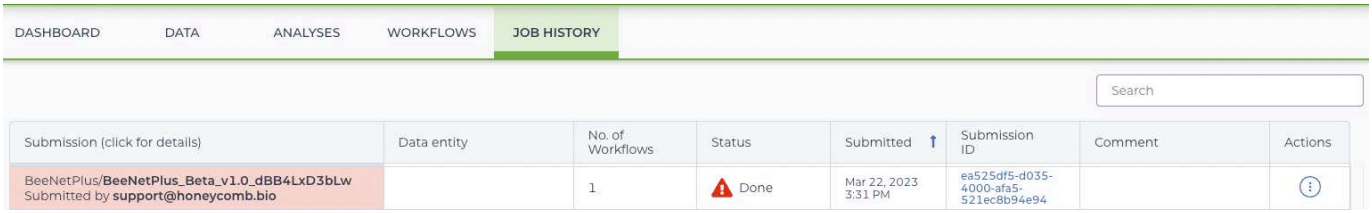
BeeNetPLUS Troubleshooting Guide

OVERVIEW

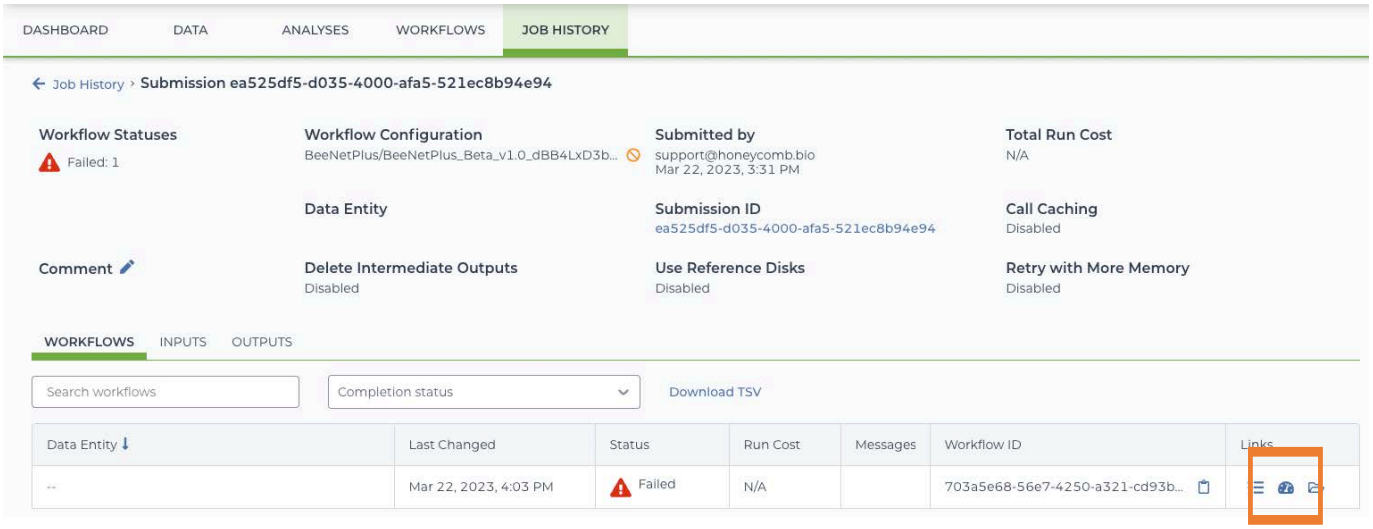
This document is to serve as a resource for the user troubleshooting during use of BeeNetPLUS hosted on Terra. For additional help, submit a support request at **support.honeycomb.bio** to be connected with a member of the Bioinformatics team.

IDENTIFYING PIPELINE FAILURE

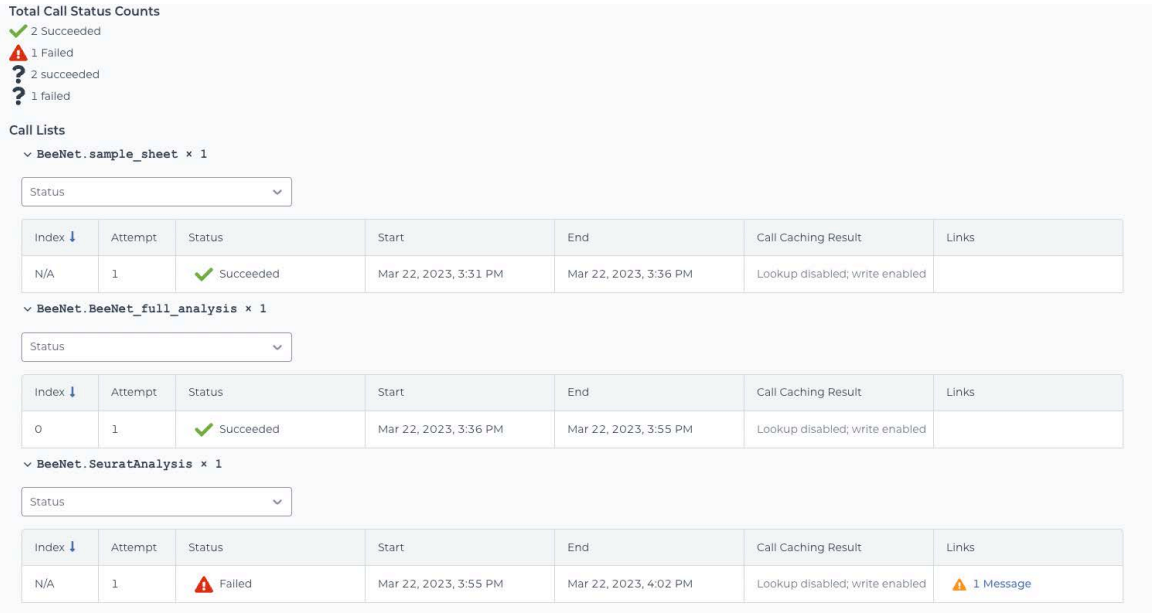
Job failures can happen at any step of the workflow.



To check the step that the pipeline failed, click on the red submission ID box which brings you to the Job Submission page for this job. Under the "Links" tab, select the Workflow Dashboard.



This page will summarize each task and demonstrate which part of the pipeline failed. Under the "Call Lists" section, if you expand each section it will show each task and its status.

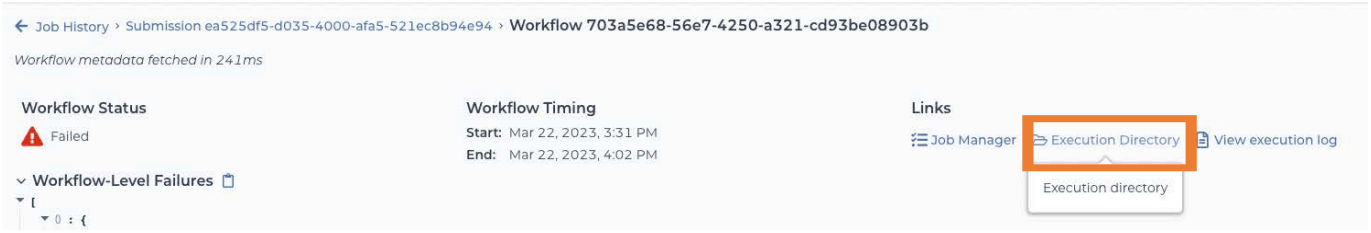


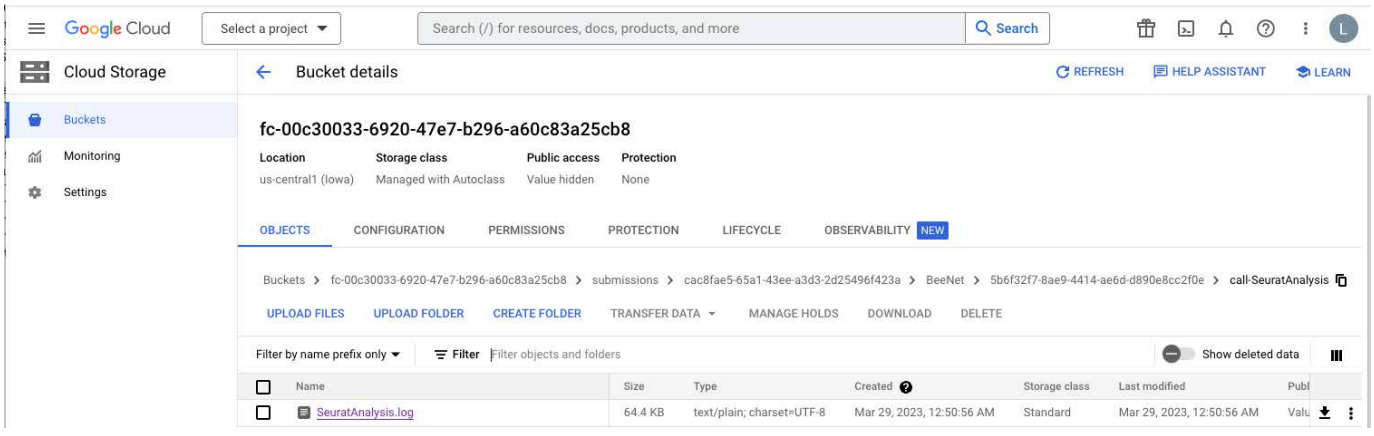
In this case, the workflow failed in downstream analysis (BeeNet.SeuratAnalysis).

FINDING THE ERROR LOG

Your error log will be output into the user-specified OutProcessed directory. Click on the log file to open it. Save it using File -> Save As on your internet browser. Save the file and include this text file in your Troubleshooting email.

Alternately, on the Workflow Dashboard, click on the "Execution Directory" Link at the top of the page. This will bring you to the Google Cloud Workflow. Click on the failed task, and locate the log file associated with that task. Click on the log file to open it. Save it using File -> Save As on your internet browser. Save the file and include this text file in your Troubleshooting email.





COMMON PIPELINE FAILURES

Task SampleSheet

Potential Cause: Insufficient Google Cloud Permissions

Recommendation: Ensure that Terra has correct permission for Google Cloud Buckets and Docker Images

Potential Cause: Error in Input Sheet.

Most likely errors involve:

- Missing data
- Missing column headers
- Misspelled column headers
- More than one value in "Dataset" column
- Spaces in Google Cloud output folder (OutProcessed file path cannot contain spaces)
- Spaces in Google Cloud data folder (Data file path cannot contain spaces)
- File Format is .csv or .xslx (must be .txt)
- File Format is not tab-separated

Recommendation: Correct errors and resubmit pipeline. Uncheck "Use Call Caching"

Task BeeNet

Potential Cause: Error in Input Sheet

Most errors at this step are caused by errors in the input sheet. Errors that cause failure at this step include:

- Misspelled reference genome information
- Reference genome does not match input tag ("hg37", "hg38", "mm10" or a custom tag from Support Team)
- R1 for both files (instead of R1/R2 pair)
- R2 for both files (instead of R1/R2 pair)
- mismatched R1/R2 pairs
- Special Characters in Sample Name
- numBC is more than the number of cells loaded into the HIVE (CellInput)

Recommendation: Correct errors and reupload sample sheet. Rerun the pipeline. Uncheck "Use Call Caching"

Task SeuratAnalysis

Potential Cause: Memory Allocation (designated by "Killed" error in log file)

Recommendation: Rerun the pipeline with increased SeuratRAM. Keep "Use Call Caching" checked to restart the pipeline at this task.

Potential Cause: Insufficient cell recovery leads to downstream code issues

Recommendation:

- Rerun with increased numBC
- Rerun with lower gene and threshold minimums

Potential Cause: Unacceptable output file path leads to no files in output bucket

Recommendation: Output file location must be a Google Cloud bucket. The text must be a file path that is wrapped in quotation marks and starts with "gs://". Example: "gs://output-file-directory". This path must not have spaces (check the last character!). Make sure that the specified output bucket also has the correct permissions.

WHEN TO USE CACHE CALLING (PIPELINE RERUN)

Cache calling is a mechanism by which the workflow can be restarted at a particular task if the previous tasks were successful. Cache calling should only be used to restart the BeeNetPLUS workflow when the pipeline fails due to memory or storage limitations. If you are rerunning a successful pipeline with different gene and transcript thresholds, the pipeline will start from the beginning whether "check call caching" is checked or not.

ANALYSIS TROUBLESHOOTING (PIPELINE SUCCESS)

Issue: Low genome mapping (< ~50% in SampleQC table)

Potential cause:

- Wrong reference genome used for alignment

Recommendation:

- Make sure to use the correct species-specific tag for alignment and annotation when running BeeNetPLUS. Contact the support team for issues with custom references.

Issue: Low exon mapping (<~20%)

Potential cause:

- Poor library quality
- Poor sequencing quality
- Low cell viability

Recommendation:

- Check sequencing quality (FastQC) and library quality (qubit and fragment analyzer)

Issue: High PolyA or primer reads (>25%)

Potential Cause:

- Library preparation errors in 1st or 2nd strand synthesis
- Failure to capture cells in the HIVE

Recommendation:

- Please refer to the HIVE CLX scRNAseq Troubleshooting Guide for more details

Issue: High BadBaseBC (in SampleQC table)

Potential cause:

- Base quality of read 1 is below Q30

Recommendation:

- Check sequencing quality (FastQC)
- Consult sequencing provider for further troubleshooting

Issue: Mostly zero values in CM Summary File

Potential cause:

- Library preparation error
- Failure to capture cells on the HIVE
- Low input cell number

Recommendation:

- Please refer to the HIVE CLX scRNAseq Troubleshooting Guide for more details

Issue: Output report is formatted incorrectly

Potential Cause:

- Report was downloaded directly from the Internet Browser using "File -> Save Page As"

Recommendation:

- The zipped "Analysis" folder contains a copy of the properly formatted Output Report

Issue: The output report does not contain cell type information

Potential cause:

- Sample type did not match auto-annotation tags
- Reference genome is incompatible with auto-annotation datasets

Recommendation:

- If you feel like your sample type should qualify for HCB data auto annotation (Filtered Blood, Bone Marrow, or PBMCs) , check the Sample Sheet for errors. Correct errors and reupload sample sheet. Rerun the pipeline. Uncheck "Use Call Caching" when restarting the pipeline to start the pipeline from the beginning.

Issue: Low cell/gene/transcript thresholds

Potential cause:

- Incorrect numBC
- Incorrect gene and/or transcript thresholds
- Library preparation error
- Failure to capture cells on the HIVE
- Low input cell number

Recommendation:

- Adjust numBC and/or cell input parameters on Sample Sheet and reupload. Rerun the pipeline. Uncheck "Use Call Caching" when restarting the pipeline to start the pipeline from the beginning.
- Restart the workflow with different gene and transcript thresholds. Uncheck "Use Call Caching" when restarting the pipeline to start the pipeline from the beginning.
- Please refer to the HIVE CLX scRNAseq Troubleshooting Guide for more details.

Issue: Number of cells recovered is equal to numBC after filtering

Potential cause:

- Cell recovery is higher than expected

Recommendation:

- Adjust numBC and/or cell input parameters on Sample Sheet and reupload. Rerun the pipeline. Uncheck "Use Call Caching" when restarting the pipeline to start the pipeline from the beginning.

Issue: Cluster expresses marker genes for >1 distinct cell type

Potential Cause:

- This cell cluster contains multiplets

Recommendation:

- Manually remove these cells from your analysis and re-cluster using the R object provided with the help of your informatics core.

Issue: Cluster is distributed diffusely across the UMAP

Potential cause:

- Low quality cluster

Recommendation:

- Restart the workflow with different gene and transcript thresholds. Uncheck "Use Call Caching" when restarting the pipeline to start the pipeline from the beginning.
- Manually remove these cells from your analysis and re-cluster using the R object provided with the help of your informatics core.