Tutorial

Modifying a Workflow

December 15, 2025

QIAGEN Aarhus A/S · Kalkværksvej 5, 11. · DK - 8000 Aarhus C · Denmark
digitalinsights.qiagen.com · ts-bioinformatics@qiagen.com

Sample to Insight

Modifying a Workflow

Customizing an existing workflow can be much faster than creating a complex workflow from scratch. This tutorial shows how to modify a workflow by updating default settings, and adding and removing tools.

Data used in this tutorial

This tutorial uses low coverage (1.5x) whole genome sequencing data for Oryza Sativa from [Wang et al., 2016]. In a 2021 Nature paper, these samples were resequenced to discover a 29nt deletion that allows rice to grow with greater nitrogen use efficiency (NUE) [Liu et al., 2021]. The deletion is common in wild rice strains, but rare in agricultural varieties. Reintroducing the deletion may lead to strains that produce more rice with less fertilizer.

We will modify the Identify DNA Germline Variants template workflow, included with *CLC Genomics Workbench*, to detect the 29nt deletion in one sample from the original low coverage genome sequencing data.

Prerequisites

For this tutorial, you must be working with *CLC Genomics Workbench* 26.0.1 or higher. Note that higher versions may produce slightly different results than those shown here.

General tips

- Throughout this tutorial, we provide links to relevant manual pages, which we recommend exploring for additional details.
- To select multiple items simultaneously in lists, or elements in a workflow editor, hold down the Ctrl key (⌘ on Mac) when selecting each item.
- The layout of a workflow in the Workflow Editor can be adjusted manually by clicking on workflow elements to select them, and then with the mouse button depressed, dragging them to the location you want. Alternatively, the workflow layout can be automatically updated by pressing the **L** key.
- There is no need to save the workflow in this tutorial, but you can use **File** | **Save** to save your progress.
- Changes can be reverted using the Undo button.

Download and import data for the tutorial

The example data for this tutorial includes the Oryza Sativa reference genome, targeted region annotations, and sequencing reads.

1. Download the example data from `https://resources.qiagenbioinformatics.com/testdata/osativa_example_data.zip`.
2. Start *CLC Genomics Workbench*.
3. Import the data by going to:

   File | **Import** ( ) | **Standard Import** ( )
4. Choose the zip file you just downloaded. Leave the Import type set to **Automatic**.

After import, the files listed in the **Navigation Area** should look like those shown in figure 1.

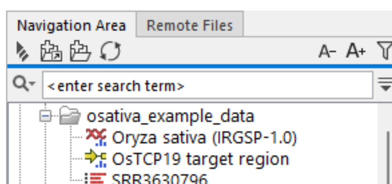

Figure 1: *Navigation area after import of files.*

View a workflow in the editor

1. Find the template workflow Identify DNA Germline Variants in the Workflows section of the Toolbox, in the bottom left side of the Workbench.
2. Right-click on the workflow name and choose "Open copy of workflow" (figure 2).

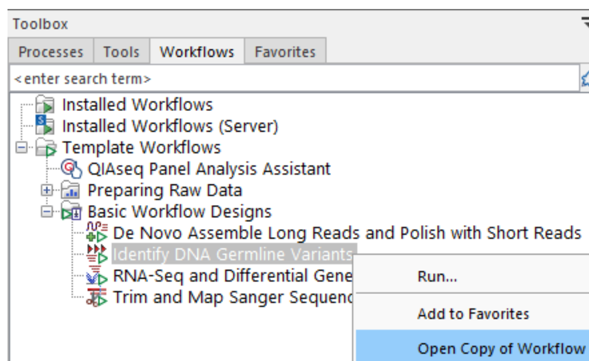

Figure 2: *Right-click to open a copy of a workflow.*

This workflow includes steps to map a set of reads to a reference sequence, carry out variant detection, create a sample report, and to gather results in a track list. Individual analysis steps have been placed into groups with names such as "Read Processing" and "Variant Detection and Filtering". By default these groups are collapsed, so it is easy to get an overview (figure 3). Each group can be expanded individually to see the analysis steps it contains, or all groups in a workflow can be expanded at once.

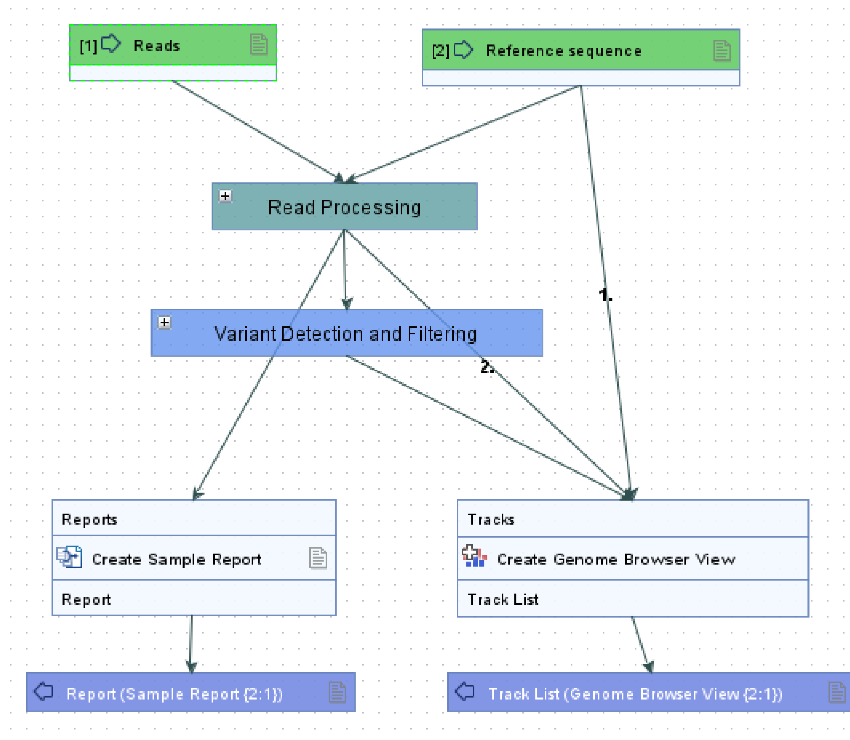

Figure 3: *The default view of a copy of the Identify DNA Germline Variants template workflow opened in the workflow editor.*

3. Left-click on the background canvas and press the **E** key on your keyboard to expand all the groups. Press the **L** key to adjust the workflow layout.

You should see that each group contains several elements with connections, shown as arrows, between them. Connections define the flow of data through the workflow, i.e. which outputs are used as inputs for downstream steps, which results are saved, and so on.

To see the entire workflow, you may need to scroll, or zoom.

Adjust element settings

As the **Identify DNA Germline Variants** workflow was not designed for ultra-low coverage data, we will change some default settings in the variant detection tools, to take into account the type of data we are working with. Our modifications will increase the sensitivity of calling, but also the number of false positives.

4. Double click on the **InDels and Structural Variants** element, in the middle of the expanded "Read Processing" group, to open the configuration wizard.

5. Set the **P-Value threshold** to 1.0 and click **Finish** (figure 4).

Notice that after modifying the P-Value threshold, the InDels and Structural Variants element body color changed from light gray to purple, indicating that there are non-default settings.

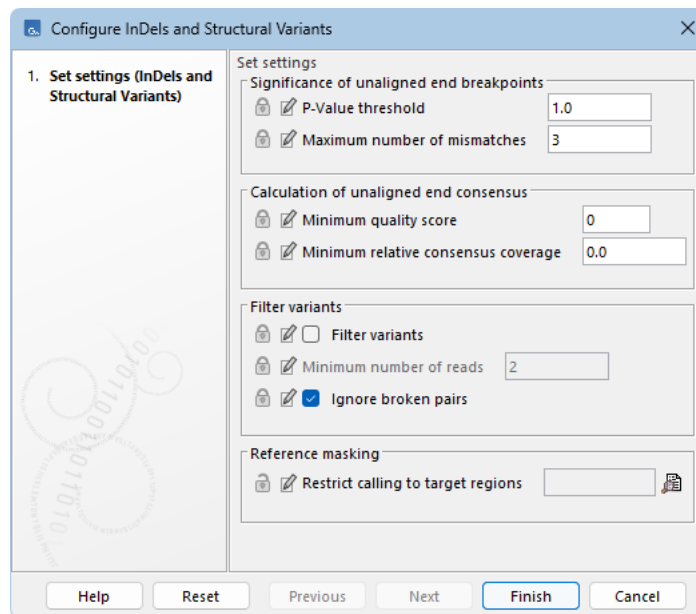

Figure 4: *Set the InDel and Structural Variants element p-value threshold.*

6. Type "Fixed Ploidy" into the Find field in the Side Panel to find the **Fixed Ploidy Variant Detection** element and bring it into view (figure 5). Double-click it to open the configuration wizard.

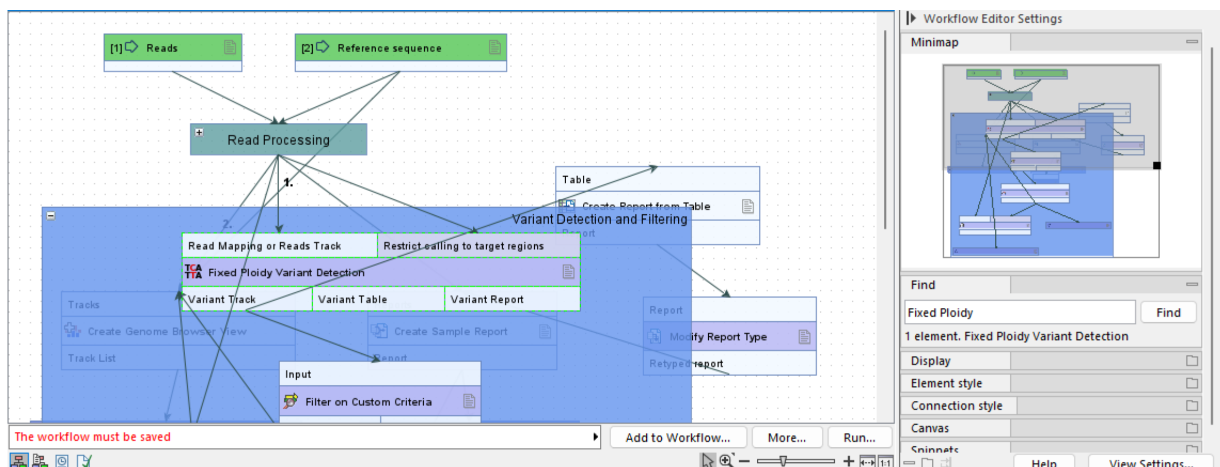

Figure 5: *The Find functionality in the Side Panel has been used to find the Fixed Ploidy Variant Detection element.*

7. Click **Next** in the wizard, and in the "General filters" step, set the **Minimum coverage** to 2.

By default, most settings in most workflow elements are locked, meaning their value cannot be changed when launching the workflow. Here, we will unlock the "Coverage and count filters" settings so that their values can be changed in the workflow launch wizard.

8. Click on the lock symbols beside each of the "Coverage and count filters" options, "Minimum coverage", "Minimum count" and "Minimum frequency (%)", to change them from locked ( ) to unlocked ( ) (figure 6).

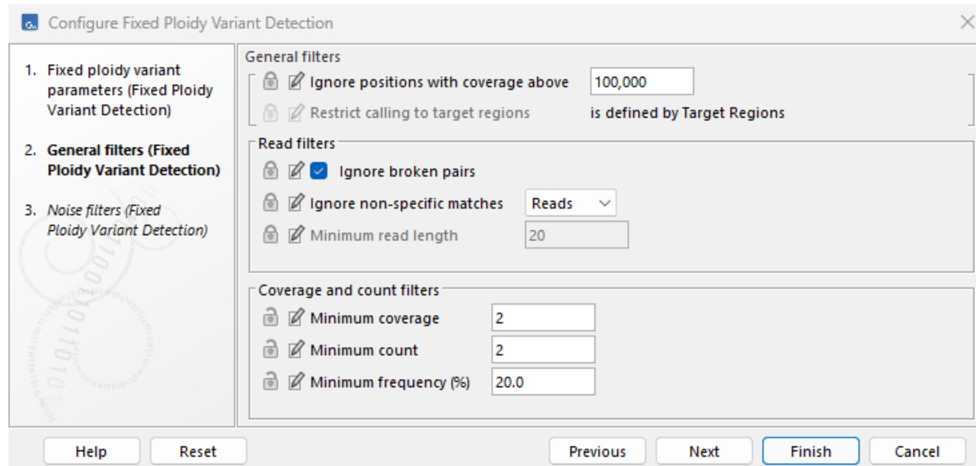

Figure 6: *The three options in the "Coverage and count filters" section are unlocked.*

9. Click **Finish**.

We will modify the **Reference sequence** Input element, which is green, so that the workflow launch wizard does not prompt us to select a Reference Data Set.

10. Double-click the **Reference sequence** element, remove "sequences" from the "Workflow role" field, and click **OK**.

See the manual for details about the connection between workflow roles and Reference Data Sets.

Add a new input

In the 2021 Nature paper, GWAS and RNA-Seq data were used to localize changes in nitrogen use efficiency to the OsTCP19 gene. We will therefore look for variants only in the OsTCP19 promoter and gene body.

To do this, we will modify the workflow so that the target regions of interest are specified once when launching the workflow, and then are used in all the relevant analysis steps.

11. Find the InDel and Structural Variants tool in the "Read Processing" group. Right-click on the "Restrict calling to target regions" input channel and choose "Connect Input Element".
12. To aid usability, rename the newly added Input element, which is green, as follows:
    (a) Left-click the body of the Input element to select just that element.
    (b) Right-click the Input element, choose "Rename Element" from the menu that appears. Assign an informative name to the element, e.g. "Target Regions", then click "OK".

The names given to Input elements are used in the launch wizard. By naming this element "Target Regions", the person launching the workflow will know what sort of data they should supply.

We now connect the Target Regions Input element to another element in the workflow.

13. Right-click the Target Regions element's output channel, (the narrow pale band, under the green body), and choose **Connect to** | **Fixed Ploidy Variant Dection** | **Restrict calling to target regions** (figure 7).

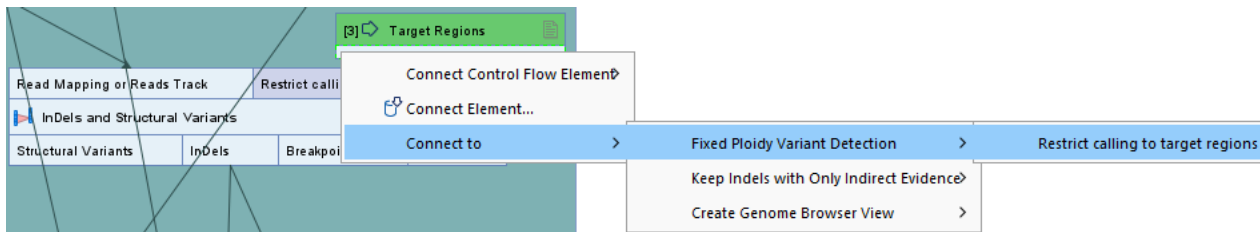

Figure 7: *Connect the Target Regions output channel to a Fixed Ploidy Variant Detection input channel.*

When launching the workflow, the wizard will now include a step for specifying a target regions track, and the analysis will only call variants in the regions defined in that track.

Remove an unnecessary element

Now that we are only calling variants in one gene, we will inspect the read mapping results visually to determine whether the data is high quality in that region. This makes the **QC for Sequencing Reads** element unnecessary, so we will remove it.

14. Right-click the body of the **QC for Sequencing Reads** element and choose **Delete Selected**.

This action removed the **QC for Sequencing Reads** element and all connections to and from that element.

Before carrying out the steps below, you may wish to press the **C** key to collapse groups, followed by the **L** key to adjust the workflow layout. The steps as described allow connections to be made to elements within the collapsed groups without needing to see or interact with those elements directly. See the manual for in more information about connecting elements.

Add elements

We would like to include details about the variants detected in the Sample Report. To do this, we will add some new elements to the workflow.

Create Sample Report accepts reports as input, so we first need to create a report from the variant track output of **Fixed Ploidy Variant Detection**. We will also make changes so that the section in the Sample Report containing variant information has a meaningful heading.

15. Click the "Add to Workflow..." button at the bottom of the Workflow Editor, or press the **A** key on your keyboard, to open the "Add to Workflow" dialog.
16. Type "Reports" in the search box (figure 8).
17. Select "Create Report from Table" and click the **Add** button.

QIAGEN Tutorial

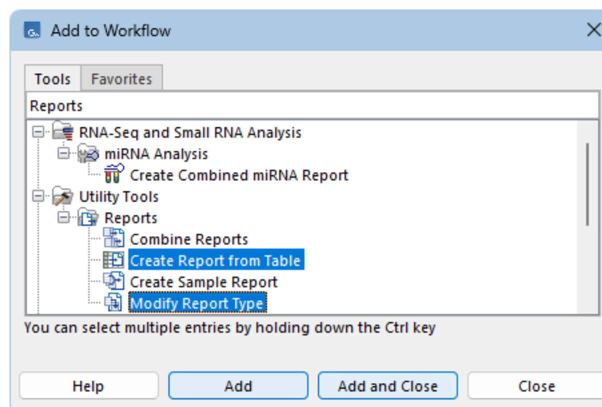

Figure 8: *The "Add to Workflow" dialog.*

18. Select "Modify Report Type" and click the **Add and Close** button.

The names of the two new elements are red, indicating that they are not yet correctly configured. The problem is that they do not yet have connections to their input and output channels. In the steps below, we add the necessary connections.

19. Right-click the "Table" input channel of **Create Report from Table** and choose **Connect to** | **Fixed Ploidy Variant Detection** | **Variant Track**.

    Note: The "Variant Track" output from **Fixed Ploidy Variant Detection** contains the tabular data we wish to include in the Sample Report. We are not using the "Variant Table" output here. See the manual for further information about these variant detection output types.

We could connect the **Create Report from Table** element directly to the **Create Sample Report** element, but if we do that, the section containing the variants in the Sample Report will be given the name "Create report from table". Using Modify Report Type, we can provide a more meaningful name.

20. Right-click the "Report" output of the **Create Report from Table** element, and choose **Connect to** | **Modify Report Type** | **Report**.

21. Similarly, connect the "Retyped report" output channel of **Modify Report Type** to the "Reports" input channel of **Create Sample Report**.

22. Double-click the **Modify Report Type** element and type a section heading for the Sample Report into the "Report type" field, for example, "Detected Variants" (figure 9). Then click **Finish**.

With these changes, the Sample Report will contain all the available information about each variant. This is more information than is needed in the Sample Report, and if that report is exported to a PDF format file, more information than can be included in a readable form.

To address this, we will configure the **Create Report from Table** element, so only the relevant variant information is included in the in the report it creates.

23. Double-click the **Create Report from Table** element.

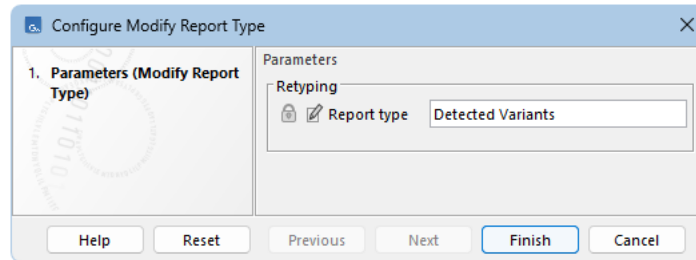

Figure 9: *Specify a name for the Sample Report section that will contain information about the variants detected.*

24. Into the Columns field, enter the following information (e.g. by copying and then pasting):
    `Chromosome, Region, Type, Reference, Allele, Length, Zygosity`

25. Click **Finish**.

Add an export element

The results generated by the workflow so far can be opened in the Workbench for viewing. To save results to third party formats, like Excel or PDF, Export elements are added. Our final modification will be to export the Sample Report to a PDF format file.

26. Right-click the "Report" output channel of the **Create Sample Report** element, and choose **Connect Element. . . .**

27. Type "PDF" in the search box to limit the number of items listed, and then choose **Export PDF**.

28. Click **Add and Close**.

At this point, the message in the validation field to the left of the "Add to Workflow..." button should contain just the message "The workflow must be saved". This means there are no configuration errors that need to be addressed, and that the workflow can be run.

Launch the modified workflow

1. Click the "Run. . . " button at the bottom right of the Workflow Editor to open the launch wizard.

2. In the "Select Reads" step, choose the sequence list "SRR3630796" and click **Next**.

3. In the "Select Reference sequence" step, choose the "Oryza sativa (IRGSP-1.0)" track and click **Next**.

4. In the "Select Target Regions" step, choose the "OsTCP19 target region" track and click **Next**.

5. In the "Trim Reads" step, keep the settings as they are and click **Next**.

6. In the "Fixed Ploidy Variant Detection" step, keep the settings as they are and click **Next**.

7. In the "Filter on Custom Criteria" step, keep the settings as they are and click **Next**.

8. In the "Export PDF" step, click **Browse** then **Select folder**, and choose a suitable location for the PDF file that will be produced to be saved to. Then click **Open** and **OK**. Then click **Next**.

9. In the "Result handling" step, keep the settings as they are and click **Next**.

10. In the "Save location for new elements" step, choose where data elements created by the workflow should be saved.

11. Click **Finish** to run the workflow.

Under the **Processes** tab in the lower left corner of the Workbench, the progress of the workflow can be monitored. By clicking the small arrow to the right of the process, you can choose to **Show Log Information** to closely monitor the progress of the analysis (figure 10).

Figure 10: *An option to view the analysis log is available for each process listed under the Processes tab. Here, the log is open for the running workflow and shows that the analysis is currently progressing through the Map Reads to Reference step.*

The first time the workflow is run, it will take a few minutes on a standard laptop. Subsequent runs are likely to be faster.

When the workflow has finished, open the Sample Report, either in the Workbench, or by opening the exported PDF file. You should see a "Detected Variants" section at the end of the report. The

4. Detected Variants

Chromosome	Region	Type	Reference	Allele	Lenght	Zygosity
6	6566662	SNV	T	C		Homozygous
6	6567056	SNV	A	T		Homozygous
6	6567607	SNV	A	G		Homozygous
6	6567739	SNV	C	T		Homozygous
6	6567740^6567741	Insertion	-	T		Homozygous
6	6567898	SNV	A	G		Homozygous
6	6568039	SNV	T	C		Homozygous
6	6568075	SNV	A	G		Homozygous
6	6568090	SNV	A	G		Homozygous
6	6568176	SNV	G	A		Homozygous
6	6568366	SNV	T	G		Homozygous
6	6568411	SNV	A	G		Homozygous
6	6568805	SNV	C	T		Homozygous
6	6568808	SNV	C	T		Homozygous
6	6568919	SNV	G	A		Homozygous
6	6569011	SNV	C	T		Homozygous
6	6569277..6569305	Deletion	CCCTTCCTATACATTTTTTTTGTGGGGA	-		Homozygous

Figure 11: *Part of the PDF report generated by the workflow.*

29nt deletion reported in the Nature paper should be present in the list presented there, along with other insertions and deletions (figure 11).

Our sample is GSOR310102. Expected variants in this sample based on higher coverage resequencing are listed in Supplementary Table 2 of the 2021 Nature paper. These variants include the 29nt deletion, which is listed as variant v3.

Bibliography

[Liu et al., 2021] Liu, Y., Wang, H., Jiang, Z., Wang, W., Xu, R., Wang, Q., Zhang, Z., Li, A., Liang, Y., Ou, S., et al. (2021). Genomic basis of geographical adaptation to soil nitrogen in rice. *Nature*, 590(7847):600–605.

[Wang et al., 2016] Wang, H., Xu, X., Vieira, F. G., Xiao, Y., Li, Z., Wang, J., Nielsen, R., and Chu, C. (2016). The power of inbreeding: Ngs-based gwas of rice reveals convergent evolution during rice domestication. *Molecular Plant*, 9(7):975–985.