July 2020

QIAGEN Clinical Insight Analyze for GeneReader 1.7 User Manual

QIAGEN Clinical Insight Analyze for GeneReader is for Research Use Only. Not for use in diagnostic procedures.

188001

QIAGEN GmbH, QIAGEN Strasse 1, D-40724 Hilden

Sample to Insight

QIAGEN®

Contents

Introduction

Thank you for choosing QIAGEN® Clinical Insight (QCI®) Analyze for GeneReader. This manual describes how to operate **QCI Analyze** software. Before using **QCI Analyze** software, it is essential that you read this user manual carefully.

About this manual

This document is the **QCI Analyze for GeneReader 1.7** User Manual for use with CLC Genomics Server Version 11.0.3, and GeneReader™ software version 1.1 or higher.

General information

Technical assistance

At QIAGEN, we pride ourselves on the quality and availability of our technical support.

If you have any questions or experience any difficulties regarding QCI Analyze software or QIAGEN products in general, do not hesitate to contact us. You may contact your regional support office or fill out the web-based contact form at `http://www.qiagen.com/support/technical-support/technical-support-form/`.

Policy statement

It is the policy of QIAGEN to improve products as new techniques and components become available. QIAGEN reserves the right to change the specifications of products at any time. In an effort to produce useful and appropriate documentation, we appreciate your comments on this user manual. Please contact QIAGEN Technical Services with any feedback.

Intended Use Statement

QIAGEN Clinical Insight Analyze is for Research Use Only. It is not intended for use in diagnostic

procedures.

Welcome to QCI Analyze

QCI Analyze is the last component of the GeneReader NGS System Sample to Insight® solution (Figure 1), performing the secondary analysis of FASTQ reads generated by GeneReader and providing a graphical interface to investigate detected variants. QCI Analyze can also export the valid variants as VCF into QIAGEN Clinical Insight (QCI) Interpret for GeneReader for interpretation.

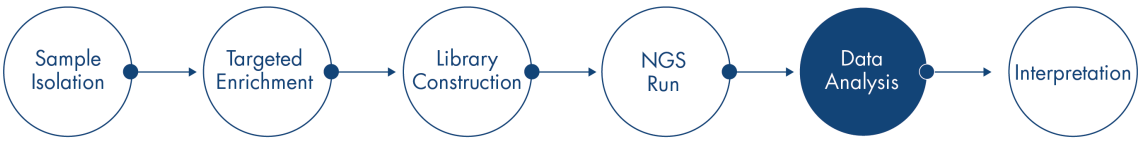


Figure 1: QCI Analyze in the NGS Sample to Insight workflow.

Training for QCI Analyze users

QCI Analyze is intended for use in combination with the GeneReader by professional users trained in molecular biology techniques and in the operation of the GeneReader.

Customers are trained by a QIAGEN representative upon installation of the GeneReader. The training covers general operation of the system, and the subsequent use of QCI Analyze.

QIAGEN can provide further training when necessary, for example after software updates, or for new laboratory personnel. Please contact QIAGEN Technical Services to get more information about retraining.

General Description

QCI Analyze is a browser-based interface that uses the capabilities of the CLC Genomics Server and QIAGEN CLC bio algorithms to analyze Next Generation Sequencing (NGS) data. QCI Analyze is integrated with QIAGEN GeneReader and GeneRead Link™, and offers preinstalled secondary analysis workflows tailored for QIAGEN GeneReader targeted panels available in the QIAGEN product portfolio.

The sample tracking environment and graphical interface minimizes the effort and time required for routine Sample to Insight workflows in a lab. The interactive report allows for efficient and qualified review of reported variants, and user activity tracking on a per sample basis ensures traceability. QCI Analyze also provides a seamless direct upload of reported variants to QCI Interpret for GeneReader.

In configurations without GeneRead Link, QCI Analyze offers a GeneReader Planner application for creating flow cell plans. Depending on configuration, GeneRead Link or GeneReader Planner enables the automatic analysis of sequencing data as soon as it is generated by the sequencing instrument.

QCI Analyze pipeline

QCI Analyze offers functionality and features supporting the full secondary analysis pipeline:

- QCI Analyze can be used to create flow cell plans and communicate them to GeneReader. After sequencing, QCI Analyze automatically retrieves the resulting FASTQ files and initiates analysis.
- Using the appropriate workflow, QCI Analyze performs the secondary analysis of the input FASTQ files and generates a report holding read quality control measures and reported variants that can be exported as VCF, Excel and PDF.
- QCI Analyze graphical interface allows for manual review of the variants in the context of the reference genome and associated annotated tracks.

- QCI Analyze offers to upload the list of valid variants to QCI Interpret for GeneReader for insight into the analysis results.

QCI Analyze secondary analysis workflows

The QCI Analyze workflows provide a streamlined and standardized approach to the analysis of NGS data. The majority of a workflow is automated, ensuring greater standardization and more accurate results. QCI Analyze offers secondary analysis workflows for both variants and copy number variations (CNV) detection in DNA samples, as well as fusion detection in RNA samples.

Variant and CNV detection

QCI Analyze offers the following secondary analysis workflows aimed at detecting variants - and in some cases CNVs - in DNA samples:

- The **AIT Basic FFPE** and **AIT Basic plasma** workflows analyze data generated using the GeneReader **QIAact Actionable Insights Tumor Panel** on FFPE or plasma samples. The workflow detects variants for the genes targeted by the panel.
- The **AIT UMI FFPE** workflow analyzes data generated using the GeneReader **QIAact AIT DNA UMI Panel** on FFPE samples. The panel uses the Unique Molecular Index (UMI) technology. The workflow supports variant detection as well as detection of copy number variations (CNV) for 12 specific genes.
- The **BRCA 1/2 Basic FFPE** workflow analyzes data generated using the GeneReader **QIAact BRCA 1/2 Panel** on FFPE samples. The workflow will report variants for the two BRCA genes.
- The **BRCA UMI FFPE** and **BRCA UMI germline** workflows analyze data generated using the GeneReader **QIAact BRCA Advanced DNA UMI Panel** on FFPE and blood samples respectively. The panel uses the Unique Molecular Index (UMI) technology. The workflows support variant detection in BRCA1, BRCA2, TP53 and PTEN.
- The **Myeloid DNA UMI** workflow analyzes data generated using the GeneReader **QIAact Myeloid DNA UMI Panel** on blood or bone marrow samples. The panel uses the Unique Molecular Index (UMI) technology. The workflow supports variant detection, including low frequency variants as well as large deletions (such as CALR deletions) and insertions (such as FLT3 internal tandem duplications).
- The **Lung DNA UMI FFPE** and **Lung DNA UMI plasma** workflows analyze data generated using the GeneReader **QIAact Lung DNA Panel** on FFPE or plasma samples. The panel

uses the Unique Molecular Index (UMI) technology. The workflows support variant detection as well as detection of copy number variations (CNV) for five specific genes.

- The **Lung Plasma Track** workflow analyzes data generated using the GeneReader **QIAact Lung Plasma Track Panel** on plasma samples. The panel uses the Unique Molecular Index (UMI) technology. The workflow supports detection of a specified list of variants, insertions and deletions, as well as detection of insertions and deletions in general for predefined regions in genes MET and EGFR. The workflow also supports detection of copy number variations (CNVs) for six specific genes.

A QCI Analyze resequencing analysis workflow for variant detection covers the following steps:

- Input reads are aligned to the reference genome, and the first alignment is adjusted based on insertions, deletions and replacements detected in the unaligned ends of mapped reads.
- The Primer regions of the reads are removed from consideration in subsequent analyses to not influence the downstream variant detection.
- The Low Frequency Variant Detection tool detects variants found within the target regions. These are then sorted according to predefined parameters: for example, variants below a range of threshold values (such as a minimum frequency or minimum average quality) are considered false positives and are removed from the results.

In parallel to the variant detection workflow, a Quality Control workflow uses a pre-defined set of thresholds to evaluate the quality of the sequencing data and analysis results.

The workflows for panels using UMI technology follow the same general steps as described above, but are preceded by a series of additional tools meant to leverage the UMI information: Per definition, reads with identical UMIs originate from the same biological fragment. Reads sharing the same UMI are thus merged into a single UMI read for each UMI group, allowing subsequent analyses to be free from artifact introduced by PCR duplication and sequencing bias.

The Myeloid DNA UMI analysis workflow supports detection of insertions that exceed the length of the sequencing reads, specifically internal tandem repeats for gene FLT3. The presence of these variants is inferred based on information on unaligned read ends. These variants are reported as type "insertion*" to differentiate them from regular, shorter insertions that are detected by the variant caller based on gaps introduced in reads during alignment.

Some workflows include CNV detection. The workflows perform an additional coverage analysis on a number of target regions defined for each of the CNV target genes or exons. The observed

coverage is compared to coverage profiles of control samples known to not have any CNVs in the relevant genes. These coverage profiles are embedded in the workflow. Based on this comparison, for each CNV target gene or exon, a fold-change and a p-value is calculated. If these pass the defined thresholds (minimum fold-change absolute value and maximum p-value), the CNV result is indicated as either Amplification and/or Deletion depending on the fold-change value.

Core concepts of variant detection workflows

Variant detection workflows are characterized by specific sets of predefined data:

- Target regions are the regions targeted by the panel.
- Regions of Interest (ROI) are the regions within the targeted regions where sampling depth and amplification efficiency are the least variable, allowing the detection of variants that consistently achieve the performance specifications for the assay at hand, i.e., the ability to detect variants with a certain allele frequency, sensitivity and specificity.

As implied by the definitions above, FFPE and plasma or germline workflows for the same panel share the same target regions but may differ in their predefined ROI. Indeed, the difference in the nature of FFPE and plasma or blood samples has consequences on sampling depth and amplification efficiency, which in turn implies differences in variants that achieve the performance specifications of the assay.

While FFPE and plasma or germline workflows for a same panel contain the same analysis steps, they differ with respect to allele frequency cut-off values since the plasma workflows intend to detect variants at lower allele frequencies while a germline workflow is not reporting on low frequency variants. Find default configuration of the parameters set for calling a variant and for defining which variant will be automatically reported as valid in Appendix A.

Fusion detection

QCI Analyze currently offers one fusion detection analysis workflow:

- The **Lung Fusion** workflow supports secondary analysis of FFPE RNA samples processed with the GeneReader **QIAact Lung Fusion Panel**. The panel uses the Unique Molecular Index (UMI) technology. The workflow supports detection of fusions in genes targeted by the panel.

A secondary analysis workflow for fusion detection includes the following steps:

- Trimming of reads: Reads are trimmed to remove poor quality bases and left-over PCR adapters, the unique molecular index, and the common sequence while retaining the UMI barcode information.
- Readmapping 1: The input reads are aligned to the reference transcriptome sequence.
- Detection of fusion genes: An initial list of fusion genes is generated based on evidence from fusion crossing and fusion spanning UMI reads. The output includes a fusion reference containing a list of fusion genes and transcripts.
- Readmapping 2: The input reads are aligned to a combination of the human reference transcriptome and the newly generated fusion reference. This second mapping enables improved mapping of fusion reads.
- Refinement of fusion genes: Fusions are reported. Values are based on UMI read counts from the second mapping.
- Filtering of fusion genes: Detected fusions are filtered to remove those failing to meet defined thresholds like the promiscuity threshold.
- Sorting: Remaining fusions are filtered into tables 3.1 and 3.2 based on the p-value and number of fusion reads.

In parallel to fusion detection, a Quality Control workflow evaluates the quality of the sequencing data and analysis results based on the detection of fusion control genes and DNA contamination.

The default configuration of the parameters set for detecting fusions are given in Appendix B.

Reference datasets and annotation information

The workflows rely on the presence of particular reference datasets (see Appendix C). This reference data must be downloaded before the workflow can be used. This is usually done during QCI Analyze installation.

User Guide

This chapter will describe the basic steps of working with QCI Analyze. Before going through this tutorial, QCI Analyze has to be properly installed, and connection to GeneReader or GeneRead Link must be configured. Detailed descriptions of the features and items provided by the QCI Analyze interface are found in later chapters.

Assigning samples to flow cells with the GeneReader Planner page

Note: For installations configured with GeneRead Link, flow cell plans are created in GeneRead Link and analyses are started from GeneRead Link, not QCI Analyze. GeneRead Link users can proceed to the next sections of this chapter.

1. Go to the **GeneReader Planner** page.
2. Click on **Add new flow cell** to open a new flow cell field.
3. Name the new flow cell.
4. Click on **Add sample** to open a row with several fields describing the samples to be processed in the flow cell (Figure 2).

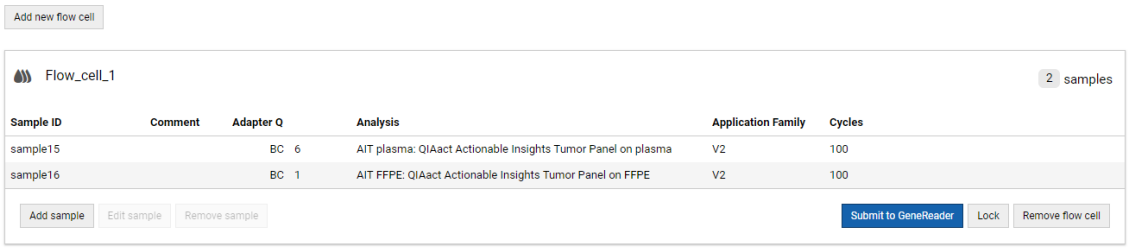


Figure 2: Creating a flow cell and adding samples.

5. Manually type in the information, or drop on the flow cell window a text file containing the sample information, including at least the Sample ID, the barcode used during sample preparation (Adapter Q), and the Analysis workflow (see the GeneReader Planner chapter for more detailed information.)
6. When entering data manually, click **Save and new** to add another sample.

7. Click **Save** when all samples have been added to the flow cell.
8. When the flow cell is complete, **Lock** it to ensure that it cannot be modified before sequencing is actually started. From the Lock dialog, it is possible to **Print** the flow cell plan to help during library preparation.
9. When the samples are ready for sequencing, **Unlock** the flow cell plan if it was previously locked, and click **Submit to GeneReader** to send it from the GeneReader Planner to GeneReader.
10. On the computer hosting the GeneReader software, accept the flow cell plan and start the sequencing.

When sequencing has completed on GeneReader, generated FASTQ files will be transferred to QCI Analyze and analysis will start automatically with the analysis workflow previously specified for each sample.

Reviewing a variant detection report

1. Go to the Analyses page.
2. Select the sample analysis of interest in the list. Inspect the **Sample analysis details** panel for a summary of Quality control and Analysis results (Figure 3).

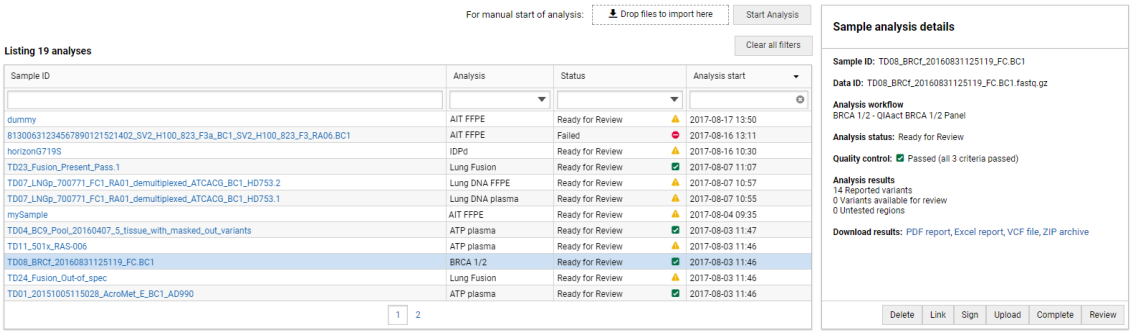


Figure 3: The sample analysis details panel contains a summary of quality control and analysis results.

- "Quality control" indicates how successful the sequencing was. This can be either yellow, *Deviations*, or green, *Passed*. If yellow, a description of the criteria for which deviations are observed is indicated.
- "Analysis results" shows an overview of the analysis results.

3. Click on the Sample ID hyperlink or the **Review** action available at the bottom of the **Sample analysis details** panel to access the results for a particular sample.

4. When the "Quality control" of the **Sample analysis details** panel is showing deviations, go to the Quality Control section of the report. For failure root cause analysis, check which of the colored statuses in the Input data summary and Secondary analysis summary sections is marked in yellow (Figure 4). Pass criteria are indicated below the tables.

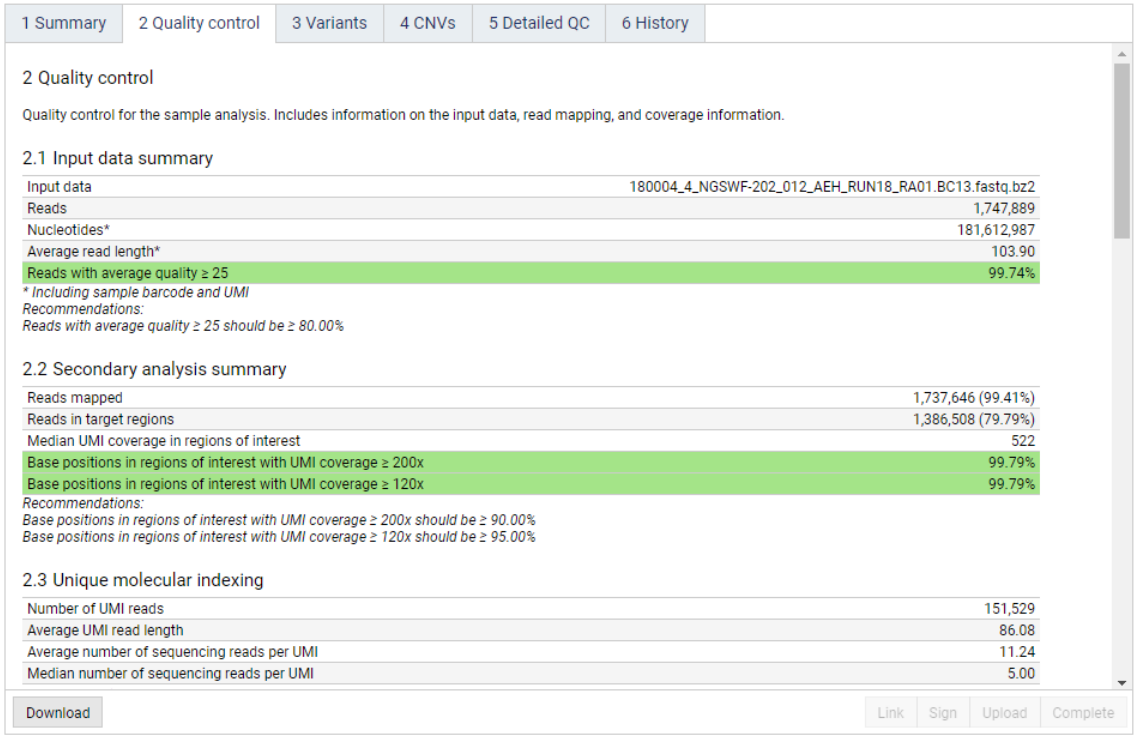
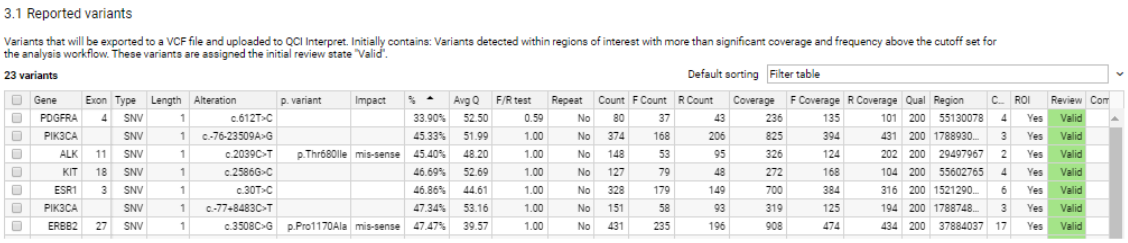


Figure 4: Quality Control section of a variant detection report.

5. The actual result of the analysis is found in the section **Variants** and, for specific panels, **CNVs**. The Variants section lists the variants present in the dataset. They are classified in two tables:

   - The **Reported variants** table (table 3.1) offers a list of the variants detected within regions of interest and with read coverage above the threshold set for "Significant coverage" (Figure 5). The state of these variants is automatically set to Valid.



Gene	Exon	Type	Length	Alteration	p. variant	Impact	%	Avg Q	F/R test	Repeat	Count	F Count	R Count	Coverage	F Coverage	R Coverage	Qual	Region	C..	ROI	Review	Com
PDGFRA	4	SNV	1	c.612T>C			33.90%	52.50	0.59	No	80	37	43	236	135	101	200	55130078	4	Yes	Valid	
PIK3CA		SNV	1	c.-76-23509A>G			45.33%	51.99	1.00	No	374	168	206	825	394	431	200	1788930...	3	Yes	Valid	
ALK	11	SNV	1	c.2039C>T	p.Thr680Ile	mis-sense	45.40%	48.20	1.00	No	148	53	95	326	124	202	200	29497967	2	Yes	Valid	
KIT	18	SNV	1	c.2586G>C			46.69%	52.69	1.00	No	127	79	48	272	168	104	200	55602765	4	Yes	Valid	
ESR1	3	SNV	1	c.30T>C			46.86%	44.61	1.00	No	328	179	149	700	384	316	200	1521290...	6	Yes	Valid	
PIK3CA		SNV	1	c.-77+8483C>T			47.34%	53.16	1.00	No	151	58	93	319	125	194	200	1788748...	3	Yes	Valid	
ERBB2	27	SNV	1	c.3508C>G	p.Pro1170Ala	mis-sense	47.47%	39.57	1.00	No	431	235	196	908	474	434	200	37884037	17	Yes	Valid	

Figure 5: The Reported variants table.

   - The **Variants available for review** table (table 3.2) includes variants detected in regions below significant read coverage, but above specified minimum coverage (Figure 6). Depending on the workflow configuration, the table may also include variants outside

of the Regions of Interest (ROI), including those with coverage above the significant coverage threshold. Users can review the variants of this table one after the other, using the data available from the Track Viewer to assess whether the variant is indeed a true variant or an artifact (see the Note on assessing variants).

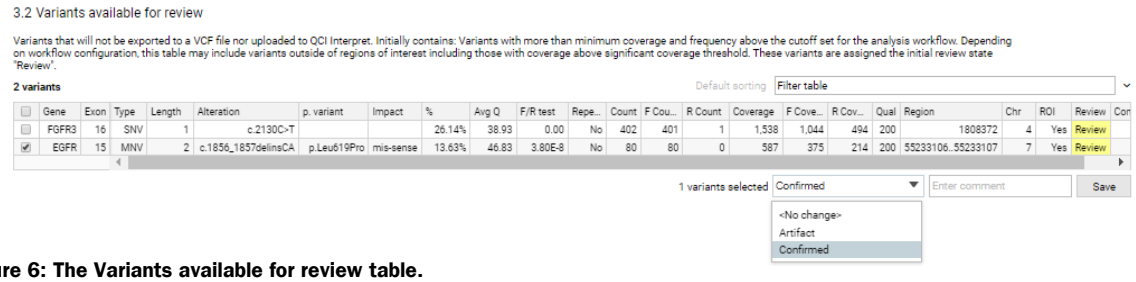


Figure 6: The Variants available for review table.

The user can manually validate one or more variants as "Confirmed" or dismiss them as an "Artifact". This is done by selecting the variants via the check-box available to the left in the table, and specifying the manual review conclusion in the drop-down menu below the table. "Artifacts" will remain in table 3.2, while "Confirmed" variants will be moved to table 3.1 Reported variants.

6. The outcome of the copy number variation detection is presented in the CNVs section of the report. For each CNV target gene, the table will indicate if a CNV was detected. CNV tables cannot be edited. For analysis workflows with exon-level CNV detection, only exons where a CNV is detected will be listed.

7. Once the analysis results have been reviewed, the user can sign the report with a comment. It is optional to sign a report, and any number of users can sign the same report.

Note on assessing variants

Some variants reported in the **Variants** section of the report may be the result of various types of artifacts derived from earlier steps of a NGS workflow: (1) random sequencing errors; (2) sequencing context dependent errors, resulting in forward reads carrying different bases than reverse reads; (3) PCR errors occurring in the early rounds of amplification; and (4) adapter/primer/ligation artifacts.

The QCI Analyze resequencing workflows aim to remove the variants introduced by these artifacts, but depending on the sensitivity required for such analyses, some artifact variants may still be reported.

To allow the user to assess the soundness of the reported variants, these are accompanied by a number of variant annotations. All annotations are displayed as a tooltip that appears when

hovering on the variants in the track viewer (Figure 7).

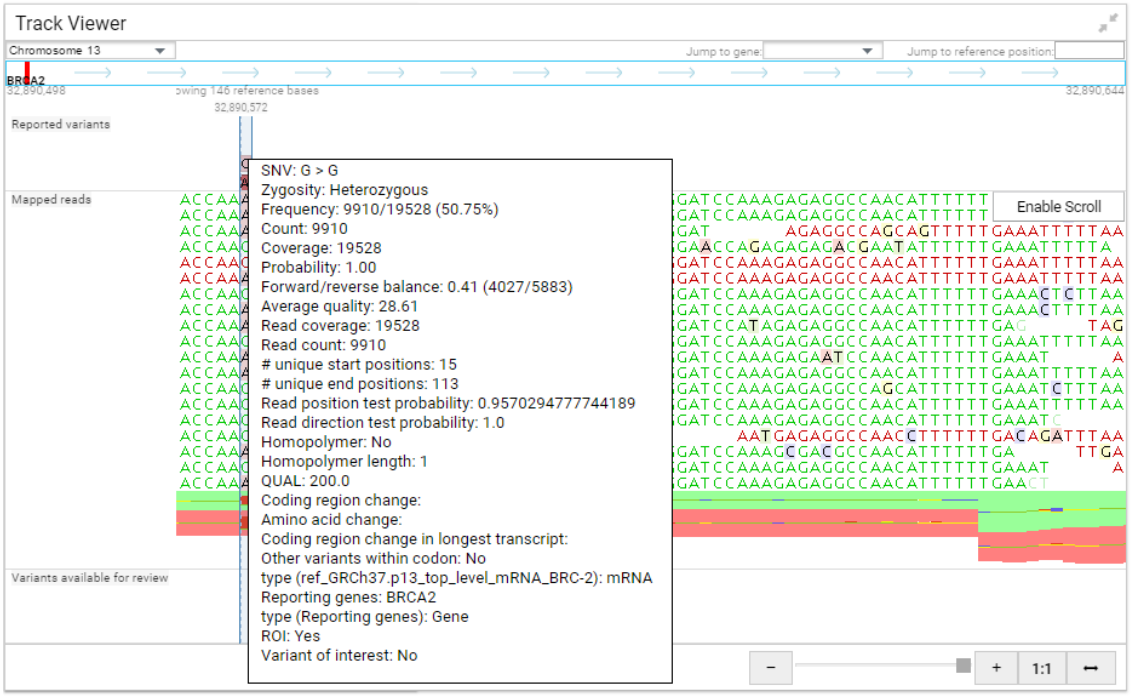


Figure 7: Annotations of a variant seen as a tooltip in the Track Viewer.

Generally speaking,

- A variant should have a high Qual score, 200 being the maximum. This value reflects the significance of a variant. The higher the score, the less likely it is that the variant is a sequencing error.
- The F/R Test reflects the relative forward/reverse read distribution, i.e., if the forward/reverse ratio for reads that support the variant (counts) is similar to the forward/reverse ratio of the total number of reads covering the position (coverage). The value ranges from 0 to 1, where values closer to 0 may indicate a sequencing artifact. With very high coverage, even small relative differences in allele frequencies (which are most pronounced for low frequency variants) may lead to an F/R test value that is 0 or very close to 0. Hence, in situations where the coverage is high and the frequency low, the value should not be used for hard filtering, but rather as a guide towards variants that deserve closer inspection.

For selected analysis workflows, additional guidance on variant assessment is given in the specific sections of Appendix A.

Reviewing a fusion detection report

1. Go to the **Analyses** page.

2. Select the sample analysis of interest in the list. Inspect the **Sample analysis details** panel for a summary of Quality control and Analysis results (Figure 8).

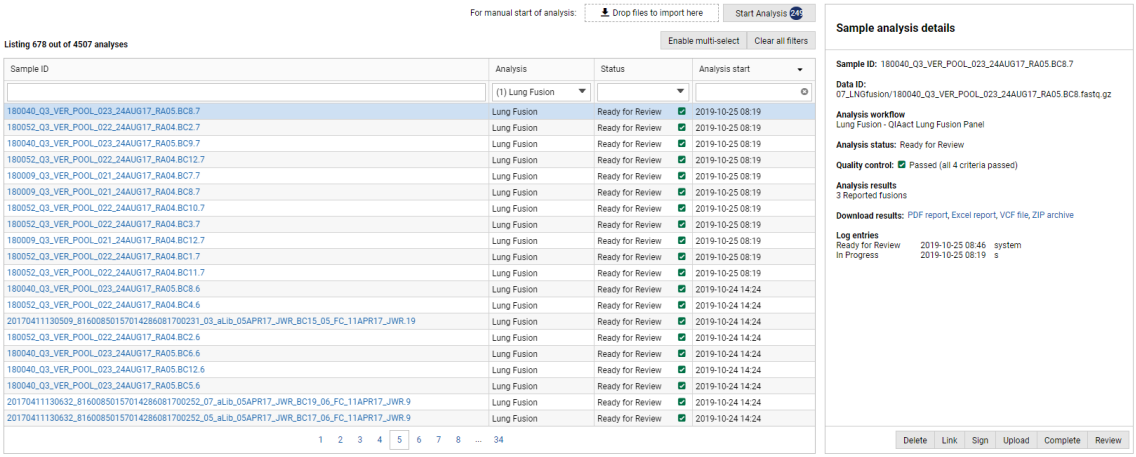


Figure 8: The sample analysis details panel contains a summary of quality control and analysis results.

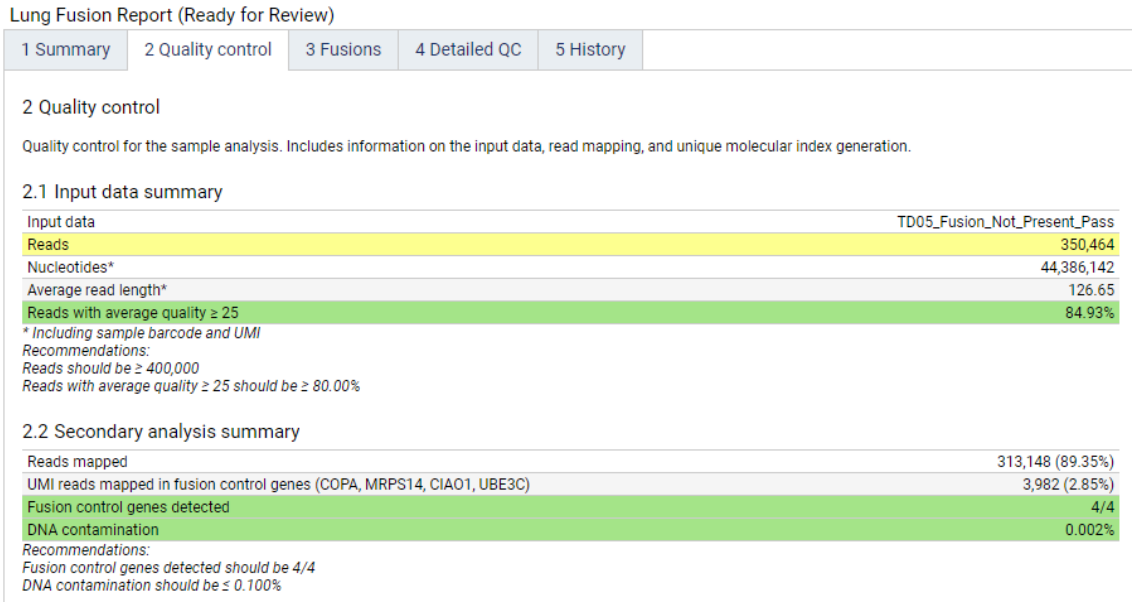
   - "Quality control" indicates how successful the sequencing was. This can be either yellow, *Deviations*, or green, *Passed*. If yellow, a description of the criteria for which deviations are observed is indicated.
   - "Analysis results" shows how many fusions were reported.

3. Click on the Sample ID hyperlink or the **Review** action available at the bottom of the **Sample analysis details** panel to access the results for a particular sample.

4. When the "Quality control" of the **Sample analysis details** panel is showing deviations, go to the Quality Control section of the report. For failure root cause analysis, check which of the colored statuses in the Input data summary and Secondary analysis summary sections is marked in yellow (Figure 4). Pass criteria are indicated below the tables. (Figure 9).

5. The actual result of the analysis are found in the section **Fusions**. This section lists the fusions present in the data set. Click on the fusion events of interest to visualize them in the Track Viewer on the right.

6. Once the analysis results have been reviewed, the user can sign the report with a comment. This step is optional, and any number of users can sign the same report.


Lung Fusion Report (Ready for Review)

1 Summary | 2 Quality control | 3 Fusions | 4 Detailed QC | 5 History

2 Quality control

Quality control for the sample analysis. Includes information on the input data, read mapping, and unique molecular index generation.

2.1 Input data summary

Input data	TD05_Fusion_Not_Present_Pass
Reads	350,464
Nucleotides*	44,386,142
Average read length*	126.65
Reads with average quality ≥ 25	84.93%

** Including sample barcode and UMI*
Recommendations:
Reads should be ≥ 400,000
Reads with average quality ≥ 25 should be ≥ 80.00%

2.2 Secondary analysis summary

Reads mapped	313,148 (89.35%)
UMI reads mapped in fusion control genes (COPA, MRPS14, CIAO1, UBE3C)	3,982 (2.85%)
Fusion control genes detected	4/4
DNA contamination	0.002%

Recommendations:
Fusion control genes detected should be 4/4
DNA contamination should be ≤ 0.100%

Figure 9: Quality Control section of a fusion detection report.

Uploading samples to QCI Interpret for GeneReader (optional)

QCI Interpret currently offers three different analysis pipelines: somatic, hereditary and hereditary genome. Which pipeline to use for the upload is specified for the analysis workflow in QCI Analyze. Table 3.1 variants, CNVs, and fusions are sent to QCI Interpret in a VCF file.

1. To upload the variants of table 3.1, CNVs when applicable, and/or the fusions to QCI Interpret for GeneReader, click **Upload**. A prompt for QCI Interpret login credentials will appear the first time a new QCI Analyze user clicks **Upload**. The credentials are saved and used for all subsequent uploads. If QCI Analyze users need to change their QCI Interpret login, the QCI Analyze administrator can use the action **Remove Interpret Access** (see the User Accounts section of this manual).

2. For uploads to the QCI Interpret somatic and hereditary genome pipeline, the QCI Interpret uploader will open in a separate browser tab. If this does not happen automatically, check that your browser allows pop-ups, and that the QCI Interpret connection is configured appropriately as described in the QCI Interpretation configuration section of this manual.

   For uploads to the QCI Interpret hereditary genome pipeline, users can use the "QCI Interpret test" link or **Go to** button to get to the QCI Interpret Hereditary genome upload page. From here samples can be associated and QCI Interpret analysis initiated.

3. In the QCI Interpret uploader, metadata can be associated with the analysis results. If a test product profile has been specified for the analysis workflow (see the QCI Interpretation configuration section of this manual), the fields will be pre-filled.

4. Go through the interview pages and finally click **Submit**. The metadata is uploaded the QCI Interpret together with the QCI Analyze results.

5. When ready, the resulting QCI Interpret test will be displayed.

Once upload to QCI Interpret for GeneReader has been initialized from QCI Analyze, the status of the sample analysis in QCI Analyze is set to Uploaded. The Sample analysis details panel will provide a link to the QCI Interpret test and display the status of that QCI Interpret test. Users can also access the QCI Interpret test by clicking **Go to** from the report (Review page). If the upload was not fully completed by clicking **Submit** on the last of the QCI Interpret uploader interview pages, the link to QCI Interpret results test will open with an error message.

For information on how to use QCI Interpret for GeneReader, please refer to the QCI Interpret for GeneReader user documentation.

Completing sample and saving analysis results

When the review of a QCI Analyze sample is done, click **Complete** on the report or from the **Sample analysis details** panel on the **Analyses** page. The sample analysis now has the status Completed, which is a read-only mode: editing variant assessment or uploading results to QCI Interpret for GeneReader is no longer possible.

For configurations with GeneRead Link, result files (VCF file and PDF report) for Completed samples will be retrieved by GeneRead Link and the sample moved to the approval section in GeneRead Link.

If automatic archiving is configured in the **Administrator** page, the final result files (VCF file, Excel and PDF report) will be zipped in a file and copied to the storage location specified for the archiving after the specified number of days, and the sample analysis removed from QCI Analyze. **Note**: the BAM and BAI file will not be automatically archived.

The result files can also be manually downloaded via the **Download results** links in the **Sample analysis details panel** (in the Analyses page) or via the "Download" options (PDF report, Excel report, VCF file in case of variant detection reports, ZIP archive as described above, and the read mapping as a BAM and corresponding BAI file) available at the bottom of the report in the **Review** page.

User Interface

The QCI Analyze web interface includes a navigation bar that allows users to easily switch between the different pages of the user interface: **GeneReader Planner**, **Analyses**, **Review** and **Comparison** (Figure 10).

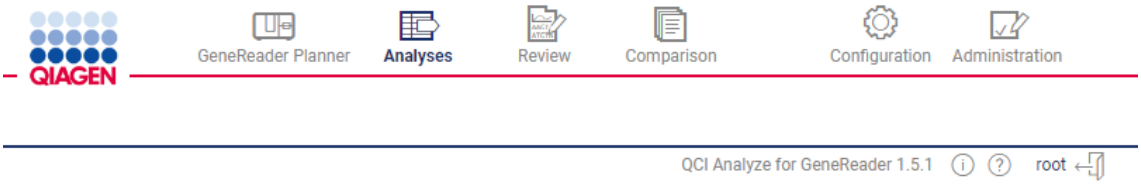


Figure 10: The navigation bar sits at the top of the user interface. Additional general information and icons are available at the bottom of the user interface.

Users with administrator privileges have two additional buttons in the navigation bar, **Configuration** and **Administration**, that provide access to admin-only functionalities. In the bottom right corner of the user interface, copyright and license information can be accessed by clicking the (ⓘ) icon. Click on the (?) to read the QCI Analyze manual and contact QIAGEN support team. The name of the user currently logged in is written to the left of the log out icon. Click on the username to change the password associated with that user. Finally, click on the (⇦) icon to log out of QCI Analyze.

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

GeneReader Planner

This page is used in configurations without GeneRead Link to upload sample information and create a flow cell plan for sequencing with GeneReader (Figure 11). When adding samples to the flow cell plan, the user will also select the workflow to be used for secondary analysis once sequencing is completed.

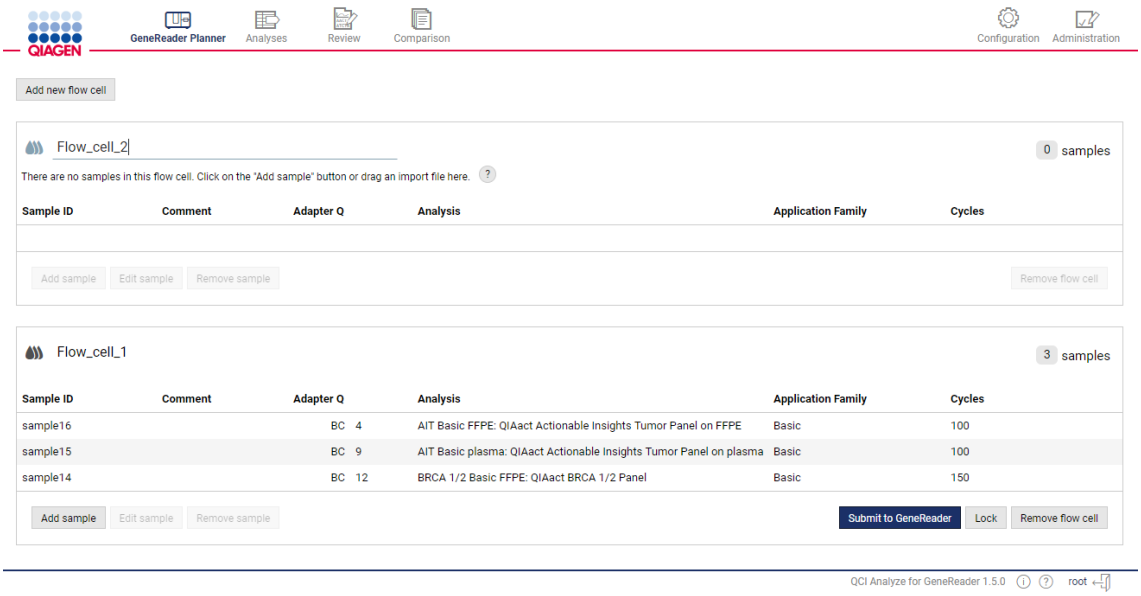


Figure 11: The GeneReader Planner page.

Note: While GeneReader and QCI Analyze exchange information dynamically to ensure that analyses start automatically when sequencing is complete, the sequencing itself does not start automatically after exporting the flow cell plan to GeneReader. It needs to be started by clicking the **Start** button in the GeneReader software.

Create a flow cell plan

Add new flow cell will open a new Flow cell field. Multiple flow cell plans can be under construction at the same time. Before sample information can be added to a flow cell plan, the name of the flow cell plan should be filled in. The name can be at most 35 characters long. A

flow cell cannot combine different application families and a message will warn users of potential incompatibilities.

Add sample information

Add sample opens a row with several fields describing sample characteristics (Figure 12).

Figure 12: Adding a sample.

There are two ways to upload sample data: the user can either manually type in the information or import a text file containing the sample information. Required input is indicated by an asterisk in the following lists.

Add samples manually

- Sample ID*. Type in a unique sample name.
- Comment.
- Adapter Q*. Can be BC or 9mBC depending on the panel used. This information is automatically filled out when choosing the analysis workflow. However, users need to choose the barcode number from the drop-down used during sample preparation.
- Analysis*. Specify which workflow should be used for secondary analysis to allow the analysis to start directly after the sequenced reads are transferred from GeneReader to QCI Analyze.
- Application family. Basic or UMI depending on the Panel used. This field is automatically filled out when choosing the secondary analysis workflow.
- Cycles. This field is automatically filled out when choosing the secondary analysis workflow.

Click **Save and new** to add another sample, or **Save** to stop editing the flow cell. Invalid values are highlighted. Hovering the mouse over an invalid field will display a description of the error. Invalid values prevent the sample from being saved.

Import samples as a text file

Another way to create a flow cell plan is to drag and drop a text file (*.csv or *.txt) into the relevant flow cell field. This file must contain the samples' IDs and additional information organized in

four columns without headers. The columns/values must be in the following order: Sample ID, Comment, Adapter Q, Analysis.

- Sample ID must be unique.
- Comment is optional (see the example below).
- Adapter Q must follow the format 1, 2, etc.
- Analysis can be designated by the full workflow name (for example AIT Basic FFPE: QIAact Actionable Insights Tumor Panel on FFPE), or just the workflow tag (AIT Basic FFPE). Workflow tags are not case-sensitive.

In the following example, there is a comment for the sample "12345" but none for "56789":

```
12345,my_comment,1,AIT Basic FFPE

56789,,2,AIT Basic plasma
```

Values can be separated by commas, semicolons or are tab delimited. Column separators are automatically detected when importing a valid file (Figure 13).

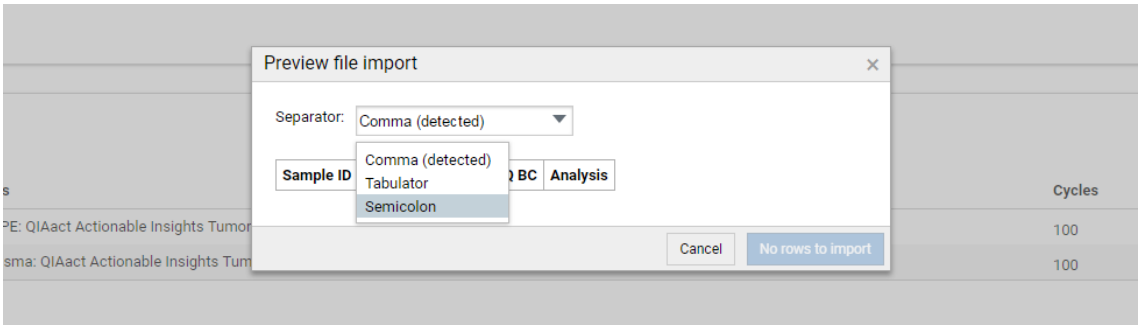


Figure 13: Preview file import dialog upon importing a TXT file.

Errors will be indicated with a pink highlight in the preview dialog (Figure 14).

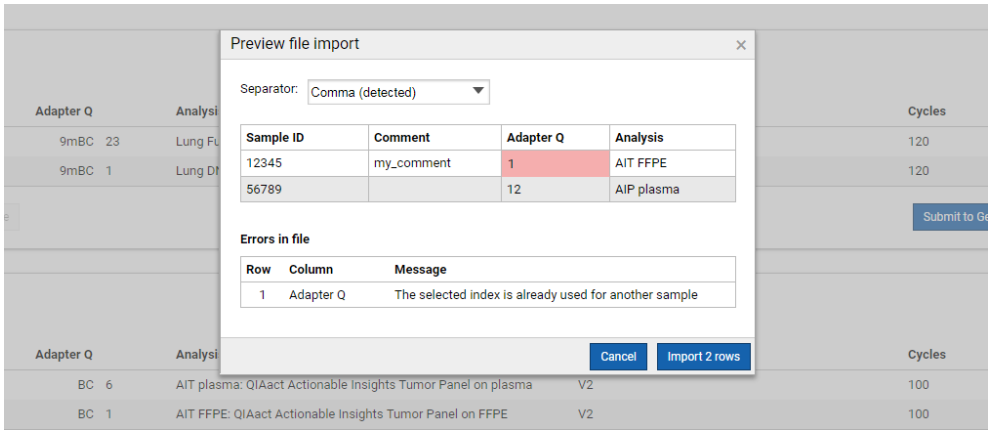


Figure 14: Preview file import dialog in case of an error in the file imported.

Edit flow cell plans

For each added sample, the following actions are available:

- Edit sample. When the Edit action is selected, the sample information, except for Sample ID, can be edited. Only one sample can be in edit mode at a time.
- Reassign via drag and drop. If information about a sample was misplaced on the flow cell plan, the sample can be "drag-and-dropped" to another available flow cell plan.
- Remove sample. When the user selects the Remove sample option, a dialog requests the user to OK or Cancel the removal. Removing a sample is non-reversible.

Lock a flow cell

When the flow cell is completed and valid, it is possible to Lock the flow cell plan to indicate to other users that it should not be altered, and that library preparation is ongoing (Figure 15). When locked, it is possible to Print the flow cell plan to generate a worksheet that can be used for reference when adding molecular indices to samples during samples preparation, or when loading samples to flow cells for sequencing.

Figure 15: Locking a flow cell.

Export a flow cell plan to GeneReader

When the samples are ready for sequencing, **Unlock** the flow cell plan if it was previously locked, and **Submit to GeneReader** to export it to GeneReader. The flowcell plan disappears from the

GeneReader Planner page in QCI Analyze, and the user needs to go to the computer hosting the GeneReader software to accept the flow cell plan and start the sequencing.

The samples included in the submitted flow cell plan are now visible on the Analyses page as "Planned". When sequencing has completed on GeneReader, generated FASTQ files will be transferred back to QCI Analyze and analysis will start automatically with the analysis workflow previously specified for each sample.

Remove a flow cell plan

The **Remove flow cell** option deletes both the flow cell plan and the assigned samples from QCI Analyze. When the user selects the **Remove flow cell** option, a dialog requests the user to **OK** or **Cancel** the removal. Removing a flow cell is non-reversible.

Analyses

The Analyses page lists all sample analyses in QCI Analyze (Figure 16).

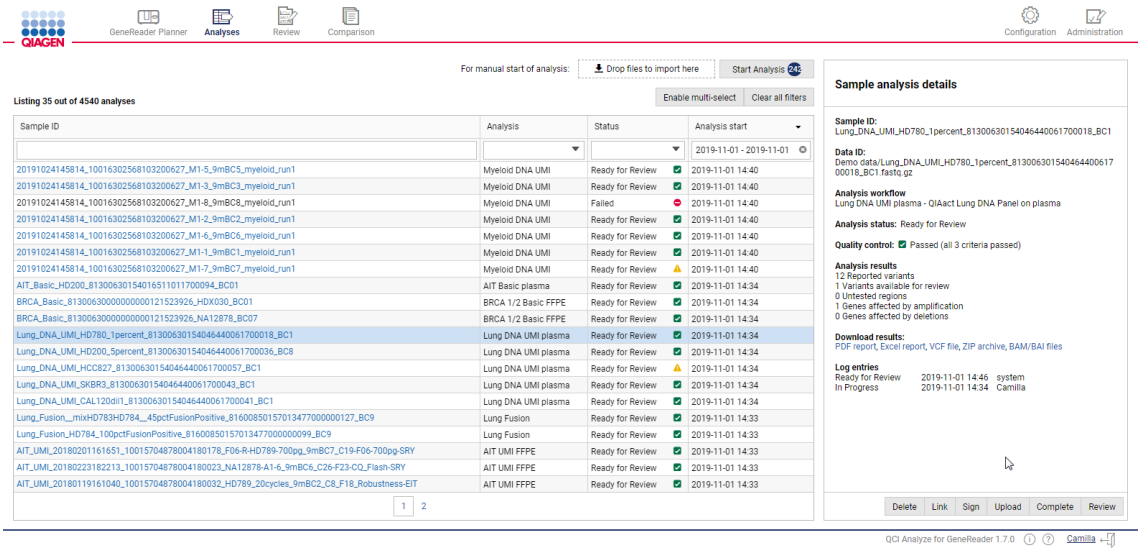


Figure 16: The Analyses page. The **Delete** button at the bottom of the Sample analysis details panel is only available to administrators.

By default, the analyses are sorted by "Analysis start" with the most recent time stamp at the top of the list. For Planned samples, the time of the flow cell plan submission to GeneReader is listed until the analysis in QCI Analyze starts. Click one of the table headers to sort the analyses based on the values in the given column. Sorting on "Status" also takes quality control status into account, placing those with quality control deviations at the top. Click twice to reverse the order. The list can also be sorted based on values in multiple columns by holding down Shift while clicking multiple column headers.

Just below the table headers, filters to apply to the analyses list are available. Depending on the column, the filtering is based on a text string, a check box selection of values, or calendars to define a time span. To the top left of the table, a number indicates how many sample analyses match the current filter. To the top right of the table, a **Clear all filters** button removes all applied filters and reestablishes the full list of sample analyses.

A sample analysis can be selected by clicking a table row. This populates the **Sample analysis**

details panel to the right of the table. Depending on the status of the analysis, different actions and indications are available via this panel.

The Sample ID in the analysis list also serves as a shortcut to open the sample analysis in **Review**.

Note: A report opened for review in another session, i.e., by a different user on a different computer, is indicated by a blue bar to the far right of the sample ID field in the Analyses overview table. In addition, the Sample analysis details panel displays the user name(s) of the current reviewer(s).

Clicking the **Enable multi-select** button to the top right of the Analyses overview will introduce a multi-select column to the far left of the Analyses overview table. When two or more sample analyses are selected, the Sample analysis details panel will display a list of selected sample analyses and users can apply the same action to all selected samples at once. Which actions are available for batch processing will depend on the status of the sample:

- In progress. The action available is **Stop**.
- Ready for Review. The actions available are **Delete**, **Upload Hereditary Genome**, **Complete** and **Sign**.
- Uploaded. The actions available are **Delete**, **Upload Hereditary Genome**, **Complete** and **Sign**.
- Completed. The actions available are **Download** and **Delete**.
- Failed. The actions available are **Rerun** and **Clear**.

The action **Delete** is only available to users with administrator privileges. The action **Rerun** will only be enabled if the selected samples were initially processed with the same analysis workflow.

If you select sample analyses with different states, only actions common to selected states will be presented.

Status of a sample analysis

Planned

Samples submitted to GeneReader using the GeneReader Planner are listed as Planned in the analyses list. Such samples remain in the Planned state until sequencing on the GeneReader is complete and data is available.

The time stamp of a Planned sample represents the time of flow cell plan submission to GeneReader. When data becomes available and the state changes to In Progress, the time stamp will be updated to be the time of analysis start.

It is possible to **Stop** a Planned analysis. This will not stop the sequencing process, but will remove the sample from the analysis queue and from the list of sample analyses. If the sequencing is later successfully completed, an analysis can be started manually as described in the Manual start of an analysis section of this manual.

In Progress

The status In Progress is given to sample analyses currently being analyzed or queuing to be analyzed. The status is indicated in the Analyses list with a progress bar and a percentage indicating the progress of the analysis (Figure 17). The progress on the queued samples remains at 1% until the analysis starts. When multiple samples are In Progress, they will run sequentially according to the analysis queue.

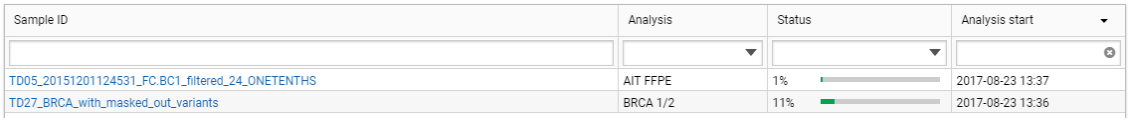


Figure 17: An analysis in the state In Progress is currently 11% complete.

The user can choose to stop the analysis by selecting the sample and clicking **Stop**, which will remove the analysis from the list of analyses and discard intermediate results. A new analysis can later be started manually (as described in the Manual start of an analysis section of this manual). In case the analysis was started from GeneRead Link, it can be restarted from GeneRead Link using the "Requeue" functionality.

Failed

A sample status is set to Failed when the analysis has failed. This is in general a technical failure, 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.

The appropriate process for restarting a failed sample will depend on the QCI Analyze setup, whether it is configured with GeneRead Link or not. For setups including GeneRead Link, use the "Requeue" option in GeneRead Link. When GeneRead Link is not available, use the **Rerun** button available at the bottom of the Sample analysis details panel to start a new analysis on the same input data.

Ready for Review and Uploaded

Sample analyses are set to Ready for Review when the analysis results are available. If upload to QCI Interpret for GeneReader has been initiated, the status is changed to Uploaded.

For sample analyses set to either of these statuses, the following actions are available:

- **Link** is used to link together a variant detection analysis and a fusion detection analysis that were run on samples from the same person. If a sample is linked, the button is renamed to **Unlink** and can be used to undo the previously established link.
- **Sign** is used to sign and optionally add a comment to the report.
- **Upload** will initiate upload of the analysis result to QCI Interpret for GeneReader.
- **Complete** renders the sample analysis read-only and ready for archiving and data clean up.
- **Review** will open the report for inspection in the **Review** page.

Actions available on analyzed samples

Link two sample analyses

It is possible to link analysis results from a DNA analysis workflow (variants, and CNVs when applicable) with analysis results from a RNA analysis workflow (fusions). The link indicates that both analyses come from the same sample material, and it ensures that results will be uploaded as one to QCI Interpret for GeneReader, thereby generating a single QCI Interpret test per sample material.

Note: It is not possible to link two variant detection results, or two fusion gene detection results with each other.

To link a sample analysis result to another, select one of the sample analyses in the list and click the **Link** action in the bottom of the **Sample analysis details** panel. This opens a dialog with the list of analysis results available in QCI Analyze. Select from the list the other analysis result to link to and click **Link**.

When a sample analysis result is linked to another, it is indicated:

- with an icon to the right hand side in the Sample ID column;
- in the **Sample analysis details** panel, where the sample ID of the linked sample analysis is available;

- in the Summary section of the report that gives the sample ID of the linked sample analysis.

In addition, the link action is documented in the History section with indication of the sample ID of the linked sample analysis.

When sample analyses are linked, the actions **Review**, **Upload** and **Complete** on one of the linked analyses affect both sample analyses. The **Review** action opens both analysis reports in the Review page. **Upload** combines the results from the two reports and upload them together. **Complete** completes both sample analyses in one action.

It is possible to unlink linked sample analyses: select one of the linked samples and click the **Unlink** action in the bottom of the **Sample analysis details** panel.

Sign a sample analysis

One or several users can approve the results of a sample analysis by clicking on the **Sign** button. They may also optionally add a comment when signing off. An icon to the right in the status column indicates when at least one user has signed the report. Information on who signed, when, and the optional comment are also logged in the History section of the report.

When a user has signed a report, he cannot sign it again unless the report is modified. If the results are edited (such as moving a variant from one table to another or adding a comment to a variant), the previous signature is considered invalid and the icon indicating the analysis results had been signed is removed from the Analyses list. However, the History section will retain all information pertaining to previous signatures.

Upload to QCI Interpret for GeneReader

The **Upload** functionality allows users to upload the list of reported variants, CNVs and fusions to QCI Interpret for GeneReader for interpretation of their clinical and biological relevance. A QCI Interpret for GeneReader account is required to make use of this functionality, and the QCI Interpret connection should be configured in the Administration page (see the QCI Interpretation configuration section of this manual).

If the user's QCI Interpret username has been specified within the QCI Analyze user account settings in the Administration page, the user is automatically logged to in QCI Interpret when doing an upload. Otherwise, the user is presented with a login screen where the username and password to QCI Interpret for GeneReader provided by QIAGEN can be specified.
Note: Because the upload interview page opens in a new browser tab, it is necessary that the browser allows for pop-ups.

In the upload interview pages, users can add metadata to the sample. For a detailed description on the individual fields and how these can be used to enhance reporting, please refer to the QCI Interpret for GeneReader user documentation.

For situations where the build-in **Upload to QCI Interpret** functionality is not available (e.g., when QCI Analyze is not set up with suitable internet access), users can instead choose to save the results as VCF and manually upload the VCF to QCI Interpret for GeneReader.

QCI Interpret test state and link

For sample analyses uploaded to QCI Interpret for GeneReader, the Sample analysis details panel will display the state of the resulting QCI Interpret test as well as a "QCI Interpret test" link. By clicking the link or the equivalent **Go to** button on the sample analysis report, the corresponding QCI Interpret test will open in a separate browser tab where the users who have permission can access it.

For information on how to use QCI Interpret, please refer to the QCI Interpret user documentation.

Complete a sample analysis

When a sample analysis report has been reviewed, and, when applicable, has been uploaded to QCI Interpret for GeneReader, the sample analysis should be completed. For Completed samples, the analysis results can still be inspected via the Review page, but the samples can no longer be edited, signed or uploaded to QCI Interpret for GeneReader. Completed sample analyses will eventually be archived as configured (see the Archiving section of this manual).

Manual import and start of analysis

For installations configured for GeneRead Link, analyses are started automatically by GeneRead Link when the sequencing has completed. For installations without GeneRead Link, GeneReader Planner is used to create flow cell plans, and sequencing reads are automatically imported and analyses started after the sequencing has completed.

However, it can be relevant in some cases to manually import and start analysis of a sample.

To manually import a sample, drag and drop the relevant FASTQ file to the "Drop files to import here" in area next to the Start Analysis button in the Analyses page. Importing starts immediately, but FASTQ files can be large and the import might take a while to complete.

Notes about importing data

- It is important to not log out while importing samples, as doing so may lead to truncated or corrupted sample files.

- When importing the same fastq file twice, the existing file will be overwritten.
- When importing data manually, the cumulative size of the data imported should not exceed 20 GB.

Once data has been imported, it is available for analysis in QCI Analyze. The value appended to the Start Analysis button indicates the number of samples for which an analysis is yet to be assigned.

To manually start an analysis:

1. Click the **Start Analysis** button in the top of the Analyses page.
2. Select the sample data to be analyzed from the list by ticking one or more checkboxes. Click **Next**. **Note**: In the "Analyzed" column, a check mark indicates if a previous analysis of the data is already available in QCI Analyze, and a cabinet icon indicates that the data has been analyzed but the analysis results were deleted and/or archived as configured.
3. The subfolder structure of the Import directory is visible from the Data ID column. As an example, `sample1` placed in the subfolder `AIT_UMI` will appear in the Data ID column as "AIT_UMI/sample1". Samples that were manually imported using drag and drop will be listed as "manual_import/". By filtering on subfolder naming, you can easily locate samples in a specific subfolder.
4. Select the analysis workflow to use from the drop-down menu (Figure 18).
5. Specify a unique sample ID for each of the selected data items, to identify the sample analyses.
6. Optionally: specify a comment to include in the analysis report(s).
7. Click **Start**.

Note for installations configured for GeneRead Link: analyses started manually in QCI Analyze will not be visible in GeneRead Link, and the analysis results cannot be retrieved by GeneRead Link. If automatic data clean up is configured and GeneRead Link used for archiving results, the results for manually started analyses need to be manually archived using the **Download** options.

Figure 18: Starting an analysis manually.

Review

Once a sample is opened for review, the analysis results are shown as a report in the **Review** page (Figure 19). Several reports can be open at the same time, and the user can navigate through these by clicking the tabs. The value appended to the Review page icon in the navigation bar indicates how many reports are currently open.

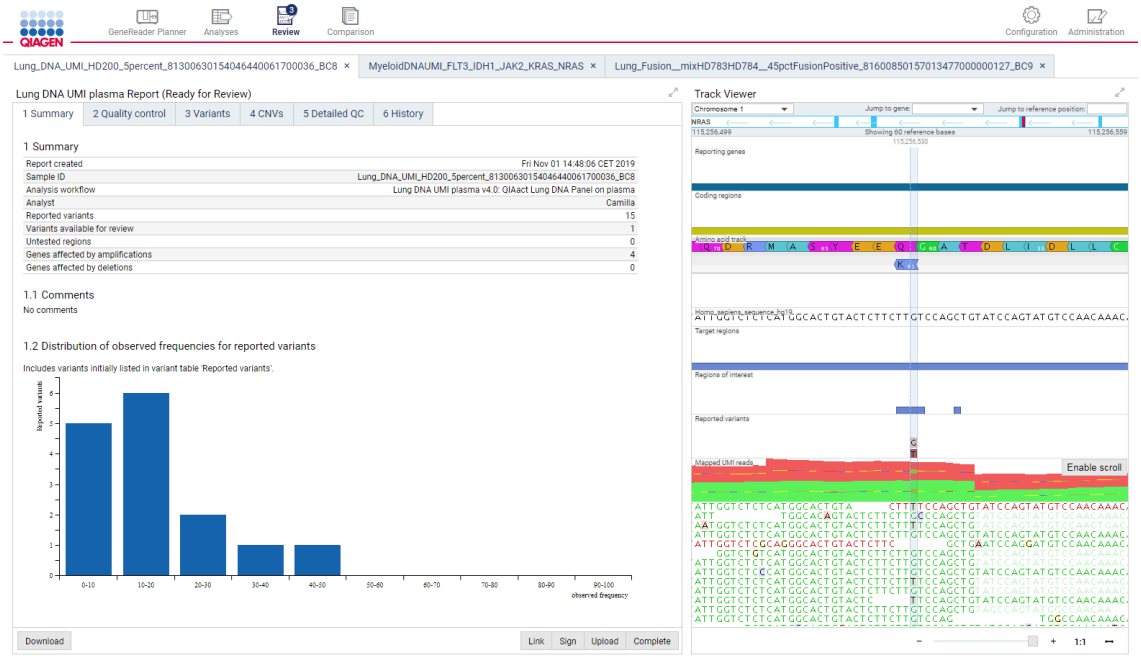


Figure 19: Review page showing the Report tile to the left and the Track Viewer tile to the right.

A report consists of two tiles, the **Report** tile and the associated **Track Viewer** tile. Both can be maximized to full screen width using the arrow icons in the upper right corner of the tile.

The report is made up of multiple sections. Users can navigate these sections by clicking the tabs at the top of the **Report** tile. The content of the Report tile can be saved as Excel or PDF file and printed, and results can be uploaded to QCI Interpret for GeneReader.

Note: A report can be opened simultaneously in different sessions, i.e., by different users on different computers. In the Review page, a blue bar to the far right of a tab indicates that the report is also open for review in another QCI Analyze session. Changes made in one session

are not automatically propagated to the other session, but closing and opening the report will update it. Editing a variant that was already edited in another session will fail without notification. Likewise, a comment on a variant can be overridden if edited in different sessions. However, all edits and comments will be logged in the History section of the report.

The Track Viewer is the graphical interface in which multiple tracks are displayed together, e.g., the reference sequence, gene annotations, and mapped reads. Tracks are the fundamental building blocks for QCI Analyze data analysis and provide a unified framework for visualization, comparison and analysis of genome-scale studies. For tracks, all information is tied to genomic positions. A central coordinate-system is provided by a reference genome, which allows different types of data to be viewed together. Tracks can be moved up and down in the viewer by simple drag and drop.

The content of the report and the track viewer differs between workflows. Currently, QCI Analyze offers two overall report types: the variant detection report and the fusion detection report.

Variant detection report

Summary

This section displays the main information about the performed analysis and a summary of the analysis results. A graph shows the variant frequency distribution (Figure 20).

Quality Control

The Quality Control tab of the Report includes a minimum of five sections, some of which are green/yellow color-coded according to "pass" thresholds indicated below each section. Some of these cutoff values can be configured by an administrator on a per workflow basis as explained in the Configuration section of this manual.

Input data summary displays details about the sequencing reads from the FASTQ file (except for the number of nucleotides and the average read length that are based on the FASTQ reads plus the number of nucleotides removed in primary analysis, i.e., 6 per read for non UMI-based panels, and 14 per read for UMI-based panels).

Secondary analysis summary gives mapping and coverage metrics, such as how many reads were mapped (note that the "Percentage of mapped reads" is based on a "total read" value consisting of the number of reads left after the initial trimming), and how much of the ROI obtained significant and minimum coverage respectively. The Myeloid DNA UMI analysis report features two additional rows. These report how much of the combined high

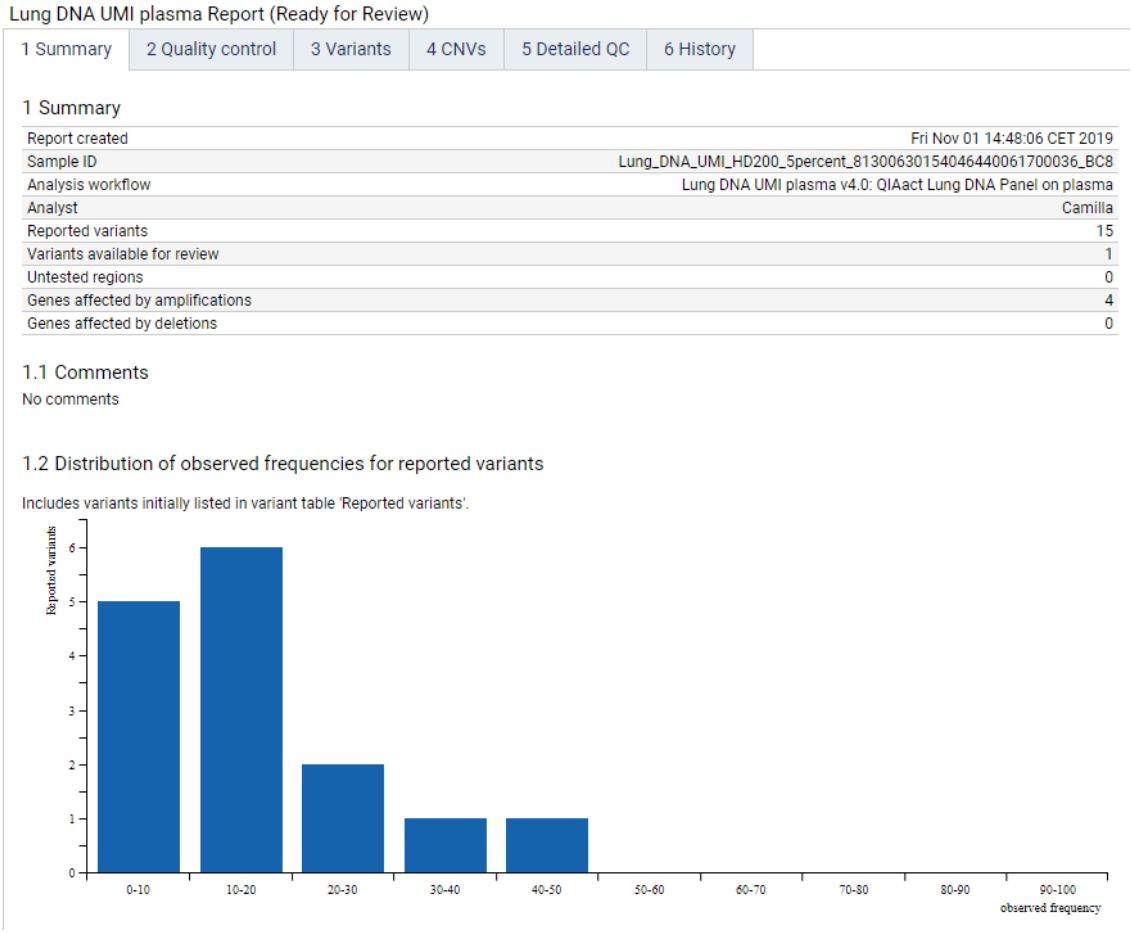


Figure 20: Summary of a report and variant frequency distribution.

sensitivity region obtained minimum and significant coverage respectively.

Unique Molecular Indexing shows metrics for the UMIs generated by the secondary analysis workflow. The "Average number of reads per Unique Molecular Index" and the "Median number of reads per Unique Molecular Index" serve as summary statistics that characterize the distribution of reads per UMI. A value of minimum 2 is needed to enable sequencing error correcting. On the other hand, an excessive amount of reads per UMI often leads to a reduction in UMI coverage, as it entails that the same fragment is sequenced many times, instead of more fragments fewer times. So while values between 6-10 are acceptable, values above 20 are typically not desirable and will often be a sign of too little DNA starting material. Finally, average and median values far apart indicate that the distribution is skewed. As an example, an Average of 50 and a Median around 4 would indicate that most UMIs are supported by a lower number of reads, and a few UMIs are supported by many reads. Please note that above values are guidelines only.

Untested regions lists regions where variants could not be tested due to insufficient coverage.

Coverage/Reads coverage gives the coverage on a per gene basis. Values are based on raw

read coverage (as opposed to unique molecular indexing, see below).

Detected variants recapitulates the detected variants on a per gene basis.

For workflows supporting panels using the UMI technology (such as the Lung FFPE and Lung plasma workflows), the report will contain metrics on unique molecular indexing:

UMI reads coverage gives UMI coverage on a per gene basis.

Some panels target genes are not of interest for variant calling, but only considered for copy number variation detection or used for normalization. Such genes are not included in the regions of interest defined for the analysis workflow, but as they are included in the panel target region they will still appear in the coverage tables mentioned above. Here, they will be listed with "ROI" and "Bases" values of zero indicating that no regions of interest are defined for these genes. The remaining fields concerning coverage for defined regions will be empty.

The track viewer in the adjacent tile is updated to display the gene corresponding to the row selected in the tables.

Variants

Detected variants are divided into two tables, "Reported variants" and "Variants available for review".

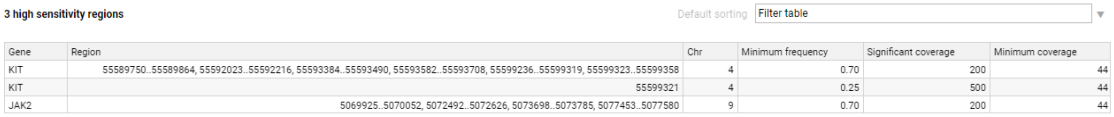
3.1 Reported variants, also described as "Valid" variants. This table lists all variants that were detected and passed the additional filtering for automatic validation of variants:

- the variant is found within the ROI;
- the variant is detected at a frequency equal to or higher than the minimum frequency threshold;
- the variant position has a coverage equal to or higher than the "significant coverage" threshold.

The thresholds for a particular analysis can be found in the History section of the report.

3.2 Variants available for review, i.e., variants that were detected but did not pass the additional filtering for automatic validation of variants. They are labeled as "Review". During review, the user can choose to change the label of a variant: if the variant is "Confirmed", it moves to table 3.1; if set to "Artifact", it remains in table 3.2.

Note: The Myeloid DNA UMI analysis report features an additional table at the top of the Variants tab, the High sensitivity regions table (Figure 21). High sensitivity regions are defined for the analysis workflow. These are regions for which detection of low frequency variants is of interest, which is why the frequency cut-off and significant coverage thresholds differ from the rest of the panel. Variants detected in high sensitivity regions are triaged for tables 3.1 and 3.2 based on frequency and coverage just as the remaining variant, only triage is based on thresholds in the high sensitivity table instead of those defined for the rest of the panel. Positions in high sensitivity regions for which coverage dropped below the minimum coverage threshold defined for that position will be listed in the Untested regions table.



3 high sensitivity regions

Default sorting | Filter table

Gene	Region	Chr	Minimum frequency	Significant coverage	Minimum coverage
KIT	55589750..55589864, 55592023..55592216, 55593384..55593490, 55593582..55593708, 55599236..55599319, 55599323..55599358	4	0.70	200	44
KIT	55599321	4	0.25	500	44
JAK2	5069925..5070052, 5072492..5072626, 5073698..5073785, 5077453..5077580	9	0.70	200	44

Figure 21: High sensitivity regions table.

Variant tables share the same column headers:

- Gene. Name of the affected gene.
- Exon. The number of the exon(s) in which the variant or region lies. Numbering is based on the applied transcript and may, unlike numbering based on the coding regions, include 5' untranslated exons. This field will be empty for variants outside of exons.
- Type. Variant type, such as Single Nucleotide Variant (SNV), Multiple Nucleotides Variant (MNV), insertion, insertion* (Myeloid DNA UMI only), deletion, and replacement.
- Length. Length of the variant.
- Alteration. Coding DNA (c. variant) or genomic reference sequence variant (g. variant) nomenclature based on the Human Genome Variation Society (HGVS) recommendations (`https://doi.org/10.1002/humu.22981`).
- p. variant. Protein sequence variant nomenclature based on the Human Genome Variation Society (HGVS) recommendations.
- Impact. Translational impact of variant.
- %. Detected variant frequency.
- Avg Q. Average quality score of the bases supporting the variant. For example, if a SNV has a read count of 3, and the Q-scores of these 3 nucleotides are 30, 32 and 35, the Avg Q of that variant will be (30 + 32 +35)/3 = 32.33.

- F/R test. Relative forward/reverse read distribution. This value indicates if the forward/reverse ratio for reads that support the variant (counts) is similar to the forward/reverse ratio of the total number of reads covering the position (coverage). The value ranges from 0 to 1, where values closer to 0 may indicate a sequencing artifact. With very high coverage, even small relative differences in allele frequencies (which are most pronounced for low frequency variants) may lead to an F/R test value that is 0 or very close to 0. Hence, in situations where the coverage is high and the frequency low, the value should not be used for hard filtering, but rather as a guide towards variants that deserve closer inspection.
- Position test (Lung Plasma Track only). Relative position of the variant in reads. This value indicates if the position of the variant allele in the reads supporting it is similar to the read position distribution for all reads covering that site. Values range from 0 to 1. Values close to 0 mean the two distributions are very different, and are suggestive of sequencing artifacts.
- Singleton UMIs (Lung Plasma Track only). The proportion of UMIs supporting the variant that are based on only one sequencing read (singletons). Values range from 0 to 1. The closer to 1, the higher the proportion of singleton UMIs, and the lower the confidence in the allele.
- Repeat. Variant is located in a low-complexity region.
- Count. Number of fragments supporting the allele. In case of single reads, this corresponds to the read count supporting the allele.
- F Count. Number of forward reads with the detected variant.
- R Count. Number of reverse reads with the detected variant.
- Coverage. Fragment coverage at this position. Overlapping paired reads have two reads in their overlap region, but only count as one fragment. In this case, overlapping paired reads contribute only 1 to the coverage.
- F Coverage. Number of forward reads covering the variant position.
- R Coverage. Number of reverse reads covering the variant position.
- Qual. Measure of the significance of a variant, i.e., a quantification of the evidence (read count) supporting the variant, relative to the coverage and what could be expected to be seen by chance, given the error rates in the data. The mathematical derivation of the value

is hence much more complicated than the "Average quality" value, because it depends on the set of probabilities of generating the nucleotide pattern observed at the variant site (1) by sequencing errors alone and (2) under the different allele models of the variant caller allows. Qual is calculated as $-10\log_{10}(1-p)$, p being the probability that a particular variant exists in the sample. Qual is capped at 200 for p=1, with 200: highly significant, 0: insignificant.

- Region. Position of the variant relative to the reference sequence.
- Chr. Affected chromosome.
- ROI. Variant located in Regions of Interest, as specified by the analysis workflow.
- Review. Status of variant review. The status "Valid" is applied to variants that passed the filtering for automatic validation. "Confirmed" and "Artifact" indicates that a variant was confirmed by review or dismissed, respectively, by the user during manual inspection.
- Comment. Remark added by user during variant review.

To find particular variants, tables can be filtered in different ways:

- Clicking on the header of a column will sort the table according to the values present in the column. Hold down the Shift key and select several columns successively to sort in the defined order of selection.
- Using the basic filter functionality will select rows containing the text typed in the filter area, regardless of which columns the text belongs to.
- Expanding the filter by clicking on the arrow to the right hand side of the filter area lets the user access more additional filter fields such as "Column" (choose which column from a drop-down menu), "Modifier" (= or contains) and "Value" (can be a numerical value or a text depending on the column chosen for the filter). Clicking the + button will add additional filtering criteria to the search, and the users can decide whether results should "Match all" filters (will display only few results), or only "Match any" one of the filters set up (will display the combined results of each filter independently).

CNVs

For workflows with CNV detection, the report includes a CNVs tab where users can access the CNV analysis results (Figure 22).

Reporting differs for gene and exon level CNV detection, as described below.

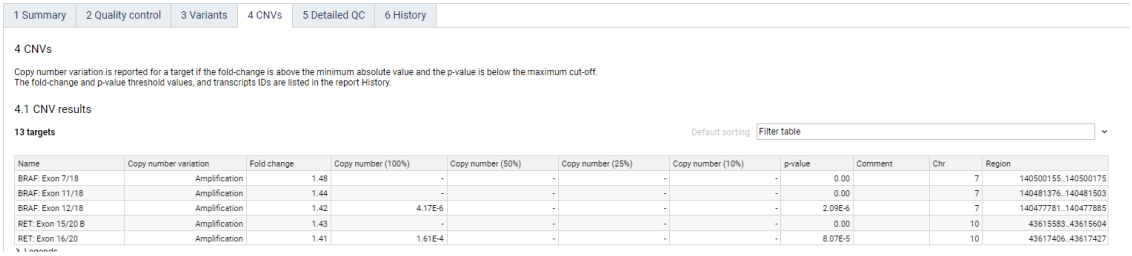


Figure 22: The CNVs tab of the report.

Gene level CNV detection

For analysis workflows with gene level CNV detection, all CNV target genes are listed in a single table with the following columns:

- Name. Name of the affected gene.
- Copy number variation. "'Amplification"' and/or "'Deletion"' depending on the fold-change value, or "'No CNV detected"' if fold-change and p-value thresholds are not met.
- Fold change. The observed fold change.
- Copy number columns.
  - Somatic analysis workflows: Four copy number columns, 100%, 50%, 25%, and 10%, corresponding to the given percentages of sample purity ("tumor content").
  - Germline analysis workflows: One copy number column corresponding to a sample purity ("tumor content") of 100%.
  - p-value. A measure of certainty of the call. The closer the value is to 0, the more certain the call.

Exon level CNV detection

For analysis workflows with exon level CNV detection, CNV calls are divided into two tables, "4.1 CNV results" and "4.2 CNVs - not testable". Table 4.1 lists CNVs that passed the defined thresholds. If the overall sample coverage is insufficient, some exons might be classified as not testable. These will be listed in table 4.2.

For panels with less than 500 targets, the fold changes for all the analyzed targets are depicted in a graph (Figure 23).

The two CNV tables have the following columns:

- Name. Name of the affected exon.

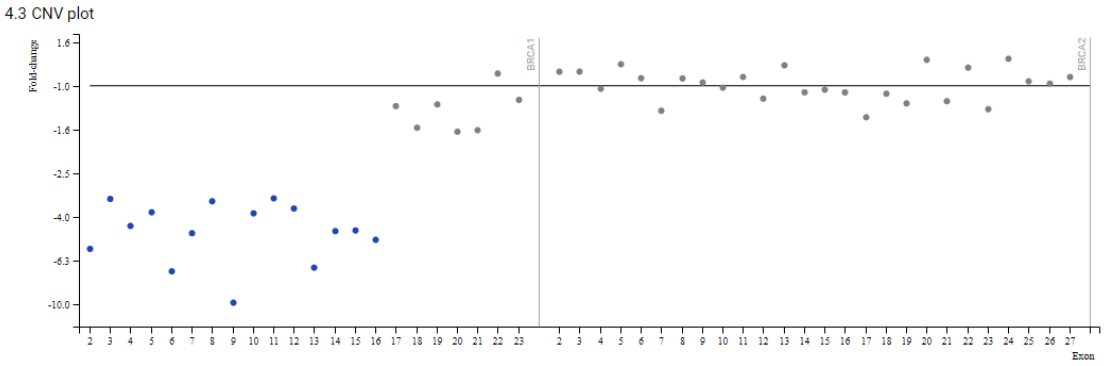


Figure 23: Graph depicting fold change for each CNV target.

- Copy number variation. "'Amplification"' and/or "'Deletion"' depending on the fold-change value.
- Fold change. The observed fold change.
- Copy number columns.
  - Somatic analysis workflows: Four copy number columns, 100%, 50%, 25%, and 10%, corresponding to the given percentages of sample purity ("tumor content").
  - Germline analysis workflows: One copy number column corresponding to a sample purity ("tumor content") of 100%.
- p-value. A measure of certainty of the call. The closer the value is to 0, the more certain the call.
- Comment. Will highlight if an observed fold change is very different from its neighboring exons ("Fold change not consistent with neighboring exons"). This is an indication that the observed CNV call is uncertain.
- Chr. Affected chromosome.
- Region. Position of the exon on the affected chromosome.

Note: The Copy number columns list the copy numbers for the cancerous part of the sample, provided that the sample tumor content is 100%, 50%, 25% and 10%, respectively. These values are based on the fold change. For a given amplification fold change, the smaller the sample tumor content (%), the higher the derived copy number is in the tumor DNA. A copy-number of "-" indicates that the given cancer percentage is incompatible with the observed fold change. For some deletion calls, the fold changes may be of such a magnitude that it is not consistent with a small cancer percentage. That is because a homozygous deletion in the cancer portion of a sample with cancer content of 10% can at most lead to a coverage reduction of 10%, corresponding to a fold change of -1.11 (= -1.0/0.9).

Detailed QC

Detailed QC provides a graphical representation of many of the key attributes of sequencing and mapping such as GC-content, ambiguities, average base quality and target regions coverage (see Appendix D for a description of the Detailed QC report).

History

History captures the progress of each sample through the analysis and review process and logs all events performed by users on the sample, such as changes to variant states, upload to QCI Interpret for GeneReader, written comments, and signatures of report. History also lists analysis workflow parameters as well as software and analysis workflow versioning information.

Variant track viewer

The **Track Viewer** tile helps visualize the individual mapped reads, enabling users to easily review identified variants. The Variant tables are linked with the track viewer. If you select a variant in a one of the tables the track viewer will navigate to the genomic position of the variant. From here, users can review the variant in the context of the following tracks:

- Chromosome ideogram. The ideogram, situated at the top of the track viewer, shows the genomic context of the current selection. It is possible to navigate between different chromosomes, genes, or other specific reference positions. Users can zoom in and out using the zoom functionality at the bottom of the track viewer. Depending on the current zoom level of the tracks, the ideogram shows a different visualization.

  At the lowest zoom mode, a **genome overview** shows chromosomes as blue rounded boxes, with the chromosome currently displayed in the track highlighted. Clicking a any of these rounded blue boxes will change the ideogram to chromosome overview mode.

  The **chromosome overview** zoom level shows the current chromosome in its entirety drawn with cytogenic bands (Figure 24). The red box marks the section of the chromosome which is shown in the tracks. Users can switch chromosome from the drop-down menu in the top left corner of the ideogram, or select a particular gene on the chromosome itself to zoom in on the gene in the tracks below the ideogram. Clicking anywhere on the chromosome visualization or submitting a position in the top right field will display the selected region in the tracks below accordingly.

  Zooming further in displays an **overview of the reporting gene** surrounding the position currently shown in the tracks (marked as a red box on Figure 25). The genes are drawn as light blue annotation bars with directional arrow-ends, and overlapping genes are

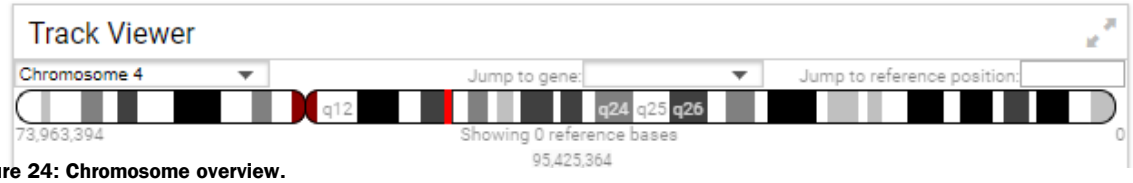


Figure 24: Chromosome overview.

drawn on top of each other. This level can be navigated in a similar way as described for the "Chromosome Overview".

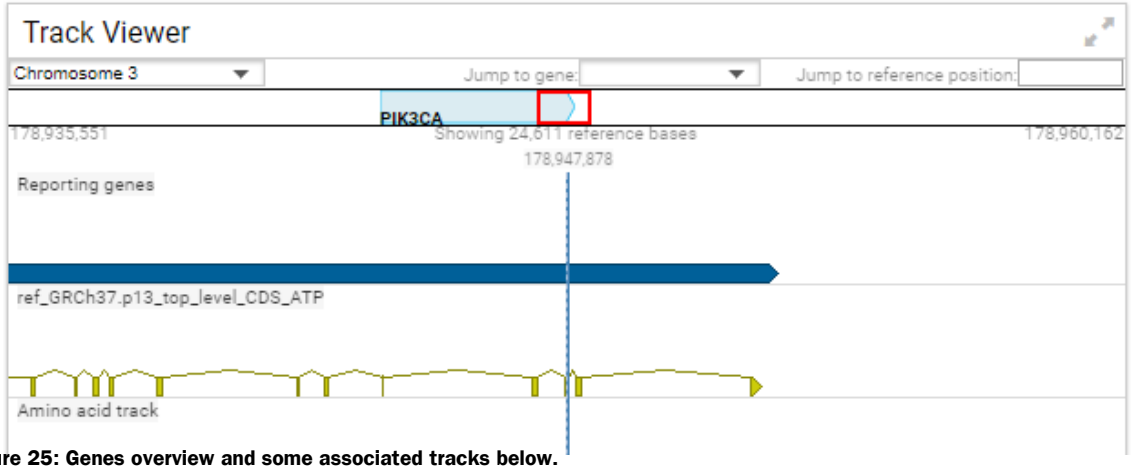


Figure 25: Genes overview and some associated tracks below.

Zooming in gives an **overview of a single gene** currently overlapping with the section of the tracks (shown as a red mark on Figure 26). If multiple genes are overlapping, only one is shown, but the names of all overlapping genes are shown in the bottom left corner of the ideogram, with the currently shown gene highlighted. It is possible to click on the gene's name to switch the gene visualized in the ideogram. Exons of the gene selected for display are drawn as yellow boxes, and the direction of the gene is indicated by arrows.

Figure 26: Single gene overview and some associated tracks below.

- Human reference sequence track hg19. This is the track containing the reference genome sequence.
- Reporting Genes, Reference CDS, Amino acid track, Target regions, Regions of interest. Each annotation track contains a specific annotation type such as gene name.
- Reported variants and Variants available for review. These correspond to the content of variant tables 3.1 and 3.2 respectively (Figure 27). Variants may be of type SNV, MNV, replacement, insertion or deletion.

Figure 27: Variant tracks in the Track Viewer.

- Mapped reads. This track displays the mapping of raw reads (Figure 28). The reads track contains all the reads mapped at a particular position. Forward reads are green, reverse reads are red. Unaligned ends of mapped reads are shown in lighter shades of the respective read color. Reads that mapped equally well to another position in the mapping, i.e., non-specific reads, are yellow. Below the reads is an overflow graph, i.e., a graphical display of the coverage at each position in the reference. The overflow graph uses the

same colors as the sequences (forward = green and reverse = red), with mismatches in reads shown as narrow horizontal lines (red = A, blue = C, yellow = G, and green = T).

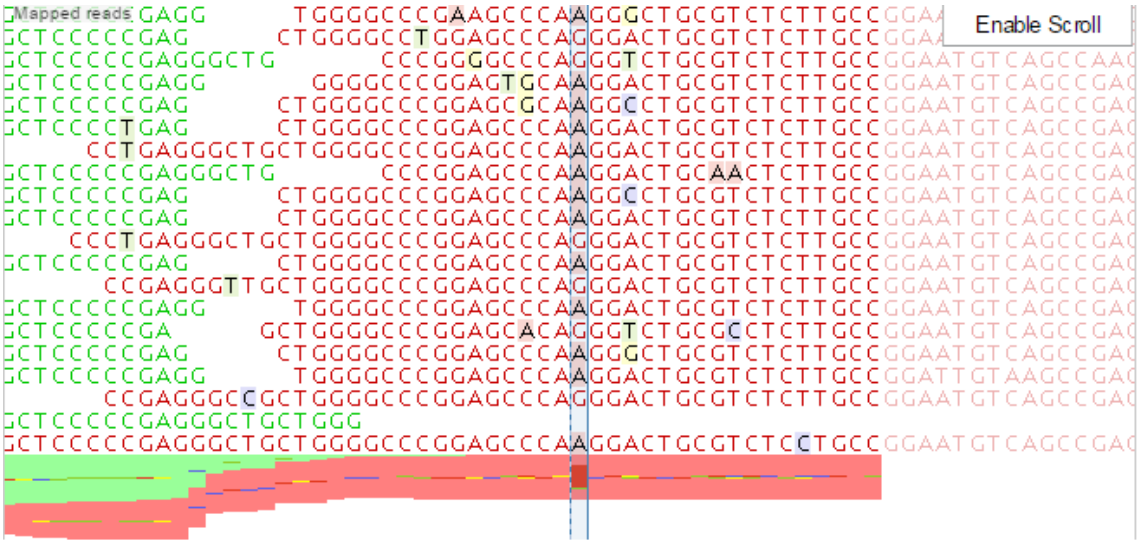


Figure 28: Mapped reads track.

Note: It is possible to scroll through all reads of a read mapping by first clicking the **Enable scroll** button. Clicking the **Disable scroll** button will display the original view of the Track Viewer.

- Mapped UMI Reads. This is the mapping track displayed for workflows using Unique Molecular Indices.

Hovering on a variant in the Reported variant track will open a popup window with a selection (depending on the analysis workflow) of the following information:

- SNV or MNV. Description of the variant.
- Zygosity. Zygosity of the variant called, as determined by the variant caller. This will be either "Homozygous", where there is only one variant called at that position or "Heterozygous", where more than one variant was called at that position.
- Frequency. "Count" divided by "Coverage".
- Count. Number of fragments supporting the allele. In case of single reads, this corresponds to the read count supporting the allele.
- Coverage. Fragment coverage at this position. Overlapping paired reads have two reads in their overlap region, but only count as one fragment. In this case, overlapping paired reads contribute only 1 to the coverage.

- Probability. For a given single site variant, the probability is calculated as the sum of probabilities for all the explanations containing that variant. So if a G variant is called, the reported probability is the sum of probabilities for these configurations: G, A/G, C/G, G/T, A/C/G, A/G/T, C/G/T, and A/C/G/T (and also all the configurations containing deletions together with G).
- Forward/reverse balance. The minimum of the fraction of forward reads and reverse reads carrying the variant among all reads carrying the variant.
- Average quality (or Avg Q in the table). Average quality score of the bases supporting the variant.
- Read coverage. Read coverage at this position. Each read of an overlapping pair contributes to the coverage.
- Read count. Number of reads supporting the allele.
- # unique start positions. Number of unique start positions for fragments that support the variant. This value can be important to look at in cases with low coverage: if all reads supporting the variant have the same start position, it could be the result of an amplification error.
- # unique end positions. Number of unique end positions for fragments that support the variant. This value can be important to look at in cases with low coverage: if all reads supporting the variant have the same end position, it could be the result of an amplification error.
- BaseQRankSum. Evaluation of the quality scores in the reads that have a called variant, compared with the quality scores of the reference allele. Variants for which there are no reads holding the corresponding reference allele do not have a BaseQRankSum value. The score is a z-score derived using the Mann-Whitney U test, so a value of -2.0 indicates that the observed qualities for the variant are two standard deviations below what would be expected if they were drawn from the same distribution as the reference allele qualities. A negative BaseQRankSum indicates a variant with lower quality than the reference variant, and a positive z-score indicates higher quality than the reference.
- Read position test probability. Tests whether the distribution of the read positions of a variant in the variant carrying reads is different from that of all the reads covering the variant position.

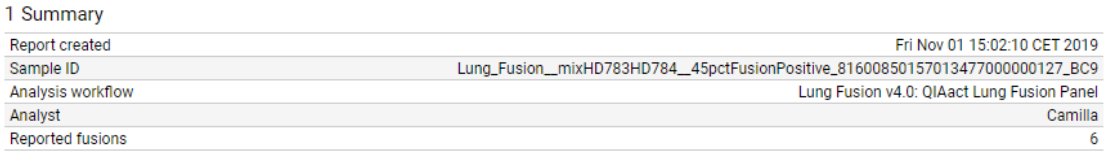
- Read direction test probability (or F/R test in the table). Tests whether the distribution among forward and reverse reads of the variant carrying reads is different from that of all the reads covering the variant position. This value reflects a balanced presence of the variant in forward and reverse reads (1: well-balanced, 0: un-balanced).
- Homopolymer (or Repeat in the table). Variant is located in a low-complexity region.
- Homopolymer length. Length of the homopolymer.
- Qual. Variant significance value calculated as $-10\log_{10}(1-p)$, p being the probability that a particular variant exists in the sample. Qual is capped at 200 for p=1, with 200: highly significant, 0: insignificant.
- Count (singleton UMI). Number of singletons UMIs supporting the variant. Singleton UMIs are UMI reads based on only one raw read.
- Count (big UMI). Number of big UMIs supporting the variant. Big UMIs are UMI reads created from 2 or more raw reads that had the same UMI.
- Sequencing context error. True when the variant is covered by reads in both directions, but only found in reads of one direction.
- PCR Error. True if the variant is covered by reads originating from at least two primers, but only found in reads originating from one primer.
- Coding region change (corresponds to Gene and Alteration in the table). Describes the relative position of the change on the coding DNA level, such that for example "c.-4A>C" would describe a SNV four bases upstream of the start codon, while "c.*4A>C" would describe a SNV four bases downstream of the stop codon. Variants inside exons and in the untranslated regions of the transcript will also be annotated with the distance to the nearest exon.
- Amino acid change (corresponds to p. variant in the table). Describes the change on the protein level. For example, single amino-acid changes caused by SNVs are listed as p.Gly261Cys, denoting that in the protein sequence (hence the "p.") the Glycine at position 261 is changed into Cysteine. Frame-shifts caused by nucleotide insertions and deletions are listed with the extension "fs", as in p.Pro244fs describing a frameshift at position 244 coding for Proline.
- Amino acid change in longest transcript. When there are many transcript variants for a protein, the longest transcript is often used for reference.

- Coding region change in longest transcript. When there are many transcript variants for a gene, the coding region change for the longest transcript is often used for reference.
- Other variants within codon. If there are other variants within the same codon, this column will have a "Yes". In this case, it should be manually investigated whether the two variants are linked also in reads.
- Reporting genes (Gene in the table). Name of the affected gene.
- Region of Interest (ROI in the table). Variant located in a Region of Interest, as specified for the analysis workflow.

Fusion detection report

Summary

This section displays the main information about the performed analysis and a summary of the analysis results (Figure 29).



1 Summary

Report created	Fri Nov 01 15:02:10 CET 2019
Sample ID	Lung_Fusion__mixHD783HD784__45pctFusionPositive_8160085015701347700000127_BC9
Analysis workflow	Lung Fusion v4.0: QIAact Lung Fusion Panel
Analyst	Camilla
Reported fusions	6

Figure 29: Summary of a fusion report.

In particular, the summary highlights how many fusions were detected by the analysis workflow.

Quality Control

The Quality Control tab of the Report includes three sections, some of which are green/yellow color-coded according to "pass" thresholds indicated below each section.

Input data summary displays details about the sequencing reads from the FASTQ file (except for the number of nucleotides and the average read length that are based on the FASTQ reads plus the number of nucleotides removed in primary analysis, i.e., 14 per read for UMI-based panels).

Secondary analysis summary lists metrics for how many reads mapped, the total number of UMI reads mapped to the fusion control genes, how many fusion control genes were detected, and the level of DNA contamination.

The panel includes primers targeting four constitutive genes, which by definition are expected to always be transcribed and hence be a source of sequencing reads. These

serve as **Fusion control genes**. A fusion control gene is detected if it has a minimum UMI read count of 5. All four of these control genes must be detected for the quality control to pass.

DNA contamination control is facilitated by control primers targeting a non-transcribed intergenic region. This target region is amplified and is a source of sequencing reads only if genomic DNA is present in the sample. The level of DNA contamination is calculated as the number of UMI reads mapping to this intergenic control region divided by the total number of UMI reads mapped to the target region.

Unique Molecular Indexing shows metrics for the UMIs generated by the secondary analysis workflow. The "Average number of sequencing reads per UMI" and the "Median number of sequencing reads per UMI" serve as summary statistics that characterize the distribution of reads per UMI. A value of minimum 2 is needed to enable sequencing error correcting. On the other hand, an excessive amount of reads per UMI often leads to a reduction in UMI coverage, as it entails that the same fragment is sequenced many times, instead of more fragments fewer times. So while values between 6-10 are acceptable, values above 20 are typically not desirable and will often be a sign of too little DNA starting material. Finally, average and median values far apart indicate that the distribution is skewed. As an example, an Average of 50 and a Median around 4 would indicate that most UMIs are supported by a lower number of reads, and a few UMIs are supported by many reads. Please note that above values are guidelines only.

Panel details lists the number of target genes, primers in target genes, primers in fusion control genes and DNA contamination primers.

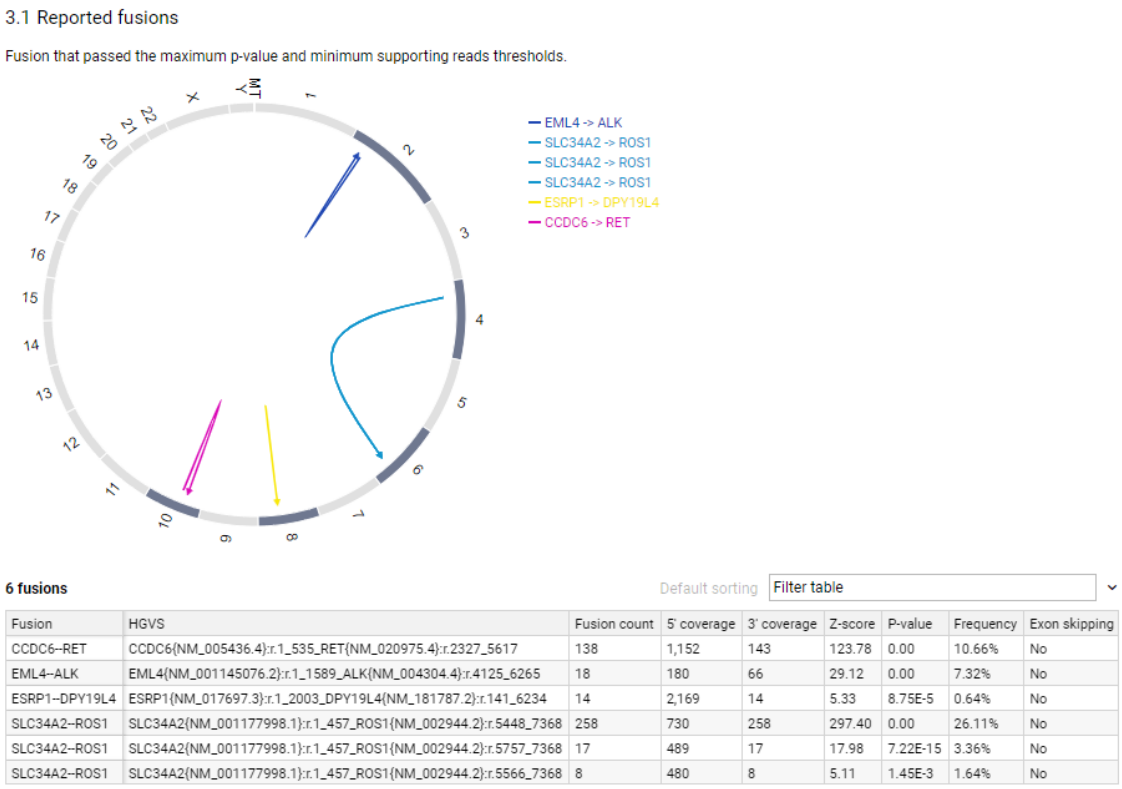
Target genes includes a table giving the names of the genes targeted by the panel primers and the maximum UMI coverage observed for the region of interest defined for that gene.

Fusions

This section presents the main result of the fusion detection by listing detected fusions in two tables.

3.1 Reported fusions lists fusions that passed the defined maximum p-value and minimum fusion count thresholds. The detected fusions are also displayed in a circos plot above the table (Figure 30).

3.2 Fusions not passing filters lists detected fusion that failed the p-value or the minimum supporting reads thresholds for table 3.1.



Fusion	HGVS	Fusion count	5' coverage	3' coverage	Z-score	P-value	Frequency	Exon skipping
CCDC6--RET	CCDC6{NM_005436.4}:r.1_535_RET{NM_020975.4}:r.2327_5617	138	1,152	143	123.78	0.00	10.66%	No
EML4--ALK	EML4{NM_001145076.2}:r.1_1589_ALK{NM_004304.4}:r.4125_6265	18	180	66	29.12	0.00	7.32%	No
ESRP1--DPY19L4	ESRP1{NM_017697.3}:r.1_2003_DPY19L4{NM_181787.2}:r.141_6234	14	2,169	14	5.33	8.75E-5	0.64%	No
SLC34A2--ROS1	SLC34A2{NM_001177998.1}:r.1_457_ROS1{NM_002944.2}:r.5448_7368	258	730	258	297.40	0.00	26.11%	No
SLC34A2--ROS1	SLC34A2{NM_001177998.1}:r.1_457_ROS1{NM_002944.2}:r.5757_7368	17	489	17	17.98	7.22E-15	3.36%	No
SLC34A2--ROS1	SLC34A2{NM_001177998.1}:r.1_457_ROS1{NM_002944.2}:r.5566_7368	8	480	8	5.11	1.45E-3	1.64%	No

Figure 30: Circos plot of fusion events.

Fusion tables share the same column headers:

- Fusion. Short name of the fusion event, 5’ gene–3’ gene.
- HGVS. The name of the fusion event as recommended by the Human Genome Variation Society.
- Fusion count. Number of unique molecular indices (UMIs) mapping to the fusion transcript reference sequence. Evidence for fusion expression.
- 5’ coverage. Number of unique molecular indices (UMIs) mapping to the 5’ wild type gene, evidence for wild type expression.
- 3’ coverage. Number of unique molecular indices (UMIs) mapping to the 3’ wild type gene, evidence for wild type expression.
- Z-score. Converted from the p-value using the inverse distribution function for a standard Gaussian distribution.
- P-value. A measure of certainty of the call. The closer the value is to 0, the more certain the call.

- Frequency. The fusion frequency calculated as the fusion count divided by the total coverage count (fusion count / (3' coverage + 5' coverage)). Low frequency fusions (frequency <0.10%) are filtered separately. For low frequency fusions, the p-value is often not a good measure of significance, why filtering for these instead is based on fusion counts. Filtering thresholds for low and high frequency fusions are configured separately in the configuration panel.
- Exon skipping. If yes, the two fusion partners are from the same gene with the 5' breakpoint found upstream of the 3' breakpoint.

Detailed QC

Detailed QC provides a graphical representation of many of the key attributes of sequencing and mapping quality such as GC-content, ambiguities, average base quality and target regions coverage (see Appendix D for a description of the Detailed QC report).

History

History captures the progress of each sample through the analysis process and logs all events performed by users, such as upload to QCI Interpret, written comments and signatures of report. History also lists analysis workflow parameters as well as software and workflow versioning information.

Fusion track viewer

The **Track Viewer** tile provides a graphical view of how reads mapped the fusion transcripts. The fusion tables are linked with the track viewer. Select a table row to navigate to the corresponding fusion transcript in the track viewer.

For the fusion report, the track viewer contains the following tracks:

- Genes. Displays gene and fusion gene annotations.
- Transcripts. Displays gene or fusion gene transcripts depending on selection. For fusions where the reads matched 2 or more wild type transcripts for e.g. the 5' gene, one fusion transcript will be listed for each of those wild type transcripts.
- Reference. Contains the reference genome or fusion sequence depending on selection.
- Fusion primers. Displays the position of the fusion event-specific primers.
- Fusion breakpoints. Indicates the fusion breakpoint, i.e., the splice site between the 5' and 3' wild type genes.

- Mapped fusion reads. This track displays the mapping of UMI reads to the transcript sequences. It contains only reads that provide conclusive evidence for or against fusion. Forward reads are green, reverse reads are red. Unaligned ends of mapped reads are shown in a lighter shade of the respective read color. Reads that mapped equally well to another position in the fusion transcript reference, i.e., non-specific reads, are yellow. Below the reads is an overflow graph, i.e., a graphical display of the coverage at each position in the reference. This overflow graph uses the same colors as the sequences (forward = green and reverse = red), with mismatches in reads shown as narrow horizontal lines (red = A, blue = C, yellow = G, and green = T).

Actions available on the Review page

Depending on the report state (Ready for Review, Uploaded, or Completed), the following options may be available at the bottom of the **Report** tile. **Note**: most of these actions are also available from the Analyses page and described in greater detail in the Actions available on analyzed samples section of this manual.

- **Review linked** is only available if the analysis result is linked with another analysis result. Clicking **Review linked** opens the linked analysis result for review. If it is already open, it brings it in focus.
- **Link** is used to link together a variant detection analysis and a fusion detection analysis that were run for the same sample material. In a linked sample, the button is called **Unlink** and used to undo the previously established link.
- **Sign** is used to sign and optionally add a comment to the report.
- **Upload** will upload the analysis result to QCI Interpret for GeneReader.
- **Complete** renders the sample analysis as read-only and ready for archiving and data clean up.
- **Go to** takes the user to the interpretation test in QCI Interpret for GeneReader in a separate browser tab. This button is only available once the analysis results have been uploaded to QCI Interpret for GeneReader.
- **Download** offers several options to export the results: the content of the Report can be exported either as PDF or Excel file. It is also possible to output a VCF file for variants detection reports. VCF result files can be used in downstream analyses such as the ones provided by the CLC workbench environment or by QCI Interpret for GeneReader. The option

ZIP archive will compile all the above download options in a single ZIP file. For variants detection reports, it is also possible to download the final read mapping as a BAM and corresponding BAI files.

Note: If automatic archiving is configured, or if QCI Analyze is configured with GeneRead Link, PDF and VCF are automatically stored or retrieved by GeneRead Link, respectively.

Comparison

To facilitate the comparison of detected variants and QC values across analyses, two tools generating comparison reports are available in the Comparison page of QCI Analyze (Figure 31). It is only possible to compare samples that were analyzed with the same variant detection workflow, and using the same parameter settings such as coverage thresholds.

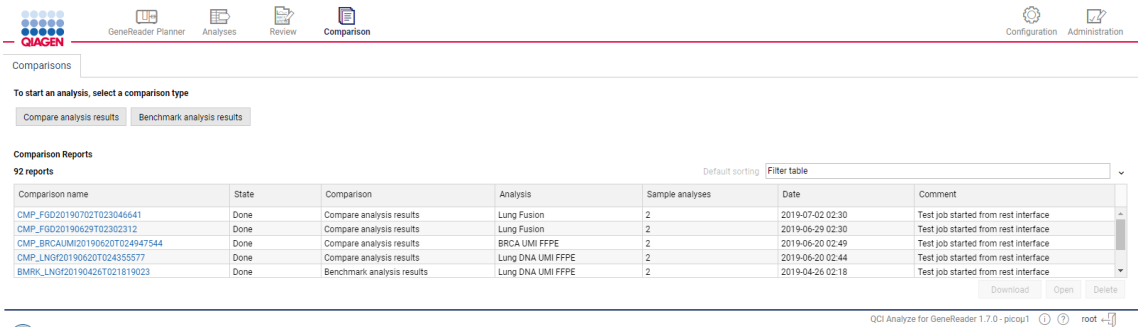


Figure 31: The Comparison page.

Use the buttons below the table to **Open**, **Delete** or **Download** (as an Excel file) the report.

Note: If the underlying analysis results of a comparison are deleted, the comparison report can still be opened, but the blue hyperlinks available in the comparison report to the deleted samples will not open a QCI Analyze report.

Compare analysis results

To start a comparison, click **Compare analysis results** (Figure 32).

1. In the **Analysis workflow** field, choose the relevant workflow from the drop-down menu. All the samples that were analyzed with the selected workflow are now listed in the table.
2. Name the comparison analysis in the **Comparison name** field. Writing a comment is optional.
3. Now use the checkboxes to select the samples to compare. To find specific samples, use the **Filter and search** options available in the headers of the table. It is possible to select

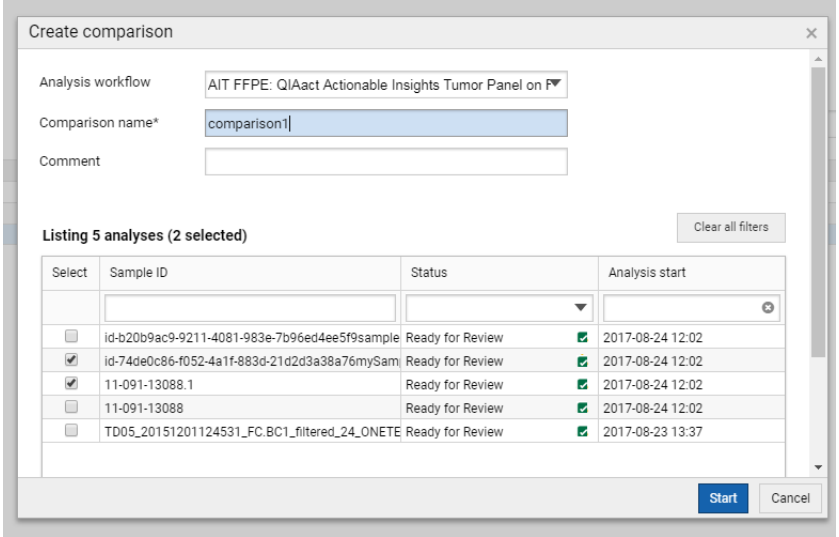


Figure 32: Compare analysis results dialog.

as many samples as necessary for the comparison. How many samples are selected is available in the text directly above the table.

4. Once all the samples to compare are selected, click **Start**.
5. A new entry will be displayed in the Comparison Reports table. It will proceed through the states "In Progress", followed by "Failed" or "Done" when the comparison analysis is completed.
6. Click on the comparison name hyperlink to open the comparison report in a new tab, or use the **Open** button to the lower right of the table.

The compare analysis results report contains the following:

Overview. Overview information about the comparison, and tables listing QC statistics for sequencing, secondary analysis and, when applicable, unique molecular indexing for each sample.

Coverage. An overview table is shown with a row for each gene showing statistics across all compared samples. For analysis workflows with both raw and UMI read coverage information, an overview table is displayed for both. When a row is selected, a table is displayed below containing a row for each sample showing coverage at the same thresholds as in the sample reports.

Reported variants, Ready for review variants and All variants. A variant summary is shown as a table. Each row describes a specific variant statistics across all compared samples. Metrics include only the samples where the variant occur, except for the column "Freq std

dev*" for which a frequency of zero is included for samples in which the variant was not detected. The All variants table includes an additional status column indicating whether a variant originally was reported as part of "Reported variants" or "Ready for review variants". Click on a row to open an additional table below the variants overview table. This table displays the variant annotations from the sample analysis reports for the selected variant in each sample.

CNVs. If the analysis workflow reports on CNVs, an overview table is shown with statistics of detected CNVs across all compared samples. When a row is selected, a table is displayed below, showing the statistics for each separate sample.

Fusions. If the analysis workflow reports on Fusions, an overview table is shown with statistics of detected fusions across all compared samples. When a row is selected, a table is displayed below, showing the statistics for each separate sample.

The Comparison reports can be exported in Excel format by clicking the "Download" action situated at the bottom of the report, or below the Comparison Reports table.

Benchmark analysis results

A "Benchmark analysis results" comparison facilitates the comparison of variants detected for a sample with the variants expected for that sample. This is especially useful for reference samples used to validate the workflow, and for verifying that results have not changed after an update to the analysis workflow. Before starting a benchmark comparison, the user must have available a VCF file with expected variants (Gold standard *.vcf file). In addition, it is possible to use a VCF file with background variants, i.e., variants confirmed for the sample material's parental cell line that should be completely ignored in the benchmark analysis.

To start benchmarking, click Benchmark analysis results (Figure 33).

Note that results from the fusion gene detection workflow cannot be used for benchmark comparisons.

1. In the **Analysis workflow** field, choose the relevant workflow from the drop-down menu. All the samples that were analyzed with the selected workflow are now listed in the table.
2. Name the benchmarking analysis in the **Comparison name** field. Writing a comment is optional.
3. Specify the Accepted frequency deviation determining the maximum accepted deviation between expected and detected frequency for the true positive variants (optional).

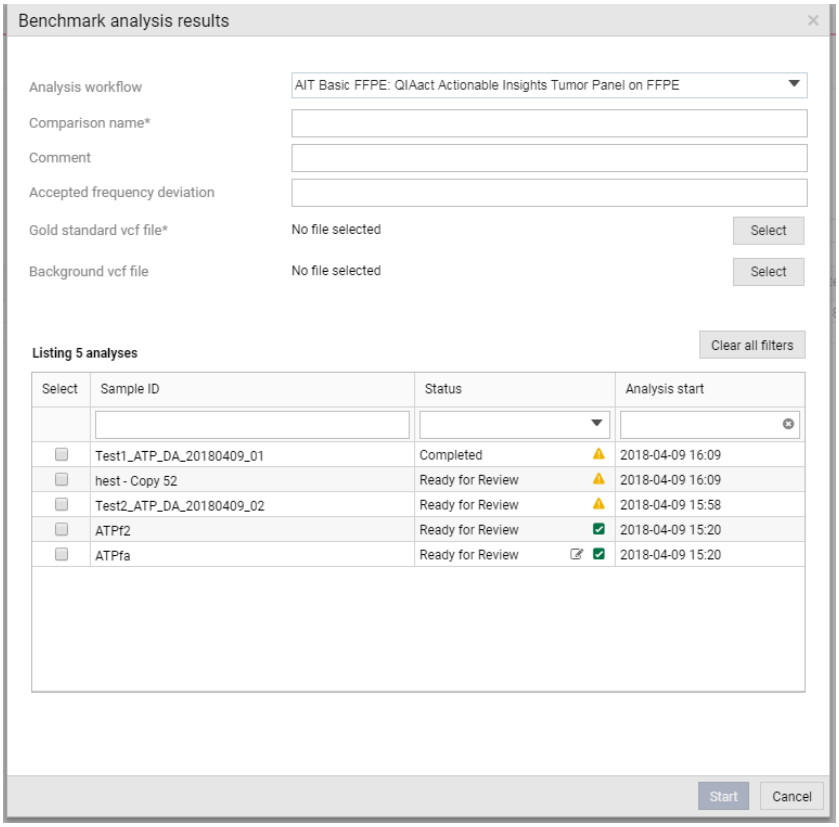


Figure 33: Compare analysis results dialog.

4. Select the *.vcf file with the expected variants to be used as Gold standard. You can also select a *.vcf file containing background variants.
   Note: The VCF file can be 5 MB at most.

5. Now use the checkboxes to select the samples to compare. To find specific samples, use the **Filter and search** options available in the headers of the table. It is possible to select as many samples as necessary for the comparison. How many samples are selected is available in the text directly above the table.

6. Once all the samples to compare are selected, click **Start**.

7. A new entry will be displayed in the Comparison Reports table. It will proceed through the states "In Progress", followed by "Failed" or "Done" when the comparison analysis is completed.

8. Click on the comparison name hyperlink to open the comparison report in a new tab, or use the **Open** button to the lower right of the table.

A benchmarking report contains the following:

Overview. Overview information about the comparison and QC statistics for sequencing and

read mapping of each sample. It is a combination of the Input data and Secondary analysis summary sections of the sample analysis reports. It also lists information about the VCF file(s) used for the benchmark.

Benchmark. A benchmark table summarizes the number of detected variants that was expected (TP = true positives), not expected (FP = false positives), and the number of missing variants (FN = false negatives). Sensitivity, specificity, precision and false positive rates are calculated based on these values. QC statistics (bases that are above minimum and significant coverage) are also shown to determine if samples with poor variant performance also had poor QC statistics.

Expected variants. Two expected variant tables are shown: detected (or true positives) and undetected (or false negatives). Tables are different when benchmarking a single sample compared to benchmarking multiple samples. In single sample mode, the statistics are shown in the main table itself. In multi-sample mode, only median and standard deviation values are shown in the main table. Selecting a variant in that table will open a sub-table below that show the statistics for each separate sample. If an expected variant is detected in some samples, and undetected in other, it will be shown in both table 3.1 and table 3.2. (Figure 34).

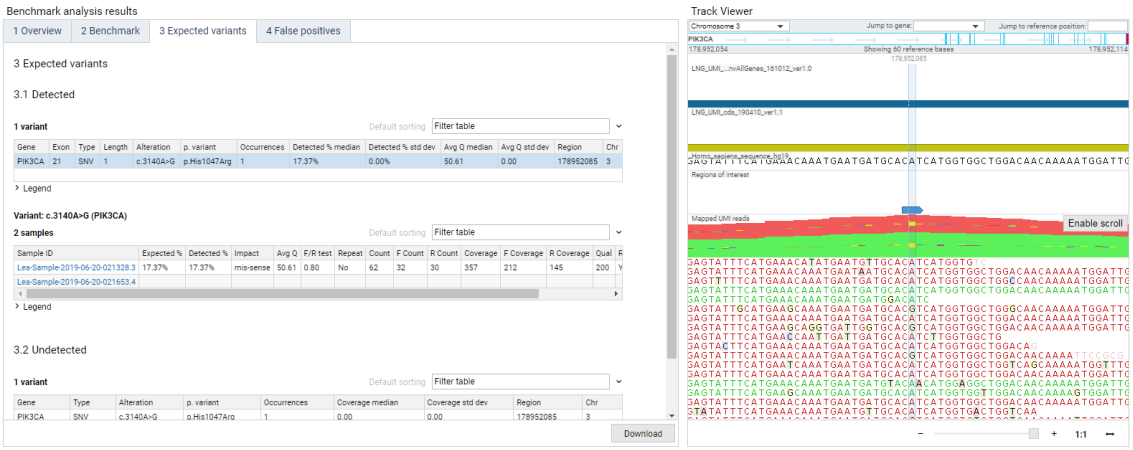


Figure 34: Compare analysis results dialog.

False positive. A false positive table lists variants that are detected in at least one of the samples, but not present in any of the uploaded VCF files. The variant table has the same single sample/multi-sample behavior as the variant tables in the Expected variants tab described above.

When a specific sample is selected for table 3.1, 3.2 or 4.1, the track viewer available in the report will reflect the sample data and zoom to the selected variant position. If the analysis

results have been deleted from QCI Analyze, the read mapping will not be available in the track viewer.

Configuration

The Configuration page (Figure 35) is only accessible to administrators. It lists all analysis workflows available for configuration and execution in QCI Analyze in a table on the left hand side of the screen. Selecting an analysis workflow in the table will display information about the configuration of this workflow in a side panel to the right hand side of the screen.

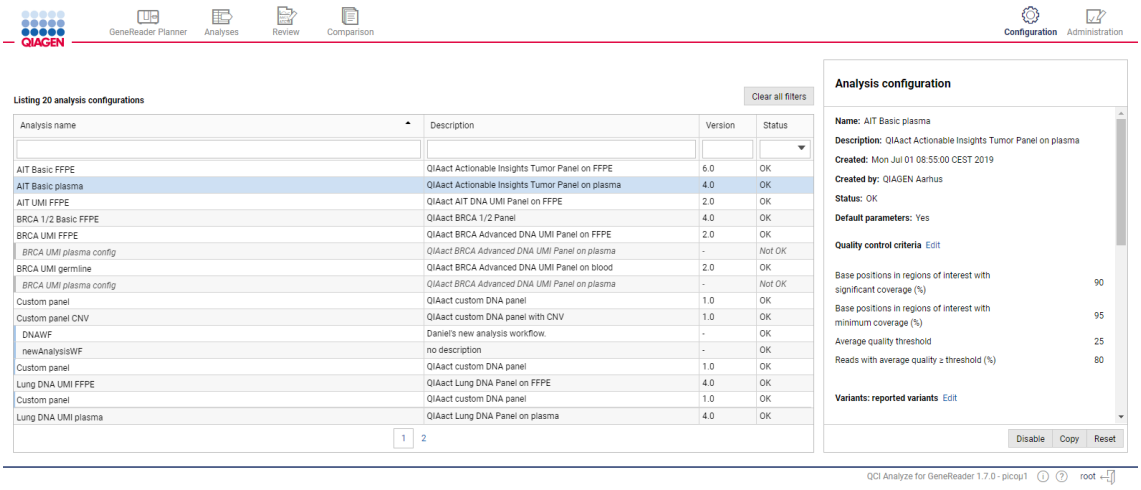


Figure 35: The Configuration page.

Analysis configuration list

In the table, each analysis is listed by the following data:

- Analysis name. This is the short name of the analysis which will be displayed in the Analyses overview in the "Analysis" column.
- Description. This is the full name of the analysis which will be displayed when selecting an analysis in the GeneReader Planner or Start Analysis functionality. This is also the name that will be specified in the final report.
- Version. The version of the analysis workflow.
- Status. If the status is **OK**, the workflow is ready to run. If the status is **Not OK** it

means the Genomics Server did not validate the workflow: verify that all plugins to the Genomics Server are updated to the relevant version required by the workflow, and that the resources bundled with the analysis workflow are available in the relevant location. A **Disabled** workflow can be enabled for use after verification of the various parameters, by using the "Enable" button at the bottom of the side panel.

Analysis configuration panel

Actions

When an analysis in the list is selected, analysis information and configuration options are available from the Analysis configuration side panel. At the bottom of the side panel, the following actions are available:

- Disable. The **Disable** action allows an administrator to disable analysis workflows that are not needed. When disabled, the workflow will no longer be available on the Analyses page and it will no longer be possible to create comparison reports based on results previously generated using that workflow. QCI Analyze reports and comparisons already created are not affected. It will still be possible to create custom regions of interest and CNV baselines using previously generated results. An Enable action is available to re-enable a disabled workflow.

- Copy. When available, the action **Copy** allows an administrator to create a copy of an analysis workflow. In the pop-up dialog, the name and description of the copy must be specified and additional items, such as custom regions of interest and a CNV baseline, can be selected. The copy will inherit the parameter settings from the original analysis workflow, but parameters can subsequently be edited in the copy. Each analysis workflow copy will obtain a unique workflow id, which will be listed in the configurations panel below the description. This id is needed when setting up a new workflow in GeneRead Link (see Analysis workflow copies and GeneRead Link configuration changes). In the analysis configurations overview table, the copy will be listed below the original analysis workflow and highlighted with a vertical blue line to the right of the name of the copy.

- Delete. This action deletes the workflow. Note that it is not possible to delete an original analysis workflow if a baseline or Region of Interest file created with this workflow has been configured for one of analysis workflow copies.

- Reset. This action resets all configurations listed in the Analysis configuration panel to the default values. The QCI Interpret configuration items will not be affected.

Analysis workflow copies and GeneRead Link configuration changes. If the GeneReader workflow is managed via GeneRead Link, to leverage a QCI Analyze analysis workflow copy the GeneRead Link test setup must be configured to point to that copy. The necessary configuration depends on which of the following two scenarios is relevant:

1. Only the analysis workflow copy should be used as it replaces the original analysis workflow. In this case, the existing test in GeneRead Link should be modified to point to the analysis workflow copy instead of the original analysis workflow.
2. It should be possible to use both the original analysis workflow and the analysis workflow copy. In this case, the existing GeneRead Link test should remain as is and a new test added that points to the analysis workflow copy.

In both scenarios, you need to log into GeneRead Link with Administrator privileges.

Scenario 1.

- In the Configuration area of GeneRead Link, go to the **Tests** tab (see GeneRead Link User Manual section 8.4.2).
- Find the name of the test you wish to configure with the new QCI Analyze analysis workflow. Make a note of which **Application** the test is based upon.
- Check to see if other tests are using the same application. All tests using that application will be affected by the change the analysis workflow configuration. If this is not the intention, follow guidance for Scenario 2 instead.
- Go to the **Workflow Management** tab (see GeneRead Link User Manual section 8.4.5).
- Access the Application configurations table as described in the GeneRead Link User Manual and locate the relevant application in the list.
- Open the application configuration as described in the Modifying an application section of the GeneRead Link User Manual.
- In the **Configure application** dialog, locate the field **QCI Analyze workflow name** and insert the unique workflow id for the newly created QCI Analyze analysis workflow (Figure 36).
- Click **OK**.

The GeneRead Link configuration is now updated, and samples processed using the affected GeneRead Link tests will be analyzed with the new analysis workflow.

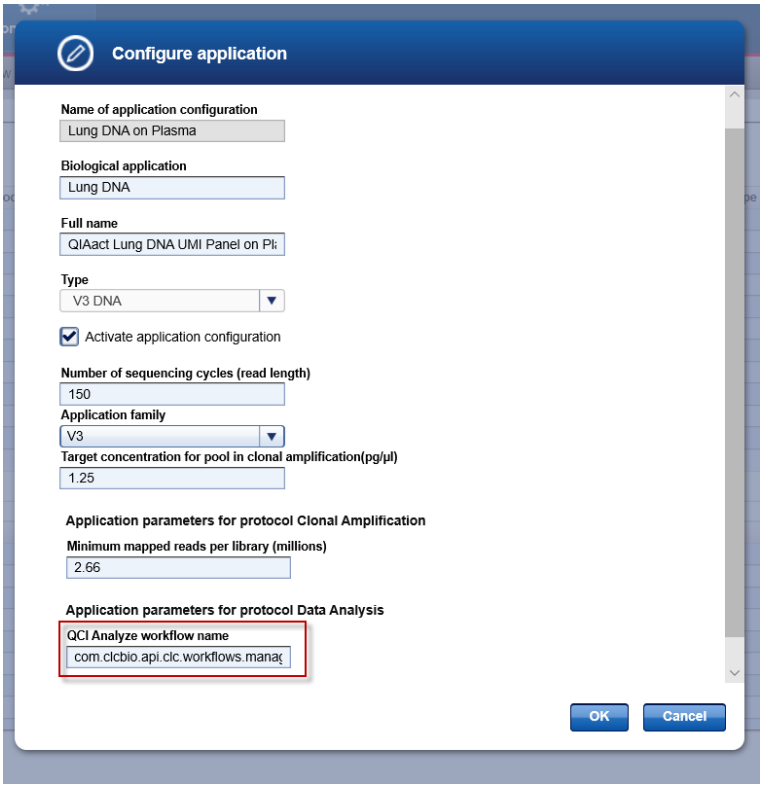


Figure 36: The GeneRead Link Configure application dialog.

Scenario 2.

- In the Configuration area of GeneRead Link, go to the **Tests** tab (see GeneRead Link User Manual section 8.4.2).
- Find the name of the test you wish to copy. Make a note of which **Application** the test is based upon.
- Go to the **Workflow Management** tab (see GeneRead Link User Manual section 8.4.5).
- Access the Application configurations table as described in the GeneRead Link User Manual and locate the relevant application in the list.
- Open the application configuration as described in the Modifying an application section of the GeneRead Link User Manual.
- In the **Configure application** dialog, make a note of all the settings except for the **QCI Analyze workflow name** and click **Cancel**.
- Click **Add new application** as described in the Adding a new application panel section.

- In the **Configure application** dialog give the application a unique name and use the configuration values of the original application noted down previously. In the **QCI Analyze workflow name** field insert the unique workflow id for the newly created QCI Analyze analysis workflow (Figure 36).
- Click **OK**
- Click **Release configuration version**
- Go back to the **Tests** tab and create a new test as described in the Creating a new test section in the GeneRead Link User Manual. In the **Application** drop-down select the newly created application.

GeneRead Link will now display the new test which enables samples to be processed and analyzed with the new QCI Analyze analysis workflow.

Analysis workflow details and parameter settings

The Analysis configuration panel lists information about the selected analysis workflow: name, description, date the analysis workflow was created and by whom, and finally if the analysis configuration is currently set to all default parameters or if one or more parameters have been altered by a user. This information will also be reported in the History section of QCI Analyze reports.

Various parameters affecting variant detection and reporting can be edited from this panel. The configuration options available depends on the analysis workflow.

Parameters for DNA panels

Quality control criteria can be edited to control the stringency of quality control.

- Base positions in regions of interest with significant coverage. This value defines what percentage of the region of interest must be covered with at least the significant coverage for quality control to pass.
- Base positions in regions of interest with minimum coverage. This value defines what percentage of the region of interest must be covered with at least the minimum coverage for the quality control to pass.
- Average quality threshold. This value defines the average quality threshold X for the quality control criteria "Reads with average quality $\geq$ X".
- Reads with average quality $\geq$ threshold. This value specifies the percentage of reads that should pass the average quality threshold X.

- Base positions in high-sensitivity regions with significant coverage. This value defines what percentage of high sensitivity regions must be covered with at least the significant coverage for quality control to pass.
- Base positions in high-sensitivity regions with minimum coverage. This value defines what percentage of high sensitivity regions must be covered with at least the minimum coverage for the quality control to pass.

Variants: reported variants. Variants that meet the defined thresholds will automatically be considered valid and reported in table 3.1.

- Significant (UMI) coverage threshold
- SNV/MNVs frequency thresholds
- Insertions, deletions and replacements frequency thresholds
- Variant quality (Qual) filter (optional, see below for details)
- Average base quality (Avg Q) filter (optional, see below for details)
- F/R test filter (optional, see below for details)

Variants: variants available for review. Variants that meet the thresholds listed below but which do not meet the thresholds for "Reported variants" described above will be available in table 3.2 for manual review.

- Minimum (UMI) coverage threshold. Regions with coverage below this threshold will be considered "untested", i.e., the coverage is too low for reliable testing. The threshold must be lower than, or equal to, the Significant coverage value.
- SNV/MNV frequency threshold. The threshold must be lower than, or equal to, the corresponding threshold for Reported variants.
- Insertions, deletions and replacements frequency threshold. The threshold must be lower than, or equal to, the corresponding threshold for Reported variants.
- Detect variants outside regions of interest. By design, table 3.1 contains only variants detected within the ROI. Setting this parameter to "Yes", will include variants detected outside of the regions of interest in table 3.2, allowing users to review them. Setting the parameter to "No" means that variants outside of ROI will not be found in any variant tables, and hence will not be included in the report.
- Variant quality (Qual) filter (optional, see below for details)
- Average base quality (Avg Q) filter (optional, see below for details)
- F/R test filter (optional, see below for details)

- Report post-filtered variants. When "Yes", table 3.2 will include variants that were not included in table 3.1 due to non-configurable filters such as a homopolymer filter. Post-filters are documented in Appendix A.

Qual, Avg Q and F/R test filters. These filters can be applied and configured to adjust the specificity of variant reporting. The values described below are reported for each variant in the variant tables, allowing users to refer to already analyzed samples to estimate how a particular filter would impact variant reporting.

- Variant quality (Qual) filter. Qual is a PHRED-scaled variant significance value. A Qual value of 10 indicates a 1 in 10 chance that the called variant is an error, while a Qual of 100 indicates a 1 in 10^{10} chance that the called variant is an error. By enabling this filter with a value of 100, all variants with Qual < 100 will be discarded and not included in the report. Qual is capped at 200 for p=1, but it is recommended to set the threshold to less than 100. When setting the filter to a higher Qual value, calculations reach the precision limits of the underlying architecture and can lead to small differences in the lists of variants resulting from the workflow.
- Average base quality (Avg Q) filter. This value is the average quality of all the bases supporting a variant. The higher this value is, the less likely it is that the variant is a sequencing error. Avg Q 30 and above are considered high quality. For example, by enabling this filter with a value of 25, all variants with Avg Q < 25 will be discarded and not included in the report.
- F/R test filter. This test assesses the relative forward/reverse read distribution, i.e., if the forward/reverse ratio for reads that support the variant (counts) is similar to the forward/reverse ratio of the total number of reads covering the position (coverage). The value ranges from 0 to 1, where values closer to 0 may indicate a sequencing artifact. With very high coverage, even small relative differences in allele frequencies (which are most pronounced for low frequency variants) may lead to an F/R test value that is 0 or very close to 0. Hence, in situations where the coverage is high and the frequency low, the value should not be used for automatic filtering, but rather as a guide towards variants that deserve closer inspection. For example, by enabling this filter with a value of 0.1, all variants with F/R Test < 0.1 will be discarded and not included in the report.
- Position test filter. The position test filter compares the position of the variant allele in the reads supporting it (the read position distribution) with the read position distribution for all reads covering that site. Values range from 0 to 1. Values close

to 0 mean the two distributions are very different, and are suggestive of sequencing artifacts.

- Singleton UMI proportion filter. The proportion of UMIs supporting the variant that are based on only one sequencing read (singletons). The value ranges from 0 to 1. The closer to 1, the higher the proportion of singleton UMIs, and the lower the confidence in the allele.

CNV detection. CNV detection configuration is only available for analysis workflows with CNV detection capability.

- Maximum p-value, amplification. Only copy number amplifications with p-values lower than this threshold be considered significant. The higher you set this value, the more CNVs will be predicted.
- Maximum p-value, deletion. Only copy number deletions with p-values lower than this threshold be considered significant. The higher you set this value, the more CNVs will be predicted.
- Minimum fold change, amplification, absolute value. For a copy number amplification to be reported, the absolute fold change must exceed this value.
- Minimum fold change, deletion, absolute value. For a copy number deletion to be reported, the absolute fold change must exceed this value.

  Guidance on copy number, fold change and fold change cutoff is provided in Appendix E.

Subset genes. This section allows the administrator to select and thereby limit reporting to a subset of the pre-configured genes. Clicking **Edit** displays a list of all genes for which the analysis workflow was set up, i.e., the genes targeted by the corresponding panel. When a gene is deselected, variants in this gene will be omitted from the report.

The upper text field can be used to filter the list for specific genes and subsequently select/deselect these.

To set up a gene subset based on a pre-defined list of genes, first deselect all genes by unchecking the select-all box at the top, enter your list of comma-separated genes into the text field, click **Enter** to filter, check the select-all box to select your filtered list of genes and **Save**.

The number of genes included in the report is indicated in the Analysis configuration side panel, as well as in the History tab of the analyzed sample.

Custom regions of interest. When configuring a workflow, it is possible to create a custom regions of interest file based on a BED file or on coverage of selected samples that were already analyzed in QCI Analyze.

- Create from BED. Browse to select a BED file on your file system and enter a name and description for the custom ROI. Once imported, a Custom regions of interest report is created and can be reviewed. The report provides information like the number of regions and total size of both the original regions of interest and the new one. From the BED file, only regions that overlap with the workflow target region will be included in the custom ROI. The report section "BED regions" provides details about this aspect for each individual BED file region. The report section "Regions of interest" lists out each of the regions included in the custom ROI.
  The BED file should be using a zero-based index as in the following examples:
  - For a custom ROI to cover position 11169419 of chromosome 1 the BED file should read: chr1 11169418 11169419.
  - For a custom ROI to cover position 32890578 to 32890684 of chromosome 13 the BED file shoul read: chr 13 32890577 32890684.
- Create from samples. Enter a name for the custom ROI, set the required coverage and the percentage of samples with more than required coverage, and select all relevant sample analyses. The resulting ROI will contain the regions that obtained the required coverage in the indicated percentage of the selected samples. Once created, a Custom regions of interest report is created and can be reviewed. The report provides information about the target region and both the original and new regions of interest. Also included is a summary of coverage and secondary analysis metrics for each of the underlying samples. The section "Target regions" provides details about coverage across samples for each of the individual target regions.
- Review. Open the Custom regions of interest report. It is also possible to download it as en Excel file.
- Delete. Remove the custom ROI file from the analysis workflow unless it is already in use in a copy.

It is possible to create multiple custom ROIs for each analysis workflow. These cannot be applied to the original analysis workflow itself, but to its copies by selecting the desired custom ROI in the Copy dialog (see Configuration section of this manual).

QCI Interpret upload. Specifies the QCI Interpret pipeline and test product profile to be used when uploading results to QCI Interpret.

Test product profiles can specify application specific values that should be pre-filled in the QCI Interpret upload interview pages. Build-in analysis workflows are pre-configured with a demo test product profile. This should be replaced for the application specific custom test product profile set up for the individual QCI Interpret account (see the QCI Interpret documentation to learn more about test product profiles).

Parameters for RNA panels

Quality control criteria can be edited to control the stringency of quality control.

- Average quality threshold. This value defines the average quality threshold X for the quality control criteria "Reads with average quality $\geq$ X".
- Reads with average quality $\geq$ threshold. This value specifies the percentage of reads that should pass the average quality threshold X.
- Read count threshold. The higher threshold, the more strict the criterion. A certain number of reads is required to confidently distinguish between a fusion event and wild type.
- Fusion control genes. A number of positive fusion controls are included in a fusion detection analysis workflow. All positive controls should be detected for this criterion to pass.
- DNA contamination threshold. The lower the threshold, the more strict the criterion.

Fusions

- Promiscuity threshold. The maximum number of fusion partners allowed for a gene.
- High frequency fusions
  - Maximum p-value. The maximum p-value for a fusion to be reported.
  - Minimum fusion count. The minimum number of reads supporting a fusion for that fusion to be reported.

  Low frequency fusions (<0.100%)
  - Maximum p-value. The maximum p-value for a fusion to be reported.
  - Minimum fusion count. The minimum number of reads supporting a fusion for that fusion to be reported.

QCI Interpret upload. Allows the user to specify which QCI Interpret pipeline to use for uploading (somatic or hereditary) and what test product profile to use for the upload.

Test product profiles can specify application specific values that should be pre-filled in the QCI Interpret upload interview pages. Build-in analysis workflows are pre-configured with a demo test product profile. This should be replaced for the application specific custom test product profile set up for the individual QCI Interpret account (see the QCI Interpret documentation to learn more about test product profiles).

Administration

The administration web interface (Figure 37) is only accessible to users with administrative privileges, and is composed of five tiles.

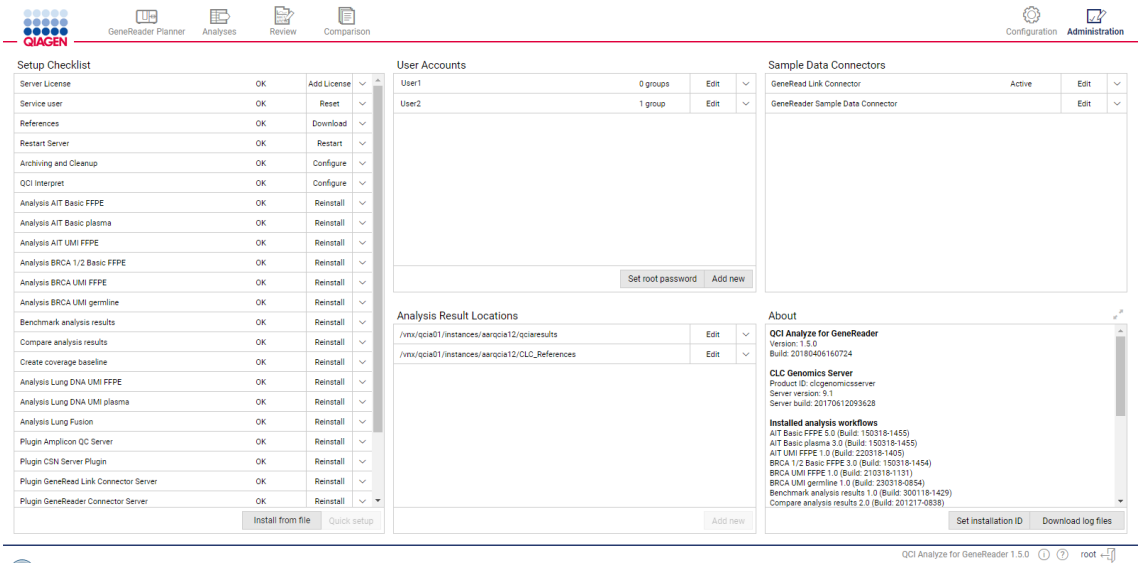


Figure 37: The Administration page.

Note: In addition to having access to this page, administrators can delete individual sample analyses using the button available at the bottom of the Sample analysis details panel in the Analyses page (see Figure 16).

Setup Checklist

This tile lists most of the components that are used to build QCI Analyze: the different plugins and connectors, analysis workflows and comparisons tools. This is also the place where the following functionalities can be managed and configured: Server License and a Service User, Restart Server, References, Archiving and Cleanup, and QCI Interpret connection.

- Server License. Field for entering the QCI Analyze license order ID to download valid license.
- Service User. Defines "seqflow" as the first user account during installation to facili-

tate communication between QCI Analyze and associated software such as GeneReader, GeneRead Link and CLC Genomics Server. Click on Create during installation of QCI Analyze, and later on click on Regenerate to reset credentials to their default values.

- References. The analysis workflows use references that are specific to each panel. Click Download in the References box each time a workflow is installed or reinstalled.
- Restart Server. Use this button to restart CLC Genomics Server. This function is mostly used during installation and when updating the system.
- Archiving and Cleanup. Specify mode of archiving and data cleanup configuration. Set as "Not OK" until automatic archiving is configured (see the Archiving section of this manual).
- QCI Interpret. Configure the connection to QCI Interpret to enable the built-in upload functionality (see the QCI Interpretation configuration section of this manual).
- From Benchmark analysis results to Plugin Variant Track Tools Server. Provide reinstall functions for different plugins and connectors.
- Following rows provide reinstall functions for the analysis workflows, as well as the comparison workflow "Compare analysis results".

Note: The button **Quick setup** allows the user to simply complete the setup procedure by installing all necessary items with just one click. The button **Install from file** can be used to add new analysis workflows to a QCI Analyze installation.

User Accounts

Administrators can add, delete and manage user accounts.

The root user account should be used exclusively for administration tasks, and not for regular use of QCI analyze. The root password can be changed by clicking on **Set root password** at the bottom of the tile.

Unfold a particular user to access additional actions:

- Remove Interpret access. Removes the user's QCI Interpret login credentials. This can be relevant if changes are made to the QCI Interpret configuration.
- Groups. Check the admin box to give the user administrator rights, including access to the Administration and Configuration pages.
- Delete. Deletes the user.

- Edit. Changes the user's password. Users can change their own password independently by clicking on their user name to the left the log out icon (⏻) at the bottom of the user interface. Note: it is not recommended to change the password of the user called "Seqflow" (created during installation), but if the administrator chooses to do so, the new credentials have to be updated on associated software as well.

Analysis Result Locations

This tile enables communication of QCI Analyze with different folders holding the data needed and generated by QCI Analyze. One field specifies the path to the folder where the data generated by the QCI Analyze analysis will be saved. The other specifies the path to the folder holding the references necessary to run the workflows. The location of the folders will be typically set during installation of QCI Analyze.

Sample Data Connectors

A connector allows the automatic import of a specific type of data. The settings will typically be set during installation of QCI Analyze. Administrators specify a standard import folder from which new sequencing data can be extracted. Select the connector to install and click **Edit**.

For GeneReader Sample Data Connector (Figure 38), enter the paths to the relevant folders created during installation.

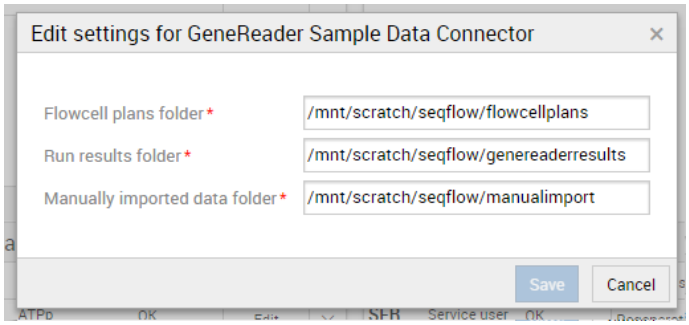


Figure 38: Editing a GeneReader Sample Data Connector.

For the GeneRead Link connector, specify a base address, username and password: these are provided to the user upon installation of GeneRead Link middleware. Also specify the path to a folder for manually imported data.

About

This tile gives information about QCI Analyze, the underlying CLC Genomics Server, a list of the different analysis workflows installed and the configuration of QCI Interpret for GeneReader.

See the software requirements for QCI Analyze for GeneReader in Appendix F.

Set installation ID allows administrators to set an installation ID, (i.e., a recognisable name) for the particular QCIA instance. This ID will appear on the login screen and at the bottom of the user interface as QCI Analyze for GeneReader [version] - [installation ID].

Download log files creates a zip file containing a bug report and all log files that were modified within the last 5 days. In case assistance from our scientific and technical support team is needed, this ZIP file will be of use in initial troubleshooting.

Archiving

QCI Analyze can be configured to automatically archive and clean up sample analysis result files (Figure 39): when a sample analysis has been in the Completed state for a defined number of days, the result files (PDF and Excel report and VCF) are saved as a ZIP file to a preselected storage location. In addition, all data related to the sample-analysis (except for the FASTQ file that was used as input for the analysis) is deleted.

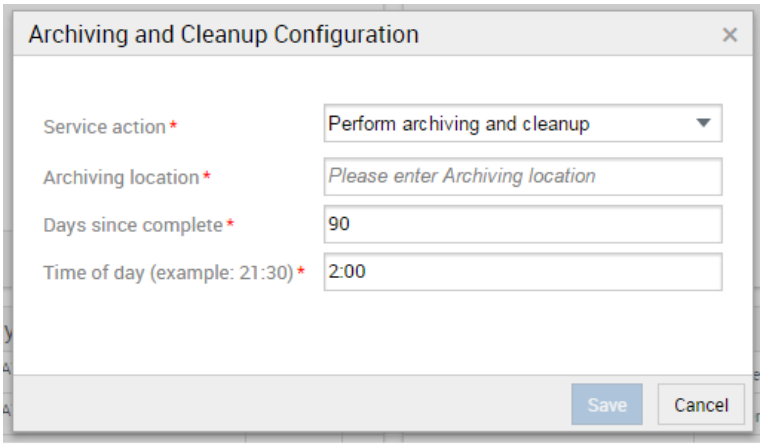


Figure 39: Editing the archiving and cleanup configuration.

The configuration of data cleanup and archiving is accessible from the Archiving and Cleanup item in the **Setup Checklist** tile. The following can be configured:

- Service action. Choose between "Perform archiving and cleanup", "Perform cleanup without archiving", and "Disable archiving and cleanup". **Note**: When QCI Analyze is configured with GeneRead Link, the result files for the analyses started by GeneRead Link are retrieved and archived by GeneRead Link. We thus recommend to choose "Perform cleanup without archiving" to avoid saving twice each sample analysis. However, in that case, results from analyses started manually in QCI Analyze will not be automatically archived before they are deleted. Use the **Download** options for the sample analysis that were started manually to

archive the results.

- Archiving location. Enter the file system path to the folder where the archive packages will be copied to.
- 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.

If archiving fails, the data will not be deleted. As the number of Completed sample analyses will keep growing, it is the user's responsibility to inspect that the archiving configuration is correct, and that enough storage space is available.

Comparison reports based on sample analyses that have been cleaned up remain available.

Linked samples are archived in separate ZIP files.

QCI Interpret configuration

QCI Analyze can be configured to allow seamless upload of analysis results to QCI Interpret for GeneReader, provided that the user has an active QCI Interpret subscription.

Click on **Register** in the QCI Interpret box of the Setup Checklist tile. The following should be specified (Figure 40):

- Server location. Specify the QCI Interpret server location for which your QCI Interpret account has been set up.
- License order ID (normally pre-filled).
- Customer ID

Figure 40: Configuring the QCI Interpret uploader.

Click on **Configure** to enable the samples status to be set to Complete once they have been signed out in QCI Interpret (Figure 41).

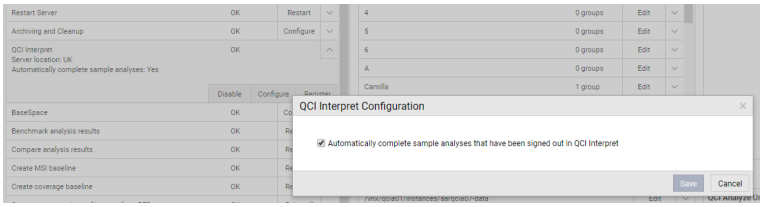


Figure 41: Automatically complete sample analyses signed out of QCI Intepret.

For additional information on QCI Interpret user accounts, group setup and the QCI Interpret uploader, please refer to the QCI Interpret documentation.

Appendix A - Variant filters and settings

The following pages recapitulate the default configuration for quality control, variant calling and filtering, CNV detection, and QCI Interpret upload.

AIT Basic FFPE and AIT Basic plasma

The AIT Basic FFPE and AIT Basic plasma workflows are based on the Hg19 reference sequence.

Table 1: Parameters for AIT Basic FFPE and AIT Basic plasma workflows

	AIT Basic FFPE	AIT Basic plasma
Quality control criteria		
Base positions in regions of interest with significant coverage*	90%	90%
Base positions in regions of interest with minimum coverage*	95%	95%
Reads with average quality ≥ threshold*	80%	80%
Average quality threshold*	25	25
Variant filters, Reported variants, table 3.1		
Significant coverage threshold*	500 reads	500 reads
Minimum frequency for SNV/MNV	4%	0.5%
Minimum frequency for insertions, deletions and replacements	4%	0.5%
Variant quality (Qual) filter*	Disabled	25
Average base quality (Avg Q)*	22	22
F/R test filter*	Disabled	Disabled
Variant filters, Variants available for review, table 3.2		
Minimum coverage threshold*	200 reads	200 reads
Minimum frequency for SNV/MNV*	4%	0.5%
Minimum frequency for insertions, deletions and replacements*	4%	0.5%
Detect variants outside regions of interest*	Yes	No
Variant quality (Qual) filter*	Disabled	25
Average base quality (Avg Q) filter*	22	22
F/R test filter*	Disabled	Disabled
Report post-filtered variants*	No	No
Post-filters**	None	None
QCI Interpret		
QCI Interpret pipeline*	somatic	somatic

* : indicates a configurable parameter, as explained in the Configuration section of this manual. The default values are listed here.

** : post-filters are non-configurable filters that have been applied in addition to those of coverage, frequency, Qual, AvgQ, and F/R Test. Post-filters always apply to table 3.1. To have post-filtered variants listed in table 3.2 choose *Report post-filtered variants: Yes*.

Important note on plasma workflows: A common use case for plasma samples in relation to cancer research is the careful tracking of drug resistance markers as opposed to reliable detection of cancer causing mutations. The analysis workflow AIT Basic plasma has been optimized for this first use case by enabling the detection of low frequency variants from plasma samples at very high sensitivity. The ability to detect low frequency variants inherently leads to lower specificity and higher levels of reported false positive variants. For instance, the AIT Basic plasma workflow with limit of detection of genomic variants down to 0.5% allelic fraction has

an estimated Positive Predictive Value (PPV = true positive / (true positive + false positive)) of 97% indicating that 3 in 100 called variants can be expected to be a false positive due to PCR or sequencing artifacts. In cases of uncertainty, QCI Analyze allows the manual curation of low frequency variants reported in tables 3.1 within the software. Users can inspect variant calls and carefully assess their credibility by considering measures such as frequency or coverage, in addition to visual inspecting the read mapping. Variants, which based on this assessment, are considered artifacts, can be moved out of table 3.1 by changing the review status to "Artifact".

AIT UMI FFPE

The AIT UMI FFPE workflow is based on the Hg19 reference sequence.

Quality control criteria

- Reads with average quality ≥ threshold*: 80%
- Average quality threshold*: 25
- Base positions in regions of interest with significant coverage*: 90%
- Base positions in regions of interest with minimum coverage*: 95%

Variant filters, Reported variants, table 3.1

- Significant coverage threshold*: 100 UMI
- Minimum frequency for SNV/MNV: 3.5%
- Minimum frequency for insertions, deletions and replacements: 3.5%
- Variant quality (Qual) filter*: 100
- Average base quality (Avg Q)*: 25
- F/R test filter*: 0.001

Variant filters, Variants available for review, table 3.2

- Minimum coverage threshold*: 60 UMI
- Minimum frequency for SNV/MNV*: 3.5%
- Minimum frequency for insertions, deletions and replacements*: 3.5%
- Detect variants outside regions of interest*: No
- Variant quality (Qual) filter*: 100
- Average base quality (Avg Q) filter*: 25
- F/R test filter*: Disabled
- Report post-filtered variants*: No
- Post-filters**:
  - Minimum count: 3
  - FGFR3:c.2130C>T: Filter allele if F/R balance = 0

CNV detection

- Maximum p-value, amplification*: 0.005

- Maximum p-value, deletion*: 0.005
- Minimum fold change, amplification, absolute value*: 2.3
- Minimum fold change, deletion, absolute value*: 2.3

QCI Interpret

- QCI Interpret pipeline*: Somatic

* : indicates a configurable parameter, as explained in the Configuration section of this manual. The default values are listed here.
**: post-filters are non-configurable filters that have been applied in addition to those of coverage, frequency, Qual, AvgQ, and F/R Test. Post-filters always apply to table 3.1. To have post-filtered variants listed in table 3.2 choose *Report post-filtered variants: Yes*.

BRCA 1/2 Basic FFPE

The BRCA 1/2 Basic FFPE workflow is based on the Hg19 reference sequence.

Quality control criteria

- Reads with average quality ≥ threshold*: 80%
- Average quality threshold*: 25
- Base positions in regions of interest with significant coverage*: 98%
- Base positions in regions of interest with minimum coverage*: 99%

Variant filters, Reported variants, table 3.1

- Significant coverage threshold*: 500 reads
- Minimum frequency for SNV/MNV: 5.0%
- Minimum frequency for insertions, deletions and replacements: 5.0%
- Variant quality (Qual) filter*: Disabled
- Average base quality (Avg Q)*: 20
- F/R test filter*: 0.0

Variant filters, Variants available for review, table 3.2

- Minimum coverage threshold*: 200 reads
- Minimum frequency for SNV/MNV*: 5.0%
- Minimum frequency for insertions, deletions and replacements*: 5.0%
- Detect variants outside regions of interest*: No
- Variant quality (Qual) filter*: Disabled
- Average base quality (Avg Q) filter*: 20
- F/R test filter*: Disabled
- Report post-filtered variants*: No
- Post-filters**:
  - Filter InDels with frequency < 8% if in homopolymer regions ≥ 7 bp
  - BRCA2:c.6336A>G: Filter allele if frequency < 10.0 or count ≤ 2

QCI Interpret

- QCI Interpret pipeline*: Somatic

* : indicates a configurable parameter, as explained in the Configuration section of this manual. The default values are listed here.

** : post-filters are non-configurable filters that have been applied in addition to those of coverage, frequency, Qual, AvgQ, and F/R Test. Post-filters always apply to table 3.1. To have post-filtered variants listed in table 3.2 choose *Report post-filtered variants: Yes*.

BRCA UMI FFPE and BRCA UMI germline

The BRCA UMI FFPE and BRCA UMI germline workflows are based on the Hg19 reference sequence.

Table 2: Parameters for BRCA UMI FFPE and BRCA UMI germline workflows

	BRCA UMI FFPE	BRCA UMI germline
Quality control criteria		
Base positions in regions of interest with significant coverage*	90%	90%
Base positions in regions of interest with minimum coverage*	95%	95%
Reads with average quality ≥ threshold*	80%	80%
Average quality threshold*	25	25
Variant filters, Reported variants, table 3.1		
Significant coverage threshold*	100 UMI	50 UMI
Minimum frequency for SNV/MNV	4%	20%
Minimum frequency for insertions, deletions and replacements	4%	20%
Variant quality (Qual) filter*	100	100
Average base quality (Avg Q)*	30	30
F/R test filter*	0.001	0.0
Variant filters, Variants available for review, table 3.2		
Minimum coverage threshold*	60 UMI	30 UMI
Minimum frequency for SNV/MNV*	4%	20%
Minimum frequency for insertions, deletions and replacements*	4%	20%
Detect variants outside regions of interest*	No	No
Variant quality (Qual) filter*	100	100
Average base quality (Avg Q) filter*	30	30
F/R test filter*	Disabled	Disabled
Report post-filtered variants*	No	No
Post-filters**		
Minimum count	3	3
Homopolymer filter	Filter InDels with frequency < 8% if in homopolymer regions ≥ 7 bp	Not applied
QCI Interpret		
QCI Interpret pipeline*	somatic	germline/ hereditary

* : indicates a configurable parameter, as explained in the Configuration section of this manual. The default values are listed here.

** : post-filters are non-configurable filters that have been applied in addition to those of coverage, frequency, Qual, AvgQ, and F/R Test. Post-filters always apply to table 3.1. To have post-filtered variants listed in table 3.2 choose *Report post-filtered variants: Yes*.

Polyclonality-derived sequencing artifacts: Due to polyclonality during clonal amplification, variants called due to sequencing artifacts may appear in variant table 3.2, and in rare cases table 3.1. These false positive variants can be easily identified, as they have the following characteristics:

- Variant allele frequency will be approximately 5%.

- Variants are found in one of the following regions:
  - chr17:g.41246229.. 41246270 (BRCA1)
  - chr17:g.7577450..7577463 (BRCA1)
  - chr13:g.32972824..32972843 (BRCA2)
- Variants are found in only one sequencing direction, i.e. either forward or reverse count will be 0, or at max 1.
- Reads support the variant contain multiple variants in the 10bp regions upstream and downstream of it. In the reads track, this can be recognized as a "noisy" looking region (see Figure 42).

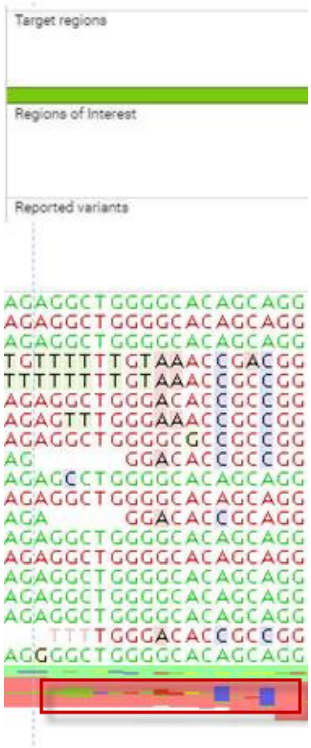


Figure 42: Polyclonality-derived sequencing artifacts are associated with "'noisy"' looking regions in the reads track.

Myeloid DNA UMI

The Myeloid DNA UMI workflow is based on the Hg19 reference sequence.

Quality control criteria

- Reads with average quality ≥ threshold*: 80%
- Average quality threshold*: 25
- Base positions in regions of interest with significant coverage*: 90%
- Base positions in regions of interest with minimum coverage*: 95%
- Base positions in high-sensitivity regions significant coverage*: 80%
- Base positions in high-sensitivity regions with minimum coverage*: 99%

Variant filters, Reported variants, table 3.1

- Significant coverage threshold*: 100 UMI
- Minimum frequency for SNV/MNV: 3.5%
- Minimum frequency for insertions, deletions and replacements: 3.5%
- Variant quality (Qual) filter*: 100
  Decreased threshold for KIT:c.2447 (KIT D816V): 10
- Average base quality (Avg Q)*: 30
  Decreased threshold for KIT:c.2447 (KIT D816V): 25
- F/R test filter*: 0.001

Variant filters, Variants available for review, table 3.2

- Minimum coverage threshold*: 44 UMI
- Minimum frequency for SNV/MNV*: 3.5%
- Minimum frequency for insertions, deletions and replacements*: 3.5%
- Detect variants outside regions of interest*: No
- Variant quality (Qual) filter*: 100
  Decreased threshold for KIT:c.2447 (KIT D816V): 10
- Average base quality (Avg Q) filter*: 30
  Decreased threshold for KIT:c.2447 (KIT D816V): 25
- F/R test filter*: Disabled
- Report post-filtered variants*: No
- Post-filters**:

- Internal tandem duplications (*insertion**)***: only include if in gene FLT3 and if:
  - the mapping signature is "'tandem duplication"' and count >9, or
  - the mapping signature is "'self-mapped"'
- Table 3.1: Filter InDels if in homopolymer regions ≥ 8 bp
- Table 3.2: Filter InDels if in homopolymer regions ≥ 8 bp and if frequency < 5%

High sensitivity regions

- The higher coverage of sensitivity regions allows for detection of variants of lower frequency than for the rest of the panel. Variants detected in high sensitivity regions are triaged for tables 3.1 and 3.2 using dedicated frequency and coverage threshold set (see Table 3). Positions in high sensitivity regions for which coverage dropped below the minimum coverage threshold defined for that position will be listed in the Untested regions table.

Table 3: Frequency and coverage parameters for high sensitivity regions

	Minimum frequency	Significant coverage	Minimum coverage
KIT	0.70%	200 UMI	44 UMI
KIT:c.2447 (KIT D816V)	0.25%	500 UMI	44 UMI
JAK2	0.70%	200 UMI	44 UMI

Note that in contrast to most of the frequency and coverage thresholds for the rest of the panel, those of the high sensitivity regions are not configurable. Analysis workflows with high sensitivity regions will also have an additional set of coverage-based Quality control criteria (see below).

QCI Interpret

- QCI Interpret pipeline*: Somatic

* : indicates a configurable parameter, as explained in the Configuration section of this manual. The default values are listed here.

** : post-filters are non-configurable filters that have been applied in addition to those of coverage, frequency, Qual, AvgQ, and F/R Test. To have post-filtered variants listed in table 3.2 choose *Report post-filtered variants: Yes*.

*** : the Myeloid DNA UMI analysis workflow supports detection of insertions that exceed the length of the sequencing reads, specifically internal tandem duplications for the gene FLT3. The presence of these structural variants is inferred based on information on unaligned read ends. These variants are reported as type "insertion*" to differentiate them from regular, shorter

insertions that are detected by the variant caller based on gaps introduced in reads during alignment. For detailed information about mapping signatures, see the relevant section of the QIAGEN CLC Genomics Workbench manual, http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Theoretically_expected_structural_variant_signatures.html.

Lung DNA FFPE and Lung DNA plasma

The Lung DNA FFPE and Lung DNA plasma workflows are based on the Hg19 reference sequence.

Table 4: Parameters for Lung DNA FFPE and Lung DNA plasma workflows

	Lung DNA UMI FFPE	Lung DNA UMI plasma
Quality control criteria		
Base positions in regions of interest with significant coverage*	90%	90%
Base positions in regions of interest with minimum coverage*	95%	95%
Reads with average quality ≥ threshold*	80%	80%
Average quality threshold*	25	25
Variant filters, Reported variants, table 3.1		
Significant coverage threshold*	100	200
Minimum frequency for SNV/MNV	4%	0.22%
Minimum frequency for insertions, deletions and replacements	4%	0.22%
Variant quality (Qual) filter*	60	60.0
Average base quality (Avg Q)*	25	25
F/R test filter*	0.0	0,001
Variant filters, Variants available for review, table 3.2		
Minimum coverage threshold*	60	120
Minimum frequency for SNV/MNV*	4%	0.22%
Minimum frequency for insertions, deletions and replacements*	4%	0.22%
Detect variants outside regions of interest*	Yes	No
Variant quality (Qual) filter*	60	60
Average base quality (Avg Q) filter*	25	25
F/R test filter*	Disabled	Disabled
Report post-filtered variants*	No	No
Post-filters**		
Homopolymer filter	Not applied	Filter InDels with frequency ≤ 1% if in homopolymer regions ≥ 3 bp
Proportion singleton UMIs 0.3	Not applied	If Proportion of singleton UMIs ≤ 0.3 then filter variants if AvgQ ≤ 30
Proportion singleton UMIs 0.5	If Proportion of singleton UMIs ≤ 0.5 then filter variants if AvgQ ≤ 30	If Proportion of singleton UMIs ≤ 0.5 then filter variants if AvgQ ≤ 25
Low count	Not applied	If count <= 2 then filter variants if Qual <= 150
CNV detection		
Maximum p-value, amplification*	0.05	0.05
Maximum p-value, deletion*	0.05	0.05
Minimum fold change, amplification, absolute value*	1.4	1.4
Minimum fold change, deletion, absolute value*	1.4***	1.4***
QCI Interpret		
QCI Interpret pipeline*	somatic	somatic

* Indicates a configurable parameter, as explained in the Configuration section of this manual. The default values are listed here.

** Post-filters are non-configurable filters that have been applied in addition to those of coverage, frequency, Qual, AvgQ, and F/R Test. Post-filters always apply to table 3.1. To have post-filtered variants listed in table 3.2 choose *Report post-filtered variants: Yes*.

*** Updated recommendation, July 2020: Minimum fold change, deletion, absolute value: 1.8

Important note on plasma workflows: A common use case for plasma samples in relation to cancer research is the careful tracking of drug resistance markers as opposed to reliable

detection of cancer causing mutations. The QCI Analyze plasma workflow Lung DNA UMI plasma (for QIAact Lung DNA Panel) has been optimized for this first use case by enabling the detection of low frequency variants from plasma samples at very high sensitivity. In cases of uncertainty, QCI Analyze allows the manual curation of low frequency variants reported in tables 3.1 within the software. Users can inspect variant calls and carefully assess their credibility by considering measures such as frequency or coverage, in addition to visual inspecting the read mapping. Variants, which based on this assessment, are considered artifacts, can be moved out of table 3.1 by changing the review status to "Artifact".

Lung Plasma Track

The Lung Plasma Track workflow is based on the Hg38 reference sequence.

The workflow supports detection of a specified list of variants, insertions and deletions, as well as insertions and deletions in general for predefined regions in the genes MET and EGFR. All other variants, insertions and deletions are removed prior to the filtering described below.

Quality control criteria

- Reads with average quality $\geq$ threshold*: 80%
- Average quality threshold*: 25
- Base positions in regions of interest with significant coverage*: 80%
- Base positions in regions of interest with minimum coverage*: 90%

Variant filters, Reported variants, table 3.1

- Significant coverage threshold*: 1000 UMI

SNVs and MNVs

- Minimum frequency*: 0.05%
  Increased threshold for variants in genes TP53 and NTRK1**: 0.15%
- Variant quality (Qual) filter*: 3.0
- Average base quality (Avg Q)*: 14.0
- F/R test filter*: Disabled
- Position test filter*: Disabled
- Singleton UMI proportion filter*: 0.7

Insertions, deletions and replacements

- Minimum frequency*: 0.025%
- Variant quality (Qual) filter*: 3.0
- Average base quality (Avg Q)*: 14.0
- F/R test filter*: Disabled
- Position test filter*: Disabled
- Singleton UMI proportion filter*: 0.7

Variant filters, Variants available for review, table 3.2

- Minimum coverage threshold*: 100 UMI

SNVs and MNVs

- Minimum frequency*: 0.05%
  Increased threshold for variants in genes TP53 and NTRK1**: 0.15%
- Variant quality (Qual) filter*: 3.0
- Average base quality (Avg Q)*: 14.0
- F/R test filter*: Disabled
- Position test filter*: Disabled
- Singleton UMI proportion filter*: 0.7

Insertions, deletions and replacements

- Minimum frequency*: 0.025%
- Variant quality (Qual) filter*: 3.0
- Average base quality (Avg Q)*: 14.0
- F/R test filter*: Disabled
- Position test filter*: Disabled
- Singleton UMI proportion filter*: 0.7

CNV detection

- Maximum p-value, amplification*: 0.005
- Maximum p-value, deletion*: 0.005
- Minimum fold change, amplification, absolute value*: 1.6
- Minimum fold change, deletion, absolute value*: 1.6

QCI Interpret

- QCI Interpret pipeline*: Somatic

* : indicates a configurable parameter, as explained in the Configuration section of this manual. The default values are listed here.

**: To ensure the desired specificity, the minimum frequency cutoff for SNVs and MNVs in genes TP53 and NTRK1 is higher than for the remaining genes. On the QCI Analyze report History, this configuration is documented as Specificity filter 1 and 2. Note that in contrast to the frequency thresholds for the rest of the panel, those of the Specificity filters are not configurable.

Appendix B - Fusion filters and settings

This page lists the default configuration for quality control, fusion detection, and QCI Interpret upload.

Lung Fusion

The Lung Fusion workflow is based on the Hg38 reference sequence.

The algorithm has been designed to detect fusions between pre-annotated intron-exon junctions. Thus, only fusions where both breakpoints sit within 8 bp of an intron/exon junction are reported.

Quality control criteria

- Reads with average quality $\geq$ threshold*: 80%
- Average quality threshold*: 25
- Read count threshold*: 400,000
- Fusion control genes*: Yes
- DNA contamination threshold*: 0.1%

Filters, Reported fusions, table 3.1

- Promiscuity threshold*: 20
- High frequency fusions:
  - Maximum p-value*: 0.05
  - Minimum fusion count*: 5
- Low frequency fusions (<0.100%):
  - Maximum p-value*: 1.0
  - Minimum fusion count*: 15

Post-filters**

- Remove exon skipping fusion events except for MET-MET
- Remove fusion event DDX11{NM_030653.3}:r.1_1493_KRAS{NM_033360.3}:r.182_5889
- Remove fusions TPM3-TPM4, TPM4-TPM3, and HACL1-COLQ

QCI Interpret

- QCI Interpret pipeline*: Somatic

* : indicates configurable parameter, as explained in the Configuration section of this manual.
**: post-filters are non-configurable filters that have been applied in addition to those of p-value and fusion count. These serve to remove e.g. fusion calls derived from common read-through mRNAs or caused by gene homology.

Appendix C - References and annotation information

Human reference sequence, ENSEMBL 682 MB

For Hg19 workflows, **Human reference sequence refseq GRCh37.p13 interim**

`ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/GRCh37.p13_interim_annotation/`

- Gene/mRNA/CDS annotations: `interim_GRCh37.p13_top_level_2017-01-13`

- Gene names have been updated to reflect those of `GRCh38.p13_109.20190607`

For Hg38 workflows, **Human reference sequence refseq GRCh38.p9**

`ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/vertebrate_mammalian/Homo_sapiens/all_assembly_versions/GCF_000001405.35_GRCh38.p9/`

- Gene/mRNA/CDS annotations: `Gene/mRNA/CDS from GRCh38.p13_109.20190607`

Transcripts and coding regions used for annotating variants with c. and p. information in DNA analysis workflows: In cases where multiple transcripts exist for a gene, the transcript most frequently reported in the literature is used, as per the QIAGEN Knowledge Base, Wonderland release.

Appendix D - Detailed QC Report

The report is divided in per-sequence (Quality control for reads) and per-base analyses. In the per-sequence analyses, some characteristic (a single value) is assessed for each sequence and contributes to the overall assessment. In per-base assessments each base position is examined and counted independently.

Quality control for reads

- Average base quality of reads calculates the amount of sequences that feature individual PHRED-scores in 64 bins from 0 to 63. The quality score of a sequence is calculated as arithmetic mean of its base qualities. PHRED-scores of 30 and above are considered high quality.
- GC content of reads counts the number of sequences that feature individual %GC-contents in 101 bins ranging from 0 to 100%. The %GC-content of a sequence is calculated by dividing the absolute number of G/C-nucleotides by the length of that sequence, and should look like a normal distribution in the range of what is expected for the human genome. If the GC-content is substantially lower (the normal distribution is shifted to the left), it may be that GC-rich areas have not been properly covered. A non-normal distribution, or one that has several peaks indicates the presence of contaminants in the reads.
- Ambiguous base content of reads counts the number of sequences that feature individual %N-contents in 101 bins ranging from 0 to 100%, where N refers to all ambiguous base-codes as specified by IUPAC. The %N-content of a sequence is calculated by dividing the absolute number of ambiguous nucleotides through the length of that sequence. This distribution should be as close to 0 as possible.

Quality control for bases

- Quality score per base position calculates the amount of bases that feature individual PHRED-scores in 64 bins from 0 to 63. This results in a three-dimensional table, where

dimension 1 refers to the base-position, dimension 2 refers to the PHRED-score and dimension 3 to amounts of bases observed at that position with that quality score. PHRED-scores above 20 are considered good quality. It is normal to see the quality dropping off near the end of reads.

- Nucleotide content per base position displays the overall percentage of A, C, T and G for each base position along the read. If there is a GC bias with changes at specific base positions along the read length this could indicate that an over-represented sequence is contaminating the original library.

Quality control of Regions of Interest positions

This graph displays coverage distribution for individual base positions by correlating base-positions with the number of sequences that supported (covered) that position. It is not included in the fusion detection report.

Quality control for Unique Molecular Indexing

For the workflows leveraging the UMI technology, the additional information is available.

- Unique Molecular Index sizes (size < 50) shows the distribution of UMI group sizes, with size being defined as the number of reads that share the same UMI. The graph is cut off at a UMI size of 50 reads to highlight the distribution of the smaller groups.
- Average quality scores of UMI reads shows the distribution of the average quality scores of UMI reads, including the ones based on single reads.
- Quality scores of UMI reads per base position displays the quality scores of UMI reads by the base-pair position.

Appendix E - Fold change cutoff

When configuring the minimum fold change thresholds for calling CNVs, it can be useful to understand the difference between copy number and fold change and the relationship between tumor fold change, sample fold change and sample purity.

Copy number and fold change

The copy number gives the number of copies of a gene. For a normal diploid sample the copy number, or ploidy, of a gene is 2.

The fold change is a measure of how much the copy number of a case sample differs from that of a normal sample. When the copy number for both the case sample and the normal sample is 2, this corresponds to a fold change of 1 (or -1).

The sample fold change can be calculated from the normal copy number and sample copy number. The formula differs for amplifications and deletions:

$$\text{Fold change, amplifications (CN(sample) >CN(normal))} = \frac{copynumber(sample)}{copynumber(normal)} \quad (1)$$

$$\text{Fold change, deletions (CN(sample) <CN(normal))} = -\frac{copynumber(normal)}{copynumber(sample)} \quad (2)$$

Fold change values for amplifications and deletions are asymmetric in that a 50% *increase* in copy number from 2 to 3 (heterozygote amplification) converts to a fold change of 1.5, whereas a 50% *decrease* in copy number from 2 to 1 (heterozygous deletion), gives a fold change of -2.0. The difference is even more pronounced if we consider what could be interpreted as a homozygote duplication (copy number 4) and a homozygote deletion (copy number 0). Here, the calculated fold changes land at 2 and -∞, respectively.

The fact that the same percent-wise change in coverage (copy number) leads to a higher fold change for deletions than for amplifications means that given the same amplification and deletion fold change cutoff there is a higher risk of calling false positive deletions than amplifications - it

takes less coverage fluctuation to pass the fold change cutoff for deletions.

For the above reasons, it is often useful to differentiate the fold change cutoff for amplifications and deletions.

	Copy number	Fold change
Amplifications		
	2	1
	3	1.5
	4	2
	6	3
	8	4
Deletions		
	2	-1
	1	-2
	0.5	-4
	0.2	-10
	0.1	-20
	0	-∞

Figure 43: The relationship between copy number and fold change for amplifications and deletions.

Tumor fold change, sample fold change and sample purity

Given a sample purity of $X\%$, and a desired detection level (absolute value of fold change in 100% pure sample) of T, the following formula gives the required fold change cutoff:

$$\text{cutoff} = \frac{X\%}{100\%} \times T + (1 - \frac{X\%}{100\%}) \tag{3}$$

For example, if the sample purity is 40%, and you want to detect 6-fold amplifications or deletions (e.g. 12 copies instead of 2, or 2 copies instead of 12), then the cutoff should be:

$$\text{cutoff} = \frac{40\%}{100\%} \times 6 + (1 - \frac{40\%}{100\%}) = 3.0. \tag{4}$$

Figure 44 shows the required fold change cutoffs needed to detect a particular degree of amplification/deletion at different sample purities. Figure 45 zooms in for low-level amplifications and deletions.

The Copy Number Variant Detection tool calls CNVs that are both global outliers on the target-level, and locally consistent on the region-level.

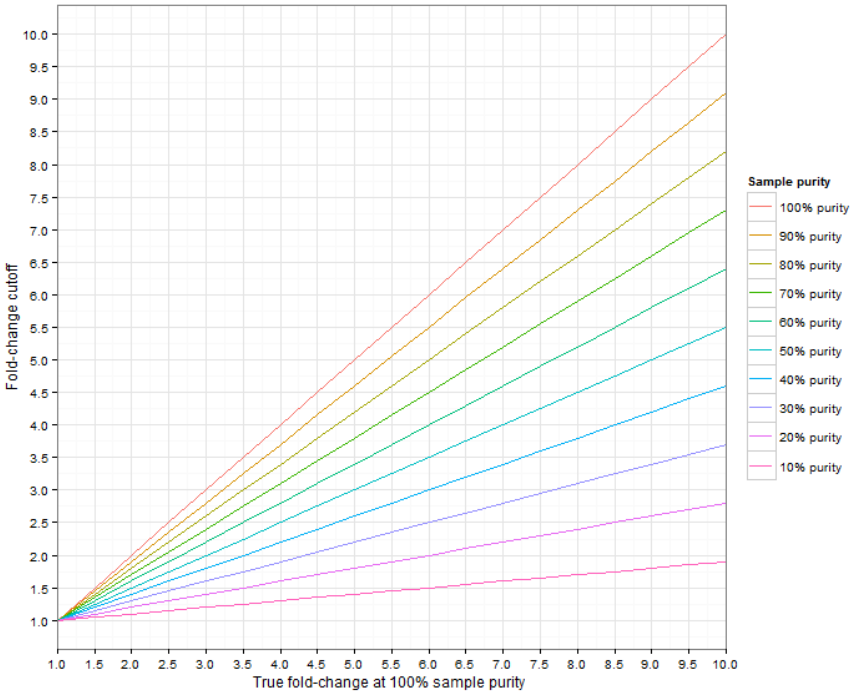


Figure 44: The required fold change cutoff to detect amplifications and deletions of different magnitudes, as a function of sample purity.

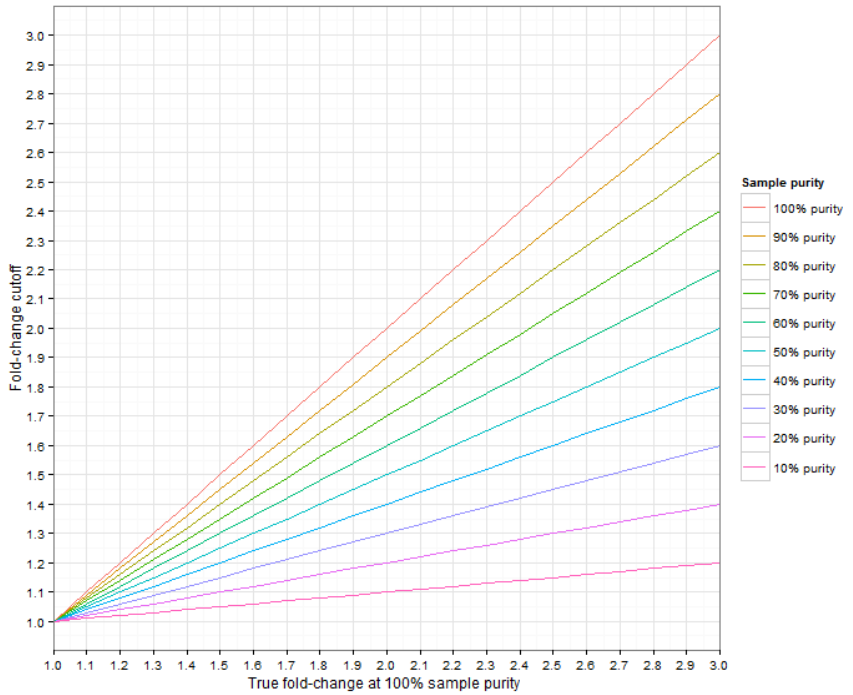


Figure 45: Low-level amplifications and deletions: The required fold change cutoff needed to detect amplifications and deletions of different magnitudes, as a function of sample purity.

Appendix F - QCI Analyze software requirement

System requirements

Here are the system requirements for the QCI Analyze for Genereader software.

Table 5: System requirements for QCI Analyze

Server	Minimum requirement
Operating system	Red Hat® Enterprise Linux® 6.x (64bits), Red Hat Enterprise Linux 7.x (64bits), CentOS® 7.x (64bit)
Processor	Intel® or AMD CPU 2 cores required; 8 cores recommended
Main memory	24 GB RAM required; 48 GB RAM recommended
Hard disk space	500 GB required

QCI Analyze client	Minimum requirement
Internet browser	Internet Explorer® 11 or a recent version of Chrome®
Monitor	Resolution at least 1024 pixels

Additional software on computers running QCI Analyze

CLC Genomics Server and QCI Analyze should be installed on dedicated machines, and no other software should be run on these machines. We strongly recommend disabling virus scanner activity during the use of QCI Analyze.

Document revision history

HB-2531_07 07/20	CNV calling removed for the BRCA UMI FFPE and BRCA UMI germline analysis workflows New recommended setting for CNV "Minimum fold change, deletion" for Lung DNA UMI FFPE and Lung DNA UMI plasma, Appendix A
HB-2531_06 05/20	New workflow included: Lung Plasma Track
HB-2531_05 03/20	Lung Fusion post-filters and intron/exon boundary limitation documented Appendix A, BRCA UMI FFPE and BRCA UMI germline, now contains paragraph on how to identify polyclonality-derived sequencing artifacts
HB-2531_04 01/20	Updates releated to QCI Analyze for GeneReader v 1.7
HB-2531_03 12/18	New workflow included: Myeloid DNA UMI Qual definition improved Percentile used for fold change calculation defined Configuration parameters are now presented in Appendix A
HB-2531_02 08/18	New workflows included: BRCA UMI FFPE and BRCA UMI germline