CloudNeo Documentation

Release 1

cloudneo

Oct 12, 2018

Contents

CHAPTER 1

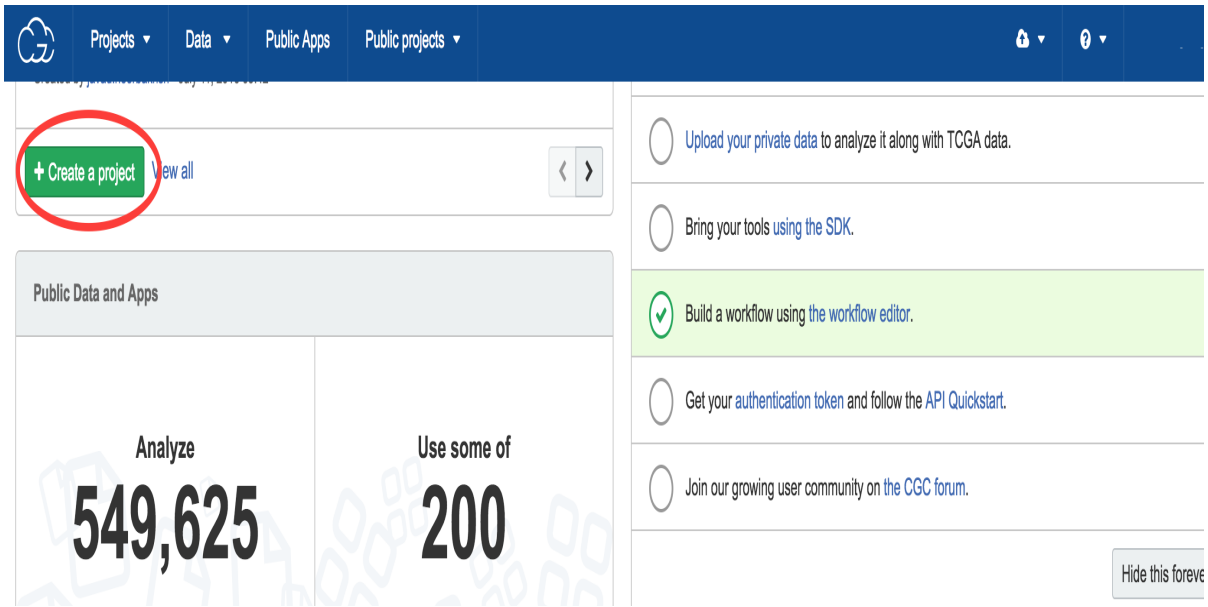
Getting Started

1.1 Register for an account

Register for a account with the Seven Bridges Cloud Platform.

1.2 Create a Project

Login to the account and create a project to do your analysis. Projects act as a container to store your data, apps, and workflows. You can create a project by clicking on 'Create a Project' button, as highlighted in the below screenshot.



See also:

More information on how to create a project is available here.

1.3 Accessing the CloudNeo App

After creating the account and the project, please contact the author Dr. Chuang for access to the CloudNeo app. Workflows and Tools are called Apps on Seven Bridges CGC platfrom. There are two ways we could provide access to the app:

- You can add one of the authors as a temporary team member to your project. The team member would then copy the app to the project.
- We could also add you as a temporary member to one of our projects. You will be able to access and copy the app to your own project.

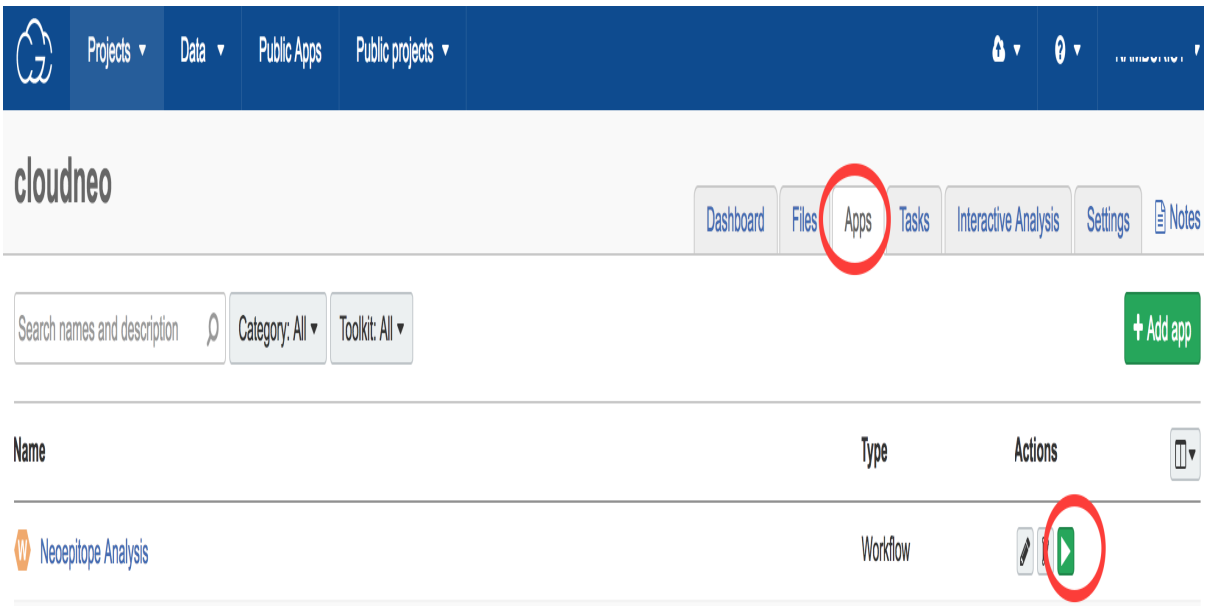
See also:

More information on adding a member to the project.

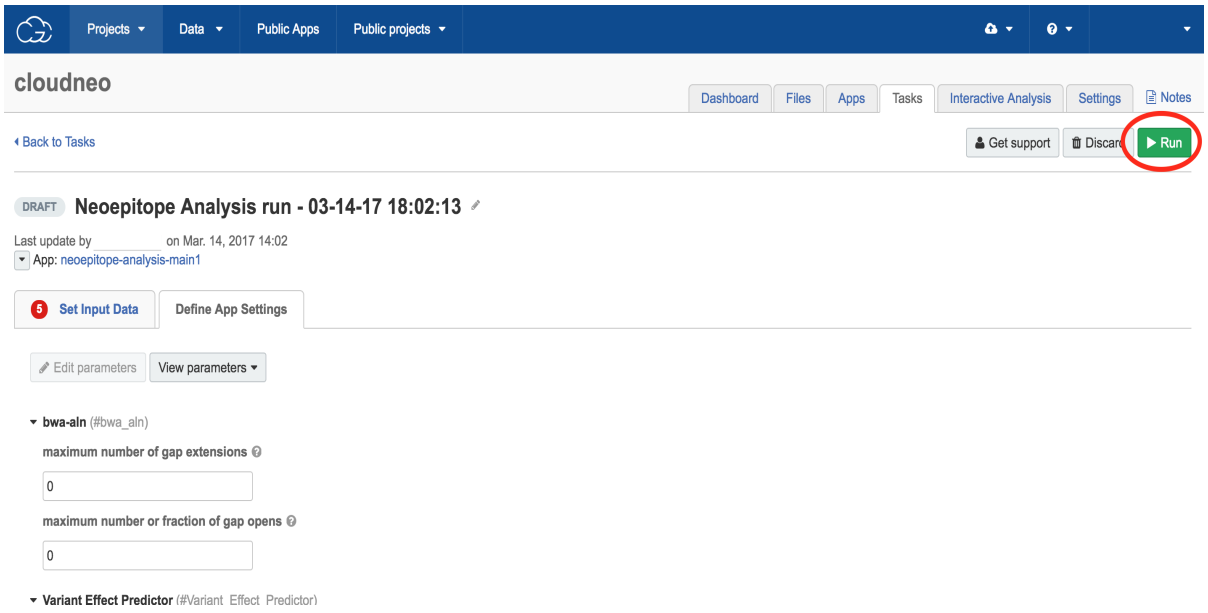
1.4 Running the CloudNeo App

1.4.1 Creating a draft task

Once the App has been copied to your project, it will be available in the 'Apps' tab as shown below in the example screenshot.



Next, click on the 'Run' button (highlighted in the screenshots) to create a draft task.

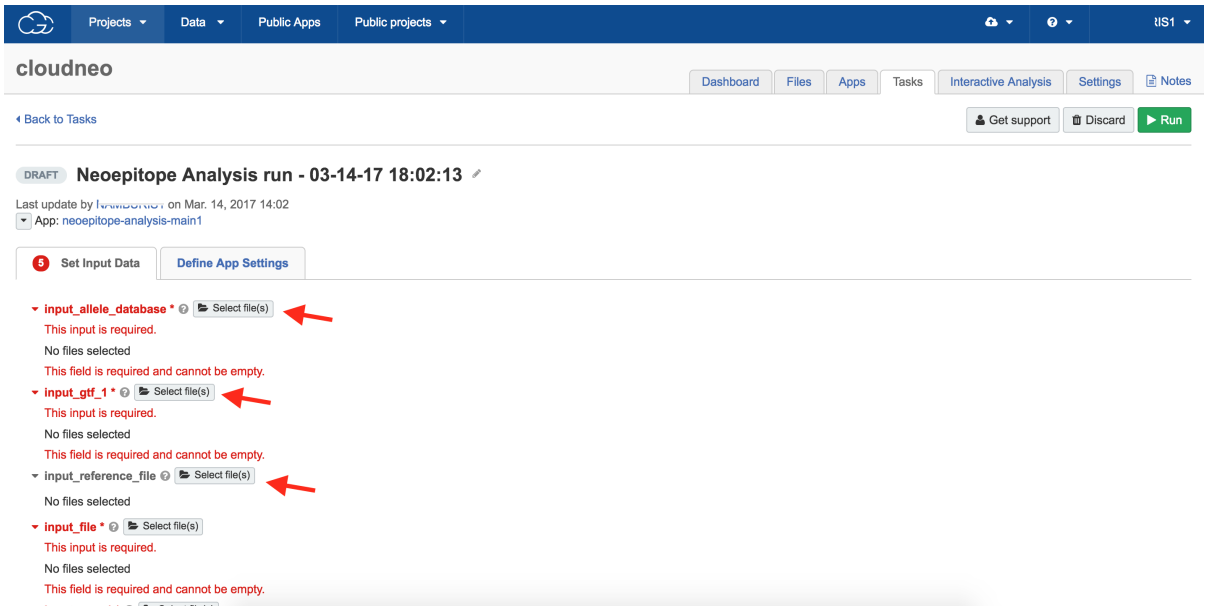


See also:

More information on how to create a draft task.

1.4.2 Select input files

Add the files that you would like to analyze to the project. The files must be in the project to select it as input. All the inputs are required for the CloudNeo app.

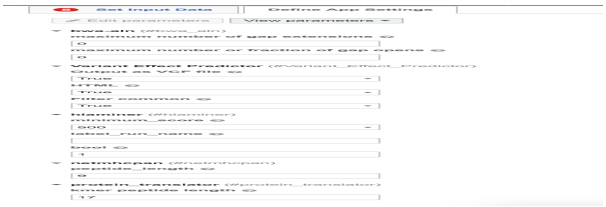


See also:

More information on uploading the files to the project.

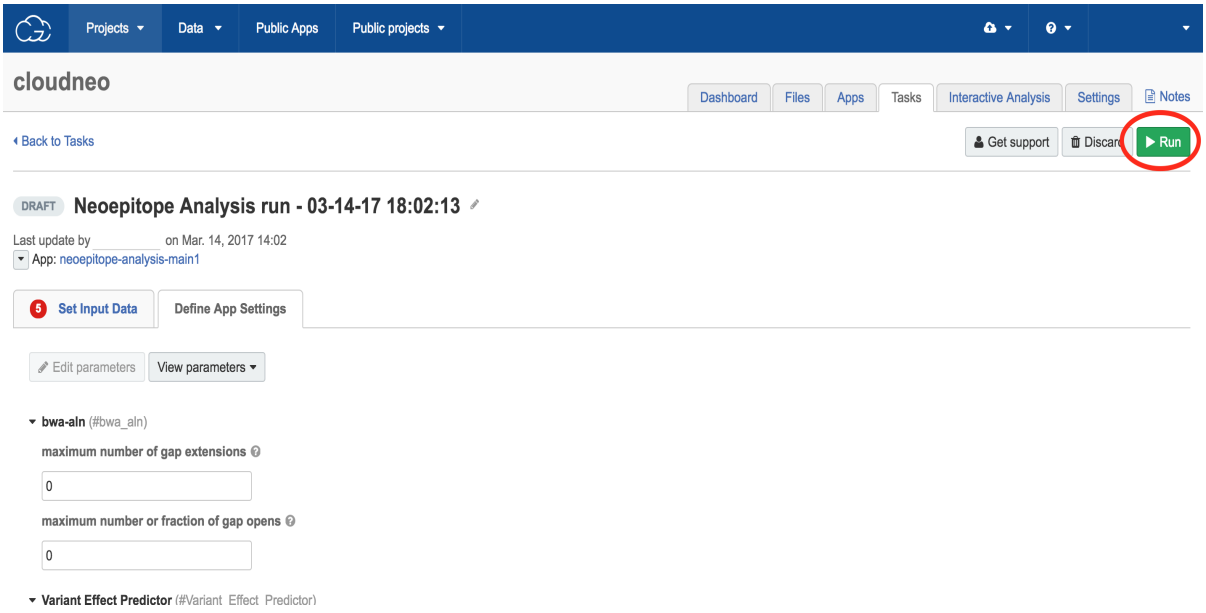
1.4.3 Define app settings

The Seven Bridges CGC interface allows you to change the app's setting before you submit the task. Click on the 'App Settings' tab to change any parameters. This is an optional step and the CloudNeo pipeline has a set of defaults that can be used.



1.4.4 Run the Analysis/Task

Click on the 'Run' button (green button - shown below) to submit the task.

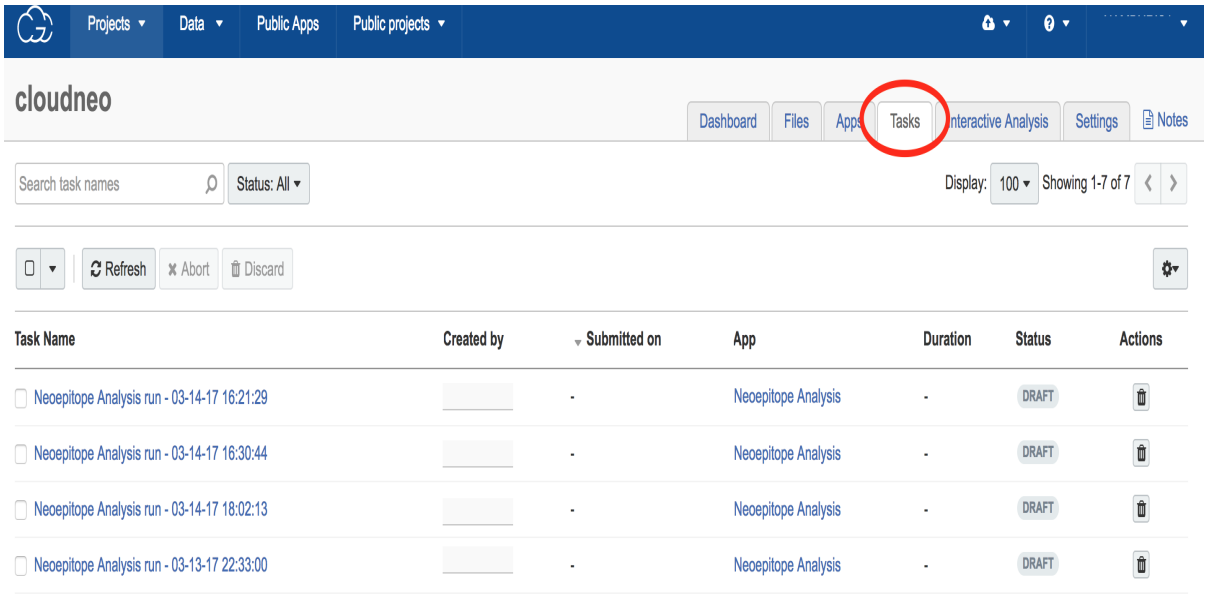


See also:

More information on how to submit a task.

1.4.5 View submitted Tasks

The tasks submitted can be viewed in the 'Tasks' tabs as shown below.

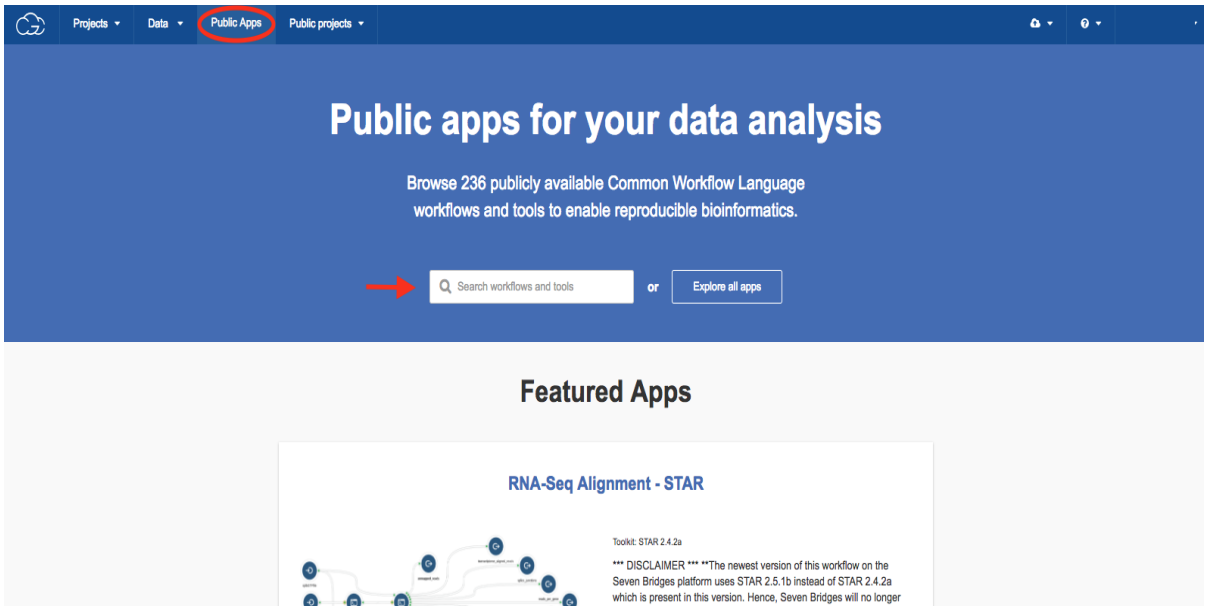


See also:

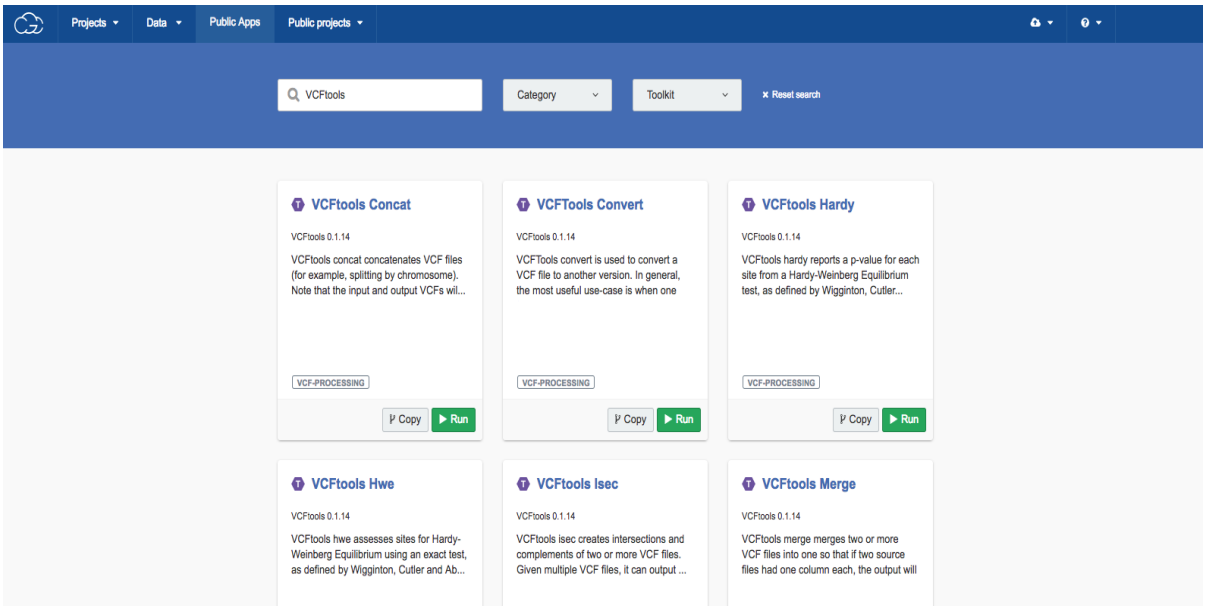
More information on how to view tasks.

1.5 Finding apps on CGC

The public apps developed by Seven Bridges can be accessed directly on the platform. Click on the 'Public Apps' tabs on the CGC platfrom as shown below. The URL can be directly accessed at: https://cgc.sbgenomics.com/public/apps.



You can search for the apps/tools on the same page. The below screenshot shows the search result for VCftools.



See also:

More information on how to search for a tool or app.

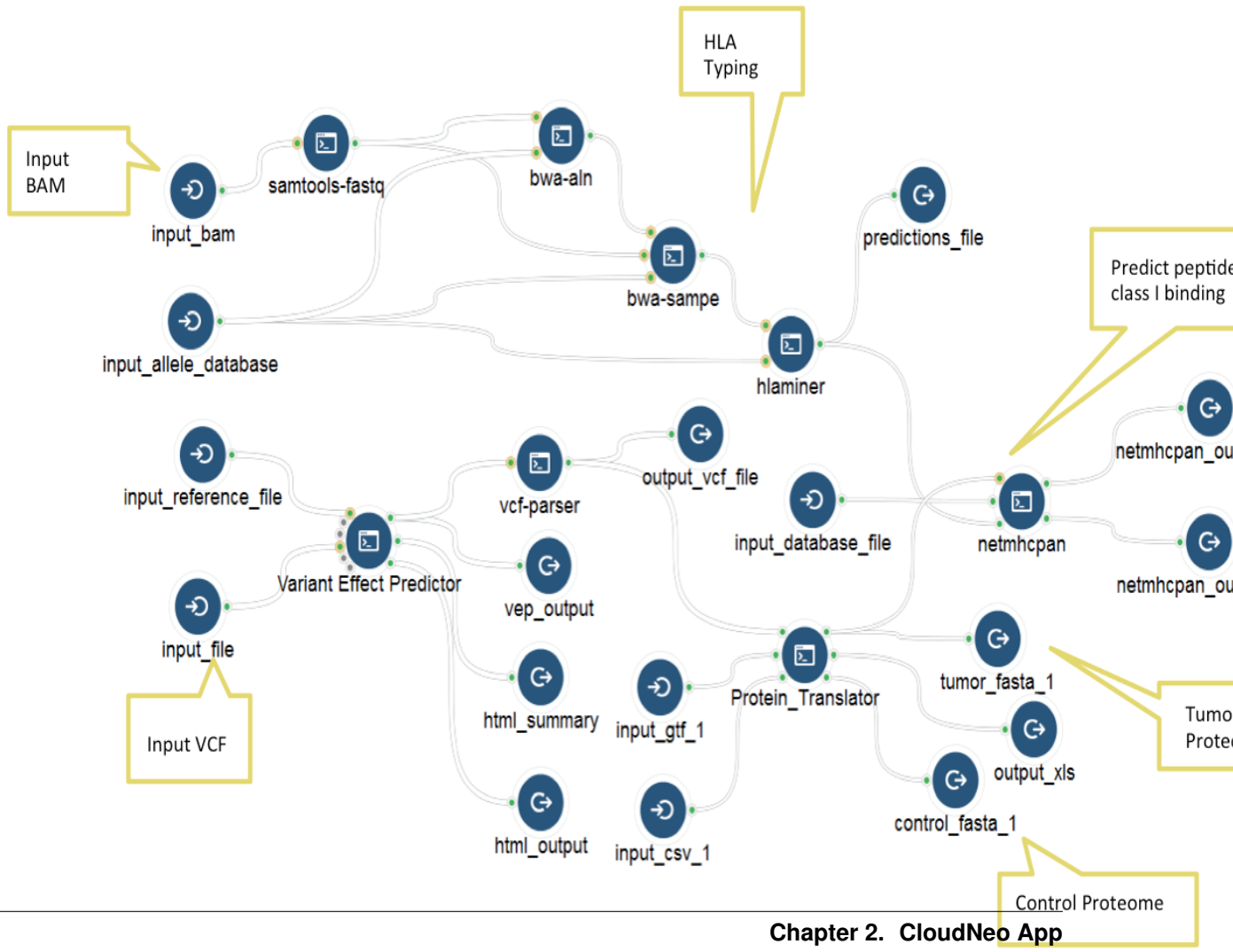
See also:

More information on how to bring your own tool to Seven Bridges Genomics Platform.

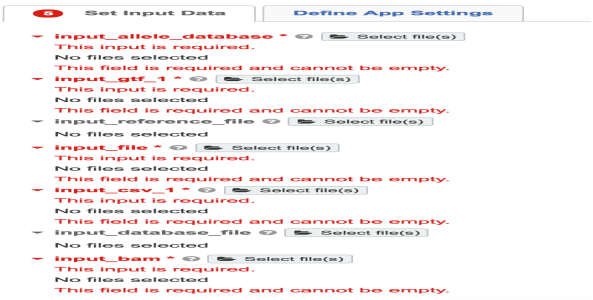
Note: The entire documentation about the Seven Bridges Genomics Platfrom is here, along with the tutorials.

CHAPTER 2

CloudNeo App



2.1 Inputs



2.1.1 Input Details

ID	Label	Type	Required	Format
input_bam	input_bam	File	Yes	BAM
input_reference_file	input_reference_file	File	Yes	TAR.GZ
input_file	input_file	File	Yes	VCF TXT
input_database_file	input_database_file	File	Yes	FASTA
input_allele_database	input_allele_database	File	Yes	fasta
input_gtf	input_gtf_1	File	Yes	GTF
input_csv	input_csv_1	File	Yes	CSV

input_bam Input BAM file to call neoepitopes

input_reference_file This is the Variant Effect Predictor reference file. This can be downloaded from the following link. homo_sapiens_vep_83_GRCh37.tar.gz

input_file Input VCF file to call neoepitopes

input_database_file This is the IMGT hla_prot.fasta file for input to netmhcpan. The input file must be formatted as below:

```
# The header in this file must be formated and renamed as netmhcpan-3.0.imgt.fasta
cat hla_prot.fasta |  perl -ne 'chomp;if(/\>\S+\s+(\S+)/){print ">$1\n";}else{print "
↪$_\n";}' > netmhcpan-3.0.imgt.fasta
```

input_allele_database This is the Hlaminer/IMGT database file, which can be downloaded from the following link: HLAminer Database. The index files can be obtained from the same link.

RNA-Seq data analysis All files starting with HLA-I_II_CDS.* must be copied for RNA-seq analysis. The HLA-I_II_CDS.fasta comprises all IMGT/HLA HLA CDS (including class I and II).

DNA-Seq data analysis All files starting with HLA-I_II_GEN.* must be copied for DNA-seq analysis. The HLA-I_II_GEN.fasta comprises all IMGT/HLA HLA genomic sequences (including class I and II).

Note: instructions to build the IMGT/HLA database are here on hlaminer github repository.

If you would like to update the coding sequences database(HLA-I_II_CDS.fasta), the instructions are provided here on the HLAminer github repository.

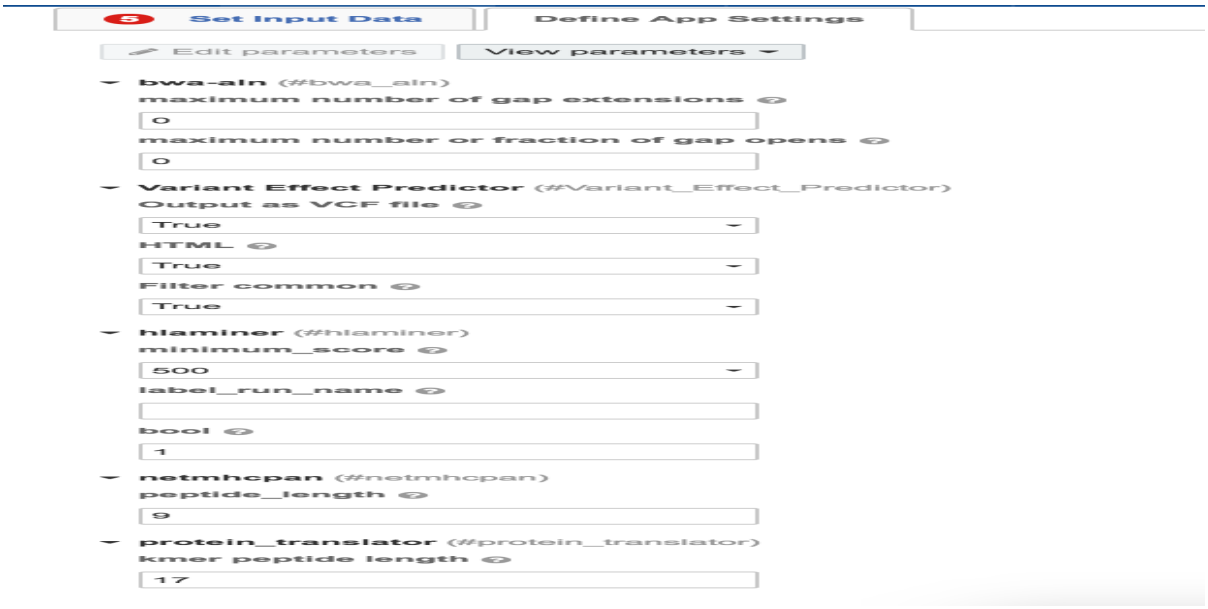
If you would like to update the genomics sequences database(HLA-I_II_GEN.fasta), the instructions are provided here on the HLAminer github repository.

---

input_gtf Input Homo sapiens GTF file from Ensembl: Homo_sapiens.GRCh37.75.gtf.gz.

input_csv Human Proteins (GRCh37.75) file: HumanProteins.GRCh37.75.csv.

2.2 App Settings



2.2.1 Settings Defaults

ID	Label	Type	Required	Prefix	Default
num_gap_extensions	maximum number of gap extensions	int	No	-e	0
maximum_number_of_gap_opens	maximum number or fraction of gap opens	int	No	-o	0
vcf	Output as VCF file	boolean	No	–vcf	TRUE
html	HTML	boolean	No	–html	TRUE
filter_common	Filter common	boolean	No		TRUE
minimum_score		enum	No	-s	500
label_run_name		string	No	-l	
bool_single_end_reads	bool	int	No	-e	1
peptide_length	peptide_length	int	No		9
kmer_peptide_length	kmer peptide length	int	No		17

2.3 Outputs

ID	Label	Type	Format
predictions_file	predictions_file	File	TXT
html_summary	html_summary	File	HTML
html_output	html_output	File	HTML
vep_output	vep_output	File	TEXT, JSON, VCF, GVF
output_vcf_file	output_vcf_file	File	VCF
netmhcpan_outputs_xls	netmhcpan_outputs_xls	File array	XLS
netmhcpan_outputs_txt	netmhcpan_outputs_txt	File array	TXT
output_xls	output_xls	File	XLS
control_fasta	control_fasta	File	FASTA
tumor_fasta	tumor_fasta	File	FASTA

2.3.1 Outputs Details

predictions_file the output file from HLAminer

html_summary The HTML output summary file from Variant Effect Predictor

vep_output The VEP annotated output VCF file

output_vcf_file Formatted VCF file as input to netmhcpan

netmhcpan_outputs_xls netmhcpan outputs in XLS format

netmhcpan_outputs_txt netmhcpan outputs in TXT format

output_xls Intermediary file from protein translator

control_fasta control file without the mutation (control Proteome)

tumor_fasta Fasta file with the mutation (tumor proteome)

CHAPTER 3

Contribute

- Issue Tracker: https://github.com/TheJacksonLaboratory/CloudNeo/issues
- Source Code: https://github.com/TheJacksonLaboratory/CloudNeo

CHAPTER 4

Support

If you are having issues, please let us know. Post your issues at CloudNeo Issues.