Baskin Engineering
UC SANTA CRUZ

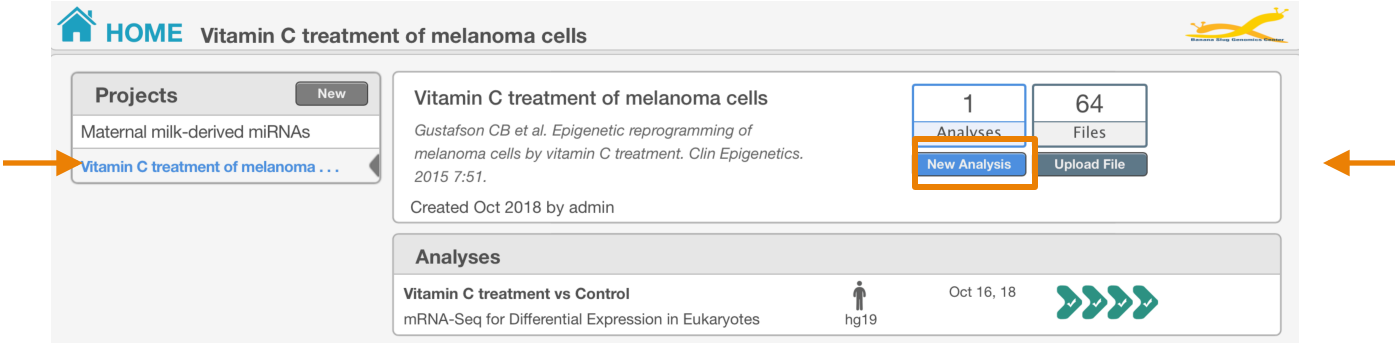
Banana Slug Analytics Platform

mRNA-Seq for Eukaryotes Analysis Setup

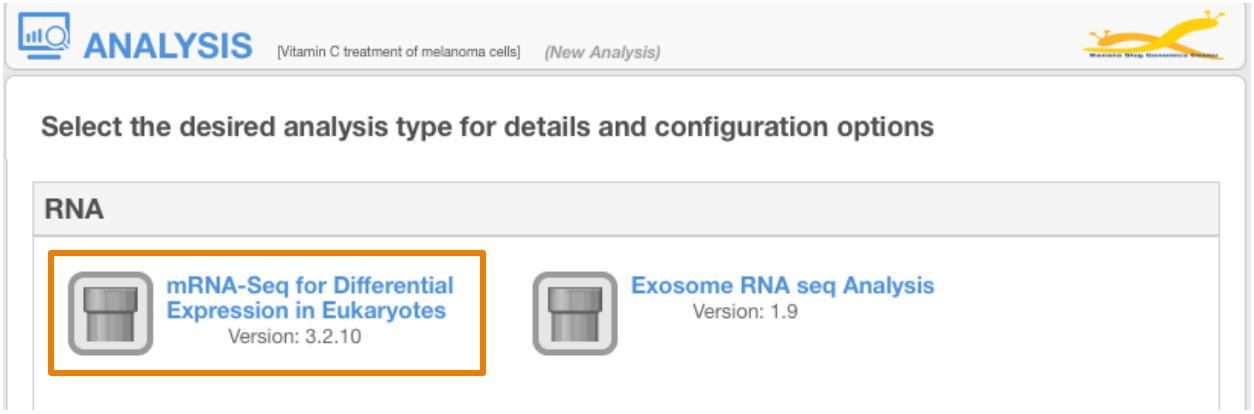
Rev. 01/2018

This document provides a guide to setup and launch the mRNA-Seq for eukaryotes analysis on the Banana Slug Analytics Platform. This procedure is performed after the FASTQ files have been uploaded to the designated project in your account.

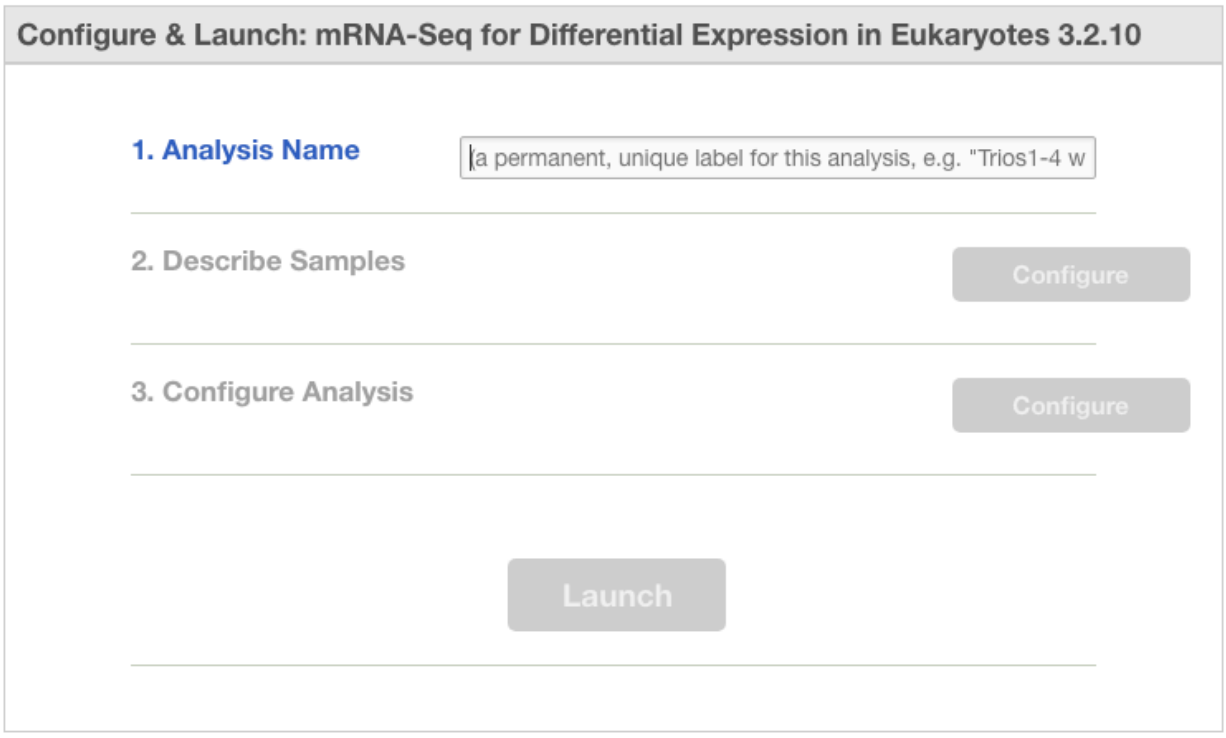
At the HOME page of your account, select the project in the Projects panel on the left for the analysis to be performed. Then click on the New Analysis button on the right panel.



The ANALYSIS page will be opened with the analysis pipeline available in your account. Click on the link for mRNA-Seq for Differential Expression in Eukaryotes.



The Configure & Launch panel will be shown that allows you to configure and launch your analysis.

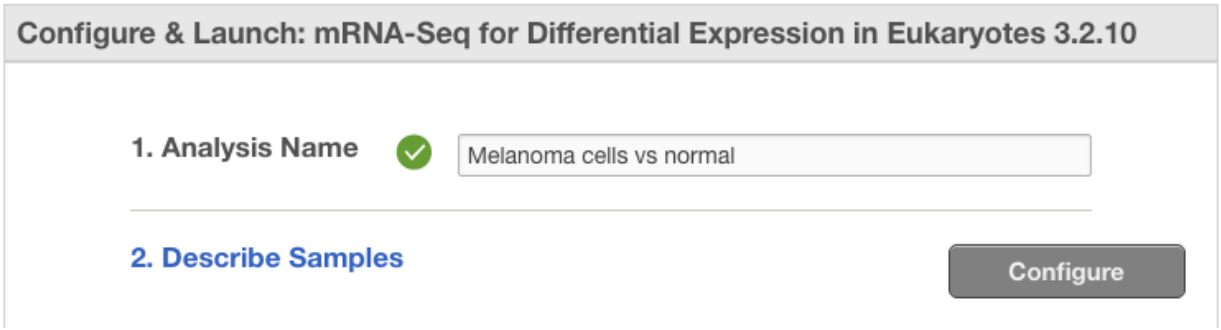


Analysis Configuration

As shown on the Configure & Launch panel, there are three steps to configure the analysis.

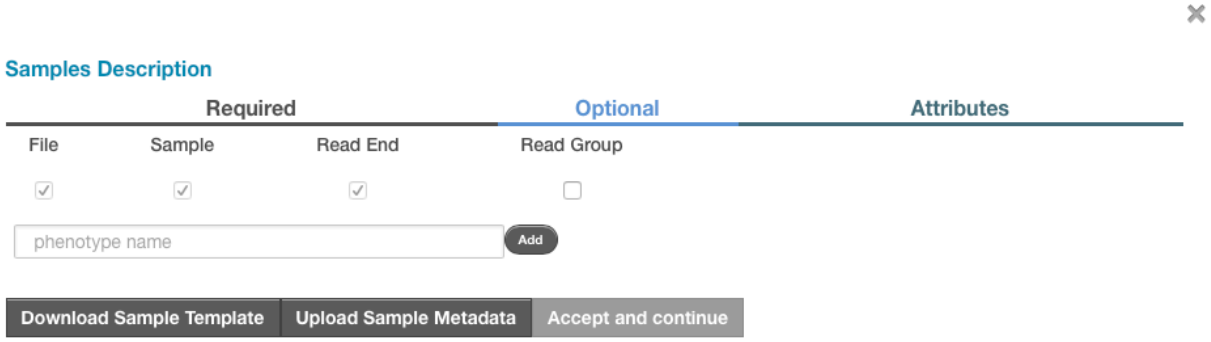
1. Analysis Name

You will have to provide a name to identify the analysis. The name has to be unique and will be listed on the home page of the analysis list and the top of each page on the platform as the currently selected analysis.



2. Describe Samples

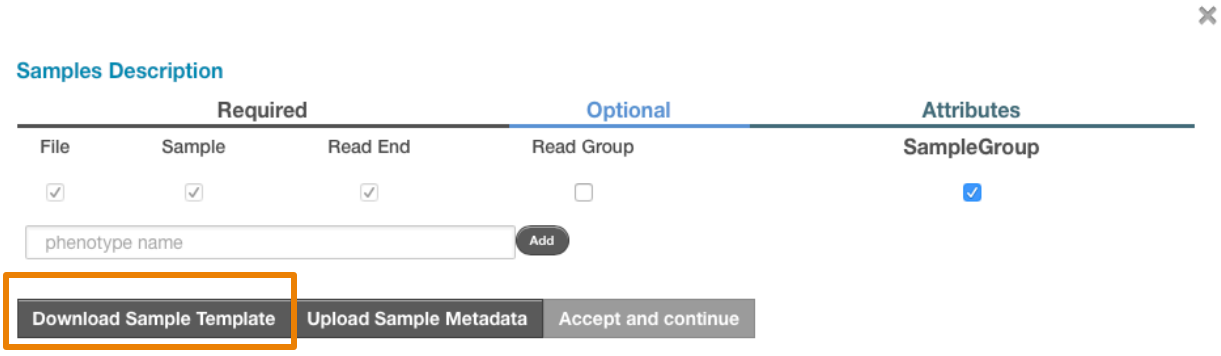
After a valid analysis name has been entered, the Configure button for the Describe Samples step will be enabled. Click on the Configure button to open the popup window for obtaining the sample template.



A sample template is a CSV file that can be opened in Microsoft Excel or text editors for including sample metadata to be used in the analysis. Two options are available to be added into the sample template.

- Read Group – If you have more than one (for single-ended reads) or one pair (for pair-ended reads) of FASTA files per sample, check the Read Group box. Otherwise, leave it unchecked.
- Attributes such as phenotype name – If the samples in your project can be grouped as replicates to represent certain conditions, treatments, or phenotypes, enter the attribute name in the text box and click the Add button.

As an example, the following image shows that an attribute named SampleGroup is added but the Read Group has left unchecked.



Click the Download Sample Template button to download the CSV file.

Here we can see the sample template CSV file in Microsoft Excel.

	A	B	C	D
1	FILE	READEND	SAMPLE	SAMPLEGROUP
2	A2058_Control_1_R1.fastq.gz			
3	A2058_Control_1_R2.fastq.gz			
4	A2058_Control_2_R1.fastq.gz			
5	A2058_Control_2_R2.fastq.gz			
6	A2058_VitC_1_R1.fastq.gz			
7	A2058_VitC_1_R2.fastq.gz			
8	A2058_VitC_2_R1.fastq.gz			
9	A2058_VitC_2_R2.fastq.gz			

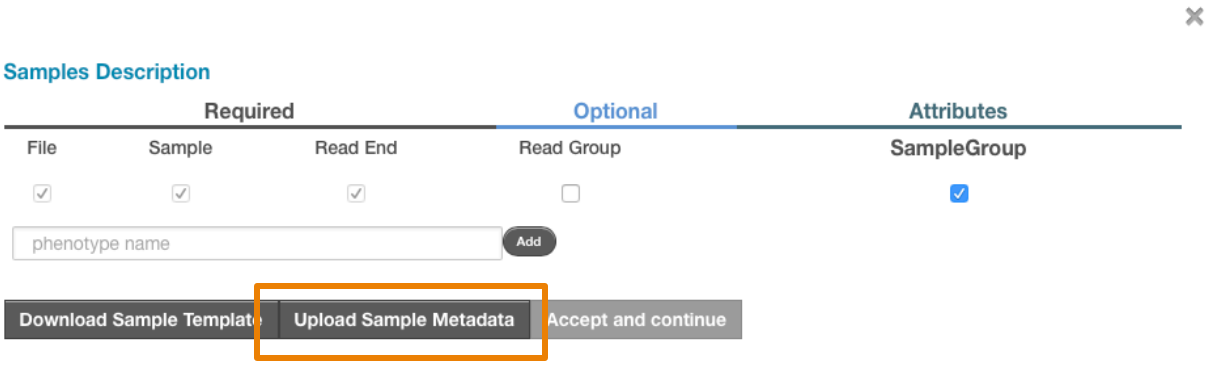
- The FASTQ files that were previously uploaded to the project are listed in FILE column of the template. If any of them should not be included in the analysis, you can remove them by deleting the rows. If no FASTQ files are included, please check the FILES page of your project to make sure that your FASTQ files have been uploaded successfully.
- READEND column is used to indicate the read end of the FASTQ files. If your samples have single-ended reads, put 1 in the READEND cells corresponding to those samples. If your samples have pair-ended reads, enter 1 for the FASTQ files representing read 1 and enter 2 for the FASTQ files representing read 2.
- SAMPLE column is the unique name of the sample. If you have pair-ended reads, this sample name should be the same for read 1 and read 2 FASTQ files that belong to the same pair. If you have single-ended reads, this sample name should be unique across the FASTQ files.
- SAMPLEGROUP column in the above figure is the attribute previously added before downloading the sample template file. It will be used to group the samples as replicates for the same conditions by having the same value in the column for those FASTQ files of the samples.

After entering all the information, the template will look like the following:

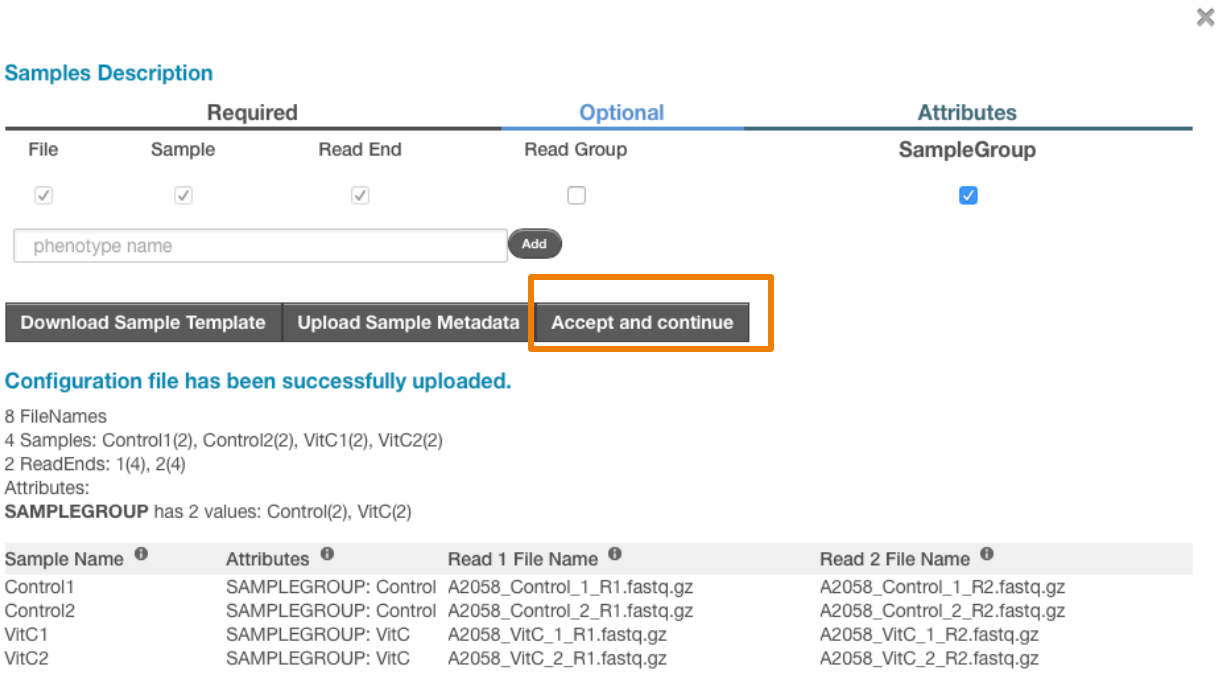
	A	B	C	D
1	FILE	READEND	SAMPLE	SAMPLEGROUP
2	A2058_Control_1_R1.fastq.gz	1	Control1	Control
3	A2058_Control_1_R2.fastq.gz	2	Control1	Control
4	A2058_Control_2_R1.fastq.gz	1	Control2	Control
5	A2058_Control_2_R2.fastq.gz	2	Control2	Control
6	A2058_VitC_1_R1.fastq.gz	1	VitC1	VitC
7	A2058_VitC_1_R2.fastq.gz	2	VitC1	VitC
8	A2058_VitC_2_R1.fastq.gz	1	VitC2	VitC
9	A2058_VitC_2_R2.fastq.gz	2	VitC2	VitC

Save template with the sample metadata. Then go back to the platform.

At the window, click the Upload Sample Metadata button to upload the CSV file with the sample information.



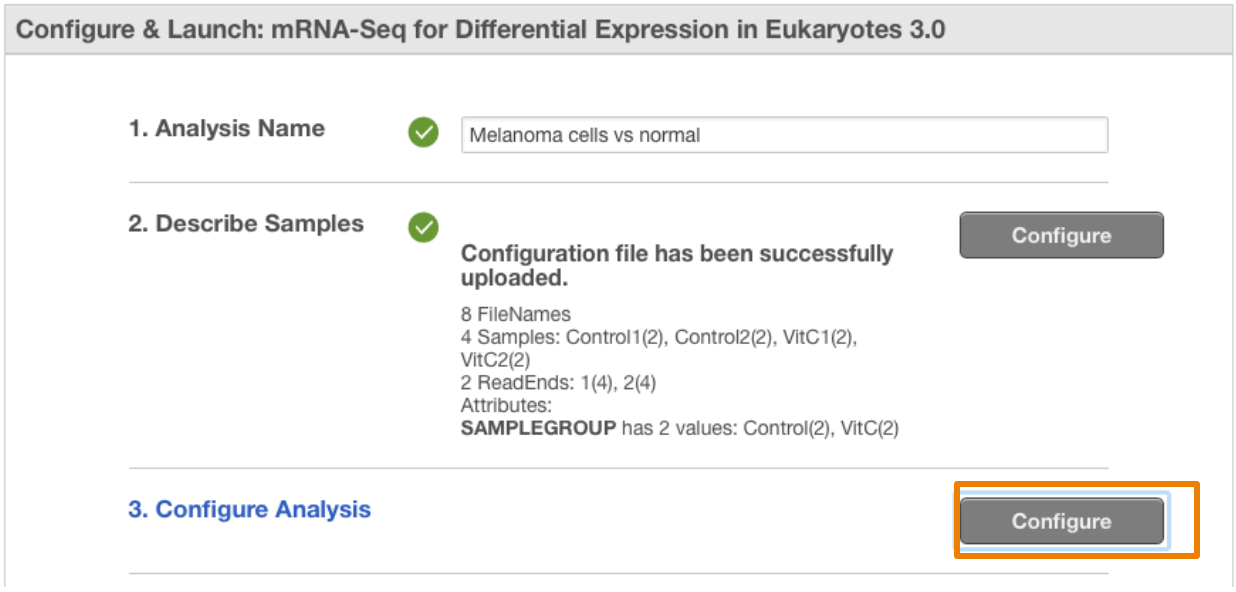
When the CSV file has been uploaded successfully, the summary of the configuration will be listed on the window as follows:



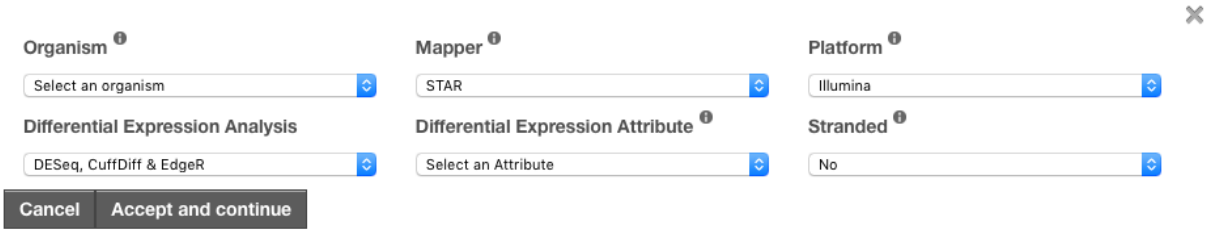
Double check to make sure that everything is correct. If not, go back to update the information in the CSV file and upload it again. If the configuration is correct, click the Accept and continue button to finish this step.

3. Configure Analysis

Now the Configure button for the last step should be enabled.



Click on it to get to the next configuration window.

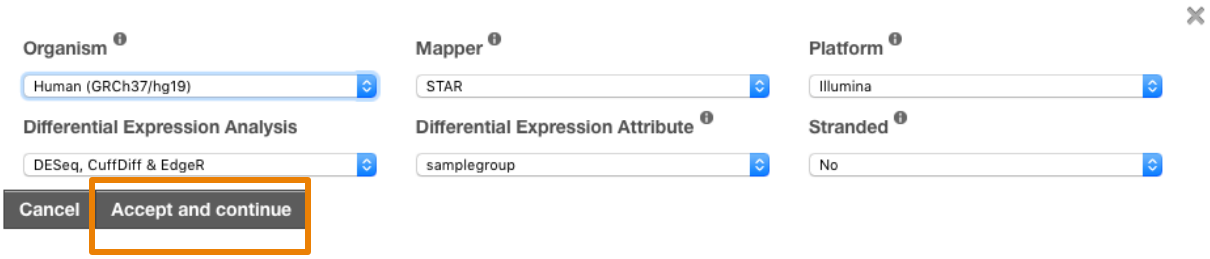


Here you can specify the analysis configuration.

- Organism – This is the reference genome that will be used for read mapping and abundance estimation at gene level. Select the genome that represents your sequencing libraries.
- Mapper – This is the choice of read mapping algorithm to be used in the analysis. STAR is selected as the default due to its speed and accuracy. But TopHat is also offered as a choice for those who prefer using the RNA-Seq Tuxedo suite.
- Platform – This is the platform used for performing the sequencing. Illumina is the default platform while Ion Torrent sequencing data is also supported as an option.
- Differential Expression Analysis – Our pipeline uses DESeq, Cuffdiff, and EdgeR separately for differential expression analysis so that you can check out the results produced by different methods.
- Differential Expression Attribute – This will include the sample attribute previously entered in Step 2 as a choice. The attribute will be used in sample grouping during differential expression analysis.

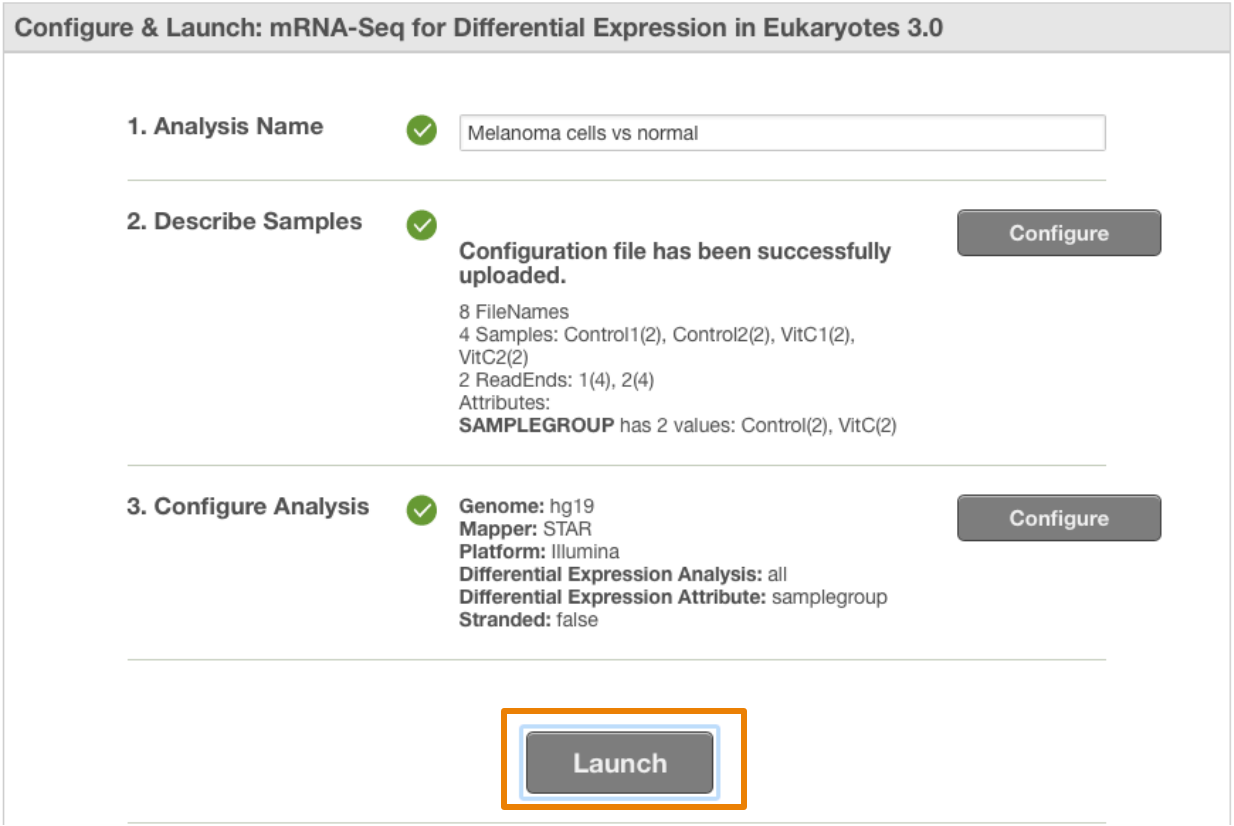
- Stranded – This field specifies whether the sequencing libraries are stranded or not.

After making selections to the fields, click the Accept and continue button.



Launching Analysis

After completing the configuration, the page will show the summary at each step as below:



If everything is correct, click on the Launch button and the analysis will be started.