

Cloud Plugin

USER MANUAL

User manual for Cloud Plugin 21.0.1

Windows, macOS and Linux

February 17, 2021

This software is for research purposes only.

QIAGEN Aarhus
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark



Contents

1 Introduction	5
1.1 Prerequisites	5
2 Configuring cloud-based data locations	7
2.1 Configuring AWS credentials	7
2.2 Configuring Illumina BaseSpace credentials	8
3 Configuring the cloud connection	10
4 Importing data from the cloud	12
5 Exporting data to the cloud	13
6 Running workflows in the cloud	14
7 Cloud Job Search	17
7.1 Downloading all results	19
7.2 Selective download of results	20
8 Using the Cloud Server Plugin with the CLC Genomics Server	22
8.1 Configuring the Cloud Server Plugin	23
8.1.1 Configuring the AWS and CLC Genomics Cloud Engine settings	23
8.2 Configuring cloud presets	24
9 External applications in the cloud	26
10 Troubleshooting	28
11 Installing and uninstalling Workbench plugins	30

11.1 Installation of Workbench plugins	30
11.2 Uninstalling Workbench plugins	31

Chapter 1

Introduction

The Cloud Plugin provides functionality to:

- Import high-throughput sequencing data from Amazon Simple Storage Service (Amazon S3) and Illumina Basespace to be used in analyses in the *CLC Workbench*.
- Export data from the *CLC Workbench* to Amazon S3.
- Run workflows in the cloud using the *CLC Genomics Cloud Engine (GCE)*.
- Find jobs and download results from jobs run in the cloud using *GCE*.

1.1 Prerequisites

Requirements for importing and exporting data to Amazon S3

To import data from and export data to Amazon S3, you need:

- An Amazon Web Services (AWS) account.
- An AWS Identity and Access Management (IAM) user with credentials for programmatic access (access key ID, secret access key). Such a user can be set up by the AWS account administrator in the AWS Management Console:

Services → IAM → Users → Add User

Programmatic access should be granted to the user when setting the AWS access type (figure 1.1). The IAM user only needs access to the S3 locations where *GCE*, the *CLC Genomics Server*, and the *CLC Workbench* should be able to access files. The IAM user does not need API access to anything else in AWS.

- The IAM User should be granted the following permission policy: *AmazonS3FullAccess*. A more limiting policy can be used if there is a need to restrict access to specific buckets in Amazon S3.

Set user details

You can add multiple users at once with the same access type and permissions. [Learn more](#)

User name*

[Add another user](#)

Select AWS access type

Select how these users will access AWS. Access keys and autogenerated passwords are provided in the last step. [Learn more](#)

Access type* **Programmatic access**
Enables an **access key ID** and **secret access key** for the AWS API, CLI, SDK, and other development tools.

AWS Management Console access
Enables a **password** that allows users to sign-in to the AWS Management Console.

Figure 1.1: *Enabling programmatic access for an AWS IAM user can be done in the AWS Management Console. Note: This screenshot is for illustration purposes only. Details may be changed at any time by AWS.*

Requirements for importing data from Illumina BaseSpace

To import data from Illumina BaseSpace, you need an Illumina BaseSpace account.

Requirements for running workflows in the cloud

To run workflows in the cloud, and to find jobs or to download results from jobs that have been run in the cloud, you also need:

- The *CLC Genomics Cloud Engine* deployed on your AWS account. Read more about the product and request a quote here: <https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/enterprise-ngs-solutions/qiagen-clc-genomics-cloud-engine/>.
- At least one Amazon S3 bucket set up to be used for caching uploaded files. The *CLC Genomics Cloud Engine* includes a tool to set up such a bucket.
- The following permission policy granted to the IAM User logging in from the *CLC Workbench*: `AWSResourceGroupsReadOnlyAccess`. This policy grants the user the rights to automatically find the cache bucket in Amazon S3.

Chapter 2

Configuring cloud-based data locations

Data can be imported from or exported to Amazon S3 buckets in one or more AWS accounts. To do this, the AWS credentials must be configured. Data can also be imported from Illumina BaseSpace accounts, which requires no prior configuration. However, a particular client id/client secret pair can be configured, if desired.

2.1 Configuring AWS credentials

To configure AWS accounts for data import and export, go to:

File | Manage Cloud Data Access (☁)

The same menu item is also available in the main toolbar, under the Cloud menu (figure 2.1).

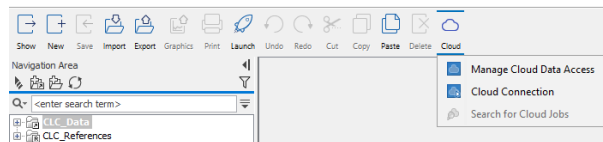


Figure 2.1: The Cloud menu provides access to the cloud configuration options, as well as the Cloud Job Search functionality.

The cloud data access configuration dialog (figure 2.2) allows you to register the credentials for one or more AWS accounts. To add an AWS account, click on **Add Amazon S3 location**. After adding one or more AWS data locations, it is possible to **Edit** or **Remove** them.

For adding or editing AWS account credentials, the following information is required (figure 2.3):

Name: A short name of your choice, identifying the AWS account. This name will be shown as the name of the data location when importing data to or exporting data from Amazon S3.

Description: An optional description of the AWS account.

AWS access key ID: The access key ID for programmatic access, set up for the AWS IAM user, as described in section 1.1.

AWS secret access key: The secret access key for programmatic access, set up for the AWS IAM user, as described in section 1.1.

AWS partition: The partition under which the AWS user is registered.

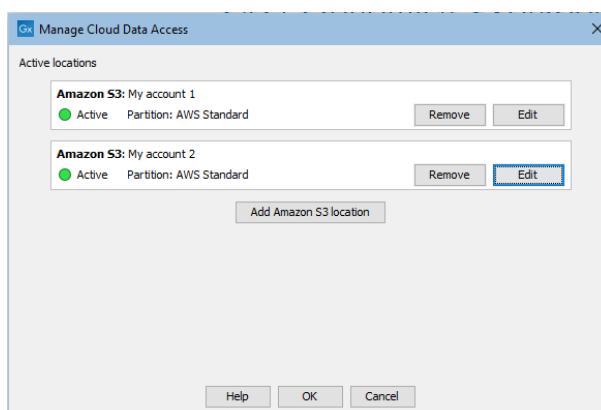


Figure 2.2: The cloud data access configuration dialog

The dialog continually validates the settings that have been entered. When the settings are valid, the Status box will contain the text "Valid" and a green icon will be shown. Click on **OK** to save the settings.

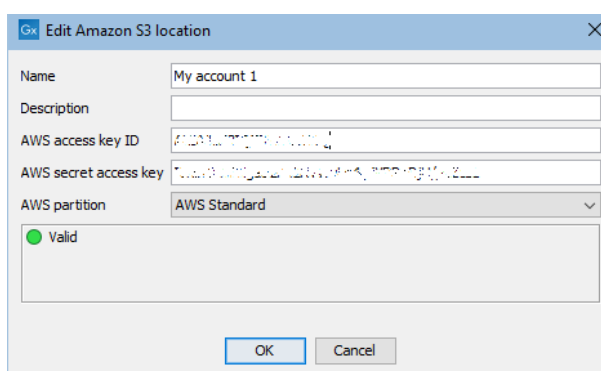



Figure 2.3: Adding an AWS account configuration dialog

When one or more AWS data locations have been added, they will be listed as data locations when importing and exporting data. You can find more information about this in chapter 4 and chapter 5.

When the cloud connection status icon at the bottom of the *CLC Workbench* looks like , a connection has been established to Amazon S3. When a connection to *CLC Genomics Cloud Engine* is also established, a different icon indicates this, as described in chapter 3.

2.2 Configuring Illumina BaseSpace credentials

Data can be imported from Illumina BaseSpace without prior configuration, as described in section 4. This section describes how to configure a specific client id/client secret pair.

To configure a client id/client secret pair for importing data from Illumina BaseSpace, go to **Edit | Preferences | Advanced** and scroll down to the **BaseSpace Configuration** section (figure 2.4). Enter the client id and client secret, and click on **OK** to save the settings.

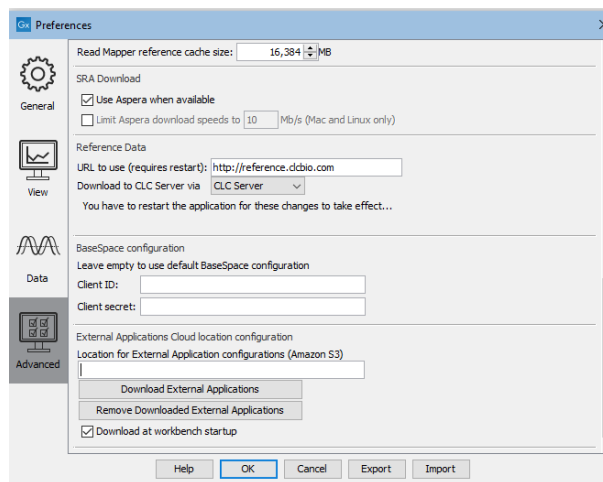


Figure 2.4: Configuring a client id/client secret pair for importing data from Illumina BaseSpace

Chapter 3

Configuring the cloud connection

To run workflows on the GCE, the *CLC Workbench* must be connected to the GCE.

A prerequisite for accessing the GCE is to set up one or more AWS data locations with credentials to access the cache bucket(s) of the GCE. Adding an AWS data location is described in chapter 2.

Once the AWS data locations have been added, the *CLC Workbench* can be connected to the GCE. To do this, go to:

File | Cloud Connection 

The same menu item is also available in the main toolbar under the Cloud menu (figure 7.1).

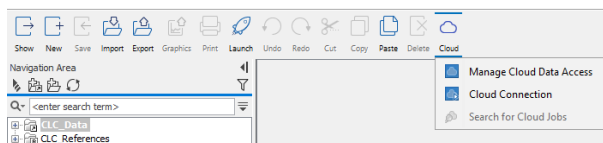


Figure 3.1: The Cloud menu provides access to the cloud configuration options, as well as the Cloud Job Search functionality.

This opens the dialog shown in figure 3.2. The following two settings must be configured:

URL: The URL pointing to the GCE service, as described in section 1.1.

AWS S3 cache bucket: The cache bucket to be used when uploading data to the cloud, as described in section 1.1.

The **Accept untrusted certificate** checkbox should be selected when GCE has been set up with a self-signed certificate.

Connecting to and disconnecting from GCE

Click on the **Log In** button in the Cloud Connection dialog to open a web browser window where you can log in to GCE using your company credentials. Once you have completed the login in the browser window, go back to the *CLC Workbench* and click on the **OK** button in the Cloud Connection dialog to complete the log in process. When the authentication succeeds, the Status box will contain the name and region of the connected user, and a green icon will be shown (figure 3.2).

Click on the **Log Out** button in the Cloud Connection dialog and then click on the **OK** button to

disconnect the *CLC Workbench* from *GCE*. You will be asked to confirm this is what you wish to do.

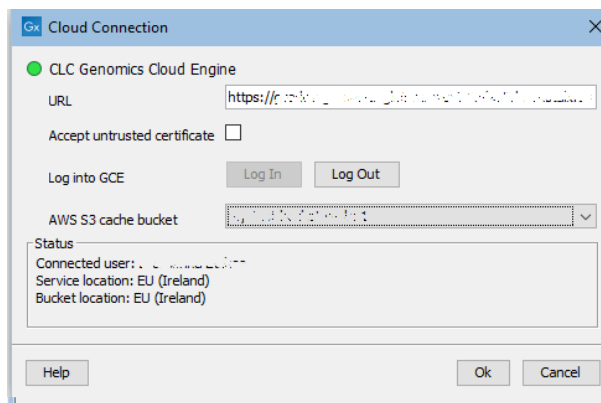



Figure 3.2: A valid **GCE** configuration is indicated by the green icon and the user details in the Status box.

When the cloud connection status icon at the bottom of the *CLC Workbench* looks like , a connection has been established to Amazon S3 and to the *CLC Genomics Cloud Engine*. With this connection status, you can import data to and export data from the cloud, and submit workflows to run on *GCE*.

Chapter 4

Importing data from the cloud

Using a *CLC Genomics Workbench*, next-generation sequencing (NGS) data can be imported from the configured cloud locations using NGS Import tools or using on-the-fly import when running a workflow. The configured cloud locations will be available to select in the import or workflow wizard (figure 4.1). When using on-the-fly import in a workflow, CLC format files can also be imported from a cloud location.

Using a *CLC Main Workbench*, sequencing trace data and CLC format files can be imported from configured cloud locations using on-the-fly import in workflows. The configured cloud locations will be available to select in the workflow wizard from a drop-down menu, similar to the drop-down menu shown at the top of figure 4.1.

Files selected in a cloud location are first downloaded to a temporary folder, and are subsequently imported into the *CLC Workbench*.

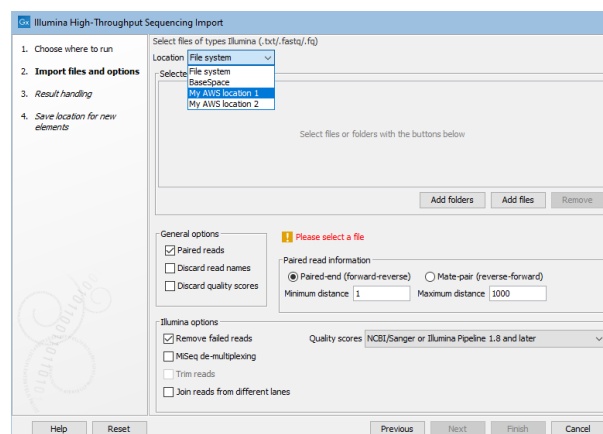


Figure 4.1: Configured cloud locations can be selected when using NGS importers.

Chapter 5

Exporting data to the cloud

To export data to a cloud location, launch the exporter, and in the final configuration step, select the desired export location from the drop-down menu (figure 5.1). When exporting to Amazon S3, new folders can be created by right-clicking the name of a folder or a bucket, and selecting "New folder".

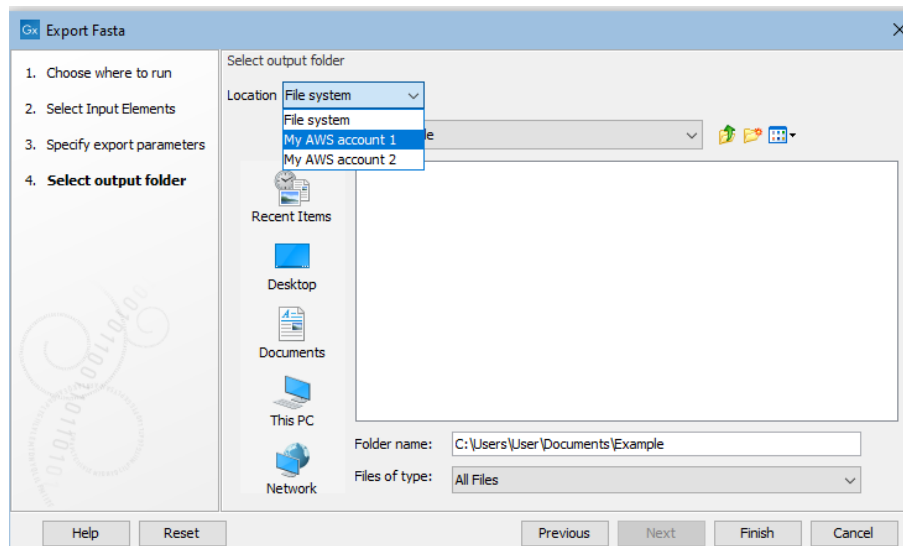


Figure 5.1: A cloud location can be selected to export data to.

Chapter 6

Running workflows in the cloud

To run workflows on the *CLC Genomics Cloud Engine*, select "CLC Genomics Cloud Engine" in the first wizard step when launching a workflow (figure 6.1).

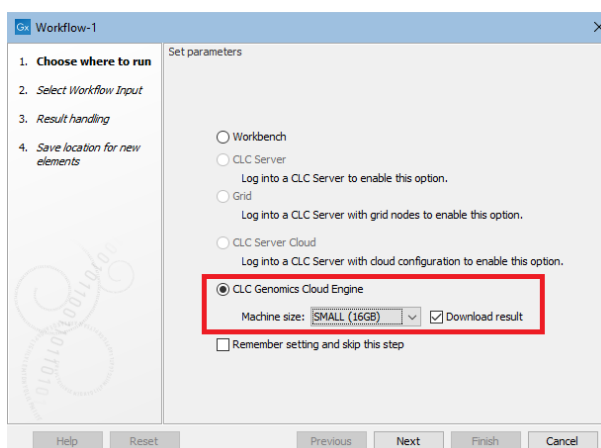


Figure 6.1: Select "CLC Genomics Cloud Engine" to run a workflow in the cloud.

Machine size and downloading results

Select the desired machine size in the drop-down menu. The available options can be configured by the administrator of the *CLC Genomics Cloud Engine*.

When running small workflows where you wish to download all the workflow results, keep the **Download result** checkbox checked. For other situations, we recommend this option is not selected. The *CLC Workbench* must be left running until the workflow completes for the results to be downloaded automatically. Workflow results can be downloaded later using the Cloud Job Search functionality, described in chapter 7.

Data inputs and outputs

The following data sources are available when launching a workflow to run on the *CLC Genomics Cloud Engine*:

- **Data already imported into a CLC location** Choose "Select from Navigation Area". The

selected data will be uploaded to the AWS S3 cache bucket configured in the GCE settings, described in chapter 3. If a given data element is already present in the cache bucket, as determined by the CLC URL of the data element and the time of its latest modification, then it will not be uploaded.

- **Locally available data to be imported on the fly when the workflow is run** Choose "Select files for import", click "Browse" and select "File system" from the drop-down menu. The data will be uploaded to the AWS S3 cache bucket. If a file is already present in the cache bucket, as determined by the full path of the file and the time of its latest modification, then it will not be uploaded.
- **Data in the cloud, to be imported on the fly when running the workflow in the cloud** Choose "Select files for import", click "Browse", and select a cloud-based location from the drop-down menu. The analysis will be carried out directly on the data in the cloud, and will not be transferred to the *CLC Workbench*.

In the "Output location in Amazon S3" wizard step, you specify an output bucket and folder in Amazon S3 to save the workflow results to. Note that workflow results will always be saved in Amazon S3, even if the "Download result" option was checked in the first wizard step. Information about data to be uploaded is also displayed here, as shown in figure 6.2. Any data already present in the cloud cache will not be uploaded.

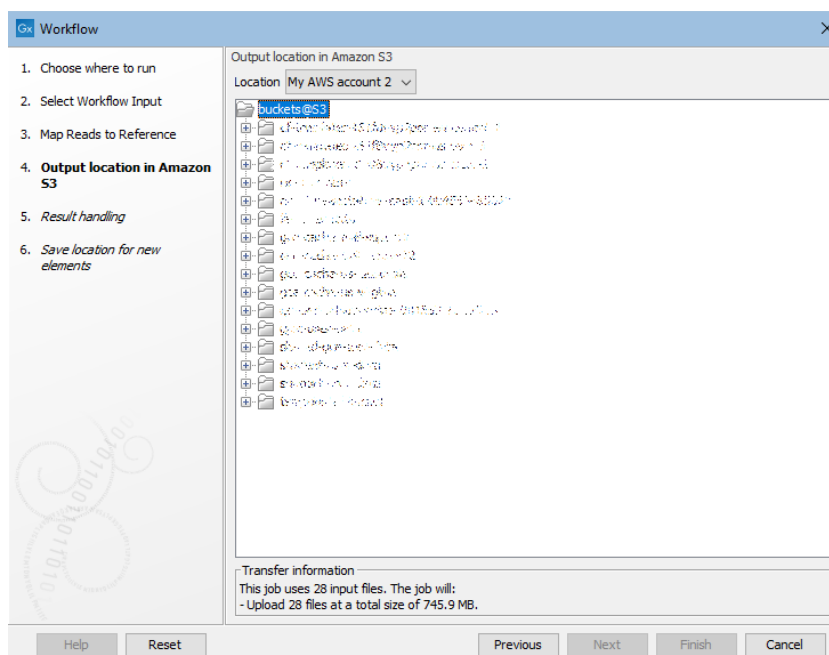


Figure 6.2: Specifying a location for saving workflow results to in Amazon S3. Information about data to be uploaded when the workflow is launched is provided near the bottom of this wizard step.

In the last wizard step, a local location must be selected for workflow outputs, even if the "Download result" checkbox was not checked in the first configuration step. This location is used to save log files to in some circumstances, for example, if a workflow run fails for particular reasons.



Following the progress of workflow jobs run on the cloud

Each workflow submitted to the cloud is submitted as a *batch* consisting of *jobs*. A batch may consist of just a single job. Multiple jobs are included in a batch when:

- The "Batch" checkbox is selected in the workflow wizard, and/or
- The workflow design includes control flow elements, as described in the *CLC Genomics Workbench* manual: https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Advanced_workflow_batching.html.

Each job within a batch is executed as a separate job in the cloud, potentially in parallel on separate server instances.

You can follow the progress of the workflow in the Processes area of the *CLC Workbench* (figure 6.3). The icon next to the process indicates the status of the job submission:

-  This icon indicates that data is being transferred to the cloud. When this icon is displayed, do not interrupt the connection to the cloud, e.g. do not disconnect from the cloud or shut down your computer.
-  This icon indicates the job submission is complete, including any data transfer. When this icon is displayed, you can safely disconnect from the cloud, and shut down your computer if you wish.

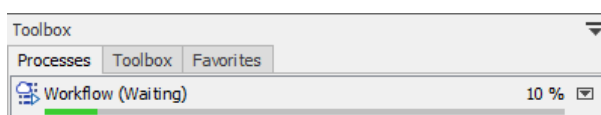


Figure 6.3: The icon next to the cloud process in the Processes area indicates the submission of this job, including data transfer, is complete.

When the job submission is complete, right-clicking the arrow next to a process and selecting "Show in Cloud Job Search" will open the batch in the Cloud Job Search (figure 6.4). See chapter 7 for further details.

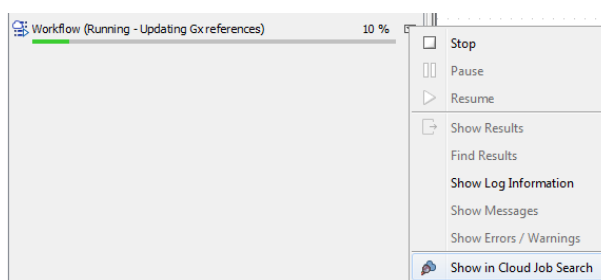


Figure 6.4: You can open an individual job in the Cloud Job Search tool by right-clicking on the arrow next to a process in the Processes area. This option is only available when the job submission to the cloud is complete, including any data transfer.

Chapter 7

Cloud Job Search

To open the Cloud Job Search tool, go to

Edit | Search for Cloud Jobs (🔍)

The same menu item is also available in the main toolbar, under the Cloud menu (figure 7.1).

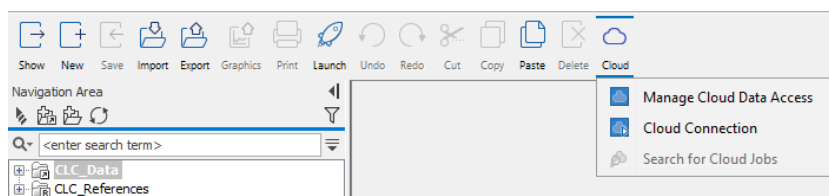


Figure 7.1: The Cloud menu provides access to the cloud configuration options, as well as the Cloud Job Search functionality.

Using Cloud Job Search, you can:

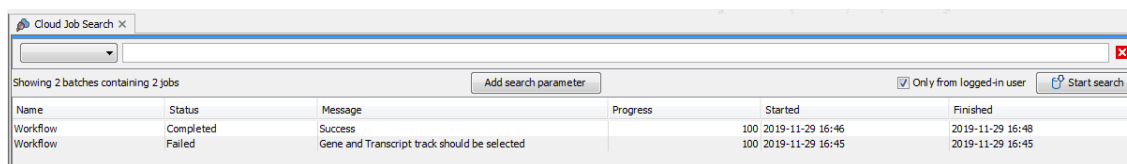
- Search for jobs submitted to the *CLC Genomics Cloud Engine*, and inspect their status, progress and other properties.
- Retrieve results produced by a job submitted to the *CLC Genomics Cloud Engine*.

If a search is run without specifying search criteria, the latest batches submitted to the *CLC Genomics Cloud Engine* will be listed (figure 7.2). A maximum of 50 results will be returned. To narrow down the list, specify search criteria using the fields at the top. To add more criteria, click on the **Add search parameter** button. To start the search, click on the **Start search** button on the top, right hand side.

With the "Only from logged-in user" checkbox selected, then the list retrieved will include only jobs submitted by the user currently logged into the *CLC Genomics Cloud Engine* from the Workbench. When this checkbox is not selected, jobs submitted by *any user* of the *CLC Genomics Cloud Engine* instance can be retrieved.

If more than 50 batches fulfill the search criteria, only the 50 most recently submitted will be listed. The search criteria must be further restricted to find older submissions of interest.

Note: The table is not refreshed automatically. The "Start search" button must be pressed to update the information, including job status information.



The screenshot shows a web application window titled 'Cloud Job Search'. At the top, there is a search bar and a 'Start search' button. Below the search bar, it says 'Showing 2 batches containing 2 jobs'. There is an 'Add search parameter' button and a checkbox for 'Only from logged-in user'. The main content is a table with the following data:

Name	Status	Message	Progress	Started	Finished
Workflow	Completed	Success	100	2019-11-29 16:46	2019-11-29 16:48
Workflow	Failed	Gene and Transcript track should be selected	100	2019-11-29 16:45	2019-11-29 16:45

Figure 7.2: The Cloud Job Search allows you to search for and inspect the properties of jobs that have been submitted to the cloud.

In the side panel settings of a Cloud Job Search, you can select the columns to display and control whether batches or individual jobs are listed in the table (figure 7.3). When the **Collapse batches** option is checked, each batch is shown as a single row in the table. When unchecked, each individual job is shown in its own row.

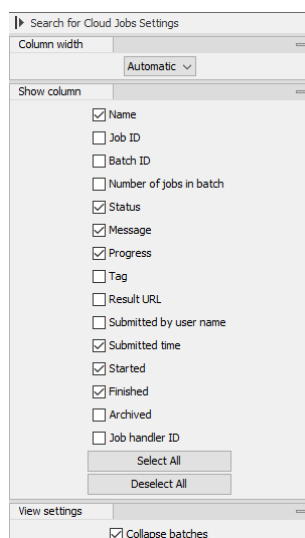


Figure 7.3: The Cloud Job Search side panel settings are used to configure what is shown in the table.

All jobs in a batch can be canceled by right-clicking on the row in the table corresponding to the batch (or a job in a batch, when displaying individual jobs), and selecting **Cancel... | Cancel all jobs in batch** from the menu that appears.

After a cloud job has completed, results can be retrieved by selecting the jobs of interest and then using the buttons at the bottom of the view (figure 7.4):

- **Download All Results** Download all results from the selected jobs, optionally including any exported files. See section 7.1 for further details.
- **Download Metadata** Download only the Workflow Result Metadata table for the selected jobs. Results can then be downloaded selectively using the Workflow Result Metadata table, as described in section 7.2.

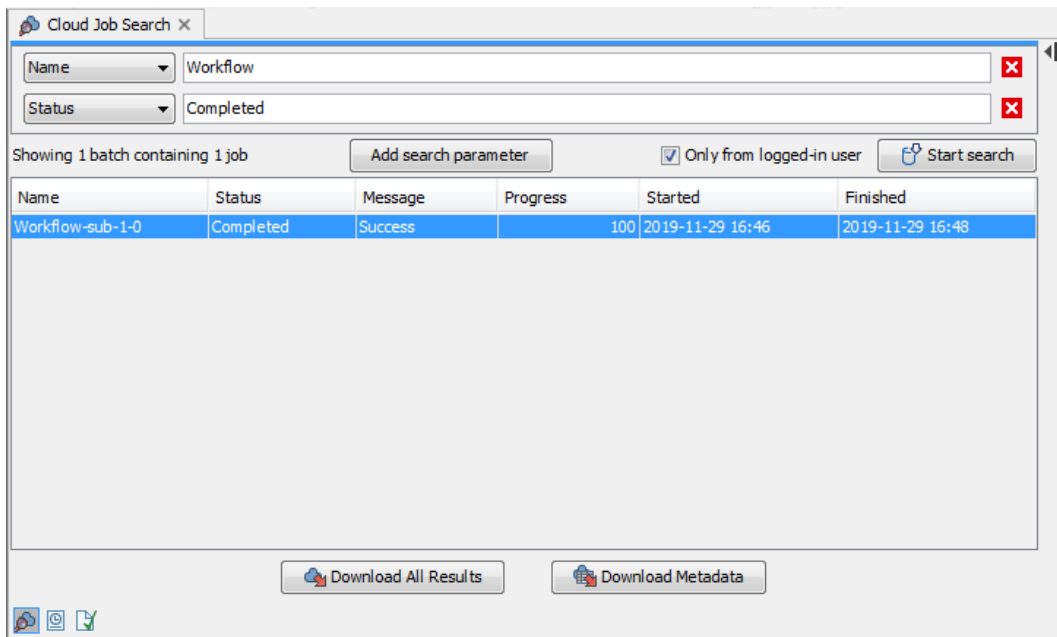


Figure 7.4: The **Download All Results** and **Download Metadata** buttons can be used to retrieve results from the cloud job submissions.

7.1 Downloading all results

To download all results, optionally including any exported data, select one or more rows in Cloud Job Search table, and click on the **Download All Results** button. Data elements will be downloaded into the Navigation Area in the folder structure specified in the workflow design for the output elements.

Check the **Also download exported files** checkbox in the Result handling step of the wizard to download exported files in addition to downloading the other results. You can specify the location to save the exported data to in the **Export directory** field. Note: Specifying a cloud location when downloading exported files is not supported.

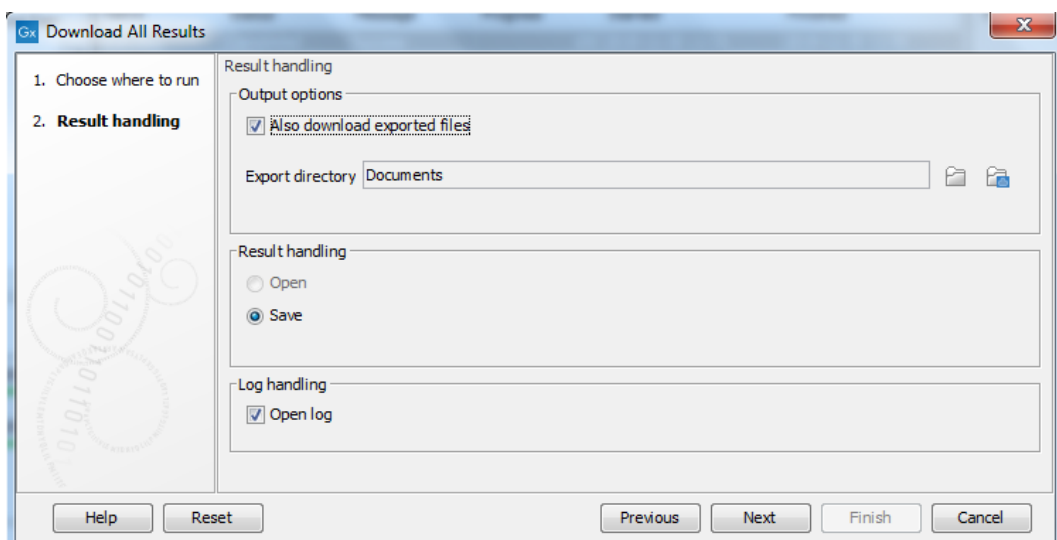


Figure 7.5: The "Also download exported files" checkbox allows you to download exported files as well as results to be saved to the **CLC Workbench** Navigation Area.

In cases where a job failed, the **Download All Results** button can be used to download logs and other technical information. More information about troubleshooting is provided in chapter 10.

7.2 Selective download of results

Individual results can be downloaded using the Workflow Result Metadata table, which is generated for each batch. This is particularly useful when the amount of data generated by a workflow execution is so large that it is not practical to download all the results at once.

In the Cloud Job Search table, download the Workflow Result Metadata table for selected jobs by clicking on the **Download Metadata** button.

In a Workflow Result Metadata table, each workflow output is shown in a separate row. The path to the output in the AWS S3 location is indicated in the column titled "External path", as shown in figure 7.6. When one or more rows in the table are selected, the **Find Associated Data** button is enabled. Clicking on this button opens a new table, where the external references (not yet downloaded) are listed, as well as any results already downloaded into the Navigation Area.

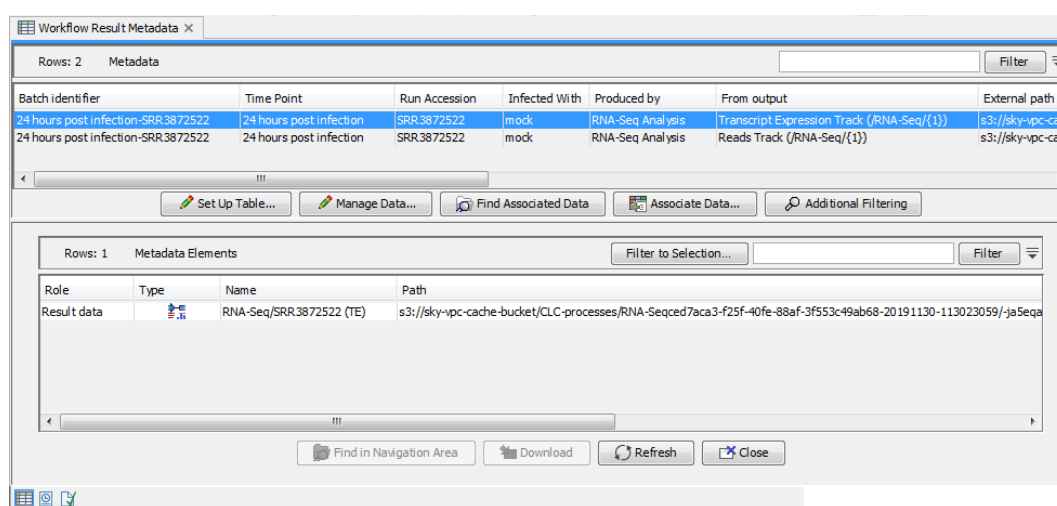


Figure 7.6: The "External path" column in the Workflow Result Metadata table shows the path to the result in AWS S3. After the "Find Associated Data" button is clicked, a Metadata Elements table which lists associated data elements and their locations. Here one element is local, and one is in the cloud.

To download particular data elements from the cloud, select the desired rows in the Metadata Elements table, and click on the **Download** button. Data elements downloaded are placed in the subfolder (if any) that was specified by the workflow design for the given output. New folders are created as required. Therefore, we recommend selecting the same folder to store different results for the same workflow. This ensures that the outputs are organized in the same folder structure that they would have been if the workflow had been executed by the *CLC Workbench*.

When an data element is downloaded, it is automatically associated with the relevant row of the Workflow Result Metadata table, and can be found by clicking on the **Find Associated Data** button again (figure 7.7). External references will not be removed from the Workflow Result Metadata table by downloading a result. Results can be downloaded this way any number of times.

A connection to the *CLC Genomics Cloud Engine* is required to find the Workflow Result Metadata

Workflow Result Metadata x

Rows: 2 Metadata Filter

Batch identifier	Time Point	Run Accession	Infected With	Produced by	From output	External path
24 hours post infection-SRR3872522	24 hours post infection	SRR3872522	mock	RNA-Seq Analysis	Transcript Expression Track (RNA-Seq/(1))	s3://sky-ipc-cad
24 hours post infection-SRR3872522	24 hours post infection	SRR3872522	mock	RNA-Seq Analysis	Reads Track (RNA-Seq/(1))	s3://sky-ipc-cad

Buttons: Set Up Table... Manage Data... Find Associated Data Associate Data... Additional Filtering

Rows: 2 Metadata Elements Filter to Selection... Filter

Role	Type	Name	Path
Result data		SRR3872522 (TE)	CLC_Data/test/RNA-Seq
Result data		RNA-Seq/SRR3872522 (TE)	s3://sky-ipc-cache-bucket/CLC-processes/RNA-Seqced7aca3-f25f-40fe-88af-3f553c49ab68-20191130-113023059/-jaSeqa

Buttons: Find in Navigation Area Download Refresh Close

Figure 7.7: When the results have been downloaded, they are automatically associated with the relevant row of the Workflow Result Metadata table, and can be located by pressing the "Find Associated Data" button again.

table and download it using the Cloud Job Search functionality. However, only the AWS S3 connection is required for downloading results via the Workflow Result Metadata table.

Thus, by saving the Workflow Result Metadata table, you can download the results from AWS S3 at a later point, as long as the data is still available in AWS S3. You do not need a connection to the CLC Genomics Cloud Engine or to use the Cloud Job Search for this activity.

Note: Exported data cannot be downloaded selectively. To download data exported by the workflow, you must use the **Download All Results** button in the Cloud Job Search tool. See section section 7.1 for further details.

Chapter 8

Using the Cloud Server Plugin with the CLC Genomics Server

When connected to a *CLC Genomics Server* where the Cloud Server Plugin has been installed and configured by the server administrator, it is possible to import data to and export data from AWS S3 and submit jobs to the *CLC Genomics Cloud Engine* through the *CLC Genomics Server*. This can be particularly useful when a large amount of data is located in a server location or a file system location available to the server. In these cases, the data will be uploaded to the cloud directly from the *CLC Genomics Server*, thereby removing the need to keep the *CLC Workbench* running during the upload.

Submitting jobs to the cloud through the *CLC Genomics Server* works in a very similar way to submitting directly from a *CLC Workbench*. In the first wizard step when running the workflow, select "CLC Server Cloud". The drop-down menu contains the cloud presets that have been configured by the server administrator. A cloud preset controls characteristics for a job submission, such as hardware requirements and whether the results should be automatically downloaded or not. The server administrator also controls the cache bucket that will be used for submitting jobs through the *CLC Genomics Server*.

When submitting jobs to the cloud through the *CLC Genomics Server*, the user account registered for the submission will be the user logged into the *CLC Genomics Server*. It is not the user from the GCE configuration settings in the cloud configuration dialog in the *CLC Workbench*. In practice, these user accounts may be the same.

When a *CLC Workbench* is connected to a cloud-enabled *CLC Genomics Server* and no configuration details have been entered in the *CLC Workbench* itself, then the *CLC Workbench* will automatically obtain the AWS and GCE settings from the *CLC Genomics Server*.

Finding jobs and results submitted via the CLC Server using Cloud Job Search

To use the Cloud Job Search tool to find jobs and results of jobs submitted to *CLC Genomics Cloud Engine* via a *CLC Genomics Server*, the AWS and GCE settings specified in the *CLC Workbench* cloud configuration dialog should either be empty, in which case the details of the configuration settings for the Cloud Server Plugin will be used, or must match those specified in the Cloud Server Plugin configuration in the *CLC Genomics Server*.

The "Only from logged-in user" option in the Cloud Job Search tool will find jobs submitted by both the user authenticated through the cloud configuration dialog and the user logged into the *CLC Genomics Server*. This is most noticeable if these user names differ.

8.1 Configuring the Cloud Server Plugin

Configuration of the Cloud Server Plugin is done using the web administrative interface of the *CLC Genomics Server*. Information about connections to AWS and *CLC Genomics Cloud Engine* are entered via plugin settings, available to configure after the plugin is installed and the *CLC Genomics Server* has been restarted.

Information about the hardware requirements of jobs sent to *CLC Genomics Cloud Engine* and how results from those jobs should be handled is configured using cloud presets. Multiple presets can be configured. Users select the preset to use when they are launching a job.

The *CLC Genomics Server* uses the OAuth2 Client Credentials Grant flow for authentication. This flow, suitable for server-to-server integration, ensures that the *CLC Genomics Server* has valid credentials that do not need to be periodically renewed. *CLC Workbench* users who connect to the *CLC Genomics Server* are granted access to the *CLC Genomics Cloud Engine* via the server's credentials, but jobs are run under the user's username.

Please note the following security-related information:

- By configuring the AWS settings in *CLC Genomics Server*, *CLC Workbench* users connecting to the server and using the Cloud Plugin will be able to see each other's data in AWS S3, as they will have access through the shared credentials stored in the server.
- There are no user permissions on jobs in *GCE*. This means that *GCE* users will be able to find each other's jobs, for example by using the Cloud Job Search functionality in their Workbench.
- It is possible to prevent a *CLC Workbench* user from being able to submit jobs to the cloud by configuring permissions for each cloud preset. (**Global permissions** | **Cloud presets**)

Information about installing server plugins can be found at https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Server_plugins.html.

8.1.1 Configuring the AWS and CLC Genomics Cloud Engine settings

To configure the Cloud Server Plugin, navigate to the **Extensions** tab and click on **Manage Plugins**. Confirm that the Cloud Server Plugin is shown in the list. Then, click on the **Manage Plugin settings...** button at the bottom of the list (figure 8.1), followed by the **Edit** button next to the Cloud Server Plugin.

This will allow you to configure the following parameters:

AWS access key ID: The access key ID for programmatic access, set up for the AWS IAM user as described in section 1.1.

AWS secret access key: The secret access key ID for programmatic access, set up for the AWS IAM user as described in section 1.1.

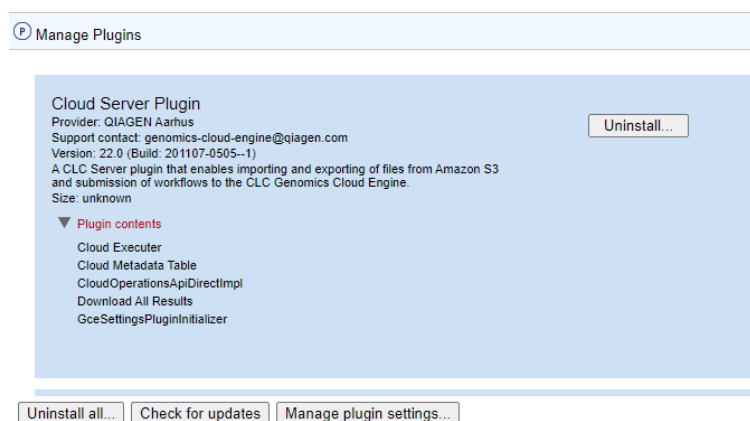


Figure 8.1: Configure the Cloud Server Plugin by clicking the Edit Plugin settings button.

AWS region: The default AWS region to connect to. We recommend choosing the region closest to the server's physical location.

S3 URL for exporting external applications: The full path to the Amazon S3 location where the external applications configuration file should be published when exporting external applications to the cloud. E.g. "s3://my-bucket/EAC/External-apps-config.xml".

GCE S3 cloud cache bucket name: The cache bucket to be used when uploading data to the cloud, as described in section 1.1.

GCE job manager rest host URI: The URL pointing to the GCE instance deployed on the specified AWS account, as described in section 1.1. Note that this URL must be the same as the "URL" setting specified in the *CLC Workbench*, under **File | Cloud Connection**.

GCE http oauth2 client id: The client ID used by the to authenticate using OAuth2 Client Credentials Grant.

GCE http oauth2 client secret: The client secret used by the *CLC Genomics Server* to authenticate using OAuth2 Client Credentials Grant.

GCE http oauth2 authorization server: The access token endpoint of the authentication server used by the *CLC Genomics Cloud Engine*.

Accept untrusted certificate: This checkbox should be selected when the GCE has been set up with a self-signed certificate. We recommend setting up the GCE with a trusted certificate.

Validate settings: This option, enabled by default, validates the AWS and GCE settings when you press OK. Uncheck this box only if you need to temporarily store invalid settings.

GLOBAL_OVERRIDABLE: This setting has no effect for the Cloud Server Plugin and we recommend leaving it at the default settings.

8.2 Configuring cloud presets

To create or edit existing cloud presets, log into the *CLC Genomics Server* web administrative interface and go to:

Job distribution | Execution node settings | Cloud settings

Click on **Add New Cloud Preset** and fill in a preset name, job tag (optional), hardware requirements and result handling strategy. The hardware requirements can be pre-filled based on available presets configured by the administrator of the *CLC Genomics Cloud Engine*.

Permissions for each cloud preset can be configured under the **Global permissions** tab.

Chapter 9

External applications in the cloud

Third party applications can be integrated into the CLC environment by configuring them as *external applications*. Containerized external applications can be used in workflows to be executed on the *CLC Genomics Cloud Engine* if the container is in the Amazon Elastic Container Repository (ECR) on the same Amazon account where the CLC Genomics Cloud Engine is deployed.

Information about configuring external applications, and exporting the configurations for use by the *CLC Genomics Cloud Engine* can be found in the **External applications** chapter of the *CLC Genomics Server* manual at https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=External_applications.html

Creating and configuring workflows with external applications

Creating and editing workflows is done in the *CLC Workbench*, as described at <https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Workflows.html>.

Workflow elements will be available for:

- External applications available on a *CLC Genomics Server*, if the Workbench is connected to one.
- External applications described in the configuration file located on S3, where that location has been entered into the Workbench Preferences, under the Advanced tab, and the relevant AWS credentials have been configured, so you have access to the S3 bucket (see section 3).

External applications configurations are exported to S3 from a *CLC Genomics Server* by an administrative user, as described at https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Import_export_external_application_configurations.html

Please ask your *CLC Genomics Server* administrator for the s3 URL to configure in the *CLC Workbench* Preferences. This is usually of particular relevance if you wish to submit workflows containing external applications to the *CLC Genomics Cloud Engine* without connecting to a *CLC Genomics Server*.

Submitting workflows with external applications to run on GCE

Workflows can be submitted for analysis on the *CLC Genomics Cloud Engine* by a *CLC Workbench*, with or without a connection to a *CLC Genomics Server*, or using the *CLC Server Command Line Tools*.

Submitting via a CLC Workbench without a connection to a CLC Server The location of the exported external application configuration file in S3 must be configured in the Workbench Preferences. All the external applications included in the workflow must be described in the configuration file in S3.

Submitting via a CLC Workbench with a connection to a CLC Server The location of the exported external application configuration file in S3 is only needed if there are external applications in the workflow that are not present on the server, (but which are described in the configuration file on S3).

Submitting via a CLC Server using the CLC Server Command Line Tools or a CLC Workbench The external applications included in the workflow must be configured on the CLC Server.

Chapter 10

Troubleshooting

If a workflow execution fails while the *CLC Workbench* is still connected to the *CLC Genomics Cloud Engine*, the workflow process will be shown in red in the Processes area, with the message: "Errors occurred: see log". An error message will also be displayed, which may contain information about the cause of failure.

If further information is required about a failure *after* the workflow submission is completed, we recommend finding the batch using the Cloud Job Search functionality (see chapter 7), and downloading all results for the entire batch. This will download any available logs related to the failure. The exact files that will be downloaded may vary, depending on the cause of the failure. An example is shown in figure 10.1.

The type of files that may be downloaded include:

- **Workflow log** CLC data elements that can be opened within a *CLC Workbench*. These correspond to the logs produced when running a workflow within a *CLC Workbench* or on a *CLC Genomics Server*.
- **Result.json** Plain text files containing information about any outputs produced in AWS S3 for a particular job, including the exact paths to the outputs.
- **gce.log** Plain text files containing information registered by the *CLC Genomics Cloud Engine* about the workflow execution process, and technical details about any errors that have occurred.

We recommend that the Workflow log is inspected first, as it may contain useful information about simple causes of failures, such as corrupt input data or errors in the workflow configuration. For more complex cases, please email our Support team, attaching the files listed above (ts-bioinformatics@qiagen.com).

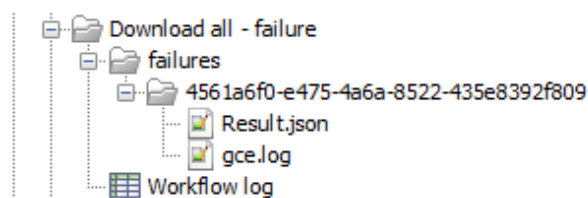


Figure 10.1: The "Download All Results" functionality in the Cloud Job Search tool allows you to download logs and other technical information if a workflow execution failed. The logs are downloaded to the Navigation Area. Some of them will be plain text files, which can be opened using any external text editing tool.

Chapter 11



Installing and uninstalling Workbench plugins

The following sections describe the installation and removal of plugins, including Cloud Plugin, on a *CLC Workbench*. For information about installing plugins on a *CLC Genomics Server*, please refer to https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Server_plugins.html and the information in this manual about configuring the Cloud Server Plugin, provided in chapter 8.1.1.

11.1 Installation of Workbench plugins

Note: In order to install plugins and modules, the *CLC Workbench* must be run in administrator mode. On Windows, you can do this by right-clicking the program shortcut and choosing "Run as Administrator". On Linux and Mac, it means you must launch the program such that it is run by an administrative user.

Plugins and modules are installed and uninstalled using the Workbench Plugin Manager. To open the Plugin Manager go to:

Help in the Menu Bar | Plugins... () or click on **Plugins** () **in the Toolbar**

The Plugin Manager has two tabs at the top:

- **Manage Plugins.** This is an overview of plugins and modules that are installed.
- **Download Plugins.** This is an overview of the plugins and modules available to download and install.

To install a plugin, click the **Download Plugins** tab. This will display an overview of the plugins that are available for download and installation (see figure 11.1).

Select an item in the list to display additional information about it on the right side of the dialog. Click on the **Download and Install** button to install the selected plugin or module.

Accepting the license agreement

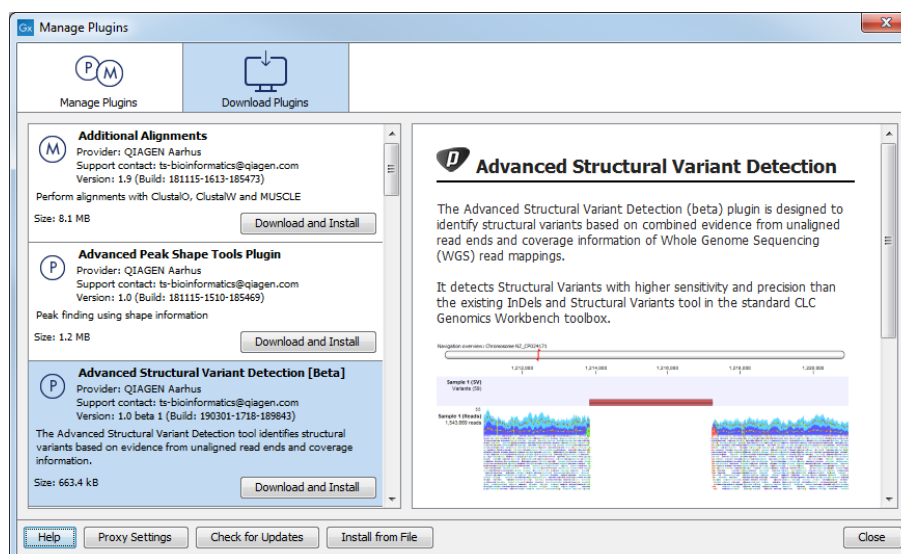


Figure 11.1: The plugins and modules that are available for download.

The End User License Agreement (EULA) must be read and accepted as part of the installation process. Please read the EULA text carefully, and if you agree to it, check the box next to the text **I accept these terms**. If further information is requested from you, please fill this in before clicking on the **Finish** button.

Installing a cpa file

If you have a .cpa installer file for Cloud Plugin downloaded to your computer, you can install it by clicking the **Install from File** button at the bottom of the dialog and specifying the .cpa file.



If you are working on a system not connected to the internet, plugin and module .cpa files can be downloaded from <https://digitalinsights.qiagen.com/products-overview/plugins/> using a networked machine, and then transferred to the non-networked machine for installation.

Restart to complete the installation

When you close Plugin Manager, a dialog will appear offering the opportunity to restart the *CLC Workbench*. Newly installed plugins and modules will be available for use after restarting.

11.2 Uninstalling Workbench plugins

Plugins and modules are uninstalled using the Workbench Plugin Manager. To open the Plugin Manager, go to:

Help in the Menu Bar | Plugins... () or click on **Plugins** () **in the Toolbar**

This will open the dialog shown in figure 11.2.

Installed plugins and modules are shown in this dialog. To uninstall, click on the item in the list to uninstall and click on the **Uninstall** button.

When you close the dialog, you will be asked whether you wish to restart the Workbench. Plugins

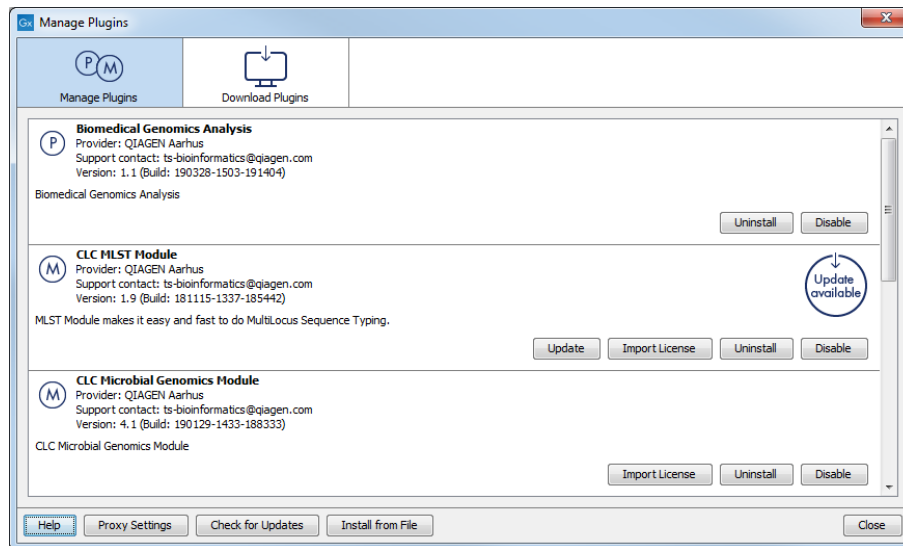


Figure 11.2: The Plugin Manager with plugins and modules installed.

and modules are not uninstalled until the Workbench is restarted.

If you do not wish to completely uninstall a plugin or module, but do not want it to be loaded the next time you start the Workbench, click the **Disable** button.