

---

# CloudNeo Documentation

*Release 1*

cloudneo

Oct 12, 2018



---

## Contents

---

<b>1 Getting Started</b>	<b>1</b>
1.1 Register for an account . . . . .	1
1.2 Create a Project . . . . .	1
1.3 Accessing the CloudNeo App . . . . .	2
1.4 Running the CloudNeo App . . . . .	2
1.4.1 Creating a draft task . . . . .	2
1.4.2 Select input files . . . . .	3
1.4.3 Define app settings . . . . .	4
1.4.4 Run the Analysis/Task . . . . .	4
1.4.5 View submitted Tasks . . . . .	4
1.5 Finding apps on CGC . . . . .	5
<b>2 CloudNeo App</b>	<b>7</b>
2.1 Inputs . . . . .	9
2.1.1 Input Details . . . . .	9
2.2 App Settings . . . . .	10
2.2.1 Settings Defaults . . . . .	10
2.3 Outputs . . . . .	11
2.3.1 Outputs Details . . . . .	11
<b>3 Contribute</b>	<b>13</b>
<b>4 Support</b>	<b>15</b>



# CHAPTER 1

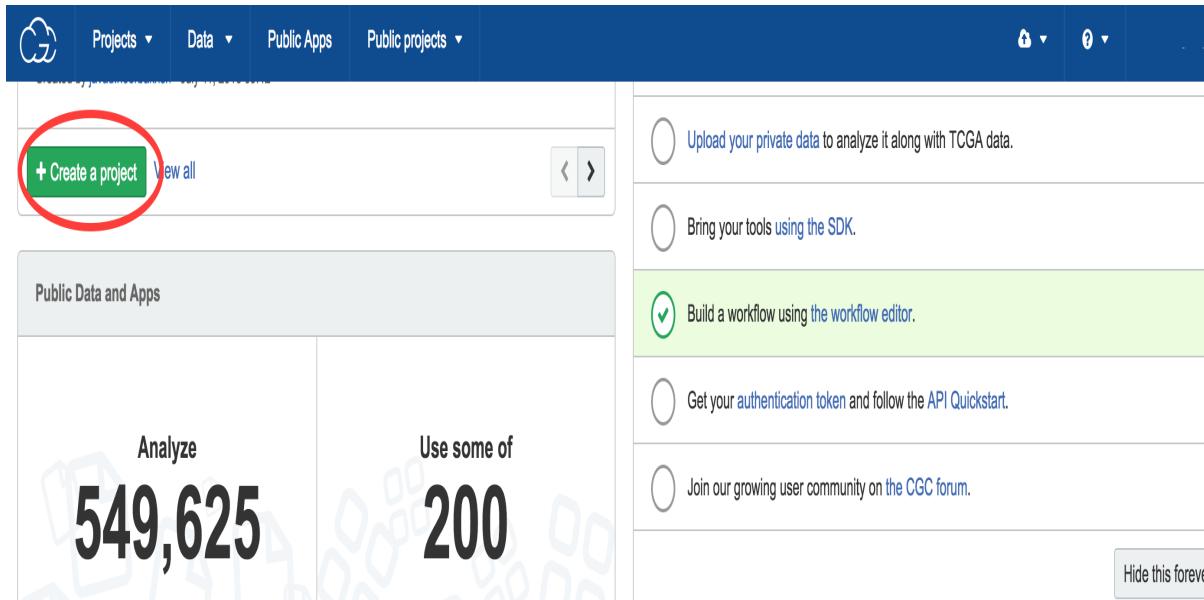
## Getting Started

### 1.1 Register for an account

Register for an 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 platform. 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.

The screenshot shows the CloudNeo web application interface. At the top, there is a navigation bar with a cloud icon, 'Projects' dropdown, 'Data' dropdown, 'Public Apps', 'Public projects' dropdown, and other user-related icons. Below the navigation bar, the main header displays the project name 'cloudneo'. Underneath the header, there is a toolbar with several tabs: 'Dashboard', 'Files', 'Apps' (which is highlighted with a red circle), 'Tasks', 'Interactive Analysis', 'Settings', and 'Notes'. Further down, there is a search bar labeled 'Search names and description' and filters for 'Category: All' and 'Toolkit: All'. On the right side of this toolbar is a green button labeled '+ Add app'. The main content area shows a table with columns 'Name', 'Type', and 'Actions'. A single row is visible, showing 'Neopeptope Analysis' under 'Name', 'Workflow' under 'Type', and three action buttons under 'Actions' (one of which is highlighted with a red circle).

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

DRAFT Neopepitope Analysis run - 03-14-17 18:02:13

Last update by [redacted] on Mar. 14, 2017 14:02  
App: neopepitope-analysis-main

**Set Input Data**

**bwa-align** (#bwa\_align)

- maximum number of gap extensions: 0
- maximum number or fraction of gap opens: 0

**Variant Effect Predictor** (#Variant Effect Predictor)

**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.

DRAFT Neopepitope Analysis run - 03-14-17 18:02:13

Last update by [redacted] on Mar. 14, 2017 14:02  
App: neopepitope-analysis-main

**Set Input Data**

**input\_allele\_database \***  This input is required.

No files selected

This field is required and cannot be empty.

**input\_gtf\_1 \***  This input is required.

No files selected

This field is required and cannot be empty.

**input\_reference\_file \***  This input is required.

No files selected

This field is required and cannot be empty.

**input\_file \***  This input is required.

No files selected

This field is required and cannot be empty.

**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.

A screenshot of the CloudNeo interface. At the top, there is a navigation bar with 'Projects', 'Data', 'Public Apps', and 'Public projects'. Below the navigation bar, the user is in a project named 'cloudneo'. In the center, there is a 'DRAFT' section for a 'Neopeptope Analysis run - 03-14-17 18:02:13'. The 'Set Input Data' tab is selected. On the right side of the draft section, there are buttons for 'Get support', 'Discard', and a green 'Run' button, which is circled in red. Below the draft section, there are buttons for 'Edit parameters' and 'View parameters'. Under the 'Set Input Data' tab, there are two sections: 'bwa-align' and 'Variant Effect Predictor'. The 'bwa-align' section has fields for 'maximum number of gap extensions' (0) and 'maximum number or fraction of gap opens' (0). The 'Variant Effect Predictor' section has a 'Predict' dropdown set to 'True'. At the bottom left, there is a 'See also:' link.

#### 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.

The screenshot shows the CloudNeo web interface. At the top, there's a navigation bar with 'Projects', 'Data', 'Public Apps', and 'Public projects'. Below the navigation bar, the title 'cloudneo' is displayed. The main content area has a header with 'Dashboard', 'Files', 'App', 'Tasks' (which is circled in red), 'Interactive Analysis', 'Settings', and 'Notes'. A search bar for 'Task names' and a status filter 'Status: All' are also present. The main table lists four task runs, all created by 'Neopeptope Analysis' and marked as 'DRAFT'. Each row includes a checkbox, a delete icon, and a refresh icon.

Task Name	Created by	Submitted on	App	Duration	Status	Actions
Neopeptope Analysis run - 03-14-17 16:21:29		-	Neopeptope Analysis	-	DRAFT	
Neopeptope Analysis run - 03-14-17 16:30:44		-	Neopeptope Analysis	-	DRAFT	
Neopeptope Analysis run - 03-14-17 18:02:13		-	Neopeptope Analysis	-	DRAFT	
Neopeptope Analysis run - 03-13-17 22:33:00		-	Neopeptope Analysis	-	DRAFT	

#### 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 platform as shown below. The URL can be directly accessed at: <https://cgc.sbggenomics.com/public/apps>.

The screenshot shows the CGC Public Apps page. At the top, there's a navigation bar with 'Projects', 'Data', 'Public Apps' (which is circled in red), and 'Public projects'. The main heading is 'Public apps for your data analysis'. Below the heading, it says 'Browse 236 publicly available Common Workflow Language workflows and tools to enable reproducible bioinformatics.' There's a search bar with a magnifying glass icon and the placeholder 'Search workflows and tools', followed by a 'Explore all apps' button. A red arrow points to the search bar.

### Featured Apps

A detailed view of a featured app card for 'RNA-Seq Alignment - STAR'. The card has a title 'RNA-Seq Alignment - STAR' and a small thumbnail image. Below the title, there's a diagram illustrating the workflow. To the right, text specifies 'Toolkit: STAR 2.4.2a' and includes a note: '\*\*\* DISCLAIMER \*\*\* The newest version of this workflow on the Seven Bridges platform uses STAR 2.5.1b instead of STAR 2.4.2a which is present in this version. Hence, Seven Bridges will no longer

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

The screenshot shows a search results page for 'VCFtools' on the Seven Bridges Genomics Platform. The top navigation bar includes 'Projects', 'Data', 'Public Apps', and 'Public projects'. The search bar contains 'VCFtools', and dropdown menus for 'Category' and 'Toolkit'. A 'Reset search' button is also present. Below the search bar, the results are displayed in a grid format:

Tool	Description	Category	Action Buttons
VCFtools Concat	VCFtools concat concatenates VCF files (for example, splitting by chromosome). Note that the input and output VCFs will...	VCF PROCESSING	Copy Run
VCFTools Convert	VCFTools convert is used to convert a VCF file to another version. In general, the most useful use-case is when one...	VCF PROCESSING	Copy Run
VCFtools Hardy	VCFtools hardy reports a p-value for each site from a Hardy-Weinberg Equilibrium test, as defined by Wigginton, Cutler...	VCF PROCESSING	Copy Run
VCFtools Hwe	VCFtools hwe assesses sites for Hardy-Weinberg Equilibrium using an exact test, as defined by Wigginton, Cutler and Ab...	VCF PROCESSING	Copy Run
VCFtools Isec	VCFtools isec creates intersections and complements of two or more VCF files. Given multiple VCF files, it can output ...	VCF PROCESSING	Copy Run
VCFtools Merge	VCFtools merge merges two or more VCF files into one so that if two source files had one column each, the output will	VCF PROCESSING	Copy Run

## 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.

---

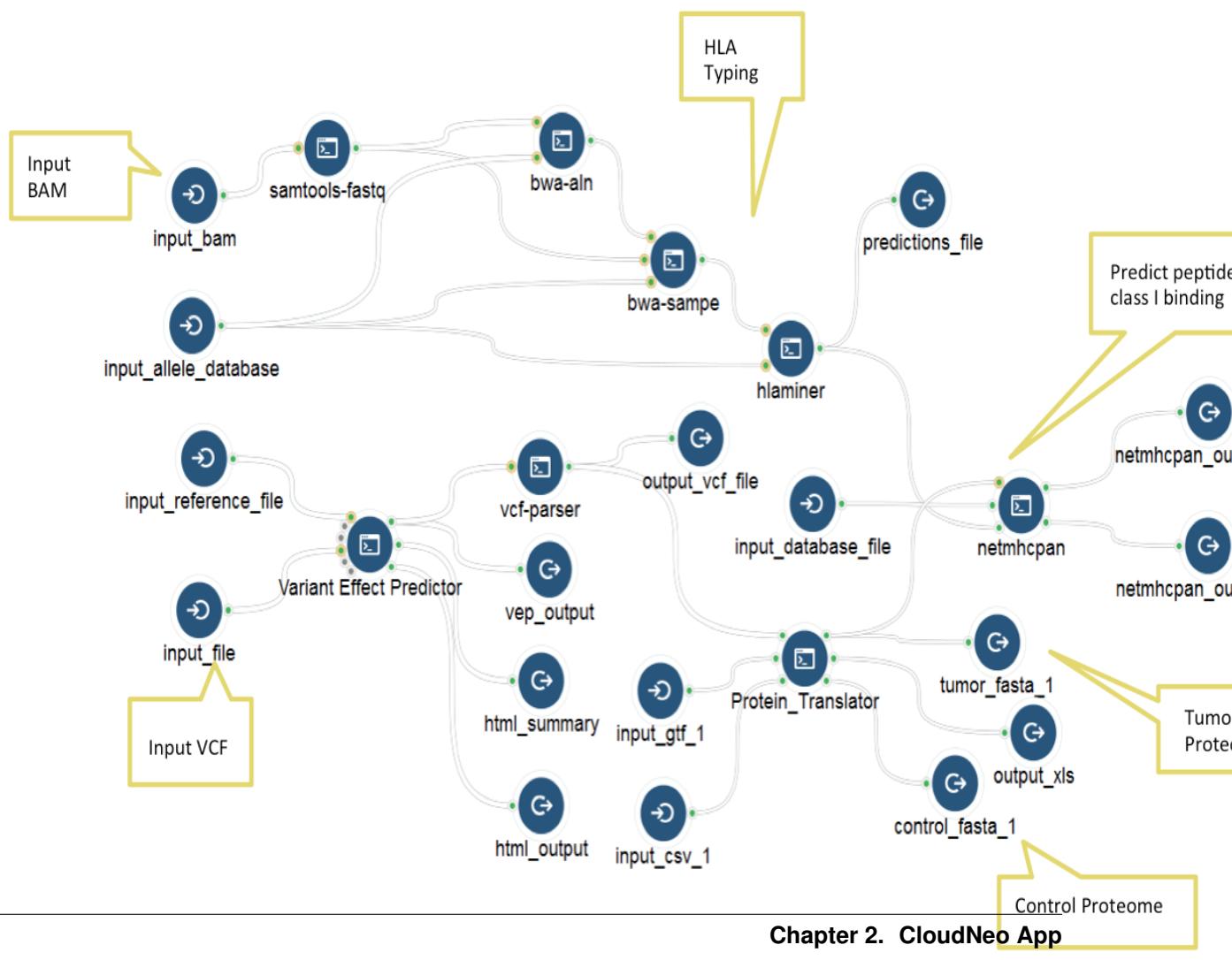
**Note:** The entire documentation about the Seven Bridges Genomics Platfrom is [here](#), along with the [tutorials](#).

---



## CHAPTER 2

## CloudNeo App



## 2.1 Inputs

The screenshot shows a configuration interface titled 'Set Input Data'. It contains several input fields with validation errors in red:

- input\_allele\_database \***: This input is required. No files selected. This field is required and cannot be empty.
- input\_gtf\_1 \***: This input is required. No files selected. This field is required and cannot be empty.
- input\_file \***: This input is required. No files selected. This field is required and cannot be empty.
- input\_csv\_1 \***: This input is required. No files selected. This field is required and cannot be empty.
- input\_database\_file**: No files selected.
- input\_bam \***: No files selected. This input is required. This field is required and cannot be empty.

### 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 netmhcpant. The input file must be formatted as below:

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

The screenshot shows a user interface for defining application settings. At the top, there are two tabs: 'Set Input Data' and 'Define App Settings'. The 'Define App Settings' tab is active. Below the tabs are two buttons: 'Edit parameters' and 'View parameters'. The main area contains several sections, each with a collapse icon (minus sign). The sections and their parameters are:

- bwa-align (#bwa\_align)**
  - maximum number of gap extensions: 0
  - maximum number or fraction of gap opens: 0
- Variant Effect Predictor (#Variant\_Effect\_Predictor)**
  - Output as VCF file: True
  - HTML: True
  - Filter common: True
- hlaaminer (#hlaaminer)**
  - minimum\_score: 500
  - label\_run\_name:
  - bool: 1
- netmhcpant (#netmhcpant)**
  - peptide\_length: 9
- protein\_translator (#protein\_translator)**
  - kmer\_peptide\_length: 17

### 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
netmhcpant_outputs_xls	netmhcpant_outputs_xls	File array	XLS
netmhcpant_outputs_txt	netmhcpant_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 HLAminder

**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 netmhcpant

**netmhcpant\_outputs\_xls** netmhcpant outputs in XLS format

**netmhcpant\_outputs\_txt** netmhcpant 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](#).