

Dataset S1. COPILOT summary files of col0 toy data, col0 full data, and PBMC data, related to Step “Run COPILOT for quality filtering”

col0_toy Summary

Processed by COPILOT

Summary

Analysis

Parameters

Iteration of Filtering	1
Mitochondrial Expression Threshold	5 %
Top High Quality Cell Filtered	1 %
Doublet Removed	No

Cell Stats

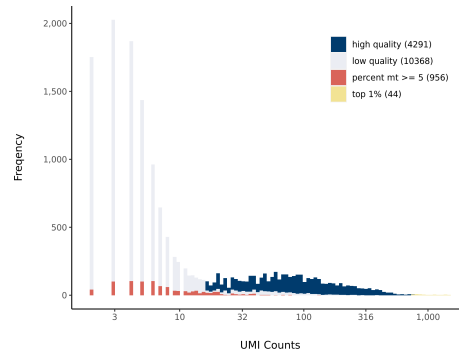
Estimated Number of High Quality Cell	4,291
High Quality Cell	27.4 %
Total UMI Counts in High Quality Cell	498,130
UMI Counts in High Quality Cell	77.89 %
Median UMI Counts per High Quality Cell	79
Median Genes per High Quality Cell	73
Total Genes Detected in High Quality Cell	15,865
Cell above Mitochondrial Expression Threshold	6.11 %
Estimated Doublet Rate in High Quality Cell	3.28 %

Sequencing Stats

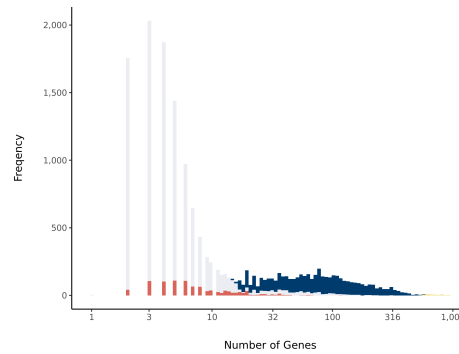
Number of Reads Processed	1e+06
Reads Pseudoaligned	90.1 %
Reads on Whitelist	94.73 %
Total UMI Counts	639,563
Sequencing Technology	10xv3
Species	<i>Arabidopsis thaliana</i>
Transcriptome	TAIR10

Sample Stats

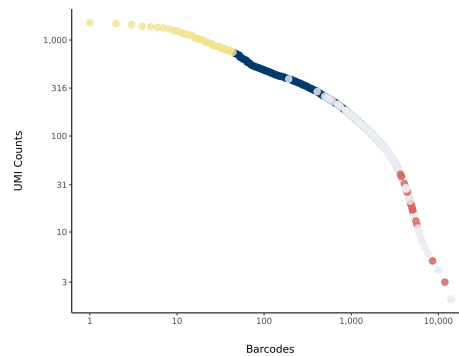
UMI Counts Histogram



Number of Genes Histogram



Barcode Rank Plot



col0 Summary

Processed by COPILOT

Summary

Analysis

Parameters

Iteration of Filtering	1
Mitochondrial Expression Threshold	5 %
Top High Quality Cell Filtered	1 %
Doublet Removed	Yes

Cell Stats

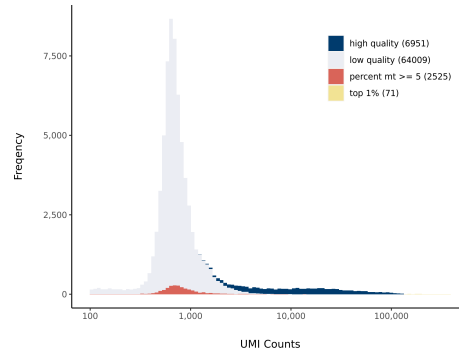
Estimated Number of High Quality Cell	6,951
High Quality Cell	9.45 %
Total UMI Counts in High Quality Cell	138,354,552
UMI Counts in High Quality Cell	67.3 %
Median UMI Counts per High Quality Cell	11,322
Median Genes per High Quality Cell	2,965
Total Genes Detected in High Quality Cell	24,702
Cell above Mitochondrial Expression Threshold	3.43 %
Estimated Doublet Rate in High Quality Cell	5.24 %

Sequencing Stats

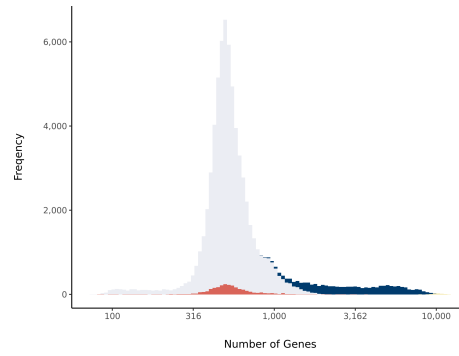
Number of Reads Processed	385,741,789
Reads Pseudoaligned	88.6 %
Reads on Whitelist	94.75 %
Total UMI Counts	205,592,307
Sequencing Technology	10xv3
Species	<i>Arabidopsis thaliana</i>
Transcriptome	TAIR10

Sample Stats

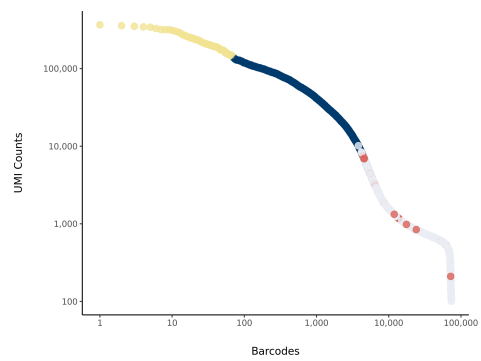
UMI Counts Histogram



Number of Genes Histogram



Barcode Rank Plot



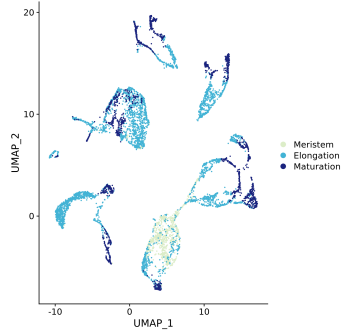
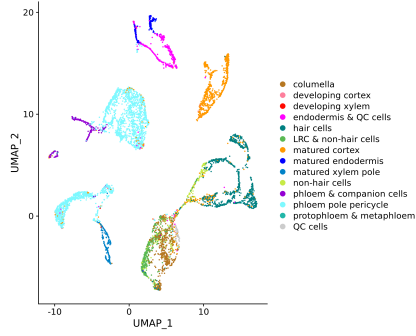
col0 Analysis

Processed by COPILOT

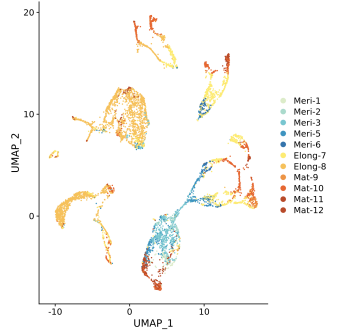
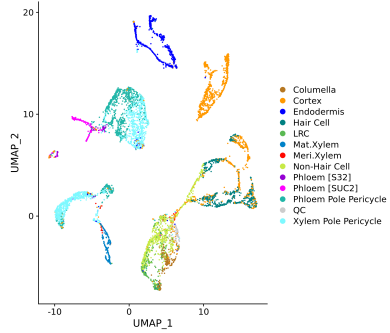
Summary

Analysis

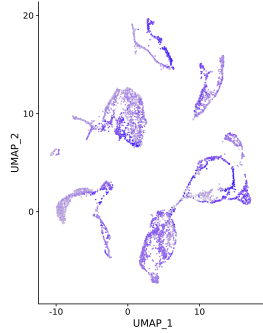
RNA Seq Annotation



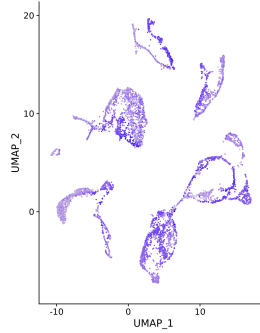
Microarray Annotation



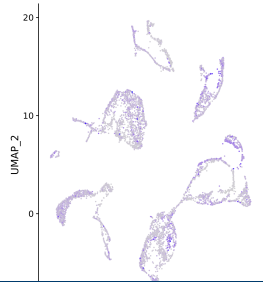
log10 UMI Counts



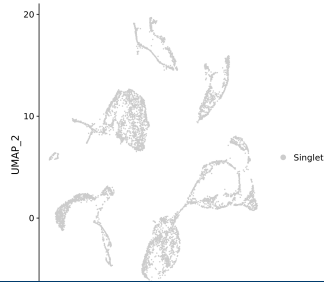
log10 Number of Genes



Percent Mitochondrial



Doublet Rate 5.24 %

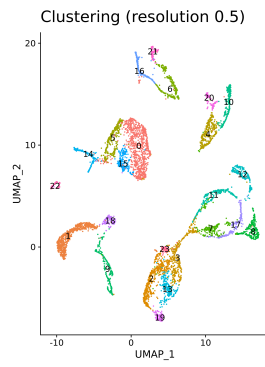


col0 Analysis

Processed by COPILOT

Summary

Analysis



pbmc_1k_v2 Summary

Processed by COPILOT

Summary

Analysis

Parameters

Iteration of Filtering	1
Mitochondrial Expression Threshold	5 %
Top High Quality Cell Filtered	1 %
Doublet Removed	No

Cell Stats

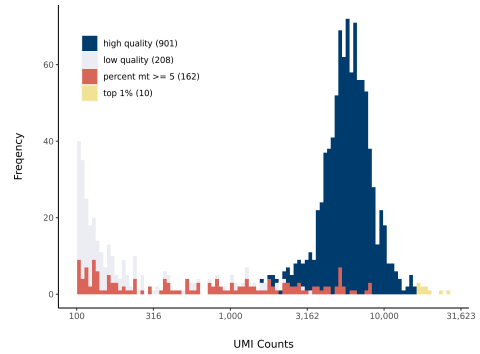
Estimated Number of High Quality Cell	901
High Quality Cell	70.34 %
Total UMI Counts in High Quality Cell	5,609,644
UMI Counts in High Quality Cell	91.86 %
Median UMI Counts per High Quality Cell	5,916
Median Genes per High Quality Cell	1,610
Total Genes Detected in High Quality Cell	15,578
Cell above Mitochondrial Expression Threshold	12.65 %
Estimated Doublet Rate in High Quality Cell	0.79 %

Sequencing Stats

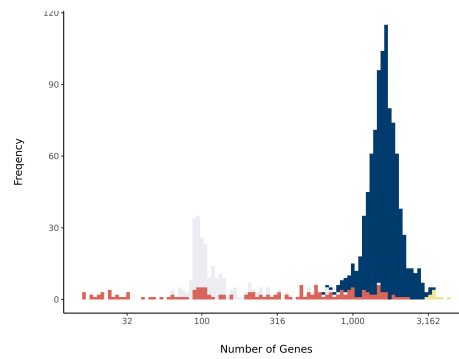
Number of Reads Processed	76,921,082
Reads Pseudoaligned	63 %
Reads on Whitelist	94.62 %
Total UMI Counts	6,106,428
Sequencing Technology	10xv2
Species	Homo sapiens
Transcriptome	hg38

Sample Stats

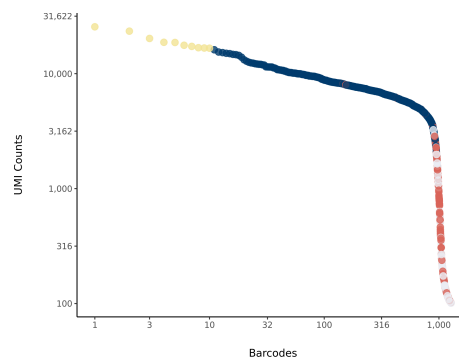
UMI Counts Histogram



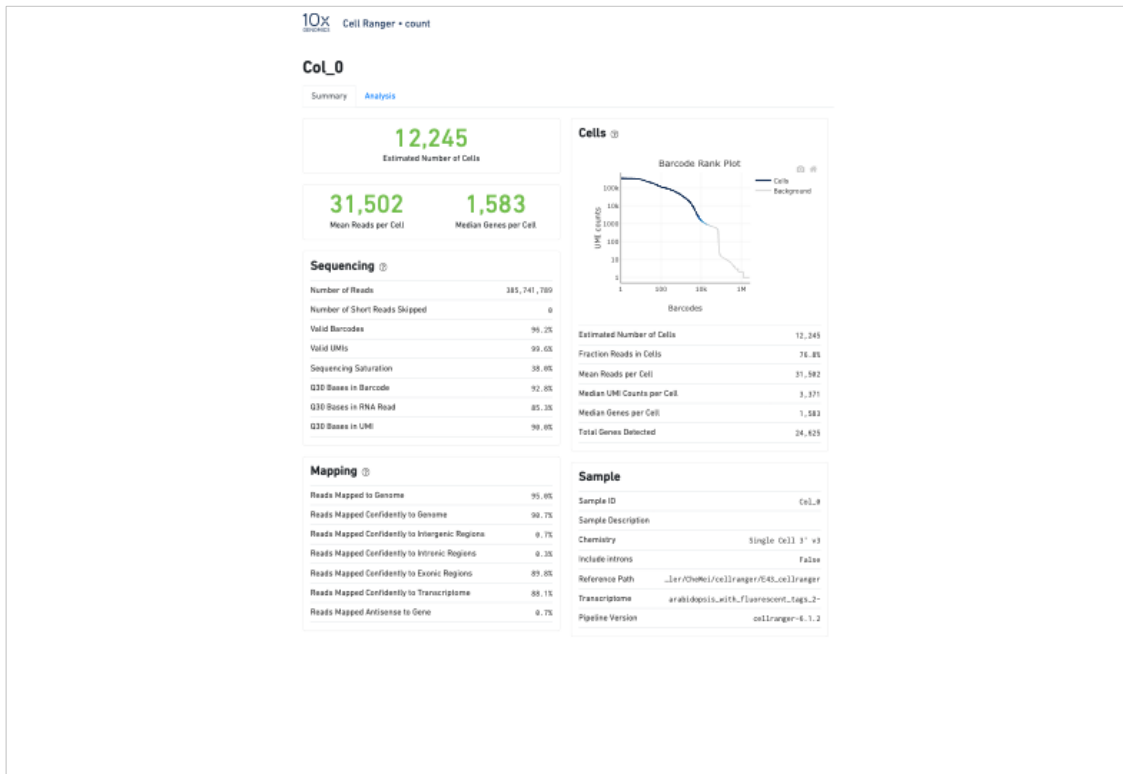
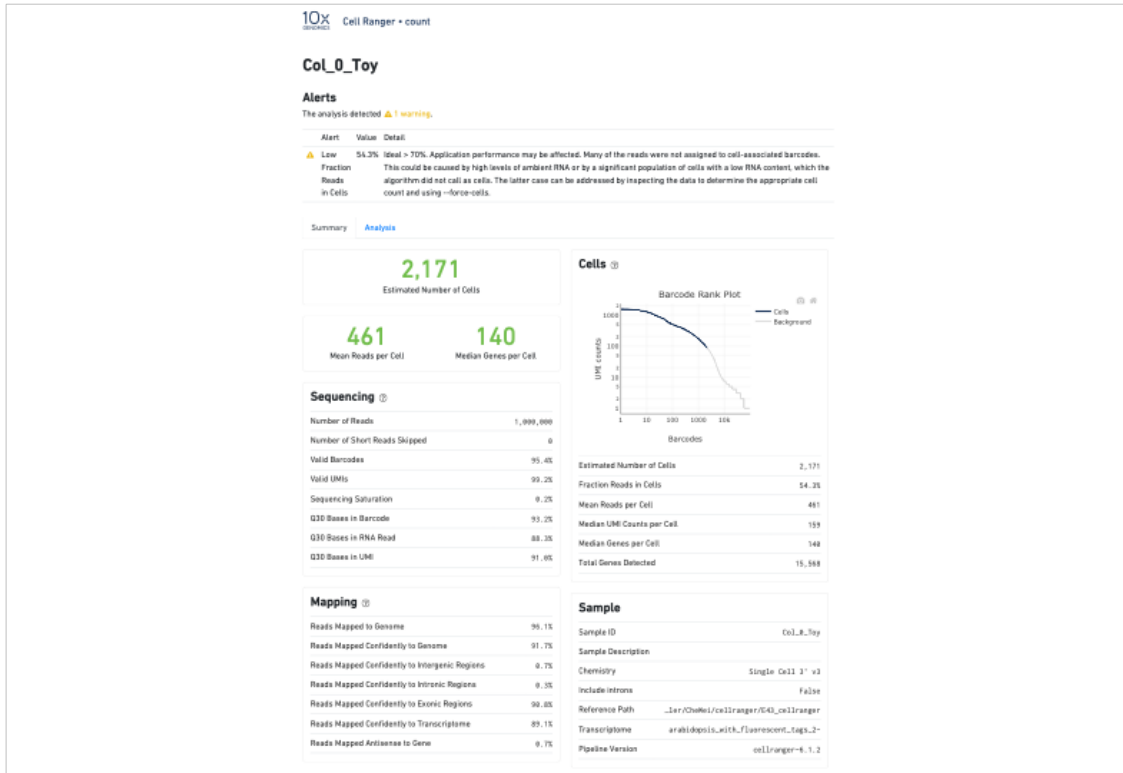
Number of Genes Histogram



Barcode Rank Plot



Dataset S2. Cell Ranger v6.1.2 summary files of col0 toy data, col0 full data, and PBMC data, related to Step “Run COPILOT for quality filtering”



pbmc_1k_v2

Summary Analysis

998
Estimated Number of Cells

77,075
Mean Reads per Cell

1,118
Median Genes per Cell

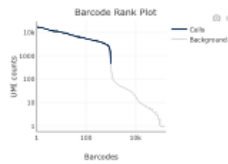
Sequencing

Number of Reads	74,921,682
Number of Short Reads Skipped	0
Valid Barcodes	99.6%
Valid UMIs	100.0%
Sequencing Saturation	99.6%
Q30 Bases in Barcode	94.6%
Q30 Bases in RNA Read	88.3%
Q30 Bases in UMI	92.8%

Mapping

Reads Mapped to Genome	93.2%
Reads Mapped Confidently to Genome	91.2%
Reads Mapped Confidently to Intergenic Regions	3.6%
Reads Mapped Confidently to Intronic Regions	22.2%
Reads Mapped Confidently to Exonic Regions	65.9%
Reads Mapped Confidently to Transcriptome	63.8%
Reads Mapped Antisense to Gene	1.1%

Cells



Estimated Number of Cells	998
Fraction Reads in Cells	84.1%
Mean Reads per Cell	77,875
Median UMI Counts per Cell	2,454
Median Genes per Cell	1,118
Total Genes Detected	17,247

Sample

Sample ID	pbmc_1k_v2
Sample Description	
Chemistry	Single Cell 3' v2
include introns	False
Reference Path	.../lib/Chrom/refs/seq-GRCh38-2020-A
Transcriptome	GRCh38-2020-A
Pipeline Version	cellranger-4.1.2