

# PROJECT MANAGEMENT

Self-contained mini-workspaces, where sample sets can be analyzed independently without interference from other data, using multiple pipelines.

- SAMPLE SET MANAGEMENT

- Project Setup

- Pipeline Management

- Favorites Pipelines
- My Pipelines
- Pipeline Creation
- Pipeline Filters
- Copy Pipeline
- De Novo Pipeline
- Pipeline Execution

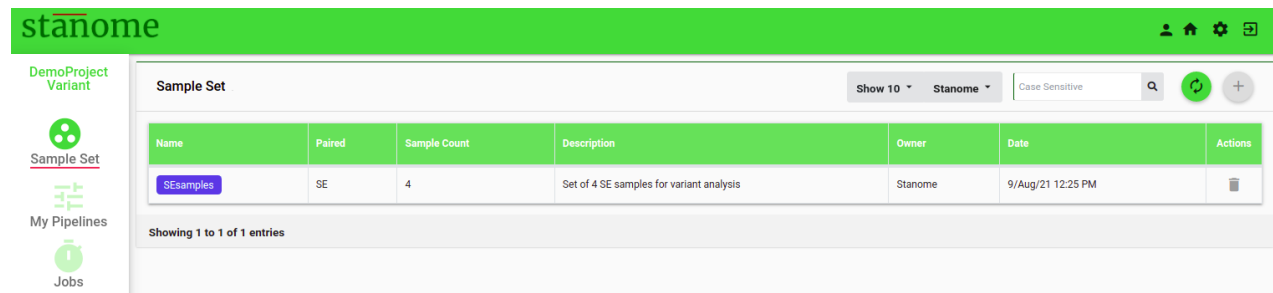
- Reports

- Overview
- Variant Report
- Transcript Report


- Jobs

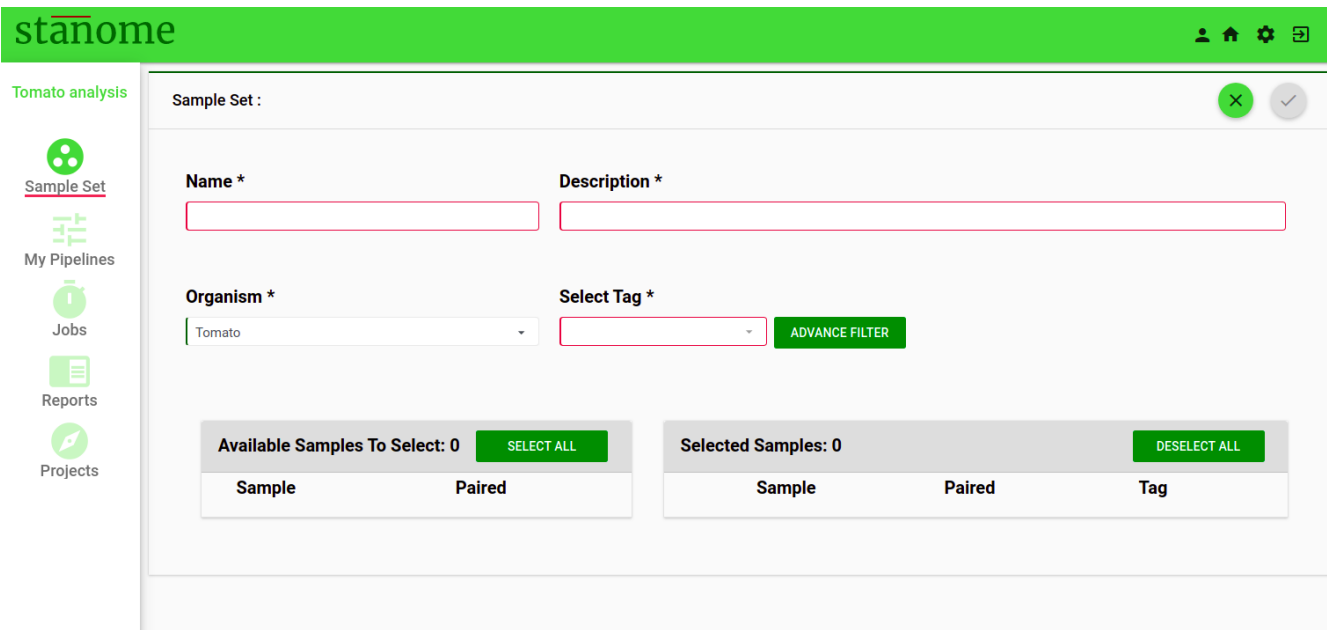
# SAMPLE SET MANAGEMENT

**Sample Set** is a group of sequence files that belong to an experiment and can be analyzed together. **Sample Set** window allows new sample set creation and provides a list view of the existing **Sample Sets** within a project (Fig. 1).



## Sample Set Creation

**Sample Set** can be created by selecting the samples available in the . Click  from the **Project Main** window or click  on the **Sample Set** window to create a new **Sample Set** (Fig. 2).



Fill in the following details on the **Sample Set** creation window:

- **Name\*** - Provide a unique name for the **Sample Set**.



Only alphanumeric characters and spaces are allowed.



**Sample Set** names are helpful while executing the pipelines.

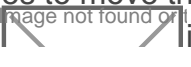
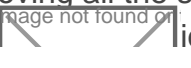
**CAUTION** - Names shouldn't be reused.

- **Description\*** - Provide a description for the **Sample Set** that can be included in the final report (Example: experiment type).
- **Organism** - Select organism name from the drop-down list.



Samples from different organisms can be selected to form a single sample set

**CAUTION** - Unless intended, samples from multiple organisms in a sample set are not recommended.

- **Select Tag** - Using the tags given during sample upload samples can be filtered. A list of samples uploaded under the corresponding tag would appear in the table on the left. Click on the samples to move them to the selected samples table and to add to the sample set (Fig. 2).  icon allows moving all the samples listed under the selected tag to the selected samples table.  icons allow deselection of the selected samples (Fig. 2).

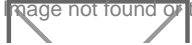


Samples from different tags can be selected to create one sample set.

**CAUTION** - Select either SE or PE samples only. The platform doesn't allow mixed selection.


## Sample Filter




Click on the  to filter samples based on the sample quality. Six filters are supported.

1. **The total number of reads:** This shows the samples with the total number of reads greater than the number entered.
2. **Number of poor quality reads:** This shows the samples which contain, number of poor quality reads less than the number entered. The quality of reads is determined by the fastqc tool.
3. **Sequence length:** Shows the samples with reads whose sequence length ranges the number entered.
4. **Per base sequence quality:** Shows the samples with reads per base sequence quality tagged as Pass/Fail/Warn.

5. **Sequence length distribution:** This shows the samples with reads sequence length distribution tagged as Pass/Fail/Warn.
6. **Adapter content:** Shows the samples with read adapter content tagged as Pass/Fail/Warn.

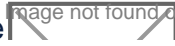

Click  on the upper right corner of the window to save the **Sample Set**. Multiple sample sets can be created under one project with different sequence files.

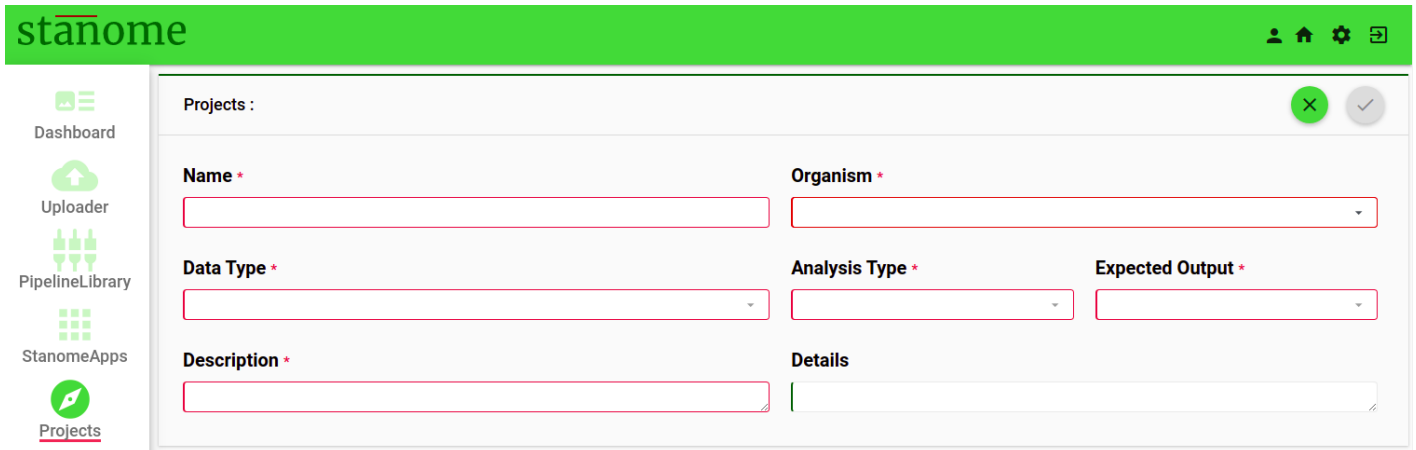
## Sample Set Deletion

A sample set can be deleted by clicking  icon on the **Sample Set** window. This action deletes the sample set only, however, the samples are still available in the **Sequence Data** for future usage.

# Project Setup

A new project can be created in two ways.

1. From the **Dashboard** window - click the  icon
2. From the **Projects** window - click  on the upper right corner



A new project can be created from the **Project** setup window (Fig. 1). Fill in the seven mandatory fields with details and appropriate descriptions. The data in this form is used for automatic reference file(s) selection and will be exported to the final report.

**CAUTION - Organism enables appropriate reference file(s) selection.**

- **Name\*** - Provide a unique name for the project.



**HINT** - Project name cannot be longer than 30 characters and alphanumeric characters and spaces are only allowed.

**CAUTION - Project name should be unique.**

- **Organism\*** - Select the organism from the drop-down menu. We currently support eight main model organisms and some custom species.



**HINT** - Select **Other** if the organism is not listed and contact Stanome technical support team to add a new organism to the list.

- **Description\*** - Provide a brief description of the project (Example: The gist of the experimental design and the aim of the study).
- **Data Type\*** - Select sample type from the drop-down list [*Microbiome*, *Targeted Genome*, *Transcriptome*, and *Whole Genome*]. The **Data type** field is tightly connected with the **Hub** field in the pipelines.



**HINT** - Carefully choose the **Data type** to get the relevant pipeline suggestions.


- **Analysis Type\*** - Select the experiment type from the drop-down list [*ABR Screening, Alignment, Annotations, Data Cleaning, Data QC, Differential Expression, Genome assembly, Genotyping, Proband, Serotyping, Species Classification, and Variant analysis*].
- **Expected output\*** - Select expected output(s) from the drop-down list [*BAM, Coverage Metrics, DE Genes (Differentially Expressed Gene list), FASTA, Genotypes, Ontology, Pathways, and VCF (Variant Calling Format)*].



Multiple options can be selected.

- **Details\*** - Provide detailed information about the project. (Example: Aim of the experiment, conditions studied, experimental design, and any other pertinent details).



Click  to save the project. The new project is saved and displays the **Project Main** window (Fig. 2), where the project-specific menu is available on the left menu. The project can be navigated with **Sample Set**, **My Pipelines**, **Jobs**, and **Reports**, and details of these four features are described in the following sections.

stanome

Project : 6663595245 , DemoProject Variant

**Name \***

DemoProject Variant

**Organism \***

Tomato

**Data Type \***

Targeted Genome

**Analysis Type \***

Variant Analysis

**Expected Output \***

Genotypes


**Description \***

This project demonstrates the complete process of running a pipeline on the platform including sample set selection. It also gives access to Report and its features generated for a small sample set.

**Details \***

The "Variant Suite 1412" pipeline provided in this project carries out alignment of reads to reference genome. A list of the variants is obtained based on target bed file. The final report provides FastQC statistics, mapping summary and variant summary alongwith quick visualization. The report also provides IGV interface for the variants.




Projects can be deleted using the  icon from the **Projects** window. All the components within a project (sample sets, pipelines, jobs, and reports) are deleted permanently.

# Pipeline Management

A set of computational tools, which run either sequentially or parallelly in order to achieve a specific data analysis objective. Tools/commands are designated as steps in a pipeline.

# Favorites Pipelines

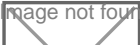

Pipelines, from any project, can be added or deleted from the favorites by clicking the  icon next to the pipeline name in the list view i.e. **My Pipelines**. Stanome owned pipelines can't be added to the favorites. Favorite pipelines are visible in the **Pipeline Library** and **My Pipelines** (filtered based on the Owner, Hub, and Category).

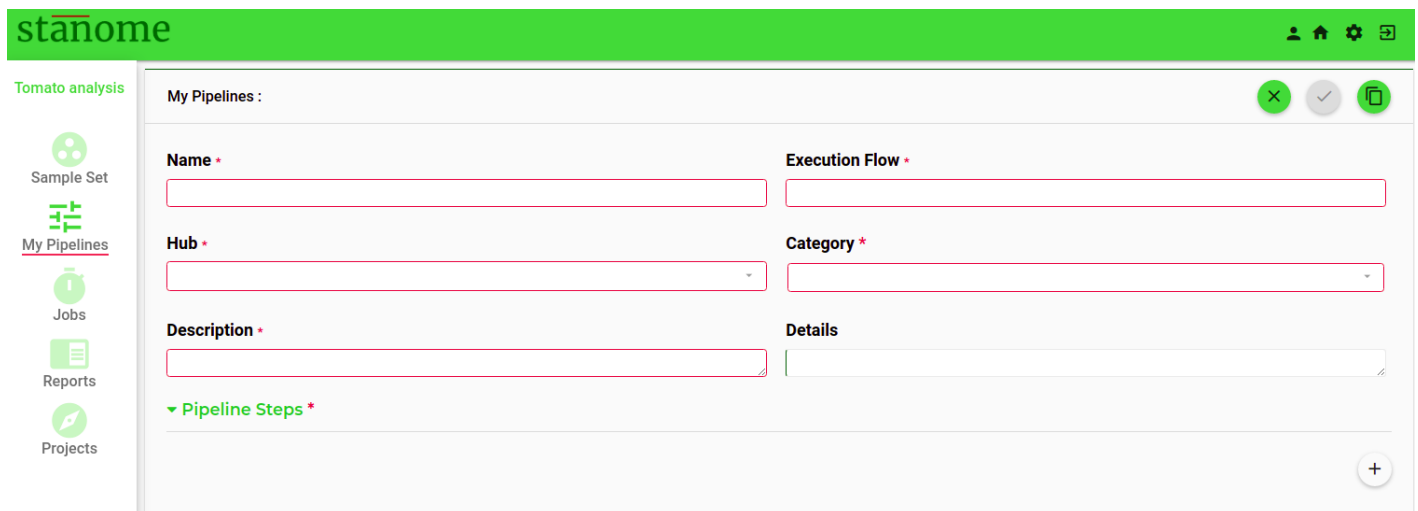


# My Pipelines

**My Pipelines** show the list of pipelines within a project. Pipelines can be created, viewed, edited, and deleted within the project scope. **My Pipelines** is empty, by default, and users can create new pipelines *de novo* or copy the pre-configured pipelines from the **Pipeline Library**. Follow the instructions in the next section to create a pipeline.

# Pipeline Creation

The new pipeline creation window is accessed from two locations. Either click  on the project window or click  on the **My Pipelines** window to create a new pipeline (Fig. 1).



The new pipeline creation window displays three icons in the upper right corner.



Exit out of the pipeline creation without saving.



Save the pipeline.



Copy pipeline.

New pipelines can be added to a project in three ways.

- Copying pre-configured pipelines
- Creating *de novo*
- Favorite pipeline

# Pipeline Filters




Pipelines on the platform are tightly connected with the projects. The **Data type** field in the projects is tied with the **Hub** field in the pipelines. Based on the **Data Type** definition in the project setup, only relevant pipelines are shown. The following table shows the corresponding terms between projects and pipelines.


Data Type in Project	Pipeline Hub
Whole Genome	Genome
Microbiome	Micro
Transcriptome	Transcript
Targeted Genome	Variant

Table 4: Terms associated between projects and pipelines

# Copy Pipeline

A new pipeline can be created by copying an existing pipeline from other projects or a pre-configured pipeline from **Pipeline Library**.

Click  in the upper right corner of the pipeline creation window to see the existing pipelines via the **Select pipeline** dialog box (Fig.). Using the project setup information, a list of prefiltered pipelines is listed. Users can use multiple field combinations to filter the pipelines. Pipelines from the  can be viewed by selecting the owner as “Stanome”. Select a pipeline and click on  to copy a pipeline into the current project. The pipeline steps, tools, parameters, and other details are auto-populated (except the pipeline name). Name the new pipeline uniquely (duplicate names are not allowed) and verify the tools and commands before saving the pipeline.

 **HINT:** Pipeline name should be less than 50 characters long and only alphanumeric characters and spaces are allowed.

stanome

Tomato analysis

My Pipelines :

×

✓

Sample Set

My Pipelines

Jobs

Reports

Projects

All Hubs ▾ All Categories ▾ Stanome ▾ All Projects ▾ Case Sensitive 🔍

	Name	Execution Flow	Hub	Category	Description	Project	Owner	Actions
<input type="checkbox"/>	Variant Suite 1412 Demo	Data Cleaning -> Alignment -> Merge -> VariantCalling	Variant	Genotyping	Data cleaning, alignment of reads to the reference, merge samples, and jointly call genotypes.	DemoProjectVariant	Stanome	
<input type="checkbox"/>	Differential Suite 1221Demo	Data Cleaning -> Quasi-Alignment -> Differential Expression -> Pathway Annotation	Transcript	Differential Expression	Data cleaning, quasi-alignment of the RNAseq reads to the reference, identification of the differentially expressed genes, and pathway enrichment analysis.	DemoProjectTranscript	Stanome	
<input type="checkbox"/>	Assembly Suite 1126	Data Cleaning -> De novo Assembly -> Assembly Correction -> Annotation	Micro	Genome Assembly	Data cleaning, de novo short-read assembly, assembly correction with long-reads, and genome annotation.	NA	Stanome	
<input type="checkbox"/>	Assembly Suite 1106	Data Cleaning -> De novo Hybrid Assembly -> Annotation	Micro	Genome Assembly	Data cleaning, de novo hybrid assembly of short and long reads, and genome annotation.	NA	Stanome	
<input type="checkbox"/>	Assembly Suite 1105	Data Cleaning -> De novo Assembly -> Annotation	Micro	Genome Assembly	Data cleaning, de novo short-read assembly, and genome annotation.	NA	Stanome	
<input type="checkbox"/>	Differential Suite 1221	Data Cleaning -> Quasi-Alignment -> Differential Expression -> Pathway Annotation	Transcript	Differential Expression	Data cleaning, quasi-alignment of the RNAseq reads to the reference, identification of the differentially expressed genes, and pathway enrichment analysis.	NA	Stanome	
		Data Cleaning -> Pseudo-			Data cleaning, pseudo-alignment of the RNAseq reads to the reference			

CANCEL

SAVE


# De Novo Pipeline

*De novo* pipeline building requires bioinformatics expertise. Please contact the technical support team for assistance.

The creation of a brand new pipeline is more challenging than copying an existing pipeline. Fill in the following details on the pipeline creation window (Fig. 1) to create a new pipeline. Mandatory fields are indicated with asterisks (\*).

- **Name\*** - Provide a unique name
- **Execution Flow\*** - Enter the tool names in the order of execution (e.g. Trimmomatic -> Salmon -> Sleuth)
- **Hub\*** - Indicate the pipeline group
- **Category\*** - Select the functional category
- **Description\*** - Provide a brief description of the pipeline's general purpose
- **Details** - Provide detailed information about tools, inputs, outputs, arguments, and other pertinent information
- **Steps\*** - Step field helps to add tools and commands to a pipeline.

At least one step is required for a functional pipeline.

Click the  icon to add a new step (or tool) to the pipeline (Fig. 1). There are eight fields in each step:

- **Number** - The number determines the step number in the pipeline. It is automatically filled when a new step is added.



**HINT:** Only positive integers are allowed

- **Name** - Provide a name to the step (e.g. Bowtie2 alignment). The outputs will override if the name field is not unique because the name is used as a directory to store the output files of the step.



**HINT:** The name should be unique.

- **Tool** - Select a tool from the drop-down menu

- **Command** - Select a command from the drop-down menu for the selected tool
- **Predecessor** - The number determines the dependency step of a step. A step executes only after the dependency step.



**HINT:** Only positive integers are allowed

- **Merge** - This is a critical variable that indicates if all the inputs need to be combined into a single output file. This is useful in some analyses where all files need to be analyzed together (Examples: Differential gene expression and Joint genotyping). Default: No.



**HINT:** The first step can't be a merge step

- **Input Source** - Determines the source of the input files for a step. Typically, input file sources are either **Data Store** (Sequence Data, References, Annotations, and Metadata) or output files from predecessor steps.






**HINT:** Input sources from multiple steps are allowed. (Example: BAM and BAI files created in different steps required for Variant calling).



**HINT:** Currently, **Data Store** is allowed for the first step only

- **Actions** - Each step allows the following actions

-  Access the **Command builder** dialog box
-  Delete a step
-  Copy a step


## 1. **Command Builder**

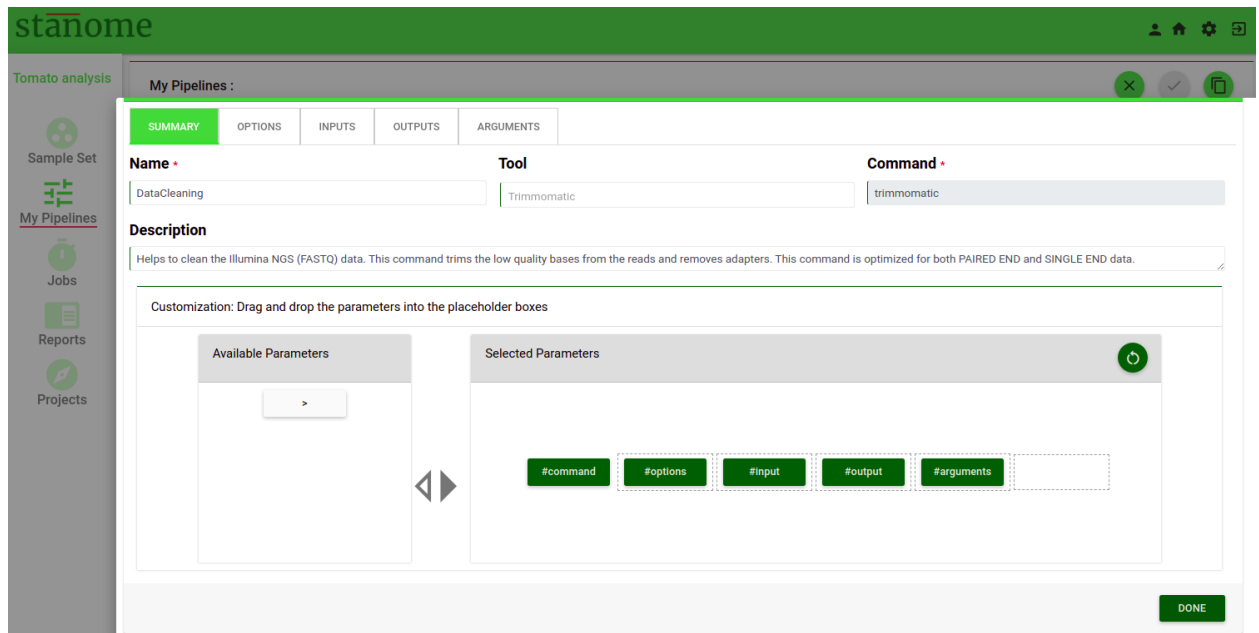
- Command building is one of the complex processes on the platform and the **Command Builder** dialog box helps to navigate the process easily.

Commands are preconfigured by the platform admin. Users can only edit the commands.

- During the pipeline development stage users define the commands and the final executable command will be dynamically created during the pipeline execution stage.

This is a generic command building process. You are NOT making the actual file selections required for the analysis. The platform does it automatically based on your definitions.

- Command Builder has two modes: View mode and Edit mode. The former allows viewing and the latter allows command modification, as described below. Access the **Command Builder** dialog box (Edit mode) (Fig. 1) by clicking  under the **Actions** column.



The first tab of the **Command Builder** describes the generic details(summary) about a command.

- **Name** - The step name as given by the user while creating the step.
- **Tool** - The selected tool (cannot be modified)
- **Command** - The actual command to be executed (cannot be modified)
- **Description** - Brief description of the command (auto-filled but can be modified by the user)
- **Build Your Command** - This box helps to build the actual command. The left box shows the available parameters and the right box (Fig. 2) shows the selected parameters. The order of the parameters is extremely important and should be maintained for the command to execute. All commands come with a default parameter sequence (Stanome defined). The parameters are prefixed with a #(hash). The default pattern can be modified by dragging and dropping the parameter buttons (green color) between the left and right boxes. Based on the selection the parameter tabs are enabled on the top.

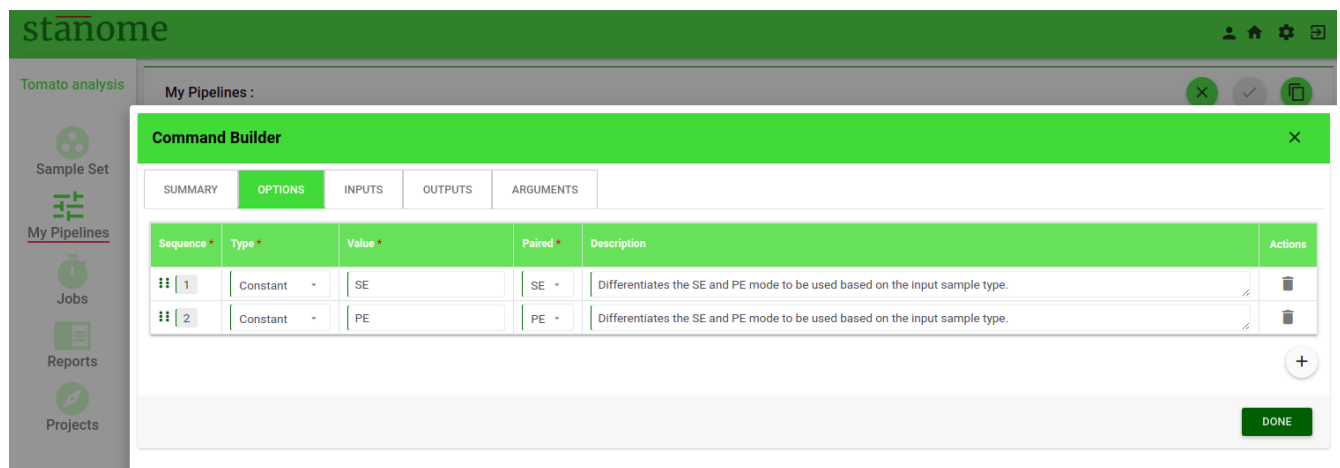
**Default pattern:** *#command #options #arguments #input #output*

The pattern should ALWAYS start with *#command* and can't be edited.

**Allowed character:** Parameter words, #, space, and >

">" is allowed preceding the *#output* ONLY

The second tab of the **Command Builder** (Fig. 2) describes the **Options** parameter. Details of the Options tab are described below:



### • Options

Single-word parameters should be defined as options (Examples: --ignore, --1, PE, SE). All the options are listed in a table format. New row(s) can be added using the '+' sign at the bottom of the table. Six fields are available under each option.

- **Sequence:** This number determines the order of the option in the command
- **Type:** Six choices are available in the drop-down. Select the type of option. (Example: Annotation, Constant, Metadata, Reference, Threshold, and Variable)
- **Value:** Based on the field **Type**, the corresponding values in the drop-down change. Select an appropriate value. See the table below for available field types and their values.

**CAUTION** - Verify usage of each option before using

Field Type	Value
Annotation	<ul style="list-style-type: none"><li>• Variant annotation files (Mills1000G_INDELS, DBSNP, 1000G_HC, 1000G_OMNI, and HAPMAP), GATK</li><li>• Pathway or GO</li><li>• VEP Cache and VEP Cache Version</li><li>• GTF</li><li>• ABR</li></ul>



Constant	<ul style="list-style-type: none"> <li>Any constant value (alphanumerics) (Examples: -o, --i, and --single)</li> </ul>
Metadata	<ul style="list-style-type: none"> <li>Experimental Design</li> <li>Targets</li> <li>Genelist</li> <li>Amplicon ranges</li> </ul>
Reference	Define references to select <ul style="list-style-type: none"> <li>References: Genome/Transcriptome</li> <li>Indexed references: BWA, Bowtie2, etc</li> </ul>
Threshold	Define threshold values to use <ul style="list-style-type: none"> <li>qvalue</li> <li>pvalue</li> </ul>
Variable	Native variables of the platform <ul style="list-style-type: none"> <li>JobID</li> <li>Organism</li> <li>Ploidy</li> <li>Sample Name</li> <li>Reference Version</li> <li>Sequencing Platform</li> </ul>

Table. Available **Field Types** and their corresponding **Values**.

- **Paired:** Indicates if an option can be used for paired-end, single-end files, or both  
(Default: All)
- **Description:** A brief description of the option functionality or usage guidelines.
- **Actions:** Allows the deletion of an option.

**CAUTION** - Please refer to the **Arguments** section for defining the parameters with key-value pairing

#### • **Inputs and Outputs**

Input and output files are defined under **INPUTS** and **OUTPUTS** tabs, respectively. Eight fields are available under each of these parameters (Fig. 3).

- **Sequence:** This number determines the order of the inputs or outputs in the command
- **Name:** The name of the value (Examples: --input, -output)

- **Type:** Input/output data can be provided to a command in three formats - File, FileList, and Directory. Select the format from the drop-down list.
  - File - Input is a single file (Example: Prefix.fastq)
  - FileList - Input is Paired-end files (Example: Prefix\_R1.fastq, Prefix\_R2.fastq)
  - Directory - Input is a directory path
- **Delimiter:** Character separating the input/output file(s) from its name in the command (Example: --input : Prefix.fastq).

**CAUTION - Allowed delimiters are =, -, :, and ;**

- **Format:** Depending on the data select file extension from the drop-down. This value is used to make the right file selections during the pipeline execution. Required for File and FileList types only and not required for Directory type.

**CAUTION - The file extensions should be precise; even the FASTQ and FQ are treated distinctly.**

- **File Name Pattern:** This field is applicable for the Inputs parameter only. Regular expressions can be used to select specific input files. This value is used in combination with the **Format** field. Few examples are provided below for an easy understanding of regular expression usage.

	Input file names	Regular expression
Example 1	castor1_R1.fastq	R1
Example 2	castor1_R1_trimmed.fastq	R1_trimmed
Example 3	abcd_1.fastq	_1
Example 4	abcd_1_R.fastq	_1_R

- **Value:** This field applies to the **Outputs** parameter only. Output value contains three parts
  - Prefix: Stanome sample variable (**\${sampleName}**)
  - Suffix: Step name
  - File extension

(Example: \${sampleName}\_trim.fastq for trimmomatic step). This helps track the files across the entire pipeline execution.

- **Paired:** Indicates if the files (Inputs/Outputs) parameter is applicable to paired-end, single-end files, or both (Default: All).
- **Actions:** Allows the deletion of an input or output.

stanome

Tomato analysis

My Pipelines :

Command Builder

SUMMARY OPTIONS **INPUTS** OUTPUTS ARGUMENTS

Sequence	Name	Type	Delimiter	Format	File Name Pattern	Paired	Orientation	Actions
1		File		FASTQ		SE	NA	
2		FileList		FASTQ		PE	NA	

DONE

stanome

Tomato analysis

My Pipelines : 2522123928 , VS1412PL8sep

Command Builder

SUMMARY OPTIONS INPUTS **OUTPUTS** ARGUMENTS

Sequence	Name	Type	Delimiter	Format	Value	Paired
1	-output	File	NA	FASTQ	#({sampleName})_Trimmed_1.fastq	PE
2	-paired-output	File	NA	FASTQ	#({sampleName})_Trimmed_2.fastq	PE
3	-output	File	NA	FASTQ	#({sampleName})_Trimmed.fastq	SE

DONE

## • Arguments

Parameters defined as a key-value pair should be defined as arguments (Fig. 4). Arguments can be used for any parameters supported by the tools and other required files (reference files, gtf or annotation files, target or hotspot files). They are defined by the following eight features:

**CAUTION** - Please refer to the **Options** section for defining the singleton parameters

- **Sequence:** This number determines the order of the argument in the command.
- **Name:** Name of the argument used by the command to identify it

Arguments are grouped into categories to support diverse tools and commands. In arguments, two fields (Type and Value) work together to define an argument.




- **Type:** Seven choices are available in the drop-down. Select the type of argument. (Example: variable, constant, and annotation\_DNAseq)
- **Value:** Based on the **Type** selected, the values in the drop-down change. Select the appropriate value. Refer to the table given in options for the available types and their values.
- **Delimiter:** Character separating the keys and values in the command (Example: --count: 10). Not all arguments require delimiters between the **Name** and the **Value** fields.

**CAUTION** - Allowed delimiters are =, -, :, %, and ,

- **Paired:** Indicates if the arguments parameter applies to paired-end, single-end files, or both (Default: All).
- **Description:** A brief description of the argument describing the function and utility of the argument.
- **Actions:** Allows the deletion of an argument.

The screenshot shows the 'stanome' web interface. On the left is a sidebar with navigation links: 'Tomato analysis', 'Sample Set', 'My Pipelines', 'Jobs', 'Reports', and 'Projects'. The main area is titled 'My Pipelines : 2522123928 , VS1412PL8sep'. Below this is a 'Command Builder' window with tabs for 'SUMMARY', 'OPTIONS', 'INPUTS', 'OUTPUTS', and 'ARGUMENTS'. The 'ARGUMENTS' tab is active, displaying a table with the following data:

Sequence	Name	Type	Value	Delimiter	Paired	Description
1	--nextseq-trim	Constant	20	=	All	Some Illumina instruments use a two-color chemistry to encode the four bases. This includes the NextSeq and the (at the time of this writing) recently announced NovaSeq. In those instruments, a 'dark cycle' (with no detected color) encodes a G. However, dark cycles also occur when sequencing "falls off" the end of the fragment. The read then contains a run of high-quality, but incorrect "G" calls <https://sequencing.qcfail.com/articles/illumina-2-colour-chemistry-can-overcall-high-confidence-g-bases/>_ at its 3' end.
2	--adapter	Constant	AGATCGGAAGAG CACACGTCTGAA CTCCAGTCA	NA	All	Sequence of an adapter that was ligated to the 3' end. The adapter itself and anything that follows is trimmed. If the adapter sequence ends with the 'S' character, the adapter is anchored to the end of the read and only found if it is a suffix of the read.
			AGATCGGAAGAG			Sequence of an adapter that was ligated to the 5' end. If the adapter sequence starts with the character "A", the adapter is 'anchored'. An anchored adapter must appear in its entirety at the 5' end of the read (it

Click  on the bottom right corner to save the changes to the command. This is the completion of the first step in the pipeline. Continue adding all the steps until the pipeline is complete. Steps can be dragged and dropped at any position with the  icon. Step number, predecessor, and input source get automatically readjusted for all the steps. Click  to save the pipeline.

# Pipeline Execution

Once a pipeline is successfully created and validated within a project, it's ready for execution. A newly created pipeline is shown in Fig. 1.

The screenshot shows the 'stanome' Pipeline Management interface. The top navigation bar is green with the 'stanome' logo and user icons. The left sidebar contains icons for 'DemoProject Variant', 'Sample Set', 'My Pipelines' (selected), 'Jobs', 'Reports', and 'Projects'. The main content area displays the details for a pipeline named 'Variant Suite 1412 Demo'.

**My Pipelines : 5708635328 , Variant Suite 1412 Demo**

**Name \***: Variant Suite 1412 Demo

**Execution Flow \***: Data Cleaning -> Alignment -> Merge -> VariantCalling

**Hub**: Variant

**Category**: Genotyping

**Date \***: 9/Aug/21 12:28 PM

**Description \***: Data cleaning, alignment of reads to the reference, merge samples, and jointly call genotypes.


**Details**: An optimized pipeline for joint genotyping from the targeted sequencing data. Sequencing reads are aligned to the reference genome with Bowtie2 and the aligned BAM files are merged to facilitate the joint genotyping of all samples. The SNP and INDEL variants, defined in the target file, are called by splitting them for parallel executions of FreeBayes.  
 Input data:  
 Reference genome (Fasta)  
 Targeted sequencing data (Fastq)  
 Targets file (BED)  
 Regions file (optional for genome browser)

**▼ Pipeline Steps**

Step	Name	Tool	Command	Predecessor	Merge	Input Source	Actions
1	DataCleaning	Trimmomatic	Trimmomatic	NA	No	Data Store	[Icon]

- The following actions are allowed on the **Pipeline** window: view/delete/edit/initialize.

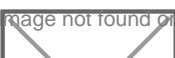
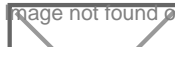
-  **Pipeline** deletion
-  **Pipeline** edit
-  Page refresh
-  Back navigation
-  **Pipeline** initialization

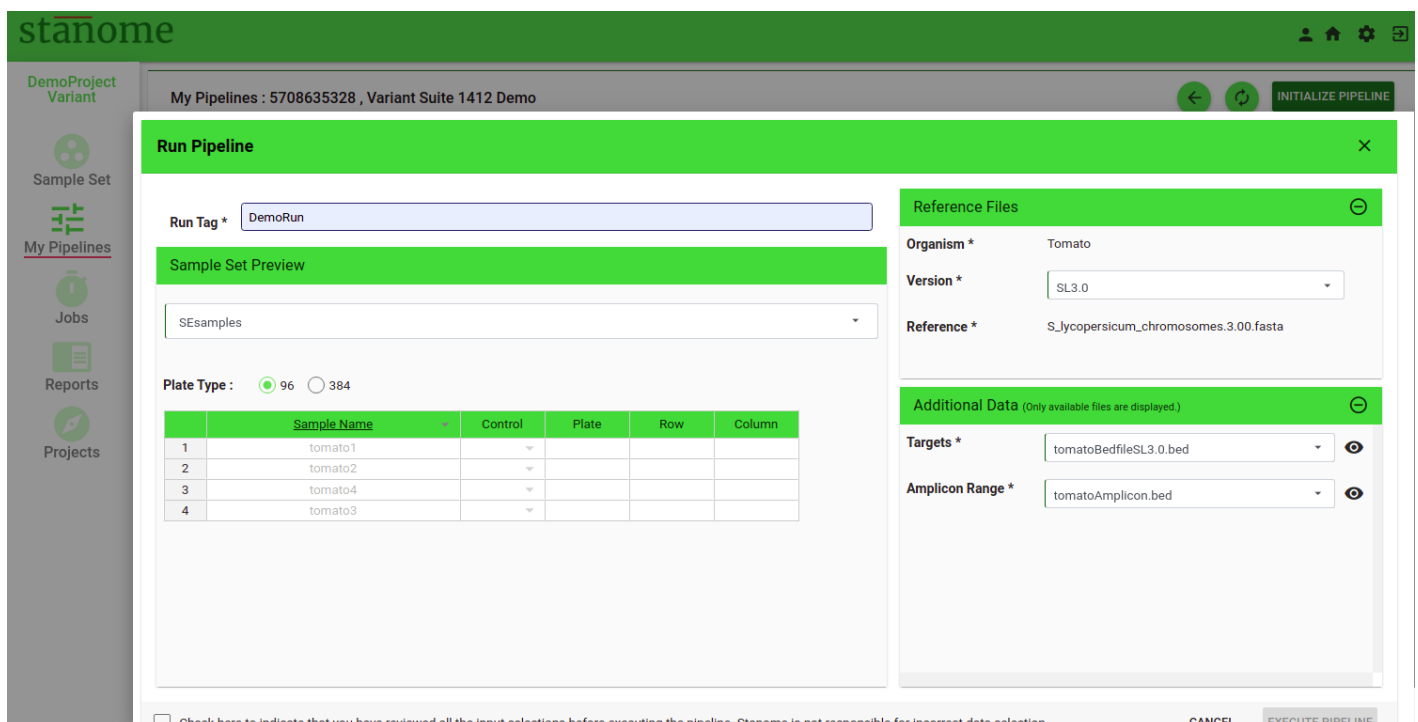
- Click  to access the **Run Pipeline** dialog box (Fig. 2). Final data selections happen during this stage and all fields are required to be filled. The contents of the dialog box change dynamically based on the analysis type and the tools in the pipeline.
- Provide a unique Run Tag
- Sample Set selection from the drop-down shows the sample names in a tabular format

- Select the plate format: 96-well or 384-well (This information is extensively used in the **Reports** to show the results in the plate format)
- Fill in the sample set table with the plate format details and control sample information.
- Latest reference genomes and corresponding additional files are preselected based on the organism of the project. Please confirm or change the selections using the drop-downs.

 **HINT:** The contents of the metadata files can be viewed by clicking the  icon

**CAUTION - Differential Suite pipelines need at least two conditions with two replicates for each. Variant Suite pipelines need properly formatted target files.**

- Agree to the terms and conditions to enable .
- Click  to run the **Pipeline**.



**stanome**

DemoProject Variant

My Pipelines : 5708635328 , Variant Suite 1412 Demo

**Run Pipeline**

Run Tag \* DemoRun

**Sample Set Preview**

SEsamples

Plate Type : ☒ 96 ☐ 384

	Sample Name	Control	Plate	Row	Column
1	tomato1				
2	tomato2				
3	tomato4				
4	tomato3				

**Reference Files**

Organism \* Tomato

Version \* SL3.0

Reference \* S\_lycopersicum\_chromosomes.3.00.fasta

**Additional Data** (Only available files are displayed.)

Targets \* tomatoBedfileSL3.0.bed

Amplicon Range \* tomatoAmplicon.bed

☐ Check here to indicate that you have reviewed all the input selections before executing the pipeline. Stanome is not responsible for incorrect data selection.


CANCEL EXECUTE PIPELINE

Computing resources are initialized upon pipeline execution. The pipeline window automatically refreshes and redirects to the jobs window. Executed jobs appear in the jobs table. Refresh the window if the job is not visible. Jobs wait in the queue until computing resources are available and the status appears as pending and changes to Running. An email is sent when the job starts and also upon completion.

## Pipeline Cancellation

Click the **STOP** button on the **JOBS** window to cancel an active pipeline execution and this will abort the run.

## Pipeline Deletion

Click  to delete a **Pipeline** from the **Pipeline** window (Fig. 1). This action deletes the pipeline records entirely from a project and can't be retrieved.

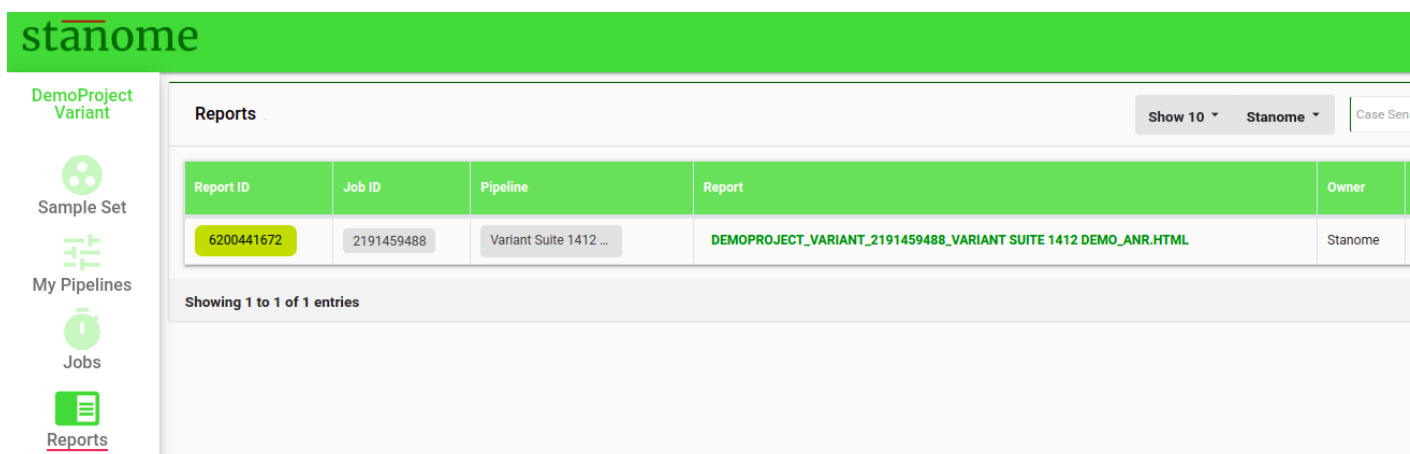


# Reports

Results of a pipeline execution are aggregated into easily understandable formats for quick viewing.

# Overview

Each pipeline execution generates an HTML report. The final report and other files can be accessed through the Reports window  (Fig.). Reports can also be accessed through the ReportID on the job details page.



Report ID	Job ID	Pipeline	Report	Owner
6200441672	2191459488	Variant Suite 1412 ...	DEMOPROJECT_VARIANT_2191459488_VARIANT_SUITE_1412_DEMO_ANR.HTML	Stanome

- Click **Report ID** to access intermediate files from a few important steps in the pipeline.
- Click the **PREFIX\_ANR.HTML** to access the final downloadable report. Default: NO REPORT AVAILABLE.

Reports are generated dynamically based on the analysis type and each report is divided into sections based on the tools used in the pipeline. The first two sections are generated for all the jobs to provide the job overview: analysis summary and sample quality.

- **Analysis Summary:**

This section consolidates the information related to the job: project, samples, and the experiment in four sub-sections:

- Project Details: This shows the information provided during the project creation and the pipeline details.
- Run Summary: Displays the job status, runtimes, and the files used for the analysis: reference, annotation or metadata files, and the samples.
- Tools: A brief summary of the tools, versions, descriptions, and citations is shown.

- Commands: A complete list of the commands used in the pipeline.

- **Sample Quality:**

Details of sample (sequencing) quality are provided in this section. It has two sub-sections:

- Metrics Table: This shows the sequence quality details of each sample in a tabular format.
- Sample Quality Plot: The average quality scores (Phred scale) of all the reads in a sample are displayed in a 96-Well plate format, each circle representing a sample. The higher the score, the better, and scores above 30 are generally considered good for the majority of the applications. Average scores are calculated using all the reads in a sample with the FastQC tool. The 96-Well plate format helps to visualize the low-quality samples, plate effects, and pooling errors easily.

The remaining sections are dynamically generated based on the pipeline type and the tools used. Two sample reports are provided below to understand the features of each report.

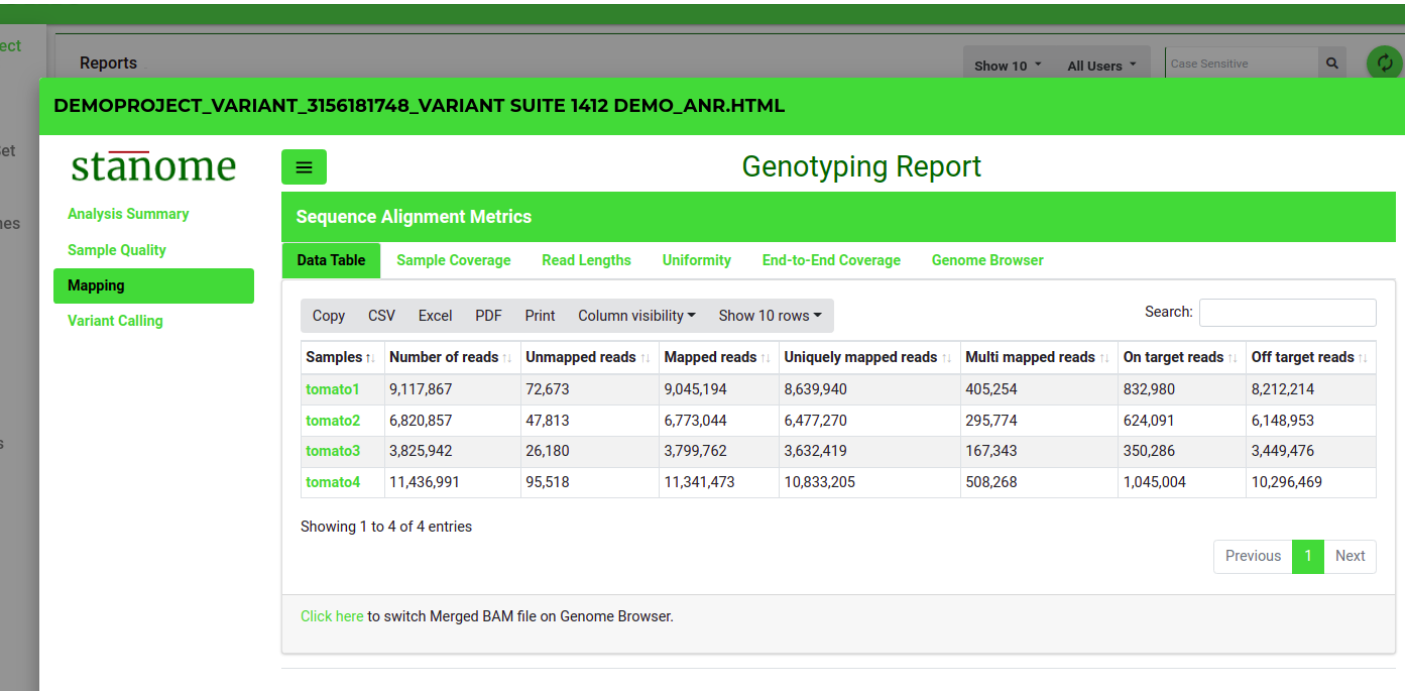
# Variant Report

Variant calling pipelines contains two exclusive sections.

## Mapping

This section provides details of the abundance/mapping statistics. This section also has three sub-sections:

- 1. Data Table: This (Fig. 1) shows several alignment statistics such as the number of total processed reads, the number of mapped or multi-mapped reads, and the uniquely mapped reads.
- 2. Sample Coverage: Allows users to explore the alignment quality through a series of plots. Sample depth of coverage in total read counts. Sample depth of coverage in percentages. On-target mapping quality in a 96-well plate format
- 3. Read Lengths: Histograms and 96-well plate plots show the read length distributions for mapped and unmapped reads.
- 4. Genome Browser: Aligned reads against the reference genome can be viewed for each sample. The genome browser is interactive and allows exploratory analysis.



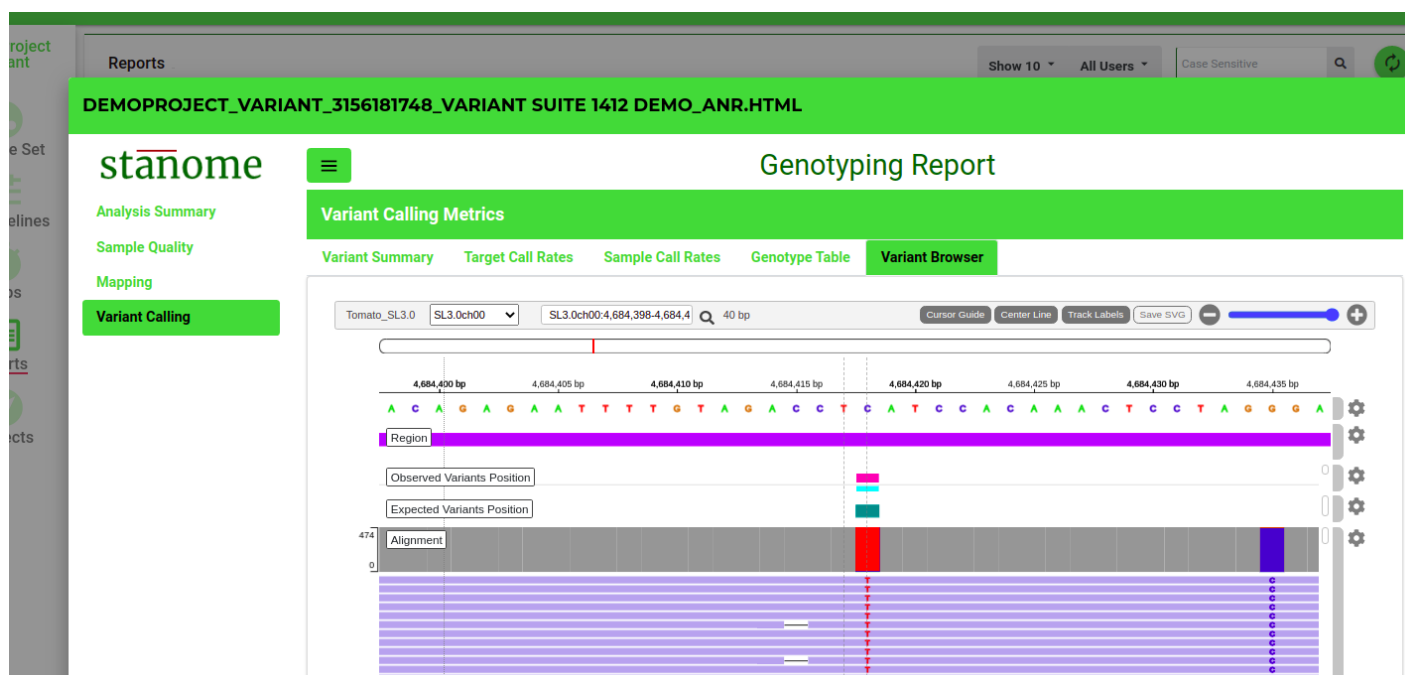
## Variant Calling

This section provides details of the abundance/mapping statistics. This section also has three sub-sections:

1. Call Rate Summary: Provides a summary of the genotype calls.
2. Target Call Rates: Genotype calling metrics for the top 100 targets are shown in the table format. The complete list can be downloaded from the **Reports** section. The position field is cross-linked to the Variant Browser.

This section also shows:

1. Genotype call distribution
2. Genotype Heatmap
3. Sample Call Rates: Histograms and 96-well plate plots show the call rate distributions from all the samples.
4. Genotypes: Shows table of genotypes obtained for each marker across all the samples
5. Variant Browser: Sequencing reads support for each variant can be viewed for each sample (Fig. 2). The genome browser is interactive and allows exploratory analysis.





# Transcript Report

This report contains three exclusive sections.

## Mapping (Salmon/Kallisto):

This section provides details of the abundance/mapping statistics. This section also has two sub-sections:

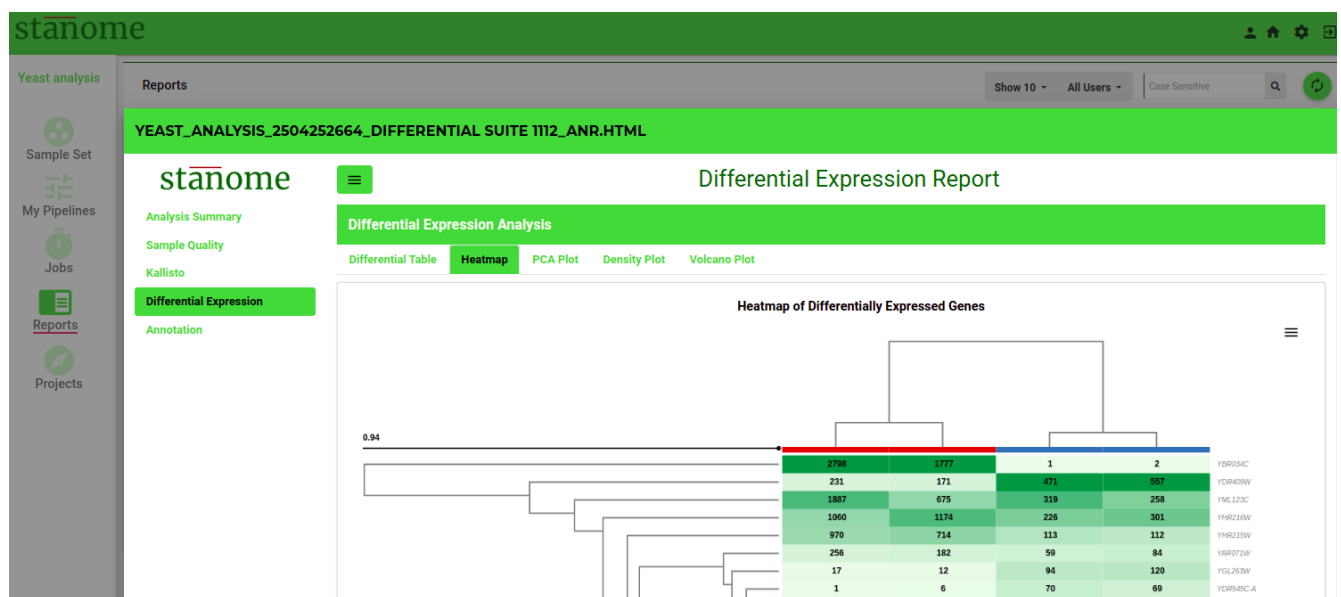
1. Data Table: This shows several quasi-alignment statistics such as the number of total processed reads, the number of mapped or multi-mapped reads, and the uniquely mapped reads.
2. Plots: Show the sample depth of coverage plots ( mapped and unmapped reads) in raw numbers and percentages.

## Differential Expression (DE):

Details of the differential expression analysis are dynamically displayed for the pipelines which contain the DE step. This has the following four sub-sections:

1. Data Table: Top 200 most significant Differentially Expressed Genes (DEGs) are listed in tabular format. The table also provides fold change values, confidence levels, and various other parameters. The tool and its various parameters that were used for identifying the DEGs are described briefly below the table. All the important parameters are also described.
2. Heatmap: The heatmap (Fig. 1) of the top 200 DEGs. Heatmap visualizes the comparison of DEGs expression across samples and within the sample. A brief description below the map enables users to understand and interpret heatmaps. The red color rectangle in the heatmap indicates the upregulation of a gene and the blue color indicates downregulation.
3. PCA & Volcano Plots: A PCA plot enables users to visualize the variability in the replicates of the two experimental conditions compared. All replicates of a condition are depicted with the same color. Grouping of samples indicates if replicates are similar among the same condition versus between the conditions.

4. The volcano plot shows significantly differentially expressed genes. It is a scatter plot between log fold change of expression among different biological conditions and the significance of the change determined from the p-value. Volcano plots enable visual inspection of expression change across all the genes.
5. Density: Shows normalization plots of samples normalized to overcome bias due to Read size and mRNA content.



## Annotation (FGSEA):

Describes *Pathway analysis* for DEGs (Fig.2). The top 100 enriched pathways are shown in the table along with their p-value and the total number of DEGs belonging to the pathway. This section also shows a bubble plot of the enriched pathways wherein the location of the bubble is determined from %DE genes in the enriched pathway to the total number of genes in the pathway.



YEAST\_ANALYSIS\_2504252664\_DIFFERENTIAL SUITE 1112\_ANR.HTML

Differential Expression Report

Analysis Summary

Sample Quality

Kallisto

Differential Expression

Annotation

Annotation

Ontology Terms

Terms vs Genes

Top GO Terms

GO Terms By Type

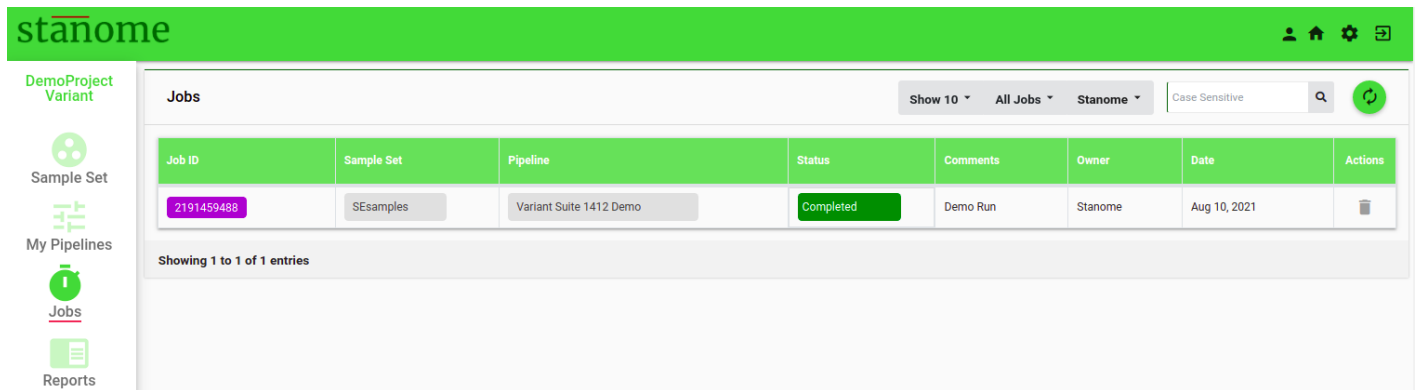
Copy CSV Excel PDF Print Column visibility ▾ Show 10 rows ▾

Search:


GO Terms	Type	Number of genes	Enrichment (Fold)	p-value	q-value
Acetyltransferase activity	molecular_function	1	0.015	0.59	0.59
Acid phosphatase activity	molecular_function	3	0.333	0.00	0.00
Active ion transmembrane transporter activity	molecular_function	2	0.021	0.09	0.09
Active transmembrane transporter activity	molecular_function	2	0.013	0.17	0.17
Adenyl nucleotide binding	molecular_function	1	0.002	0.86	0.86
Adenyl ribonucleotide binding	molecular_function	1	0.002	0.86	0.86
Aging	biological_process	1	0.012	0.25	0.25


# Jobs

Completed or unfinished pipeline runs are listed in the jobs table on the **Jobs** window (Fig. 1).



The screenshot shows the Stanome web interface. The top navigation bar is green with the 'stanome' logo and user icons. The left sidebar contains icons for 'DemoProject Variant', 'Sample Set', 'My Pipelines', 'Jobs' (highlighted), and 'Reports'. The main content area is titled 'Jobs' and includes a table with one entry. The table has columns for Job ID, Sample Set, Pipeline, Status, Comments, Owner, Date, and Actions. The entry shows Job ID 2191459488, Sample Set SEsamples, Pipeline Variant Suite 1412 Demo, Status Completed, Comments Demo Run, Owner Stanome, and Date Aug 10, 2021. Below the table, it says 'Showing 1 to 1 of 1 entries'.

Job ID	Sample Set	Pipeline	Status	Comments	Owner	Date	Actions
2191459488	SEsamples	Variant Suite 1412 Demo	Completed	Demo Run	Stanome	Aug 10, 2021	

- Click on the jobID to access job details.
- Overview tab lists details of Pipeline run i.e. Job ID, pipeline name and sample set, start and end time of execution, and status.
- Sample Deep-dive (Fig. 2) lists all the sample names.
- Click one sample to check the status or log files for that sample.
- Click the  icon on the upper right corner of the jobs details window, to delete the job.
- Stanome generates a consolidated report for each job. The report can be accessed directly by clicking the ReportID on the job details page or through the **Reports** section from the left-hand menu.

DemoProject  
Variant

Sample Set

My Pipelines

Jobs

Reports

Projects

Jobs : 2191459488 , Demo Run



Job ID	Pipeline	Sample Set *	Report ID
2191459488	Variant Suite 1412 Demo	SEsamples	6200441672
Executed on	Completed on	Status	Comments
10/Aug/21 9:26 AM	10/Aug/21 9:43 AM	Completed	Demo Run

4 Completed 0 Failed 0 Running 0 Pending

tomato1	tomato2	tomato3	tomato4
---------	---------	---------	---------