ARCHER®

Manual

Archer® Analysis 4.1

User Manual

CS001-02 Revision 2

August 26, 2016

For use with versions 4.1 and above

View videos and additional resources for Archer products at:
http://archerdx.com/support

Table of Contents

1 Introduction and Summary

1.1 Introduction

This document describes the installation process and running procedure for the Archer Analysis Software. The Analysis software is available as a fully contained VM (Virtual Machine), which provides a web-based interface to the analysis software. Instructions for downloading and installing the VM are contained in a separate Installation Guide.

2 Analyzing Samples

This section describes the execution of an analysis on the Archer Analysis Software.

2.1 Run an analysis using the web interface

The Archer Analysis Software can be accessed through most modern Web Browsers, with the exception of Internet Explorer® version 7 (Windows® XP) which has been shown to have some problems. Google Chrome™ is the recommended web browser. Refer to the **Archer Analysis Virtual Machine Installation Guide** for instructions on how to discover the IP address of the Archer Analysis software. Start a web browser application and enter the IP address for the virtual machine in the address box.

The login screen for the Archer Analysis Software will appear as shown below:

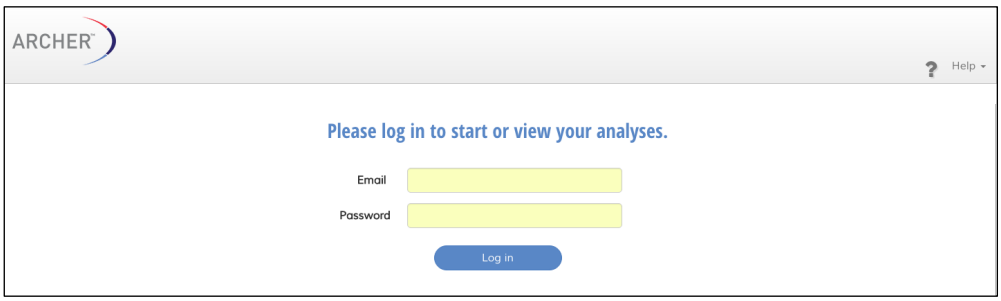


Figure 1. Archer login page

If the web page does not appear, check that the virtual machine is running (see the **Archer Analysis Virtual Machine Installation Guide** for instructions on how to install the virtual machine).

2.1.1 Create a login account

The Archer Analysis Software is a fully contained and secure environment that allows users to run their analyses under their own account. Create a login account by selecting the “Create Account” link on the bottom left corner of the screen. The account creation webpage will appear as shown below:

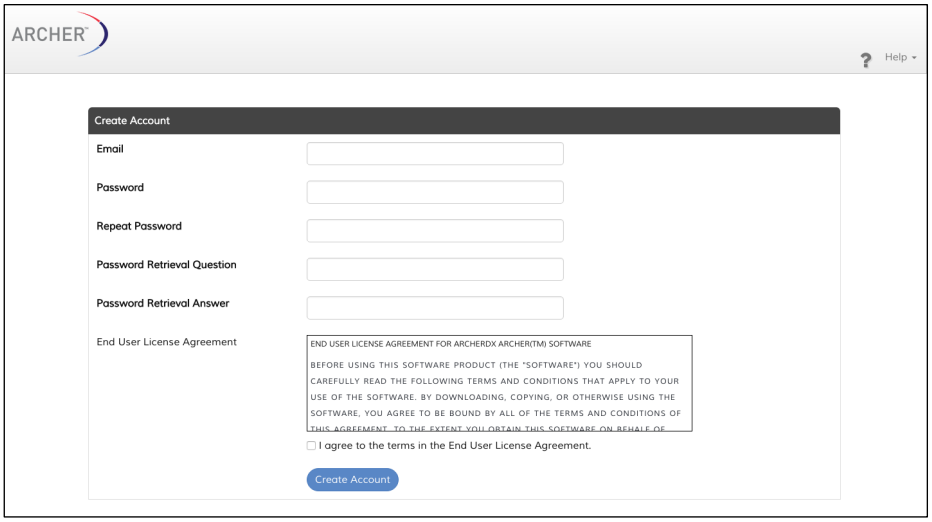


Figure 2. Archer Analysis account creation page

Use your email address as your login and create a new password. Alternatively, your site administrator can create a login for you with your e-mail address. See section 7.2.1 for more details.

At this time email cannot be sent from the VM. Enter a Password Retrieval Question and the appropriate answer. Later, if you lose your password, you will be asked the Password Retrieval Question to reset your password.

Read the End User License Agreement (EULA) and check the box to indicate you have read and agree to the EULA, then click the "**Create Account**" button to create your account. This will automatically log you in.

After successful login, the home screen will be displayed as shown below:

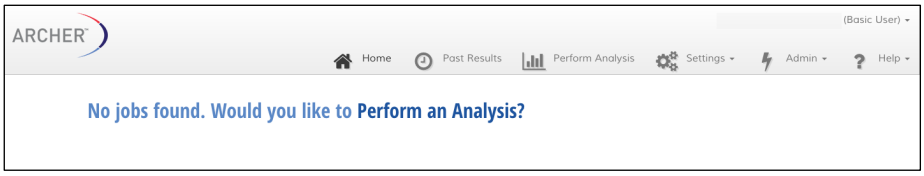


Figure 3. Archer home screen

2.1.2 Run an analysis

To start a new analysis, click the "**Perform Analysis**" button on the top right side of the screen. The Archer run analysis page will then appear:

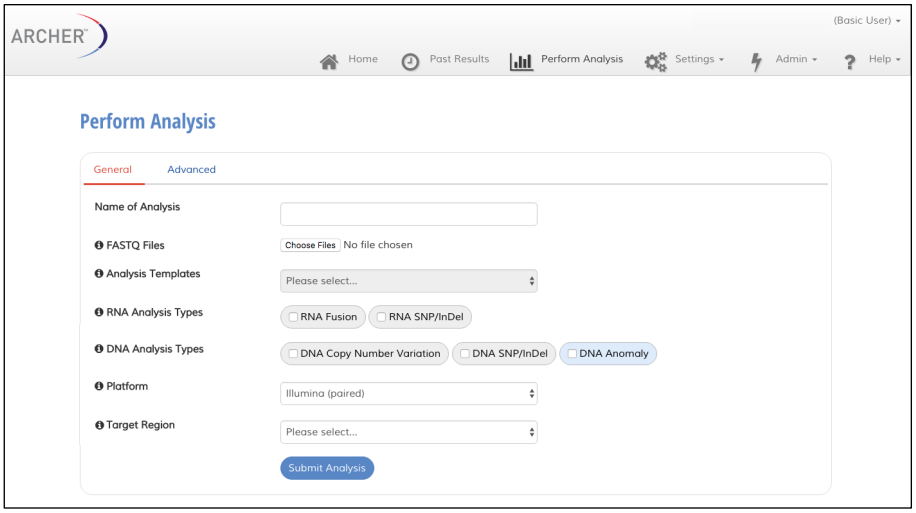


Figure 4. Archer run analysis page

2.1.2.1 Name the analysis

Give the analysis a name in the "**Name of Analysis**" box. The name is not required to be unique.

2.1.2.2 Select FASTQ files for analysis

Provide the demultiplexed FASTQ files for your sample(s) in the "**FASTQ Files**" field by selecting the "**Choose Files**" button. This will open a file selection dialog box to allow you to select the FASTQ files as shown below. **NOTE:** All files must be from a single folder.

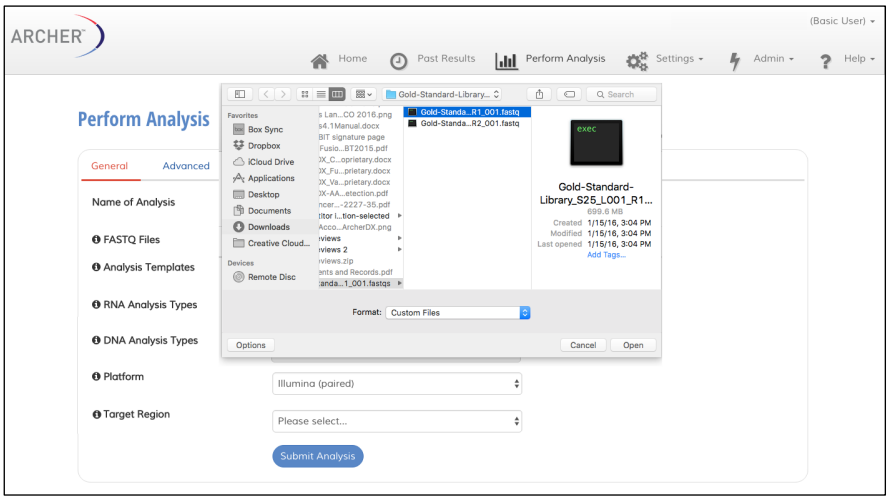


Figure 5. Archer select FASTQ file dialog box

Select one or more files and click "**Open**".

The number of files and the total size in MB or GB of the combined set of files is listed below the section and this can be used to predict the time it will take to upload the files to the server.

NOTE: The FASTQ files can be either uncompressed (and have the extension "**.fastq**" or "**.fq**") *or* compressed (using the GZIP algorithm, and have the extension "**.gz**"). The ZIP compression file format is currently not supported.

2.1.2.3 Choose an Analysis Template

If you have a non-default set of analysis settings that you would like to routinely run, you may save them as an Analysis Template by selecting "**New**" from the "**Analysis Templates**" pull-down list, and providing a name for it in the dialog box as shown below.

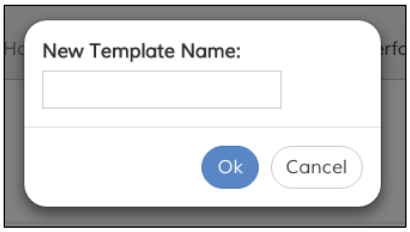


Figure 6. New Template Name dialog box for naming a new Analysis Template

Subsequently you can load it when setting up future analyses by selecting it from the "**Analysis Templates**" pull-down list.

2.1.2.4 Select the Assay Type

There are a number of different Archer assays that can be analyzed, based on the origin of the nucleic acids; i.e RNA vs DNA assay. For each assay, multiple analyses can be run at the same time but RNA analyses and DNA analyses cannot be run simultaneously.

RNA Analyses

RNA Fusion detection will detect gene fusion events by annotating the de-novo assembled mRNA reads with BLAST. Archer specific filtering logic is used to reduce misalignments and false positives. SNP and Indel analysis can also be selected to include the detection of small variations with respect to the human hg19 reference.

Select the required analyses by clicking the checkbox.

DNA Analyses

For DNA samples, small variation, DNA anomaly, and CNV (Copy Number Variation), analyses are available. You can also now monitor DNA target coverage.

Select the required analyses by clicking the checkbox.

The DNA small variation workflow detects SNPs (single nucleotide polymorphisms) and Indels (insertions and deletions) in your data. This can be utilized in addition to other workflows. The software will produce a variant summary with customizable filters that can be used to filter variants of interest. These filters can be adjusted in the interactive table and saved for future analyses. The user may select which attributes of the variants to show using basic logic gates provided in the software. When satisfied with the results, the data may be exported into a PDF report or downloaded as a TSV (tab separated values) or PDF file.

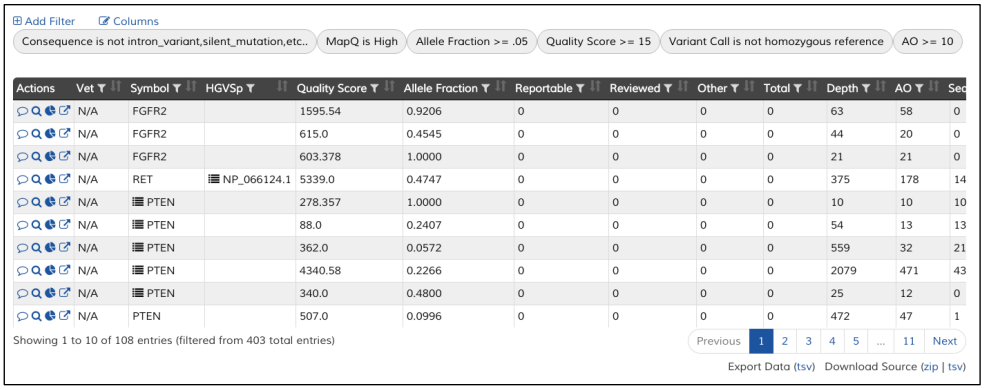


Figure 7. Archer Analysis Variant Summary

The DNA anomaly analysis is a new workflow starting in Archer Analysis 4.0, and features a streamlined and powerful new variant grid (Figure 7 above). This is similar to our RNA Fusion pipeline in that DNA reads are de-novo assembled and annotated. This allows us to detect when exons, introns or entire chromosomal regions are skipped or rearranged. For example, FLT3 internal tandem duplications (ITDs) and entire losses of genes can be detected.

CNV detection can be selected in addition to other assay types if the user has a kit that is designed for CNV detection. This assay uses one or more control samples, if present, to compare expression levels in normal versus diseased samples. If no controls are present, all samples are normalized and used as a baseline. Selection of the proper normal sample(s) is described in 2.1.3. The analysis software uses this information to produce both a readout and a visualization of expression levels, along with a p value to convey confidence in the calls.

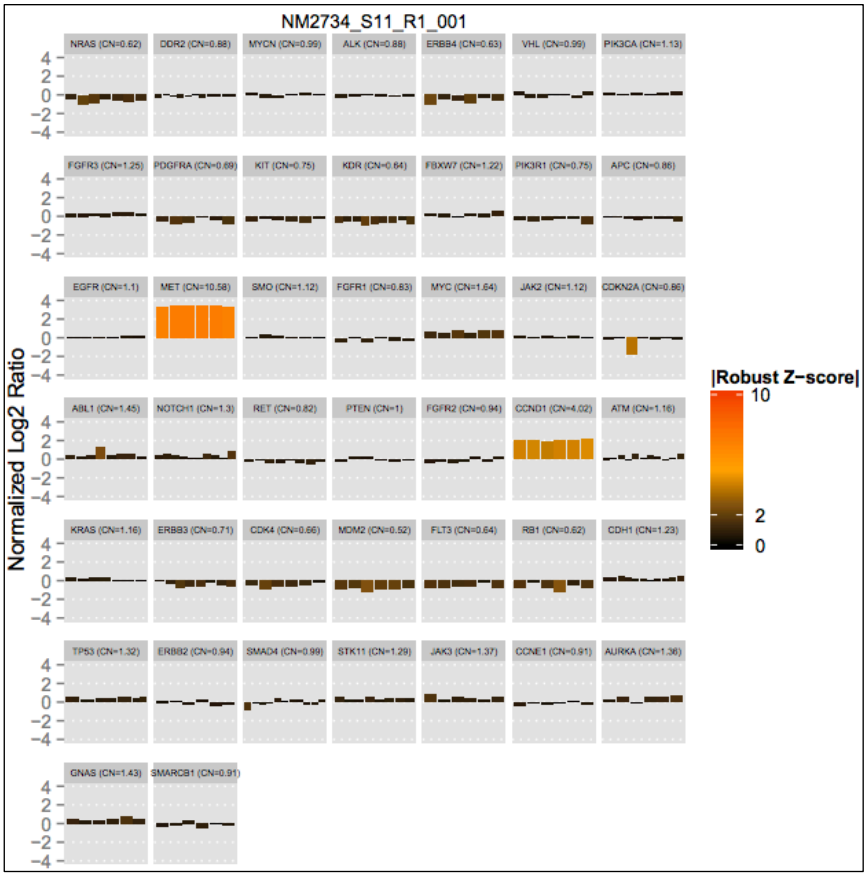


Figure 8. Archer Analysis CNV report

2.1.2.5 Platform

The Archer Analysis Software supports the Illumina MiSeq®/NextSeq®/HiSeq® platform as well as the ThermoFisher Scientific Ion PGM™/Proton™ platforms.

Select the appropriate platform as well as the library type (single read vs paired-end reads for the Illumina MiSeq/NextSeq/HiSeq platform or "demultiplexed" for the Ion Torrent platform) from the dropdown menu.

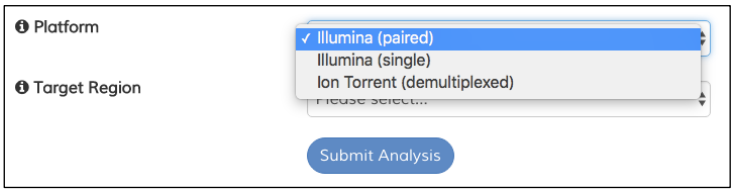


Figure 9. Platform selection options

2.1.2.6 Target region

Select the appropriate target region corresponding to the Archer kit/panel.

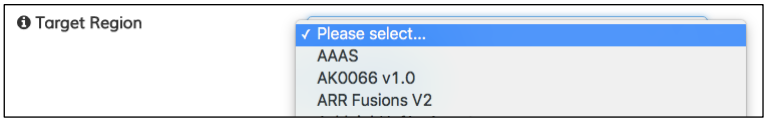


Figure 10. Target region selection options

To add a different target region for a custom Archer kit, section 3.2.17, which also includes further information about adding custom content.

2.1.2.7 Targeted Mutation (Optional)

When selecting the RNA or DNA SNPs/InDels assay type, it is possible to instruct the analysis software to *only* report on the mutations in a Targeted Mutation file (in VCF format).

When a targeted mutation file is selected, and "**Include Non-Targeted Variants**" is set to "**Off**", the Variant Summary page will show the results for the targeted mutations ***only*** and not report on any other variants/mutations that may be present in the sample.

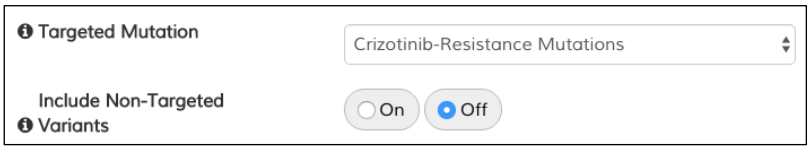


Figure 11. Targeted Mutation file options

When a targeted mutation file is selected, and "**Include Non-Targeted Variants**" is set to "**On**", the Variant Summary page will show the results for the targeted mutations ***in addition to*** any other variants/mutations that may be present in the sample.

NOTE: Only variants within 1kb downstream of any GSP2 will be detected.

Select the desired targeted mutation file from the dropdown menu or leave at "**(optional)**" if no targeted mutation file is desired.

See section 3.2.18 for instructions on how to add/manage the set of targeted mutations files available in the VM.

To start the analysis, select the "**Submit Analysis**" button.

2.1.3 Selecting the CONTROL samples for CNV runs

If the DNA Copy Number Variation assay type was selected, a dialog box will appear that will allow the selection of the CONTROL or NORMAL sample. CNV analyses will be far more sensitive if a matched CONTROL sample is available for each CASE (tumor) sample although it is not required to be a matched control sample. It is even possible to have no control sample whatsoever.

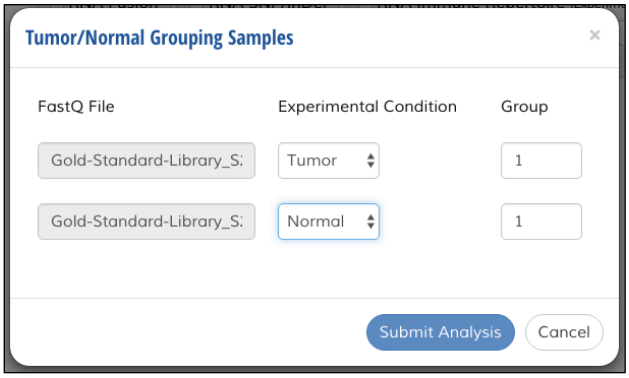


Figure 12. Example of a copy number job with a single normal/control sample

The dialog box lists each of the samples and the "Experimental Condition". The Experimental Condition is used to indicate if the sample is a "Normal" sample, or if it is a "Tumor" sample. If the sample name has the word "Normal" (case insensitive) in the file name, the experimental condition is automatically set to "Normal".

Use the dropdown menu to change the experimental condition.

If only a single control sample is available, select that sample as the control sample by selecting "**Normal**" from the "**Experimental Condition**" dropdown menu and ensure the Group identifier is 1 for the control and all case (tumor) samples.

2.1.3.1 Using multiple control samples as a group to increase sensitivity

If no matched case-control samples are available, the use of multiple control samples could increase the sensitivity. The group of control samples will inform the algorithm about the natural variability of the coverage and GSP behavior. This could result in more accurate CNV calling.

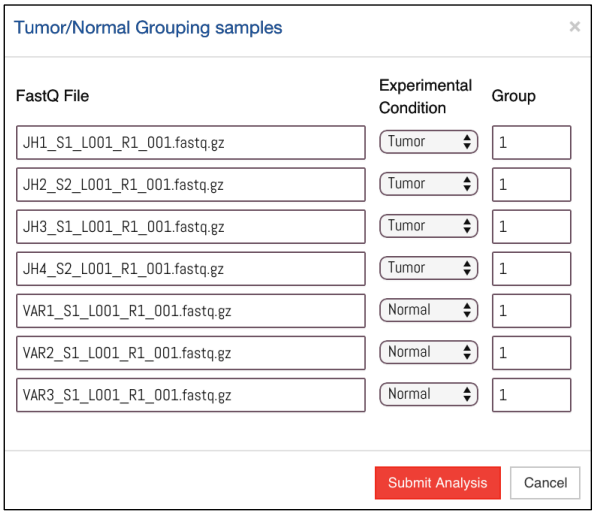


Figure 13. Example of a CNV run with a group of normals (VAR* samples) to increase sensitivity

In the above example the samples starting the JH* are the case samples for which no matched control was available. The samples starting with VAR* are used as the set of control samples. Since all samples are in the same group (1) the control

samples are combined in the algorithm and the four case samples are compared with the group of control samples.

2.1.3.2 Using matched case-control samples

In the situation where there are matched case and control samples from the same individual, the control sample must be matched to the case sample by using the GROUP identification. The case and control sample should have the same group identifier to ensure the correct control is matched to the correct case sample.

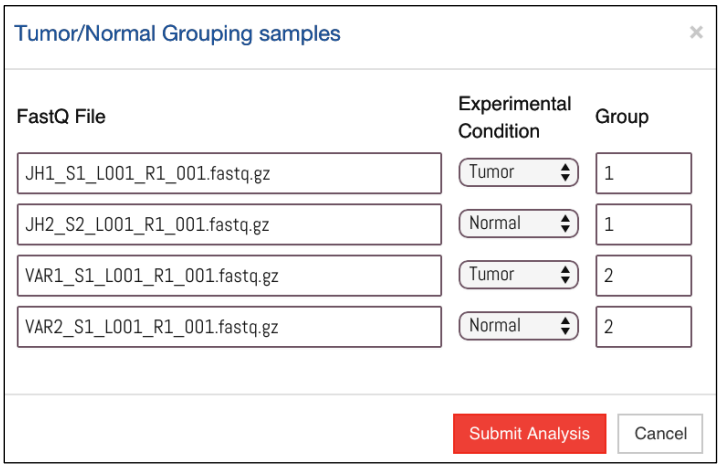


Figure 14. The group identifier matches case and control samples from the same individual

In the example shown above, sample JH1 and JH2 are the case and the control sample for individual 1 and are assigned to group 1 and sample VAR1 and VAR2 are the case and control samples for individual 2 and are assigned to group 2.

2.1.4 Advanced Options

Clicking on the "**Advanced**" option, will bring you to the Advanced Options page as shown below.

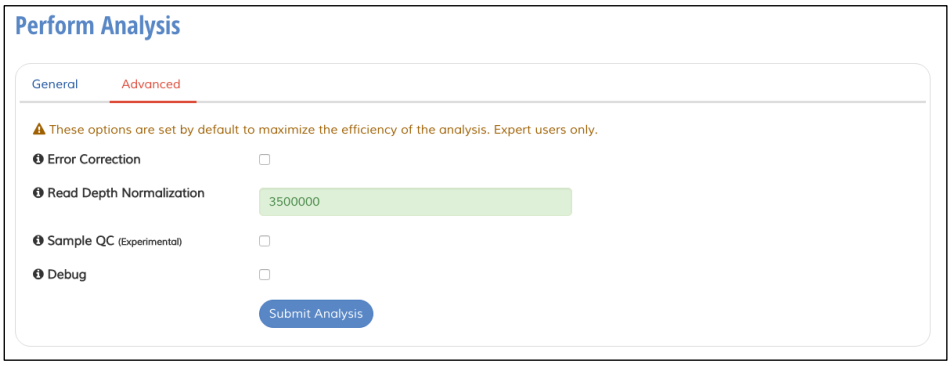


Figure 15. Advanced Options

2.1.4.1 Error Correction

Error correction is optionally employed during read de-duplication to remove likely sequencing errors based on discrepancies between read originating from the same

molecular bin. Turn it on by checking the "**Error Correction**" checkbox. This will lead to a significant increase in processing time.

2.1.4.2 Read Depth Normalization

More reads per sample analysis requires longer processing time and more RAM. In general, increasing the number of reads above 3.5 million will not improve the quality of results using the currently available ArcherDX FusionPlex™ and VariantPlex™ panels. Therefore, the default number of reads per sample is set to the number of reads provided in the sample's FASTQ file or 3.5 million, whichever is lower. Change this maximum by providing a different number in the "**Read Depth Normalization**" field.

NOTE: Setting to 0 will disable normalization, and setting to more reads than are in teh sample will have no effect on the sample.

2.1.5 Progress Bar

A progress bar, as shown below, will indicate the upload progress of FASTQ files to the analysis server. The speed of the upload depends on the network speed between the client and the host. If the client and the host are located on the same machine, the upload will progress very quickly, even for many GB of data. If the client machine is accessing a host that is on a network or on the Internet, the upload speed is determined by the network speed. The network upload speed can be verified with services such as http://speedtest.net.

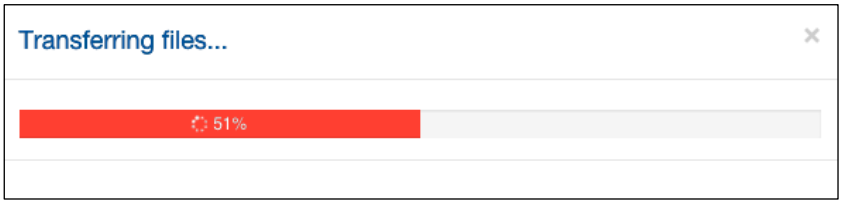


Figure 16. Archer Analysis progress bar shows the progress of the *upload* of the FASTQ file to the server

Once upload is complete, the screen will indicate that the job has been submitted for analysis. The job status will show the analysis progress of this job, which can be "NEW", "QUEUED", "RUNNING", "HALTED", "COMPLETED_ERROR" or "COMPLETED_OKAY".

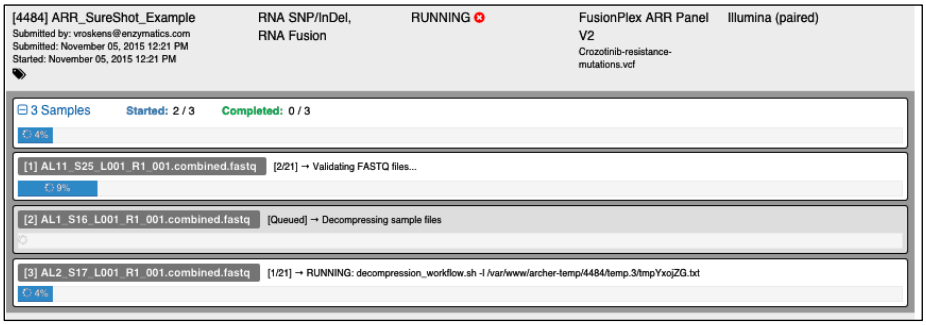


Figure 17. Archer home screen showing a running job

At this point the user can logout of the system. The analysis will continue in the background.

NOTE: DO NOT turn off the Virtual Machine (Appliance) or shutdown the host computer; this will disrupt the analysis and most likely result in failure.

2.1.5.1 Duration of an analysis run

The analysis of a single set of FASTQ files (~250,000 reads, or 50-100 MB of data) should only take a few minutes. Larger runs of multiple samples can take a few hours, depending on the number of samples. Any job that is running longer than two days will automatically be killed to allow other jobs to run.

2.1.6 Progress bar and status updates

A progress bar will indicate the job execution progress. In the example shown below, "Completed: 4/11" indicates that 4 samples out of 11 total samples are finished processing. If you expand the job by clicking on the "**+/-**" symbol next to the job name, each sample is shown with its own progress bar. Below the progress bar, the steps of the analysis are indicated as well as the position in the total progress of the current step. In the example shown below, "[12/16]" indicates that step 12 of a total of 16 steps is being executed for sample [2].

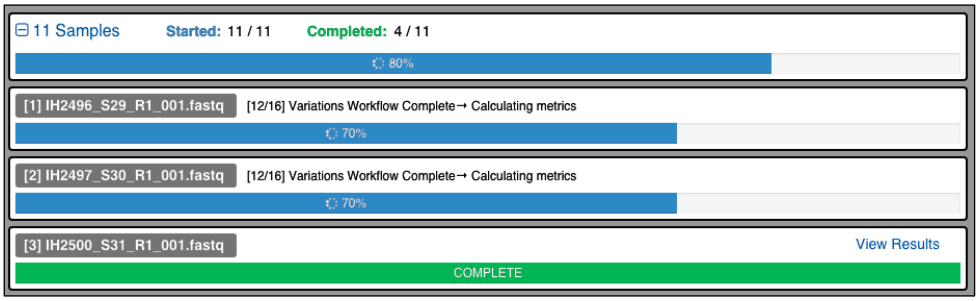


Figure 18. Progress bar indicates the progression of the analysis job

The default setting of the virtual machine is to process only one sample at a time due to the memory requirements for certain steps. The virtual machine can be configured to allow more than one job to run simultaneously if the VM is running on a host with more than 24 GB of memory. See the **Archer Analysis Virtual Machine Installation Guide** for more information on how to set up the VM for multiple sample processing.

2.1.7 Analyzing data from multiple runs

The Archer Analysis Software can analyze data from multiple runs, either in a single job or split over multiple jobs. When a job has been started, another job can be initiated immediately without having to wait for the previous job to finish. In the default situation only one sample (and therefore job) will be running at once. Any new jobs that are started will be placed in a queue, as will the samples contained in that job.

When there are multiple jobs queued up, the location of the job in the queue will be indicated with a number as shown below. In this example, the job with the number "2" is the next in line to be executed when an available slot for the sample(s) within it becomes available.

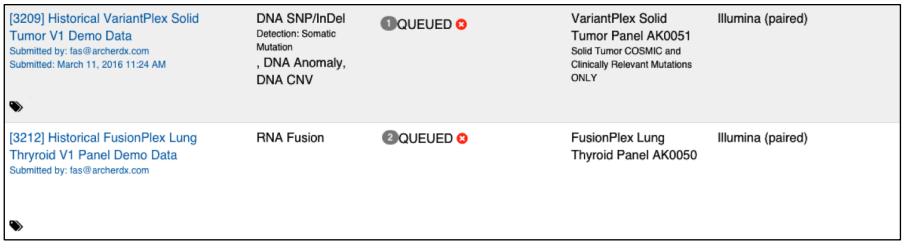


Figure 19. Screen shows running and queued jobs

2.1.8 Stopping a running or queued job

An analysis job that is currently running can be stopped by clicking the " " icon on the home page as shown above.

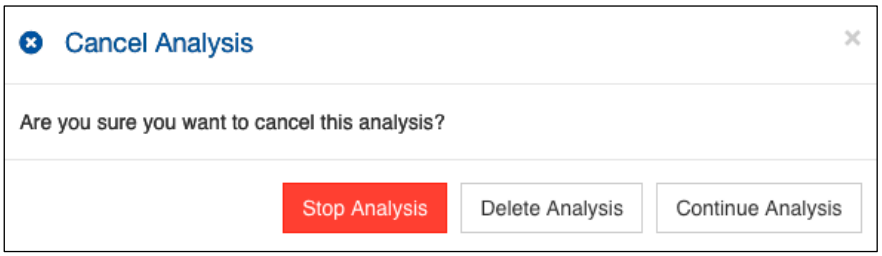


Figure 20. Cancel analysis screen from the home page

There are two options for stopping a running job:

Stop Analysis – this will stop the job but leave the FASTQ file(s) on the server.

Delete Analysis – this will stop the job and remove all the data and results from the server.

To remove a job that is in the queue but not yet running, click the " " icon. The queue position of the other jobs will be adjusted accordingly.

3 Understanding Results

3.1 Home screen

After an analysis has completed, its status is indicated on the Home screen:

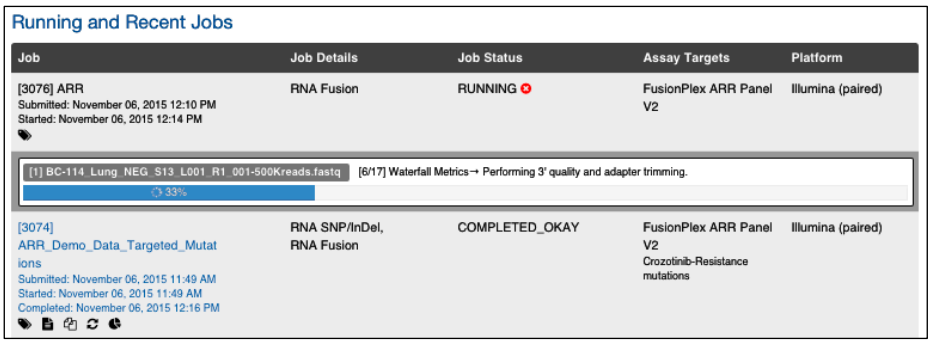


Figure 21. Archer home screen showing a completed job (3074) and a still running job (3076)

The "**Past Results**" page also shows a list of completed jobs, including any jobs run by other users that the current user has permission to view (see Chapter 7).

Specific jobs can be located on the "**Past Results**" page by typing a query into the search box (Figure 22).

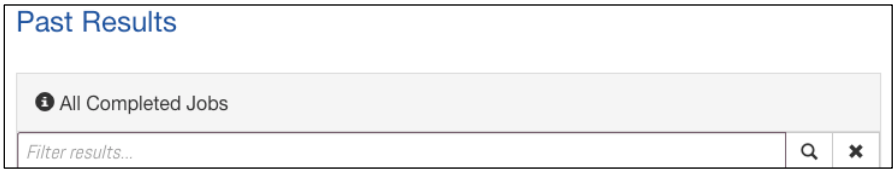


Figure 22. The Past Results search box

The job list (Figure 21) includes descriptive information about each job.

Job: Job ID, job name, and timestamps for when each job was submitted, started, and completed. This column also contains several clickable icons whose functions are described in the following sections.

Job Details: The type of analyses that were chosen when setting up the job

Job Status: One of "NEW", "QUEUED", "RUNNING", "HALTED", "COMPLETED_ERROR" and "COMPLETED_OKAY".

Assay Targets: The panel that was chosen when setting up the job. If a targeted mutations file was chosen, it is also shown here.

Platform: The sequencing technology that was chosen when setting up the analysis.

3.1.1 Reviewing the analysis settings

Click on the Analysis Settings icon ( ) to see what parameter values (user settings) were used for the processing of the job. The default values of each parameter are also shown. If any settings differ from the default, they will be highlighted (Figure 23).

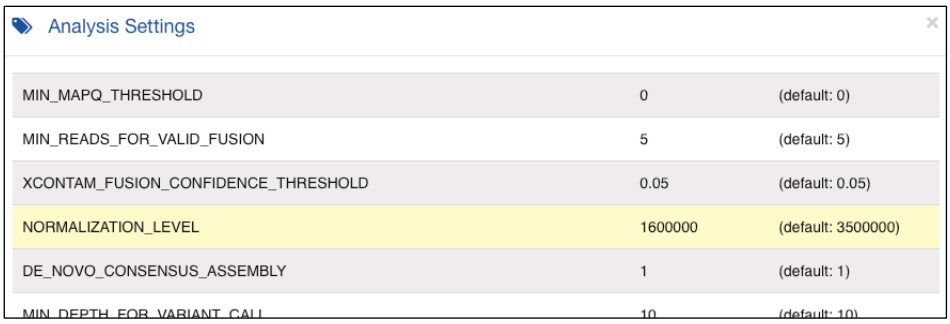


Figure 23. Analysis Settings, illustrating one parameter (NORMALIZATION_LEVEL) that differs from the default

3.1.2 Reviewing the (error) logs

After a run has completed the status of the job should be "COMPLETED_OKAY". If this is not the case, review the error logs to determine the cause of the failure.

To review the error log, select the file icon () below the job details for the run. A tabbed page will show the Workflow Logs, the Torque STDERR (Standard Error) log and the Torque STDOUT (Standard Out) log. Review the Torque STDOUT log file first, since this could also contain error messages.

In addition, click the "**View Raw File**" link to download the main log file to the local computer.

Contact tech@archerdx.comfor further assistance and troubleshooting.

3.1.3 Rerun or clone a previous run

If the results of a run are not satisfactory, it may be useful to re-run an analysis with different analysis settings. This can be accomplished by either making a clone of the original analysis or by re-running the original analysis. The difference between the two options is that a clone will be a copy of the original run, leaving the old run intact, while re-running an analysis will remove the old results and replace it with the new one (i.e., all existing data except the fastqs will be deleted for the job).

To clone an analysis, select the “Clone Job” icon ().

To re-run an analysis, select the “Rerun Analysis” icon ().

3.1.4 View overview of results for the entire job

To generate a PDF report showing the Fusion QC results and a summary of the analysis results for each sample, select the "Sample Overview" icon ().

3.2 Job detail page

From either the Home Screen or the Past Results screen, click on the analysis ID to see the job detail page for this analysis. The page will show a list of samples with various clickable icons below them as well as the “Assay Result” and “QC Result” for each sample. The specific Assay Result categories that are shown will depend on how the analysis was set up in section 2.1.2.

All files, including summaries, etc., that were generated for the job can be downloaded in batch by clicking on the "**Download All Files**" icon ().

Figure 24. Archer job detail page showing gene fusion results for two samples

3.2.1 Sample Name column

3.2.1.1 Detailed Summary

Below each sample name is a link that will take you to the "Detailed Summary" for that sample (section 3.2.5). Clicking on "**Detailed Summary"** will show the summary results pages with more details. Clicking on a specific fusion of interest will take you directly to the evidence for that fusion on the Detailed Summary page.

3.2.1.2 Processing Log

Clicking on the Processing Log icon ( ) will allow you to load the processing log into the browser window, which you can then save as a text file if desired. Generally this is only informative in the case of an error in processing, in which case it is recommended to contact tech@archerdx.com for assistance with troubleshooting.

3.2.1.3 VEP Variant Overview

For analyses that include SNP/Indel detection, a link will be available ( ) that redirects to the Ensembl Variant Effect Predictor results for that sample. For further information on variant results refer to section 3.2.12.

3.2.2 Assay Result column

3.2.2.1 Fusion calls

In Figure 24 above, there are two samples. A triple-positive sample where three strong evidence gene fusions (SLC34A2-ROS1, CCDC6-RET and EML4-ALK) were detected, while the other (SureShot fusion negative control) did not show any strong evidence fusions.

3.2.2.2 *Targeted Mutations Results*

If a targeted mutations file was selected for a run, the job detail page would list the targeted mutations that were detected in the “Variants Found” section as shown below:

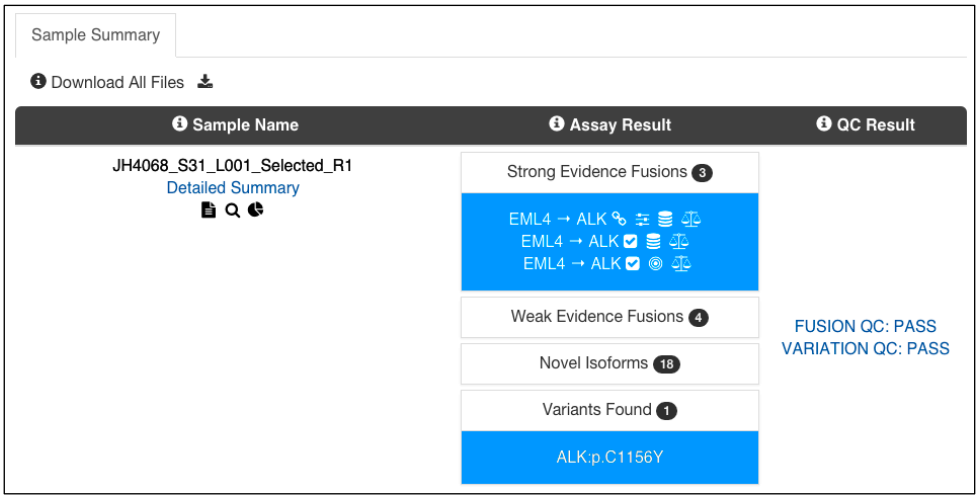


Figure 25. The result of a targeted mutation shows 1 mutation under the “Variants Found” section for this sample in the ALK gene.

In the "**Detailed Summary**" section “Variant Summary”, each of the targeted mutations will be listed – not just those that were found to be actually present in the sample – with their status listed as “Present”, “Not Present”, or “No Call”.

3.2.2.3 *CNV Analysis Results*

For DNA Copy Number Variation analyses, the summary page will list the results of the CNV analysis by highlighting the genes that show significant copy gains or losses.

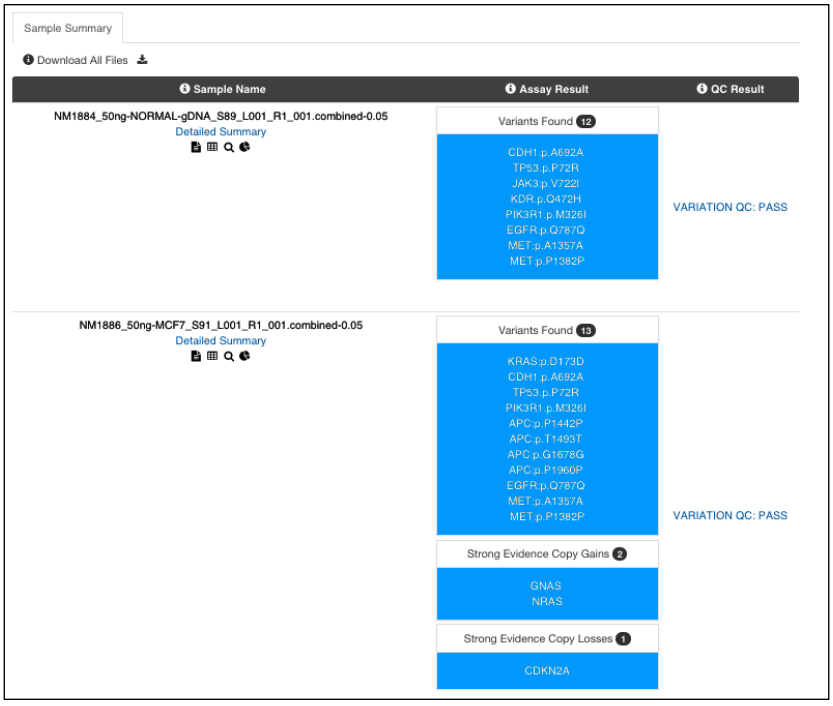


Figure 26. Sample Summary page for a CNV analysis showing samples with significant gains and losses

Only the samples that were marked as *case* (tumor) samples will show CNV results. Any sample that does not show any data in the "**Assay Results**" column were marked as *control* (normal) samples.

3.2.3 QC Result column

This column conveys pass/fail information on various QC filters that may used to screen sample data, in order to have statistical confidence in the analysis results. Which filters get applied depends on the type of sample (RNA vs DNA) and the type of pipeline run:

Input QC: Pass/Fail for Input data. Fastq files will fail if they are empty or all reads are removed during data preprocessing. This filter is applied to all samples

Fusion QC: Pass/Fail for the Fusion pipeline. Marked as failure if average unique reads per GSP2 is too low or average unique start sites per GSP2 controls is too low. This filter is applied on all RNA samples, and is configurable.

Anomaly QC: Pass/Fail for the Anomaly pipeline. Marked as failure if average unique reads per GSP2 is too low or average unique start sites per GSP2 controls is too low. This filter is applies to DNA samples, if DNA Anomaly was selected during setup of the analysis. It is also configurable.

Variations QC: Pass/Fail for the SNV/Indel pipeline. Marked as failure if there are too few unique reads per GSP2, if there are too few unique RNA start sites per GSP2 control (RNA only), or if there are too few unique DNA/AMBIG start sites per GSP2 (DNA only).

3.2.4 Visualize Sample Data in JBrowse

From the Sample Summary page, the "**Visualize Sample Data**" link ( )will open a window with the JBrowse genome browser showing all the reads aligned to the human genome that formed the basis for the Gene fusion and/or Mutation analysis.

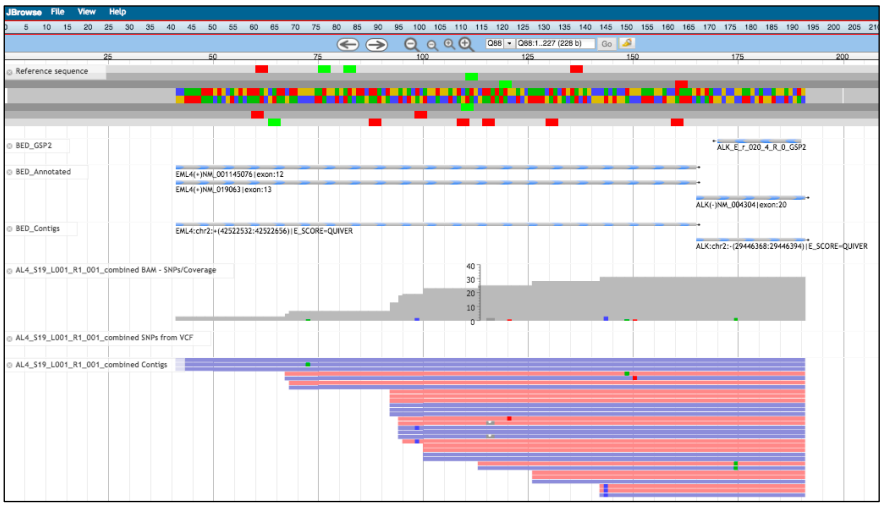


Figure 27. Visualizing the unique reads mapping the hg19 genome

The genome browser shows a number of different tracks, a description of which can be found below

Reference Sequence

A track showing the hg19 reference sequence with both strands and the 6-frame translation as well

refGene

A track showing the genes and each of their transcripts from the RefSeq database

Right-click on the transcript to see its name and follow the link for a search of the transcript at the NCBI website

Target Region GSP2

A track showing the position and direction of the target gene specific primers

cosmic

A track showing the location and identifiers from the COSMIC mutation database

Right-click on the feature to follow the link of the identifier at the Sanger website

Read1(2)_contigs_coverage

The coverage plot of the read coverage and mismatch ratio for reads 1 and 2 separately

Read1(2) Contigs

The mapped reads for reads 1 and 2 separately

The search box at the top can be used to navigate to a location directly by typing in the location OR can be used to enter a gene name.

For more information on usage of JBrowse, see the manual for JBrowse help links in the JBrowse window

3.2.5 Detailed Summary pages

The Summary Results pages display the detailed results of the analysis and QC metrics. Depending on the selected Assay Types on the "**Perform Analysis**" page one or more of the following tabs are shown:

Strong Fusion Candidates

Shows the information about the fusion candidates, novel isoforms and the non-fusion (wild-type) spliced genes with strong evidence supporting the call.

Weak Fusion Candidates

Shows the information about the fusion candidates, novel isoforms and the non-fusion (wild-type) spliced genes with weak evidence supporting the call. The tab with the list of weak evidence calls is collapsed/closed by default. Click on the header text "Weak Evidence Fusions" to open the section to reveal the weak evidence fusions.

Read Statistics

Shows the high-level metrics of the assay such as total reads, on-target read percentage etc.

Assay Targets

Shows detailed metrics of the assay such as the coverage for each target region.

CNV Summary

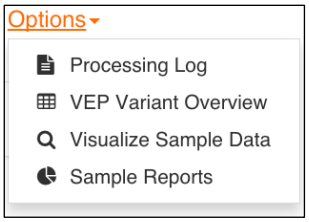
Shows detailed results for the copy number assay.

Variant Summary

Shows a summary of the SNP's and InDels found for the sample.

Options

Menu for more options such as viewing the processing log, the VEP results, visualizing the reads mapped to hg19 and the Sample PDF reports.



3.2.6 Strong Evidence and Weak Evidence Criteria

Gene fusions, oncogenic isoforms, wild type isoforms, and novel isoforms are separated into two categories:

1. Those with strong support for the call
2. Those with either weak support for the call, or with characteristics indicative of a false positive call

The software provides a Strong Evidence tab and a Weak Evidence tab, so that all candidates can be viewed, regardless of whether they are called strong or not. The Strong Evidence calls should be considered the calls that are being made by the software. Weak Evidence calls are primarily used for troubleshooting false negatives in the Strong Evidence category and are not to be considered positive calls. The information presented in the Strong and Weak evidence tabs shares the same format. There are a number of different criteria for the categorization of weak vs. strong, as described in subsequent sections.

3.2.6.1 Positive Evidence

In order to be called a strong fusion, the following criteria must be satisfied.

1. Minimum number of reads - In order for any candidate to be considered at all, there must be at least 5 breakpoint spanning reads that support the candidate. If there are not at least 5, it will not be further evaluated, and it will not be classified as either strong or weak (i.e. it will be absent from the results entirely). This cutoff can be adjusted by changing the MIN_READS_FOR_STRONG_FUSION parameter under "General Analysis Settings."

2. Presence of fusion in Quiver - If the fusion is found in the database of known fusions, called Quiver (http://quiver.archerdx.com), it will be called Strong.

The fusion is indicated with the bulls eye icon () to indicate the fusion is in the Quiver database and the breakpoint is an exact match.

The fusion is indicated with this icon () to indicate it is a known fusion but the breakpoint is not an exact match.

NOTE: If a fusion is found in Quiver this overrides all other criteria and it will be reported in the Strong Evidence tab regardless of how weak the evidence might otherwise be, or what other criteria for a Strong fusion fail to be met. If the fusion is not found in Quiver, then the candidate can still be called Strong if it satisfies the rest of the criteria.

3. Percent GSP2 - Percent GSP2 is the proportion of breakpoint spanning reads that support the candidate relative to the total number of reads spanning the breakpoint.

Percent GSP2 needs to be at least 10% in order for the fusion to be considered Strong.

Changing the FUSION_PERCENT_OF_GSP2_READS parameter under "Fusion UI Settings" adjusts this cutoff.

If a candidate fails to meet the Percent GSP2 cutoff, it will be annotated with the () icon and be placed in the Weak category.

4. Minimum unique start sites - Within the population of breakpoint spanning reads that support the candidate, there will be a distribution of unique start sites.

There must be at least 3 unique start sites to be considered Strong.

Changing the MIN_UNIQUE_START_SITES_FOR_STRONG_FUSION parameter under "Fusion UI Settings" adjusts this cutoff.

If a candidate fails to meet the unique start site cutoff, it will be annotated with the () icon and placed in the Weak category.

5. Unless it is found in Quiver, a candidate must not trigger any of the conditions found in the Absence of Negative Evidence section to be called Strong Evidence. With one exception (detailed in section 3.2.6.2), if any of the negative evidence criteria are met, the candidate will be called Weak.

3.2.6.2 Negative Evidence

In order to be called a strong fusion, the following criteria must **not** be met (unless the fusion is present in Quiver).

Exon-intron fusion. If the fusion sequence on one side of the breakpoint is found to be entirely intronic (which is indicative of a DNA mispriming event), the fusion will be classified as exon-intronic. This is to distinguish such events from those that utilize an intronic cryptic splice site resulting in just an internal portion of the fusion sequence corresponding to an intron, such as can be found in a common ALK-EML4 variant.

Exon-intronic fusions are indicated with the () icon and placed in the Weak category.

Mispriming. If significant sequence similarity is found between the fusion partners, the event is likely to be due to mispriming. Additionally, if the fusion breakpoint is less then 5bp from the GPS2, then similarity is assumed.

Likely off-target mispriming events are indicated with the () icon and placed in the Weak category.

Paralogs. Archer Analysis compares the identities of the fusion partners with a list of known paralogs taken from the Ensembl database.

Known paralogs are indicated with the () icon and placed in the Weak category.

Low confidence. Annotation of each fusion consensus is done by aligning the sequencing to the human genome with BLAST. The quality of these alignments, and the confidence of resulting annotation, depends on alignment length and repeat content of the sequence.

Events with a low confidence annotation are indicated with the () icon and placed in the Weak category.

Cross-contamination. If a low expressing fusion candidate shows significant similarity to a high expressing fusion in the same analysis, it will be considered the result of intra-run cross-contamination.

Likely intra-run cross-contamination events are indicated with the ( ) icon and placed in the Weak category.

Transcriptional readthrough. Fusion transcripts of interest are generally derived from a genomic translocation event. However, fusion transcripts can also arise from failure to properly terminate transcription from a gene such that transcription continues on into the next gene downstream (if it is on the same strand).

Transcriptional readthrough events are indicated with the ( ) icon and placed, by default in the **STRONG** category.

Transcriptional readthrough events are placed in the **STRONG** category by default because by representing actual molecules produced in cells, they are technically not false positives.

Transcriptional readthrough events can be made to appear in the **WEAK** category by changing the value from "OFF" to "ON of the parameter CALL_TRANSCRIPTIONAL_READTHROUGH_EVENTS_WEAK under the “Fusion UI Settings.”

This is the only criterion in the "Negative Evidence" category that can be configured in this way.

Finally, if a fusion has been found to pass all the Positive Evidence criteria (except for the presence in Quiver, which is optional), and does not trigger any of the Negative Evidence criteria, it will be indicated with the ( ) icon.

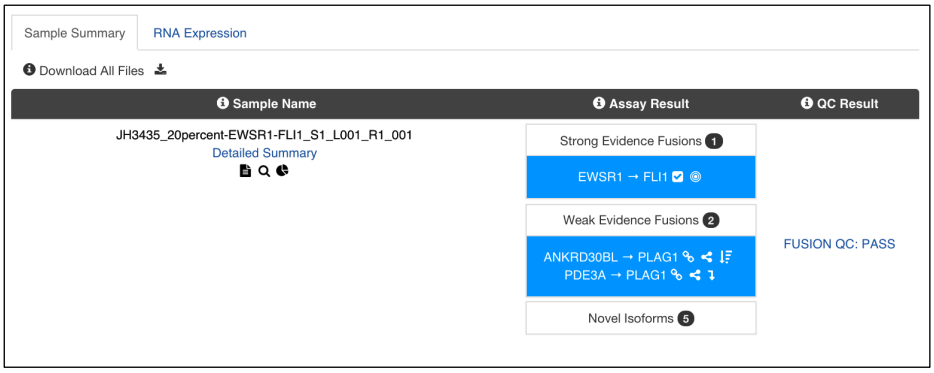


Figure 28. Example of a sample with both strong and weak evidence fusions.

3.2.7 Strong Evidence and Weak Evidence Tabs

In addition to the detailed information about the fusions, the links below each of the fusion candidates provide even more information about the fusion candidates.

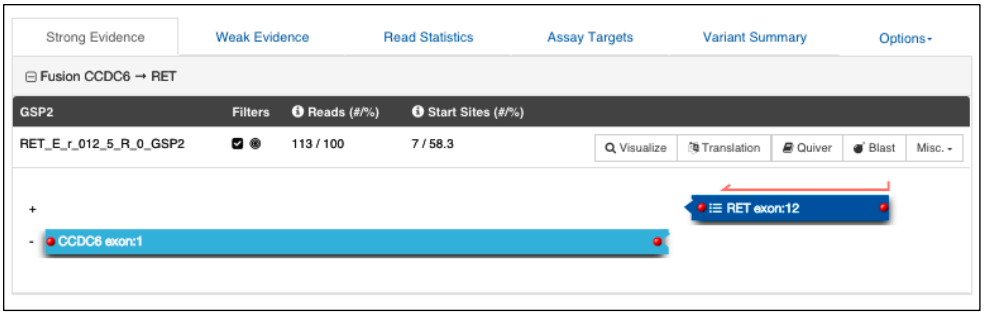


Figure 29. Full Summary page showing the Strong Fusion Candidates tab

Visualize

Opens a new window with the reads supporting the gene fusion

Blast

Opens a new window with the results of a BLAST search against the human database of the DNA sequence of the consensus

Translation

Opens a new window showing the results of the protein translation prediction of the gene fusion product

A link on this page, marked Blast will open a new window with the results of a BLAST search against the human *protein* database

3.2.7.1 Visualization of fusion candidate-supporting reads in JBrowse

To verify the accuracy of the fusion calls, it is possible to visualize the reads supporting the breakpoint and fusion call (or wildtype call). Click the " Visualize "link beneath the fusion of interest and a JBrowse window will open up with the supporting data.

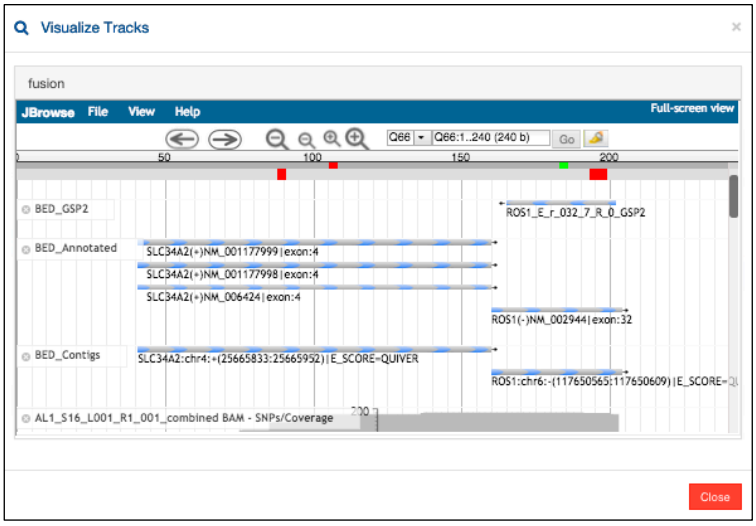


Figure 30. Reads supporting the fusion call between SLC34A2 and ROS1

The JBrowse viewer is provided in a small dialog box but clicking the "**Full-Screen View**" in the top right corner will open the genome browser in its own window or tab in the web browser.

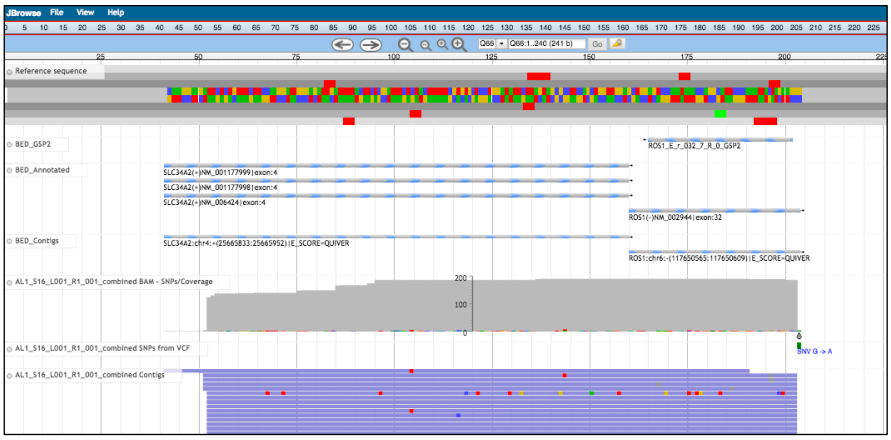


Figure 31. Full screen view of the SLC34A2-ROS1 fusion in JBrowse

The genome browser shows the constructed consensus sequence of the fusions as the reference sequence to which the de-duplicated reads from the original FASTQ files are mapped.

3.2.7.2 Tracks in the JBrowse view

There are several tracks in the JBrowse view and a description of each is provided below.

BED_GSP2

A track showing the location of the gene specific primer 2 (GSP2) used to detect the gene fusion

BED_Annotated

A track showing the gene and exon number annotation for each of the fusion partners.

If a gene has multiple transcripts/isoforms, each of the possible isoform annotation is shown as a separate line

BED_Contigs

The different contigs that make up this fusion

The annotation of each of the regions in this track contains the name, the location on the HG19 genome as well as the BLAST E score on which the gene annotation was based. **NOTE**: Fusions are not aligned against HG19 directly. First, a fusion reference is created from HG19 using annotations. This is done in order to see reads aligned across the breakpoint.

*_coverage

The coverage track for the reads supporting the fusion

* Contigs

The unique (de-duplicated) reads supporting the fusion

SNPs from VCF

Any differences in the reads from the reference are listed here as a VCF track (produced by Freebayes variant caller)

3.2.7.3 Protein Translation prediction

For gene fusions, the Archer Analysis software determines if the resulting protein is in or out of frame with the 5'-gene fusions candidate transcripts (all possible isoforms) and creates a amino acid sequence that can then be subsequently BLASTed against the human protein database using the web interface to the NCBI BLAST application, to provide more information about the fusion protein.

Clicking the Translation box provides a new window with the prediction results and a link to the BLAST application.

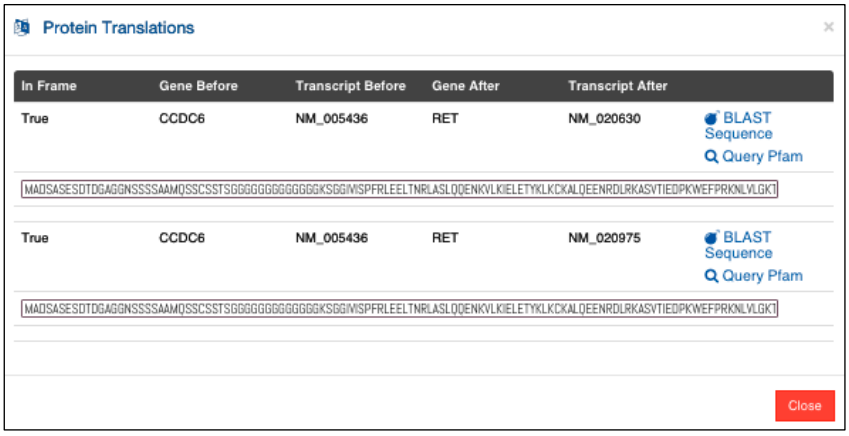


Figure 32. Result of protein translation prediction of gene fusion candidates

3.2.8 Read Statistics Page

The reads statistics page contains basic metrics for the sample library, such as mapping percentages, on target percentages and DNA/RNA statistics.

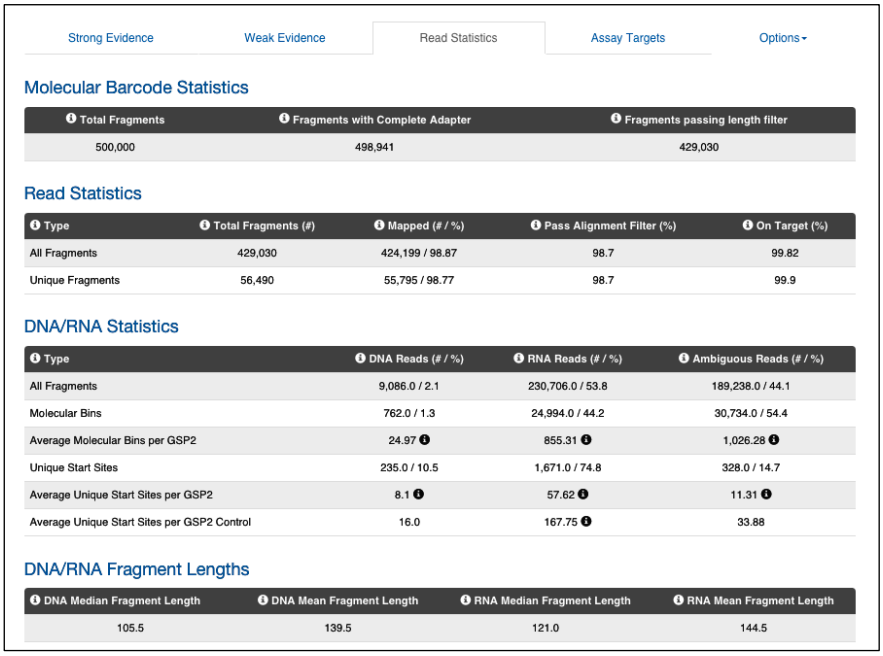


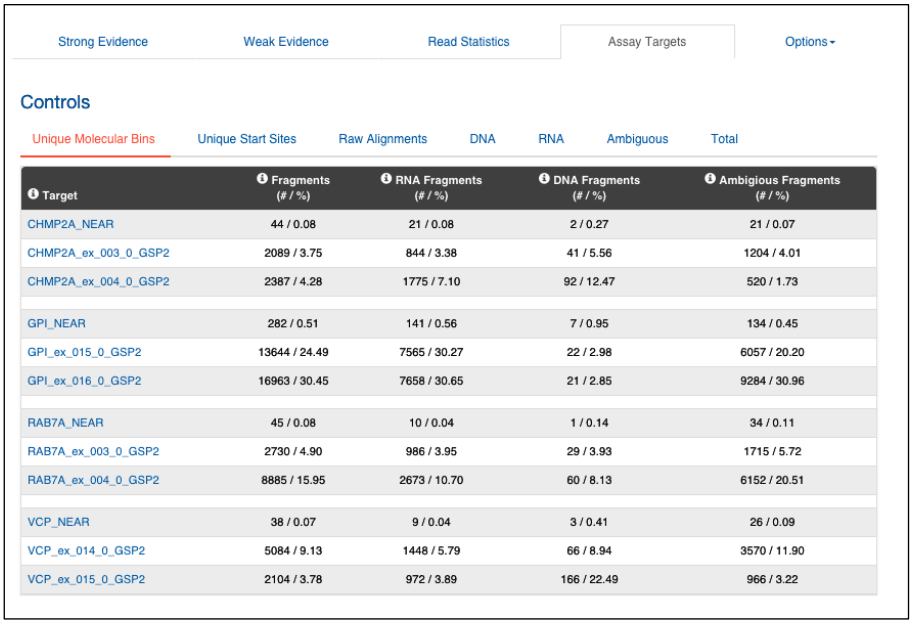
Figure 33. Read Statistics page contains basic metrics on the sample library

Most of the metrics contain help texts (mouseover the icons) and a full description of the fields can be found in section 8.3 on page 82.

3.2.9 Assay Targets Page

The assay targets page contains detailed coverage information for each of GSP2 in the selected target region. It contains information about the number of unique fragments based on the number of unique molecular barcodes or bins, as well as the number of unique fragments based on the number of unique start sites. The data is also separated by the type of molecule the fragments seem to originate from (DNA, RNA or AMBIGUOUS).

In addition to the coverage for the actual targets/primers, coverage data is also provided for those reads that fall anywhere inside the gene locus. Those targets can be recognized by the “NEAR” tag, as shown in Figure 34.



Target	Fragments (# / %)	RNA Fragments (# / %)	DNA Fragments (# / %)	Ambigious Fragments (# / %)
CHMP2A_NEAR	44 / 0.08	21 / 0.08	2 / 0.27	21 / 0.07
CHMP2A_ex_003_0_GSP2	2089 / 3.75	844 / 3.38	41 / 5.56	1204 / 4.01
CHMP2A_ex_004_0_GSP2	2387 / 4.28	1775 / 7.10	92 / 12.47	520 / 1.73
GPI_NEAR	282 / 0.51	141 / 0.56	7 / 0.95	134 / 0.45
GPI_ex_015_0_GSP2	13644 / 24.49	7565 / 30.27	22 / 2.98	6057 / 20.20
GPI_ex_016_0_GSP2	16963 / 30.45	7658 / 30.65	21 / 2.85	9284 / 30.96
RAB7A_NEAR	45 / 0.08	10 / 0.04	1 / 0.14	34 / 0.11
RAB7A_ex_003_0_GSP2	2730 / 4.90	986 / 3.95	29 / 3.93	1715 / 5.72
RAB7A_ex_004_0_GSP2	8885 / 15.95	2673 / 10.70	60 / 8.13	6152 / 20.51
VCP_NEAR	38 / 0.07	9 / 0.04	3 / 0.41	26 / 0.09
VCP_ex_014_0_GSP2	5084 / 9.13	1448 / 5.79	66 / 8.94	3570 / 11.90
VCP_ex_015_0_GSP2	2104 / 3.78	972 / 3.89	166 / 22.49	966 / 3.22

Figure 34. Example of an Assay Targets page. The page contains 7 different tabs with more detailed information

For a full description of the tabs and fields on this page, see section 8.4 on page 85.

3.2.10 Definition of RNA, DNA and AMBIGUOUS fragments

Since the Archer Anchored Multiplex PCR assay is usually performed on total nucleic acid samples, it is possible that the assay contains both RNA and DNA molecules, which is reflected in the final library.

For successful detection of gene fusions, a large percent of the reads should originate from RNA molecules. The definition of each type of read is provided below:

RNA Reads

Reads that have a greater than 100 bp gap in the alignment to hg19

Reads that are split, and thus have intronic and/or multiple exon content.

DNA Reads

Reads that do not have a greater than 100 bp gap in the alignment to hg19

Include an intron region that is at least 10% of the read length

AMBIGUOUS Reads

Reads that fall *completely inside an exon* making it impossible to distinguish RNA reads from DNA reads

Only RNA reads will lead to a successful gene fusion analysis; a low amount of RNA reads in the library often means low sample quality. The QC criteria that is used for RNA sample passing is an average of 10 unique start sites per control gene.

3.2.11 CNV Summary results

If the DNA Copy Number Variation assay was selected, the CNV Summary page will show the detailed copy number results for all the genes in the assay as well as visual summary of the results. The page is separated into two sub-tabs as well:

- Data Summary
- Visual Summary

3.2.11.1 Data Summary

The Data summary page shows the copy number for each of the genes in the assay that are marked as CNV targets and is separated into three sections

- Strong Evidence Copy Gains
- Strong Evidence Copy Losses
- Below Set Threshold

Genes that show significant copy gains or losses are defined as those genes that fall above or below the cutoff set in the Analysis Settings page which defines both a copy number threshold *and* minimum p-value threshold.

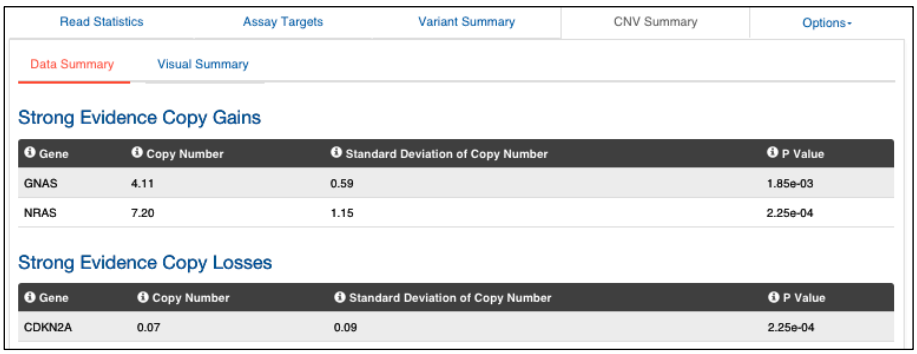


Figure 35. Strong evidence copy gains and losses results in the data summary sub-tab

The results for all other genes that were marked as CNV targets is listed in the "**Below Set Threshold**" section.

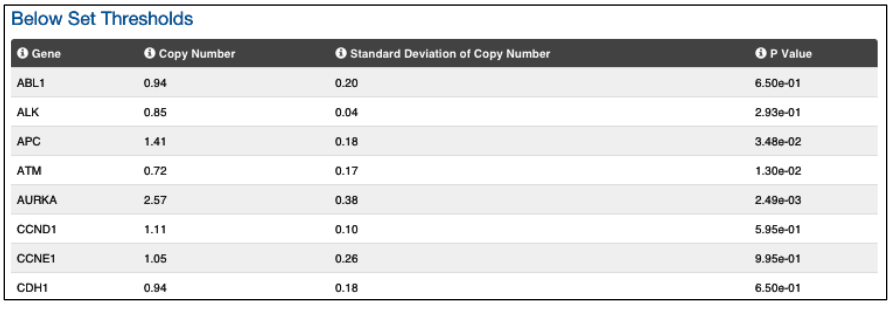


Figure 36. Subset of the copy number results for genes that fall outside the threshold range

3.2.11.2 *Visual Summary*

The visual summary sub-tab contains a PDF file with the results for each GSP2. There are usually multiple primers for each CNV target gene and the consistent results for each of the individual primers is a good indication that the assay was successful.

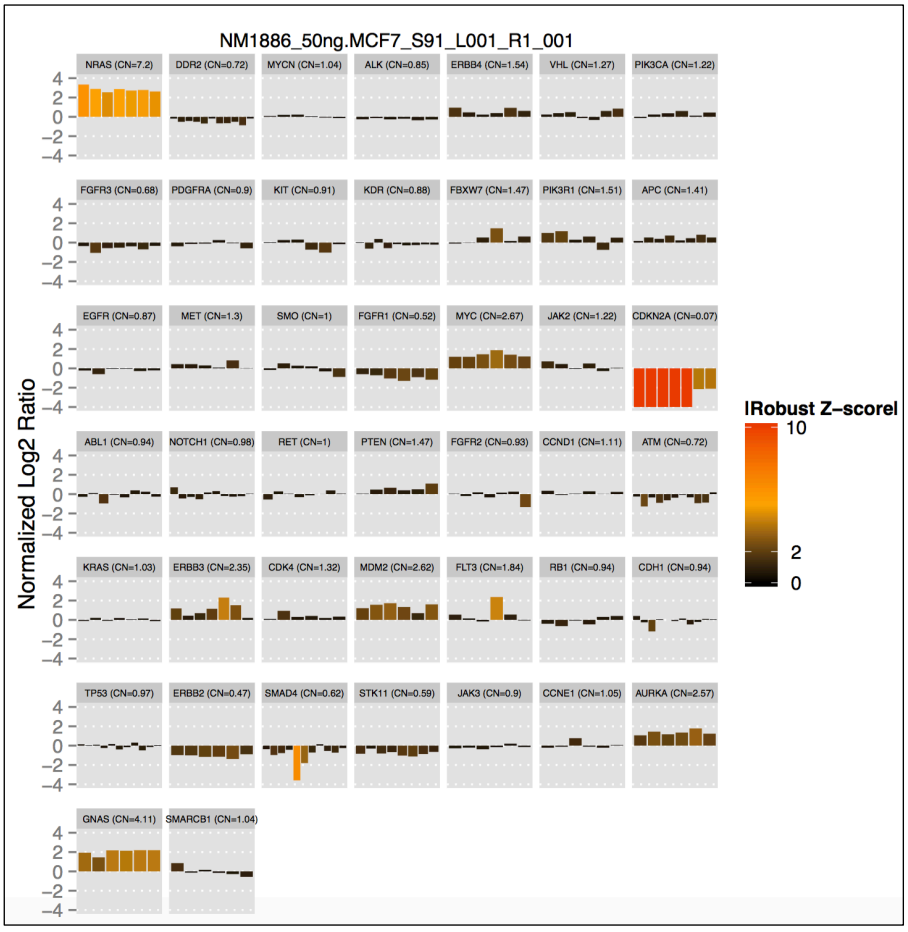


Figure 37. PDF with the visual representation of the individual target primers for the Copy Number assay. The bar represents the Log2 value of the normalized copy number ratio (0 = normal for that gene). The color represents the Z-score for the target primer

NOTE: In some occasions the target primers do not show roughly equal results for the copy number. In the example in Figure 37, the gene for CDKN2A shows that four of the primers are complete without any coverage (complete deletion) but two of the primers show some coverage and are clearly not completely deleted. This is an indication that there was a heterozygous deletion of gene CDKN2A and the breakpoint of the deletion was somewhere between primer 4 and 5 (counting from the left). This result was confirmed with Sanger sequencing (Results not shown).

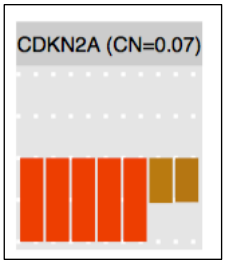


Figure 38. Evidence for a heterozygous partial deletion of gene CDKN2A in sample for MCF-7 cell line

3.2.12 SNP/InDel detection

Archer Analysis 4.1 offers the ability to search for variants with both RNA and DNA assays. You can choose to search for non-targeted mutations, targeted mutations, or both. See section 3.2.12.1 for further details.

3.2.12.1 Targeted mutations results

The Variant Summary page for a sample that used the Targeted Mutation option shows only those variants that were defined in the target variant call format (VCF) file and if "Show All" is turned OFF. The names of the variants consist of the gene name and the amino acid mutation as defined in the target VCF file.

For each variant there are three possible outcomes:

IS PRESENT

Indicates that the mutation was found, either as a heterozygous or homozygous *alt* call (0/1 or 1/1)

NOT PRESENT

Indicates that the mutation was not present in this sample. The location was found to be homozygous reference (0/0).

NO CALL

Indicates that there was not enough coverage (or coverage that was of low quality) to make a call for homozygous reference or a mutation call.

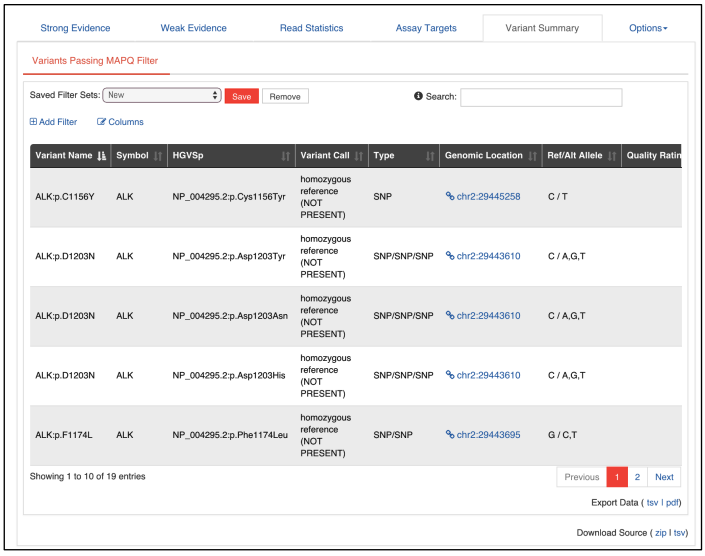


Figure 39. The result of a targeted mutation analysis shows the results for all targeted mutation, including when there is insufficient coverage or No Call.

3.2.12.2 Non targeted mutations results

If a targeted mutation file was not selected for a SNPs/InDels analysis job, the variant caller will call all variants that pass the filtering parameters. See analysis settings section 3.2.19 for more information about changing the filtering parameters.

The Variant Summary page in the **Detailed Summary** section will show the results of the variant detection as shown below:

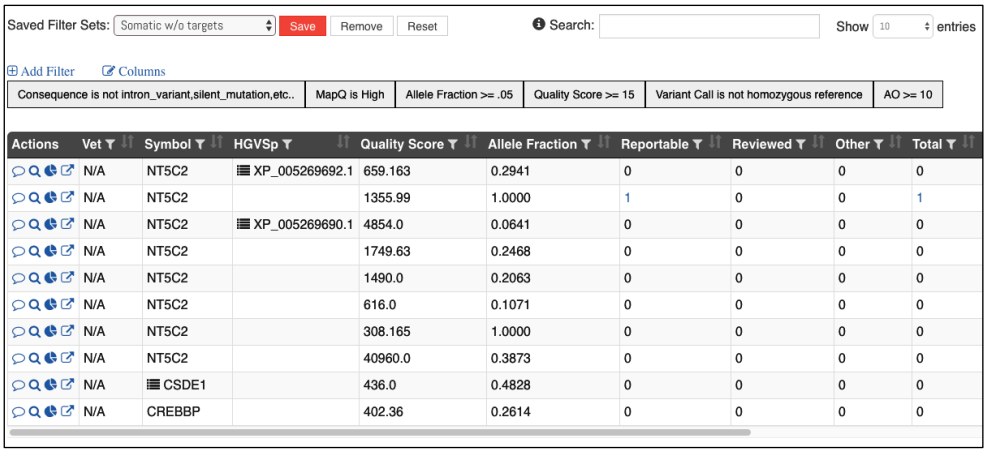
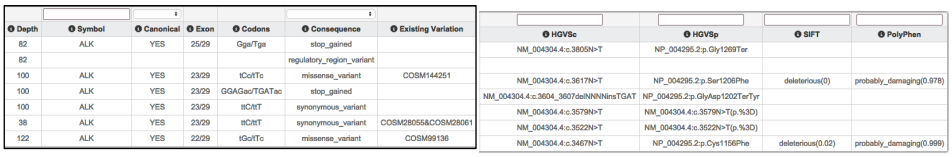


Figure 40. Variant detection results of a non-targeted mutation analysis. The link in the Genomic Location column will show the read coverage at the position using JBrowse

The "**Variant Name**" column will be empty for variants that were not detected using targeted mutation analysis.

In addition to the variant calls and the basic statistics about the coverage etc. the Archer Analysis software also provides more detailed information about the potential effect of the variant. This is achieved by using the Variant Effect Predictor tool, developed by the EBI at ENSEMBL (http://www.ensembl.org/info/docs/tools/vep).



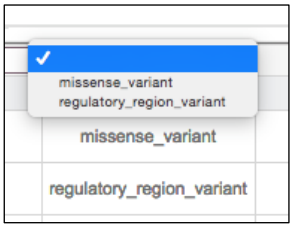
Depth	Symbol	Canonical	Exon	Codons	Consequence	Existing Variation	HGVSc	HGVSp	SIFT	PolyPhen
82	ALK	YES	25/29	Gga/Tga	stop_gained		NM_004304.4:c.3805N>T	NP_004295.2:p.Gly1269Ter		
82					regulatory_region_variant					
100	ALK	YES	23/29	tCc/tTc	missense_variant	COSM144251	NM_004304.4:c.3617N>T	NP_004295.2:p.Ser1206Phe	deleterious(0)	probably_damaging(0.978)
100	ALK	YES	23/29	GGAGac/TGATac	stop_gained		NM_004304.4:c.3604_3607delNNNNinsTGAT	NP_004295.2:p.GlyAsp1202TerTyr		
100	ALK	YES	23/29	ttC/ttT	synonymous_variant		NM_004304.4:c.3579N>T	NM_004304.4:c.3579N>T(p.%3D)		
38	ALK	YES	23/29	ttC/ttT	synonymous_variant	COSM28055&COSM28061	NM_004304.4:c.3622N>T	NM_004304.4:c.3622N>T(p.%3D)		
122	ALK	YES	22/29	tGc/tTc	missense_variant	COSM99136	NM_004304.4:c.3467N>T	NP_004295.2:p.Cys1156Phe	deleterious(0.02)	probably_damaging(0.999)

Figure 41. Annotation of the variants with the predicted effect of the variant on the protein function using VEP and ClinVar

3.2.13 Filtering of results

Each of the columns from the Variant Summary can be filtered either by selecting a value from the drop down menu OR by typing in the search commands in the text box

Columns that use the drop down menu show a box with arrows and selecting the dropdown shows the values that are found in the columns.



The columns with text boxes can be used to enter a search text and any rows NOT containing the search text will be removed from view.

The search text can contain special operators to make more complex search queries such as combining fields with Boolean statements or, for columns containing numerical values, the rows that contain a value that is less than or more than some search value can be selected as well.

More information about the search operators can be found in Figure 42.

Operator	Description	Type	Example
<	Values lower than search term are matched	number	<1412
<=	Values lower than or equal to search term are matched	number	<=1412
>	Values greater than search term are matched	number	>1412
>=	Values greater than or equal to search term are matched	number	>=1412
=	Exact match search: only the whole search term(s) is matched	string / number	=Sydney
*	Data containing search term(s) is matched (default operator)	string / number	*Syd
!	Data that doesn't contain search term(s) is matched	string / number	!Sydney
{	Data starting with search term is matched	string / number	{S
}	Data ending with search term is matched	string / number	}y
||	Data containing at least one of the search terms is matched	string / number	Sydney || Adelaide
&&	Data containing search terms is matched	string / number	>4.3 && <25.3
[empty]	Empty data is matched		[empty]
[nonempty]	Data which is not empty is matched		[nonempty]
rgx:	A regular expression is used to match data		rgx:de$

Figure 42. Search operators for the Variant Summary search boxes

3.2.13.1 Definition of the columns in the variant summary

See section 8.4.8 for a definition of the values in each of the columns.

3.2.13.2 Download the VCF or Text file for the Variant Summary

Filtering variants in the web browser can be somewhat difficult if there are thousands of variants, due to the limited space on the page. It may be necessary to use the variant results in another filtering and reporting tool.

The VCF file with the results for the sample or a tab delimited file with the information can be downloaded by selecting the "**Download Source (vcf | tab)**" option below the variant summary page.

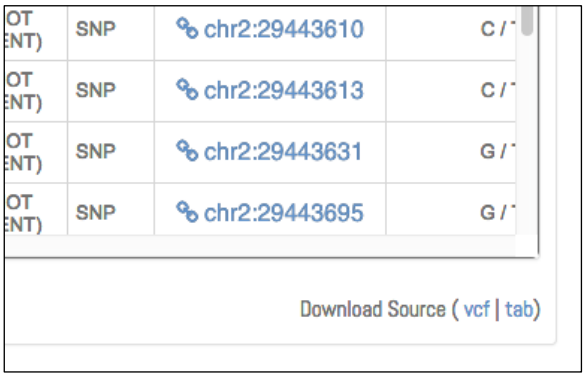


Figure 43. To download the VCF or tab delimited results for the variants, select the option at the bottom right side of the Variant Summary page

3.2.14 Visualization of the variants using JBrowse

The reads supporting the variants can be directly visualized from within the Archer Analysis software using JBrowse, by clicking the location link in the Genomic

Location column (e.g. chr2:1481231). This brings up a separate window with JBrowse, focused on the selected location.

Figure 44. Visualization of variants in JBrowse

There are a number of tracks in the JBrowse window:

refGene

A track showing the genes and each of their transcripts from the RefSeq database

Right-click on the transcript to see its name and follow the link for a search of the transcript at the NCBI website

cosmic

A track showing the location and identifiers from the COSMIC mutation database

Right-click on the feature to follow the link of the identifier at the Sanger website

SNPs from VCF

A track showing the variants detected (or targeted)

Read1(2)_contigs_coverage

The coverage plot of the read coverage and mismatch ratio for reads 1 and 2 separately

Read1(2) Contigs

The mapped reads for reads 1 and 2 separately

3.2.15 Customizable and printable PDF reports

In many occasions it is required to have a physical copy of the results of the analysis and this can be accomplished with the Sample and Run level reports.

For each sample a report can be generated of all or a subset of the results. The sample report generator contains a filtering option to only select the results of interest for the report.

To create a Sample Level report, select the "**Sample Reports**" icon ( ) from the **Sample Summary** page or select menu option "**Sample Reports**" from the Options menu on the **Detailed Summary** page.

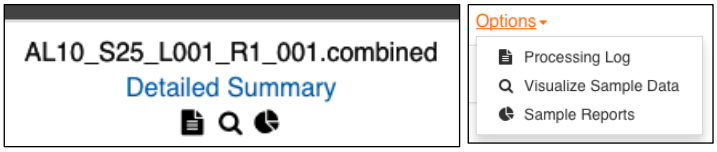


Figure 45. Create Sample Report by selecting the Sample Report icon on the Sample Summary page or the Sample Summary Options menu from the Details Summary page

A dialog box will appear with a list of all the available sample-level reports. By default only one report is available, but other custom reports can be added (See section 3.2.19 "Adding custom reports" for more information on how to add a custom report)

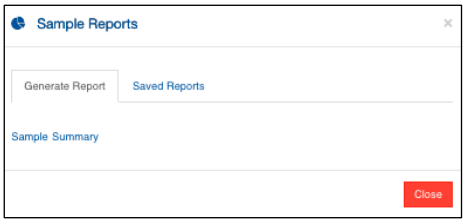


Figure 46. By default only Sample Summary is listed

Select the "**Sample Summary**" report. A new window will appear which is the filtering stage of the report creation. Select the desired information for the report:

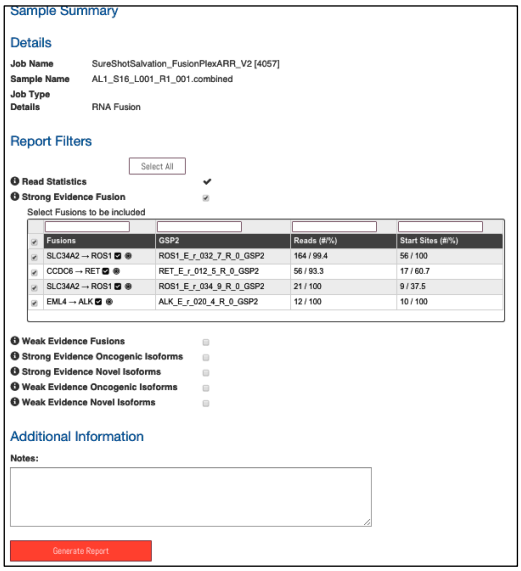


Figure 47. Summary report with default filter selection options

By default, only a subset of the data is automatically selected for inclusion in the report, but other data can be included as well by selecting the checkbox for the type of data. For instance, if strong novel isoforms should be included in the report, select

the checkbox for "**Strong Evidence Novel Isoforms**". A list of the strong novel isoforms will appear that can be used to select the specific isoforms that should be added.

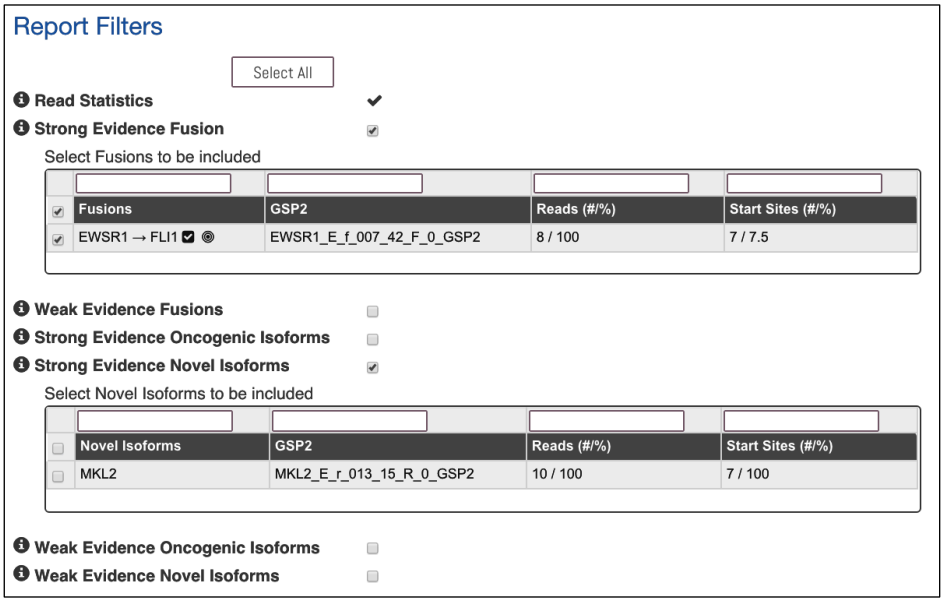


Figure 48. To add a section to the report, select the data type for that section and a list of the data type will appear from which a selection can be made

The items in the default sections are always selected but for any new section that is added the items are NOT selected. Click the box next to the item to include the item on the list.

After the filtering and selection is finished, click the "**Generate Report**" button to create the report PDF.

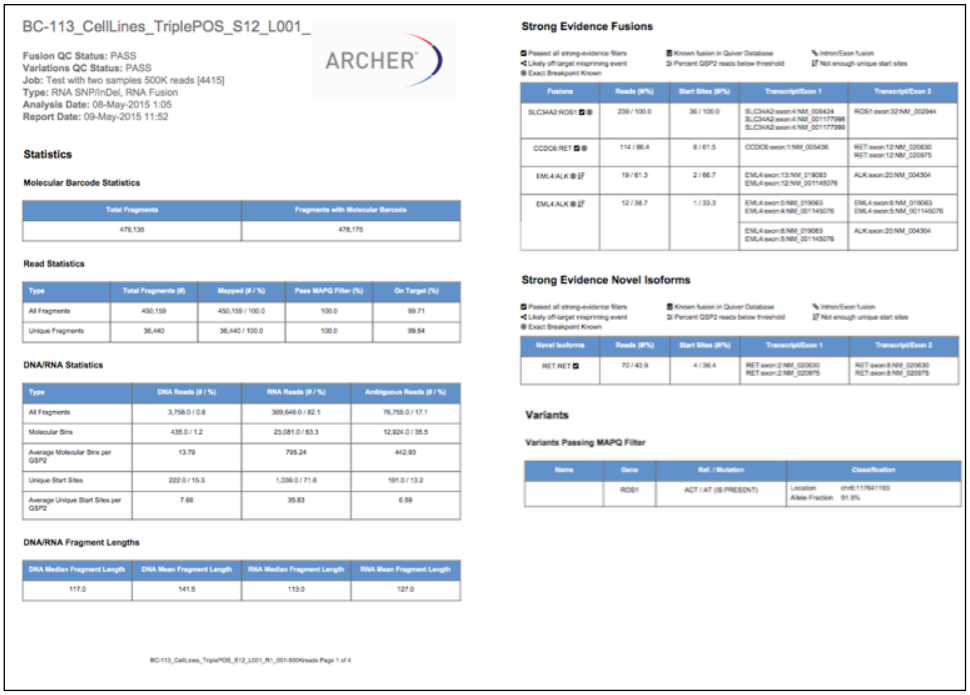


Figure 49. Example of a PDF sample report

3.2.15.1 Filtering of Variants for the report

In the case that the analysis includes germline or somatic variants, the variants can be filtered in the same way as on the "**Variant Summary**" page. See section 3.2.13 for more information on how to filter the variants.

3.2.15.2 The PDF reports are saved with the sample run

Each report that is generated is automatically saved with the sample run. This allows the report to be re-printed in the future should the need arise.

In the Sample Report dialog box, select the "Saved Reports" tab to see previously saved reports.

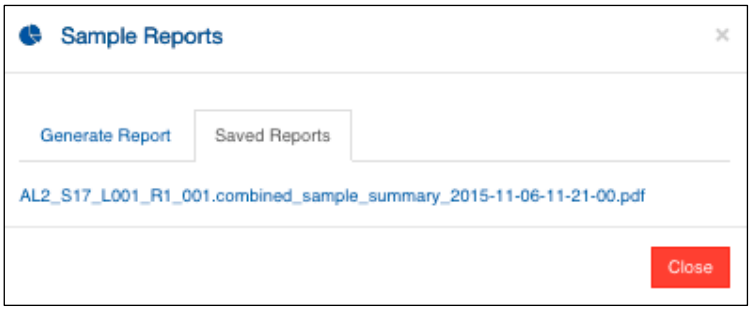


Figure 50. The "Saved Reports" tab contains previously saved reports

Click on the PDF of interest to open the report. The name of the report contains the date and time the PDF was created as the suffix in the format "YYYY-MM-DD-HH-MM-SS".

NOTE: immediately after generating the report, the report does not always appear in the "**Saved Reports**" section. The report IS saved, but the web page has not been refreshed. To see the report appear in the "**Saved Reports**" section RELOAD the page.

3.2.16 Jobs that complete with errors

It is possible that a job will finish with errors. This is often due to the analysis of very large FASTQ files that can result in the job running out of memory. If a job fails during a run, the sample(s) that fail will be indicated with a red bar as shown below.

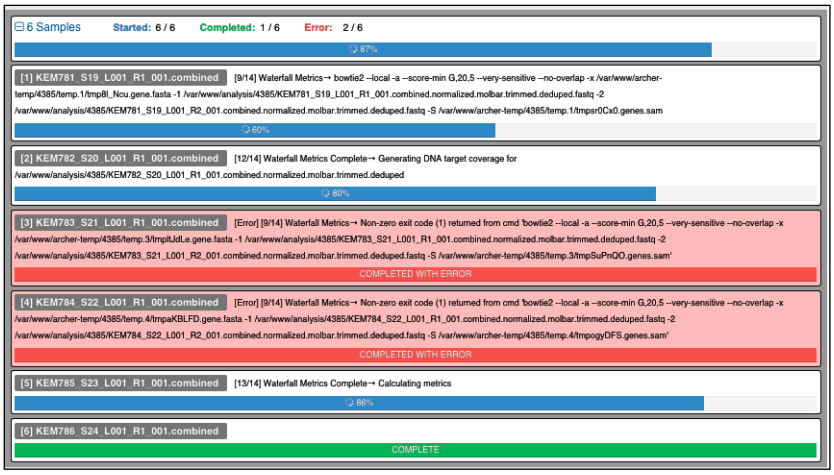


Figure 51. Sample 3 and 4 have completed with errors

After all the samples have finished processing, the job will be marked as "**COMPLETED_ERROR**". Consult the log files (icon) to determine the reason for the failed job and contact tech@archerdx.com for assistance.

The number in the job queue ([3] and [4] in the example above) corresponds to the log file with the name "2.log.stderr.txt". Find that log file in the list of log files or click the "**[Processing Log]**" link in the Summary Page or select "Processing Log" from the Options menu to download the log file to the local computer.

In some cases, the error prevented the sample from being processed completely and this is indicated with the message "**Sample processed with errors**" in the job summary page.

Figure 52. Sample processed with errors message of a sample that failed.

3.2.17 Adding and managing custom content

The Archer Gene Fusion kits can be used with different targets from those provided in the FusionPlex Panel kits. A design can be created from the Assay Designer website: http://assay.archerdx.com, as well as by the Designer Pro team.

Once a design is created, the target regions used for the design will be returned as an email attachment. This file can be uploaded to the Archer Analysis Software by selecting the "**Custom Targets**" menu item in the "**Admin**" menu.

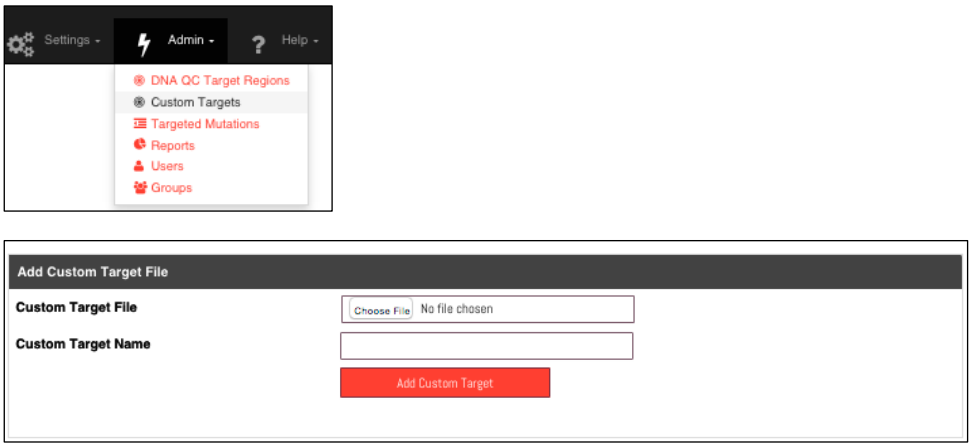


Figure 53. Adding Custom Targets

Select the "Choose File" option and then select the GTF-formatted file provided by the primer design website, or the Designer Pro team. Provide a name for the target and press "**Add Custom Target**" to add the assay to the analysis software.

3.2.17.1 Display, download or delete target files

To display the contents of target file, click the display icon ( ).

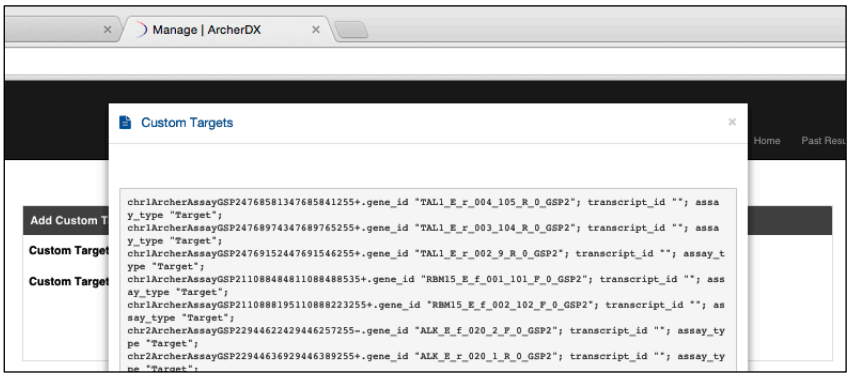


Figure 54. Content of the target file is displayed in the web browser for review

To download the Gene Transfer Format (GTF) file to the local computer, click the download icon ( ). The original GTF file, with the name of the original GTF file will be downloaded to the local computer.

To delete a target file, click the delete icon ( ). This will only be possible if the target file is not used in a previous job. To remove the target file, first remove all the jobs that use this custom target file in the "Past Results" section.

3.2.17.2 Format of the target region file

The target region file is a GTF-formatted file containing the names and locations of the primers. The format is described at this site: http://genome.ucsc.edu/FAQ/FAQformat.html#format4

The ***gene_id*** tag contains the name of the target region. **NOTE: The format of this identifier is important;** The name of the target gene should be the first part of the name, followed by an underscore "_" character. The rest of the identifier is not important.

The ***transcript_id*** tag contains the name of the transcripts the target is derived from (can often be "empty ").

The ***assay_type*** tag contains the type of target, can either be "Target" or "Control" (case insensitive) to indicate the assay target and the control targets (RNA only) respectively.

```
chr2    ArcherAssay GSP2    29432692    29432715    0   -   .   gene_id "ALK_E_025_GSP2_CRIZ"; transcript_id ""; assay_type "Target";
chr2    ArcherAssay GSP2    29436923    29436947    0   +   .   gene_id "ALK_E_024_GSP2_CRIZ"; transcript_id ""; assay_type "Target";
chr2    ArcherAssay GSP2    29443659    29443686    0   +   .   gene_id "ALK_E_023_GSP2_CRIZ"; transcript_id ""; assay_type "Target";
chr2    ArcherAssay GSP2    29445218    29445254    0   +   .   gene_id "ALK_E_022_GSP2_Rep"; transcript_id ""; assay_type "Target";
chr2    ArcherAssay GSP2    29445436    29445467    0   +   .   gene_id "ALK_E_r_021_3_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr2    ArcherAssay GSP2    29446369    29446389    0   +   .   gene_id "ALK_E_r_020_4_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr2    ArcherAssay GSP2    29448407    29448426    0   +   .   gene_id "ALK_E_r_019_13_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr6    ArcherAssay GSP2    117639379   117639410   0   +   .   gene_id "ROS1_E_r_037_12_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr6    ArcherAssay GSP2    117641155   117641188   0   +   .   gene_id "ROS1_E_r_036_11_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr6    ArcherAssay GSP2    117642505   117642531   0   +   .   gene_id "ROS1_E_r_035_10_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr6    ArcherAssay GSP2    117645531   117645568   0   +   .   gene_id "ROS1_E_r_034_9_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr6    ArcherAssay GSP2    117647535   117647572   0   +   .   gene_id "ROS1_E_r_033_8_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr6    ArcherAssay GSP2    117650567   117650604   0   +   .   gene_id "ROS1_E_r_032_7_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr6    ArcherAssay GSP2    117658461   117658498   0   +   .   gene_id "ROS1_E_r_031_6_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr10   ArcherAssay GSP2    43607570    43607594    0   -   .   gene_id "RET_E_r_008_1_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr10   ArcherAssay GSP2    43608306    43608331    0   -   .   gene_id "RET_E_r_009_15_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr10   ArcherAssay GSP2    43609019    43609038    0   -   .   gene_id "RET_E_r_010_17_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr10   ArcherAssay GSP2    43609933    43609953    0   -   .   gene_id "RET_E_r_011_16_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr10   ArcherAssay GSP2    43612037    43612067    0   -   .   gene_id "RET_E_r_012_5_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr10   ArcherAssay GSP2    43613826    43613847    0   -   .   gene_id "RET_E_r_013_14_R_0_GSP2"; transcript_id ""; assay_type "Target";
chr10   ArcherAssay GSP2    43615119    43615139    0   -   .   gene_id "RET_E_14_GSP2_CRIZ"; transcript_id ""; assay_type "Target";
chr3    ArcherAssay GSP2    128516851   128516879   255 +   .   gene_id "RAB7A_ex_003_0_GSP2"; transcript_id ""; assay_type "CONTROL";
chr3    ArcherAssay GSP2    128525364   128525388   255 +   .   gene_id "RAB7A_ex_004_0_GSP2"; transcript_id ""; assay_type "CONTROL";
chr9    ArcherAssay GSP2    35059530    35059555    255 -   .   gene_id "VCP_ex_014_0_GSP2"; transcript_id ""; assay_type "CONTROL";
chr9    ArcherAssay GSP2    35059090    35059118    255 -   .   gene_id "VCP_ex_015_0_GSP2"; transcript_id ""; assay_type "CONTROL";
chr19   ArcherAssay GSP2    34890195    34890214    255 +   .   gene_id "GPI_ex_015_0_GSP2"; transcript_id ""; assay_type "CONTROL";
chr19   ArcherAssay GSP2    34890484    34890513    255 +   .   gene_id "GPI_ex_016_0_GSP2"; transcript_id ""; assay_type "CONTROL";
chr19   ArcherAssay GSP2    59063661    59063682    255 -   .   gene_id "CHMP2A_ex_003_0_GSP2"; transcript_id ""; assay_type "CONTROL";
chr19   ArcherAssay GSP2    59063444    59063466    255 -   .   gene_id "CHMP2A_ex_004_0_GSP2"; transcript_id ""; assay_type "CONTROL";
```

Figure 55. Example of target region file

3.2.18 Adding custom targeted mutation sets

Target mutations can be defined by providing the analysis software a VCF formatted file with all the targeted mutations. Select the "**Targeted Mutations**" menu item from the "**Admin**" menu.

Provide the VCF formatted file and enter a name for the targeted mutation file then click the "**Add Targeted Mutation**" button to upload the file. This file must be in VCF format.

3.2.18.1 Display, download or delete targeted mutation files

To display the contents of targeted mutations file, click the display icon ( ).

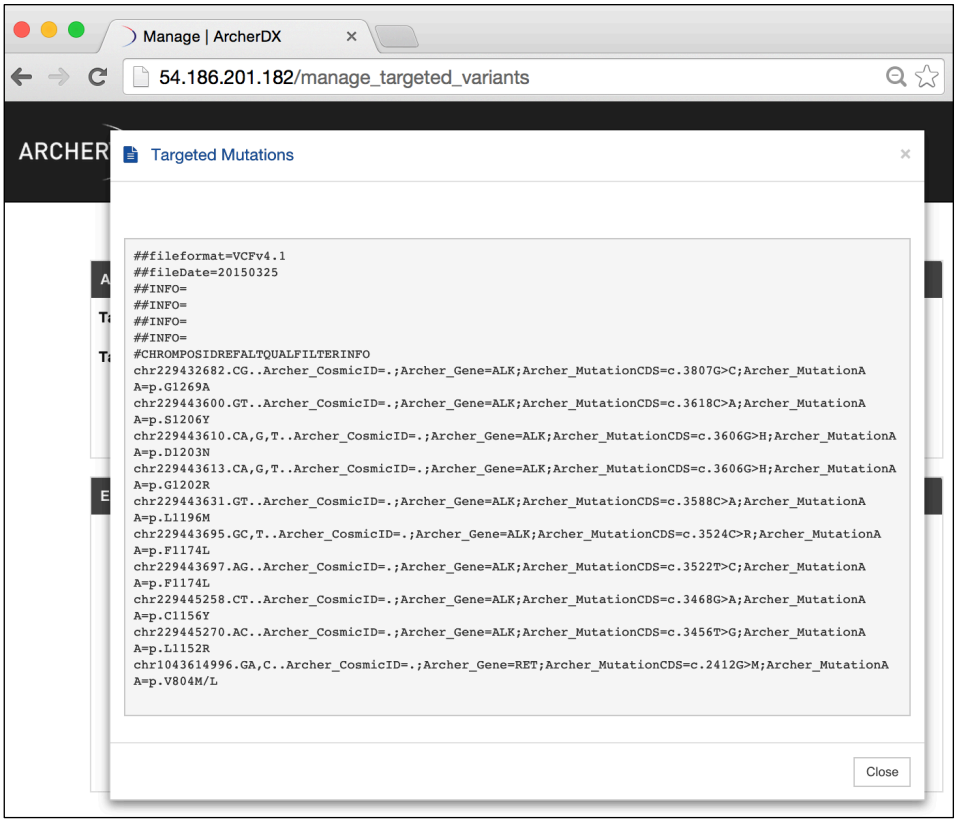


Figure 56. Content of the targeted mutations file is displayed in the web browser for review

To download the VCF file to the local computer, click the download icon ( ). The VCF file will be downloaded to the local computer.

To delete a target file, click the delete icon ( ). This will only be possible if the target file is not used in a previous job. To remove the target file, first remove all the jobs that use this custom target file in the "**Past Results**" section.

3.2.18.2 Format of the targeted mutation File

The format of the targeted mutation file is a standard VCF formatted file (Version 4.1) with some special INFO tags.

For more information about the format of VCF file, see the following web page:

https://samtools.github.io/hts-specs/VCFv4.1.pdf

The special INFO tags are defined in the Figure 57:

INFO Tag	Description
Archer_CosmicID	The mutation identifier from the COSMIC database
Archer_Gene	The gene the mutation is found in
Archer_MutationCDS	The location and mutation in HGVS format with respect to the CDS. Eg. C.3522G>C
Archer_MutationAA	The location and mutation in HGVS format with respect to the Amino Acid sequence. eg p.F1174L

Figure 57. The definition of the special INFO tags defined in the Archer Targeted Mutations VCF file

An example of a targeted mutation file is provided below:

```
##fileformat=VCFv4.1
##fileDate=20150325
##INFO=<ID=Archer_CosmicID,Number=1,Type=String,Description="COSMIC Mutation Database identifier. Provided by Archer Analysis">
##INFO=<ID=Archer_Gene,Number=1,Type=String,Description="Gene name. Provided by Archer Analysis">
##INFO=<ID=Archer_MutationCDS,Number=1,Type=String,Description="Mutation in the CDS. Provided by Archer Analysis">
##INFO=<ID=Archer_MutationAA,Number=1,Type=String,Description="Amino Acid mutation in the protein. Provided by Archer Analysis">
#CHROM  POS     ID      REF     ALT     QUAL    FILTER  INFO
chr2    29432682        .       C       G       .       .       Archer_CosmicID=.;Archer_Gene=ALK;Archer_MutationCDS=c.3807G>C;Archer_MutationAA=p.G1269A
chr2    29443600        .       G       T       .       .       Archer_CosmicID=.;Archer_Gene=ALK;Archer_MutationCDS=c.3618C>A;Archer_MutationAA=p.S1206Y
chr2    29443610        .       C       A,G,T   .       .       Archer_CosmicID=.;Archer_Gene=ALK;Archer_MutationCDS=c.3606G>H;Archer_MutationAA=p.D1203N
chr2    29443613        .       C       A,G,T   .       .       Archer_CosmicID=.;Archer_Gene=ALK;Archer_MutationCDS=c.3606G>H;Archer_MutationAA=p.G1202R
chr2    29443631        .       G       T       .       .       Archer_CosmicID=.;Archer_Gene=ALK;Archer_MutationCDS=c.3588C>A;Archer_MutationAA=p.L1196M
chr2    29443695        .       G       C,T     .       .       Archer_CosmicID=.;Archer_Gene=ALK;Archer_MutationCDS=c.3524C>R;Archer_MutationAA=p.F1174L
chr2    29443697        .       A       G       .       .       Archer_CosmicID=.;Archer_Gene=ALK;Archer_MutationCDS=c.3522T>C;Archer_MutationAA=p.F1174L
chr2    29445258        .       C       T       .       .       Archer_CosmicID=.;Archer_Gene=ALK;Archer_MutationCDS=c.3468G>A;Archer_MutationAA=p.C1156Y
chr2    29445270        .       A       C       .       .       Archer_CosmicID=.;Archer_Gene=ALK;Archer_MutationCDS=c.3456T>G;Archer_MutationAA=p.L1152R
chr10   43614996        .       G       A,C     .       .       Archer_CosmicID=.;Archer_Gene=RET;Archer_MutationCDS=c.2412G>M;Archer_MutationAA=p.V804M/L
```

Figure 58. Example of a targeted mutations file in VCF format

NOTE: It is important that the REF column contains the correct reference base. If this is not correct or "." the analysis will fail since the analysis software verifies the expected reference base.

When creating a new VCF file for use in the Archer Analysis software, it is recommended to use the provided VCF file as a guide to ensure the proper format of the file is maintained. It is also important that the INFO header fields are added to any custom VCF file, since these are required to be present for a valid VCF format.

The result of a targeted mutation analysis is shown below:

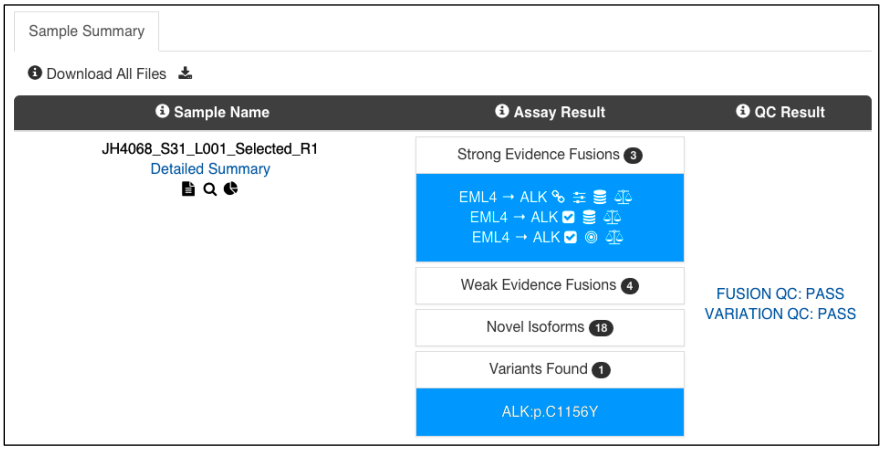
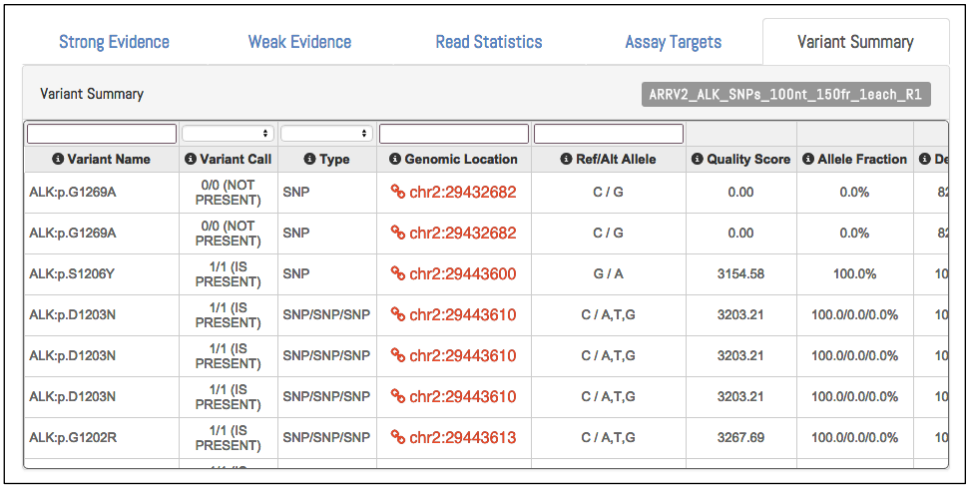


Figure 59. The result of a targeted mutation shows 1 mutation under the "Variants Found" section



Variant Name	Variant Call	Type	Genomic Location	Ref/Alt Allele	Quality Score	Allele Fraction	De
ALK:p.G1269A	0/0 (NOT PRESENT)	SNP	chr2:29432682	C / G	0.00	0.0%	82
ALK:p.G1269A	0/0 (NOT PRESENT)	SNP	chr2:29432682	C / G	0.00	0.0%	82
ALK:p.S1206Y	1/1 (IS PRESENT)	SNP	chr2:29443600	G / A	3154.58	100.0%	10
ALK:p.D1203N	1/1 (IS PRESENT)	SNP/SNP/SNP	chr2:29443610	C / A,T,G	3203.21	100.0/0.0/0.0%	10
ALK:p.D1203N	1/1 (IS PRESENT)	SNP/SNP/SNP	chr2:29443610	C / A,T,G	3203.21	100.0/0.0/0.0%	10
ALK:p.D1203N	1/1 (IS PRESENT)	SNP/SNP/SNP	chr2:29443610	C / A,T,G	3203.21	100.0/0.0/0.0%	10
ALK:p.G1202R	1/1 (IS PRESENT)	SNP/SNP/SNP	chr2:29443613	C / A,T,G	3267.69	100.0/0.0/0.0%	10

Figure 60. The result of a targeted mutation analysis shows all target mutations, regardless of variant call status ("Present", "Not Present", "No Call")

3.2.19 Adding custom reports

In addition to the standard sample reports provided with Archer Analysis, it is also possible to create custom reports that can be adjusted and branded to your own specifications. The reports can be made at the run level or at the sample level and consist of HTML templates that contain special fields that represent the data for the run or the sample. An example for a custom run level report is shown in Figure 61.

```
<html>
<head>
    <title>Job Overview</title>
    <meta name="report_level" content="run"/>
    <!-- Remove the _ from save_to_file to make this a pdf report -->
    <meta name="_save_to_file" content="job_overview.pdf"/>

    <style>
    #report_title {font-size:25px;margin-bottom:8px;}
    #inline_meta label { display:inline-block;width:150px; }
    #logo {float:right}
    </style>

</head>
<body>
    <div style="width:700px">
        <div id="logo">
            <img src="/static/images/archerdx_logo_reporting.png" width="250px"/>
        </div>
         <div id="inline_meta">
            <br/>
            <div id="report_title">{{job_name}}</div>
            <br/>
            <label>Job ID:</label> {{job_id}}
            <br/>
            <label>Type:</label> <span>{{job_type_description|removetags:"br div"}}</span>
            <br/>
            <label>Start Date:</label> {{ start_time|date:"d-M-Y" }} {{ start_time|time:"G:i" }}
            <br/>
            <label>Complete Date:</label> {{ complete_time|date:"d-M-Y" }} {{ complete_time|time:"G:i" }}
            <br/>
            <label>Submitted By:</label> {{user_email}}
            <br/>
            <label>Job Status:</label> {{job_status_name}}
            <br/>
            <label>Platform:</label> {{platform}}
            {% if target_region_name %}
                <br/>
                <label>Target Region:</label> {{target_region_name}}
            {% endif %}
            {% if targeted_variant_name %}
                <br/>
                <label>Target Variant:</label>  {{targeted_variant_name}}
            {%endif %}

        </div>


    </div>
    <br/>
     <label>Sample Names</label>
     {% for sample in samples %}
         <ul>
             <li>{{sample}}</li>
         </ul>
     {% endfor %}
</body>
</html>
```

Figure 61. Example of HTML template to create a run level custom report

For more information on custom reports, please contact tech@archerdx.com.

3.2.20 Changing the analysis settings

The analyses use a set of parameters that determine the sensitivity and QC Metrics cutoffs for each analysis. The analysis parameters can be adjusted in the "User Settings" menu item under "Settings."

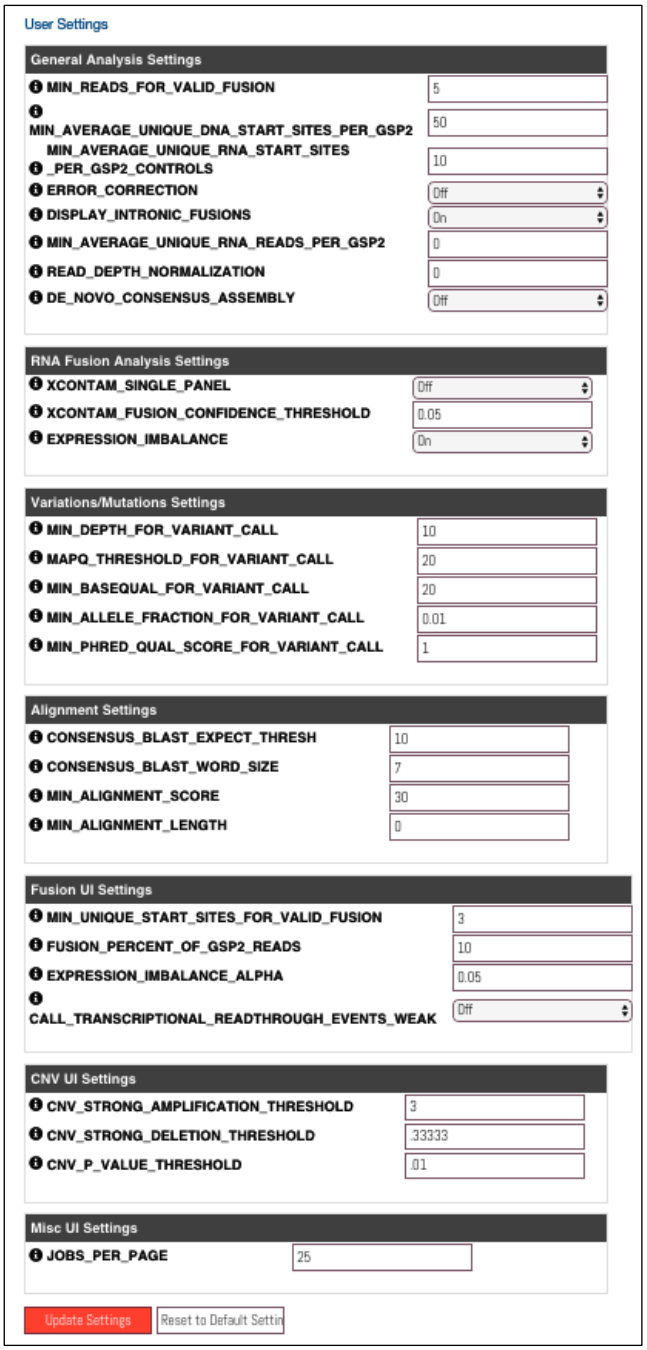


Figure 62. The analysis setting page allows the parameter defaults to be changed. The changes are specific for the user and are NOT global

The definition for each of the analysis settings can be found in 8.1 on page 79. In addition, the popover help text can be accessed by moving the mouse over the small "**I**" icon ( ) for more information.

The analysis settings are unique for each user and changing them will only affect the analysis settings for that user.

To review the analysis settings that were used for a run, select the tag icon ( ) below the job details for the run. This will display the settings used as compared to the default values.

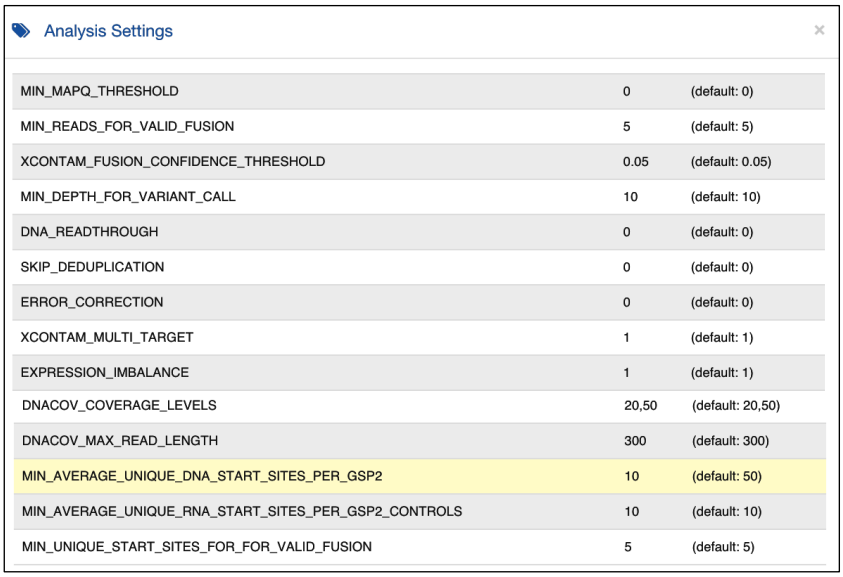


Figure 63. Analysis setting used in a job. The yellow box indicates the setting that was changed from the default.

4 Description of output files and formats

The Archer Analysis Software produces a number of result files for each sample (see section 3.2 for download these files). These files are stored in the jobs' output directory which on the virtual machine will be a location such as: /var/www/analysis/job_id, where job_id is the job number as reported on the Past Results page in the Analysis GUI.

The following set of files are created:

Table 1. Description of the files created in the Archer Analysis Software. [SAMPLE1] represents read 1, for the sample named SAMPLE. (".molbar" in the name of the file indicates that molecular barcode based de-duplication was used.)

File	Description
[SAMPLE1]_full_results.txt	Results summary file. See description

	of the format in section 4.1
summaries/ Summary.r_and_d_results.txt	A subset of the Results Summary file metrics for all samples of the run in a table. See description of the subset of metrics in section 4.1
Summary.non_deduped.counts.txt	Coverage data for each target for all samples for read1 and read2 combined
`Summary.non_deduped.counts_pct.txt`	Coverage data for each target for all samples for read1 and read2 combined as a percentage of the total reads for that sample
`Summary.deduped.counts.txt`	Coverage data for each target for all samples for read1 and read2 combined using unique (de-duped) reads
`Summary.deduped.counts_pct.txt`	Coverage data for each target for all samples for read1 and read2 combined using unique (de-duped) reads as a percentage of the total reads for that sample
Summary.non_deduped.readN.counts.txt	Coverage data for each target for all samples for read1 or read2
`Summary.non_deduped.readN.counts_pct.txt`	Coverage data for each target for all samples for read1 or read2 as a percentage of the total reads for that sample
`Summary.deduped.readN.counts.txt`	Coverage data for each target for all samples for read1 or read2 using unique (de-duped) reads
`Summary.deduped.readN.counts_pct.txt`	Coverage data for each target for all samples for read1 or read2 using unique (de-duped) reads as a percentage of the total reads for that sample
`Summary-All-Variants.vcf`	Variant summary file in VCF format, containing all called variants for all files (Only present for SNP/InDel assay types)
`[0-9]*.log.stderr.txt`	Log file for the analysis workflow. The number [0-9]* indicates there are as many configuration files as there are samples. The number indicates the sample number.
`workflow.config.[0-9]*`	Configuration file for the analysis

	workflow. Contains the analysis parameters used in the analysis. The number [0-9]* indicates there are as many configuration files as there are samples. The number indicates the sample number.
cnv_sample_sheet.tsv	Tab delimited file describing the sample-designation for the CNV analysis. Samples are designated either as CASE or CONTROL and the group identifier allows for matched CASE-CONTROL analyses. The REPLICATE column is not used and is there for backward compatibility.
[SAMPLE1].molbar.trimmed.deduped.bam	BAM file of the de-duplicated reads mapped to the genome
[SAMPLE1].molbar.trimmed.deduped.bam.bai	Index file for the BAM file.
plots/ [SAMPLE1].molbar.trimmed.deduped.pdf|svg	Histogram of the fragment length distribution in PDF or SVG format of the de-duped reads
plots/ [SAMPLE1].molbar.trimmed.deduped_DNA|_RNA.pdf|.svg	Histogram of the fragment length distribution in PDF or SVG format of the de-duped reads, separated by the classification of the reads (DNA or RNA reads; RNA reads span an exon-exon boundary)
[SAMPLE1][.molbar].trimmed.deduped.vcf	Variant summary for sample SAMPLE1 in VCF version 4.1 format
consensus_sequence_info/ [SAMPLE1].molbar.trimmed.deduped.fusion.contigs.fasta	File containing the refernce sequence(s) for fusion candidates in FASTA format. Each FASTA contig is for one fusion candidate. The consensus is then constructed against this reference. (20 N bases padded on each side for visualization in JBrowse)
consensus_sequence_info/ [SAMPLE1].molbar.trimmed.deduped.fusion.contigs.bed	File containing the JBrowse references and BED tracks. The locations are relative to the contigs.fasta file
VCFs/ [SAMPLE1].molbar.trimmed.deduped.fusion.contig.vcf	A VCF file containing the variants found when mapping the original FASTQ files to the consensus sequence of the fusion transcript (fusion.contigs.fasta file). This will show the difference between the HG19 and actual fusion consensus.
consensus_sequence_info/ [SAMPLE1].molbar.trimmed.deduped.consensus.fusions.f	File containing the consensus sequence(s) of the fusion candidates

asta	in FASTA format. Each FASTA contig represents one fusion candidate. The consensus is constructed from the FASTQ files and represents the actual consensus sequence of the fusion candidate (this is in contrast to the file *.fusion.contigs.fasta which is merely a concatenation of the reference sequence of the fusion partners)
[SAMPLE1].molbar].trimmed.deduped.consensus.fusions.global.bed	File containing the location of the gene fusions. Each line in the BED file refers to the location of the region in the hg19 genome, using the base pair numbering of the genome sequence.
[SAMPLE1].molbar.trimmed.deduped.consensus.fusions.local.bed	File containing the location of the gene fusions. Each line in the BED file refers to the location of the region in the consensus file, using the base pair numbering of the consensus sequence. The name of the region represents the location on the genome
[SAMPLE1].molbar.trimmed.deduped.consensus.fusions.ann.gtf	File containing the location of the gene fusions. Each line in the GTF file refers to the location of the regions of the gene fusions in reference genome coordinates
wildtype	Files with ***.wildtype.*** instead of ***.fusions.*** are similar to the fusion files described above
[SAMPLE1].[READ1|2].molbar.trimmed.depthhist.png	Histogram plot of the duplication rate for read1 and read2 separately. X-axis represents the depth and the Y-axis represents the number of barcodes with that specific number of reads (depth)
counts/[SAMPLE1].molbar.trimmed.deduped.[READ_n|READ_EITHER].counts	The coverage at each target or gene *before* de-duplication (the “deduped” in the file name is an error)
counts/[SAMPLE1].molbar.trimmed.deduped.UNIQUE_[READ_n|READ_EITHER].counts	The coverage at each target or gene *after* de-duplication for read1 or read2 or either read1 *and* read2)
[SAMPLE1].molbar.trimmed.deduped.UNIQUE_READ_EITHER.counts.csv	The results of the Copy Number Variation analysis. For a description of the format, see Table 3

4.1 Full results summary

The results of the Archer Analysis Software for each sample is provided in a single file containing all relevant metrics and some assay results (fusions).

The filename below follows the following structure where SAMPLE1 is the name of the original FASTQ file for READ1 without the ".fastq" extension, respectively.

[SAMPLE1]_full_results.txt

Below is the result of the sample with the FASTQ file "1305_S12_L001_R1_001.fastq".

1305_S12_L001_R1_001_full_results.txt

The format of the file is a simple KEY VALUE pair, where the KEY and VALUE are separated by a TAB character.

Table 2. The definition of the KEY values

Key	Description of the value
SAMPLE_NAME	The name of the sample, which is the name of the FASTQ file, without the ".fastq" extension. For paired end reads it is the concatenated names of the original FASTQ files separated by an underscore (_).
FUSION_QC_FILTER	The results of the Fusion Quality Control filter. Will be PASS if it passes all QC filter settings or will indicate one or more values, indicating a potential issue with the library. Specific for the RNA Fusion type of analysis
VARIATIONS_QC_FILTER	The results of the Variations Quality Control filter. Will be PASS if it passes all QC filter settings or will indicate one or more values, indicating a potential issue with the library. Specific for the SNP/InDel types of analysis
MOLBAR_TOTAL_NUM_READS	Total number of read (pairs) in the FASTQ file used as input
MOLBAR_READS_WITH_CORRECT_COMMON_REGION	Total number of reads with the correct "common region" (no mismatches allowed)
MOLBAR_FRACTION_OF_TOTAL	Fraction of reads that have a perfect or near perfect common region (MOLBAR_READS_WITH_CORRECT_COMMON_REGION + MOLBAR_READS_WITH_CLOSE_COMMON_REGION)/ MOLBAR_TOTAL_NUM_READS
JUNK_PERCENT	The percentage of total reads that appear to be random sequence and not the result of the Archer AMP technology. Defined as the fraction of reads that don't align to the genome

	MOLBAR_TOTAL_NUM_READS - FRAGMENTS_ALIGNED_FILTERED)/ MOLBAR_TOTAL_NUM_READS
FRAGMENT_TOTAL	The total number of reads (pairs) for this sample
FRAGMENT_ALIGNED	The total number of reads (pairs) that align with the human genome (hg19)
FRAGMENT_ALIGNED_PERCENT	100*FRAGMENT_ALIGNED / FRAGMENT_TOTAL
FRAGMENT_ALIGNED_FILTERED	Total number of reads (pairs) that align and also pass the mapping quality/alignment score filtering step
FRAGMENT_ALIGNED_FILTERED_ON_TARGET	Number of reads that have at least one of the pair aligned, pass alignment filtering, and is on-target with at least 1 base pair overlap
FRAGMENT_ALIGNED_FILTERED_ON_TARGET_PERCENT	100*FRAGMENT_ALIGNED_FILTERED_ON_TARGET / FRAGMENT_ALIGNED_FILTERED
FRAGMENT_ALIGNED_FILTERED_OFF_TARGET	Number of reads where at least one of the pair is determined to be off-target and the other is not on-target
FRAGMENT_ALIGNED_FILTERED_OFF_TARGET_PERCENT	100*FRAGMENT_ALIGNED_FILTERED_OFF_TARGET / FRAGMENT_ALIGNED_FILTERED
READ_n_TOTAL	The total number (n) of reads for this sample.
READ_n _ALIGNED	The total number of reads that align to the reference genome
READ_n _ALIGNED_PERCENT	The percentage of aligned reads, compared to the number of total reads (READ_n_TOTAL)
READ_n_ALIGNED_FILTERED	Number of reads with a mapping quality ≥ 35. Reads with a mapping quality below that value are removed from the analysis
READ_n_ALIGNED_FILTERED_ALONE_ON_TARGET	The number (n) of reads (after alignment score filtering) that align with at least 1 base on the define target region (controls and fusion gene candidates).
READ_n_ALIGNED_FILTERED_ALONE_ON_TARGET_PERCENT	The percentage of on-target reads, relative to the Filtered Total Molecules (READ_n_ALIGNED_FILTERED)
READ_n_ALIGNED_FILTERED_ALONE_OFF_TARGET	The number of read n reads (after alignment score filtering) that align and do not fall inside the defined target region (controls and fusion gene candidates)
READ_n_ALIGNED_FILTERED_ALONE_OFF_TARGET_PERCENT	The percentage of off-target reads is relative to the Filtered Total Molecules (READ_n_ALIGNED_FILTERED)

READ_n_ALIGNED_FILTERED_EITHER_ON_TARGET	The number of read1 *or* read2 reads (after alignment score filtering) that align with at least 1 base on the define target region (controls and fusion gene candidates).
READ_n_ALIGNED_FILTERED_EITHER_ON_TARGET_PERCENT	The percentage of on-target reads is relative to the Filtered Total Molecules (READ_n_ALIGNED_FILTERED)
READ_n_ALIGNED_FILTERED_EITHER_OFF_TARGET	The number of read1 *or* read2 reads (after alignment score filtering) that align and do not fall inside the defined target region (controls and fusion gene candidates)
READ_n_ALIGNED_FILTERED_EITHER_OFF_TARGET_PERCENT	The percentage of off-target reads is relative to the Filtered Total Molecules (READ_n_ALIGNED_FILTERED)
UNIQUE_*	The keys described above prefixed with UNIQUE have the same definition, but refer to the reads AFTER de-duplication (either alignment- or Molecular Barcoded-based)
FRAGMENT_EXON	The number of fragments (read-pairs) that contain a split alignment
FRAGMENT_INTRON	The number of fragments (read-pairs) that have at least 10% of the read covering an intron
FRAGMENT_EXON_PERCENT	100*FRAGMENT_EXON/(FRAGMENT_EXON+FRAGMENT_intron). NOTE: a read could be counted TWICE if it both covers an exon AND an intron for at least 10%
FRAGMENT_MEAN_LENGTH	The average (deduced) length of the fragment (read pair)
FRAGMENT_MEDIAN_LENGTH	The median (deduced) length of the fragment (read pair)
RNA_FRAGMENT_MEAN_LENGTH	The average (deduced) length of the fragment (read pair) that are unambiguously categorized as RNA (span exon/exon boundaries)
RNA_FRAGMENT_MEDIAN_LENGTH	The median (deduced) length of the fragment (read pair) that are unambiguously categorized as RNA (span exon/exon boundaries)
DNA_FRAGMENT_MEAN_LENGTH	The average (deduced) length of the fragment (read pairs) that are unambiguously categorized as DNA (reads that cross from exon into intron without being split)
DNA_FRAGMENT_MEDIAN_LENGTH	The median (deduced) length of the fragment (read pair) that are unambiguously categorized as DNA (reads that cross from exon into intron without being split)

TOTAL_DNA_READS	Reads that are unambiguously categorized as DNA reads (reads that cross from exon into intron without being split)
TOTAL_RNA_READS	Reads that are unambiguously categorized as RNA reads, (reads that span an exon/exon boundary)
TOTAL_AMBIG_READS	Reads that fall completely inside an exon or intron and therefore cannot be categorized as either DNA or RNA
RNA_FRAGMENT_MEAN_LENGTH	The average length of fragments for reads classified as RNA reads (spanning at least two exons)
RNA_FRAGMENT_MEDIAN_LENGTH	The median length of fragments for reads classified as RNA reads (spanning at least two exons)
DNA_FRAGMENT_MEAN_LENGTH	The average length of fragments for reads classified as DNA reads (read that runs from exon to its next intron)
DNA_FRAGMENT_MEDIAN_LENGTH	The median length of fragments for reads classified as DNA reads (read that runs from exon to its next intron)
AMBIG_FRAGMENT_MEAN_LENGTH	The average length of fragments for reads classified as Ambiguous reads
AMBIG_FRAGMENT_MEDIAN_LENGTH	The median length of fragments for reads classified as Ambiguous reads
UNIQUE	The six keys described above that contain the string UNIQUE have the same definition, but refer to the reads AFTER de-duplication (Molecular Barcoded)
READS_PER_TARGET_n_TARGET_GENE	The name of the target region for which the data is relevant. n is the target number, starting at 0
READS_PER_TARGET_n_TYPE	The type of target. Can be either “TARGET” or “CONTROL”.
READS_PER_TARGET_n_READ_1	Reads covering at least 1 bp of the target for read 1
READS_PER_TARGET_n_READ_2	Reads covering at least 1 bp of the target for read 2 (for paired-end reads only)
READS_PER_TARGET_n_READ_EITHER	Reads covering at least 1 bp of the target for read 1 or read2
READS_PER_TARGET_n_READ_1_PERCENT	Percentage of the reads covering at least 1 bp of the target for read 1, compared to the total number of reads (READ_1_ALIGNED_FILTERED)
READS_PER_TARGET_n_READ_2_PERCENT	Percentage of the reads covering at least 1 bp of the target for read 2, compared to the total number

	of reads (READ_2_ALIGNED_FILTERED)
READS_PER_TARGET_n_READ_EITHER_PERCENT	Percentage of the reads covering at least 1 bp of the target for read 1 or read 2, compared to the total number of fragments (read pairs) (FRAGMENT_ALIGNED_FILTERED)
READS_PER_TARGET_n_RNA_READS	Number of reads classified as RNA reads (spanning two or more exons) that map to this GSP2
READS_PER_TARGET_n_RNA_READS_PERCENT	The percentage of the reads classified as RNA reads (spanning two or more exons) that map to this GSP2
READS_PER_TARGET_n_DNA_READS	Number of reads classified as DNA reads (reads that span a consecutive intron/exon boundary) that map to this GSP2
READS_PER_TARGET_n_DNA_READS_PERCENT	The percentage of the reads classified as DNA reads (reads that span a consecutive intron/exon boundary) that map to this GSP2
READS_PER_TARGET_n_NUCLEIC_ACID_READS	Number of total reads (DNA + RNA + Amb.) that map to this GSP2
EADS_PER_TARGET_n_NUCLEIC_ACID_READS_PERCENT	The percentage of total reads (DNA + RNA + Amb.) that map to this GSP2
READS_PER_TARGET_n_AMBIG_READS	Number of reads classified as AMBIGUOUS reads (reads falling entirely inside a single exons) that map to this GSP2
READS_PER_TARGET_n_AMBIG_READS_PERCENT	The percentage of the reads classified as AMBIGUOUS reads (reads falling entirely inside a single exons) that map to this GSP2
DNA_FRAGMENT_GSP2_n_MEAN_LENGTH	The mean length of the fragments for reads classified as DNA reads (reads spanning a consecutive intron/exon boundary) that map to this GSP2
DNA_FRAGMENT_GSP2_n_MEDIAN_LENGTH	The median length of the fragments for reads classified as DNA reads (reads that span a consecutive intron/exon boundary) that map to this GSP2
RNA_FRAGMENT_GSP2_n_MEAN_LENGTH	The mean length of the fragments for reads classified as RNA reads (reads spanning two or more exons) that map to this GSP2
RNA_FRAGMENT_GSP2_n_MEDIAN_LENGTH	The median length of the fragments for reads classified as RNA reads (reads spanning two or more exons) that map to this GSP2
AMBIG_FRAGMENT_GSP2_n_MEAN_LENGTH	The mean length of the fragments for reads classified as AMBIGUOUS reads (reads falling entirely inside a single exons) that map to this

	GSP2
AMBIG_FRAGMENT_GSP2_n_MEDIAN_LENGTH	The median length of the fragments for reads classified as AMBIGUOUS reads (reads falling entirely inside a single exons) that map to this GSP2
UNIQUE_*	The keys described above prefixed with UNIQUE have the same definition, but refer to the reads AFTER de-duplication taking into account only reads with unique start sites
UNIQUE_START_SITES_*	The keys described above prefixed with UNIQUE_START_SITES have the same definition, but refer to the reads AFTER de-duplication taking into account only reads with unique molecular barcodes (bins)
TOTAL_UNIQUE_DNA_READS	Total number of unique DNA reads based on the molecular barcode (bins)
TOTAL_RAW_DNA_READS	Total number of DNA reads based before deduplication
TOTAL_UNIQUE_DNA_START_SITES	Total number of unique DNA reads based on the unique start sites
AVERAGE_UNIQUE_DNA_READS_PER_GSP2	Average number of unique DNA reads per GSP2 based on the molecular barcode (bins)
AVERAGE_DNA_READS_PER_GSP2	Average number of DNA reads per GSP2 based before deduplication
AVERAGE_UNIQUE_DNA_START_SITES_PER_GSP2	Average number of unique DNA reads per GSP2 based on the unique start sites
TOTAL_UNIQUE_RNA_READS	Total number of unique RNA reads based on the molecular barcode (bins)
TOTAL_RAW_RNA_READS	Total number of RNA reads based before deduplication
TOTAL_UNIQUE_RNA_START_SITES	Total number of unique RNA reads based on the unique start sites
AVERAGE_UNIQUE_RNA_READS_PER_GSP2	Average number of unique RNA reads per GSP2 based on the molecular barcode (bins)
AVERAGE_RNA_READS_PER_GSP2	Average number of RNA reads per GSP2 based before deduplication
AVERAGE_UNIQUE_RNA_START_SITES_PER_GSP2	Average number of unique RNA reads per GSP2 based on the unique start sites
TOTAL_UNIQUE_AMBIG_READS	Total number of unique ambiguous reads based on the molecular barcode (bins)

TOTAL_RAW_ AMBIG _READS	Total number of ambiguous reads based before deduplication
TOTAL_UNIQUE_ AMBIG _START_SITES	Total number of unique ambiguous reads based on the unique start sites
AVERAGE_UNIQUE_ AMBIG _READS_PER_GSP2	Average number of unique ambiguous reads per GSP2 based on the molecular barcode (bins)
AVERAGE_ AMBIG _READS_PER_GSP2	Average number of ambiguous reads per GSP2 based before deduplication
AVERAGE_UNIQUE_ NUCLEIC_ACID _START_SITES_PER_GSP2	Average number of unique ambiguous reads per GSP2 based on the unique start sites
TOTAL_UNIQUE_ NUCLEIC_ACID _READS	Total number of unique total nucleic acid (DNA = RNA + Ambig) reads based on the molecular barcode (bins)
TOTAL_RAW_ NUCLEIC_ACID _READS	Total number of total nucleic acid (DNA = RNA + Ambig)reads based before deduplication
TOTAL_UNIQUE_ NUCLEIC_ACID _START_SITES	Total number of unique total nucleic acid (DNA = RNA + Ambig) reads based on the unique start sites
AVERAGE_UNIQUE_ NUCLEIC_ACID _READS_PER_GSP2	Average number of unique total nucleic acid (DNA = RNA + Ambig)reads per GSP2 based on the molecular barcode (bins)
AVERAGE_ NUCLEIC_ACID _READS_PER_GSP2	Average number of total nucleic acid (DNA +RNA + Ambig) reads per GSP2 based before deduplication
AVERAGE_UNIQUE_ NUCLEIC_ACID _START_SITES_PER_GSP2	Average number of unique total nucleic acid (DNA +RNA + Ambig) reads per GSP2 based on the unique start sites
*_CONTROL	The 24 keys described above with the PREFIX _CONTROL have the same definition but are limited to the control targets only.
FRAGMENT_GSP2_n_NAME	The name of the target region for which the data is relevant. n is the target number, starting at 0
FRAGMENT_GSP2_n_MEAN_LENGTH	The average (deduced) length of the fragment (read pair) for reads for this specific target [n]
FRAGMENT_GSP2_n_MEDIAN_LENGTH	The median (deduced) length of the fragment (read pair) for reads for this specific target [n]
READS_PER_TARGET_n_RNA_READS	The total number of RNA reads for this specific target [n]
UNIQUE_*	The keys described above prefixed with UNIQUE have the same definition, but refer to the reads AFTER de-duplication (either alignment- or Molecular Barcoded-based). Most metrics are deduced from the UNIQUE metrics using the

	Molecular barcode bin sizes to calculate the non-de-duplicated metrics.
AVERAGE_UNIQUE_RNA_READS_PER_GSP2	The average number of reads classified as RNA reads (spanning two or more exons) that map to a GSP2.
COVERAGE_000_GENE	The coverage metrics for the bases in all targeted exons of the gene. "Summary" indicates the results are for all exons covered by GSP2's, combined
COVERAGE_000_MIN_COV	The minimum coverage
COVERAGE_000_MAX_COV	The maximum coverage
COVERAGE_000_MEAN_COV	The mean coverage
COVERAGE_000_MEDIAN_COV	The median coverage
COVERAGE_000_PERCENT_10X_OR_GREATER	The percent of bases in the exons that have at least 10X coverage
COVERAGE_000_PERCENT_100X_OR_GREATER	The percent of bases in the exons that have at least 100X coverage
COVERAGE_000_PERCENT_1000X_OR_GREATER	The percent of bases in the exons that have at least 1000X coverage
COVERAGE_000_PERCENT_BASES_GT_20_PRCT_MEAN	The percentage of bases that have coverage at least 20% of the mean coverage (COVERAGE_000_MEAN_COV)
COVERAGE_000_TOTAL_BASES	Total number of bases covered by the targeted exons
COVERAGE_000_MIN_FOLD_CHANGE_70%	The ratio between the 70th percentile and the 30th percentile coverage value. This ratio is an indication of the "evenness of coverage"
COVERAGE_000_MIN_FOLD_CHANGE_90%	The ratio between the 90th percentile and the 10th percentile coverage value. This ratio is an indication of the "evenness of coverage"
FC_n_GSP2	Name of the target region associated with this Fusion Candidate (FC). n represents the fusion candidate number (starting at 1)
FC_n_GENES	The two (or more) genes participating in this gene fusion.
FC_n_KNOWN_FUSION	Indicates if the gene fusion candidate is a KNOWN fusion or not. The Archer Quiver database is used as the source. Values can be TRUE or FALSE
FC_n_INTRON_EXON_FUSION	Indicates if the gene fusion candidate is a fusion between an exon and an intron, often a sign of a false positive finding. Values can be TRUE or FALSE

FC_2_MISPRIMING_BASED_OFF_TARGET	Indicates if the two fusion partners share significant sequence similarity which is often caused by the primer mispriming of the not originally targeted gene. Values can be TRUE or FALSE
FC_n_ANNOTATION_1	Annotation of the first fusion partner. Format: [GENE_NAME]([STRAND])|[exon|intron]:[0-9]*|[CHROM]:[STARTOFCONSENSUS],[CHROM]:[BREAKPOINT]
FC_n_ANNOTATION_2	Annotation of the second fusion partner. Format: [GENE_NAME]([STRAND])|[exon|intron]:[0-9]*|[CHROM]:[STARTOFBREAKPOINT],[CHROM]:[STARTOFSECONDBREAKPOINT]
FC_n_BARCODE_ID	The identifier used in the consensus FASTA file. This identifier links the Fusion Candidate to the "molbar" identifier, used in the FASTA consensus sequence and BED file. i.e If the FC_n_BARCODE_ID is 2, the identifier used in the consensus FASTA file is "2(GENE1:GENE2)_molbar_nn" where nn is some random number and (GENE1:GENE2) indicate the two gene fusion partners
FC_n_R1_COUNT	The number of filtered, non-redundant read 1 reads supporting this fusion candidate.
FC_n_R2_COUNT	The number of filtered, non-redundant read 2 reads supporting this fusion candidate.
FC_n_EITHER_R1_OR_R2	The number of filtered, non-redundant read 1 OR 2 reads supporting this fusion candidate.
FC_n_BOTH_R1_AND_R2	The number of filtered, non-redundant fragments (read-pairs) supporting this fusion candidate.
FC_1_UNIQUE_START_SITES	The number of reads with unique start sites that support the fusion candidate
FC_n_PROTEIN_TRANSLATION_x_GENE_BEFORE	The name of the fusion partner on the 5' side of the breakpoint (on RNA)
FC_n_PROTEIN_TRANSLATION_x_TRANSCRIPT_BEFORE	The (s) transcript for the 5' gene fusion partner (separated by a forward slash (/) if there is more than one) that could result in the sequence provided in the SEQUENCE field for this protein translation. If multiple transcripts are listed, it means that the AA sequence for these transcripts was the same.
FC_n_PROTEIN_TRANSLATION_x_GENE_AFTER	The name of the fusion partner at the 3' side of the break point (on RNA)
FC_n_PROTEIN_TRANSLATION_x_TRANSCRIPT_AFTER	The (s) transcript for the 3' gene fusion partner (separated by a forward slash (/) if there is more than one) that could result in the sequence provided in the SEQUENCE field for this protein translation. If

	multiple transcripts are listed, it means that the AA sequence for these transcripts was the same.
FC_n_PROTEIN_TRANSLATION_x_INFRAME	Indicates if the 3' fusion transcript is in frame with the 5' fusion partner (Yes or No)
FC_n_PROTEIN_TRANSLATION_x_SEQUENCE	The deduced AA sequence for this fusion partner/transcript combination. When out of frame and not stop codon is found before the last codon of the transcript, the last AA is the last AA of the last exon.
WT_n_GSP2	Name of the target region associated with this Wild Type (non-fusion) isoform (WT)
WT_n_NOVEL	Indicates the isoform is a NOVEL isoform or not. Novel isoforms are defined as those isoforms that have non-consecutive exon/intron numbering in the annotation, suggesting exon skipping events
MOLBAR_TOTAL_NUM_READS	The total number of read (pairs) in the FASTQ files (only provided when using the Molecular Barcode based de-duplication)
MOLBAR_READS_WITH_CORRECT_COMMON_REGION	The number of read (pairs) in the FASTQ file that contain the common-region following the random Molecular Barcode. Only those reads with a perfect match to the common region sequence pass this filter.

4.1.1 CNV Results file format

The Copy Number changes are not part of the full_results.txt summary file for a sample, but are provided as a separate comma separated results file

[SAMPLE1].molbar.trimmed.deduped.UNIQUE_READ_EITHER.counts.csv

The file consists of four columns, the description of each is provided in Table 3.

Table 3. Format of the CNV results file

Colum	Description
GSP2	The name of the target/primer
Gene	The name of the targeted gene followed by the normalized copy number of the gene. 1 represents "same as normal" which, for a gene on a diploid autosome, represents two copies
NormalizedLog2Value	The Log2 value of the normalized copy number change for the

	target/primer, compared to the “control: sample.
Zscore	The Z score for the normalized copy number change for the target/primer

5 Workflow Automation

This section describes the setup of the workflow automation for Archer Analysis. The workflow automation allows the automatic execution of a predefined workflow whenever a set of FASTQ files are moved to a special watch directory.

There are three steps for the setup of the workflow automation:

Set up a watched folder in the web interface

Set up workflow automation definition in the web interface

Develop a script or procedure to move FASTQ files to the watch directory

Each of these steps is described in the sections below.

5.1 Set up a watched folder in the web interface

To allow Archer Analysis to automatically execute a workflow when (links to) FASTQ files are placed in a special watch-folder, use the web interface to create a folder on the VM (or on a file server that is accessible to the VM).

Select the "**Settings -> Automation Settings**" menu to reach the workflow automation page.

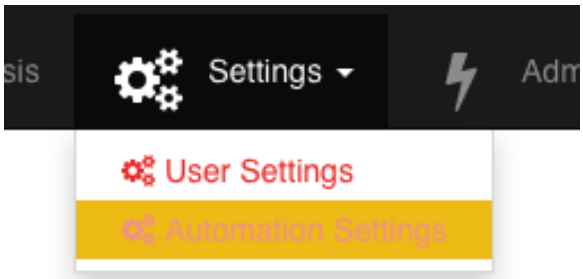


Figure 64. Automation settings found in Setting dropdown menu

This will also show you any watched folders that may already exist.

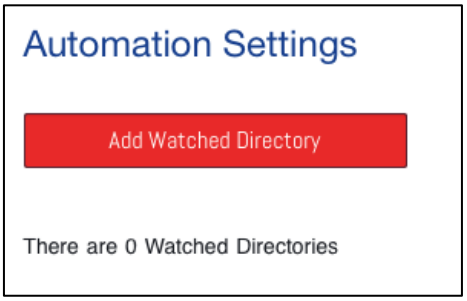


Figure 65. Automation worfklow setup shows there currently are no watched directories

To add a new watch folder click the "**Add Watched Directory**" button.

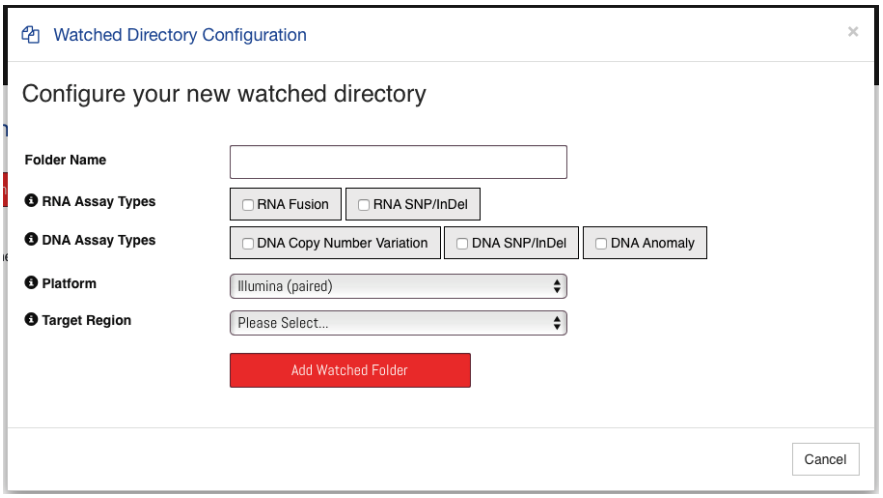


Figure 66. Workflow automation dialog box

The dialog box is similar to the regular dialog box used to start a new analysis, with the exception of the “Folder Name” text box.

Type in the desired name of the watched directory in the Folder Name box (do not use spaces or symbols other than hyphen or underscore).

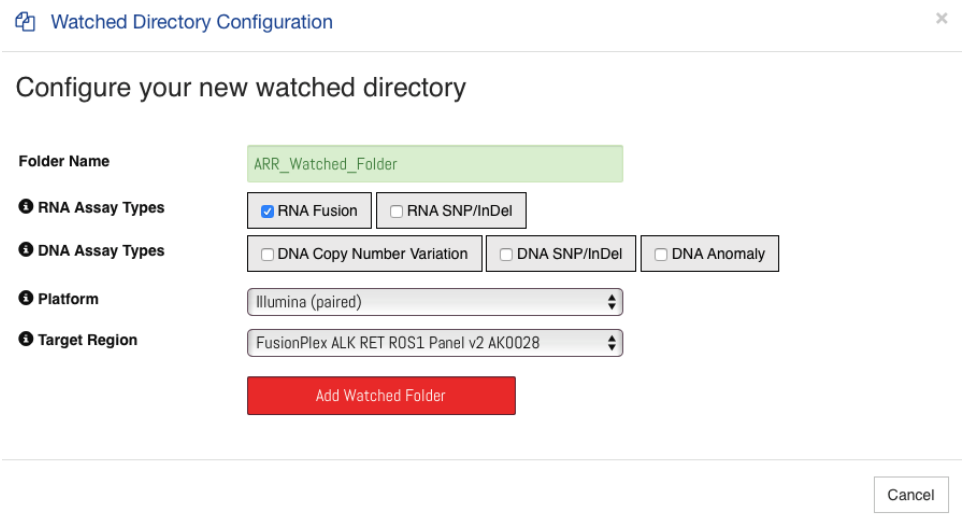


Figure 67. Fully configured workflow automation definition

Configure assay type, platform and target region for the watch folder in the same way as it is done for a manual analysis. Each watched folder has to be unique and can only perform one specific workflow.

Upon successful creation of the watched folder, you will now see it listed in the Watched Directories list.

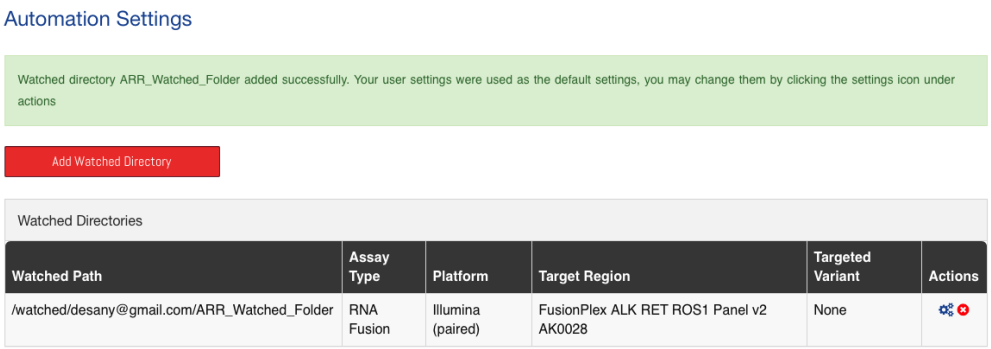


Figure 68. Automation Settings

5.2 Setup workflow automation definition in the web interface

5.2.1 Analysis settings for the executed workflows

Click on the Gear Icon in the Actions column to open the "Watched Folder Settings" Dialog.

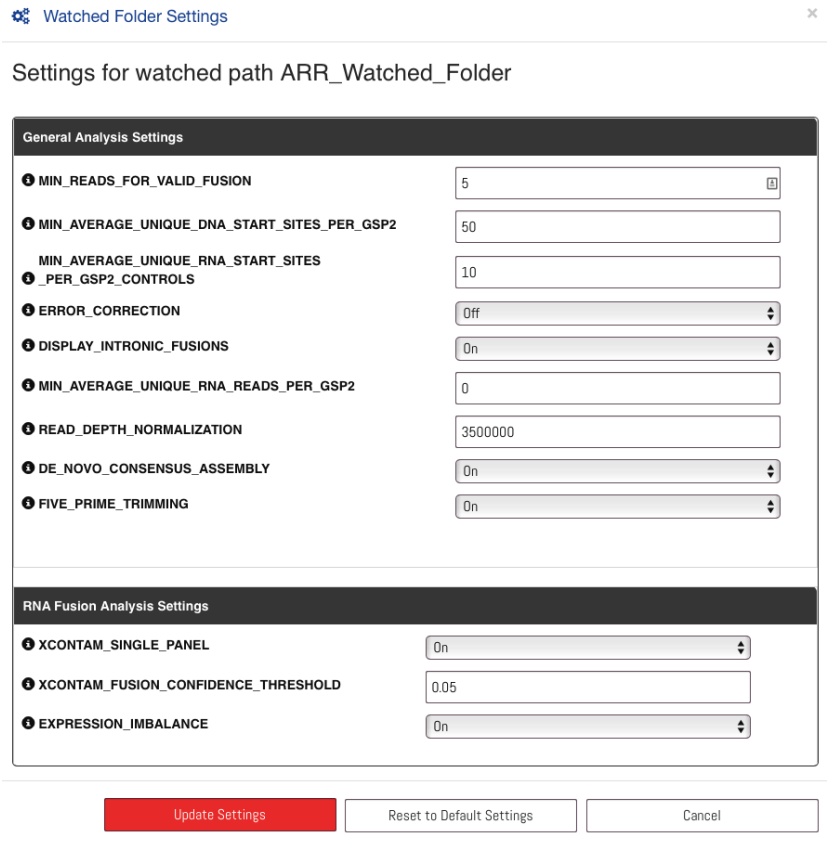


Figure 69. Watched Folder Settings dialog box

Make any desired changes to these setting and click "Update Settings."

The analysis settings for the executed workflows are applied at running time and NOT at the time that the workflow is added to the list of watched directories. Ensure that the analysis settings for the admin user are correct and as intended.

5.2.2 Removing a watched directory

To remove a watch directory for future consideration, click the X icon under Actions.

A warning message will appear to ensure this was the intended behavior.

Removing a watch directory will NOT interfere or stop any currently running job, but will only avoid this directory from being considered for any future workflow execution.

Removing a watch directory will NOT delete the folder on the server but will only remove it from future consideration.

5.2.3 Removing target region files that are used in watched directory workflows

When a target region is defined as the target region for one or more automated workflow definitions, it cannot be removed until the watch directory itself is removed. This ensures that workflows that rely on the target region can be executed correctly.

The “Existing Custom Target” list will show if a target file is used in one or more workflow definitions.

Figure 70. Existing custom target list shows “FusionPlex ARR Panel V1” is used by a workflow automation defini- tion and cannot be removed

To remove a target file, first remove all the analyses AND remove any watched directory definitions. After the removal, the delete icon will appear and the target region can be removed.

Figure 71. Watched folder using the “FusionPlex ARR Panel V1” has been removed and the target file can now be removed

5.2.4 Develop procedure/script for the movement of FASTQ files

Once the watch folder has been created and the system has been setup to automatically execute a predefined workflow, a procedure or script needs to be developed to move, copy or link the required FASTQ files to the watch folder.

If the “archer” group has been created to allow only the archer_daemon and apache user to have write permission for the watch folder, ensure that whatever automation script is executed and actually fills the watch folder with the FASTQ files has the correct permissions for the watch folder.

5.2.5 How the systems knows when to execute a workflow

The workflow automation system continuously monitors each of the define watch directory and looks for following files:

A file with the extension “[SAMPLE_NAME].completed”

A folder containing all the FASTQ files to be analyzed PLUS a file with the name “[FOLDER_NAME].completed” in the top level of the watch folder

In the first case, it will create a job in the Archer™ Analysis system and will run the workflow on the (pair of) FASTQ file(s) with the name [SAMPLE_NAME]*.fastq or [SAMPLE_NAME]*.fastq.gz (for the uncompressed and compressed version, respectively. In the second case it will create a job in the analysis system for ALL FASTQ files in the folder. The latter case will allow samples that are related (such as in CNV type of analysis, where there are case/control cases) to be run together.

Note the use of the ‘*’ (asterisk) wildcard symbol. In the case of Illumina paired-end sequence data, the names of the two FASTQ files are typically something like:

[SAMPLE_NAME]_R1_001.fastq

SAMPLE_NAME]_R2_001.fastq

The suffixes “_R1_001” and “_R2_001” indicate the first and second read in the read pair data and the use of the ‘*’ character ensures that BOTH files are picked up for the workflow automation. For single read technologies such as Life Tech- nologies’ Ion Torrent PGM system or the Illumina single read libraries, create a file with the structure “[SAMPLE_NAME].completed”.

5.2.6 Example for job containing a single sample

Here is an example of how to create a single sample for automatic execution. Content of the watch folder:

/var/www/analysis/watch_folder_ARR

|

+ - BC-112_NA1473-FFPE_S11_L001_R1_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001_R2_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001.completed

Placing these three files in the watch directory will result in a job with the name “**BC-112_NA1473-FFPE_S11_L001**”

NOTE: always ensure that the FASTQ files are moved/copied to the watch folder BEFORE the “.completed” file is created. If the FASTQ files are NOT present, the job will produce an error.

5.2.7 Example for job containing multiple samples Contents of the watch folder

/var/www/analysis/watch_folder_ARR

|

+ - 4_Samples.completed

+ - 4_Samples

|

+ - BC-112_NA1473-FFPE_S11_L001_R1_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001_R2_001.fastq

+ - BC-113_TriplePOS_S12_L001_R1_001.fastq

+ - BC-113_TriplePOS_S12_L001_R2_001.fastq

+ - BC-114_AmbionNEG_S13_L001_R1_001.fastq

+ - BC-114_AmbionNEG_S13_L001_R2_001.fastq

+ - BC-115_Water_S14_L001_R1_001.fastq

+ - BC-115_Water_S14_L001_R2_001.fastq

This will result in a job with the name "**4_Samples**" containing all 4 samples being analyzed at the same time. Note that the file "4_samples.completed" is located in the top level of the watch folder AND is created AFTER all the FASTQ files have been moved/copied/linked, to avoid errors in the job.

5.2.8 Fate of the files in the watch directory

Files (or links to files) that are placed in the watch directory will be removed from the watch directory and placed in the special directory "picked_up_files" (in the top level of the watch directory) where they will remain for the rest of their lives. The automation engine will create a symbolic link from this directory to the analysis directory (typically in "/var/www/analysis/ [JOB_NUMBER]").

This is the structure of the watch directory BEFORE the files are picked up for analysis:

/var/www/analysis/watch_folder_ARR

|

+ - BC-112_NA1473-FFPE_S11_L001_R1_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001_R2_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001.completed

This is the structure of the watch directory AFTER the files are picked up for analysis:

/var/www/analysis/watch_folder_ARR

|

+ - picked_up_files

|

+ - BC-112_NA1473-FFPE_S11_L001_R1_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001_R2_001.fastq

and this is the structure of the analysis folder:

/var/www/analysis/[JOB_NUMBER]

|

|

+ - BC-112_NA1473-FFPE_S11_L001_R1_001.fastq ->

/var/www/analysis/watch_folder_ARR/picked_up_files/BC-112_NA1473-FFPE_S11_L001_R1_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001_R2_001.fastq ->

/var/www/analysis/watch_folder_ARR/picked_up_files/BC-112_NA1473-FFPE_S11_L001_R1_001.fastq

The analysis folder/[JOB_NUMBER] contains symbolic links to the FASTQ files in the "picked_up_files" directory of the watch folder. It is imperative that the files in the "picked_ up_files" folder are NOT removed, since this will prevent the jobs from being re-run or cloned.

5.2.9 File collisions

If the FASTQ files for a sample that had been previously been analyzed by the workflow automation engine are placed in the watch directory again, the system will avoid the files from being overwritten in the "**picked_up_files**" directory by pre-pending the new files with the date and time and placing the new files in a special sub-directory of "**picked_up_files**" called "**collision_files**".

Here's an example of the situation BEFORE the sample BC-112 is run again:

/var/www/analysis/watch_folder_ARR

|

+ - BC-112_NA1473-FFPE_S11_L001_R1_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001_R2_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001.completed

|

+ - picked_up_files

|

+ - BC-112_NA1473-FFPE_S11_L001_R1_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001_R2_001.fastq

This is the situation after the files have been picked up again to be analyzed:

/var/www/analysis/watch_folder_ARR

|

+ - picked_up_files

|

+ - BC-112_NA1473-FFPE_S11_L001_R1_001.fastq

+ - BC-112_NA1473-FFPE_S11_L001_R2_001.fastq

+ - collision_files

|

+ - 01_06_2015_11_21_00_BC-112_NA1473-FFPE_S11_L001_

R1_001.fastq

+ - 01_06_2015_11_21_00_BC-112_NA1473-FFPE_S11_L001_

R2_001.fastq

The job will still be executed and the job will have the same name as the sample, but the FASTQ file name will be changed. The analysis/[JOB_NAME] folder will contain a link to the file in the **collision_files** directory.

6 Troubleshooting

This section contains information on dealing with problems that may arise during the execution of the Archer Analysis Software (for problems that arise during installation, please see the Installation Guide).

6.1 Job finished with job status "COMPLETED_ERROR"

This status indicates that one or more of the samples showed some sort of error during processing. The Summary Page may show one or more jobs with the message "Sample processed with errors".

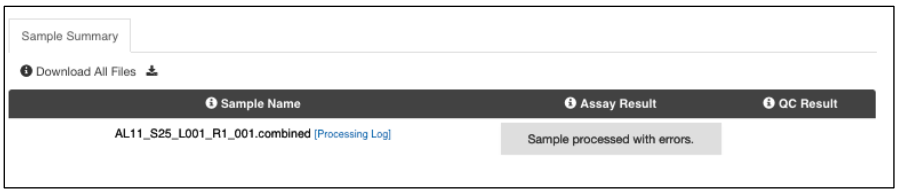


Figure 72. One of the jobs finished with errors

If only one or a few of the jobs finished with errors, it is likely that the error happened during the processing of the sample. Click the "Processing log" link to see the log file to determine the location of the error. The easiest way to find the error is to search for the text "[ERROR]" in a text editor.

A common reason for failure of an individual job is running out of disk space:

```
[20:32:24] waterfall.sh [INFO] > Checking for required files...
[20:32:24] waterfall.sh [ERROR] > Can't find required file /var/www/analysis/1468/NA13-865_S10_L001_R2_001.molbar.trimmed.deduped.fastq
[20:32:24] waterfall.sh [ERROR] > Aborting.
/var/www/archer/analysis/shared_scripts/waterfall.sh: line 25: cannot create temp file for here-document: No space left on device
[20:32:24] run_waterfall_metrics_workflow [ERROR] > Non-zero exit code (1) from workflow, aborting.
```

Figure 73. Job failure due to lack of disk space

This may occur if the /var/www/analysis directory on the virtual machine has reached its defined space. Deleting analyses that are no longer needed will free up some space. Alternatively, provision the virtual machine with more disk space. See the manual for the virtualization software for more information.

If the error is not pointing to an obvious solution, send the log file to tech@archerdx.com for support.

6.2 Job is stuck with job status "NEW"

This indicates that the queue manager software ("poller") is not running. The easiest way to restart the poller is to restart the complete virtual machine.

Alternatively, log in to the virtual machine (See section **Error! Reference source not found. Error! Reference source not found.** for the credentials) and execute the following command in the **root home directory**:

$ ./restart-poller.sh

This should restart the poller and the job that was stuck in "NEW" should start running. To verify the poller is running, enter the following command:

$ ps –ef | grep poller

This should show a running python2.7 process called "poller.py"

```
[root@analysis ~]# ps -ef | grep poller
500       2164     1  0 Aug04 ?        00:00:46 python2.7 /var/www/html/archer_web/analysis/daemon/poller.py start
root     11199 11065  0 12:04 tty1     00:00:00 grep poller
[root@analysis ~]# _
```

If the poller is still not running, check the poller log file (stderr.poller) for more information.

If the VM was restarted it is likely that the job will show job status "COMPLETED_ERROR". Re-run the job by clicking the rerun icon ( ) to restart the job with the same settings.

6.3 Job is stuck with job status "HOLD"

This status shows that the job is still being processed although the processing step could have crashed. A job will show "HOLD" when it is processing the FASTQ files after upload.

If the job continues to show "HOLD" (for large files this could take up to an hour), restart the VM or restart the poller as shown in section 6.2 'Job is stuck with job status "NEW"'.

7 Managing Users And Groups

This section describes the user and group management functionality of Archer Analysis 4.1. Key concepts are that users are organized into groups, and that software permissions can be applied on a per-user and a per-group basis. Users can be a member of multiple groups, and permissions are additive for the user and any groups it is a member of.

7.1 The admin user

Archer Analysis 4.1 comes configured with a default user called "admin" (password = password123), that is a member of the "Basic Users" and "Experimental" groups. The admin user has a special permission called "Is System Admin", which is required in order to perform any of the user, group, and permissions management described in this section. This functionality is accessed via the "Admin" menu, by selecting either "Users" or "Groups".

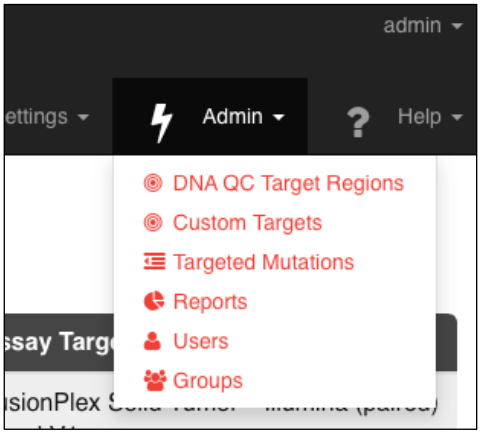


Figure 74. Admin Menu that can be used to set User and Group permissions

7.2 Users

Selecting the "**Admin->Users**" menu item takes you to a table containing a list of all the users in the system. The columns indicate the active status of the users ("Active"), whether they have the special "Is System Admin" status, the "Groups" they belong to, how many "Ownerships" they have of the Analyses and Watched Folders present on the system.

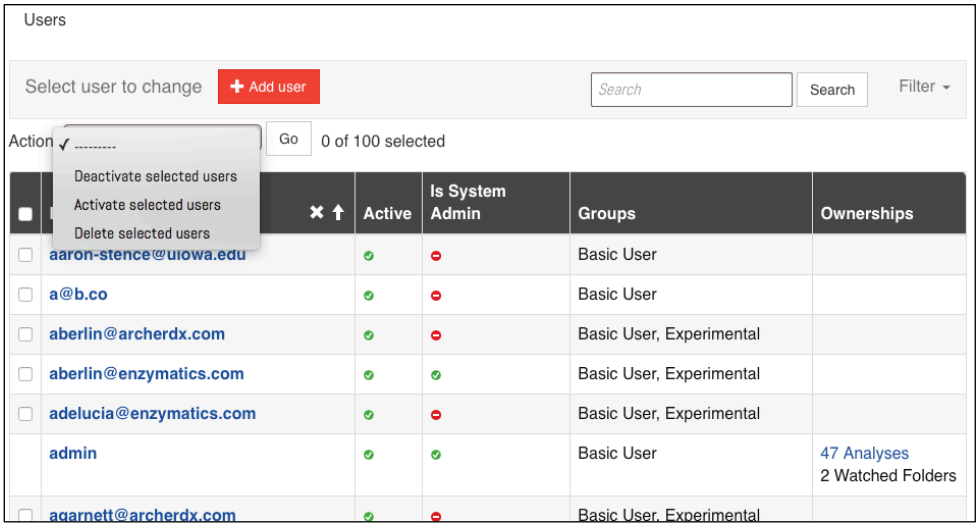


Figure 75. Users Table accessed via the Admin Menu

Multiple users can be selected by using the checkboxes in the left-most column, at which point they can be activated, deactivated, or deleted all at once.

You can search for users by name by using the "**Search**" box.

Users can be filtered by group membership by using the "**Filter**" drop-down.

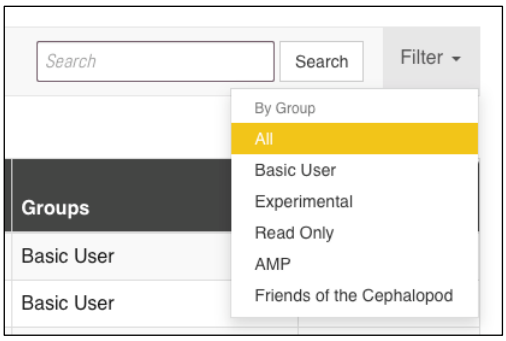


Figure 76. Filter drop-down on the Users Table page.

7.2.1 Adding users

Click on the "**Add User**" button to go to the "**Add User**" screen.

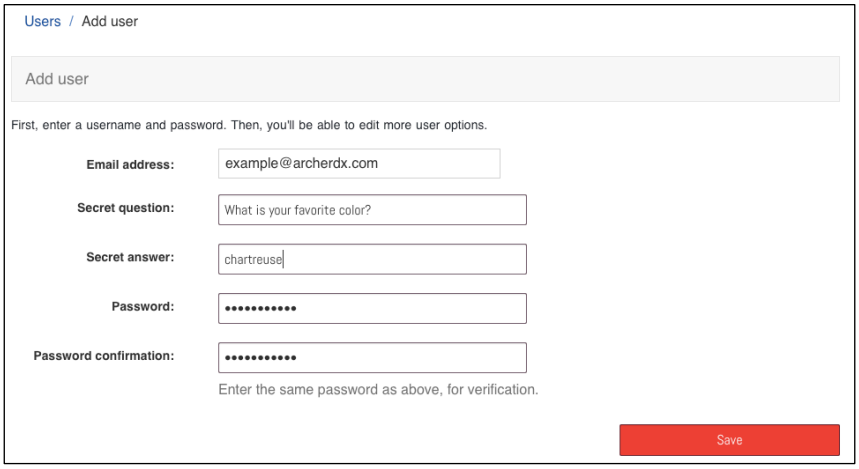


Figure 77. Add User screen

Provide the required fields and click the "**Save**" button.

This takes you to the "**Edit User**" screen for the new user, where you can edit the user.

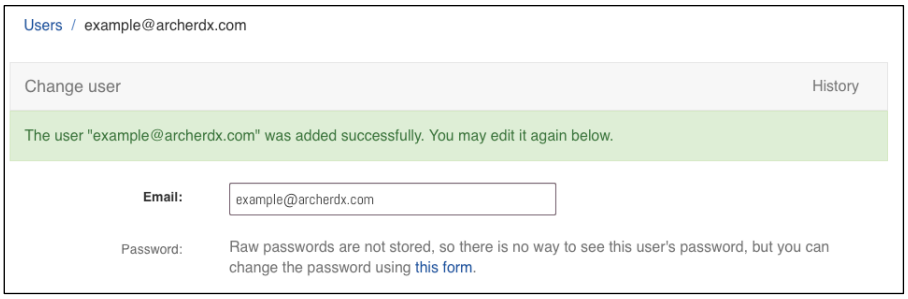
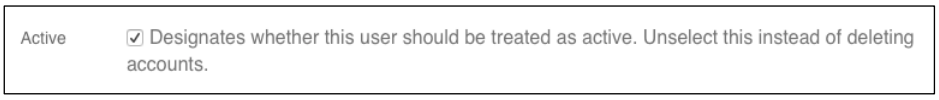


Figure 78. Edit User screen

7.2.2 Editing users

The "**Edit User**" screen can be reached by either adding a new user, or clicking on the user name in the "**Users**" screen.

Users can be deactivated by unchecking the "**Active**" checkbox.



Users can be granted the ability to manage users and groups by checking the "**Is System Admin**" checkbox.

Group membership is controlled by choosing Available Groups from the Group Selection List and moving them to the Chosen Groups list.

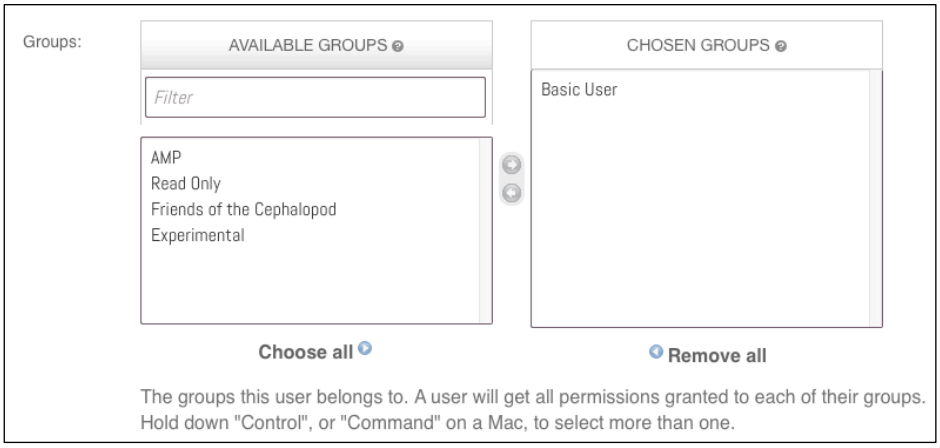


Figure 79. Group Membership screen

Permissions inherited by the Chosen Groups will then be shown:

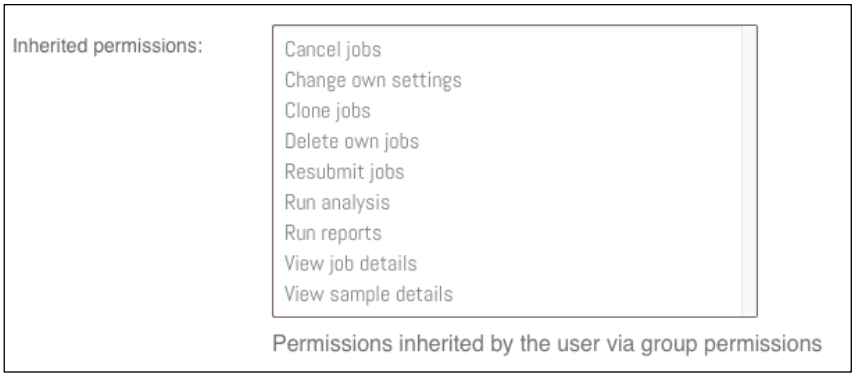


Figure 80. Inherited Permissions list

Additional permissions can be granted by choosing from the Available Permissions in the Permissions Selection List and moving them to the Chosen Permissions list.

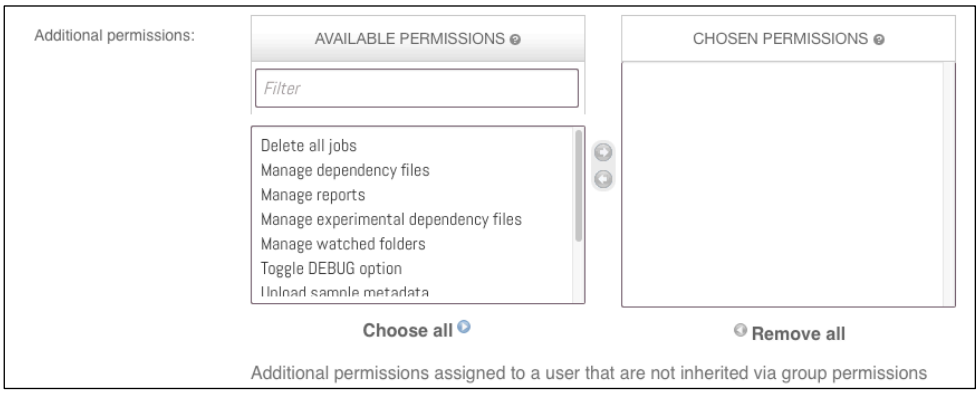


Figure 81. Additional Permissions selection screen

When you are finished editing the user, click on the "**Save**" button, located under the "Important Dates" section.

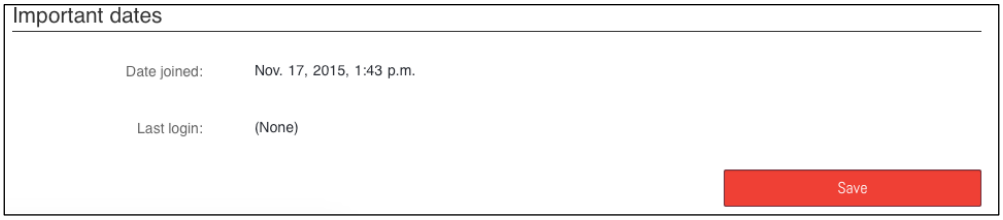


Figure 82. Final step in editing a user

7.3 Groups

Groups serve two purposes. One is to conveniently manage groups of permissions, so that large numbers of individual permissions do not have to be manually assigned to individual users. The other is to allow users belonging to a given group to have visibility on the analyses performed by other group members, a functionality called Job Sharing.

Selecting the "**Admin->Groups**" menu item takes you to a table containing a list of all the groups in the system. The columns indicate the name of the Group and whether Job Sharing is enabled for the group.

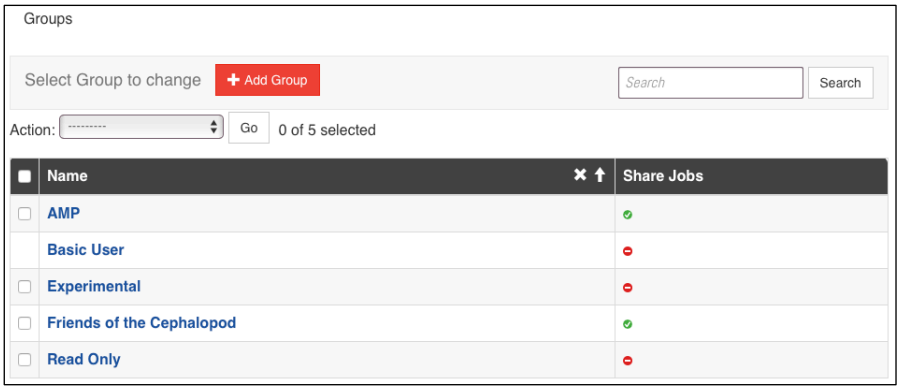


Figure 83. Groups List screen, accessed via the Admin Menu

Multiple groups can be selected by using the checkboxes in the left-most column, at which point the selected groups can be deleted all at once.

You can search for groups by name by using the "**Search**" box.

7.3.1 Adding groups

Click on the "**Add Group**" button to go to the "**Add Group**" screen.

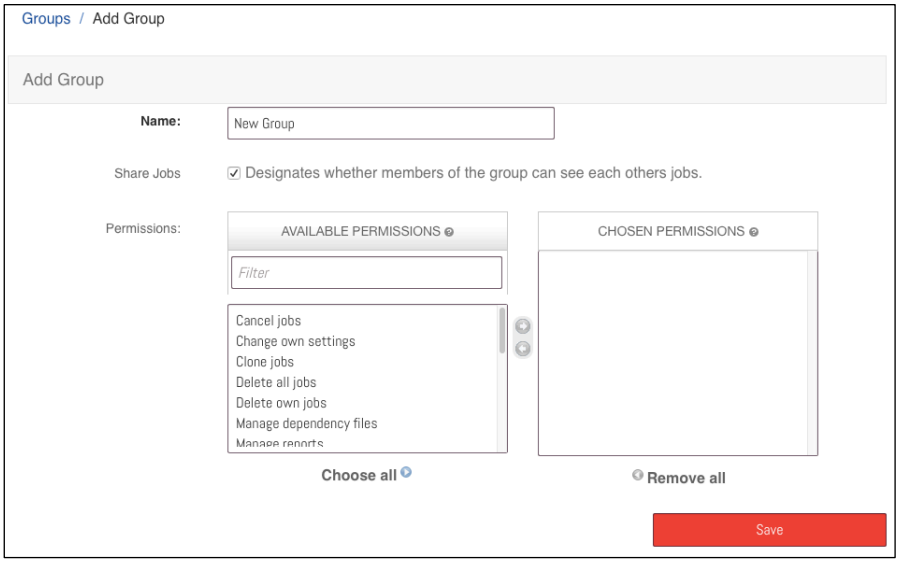


Figure 84. Add Group screen

Type the name of the new group in the "**Name**" field, and determine whether Job Sharing is enabled by checking or unchecking the "**Share Jobs**" checkbox.

Permissions of the group can be granted by choosing from the Available Permissions in the Permissions Selection List and moving them to the Chosen Permissions list.

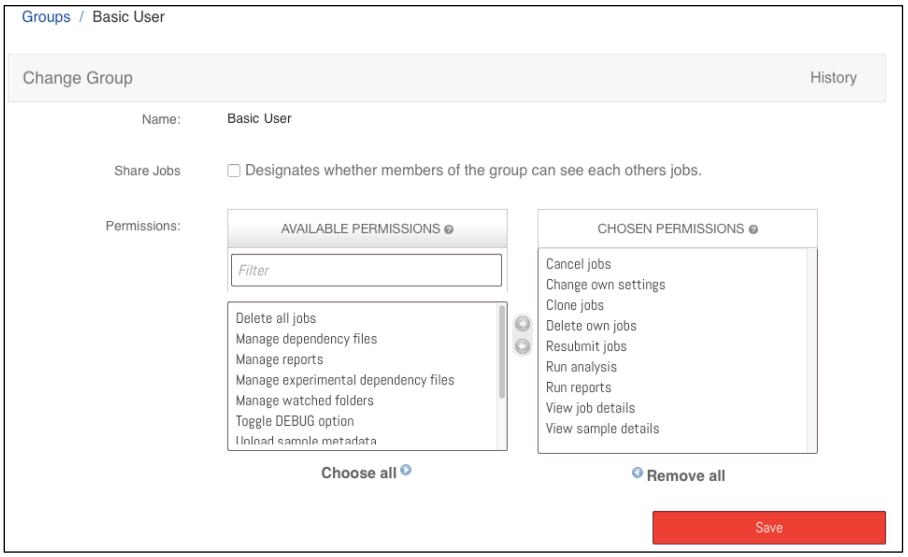


Figure 85. Group Permissions selection screen

A new group is being created to enable a group of collaborating users to see each other's analyses, then adding additional permissions will not be necessary. All new users are members of the "Basic User" group by default, which has all the permissions necessary to run analyses.

When you have finished configuring the new group, click on the "**Save**" button to save it. You can then assign membership of this group to individual users, by editing the users (section 7.2.2).

7.3.2 Editing groups

The "**Edit Group**" screen can be reached by either adding a new group (as described above), or by clicking on the group name in the "**Groups**" screen.

8 Appendix

This section contains a description of the various metrics and results statistics for the results pages.

8.1 Settings Menu – Analysis Settings

Table 4. Definition of the analysis settings

Key	Description
Analysis and QC filter settings	
MIN_READS_FOR_VALID_FUSION	Minimum number of breakpoint-spanning reads required to support a gene fusion (Default = 5)
MIN_AVERAGE_UNIQUE_DNA_START_SITES_PER_GSP2	Minimum number of DNA reads with a unique start site, required per GSP2. If less than this number is found.
MIN_AVERAGE_UNIQUE_RNA_START_SITES_PER_GSP2_CONTROLS	Minimum number of RNA reads with a unique start site, required to fall on one or more of the CONTROL targets. If less than this number fall on a control target, the QC STATUS will indicate "NOT ENOUGH RNA READS" (Default = 10 for RNA assays only)
MIN_UNIQUE_START_SITES_FOR_VALID_FUSION	Minimum number of unique start sites required for a valid fusion. If less than this number of reads have a unique start site, the fusion is marked with the icon and, unless it is a known fusion, will place the fusion in the WEAK EVIDENCE bin
ERROR_CORRECTION	Variable indicating whether or not the Molecular Barcode-based error correction should be applied. NOTE: Experimental – May greatly increase the running time. Default = 0 (Set to 1 to turn on).
DISPLAY_INTRONIC_FUSIONS	Variable indicating if INTRON-EXON fusions should be reported. Intron-exon fusion are often, but not always false positives and this options allows the suppression of those fusions.
FUSION_PERCENT_OF_GSP2_READS	The minimum percent of reads supporting the fusion (as opposed to the wild type transcript) to qualify for strong evidence fusion. If less that this percent of reads support the fusion, the fusion is marked by the icon and, unless it is a known fusion, will place the fusion in the

	WEAK EVIDENCE bin
Variant analysis settings (Freebayes)	
MIN_DEPTH_FOR_VARIANT_CALL	Minimum number of reads supporting the alternative allele for variant detection to be considered at this position (Default = 10)
MIN_BASEQUAL_FOR_VARIANT_CALL	Exclude alleles from SNPs/InDel analysis if their supporting base quality is less than this value
MIN_ALLELE_FRACTION_FOR_VARIANT_CALL	Require at least this fraction of observations supporting an alternate allele within a single individual in order to evaluate the position in SNPs/InDel analysis. Default = 2% (0.02). NOTE: The default setting for Freebayes is 20% (0.2), Archer Analysis is set to 2% to allow for the detection of low allele fraction mutations.
MIN_PHRED_QUAL_SCORE_FOR_VARIANT_CALL	The minimum QUAL score for a variant to pass this filter. Default = 1
CNV Summary Settings	
CNV_STRONG_AMPLIFICATION_THRESHOLD	Copy gains must be above this fold change threshold to be categorized as strong evidence CNV copy gains NOTE: This setting immediately affects the the CNV reporting and the job does NOT have to be re-run for these changes to take effect
CNV_STRONG_DELETION_THRESHOLD	Copy loss must be below this fold change threshold to be categorized as strong evidence CNV copy loss. NOTE: This setting immediately affects the the CNV reporting and the job does NOT have to be re-run for these changes to take effect
CNV_P_VALUE_THRESHOLD	Copy gains and losses must have a P value below this value to be categorized as strong evidence CNV copy gain or loss. NOTE: This setting immediately affects the the CNV reporting and the job does NOT have to be re-run for these changes to take effect
Alignment Settings	

CONSENSUS_BLAST_EXPECT_THRESH	Statistical significance threshold for BLASTN used for gene fusion annotations. This number of matches are expected to be found by chance
CONSENSUS_BLAST_WORD_SIZE	The initial alignment exact match word size for BLASTN
MIN_ALIGNMENT_SCORE	Minimum alignment score for a read to be considered valid. Reads with an alignment score below this number are removed from consideration (but will be retained in the BAM files).
MIN_ALIGNMENT_LENGTH	Any reads with an alignment length less than this number will be removed. 0 will allow any length alignment to pass
UI Settings	
JOBS_PER_PAGE	The number of jobs that will be displayed on the Home screen or the Past Results screen

8.2 Detailed Summary - Strong and Weak evidence tabs

Table 5. Definition of the strong and weak fusion candidate metrics

Metric	Description
Reads (#/%)	The number and percentage of unique reads supporting this gene fusion based on the molecular barcode. Only reads spanning the breakpoint are considered to support the fusion. Paired reads where both reads completely cover only one of the genes are NOT considered as supporting a gene fusion. The percentage is calculated in reference to the total number of reads covering this target, including wild-type transcripts. The read must extend 5 bp past GSP2 in order for the fusion to be called on the GSP2 side **&** this read must read 30 bp into the fusion partner for this to be called. Also both sides of the fusion must have an alignment score greater or equal to the default cutoff of 30
Start Sites (#/%)	The number and percentage of unique reads supporting this gene fusion based on the unique start sites. Only reads spanning the breakpoint are considered to support the fusion. Paired reads where both reads completely cover only one of the genes are NOT considered as supporting a gene fusion. The percentage is calculated in reference to the total number of reads covering this target, including wild-type

	transcripts. Need to read 5 bp past GSP2 in order for the fusion to be called on the GSP2 side **&** this read must read 30 bp into the fusion partner for this to be called. Also both side of the fusion must have an alignment score greater or equal to the default cutoff of 30
Annotation 1-Gene:Transcript	The gene name, transcript and exon/intron number for the fusion partner. If there is more than one possible transcript that contains the participating exon, the dropdown list shows all possible transcripts and exon/intron number
Annotation 2-Gene:Transcript	The gene name, transcript and exon/intron number for the fusion partner. If there is more than one possible transcript that contains the participating exon, the dropdown list shows all possible transcripts and exon/intron number
Breakpoint 1	The chromosomal breakpoint location for Annotation 1 as deduced from the RNA. This does NOT represent the exact breakpoint on the DNA level
Breakpoint 2	The chromosomal breakpoint location for Annotation 2 as deduced from the RNA. This does NOT represent the exact breakpoint on the DNA level

8.3 Detailed Summary – Read Statistics tab

8.3.1 Molecular Barcode Statistics

Table 6. Definition of the Molecular Barcode QC statistics

Metric	Description
Total Fragments	Total number of read (pairs) that were present in the original FASTQ file.
Fragments with Complete Adapter	Total number of reads that contained the common region and 8mer molecular barcode. This also includes reads that may later be removed because they were too short.
Number of Reads After Trimming Adapter	Total number of reads (pairs) that are greater than 35bp after trimming the adapters, the common region and 8mer barcode.

8.3.2 Read Statistics

Table 7. Definition of the read statistics QC metrics

Metric	Description

Type	Indicates the type of fragment (read pairs) the QC metrics are reported for. All Fragments indicates the non-deduplicated raw read(s) or read pairs. Unique fragments indicate the de-duplicated reads, based on either the alignment or the Molecular Bar Code (depends on the selected option in the analysis)
Total Fragments (#)	Total number of fragments (read pairs) that pass the initial quality filter
Mapped (#/%)	Total number and percentage of fragments (read pairs) that map to the genome. Percentage is compared to the total number of fragments that pass the initial quality filter. [NUMBER OF MAPPED READS/NUMBER OF TOTAL READS x 100= % MAPPED]
Passed Alignment File	Percentage of fragments (read pairs) that pass the Alignment Score filter compared to the total number of fragments that map. Alignment Score setting used can be found in the Analysis Settings page. **Default Setting = 30**
On Target (%)	Percentage of fragments (read-pairs) that overlap with at least one base pair with the target area (GSP2), compared to the total number of fragments that pass the mapping filter. Low % on target can be caused by a bad .gtf, promiscuous primers or ribosomal RNA.

8.3.3 DNA/RNA Statistics

Table 8. Definition of the DNA/RNA statistics QC metrics

Metric	Description
Type	Indicates the type of fragment (read pairs) the QC metrics are reported for. All Fragments indicates the non-deduplicated raw read(s) or read pairs. Unique fragments indicate the de-duplicated reads.
Molecular Bins	Molecular bins are defined as reads having the same random molecular barcode (8-mer) in the ligated adapter.
Average Molecular Bins per GSP2	The total number of molecular bins divided by the total number of target GSP2 (Gene Specific Primer 2). This metric is used to determine if a library passed or failed QC, but this metric is replaced by the Average Unique start sites per GSP2 for the control genes, which has been shown to be a

	more accurate predictor of sample quality
Unique Start Sites	Unique start sites are defined as read 1 having a unique start site.
Average Unique Start Sites per GSP2	The total number of reads with a unique start site divided by the total number of target GSP2 (Gene Specific Primer 2).
DNA Reads (#/%)	Total reads that likely come from a DNA source; reads that at least partially map to introns. Reads will be put into this category if they: 1. Do not have a break in alignment to hg19 of more than 100 bp (aka a split) 2. Include an intron region that must be at least 10% of the read length
RNA Reads (#/%)	Total reads that likely come from an RNA source (reads that span exon-exon splice junctions). Reads will be put into this category if they contain a split (a greater than 100 bp gap in alignment to hg19,which will occur when introns are spliced out)
Ambiguous Reads (#/%)	Total reads that map completely within exons: reads that are ambiguous as to their source. Reads will be put into this category if they do not have enough information to be placed in either of the other categories. (For example if a read does not contain a split, but also does not contain an intron region that is 10% of the length of the read.)

8.3.4 DNA/RNA Fragment Lengths

Table 9. Definition of the DNA/RNA fragment length QC metrics

Metric	Description
Mean Length (bp)	Apparent average fragment length as calculated by the mean of the total number of mapped reads. This will cap the apparent fragment length to 2 x [READ LENGTH] and should be considered an underestimate of the actual fragment length
Mean Length (bp)	Median fragment length as calculated by the median of the total number of mapped reads. This is a better estimate of the actual fragment length since the fragment length is capped at 2X [READ LENGTH] and the median is less sensitive to this capping

8.4 Detailed Summary – Assay Targets tab

The Assay Targets tab is itself divided into seven different sections.

Each sub-tab is divided by Controls and Assay Targets section. Controls are usually only defined for RNA gene fusion assays and represent the 8 standard targets Archer uses as targets that are typically moderately expressed in most tissues.

8.4.1 Unique Molecular Bins sub-tab

Molecular bins are defined as reads containing the same unique molecular barcode (8-mer). The data is separated by Total, RNA, DNA and AMBIGUOUS reads. See section 3.2.10 on page 30 for a definition of the various read types.

Table 10. Definition of the assay targets QC metrics

Metric	Description
Target	The name of the target
Fragments (#/%)	The number and percentage of unique (de-duped) reads that meet the pre-defined defintions of RNA, DNA, and Ambigious for each target (or GSP2).
RNA/DNA/Ambiguous Fragments (#/%)	The number and percentage of unique (de-duped) reads that meet the pre-defined definitions of RNA, DNA, and Ambiguous fragments for each target (or GSP2)

8.4.2 Unique Start Sites sub-tab

Unique start sites are defined as reads that have a unique start site. The data is separated by Total, RNA, DNA and AMBIGUOUS reads. See section 3.2.10 on page 30 for a definition of the various read types. See Appendix Table 10. Definition of the assay targets QC metrics or a description of the various fields.

8.4.3 Raw Alignment sub-tab

Raw alignments represent the coverage data BEFORE de-duplication. The data is separated by Total, RNA, DNA and AMBIGUOUS reads. See section 3.2.10 on page 30 for a definition of the various read types. See Appendix Table 10. Definition of the assay targets QC metrics or a description of the various fields.

8.4.4 DNA sub-tab

The DNA sub-tab represents the data for DNA reads only and further separated by DNA reads that have a unique molecular bin and those DNA reads that have unique start sites. Raw Alignments represent the coverage data BEFORE de-duplication. See section 3.2.10 on page 30 for a definition of the DNA read types.

8.4.5 RNA sub-tab

The RNA sub-tab represents the data for RNA reads only and further separated by RNA reads that have a unique molecular bin and those RNA reads that have unique start sites. Raw Alignments represent the coverage data BEFORE de-duplication. See section 3.2.10 on page 30 for a definition of the RNA read types.

8.4.6 Ambiguous sub-tab

The Ambiguous sub-tab represents the data for ambiguous reads only and further separated by ambiguous reads that have a unique molecular bin and those ambiguous reads that have unique start sites. Raw Alignments represent the coverage data BEFORE de-duplication. See section 3.2.10 on page 30 for a definition of the ambiguous read types.

8.4.7 Total sub-tab

The Total sub-tab represents the data for all reads combined (RNA, DNA and Ambiguous reads) and further separated by reads that have a unique molecular bin and those reads that have unique start sites. Raw Alignments represent the coverage data BEFORE de-duplication. See section 3.2.10 on page 30 for a definition of the various read types.

8.4.8 Details Summary – Variant Summary tab

Metric	Description
Variant Name	The variant name consists of the gene name and the amino acid mutation. This information is constructed from a set of special INFO fields in the provided mutations VCF file (Archer_Gene and Archer_MutationAA)
Symbol	The Gene symbol for the gene located at this position (Empty if variant is found in the intergenic region)
HGVSp	The mutation at the Protein level in the format from the Human Genome Variant Society
Variant Call	The variant call for this location. 0/0 represents a homozygous reference call. 0/1 and 1/1 represent heterozygous and homozygous alternative allele calls, respectively. For somatic mutations there can be 4 fields (i.e 0/1/1/1) since for somatic mutations a ploidy of 4 is assumed)
Type	Type of variant detected. Can be Single Nucleotide Polymorphism (SNP), Insertion or Deletion (InDel) or Complex, if more than one position is different
Genomic Location	Start position of the variant call (1- based, closed notation)
Ref/Alt Allele	The two (or more) alleles at this position, separated by the slash (/) symbol. The first allele is the reference allele, subsequent alleles are the alternative alleles. More than one alternative allele is possible
Quality Score	The PHRED based quality score of the variant

	call.
Clinical Significance	The clinical relevance of the variant according to the CLINVAR database
Variant disease name	The disease name this variant could have clinical relevance for according to the CLINVAR database
Alelle Fraction	The percent of the reads supporting the alternative allele
Depth	The total sequence depth at this position
Alt reads +	The total number of reads supporting a non-reference allele on the + strand. NOTE: The non-reference allele includes ALL non-reference alleles, not only the reported minor non-reference allele
Ref reads +	The total number of reads supporting the reference allele on the + strand
Alt reads -	The total number of reads supporting a non-reference allele on the - strand. NOTE: The non-reference allele includes ALL non-reference alleles, not only the reported minor non-reference allele
Ref reads -	The total number of reads supporting the reference allele on the - strand
Fisher Test Ratio	The fisher exact test ratio of the Alt +, Alt -, Ref + and Ref - measurements
Fisher P value	The P value for the fisher exact test of the Alt +, Alt -, Ref + and Ref – measurements. A value closer to 0 indicates significant strand bias, that is, most of the reads that support the alternate allele are found on only ONE of the strands, indicating possible strand bias and likely a false positive call.
Canonical	Indicates “YES” if this transcript is considered the canonical transcripts. The canonical transcript is defined as either the longest CDS, if the gene has translated transcripts, or the longest cDNA
Exon	The exon that contains the variant (exon/total_exons)
Codons	The three letter sequence for the codon the variant is found in. The variant base is shown as a capital letter
Consequence	The calculated consequence of the variation. See the ENSEMBL VEP page for more information

	about the various classes of consequence
Existing Variations	Identifier of any existing variations at this positions in dbSNP (rs…) or Cosmic (COSM…)
HGVSc	The mutation at the RNA coding sequence level in the format from the Human Genome Variant Society
SIFT	SIFT predicts whether an amino acid substitution affects protein function. The amino acid substitution is predicted to be damaging if the score is ≤0.05 and tolerated if the score is > 0.05 (note that is the opposite to the PolyPhen score, where higher scores are considered deleterious)
PolyPhen	PolypPhen predicts whether an amino acid substitution affects protein function. The PolyPhen score represents the probability that a substitution is damaging. Values nearer 1 are more confidently predicted to be deleterious (note that this is the opposite to the SIFT score, where lower scores are considered deleterious)
GMAF	The total allele frequency of the minor allele in the 1000 genomes project dataset (0-1)
MAF_AFR	The minor allele frequency in the 1000 genomes project dataset for the African population (0-1)
MAF_AMR	The minor allele frequency in the 1000 genomes project dataset for the American population (0-1)
MAF_ASN	The minor allele frequency in the 1000 genomes project dataset for the Asian population (0-1)
MAF_EUR	The minor allele frequency in the 1000 genomes project dataset for the European population (0-1)
MAF_AA	The minor allele frequency in the NHLBI-ESP project dataset for the African American population (0-1)
MAF_EA	The minor allele frequency in the NHLBI-ESP project dataset for the African population (0-1)

Table 11. Definition of the variant summary statistics

Limitations of Use
This product was developed, manufactured, and sold for *in vitro* use only. The product is not suitable for administration to humans or animals. SDS sheets relevant to this product are available upon request.

For more information please visit **www.archerdx.com**

ArcherDX
2477 55th Street, Suite 202
Boulder, CO 80301
(877)-771-1093
www.archerdx.com