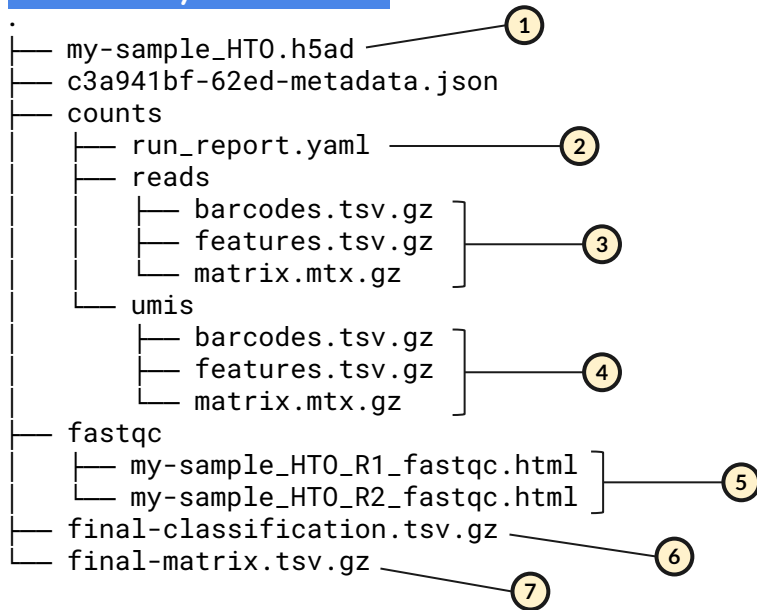


Understanding Outputs

In addition to cell barcode/UMI correction and quantification, cells are demultiplexed to their sample-of-origin and cross-sample doublets are identified.

Directory Structure



1. AnnData (h5ad)
2. Mapping rate, error correction, and a few other information
3. Read counts matrix
4. UMI counts matrix
5. Quality control checks on raw sequence data
6. Hashtag classification
7. Gene expression matrix + hashtag classification

AnnData (*.h5ad)

adata.obs

	unmapped	barcode_sequence	hash_id
227303201131438	27	CTGTGAACATCAGCGC	Doublet
191766864546027	16	GTACAACTCGCAATGT	A0252
227434332055453	7	CTGGACGCAAACTCTG	A0256
133921842280299	14	TCTACATTCGATTGGT	A0255
165239984667435	21	AGAAATGCAGGTGAGT	A0257
...
197281467784107	24	GACCCAGTCTGGGCGT	A0251
205861516597556	13	GCCTGTTCAATGAACA	A0251
196576840146219	30	GAGAAATAGCAGCAGT	Doublet
162072265964956	17	AACCTGATCGCTCCTA	A0257
122490813798755	9	TTCCTCTGTATGAGAT	A0251

10957 rows x 3 columns

adata.var

	feature_name
A0251-GTCAACTCTTTAGCG	CD56brightA+C-
A0252-TGATGGCCTATTGGG	CD56brightA+C+
A0253-TTCCGCCTCTCTTTG	CD56dimA+C+
A0254-AGTAAGTTCAGCGTA	CD56dimA+C-
A0255-AAGTATCGTTTCGCA	CD56dimA-C+NKp30hi
A0256-GGTTGCCAGATGTCA	CD56dimA-C+NKp30-
A0257-TGTCTTCTCGCCAG	CD56dimA-C-

adata.X

	A0251- GTCAACTCTTTAGCG	A0252- TGATGGCCTATTGGG	A0253- TTCCGCCTCTCTTTG	AGTAAGTTCA
227303201131438	0	408	329	
191766864546027	0	533	1	
227434332055453	0	0	0	
133921842280299	0	0	1	
165239984667435	1	1	2	
...	
197281467784107	423	0	2	
205861516597556	324	4	2	
196576840146219	0	572	2	
162072265964956	1	1	0	
122490813798755	259	1	2	

10957 rows x 7 columns

The cellular barcode "AACCTGATCGCTCCTA" is assigned to the hash ID "A0257" which is "CD56dimA-C-" in this example.

Hashtag Classification:
the hash ID assigned by the classification algorithm for the given cellular barcode.

Understanding Outputs

final-matrix.tsv.gz

Cell Barcodes

Genes

Classification

	CLUSTER	GNAI3	PBSN	CDC45	SCML2	NARF	CAV2	KLF6	SCMH1	COX5A	...	OLFR1246	RP23- 242K3.5	RP23- 181A8.2	RP23- 218I3.4	RP23- 13G1.8	RP23- 292G1.2	RP23- 13G1.9	WI1- 1735G24.3	RP24- 275P22.1	hashID
120703455222685	5	4	0	0	0	0	0	3	0	4	...	0	1	0	0	0	0	0	0	0	A0301
120726896884636	2	0	0	0	0	0	0	2	1	3	...	0	0	0	0	0	0	0	0	0	A0305
120728775940318	3	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	Doublet
120728791141286	0	0	0	0	0	0	0	114	0	6	...	0	0	0	0	0	0	0	0	0	A0302

Cluster

a cell x gene matrix from scRNA-seq
(pre-processed by SEQC)

Understanding Outputs

Cell Barcodes

Classification

hashID
120703408860404 HTO-301
120703409081700 HTO-302
120703409146292 HTO-301
120703424026909 HTO-301
120703436118757 HTO-301
120703436151085 HTO-301
120703436602228 HTO-301
120703455287707 HTO-301
120703455705899 HTO-302
120718441167717 HTO-301
120718441236211 HTO-301
120718441274589 HTO-301
120718455916774 HTO-301
120718456470838 Doublet
120718487414510 Doublet
120726896888174 HTO-301

“Doublet” here
could actually
mean two things.

①

Doublet or Multiplet:

e.g. Difficult to pick one (it could be either HTO_301 or HTO_302)

	HTO_301	HTO_302	HTO_303	HTO_304
GCGATCGCAGCTATAC	2052	3235	2	1
AAACGCTCATACATCG	2074	1088	1	0
ATGAGTCTCACGACTA	2129	2415	3	0
CTTGAGAAGGCACAAC	1946	2539	0	0

②

Negative:

e.g. Difficult to pick one because of the background noise or low number of reads/UMIs present

	HTO_301	HTO_302	HTO_303	HTO_304
AACCACACAAATACAG	17	4	19	1
TGTGATGAGTAGGGTC	22	10	18	0

final-classification.tsv.gz