QIAGEN

Tutorial

How to Configure a Workflow to Calculate an HRD Score

September 22, 2022

Sample to Insight

QIAGEN Aarhus · Silkeborgvej 2 · Prismet · 8000 Aarhus C · Denmark
digitalinsights.qiagen.com · ts-bioinformatics@qiagen.com

How to Configure a Workflow to Calculate an HRD Score

This tutorial demonstrates how a template workflow can be configured to calculate a Homologue Recombination Deficiency (HRD) score by adding the tools Detect Regional Ploidy and Calculate HRD Score (beta).

In more detail, during this tutorial, we will:

- Use the template workflow **Identify Variants (TAS)** as an example.
- Add the tool **Copy Number Variant Detection (CNVs)** to the workflow.
- Add the tool **Detect Regional Ploidy**.
- Add the tool **Calculate HRD Score (beta)** to the workflow.

Note that any template workflow that is used to identify somatic variants from targeted DNA sequencing can be used instead of Identify Variants (TAS).

Prerequisites For this tutorial, you must have installed the Biomedical Genomics Analysis plugin version 22.1 or newer. How to install plugins is described here: http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Install.html.

General notes In this tutorial, we use the tool Calculate HRD Score (beta) to provide an HRD score. The tool is designed to calculate an HRD score from targeted research resequencing experiments and is described in detail in the manual: https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/current/index.php?manual=Calculate_HRD_Score_beta.html.

Copy Number Variant Detection (CNVs) is used to produce a Target-level Annotation Track which is needed as input to Detect Regional Ploidy. Detect Regional Ploidy in turn produces a Ploidy Target Track which is needed at input to Calculate HRD Score (beta).

Control samples that are used to generate a baseline for copy number variant detection must be available. Control samples must not contain any copy number variants and it is important that control samples are prepared with the same panel, the same library preparation protocol and the same sequencing technology as case samples. To generate control read mappings, the control samples should be analyzed with the same workflow that is used as a basis for HRD calculation. Read more about copy number variant detection here: https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Copy_Number_Variant_Detection.html.

Calculation of an HRD score is limited to targeted DNA sequencing data, hence whole genome sequencing data cannot be used. This is mainly due to a limitation in the tools used for copy number variant detection, as they are only able to handle targeted data. In addition, targets must be dispersed over a large part of the genome, to provide the Calculate HRD Score (beta) tool with information needed to detect large chromosomal aberrations.

Workflow elements can be renamed by right-clicking on the central part of the elements in the workflow. We will not do that in this tutorial, but renaming for example workflow inputs can make it easier to navigate in the workflow. In addition, workflow outputs can be configured to use custom names including placeholders and redirected to specific destinations. A detailed description of customization of input and output elements is available

here: https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Configuring_input_output_elements.html.

Make a copy of the workflow which will be adapted to provide an HRD score

First, make a copy of the Identify Variants (TAS) workflow. Navigate to the workflow in the toolbox:

> **Template Workflows** | **Biomedical Workflows** ( ) | **Targeted Amplicon Sequencing** ( ) | **Somatic Cancer** ( ) | **Identify Variants (TAS)** ( )

Alternatively, type identify variants (tas) in the search box at the top of the toolbox (figure 1).

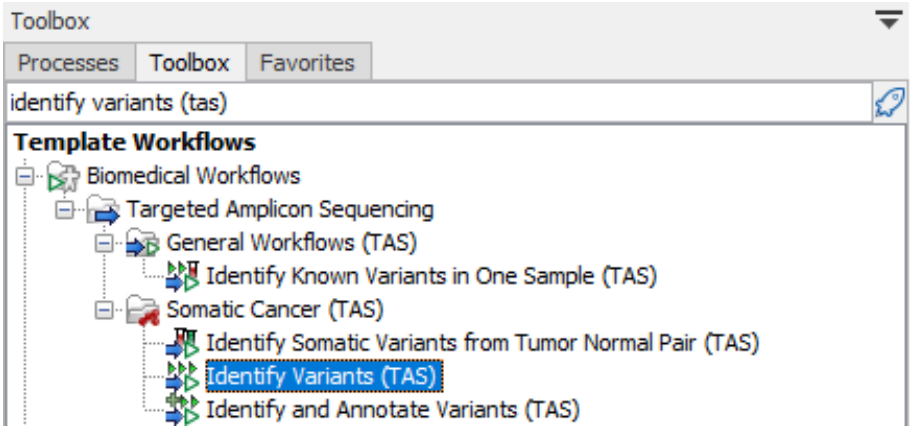

Figure 1: *Locate Identify Variants (TAS) in the Toolbox.*

Right click on **Identify Variants (TAS)** and choose **Open Copy of Workflow**. The opened workflow is now visible in View area. From there, save the workflow by right clicking on the workflow tab and choosing **Save As** (figure 2).

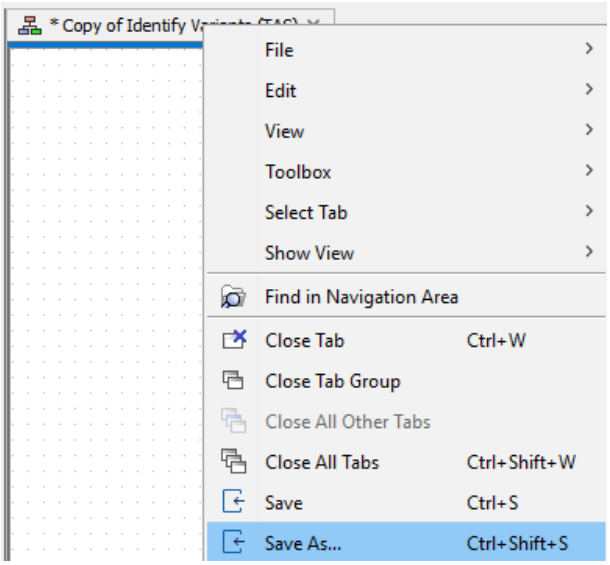

Figure 2: *Save a copy of the workflow Identify Variants (TAS).*

Add Copy Number Variant Detection (CNVs)

As calculation of an HRD requires each of the three tools **Copy Number Variant Detection (CNVs)**, **Detect Regional Ploidy** and **Calculate HRD Score (beta)**, we will first add **Copy Number Variant Detection (CNVs)** to the workflow. If a workflow that already contains **Copy Number Variant Detection (CNVs)** is used, these steps can be skipped, and you can go straight to the section Add Detect Regional Ploidy.
Before continuing, the workflow must be open in the view area. If you closed the workflow after saving it, you can re-open by double-clicking on the name of the workflow in the navigation area.
To add **Copy Number Variant Detection (CNVs)**, click **Add Element** in the bottom left corner of the workflow. In the resulting wizard, type copy in the search bar, select **Copy Number Variant Detection (CNVs)** and press **OK** (figure 3).

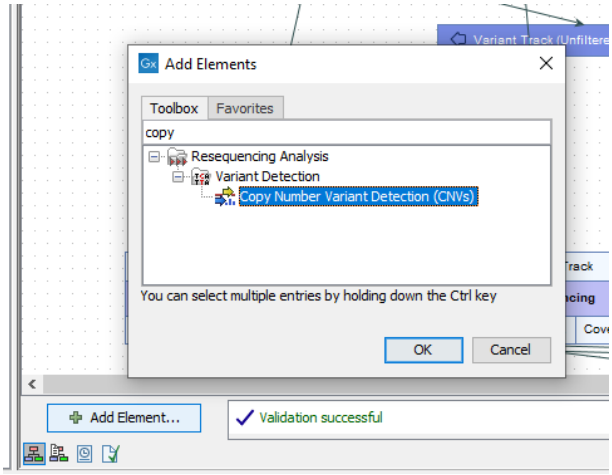

Figure 3: *Add Copy Number Variant Detection (CNVs) to the workflow.*

Copy Number Variant Detection (CNVs) is now visible in the workflow View area. The text in the tool is red, because the tool needs to be connected to the other workflow elements (figure 4).

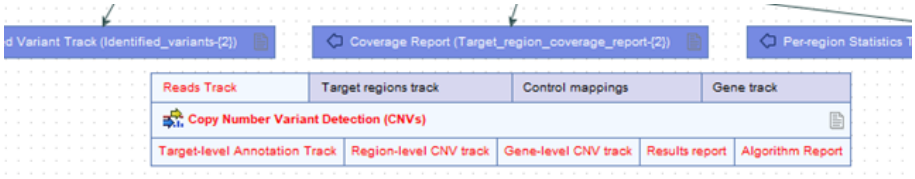

Figure 4: *The Copy Number Variant Detection (CNVs) element in the workflow.*

Drag **Copy Number Variant Detection (CNVs)** down to the bottom part of the workflow.

In the next steps, we will configure **Copy Number Variant Detection (CNVs)** with required inputs. First, add a reads track to Copy Number Variant Detection (CNVs). The reads track (or read mapping) provided to **Copy Number Variant Detection (CNVs)** must be the final processed reads track. In this workflow, this reads track is produced by the tool **Local Realignment**.
Other workflows may have additional steps that modify the read mapping, in that case, identify the reads track or read mapping output (blue box) in the workflow. The tool producing the output is the tool that should be connected to **Copy Number Variant Detection (CNVs)**.

- Find the workflow element **Local Realignment**, you can use the search field to the right of the workflow to easily find the tool. Type local and press enter, the **Local Realignment element** will now be highlighted with a green border (figure 5).

- Click on the **Reads Track** output in the **Local Realignment** element and keep the mouse button pressed. Now drag an arrow down to the **Reads Track** input in **Copy Number Variant Detection (CNVs)** (figure 6).
- Release the mouse button. There should now be an arrow between **Local Realignment** and **Copy Number Variant Detection (CNVs)**.

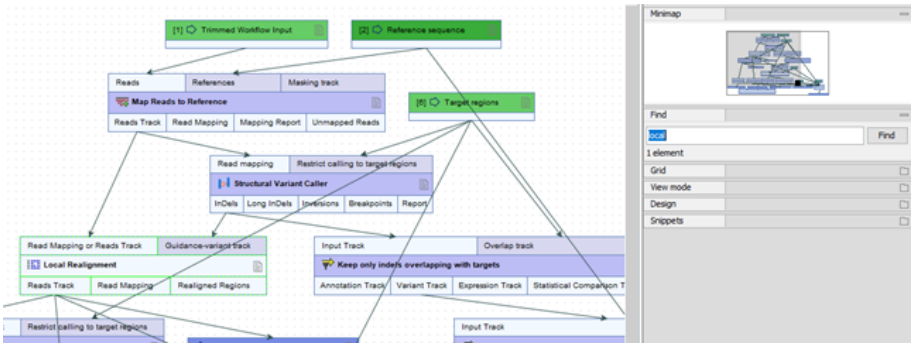

Figure 5: *Find the workflow element Local Realignment.*

Next, we will provide the **Target regions track** and the **Gene track** inputs to **Copy Number Variant Detection (CNVs)**.

- Locate the green **Target Regions** input element. Again, you can choose to use the search functionality.
- When the input element has been located, click on the bottom grey bar of the element and keeping the mouse button pressed, drag an arrow to the **Target regions track** input of **Copy Number Variant Detection (CNVs)**.
- Repeat these steps to connect the green **Genes** input element to **Gene track** input in **Copy Number Variant Detection (CNVs)** (figure 7).

Go through the following steps to configure **Copy Number Variant Detection (CNVs)** with control mappings that will be used to create a baseline:

- Right click on the **Control Mappings** input of **Copy Number Variant Detection (CNVs)** and choose **Connect to Workflow Input** (figure 8).
- Double click on the resulting green **Workflow Input** Element. This will open a wizard that allows selection of baseline control mappings (figure 9).
- Click on the folder icon to the right of the workflow input.
- In the resulting window, find the control mappings under navigation area and select them for use in the workflow by double clicking on them (figure 10).

Click OK to close the wizard. Note, if you do not have the control mappings ready at this time, it is always possible to go back and do these steps at a later point.

The text in the **Copy Number Variant Detection (CNVs)** element is still red, this is because we have not yet defined the output from the tool. We will do that, once the tool **Detect Regional Ploidy** has also been added to the workflow.

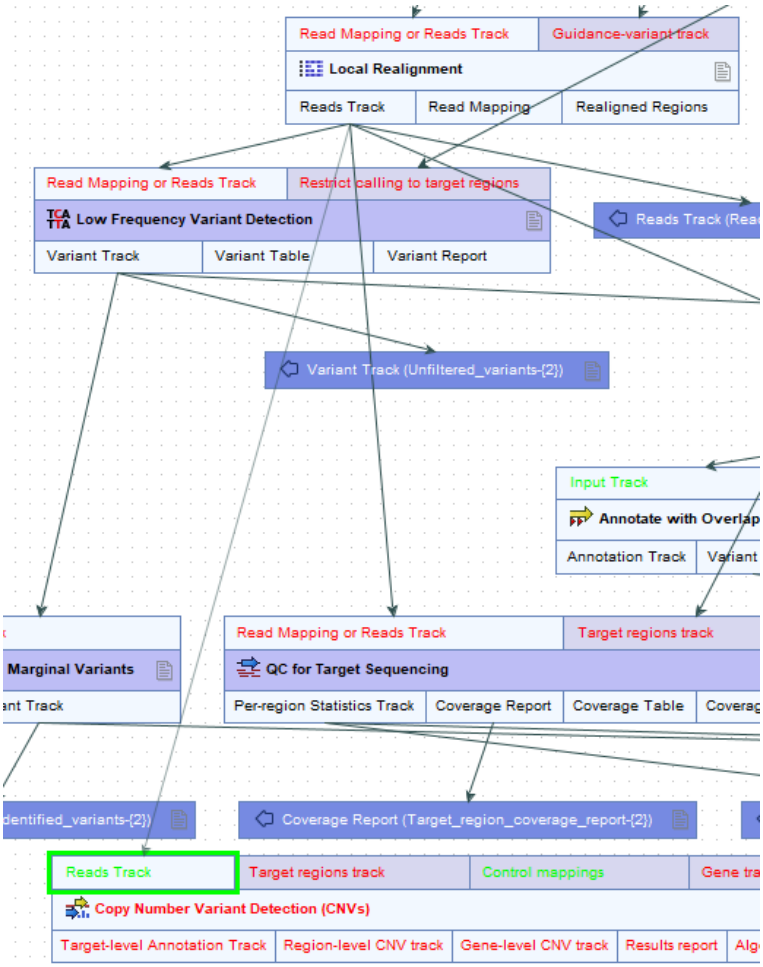

Figure 6: *Connect the Reads Track output from Local Realignment to the Reads Track input in Copy Number Variant Detection (CNVs). Note that when dragging the arrow from one workflow element to another, input channels that you can connect to will be in green, and input channels that you cannot connect to are in red.*

Add Detect Regional Ploidy (beta)

Next, we will add **Detect Regional Ploidy** to the workflow.

To add **Detect Regional Ploidy**, click **Add Element** in the bottom left corner of the workflow. In the resulting wizard, type regional in the search bar, select **Detect Regional Ploidy** and press **OK**. The tool is now visible in the workflow View area. If you cannot see the tool, you can find it by searching for ploidy in the search box to the right of the workflow. The text in the tool is red, because the tool needs to be connected to the other workflow elements.

Drag **Detect Regional Ploidy** down to the bottom part of the workflow if it is not already there.

In the next steps, we will configure **Detect Regional Ploidy** with required inputs.

First, configure **Detect Regional Ploidy** with the **Target-level Annotation Track**. This is the target level output from **Copy Number Variant Detection (CNVs)**.

- Locate **Copy Number Variant Detection (CNVs)** in the workflow. You can use the search field to the right of the workflow to easily find the tool.

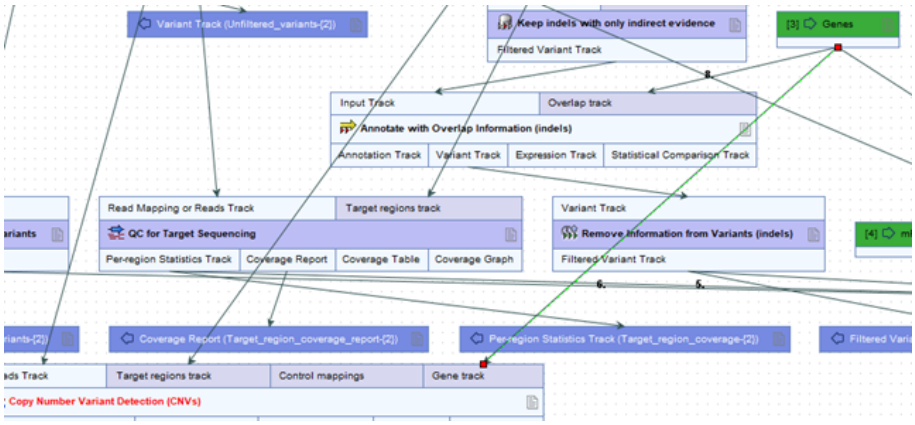

Figure 7: *Connect the Genes Track to the Gene Track input in Copy Number Variant Detection (CNVs).*

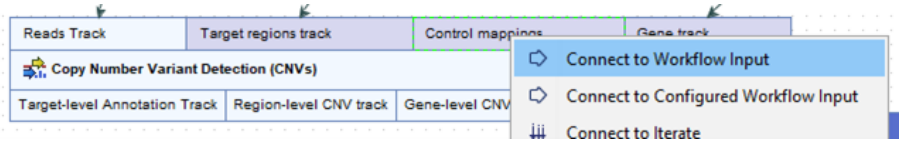

Figure 8: *Add input to Copy Number Variant Detection (CNVs) in order to provide the tool with control mappings.*

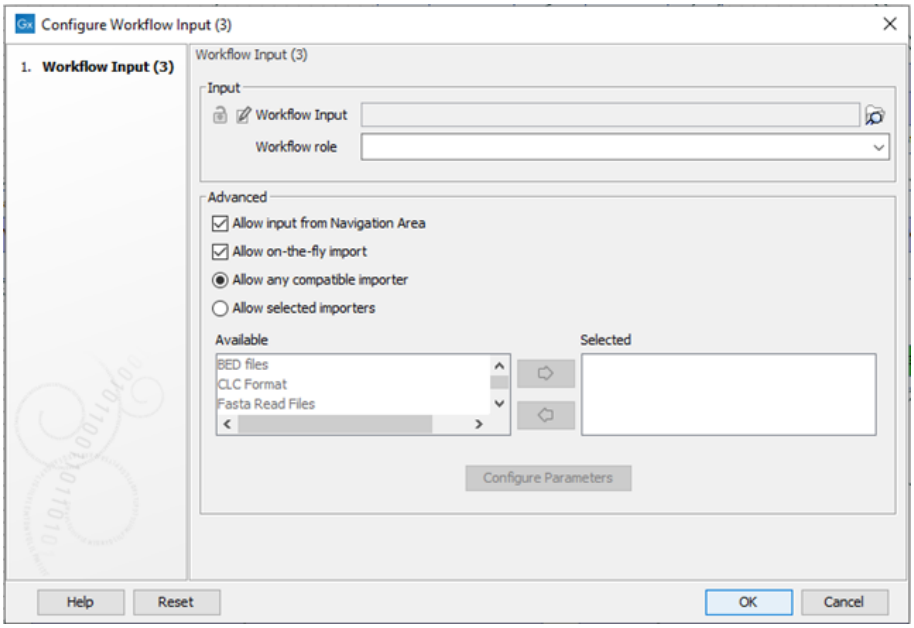

Figure 9: *To provide Copy Number Variant Detection (CNVs) with control mappings or coverage tables, press the folder symbol to the right of Workflow Input.*

- Click on the **Target-level Annotation track** output in the **Copy Number Variant Detection (CNVs)** element and keep the mouse button pressed.
- Now drag an arrow down to the **CNV Target-level Annotation track** input in **Detect Regional Ploidy**. Release the mouse button (figure 11).

The two tools should now be connected by an arrow between the **Target-level Annotation Track** output from **Copy Number Variant Detection (CNVs)** and the **Target-level Annotation Track** input

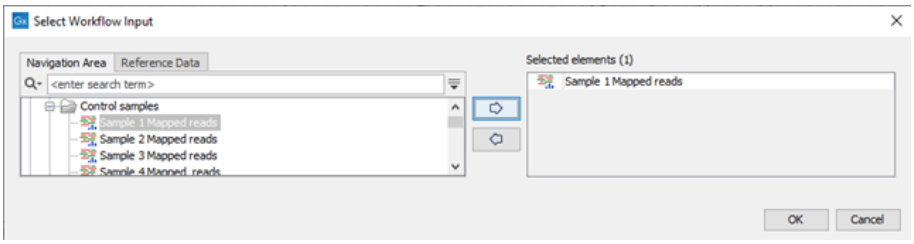

Figure 10: *Select control mappings or control coverage tables for copy number variant detection.*

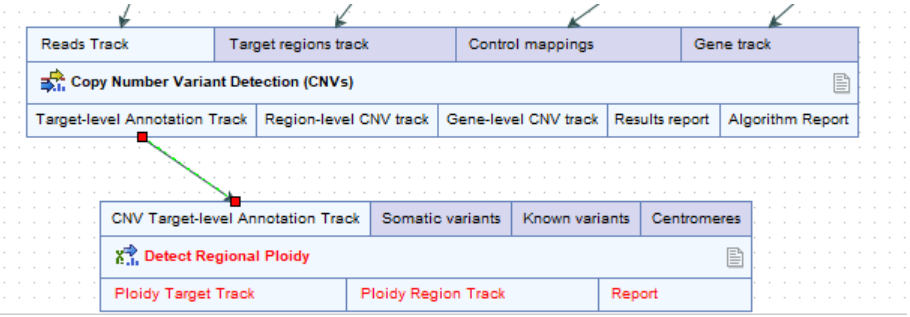

Figure 11: *Provide the Target-level Annotation Track output from Copy Number Variant Detection (CNVs) to Detect Regional Ploidy.*

in **Detect Regional Ploidy**.

Next, provide the Somatic variants input to Detect Regional Ploidy. In general, an unfiltered variant track is preferred, we will therefore provide the track directly from **Low Frequency Variant Detection**.

- Locate **Low Frequency Variant Detection in the workflow**. If needed, use the search functionality in the right side bar.
- Connect the **Variant Track** output from **Low Frequency Variant Detection** to the **Somatic Variants** input in **Detect Regional Ploidy** (figure 12).

We want to use dbSNP as our **Known variants** input in **Detect Regional Ploidy**. The dbSNP database is available in the CLC reference data. To download dbSNP:

- Go to **References** in the top right corner of the workbench.
- Under **Download Genomes** find the **Homo sapiens - hg19** reference data set. Note, if you are working with a panel designed against hg38, instead use a hg38 reference data set, the dbSNP database must match the reference sequence used.
- In **Homo sapiens - hg19**, choose **Dbsnp (common) variants** and press download in the bottom right corner (figure 13). Note, that you can choose between downloading reference data locally and on a server using the drop-down list in the top right corner of the **Manage Reference Data**. If you will be running the workflow on a server, choose to download dbSNP to a server instead of locally.

Set dbSNP as the **Known variants** input in **Detect Regional Ploidy**:

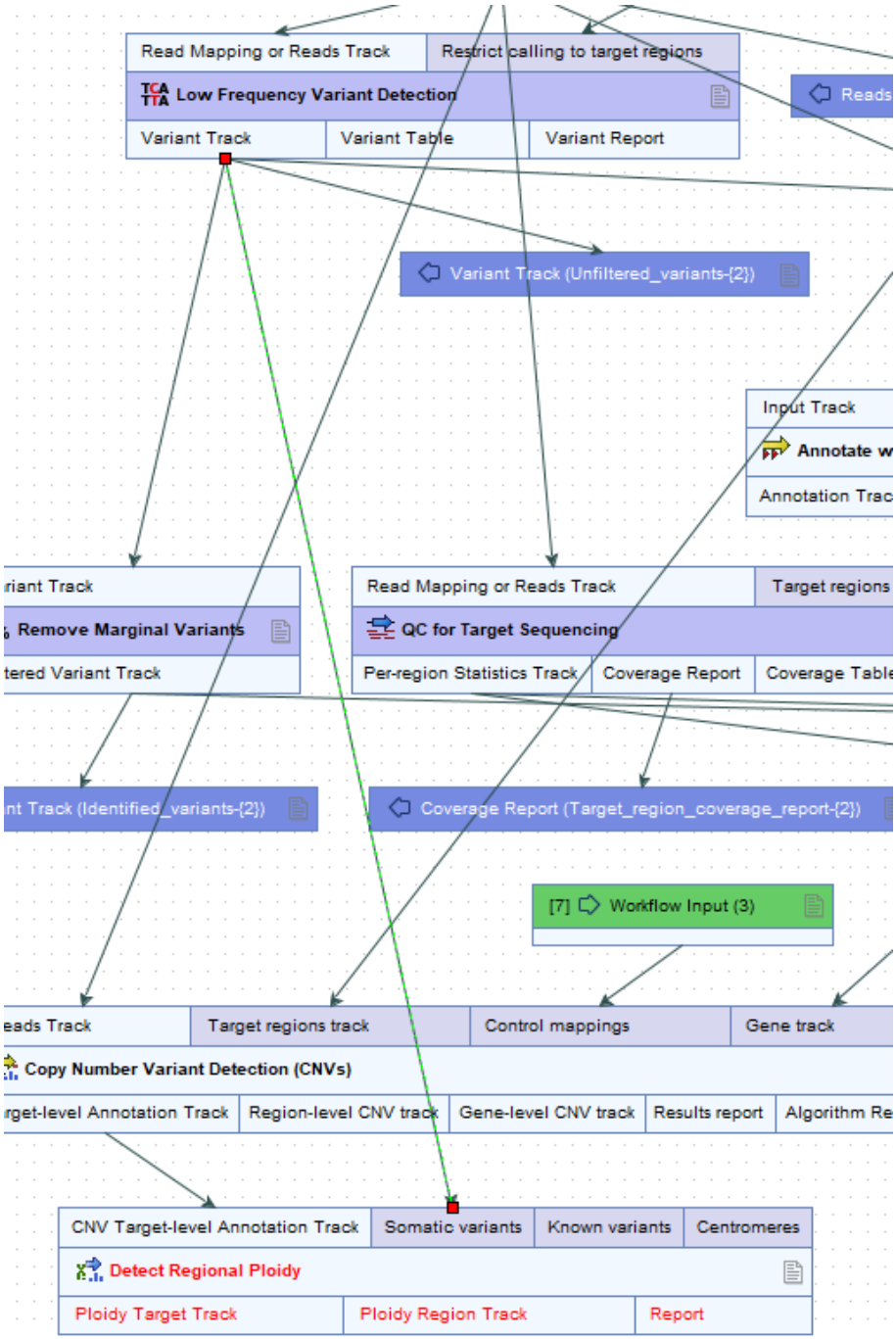

Figure 12: *Connect the Variant Track output from Low Frequency Variant Detection to the Somatic variants in Detect Regional Ploidy.*

- Right click on the **Known variants** input in **Detect Regional Ploidy** and choose **Connect to Workflow Input**.
- Double click on the resulting green **Workflow Input** element.
- In the resulting wizard, click on the folder icon to the right of **Workflow Input**.
- In the dialog, locate dbSNP under CLC_References/Genomes/Homo_sapiens_hg19-download-date/Homo sapiens (hg19) dbsnp (common) variants and select it by double clicking on

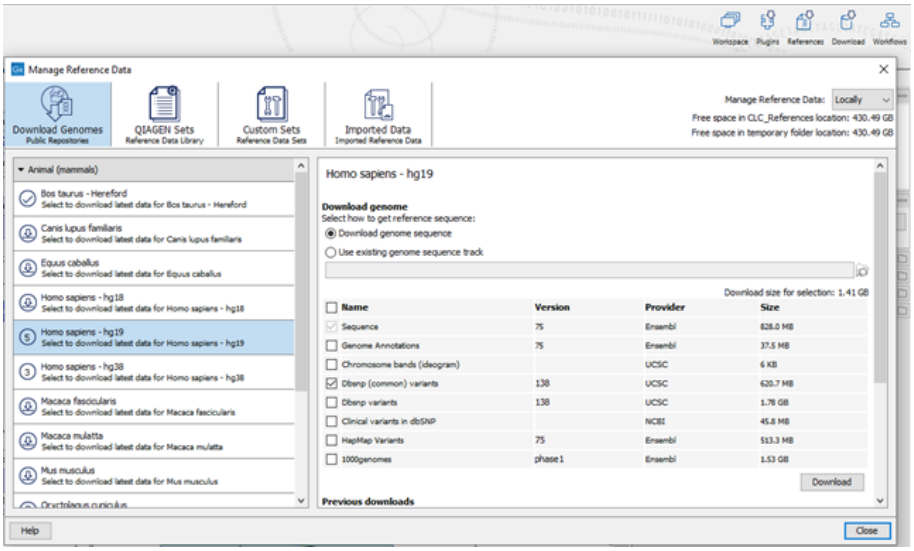

Figure 13: *Download dbSNP common from the CLC reference data.*

the file.

- Click **OK** twice to close the wizard.

Similarly to dbSNP, we will download a centromere file that will be used as the **Centromere** input to **Detect Regional Ploidy**.

- Go to References in the top right corner of the workbench.
- Under **QIAGEN Sets**, find **Reference Data Elements** (figure 14)
- Locate **Centromeres hg19**. Note that a centromere file is also available for hg38 if needed.
- Click **OK** twice to close the wizard.

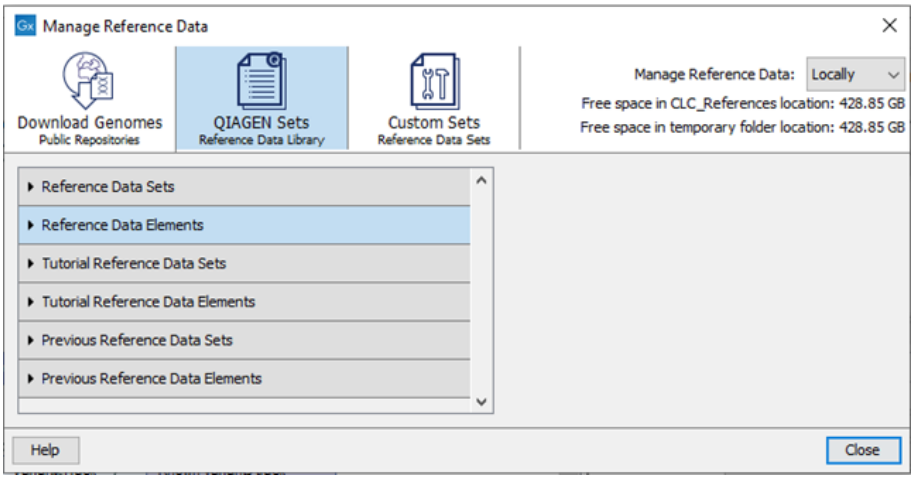

Figure 14: *The folder in the Manage Reference Data dialog where Centromeres are located. You may need to close the folder Reference Data Sets to obtain a view that is similar to the one in the screenshot. To find the file, open the folder in blue and scroll down to Centromeres.*

Once the centromere file has been downloaded we will use it as input to **Detect Regional Ploidy**:

- In the workflow, right click on the **Centromeres** input in the **Detect Regional Ploidy** workflow element and choose **Connect to Workflow Input**.
- Double click on the resulting green **Workflow Input** element.
- In the dialog that opens up, first click on the folder icon to the right of Workflow Input, then under the tab **Navigation Area** navigate to CLC_References/homo_sapiens/centromeres (figure 15).
- Double click on the centromere file to select it.
- Click **OK** twice to close the wizard.

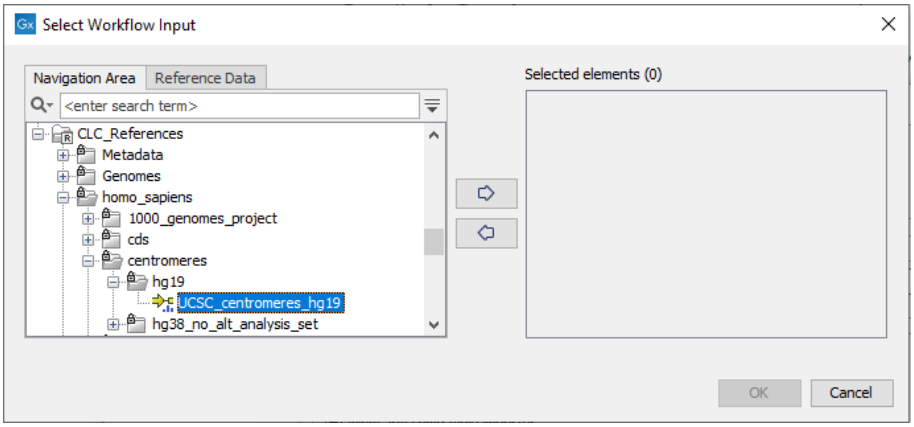

Figure 15: *Choose UCSC_centromeres_hg19 as input to Detect Regional Ploidy.*

All of the necessary inputs have now been provided to **Detect Regional Ploidy**.

Add Calculate HRD Score (beta)

To add **Calculate HRD Score (beta)**, click **Add Element** in the bottom left corner of the workflow. In the resulting wizard, type hrd in the search bar, select **Calculate HRD Score (beta)** and press **OK** (figure 16).

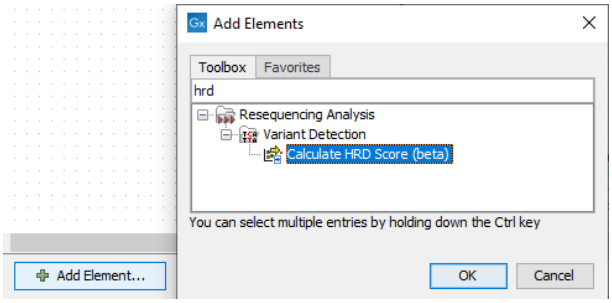

Figure 16: *Add the tool Calculate HRD Score (beta) to the workflow.*

Calculate HRD Score (beta) is now visible as a workflow element in the workflow View area. If you cannot see the tool, you can find it by searching for hrd in the search box to the right of the workflow. The text in the tool is red, because the tool needs to be connected to the other workflow elements (figure 17).

Drag **Calculate HRD Score (beta)** down to the bottom part of the workflow.

Figure 17: *Calculate HRD Score (beta) after it has been inserted in the workflow. The text in the element is red because connections and outputs have not been configured.*

First, configure **Calculate HRD Score (beta)** with the **Ploidy Target Track** from **Detect Regional Ploidy**.

- Locate **Detect Regional Ploidy** in the workflow.
- Click on the **Ploidy Target Track** output in the **Detect Regional Ploidy** element and keep the mouse button pressed.
- Now drag an arrow down to the **Ploidy states target level track** input in **Calculate HRD Score (beta)**. Release the mouse button (figure 18).

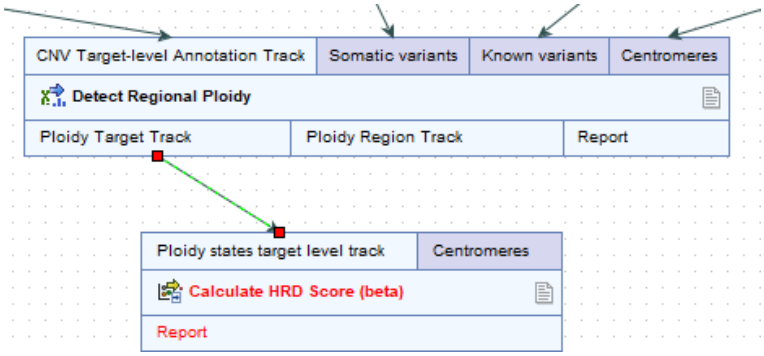

Figure 18: *Provide the Ploidy Target Track output from Detect Regional Ploidy to Calculate HRD Score (beta).*

The two tools should now be connected by an arrow between the **Ploidy Target Track** output from **Detect Regional Ploidy** and the **Ploidy states target level track** input in **Calculate HRD Score (beta)**.

Next set the **Centromere** input to **Calculate HRD Score (beta)**.

- In the workflow, right click on the bottom grey bar in the **UCSC_centromeres_hg19** input to the **Detect Regional Ploidy** workflow element.
- Keeping the mouse button pressed, drag an arrow to the **Centromeres** input of **Calculate HRD Score (beta)** (figure 19).

The calculated HRD score is listed in the report produced by Calculate HRD Score (beta). To output the report right-click on the **Report** output of the **Calculate HRD Score (beta)** workflow element and choose **Use as Workflow Output**.

All of the necessary tools and connections necessary for calculation of an HRD score have now been added to the workflow.

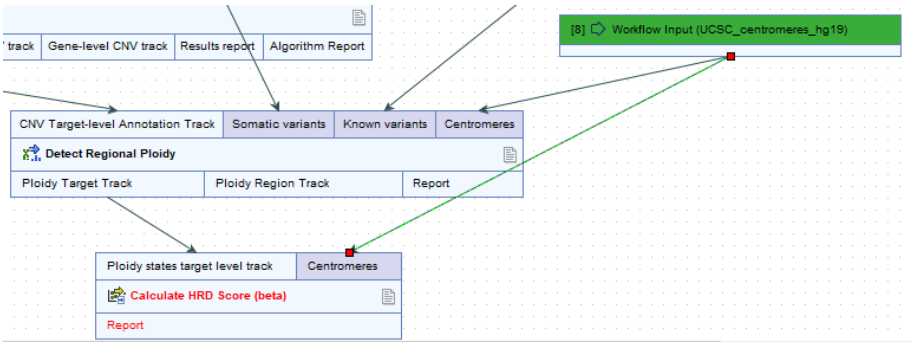

Figure 19: *Choose UCSC_centromeres_hg19 as input to Calculate HRD Score (beta).*

Settings

The default settings of the **Calculate HRD Score (beta)** tool have been set based on a limited number of samples, hence the settings may need to be adjusted depending on the application. To adjust the settings of **Calculate HRD Score (beta)**, double click on the central part of Calculate HRD Score (beta) workflow element. This will open up a wizard that allows adjustment of settings. Please see the manual for details about individual parameters: https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/current/index.php?manual=Calculate_HRD_Score_beta.html.

In addition, the settings used for CNV detection and as well as regional ploidy estimation may affect the calculated HRD score. In initial experiments, we found that setting the parameter **Graining level for the prediction of CNV regions** in **Copy Number Variant Detection (CNVs)** to fine, gave slightly better results. However, the setting may need to be adjusted depending on the application.
To adjust the settings in **Copy Number Variant Detection (CNVs)**, double click on the central part of the workflow element. This will open up a wizard that allows adjustment of all available settings.