December 2017

QIAGEN Clinical Insight Analyze 1.4.5 User Manual

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

Sample to Insight

QIAGEN

Contents

Introduction

Thank you for choosing QIAGEN® Clinical Insight (QCI™) **Analyze**. 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 1.4.5** User Manual for use with CLC Genomics Server Version 9.1.0 or higher, 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.

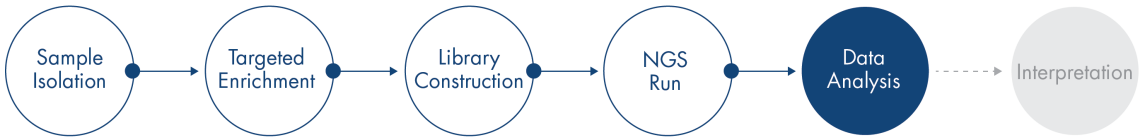


Figure 1: QCI Analyze in the NGS Sample to Insight workflow.

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 event detection in RNA samples.

Variant and CNV detection

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

- The **AIT FFPE** and **AIT 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 **BRCA 1/2 FFPE** workflows analyze data generated using the GeneReader **QIAact BRCA 1/2 Panel** on FFPE samples. The workflow will report variants for the two BRCA genes.
- The **Lung DNA FFPE** and **Lung DNA 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.

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 threshold or, in the case of

the BRCA 1/2 FFPE workflow, confidence scoring) 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 Lung DNA FFPE and Lung DNA plasma workflows 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.

In addition, the Lung DNA FFPE and Lung DNA plasma 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. 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 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 or Deletion depending on the fold-change value. If the fold-change and p-value thresholds are not met, this means that no copy number variation could be detected. This will be indicated as "No CNV detected".

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.
- Some workflows also define Variants of Interest (VOI) as variants that the targeted resequencing protocol was designed to capture, and for which reporting is of particular interest.

As implied by the definitions above, FFPE and plasma workflows for the same panel share the exact same target regions but may differ in their predefined VOI and ROI. Indeed, the difference

in the nature of the FFPE and plasma samples has consequences on sampling depth and amplification efficiency, which in turn implies differences in variants that achieve the performance specifications of the assay.

Configuration of the workflows parameters

While FFPE and plasma 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. The following tables recapitulate the default configuration of the parameters set for calling a variant (Table 1) and for defining which variant will be automatically reported as valid (Table 2).

Table 1: Comparisons of the parameters that differ between workflows

	AIT FFPE	AIT plasma	BRCA 1/2	Lung DNA FFPE	Lung DNA plasma
Detect variants outside of regions of interest*	Yes	No	No	Yes	No
Minimum coverage*	200 reads	200 reads	200 reads	60 UMI	120 UMI
Frequency threshold*	4%	0.5%	5%	4%	0.22%
Minimum count threshold for Low Frequency Variant caller	8 reads	5 reads	8 reads	1 UMI	1 UMI
Minimum C-score	N/A	N/A	0.4	N/A	N/A
Min Qual filter for SNV/MNV	Not applied	25	Not applied	100	60
Min Qual filter for insertions, deletions and replacements	Not applied	25	Not applied	100	100
Minimum average base quality (Avg Q) filter	22	22	15	36	36
Homopolymer region filter	No	No	No	No	Yes[1]
Minimum F/R test filter	No	No	No	>0	>0.01

UMI: Unique Molecular Indices

*: indicates configurable parameters, as explained in the "Analysis Configurations" section of this manual. The default values are listed in this table.

[1]: If the variant is an homopolymeric insertion or deletion ≥3bp, this filter requires that the variant frequency is >1.0 for the variant to pass.

Table 2: Additional filtering parameters for automatic validation of variants (variants assigned to table 3.1 Reported variants)

	AIT FFPE	AIT plasma	BRCA 1/2	Lung DNA FFPE	Lung DNA plasma
Detect variants outside regions of interest	No	No	No	No	No
Significant coverage*	500 reads	500 reads	500 reads	100 UMI	200 UMI
Frequency threshold	4%	0.5%	5%	4%	0.22%

UMI: Unique Molecular Indices

*: indicates configurable parameters, as explained in the "Analysis Configurations" section of this manual. The default values are listed in this table.

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 workflows AIT plasma and Lung DNA plasma (respectively Actionable Insights Tumor GeneRead Panel and QIAact Lung DNA Panel) have 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 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".

Fusion detection

QCI Analyze currently offers one fusion detection analysis workflow:

- The **Lung Fusion** supports secondary analysis of FFPE RNA samples run with the GeneReader **QIAact Lung Fusion Panel**. This targeted RNA panel allows for detection of the 79 predefined fusion transcripts. The panel uses the Unique Molecular Index (UMI) technology.

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

- Reads are mapped to the fusion transcript reference. Subsequently, unmapped reads are mapped to the wild type reference.
- The initial sequencing reads with identical UMI are grouped to create one UMI read for each UMI group.
- The primer part of mapped reads are removed from consideration to not influence the downstream fusion detection.
- Based on information about fusion breakpoints and primer position and orientation, the workflow determines whether the individual UMI reads represent evidence for wild type or fusion transcription. Reads for which evidence for wild type versus fusion is inconclusive are also counted.

- A p-value for each fusion event is calculated as a standard one-tailed binomial test, taking into account the number of reads mapping to the particular fusion transcript and the total number of reads mapping to either the fusion or wild type transcripts for that fusion. A small p-value indicates the presence of a fusion. Based on the p-value and coverage, the individual fusion events are placed in one of the three fusion tables in the report.

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

- Four constitutive genes serve as fusion event positive controls. Control primers target artificial fusion breakpoints in the genes. The resulting "fusion transcripts" and mapping characteristics mimic a true fusion event, and should per definition always be present in a sample. All four controls must be detected for the QC to pass.
- DNA contamination control is facilitated with the use of control primers targeting a non-transcribed intergenic region. This target region is amplified and generates sequencing reads only if genomic DNA is present in the sample. Reads that mapped to neither the fusion nor the wild type transcript references are mapped to the DNA contamination sequence reference. The level of DNA contamination is calculated as the number of reads mapping to the 1000 bp control region divided by the total number of reads. This value must be below 0.1% for QC to pass.

Reference datasets and annotation information

The workflows rely on the presence of particular reference datasets (see Appendix A). 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).

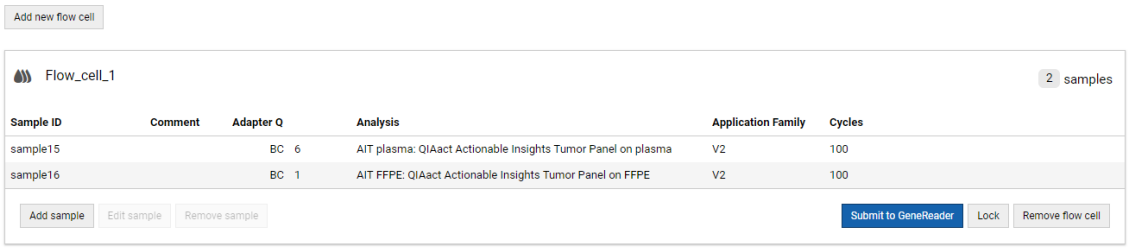


Figure 2: Creating a flow cell and adding samples.

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).

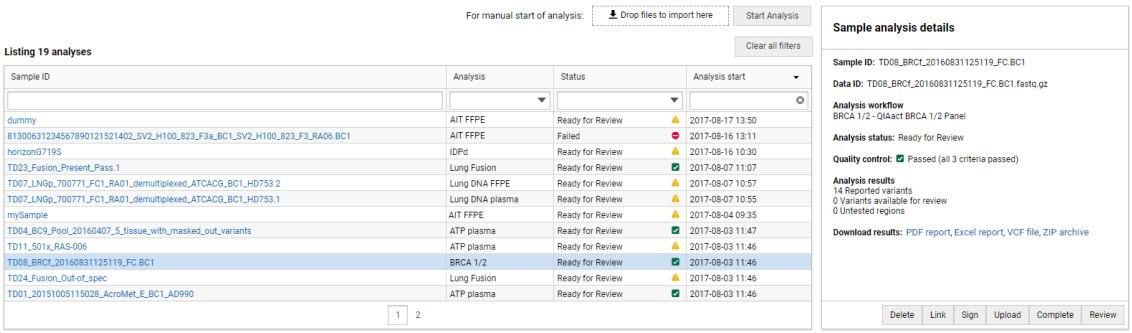


Figure 3: The sample analysis details panel contains a summary of quality control and analysis results.

- "Quality control" indicates how successful the sequencing was. This can be either yellow, *Deviations*, or green, *Passed*. If yellow, the number of criteria for which deviations are observed is indicated.
- "Analysis results" shows how many variants were found and sorted in the three variant tables of the report and, if applicable, how many genes were affected by amplification or deletion.

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 three colored statuses in the Fastq and Secondary analysis summary sections is marked in yellow (Figure 4). Pass criteria are indicated below the tables.

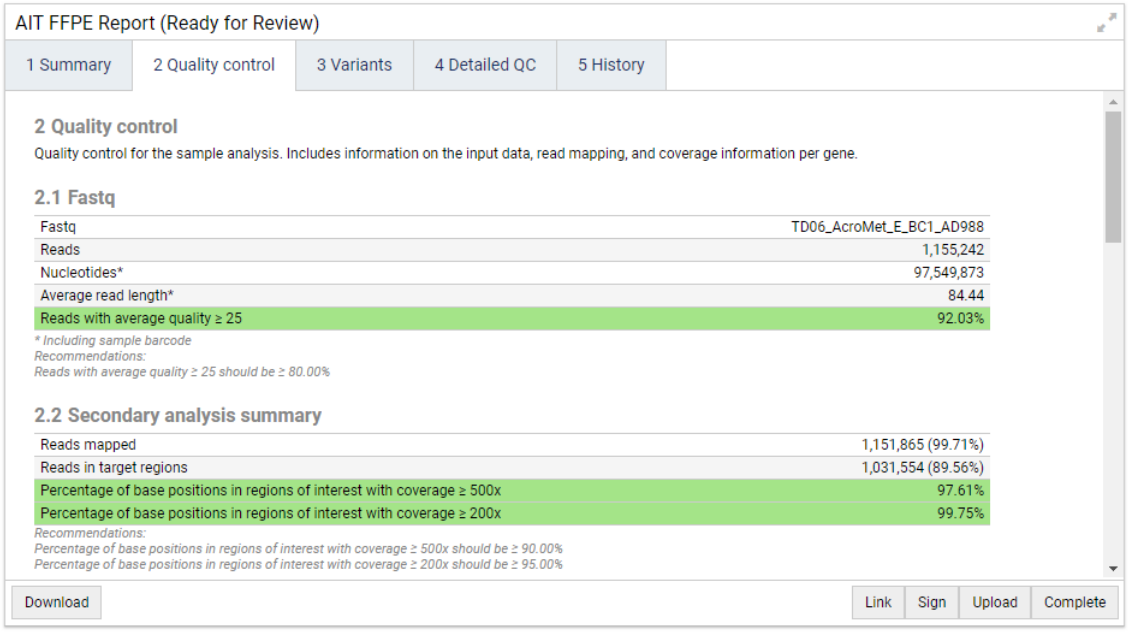


Figure 4: Quality Control section of a variant detection report.

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 three 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.

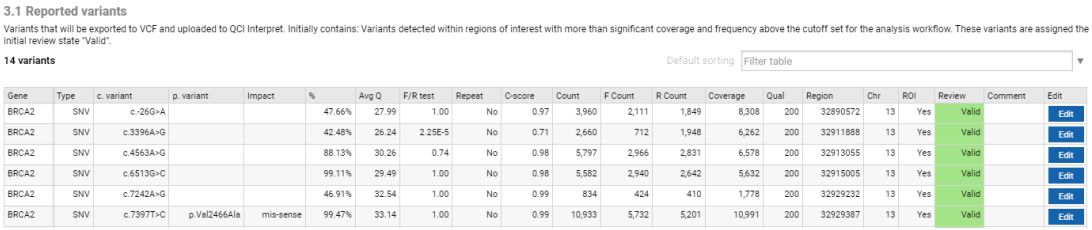
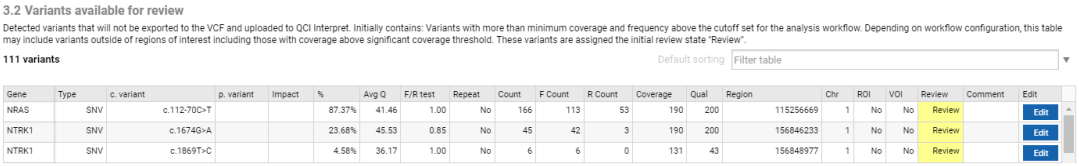


Figure 5: The reported variants table.

   - 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). Extra care should be given to variants marked as Variants of Interest (VOI) - if such an annotation is available in the table.



3.2 Variants available for review

Detected variants that will not be exported to the VCF and uploaded to QCI Interpret. Initially contains: Variants with more than minimum coverage and frequency above the cutoff set for the analysis workflow. Depending on workflow configuration, this table may include variants outside of regions of interest including those with coverage above significant coverage threshold. These variants are assigned the initial review state "Review".

111 variants

Default sorting | Filter table

Gene	Type	c. variant	p. variant	Impact	%	Avg Q	F/R test	Repeat	Count	F Count	R Count	Coverage	Qual	Region	Chr	ROI	VOI	Review	Comment	Edit
NRAS	SNV	c.112-70C>T			87.37%	41.46	1.00	No	166	113	53	190	200	115256669	1	No	No	Review		Edit
NTRK1	SNV	c.1674G>A			23.68%	45.53	0.85	No	45	42	3	190	200	156846233	1	No	No	Review		Edit
NTRK1	SNV	c.1869T>C			4.58%	36.17	1.00	No	6	6	0	131	43	156848977	1	No	No	Review		Edit

Figure 6: The Variants available for review table.

Clicking **Edit** at the end of each row allows the user to validate a variant as "Confirmed by review", or dismiss it as an "Artifact" (Figure 7). "Artifacts" will remain in table 3.2, while "Confirmed" variants will be moved to table 3.1 Reported variants.

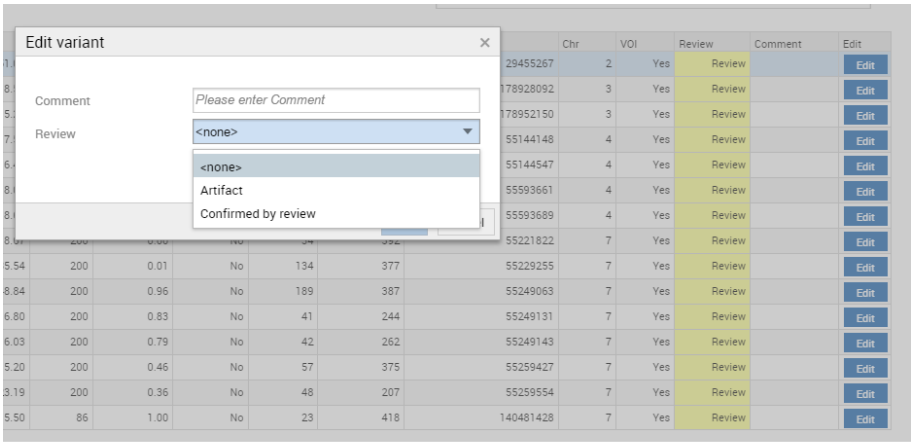


Figure 7: Editing a variant.

- The **Untested variants/Untested regions** (table 3.3) contains variants or regions insufficiently covered to provide a valid test result.

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. This table cannot be edited.

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 resequencing workflow used by QCI Analyze aims to bioinformatically remove the variants introduced by these artifacts, but in order to protect the high sensitivity required for such analyses, some artifact variants may still be reported.

In order 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 8).

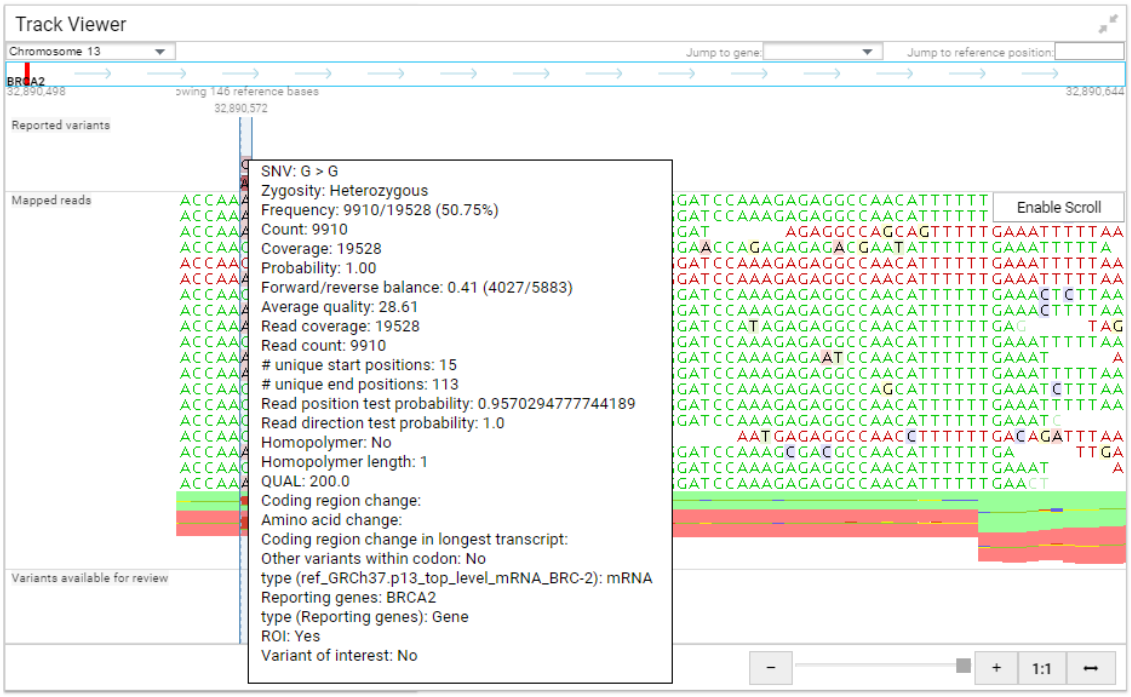


Figure 8: Annotations of a variant seen as a tooltip in the Track Viewer.

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 value should be larger than 0. This 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 zero to 1, where values closer to zero 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 zero. 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.

- The BRCA 1/2 FFPE workflow includes the calculation of a confidence score (C-score) for each variant. The C-score is a probability calculated from the variant's annotations by applying a Multiple logistic regression function that has been trained on a large set of True positive (TP) and False positive (FP) variant calls. The value reflects the probability that the variant is a true variant and not an artifact, as judged from its annotations and the annotations that were observed on the set of TP and FP variants. A C-score value ranges from 0 to 1, with 1 indicating a high-confidence variant.

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 9).

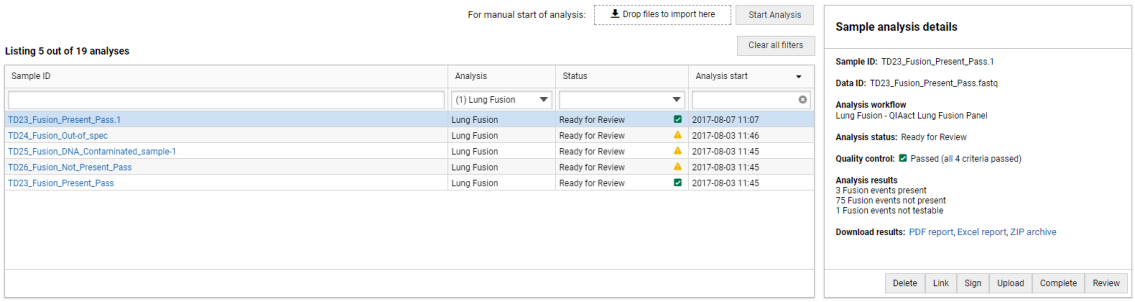


Figure 9: The sample analysis details panel contains a summary of quality control and analysis results.

   - "Quality control" indicates how successful the sequencing was. This can be either yellow, *Deviations*, or green, *Passed*. If yellow, the number of criteria for which deviations are observed is indicated.
   - "Analysis results" shows how many fusions events were found and sorted in the three fusion tables of the report.

3. Click on the Sample ID hyperlink or the **Review** action available in the **Sample analysis details** panel to access the results for the sample.

4. In the Quality Control section, you can check the number and quality of the reads, potential DNA contamination and if the fusion controls were detected (Figure 10).

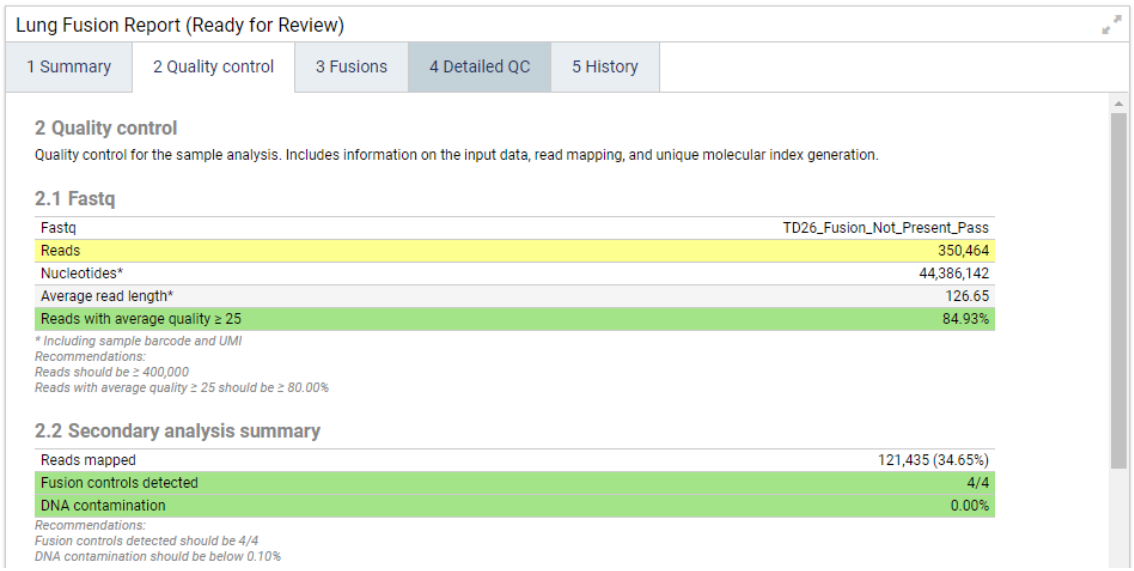


Figure 10: Quality Control section of a fusion detection report.

5. Review the Fusions section of the report. Click on the fusion events of interest to see them in the Track Viewer to the right hand side of the Review page. You can also save the report as a PDF.

Uploading samples to QCI Interpret for GeneReader (optional)

1. To upload the variants of table 3.1, the detected CNVs when applicable, and/or the fusion events to QCI Interpret for GeneReader, click **Upload**.
2. If requested, enter the QCI Interpret for GeneReader credentials provided by QIAGEN and click **Connect to QCI Interpret**.
3. In the following dialog, metadata can be added to the sample. Mandatory fields are indicated with an asterisk.
4. Click **Upload**.

Once uploaded to QCI Interpret for GeneReader, the status of the sample analysis in QCI Analyze will be set to Uploaded to QCI Interpret. Users can access the interpretation report by clicking **Go to** from the report (Review page), or by following the "QCI Interpret test" hyperlink in the Sample analysis details panel (Analyses page). 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 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, the 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 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.

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, and ZIP archive) 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 11).

Figure 11: The navigation bar.

In the top right of the user interface, the name of the user currently logged in is displayed. Click the user icon to log out of QCI Analyze.

Users with administrator privileges have an additional button **Administration** in the navigation bar to provide access to admin-only functionalities.

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 12). 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.

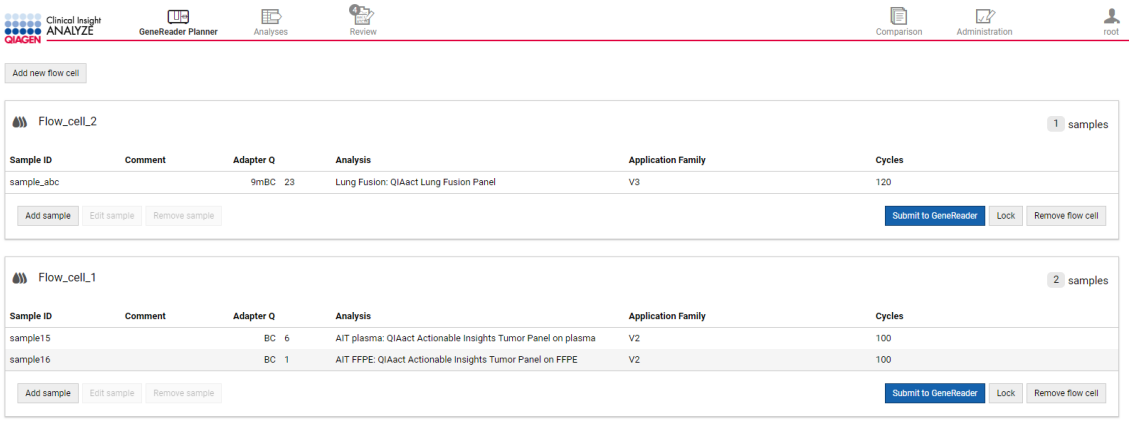


Figure 12: The GeneReader Planner page.

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 13).

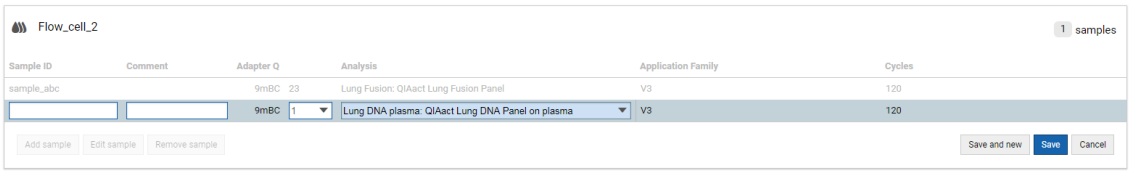


Figure 13: Adding a sample.

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. V2 or V3 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 FFPE: QIAact Actionable Insights Tumor Panel on FFPE), or just the workflow tag (AIT 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 FFPE

56789,,2,AIT plasma
```

Values can be separated by commas, semicolons or are tab delimited. Column separators are automatically detected when importing a valid file (Figure 14).

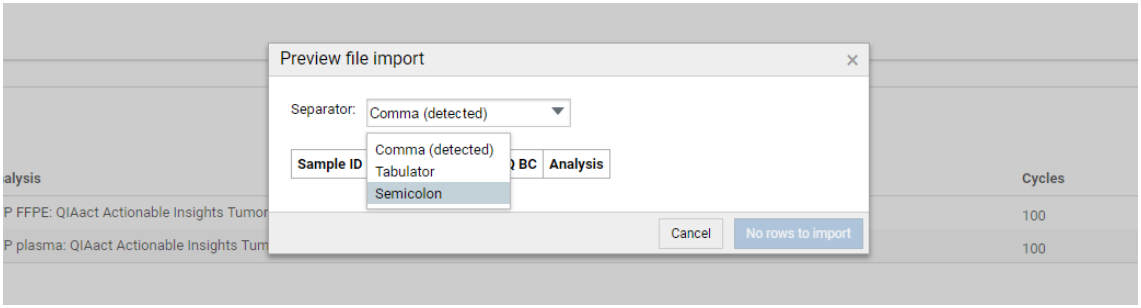


Figure 14: Preview file import dialog upon importing a TXT file.

Errors will be indicated with a pink highlight in the preview dialog (Figure 15).

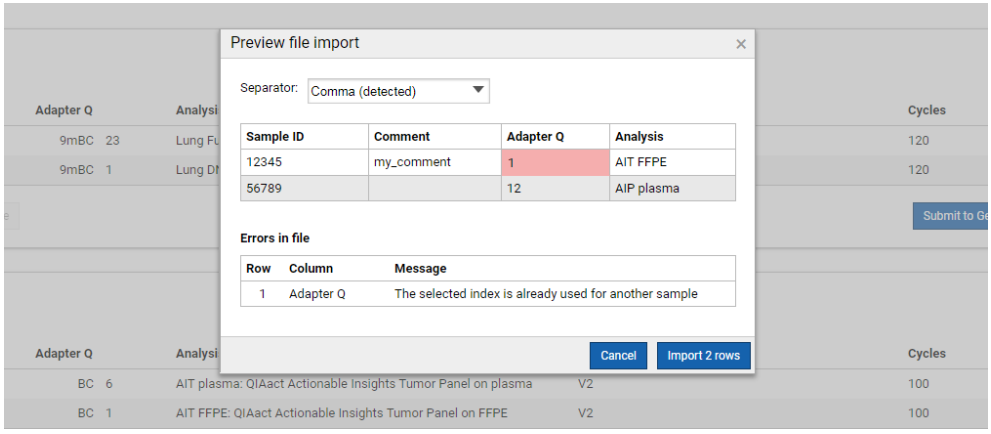


Figure 15: Preview file import dialog in case of an error in the file imported.

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 16). 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 16: 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 17).

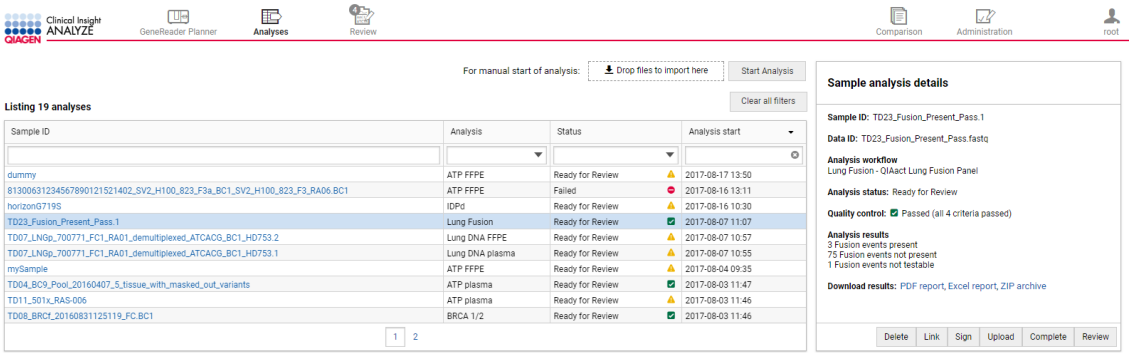


Figure 17: The Analyses page. Note that the **Delete** button at the bottom of the Sample analysis details panel is only available to administrators.

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. 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**.

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.

Clicking **Review** on a Planned sample shows the analysis progress at 1%. It remains as such until the analysis actually starts.

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 18). 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.

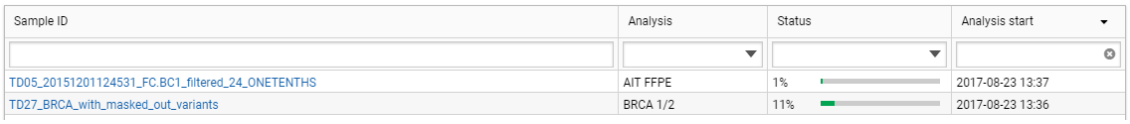


Figure 18: An analysis in the state In Progress is currently 11% complete.

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, press **Clear** to remove the failed analysis data for that sample and start the analysis again via the Start Analysis functionality (described in the Manual start of an analysis section of this manual).

Ready for Review and Uploaded to QCI Interpret

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

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 upload 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 (fusion gene events). The link indicates that both analyses come from the same person, and it ensures that results will be uploaded as one to QCI Interpret for GeneReader, thus generating a single QCI Interpret test per person.

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

The **Upload** functionality allows users to upload the list of reported variants listed in variant table 3.1, CNV information and detected fusion gene events 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. Enter the username and password for QCI Interpret for GeneReader provided by QIAGEN in the relevant fields of the pop-up window that opens upon clicking **Upload**, and click **Connect to QCI Interpret**.

In the following dialog, users can add metadata to the sample. Mandatory fields are marked with an asterisk. 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.

Go to QCI Interpret for GeneReader

For sample analyses uploaded to QCI Interpret for GeneReader, users can access the QCI Interpret test content by clicking the **Go to** button on the QCI Analyze report or follow the "QCI Interpret test" link from the Sample analysis details panel. This will take users to the QCI Interpret for GeneReader website, where log in is required. For information on how to use QCI Interpret for GeneReader, please refer to the QCI Interpret for GeneReader 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 start of an 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 start an analysis. The **Start Analysis** button available in the top of the Analyses page is used for this purpose (see Figure 17 to locate the button). Clicking the button opens a list of sample data that have been imported and are available in QCI Analyze.

If the relevant sample data is not already available in the list, it should first be imported by drag and drop of the relevant FASTQ files to the "Drop files to import here" area next to the Start

Analysis button. Importing starts immediately, but FASTQ files can be large and the import might take a while to complete. Once data import has completed, the sample is available for manual start of analysis.

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

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. Click **OK**. **Note**: in the first column, a check mark indicates if a previous analysis of the data is already available in QCI Analyze.
3. Select the analysis workflow to use from the drop-down menu (Figure 19).
4. Specify a unique sample ID for the sample analysis.
5. Optionally: specify a comment to include in the analysis report.
6. Click **Start**.

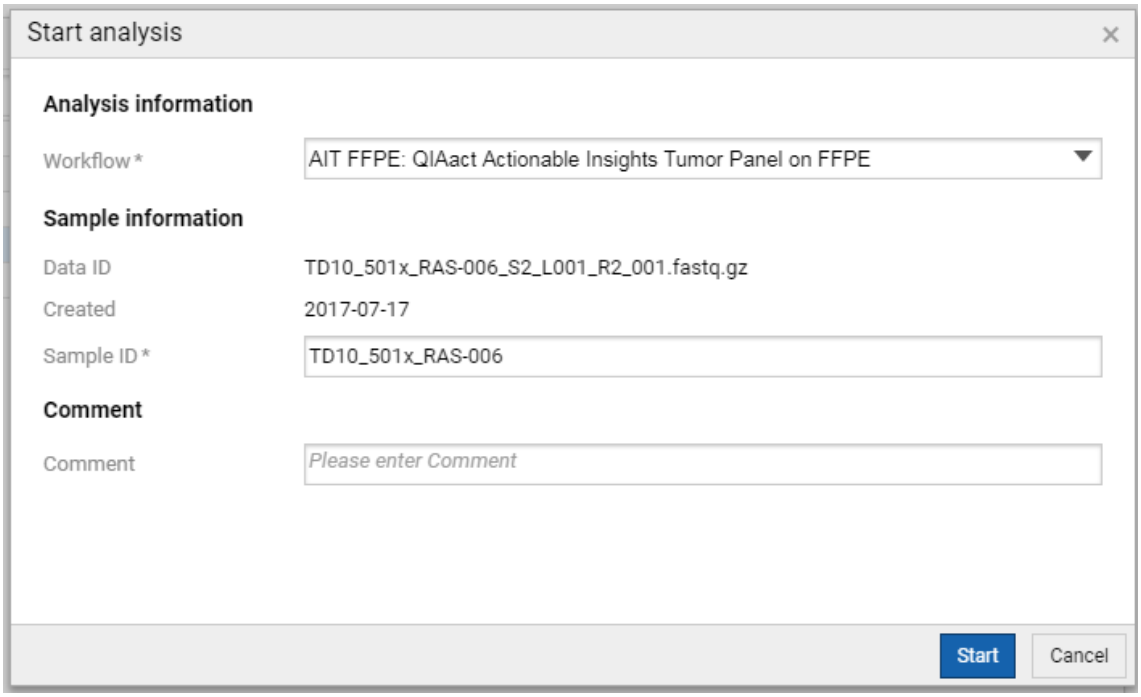


Figure 19: Starting an analysis manually.

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.

Review

Once a sample is opened for review, the analysis results are shown as a report in the **Review** page (Figure 20). 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.

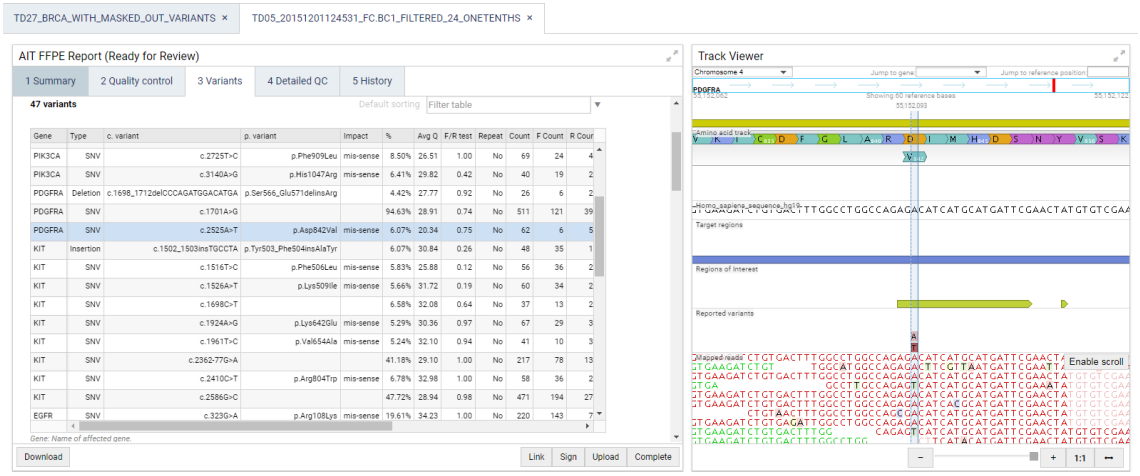


Figure 20: Review page showing the Report tile to the left and the Track Viewer tile to the right.

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. 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 21).

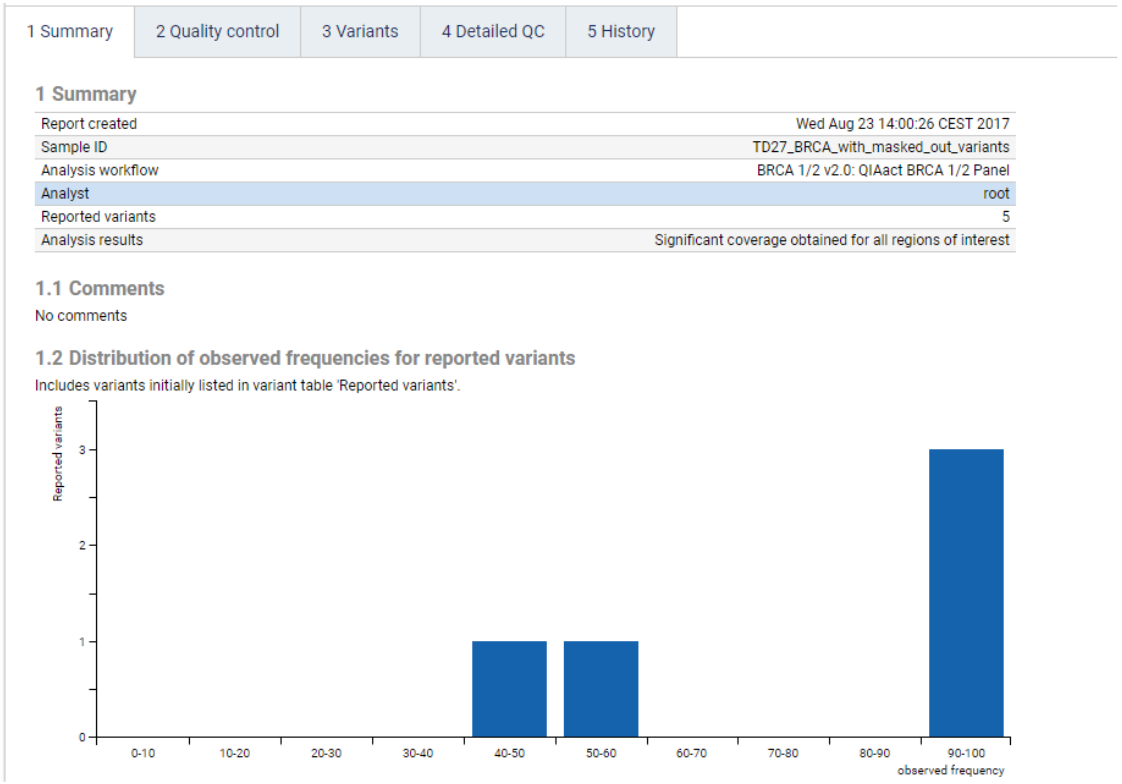


Figure 21: Summary of a report and variant frequency distribution.

Analysis results

- *Significant coverage obtained for all variants of interest*, indicates that the test was successful in providing significant coverage for all variants of interest. No variants are

left for manual review and no variants or positions of interest were left untested, i.e., variant tables 3.2 and 3.3 are empty. Action: No additional action required.

- *Variants available for review*, reflects that minimum coverage was obtained for all variants of interest, but that one or more positions are below significant coverage and are available for manual review, or variants were detected outside Regions of Interest. These variants are listed in variant table 3.2. No variants/regions were left untested, i.e., variant table 3.3 is empty. Action: It is possible to edit the status of variants from table 3.3 after careful review (see the section Variants of this manual).
- *Untested variants/Untested regions*. The test did not cover all positions. Some variants or positions of interest were left untested, i.e., coverage was below the minimum coverage threshold. Action: Re-sequencing will be required to obtain sufficient coverage for all variants/positions of interest.

Quality Control

The Quality Control tab of the Report includes a minimum of four 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 Analysis Configuration section of this manual.

Fastq 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.

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:

Unique Molecular Indexing shows metrics for the UMIs generated by the secondary analysis workflow.

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

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". A third table lists "Untested variants" or "Untested regions".

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 by review", it moves to table 3.1; if set to "Artifact", it remains in table 3.2.

3.3 Untested variants/Untested regions For analysis workflows where a list of "variants of interest" has been defined, table 3.3 lists *Variants of Interest* that could not be tested due to insufficient coverage. For analysis workflows where a list of "variants of interest" has not been defined, table 3.3 lists *regions* where variants could not be tested due to insufficient coverage. Both variants and regions are labeled "Untested".

All variants tables share the same column headers:

- Gene. Name of the affected gene.
- Type. Variant type, such as Single Nucleotide Variant (SNV), Multiple Nucleotides Variant (MNV), insertion, deletion, and replacement.

- c. variant. Coding DNA sequence variant nomenclature based on the Human Genome Variation Society (HGVS) recommendations.
- 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.
- 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 zero to 1, where values closer to zero 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 zero. 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.
- Repeat. Variant is located in a low-complexity region.
- C-score. Value reflecting the confidence that can be put in the variant call, judging from its overall annotations (1: high confidence, 0: low confidence). It is only available for selected analysis workflows.
- 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.
- 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. In rare cases, the Qual value cannot be calculated for specific variant and as a result the Qual field will be empty.

- 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.
- VOI. Variant of interest, as specified by the analysis workflow.
- Review. Status of variant review.
- Comment. Remark added by user during variant review.
- Edit. Allows the user to change the state of a variant or to add a comment.

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 apply the respective filters 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).

Note: A VOI position with a coverage above the significant threshold but showing no allelic difference between reads and reference will not be included in any of the tables as they are safely dismissed as non-variants for the dataset.

CNVs

For workflows including CNV detection, the report includes a CNVs tab where users can access the CNV analysis results (Figure 22). All CNV target genes are listed in a table together with the observed fold change, four Copy number columns (100%, 50%, 25%, 10%), and finally the

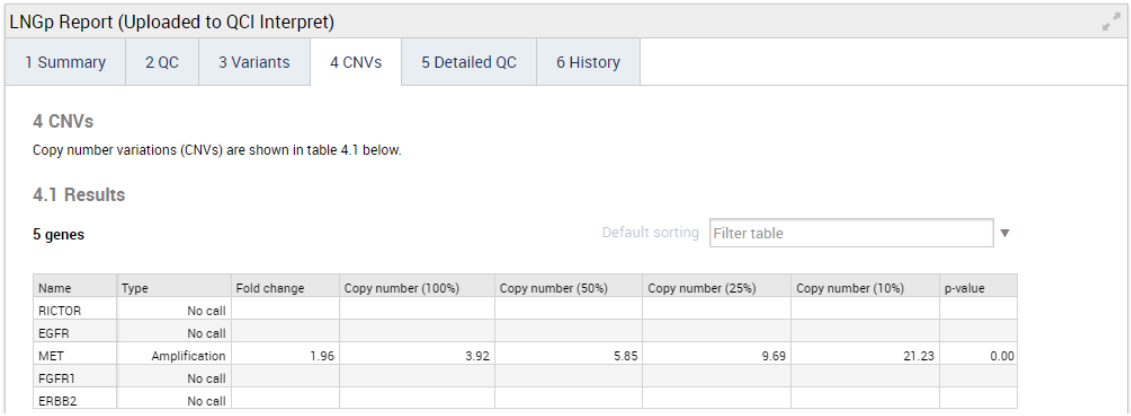


Figure 22: The CNV tab of the report.

p-value, which is to be seen as a measure of certainty of the call. The closer the value is to zero, the more certain is the call.

CNV result is indicated as either Amplification or Deletion depending on the fold-change value. If the fold-change and p-value thresholds are not met, this means that no copy number variation could be detected and it will be indicated in the table as No CNV detected.

The track viewer contains a CNV track indicating the region spanning the affected gene. Click a table row to display the corresponding gene in the Track Viewer.

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 B 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 Variants and CNVs tables of the report are linked with the track viewer. Select a variant in a table and 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 23). 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.

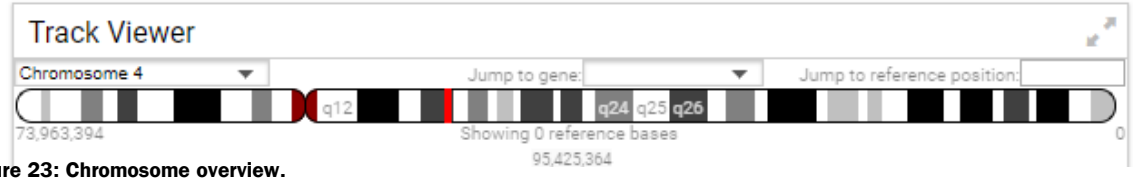


Figure 23: Chromosome overview.

  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 24). The genes are drawn as light blue annotation bars with directional arrow-ends, and overlapping genes are drawn on top of each other. This level can be navigated in a similar way as described for the "Chromosome Overview".

  Zooming in gives an **overview of a single gene** currently overlapping with the section of

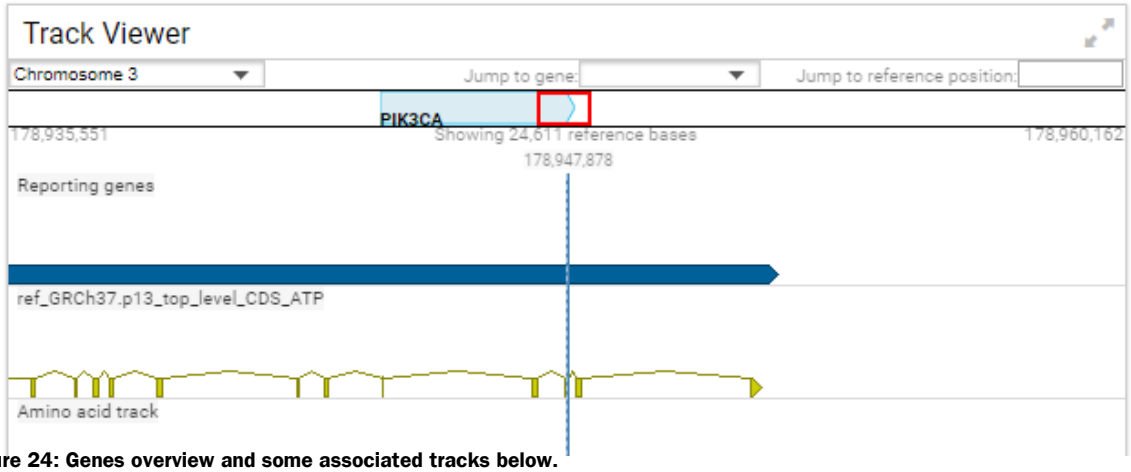


Figure 24: Genes overview and some associated tracks below.

the tracks (shown as a red mark on Figure 25). 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.

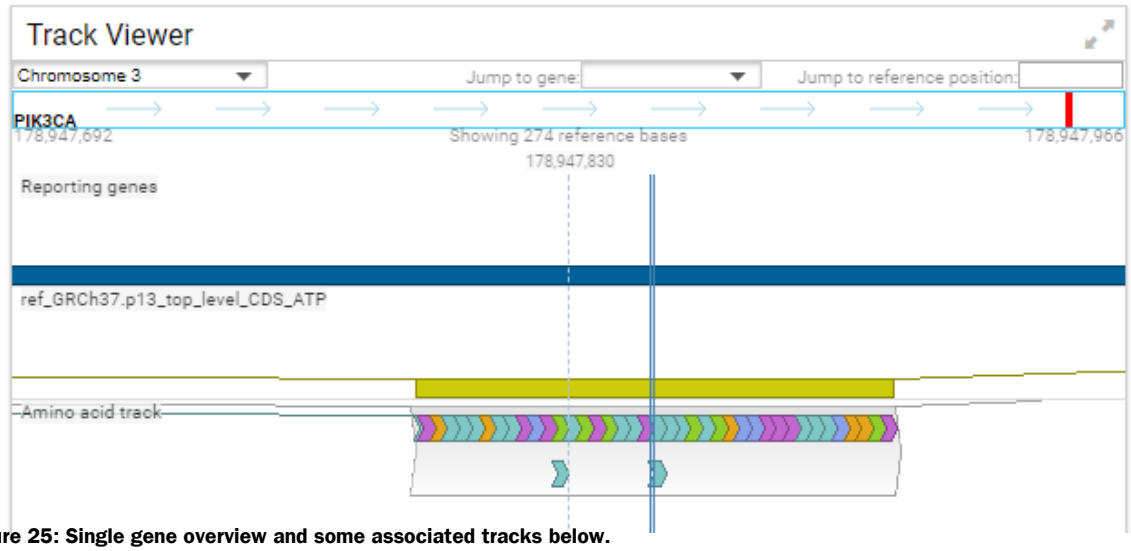


Figure 25: Single gene overview and some associated tracks below.

- 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 26). Variants may be of type SNV, MNV, replacement, insertion or deletion.

Figure 26: Detail of a selected variant in the Track Viewer.

- Mapped reads. This track displays the mapping of raw reads (Figure 27). The reads track contains all the reads mapped at a particular position. Forward reads are green, reverse reads are red. Un-aligned 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 in 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).

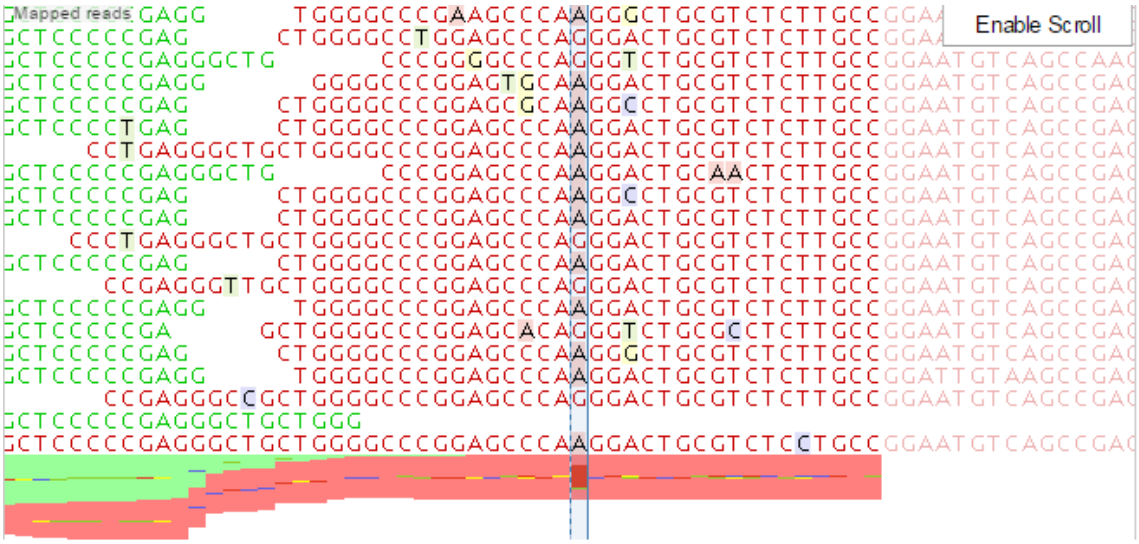


Figure 27: Mapped reads track.

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.
- CNV detection result track. This track is available for workflows with CNV detection. The track indicates the region spanning the affected gene, and can be navigated by clicking the individual rows in the CNV table.

Hovering on a variant in the track viewer will open a popup window with a selection (depending on the analysis workflow) of the following information (some of which is already available in the Variants tables):

- 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.

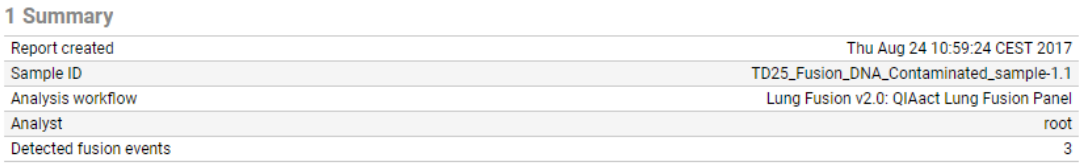
- Read coverage. Read coverage at this position. Each read of an overlapping pair contributes to the coverage.
- Read count. Number of reads supporting the allele.
- # unique start positions. Number of unique start positions for fragments that support the variant. This value can be important to look 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.

- 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 c. variant 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.
- Variant of Interest (VOI in the table). Variant of interest, as specified for the analysis workflow.
- 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 28).



1 Summary

Report created	Thu Aug 24 10:59:24 CEST 2017
Sample ID	TD25_Fusion_DNA_Contaminated_sample-1.1
Analysis workflow	Lung Fusion v2.0: QIAact Lung Fusion Panel
Analyst	root
Detected fusion events	3

Figure 28: Summary of a fusion report.

In particular, the summary highlights how many fusion events 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.

Fastq 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, how many fusion controls were detected, and the level of DNA contamination. In general, the number of reads mapped is low relative to that of the variant detection workflows. Unmapped reads can originate from unspliced primary transcripts of the target genes, or for a few of them off-target genes and rRNA.

Unique Molecular Indexing shows metrics for the UMIs generated by the secondary analysis workflow.

Fusions

This section presents the main result of the fusion detection by listing the predefined fusion events in three tables.

3.1 Fusion events - present lists fusion transcripts detected with a p-value below the specified threshold ($p < 0.005$). The number of reads that match to the fusion transcript and wild type transcript reference sequences are provided.

3.2 Fusion events - not present lists fusion transcripts that had sufficient read coverage to be tested, but that were found to have p-values above the specified threshold.

3.3 Fusion events - not testable Fusion transcripts that could not be tested due to insufficient number of reads mapping to the fusion or wild type transcript reference sequences. This can be the case if the involved genes are not expressed. At least five reads mapping to either of the reference sequences (fusion transcript or wild type transcript) are required for a fusion event to be testable.

All fusion events 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.
- 5' wild type. The 5' gene and position of the 5' breakpoint relative to the start of the gene.
- 3' wild type. The 3' gene and position of the 3' breakpoint relative to the start of the gene.
- 5' count. Number of unique molecular indices (UMIs) mapping to the 5' wild type gene, evidence for wild type expression.
- 3' count. Number of unique molecular indices (UMIs) mapping to the 3' wild type gene, evidence for wild type expression.
- Total wild type count. Total number of unique molecular indices (UMIs) mapping to the wild type genes. Evidence for wild type expression.
- Fusion count. Number of unique molecular indices (UMIs) mapping to the fusion transcript reference sequence. Evidence for fusion expression.
- Inconclusive count. Number of unique molecular indices (UMIs) mapping to either wild type or fusion transcript sequences, but for which evidence for wild type vs. fusion was inconclusive.
- Total count. Total number of unique molecular indices (UMIs) mapping to either wild type or fusion transcripts.
- p-value. A measure of certainty of the call. The closer the value is to 0, the more certain the call.

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 B 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:

- 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.
- Fusion transcript sequences. This track contains the fusion transcript reference sequences.
- 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. Inconclusive reads were removed. Forward reads are green, reverse reads are red. Un-aligned 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 in 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 in the Report tile

Depending on the report state (Ready for Review, Uploaded to QCI Interpret, 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 person. 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 three or four 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 other download options in a single ZIP file.

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, a tool generating comparison reports is available in the Comparison page of QCI Analyze (Figure 29). 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.

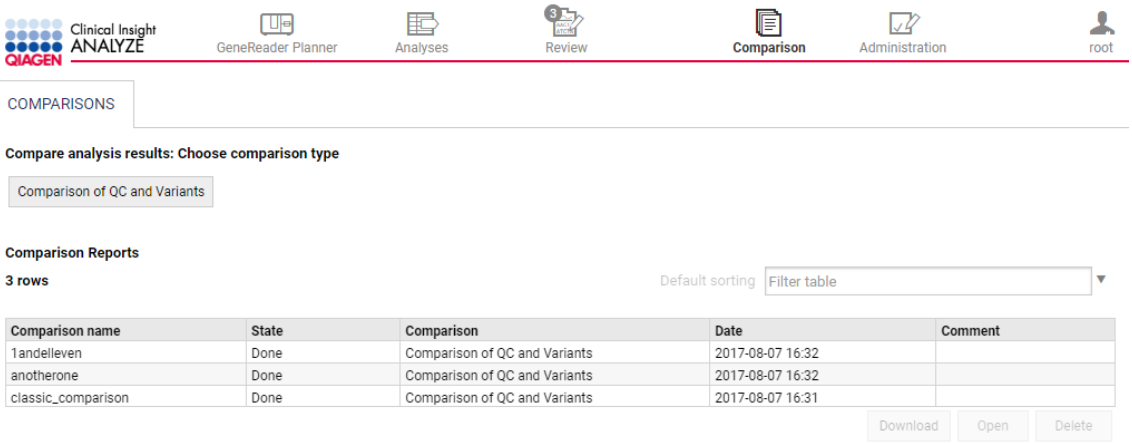


Figure 29: The Comparison page.

To navigate to the **Comparison** page, click on the corresponding icon in the Navigation bar. To start a comparison, click Comparison of QC and Variants (Figure 30).

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 as many samples as necessary for the comparison. How many samples are selected is available directly above the table.
4. Once all the samples to compare are selected, click **Start**.

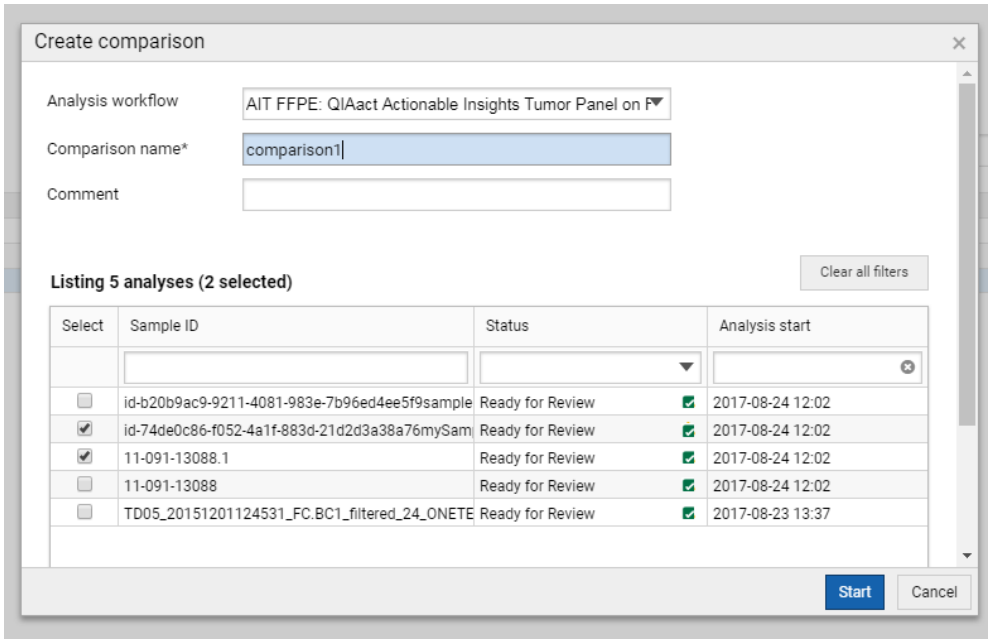


Figure 30: Create comparison dialog.

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 analysis is completed.

6. Double-click on the relevant row in the table to open the comparison report in a new tab, or use the **Open** button to the lower right of the table.

To delete a comparison from QCI Analyze, select the corresponding row in the table and click the **Delete** button to the lower right of the table.

Comparison report

A comparison report contains the following:

Overview: Overview information about the comparison, and QC statistics for sequencing and read mapping of each sample. It's a combination of the Fastq and Secondary analysis summary sections of the sample reports.

Gene coverage: For each gene, a table with sample-specific rows shows coverage at similar coverage thresholds.

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. Click on a row to open an additional table below the

variants overview table. This table displays frequency and quality values for the selected variant in each 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.

Administration

The administration web interface (Figure 31) is only accessible to users with administrative privileges, and is composed of eight tiles. The value to the left of the tile headings indicate the number of items in each tile.

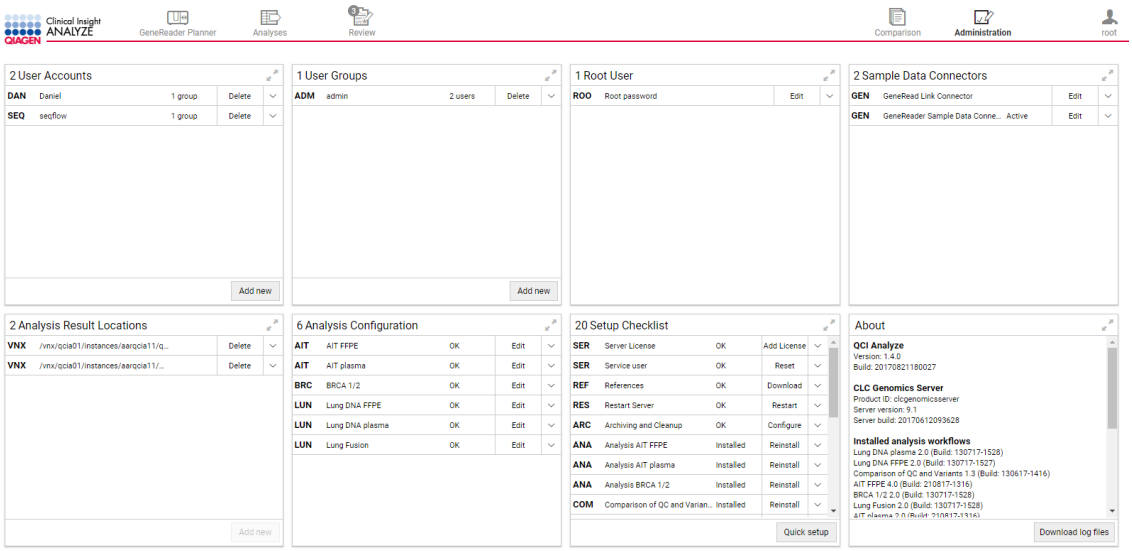


Figure 31: The Administration page.

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 17).

User Accounts

This tile is used to manage user accounts. Administrators can add and delete users. Click on a user name to access the button Edit and change the user's password. We do not recommend 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.

User Groups

Administrators can define groups of users, as well as a group of administrators, and add existing users to these. When creating a new group, choose a name for the group (no space, no special characters) and click on the group name to expand the box. Click on **Edit** to add users to a group. A group of administrators admin (ADM) is preinstalled. Add selected users from the drop-down menu to give them access to the Administration page.

Root User

This tile is meant to specify the root user and allows administrators to edit its password.

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 32), enter the paths to the relevant folders created during installation.

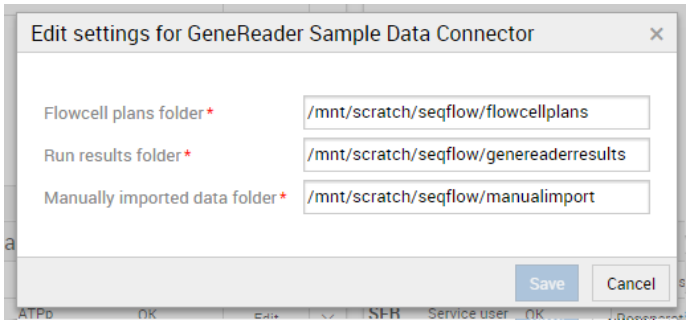


Figure 32: Editing a GeneReader Sample Data Connector.

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.

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.

Analysis Configuration

This tile lists the analysis workflows available. Two actions are available for the analysis workflows: **Edit** and **Disable/Enable**.

Click **Edit** to modify the parameters for the variant detection analysis workflows (Figure 33). Refer to Table 1 and Table 2 for a list of the default values for the different parameters.

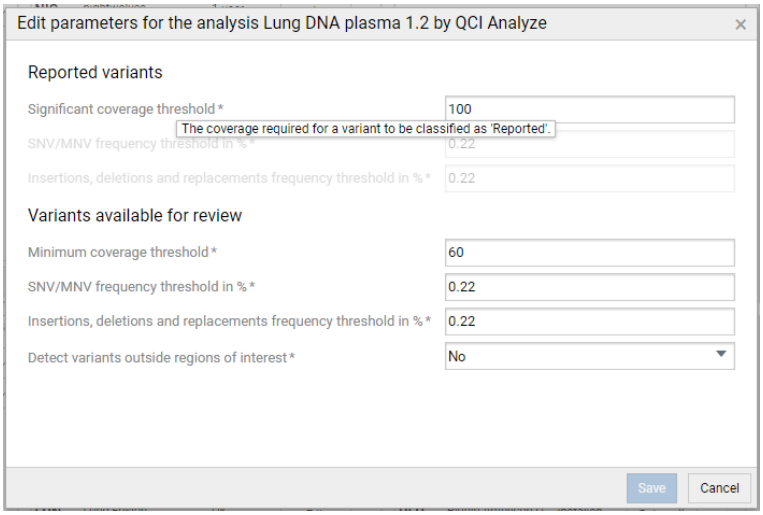


Figure 33: Editing parameters of a variant detection analysis workflow.

In the Reported variants section, it is possible to edit the thresholds that will define the variants automatically considered valid and reported in table 3.1.

- Significant coverage. Only variants with coverage above this threshold will be considered valid and reported in table 3.1.

The frequency thresholds for SNV/MNVs, and insertions, deletions and replacements are locked (see Table 2).

In the Variants available for review section, it is possible to edit the lower thresholds for detected variants. Variants, which make this cut-off but not that for "Reported variants" (see above), will be available in table 3.2 for manual review (see Table 1):

- Minimum coverage. Variants and regions with coverage below this threshold will be considered "untested", i.e., the coverage is too low for reliable testing. The threshold must be set lower or equal to the Significant coverage parameter.
- SNV/MNV frequency threshold. SNV/MNV variants detected with frequency above this threshold will be included in table 3.2. The threshold must be equal to, or lower than the corresponding threshold for Reported variants.

- Insertions, deletions and replacements frequency threshold. Insertions, deletions and replacements detected with frequency above this threshold will be included in table 3.2. The threshold must be set lower than the corresponding threshold for Reported variants.
- Detect variants outside of regions of interest. By design, Table 3.1 contains only variants detected within the ROI. Setting this parameter to "Yes", will include variants detected inside of the target region but outside of ROI in table 3.2 allowing users to review them. Setting the parameter to "No" means that variants outside of ROI will be included in neither variant table and hence will not be included in the report at all.

If an analysis workflow is disabled, it will not show up in the list of options when starting an analysis, creating a flow cell plan or a comparison report. It will also be absent from the drop-down menu filter options for analysis workflows.

Setup Checklist

This tile lists the different plugins and connectors (PLU), analysis workflows (ANA) and comparisons (COM) available. There are also a Server License and a Service User (SER) box, a Restart Server box (RES), a References box (REF), and an Archiving and Cleanup configuration box (ARC).

- SER Server License. Field for entering the QCI Analyze license order ID in order to download valid license.
- SER Service User. Defines seqflow as the first user account during installation to facilitate 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.
- REF. The analysis workflows use references that are specific to each panel. Click Download in the box REF each time a workflow is installed or reinstalled.
- RES. Use this button to restart CLC Genomics Server. This function is mostly used during installation and when updating the system.
- ARC. Specify mode of archiving and data cleanup configuration. Set as "Not OK" until automatic archiving is configured (see the Archiving section of this manual).
- PLU. Provide reinstall functions for different plugins and connectors.
- ANA and COM. Provide reinstall functions for the analysis workflows, as well as the comparison workflow "Comparison of QC and Variants".

Note: The button **Quick setup** allows the user to simply complete the setup procedure by installing all necessary items with just one click.

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.

Download bug report 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 34): 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.

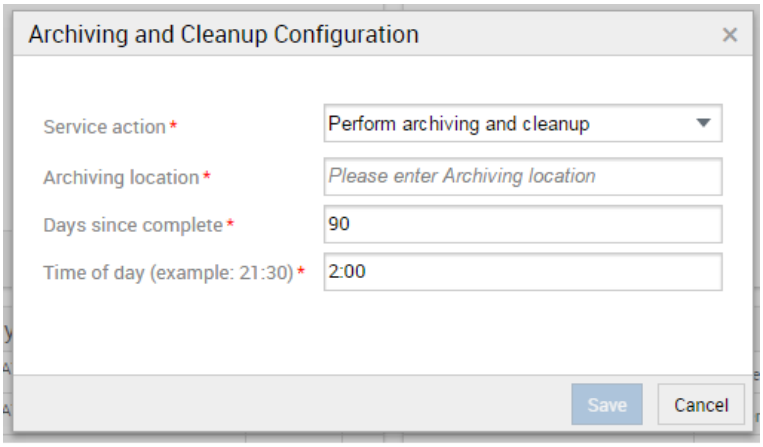


Figure 34: Editing the archiving and cleanup configuration.

The configuration of data cleanup and archiving is accessible from the ARC 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.

Appendix A - References and annotation information

Human reference sequence, ENSEMBL 682 MB

`ftp://ftp.ensembl.org/pub/current_fasta`

Chromosomes 1-22, X, Y and M human reference DNA sequence GRCh37(HG19).

Human genes, coding sequences and transcripts, ENSEMBL 25 MB

`ftp://ftp.ensembl.org/pub/current_gtf`

All annotated protein coding genes for human reference sequence GRCh37(HG19). The annotation was done by ENSEMBL and includes annotations from RefSeq, CDS as well as ENSEMBL itself. However, the transcripts and coding regions used for interpreting variants in the workflows are from RefSeq release 105 exclusively. In cases where there are multiple transcripts for a gene, the transcript included is the most frequently reported in the literature.

Target primers and target regions

Primers and regions are defined by the respective QIAGEN panels. Please refer to the QIAGEN website for additional information.

Appendix B - Detailed QC Report

The report is divided in per-sequence (QC 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.

QC 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.

QC 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.

- GC-content per base position calculates absolute coverages of C's + G's for each base position in the sequences. 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.
- Ambiguous base content per base position calculates absolute coverages of N's, for each base position in the sequences, where N refers to all ambiguous base-codes as specified by IUPAC.

Coverage 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.

Mean coverage of relative position in Regions of Interest

This graph shows the mean coverage for the first 10% of base positions in the ROI. When regions of interest are one base long, i.e., the first base position of the ROI represents 100% of that ROI, this graph will display flat lines. It is not included in the fusion detection report.

QC for UMI

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

- Unique Molecular Index sizes shows the distribution of UMI group sizes, with size being defined as the number of reads that share the same UMI.
- Unique Molecular Index sizes (size < 50) is a detailed portion of the above graph to highlight the distribution of the smaller groups.
- Average quality scores of all 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 by read position displays the quality scores of UMI reads by the base-pair position.

Appendix C - QCI Analyze software requirement

System requirements

Table 3: 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 and most recent versions 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.