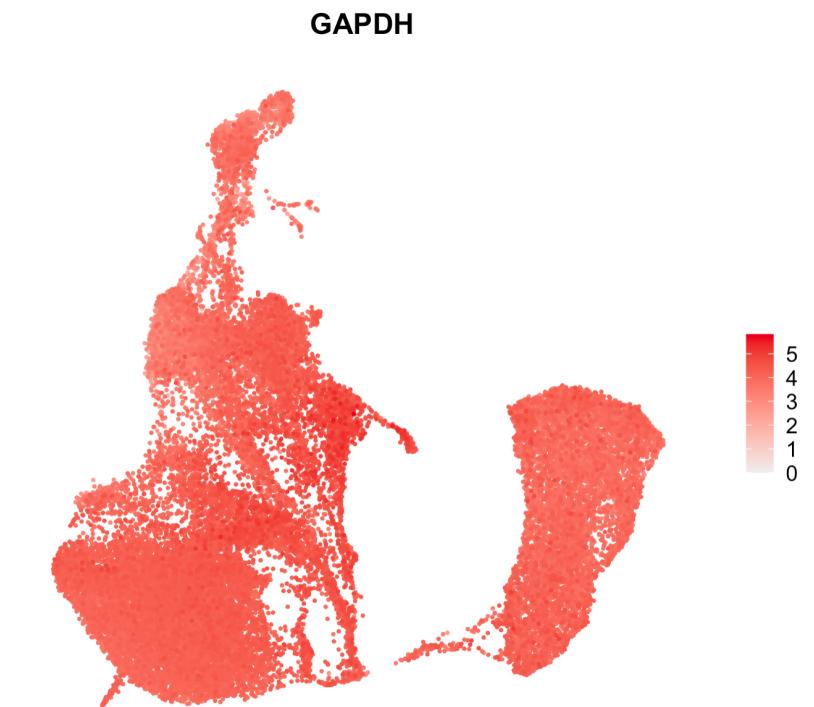
High expression levels of mitochondrial genes could be an indicator of:

1. Poor sample quality, leading to a high fraction of apoptotic or lysing cells.
2. Biology of the particular sample, for example tumor biopsies, which may have increased mitochondrial gene expression due to metabolic activity and/or necrosis.

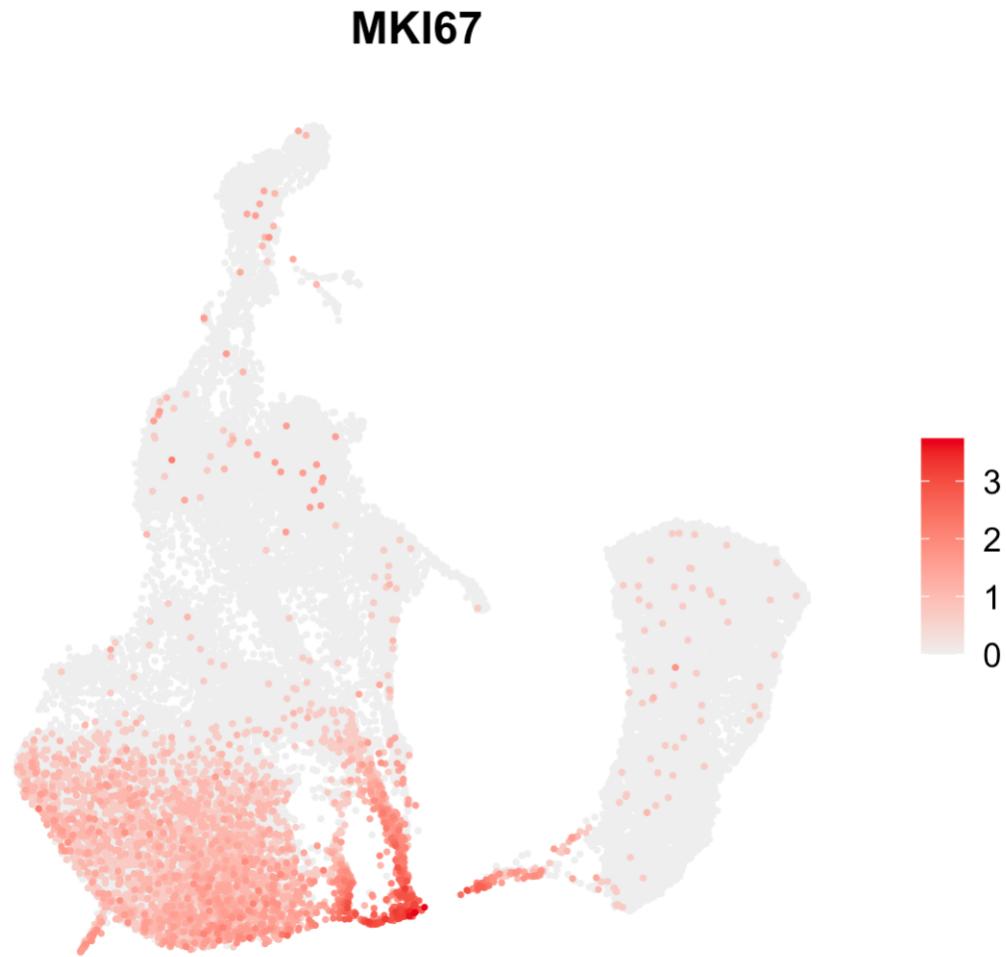
Cluster 7 has low expression of nuclear lncRNA MALAT1 and NEAT1, may be representing a poor nuclear lysis cell cluster



BMR-32 457 cells
BMS-32 133 cells
BMY-32 914 cells

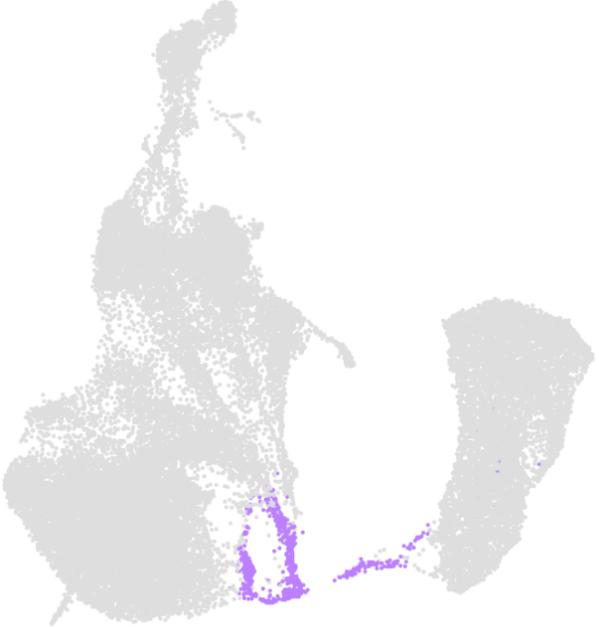


Proliferation marker Ki67 marks specific clusters

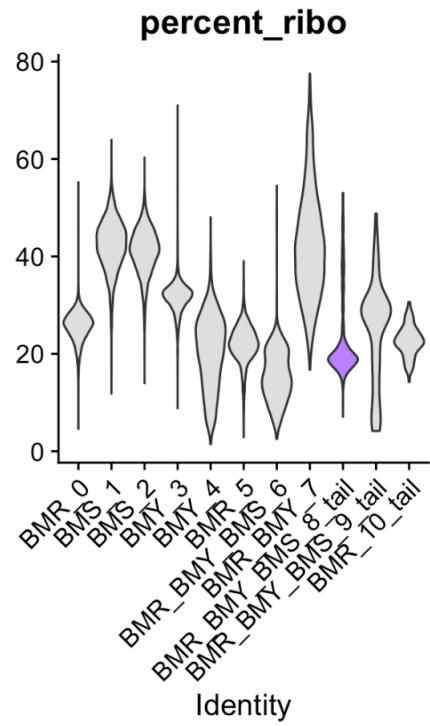
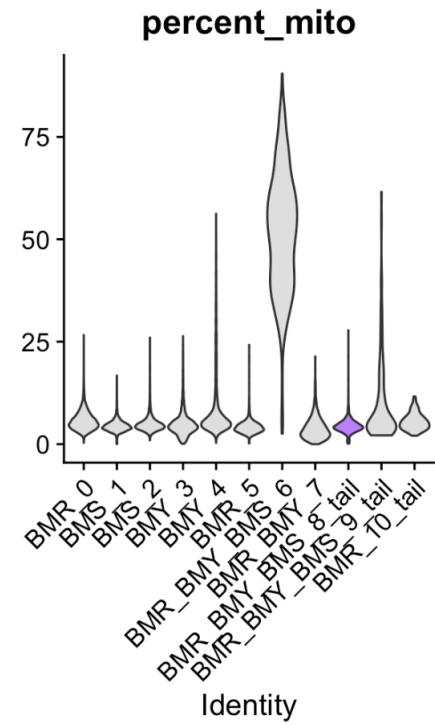
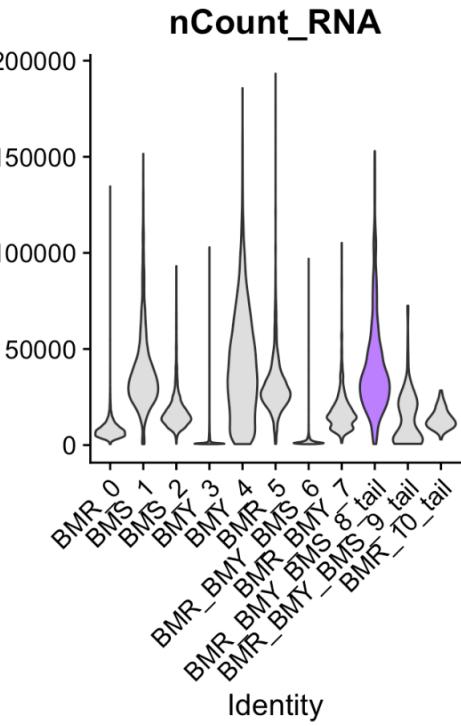
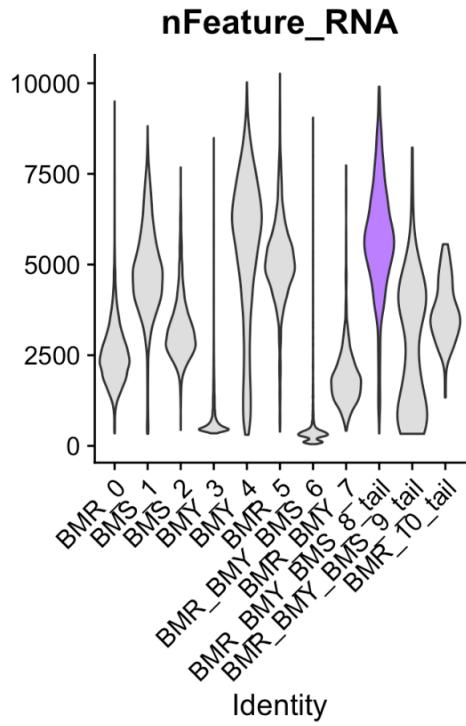


Cluster 8 may be a proliferating cell cluster

Top markers: CENPF, HIST1H4C, H2AFZ, MKI67, CCNB1, UBE2S, STMN1, CKS2, PTTG1, HIST1H1B

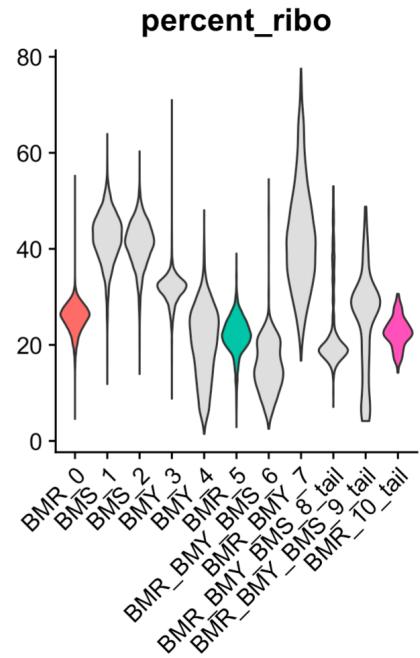
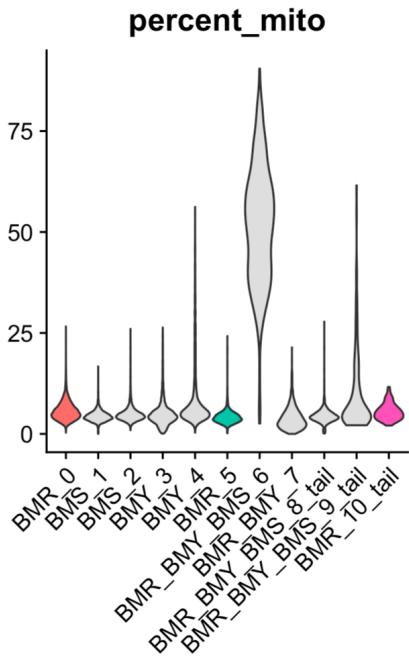
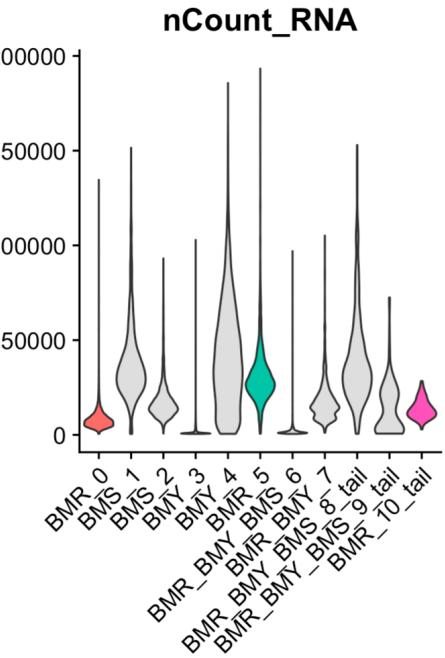
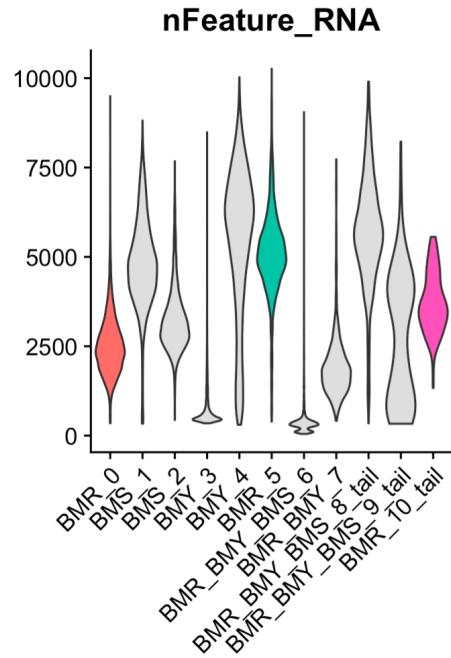
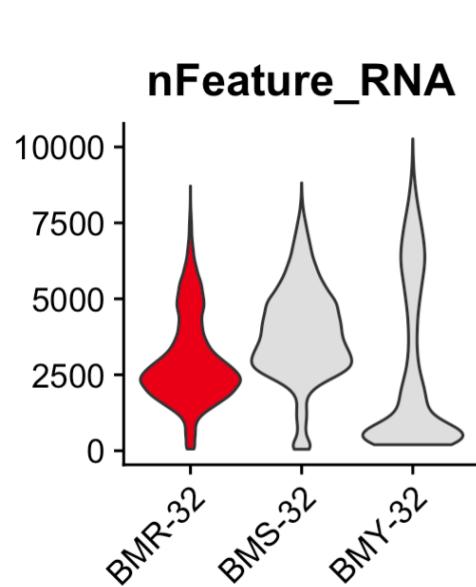
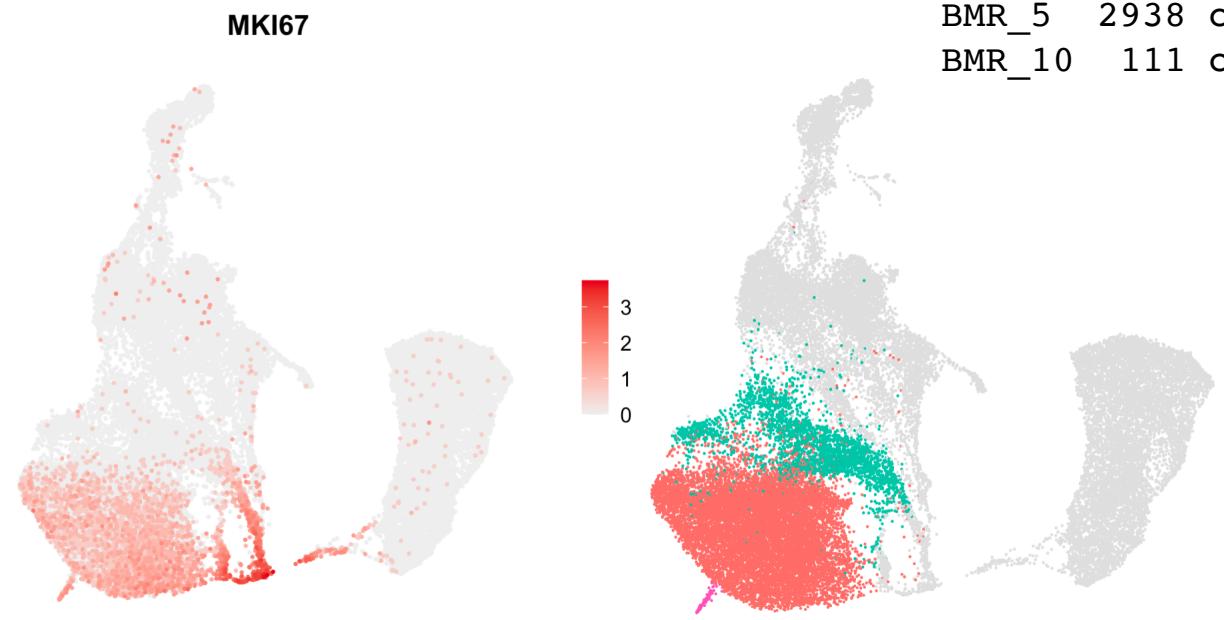


BMR-32 488 cells
BMS-32 135 cells
BMY-32 142 cells



BMR-32 clusters may be distinguished by proliferation status

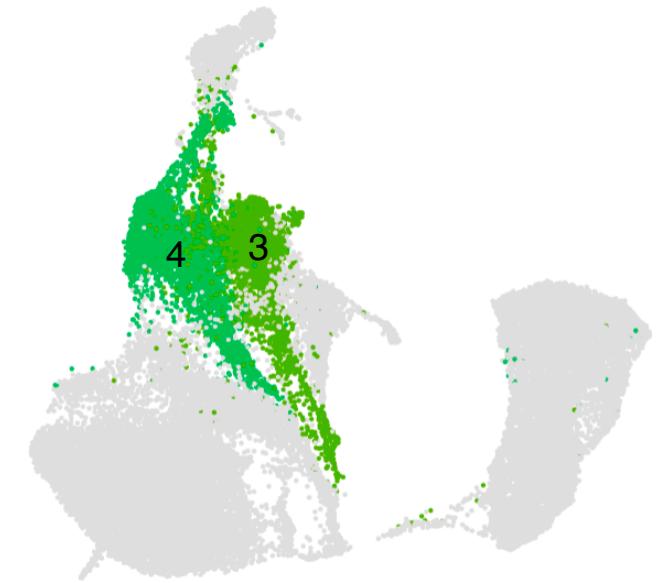
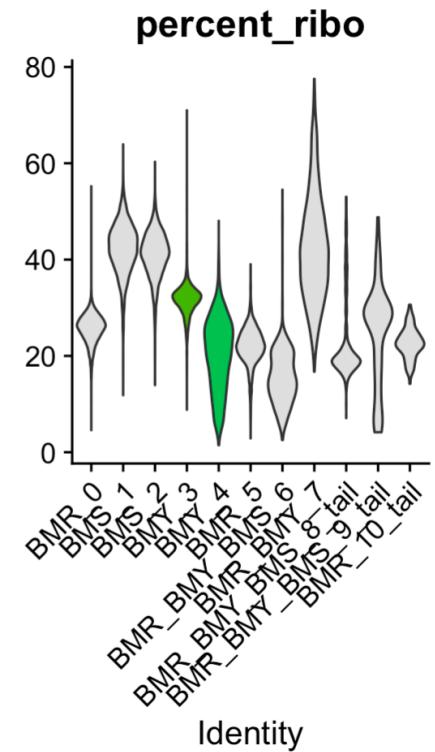
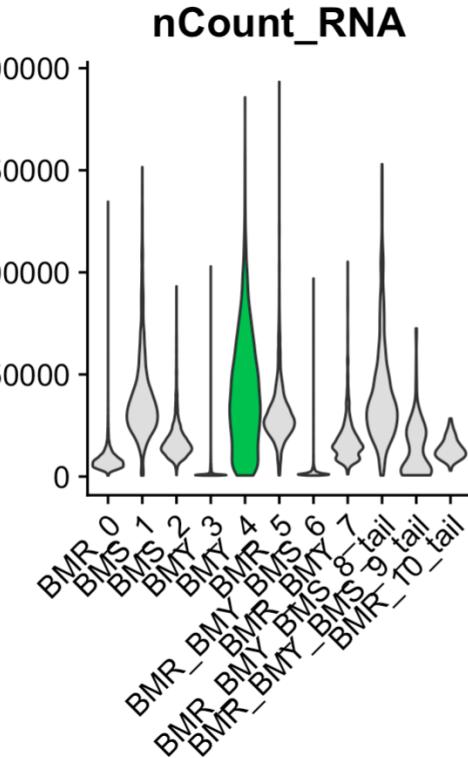
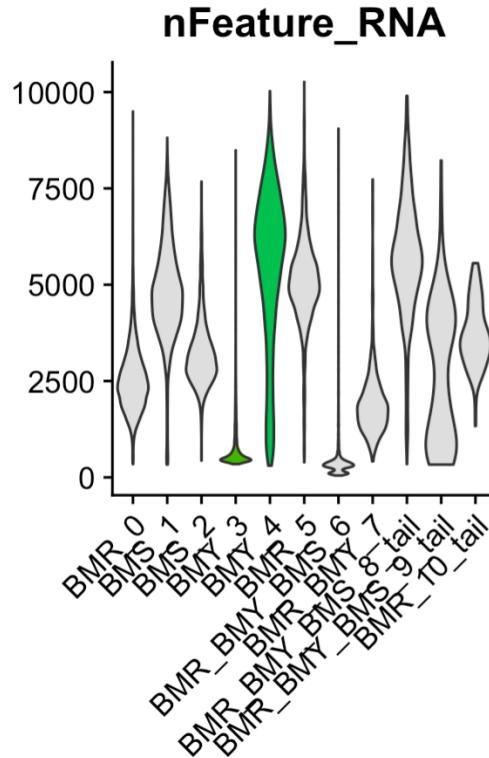
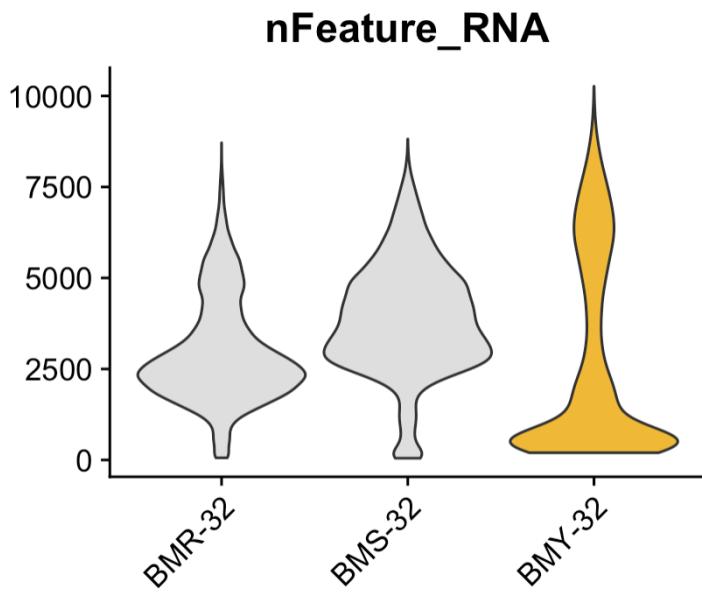
BMR_0 11894 cells
BMR_5 2938 cells
BMR_10 111 cells



BMY-32 has a relatively extreme bimodal distribution of genes / reads per cell

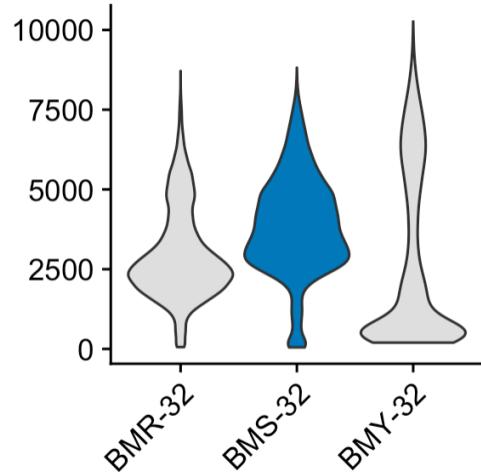
BMY-32 has the deepest sequencing depth among the 3 samples

BMY_3 4046 cells
BMY_4 3755 cells

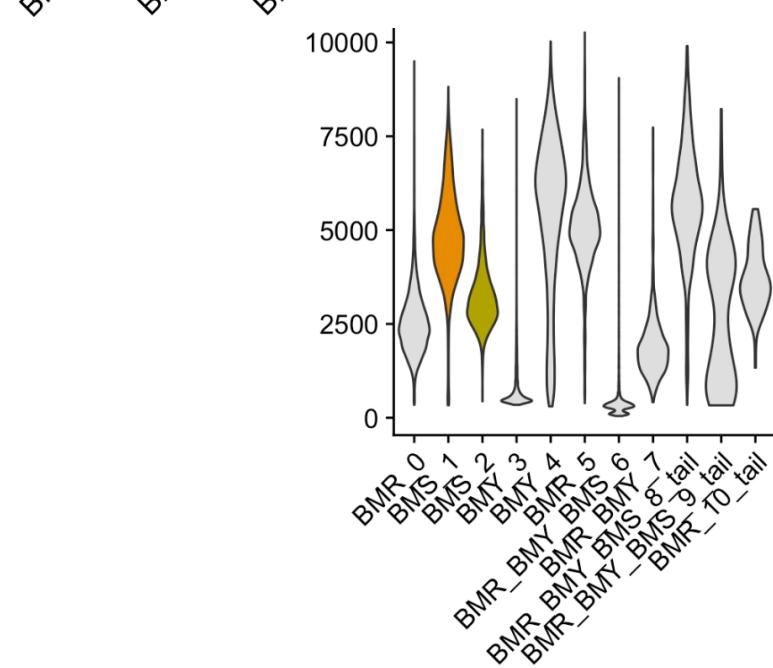


BMS-32

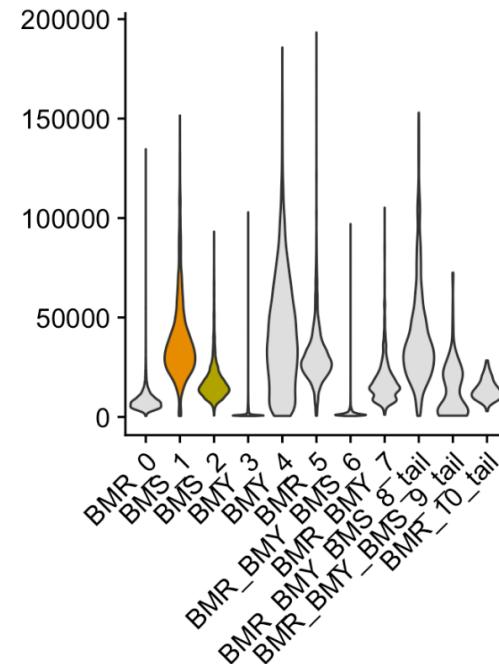
nFeature_RNA



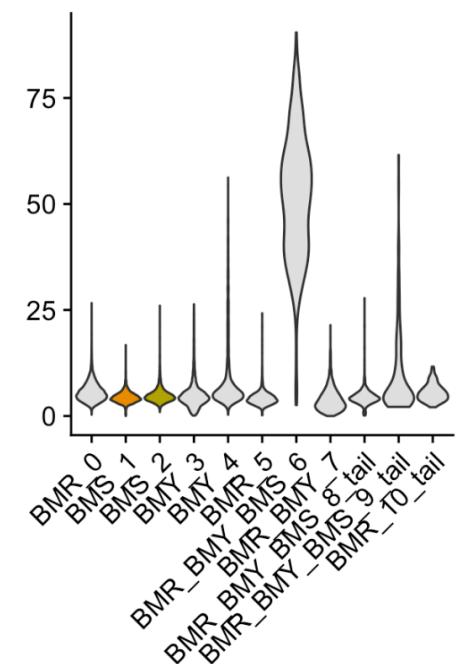
nFeature_RNA



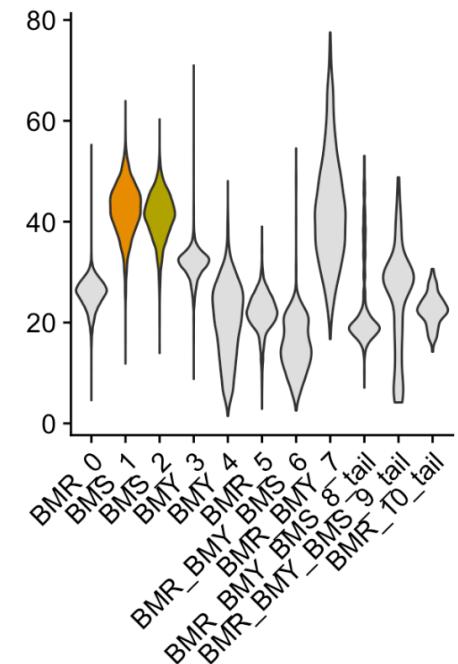
nCount_RNA



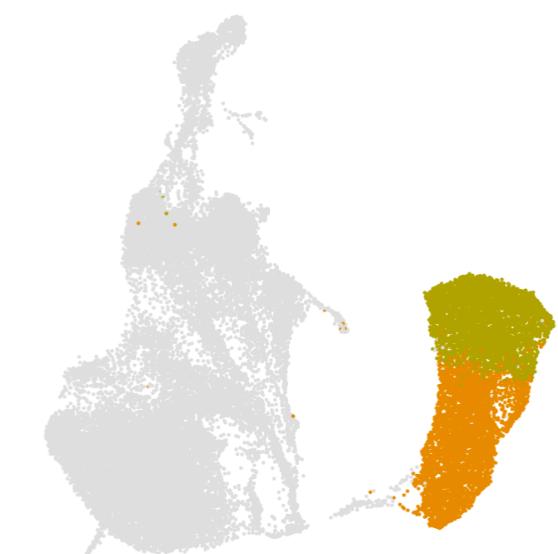
percent_mito



percent_ribo

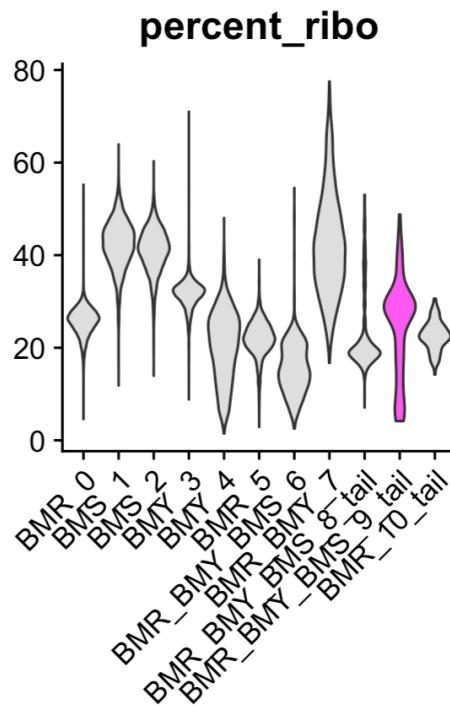
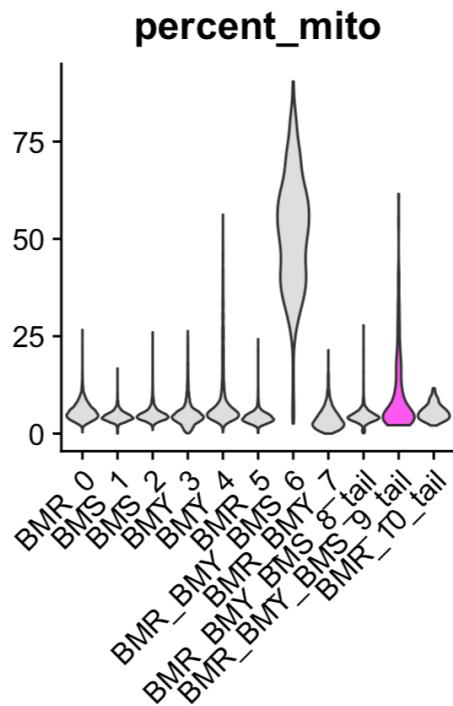
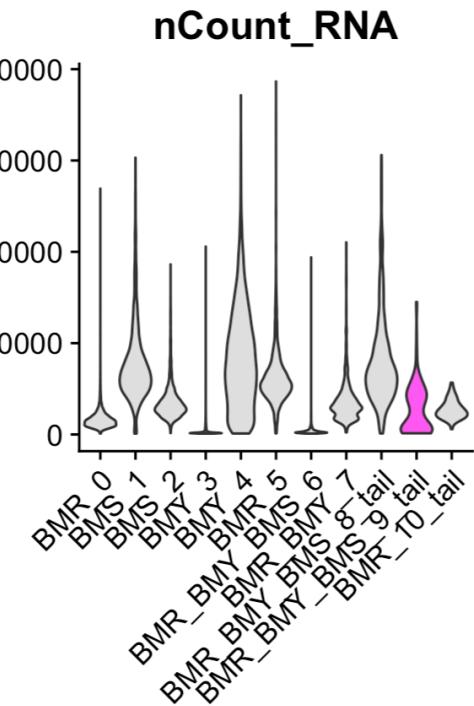
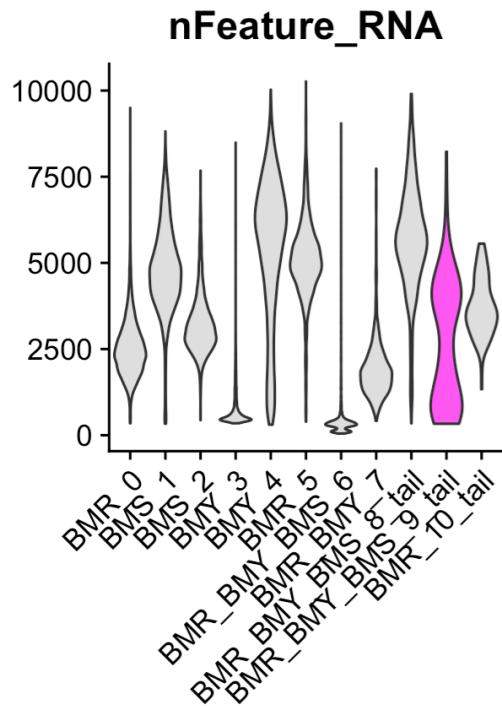
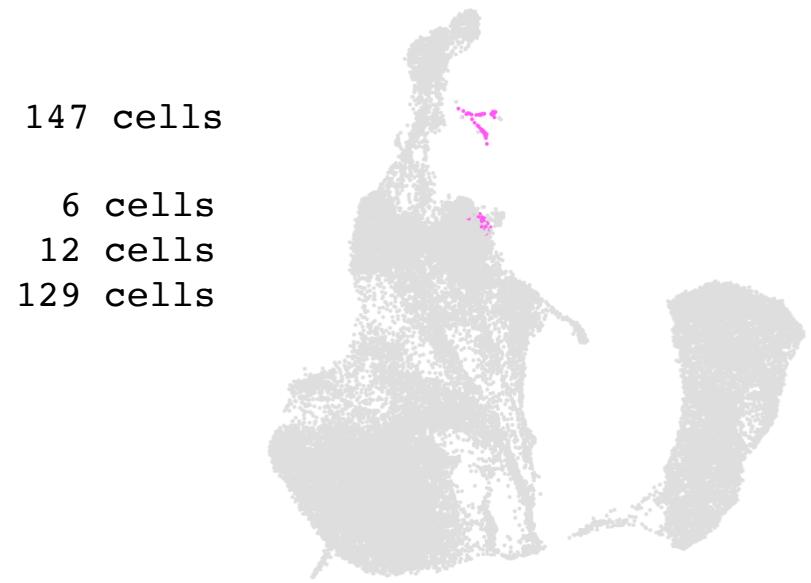


BMS_1 4393 cells
BMS_2 4349 cells



Cluster 9

Top markers: C1QA, C1QB, SPP1, CD163, RNASE1, CCL18, VSIG4, LGMN, CTSB, CTSS



Discussion

- **EMY-32**
 - Highest contributor to Cluster 6 – high mitochondrial gene cluster
 - Highest contributor to Cluster 7 – low nuclear gene cluster
 - Is cluster 3 a cluster of less-than-ideal quality cells?
 - Cluster 3 may be mostly eliminated if imposing a minimum gene detection threshold
- **Technical artifact clusters**
 - Cluster 6 is a low count, high mitochondrial gene cluster
 - Cluster 7 is likely a cluster of low nuclear lysis cells
- This analysis does not review clustering at higher resolution
 - Heterogeneity could be further explored
- The unfiltered Seurat object size is 1.82 GB
 - 25570 genes x 35572 cells
- **Imposing threshold?**
 - Limit to genes expressed in min number of cells
 - Limit to cells with min number of genes detected

