Genomics Analysis Portal

USER MANUAL

Manual for
Genomics Analysis Portal for CLC Genomics Server
Windows, macOS and Linux

August 16, 2021

This software is for research purposes only.

Contents

Chapter 1

Introduction to the Genomics Analysis Portal

This manual is for the *Genomics Analysis Portal* distributed with *CLC Genomics Server 21.0.5*.

The *Genomics Analysis Portal* is a browser-based, graphical client for the *CLC Genomics Server*, offering a controlled environment for processing samples using workflows. Using the *Genomics Analysis Portal*, workflows can be launched using fastq format files as input, their progress can be monitored, and results can be accessed. An overview of previous analyses is also provided.

The workflows accessible using *Genomics Analysis Portal* are made available by the *CLC Genomics Server* administrator. Options in these workflows are locked; no settings can be adjusted when the workflow is launched.

Configuring workflows and opening analysis results is done using a *CLC Genomics Workbench* compatible with the version of the *CLC Genomics Server* being used. As a rule of thumb, if the *CLC Genomics Workbench* has the same version number as the *CLC Genomics Server*, it is compatible. Full compatibility information is provided with the *CLC Genomics Server* latest improvements at https://digitalinsights.qiagen.com/products/qiagen-clc-genomics-server/latest-improvements/current-line/.

The latest version of the user manual can also be found in pdf format at https://digitalinsights.qiagen.com/technical-support/manuals/.

Accessing the Genomics Analysis Portal

To use the *Genomics Analysis Portal*, your server administrator must enable the functionality in the *CLC Genomics Server* and you must be a registered user of the *CLC Genomics Server*. Your server administrator will be able to provide you with the relevant URL for the *Genomics Analysis Portal* you have access to.

The *CLC Genomics Server* admin manual is available at:

https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Introduction.html.

1.1 Contact information

Genomics Analysis Portal is developed by:

QIAGEN Aarhus
Silkeborgvej 2
Prismet
8000 Aarhus C
Denmark

https://digitalinsights.qiagen.com/

Email: ts-bioinformatics@qiagen.com

The QIAGEN Aarhus team continuously improves products with your interests in mind. We welcome feedback and suggestions for new features or improvements. How to contact us is described at: http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Contact_information_citation.html.

You can also make use of our online documentation resources, including:

- Core product manuals https://digitalinsights.qiagen.com/technical-support/manuals/

- Plugin manuals https://digitalinsights.qiagen.com/products-overview/plugins/

- Tutorials https://digitalinsights.qiagen.com/support/tutorials/

- Frequently Asked Questions http://helpdesk.clcbio.com/index.php?pg=kb

How to cite us

To cite a CLC Workbench or Server product, use the name of the product, the version number and add (QIAGEN) to it. For example CLC Server 21.0 (QIAGEN) or CLC Genomics Workbench 20.0 (QIAGEN). If a location is required by the publisher of the publication, use (QIAGEN, Aarhus, Denmark).

In the References, cite our web site https://digitalinsights.qiagen.com/ according to the preferences of the particular journal you will publish in.

Chapter 2

Genomics Analysis Portal interface

The *Genomics Analysis Portal* provides a graphical interface for launching workflows and viewing information about analyses run, including progress and the details about individual analyses (figure 2.1). In addition, if BaseSpace has been configured, data stored in the Illumina$^{\circledR}$ BaseSpace environment can be accessed for analysis.

Note: The user interface is optimized for a minimum screen width of 1024 pixels.

Figure 2.1: *The **Genomics Analysis Portal** offers an overview of analyses that have been run, and the progress and details of individual analyses.*

Key areas of the *Genomics Analysis Portal* interface are listed below, along with links to more detailed information.

- **Analyses table** Analyses run via the *Genomics Analysis Portal* are listed in the **Analyses** table (section 2.3).

- **Sample analysis details panel** Analysis details, links, and actions for the selected analysis are presented in this panel (section 2.4).

- **Start Analysis** Use this button to start an analysis using input data stored locally or on a network drive (section 2.1).

- **Access BaseSpace** Use this button if the BaseSpace connection has been configured, and you wish to use data located in the Illumina$^{\circledR}$ BaseSpace environment as input to an analysis. (section 2.2).

- **Help, Info and Logout** Icons at the bottom of the interface can be clicked upon to:

  - Access the manual (⑦)
  - Access copyright information (ⓘ)
  - Logout of the *Genomics Analysis Portal* (⤶)

2.1   Start an analysis using data stored locally

To start an analysis from the Genomics Analysis Portal using input data stored locally or on a network drive:

1. Click on the **Start Analysis** button at the top of the **Analyses** table.

2. Browse the file location and select the data to be analyzed. You can select either individual fastq files or folders containing fastq files.

3. In the next step, select the workflow to launch from the drop-down menu (figure 2.2).

4. Specify a unique sample ID for each of the selected data items, or use the one suggested.

5. Include a comment, if desired. Comments are visible in the **Sample analysis details** panel.

6. Click on the **Start** button when you are ready to submit the workflow.

Figure 2.2: *Select a workflow from the drop down list. Clicking the Start button submits the workflow to run using the specified data inputs.*

2.2   Start an analysis using data stored in BaseSpace

Data located in the Illumina$^{\circledR}$ BaseSpace environment can be used as input to workflows launched using the *Genomics Analysis Portal*. The BaseSpace connection must first be configured, as described in section 4.1.

To start an analysis using data in the Illumina$^{\circledR}$ BaseSpace environment:

1. Click on the **Access BaseSpace** button at the top of the **Analyses** table. This redirects you to BaseSpace.

2. Log in to the relevant BaseSpace account. The first time you log in, you will be asked to grant the *Genomics Analysis Portal* permission to access your BaseSpace account (basic information and names of projects, samples, etc.) Click on **Accept**.

3. Select the BaseSpace projects to access and click on **Next**. The first time you access a particular project, you will be asked to grant the *Genomics Analysis Portal* permission to read data in that project (more specifically to view and download the content of the project). Click on **Accept**.

4. Select the samples to be analyzed.

5. In the next step, select the analysis workflow from the drop-down menu.

6. Specify a unique sample ID for each of the selected data items, or use the one suggested.

7. If you want, you can include a comment. This will be visible from the **Sample analysis details** panel.

8. Click on the **Start** button when you are ready to submit the workflow.

2.3 Analyses table

Analyses submitted from the *Genomics Analysis Portal* are listed in the **Analyses** table, which makes up the main part of the view area. Information in this table includes the sample ID, the analysis workflow used, the status of the analysis, and the time the analysis started.

Sorting the Analyses table

By default, the **Analyses** table is sorted according to the **Analysis start** time, with the item with the most recent time stamp at the top of the table.

You can sort according to other column values by clicking on the relevant column header. If you sort based on "Status", the sorting takes quality control status into account, placing analyses with quality control deviations at the top.

Click on a column heading twice to reverse the order. The table can be sorted based on values in multiple columns by holding down the Shift key while clicking on the relevant column headers.

Filtering for particular analyses in the Analyses table

The **Analyses** table can be filtered using the fields just below the table headers. The number of analyses listed after filtering is reported in the top, left side of the table. Clicking on the **Clear all filters** button at the top, right side of the table removes all the applied filters, so the full list of analyses is once again visible.

A particular analysis can be selected by clicking a table row. The **Sample analysis details** panel on the right will then show details about that analysis. Depending on the status of the analysis, different actions and information will be available. See section 2.4 for further details.

Clicking the **Enable multi-select** button above the table introduces a multi-select column on the left hand-side of the **Analyses** table. When two or more analyses are selected, the **Sample analysis details** panel displays a list of the selected analyses and the available actions. If you select analyses with different states (figure 2.3), only actions common to the selected states will be available (figure 2.4).

Figure 2.3: *Analyses can be listed according to their state.*

Figure 2.4: *By enabling multi-select, you can select and apply the same action to multiple analyses at the same time.*

Status of sample analyses

The **Status** column contains information about the analysis state, as described below. A quality control indicator is displayed for analysis workflows with quality control criteria configured (section 3.1), supporting the quick identification of analyses that pass or fail quality control checks.

In Progress For samples still being analyzed, the status field displays a progress bar (figure 2.5). The progress of queued samples remains at 1% until the analysis starts. When multiple samples are in the "In Progress" state, they are run sequentially, according to their order in the analysis queue.

You can stop an analysis by selecting the sample and clicking on **Stop**. This will remove the analysis from the list and discard any intermediate results.

Figure 2.5: *Two analyses are in the "In Progress" state, at 4% and 6% complete.*

Ready for Review Analyses are set to Ready for Review when the analysis results are available. In this state results can be opened or downloaded via links in the **Sample analysis details** panel.

Users included in the *CLC Genomics Server* admin group will also see a **Delete** button. Clicking on this removes the analysis from the list and deletes results from the *CLC Genomics Server*.

Completed Clicking on the **Complete** button moves an analysis into the Completed state. This can be used as an indication that results have been reviewed.

If archiving has been set up (section 4.3), completed analyses will be archived according to the archiving configuration. Archived analyses are not listed in the **Analyses** table.

Users included in the *CLC Genomics Server* admin group are able to **Delete** completed analyses. This removes analyses from the list and deletes results from the *CLC Genomics Server*.

Failed An analysis failure is most often caused by technical issues, but can also be due to invalid or corrupt fastq files. It is not related to the quality of the sequencing data or the analysis results. To learn more about the cause of failure, select the sample analysis in the list and inspect the failure message displayed in the **Sample analysis details** panel.

To restart a failed sample, that is, to run the analysis again using the same input data and workflow, click on the **Rerun** button available at the bottom of the **Sample analysis details** panel.

2.4   Sample analysis details panel

When an analysis is selected in the **Analyses** table, information about it is displayed in the **Sample analysis details** panel to the right. This includes basic information such as the sample ID, the name of the analysis workflow, and log entries (figure 2.6), as well as more detailed information, described below.

Quality control

The **Quality control** section reports on the criteria and thresholds configured for the analysis workflow. A green icon (✅) signals that all QC criteria passed. A yellow icon (⚠️) flags a potential quality control issue. The text underneath lists the deviating criteria. The icon shown aligns with that presented in the **Analyses** table Status column.

Open workflow output

Under the **Open workflow output** heading are links to workflow outputs. Clicking on the links to these CLC format elements opens them in a *CLC Genomics Workbench*.

To open CLC format files by clicking on links, the following must be in place:

- The *CLC Genomics Workbench* is installed on the system you are working on, and it is a version compatible with the *CLC Genomics Server* that the *Genomics Analysis Portal* is running on. Information about compatible software versions is provided on the latest

Figure 2.6: *The Sample analysis details panel lists information relating to a selected analysis.*

improvements information for the *CLC Genomics Server*: `https://digitalinsights.qiagen.com/products/qiagen-clc-genomics-server/latest-improvements/current-line/`.

- The *CLC Genomics Workbench* is set as the default program for opening .clc files.

When you click a workflow output link, a dialog opens asking you to confirm you wish to open this file with the associated *CLC Genomics Workbench* (figure 2.7).
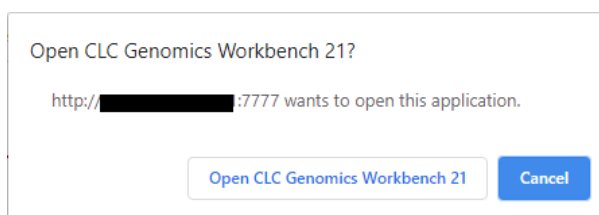
Figure 2.7: *When you click on a workflow output link, you are asked to confirm that you wish to open the file using the relevant program for opening .clc files.*

If not already running, the *CLC Genomics Workbench* will open and automatically log into the *CLC Genomics Server* sitting underneath the *Genomics Analysis Portal*.

Download workflow exports

If your workflow contains export elements, the **Download workflow exports** section will list the exported files. These are typically formats other than CLC, for example VCF, PDF, etc. If you click on an export link, the file will be saved to the location of your choice.

Chapter 3

Workflows for the *Genomics Analysis Portal*

For a workflow to become available in *Genomics Analysis Portal* the following must be in place:

- The workflow must be created and appropriately configured using a compatible *CLC Genomics Workbench*.

- The workflow must be installed on the *CLC Genomics Server*.

- The workflow permissions setting must specify a group to which the *Genomics Analysis Portal* users belong.

3.1 Workflow creation and configuration

Creating a workflow

To create a workflow for *Genomics Analysis Portal* on *Genomics Analysis Portal for CLC Genomics Server*, you must be using a licensed version of a corresponding version of the *CLC Genomics Workbench*. To use functionality provided by plugins, those plugins must be installed on both the *CLC Genomics Workbench* and *CLC Genomics Server*. For commercial plugins, licenses must be present on both systems.

The information in this section focuses on special requirements when creating workflows to be launched using the *Genomics Analysis Portal*.

Information about creating workflows is provided in the *CLC Genomics Workbench* manual at `https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Creating_workflow.html`. How to modify existing workflows using the **Open Copy of Workflow** function is also described there.

Note: When users start an analysis from the *Genomics Analysis Portal*, all workflow parameters will be locked regardless of whether they are locked in the original workflow design.

Configuring quality control criteria

If you wish to use the *Genomics Analysis Portal* quality control functionality for your workflow, insert the workflow element **Create Sample Report** in the workflow design and configure the quality control report section. Instructions for this are given in the *CLC Genomics Workbench* manual at https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Create_Sample_Report.html.

Link an output element to the "Report" output channel of the **Create Sample Report** element.

Tip: Even if you do not wish to use quality control reporting, **Create Sample Report** is generally useful for creating one combined sample report that can be exported as a PDF format file.

Once all workflow elements are in place, the workflow input elements need to be configured.

Configuring workflow input elements

Configuration of workflow input elements is described in the *CLC Genomics Workbench* manual at https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Configuring_input_output_elements.html.

It is not necessary to lock any parameters. When a workflow is started using the *Genomics Analysis Portal*, all parameters will be locked regardless.

For workflows to be launched using the *Genomics Analysis Portal*, take the following steps for workflow input elements:

- **For the primary input of sample data**
    - Enable import of raw data using on-the-fly import only.
    - Restrict the available importers to the one importer relevant for your data, e.g. Illumina.
    - Configure the selected importer by clicking on the **Configure Parameters** button to open the "Configure parameters" dialog (figure 3.1).

  Only one Input element for importing data is expected in workflows to be submitted via the *Genomics Analysis Portal*. This Input element must be first in the ordered list, i.e. [1] appears in the top left corner of the element in the workflow design. If another number is present there, please re-order your workflow input elements, as described at https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Ordering_inputs.html.

- **For parameter inputs, e.g. configuring a reference data input**
    - For each workflow input element connected to a parameter input channel, select the relevant data element, e.g. a reference sequence.
      Note: It is not possible to use workflow roles when configuring parameter inputs in workflows to be launched using *Genomics Analysis Portal*, as it is not possible to select a Reference Data Set when starting the analysis.

Workflow output and export elements

Connect an *output element* to each output channel generating results you want to have available as a workflow output that can be opened in a *CLC Genomics Workbench*.
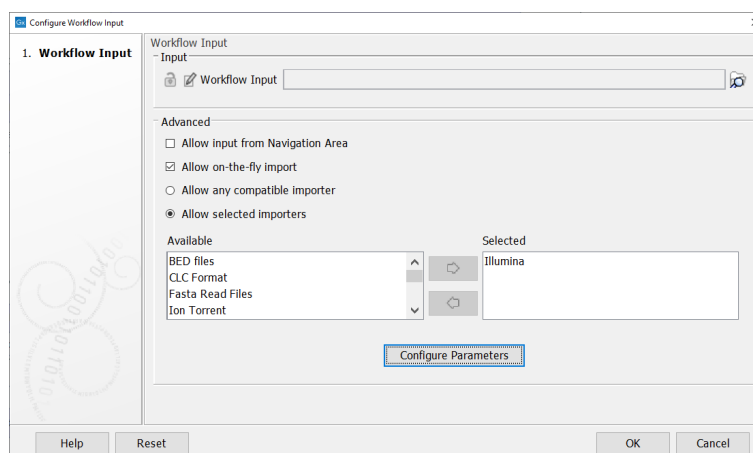
Figure 3.1: *Configure the importer to be used for on-the-fly import.*

Connect an *export element* to each output channel generating results you want to have available for download and use with other tools, for example VCF or BED format files. For export elements, leave the "Export destination" **empty**. The *Genomics Analysis Portal* export location is set in the *CLC Genomics Server* administration interface rather than in the workflow export elements.

Checking your workflow configuration

To quickly check all the required inputs have been configured, step through the workflow wizard. If you are able to proceed all the way to the last wizard step without having to select input other than the initial fastq and a (dummy) destination for export elements, your workflow can be installed on the *CLC Server*.

To step through the workflow wizard:

1. Click on the **Run** button at the bottom of the Workflow Editor.

2. Choose to run the job on the *CLC Server* (or Grid if relevant for your setup). Click on **Next**.

3. Browse to select a random fastq file on your system as the primary workflow input. You will not actually be analyzing the fastq file, you just need to select one to proceed in the wizard. Click on **Next**.

4. Go through the remaining steps to check that input has been selected for all required input fields.

   Note: If your workflow includes export elements, you will need to select an export destination for the purpose of proceeding through the wizard steps. When run on *Genomics Analysis Portal* the workflow will use the export destination that was configured by the *CLC Server* administrator.

5. When you get to the last wizard step, click on **Cancel**.

3.2   Workflow Installation

Once your have created and configured your workflow, you can install it on the *CLC Genomics Server*. For this, you need to be member of the *CLC Server* admin group.

Workflows can be installed on the *CLC Server* from a *CLC Genomics Workbench* or through the web administrative interface of the *CLC Genomics Server*.

After installing a workflow on the *CLC Server*, you also need to set permissions to make it available through the *Genomics Analysis Portal*, as described in section 3.3.

Installing a workflow on the Server from the Workbench

Log into the *CLC Genomics Server* from the *CLC Genomics Workbench* as an administrative user and use the Workflow Manager to install the workflow.

Logging into the *CLC Genomics Server* is described at `https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=CLC_Server_connection.html`.

Installing workflows using the Workflow Manager is described at `https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Installing_workflow.html`.

Installing a workflow on the Server using the web administrative interface

Workflows are installed using a .cpw file created using a *CLC Genomics Workbench*. To install the .cpw file, log into the *CLC Genomics Server* web administrative interface as an administrative user and then use functionality under
Admin | Workflows.

Creating workflow installer (.cpw) files is described at `https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Installing_workflow.html`.

Installing workflows on the *CLC Genomics Server* using the web administrative interface is described at `https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Installing_configuring_workflows.html`.

3.3   Workflow permissions settings

For a workflow to become available from within the *Genomics Analysis Portal*, you need to authorize it for use by *Genomics Analysis Portal* users.

Workflow permissions are configured in the *CLC Genomics Server* web administrative interface by an administrator.

1. Log into to the web administrative interface, as described in the *CLC Server* administrator manual at `https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Logging_into_administrative_interface.html`.

2. Go to **Admin ( )** | **Global permissions | Workflows**

3. Locate your newly installed workflow and click on **Edit permissions**.

4. Choose **Only authorized users from selected groups** and select your *Genomics Analysis Portal* user group(s). (Figure 3.2)

You can give access to different workflows to different groups, allowing you to tailor which workflows available via *Genomics Analysis Portal* for particular users.

Figure 3.2: **Genomics Analysis Portal** *workflows must be authorized for one or more groups, including the one holding your* **Genomics Analysis Portal** *users.*

Note that all *CLC Server* users are able to log into the *Genomics Analysis Portal*, but in the **Start analysis** workflow drop-down, only the workflows they have been authorized to use will be visible.

Chapter 4

Administration

The *Genomics Analysis Portal* administration page (Figure 4.1) is only accessible to the root user. It contains two tiles. The **About** tile lists the version of the application. The **Configuration** tile provides access to the *Genomics Analysis Portal*-specific configuration items listed below.
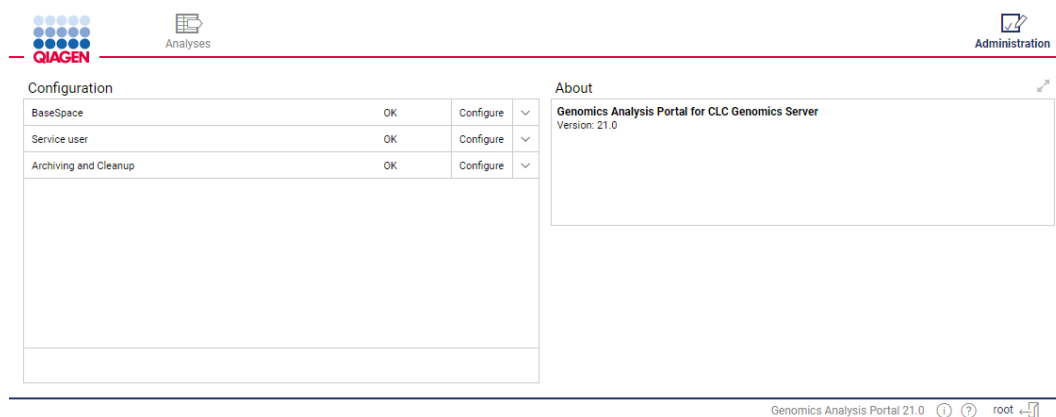
Figure 4.1: *The Administration page is available to the root user only.*

4.1 BaseSpace configuration

A requirement for integration with BaseSpace is that your *CLC Server* is https enabled. Details about how to do this can be found in the administrator manual at `https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=SSL_encryption.html`

Some organizations may prefer to use a reverse proxy instead. The following section provides information relating to this.

HTTPS using Apache 2.4.x as a reverse proxy

Here we provide an example of a typical Apache 2.4.x configuration. The example assumes the *CLC Genomics Server* is installed on a machine with the IP address 10.108.241.134. The following would be added to the Apache configuration file on that system. The keys SSLCertificateFile and SSLCertificateKeyFile are paired with the locations of the relevant certificate files.

17

```
<VirtualHost *:443>
        SSLEngine on
        SSLCertificateFile /etc/ssl/certs/clcserver01.crt
        SSLCertificateKeyFile /etc/ssl/certs/clcserver01.key

        ProxyPass "/" "http://10.108.241.134:8877/"
        ProxyPassReverse "/" "http://10.108.241.134:8877/"
</VirtualHost>
```

Using the *Genomics Analysis Portal* built-in support for HTTPS

An SSL certificate is required before HTTPS can be enabled on the *Genomics Analysis Portal*. This is usually obtained from a Certificate Authority like Thawte or Verisign (see http://en. wikipedia.org/wiki/Certificate_authorities).

A signed certificate in a pkcs12 keystore file is also needed. The keystore file is either provided by the Certificate Authority or it can be generated from the private key used to request the certificate and the signed-certificate file from the Certificate Authority (see Creating a PKCS12 keystore file).

1. Copy the keystore file to the conf subfolder in the *CLC Genomics Server* installation directory.

2. Next, the server.xml file in the conf subfolder in the *CLC Genomics Server* installation directory must be edited to enable SSL-connections. Add text like the following text to the server.xml file:

```
<Connector port="8443" protocol="HTTP/1.1" SSLEnabled="true"
            maxThreads="150" scheme="https" secure="true"
            clientAuth="false" sslProtocol="TLS"
            keystoreFile="conf/keystore.pkcs12" keystorePass="tomcat"
            keystoreType="PKCS12"
/>
```

Replace keystore.pkcs12 with the name of your keystore file, and replace tomcat with the password for your keystore. The above settings will make SSL available on port 8443. The standard (non-SSL) port would be 8877, or whichever port number you have configured it to.

Self-signed certificates can be generated if only connection encryption is needed. See http://www.akadia.com/services/ssh_test_certificate.html for further details.

Creating a PKCS12 keystore file

If the certificate is not supplied in a pkcs12 keystore file, it can be put into one by combining the private key and the signed certificate obtained from the Certificate Authority by using openssl:

```
openssl pkcs12 -export -out keystore.pkcs12 -inkey private.key -in certificate.crt -name "tomcat"
```

This will take the private key from the file private.key and the signed certificate from certificate.crt and generate a pkcs12-store in the keystore.pkcs12 file.

Create a new BaseSpace application

1. Login to BaseSpace, https://developer.basespace.illumina.com/.

2. Select My Apps.

3. Create a new application and fill in all the necessary details.

   - For App Type, make sure you select "Web/Desktop Visualization"
   - The redirect URI should point to your *Genomics Analysis Portal* installation with /basespace appended (for example https://clcserver.yourdomain.com:8443/gap/basespace).

4. Find your Client Id and Client Secret under Credentials. Copy these.

Configure BaseSpace connection from within *Genomics Analysis Portal*

1. Log into *Genomics Analysis Portal* as root.

2. Go to the Administration page, locate BaseSpace in the **Configuration** tile, and click on **Configure**.

3. Fill in the Client Id and Client Secret from above and set the redirect URI to point to your *Genomics Analysis Portal* installation (for example https://clcserver.yourdomain.com:8443/gap). Click on **Save**.

4. Restart the *CLC Genomics Server*. If jobs are running, you may wish to put the server in Maintenance Mode first, as described in the administration manual at https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Server_maintenance.html.

When using the secure *Genomics Analysis Portal* link, *Genomics Analysis Portal* users will now see a button providing access to BaseSpace.

Note that when samples are selected for analysis in the *Genomics Analysis Portal*, the necessary files are downloaded from BaseSpace into the *CLC Genomics Server* temporary folder, and are deleted when they are no longer needed, i.e. when the *Genomics Analysis Portal* analysis is done. Due to the amount of space that can be taken up by these temporary files, it can be useful to specify an alternative, larger, disk area where temporary files created by the *CLC Genomics Server* can be written. Instructions to do this are provided in the server administration manual at http://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Changing_tmp_directory.html.

4.2 Service user

The service user is used for running all the analyses launched using *Genomics Analysis Portal* on the *CLC Genomics Server*.

The user name and password for the service user can be changed in two places:

- From the **Service user** item the **Configuration** tile, visible when you are logged into *Genomics Analysis Portal* as root.

- From the *Genomics Analysis Portal* tab of the *CLC Genomics Server* web administrative interface, under the Extensions tab.

4.3 Archiving and Cleanup

The *Genomics Analysis Portal* can be configured to automatically archive and clean up analysis result files (figure 4.2). When an analysis has been in the Completed state for a defined number of days, any workflow-exported external files, e.g. PDF and VCF files, are saved as a ZIP archive to a preselected storage location. Once done, all remaining data related to that analysis is deleted. Note that the input FASTQ files are not touched by *Genomics Analysis Portal* and are therefore not affected by this functionality.
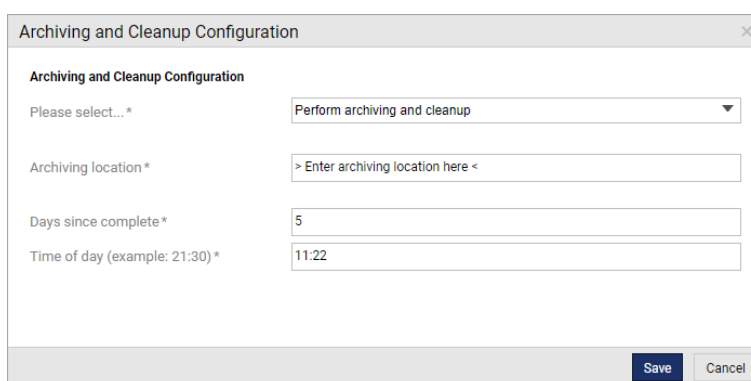

Figure 4.2: *Editing the archiving and cleanup configuration.*

The configuration of this feature is accessible from the **Archiving and Cleanup** item in the **Configuration** tile:

- Service action. Choose between "Perform archiving and cleanup", "Perform cleanup without archiving", and "Disable archiving and cleanup".

- Archiving location. Enter the file system path to the folder where the archive packages should be stored.

- Days since complete. Number of days between a sample is put in a final state (Completed or Failed) and the archiving.

- Time of day. Time of day for cleanup and archiving.

For this functionality to work as described, the administrator must ensure the archiving configuration is correct, and that enough storage space is available. If archiving fails, the data will not be deleted, and the space taken by the outputs generated by completed analyses will keep growing.

4.4 *CLC Genomics Server* **administration for the** *Genomics Analysis Portal*

There are *CLC Genomics Server* administrative items to be aware of in the context of the *Genomics Analysis Portal*.

Log into the *CLC Genomics Server* administrative interface as an administrative user to access these.

Information on how to log into to the administrative interface is found in the *CLC Server* administrator manual at `https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Logging_into_administrative_interface.html`.

The sections on the *CLC Genomics Server* administrative interface that are of particular relevance for the *Genomics Analysis Portal* are:

- The *Genomics Analysis Portal* section on the **Extensions (** ⚎ **)** tab. This is where you enable the *Genomics Analysis Portal* and also where you find the url that users will need for accessing *Genomics Analysis Portal*. See more in the *CLC Server* administrator manual at `https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Genomics_Analysis_Portal.html`.

- The **Users and Groups** section on the **Admin (** ⚙ **)** tab. You will need to create users and include them into a user group. The user group is needed for setting permissions on and thereby *Genomics Analysis Portal* authorizing users for *Genomics Analysis Portal* workflows. Note that all *CLC Server* users will be able to log into the *Genomics Analysis Portal*, but in the **Start analysis** workflow drop-down they will only be able to see the workflows for which they have been authorized.

- The **Global permissions** section on the **Admin (** ⚙ **)** tab. You must authorize your *Genomics Analysis Portal* user group for each of the workflows that should be available from within the *Genomics Analysis Portal*.

If your *CLC Server* is configured to use a non-default listening port, please also refer to the information at `https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Changing_listening_port.html`.