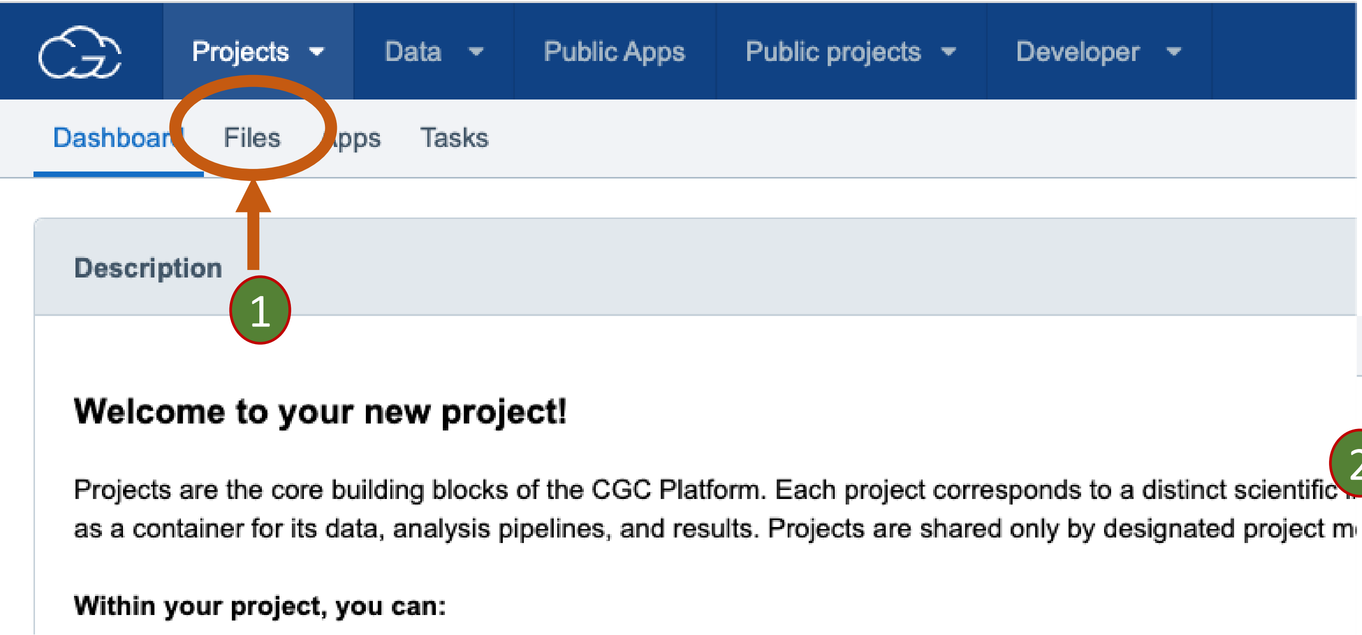
Setup

We’re aware some of your tasks have failed and have determined that the issue was the number of duplicate files in the main project which were copied over

This can be resolved by re-running the task with fewer files and have made instructions on how to do this in your projects

Delete Old Files

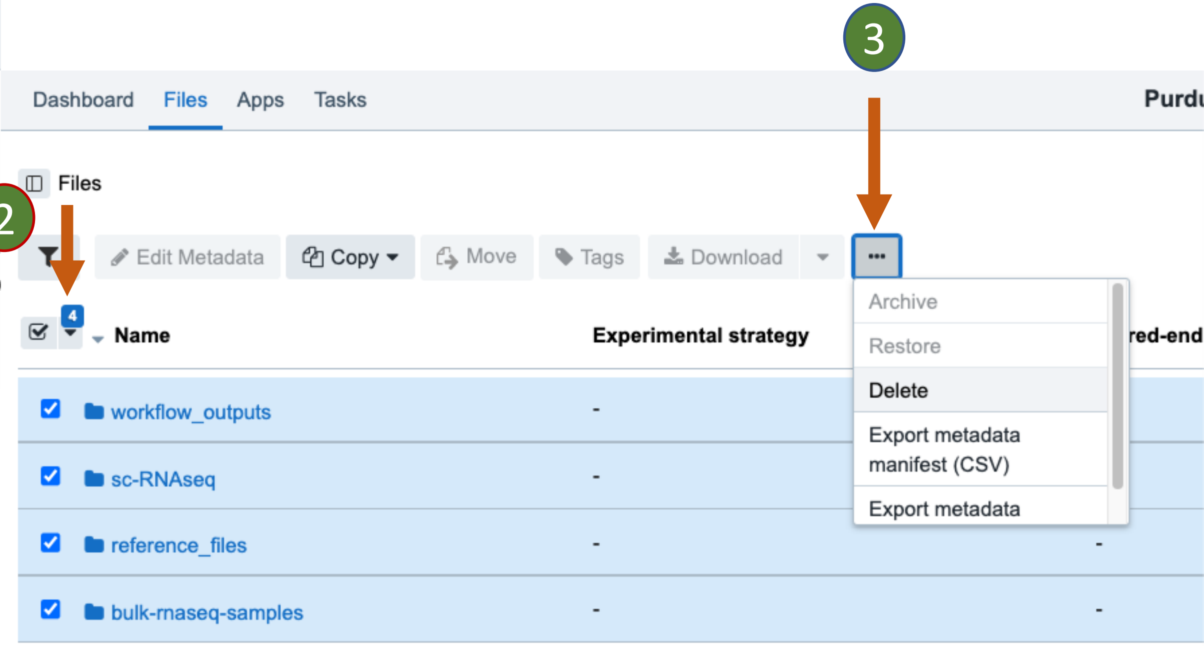
1. Navigate to your project and click on the **Files** tab beneath the blue navigation bar
2. Select all of the folders and click on the **three dots** next to the download button



Welcome to your new project!

Projects are the core building blocks of the CGC Platform. Each project corresponds to a distinct scientific i
as a container for its data, analysis pipelines, and results. Projects are shared only by designated project m

Within your project, you can:



	Name	Experimental strategy	
☑	workflow_outputs	-	
☑	sc-RNAseq	-	
☑	reference_files	-	-
☑	bulk-rnaseq-samples	-	-

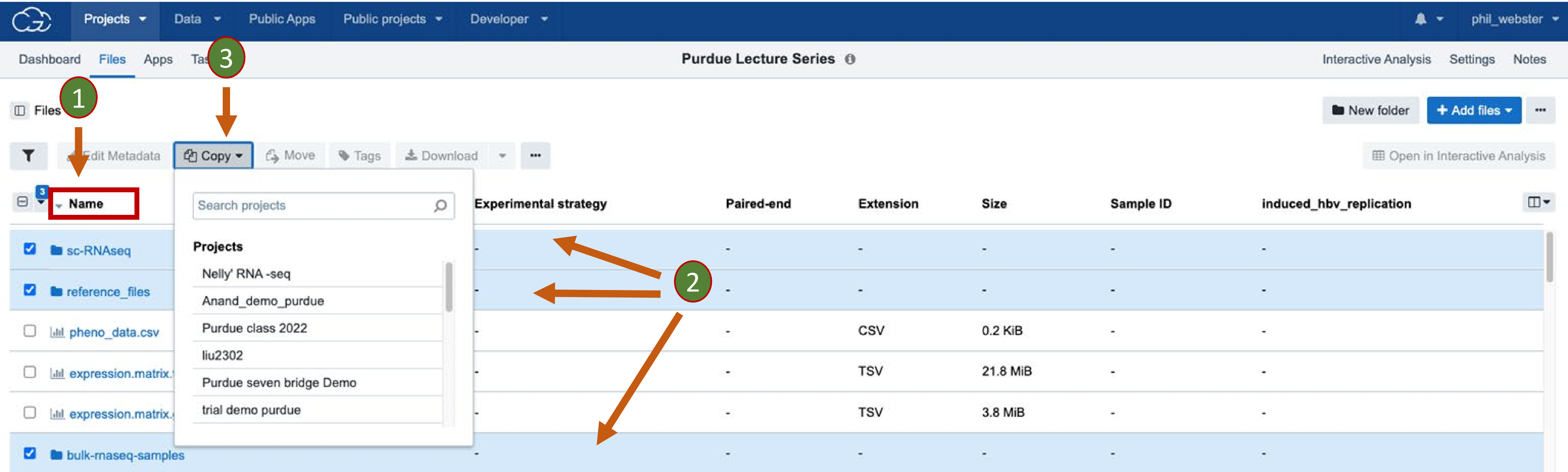
Copy Files Over To Your Project From Purdue Lecture Series

1. Navigate to the project **Purdue Lecture Series**

2. Sort by **Name**

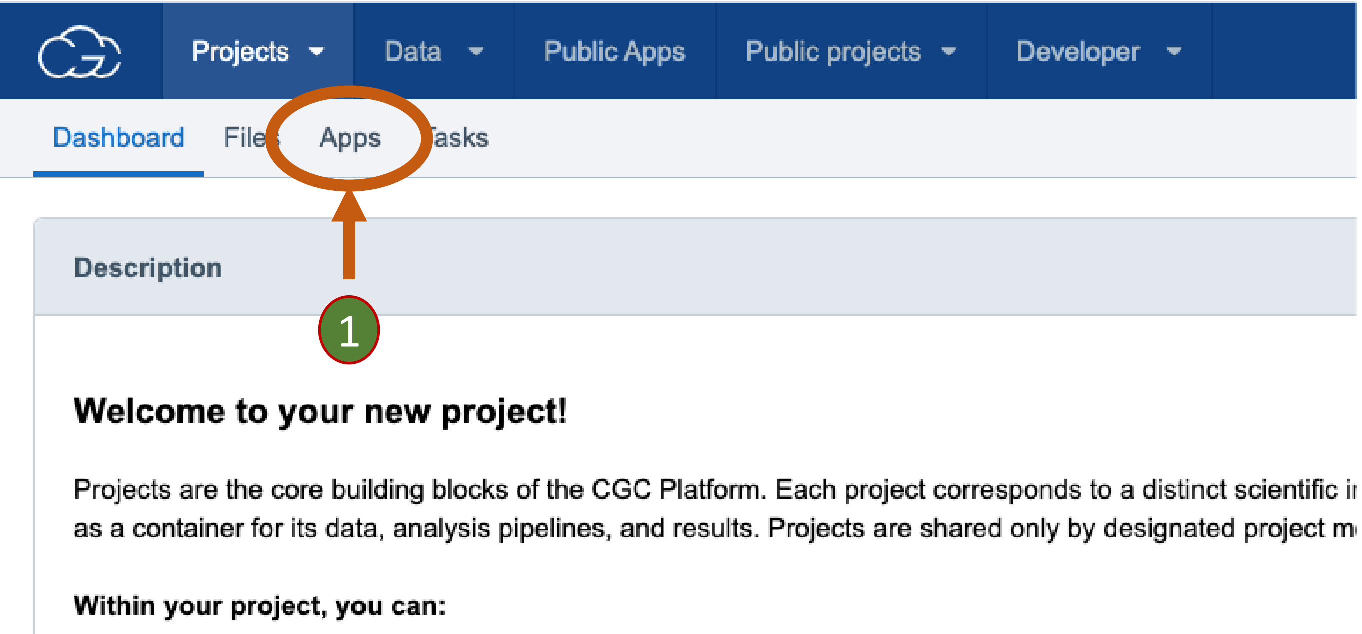
3. Select the **sc-RNAseq, reference_files,** and **bulk-rnaseq-samples** folders

4. Click **copy** then select your project



	Name	Experimental strategy	Paired-end	Extension	Size	Sample ID	induced_hbv_replication
☑	sc-RNAseq	-	-	-	-	-	-
☑	reference_files	-	-	-	-	-	-
☐	pheno_data.csv	-	-	CSV	0.2 KiB	-	-
☐	expression.matrix.	-	-	TSV	21.8 MiB	-	-
☐	expression.matrix.	-	-	TSV	3.8 MiB	-	-
☑	bulk-rnaseq-samples	-	-	-	-	-	-

Navigating to the Workflow



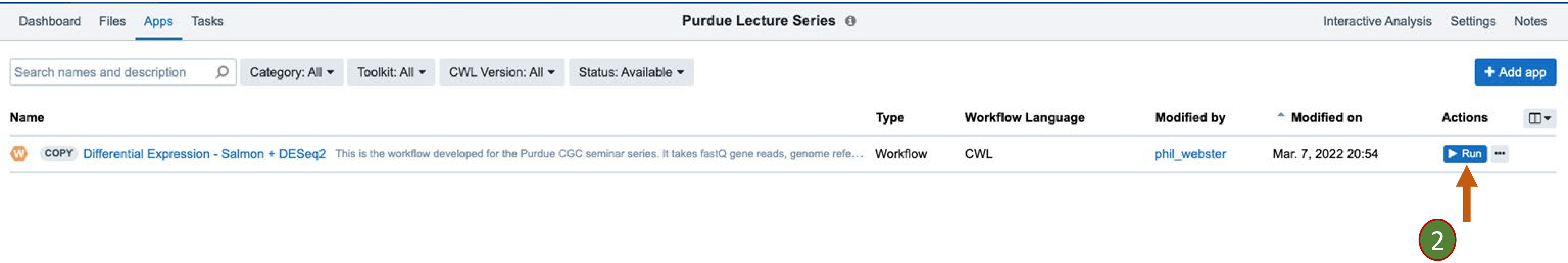
Description

Welcome to your new project!

Projects are the core building blocks of the CGC Platform. Each project corresponds to a distinct scientific i
as a container for its data, analysis pipelines, and results. Projects are shared only by designated project m

Within your project, you can:

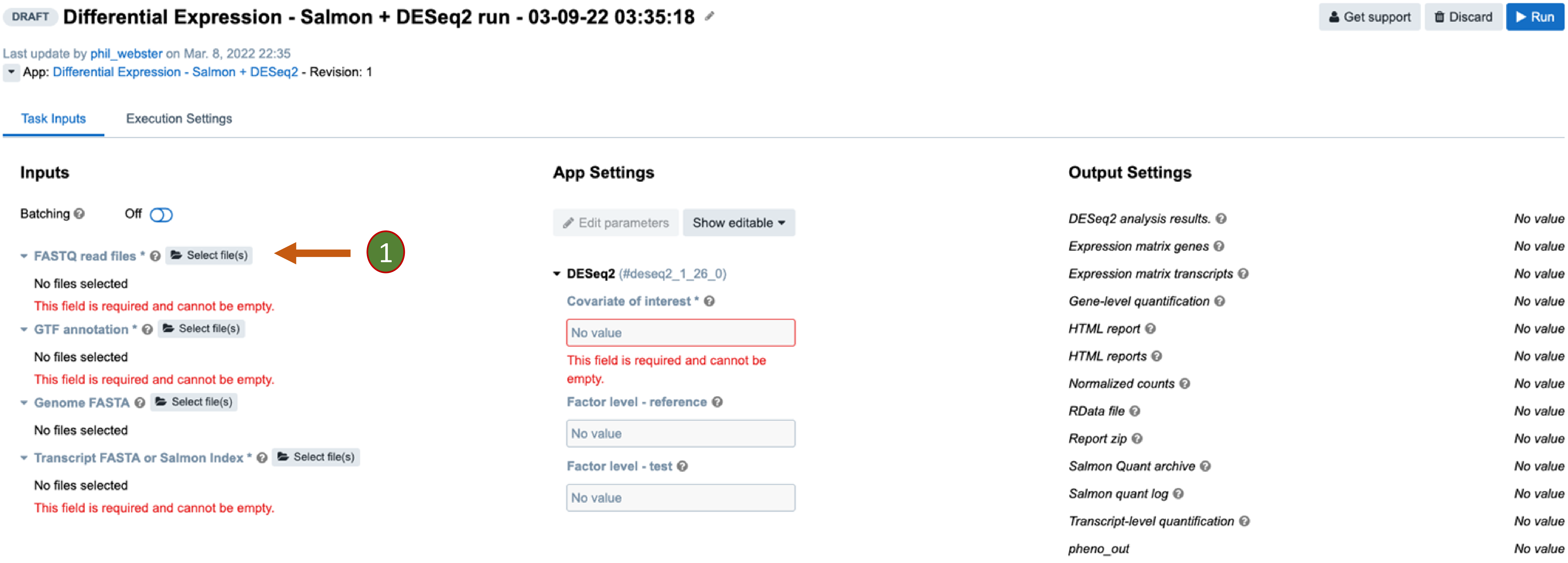
1. Navigate to **your project**. (Not **Purdue Lecture Series** project).
2. Click on the **Apps** tab under the blue navigation bar. The workflow should already be there from class.
3. Then click the **Run** button next to the workflow we copied over to our projects



Name	Type	Workflow Language	Modified by	Modified on	Actions
COPY Differential Expression - Salmon + DESeq2 This is the workflow developed for the Purdue CGC seminar series. It takes fastQ gene reads, genome refe…	Workflow	CWL	phil_webster	Mar. 7, 2022 20:54	▶ Run ···

Setup The Workflow: FastQ Files

1. We will fill out our three inputs, begin by clicking **Select File(s)** next to **FASTQ read files**



DRAFT **Differential Expression - Salmon + DESeq2 run - 03-09-22 03:35:18**

Last update by phil_webster on Mar. 8, 2022 22:35

App: Differential Expression - Salmon + DESeq2 - Revision: 1

Inputs

Batching Off

FASTQ read files * Select file(s)

No files selected

This field is required and cannot be empty.

GTF annotation * Select file(s)

No files selected

This field is required and cannot be empty.

Genome FASTA Select file(s)

No files selected

Transcript FASTA or Salmon Index * Select file(s)

No files selected

This field is required and cannot be empty.

App Settings

DESeq2 (#deseq2_1_26_0)

Covariate of interest *: No value

This field is required and cannot be empty.

Factor level - reference: No value

Factor level - test: No value

Output Settings

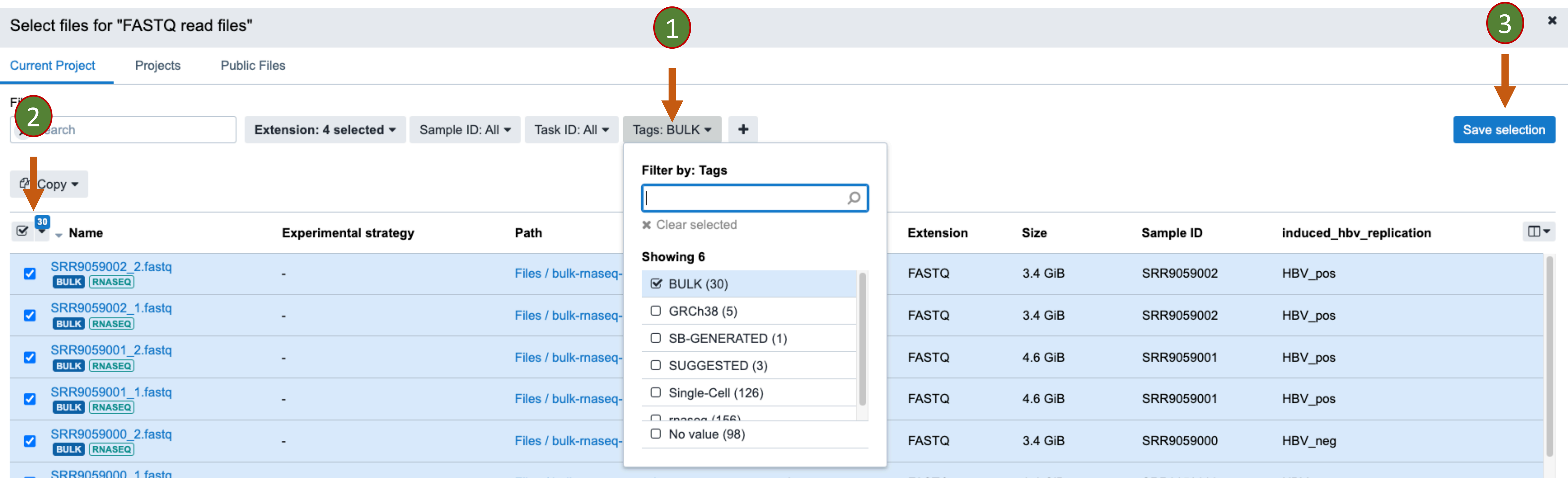
DESeq2 analysis results.	*No value*
Expression matrix genes	*No value*
Expression matrix transcripts	*No value*
Gene-level quantification	*No value*
HTML report	*No value*
HTML reports	*No value*
Normalized counts	*No value*
RData file	*No value*
Report zip	*No value*
Salmon Quant archive	*No value*
Salmon quant log	*No value*
Transcript-level quantification	*No value*
pheno_out	*No value*

Selecting FastQ Files

1. When selecting the FastQ files first click on **Tags** at the end of the search bar next to the **+** symbol and select **BULK**

2. The 154 files should filter down to 30

3. Select all files and click **Save selection**

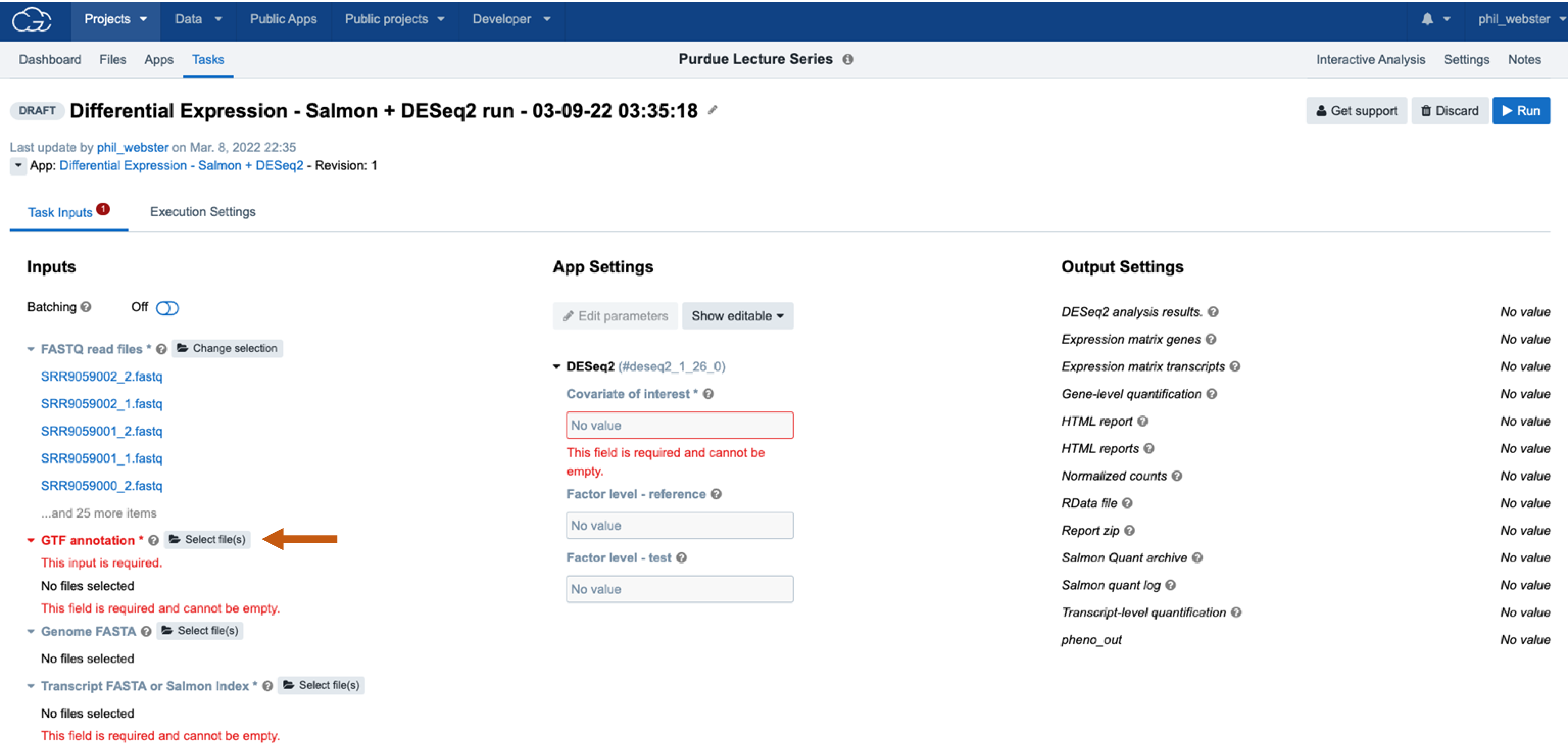


Extension: 4 selected
Sample ID: All
Task ID: All
Tags: BULK
+

Showing 6
☑ BULK (30)
☐ GRCh38 (5)
☐ SB-GENERATED (1)
☐ SUGGESTED (3)
☐ Single-Cell (126)
☐ rnaseq (156)
☐ No value (98)

☑	Name	Experimental strategy	Path	Extension	Size	Sample ID	induced_hbv_replication
☑	SRR9059002_2.fastq BULK RNASEQ	-	Files / bulk-rnaseq-	FASTQ	3.4 GiB	SRR9059002	HBV_pos
☑	SRR9059002_1.fastq BULK RNASEQ	-	Files / bulk-rnaseq-	FASTQ	3.4 GiB	SRR9059002	HBV_pos
☑	SRR9059001_2.fastq BULK RNASEQ	-	Files / bulk-rnaseq-	FASTQ	4.6 GiB	SRR9059001	HBV_pos
☑	SRR9059001_1.fastq BULK RNASEQ	-	Files / bulk-rnaseq-	FASTQ	4.6 GiB	SRR9059001	HBV_pos
☑	SRR9059000_2.fastq BULK RNASEQ	-	Files / bulk-rnaseq-	FASTQ	3.4 GiB	SRR9059000	HBV_neg
	SRR9059000_1.fastq						

Selecting GTF Annotation Files



Purdue Lecture Series

DRAFT **Differential Expression - Salmon + DESeq2 run - 03-09-22 03:35:18**

Last update by phil_webster on Mar. 8, 2022 22:35

App: Differential Expression - Salmon + DESeq2 - Revision: 1

Task Inputs 1 | Execution Settings

Inputs

Batching Off

FASTQ read files *

SRR9059002_2.fastq
SRR9059002_1.fastq
SRR9059001_2.fastq
SRR9059001_1.fastq
SRR9059000_2.fastq
...and 25 more items

GTF annotation *

This input is required.
No files selected
This field is required and cannot be empty.

Genome FASTA

No files selected

Transcript FASTA or Salmon Index *

No files selected
This field is required and cannot be empty.

App Settings

DESeq2 (#deseq2_1_26_0)

Covariate of interest *: No value
This field is required and cannot be empty.

Factor level - reference: No value

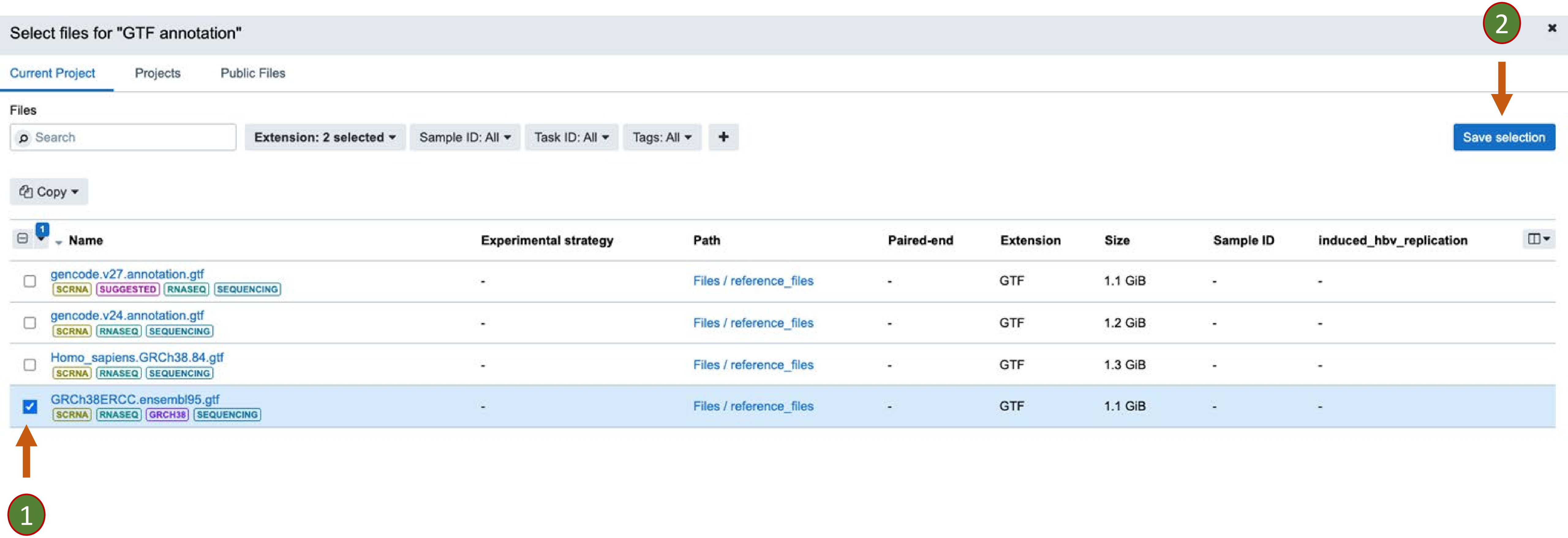
Factor level - test: No value

Output Settings

DESeq2 analysis results.	*No value*
Expression matrix genes	*No value*
Expression matrix transcripts	*No value*
Gene-level quantification	*No value*
HTML report	*No value*
HTML reports	*No value*
Normalized counts	*No value*
RData file	*No value*
Report zip	*No value*
Salmon Quant archive	*No value*
Salmon quant log	*No value*
Transcript-level quantification	*No value*
pheno_out	*No value*

Selecting GTF Annotation Files

1. Select **GRCh38ERCC.endemble95.gtf**
2. click **Save selection**



Select files for "GTF annotation" ×

Current Project | Projects | Public Files

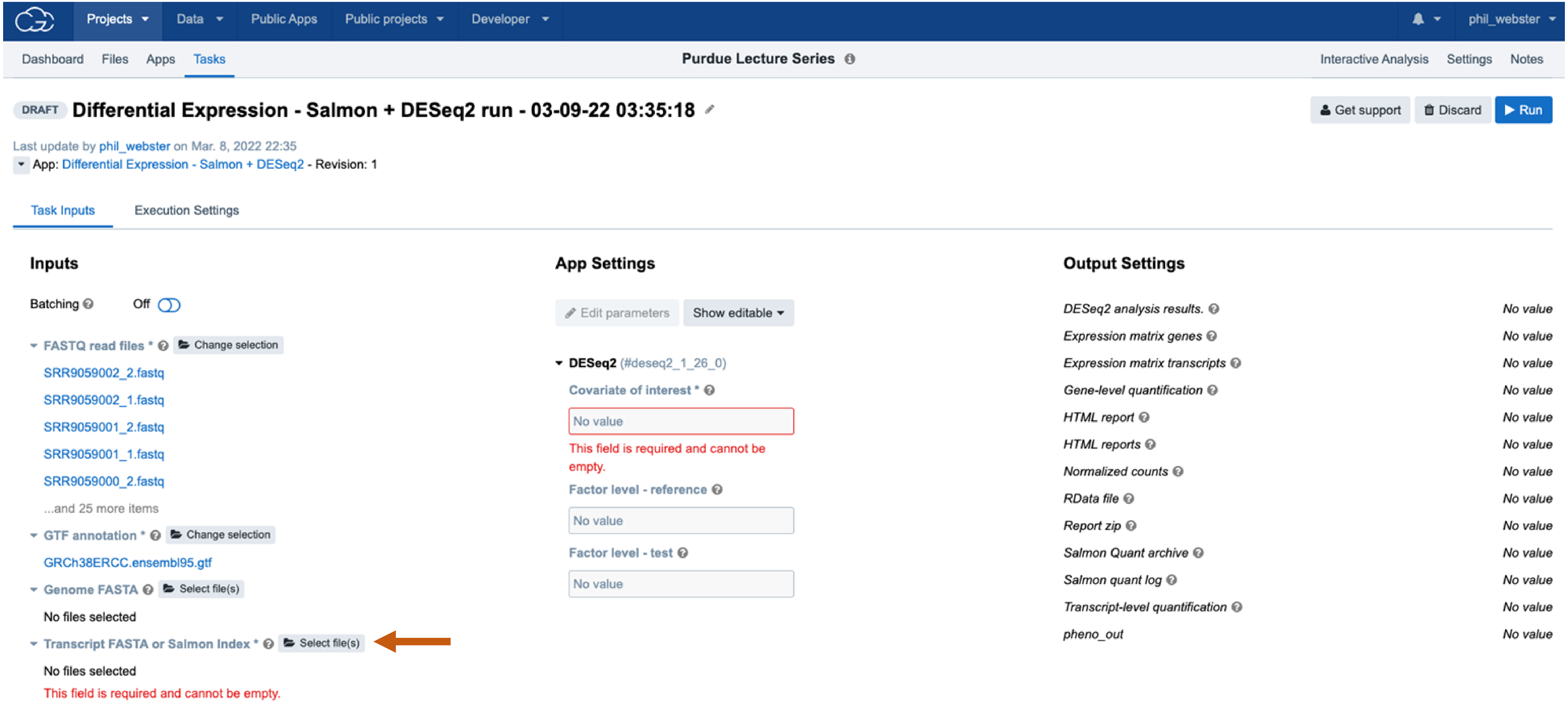
Files

Search | Extension: 2 selected | Sample ID: All | Task ID: All | Tags: All | + | Save selection

Copy

☐ 1	Name	Experimental strategy	Path	Paired-end	Extension	Size	Sample ID	induced_hbv_replication
☐	gencode.v27.annotation.gtf SCRNA SUGGESTED RNASEQ SEQUENCING	-	Files / reference_files	-	GTF	1.1 GiB	-	-
☐	gencode.v24.annotation.gtf SCRNA RNASEQ SEQUENCING	-	Files / reference_files	-	GTF	1.2 GiB	-	-
☐	Homo_sapiens.GRCh38.84.gtf SCRNA RNASEQ SEQUENCING	-	Files / reference_files	-	GTF	1.3 GiB	-	-
☑	GRCh38ERCC.ensembl95.gtf SCRNA RNASEQ GRCH38 SEQUENCING	-	Files / reference_files	-	GTF	1.1 GiB	-	-

Selecting Transcript Salmon Index File



DRAFT **Differential Expression - Salmon + DESeq2 run - 03-09-22 03:35:18**

Last update by phil_webster on Mar. 8, 2022 22:35

App: Differential Expression - Salmon + DESeq2 - Revision: 1

Inputs

Batching Off

FASTQ read files * Change selection

- SRR9059002_2.fastq
- SRR9059002_1.fastq
- SRR9059001_2.fastq
- SRR9059001_1.fastq
- SRR9059000_2.fastq
- ...and 25 more items

GTF annotation * Change selection

GRCh38ERCC.ensembl95.gtf

Genome FASTA Select file(s)

No files selected

Transcript FASTA or Salmon Index * Select file(s)

No files selected

This field is required and cannot be empty.

App Settings

DESeq2 (#deseq2_1_26_0)

Covariate of interest *: No value

This field is required and cannot be empty.

Factor level - reference: No value

Factor level - test: No value

Output Settings

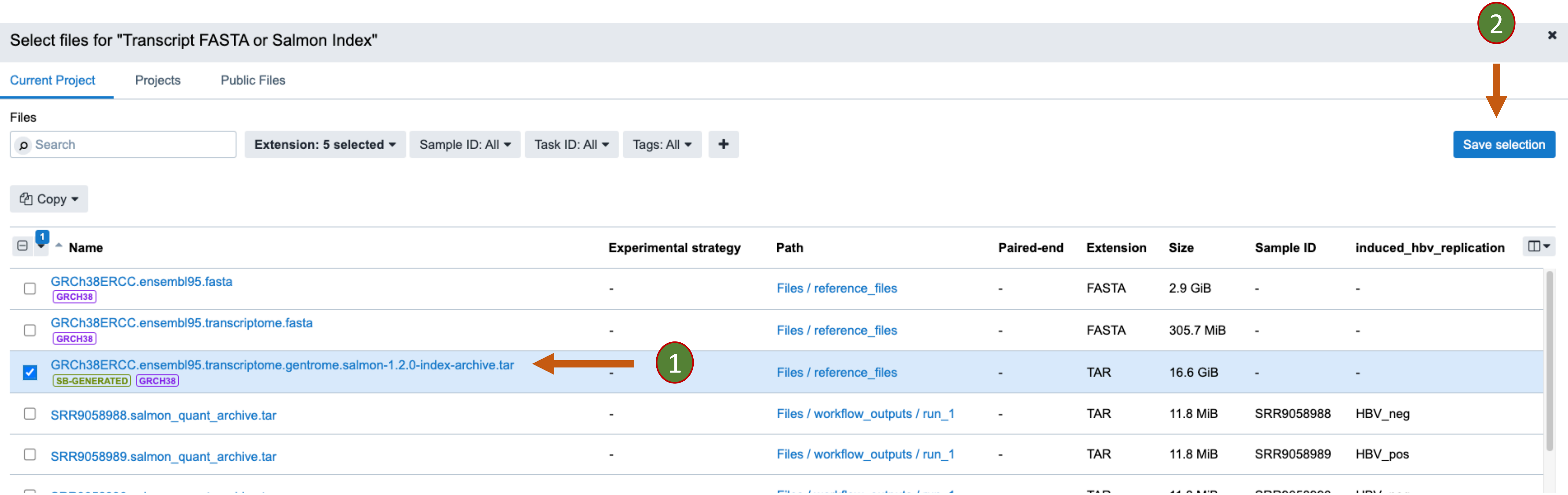
DESeq2 analysis results.	*No value*
Expression matrix genes	*No value*
Expression matrix transcripts	*No value*
Gene-level quantification	*No value*
HTML report	*No value*
HTML reports	*No value*
Normalized counts	*No value*
RData file	*No value*
Report zip	*No value*
Salmon Quant archive	*No value*
Salmon quant log	*No value*
Transcript-level quantification	*No value*
pheno_out	*No value*

Selecting Transcript FASTA/Salmon Index File

1. Sort by **Name**

2. Select **GRCh38ERCC.ensembl95.transcriptome.gentrome.salmon-1.2.0-index-archive.tar**

3. click **Save selection**



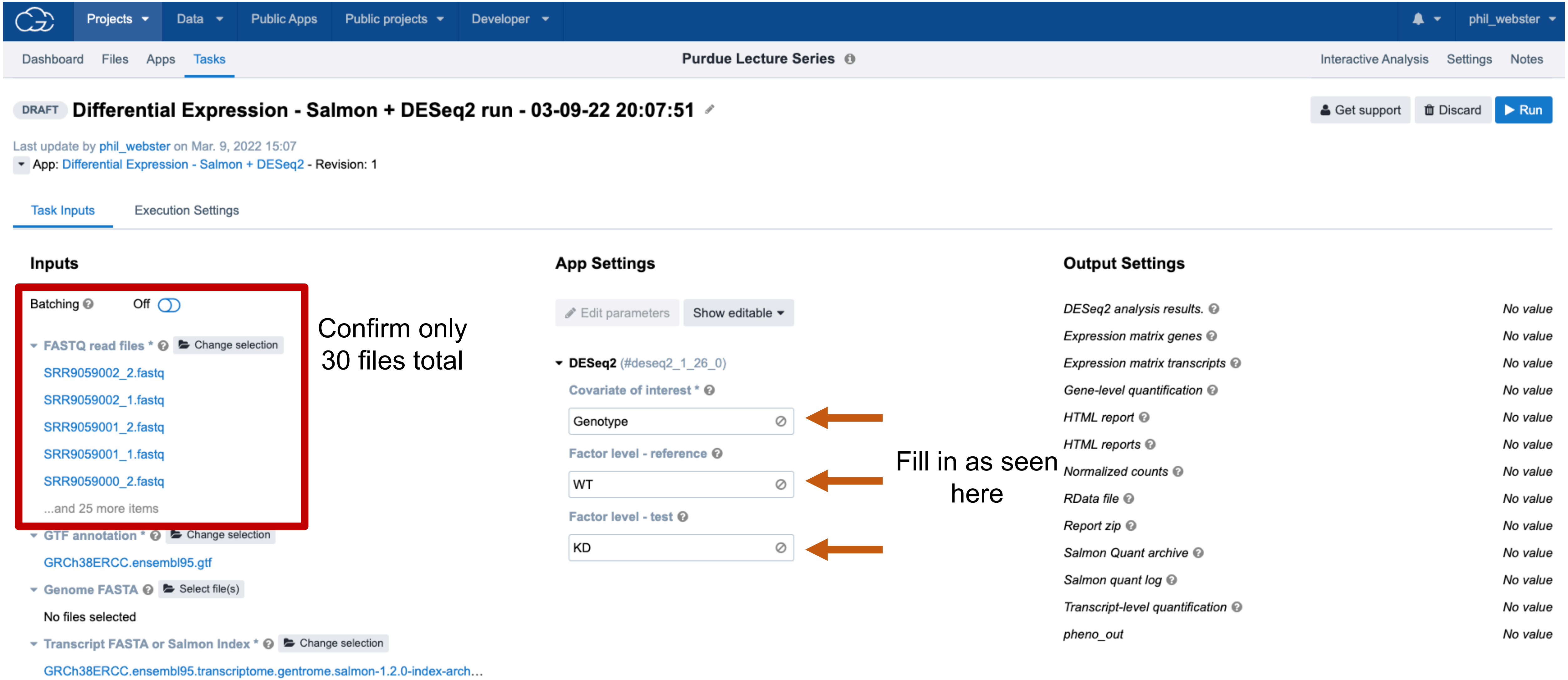
Select files for "Transcript FASTA or Salmon Index"

☐	Name	Experimental strategy	Path	Paired-end	Extension	Size	Sample ID	induced_hbv_replication
☐	GRCh38ERCC.ensembl95.fasta GRCH38	-	Files / reference_files	-	FASTA	2.9 GiB	-	-
☐	GRCh38ERCC.ensembl95.transcriptome.fasta GRCH38	-	Files / reference_files	-	FASTA	305.7 MiB	-	-
☑	GRCh38ERCC.ensembl95.transcriptome.gentrome.salmon-1.2.0-index-archive.tar SB-GENERATED GRCH38	-	Files / reference_files	-	TAR	16.6 GiB	-	-
☐	SRR9058988.salmon_quant_archive.tar	-	Files / workflow_outputs / run_1	-	TAR	11.8 MiB	SRR9058988	HBV_neg
☐	SRR9058989.salmon_quant_archive.tar	-	Files / workflow_outputs / run_1	-	TAR	11.8 MiB	SRR9058989	HBV_pos

Fill In App Settings

Verify that everything is set up as shown here, and the draft task is in **your own project**.

There should be a total of 30 input fastq files and the DESeq2 settings should be as shown below



Purdue Lecture Series

DRAFT **Differential Expression - Salmon + DESeq2 run - 03-09-22 20:07:51**

Last update by phil_webster on Mar. 9, 2022 15:07

App: Differential Expression - Salmon + DESeq2 - Revision: 1

Task Inputs | Execution Settings

Inputs

Batching: Off

FASTQ read files *
- SRR9059002_2.fastq
- SRR9059002_1.fastq
- SRR9059001_2.fastq
- SRR9059001_1.fastq
- SRR9059000_2.fastq
- ...and 25 more items

GTF annotation *
- GRCh38ERCC.ensembl95.gtf

Genome FASTA
- No files selected

Transcript FASTA or Salmon Index *
- GRCh38ERCC.ensembl95.transcriptome.gentrome.salmon-1.2.0-index-arch...

Confirm only 30 files total

App Settings

DESeq2 (#deseq2_1_26_0)

Covariate of interest *: Genotype

Factor level - reference: WT

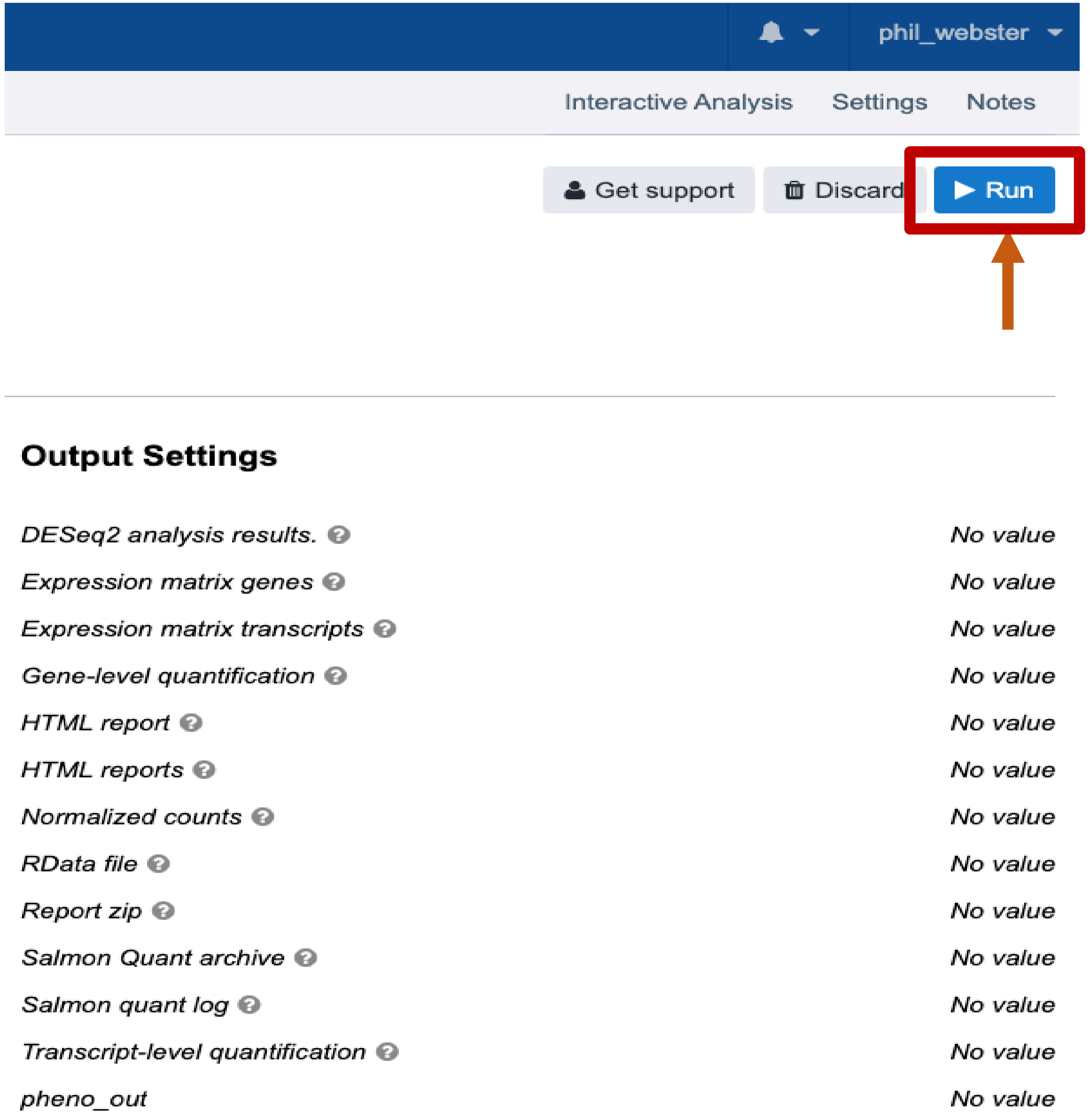
Factor level - test: KD

Fill in as seen here

Output Settings

DESeq2 analysis results.	*No value*
Expression matrix genes	*No value*
Expression matrix transcripts	*No value*
Gene-level quantification	*No value*
HTML report	*No value*
HTML reports	*No value*
Normalized counts	*No value*
RData file	*No value*
Report zip	*No value*
Salmon Quant archive	*No value*
Salmon quant log	*No value*
Transcript-level quantification	*No value*
pheno_out	*No value*

Run!



Output Settings

DESeq2 analysis results.	*No value*
Expression matrix genes	*No value*
Expression matrix transcripts	*No value*
Gene-level quantification	*No value*
HTML report	*No value*
HTML reports	*No value*
Normalized counts	*No value*
RData file	*No value*
Report zip	*No value*
Salmon Quant archive	*No value*
Salmon quant log	*No value*
Transcript-level quantification	*No value*
pheno_out	*No value*

Please Contact Us With Further Questions

https://dotandline.net/my-hero-academia-all-for-one-0309/