

# 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

### Harmonized Cohort Data

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

File	Type	Description	Location
Sample List	TSV	List of all samples contributing to the cohort level curated data	<code>cohort_analysis/asap-cohort.sample_list.tsv</code>
Cell Metadata	CSV	Full cell metadata from analyses including typing	<code>cohort_analysis/asap-cohort.final_metadata.csv</code>
Harmony integrated Seurat Object	RDS	Harmony integrated Seurat Object. Intermediate output of cluster_data sub-workflow.	<code>preprocess/asap-cohort.seurat_object.harmony_integrated_04.rds</code>

Harmony integrated Seurat Object + neighbors	RDS	Harmony integrated Seurat Object with Neighbor Graph. Intermediate output of cluster_data sub-workflow.	preprocess/asap-cohort.seurat_object.harmony_integrated_neighbors_05.rds
Harmony integrated Seurat Object + UMAP	RDS	Harmony integrated Seurat Object with UMAP. Intermediate output of cluster_data sub-workflow.	preprocess/asap-cohort.seurat_object.harmony_integrated_neighbors_umap_06.rds
Harmony Integrated and Typed Data Object	RDS	UMAP Clustered Harmony integrated Seurat Object. Final output of cohort_analysis workflow and cluster_data sub-workflow.	cohort_analysis/asap-cohort.seurat_object.harmony_integrated_neighbors_umap_cluster_07.rds

## ASAP Team Level Data

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

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

Raw counts	RDS	Seurat object capturing raw counts. Output the preprocessing workflow.	preprocess/{sample_id}.seurat_object.preprocessed_01.rds
Filtered counts	RDS	Filtered Seurat object. Intermediate output of the filter_data sub-workflow.	preprocess/{sample_id}.seurat_object.preprocessed_filtered_02.rds
Filtered normalized counts	RDS	Filtered and Normalized Seurat object. Output of the filter_data sub-workflow.	preprocess/{sample_id}.seurat_object.preprocessed_filtered_normalized_03.rds
Harmony integrated	RDS	Harmony integrated Seurat Object. Intermediate output of cluster_data sub-workflow.	preprocess/{team}.seurat_object.harmony_integrated_04.rd
Harmony integrated with neighbors	RDS	Harmony integrated Seurat Object with Neighbor Graph. Intermediate output of cluster_data sub-workflow.	preprocess/{team}.seurat_object.harmony_integrated_neighbors_05.rd
UMAP of Harmony integrated neighbors	RDS	Harmony integrated Seurat Object with UMAP embedding. Intermediate output of cluster_data sub-workflow.	preprocess/{team}.seurat_object.harmony_integrated_neighbors_umap_06.rds
Integrated and Typed Data Object	RDS	UMAP Clustered Harmony integrated Seurat Object. Final output of cohort_analysis workflow and cluster_data sub-workflow.	cohort_analysis/{team}.seurat_object.harmony_integrated_neighbors_umap_cluster_07.rds
Cell Metadata	CSV	Full cell metadata from analyses including typing	cohort_analysis/{team}.final_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	FAST Q	NOTE: <b>*Requester Pays</b> 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
Batch group UMAP	PDF	Batch labeled UMAP cell embedding.	cohort_analysis/{cohort   team}.batch_group_umap.png
Doublet scores (QC) UMAP	PDF	Doublet score labeled UMAP cell embedding.	NOT platformed
Cell types UMAP	PDF	Major cell type labeled UMAP cell embedding.	cohort_analysis/{cohort   team}.major_type_module_umap.pdf
nCount RNA UMAP	PDF	nCount labeled UMAP cell embedding.	cohort_analysis/{cohort   team}.nCount_RNA_feature_umap.png
nFeature RNA UMAP	PDF	nFeature labeled UMAP cell embedding.	cohort_analysis/{cohort   team}.nFeature_RNA_feature_umap.png
Mitochondrial percentage (QC) UMAP	PDF	mp labeled UMAP plot from QC.	cohort_analysis/{cohort   team}.mt_feature_umap.png
Ribosomal percentage (QC) UMAP	PDF	rp labeled UMAP plot from QC.	cohort_analysis/{cohort   team}.rb_feature_umap.png
QC UMIs vs Genes Plot	PDF	Quality control UMI count vs Gene count scatter.	cohort_analysis/{cohort   team}.umis_genes_plot.png
QC Violin Distribution Plots	PDF	Violin plots for cell QC (nCount_RNA, nFeature_RNA, pct.mt, pct.rb, doublet_scores)	cohort_analysis/{cohort   team}.qc.violin_plots.png
Sample Grouping UMAP	PDF	UMAP plot showing sample grouping.	cohort_analysis/{cohort   team}.sample_group_umap.png
Seurat Clusters UMAP	PDF	UMAP visualization of Seurat clusters.	cohort_analysis/{cohort   team}.seurat_clusters_group_umap.png

ETC.

File	Type	Description	Location
Reproducibility code	ZIP	WDL workflows and scripts (R and python). From ASAP CRN github <a href="#">repo</a>	<a href="https://github.com/ASAP-CRN/harmonized-wf-dev/">https://github.com/ASAP-CRN/harmonized-wf-dev/</a>
ASAP CRN Cloud File Manifest	PDF	this document	<a href="https://github.com/ASAP-CRN/harmonized-wf-dev/ASAP_CRN_file_manifest.pdf">https://github.com/ASAP-CRN/harmonized-wf-dev/ASAP_CRN_file_manifest.pdf</a>
Data Dictionary	PDF	ASAP CRN Data Dictionary, detailing the common data elements.	<a href="https://github.com/ASAP-CRN/harmonized-wf-dev/ASAP_CRN_data_dictionary.pdf">https://github.com/ASAP-CRN/harmonized-wf-dev/ASAP_CRN_data_dictionary.pdf</a>

## 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}