

ASAP CRN Cloud Data File Manifest

Overview

This manifest provides a detailed summary of the outputs generated by the harmonized postmortem-derived brain sequencing (PMDBS) workflow. The workflow processes PMDBS data to harmonize across ASAP CRN teams, generating a range of data files and visualizations for scientific exploration. Currently there are 88 samples harmonized over two teams (Team Lee, Team Hafler) that can be found via the Verily workbench resource browser, or directly from Google Cloud Platform via `gsutil` or GCP console.

Files

The artifacts can be categorized “*preprocessed*” and “processed” , and “*integrated*”.

Harmonized Cohort Data

Cohort-level-curated-data :: Cloud ID `gs://asap-curated-data-cohort`

File	Type	Description	Location
<i>Integrated</i>			
Sample List	TSV	List of all samples contributing to the cohort level curated data	cohort_analysis/asap-cohort.sample_list.tsv
Cell Metadata	CSV	Full cell metadata from analyses including typing	cohort_analysis/asap-cohort.annotate_cells.metadata.csv

Integrated Clustered Annotated AnnData Object	H5AD	scVi integrated AnnData Object. Final output of cluster_data sub-workflow.	cohort_analysis/asap-cohort.adata_object.scvi_integrated.umap_cluster.annotate_cells.h5ad
Cell Types	CSV	Inferred cell-types	cohort_analysis/asap-cohort.asap-cohort.cell_types.csv
scVI Model	TAR. GZ	Trained scVI model	cohort_analysis/asap-cohort_scvi_model.tar.gz

ASAP Team Level Data

{team}-level-curated-data :: Cloud ID **asap-curated-data-{team}**

File	Type	Description	Location
<i>Preprocessed</i>			
Gene expression matrix (raw)	H5	Raw gene expression matrix produced during the Cellranger analysis in the preprocessing step.	preprocess/{ASAP_sample_id}.raw_feature_bc_matrix.h5
Gene expression matrix (filtered)	H5	Filtered gene expression matrix generated after preprocessing with Cellranger.	preprocess/{ASAP_sample_id}.filtered_feature_bc_matrix.h5
Molecule information	H5	Detailed molecular data generated during the Cellranger analysis.	preprocess/{ASAP_sample_id}.molecule_info.h5
Cellranger metrics	CSV	Metrics summary produced during the Cellranger analysis in the preprocessing step.	preprocess/{ASAP_sample_id}.metrics_summary.csv

Cell Bender	H5	Ambient RNA removed cell counts. (+ auxiliary cellbender artifacts)	preprocess/{ASAP_sample_id}.cellbender_filtered.h5
Raw counts	H5AD	AnnData object capturing raw counts. Output the preprocessing workflow.	preprocess/{ASAP_sample_id}.adata_object.h5ad
Processed			
Merged counts	H5AD	Merged AnnData object. Intermediate output of the filter_data sub-workflow.	preprocess/{ASAP_sample_id}.merged_adata_object.h5ad
Filtered counts	H5AD	Filtered AnnData object. Intermediate output of the filter_data sub-workflow.	preprocess/{ASAP_sample_id}.merged_adata_object_filtered.h5ad
Integrated			
integrated	H5AD	scVI integrated AnnData Object. Intermediate output of cluster_data sub-workflow.	preprocess/{team}.adata_object.scvi_integrated.h5ad
UMAP integrated	H5AD	scVi integrated AnnData Object with UMAP embedding. Intermediate output of cluster_data sub-workflow.	preprocess/{team}.adata_object.scvi_integrated.umap_cluster.h5ad
Cell Metadata	CSV	Full cell metadata from analyses including typing	cohort_analysis/{team}.annotate_cells.metadata.csv
Sample List	TSV	List of all samples contributing to the asap-cohort.sample_list.tsv	cohort_analysis/{team}.sample_list.tsv
Raw Fastqs	FASTQ	NOTE: *Requester Pays for data accessibility of the raw fastqs	/asap-raw-data-{team}/fastqs/
Additional artifacts	various	TBD	

Data Summary Plots (Harmonized Cohort & Team Level)

File	Type	Description	Location
Groups UMAP	PNG	`sample`, `batch`, `cell_type` labeled UMAP cell embedding.	cohort_analysis/{cohort team}.groups.umap.png
Features UMAP	PNG	`n_genes_by_counts`, `total_counts`, `pct_counts_mt`, `pct_counts_rb`, `doublet_score`, `S_score`, `G2M_score` labeled UMAP cell embedding.	cohort_analysis/{cohort team}.features.umap.png
QC N Genes by Count	PNG	QC violin plots for cell QC - n_genes_by_counts	cohort_analysis/{cohort team}.n_genes_by_counts.violin.png
QC Total Counts	PNG	QC violin plots for cell QC - total_counts	cohort_analysis/{cohort team}.total_counts.violin.png
QC Percent Mitochondrial	PNG	QC violin plots for cell QC - pct_counts_mt	cohort_analysis/{cohort team}.pct_counts_mt.violin.png
QC Percent Ribosomal	PNG	QC violin plots for cell QC - pct_counts_rb	cohort_analysis/{cohort team}.pct_counts_rb.violin.png
QC Doublet Scores	PNG	QC violin plots for cell QC - doublet_score	cohort_analysis/{cohort team}.doublet_score.violin.png

ETC.

File	Type	Description	Location
Reproducibility	ZIP	WDL workflows and scripts (R and python).	https://github.com/ASAP-CRN/harmonized-wf-dev/

code		From ASAP CRN github repo	
ASAP CRN Cloud File Manifest	PDF	this document	https://github.com/ASAP-CRN/harmonized-wf-dev/ASAP_CRN_file_manifest.pdf (UPDATE to v2)
Data Dictionary	PDF	ASAP CRN Data Dictionary, detailing the common data elements.	https://github.com/ASAP-CRN/harmonized-wf-dev/ASAP_CRN_data_dictionary.pdf (UPDATE to v2)

Buckets

Harmonized Cohort Data

Cohort-level-curated-data :: Cloud ID gs://asap-curated-data-cohort

Team-level Data

Team-level-curated-data :: Cloud ID gs://asap-curated-data-{team}

Team-level-raw-data :: Cloud ID gs://asap-raw-data-{team}

Version 1.1