Tutorial

Detect Fusions and Actionable Variants Using QIAseq Multimodal Panels

January 7, 2020

Sample to Insight

QIAGEN Aarhus · Silkeborgvej 2 · Prismet · 8000 Aarhus C · Denmark
Telephone: +45 70 22 32 44 · www.qiagenbioinformatics.com · ts-bioinformatics@qiagen.com

Detect Fusions and Actionable Variants Using QIAseq Multimodal Panels

This tutorial uses the capabilities of *CLC Genomics Workbench* and Biomedical Genomics Analysis plugin to detect fusion events and find actionable variants in reads generated on a custom QIAseq Multimodal panel using SeraCare Seraseq ® Myeloid DNA and RNA Fusion Reference Materials.

This tutorial covers the following in just a few steps:

- Import data required for the analysis.
- Create a custom panel analysis.
- Import Illumina paired reads in the workbench.
- Find fusion events and low frequency variants with the **Analyze QIAseq Panels** guide.
- (Optionally) fine-tune your analysis by running the analysis directly.
- (Optionally) upload fusion events and variants to QCI Interpret.

Prerequisites

For this tutorial, you must be working with *CLC Genomics Workbench* 20 (or higher) with the Biomedical Genomics Analysis plugin installed. How to install plugins is described here: `http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Install.html`.

For the optional section, "Upload your results to QCI Interpret", you will need an active QCI Interpret subscription (not a trial license), and a QCI Interpret account with API license privileges.

Import the panel data

1. First, download the sample data from our `website`. This file contains the sequencing reads for the sample we will analyze and primers and target regions for the custom QIAseq Multimodal Panel. Unzip it in a location of your choice on your computer.
2. Start *CLC Genomics Workbench*.
3. In *CLC Genomics Workbench*, import the data by going to: **File** | **Import** () | **Standard Import** (). From among the files just unzipped, choose those with names ending in ".clc". Leave the import type set to Automatic as shown in figure 1. Click Next and then select a folder to save the imported data to. A new folder can be created by clicking on the "New Folder" option at the top of the window. Once the import is completed, you should see the folder and files in the Navigation Area as shown in figure 2.
4. Open the Analyze QIAseq Panels guide and go to the Multimodal tab:
   Ready-to-Use Workflows | **QIAseq Panel Analysis** () | **Analyze QIAseq Panels**
5. Reference data is required for the analysis of QIAseq Multimodal panels. These data include the human reference genome, gene and mRNA annotations suitable for annotating fusion calls, and trim adapters specific to the panels. An easy way to download these is to click **Run** on one of the existing panels. In the first dialog choose to "Download to Workbench" (figure 3). If you do not see the dialog, then you already have the reference data and can skip this step. Once the data has been downloaded, click **Cancel**.

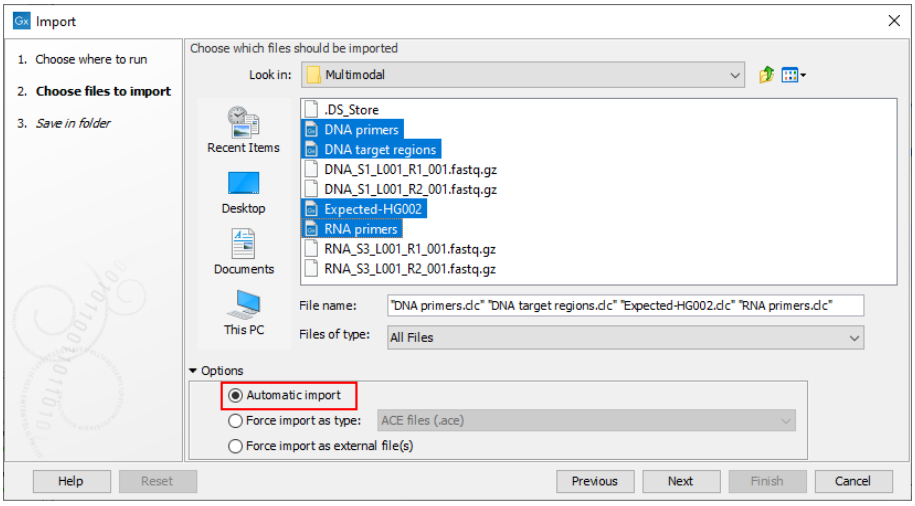


Figure 1: *Import the ".clc" files for this tutorial*

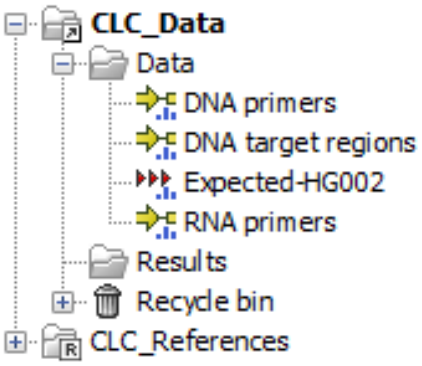


Figure 2: *The imported CLC files*

Create a custom panel analysis

The Analyze QIAseq Panels guide helps you run an analysis for a given panel in a streamlined way. Now that the reference data is present we will add our custom panel to the guide.

1. Click **Add custom panel** at the bottom of the Analyze QIAseq Panels guide window.
2. Choose to **Copy from** and select any of the existing panels in the drop down menu (for instance the Human Lung Cancer Panel (UHS-005Z)). The panel type should be set to Multimodal (figure 4). You should see that the "Primer annotation track (DNA)", "Primer annotation track (RNA)", "Target regions track", and "Gene-pseudogene track" items become populated. If this is not the case, check that you downloaded the reference data as shown in figure 3.
3. Replace the "Primer annotation track (DNA)", "Primer annotation track (RNA)" and "Target regions track" items with the files you imported.
4. Add a name for your panel and click **Save**.

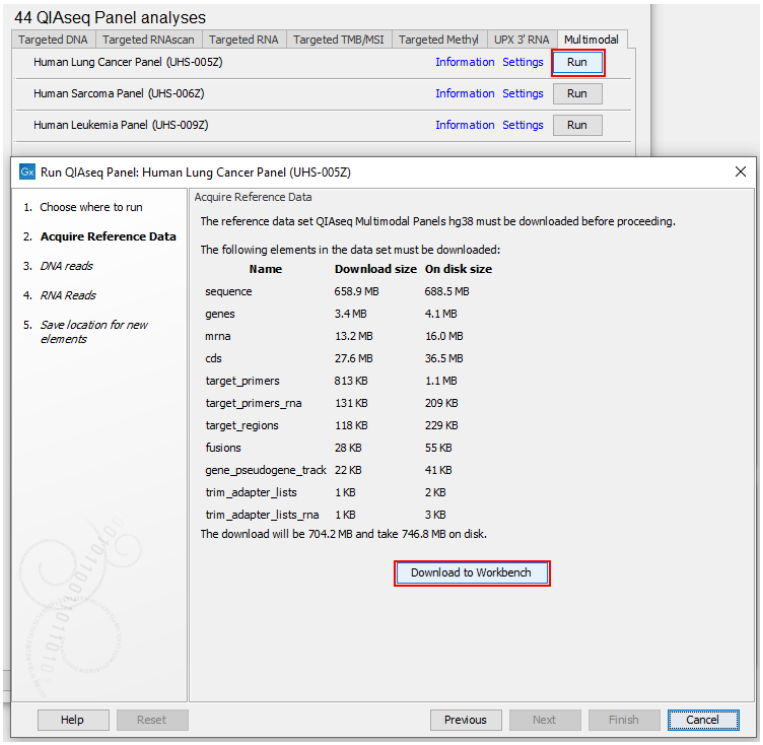


Figure 3: *Download the reference data for the analysis of QIAseq Multimodal panels*

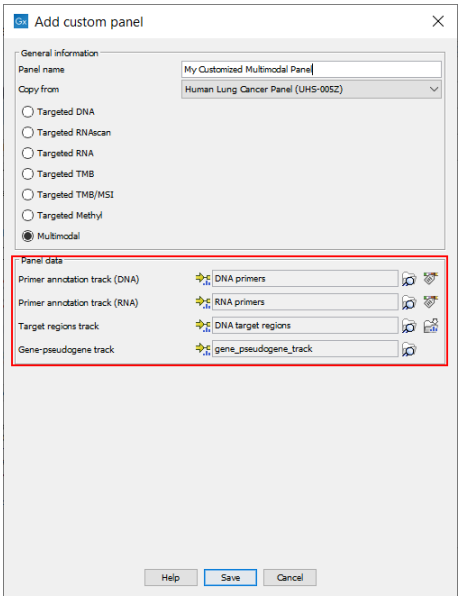


Figure 4: *Configure a custom panel*

Your custom panel has now been added to the list of available panels in the Analyze QIAseq Panels guide.

The steps up to this point have been to configure a custom panel analysis, and they only need to be performed once per panel. From this point on, steps are performed once per sample.

Import the multimodal reads

1. Within the Analyze QIAseq Panels guide, click on **Import Illumina reads**, then the folder icon next to Select files. Select **Add folders** and choose the folder you unzipped on your computer. You should see that 4 files are selected. Set up the importer configuration as seen in figure 5: all parameters remain as they are set by default, except for "Paired-end reads", which needs to be checked.

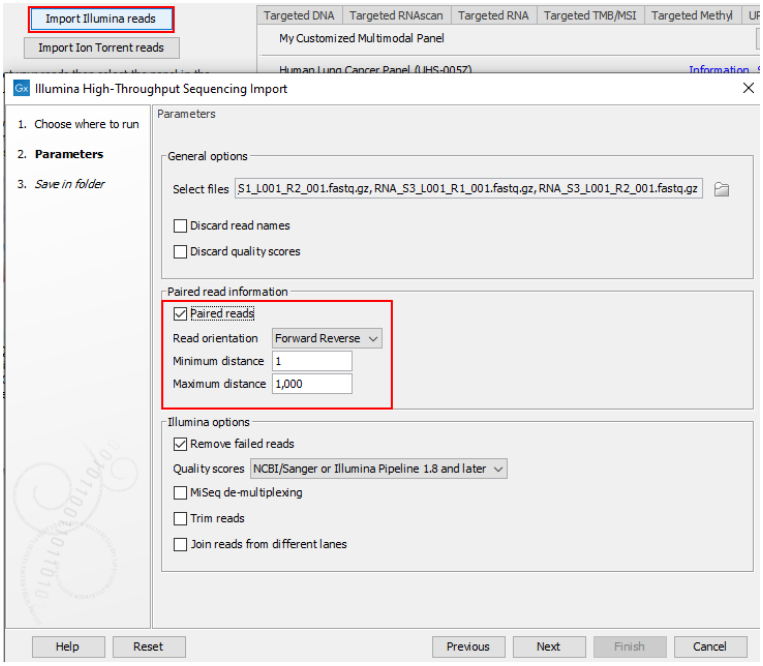


Figure 5: *Import the reads*

2. Click **Next** and choose the folder you wish to save the reads in. Then click **Finish**.

   Each pair of fastq files is imported into a single sequence list that contains both members of each read pair.

Run the analysis

We will now analyze this custom panel using the Analyze QIAseq Panels guide.

1. Click the **Run** button of your custom multimodal panel. If you are connected to a *CLC Genomics Server*, you will be asked where to run the workflow. We recommend that you choose the *CLC Genomics Server* when possible.
2. Select the DNA sequencing reads that you just imported (figure 6) and click **Next**.
3. Now select the RNA sequencing reads that you just imported and click **Next**.
4. In the final wizard step, choose to **Save** the results of the workflow and specify the location to save to before before clicking **Finish**.

This analysis takes approximately 90 minutes on a 2014 MacBook Pro with 16GB of RAM with no other heavy tasks running.

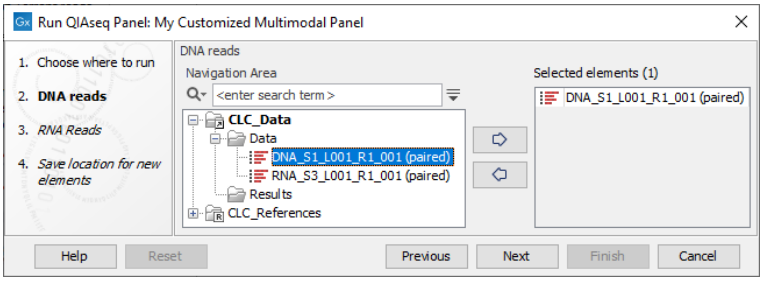


Figure 6: *After clicking Run, select the sequencing reads by double-clicking on the file name or by clicking once on the file name and then on the arrow pointing to the right hand side.*

Output from the QIAseq Multimodal Panel Analysis

The Perform QIAseq Multimodal Analysis (Illumina) workflow produces the following top-level outputs as shown in figure 7:

- A Workflow Result Metadata table keeping track of all generated output.
- A DNA Combined QC Report ( ) summarizing important QC values for the DNA run.
- A RNA Combined QC Report ( ) summarizing important QC values for the RNA run.
- A Gene Expression Track ( ) with gene expression counts
- A Fusion Report ( ) with graphical representations of the fusions found. Double-clicking on a fusion plot in the report will open the plot in a view that allows it to be exported as a high resolution image.
- A Genome Browser View (WT) ( ) containing DNA and RNA read mappings and variant callings.
- A Genome Browser View (Fusions) ( ) containing only the fusion chromosomes, which are used for refining the fusions.

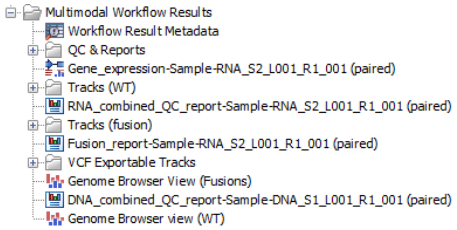


Figure 7: *Output from the Perform QIAseq Multimodal Analysis (Illumina) workflow.*

Four subfolders (QC & Reports, Tracks (WT), Tracks (fusion), and VCF Exportable Tracks) contain outputs that are either referenced by a Genome Browser View, or provide more detail for the top-level outputs.

Reviewing QC Reports First, it is important to check the Combined QC Reports to see if there are problems with the samples. There is one report section for each type of tool in the workflow, and the order of sections matches the order in which the tools are run. One way of reading the DNA Combined QC report might be as follows:

- "Trim Reads | Detailed Trim Results | Trim on Quality": Few reads were trimmed for low read quality
- "Reads mapping summary | Reads summary": Almost all reads mapped to the reference
- "Reads mapping summary | Mapped paired reads": The average fragment size was close to 200nt, with a standard deviation close to 100nt
- "Calculate unique molecular index groups | Group sizes | Median": Most of the reads had only a single UMI. Ideally there would be 2-4 reads per UMI such that most reads could be error-corrected by making a consensus UMI read. Extremely low frequency variant calls on this sample are less likely to be correct.
- "Create UMI reads from grouped reads | Median Q scores for UMI reads": This value is under 44. Stringent filtering in the workflow filters away variant calls that are supported by many consensus UMI reads where the average Q score is <44. This sample may have missing variant calls.
- "QC for targeted sequencing | Fractions of targets with coverage above threshold": Approximately half the target regions have >100x coverage over 90% of the target region.

A similar review of the RNA Combined QC Report should be carried out. The contents of this report are different, as the analysis uses different tools. One cause of concern is that a cell in the "Strand specificity" section is pink, indicating a suspected problem. The warning shows that >90% of the RNA reads mapped sense to annotated transcripts, which might suggest that the data was generated with a strand-specific protocol and should be analyzed accordingly. In the case of Multimodal panels, however, this number is determined by primer design and it is correct to ignore the warning.

Reviewing variant calls

To investigate variant calls, double-click on the Genome Browser View (WT) output. The file opens as a split screen with both Track List and the corresponding variant table. In the table, filter for variants for which the allele is not identical to the reference: open the advanced filter by clicking on the arrow in the top right corner of the table. Fill in the filtering fields with "**Reference allele**", "**=**" and type "**No**" (figure 8).

Clicking on a row in the variant table causes the Genome Browser to zoom in to the nucleotide level with the focus on that variant position. This allows you to view the coverage and the context of this variant in the various tracks, e.g., Mapped UMI Reads, Amino acid track, and so on.

This sample consists of a germline background supplemented by synthetic spike-ins of variants at somatic frequencies. For some regions of the genome, the calls in the germline background have been well-characterized by the GIAB project (Zook et al. 2019). To compare your calls with the background set, right-click within the Track List, choose **Include More Tracks...**, and select the Expected-HG002 track. There are 148 expected background variants. Of these,

Chromo...	Region	Type	Reference	Allele	Referenc...	Length	Zygosity	Count	Coverage	Frequency	Probability
1	12861477	SNV	T	C	No	1	Heterozygous	20	259	7.72	1.
1	12861780	SNV	G	A	No	1	Heterozygous	31	194	15.98	1.
1	36467833	SNV	G	A	No	1	Heterozygous	22	276	7.97	1.
1	36471458	SNV	A	G	No	1	Homozygous	193	193	100.00	1.
1	43349338	SNV	G	T	No	1	Heterozygous	9	110	8.18	1
1	64844806	SNV	T	C	No	1	Heterozygous	113	213	53.05	1.
10	27014676	SNV	A	T	No	1	Heterozygous	32	101	31.68	1.
10	27028911	SNV	C	T	No	1	Heterozygous	82	156	52.56	1.
10	27044193	Deletion	A	-	No	1	Heterozygous	38	67	56.72	1.
10	27046470	SNV	C	T	No	1	Heterozygous	30	68	44.12	1.
10	27048877..27048879	Deletion	CAT	-	No	3	Heterozygous	3	177	1.69	1.
10	27064078	SNV	T	C	No	1	Heterozygous	29	64	45.31	1.
10	27092420	SNV	T	C	No	1	Heterozygous	118	200	59.00	1.

Figure 8: *Genome Browser View after filtering to show only alt alleles. The selected variant is from SeraCare Seraseqï¿½ Myeloid DNA and has an expected frequency of 5%.*

approximately 130 will be present in your calls. The missing calls are mostly filtered out due to the "Median Q scores for UMI reads" being lower than ideal.

Your call set will contain additional variants corresponding to i) the germline background in regions that are not yet well-characterized (outside GIAB high confidence regions), ii) the synthetic spike-ins, and iii) some false positive calls, which are almost exclusively at frequencies <5%.

An example of a synthetic spike-in call is a 28nt duplication in FLT3 with an additional "GCCC" between the duplicated and native sequence. This is represented as a single insertion of length 33nt (figure 9). To find this in your own data, sort the "**Length**" column in the variant table in descending order by clicking on the header of that column

Calls down to a frequency of 0.5% are produced by the workflow, but the lower the frequency, the greater the importance of having several reads per UMI to correct for error. In this sample, there is usually only one read per UMI and so most reads are uncorrected. To find low frequency calls, sort the "**Frequency**" column in the variant table in ascending order by clicking on the header of that column.

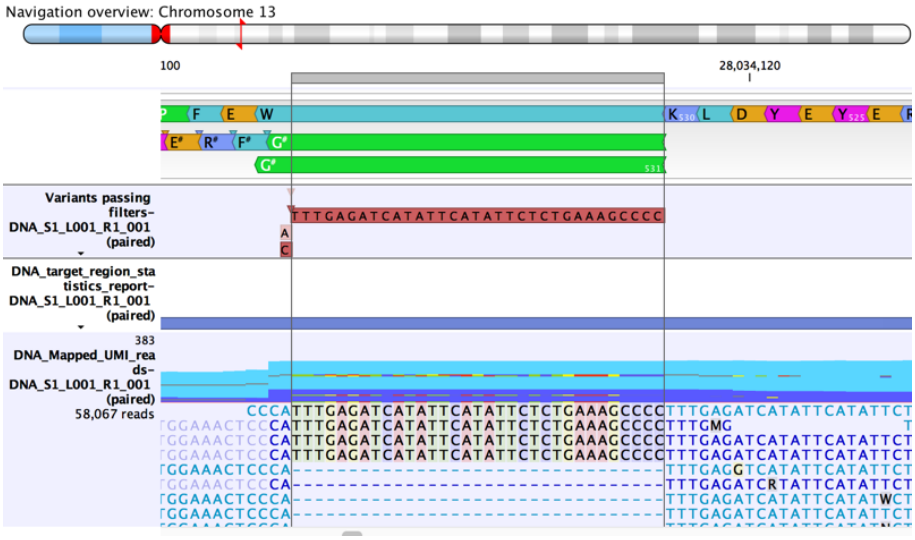


Figure 9: *An example of a synthetic spike-in call: a 28nt duplication in FLT3 with an additional "GCCC" between the duplicated and native sequence.*

Reviewing fusion calls To investigate fusion calls, double-click on the Fusion Report output. The report has a section for each detected fusion that passes quality filters, and the sections are ordered from the most certain to the least certain calls.

Quality control of fusion calls is described in detail here http://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/current/index.php?manual=Interpretation_fusion_results.html

In the report, 9 fusions are reported (figure 10). The first 8 calls are all the fusions expected to be present in this SeraCare sample. The last call, CD74-RNA18SN4, is a false positive. There are several indications that this might be a false positive:

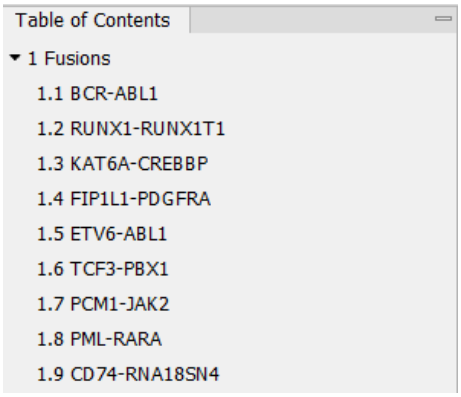


Figure 10: *Contents of the fusion report. Each fusion has its own section.*

- Fusion breakpoints are most commonly observed at exon boundaries, but here most reads extend from the middle of an exon in CD74 to the middle of an exon in RNA18SN4, as shown in figure 11.
- The fusion is the last reported one, and therefore has least evidence among the fusions passing filters.

- CD74 is the most expressed gene in the sample, and RNA18SN4 codes for ribosomal RNA. This gives a possible mechanism for the two being in close spatial proximity during transcription.
- There is no primer for RNA18SN4 in the panel. This is not conclusive, but suggests that the fusion is unexpected.

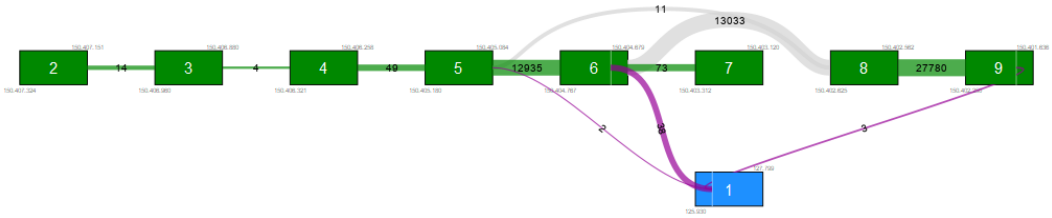


Figure 11: *The CD74-RNA18SN4 fusion is a false positive. Fusion breakpoints are most commonly observed at exon boundaries, but here most fusion reads (purple) have both breakpoints in the middle of exons (white vertical lines).*

(Optional) Fine-tune your analysis

The above analysis used the Analyze QIAseq Panels guide to run the Perform QIAseq Multimodal Analysis (Illumina) workflow with default settings. It is also possible to run the workflow directly, which comes with several advantages:

- The workflow can take fastq.gz files as input, thereby skipping the need to import reads.
- The workflow can process multiple samples at once, using a sample sheet to determine which read files correspond to a sample https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/current/index.php?manual=Running_workflow_in_batch_using_Metadata.html
- The filtering of variants can be fine-tuned.
- Options become available to call CNVs and exon-skippings.

Here we show how to run the workflow with less stringent average Q score filtering. As noted in the section **Reviewing QC Reports**, this sample has lower median Q scores than expected, so the default settings are likely to have filtered away some variant calls where the average Q score is less than 44 when the call is supported by many consensus UMI reads.

To run the Perform QIAseq Multimodal Analysis (Illumina) workflow directly:

> **Ready-to-Use Workflows** | **QIAseq Panel Analysis** () | **QIAseq Analysis workflows** () | **Perform QIAseq Multimodal Analysis (Illumina)** ()

The first two dialogs require the DNA and RNA reads to be selected as before. However, an additional option is now present to "Select files for import: Illumina" (figure 12). If selected, this allows reads to be supplied as fastq files, so that they do not need to be imported before the workflow is run.

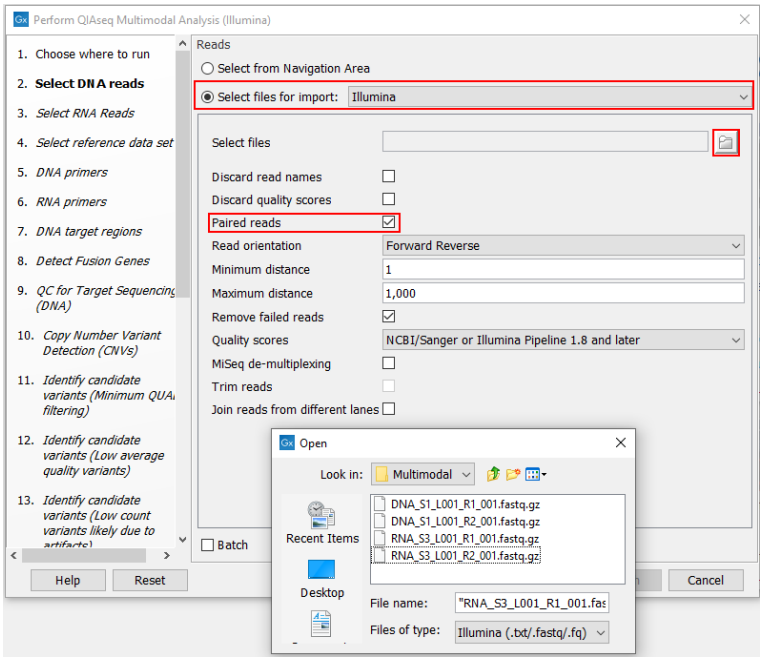


Figure 12: *Running the Perform QIAseq Multimodal Analysis (Illumina) workflow with fastq.gz files as input.*

In the next dialog choose to **Use the default reference data**. This allows the DNA primers, RNA primers, and DNA target regions to be specified in subsequent steps for our custom panel.

The **Detect Fusion Genes** dialog allows exon-skipping events to be detected. By default these are not reported, and so exon-skipping events are not reported when running the workflow from the Analyze QIAseq Panels guide.

The **Copy Number Variant Detection (CNVs)** dialog allows read mappings to be supplied as a baseline for CNV calling. The control mappings should be from normal tissue, sequenced on the same sequencer and panel.

The remaining dialogs control how variants are filtered and annotated. In the **Identify candidate variants (Low average quality variants)** dialog, make the Q score filtering less stringent by reducing the "Average quality" from 44 to 42.

(Optional) Upload your results to QCI Interpret

It is now possible to upload your results to QCI Interpret. For this section, you need an active QCI Interpret subscription (not a trial license), and a QCI Interpret account with API license privileges.

1. To start the upload, open the Analyze QIAseq Panels guide and click on **Upload to QCI Interpret**.

2. In the first dialog (figure 13), select the variant track and fusion track from the "VCF Exportable" subfolder of your results. Then change the reference sequence to hg38. This sequence can be found in the Reference Data tab, under the QIAGEN Active sets, in the QIAseq Multimodal Panels hg38 folder. Finally, choose which application you would like for the QCI Interpret results (Somatic or Hereditary).

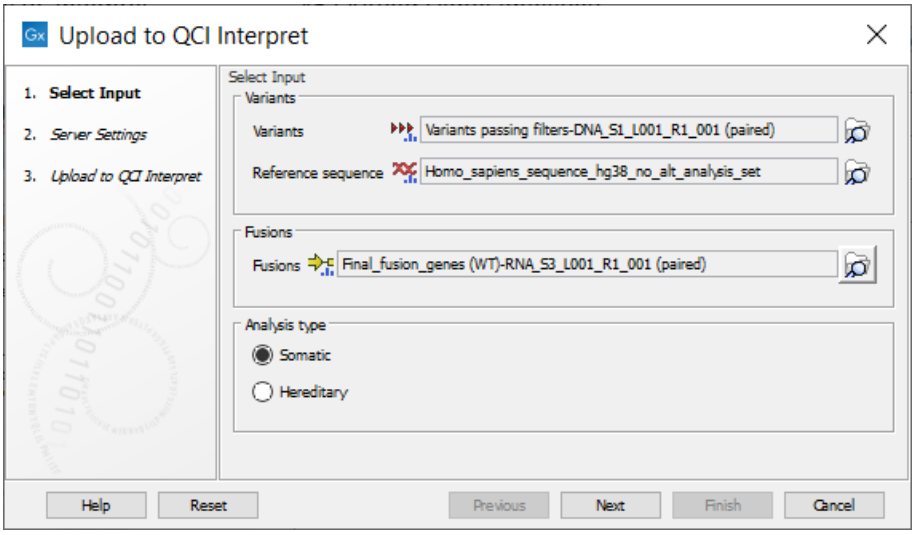


Figure 13: *Select the variants and fusion files you would like to interpret.*

3. In the Server Settings dialog (figure 14), fill in the fields with the following information:

   - Server location. QCI Interpret is hosted in multiple locations. Your QCI Interpret account is created for a specific QCI Interpret server that should be specified here.
   - API key ID and API key secret. If you do not know your API credentials, contact support-license@qiagen.com.
   - Username. Insert the username associated with the QCI Interpret account.

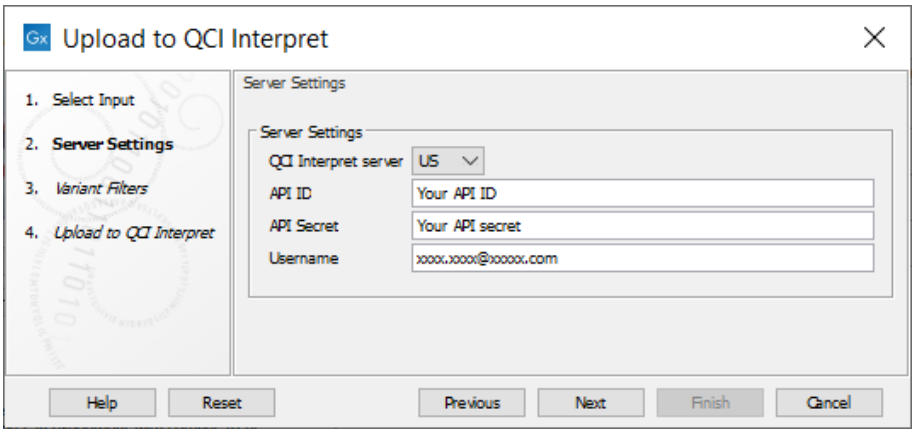


Figure 14: *Configure the connection to the QCI Interpret server dialog.*

4. It is then possible to set up filters to reduce the amount of variants to send to QCI Interpret, if the input variants have not already been filtered (figure 15). In this tutorial, we choose to leave all filters at their default value, i.e. we do not set any filter.

Clicking on **Finish** only initiates the transfer of variants or fusions to QCI Interpret: the tool will open a browser window with the QCI Interpret interview page. This page must be filled in and the button **Submit** must be pressed to complete the upload of the files and metadata to QCI

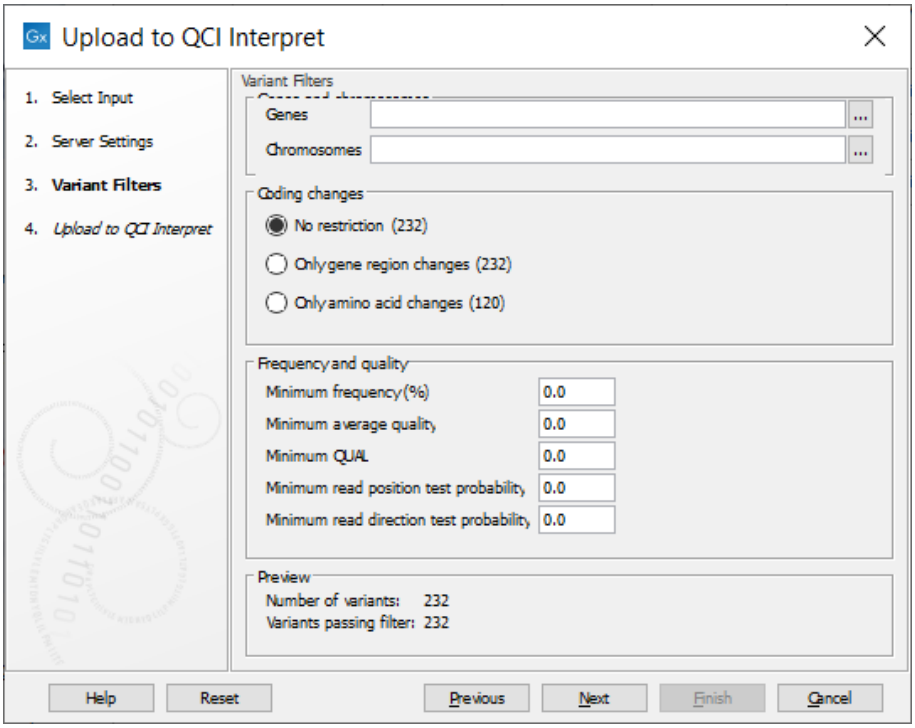
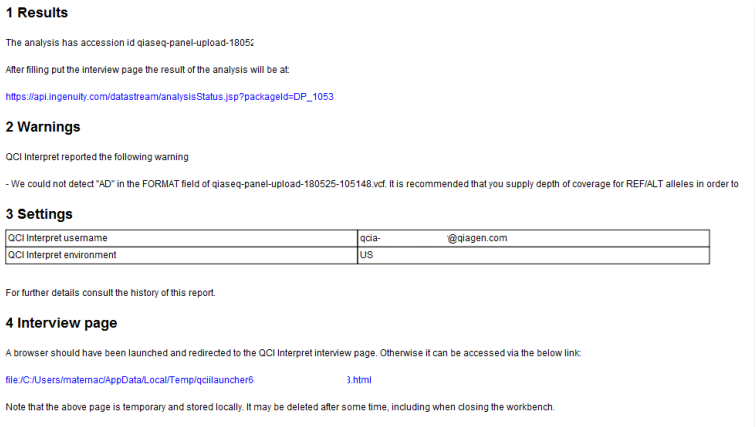


Figure 15: *Filters can be set to reduce the number of variants sent to interpretation.*



1 Results

The analysis has accession id qiaseq-panel-upload-18052

After filling put the interview page the result of the analysis will be at:

https://api.ingenuity.com/datastream/analysisStatus.jsp?packageId=DP_1053

2 Warnings

QCI Interpret reported the following warning

- We could not detect "AD" in the FORMAT field of qiaseq-panel-upload-180525-105148.vcf. It is recommended that you supply depth of coverage for REF/ALT alleles in order to

3 Settings

QCI Interpret username	qcia- @qiagen.com
QCI Interpret environment	US

For further details consult the history of this report.

4 Interview page

A browser should have been launched and redirected to the QCI Interpret interview page. Otherwise it can be accessed via the below link:

file:/C:/Users/maternac/AppData/Local/Temp/qciilauncher6 3.html

Note that the above page is temporary and stored locally. It may be deleted after some time, including when closing the workbench.

Figure 16: *Example of a report containing a link to the QCI Interpret Interview page.*

Interpret. In cases where a browser window fails to open automatically, the tool will also output a report containing the link to the QCI Interpret interview page (figure 16).

Please refer to the QCI Interpret documentation to learn more about the interpretation of the results (figure 17).

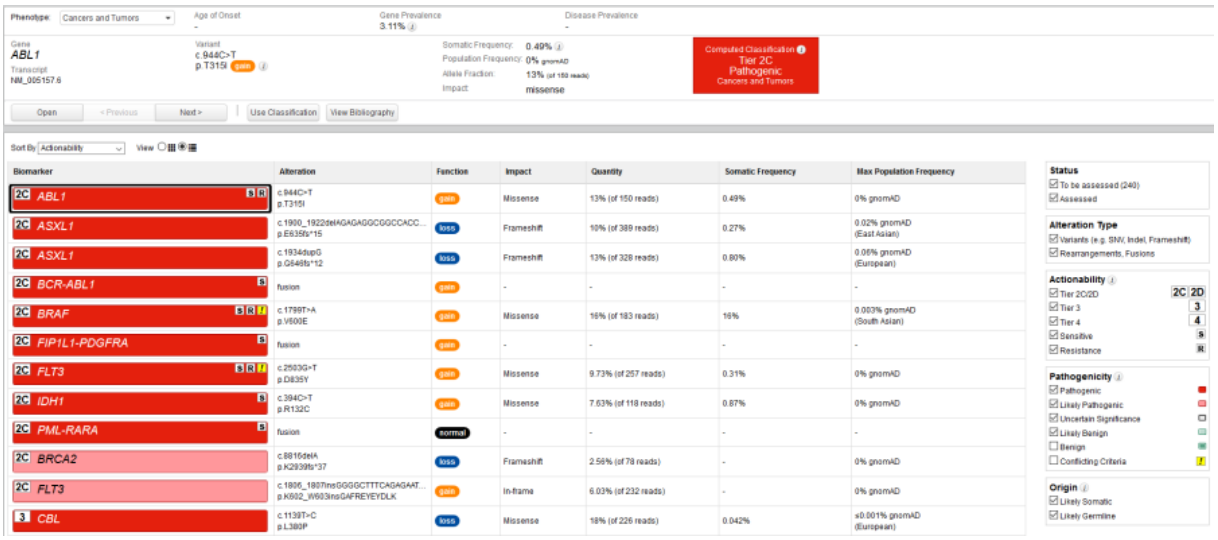


Figure 17: *Results of the interpretation analysis from the QCI Interpret website.*