

# Invitae's Archer® Analysis 7.0 User Manual

For software version(s): 7.0

## Notices

### Limitations of Use

**For Research Use Only.** Not intended to be used for diagnosis or treatment of human or animal diseases. This product was developed, manufactured, and sold for research use only.

### Symbols Used In This Manual



**Attention** symbols denote critical information about additional resources needed, key workflow steps, limitations of use, or where risk of assay failure could occur if not carefully observed.



**Reminder** symbols call attention to minor details that may be easily overlooked and compromise the procedure resulting in decreased assay performance.

## In This User Manual

This manual is a guide to using Archer Analysis, a comprehensive bioinformatics suite designed to be used in conjunction with Archer library preparation kits and gene enrichment panels for targeted sequencing of select genes and regions of interest using next-generation sequencing (NGS).

### Overview – Page 3

This section contains the intended use statement, test principle, an overview of the scope of the software, and a description of how the software can be accessed.

### Before Getting Started – Page 4

This section contains critical guidance for the successful implementation of the functions and workflows described in this manual, including additional software that may be required but not supplied by Archer. This should be read and understood before data analysis is initiated.

### Detailed Software Features and Workflows – Page 6

This section provides detailed definitions and step-by-step instructions for the features and workflows available in the software. A table of contents is provided for quick reference to specific headings.

## Additional Resources

View videos and additional resources for Archer products at:

<https://support.archerdx.com>

## Technical Support

Visit <https://support.archerdx.com> for a list of helpful answers to frequently asked questions or contact us directly at [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com).

## Overview

### Intended Use

Invitae's Archer Analysis is a comprehensive bioinformatics suite intended for use only with next-generation sequencing (NGS) data generated by Archer reagent kits and corresponding target-enrichment panels on either Ion Torrent™ or Illumina® sequencing platforms. Archer Analysis provides functions for all steps of secondary analysis, (read cleaning, alignment, and variant calling) and some steps for tertiary analysis. The software can be used with any Archer panel, without need for installing additional modules or other special configuration.

### Test Principle

Anchored Multiplex PCR (AMP) is a rapid and scalable method of generating target-enriched libraries for NGS. This technology can be used for applications in targeted RNA sequencing, genomic DNA sequencing, and genotyping applications to generate a sequencing library in a matter of hours. Designed for low nucleic acid input, this process delivers robust performance across a variety of sample types.

AMP utilizes unidirectional gene-specific primers (GSPs) that enrich for both known and unknown variants. During library preparations, adapters containing molecular barcodes and sample indices are ligated to nucleic acid fragments.

The Archer Analysis software utilizes these molecular barcodes for error correction and read deduplication to support quantitative multiplex data analysis and confident variant detection. Archer Analysis reports both sequencing metrics and number of unique observations supporting called variants.

### Access the Software

Archer Analysis is a web-based software, and Google Chrome™ is the recommended web browser. A hosted version of the software (private to your organization) is available via Invitae's Archer Analysis Unlimited service. A free demo of this service can be accessed at <https://analysis.archerdx.com/>.

If local installation is required, your lab *must* secure the requisite computing resources as the Archer Analysis server is packaged as a virtual machine (i.e., must run in a virtualization environment such as VMware ESXi®.) View *Before Getting Started* for additional information.

Please visit <https://archerdx.com/technology-platform/analysis/>, contact your field team, or [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com) for further details and assistance in accessing any of the options listed above.

## Before Getting Started

### Resources Required To Operate Archer Analysis

While the user interacts with Analysis via a standard web browser, data is processed, stored, and served via the Analysis Server package, and this server must be accessible to the user's web browser and running on one of these options:

1. The user's local machine
2. other hardware on the local intranet
3. a remotely hosted location on the internet

Archer provides option 3 as a private, paid service for organizations through the Analysis Unlimited service which automatically and dynamically allocates computing resources to process any number of samples submitted by the user.

If option 1 or 2 is chosen, the Analysis server must be installed on a VMware-based hypervisor. Of those, we recommend ESXi for best performance, but we also support Workstation and Fusion.



Archer Analysis does not provide virtualization software to run the Analysis Server VM locally; this must be obtained separately by the organization. VMware is recommended.

The computing resources required to run Archer Analysis depend on the lab's required throughput, i.e., the number of samples and total processing time required per batch. For detailed instructions on installation and system requirements of the VM please contact [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com).

### Preparing Your Data For Archer Analysis

Samples are processed in Archer Analysis as a job. One or multiple samples can be run together in a single job. Note these factors when preparing data for an Analysis job:

- All samples must have been prepped using the same Archer panel (but not necessarily at the same time).
- All samples in a job will be processed using the same Analysis pipelines and settings that are assigned at job setup.

Archer Analysis accepts four types of data files as input when setting up a new job, two required and two optional.

### Demultiplexed FASTQ (Illumina) or BAM (Ion Torrent) file(s) – REQUIRED

For each sample to be processed in a job, the corresponding FASTQ (for Illumina data) or BAM (for Ion Torrent data) file must be uploaded to Archer Analysis (both R1 and R2 files are required for samples that are paired-end sequenced).



Note these factors when preparing FASTQ files for a job:

- Software to demultiplex FASTQ files is NOT provided by Archer but provided by the sequencing instrument vendor.
- Aside from demultiplexing, no other preprocessing of the sequencing data is required. (Analysis performs all read trimming, cleaning, etc.)

### Panels file – REQUIRED

This is a GTF file unique to the Archer panel used for library preparation of the sample(s) to be processed in a job. It is supplied by Archer and is packaged in Analysis while custom GTF files are generated ad-hoc and require upload after installation as described in Managing and Adding GTF Files. Consult the product literature to determine which file should be used and where to source it. Contact [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com) for questions.



### **Targeted Mutations file – OPTIONAL**

If there are specific mutations of interest when running the SNP/InDel pipeline (for either RNA or DNA panels), then a list of these can be compiled in Variant Call Format (VCF, v4.1) and supplied to the software before processing a batch of samples. This forces the software to report on each mutation in the list regardless of being called by the pipeline.



Using a Targeted Mutations file with the SNP/InDel pipeline is an excellent way to quickly surface information on mutations of known significance for your application. It can be used while simultaneously searching for non-targeted variants, so no stone is left unturned.

### **Regions of Interest file – OPTIONAL**

This file defines the targeted regions of interest assessed when running the DNA Target Coverage pipeline (for VariantPlex panels) in BED Format (requiring chromosome, start position and stop position). This file is supplied to the software before processing a batch of samples. The software will report coverage metrics on each location specified.

## Detailed Software Features and Workflows

<b>PIPELINES INCLUDED IN ARCHER ANALYSIS</b> .....	<b>11</b>
<b>FUSIONPLEX® (RNA) ASSAYS</b> .....	<b>11</b>
<b>VARIANTPLEX® (DNA) ASSAYS</b> .....	<b>11</b>
<b>LIQUIDPLEX® (CTDNA™) ASSAYS</b> .....	<b>11</b>
<b>IMMUNOVERSE™ (RNA) ASSAYS</b> .....	<b>11</b>
<b>PRIMARY NAVIGATIONAL FEATURES</b> .....	<b>11</b>
<b>MAIN HEADER AND MAIN MENU</b> .....	<b>11</b>
<b>DROPDOWN MENUS</b> .....	<b>11</b>
<b>FOOTER: LEGAL AND ABOUT ARCHER ANALYSIS PAGES</b> .....	<b>12</b>
<b>INFO POP-UPS</b> .....	<b>12</b>
<b>INTRODUCTION</b> .....	<b>12</b>
<b>GUIDED TUTORIALS</b> .....	<b>13</b>
<b>OUTPUT OPTIONS, SYSTEM ARCHITECTURE, AND ALGORITHM DETAILS</b> .....	<b>13</b>
<b>LOG INTO ANALYSIS</b> .....	<b>14</b>
<b>CREATE AN ACCOUNT</b> .....	<b>15</b>
<b>THE LANDING “HOME” PAGE</b> .....	<b>15</b>
<b>LOG OUT OF ANALYSIS</b> .....	<b>16</b>
<b>MANAGE USERS AND GROUPS</b> .....	<b>16</b>
<b>ADMIN USERS</b> .....	<b>16</b>
<b>USERS</b> .....	<b>17</b>
Add and/or Edit Users .....	18
Delete Users .....	19
<b>GROUPS</b> .....	<b>20</b>
Add and Edit Groups .....	20
Delete Groups .....	21
<b>MANAGE SUPPORTING FILES (GTF, VCF, AND BED) FILES</b> .....	<b>21</b>
<b>ADD AND MANAGE FILES</b> .....	<b>22</b>
<b>FILE ACTIONS (VIEW, DOWNLOAD, VISIBILITY, AND OWNERSHIP)</b> .....	<b>22</b>
Delete Files .....	23
Change Visibility .....	23
Change Ownership .....	24
<b>USER SETTINGS (APPLIED ACROSS ALL JOBS)</b> .....	<b>24</b>
<b>VIEW AND MANAGE USER SETTINGS</b> .....	<b>25</b>
<b>GENERAL ANALYSIS SETTINGS</b> .....	<b>25</b>
Min Average Unique DNA and AMBIG Start Sites per GSP2 .....	25
Min Average Unique RNA Start Sites per GSP2 Controls .....	26
Error Correction .....	26
Read Depth Normalization .....	26
Barcode Length .....	26
Common Region Set .....	26
Pre-/Post-Alignment Read Cleaner .....	26
Merge Reads in Read Cleaner .....	26

Debug .....	27
<b>STRUCTURAL VARIANT ANALYSIS SETTINGS .....</b>	<b>28</b>
<b>DISPLAY INTRONIC STRUCTURAL VARIANT</b> .....	<b>28</b>
<b>MIN READS FOR VALID STRUCTURAL VARIANT</b> .....	<b>28</b>
<b>CROSS-CONTAMINATION SINGLE PANEL</b> .....	<b>28</b>
<b>CROSS-CONTAMINATION FUSION CONFIDENCE THRESHOLD</b> .....	<b>29</b>
<b>EXPRESSION IMBALANCE</b> .....	<b>29</b>
<b>MIN UNIQUE START SITES FOR VALID STRUCTURAL VARIANT</b> .....	<b>29</b>
<b>STRUCTURAL VARIANT PERCENT OF GSP2 READS</b> .....	<b>29</b>
<b>EXPRESSION IMBALANCE ALPHA</b> .....	<b>29</b>
<b>CALL TRANSCRIPTIONAL READ THROUGH EVENTS WEAK</b> .....	<b>29</b>
<b>RNA EXPRESSION VISUALIZATION</b> .....	<b>29</b>
<b>MIN RNA RELATIVE EXPRESSION VALUE</b> .....	<b>29</b>
<b>MAX RNA RELATIVE EXPRESSION VALUE</b> .....	<b>29</b>
<b>RNA EXPRESSION OUTLIERS</b> .....	<b>29</b>
<b>MIN REPORTABLE RNA WT PERCENT OF READS</b> .....	<b>29</b>
<b>VARIATIONS/MUTATIONS SETTINGS .....</b>	<b>30</b>
<b>MIN DEPTH FOR VARIANT CALL</b> .....	<b>30</b>
<b>MAPQ THRESHOLD FOR VARIANT CALL</b> .....	<b>30</b>
<b>MIN BASEQUAL FOR VARIANT CALL</b> .....	<b>30</b>
<b>MIN ALLELE FRACTION FOR VARIANT CALL</b> .....	<b>30</b>
<b>MIN PHRED QUAL SCORE FOR VARIANT CALL</b> .....	<b>31</b>
<b>VARIANT DOWNSTREAM ROI SIZE</b> .....	<b>31</b>
<b>OUTLIER DETECTION</b> .....	<b>31</b>
<b>MIN DAO FOR OUTLIER DETECTION CALCULATION</b> .....	<b>31</b>
<b>DEEP SHALLOW THRESHOLD</b> .....	<b>31</b>
<b>CNV PIPELINE SETTINGS .....</b>	<b>31</b>
<b>CNV STRONG AMPLICATION THRESHOLD</b> .....	<b>31</b>
<b>CNV STRONG DELETION THRESHOLD</b> .....	<b>31</b>
<b>CNV P-VALUE THRESHOLD</b> .....	<b>31</b>
<b>MSI PIPELINE SETTINGS .....</b>	<b>32</b>
<b>MINIMUM NUMBER OF LOCI FOR MSI</b> .....	<b>32</b>
<b>MINIMUM DEPTH PER SAMPLE FOR MSI</b> .....	<b>32</b>
<b>STABLE THRESHOLD FOR MSI</b> .....	<b>32</b>
<b>HIGH THRESHOLD FOR MSI</b> .....	<b>32</b>
<b>MSI ALPHA</b> .....	<b>32</b>
<b>MINIMUM DEPTH PER LOCUS</b> .....	<b>32</b>
<b>MISCELLANEOUS SYSTEM SETTINGS .....</b>	<b>32</b>
<b>ITEMS PER PAGE</b> .....	<b>33</b>
<b>LANDING "HOME" PAGE</b> .....	<b>33</b>
<b>AUTOMATICALLY EXPAND AGGREGATE ISOFORM VISUALIZATIONS</b> .....	<b>33</b>

<b>SAVE NEW USER SETTINGS OR RESTORE DEFAULTS .....</b>	<b>33</b>
<b>IDENTIFY NON-DEFAULT SETTINGS .....</b>	<b>33</b>
<b>JOB SETUP AND MANAGEMENT .....</b>	<b>33</b>
<b>STARTING A JOB: NEW JOB PAGE .....</b>	<b>34</b>
Section 1: Set Up.....	34
<b>CHOOSING THE PIPELINE(S) .....</b>	<b>34</b>
Section 2: Pipeline.....	35
Common Analysis Types.....	35
RNA Analysis Types .....	35
DNA and cfDNA Analysis Types.....	37
Section 3: Settings.....	41
Section 4: Hooks.....	41
Section 5: Batch Sign-out.....	42
Section 6: Sequenced Libraries .....	43
Selecting Normalization Parameters for CNV.....	44
<b>RUNNING JOBS .....</b>	<b>45</b>
Progress Bar and Status Updates .....	47
Analyzing Multiple Samples & Jobs.....	48
Stopping a Running or Queued Job .....	48
Jobs That Complete with Errors.....	49
<b>JOB HOOKS .....</b>	<b>52</b>
<b>ADD A JOB HOOK .....</b>	<b>52</b>
<b>MANAGING EXISTING HOOKS .....</b>	<b>54</b>
Editing Existing Hooks .....	54
Enabling Hooks on a Per Job Basis .....	55
Deleting Existing Hooks.....	56
<b>ANALYTICAL VALIDATION SETS .....</b>	<b>56</b>
<b>MARK SAMPLES FOR ANALYTICAL VALIDATION SET .....</b>	<b>56</b>
<b>CREATE AN ANALYTICAL VALIDATION SET .....</b>	<b>57</b>
<b>DELETE AN ANALYTICAL VALIDATION SET .....</b>	<b>58</b>
<b>RUN AN ANALYTICAL VALIDATION SET .....</b>	<b>58</b>
<b>VIEW ANALYTICAL VALIDATION SAMPLES .....</b>	<b>58</b>
<b>ANALYTICAL VALIDATION REPORT .....</b>	<b>59</b>
<b>RUN HISTORY .....</b>	<b>60</b>
<b>ACCESSING AND INTERPRETING RESULTS .....</b>	<b>60</b>
<b>KEY ASSAY CONCEPTS .....</b>	<b>60</b>
Genome Reference.....	60
Raw Reads, Unique Fragments, and Unique Start Sites.....	61
RNA vs. DNA vs. Ambiguous Reads.....	61
Control Genes (FusionPlex Only).....	62
<b>GENERAL RESULTS REVIEW &amp; INTERPRETATION WORKFLOW .....</b>	<b>63</b>
<b>SIGN-OUT BATCHES .....</b>	<b>63</b>
<b>ANALYZED SAMPLES .....</b>	<b>63</b>
Sample Summary.....	65
QC Results .....	66
Viewing and Overriding QC Pass/Fail Threshold.....	67
Results Review and Reporting .....	68
<b>JOBS .....</b>	<b>68</b>
Configuration .....	69
View Logs .....	70
Clone Job and Rerun Job.....	71

Job Reports .....	71
Download Results.....	71
<b>JOB DETAILS PAGE (SAMPLE SUMMARY)</b> .....	<b>71</b>
Visualize Sample Data.....	73
<b>SAMPLE DETAILS PAGE</b> .....	<b>73</b>
Read Statistics .....	73
Assay Targets .....	74
Pipeline-Specific Links.....	75
<b>PIPELINE-SPECIFIC INFORMATION</b> .....	<b>75</b>
Fusions (RNA Only) .....	75
RNA Expression Pipeline Outputs .....	81
SNP/InDel Variants (RNA and DNA).....	84
CNVs (DNA Only).....	85
MSI (DNA Only).....	87
Outlier Detection .....	90
Lymphoma Cell of Origin (Experimental).....	92
Immune Repertoire (RNA Only).....	93
<b>NORMAL DATA SETS (SNP/INDEL AND CNV)</b> .....	<b>99</b>
<b>NORMAL DATA SETS TERMINOLOGY</b> .....	<b>99</b>
<b>CREATING NORMAL DATA SETS</b> .....	<b>100</b>
<b>USER SETTINGS FOR NORMAL DATA SETS</b> .....	<b>100</b>
<b>CREATING A NORMAL DATA SET</b> .....	<b>101</b>
<b>SAMPLE VISIBILITY</b> .....	<b>101</b>
<b>ADDING A SAMPLE TO AN NORMAL DATA SET FROM JOB DETAILS</b> .....	<b>102</b>
<b>EXISTING NORMAL DATA SETS</b> .....	<b>103</b>
<b>SUBMITTING A NEW JOB WITH NORMAL DATA SETS</b> .....	<b>104</b>
<b>PROTOCOLS AND NORMAL DATA SETS</b> .....	<b>105</b>
<b>DEPRECATED AND UPGRADING NORMAL DATA SETS</b> .....	<b>105</b>
<b>WORKFLOW AUTOMATION AND INTEGRATIONS</b> .....	<b>106</b>
<b>PROTOCOLS AND WATCHED FOLDERS</b> .....	<b>106</b>
<b>CREATE A NEW PROTOCOL</b> .....	<b>107</b>
<b>START A NEW JOB USING AN EXISTING PROTOCOL</b> .....	<b>110</b>
<b>SEARCH AND FILTER EXISTING PROTOCOLS</b> .....	<b>110</b>
<b>CONFIGURE EXISTING PROTOCOLS</b> .....	<b>112</b>
<b>SET UP WATCHED FOLDERS</b> .....	<b>112</b>
<b>ADD A NEW WATCHED FOLDER</b> .....	<b>112</b>
<b>REMOVE A WATCHED FOLDER</b> .....	<b>113</b>
<b>REMOVE ASSETS USED IN A PROTOCOL</b> .....	<b>114</b>
<b>DEVELOP PROCEDURE/SCRIPT FOR THE MOVEMENT OF FASTQ OR BAM FILES</b> .....	<b>114</b>
<b>HOW ARCHER ANALYSIS KNOWS WHEN TO EXECUTE A WORKFLOW</b> .....	<b>114</b>
Example for a Job Containing a Single Sample .....	114
Example for a Job Containing Multiple Samples.....	115
Location of Files in the Watched Folder upon Analysis.....	115
File Name Collisions .....	115
S3 Watched Folders.....	116
SSL-Enabled Linked Analysis Servers.....	116
<b>ADD LINKED ANALYSIS SERVER</b> .....	<b>116</b>
<b>ENABLE LINKED ANALYSIS SERVER</b> .....	<b>117</b>
<b>TROUBLESHOOTING</b> .....	<b>118</b>

JOB FINISHED WITH JOB STATUS “COMPLETED_ERROR”	118
JOB IS STUCK WITH JOB STATUS “NEW”	118
JOB IS STUCK WITH JOB STATUS “HOLD”	118
APPENDIX.....	119
GLOSSARY OF GENERAL TERMS	119
FORMAT OF THE GTF FILES	119
FORMAT OF THE TARGETED MUTATION FILE (TMF)	120
TECHNOLOGY LICENSES.....	121

## Pipelines Included in Archer Analysis

The pipelines available in the software are listed below, and pipelines within a given product group can be run in parallel.

### FusionPlex® (RNA) Assays

- **Fusions/Isoforms** – to call fusions and other structural rearrangements
- **SNP/InDel** – to call point mutations and insertions/deletions up to ~30bp in length
- **Expression** – for determining relative expression levels and imbalances
- **Lymphoma Cell of Origin (COO)** – experimental feature to identify the cell of origin subtype from a diffuse large B-cell lymphoma (DLBCL) sample

### VariantPlex® (DNA) Assays

- **Structural Variations** – to call exon skipping, internal tandem duplications (ITDs), and other structural rearrangements.
- **SNP/InDel** – to call point mutations and insertions/deletions up to ~30bp in length
- **CNV** – to identify relative copy number gains/losses at both the primer and gene-level
- **MSI** – to identify microsatellite instability at targeted regions

### LiquidPlex® (ctDNA™) Assays

- **Structural Variations** – to call exon skipping, internal tandem duplications (ITDs), and other structural rearrangements
- **SNP/InDel** – to call point mutations and insertions/deletions up to ~30bp in length
- **CNV** – to identify relative copy number gains/losses at both the primer and gene-level

### Immunoverse™ (RNA) Assays

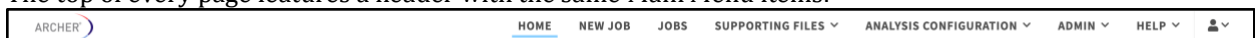
- **Immune Repertoire** – to identify total clonotypes (diversity) and specific clonal frequencies for TCR and BCR populations, as well as the hypermutation status for BCR populations.

## Primary Navigational Features

Archer Analysis features the same core navigational elements, across the entire system, to ensure a consistent and intuitive experience. Additionally, once logged in, every page has the same core structural elements, containing either links to key parts of the system and/or informational elements intended to guide use of the system. These navigational and structural elements are listed below.

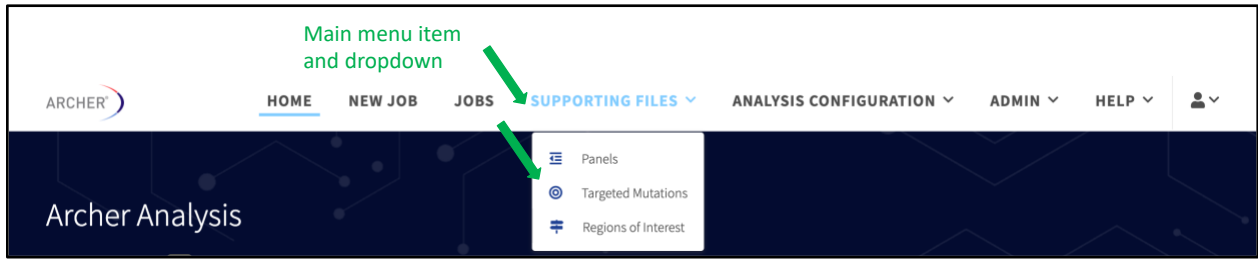
### Main Header and Main Menu

The top of every page features a header with the same Main Menu items:



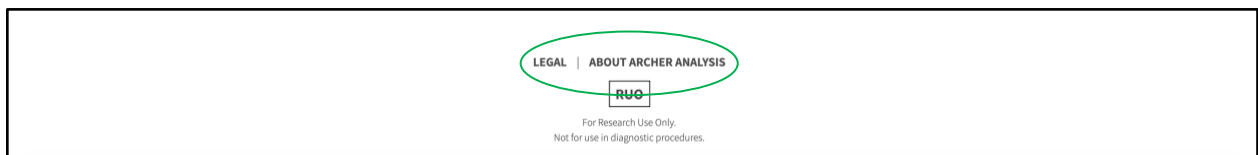
### Dropdown Menus

Dropdown menus are used throughout the system. The following shows the options available from the Supporting Files dropdown in the Main Menu:




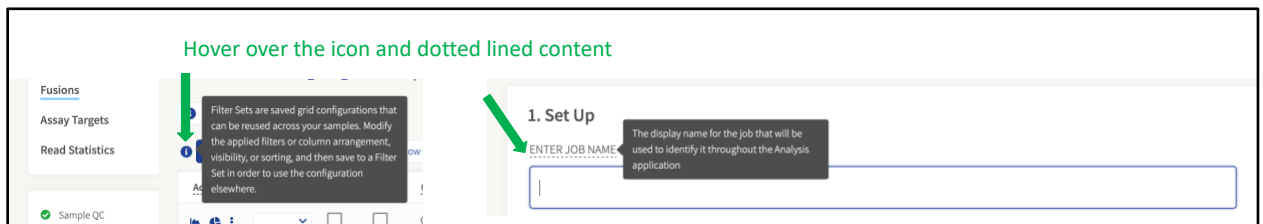
## Footer: Legal and About Archer Analysis Pages

The bottom of every page features a footer that includes links to the Legal and About Archer Analysis pages which detail information on the exact version of Archer Analysis, which third-party components are used, etc.:



## Info Pop-ups

Whenever an Information Icon  appears next to terms and metrics throughout Archer Analysis, users can click or hover over the symbol to view a descriptive informational pop-up. There are also many elements in the application with a 'dotted underline' as shown below. When these are available, hover to display the pop-up content. Make use of these for a quick reference for definitions of terms, fields, or explanations of how parameters are being used in the software.



## Introduction

Invitae's Archer® Analysis is a comprehensive bioinformatics suite that includes pipelines for all of Invitae Oncology's targeted NGS assays, both catalog and custom. The software covers all secondary analysis (e.g., read trimming, QC, alignment, and mutation calling), as well as initial steps of tertiary analysis (e.g., annotations from Ensembl Variant Effect Predictor (VEP) and ClinVar):

- <http://uswest.ensembl.org/info/docs/tools/vep/index.html>
- <https://www.ncbi.nlm.nih.gov/clinvar/>



Technical specialists (adx-tech-support@invitae.com) are available to assist in this process. Once optimized, Analysis is an excellent tool for routine use by both expert and novice users.



## Guided Tutorials

In-app guided tutorials are available throughout Archer Analysis that are intended to highlight new and improved features as well as to help use the software more efficiently. When logging in for the first time, pop ups will appear that inform, point to, or highlight parts of the application. We recommend following along and clicking next when possible; even power users may find additional features or faster ways of accomplishing tasks. There are several screens and processes with permanently placed info icons in the upper left corner. These circular icons are light blue. When these icons are clicked, a walk through will begin that *walks you through* the process specific to that screen or set of elements on that screen.



In the lower right corner of the screen is a larger question mark icon. When clicked, the Help Menu will display. This menu has walk throughs available to help with common tasks for working quickly and efficiently in Archer Analysis.



View *Primary Navigational Features* for additional information.

## Output Options, System Architecture, and Algorithm Details

Most of the results from Archer Analysis can be viewed in its graphical user interface (GUI) accessed via a standard web browser. Full results, including processing logs, can also be accessed as a combination of flat text, BAM, image, and PDF files, as well as via a REST API. This manual focuses on the GUI. View *Reports and Other Output Files* and *Output from Download All Files* for additional information regarding these features.

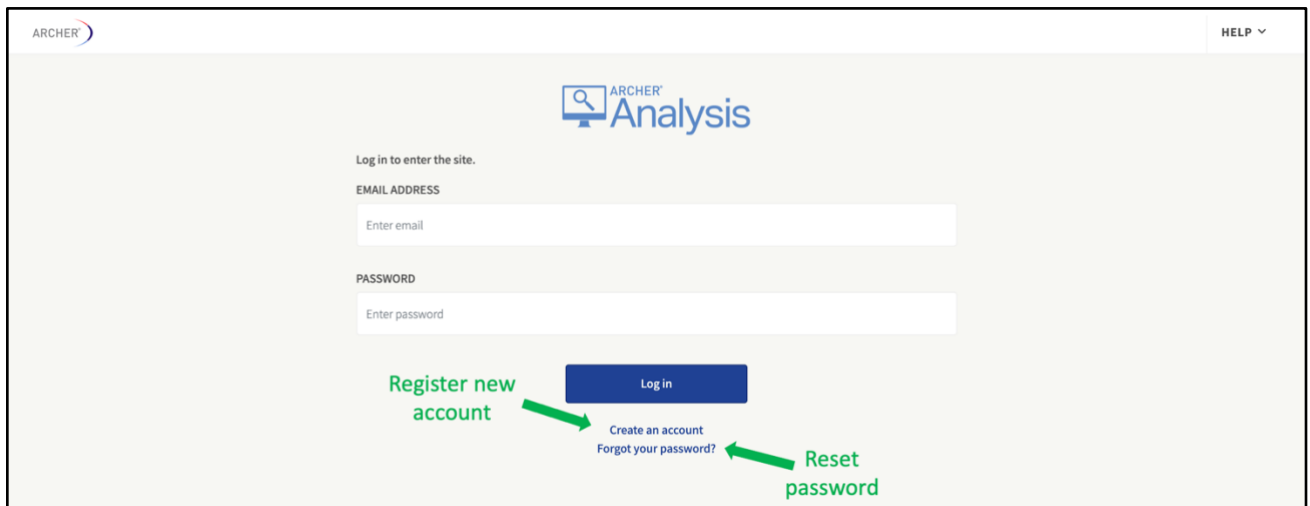
Please note this manual is intended primarily to guide routine use of the software and does not provide an exhaustive explanation of the methods and algorithms underpinning the various pipelines and functions available. If needed for security, privacy, or other concerns of your lab, additional documentation is available on the below topics and can be obtained by contacting your field team or [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com) for assistance.

- How to install the software locally as a VM
- General system architecture
- Details of each bioinformatics pipeline
- PDF Reporting API
- REST API
- Sample processing logs

## Log Into Analysis

Archer Analysis is accessible via most web browsers except for Internet Explorer® version 7 which is incompatible and unsupported. Google Chrome™ is the recommended web browser. When using a Virtual Machine, refer to the *Archer Analysis Virtual Machine Installation Guide* for instructions on how to discover the IP address of the Archer Analysis software. Once identified, open a web browser, and enter the **IP address** for the virtual machine.

The Archer Analysis Software is a fully contained and secure environment, allowing users to run their analyses under their own account. The login screen allows users to sign into their account, register for a new account, and reset their password.



If the login page does not appear, ensure you are connected to the internet, or that the virtual machine is running and accessible by the IP address or URL designated by your site administrator.



There is a limit of five (5) login attempts for incorrect passwords, after which the account will be deactivated. Users will then be required to supply answers to two security questions specified at account registration to re-assign a new password. Alternatively, an admin-level user can manually re-activate accounts and assign temporary passwords.

## Create an Account

1. From the Login screen, click **Create an account**. The Account Creation page will appear.
2. Enter your email address as the login, create a password, and specify two security questions/answers. For security reasons, passwords require the following complexity:
  - Password cannot be similar to personal information
  - Password must contain at least nine (9) characters
  - Password cannot be a commonly used password
  - Password must contain one (1) uppercase letter, one (1) lowercase letter, one (1) digit, and (1) special character.
3. Read the End User License Agreement (EULA) and **check the box**, to indicate the EULA has been read and agreed to.
4. Click **Create Account**. You will automatically be logged in. Alternatively, your system administrator can create a login for you. View *Adding and/or Editing Users* for more details, though the above steps are still required upon first login.

Account Information

Email

Password

Repeat Password

---

Forgot Password Security

In case you forget your password you can click forgot password and answer 2 security questions to reset your password. Please select your preferred questions and answers.

Security Question 1

Security Answer

Security Question 2

Security Answer

End User License Agreement

END USER LICENSE AGREEMENT FOR ARCHERDX ARCHER(TM) SOFTWARE

BEFORE USING THIS SOFTWARE PRODUCT (THE "SOFTWARE") YOU SHOULD CAREFULLY READ THE FOLLOWING TERMS AND CONDITIONS THAT APPLY TO YOUR USE OF THE SOFTWARE. BY DOWNLOADING, COPYING, OR OTHERWISE USING THE SOFTWARE, YOU AGREE TO BE BOUND BY ALL OF THE TERMS AND CONDITIONS OF THIS AGREEMENT. TO THE EXTENT YOU OBTAIN THIS SOFTWARE ON BEHALF OF ANOTHER PERSON OR ENTITY, YOU AGREE TO BIND THAT PARTY, AND THROUGH THAT PARTY ITS EMPLOYEES, CONSULTANTS, AND OTHER AGENTS, TO THESE TERMS AND CONDITIONS AND YOU REPRESENT, WARRANT AND COVENANT THAT YOU HAVE THE AUTHORITY TO DO SO. IF YOU DO NOT ACCEPT THESE TERMS AND CONDITIONS, THEN YOU MAY NOT USE THIS SOFTWARE.

RESTRICTIONS OF USE.

I agree to the terms in the End User License Agreement.

[Create Account](#)

## The Landing "Home" Page

After successfully logging in, the Archer Analysis home page will display.

**Note:** If a user has been assigned admin privileges or assigned to a User Group with jobs sharing enabled, they may see jobs already started by other users.



The landing page is configurable within the user settings indicated above. Click **user icon** (as shown above by green arrow) → **Settings** → **Landing Page**.

## Log Out of Analysis

1. Click **user icon** → **Logout**.



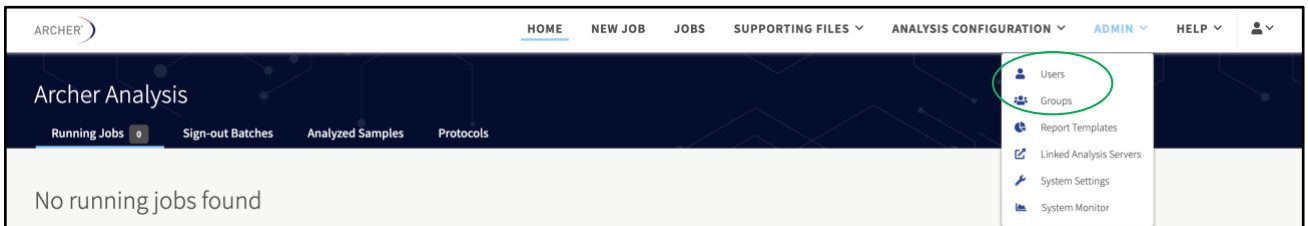
Users are automatically logged out after thirty minutes of inactivity.

## Manage Users and Groups

Archer Analysis utilizes user groups to control the visibility of job and sample data, as well as access to global system settings. There are two user types: admins and basic users. Additional permissions and data access granularity can be defined via custom user groups. 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 of their groups.

### Admin Users

By default, the admin user is a member of the Basic Users and Experimental groups, and has a special permission called Is System Admin, which is required to perform any of the user, group, and permissions management described in this section. This functionality is accessed by clicking **ADMIN** from the main header menu and selecting either **Users** or **Groups**.





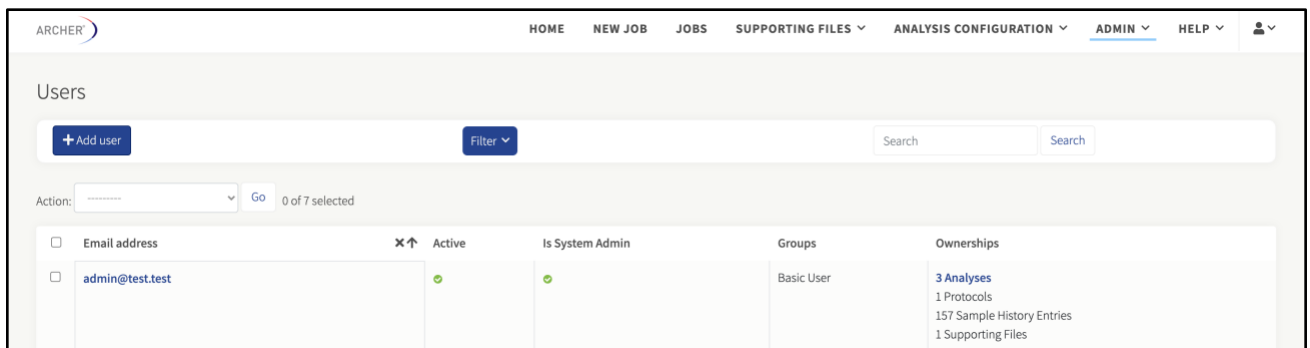
**Note:** Archer Analysis does not generally come configured with a default admin user. One of the primary steps in installing Analysis is to set up the first admin account via the command line interface (CLI). Please refer to the *Archer Analysis Virtual Machine Installation Guide* for further details. For Archer Analysis Unlimited (AAU) instances, admin user(s) will be added by Invitae engineering team through the provided questionnaire during AAU instance setup.



Admin users automatically have many privileges over basic users. This includes but is not limited to: (1) user and group management, (2) access to experimental features, (3) ability to view all jobs in the system, (4) ability to change ownership of jobs to any user, and (5) management of Targeted Mutations (VCF) and Assay Targets (GTF) files.

## Users

Selecting **ADMIN** → **Users** will navigate to a table containing a list of all the users in the system. The columns indicate the status of each user, whether the account is active, whether they are a designated system admin, the groups to which they belong, and how many jobs and supporting files they own.



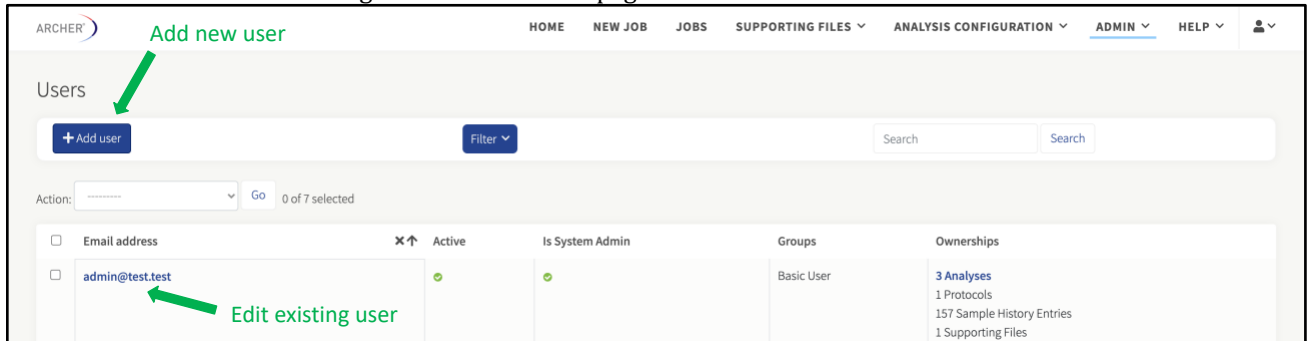
<input type="checkbox"/>	Email address	Active	Is System Admin	Groups	Ownerships
<input type="checkbox"/>	admin@test.test	Active	Is System Admin	Basic User	<b>3 Analyses</b> 1 Protocols 157 Sample History Entries 1 Supporting Files

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

Users can search for users by entering the entire or part of the email address in the Search field.

## Add and/or Edit Users

1. Click **Add User** to navigate to the Add User page.



ARCHER® HOME NEW JOB JOBS SUPPORTING FILES ANALYSIS CONFIGURATION ADMIN HELP

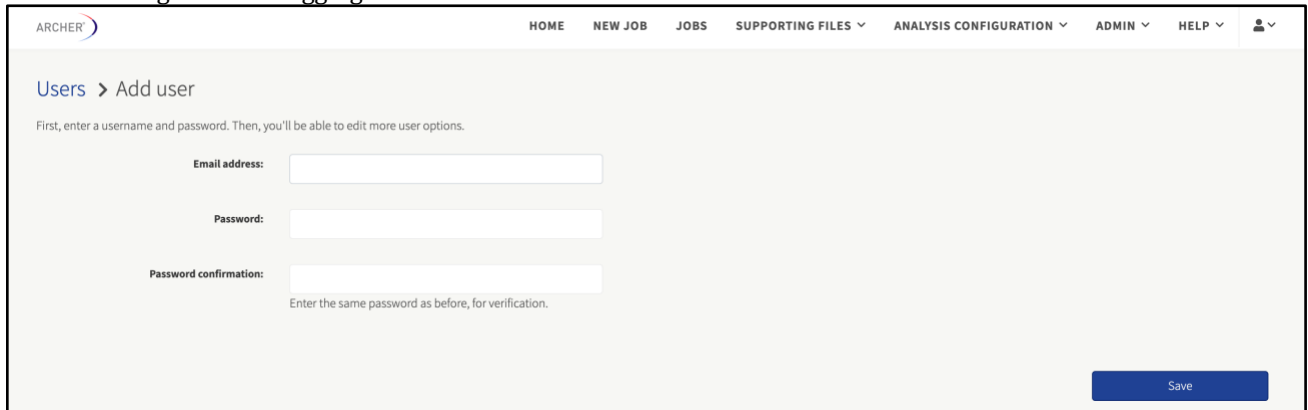
Users

+ Add user Filter Search Search

Action: Go 0 of 7 selected

<input type="checkbox"/>	Email address	Active	Is System Admin	Groups	Ownerships
<input type="checkbox"/>	admin@test.test	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Basic User	<b>3 Analyses</b> 1 Protocols 157 Sample History Entries 1 Supporting Files

2. Enter the required information and click **Save**. This is a temporary password and users will be forced to change it when logging in for the first time.



ARCHER® HOME NEW JOB JOBS SUPPORTING FILES ANALYSIS CONFIGURATION ADMIN HELP

Users > Add user

First, enter a username and password. Then, you'll be able to edit more user options.

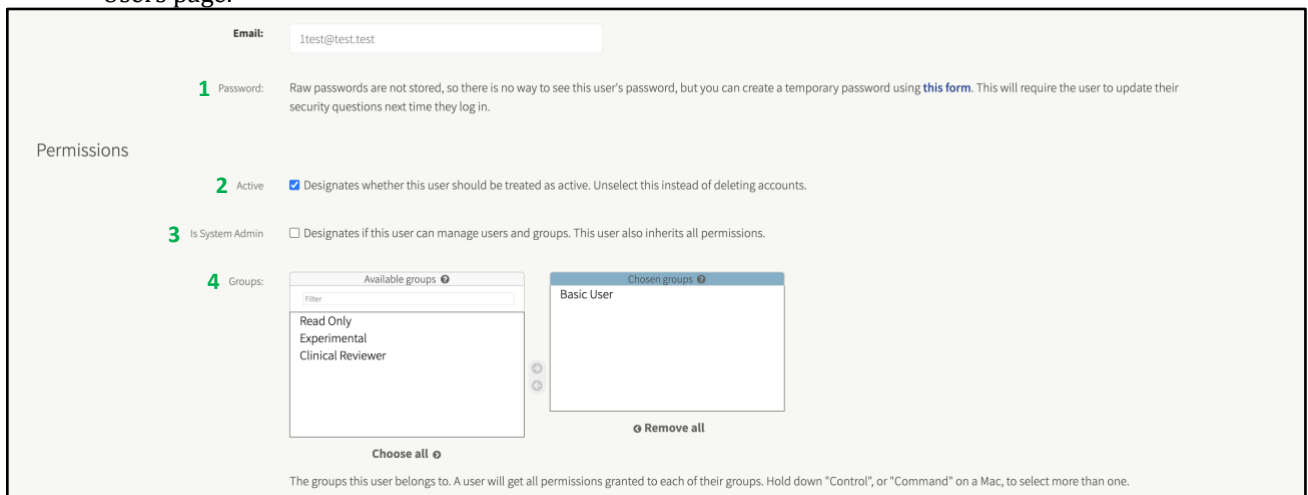
Email address:

Password:

Password confirmation:   
Enter the same password as before, for verification.

Save

3. As shown below, the Edit User page can be reached by adding a new user or clicking the username in the Users page.



Email:

**1** Password: Raw passwords are not stored, so there is no way to see this user's password, but you can create a temporary password using [this form](#). This will require the user to update their security questions next time they log in.

Permissions

**2** Active  Designates whether this user should be treated as active. Unselect this instead of deleting accounts.

**3** Is System Admin  Designates if this user can manage users and groups. This user also inherits all permissions.

**4** Groups:

Available groups

Filter

- Read Only
- Experimental
- Clinical Reviewer

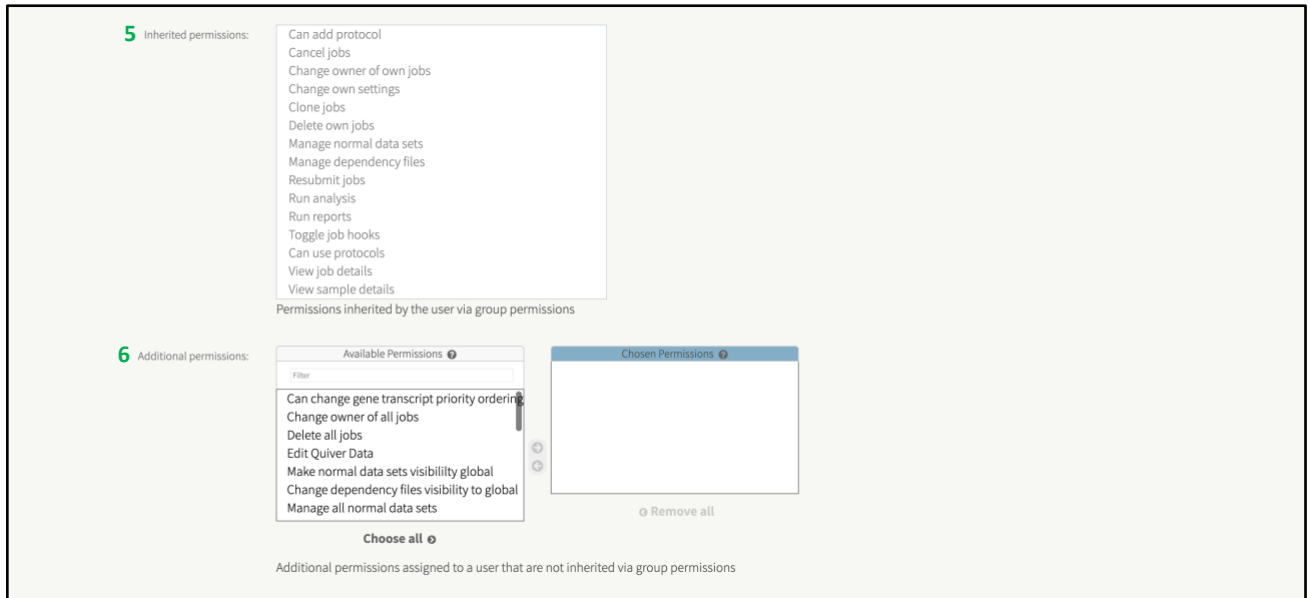
Choose all

Chosen groups

- Basic User

Remove all

The groups this user belongs to. A user will get all permissions granted to each of their groups. Hold down "Control", or "Command" on a Mac, to select more than one.



- A new, temporary password can be assigned to the user.
- 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.

Detailed permissions can also be viewed and customized on the Edit User page, as shown below.

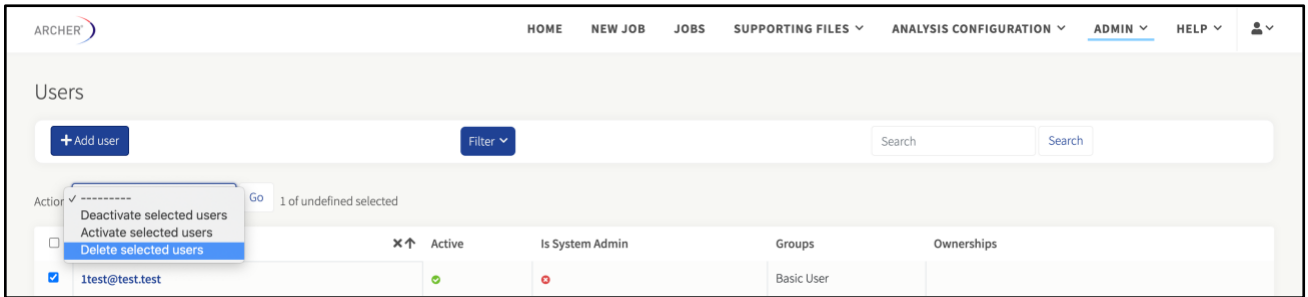
- Permissions inherited by the chosen Groups are shown.
- Additional permissions can be granted by choosing from the Available Permissions list and moving them to the Chosen Permissions list.

When finished editing the user, click **Save**.



### Delete Users

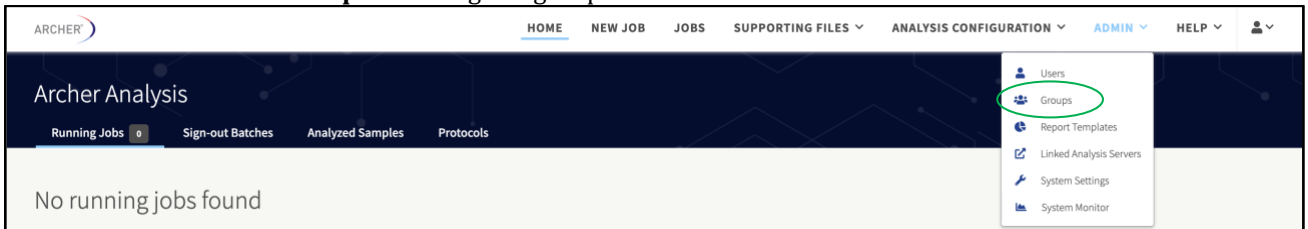
Users can be deleted from the Users page by selecting the **checkbox next to each username** and then using the Action box.



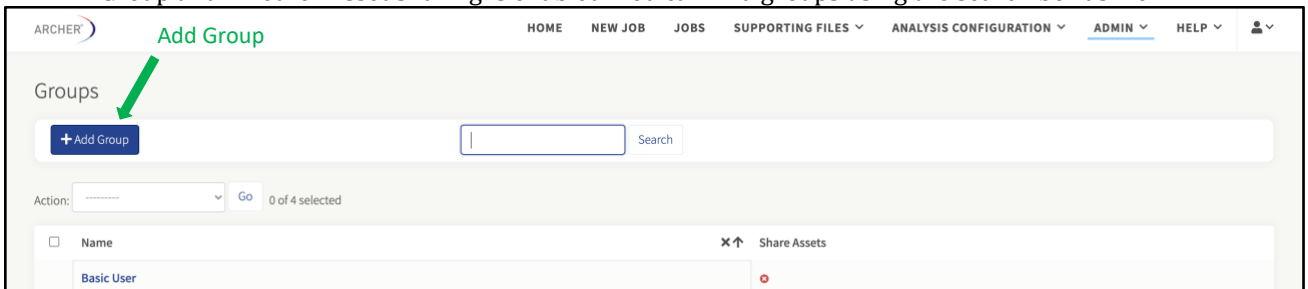
## Groups

Groups serve two purposes: (1) managing groups so many individual permissions do not have to be manually assigned to individual users, and (2) have visibility into Shared Assets.

1. Click **Admin** → **Groups** to manage all groups.



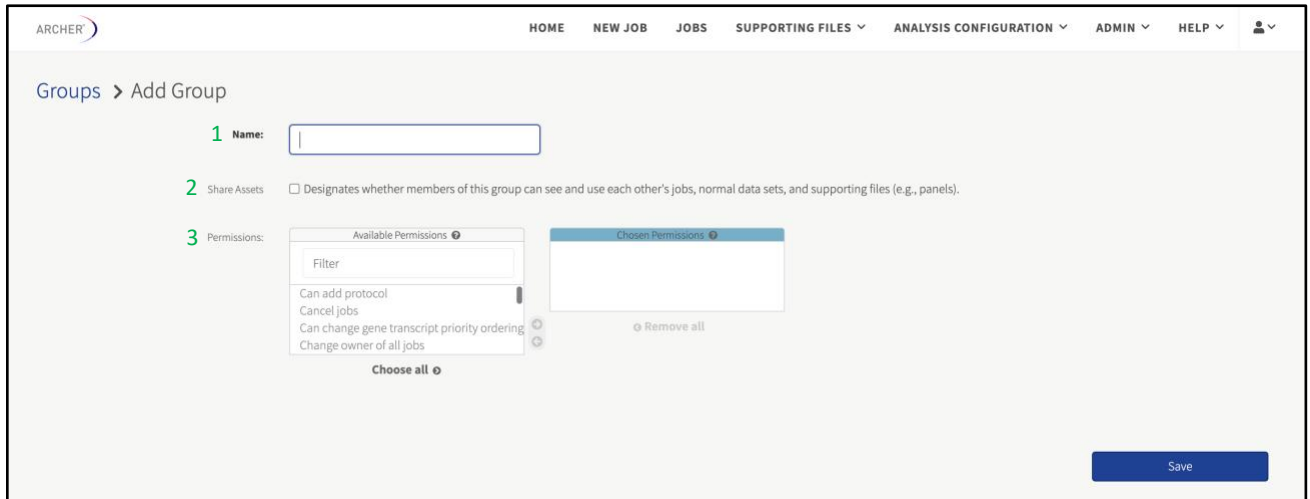
2. The Groups page displays a list of all the groups in the system. The columns indicate the name of the Group and whether Asset Sharing is enabled. You can find groups using the search box as well.



## Add and Edit Groups

1. Click **Admin** → **Groups** and enter a **group name**.





2. If asset sharing should be enabled so group members can see each other's jobs, select the **Share Assets** checkbox.
3. Select any additional permissions that should be enabled for the group.
4. Enter the **required fields** and click **Save**.

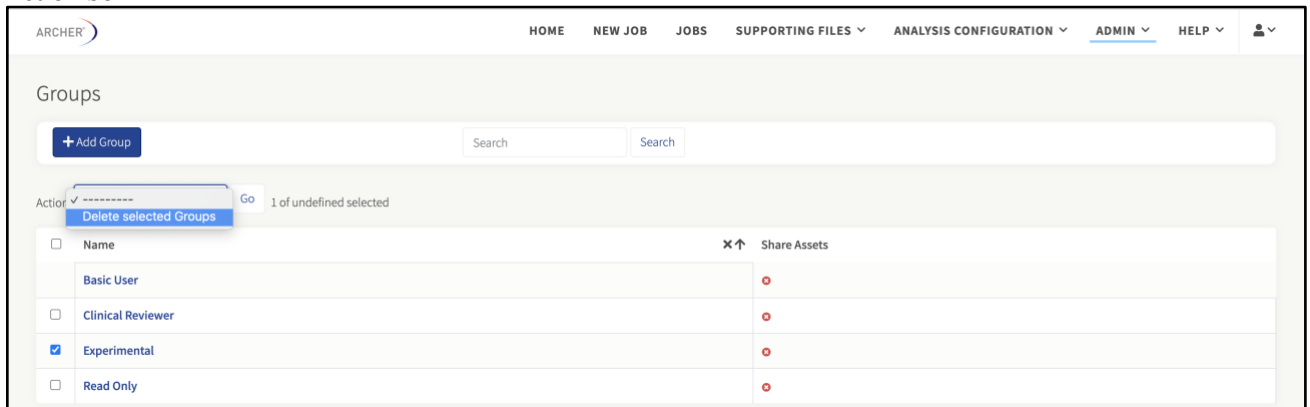


If Group Sharing is enabled, all members of the group will be able to see **all** jobs belonging to every member – including those intended for another group. Best practice is to only assign users to multiple groups if they need to view those jobs and not own any themselves. If a user is an admin, they can run jobs and transfer ownership to other users.

Existing groups can be edited by clicking the **group name**, which brings you to the Change Group page. The options on this page are identical to that of the Add Group page (see above).

### Delete Groups

Groups can be deleted from the Groups page by selecting the **checkbox next to each group name** and using the Action box.



<input type="checkbox"/>	Name	Share Assets
<input type="checkbox"/>	Basic User	<input type="radio"/>
<input type="checkbox"/>	Clinical Reviewer	<input type="radio"/>
<input checked="" type="checkbox"/>	Experimental	<input type="radio"/>
<input type="checkbox"/>	Read Only	<input type="radio"/>

## Manage Supporting Files (GTF, VCF, and BED) Files

There are three types of files specified during the set-up of an Archer Analysis job:

- Panel file in Gene Transfer Format (GTF) – REQUIRED
- Targeted Mutations file in Variant Call Format (VCF, v4.1) – OPTIONAL
- Regions of Interest file in BED Format – OPTIONAL

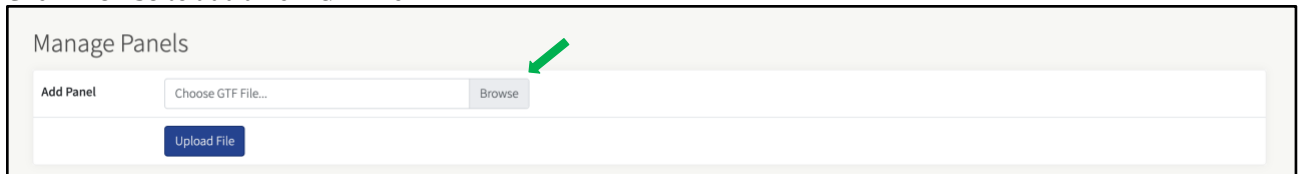
These files must be uploaded and stored in Archer Analysis before being used in a job. (Several are already packaged with Archer Analysis for catalog panels.) These files can be owned by the system or by individual users and in turn, shared with all groups to which the user belongs. By default, new files are owned by the user who uploaded them, regardless of permission level.

## Add and Manage Files

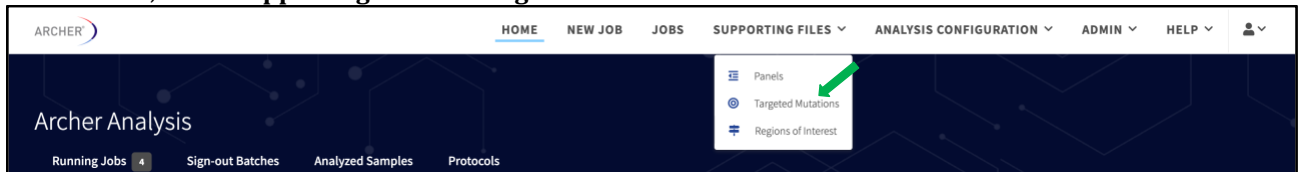
For GTF files, select **Supporting Files** → **Panels**. The Manage Panels page will open where users can add new GTF files.



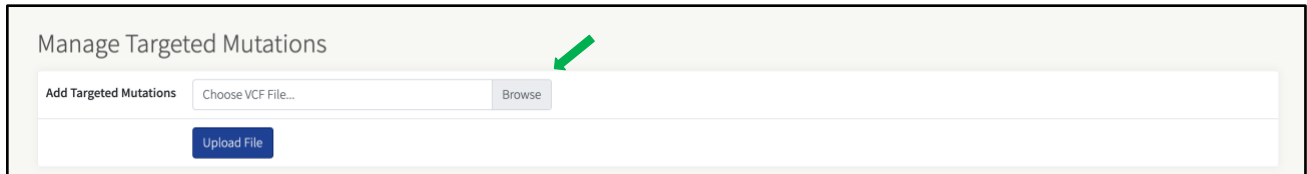
Click **Browse** to add a new GTF file.



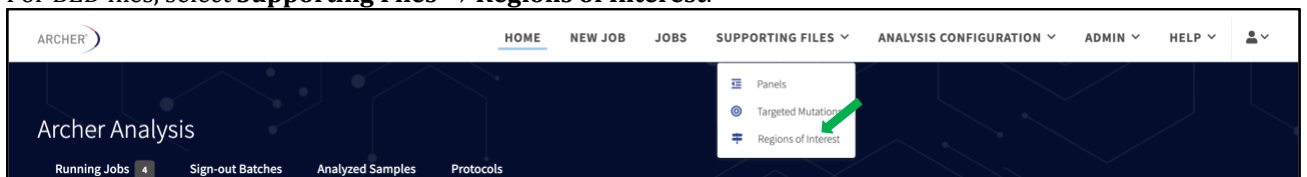
For VCF files, select **Supporting Files** → **Targeted Mutations**.



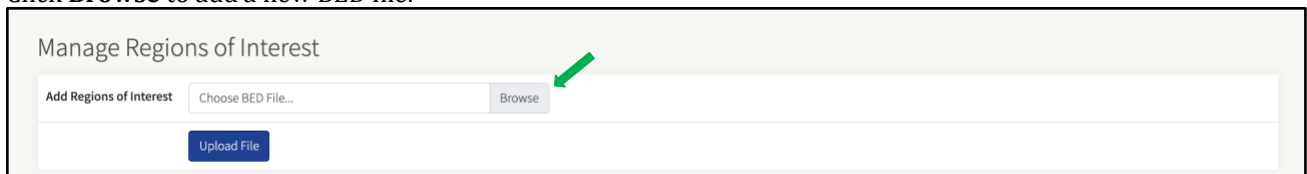
Click **Browse** to add a new VCF file.



For BED files, select **Supporting Files** → **Regions of Interest**.



Click **Browse** to add a new BED file.



## File Actions (View, Download, Visibility, and Ownership)

Use the **more actions** dropdown for options available under the Name column of any supporting files page to (1) view the file within the browser, (2) download a copy of the file, (3) edit the visibility or (4) change ownership:

## Delete Files

If a file does not have any jobs or protocols with which it's associated, it can be deleted.

Click the **Delete** option in the corresponding menu.

In these tables, users can view existing file names, visibility, owner, etc. Usage, as shown below, indicates how many jobs have used the file and if it's part of a protocol workflow.

### Existing Panels

Existing Panels						
is:valid <input type="text"/> <input type="button" value="Q"/> <input type="button" value="x"/>						
Name	Visibility	Usage	Owner	Group Sharing		
cGSP5026 v1.1		12 jobs 5 protocols		N/A		
<ul style="list-style-type: none"> <li> View Panel</li> <li> Download Panel</li> <li> Change Visibility</li> <li> Change Owner</li> <li> Delete</li> </ul>		1 job 3 protocols		N/A		
	amo ltd 18895 v1.0		15 jobs 3 protocols		N/A	
			3 jobs 3 protocols		N/A	
					N/A	
					N/A	

### Existing Targeted Mutations






Existing Targeted Mutations						
is:valid <input type="text"/> <input type="button" value="Q"/> <input type="button" value="x"/>						
Name	Visibility	Usage	Owner	Group Sharing		
AK0079 v1.0 comprehensive27DEC16				N/A		
<ul style="list-style-type: none"> <li> View Targeted Mutations</li> <li> Download Targeted Mutations</li> <li> Change Visibility</li> <li> Change Owner</li> <li> Delete</li> </ul>	IIH v1.3.1	2 jobs		N/A		
	IIH v1.3.1	1 job		N/A		
	IIH v1.3.1			N/A		
	1.0		1 job		N/A	
					N/A	

### Existing Regions of Interest

Existing Regions of Interest						
is:valid <input type="text"/> <input type="button" value="Q"/> <input type="button" value="x"/>						
Name	Visibility	Usage	Owner	Group Sharing		
17627 VariantPlex Expanded Carrier Panel primers targets v1.0				N/A		
<ul style="list-style-type: none"> <li> View Regions of Interest</li> <li> Download Regions of Interest</li> <li> Change Visibility</li> <li> Change Owner</li> <li> Delete</li> </ul>	ocus v2 v10 primers targets v1.0	1 job		N/A		
	mor v10 primers targets v1.0			N/A		
	lid Tumor v10 primers targets v1.0		1 job		N/A	
	nor v10 primers targets v1.0		1 job		N/A	
					N/A	

## Change Visibility

Selecting the **Change Visibility** option in the menu will bring up a modal to edit visibility for all supporting files.

Select	Symbol	Name	Description
		Owner	Only the owner will see the file.
Select		Group	All users who belong to any of the owner's groups that have sharing enabled will see the file.
Select		Global	All users on the system will be able to see the file.
Select		Global / Owned by System	All users on the system will be able to see the file and the file will not be associated to any particular user.
Select		Hidden	No one (including the owner) will be able to see the file when performing new analyses.

### Change Ownership




Any user with the Manage all Supporting Files permission may change a supporting file's ownership. View the *Managing Users and Groups* section for additional information.

Selecting the **Change Owner** option in the menu will bring up a modal to change the owner for this file. Enter the **new owner's email** into the modal's input field.

Existing Regions of Interest

is:invalid

Name	Visibility	Usage	Owner	Group Sharing
17627 VariantPlex Expanded Carrier Panel primers targets v1.0				N/A
View Regions of Interest				
Download Regions of Interest				
Change Visibility				
Change Owner				
Delete				

**Change File Ownership** ✕

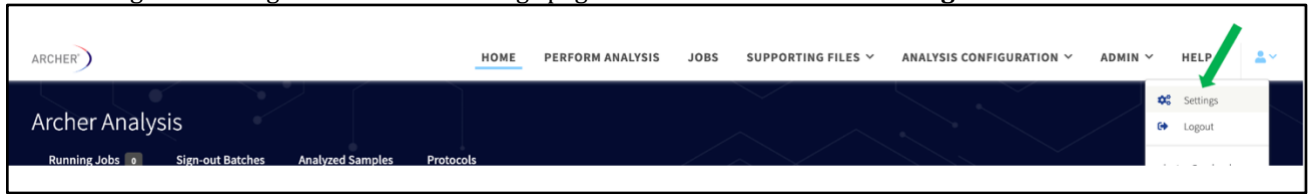
New owner:

## User Settings (Applied Across All Jobs)

User settings apply to every future job started by a user. These settings control processes that are run for all samples (i.e., read cleaning, normalization) as well as for individual processing pipelines (structural variant, SNP/InDel, etc.). The default values have been carefully selected and should not be changed without extensive experience with the software or consulting an Invitae Oncology bioinformatics specialist. Please contact your field team, or [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com).

## View and Manage User Settings

User settings are managed on the User Settings page. Click the **User icon** → **Settings**.



User settings will only affect future jobs started by the current user. Changing settings will not affect prior jobs. If an admin changes its settings, it will not affect any other user in the system.

## General Analysis Settings

The first section in the User Settings page contains parameters include adjustments to sample QC thresholds and to read pre-processing behavior:



### Min Average Unique DNA and AMBIG Start Sites per GSP2

This is the key sample QC metric for the SNP/InDel pipeline identifying the pass/fail criteria. It is defined as the minimum required average number of unique starts sites (**NOT** unique fragments) found across all GSP2s in the panel. The unique start sites are only from reads that have been identified as DNA or ambiguous, and do not include RNA reads. View [RNA vs. DNA vs. Ambiguous Reads](#) for further details on how reads are classified as RNA/DNA/ambiguous.

**The default threshold is 50.** This value is based on an in-house study aiming to achieve 90-95% sensitivity over a range of variants.



While the default value is a great starting point, it is highly recommended a custom calibration be done at your own lab, to determine the optimum QC threshold reflecting the specific requirements of your lab.

### Min Average Unique RNA Start Sites per GSP2 Controls

This is the key sample QC metric for the Fusion/Isoform pipeline identifying the pass/fail criteria. It is defined as the minimum required average number of unique start sites (note, **NOT** unique fragments) found across the control GSP2s in the panel. Control GSP2s in FusionPlex assays target genes that reliably express across any tissue type in the body. The unique start sites are only those from reads that have been identified as RNA reads. View *Key Assay Concepts* for further explanation of unique start sites, control GSP2s, and how reads are classified as RNA/DNA/ambiguous.

**The default threshold is 10.** This value is based on an in-house study aiming to achieve 90-95% sensitivity over a range of fusions.



While the default value is a great starting point, it is highly recommended a custom calibration is done at your own lab, to determine the optimum QC threshold reflecting the specific requirements of your lab.

### Error Correction

This setting turns on/off the use of Molecular Barcode (MBC) based error correction. All reads associated with a single MBC are compared and condensed into a single consensus read, using the base call quality scores from primary analysis, and listed in the FASTQ.



Error correction is On by default. If false positives are of concern in your data, it is highly recommended this setting remains On.

### Read Depth Normalization

All samples in all jobs (regardless of pipeline(s) run) are randomly subsampled to the number of reads specified in reads per sample analysis requires longer processing time and more RAM. "0" represents NO this setting. More subsampling and all reads in the Fastq file will be used. 3,500,000 is the default setting. Note, this metric setting's default is 10,000,000 cfDNA is specified as the assay's input material. View cfDNA Settings for additional information.

### Barcode Length

The number of nucleotides in molecular barcodes.

### Common Region Set

This setting specifies which common region set to use. The default setting is to dynamically detect the correct set. Other settings include:

- Standard Illumina/Ion Torrent Set
- Illumina Expanded Set Option 2
- Legacy Ion Torrent Set
- Illumina Expanded Set Option 1

### Pre-/Post-Alignment Read Cleaner

Turn on pre-/post-alignment read cleaner, to speed up read cleaning. Default setting is Off.

### Merge Reads in Read Cleaner

Merge reads for fusion calling in read cleaner instead of using pmerge. Default setting is Off.

### Debug

If debug is turned on, all temporary files will be kept within the job directory for debugging purposes. If turned off, temp files will be removed after the job is completed. Debug is Off by default.

## Structural Variant Analysis Settings

This section of the User Settings page contains parameters that affect only the Structural Variant pipeline:

Structural Variation Analysis Settings	
Display Intronic Structural Variations	On
Min Reads for Valid Structural Variation	5
Cross-Contamination Single Panel	On
Cross-Contamination Fusion Confidence Threshold	0.05
Expression Imbalance	On
Min Unique Start Sites for Valid Structural Variation	3
Structural Variation Percent of GSP2 Reads	10.0
Expression Imbalance Alpha	0.05
Call Transcriptional Read Through Events Weak	Off
RNA Expression Visualization	On
Min RNA Relative Expression Value	-6
Max RNA Relative Expression Value	6
RNA Expression Outliers	Off
Min Reportable RNA WT Percent of Reads	0.0

### Display Intronic Structural Variant

This setting configures the pipeline to display and report intronic structural variants. Intronic structural variants are defined as those where an entire leg of the variant aligns solely to an intronic region.

### Min Reads for Valid Structural Variant

This setting specifies the minimal number of breakpoint-spanning unique fragments (molecular bins) required to support a structural variant. Any structural variant not meeting this threshold will not be evaluated further by the pipeline, and thus will not be displayed in the Graphical User Interface (GUI) or other file outputs.

### Cross-Contamination Single Panel

This setting toggles the algorithm ON/OFF that can detect cross-contamination between samples run with the same panel and run in the same job.

Cross-contamination is determined using a probability model, the basis of which is that duplicate fusions called across multiple samples (but within a single job) would also feature the same underlying read metrics. A scoring system incorporates the number of unique start sites, unique reads, raw reads, as well as average read depth (which is just the number of raw reads divided by the number of unique reads) for each fusion call. Each of these metrics is assigned a weight used to produce a final score (p-value). If that score surpasses a pre-determined threshold (empirically determined by training the model on both in-silico and known contamination samples), then cross-contamination is flagged.



### Cross-Contamination Fusion Confidence Threshold

This setting specifies the cross-contamination score (p-value) metric below which a Fusion Candidate will be flagged as likely contamination, and thus downgraded to the Weak Evidence bin.

### Expression Imbalance

This setting toggles ON/OFF the detection of expression imbalance across transcripts found from a given gene.

### Min Unique Start Sites for Valid Structural Variant

This setting specifies the minimal number of breakpoint-spanning fragments with unique start sites (a subset of unique fragments) required to support a structural variant. Any structural variant that does not meet this threshold will not be evaluated further by the pipeline, and thus will not be displayed in the GUI or other file outputs.


### Structural Variant Percent of GSP2 Reads

This setting specifies the minimal percentage of unique fragments from the GSP2s associated with a structural variant that must be met in order to be classified as Strong Evidence. Otherwise, the structural variant will be placed in the Weak Evidence bin.

### Expression Imbalance Alpha

This setting specifies the significance threshold for determining if a gene has expression imbalance due to the partial RNA expression of a fusion candidate. If below this threshold, the fusion candidate will feature an Expression Imbalance icon (⚠️).

### Call Transcriptional Read Through Events Weak

This setting toggles the rule that, if a fusion candidate has been identified as a transcriptional read through event (as denoted by the  icon), regardless of any other criteria, the fusion candidate will be placed in the Weak Evidence bin. Transcriptional readthrough is defined as a fusion between two adjacent genes on the same strand, occurring in that same natural order.

### RNA Expression Visualization

This setting turns ON/OFF the generation and presentation of a heat map visualization of the relative RNA expression for all samples in the run.

### Min RNA Relative Expression Value

This is the minimum value used in RNA expression visualization. Any value below this limit will be truncated.

### Max RNA Relative Expression Value

This is the maximum value used in RNA expression visualization. Any value above this limit will be truncated.

### RNA Expression Outliers


This setting turns ON/OFF the detection of abnormally high RNA expression levels.

### Min Reportable RNA WT Percent of Reads

The minimum reportable percentage of reads for WT RNA candidates.

## Variations/Mutations Settings

This section of the User Settings page contains parameters affecting only the SNP/InDel pipeline:

Variations/Mutations Settings	
<u>Min Depth for Variant Call</u>	<input type="text" value="2"/>
 <u>MAPQ Threshold for Variant Call</u>	<input type="text" value="0"/>
 <u>Min Basequal for Variant Call</u>	<input type="text" value="20"/>
<u>Min Allele Fraction for Variant Call</u>	<input type="text" value="0.001"/>
<u>Min Phred Qual Score for Variant Call</u>	<input type="text" value="1"/>
 <u>Variant Downstream ROI Size</u>	<input type="text" value="400"/>
<u>Outlier Detection</u>	<input type="text" value="On"/> 
<u>Min DAO for Outlier Detection Calculation</u>	<input type="text" value="3"/>
 <u>Deep Shallow Threshold</u>	<input type="text" value="5"/>

### Min Depth for Variant Call

This setting specifies the minimum number of alternate observations (AO) supporting an allele in order to qualify as a potential variant (SNV or InDel) call. Any potential variants with AO below this threshold will not be reported by Analysis.

**Note: this setting DOES NOT apply to Targeted Mutation calling (Vision).**

### MAPQ Threshold for Variant Call

This setting specifies the minimum MAPQ (Phred-scale) score of any AO supporting an allele in order to qualify as a potential variant (SNV or InDel) call. Any supporting reads with MAPQ below this threshold will not be used by Analysis to support a variant call.

It is recommended to leave this at the default value of 0 as other mechanisms of the SNP/InDel pipeline can be used to adequately screen for alignment and general call quality.

### Min Basequal for Variant Call

This setting specifies the minimum base call quality (Phred-scale) score of any AO supporting an allele in order to qualify as a potential variant (SNV or InDel) call. Any supporting AO with base quality below this threshold will not be used by Analysis to support a variant call.

### Min Allele Fraction for Variant Call

This setting specifies the minimum allele fraction (AF = AO/DP) supporting an allele in order to qualify as a potential variant (SNV or InDel) call. Any potential variants with AF below this threshold will not be reported by Analysis.

**Note: this setting DOES NOT apply to Targeted Mutation calling (Vision).**

### Min Phred Qual Score for Variant Call

This setting specifies the minimum variant call score (as reported by Freebayes or Lofreq) required to report the variant in the SNP/InDel pipeline. Any potential variants with a quality score below this threshold will not be reported by Analysis.

**Note: this setting DOES NOT apply to Targeted Mutation calling (Vision).**

### Variant Downstream ROI Size

This setting specifies the number of bases downstream of the targeted GSP2s that are considered when looking for variants. Only applies when not already specified by the Target Region GTF file of the Archer assay. Targeted Mutations must be within the specified region of interest (ROI) to be reported.

### Outlier Detection

When enabled intra-job samples will be used to calculate the sequencing noise at a given position. The calculation will set the confidence in low AF variant calls.

### Min DAO for Outlier Detection Calculation

The minimum number of Deep Alternate Observations on a variant required to perform Allele Fraction Outlier P-value calculations. Decreasing this parameter below 3 will increase analysis runtime.

### Deep Shallow Threshold

The depth at which a molecular bin will be considered a 'deep' bin. Deep bins are considered higher confident reads as they have been assembled from multiple PCR duplicates.

## CNV Pipeline Settings

This section of the User Settings page contains parameters that only affect the CNV pipeline in Analysis.

CNV Settings	
CNV Strong Amplification Threshold	3.0
CNV Strong Deletion Threshold	0.33333
CNV P-Value Threshold	0.01

### CNV Strong Amplification Threshold

This setting specifies the X-fold increase in Copy Number (CN) relative to the calculated baseline for the job, to be categorized as Strong Evidence Copy Gain.

### CNV Strong Deletion Threshold

This setting specifies the X-fold decrease in CN relative to the calculated baseline for the job, to be categorized as Strong Evidence Copy Loss.

### CNV P-Value Threshold

In addition to the settings above, this setting specifies the confidence threshold (p-value), which a CN must be at or below, to be categorized as a Strong Evidence Copy Gain or Loss.

## MSI Pipeline Settings

This section of the User Settings page contains parameters that only affect the CNV pipeline in Analysis.

MSI Settings	
Minimum number of loci for MSI	25
Minimum Depth Per Sample for MSI	35000
Stable Threshold for MSI	20
High Threshold for MSI	30
MSI Alpha	0.05
Minimum Depth Per Locus	25

### Minimum number of loci for MSI

The minimum number of targeted microsatellite loci that must pass QC for a sample to get a percent unstable value and an MSI status call.

### Minimum Depth Per Sample for MSI

The minimum total useable raw reads depth a sample must have across all microsatellites for microsatellite instability to be determined. Minimum read depth required for the sample to pass QC).

### Stable Threshold for MSI

The maximum percentage of unstable loci for a sample to be considered MSI-Stable. Unstable loci percentages above this, and less than MSI\_HIGH\_THRESHOLD, will be considered MSI-Intermediate.

### High Threshold for MSI

The minimum percentage of unstable loci for a sample to be considered MSI-High. Unstable loci percentages less than this, and above MSI\_STABLE\_THRESHOLD, will be considered MSI-Intermediate.

**Note:** The high and stable threshold can be set to equal each other, thereby removing the MSI-Intermediate call.

### MSI Alpha

The maximum pval for a locus to be considered unstable.

### Minimum Depth per Locus

The minimum read depth a microsatellite locus must have to generate a locus instability score.

## Miscellaneous System Settings

This section of the User Settings page contains parameters that either do not affect data analysis, or do not fit in the other page sections described above:

Misc Settings	
Items Per Page	<input type="text" value="25"/>
Landing Page	<input type="text" value="Running Jobs"/>
Automatically Expand Aggregate Isoform Visualizations	<input type="text" value="Off"/>

### Items Per Page

This setting specifies the number of items that will display per page on paginated pages (e.g., Jobs, Protocols, Panels, etc.).

### Landing “Home” Page

This setting specifies the Home page of Analysis. The default is the Running Jobs page as illustrated in the Home Page section above. However, this can also be changed to the New Job or Jobs page, or the Sign-out Batches, Protocols, or the Analyzed Samples tab of the Home page.

### Automatically Expand Aggregate Isoform Visualizations

This setting specifies whether to automatically expand all aggregate visualizations when loading the Fusion or Structural Variation grids.

## Save New User Settings or Restore Defaults

After changes are made on the User Settings page, click **Update Settings** at the bottom of the page. If the goal is to restore all original default settings in the system, as captured in the sections above, instead click **Reset to Defaults**:

<input type="button" value="Reset to Defaults"/>	<input type="button" value="Update Settings"/>
--	--

## Identify Non-default Settings

The User Settings page makes it simple to determine which setting(s) have been changed from their default value(s) by highlighting those parameters in yellow and including the default value.

<input type="checkbox"/> Min Basequal for Variant Call		<input type="text" value="20"/>
Min Allele Fraction for Variant Call	(default: 0.001)	<input type="text" value="0.0008"/>
Min Phred Qual Score for Variant Call		<input type="text" value="1"/>

## Job Setup and Management

Jobs are fundamental to the Archer Analysis workflow. All samples must be processed as a part of a job, even if processed individually. All sample data is organized according to the job ID which is auto-assigned. Each job processes samples for a single Archer assay, but can include any number of samples, as well as run multiple pipelines in parallel (so long as they are compatible with the assay).

## Starting a Job: New Job Page

Click on “New Job” in the Main Menu in order to set up a new job; this will load the New Job page.



The New Job process is separated into six sections. The software will validate each step when the ‘Next’ button is clicked. After each validation, the next section of the flow will display. All fields are required unless indicated.

### Section 1: Set Up

Input a Job Name and select the desired Panel (GTF file) to run with this job. Custom panels require uploading of a dedicated GTF file, as illustrated in **Managing and Adding GTF Files**.



If the selected GTF file does not correspond to the Archer panel used to process the sample, there will be no automated flags from the system. However, analysis failure may be evident via the “On Target %” metric found in the Read Statistics tab of the **Sample Details** page. See [Error! Reference source not found.](#) for further details.

## New Job

### 1. Set Up

ENTER JOB NAME

SELECT PANEL

Select
▼

### Choosing the pipeline(s)

Depending on the type of Archer assay, a job may run multiple pipelines (a.k.a. Analysis Types). Only certain pipelines can be run together:

## Section 2: Pipeline

When this section loads, choose the Sequencing Platform (Illumina or Ion Torrent) and Input Material. The input material chosen will determine which pipeline options are available. Certain pipelines can be run together, and depending on which are chosen, additional setup steps may be required. In some cases, the Analysis Type Option Sets show which pipelines can and cannot be combined.



**2. Pipeline**

SELECT SEQUENCING PLATFORM

Illumina     Ion Torrent

SELECT INPUT MATERIAL

RNA     DNA     cDNA

### Common Analysis Types

There is a base pipeline available to both RNA/DNA input material:

**Read Processing Only** - Select this to only run the preprocessing steps (read cleaning, deduplication) on the input data, with no additional analysis. Generating read alignments along with the preprocessing steps is a configuration option.

### RNA Analysis Types

There are four pipelines for Archer RNA assays:

- **Fusion** – select this to run the Fusion/Isoform pipeline for FusionPlex assays.
- **SNP/InDel** – select this to run the Variants pipeline for FusionPlex assays. If selected, an additional field will display the option to select a Targeted Mutation File (TMF). Further, if a TMF is selected, another option will display to toggle Non-Targeted calling ON/OFF as well. Non-Targeted calling is what would run if no TMF were selected. See the Before Getting Started section of this manual for more information on TMFs. Note that RNA SNV/InDel mutation detection is not supported on the Ion Torrent Sequencing Platform.
- **Lymphoma** – select this to run the experimental pipeline to identify the cell of origin subtype of a diffuse large B-cell lymphoma (DLBCL) sample.
- **Immune Repertoire** – select this to run the immune profiling pipeline for Immuniverse assays. No other pipelines can be run in the same job (see [Error! Reference source not found.](#) for further details). This analysis type is only available for samples sequenced on the Illumina (TM) platform.

Note that the RNA Expression pipeline will run for all FusionPlex jobs for which the User > Settings “RNA Expression Visualization” is set to “On” (see the screenshot below). The range of “Relative Expression (Log Base 2)” values displayed in the heatmap visualizations are adjustable via the User > Settings page or when creating a New Job. These settings are available in the third step of the New Job flow.

RNA Expression Visualization	On
Min RNA Relative Expression Value	-6
Max RNA Relative Expression Value	6
RNA Expression Outliers	Off
Min Reportable RNA WT Percent of Reads	0.0

## 2. Pipeline

SELECT SEQUENCING PLATFORM

Illumina    Ion Torrent

SELECT INPUT MATERIAL

RNA    DNA    cfDNA

SELECT ANALYSIS TYPE

**Option Set A**

Fusion    SNP/InDel    Lymphoma

**Option Set B**

Immune Repertoire

**Option Set C**

Read Processing Only

TARGETED MUTATIONS

Select ▼

Optional

*Available pipeline settings for FusionPlex assays*



RNA Fusion and RNA SNP/InDel analysis can be run in the same job. Both Germline and Somatic SNP/InDel detection are performed for RNA panels – see the **Error! Reference source not found.** section below for more details on the specific SNP/InDel Detection Types.



## 2. Pipeline

SELECT SEQUENCING PLATFORM

Illumina     Ion Torrent

SELECT INPUT MATERIAL

RNA     DNA     cfDNA

SELECT ANALYSIS TYPE

**Option Set A**

Fusion     SNP/InDel     Lymphoma

**Option Set B**

Immune Repertoire

**Option Set C**

Read Processing Only

IGHV HYPERMUTATION

On     Off

*Available pipeline settings for Immunoverse assays*

### DNA and cfDNA Analysis Types

There are five pipelines for Archer DNA and cfDNA assays (VariantPlex and LiquidPlex, respectively). All can be run in the same job:

- **SNP/InDel** – select this to run the Variants pipeline. If selected, an additional field will display to select the Detection Type (Somatic, Germline, or cfDNA – see DNA SNP/InDel Detection Types below for further details). Another field will also display for the option to select a Targeted Mutation File (TMF). Further, if a TMF is selected, another option will display to toggle Non-Targeted calling ON/OFF as well. Non-Targeted calling is what would run if no TMF were selected. See the [Before Getting Started](#) section of this manual for more information on TMFs.
- **Copy Number Variation** – select this to run the CNV pipeline. A further step is required to identify normalization parameters in the Sequenced Libraries section, as detailed in [Error! Reference source not found.](#)
- **Structural Variations** – select this to run the Structural Variations pipeline, to detect large deletions, exon skipping events, internal tandem duplications (ITDs), and other structural rearrangements. This pipeline is analogous to the RNA Fusion/Isoforms pipeline and utilizes de novo assembly.
- **MSI** – select this to run the Microsatellite Instability pipeline. If selected, an additional field will display to select the Illumina instrument on which the sample(s) were sequenced. **Note:** The MSI pipeline is unsupported with cfDNA inputs.
- **Target Coverage** – select this to run the experimental feature to report coverage metrics (i.e., depth at each base) for target regions of interest. If chosen, then an additional field will become visible for selecting a QC Target Regions BED file.

### Select DNA SNP Detection

#### Germline Variation

Select this to detect only variants at relatively high allele fraction (AF). The germline variant detection algorithm will not reliably detect variants below AF of ~0.20. Also, the Germline filter set will be applied by

default to the Variant Summary Grid, although this can then be changed at any time after the job is processed (see Filtering Results for further details).

### **Somatic Variation**

Select this to detect variants at any allele fraction (AF). The Somatic filter set will be applied by default to the Variant Summary Grid, although this can then be changed at any time after the job is processed (see Filtering Results for further details).



**For LiquidPlex ctDNA panels:** When cfDNA is selected as input material, job settings will use adjusted defaults designed to improve sensitivity for very low allele fraction (AF) mutations in cell-free DNA from liquid biopsies. These adjusted settings are: (1) read depth normalization is set to 10 million, (2) Variant Downstream ROI Size is set to 150bp, and (3) Error Correction is turned ON. These settings may be adjusted in the cfDNA section of the Settings step. Additionally, the cfDNA filter set is applied by default in the Variant Grid (see Filtering Results for further details).



The Vision caller is always used for Targeted Mutations (i.e., if a TMF is selected) regardless of the Detection Type selected.

## **Supporting Files and Normal Data Sets**

Some pipeline configurations may be further refined through the use of Supporting Files and Normal Data Sets:

### **Targeted Mutations**

As explained in Before Getting Started, this option directs the pipeline to report on a set of mutations of interest regardless of whether they were detected in the sample.

### **Regions of Interest**

As explained in Before Getting Started, this option directs the pipeline to report coverage metrics for specified locations.

### **Variant Normal Data Set**

If a Variant Normal Data Set has been established for the chosen Panel, it can be enabled here. Please refer to the “Normal Data Sets” section for more information.

### **CNV Normal Data Set**

If a CNV Normal Data Set has been established for the chosen Panel, it can be enabled here. Please refer to the “Normal Data Sets” section for more information.

## **Instrument**

The MSI pipeline requires selection of the Illumina instrument that was used to sequence the sample. Only the following Illumina instrument types are supported for MSI analysis:

- MiSeq
- NextSeq 500/550
- NovaSeq

It is important to choose the correct instrument that was used to generate the sequencing data for the pipeline to produce the most accurate results. If multiple samples are selected for a single job, they must all have been sequenced using the same Illumina instrument type.

## 2. Pipeline

SELECT SEQUENCING PLATFORM

Illumina     Ion Torrent

SELECT INPUT MATERIAL

RNA     DNA     cfDNA

SELECT ANALYSIS TYPE

**Option Set A**

SNP/InDel     Copy Number Variation     Structural Variation     MSI     Target Coverage

**Option Set B**

Read Processing Only

TARGETED MUTATIONS

Archer Comprehensive Targets v1.6 ▼

**Optional**

INCLUDE NON-TARGETED VARIANTS

No     Yes

SELECT DNA SNP DETECTION

Somatic Mutation     Germline Variation

VARIANT NORMAL DATA SET

Select ▼

**Optional**

CNV NORMAL DATA SET

Select ▼

**Optional**

REGIONS OF INTEREST

Select ▼

INSTRUMENT

MiSeq     NextSeq 500/550     NovaSeq

*Available pipeline settings for analyzing VariantPlex assays*

## 2. Pipeline

### SELECT SEQUENCING PLATFORM

Illumina
  Ion Torrent

### SELECT INPUT MATERIAL

RNA
  DNA
  cfDNA

### SELECT ANALYSIS TYPE

#### Option Set A

SNP/InDel
  Copy Number Variation
  Structural Variation
  MSI
  Target Coverage

#### Option Set B

Read Processing Only

### TARGETED MUTATIONS

Select

Optional

### VARIANT NORMAL DATA SET

Select

Optional

### CNV NORMAL DATA SET

Select

Optional



Running MSI analysis with cfDNA input is experimental and not fully supported by Archer Analysis.

### REGIONS OF INTEREST

Select

### INSTRUMENT

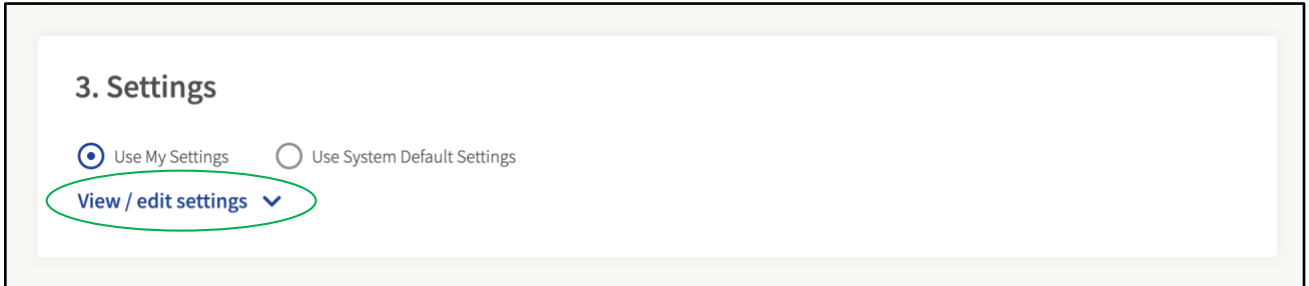
MiSeq
  NextSeq 500/550
  NovaSeq

*Available pipeline settings for analyzing LiquidPlex assay*

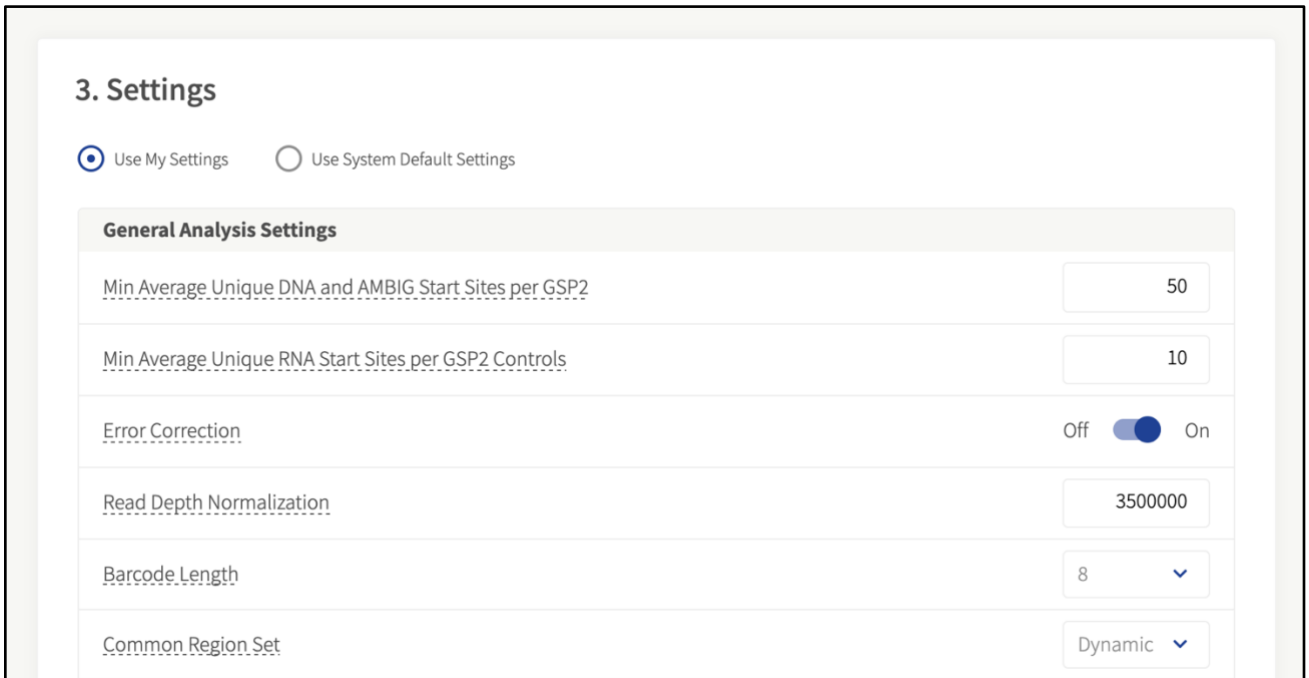
### Section 3: Settings

In this section users have the option of using their personal job settings, using the system defaults, or making adjustments to either of these by changing individual settings. Click the 'View / edit settings' button to access the job's settings.

**Note:** Available settings will change based on what kind of job the user is setting up.

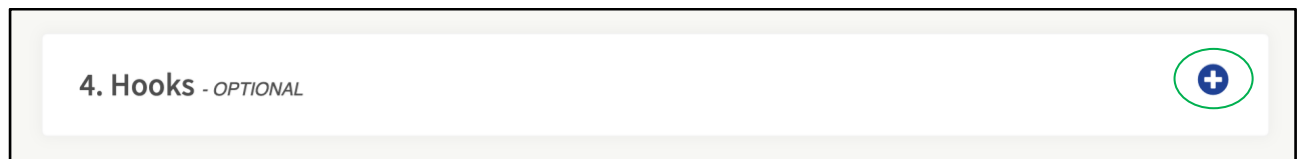


View of the Settings section after the 'View / edit settings' button is clicked. **Note:** This screenshot is shortened for brevity.




### Section 4: Hooks

By default, the Hooks section is minimized. To open this section, click the 'Plus' icon in the right corner. Available hooks will display in the opened box. This step is optional. See the Job Hooks section for more information.



#### 4. Hooks - *OPTIONAL*

 Enabled hooks were set by an administrator on this system. Changing which ones are enabled is for expert users only.

Sample Pre Hooks

Lower Case to Capital in FASTQ

Sample Post Hooks

SampleDoneHook\_test

Job Post Hooks

JobDoneHook

python\_posthook\_summarize\_full\_results

#### Section 5: Batch Sign-out

To facilitate downstream analysis, the laboratory/sequencing **Batch ID** can be added, and results review, and reporting can be **assigned to** any current user in the system. A **Batch due date** for review/reporting can be specified as well. This step is optional.

By default, the Batch Sign-Out section is minimized. To open this section, click the 'Plus' icon in the right corner.

#### 5. Batch Sign-Out - *OPTIONAL*



In this screenshot, users may interact with the input fields as described above.

### 5. Batch Sign-Out - *OPTIONAL*

Batch ID

Assign To


Batch Due Date

## Section 6: Sequenced Libraries

This last section is where the sequence files are uploaded. If the Illumina (TM) platform is chosen during set up, then FASTQ files for all samples to be analyzed must be uploaded. Both R1 and R2 files must be selected for each sample. If the Ion Torrent (TM) platform is chosen during set up, then one BAM file per sample must be uploaded.

### 6. Sequenced Libraries

Upload the sequence files for the libraries to be analyzed.  
Because this job is configured for the Illumina platform, only .fastq and .fastq.gz files may be uploaded.

 [Drag and drop your FASTQ files here or click to browse.](#)



FASTQ or BAM files to be analyzed in a job may derive from different folders on the user's device. Users may also upload different sets of files in multiple 'clicks' / 'drag-and-drops, or 'browse to' actions. The UI will display the selected files in a list to reflect which files are ready for upload.

The files will be copied to and permanently saved in the system. Also, FASTQ files can be either uncompressed (with extension ".fastq") but it is **recommended to use compression** (using the GZIP algorithm and have the extension ".fastq.gz"). ZIP compression is not supported.

Files can be removed from the queue by clicking the red 'x' button in the right side of the row.

## 6. Sequenced Libraries

Upload the sequence files for the libraries to be analyzed.  
 Because this job is configured for the Illumina platform, only .fastq and .fastq.gz files may be uploaded.

 Drag and drop your FASTQ files here or click to browse.

Sample Name	Sequence Files	File Size
AAAS_deletion_0.2_R1	AAAS_deletion_0.2_R1.fastq	223.53KB
	AAAS_deletion_0.2_R2.fastq	223.53KB



### Selecting Normalization Parameters for CNV

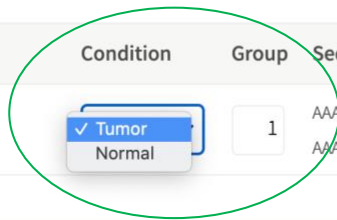
If the CNV pipeline is selected, the Sequenced Libraries section requires specification of how normalization in the algorithm should be done for the job. This is done by selecting either Normal or Tumor from the Condition dropdown. CNV analyses will be more sensitive if a matched Normal sample is available for each Tumor sample (multiple Sample Groups can be identified within each job, if there are multiple Normal-Tumor matches):

## 6. Sequenced Libraries

Upload the sequence files for the libraries to be analyzed.  
 Because this job is configured for the Illumina platform, only .fastq and .fastq.gz files may be uploaded.

 Drag and drop your FASTQ files here or click to browse.

Sample Name	Condition	Group	Sequence Files	File Size
AAAS_deletion_0.2_R1	<input checked="" type="checkbox"/> Tumor	1	AAAS_deletion_0.2_R1.fastq	223.53KB
	<input type="checkbox"/> Normal		AAAS_deletion_0.2_R2.fastq	223.53KB

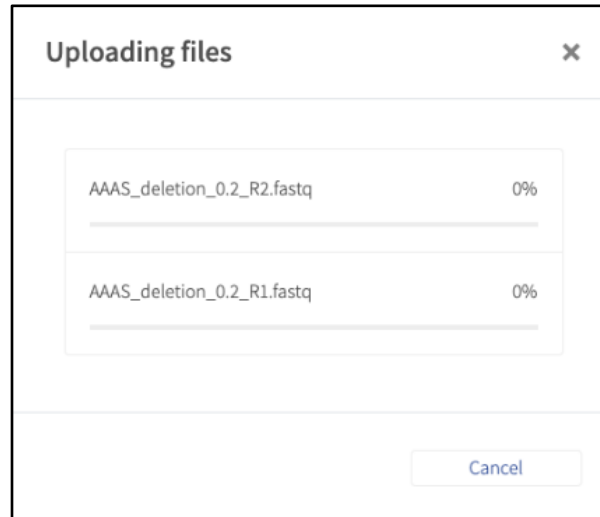






Although recommended for maximum performance, it is not required to identify NORMAL samples in the job. Simply keep all samples labeled as TUMOR, and then all samples in each Group will be used collectively for normalization.

After at least one pair of sample files has been added, the 'Submit Job' button will be clickable. When this button is clicked, the 'Submit Job' function will begin. A progress bar, as shown below, will indicate the upload status.



The speed of the upload depends on the network speed between the client and the Analysis server. If the client and the server are located on the same machine, the upload will progress very quickly, even for many GB of data. If the client is accessing a server 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>. After file upload is complete, the job will be entered into the queue, and processed as soon as system resources are available. At this point, the user can safely log out of the system.

## Running Jobs

Once a job has been submitted and sequencing files uploaded, it will be entered into the job queue, which can be viewed from the **Running Jobs** page (the default **Home** page – to change this, refer to MSI Pipeline Settings). This section of the User Settings page contains parameters that only affect the CNV pipeline in Analysis.

MSI Settings	
Minimum number of loci for MSI	25
Minimum Depth Per Sample for MSI	35000
Stable Threshold for MSI	20
High Threshold for MSI	30
MSI Alpha	0.05
Minimum Depth Per Locus	25

### Minimum number of loci for MSI

The minimum number of targeted microsatellite loci that must pass QC for a sample to get a percent unstable value and an MSI status call.

### Minimum Depth Per Sample for MSI

The minimum total useable raw reads depth a sample must have across all microsatellites for microsatellite instability to be determined. Minimum read depth required for the sample to pass QC).

### Stable Threshold for MSI

The maximum percentage of unstable loci for a sample to be considered MSI-Stable. Unstable loci percentages above this, and less than MSI\_HIGH\_THRESHOLD, will be considered MSI-Intermediate.

### High Threshold for MSI

The minimum percentage of unstable loci for a sample to be considered MSI-High. Unstable loci percentages less than this, and above MSI\_STABLE\_THRESHOLD, will be considered MSI-Intermediate.

**Note:** The high and stable threshold can be set to equal each other, thereby removing the MSI-Intermediate call.

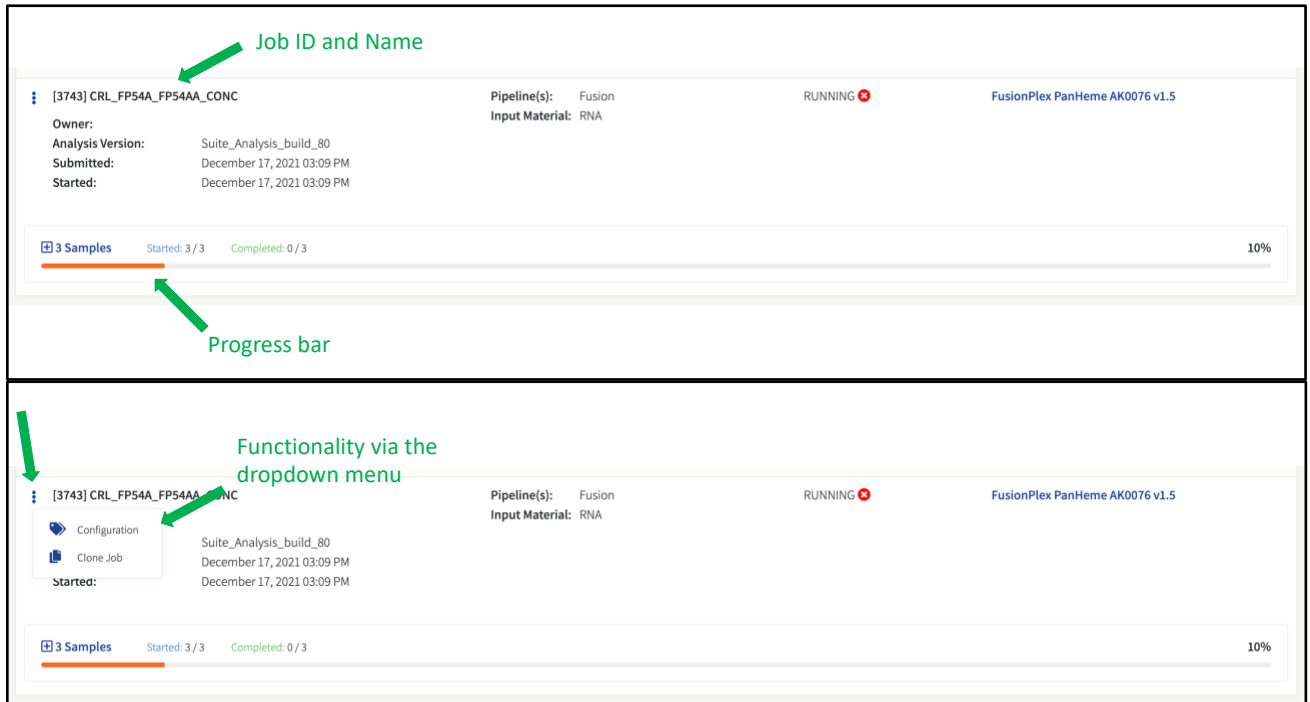
### MSI Alpha

The maximum pval for a locus to be considered unstable.



### Minimum Depth per Locus

The minimum read depth a microsatellite locus must have to generate a locus instability score.

Miscellaneous System Settings):



The screenshot displays a job entry for [3743] CRL\_FP54A\_FP54AA\_CONC. The job is in a RUNNING state, indicated by a red stop sign icon. The pipeline is Fusion, and the input material is RNA. The job was submitted and started on December 17, 2021, at 03:09 PM. The progress bar shows that 3 out of 3 samples have started, and 0 out of 3 have completed, with an overall progress of 10%. A dropdown menu is visible next to the job ID, containing options for Configuration and Clone Job.

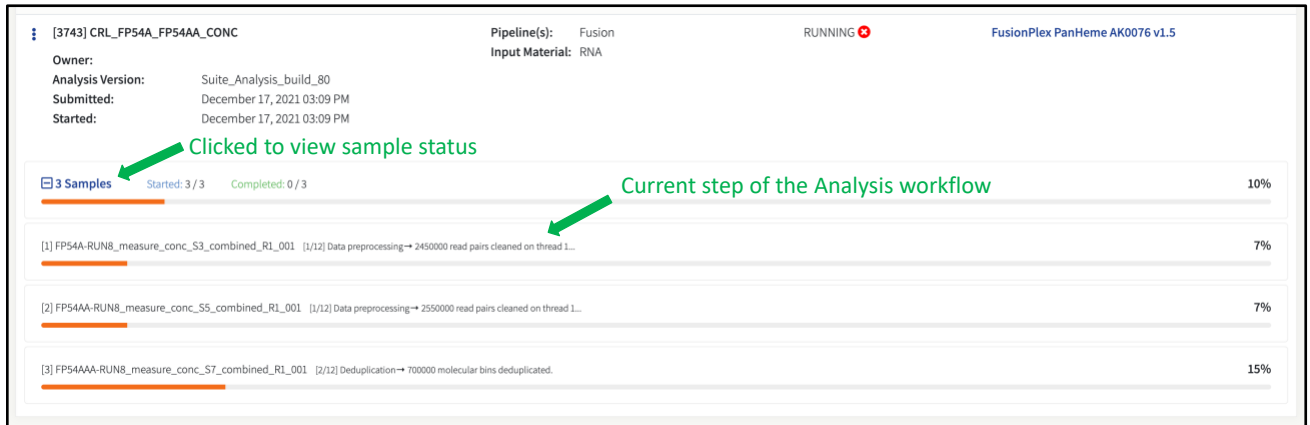
The **Running Jobs** page will display comprehensive information about every job currently in the queue, including Job Status, which can be: “NEW”, “QUEUED”, or “RUNNING”. From this view, the detailed settings for the job can be viewed by clicking on the  Configuration item, and the job can be cloned by clicking on the  Clone Job item. See [Error! Reference source not found.](#) for further details on reanalyzing jobs.



**DO NOT** turn off the system or shutdown the machine hosting the Analysis server; this will disrupt currently running jobs, and most likely result in job failures.

### Progress Bar and Status Updates

The progress bar indicates overall status of the job. This can be expanded to view progress bars for each sample in the job:



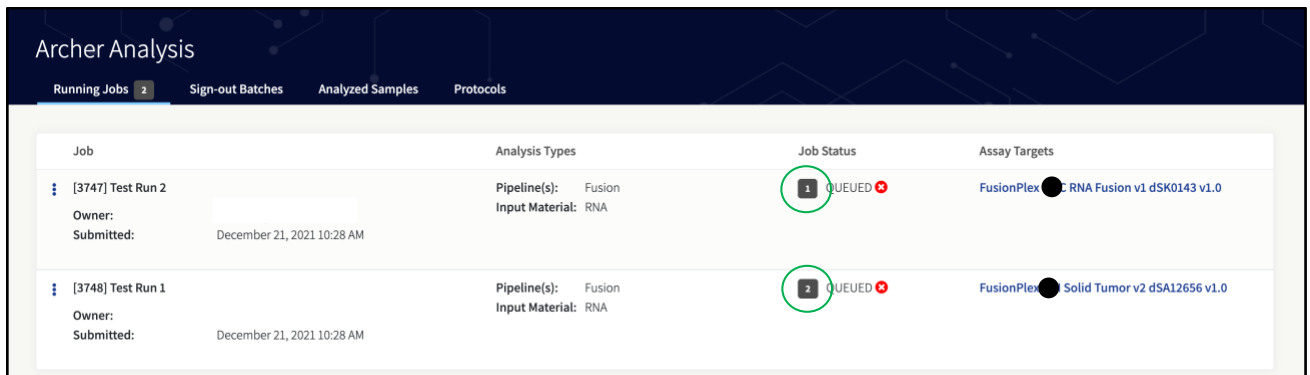
The screenshot shows a job titled "[3743] CRL\_FP54A\_FP54AA\_CONC" with a status of "RUNNING". The pipeline is "Fusion" and the input material is "RNA". The analysis version is "Suite\_Analysis\_build\_80", submitted on December 17, 2021 at 03:09 PM, and started at the same time. Below the job details, there are three sample progress bars. The first sample, "[3] Samples", is 10% complete. The second sample, "[1] FP54A-RUN8\_measure\_conc\_S3\_combined\_R1\_001", is 7% complete and is the current step of the analysis workflow. The third sample, "[2] FP54AA-RUN8\_measure\_conc\_S5\_combined\_R1\_001", is also 7% complete. The fourth sample, "[3] FP54AAA-RUN8\_measure\_conc\_S7\_combined\_R1\_001", is 15% complete. Each sample's progress bar is followed by a list of workflow steps, such as "Data preprocessing" and "Deduplication".

Below each sample's progress bar, the steps of the analysis workflow are indicated. In the example above, "[1/12]" indicates that step 1 of a total of 12 steps is being executed for sample [1].

### Analyzing Multiple Samples & Jobs

The system can queue multiple samples simultaneously, 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. For local virtual machine (VM) deployments, the default configuration is 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 "1" is the next in line to be executed when an available slot for the sample(s) within it becomes available.




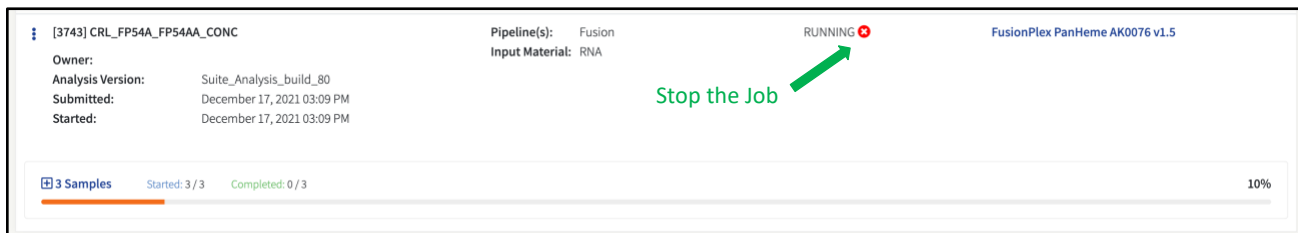
The screenshot shows the "Archer Analysis" interface with a "Running Jobs" tab selected. There are two jobs listed: "[3747] Test Run 2" and "[3748] Test Run 1". Both jobs are in a "QUEUED" status. The first job is the next in line (indicated by a "1" in a circle) and is using the "Fusion" pipeline with "RNA" input material. The second job is also in a "QUEUED" status (indicated by a "2" in a circle) and is using the "Fusion" pipeline with "RNA" input material. The assay targets for the first job are "FusionPlex" and "RNA Fusion v1 dSK0143 v1.0". The assay targets for the second job are "FusionPlex" and "Solid Tumor v2 dSA12656 v1.0".



Note that the VM can be configured to process >1 sample simultaneously if the VM is running on a host with sufficient resources. See the **Archer Analysis Virtual Machine Installation Guide** for further information.

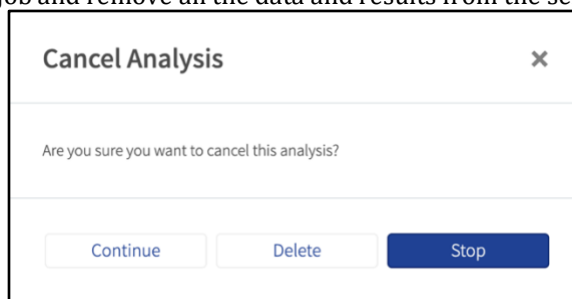
### Stopping a Running or Queued Job

An analysis job that is queued or currently running can be stopped by clicking the  icon:



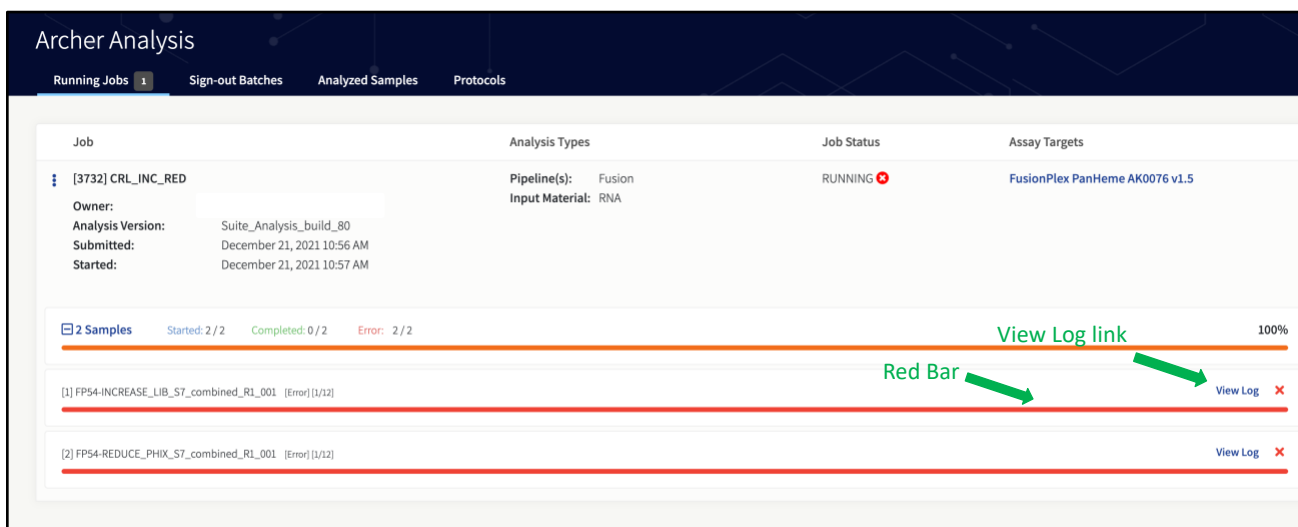
There are two options for canceling a running job:

- **Stop** – will stop the job but leave the sequencing file(s) on the server and allow the samples to be rerun in the future.
- **Delete** – will stop the job and remove all the data and results from the server.



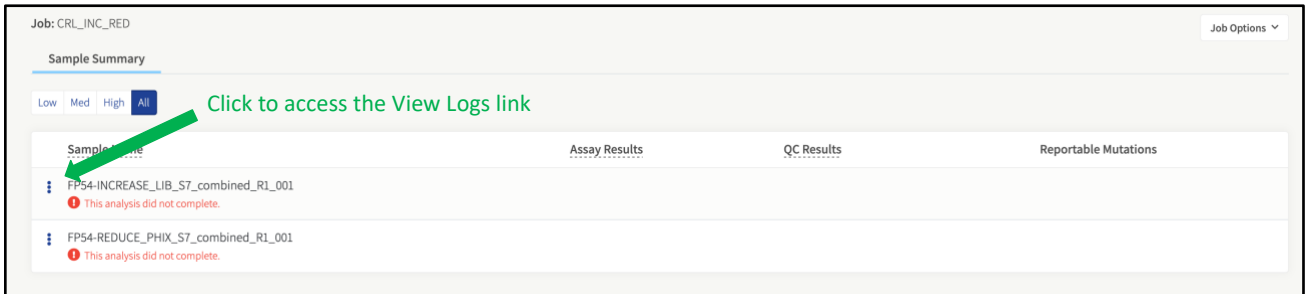
### Jobs That Complete with Errors

It is possible that a job will finish with errors. If a job fails during a run, the sample(s) that fail will be indicated with a red bar. Click the 'View Log' link to troubleshoot that particular sample.

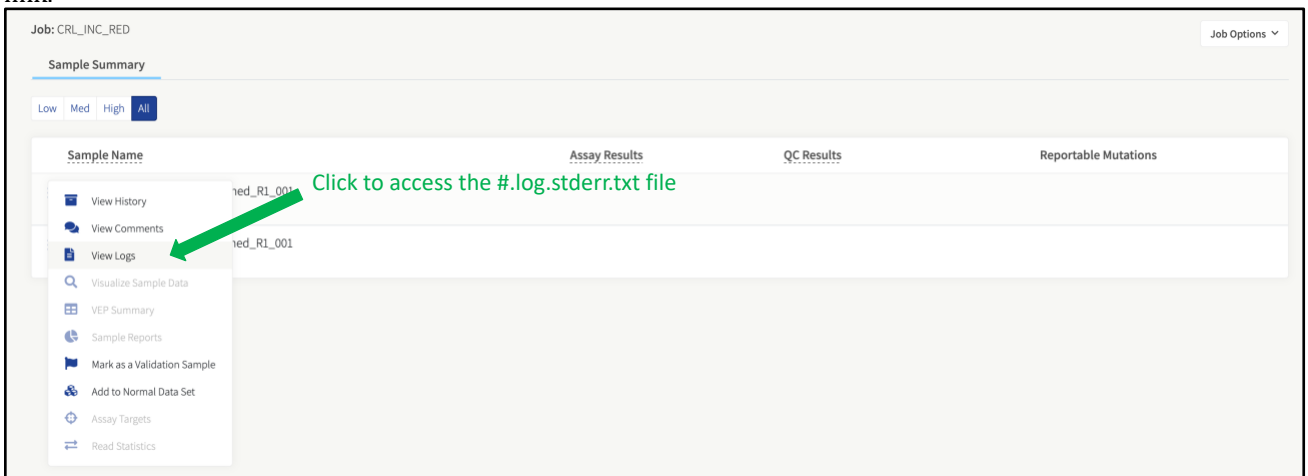


After all the samples have finished processing, the job will be marked as “COMPLETED\_ERROR”. Consult the log files from the job's ellipsis menu, to determine the reason for the failed job and/or contact [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com) for assistance.

In some cases, the error prevents the sample from being processed completely and this is indicated with the message “This analysis did not complete” in the **Job Details** page:



To troubleshoot these errors, users will need to view the Logs. Click the ellipsis menu icon to access the View Logs link.



After clicking that link, the `#.log.stderr.txt` file will be accessible to view and download in the modal. See below:

### Analysis Logs: 3732 ✕

Primary Workflow
Summarization and Staging
Workflow Configs

1.log.stderr.txt ← #.log.stderr.txt content
⋮

```

[2021-12-21 10:57:53,608] master_workflow.py [CHECKPOINT] > Archer Analysis Pipeline
[2021-12-21 10:57:53,608] master_workflow.py [STATUS] > Preparing pipeline
[2021-12-21 10:57:53,608] master_workflow.py [INFO] > Analysis Version: Suite_Analysis_build_80
[2021-12-21 10:57:53,644] master_workflow.py [INFO] > Quiver Version: 6.0.8
[TOTAL STEPS] : 11
Running on host: ip-10-159-52-34 [10.159.52.34]
[2021-12-21 10:57:53,667] workflow.py [CHECKPOINT] > Reference file preparation
[2021-12-21 10:57:53,667] workflow.py [STATUS] > Starting Workflow
[2021-12-21 10:57:53,667] prep.py [STATUS] > Preparing reference files...
[2021-12-21 10:58:28,351] bowtie2.py [INFO] > bowtie2-build --quiet /var/www/archer-temp/3732/temp.1/tmp_3c2uc.gene.fasta /var/www/archer-temp/3732/temp.1/tmp_3c2uc.gene.fasta
Traceback (most recent call last):
  File "/var/www/archer/analysis/workflows/master_workflow.py", line 1493, in main
    run_workflows(workflows, variations_timing_file, mw_config)
  File "/var/www/archer/analysis/workflows/master_workflow.py", line 1104, in run_workflows
    results_loader = AnalysisResultsLoader(output_dir, mw_config.SAMPLE)
  File "/var/www/archer/analysis/reporting/analysis_results_loader.py", line 34, in __init__
    self.sample_results_loader = self._generateSampleResultLoader(sample_files)
  File "/var/www/archer/analysis/reporting/analysis_results_loader.py", line 63, in _generateSampleResultLoader
    sample_name = AnalysisResultsLoader.generateSampleName(sample_files[0])
IndexError: list index out of range

[2021-12-21 10:58:52,705] master_workflow.py [ERROR] > Fatal error in master_workflow.py.

```

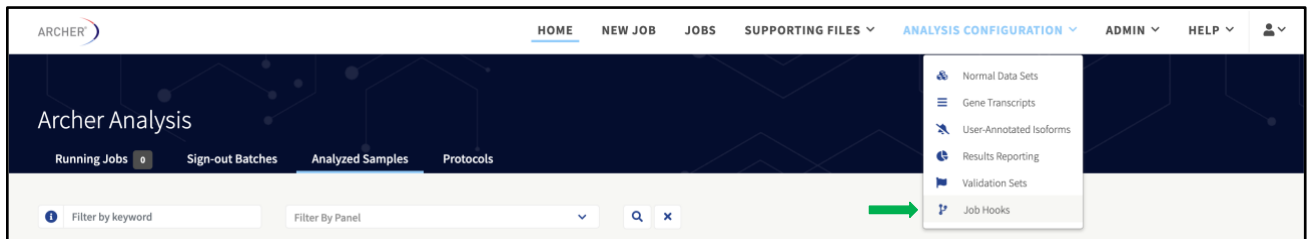
[Download Raw File](#)

## Job Hooks

Job hooks allow users to execute custom scripts at specific points in the analysis pipeline. This mechanism allows users to perform custom preprocessing steps, or custom tertiary analysis and reporting. The hook architecture provides users with access to job identifying information and pipeline results through system level variables.

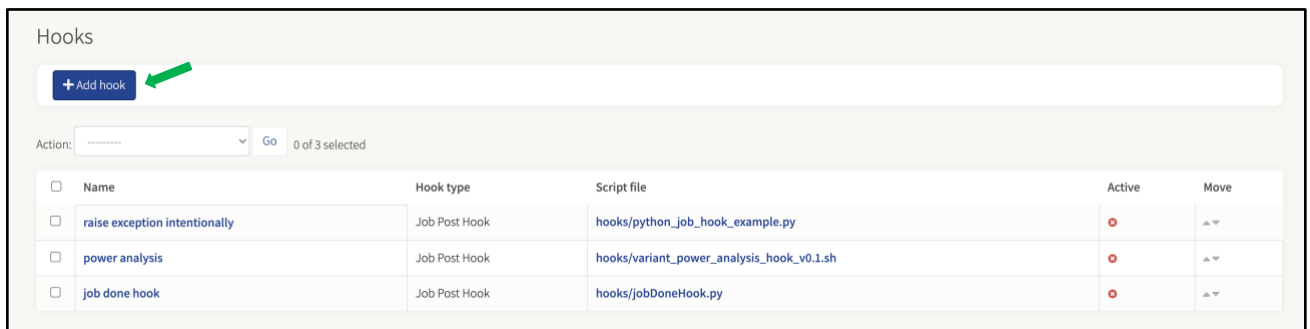
Note that users must have Admin privileges to create job hooks. However, standard users may utilize job hooks once created.

Select **Analysis Configuration** → **Job Hooks** to add/manage all job hooks.



### Add a Job Hook

Click **Add hook** to navigate to the **Add hook** page.



From the Add Hook page, to create a new job hook, the user will need to provide a name, description, hook type, script file and script arguments.



Hooks > Add hook

**Name:**   
A name for the hook

**Description:**   
Explanation of the hook's purpose

**Hook type:**   
Type of hook -- pre/post + sample/job

[Example Scripts](#)  
[python\\_job\\_hook\\_example.py](#)  
[bash\\_job\\_hook\\_example.sh](#)  
[python\\_sample\\_hook\\_example.py](#)  
[python\\_validation\\_hook\\_example.py](#)

**Script file:**  No file chosen

**Args:**   
 Args to pass to the script.  
 Include an associated option character if desired. e.g.,  
 -j \${JOB\_ID}

**Hook Name:** The hook name is important when differentiating between various hooks shown when setting up an analysis. The hook name must be unique to an analysis instance.

**Hook Description:** An explanation of what the hook does and how it fits with the analysis pipeline.

**Hook Types:** There are five hook types that define when the hook will be executed:

- **Job Pre Hook** – This hook type is implemented prior to a job running on the analysis pipeline.
- **Sample Pre Hook** – This hook type is implemented prior to each sample being run within a job in the analysis pipeline.
- **Job Post Hook** – This hook type is implemented after a job has completed the analysis pipeline.
- **Sample Post Hook** – This hook type is implemented after each sample has completed the analysis pipeline.
- **Validation Hook** – This hook type is used for software validation jobs being run on the analysis pipeline.

**Hook script file:** The script to be executed. This script can be implemented using Python version 2.7 or Bash (.py or .sh file extensions). Example hook scripts are located on the **Add hook** page and can be used as a base to develop a customized hook.

**Script arguments:** These variables allow information to be passed to the script. The only required system variable is `HOOK_OUTPUT_DIR`. All other system variables are optional.

- **Required:**
  - `HOOK_OUTPUT_DIR` – This variable represents a location for output files to be displayed by the web UI.
- **Available Job Hook Variable Names:**
  - `JOB_DIR` – This variable represents the job directory.
  - `JOB_ID` – This variable represents the job ID.
  - `SAMPLE_NAMES_COMMAS` – This variable represents sample names separated by commas.
  - `SAMPLE_NAMES_SPACED` – This variable represents sample names separated by spaces.
- **Available Validation Hook Variable Names:**

Optional arguments include the above job hook args, plus:

  - `ORACLE_DATA_PATH` – This variable validates against expected results.
- **Available Sample Hook Variable Names:**
  - `JOB_ID` – This variable represents the job ID.
  - `SAMPLE_INPUT_DIR` – This variable represents the directory containing the input data.
  - `SAMPLE_OUTPUT_DIR` – This variable represents the directory containing the output file for a sample.
  - `PBS_ARRAY_ID` – This variable identifies the torque job scheduler ID.
  - `SAMPLE_NAME` – This variable represents the sample name.
  - `SAMPLE_ID` – This variable represents the sample ID.
  - `LOG_FILE` – This variable represents a log file.
  - `SEQUENCE_FILES_COMMAS` – This variable represents sequence file names, separated by commas.
  - `SEQUENCE_FILES_SPACED` – This variable represents sequence file names, separated by spaces.
  - `SAMPLE_VCF` – This variable represents the sample VCF file.
  - `FULL_RESULTS_FILE` – This variable represents the full results file.
  - `CNV_GENE_SUMMARY` – This variable represents the CNV gene summary file.
  - `CNV_PRIMER_SUMMARY` – This variable represents the CNV primer summary file.
  - `VARIANT_SUMMARY` – This variable represents the variant summary file.
  - `ALL_VARIANTS_VCF` – This variable represents the job's unified VCF file.

To configure the job hook to execute by default for every new job, check the “Hook Enabled” box on the “Add Hook” page.

Once the required field is completed, click the **Save** button.

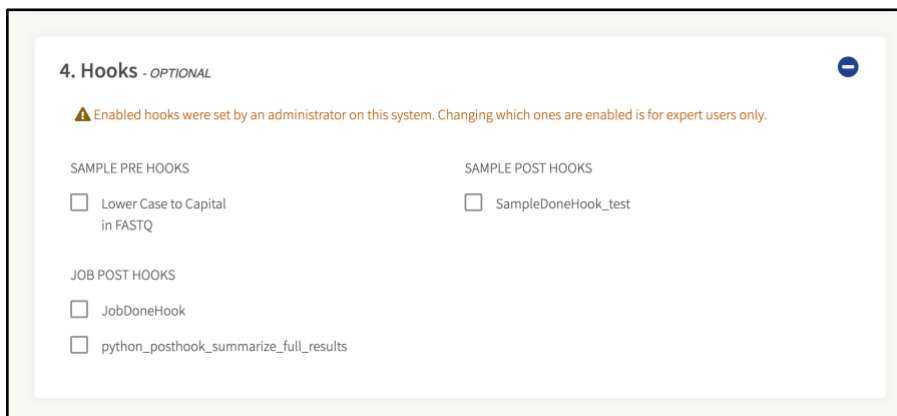
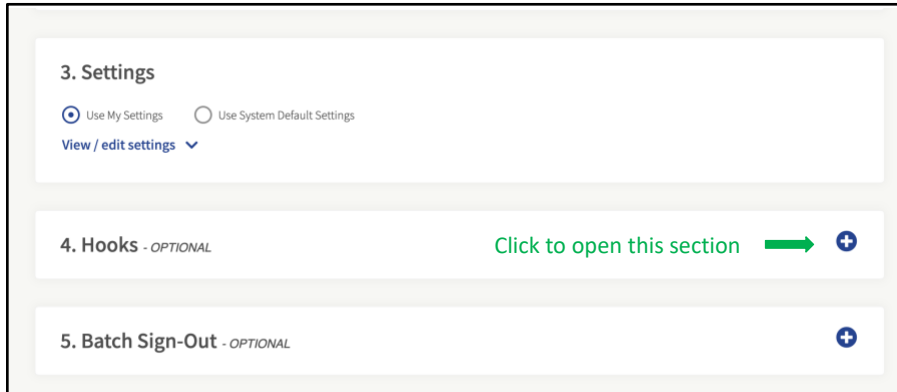
## Managing Existing Hooks

### Editing Existing Hooks

Job hooks can be modified within the **Job Hooks** page by clicking the desired hook name to edit. This will open in the **Change hook** page where modifications to the hook name, hook description, hook type, script file, hook arguments and enabled state can be made.

### Enabling Hooks on a Per Job Basis

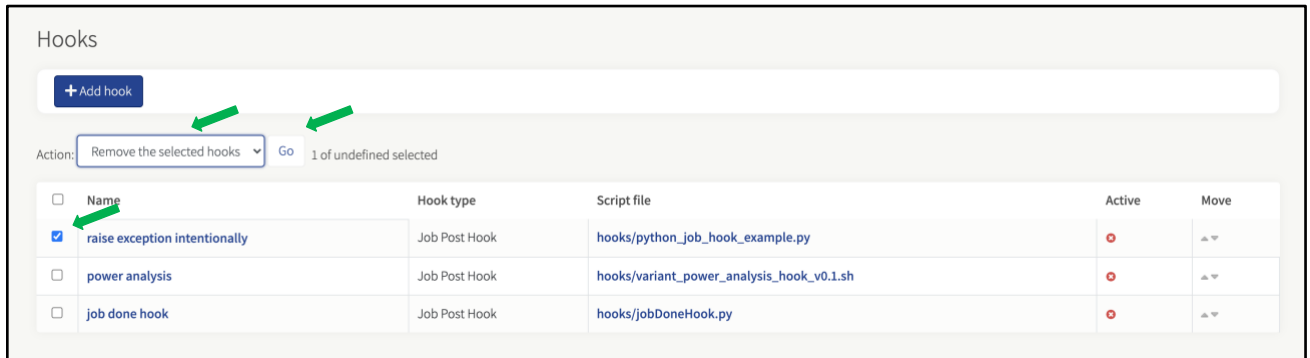
Job hooks can be enabled/disabled from the **New Job** page, when performing an analysis, by opening the **Hooks** section on the **New Job** page and checking/unchecking all desired hooks to enable/disable.



Once job setup is complete, click Submit Analysis to begin job processing. Please see **Job Setup and Management** section for more details.

## Deleting Existing Hooks

To delete a job hook, check the box to the left of the job hook name, then select **Remove the selected hooks** from the **Action** dropdown menu and click **Go**.



Hooks

+ Add hook

Action: Remove the selected hooks Go 1 of undefined selected

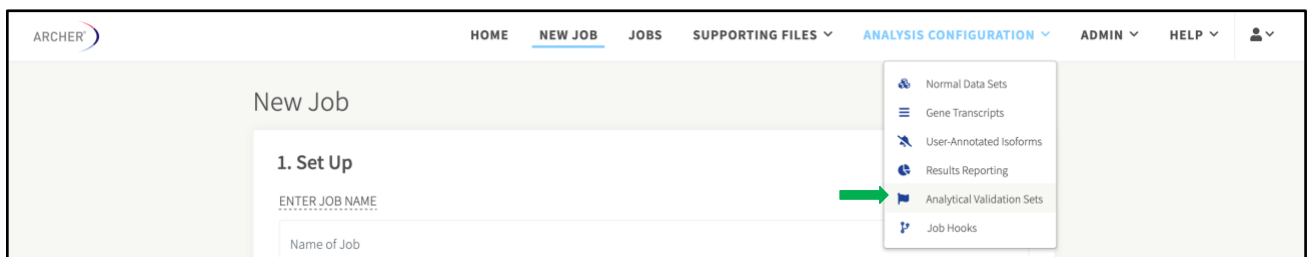
<input type="checkbox"/>	Name	Hook type	Script file	Active	Move
<input checked="" type="checkbox"/>	raise exception intentionally	Job Post Hook	hooks/python_job_hook_example.py	<span style="color: red;">●</span>	▲▼
<input type="checkbox"/>	power analysis	Job Post Hook	hooks/variant_power_analysis_hook_v0.1.sh	<span style="color: red;">●</span>	▲▼
<input type="checkbox"/>	job done hook	Job Post Hook	hooks/jobDoneHook.py	<span style="color: red;">●</span>	▲▼

## Analytical Validation Sets

The use of analytical validation sets is important to users who want to validate the version of Archer Analysis they are running. Users can select any of their completed samples to produce an analytical validation set that executes and automatically checks to confirm the software is working properly. This feature not only notifies the user an analytical validation set passed/failed but provides a detailed downloadable PDF report.

Only Admin users can run analytical validation sets.

To access Analytical Validation Sets, select the **ANALYSIS CONFIGURATION → Analytical Validation Sets** menu to add/manage Analytical Validation Sets.



ARCHER

HOME NEW JOB JOBS SUPPORTING FILES ANALYSIS CONFIGURATION ADMIN HELP

New Job

1. Set Up


ENTER JOB NAME

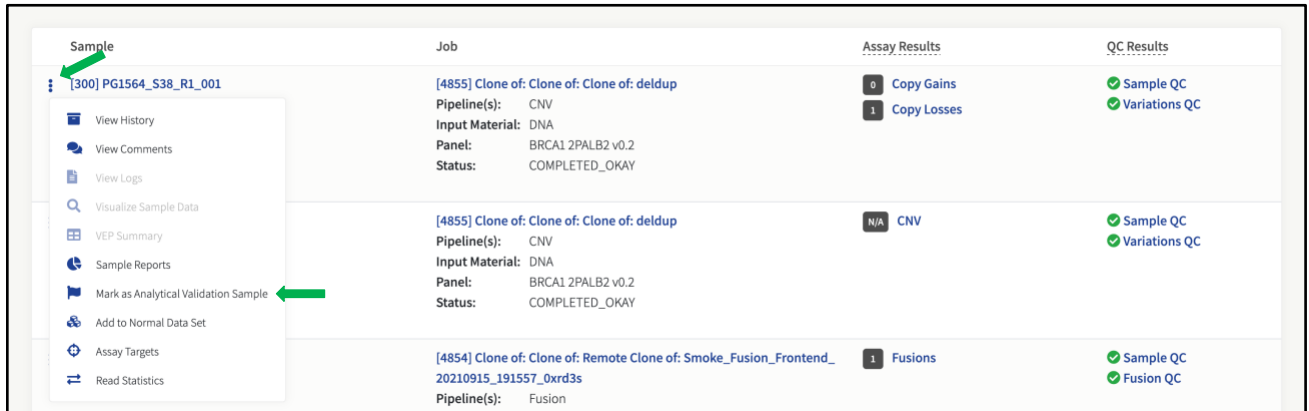
Name of Job

- Normal Data Sets
- Gene Transcripts
- User-Annotated Isoforms
- Results Reporting
- Analytical Validation Sets
- Job Hooks

## Mark Samples for Analytical Validation Set

To mark samples for an analytical validation set, navigate to the Jobs page, select a specific job, and locate the sample(s) to incorporate on the **Job Details** page.

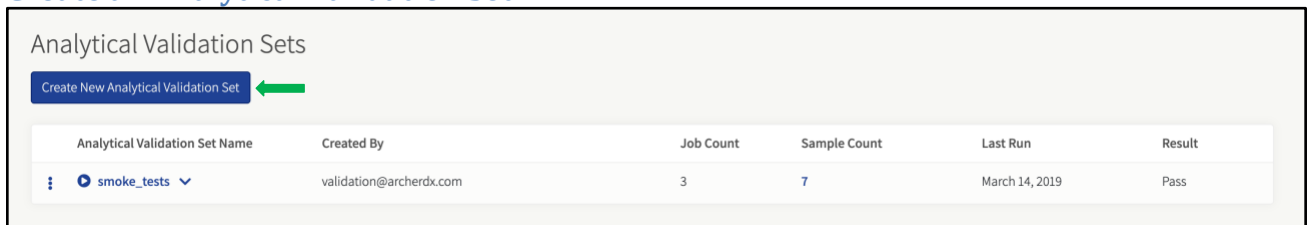
After locating a sample to add, click the  icon and select **Mark as an Analytical Validation Sample**.



Sample	Job	Assay Results	QC Results
[300] PG1564_S38_R1_001	[4855] Clone of: Clone of: Clone of: deldup Pipeline(s): CNV Input Material: DNA Panel: BRCA1_2PALB2 v0.2 Status: COMPLETED_OKAY	0 Copy Gains 1 Copy Losses	Sample QC Variations QC
	[4855] Clone of: Clone of: Clone of: deldup Pipeline(s): CNV Input Material: DNA Panel: BRCA1_2PALB2 v0.2 Status: COMPLETED_OKAY	N/A CNV	Sample QC Variations QC
	[4854] Clone of: Clone of: Remote Clone of: Smoke_Fusion_Frontend_20210915_191557_0xrd3s Pipeline(s): Fusion	1 Fusions	Sample QC Fusion QC

After marking all the samples from a single job, or multiple jobs, to include in an analytical validation set, create a new analytical validation set.

### Create an Analytical Validation Set

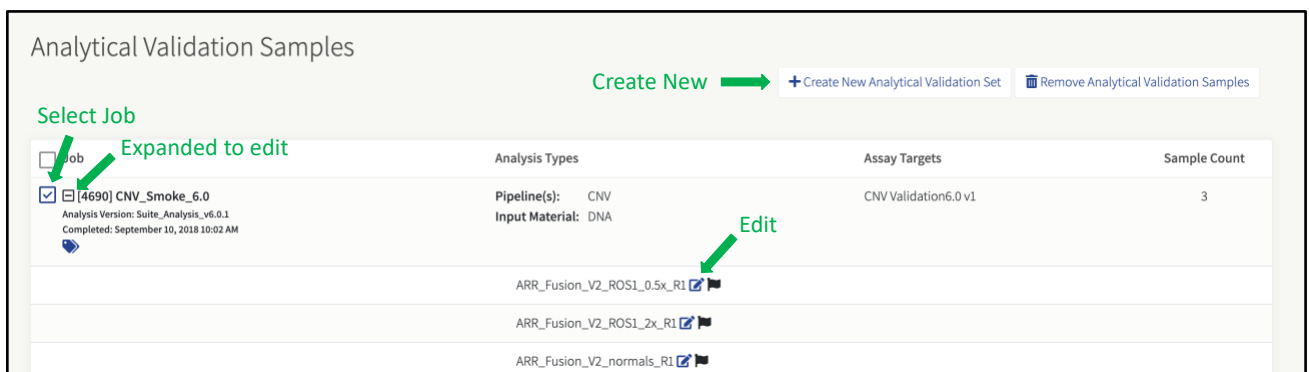


Analytical Validation Sets

[Create New Analytical Validation Set](#)

Analytical Validation Set Name	Created By	Job Count	Sample Count	Last Run	Result
smoke_tests	validation@archerdx.com	3	7	March 14, 2019	Pass

Once on the Analytical Validation Sets page, click **Create New Analytical Validation Set**. The Analytical Validation Samples page will appear.



Analytical Validation Samples

[Create New](#) [+ Create New Analytical Validation Set](#) [Remove Analytical Validation Samples](#)

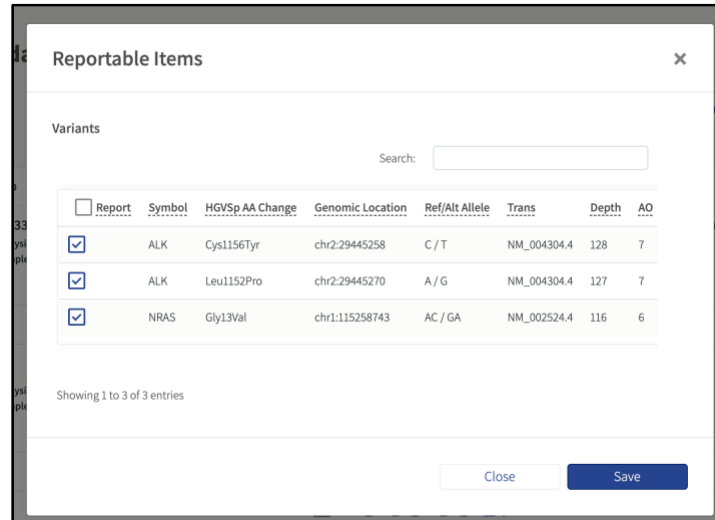
Select Job

Job	Analysis Types	Assay Targets	Sample Count
<input checked="" type="checkbox"/> [4690] CNV_Smoke_6.0 Analysis Version: Suite_Analysis_v6.0.1 Completed: September 10, 2018 10:02 AM	Pipeline(s): CNV Input Material: DNA	CNV Validation6.0 v1	3
	ARR_Fusion_V2_ROS1_0.5x_R1		
	ARR_Fusion_V2_ROS1_2x_R1		
	ARR_Fusion_V2_normals_R1		


Next, select **all the jobs/samples to include in the set** and click the **+ Create New Analytical Validation Set** button.

Name the analytical validation set and click **Ok**. The **Analytical Validations Sets** page appears to view the newly created analytical validation set.

Users can edit each analytical validation sample to only report specific variant types found within a sample. After clicking the edit box icon , a Reportable Items box appears allowing users to select the specific variants to include for analytical validation testing.




## Delete an Analytical Validation Set

To delete an analytical validation set(s) from the Analytical Validation Sets page, click the  icon to the left of the analytical validation set and select **Delete**. Analytical validation sets that have finished running at least once are retained as a record and cannot be deleted.

Analytical Validation Set Name	Created By	Job Count	Sample Count	Last Run	Result
 Test Validation Set	[redacted]@invitae.com	2	5	N/A	
 Delete	[redacted]@invitae.com	1	2	N/A	

## Run an Analytical Validation Set

After creating a new analytical validation set, run the set to ensure the software is working properly. Select the icon shown below to begin running the analytical validation process.

Analytical Validation Set Name	Created By	Job Count	Sample Count	Last Run	Result
 smoke_tests 	validation@archerdx.com	3	7	March 14, 2019	Pass

## View Analytical Validation Samples

Users can view analytical validation jobs/samples contained within an analytical validation set.

Click the **number of samples shown** under the Sample Count column to access the jobs/samples on the Analytical Validation Samples page.

### Analytical Validation Samples



Included samples for validation set: Test Validation Set

Job	Analysis Types	Assay Targets	Sample Count
[3385] Smoke variant 2017_05_16 <small>Analysis Version: Suite_Analysis_v5.1_RC.1            Completed: May 16, 2017 11:39 AM</small>	<b>Pipeline(s):</b> SNP/InDel <b>Input Material:</b> DNA <b>Mutation Detection Type:</b> Somatic Mutation	VariantPlex SolidTumor v1	2
[3389] Smoke CNV resubmit <small>Analysis Version: Suite_Analysis_v5.1.2            Completed: August 28, 2017 09:07 AM</small>	<b>Pipeline(s):</b> CNV <b>Input Material:</b> DNA	CNV Validation v1	3

From here, users can view the details about the jobs/samples, analysis types, assay targets and sample count. Users can also add/remove samples from this this page if needed.

## Analytical Validation Report

Regardless of the analytical validation set passing or failing, a detailed analytical validation report PDF provides information about the status of the analytical validation run. Details for each sample include the name, job type, QC status, overall results, and what was detected/expected.

Analytical Validation Set	smoke_tests		
Overall Status	PASS		
Started	March 14, 2019 12:14 PM		
Completed	March 14, 2019 12:57 PM		
Software Version	Suite_Analysis_v6.0.4		
Sample Count	7		
<b>ARR_Fusion_V2_ROS1_0.5x_R1</b>			
Sample Links	<b>observed / expected</b>	Job Type	<b>Pipeline(s):</b> CNV <b>Input Material:</b> DNA
QC	Sample QC: PASS Variations QC: PASS	Job Status	COMPLETED_OKAY 
Overall Result	PASS		
Losses Found	1 (1 expected)		
<b>Details</b>			
<b>ARR_Fusion_V2_normals_R1</b>			
Sample Links	<b>observed / expected</b>	Job Type	<b>Pipeline(s):</b> CNV <b>Input Material:</b> DNA
QC	Sample QC: PASS Variations QC: PASS	Job Status	COMPLETED_OKAY 
Overall Result	PASS		
CNVs	0 (0 expected)		

To download the Analytical Validation Report, navigate to the Analytical Validation Sets page, click the **analytical validation set**, then click **Download PDF Report** next to the date for which test was run.

Analytical Validation Sets

[Create New Analytical Validation Set](#)

Analytical Validation Set Name	Created By	Job Count	Sample Count	Last Run	Result
Test 2	test@test.test	1	3	January 19, 2022	Fail
smoke_tests	validation@archerdx.com	3	7	March 14, 2019	Pass

[View Run](#) [Download PDF Report](#)

**Started By:** validation@archerdx.com  
**Software Version:** Suite\_Analysis\_v6.0.4  
**Started:** March 14, 2019, 12:14 p.m.  
**Completed:** March 14, 2019, 12:57 p.m.  
**Result:** Pass

## Run History

View the analytical validation set run history on the Analytical Validation Sets page. This information includes details such as the name of the analytical validation set, who created the set, job count, sample count, when it was last run, and by whom.

Analytical Validation Sets

[Create New Analytical Validation Set](#)

Analytical Validation Set Name	Created By	Job Count	Sample Count	Last Run	Result
Test 2	test@test.test	1	3	January 19, 2022	Fail
smoke_tests	validation@archerdx.com	3	7	March 14, 2019	Pass

[View Run](#) [Download PDF Report](#)

**Started By:** validation@archerdx.com  
**Software Version:** Suite\_Analysis\_v6.0.4  
**Started:** March 14, 2019, 12:14 p.m.  
**Completed:** March 14, 2019, 12:57 p.m.  
**Result:** Pass

[View Run](#) [Download PDF Report](#)

**Started By:** validation@archerdx.com  
**Software Version:** Suite\_Analysis\_v6.0.3.2  
**Started:** Sept. 21, 2018, 10:07 a.m.  
**Completed:** Sept. 21, 2018, 10:47 a.m.  
**Result:** Pass

[View Run](#) [Download PDF Report](#)

**Started By:** validation@archerdx.com  
**Software Version:** Suite\_Analysis\_v6.0.3.1  
**Started:** Sept. 20, 2018, 2:52 p.m.  
**Completed:** Sept. 20, 2018, 3:34 p.m.  
**Result:** Pass

[View Run](#) [Download PDF Report](#)

**Started By:** validation@archerdx.com  
**Software Version:** Suite\_Analysis\_v6.0.2.1  
**Started:** Sept. 18, 2018, 12:15 p.m.  
**Completed:** Sept. 18, 2018, 1 p.m.  
**Result:** Pass

[View Run](#) [Download PDF Report](#)

**Started By:** validation@archerdx.com  
**Software Version:** Suite\_Analysis\_v6.0.1  
**Started:** Sept. 10, 2018, 6:16 p.m.  
**Completed:** Sept. 10, 2018, 6:57 p.m.  
**Result:** Pass

## Accessing and Interpreting Results

### Key Assay Concepts

There are several key concepts that should be understood when working with Anchored Multiplex PCR (AMP) chemistry, in order to understand many of the terms and metrics below, as well as to properly interpret results from all Archer assays. These concepts are reviewed in the sections below.

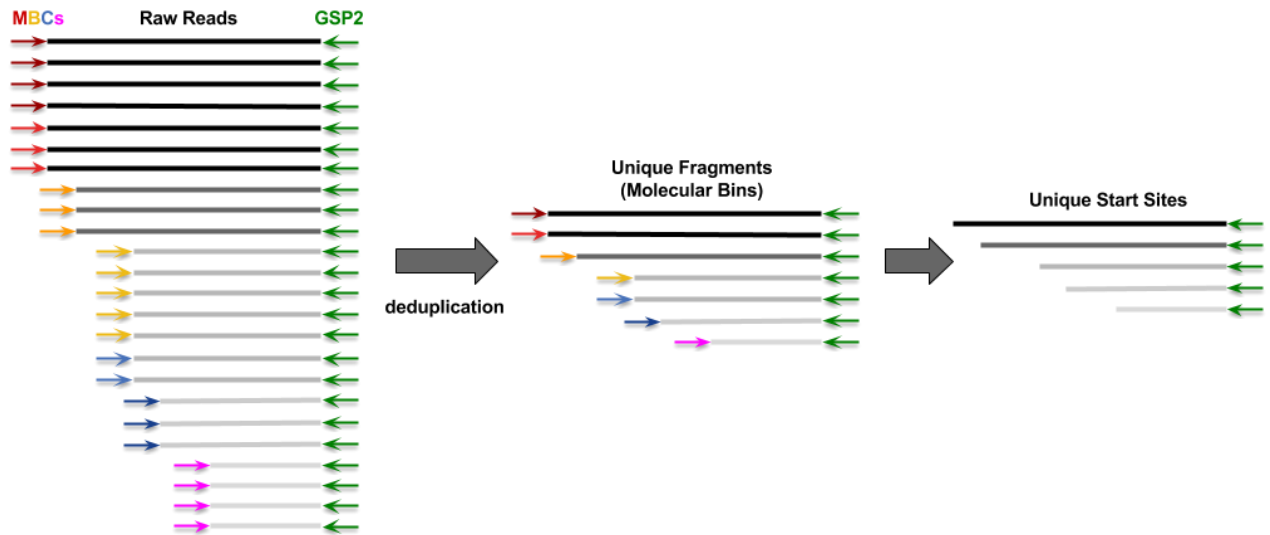
### Genome Reference

All Archer panels are designed for, and results presented relative to, hg19 (GRCh37).



### Raw Reads, Unique Fragments, and Unique Start Sites

The graphic below illustrates how Analysis leverages the Molecular Barcodes (MBCs) of AMP in order to work backward from all reads reported from the sequencer (i.e., in the sequencer output files uploaded at job setup), in order to deduce the original molecules that were present prior to the target enrichment protocol.



**Raw Reads** (a.k.a. Total Fragments) – this is the total number of reads (or read pairs) reported from the sequencer.

**Unique Fragments** (a.k.a. Molecular Bins) – this is the total number of original fragments extracted from your sample, i.e., after fragmentation but before any PCR steps. These are determined by finding all raw reads (or read pairs) that share the same MBC, and deduplicating these into a single consensus sequence (note, low quality raw reads are filtered out before this step. See [Read Statistics](#) for further details).

**Unique Start Sites** – this is a subset of Unique Fragments and represents the total number of unique fragment lengths extracted from your sample (since each fragment is anchored by a GSP2 on one side, length can only be affected by the random start site on the opposite end of the fragment). Unique start sites can manifest as multiple physical molecules, which is why there are typically more Unique Fragments than Unique Start Sites.



Unique Fragments tell you the true input quantity from your sample and allow for accurate detection of allele fractions (AF). Unique Fragments and Unique Start Sites together provide a reliable measure of sample complexity, which can be correlated to assay sensitivity. These metrics also help guard potential false positives arising from PCR and sequencing error.

### RNA vs. DNA vs. Ambiguous Reads

In RNA-based assays, it is likely that some fraction of reads originated from DNA, and it is important to differentiate which reads represent transcripts. To accomplish this, Unique Fragments that pass quality control (see [Read Statistics](#) for further details) are parsed further into three categories:

**RNA Reads** – those that span an exon-exon junction, as only transcripts would have such content.

**DNA Reads** – those that span an exon-intron junction, since RNA would rarely have such content.

**Ambiguous Reads** – those that are contained within a single exon (which makes it impossible to apply either of the rules above), or split reads that align to different genes that are not in Quiver.



The SNP/InDel pipeline (for both DNA and RNA) uses all reads from GSP2s marked SNV in the target GTF. The CNV Pipeline uses all reads from GSP2s marked CNV in the target GTF. (There is no reverse-transcription step in DNA assays, thus it is virtually impossible to get RNA reads from VariantPlex and LiquidPlex assays).

The Immune Repertoire pipeline does not employ the above rules to differentiate RNA vs. DNA reads, and instead utilizes a third-party tool, MiXCR. See Immune Repertoire (RNA only) for more details.

Two sample tracking pipelines are offered that generate additional output files and enable users to predict sample sex and report a string of SNP genotype calls which are likely to be unique to each sample donor. The sex enumeration and string representing the genotype of all SNP\_ID targeted variants can be used to troubleshoot sample swapping. The results are not intended to be used to assess low-level sample cross-contamination. The pipelines that yield the previously mentioned identifiers are referred to as SEX\_ID and SNP\_ID and utilize primers that target the X and Y chromosomes and a series of highly polymorphic genomic positions respectively. Only Archer panels that contain these primers are capable of reporting this data. The pipeline is evoked via function flags in the GTF for these primers: function "SEX\_ID" and function "SNP\_ID".

Note, these primers target non-expressed DNA sequence (intronic or intergenic). As such, input samples must contain DNA, otherwise insufficient coverage will be generated from these primers. When using SEX\_ID and SNP\_ID with FusionPlex panels, do not use DNase treated RNA as an input into library preparation.

#### **SEX\_ID**

Some Archer panels contain SEX\_ID primers that can be used to determine the sex of the donor. If a panel containing SEX\_ID primers and the appropriate function flag in the GTF is used, there will be an output file that has "\*molbar.trimmed.deduped.snp\_id.txt" as a suffix. This file reports the sex of the donor based on data from two primers. The primers bind ambiguously to both X and Y chromosomes however the reads that extend off these primers generate sequence that unambiguously maps downstream from the primer. This provides the ability to determine the chromosome of origin of the resulting fragments. There is a read depth requirement of 20x total unique reads (summed across the two primers) to make a call; if this is not met, the algorithm results in Unknown. If coverage on chrY (>20x unique reads) exists, the initial designation is male. If no coverage exists on chrY, the algorithm reports female. If there is coverage in Y, and also at least one heterozygous SNP call in X, the call is changed to Undetermined.

#### **SNP\_ID**

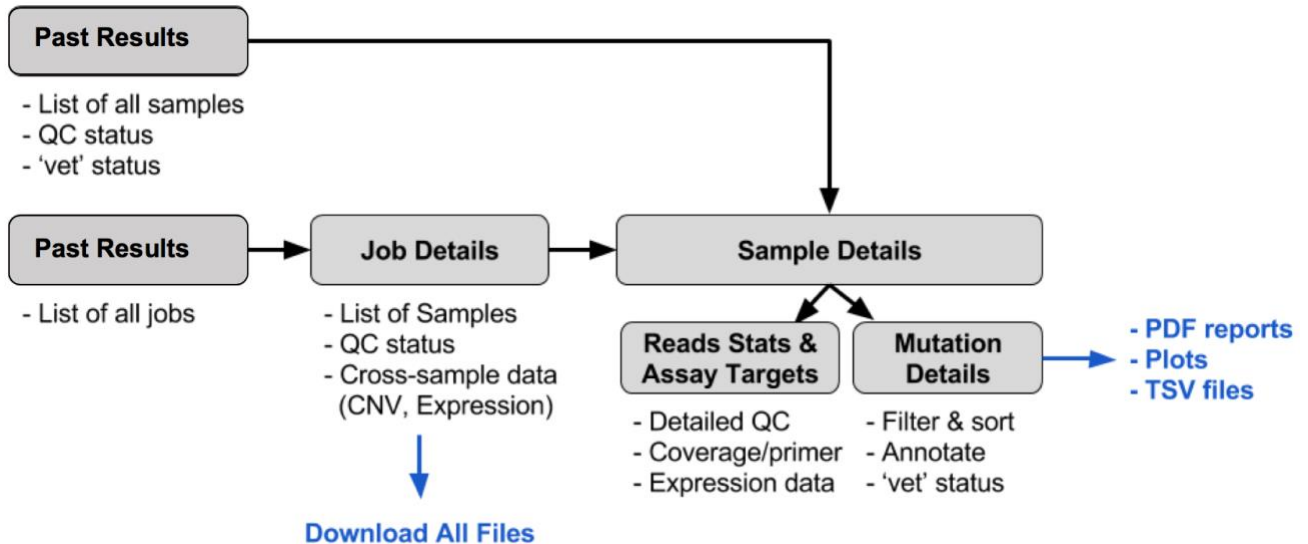
Some Archer panels contain SNP\_ID primers that can be used to report genotype across a series of highly polymorphic genomic positions in mostly intronic and intergenic regions. If a panel containing SNP\_ID primers and the appropriate function flag in the GTF is used, there will be an output file that has "\*molbar.trimmed.deduped.snp\_id.txt" as a suffix, which contains the genotype report. Our in-house targeted variant caller, VISION is used to determine the genotype of a SNP call across the targets and the following is reported for each target in the snp\_id.txt file: chromosome, position, Ref, Alt, genotype, unique AF, unique RO, and unique AO. If the sum of unique AOs and unique ROs is less than 20 the genotype assignment is deferred. The genotype call for each position is classified as one of 4 types: HOMOZYGOUS REF, HETEROZYGOUS, HOMOZYGOUS ALT, or UNDETERMINED. These calls are abbreviated as "r", "h", "A" or ".", respectively. The first line of the snp\_id.txt file contains the series of abbreviations as a sample ID string.

#### **Control Genes (FusionPlex Only)**

Since RNA expression can vary widely based on a number of biological factors, Control GSP2s targeting 4 genes are included in FusionPlex assays. These 4 genes, CHMP2A, GPI, RAB7A, and VCP, are known to be consistently moderately expressed across most tissue types, and thus collectively serve as a reliable (1) indicator of overall RNA quality and content in the sample, and (2) benchmark for normalization in the RNA Expression pipeline. See [Sample QC Fusions QC and Relative Expression for further details on these points.](#)

## General Results Review & Interpretation Workflow

The GUI is designed to present all information needed to complete secondary analysis (i.e., true/false positive determination), as well as some amount of tertiary analysis (e.g., using Clinvar annotations), but all data is accessible for download and custom processing. There are two primary ways to access results of completed jobs in Archer Analysis, by sample (**Analyzed Samples**) and by job (**Jobs**). The diagram below illustrates the basic job/sample/mutation review and sign-out workflow. There are several points throughout the workflow to extract results, either for final reporting (PDFs and image files) or for further data processing (BAMs, raw text files, etc.):



### Sign-out Batches

The **Sign-out Batches** view is intended as a quick way to screen through a daily caseload and is a means to access all batches that you either own or have been assigned to you. All batches are listed, and batches that have completed review are marked accordingly. The view is accessed by selecting the “**Sign-out Batches**” tab from the default Home page (which can be changed to this view, see [Miscellaneous System Settings](#) for further details):

### Analyzed Samples

The **Analyzed Samples** view is intended as a quick way to screen through a daily caseload and is a means to access all samples in the system. The view is accessed by selecting the “**Analyzed Samples**” tab from the default Home page (which can be changed to this view, see [MSI Pipeline Settings](#))

This section of the User Settings page contains parameters that only affect the CNV pipeline in Analysis.

MSI Settings	
Minimum number of loci for MSI	25
Minimum Depth Per Sample for MSI	35000
Stable Threshold for MSI	20
High Threshold for MSI	30
MSI Alpha	0.05
Minimum Depth Per Locus	25

### Minimum number of loci for MSI

The minimum number of targeted microsatellite loci that must pass QC for a sample to get a percent unstable value and an MSI status call.

### Minimum Depth Per Sample for MSI

The minimum total useable raw reads depth a sample must have across all microsatellites for microsatellite instability to be determined. Minimum read depth required for the sample to pass QC).

### Stable Threshold for MSI

The maximum percentage of unstable loci for a sample to be considered MSI-Stable. Unstable loci percentages above this, and less than MSI\_HIGH\_THRESHOLD, will be considered MSI-Intermediate.

### High Threshold for MSI

The minimum percentage of unstable loci for a sample to be considered MSI-High. Unstable loci percentages less than this, and above MSI\_STABLE\_THRESHOLD, will be considered MSI-Intermediate.

**Note:** The high and stable threshold can be set to equal each other, thereby removing the MSI-Intermediate call.

### MSI Alpha

The maximum pval for a locus to be considered unstable.

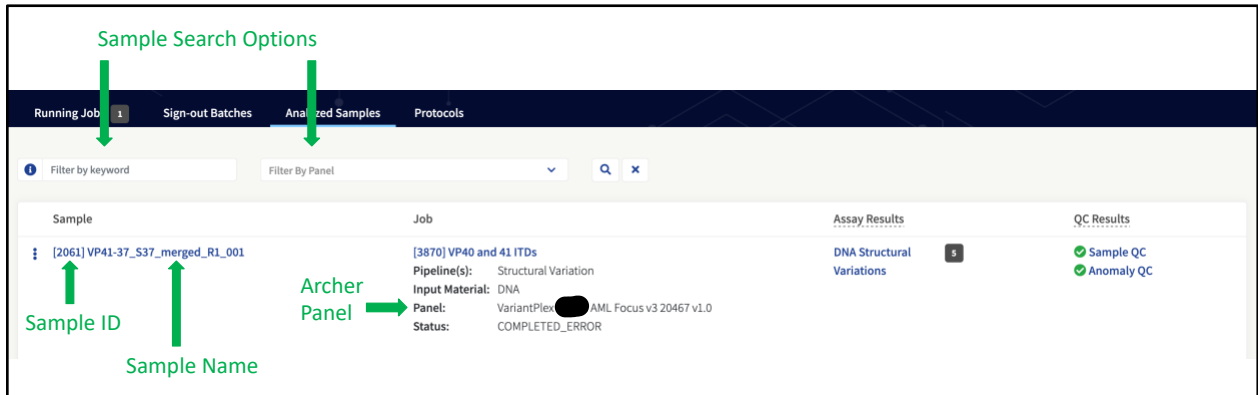
### Minimum Depth per Locus

The minimum read depth a microsatellite locus must have to generate a locus instability score. Miscellaneous System Settings for further details).

By default, the samples are listed according to:

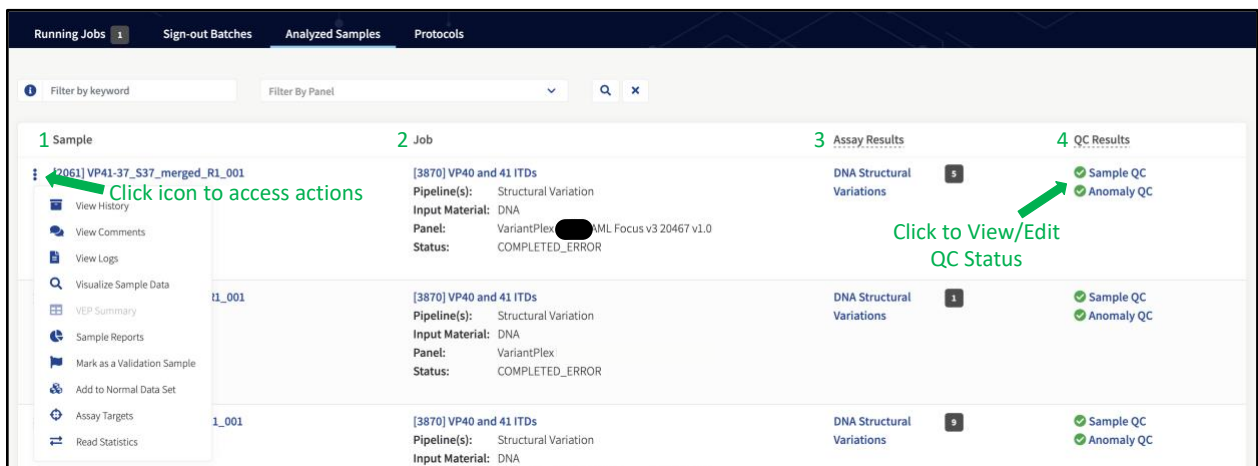
1. Most recent completed job
2. Then by the order the sample was processed (i.e., ascending sample ID)



Samples can also be found by searching by keyword (e.g., sample name, sample ID, job name, or job ID) and by panel name:



## Sample Summary

The **Analyzed Samples** page provides a summary of key sample data, as well as links to take action (i.e., review, sign out) on each sample:



1. Click the sample ID/name under the **Sample** column to jump to the **Sample Details** page and get an in-depth view of all results for the sample (see [Error! Reference source not found.](#) for further details).
2. Click the job ID/name under the **Job** column to jump to the Job Details (Sample Summary) page of the job the sample was run in (see [Error! Reference source not found.](#) for further details).
3. Click on any of the links under the **Results** column in order to jump right to the pipeline-specific results (e.g., Fusions, Variants) within the **Sample Details** page. See [Error! Reference source not found.](#) for further details.
4. Click on either the name of the QC type (e.g., Fusion QC) or the pass/fail icon (  /  ) to view the **Sample QC History** popup. See [Error! Reference source not found.](#) for further details.

These actions are also available in the more-actions menu for each sample in the **Analyzed Samples** page:

- View History
- View Comments
- View Logs
- Visualize Sample Data
- VEP Summary
- Sample Reports
- Mark as a Validation Sample

- Add to Normal Data Set
- Assay Targets
- Read Statistics

### QC Results

There are four pipeline-specific Quality Control (QC) metrics reported at the sample level to determine if there is any potential issue with the sample related to the extraction, library prep, and/or sequencing: Sample QC, Anomaly QC, Fusions QC and Variations QC.

#### *Sample QC*

This is the overall QC result for a sample. The value is the aggregation of the individual metric QC results. Please note that Immune Repertoire do not have a Sample QC metric.

#### *Anomaly QC*

The Anomaly QC metric will report pass/fail only if the Structural Variations pipeline is selected for. It is defined as the average number of unique start sites (from DNA and ambiguous reads) calculated per GSP2 across the entire panel. (See *Key Assay Concepts* for explanations of unique start sites and DNA/ambiguous reads).

#### *Fusions QC*

The Fusions QC metric will report pass/fail only if the Fusions/Isoforms pipeline is selected for FusionPlex assays (i.e., “RNA Fusion” Analysis Type). It is defined as the average number of unique start sites (from RNA reads) calculated per control GSP2 (see *Key Assay Concepts* for explanations of unique start sites, RNA reads, and control GSP2s).



This metric is considered a reliable measure of sample complexity (i.e., information content) of the sequencing library, with low values indicating that either (1) sample input was insufficient quality and/or quantity, or (2) there was some issue with extraction, library prep, and/or sequencing.



This metric is user-configurable and should be calibrated per your lab’s specific requirements and operating environment. See *Error! Reference source not found.* (Min Average Unique RNA Start Sites per GSP2 Controls) for further details.



The PreSeq™ RNA QC Assay is correlated to this Analysis QC metric, and thus provides a means of rescuing samples early in library preparation from further processing. Consult the FusionPlex protocol for your assay for further details on using PreSeq.

#### *Variations QC*

The Variations QC metric will report pass/fail if the SNP/InDel and/or CNV pipelines are selected. It is defined as the average number of unique start sites (from DNA and ambiguous reads) calculated per GSP2 across the entire panel. (See *Key Assay Concepts* for explanations of unique start sites and DNA/ambiguous reads).





This metric is considered a reliable measure of sample complexity (i.e., information content) of the sequencing library, with low values indicating that either (1) sample input was insufficient quality and/or quantity, or (2) there was some issue with extraction, library prep, and/or sequencing.



This metric is user-configurable and should be calibrated per your lab's specific requirements and operating environment. See *General Analysis Settings (Min Average Unique DNA and AMBIG Start Sites per GSP2)* for further details.

### Viewing and Overriding QC Pass/Fail Threshold

Click on either the name of the QC type (e.g., Fusion QC) or the pass/fail icon (  /  ) to view the **Sample QC History** popup, which shows not only the QC metric for the sample, but also what the threshold is (as defined by User Settings at the time of job setup; see *General Analysis Settings* for further details). This popup also allows for manual override of the pass/fail call, along with commenting on the sample – all actions are recorded and viewable in the History section at the top:

QC Metrics
✕

History of comments & pass/fail alterations


History

Reviewer	Date	Type	Action	Comment
No history found.				

From User Settings

Metrics

Sample QC
Anomaly QC

Metric	Value	Threshold	Pass
Average Unique DNA Start Sites per GSP2	645.08	>= 50.0	

Overall Anomaly QC:

Pass ▼

Comment

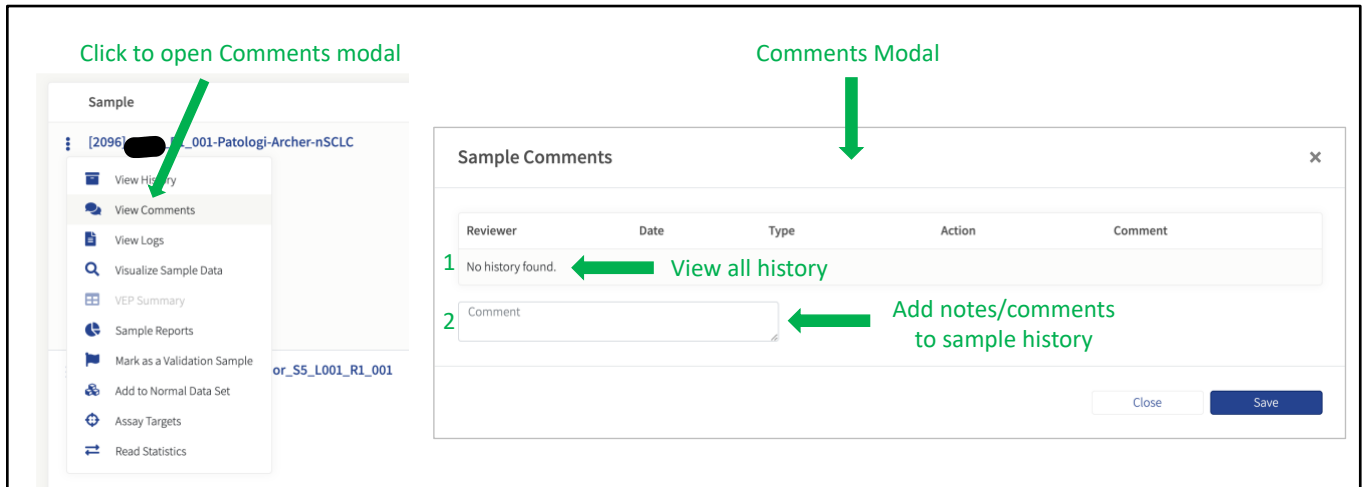
Comment and/or override Pass/Fail status

Close
Save



## Results Review and Reporting

The core intent of the **Sign-out Batches** and **Analyzed Samples** page is to review and report on each sample. A complete history of any comments and changes for a sample is tracked in the View History and View Comments sections in the more-actions menu for each sample:



Click to open Comments modal

Comments Modal

View all history

Add notes/comments to sample history

1. View entire review history. This includes change in status and/or comments added. Previous additions cannot be modified.
2. Add a comment to the sample history. This can be done independently of review status change.

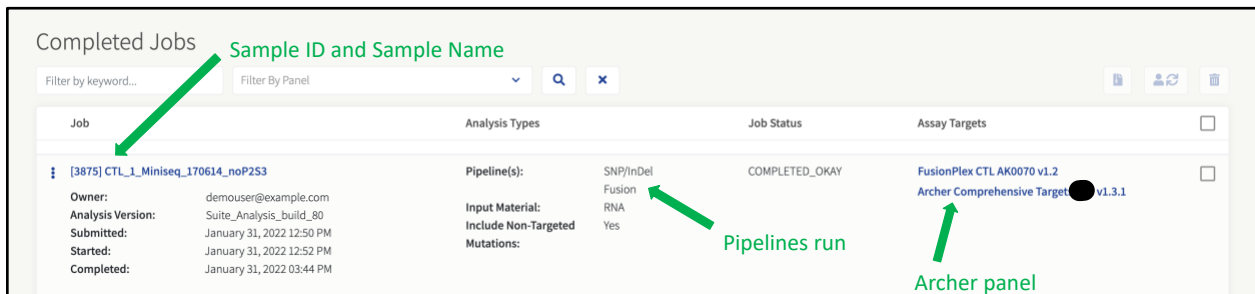
## Jobs

Access the **Jobs** page via the main menu:



Access all completed jobs

The **Jobs** page also shows a list of all completed jobs, including any jobs run by other users that the current user has permission to view (see Manage Users and Groups for further details) Specific jobs can be located by typing a query into the search box, or by filtering by any GTF files currently loaded into the system that have associated jobs:



Completed Jobs

Sample ID and Sample Name

Filter by keyword...

Filter By Panel

Job	Analysis Types	Job Status	Assay Targets
[3875] CTL_L_Miniseq_170614_noP253 Owner: demouser@example.com Analysis Version: Suite_Analysis_build_80 Submitted: January 31, 2022 12:50 PM Started: January 31, 2022 12:52 PM Completed: January 31, 2022 03:44 PM	Pipeline(s): SNP/InDel Input Material: Fusion RNA Include Non-Targeted: Yes Mutations: Yes	COMPLETED_OKAY	FusionPlex CTL AK0070 v1.2 Archer Comprehensive Target v1.3.1

Pipelines run

Archer panel

The job list includes descriptive information about each job:

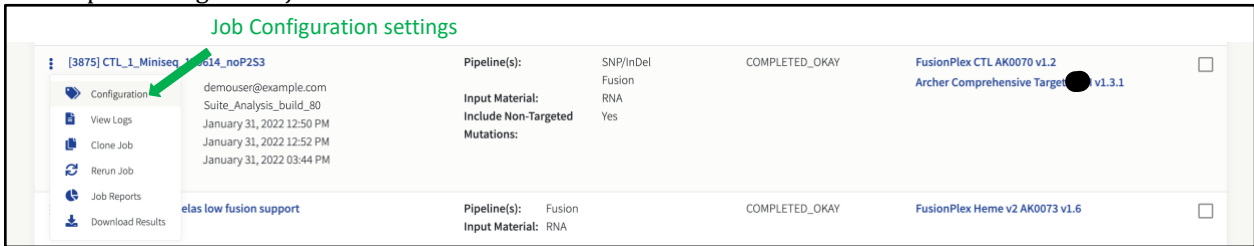
- **Job** - Job ID, job name, owner, analysis version, and timestamps for when each job was submitted, started, and completed.
- **Analysis Types** - The pipeline(s) that were chosen when setting up the job.



- **Job Status** - One of “HALTED”, “COMPLETED\_ERROR” and “COMPLETED\_OKAY”.
- **Assay Targets** - The panel (GTF file) that was chosen when setting up the job. If a Targeted Mutations file was chosen, it is also shown here.

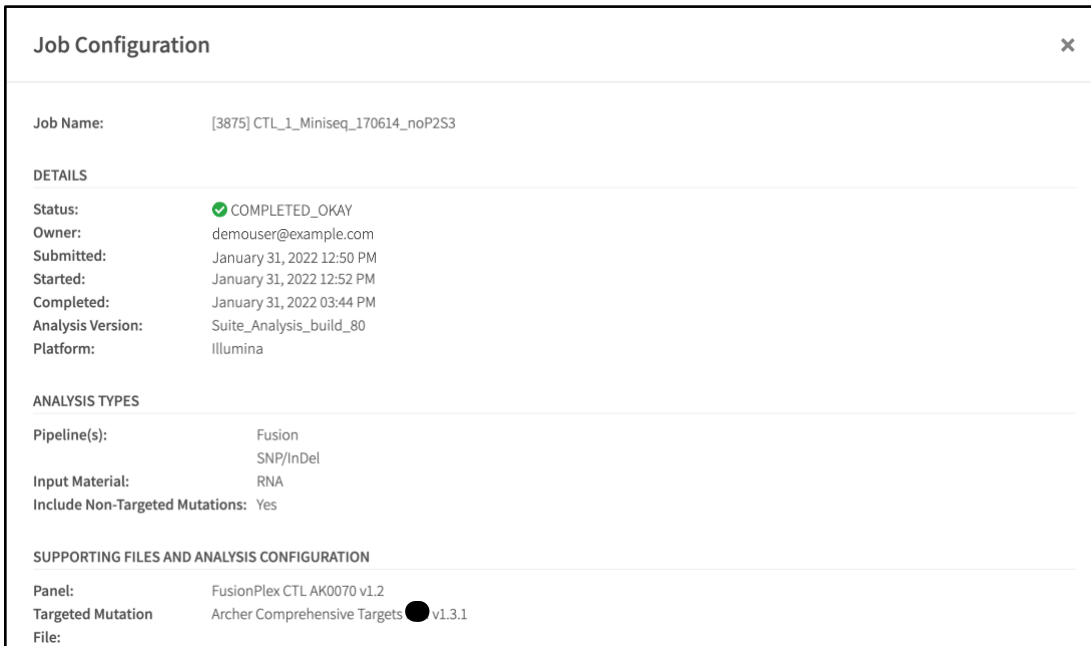
### Configuration

Click on “Configuration” (🔗) in the more-actions menu to see what parameter values (user settings) were used for the processing of the job:



Job Configuration settings			
[3875] CTL_1_Miniseq_170614_noP253	demouser@example.com	Pipeline(s): SNP/InDel	COMPLETED_OKAY
Configuration	Suite_Analysis_build_80	Input Material: Fusion	FusionPlex CTL AK0070 v1.2
View Logs	January 31, 2022 12:50 PM	Include Non-Targeted: Yes	Archer Comprehensive Targets v1.3.1
Clone Job	January 31, 2022 12:52 PM	Mutations:	
Rerun Job	January 31, 2022 03:44 PM		
Job Reports	elas low fusion support	Pipeline(s): Fusion	COMPLETED_OKAY
Download Results		Input Material: RNA	FusionPlex Heme v2 AK0073 v1.6

### Modal View:



**Job Configuration** ✕

---

Job Name: [3875] CTL\_1\_Miniseq\_170614\_noP253

**DETAILS**

Status: ✔ COMPLETED\_OKAY

Owner: demouser@example.com

Submitted: January 31, 2022 12:50 PM

Started: January 31, 2022 12:52 PM

Completed: January 31, 2022 03:44 PM

Analysis Version: Suite\_Analysis\_build\_80

Platform: Illumina

**ANALYSIS TYPES**

Pipeline(s): Fusion  
SNP/InDel

Input Material: RNA

Include Non-Targeted Mutations: Yes

**SUPPORTING FILES AND ANALYSIS CONFIGURATION**

Panel: FusionPlex CTL AK0070 v1.2

Targeted Mutation: Archer Comprehensive Targets v1.3.1

File:

The default values of each parameter are also shown. If any settings differ from the default, they will be indicated (see Identify Non-default Settings for further details):

ANALYSIS SETTINGS	
<b>General Analysis Settings</b>	
Min Average Unique DNA and AMBIG Start Sites per GSP2	50
Min Average Unique RNA Start Sites per GSP2 Controls	10
Error Correction	On
Read Depth Normalization	System Default: 3500000 0
Barcode Length	8

Abbreviated list shown here

### View 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 "View Logs" (📄) in the three dots menu:

Job	Analysis Types	Job Status	Assay Targets
<b>[3875] CTL_1_Miniseq_0614_noP2S3</b> Configuration: demouser@example.com View Logs: Suite_Analysis_build_80, January 31, 2022 12:50 PM Clone Job: January 31, 2022 12:52 PM Rerun Job: January 31, 2022 03:44 PM Job Reports Download Results	<b>Pipeline(s):</b> SNP/InDel, Fusion, RNA <b>Input Material:</b> RNA <b>Include Non-Targeted Mutations:</b> Yes	COMPLETED_OKAY	FusionPlex CTL AK0070 v1.2 Archer Comprehensive Targets v1.3.1
elas low fusion support	<b>Pipeline(s):</b> Fusion <b>Input Material:</b> RNA	COMPLETED_OKAY	FusionPlex Heme v2 AK0073 v1.6

A tabbed page will show the Primary Workflow log, the Summarization and Staging log, and the Workflow Configs log. Review the "Primary Workflow" log files first, and then the "Summarization and Staging" log files, since this could contain error messages.

In addition, click the "Download Raw File" link to download the main log file to the local computer:

### Analysis Logs: 3875

Primary Workflow
Summarization and Staging
Workflow Configs
Download main job log file

1.log.stderr.txt -

Download Raw File

SHOWING THE LAST 40 LINES

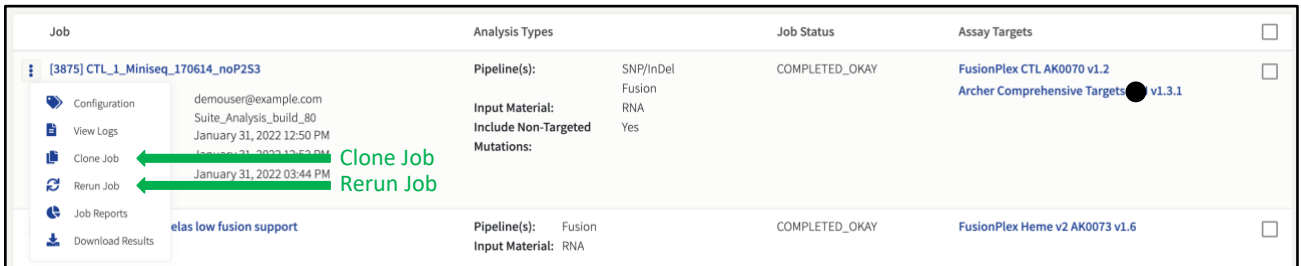
```
[2022-01-31 13:35:23,934] master_workflow.py [CHECKPOINT] > Archer Analysis Pipeline
[2022-01-31 13:35:23,948] master_workflow.py [CHECKPOINT] > Sample QC
[2022-01-31 13:35:24,200] chewbacca_analysis.py [INFO] > Look for GSP2 at read start with 0bp chew back
[2022-01-31 13:35:28,991] chewbacca_analysis.py [INFO] > Look for GSP2 at read start with 1bp chew back
[2022-01-31 13:35:33,856] chewbacca_analysis.py [INFO] > Look for GSP2 at read start with 2bp chew back
[2022-01-31 13:35:38,619] chewbacca_analysis.py [INFO] > Look for GSP2 at read start with 3bp chew back
[2022-01-31 13:35:43,392] chewbacca_analysis.py [INFO] > Look for GSP2 at read start with 4bp chew back
[2022-01-31 13:35:47,911] chewbacca_analysis.py [INFO] > Look for GSP2 at read start with 5bp chew back
[2022-01-31 13:35:52,091] chewbacca_analysis.py [INFO] > Summarizing and plotting
```

Contact [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com) for further assistance and troubleshooting.

### Clone Job and Rerun Job

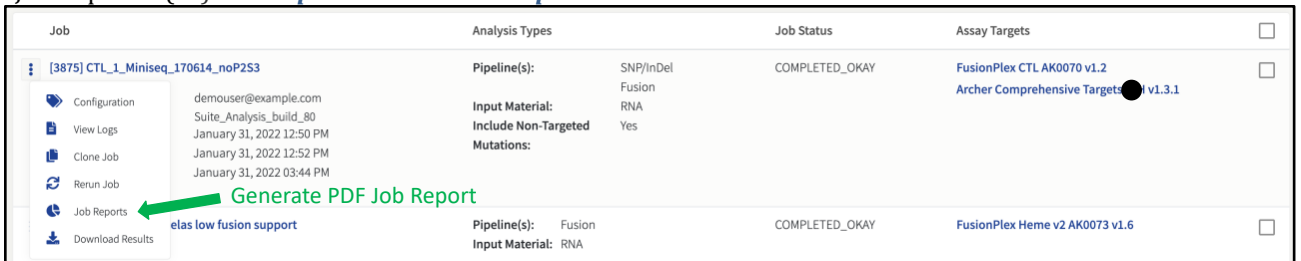
If the results of a run are not satisfactory, it may be useful to re-run an analysis with different analysis settings. This can be accomplished by either making a clone of the original analysis or by re-running the original analysis. The difference between the two options is that a clone will be a copy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

To clone an analysis, select “Clone Job” (📄), and to re-run an analysis, select “Rerun Job” (🔄) in the three dot menu:



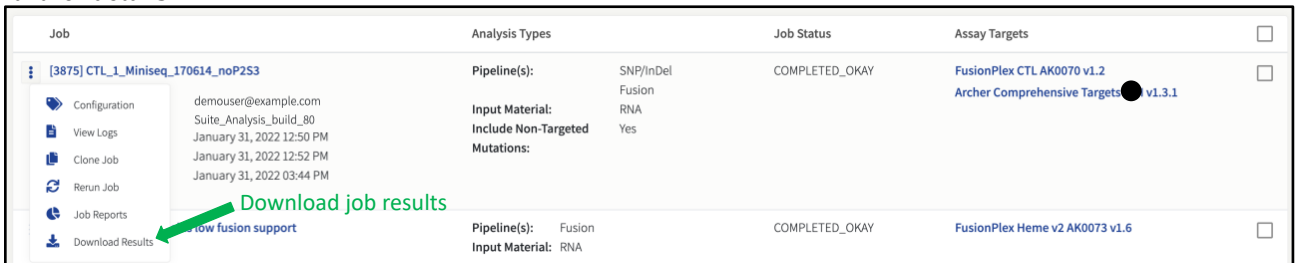
### Job Reports

To generate a PDF report showing the QC results and a summary of the analysis results for each sample, select “Job Reports” (📄). See [Reports and Other Output Files](#) for further details:



### Download Results

To download analysis results for the job, select “Download Results” (📄). See [Output From ‘Download All Files’](#) for further details:



### Job Details Page (Sample Summary)

From the **Sign-out Batches, Analyzed Samples, or Jobs** page, click on the job ID/name link to see the Job Details page. The page will show a list of all samples in the jobs, 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 job was set up (see [Starting a Job: New Job Page](#) for further details):

Job: MP22-66 Job Options ▾

Sample Summary RNA Expression

Low Med High All

Sample Name	Assay Results	QC Results	Reportable Mutations
MP22-66F_S13_R1_001 Low Med High Unassigned <input type="checkbox"/> Review Complete	Fusions 21 Oncogenic 2 Isoforms	Sample QC Fusion QC	

## Download Files

To download files for the job, select **“Download Results”** (📄). This will open a pop-up screen that will allow manual selection of specific (or all) files that can then be downloaded as a single ZIP.

All files, including summaries, logs, etc., that were generated for the job are available to be downloaded as specified by the user, by selecting **“Download Results”** (📄). See [Downloadable Data](#) for further details.

Job: MP22-66 Job Options ▾

Sample Summary RNA Expression

Low Med High All

Sample Name	Assay Results	QC Results	Reportable Mutations
MP22-66F_S13_R1_001 Low Med High Unassigned <input type="checkbox"/> Review Complete	Fusions 21 Oncogenic 2 Isoforms	Sample QC Fusion QC	

Download results

Configuration  
 View Logs  
 Clone Job  
 Rerun Job  
 Job Reports  
Download Results  
 Archive Job

From this view all files produced by the job can be browsed and selected, pre-organized into tabs with commonly downloaded files. After selecting at least one file, the selection can then be downloaded as a single ZIP file.

### Download Files ✕

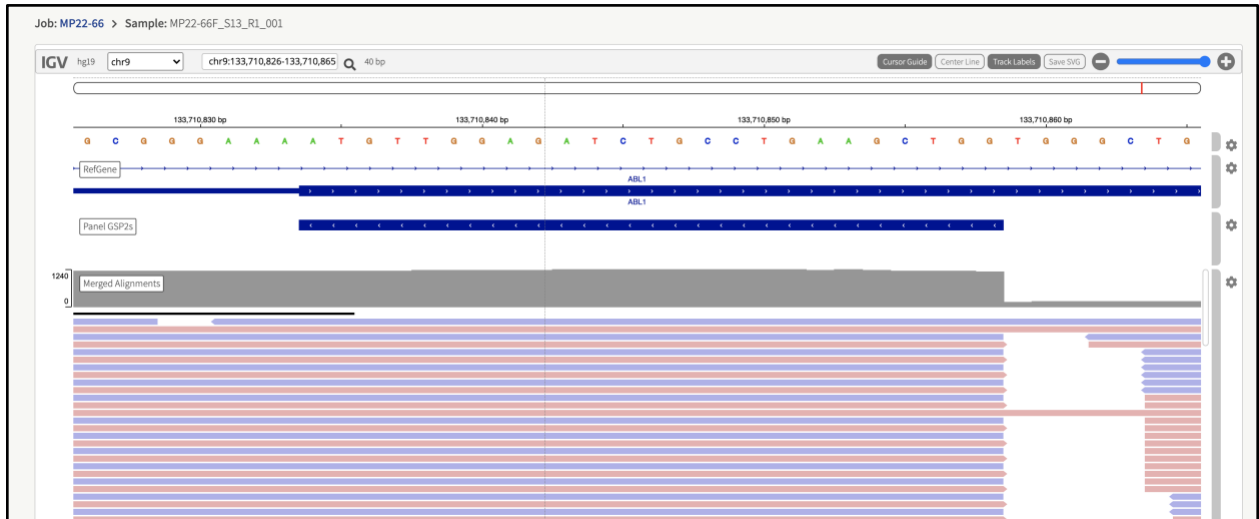
Results Files
Aggregated Results Files
Visualization Files
Diagnostic Files
All

- Select All
- MP22-66F\_S13\_R1\_001\_full\_results.txt
- summaries/Summary.deduped.counts.txt
- summaries/Summary.deduped.counts\_pct.txt
- summaries/Summary.expression\_summary\_table.txt
- summaries/Summary.non\_deduped\_counts.txt
- summaries/Summary.non\_deduped\_counts\_pct.txt
- summaries/Summary.per\_sample\_columns\_expression\_summary\_table.txt
- summaries/Summary.r\_and\_d\_results.txt
- summaries/Summary.run\_agged\_fusion\_metrics.txt
- summaries/Summary.run\_fusion\_metrics.txt

Close
Download

## Visualize Sample Data

From the **Sample Summary** page, the “Visualize Sample Data” link (🔍) will open a window with IGV.js genome browser showing all the reads aligned to the human genome (hg19) that support the gene fusion and/or mutation calls:



The genome browser shows several different tracks:

**Reference Sequence** – A track showing the hg19 reference sequence with either the forward or reverse strand and option to show the three-frame translation for the current strand.

**RefGene** – A track showing the genes and each of their transcripts from the RefSeq database.

**Panel GSP2** – A track showing the position and direction of the gene specific primers.

**Variant ROI** – A track showing the defined region of interest for the sample (not available for fusions and structural variants).

**Preprocessed Alignments** – BAM coverage after filtering, InDel realignment and deduplication (not available for fusions and structural variants).

**Merged Alignments** – merged BAM coverage after filtering and duplication for fusions and structural variants.

**Cosmic** – A track showing the location and identifiers from the COSMIC mutation database.

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

For more information on usage, please refer to online documentation for IGV.js.

## Sample Details Page

Each sample name in the **Job Details** page is a link that will take you to the **Sample Details** page for that sample. Clicking on this link will show the summary results pages with more details. This page displays the detailed results of the analysis and QC metrics. A left-hand side menu provides all links to all results, review/report status, and any other options relevant to the sample:

## Read Statistics

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

Read Statistics				
Total Fragments	Fragments with Complete Adapter		Number of Reads After Trimming Adapters	
3,500,000	3,341,193 (95.5%)		2,918,781	
<a href="#">Export Data (tsv)</a>				
Fragment Type	Total Fragments	Mapped	Pass Alignment Filter	On Target %
All Fragments	2,918,781 (100.0%)	2,912,359 (99.8%)	99.8%	96.2%
Unique Fragments	1,807,808 (61.9%)	1,805,453 (99.9%)	99.9%	95.0%
<a href="#">Export Data (tsv)</a>				
DNA/RNA Statistics				
Type	DNA Reads	RNA Reads	Ambiguous Reads	
All Fragments	191,115 (6.8%)	2,132,901 (76.1%)	477,946 (17.1%)	
Molecular Bins	137,940 (8.0%)	1,323,653 (77.2%)	254,033 (14.8%)	
Average Molecular Bins per GSP2	127.46	1,253.46	240.56	
Unique Start Sites	48,060 (26.7%)	127,492 (70.9%)	20,769 (11.5%)	
Average Unique Start Sites per GSP2	45.82	185.12	21.15	
Average Unique Start Sites per GSP2 Control	51.62	261.75	20.12	
<a href="#">Export Data (tsv)</a>				

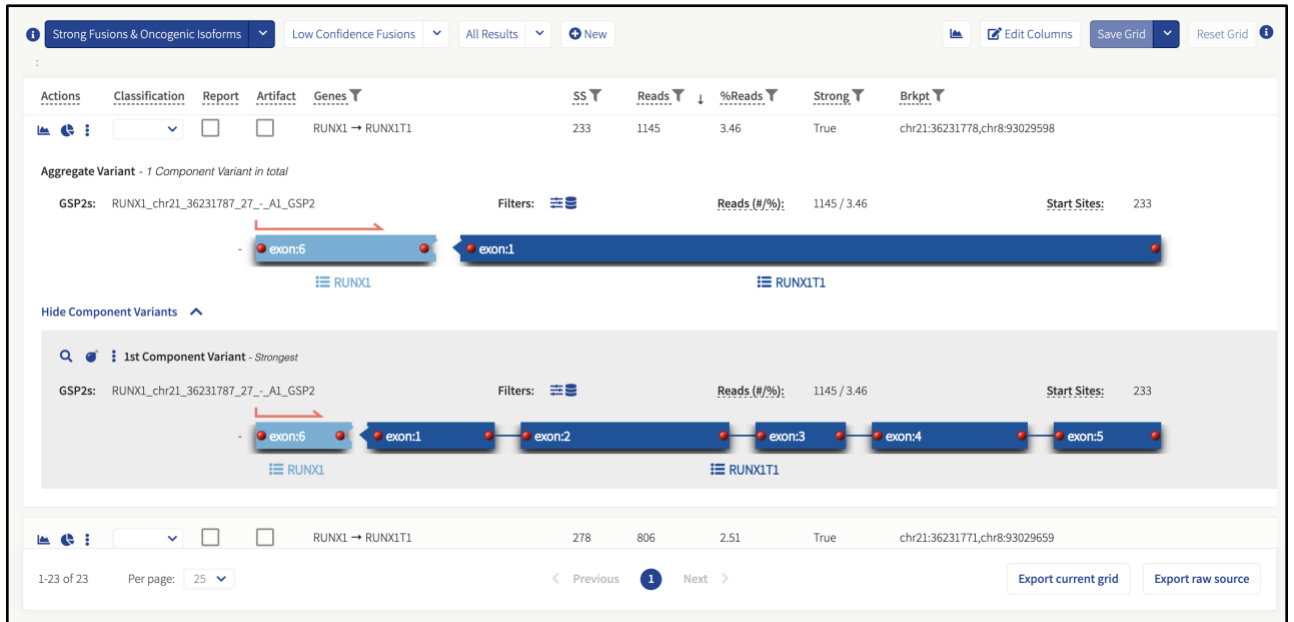
QC Statistics			
Avg. Unique DNA And Ambiguous Start Sites Per GSP2	Avg. Unique RNA Start Sites Per GSP2 Control		
67.69	261.75 <span style="color: green;">✔</span>		
Miscellaneous Statistics			
On Target Deduplication Ratio			
1.63:1			
<a href="#">Export Data (tsv)</a>			
DNA/RNA Fragment Lengths			
DNA Fragment Median Length (bp)	DNA Fragment Mean Length (bp)	RNA Fragment Median Length (bp)	RNA Fragment Mean Length (bp)
159.0	199.3	183.0	183.1
<a href="#">Export Data (tsv)</a>			

Most of the metrics contain help texts (mouse-over the metrics) and a full description can be found in [Error! Reference source not found.](#)

### Assay Targets

The assay targets page contains detailed coverage information for each GSP2 in the panel. It contains information about the number of unique fragments based on the number of unique molecular barcodes, as well as the number of unique start sites, and lastly raw (non-deduplicated) reads. The data are also separated by the type of molecule from which the fragments originate (DNA, RNA or AMBIGUOUS). In addition to the coverage for the actual 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:





The initial displayed fusion is aggregated by genomic breakpoint, showing where the GSP2 primer(s) are positioned on the exon. This fusion can be viewed by component variants. The aggregated visualization highlights the breakpoint by showing the 2 contigs adjacent to the breakpoint and lists all the primer(s) that support the fusion. Additional details are shown in the per fusion consensus visualizations listed under the aggregated fusion.

For each aggregate variant, the following options are available in the **Fusion Grid** and in the more-actions menu:

- Expand structural variant – Toggles the display of the Aggregate Variant view.
- Create isoform report – Shows a report of the isoform, including information on category, gene, breakpoints, whether the isoform is found in Quiver, how many samples this isoform has been observed and links to the jobs, and a link to export the report as a PDF.
- View History – View entire review history. This includes change in status and/or comments added. Previous additions cannot be modified.
- View Comments – View all comments. Previous comments cannot be modified.
- Show Component Variants – Displays component variants for a given aggregate variant.

For each component variant, the following options are available in the **Fusion Grid** and in the more-actions menu:

- 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 protein database.
- Translation - Opens a new window showing the results of the protein translation prediction of the gene fusion product.
- View Evidence – Opens a new window with information on the GSP2's that support the component variant.
- Export Visualization – Exports a png of the aggregate variant.
- Mark as False Positive – Opens a confirmation window to indicate that isoforms with this signature (including this one, others in this job's samples, and in all future jobs) will no longer show up in the strong or weak evidence category and will instead be filed under “User-Annotated False Positives” in the appropriate strong or weak tab.



- Mark as True Positive – Opens a confirmation window to indicate that isoforms with this signature (including this one, others in this job's samples, and in all future jobs) will always show up in the strong evidence category regardless of any other metrics associated with the isoform.

### *Key Metrics for Strong and Low Confidence Fusion/Isoform Candidates*

**Reads and %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.

**SS (Start Sites)** – The number 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.

**Brkpt (Breakpoint)** – The hg19 chromosomal breakpoint locations for the fusion as deduced from the RNA. This does NOT represent the exact breakpoint at the DNA level. (Coordinates at the start of chr1 are placeholders for unaligned bases).

### *Strong Fusions & Oncogenic Isoforms and Low Confidence Fusions*

Gene fusions, oncogenic 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 Fusions & Oncogenic Isoforms filter set and a Low Confidence Fusions filter set, so that all candidates can be viewed, regardless of whether they are called strong or not. The Strong Fusions & Oncogenic Isoforms calls should be considered the calls that are being made by the software. Low Confidence Fusions calls are primarily used for troubleshooting false negatives in the Strong Fusions & Oncogenic Isoforms category and are not to be considered positive calls. There are several different criteria for the categorization of weak vs. strong, as described in the subsequent sections.

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 unique (de-duplicated) 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 Valid Structural Variation” setting 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 marked indicated with the bull’s eye icon (👁) to indicate the fusion is in the Quiver database and the breakpoint is an exact match.

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



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 RNA reads spanning the breakpoint.

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

Changing the “Structural Variation Percent of GSP2 Reads” setting under “Structural Variation 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 Valid Structural Variation" parameter under "Structural Variation 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 Negative Evidence Criteria section to be called Strong Evidence. With this exception, if any of the negative evidence criteria are met, the candidate will be called Weak.

### *Negative Evidence Criteria*

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 than 3bp from the GSP2, 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 performed 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 Read-Through** – Fusion transcripts of interest are generally derived from a genomic translocation event. However, fusion transcripts can also arise from failure of the cell 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 read-through events are indicated with the (⊙) icon and placed, by default in the **STRONG** category.
  - Transcriptional read-through events are placed in the **STRONG** category by default because by representing actual molecules produced in cells, they are technically not false positives.
  - Transcriptional read-through 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 “Structural Variation Analysis 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.

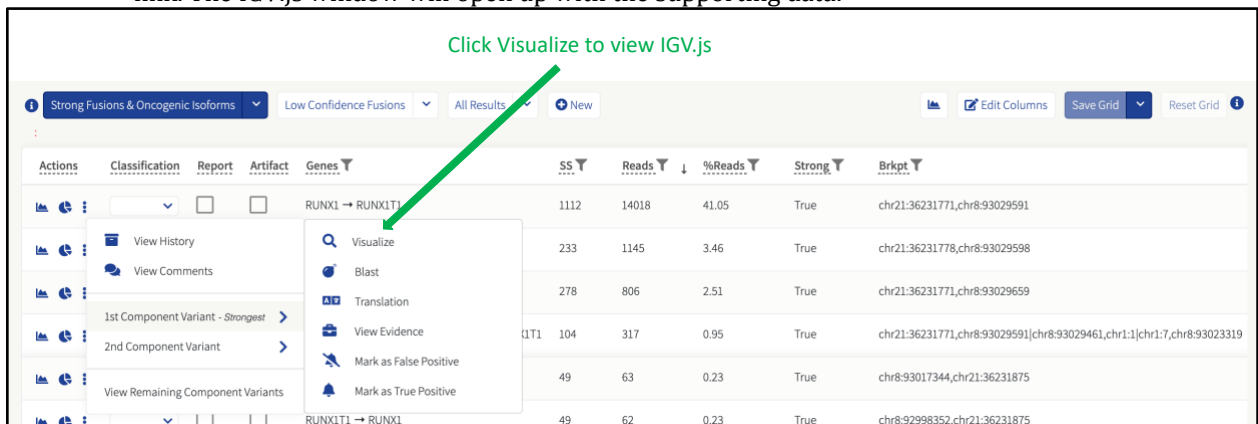
### User-Annotated False and True Positives

It is possible to override the positive and negative evidence criteria by annotating specific fusions (according to breakpoint). Such annotation will be recorded and used in a manner similar to Quiver entries, and will trump all other criteria:



### Visualization of Fusion Candidate-Supporting Reads in IGV.js

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 three-dot menu, hover over the preferred Component Variant and then the [Visualize](#) link. The IGV.js window will open up with the supporting data.



The IGV.js genome browser will open in its own tab in the web browser.



The genome browser shows the constructed consensus sequence of the fusions as the reference sequence to which the de-duplicated reads supporting the fusion are mapped.

### *Tracks in the IGV.js View for Fusions*

There are several tracks in IGV.js 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 Annotations** – 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.




Fusions are not aligned against hg19 directly. First, a fusion reference is created from the hg19 annotations. This is done in order to see reads aligned across the breakpoint.

**Variants Observed** – Any differences between the consensus sequence and the reference are listed here as a VCF track. In some cases, there can be a large insertion in the alignments that typically corresponds to a segment of the fusion contig that could not be annotated to the genomic reference.

**Alignments** – The unique (de-duplicated) reads supporting the fusion.

### *Protein Translations*

For gene fusions, the Archer Analysis software determines if the resulting protein is in or out of frame using all combinations of the fusions candidate isoforms and creates an amino acid sequence that can then be subsequently BLAST'ed 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.

In Frame	5' Gene	5' Transcripts	3' Gene	3' Transcripts
True	RUNX1	XM_005261068.3	RUNX1T1	XM_017013938.1
<a href="#">BLAST Sequence</a> <a href="#">Query Pfam</a> MPAAPRGPQAQGEAAARTSRDASTSRRFTPPSTALSPGKMSEALPLGAPDAGAALAGKLRSGDRSMVEVLADHPGELVRTDSPNFLCSVLPHTWRCNKTLPIAFK				
True	RUNX1	XM_005261068.3	RUNX1T1	XM_017013940.1
<a href="#">BLAST Sequence</a> <a href="#">Query Pfam</a> MPAAPRGPQAQGEAAARTSRDASTSRRFTPPSTALSPGKMSEALPLGAPDAGAALAGKLRSGDRSMVEVLADHPGELVRTDSPNFLCSVLPHTWRCNKTLPIAFK				
True	RUNX1	XM_017028487.1	RUNX1T1	NM_001198629.2 NM_001198634.2 NM_001198630.2 XM_024447317.1 NM_001198631.1 NM_175636.2

## RNA Expression Pipeline Outputs

### Outputs of the RNA Expression Pipeline

1. A relative expression heat map
2. A log<sub>2</sub> transformed .tsv data frame of the values used to generate the heat map
  - a. logshift\_rna\_expression\_visualization.tsv—available for download via the “Download Source (TSV)” link at the bottom of the RNA Expression tab on each job’s landing page
3. An untransformed .tsv data frame of the per sample normalized expression values
  - a. rna\_expression\_visualization.tsv—available for download via the full job folder: Job Options > Download Results > All > (Expand Job Number)

### Per Primer Pair RNA Expression Normalization

Normalized RNA expression values are calculated for each sample individually on a per primer pair basis by dividing the unique RNA reads for each GSP2 by a normalization value. Normalized values can be compared across analysis jobs.

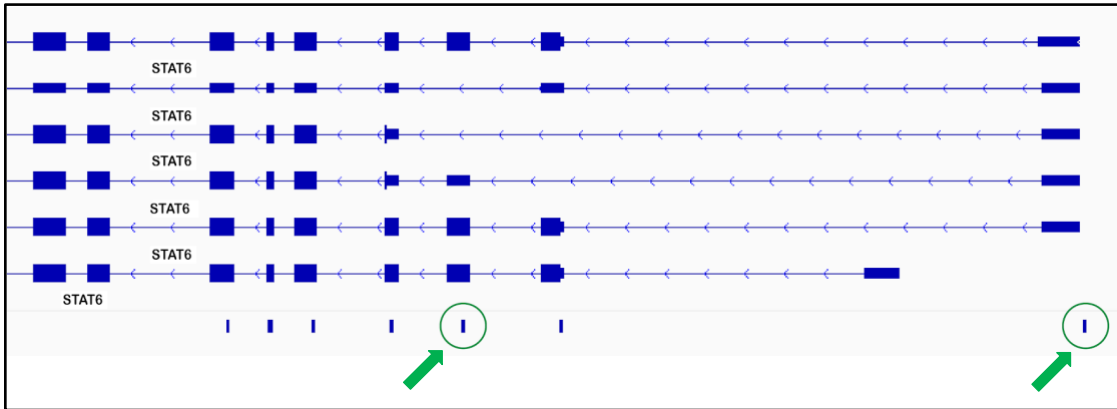
The normalization value is based on the control genes included in the panel (CHMP2A, GPI, RAB7A, and VCP). It is calculated by taking the geometric mean of the median unique RNA reads +1 (to protect against 0-value errors in the geometric mean calculation) for all GSP2s in each control gene individually.

The normalized per GSP2 values from this calculation are written out to the rna\_expression\_visualization.tsv data frame. These values are log transformed (logshift\_rna\_expression\_visualization.tsv) and binned to generate the heatmap visualizations.

### RNA Expression Normalization Primer Pair Selection

GSP2s are not flagged specifically for inclusion in the RNA Expression outputs (heatmaps, logshift\_rna\_expression\_visualization.tsv, rna\_expression\_visualization.tsv) at this time. This means that because of the way that Archer Analysis bins reads (RNA, DNA, AMBIG), a subset of GSP2s are **expected** to consistently display very low, 0, or near 0 normalized values. The most common reason for low values are GSP2s that do not generate unique RNA reads. RNA reads must span an exon junction to be classified as such (see below).

Read count variability across GSP2s can also sometimes be explained biologically. For instance, primers that target specific isoforms (see below) will only generate reads if that isoform is expressed.

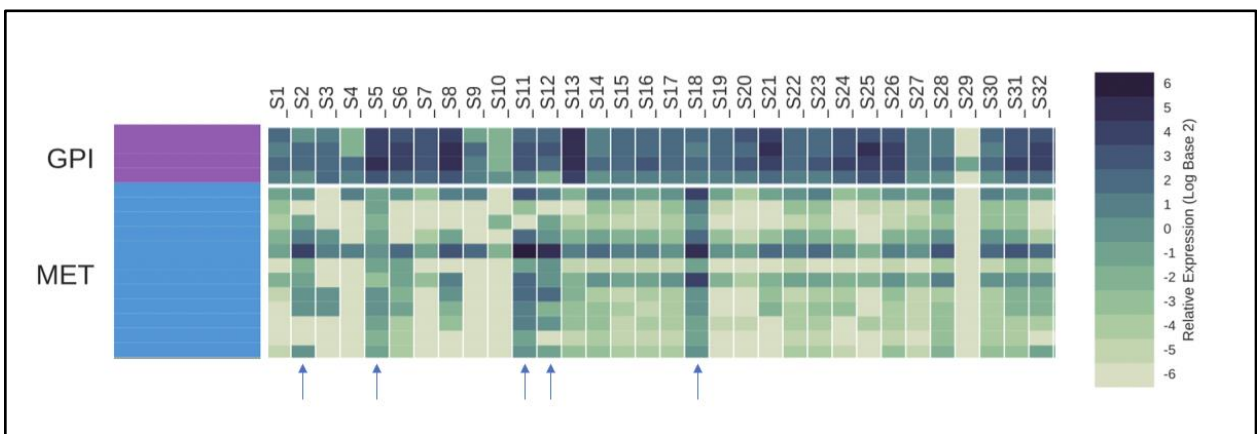


This screenshot shows all 5 STAT6 RefSeq isoforms. In green circles are two primers that target a subset of the isoforms. The primer in the second green circle (on the right) will always generate 0 (or very few) RNA reads because it amplifies into the 5' UTR (its amplicon does not cross an exon-junction).

### Qualitative Analyses

The RNA Expression tab on the Job Summary page contains a heatmap that visualizes normalized per sample RNA expression.

- Samples are organized by column
- Genes are organized by row
  - Rows within a gene display the binned, normalized expression value for each GSP2 targeting that gene, sorted 5' to 3' by transcript (see below)



Heatmap representing the Log<sub>2</sub> transformed, normalized RNA expression values.

### Quantitative Analyses



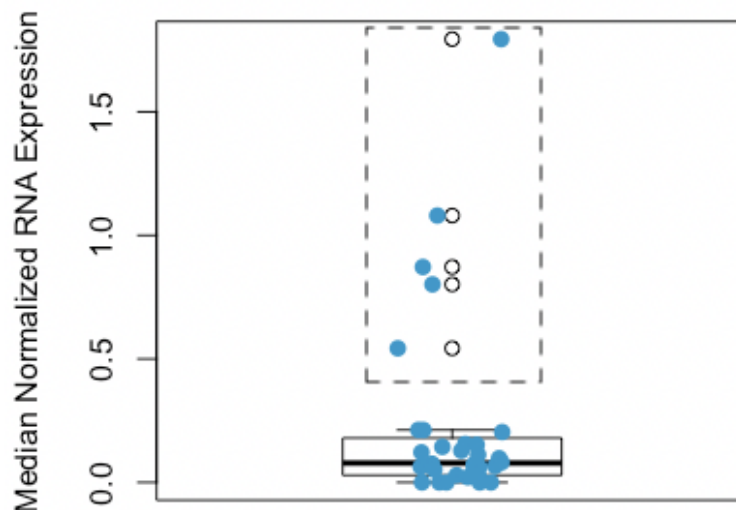
Customers and internal teams exploring custom RNA expression analysis pipelines often begin with the `rna_expression_visualization.tsv` data frame. This output contains the normalized per GSP2 expression values for each primer, and each sample.

One example of how this data has been manipulated internally is below:

1. Per gene values were consolidated down to a single value by determining the median for all GSP2s that target a single gene (MET in this example).

Gene	Primer	Sample_1	Sample_2	Sample_3	Sample_4	Sample_5	Sample_6	Sample_7
MET	MET_chr7_116339157_20_-_A1_GSP2	0.5126573	0.9619538	0	1.4107311	0.651508	1.2209975	0.116598
MET	MET_chr7_116340225_29_+_A1_GSP2	0.1386073	0	0	0	0.488631	0	0.0062186
MET	MET_chr7_116380008_24_-_A1_GSP2	0.0868669	0.6413025	0	0	0.325754	0	0.0202103
MET	MET_chr7_116380910_24_-_A1_GSP2	0.2264236	1.9239075	0.8086369	0	0.5972157	0	0.1041609
MET	MET_chr7_116395414_23_-_A1_GSP2	0.6170875	12.184748	2.4259108	2.4687795	0.8686773	4.7822402	0.6389572
MET	MET_chr7_116411556_24_-_A1_GSP2	0.0251582	0.3206513	0	0	0.651508	0.6104987	0.021765
MET	MET_chr7_116411645_28_+_A1_GSP2	0.3270564	0.6413025	0	0	0.1085847	0.7122485	0.1274805
MET	MET_chr7_116411907_25_-_A1_GSP2	0.0370252	1.282605	0.8086369	0	1.0315543	0.2034996	0.0279835
MET	MET_chr7_116414939_29_-_A1_GSP2	0.0180379	0.9619538	0.8086369	0	0.7600927	0.4069992	0.017101
MET	MET_chr7_116417450_29_-_A1_GSP2	0.0213607	0	0	0	0.4343387	0.1017498	0.0202103
MET	MET_chr7_116418834_27_-_A1_GSP2	0.0204114	0	0	0	0.325754	0.1017498	0.0139918
MET	MET_chr7_116435945_25_-_A1_GSP2	0.0379746	0.9619538	0	0	0.4343387	0.1017498	0.0202103
		Sample_1	Sample_2	Sample_3	Sample_4	Sample_5	Sample_6	Sample_7
	MET_MEDIAN	0.0624208	0.8016281	0	0	0.5429233	0.1526247	0.0209876

2. In the absence of a normal cohort, values were compared across the entire cohort.
  - a) Given a high enough n, a clinical cohort can be used to establish a threshold for “increased” expression.
  - b) The threshold was defined in this example simply as the upper whisker ( $=\min(\max(x), Q_3 + 1.5 * IQR)$ ).



3. Samples that display expression above this threshold (dashed-line box above, indicated with arrows in Figure 2) display RNA expression consistent with MET amplification (in this example) and were investigated further.
  - a) Note: comparator data was not used to refine the threshold for this example set. However, this will likely be required to balance sensitivity/specificity for expression-based analyses.

### SNP/InDel Variants (RNA and DNA)

The 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).

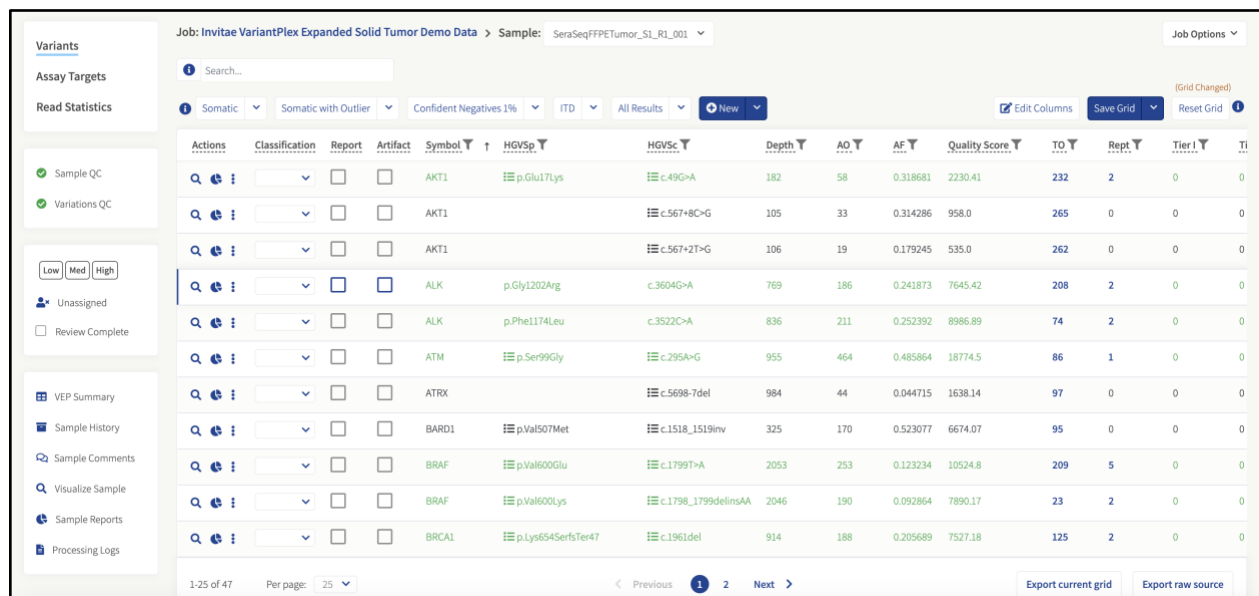
### Targeted vs. Non-Targeted Calling

In the “Confident Negative 1%” and “All Results” sections of “Variants”, each of the targeted mutations will be listed – not just those that were found to be present in the sample – with their Classification, Report, and Artifact values as “NA”.

The variant grid 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 “Include Non-targeted” is turned OFF. The names of the variants consist of the gene name and amino acid mutation as defined in the target VCF file.

### General Information About the Variant Grid

The Variants tab in the **Sample Details** page will show the results of the variant detection as shown below:



Actions	Classification	Report	Artifact	Symbol	HGVS	HGVS	Depth	AO	AF	Quality Score	TO	Rept	Tier 1	TI
				AKT1	p.Glu17Lys	c.49G>A	182	58	0.318681	2230.41	232	2	0	0
				AKT1		c.567+8C>G	105	33	0.314286	958.0	265	0	0	0
				AKT1		c.567+2T>G	106	19	0.179245	535.0	262	0	0	0
				ALK	p.Gly1202Arg	c.3604G>A	769	186	0.241873	7645.42	208	2	0	0
				ALK	p.Phe1174Leu	c.3522C>A	836	211	0.252392	8986.89	74	2	0	0
				ATM	p.Ser99Gly	c.295A>G	955	464	0.485864	18774.5	86	1	0	0
				ATRX		c.5698-7del	984	44	0.044715	1638.14	97	0	0	0
				BARD1	p.Val507Met	c.1518_1519inv	325	170	0.523077	6674.07	95	0	0	0
				BRAF	p.Val600Glu	c.1799T>A	2053	253	0.123234	10524.8	209	5	0	0
				BRAF	p.Val600Lys	c.1798_1799delinsAA	2046	190	0.092864	7890.17	23	2	0	0
				BRCA1	p.Lys654SerfsTer47	c.1961del	914	188	0.205689	7527.18	125	2	0	0

The HGVS (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 using the Variant Effect Predictor tool, developed by the EBI at ENSEMBL (<http://www.ensembl.org/info/docs/tools/vep>).



### CNVs (DNA Only)

CNV detection can be selected in addition to other assay types if the user has a panel that is designed for CNV detection. This assay uses one or more control samples, if present, to compare copy number in normal versus diseased samples. If no controls are present, all samples within the job are used as a baseline. Selection of the proper normal sample(s) is described in [Selecting Normalization Parameters for CNV](#) section. The analysis software uses this information to produce both a readout and a visualization of copy number, along with a p-value to convey confidence in the calls.

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, and could result in more accurate CNV calling. Only the samples that are marked as tumor samples will show CNV results. Samples marked as normal will not display data in the “**Assay Results**” column. For DNA Copy Number Variation analyses, the **Job Detail** page will list the results of the CNV analysis by highlighting the genes that show significant copy gains or losses:

Sample Name	Assay Results	QC Results	Reportable M
21-SeraSeq-6CNV-DNA_S21_L001_R1_001 <input type="button" value="Low"/> <input type="button" value="Med"/> <input type="button" value="High"/> <input checked="" type="checkbox"/> Unassigned <input type="checkbox"/> Review Complete	Variants: 5 DNA Structural: 0 Variations: 2 Copy Gains: 2 Copy Losses: 0	<input checked="" type="checkbox"/> Sample QC <input checked="" type="checkbox"/> Variations QC <input checked="" type="checkbox"/> Anomaly QC	

### CNV Gene/Group Plots

Copy number calculations are performed for every primer which has CNV enabled in a panel, and each primer is shown as a dot in the CNV grid. These primer-by-primer results are then summarized and displayed for larger regions of the genome. By default, the plotted data is color coded by gene.



*CNV plots with primers grouped by gene*

If the gene/group toggle is switched to group then in addition to the by gene color scheme, the custom-made primer groups are split into a separate color. CNV groups are defined in the panel gtf file under the attribute

column as: CNV\_group "groupName". CNV primer groups are classified as Gain or Loss for the entire group, or Partial Gain or Loss where at least three primers in a group show significant copy number variations.



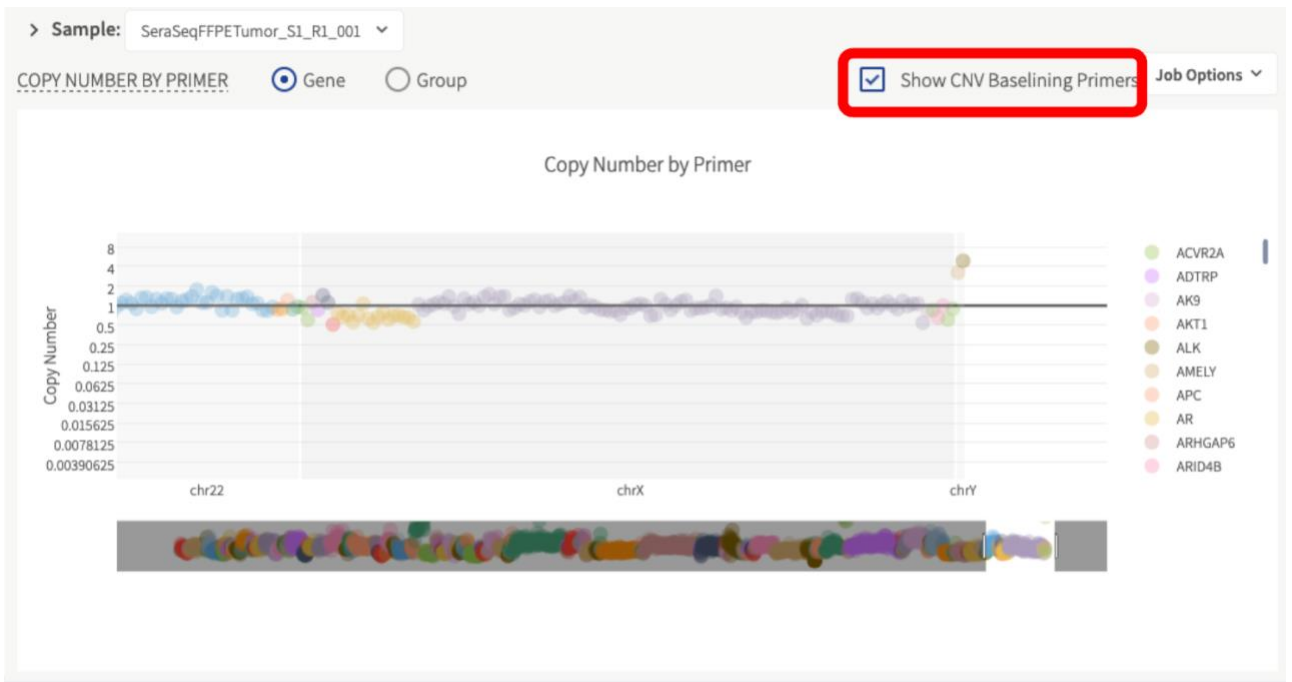
*CNV plots with primers grouped by user-defined groupings*

The CNV grid contains the summarized CNV results for both genes and groups regardless of the toggle for the plot.

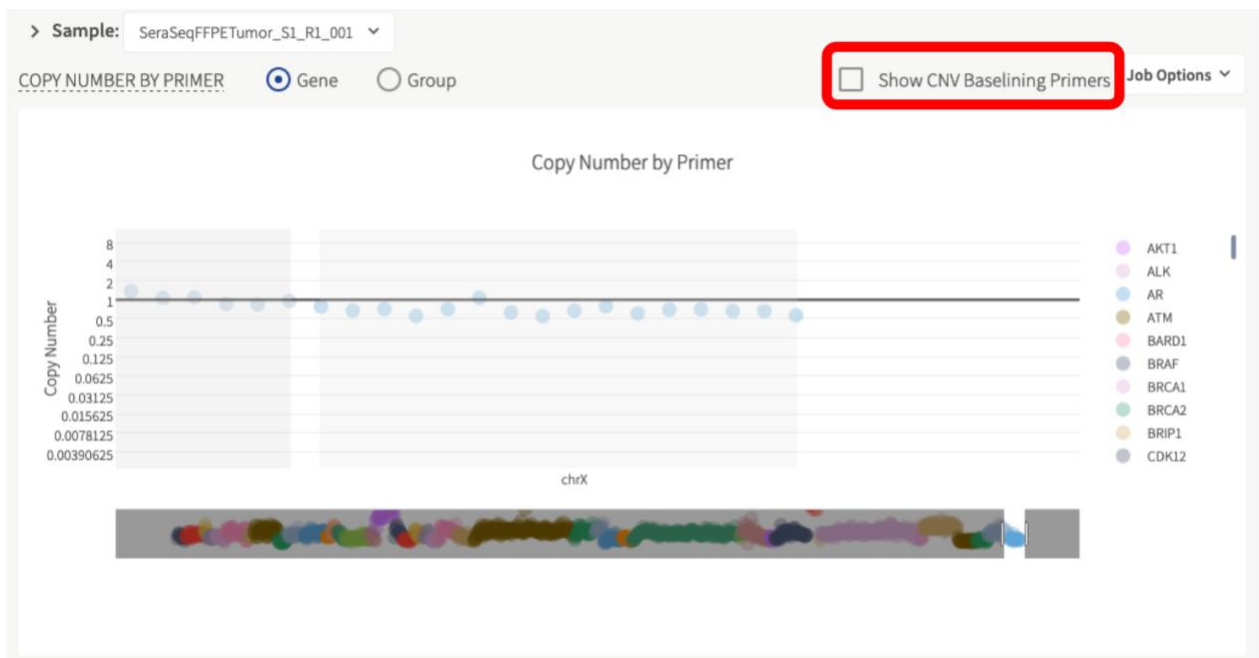
### Showing CNV Baseline Primers

For the primer-by-primer copy number plot there is an additional "Show CNV Baseline Primers" option. The CNV relative copy number calculation is made more robust by including many data points for the baseline. Because of this, CNV information is calculated across regions of the panel that may not be of interest to the user. This additional CNV data can be displayed in the plot by turning on the "Show CNV Baseline Primers" option.

Whether or not the primer will be displayed by default or not is dictated by information in panel gtf attribute section. Here, each primer which is used in CNV calculations contains a value in the function tag of either "CNV" alone or both "CNV" and "CNV\_NO\_DISPLAY". Those primers with both "CNV" and "CNV\_NO\_DISPLAY" tag will only appear in the plot when "Show CNV Baseline Primers" is selected. CNV calls will not be made for primers with the additional "CNV\_NO\_DISPLAY" tag, but they will be used in creating the CNV baseline.



*CNV plots with baselining primers included in the visualization*



*CNV plots with baselining primers hidden from the visualization*

### MSI (DNA Only)

MSI detection can be selected in addition to other assay types if the user has a panel that is designed for MSI detection. The MSI algorithm calculates the Shannon diversity (instability score) of microsatellite repeat lengths in unique reads for each of 114 microsatellite loci. Stability of each locus is determined by comparing the loci's instability score against instability scores from a baseline cohort of stable samples that are packaged with the

pipeline. The baseline samples are scaled to match sequencing depth at each loci of the interrogated sample. The proportion of loci determined to be unstable determines categorization of samples as microsatellite instability high (MSI-H), microsatellite instability intermediate, or microsatellite stable (MSS) based on the user's settings.

The MSI pipeline makes a call on microsatellite stability with for each sample based on the overall percentage of unstable microsatellite (MS) loci:

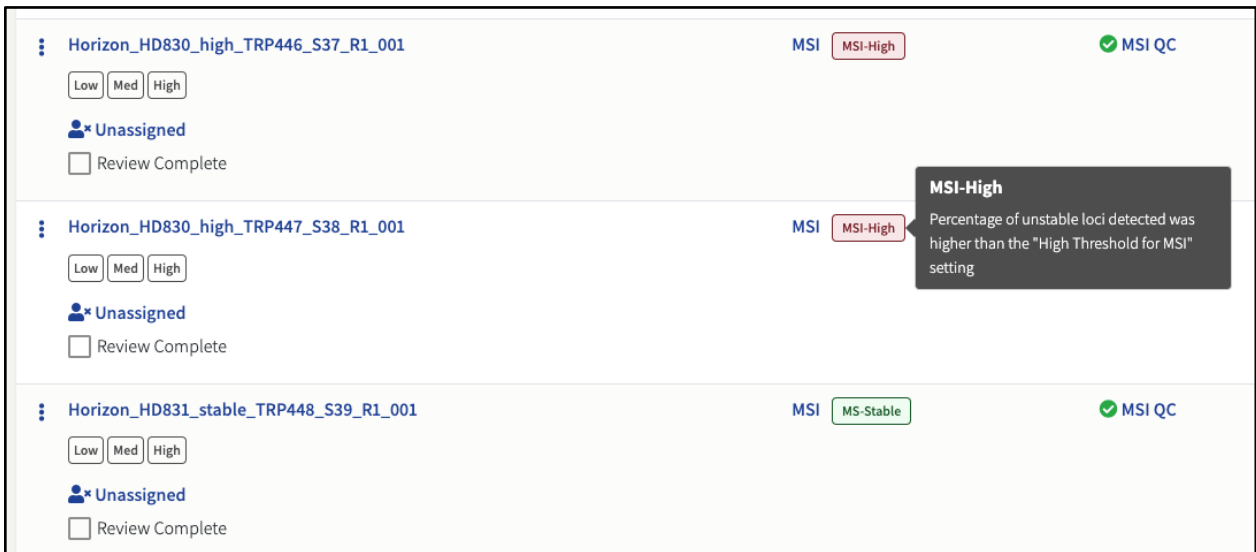
- **MS-Stable:** < 20% loci unstable
- **MS-Intermediate:** 20-30% loci unstable
- **MS-High:** >30% loci unstable

A per-MS loci read depth quality control check is made to ensure each loci meets a minimum 25 read depth threshold. An additional QC check is made to ensure that each sample meets a minimum of 35,000 total usable MS reads. Samples that fail to meet either of these QC checks will be called as **MSI-Unknown**.

**Note:** The default thresholds for the QC checks and MSI calling can be configured to user-specified values. See [MSI Pipeline Settings](#) for more information.

The analysis software produces readouts and visualizations of per MS loci classifications with p-values to convey confidence in the calls, as well as the overall sample classification for MSI status.

The **Job Detail** page will display the MSI call highlighted in green (MS-Stable), red (MSI-High), yellow (MSI-Intermediate), and the MSI specific QC status.



The screenshot displays a table of sample results with the following details:

Sample ID	MSI Status	MSI-QC Status
Horizon_HD830_high_TRP446_S37_R1_001	MSI-High	MSI QC (Pass)
Horizon_HD830_high_TRP447_S38_R1_001	MSI-High	MSI QC (Pass)
Horizon_HD831_stable_TRP448_S39_R1_001	MS-Stable	MSI QC (Pass)

A tooltip for the 'MSI-High' status indicates: "Percentage of unstable loci detected was higher than the 'High Threshold for MSI' setting".

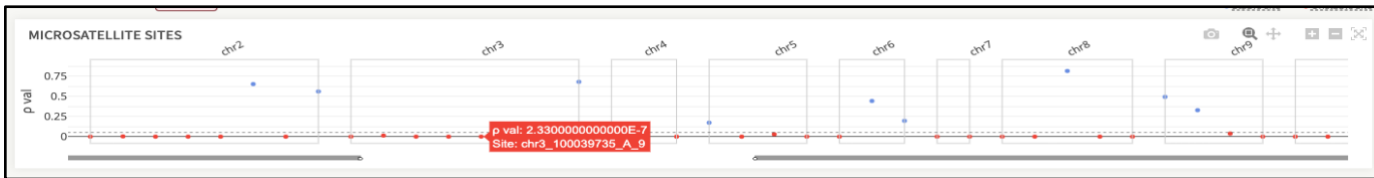
### MSI Plots and Grid

For a single sample, the MSI loci grid shows each locus based on its position in the chromosome, compared to the pval of the statistical test. The dotted line shows the pval alpha (currently set to 0.05).

Loci that pass the statistical test are shown in red indicating an unstable locus. Stable loci are shown in blue, and loci that did not pass QC (and not included in the analysis) are not shown.



Users can also zoom in on the MSI loci grid.



The results of each locus are shown in table format below the MSI loci grid.

Actions	Microsatellite	P Val	Stability
Q	chr4_55598211_T_25	0.0163	Unstable
Q	chr2_165551295_A_9	0.0467	Unstable
Q	chr14_73959703_T_16	0.0755	Stable
Q	chr2_47641559_A_27	0.0877	Stable
Q	chr17_27081792_T_16	0.105	Stable

## Output Files

There are two types of tab-separated files, generated as results of the MSI calling algorithm. Per job, there is a single file for all samples generated, titled *overall\_sample\_results.tsv*.

There is also *{sample\_name}\_loci\_results.tsv*, generated per sample. Both files can be found in the Results Files tab of the Download Results feature.

## Overall Aggregate Sample Results

The *overall\_sample\_results.tsv* file has a single row per sample and the following columns for metrics:

- **sample:** The name of the sample as gleaned from the R1 fastq.
- **unstable\_loci:** The number of loci that were flagged as unstable.
- **total\_loci:** The total number of loci used post-QC.
- **percent\_unstable:** The percent of unstable loci over the total observed.

- $(\text{unstable\_loci} / \text{total\_loci}) * 100\%$
- **QC:** Whether the sample passed or failed QC. See *Appendix* for more details on how QC is assessed.
- **MSI\_call:** Overall enumeration for the sample. Whether the sample in aggregate is MSI High, MSI Indeterminant, or MSI Stable. See *Appendix* for more details on how the MSI call is made.

### Individual Sample Per Locus Results

The `{sample_name}_loci_results.tsv` file has a single row per loci observed with the following columns for metrics:

- **MS\_loci:** The name of the locus.
- **pval:** The p-value assigned to that particular locus. The lower the p-value the more unstable it is (we currently use an alpha of 0.05 as a cutoff).
- **instability\_score:** The raw score assigned to the locus. The higher the score, the more unstable it is.
- **depth:** The number of MS usable reads per loci.
- **stability:** Locus stability call based on the user settings at the time of analysis.

### Outlier Detection

Outlier Detection is an algorithm that identifies variants with allele fractions that are statistically significant in a sample compared to a set of samples. If Normal Data Sets are not used, Outlier Detection will first identify positions in samples with unusually high allele fractions and reserve them from noise profiling on the hypothesis that these data are true positives. The algorithm then models the noise on a per-position basis across samples and then compares specific sample variant calls to the noise profile to generate statistics to discern true variants from noise. This feature helps provide greater confidence in variant calling and increased sensitivity at low allele fractions while increasing specificity, providing confidence in variant calling.

When Outlier Detection is enabled, the feature can be used in two different ways: intra-job and Normal Data Sets. Intra-job models the noise at each sequenced position across all samples within a job. Normal Data Sets allows a user to designate normal samples so that noise is modeled only using these samples.

When designating a Normal Data Set of samples, it is critical to ensure that those samples do NOT contain the variations you wish to measure in the assay. If they are present, the ability to accurately estimate the background noise of those variations will be diminished - subsequently lowering the sensitivity of those variations in the unknown samples. It is recommended to use multiple samples from multiple sources, so that variants or SNPs specific to individual samples do not skew the noise modeling.

### Enabling/Disabling Outlier Detection

From the menu bar, select **“User”** > **“Settings”** to access the **“User Settings”** page. Scroll down to the **“Variations/Mutations Settings”** section where the toggle for **“OUTLIER\_DETECTION”** can be found.

**Variations/Mutations Settings**

Min Depth for Variant Call

MAPO Threshold for Variant Call

Min Basequal for Variant Call

Min Allele Fraction for Variant Call

Min Phred Qual Score for Variant Call

Variant Downstream ROI Size

Outlier Detection Off  
✓ On

Min DAO for Outlier Detection Calculation

Deep Shallow Threshold

Outlier Detection is enabled by default but can be turned off if the user wants to disable the feature. If you make a change to the settings, you will want to click **“Update Settings”** located at the bottom of the **“User Settings”** page.

### Using Outlier Detection in the Variant Grid

For variant calling, users can filter variants using Outlier Detection p-value metric. By default, this filter set is the same as the original variant filter set except the software adds the **“AF Outlier P Value”** with a default p-value. By selecting the filter set **“Somatic with Outlier”**, only variants with significant p-values will be reported in the Variant Grid.

Actions	Classification	Report	Artifact	Symbol	HGVSp	HGVSc	Depth	AO	AF	Quality Score	TO	Rept	Tier I	Ti
	<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	AKT1	p.Glu17Lys	c.49G>A	182	58	0.318681	2230.41	232	2	0	0
	<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	ALK	p.Gly1202Arg	c.3604G>A	769	186	0.241873	7645.42	208	2	0	0
	<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	ALK	p.Phe1174Leu	c.3522C>A	836	211	0.252392	8986.89	74	2	0	0
	<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	ATM	p.Ser99Gly	c.295A>G	955	464	0.485864	18774.5	86	1	0	0

Note that Outlier Detection will not be beneficial to run on Germline panels. To save analysis runtime, manually disable the feature in **“User Settings”**.

The **“Confident Negatives 1%”** filter set is useful to determine which genomic positions can be reported as reference. By default, this filter set displays positions with 95MDAF below 1% that also have insignificant AF Outlier p-values.

### Key Outlier Detection Metrics

**Intra-Job AF Outlier P Value (Intra-Job Allele Fraction Outlier P-value)** - The probability this mutation was due to background noise (BN), as estimated across all samples in the same job. A low p-value indicates the AF of the called variant is significantly outside the BN, providing confidence the mutation is real. For each individual variant, outlier detection metrics will be reported if another sample has coverage over the variant. In addition, if the variant is not included in the targeted mutation file, it must also have 3 or more Deep Alternate Observations

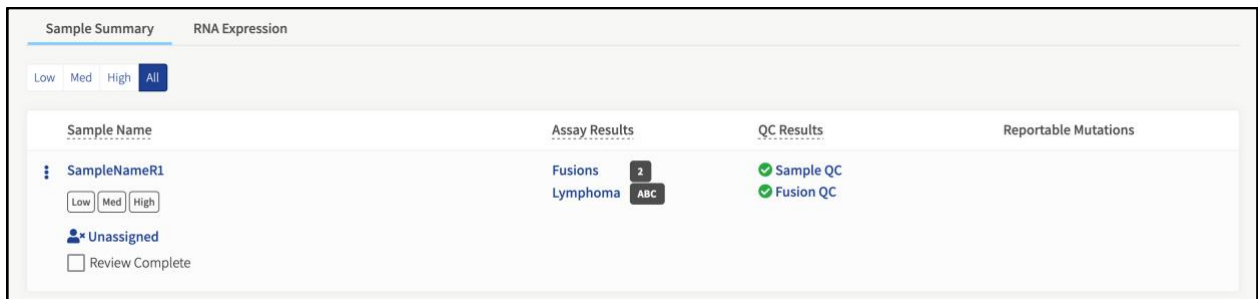
(DAO) for outlier detection metrics to be reported. Variants in the targeted mutation file do not have a DAO requirement for outlier detection metrics to be reported.

**Minimal Detectable Allele Fraction 95% (95MDAF)** - The allele fraction (AF) at which a variant can be called with high statistical power. If the true AF in the sample is at least this, and this identical experiment were run multiple times, 95% of the time there would be sufficient signal to capture this variant. Note, to show a value here, at least one other sample in the job must have the same variant called. This value will be the 95MDAF given the background of the normal data set if one was provided, otherwise it will be the 95MDAF of the intra-job comparison if outlier detection is on.

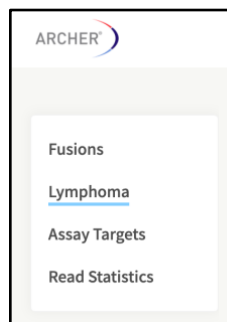
### Lymphoma Cell of Origin (Experimental)

This pipeline is designed to identify the cell of origin subtype from a diffuse large B-cell lymphoma (DLBCL) sample. It is intended to run on known lymphoma samples, not determine if an unknown sample is a lymphoma sample. The classification is based on the work in “Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue” (PMC3931191). The analysis pipeline will occur after the standard primary analysis and is based on calculated gene expression. The pipeline is automatically enabled for panels designed to detect lymphoma cell of origin (COO).

The results of the pipeline displayed in the “**Assay Results**” column of samples is located on the “**Sample Summary**” page and will provide the Lymphoma results: ABC, GCB or Unclassified. Select the link in the “**Assay Results**” column for more detailed information on the results.



By selecting “**Lymphoma**”, this will take you to the lymphoma “**Overview**” page that displays the subtype metric, subtype probability and a lymphoma gene expression heat map. You can also navigate to the Lymphoma overview page by selecting “**Lymphoma**” on the left side of the sample details page.



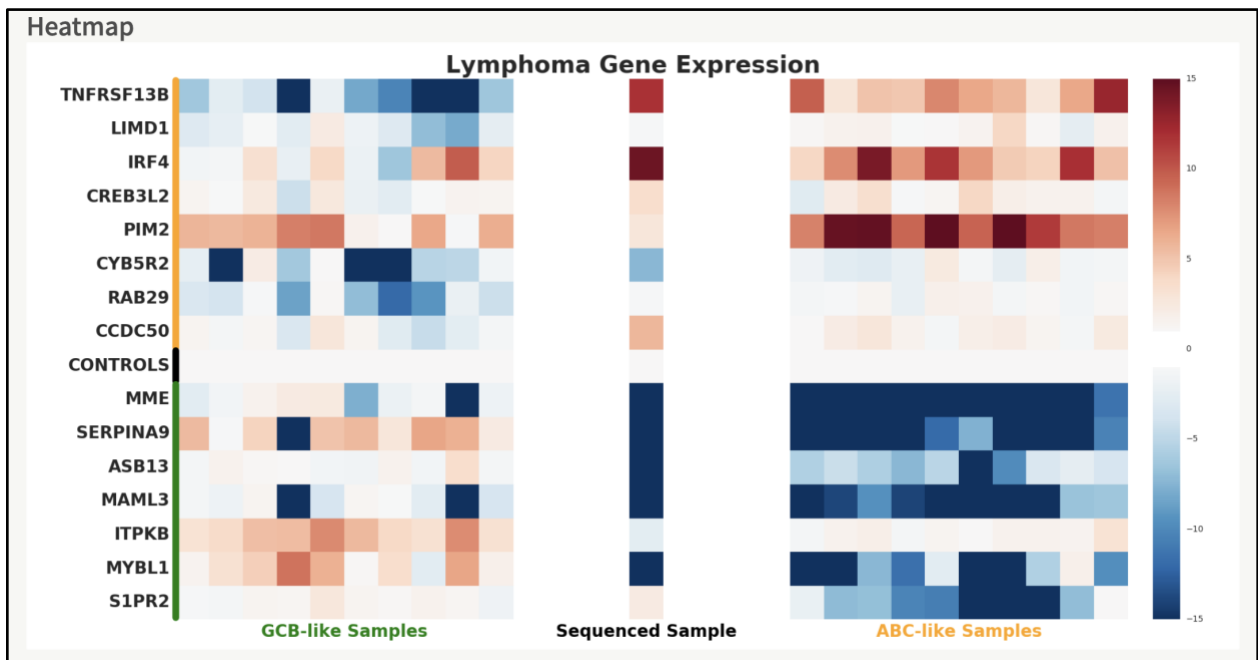
Lymphoma COO Metrics:



- **Subtype Metric** – The reported lymphoma cell-of-origin subtype (GCB, ABC or Unclassified).
- **Subtype Probability** – Estimated probability of subtype classification.
- **Lymphoma Gene Expression Heat map** – This is a heat map of the gene expression considered in the COO algorithm. The results of the sequenced sample are plotted against known ABC and GCB reference samples.

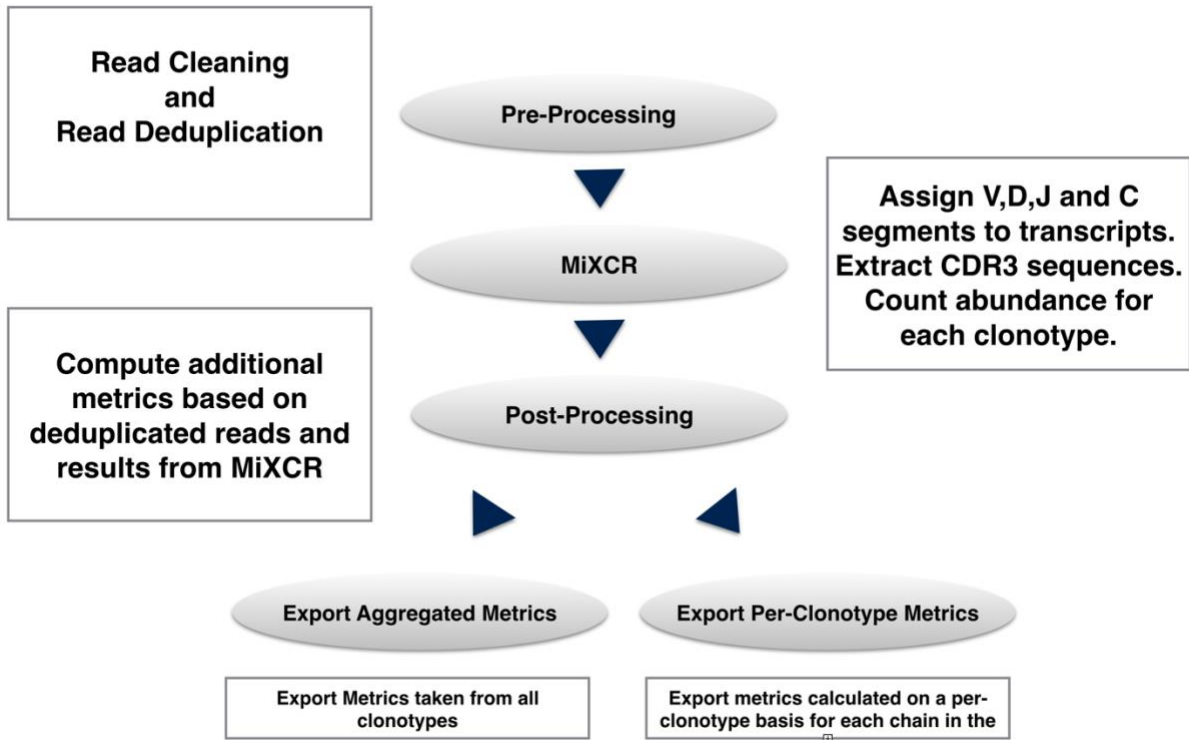
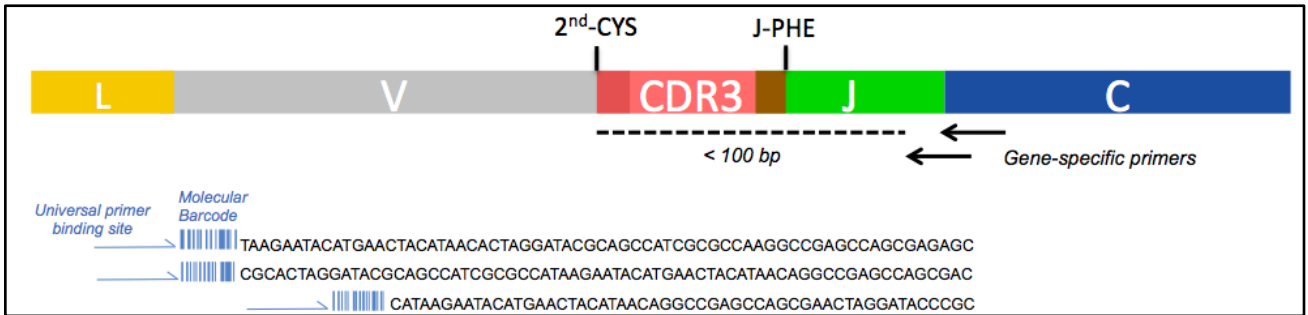
Overview	
Subtype Metric	Subtype Probability
ABC	0.999

[Export Data \(tsv\)](#)



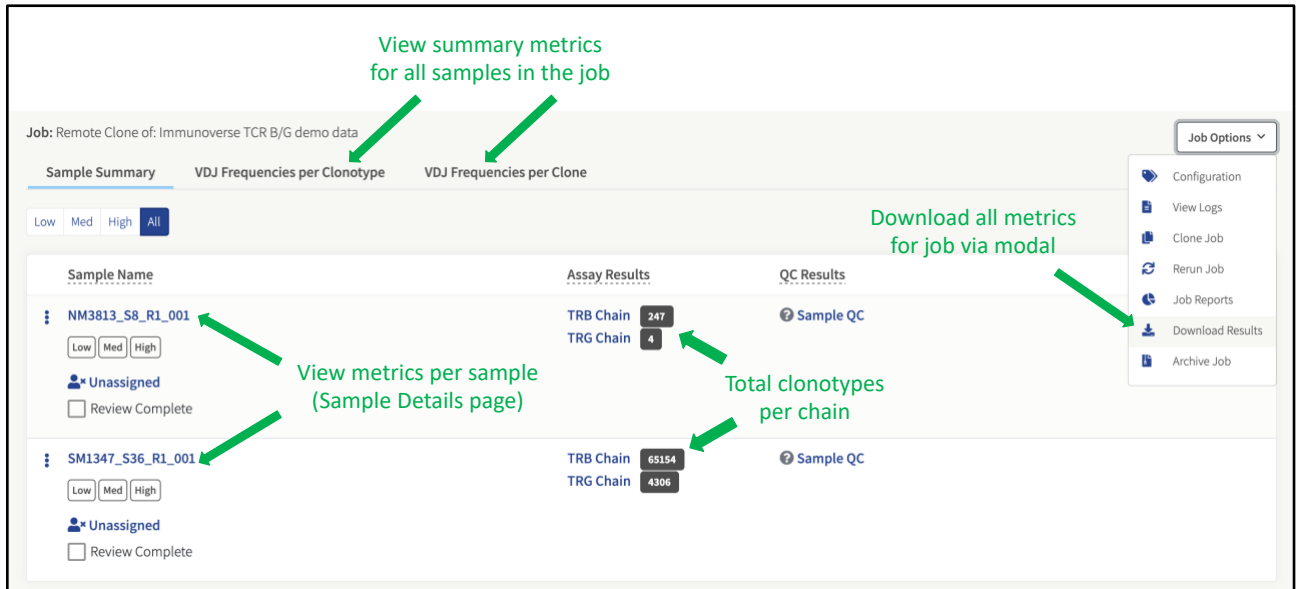
### Immune Repertoire (RNA Only)

The Immune Repertoire pipeline provides an overview of the immune receptor population expressed by IG and TCR clonotypes within a sample. Reads are pre-processed (i.e., adapter trimming, deduplication, error correction), and then profiled using a third-party tool (MiXCR) with some post-processing. Please note, error correction is automatically enabled for the pre-processing of this pipeline. MiXCR processes Archer Immuniverse data to assign V, D, J and C regions to the unique fragments, as well as extracts the primary structure of the CDR3 sequence. It then clusters fragments with similar CDR3 sequences together thus providing a consensus sequence for the CDR3 as well as an abundance count of number unique fragments supporting the CDR3.



### Job-level Results

After the job is processed, results can be viewed and downloaded. The GUI reports many, but not all, of the MiXCR and post-processed metrics, on both per job and per sample basis. Comprehensive results, in tab-delimited format, can be downloaded by selecting “Download Results” from the more-actions menu for each job in the **Jobs** page, or under the Job Options dropdown menu in **Job Details** and **Sample Details** page:



Job: Remote Clone of: Immuniverse TCR B/G demo data

Sample Summary | VDJ Frequencies per Clonotype | VDJ Frequencies per Clone

Low | Med | High | All

Sample Name	Assay Results	QC Results
NM3813_S8_R1_001 Unassigned <input type="checkbox"/> Review Complete	TRB Chain 247 TRG Chain 4	Sample QC
SM1347_S36_R1_001 Unassigned <input type="checkbox"/> Review Complete	TRB Chain 65154 TRG Chain 4306	Sample QC

Job Options

- Configuration
- View Logs
- Clone Job
- Rerun Job
- Job Reports
- Download Results
- Archive Job

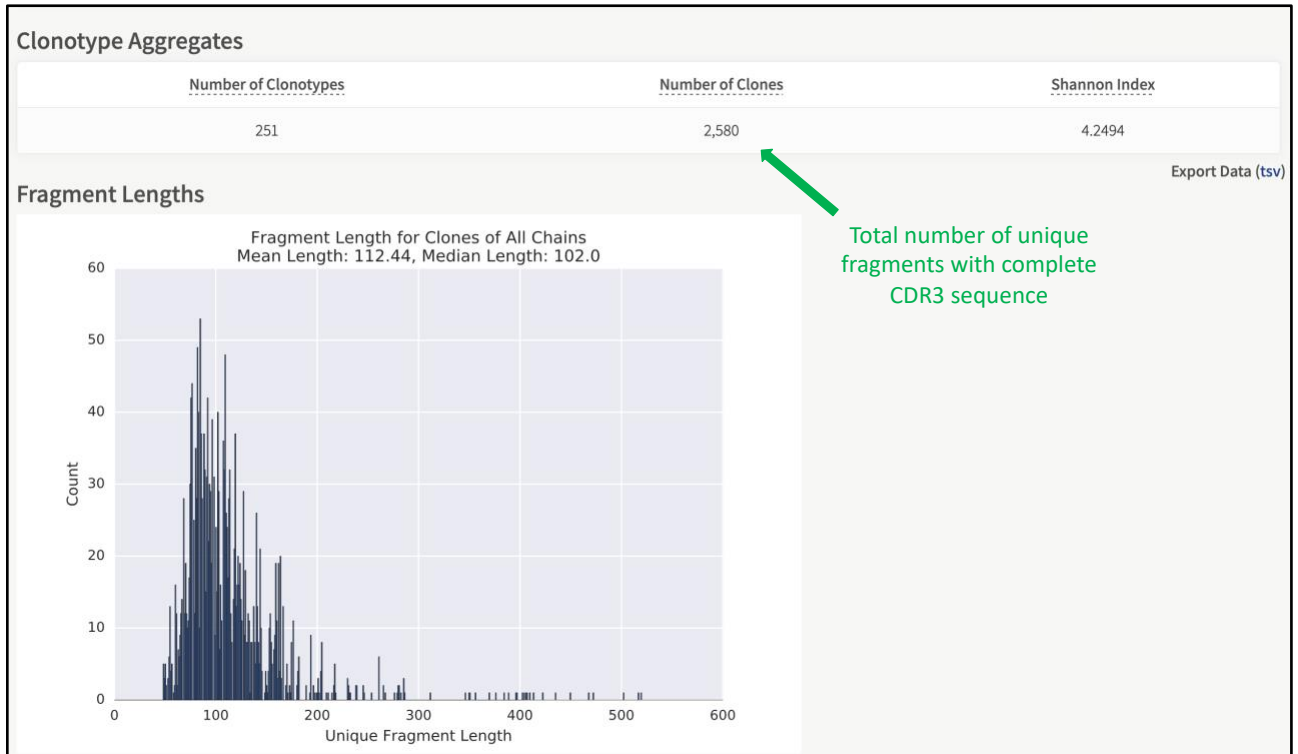
### Sample-level Results

Access the **Sample Details** page by clicking on sample name on the **Job Summary** page. There are two tabs on this page:

- **“Read Statistics”** overviews key metrics before and after processing read data
- **“Clonotypes”** provides an in depth look at clones and clonotypes

#### Sample Details page → Read Statistics tab

The top sections of the “Read Statistics” tab provide metrics similar to other Analysis pipelines. Unique information for this pipeline is provided in two sections at the bottom of this tab (“Clonotype Aggregates” and “Fragment Lengths”):



*Sample Details page* → *Clonotype Summary tab*

Clicking on the metrics under the “**Clonotypes**” Column brings up a new page with new tabs for each chain included in the Archer panel relevant to the current sample. Each of these tabs provides a grid to investigate all clonotypes (CDR3 sequences) found in the sample.

The clonotype grid functions in a manner analogous the variant grid of the **SNP/InDel** pipeline of Archer Analysis; each column represents a metric that can be filtered/sorted against, and visibility toggled on/off. Additionally, data from the grid can be exported as a tab-delimited file, by clicking links at the bottom right-hand corner:

- “**Export current grid**” exports only what is currently displayed in the grid
- “**Export raw source**” downloads all clonotype data (i.e., ignoring column filtering and visibility).

In addition to the grid, plots (histograms and heat maps) of clonotypes, associated clones, as well as V, D, and J segments, can be accessed on this page as well:

TRB TRG ← Metrics are reported on a per chain basis

Data Summary Frequencies per Clonotype Frequencies per Clone Other Plots ← View plots relating to the sample (tabs)

Search...

Default Hypermutation Default New ← Filtering options Edit Columns Save Grid (Grid Changed) Reset Grid

Actions	CDR3 Sequence	Clone Frequency	Clone Abundance	Raw Fragment Abundance	Predicted V Region	Pre
	TGCAGCGTCTGGACAGGGTCTAAGTTCTTC	0.155192532089	399	97575	TRBV29-1	TRE
	TGTGCCAGCAGATGGACTCATGCAAGCACAGATACGCAGTATTTT	0.073512252042	189	25603	TRBV6-5	TRE
	TGTGCCAGCAGCTTCAGGAGGGGTGCCAATGAGCAGTCTTC	0.0688448074679	177	38000	TRBV5-6	TRE
	TGTGCCAGCAGCATACTAGCGGGACAGCAAGAGACCCAGTACTTC	0.0665110851809	171	19255	TRBV2	TRE
	TGTGCCAGCAGCCTCCTTAGCGGGGACCTAGGGAGCAGTCTTC	0.0486192143135	125	23280	TRBV7-8	TRE
	TGTGCCAGCAGAACCGGGACAGGGAACCTCTACGAGCAGTACTTC	0.0264488525865	68	8659	TRBV27	TRE
	TGTGCCAGCAGCACACCGGGACAGTGACCCGGGGGAATCAGCCCCAGCATTTT	0.0186697782964	48	11037	TRBV21-1	TRE
	TGTGCCAGCAGCTTTAGCACCGGGAGCTGTTTTT	0.0108907040062	28	4128	AMBIGUOUS	IND
	TGTGCCAGCAGCTCTAGTCTAGAAGCACTGAAGCTTTCTTT	0.00933488914819	24	3239	TRBV7-9	IND
	TGTGCCAGCAGCCCCAGTGGACTAGCGGGGGCAAGGATACGCAGTATTTT	0.00894593543368	23	2023	TRBV9	TRE

1-25 of 247 Per page: 25 < Previous 1 2 3 4 5 ... 10 Next > Export current grid Export raw source

### Per Sample Plots

Below details the per-sample plots that can be accessed via the “Clonotypes” section:

In the “Clonotypes” section, there are four tabs:

- 1) Data Summary: Contains the grid with metrics available on a per-clonotype basis for the selected chain.
- 2) Frequencies per Clonotype: Contains visual displays of segment usage frequencies of each clonotype:
  - a) Predicted V-Region Frequency: frequency among all V-regions found per-clonotype.
  - b) Predicted D-Region Frequency: frequency among all D-regions found per-clonotype.
  - c) Predicted J-Region Frequency: frequency among all J-regions found per-clonotype.
  - d) Predicted C-Region Frequency: frequency among all C-regions found per-clonotype.
  - e) CDR3 Translation Frequency: Frequency of CDR3 translations amongst all clonotypes of the specified chain
  - f) Productive CDR3 Translation Frequency: Frequency of CDR3 translations amongst all clonotypes of the specified chain of only productive CDR3 translations.
  - g) CDR3 Sequence Frequency: Frequency of CDR3 sequences amongst all clonotypes of the specified chain
- 3) Frequencies per Clone: Contains visual displays of segment usage frequencies of each clonotype the plots that can be selected are:
  - a) Predicted V-Region Frequency: frequency among all V-regions found per-clone.
  - b) Predicted D-Region Frequency: frequency among all D-regions found per-clone.
  - c) Predicted J-Region Frequency: frequency among all J-regions found per-clone.
  - d) Predicted C-Region Frequency: frequency among all C-regions found per-clone.
  - e) CDR3 Translation Frequency: Frequency of CDR3 translations amongst all clones of the specified chain
  - f) Productive CDR3 Translation Frequency: Frequency of CDR3 translations amongst all clones of the specified chain of only productive CDR3 translations.
  - g) CDR3 Sequence Frequency: Frequency of CDR3 sequences amongst all clones of the specified chain

#### 4) Other Plots:

- a) CDR Length: A sequence length histogram of all unique CDR3 sections found in the chain
- b) Raw Fragment: A fragment length histogram of CDR3 positive reads.
- c) Raw V segment: Histogram of the number of base pairs covered in the V segment per read
- d) Unique Fragment: A fragment length histogram of CDR3 positive deduplicated reads.
- e) Unique V segment: Histogram of the number of base pairs covered in the V segment per deduplicated read.
- f) V & J usage: A heatmap detailing co-occurrences of V and J usage per clonotype.
- g) V usage: A histogram of V segment usage of the 100 most abundant clonotypes
- h) D usage: A histogram of D segment usage of the 100 most abundant clonotypes
- i) J usage: A histogram of J segment usage of the 100 most abundant clonotypes
- j) C usage: A histogram of C segment usage of the 100 most abundant clonotypes

#### *Post-processing after Archer Analysis*

All data presented in the Archer Analysis GUI, as well as all metrics generated by MiXCR, can be exported for further analysis (e.g., via R, Excel, etc.). This data can be accessed by clicking the “Job Options” > “Download Results” link on the **Job Summary** page. This will initiate download of a zip file to your local machine, named after the job number. After extracting the archive, data for each sample can be found in its own subfolder, named as (with “\*” as a placeholder for the sample name):

**\*.molbar.trimmed.deduped.Immune\_Repertoire\_Results**

Each sample subfolder contains all plots (as PDFs) and raw text output (as tab-delimited files).

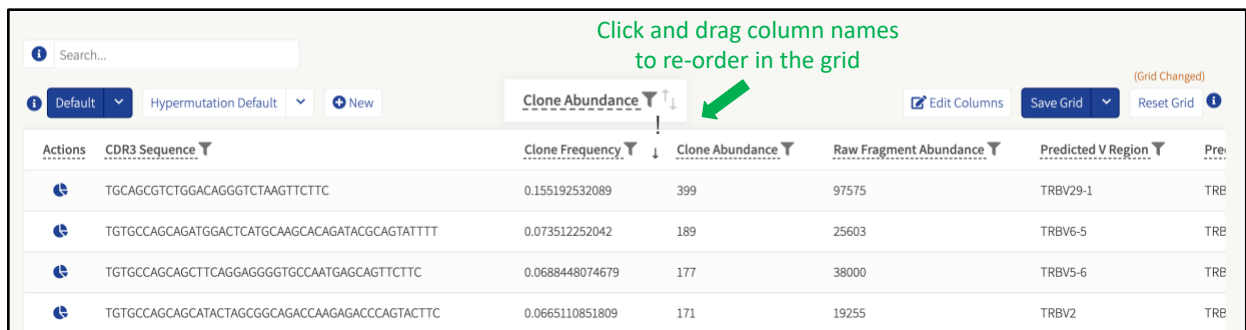
#### *Filtering Results*

Each of the columns in the Variant, Fusion, Structural Variant, CNV, and Clonotypes grids can be filtered by selecting an operator and entering a custom value from the drop-down menu. 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.

The information currently displayed in the Variant Summary can be downloaded by selecting the “**Export current grid**”, or the unfiltered data is available by selecting “**Export raw source**”, below the variant summary page.

#### *Editing and Saving Grid Columns*

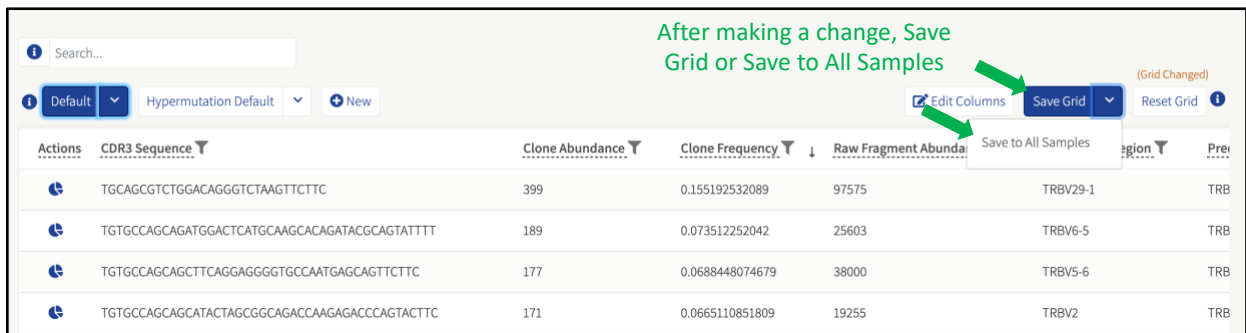
Users can configure grids by showing/hiding and rearranging the column headers as they see fit. Once the grid columns are rearranged, these settings can be saved across a single sample (by clicking on Save Grid) or multiple samples within a job (by clicking on the dropdown and selecting Save to All Samples).



Click and drag column names to re-order in the grid

Actions	CDR3 Sequence	Clone Frequency	Clone Abundance	Raw Fragment Abundance	Predicted V Region	Pre
	TGCAGCGTCTGGACAGGGTCTAAGTTCTTC	0.155192532089	399	97575	TRBV29-1	TRB
	TGTGCCAGCAGATGGACTCATGCAAGCACAGATACGCAATATTT	0.073512252042	189	25603	TRBV6-5	TRB
	TGTGCCAGCAGCTTCAGGAGGGGTGCCAATGAGCAGTTCTTC	0.0688448074679	177	38000	TRBV5-6	TRB
	TGTGCCAGCAGCATACTAGCCGGCAGACCAAGAGACCCAGTACTTC	0.0665110851809	171	19255	TRBV2	TRB

After making a change, Save Grid or Save to All Samples



Actions	CDR3 Sequence	Clone Abundance	Clone Frequency	Raw Fragment Abundance	Save to All Samples	Region	Pre
	TGCAGCGTCTGGACAGGGTCTAAGTTCTTC	399	0.155192532089	97575	TRBV29-1	TRB	
	TGTGCCAGCAGATGGACTCATGCAAGCACAGATACGCAGTATTTT	189	0.073512252042	25603	TRBV6-5	TRB	
	TGTGCCAGCAGCTTCAAGGGGGTGCCAATGAGCAGTCTTC	177	0.0688448074679	38000	TRBV5-6	TRB	
	TGTGCCAGCAGCATACTAGCGGCAGACCAAGAGACCCAGTACTTC	171	0.0665110851809	19255	TRBV2	TRB	

## Normal Data Sets (SNP/InDel and CNV)

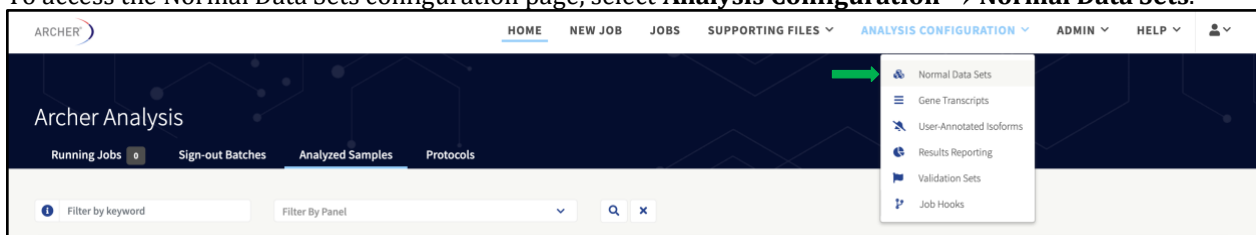
This feature supports the designation of a Normal Data Set of samples. These samples will be used to calculate background noise that is due to sequencing and enzymatic error.

When designating a Normal Data Set of samples, it is critical to ensure those samples do NOT contain the variations to measure in the assay. If they are present, the ability to accurately estimate the background noise of those variations will be diminished — subsequently lowering the sensitivity of those variations in unknown samples.

When it is not possible to get perfectly normal tissue, it is desirable to have multiple samples from multiple sources, so variations or SNPs specific to individual samples can be averaged out.

For SNP/InDel Normal Data Sets, it is important to ensure the total depth of coverage among all samples is sufficient. For example, if a genomic coordinate only has 50x total coverage calculated among all samples in the Normal Data Set, Outlier Detection will not be sufficiently powered for variants with an allele fraction of one in a hundred.

To access the Normal Data Sets configuration page, select **Analysis Configuration** → **Normal Data Sets**.



## Normal Data Sets Terminology

**Committed Normal Data Set** - A committed Normal Data Set is one which has been used in a job on the instance. Once a Normal Data Set is committed, it can no longer be removed or altered by adding or removing samples. If the user wishes to add/subtract samples from a committed Normal Data Set, they will first need to clone the original Normal Data Set. Cloning creates an uncommitted copy of the original dataset, to which samples can be added or removed.

**Deprecated Normal Data Set** - A Normal Data Set that was functional on a previous version of the software but needs to update to work on the current version due to a system upgrade. No new jobs can be submitted using this

Normal Data Set and no existing jobs that used this normal data set can be resubmitted or cloned until the normal data set is updated. Any protocols that use this normal data set will also be temporarily disabled.

The Normal Data Set can be upgraded by (re-)running the samples in the Normal Data Set on the new version of the Archer Analysis software. See [Deprecated and Upgrading Normal Data Sets](#) section.

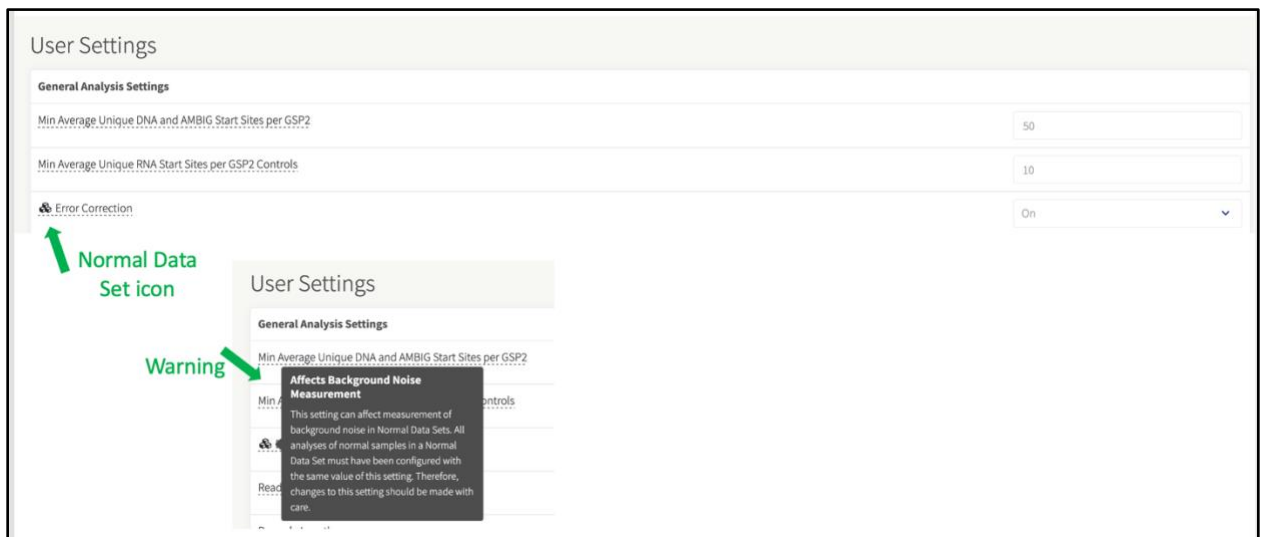
## Creating Normal Data Sets

To establish a Normal Data Set the following practices should be considered:

1. Normal inputs should be of similar quality to the typical input for the panel,
2. Similar amounts of input should be used for normal inputs as for sample(s) of interest,
3. Inputs from a variety of samples should be used,
4. Normal Data Sets should be specific to instrument type and library preparation chemistry, and
5. Libraries can be sequenced on sequencing runs with a range of acceptable cluster densities, to account for variability of noise profiles across cluster densities.

It can be difficult to acquire tumor/normal paired samples, so it is acceptable to utilize other normal control samples such as Genome in a Bottle (GIAB) when creating a Normal Data Set. This way users can utilize the same Normal Data Set in subsequent analyses.

Settings that must be the same are defined in the next section. To help users remember what these settings are, they are marked within the User Settings page with a Normal Data Set icon.



## User Settings for Normal Data Sets

To add samples to a Normal Data Set, these User Settings must be the same across the samples:

- "Error Correction"
- "Skip Deduplication"
- "Barcode Hamming Distance"

For a SNP/InDel Variant Normal Data Set in particular, these additional User Settings need to be the same as well:

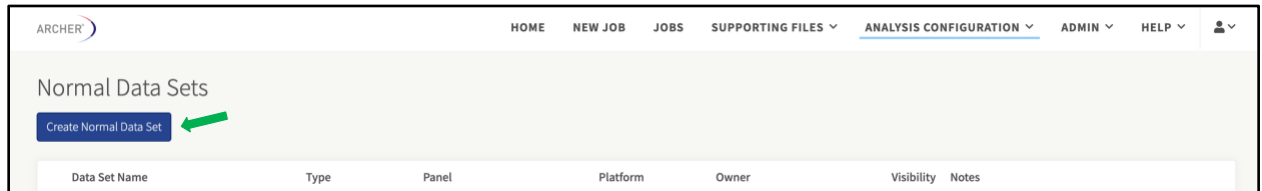
- "MAPQ Threshold for Variant Call"
- "Min Basequal for Variant Call"
- "Variant Downstream ROI Size"



- "Deep Shallow Threshold"

## Creating a Normal Data Set

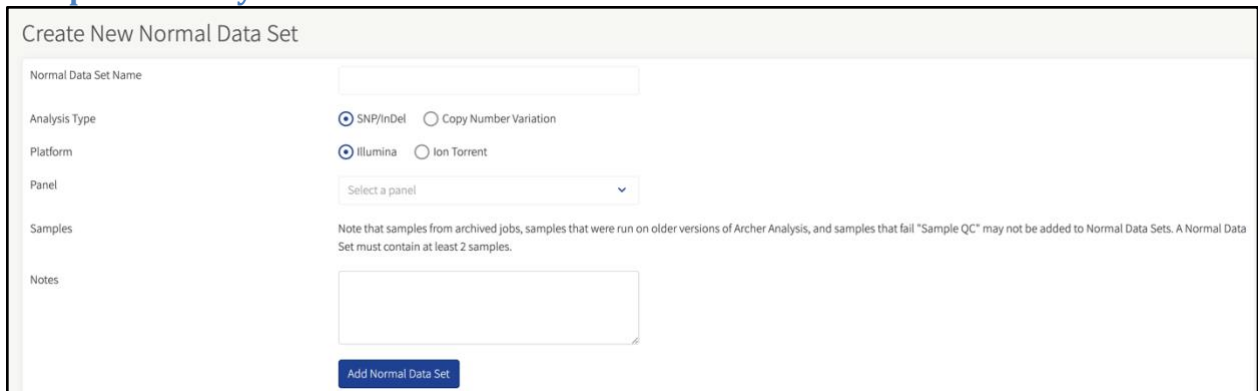
To create a Normal Data Set, select **Analysis Configuration** → **Normal Data Sets** → **Create Normal Data Set**.



When creating a Normal Data Set, users must provide:

- Normal Data Set Name
- Mutation Type (SNP/InDel or Copy Number Variation)
  - If samples were initially run with both SNP/InDel and DNA CNV analyses, and the user wants to create a Normal Data Set for both analysis types, the user must create two separate Normal Data Sets.
- Platform
- Panel
- Samples
  - Samples matching the chosen Analysis Type, Platform, and Panel can be selected.
  - Samples are grouped by user settings. See below for sample visibility information.
  - After some in house studies, our Bioinformatics R&D team determined an acceptable range of samples to add to a Normal Data Set for the SNP/indel pipeline is 3-7. This is acceptable for the CNV pipeline, however, increasing the sample numbers included will improve the quality of the data.
- Notes (optional)

## Sample Visibility



The screenshot shows the 'Create New Normal Data Set' form. It includes the following fields and options:

- Normal Data Set Name:** A text input field.
- Analysis Type:** Radio buttons for  SNP/InDel and  Copy Number Variation.
- Platform:** Radio buttons for  Illumina and  Ion Torrent.
- Panel:** A dropdown menu with the text 'Select a panel'.
- Samples:** A text area with a note: 'Note that samples from archived jobs, samples that were run on older versions of Archer Analysis, and samples that fail "Sample QC" may not be added to Normal Data Sets. A Normal Data Set must contain at least 2 samples.'
- Notes:** A text area.
- Buttons:** An 'Add Normal Data Set' button at the bottom.

Samples will be shown for selection once the panel is selected. To be able to select a particular sample for a Normal Data Set, the following requirements must be met:

- The analysis type, platform, and panel must match that of the Normal Data Set containing the sample.
- The sample must have processed without errors.
- Overall sample QC must have passed for the sample.
- All selected samples must be run on the same panel and version of the software.

- Any job containing a sample(s) included in a Normal Data Set should not be archived.

Samples will be grouped together based on the values of the concordant settings listed above in the “User Settings for Normal Data Sets” section. If more sample groups are displayed, hovering over the setting icon will show the values and how they differ from the default. Samples can only be selected from within a single group. The following section explains an alternative way to add samples to a Normal Data Set which is informative when determining why a particular sample cannot be added to a Normal Data Set.

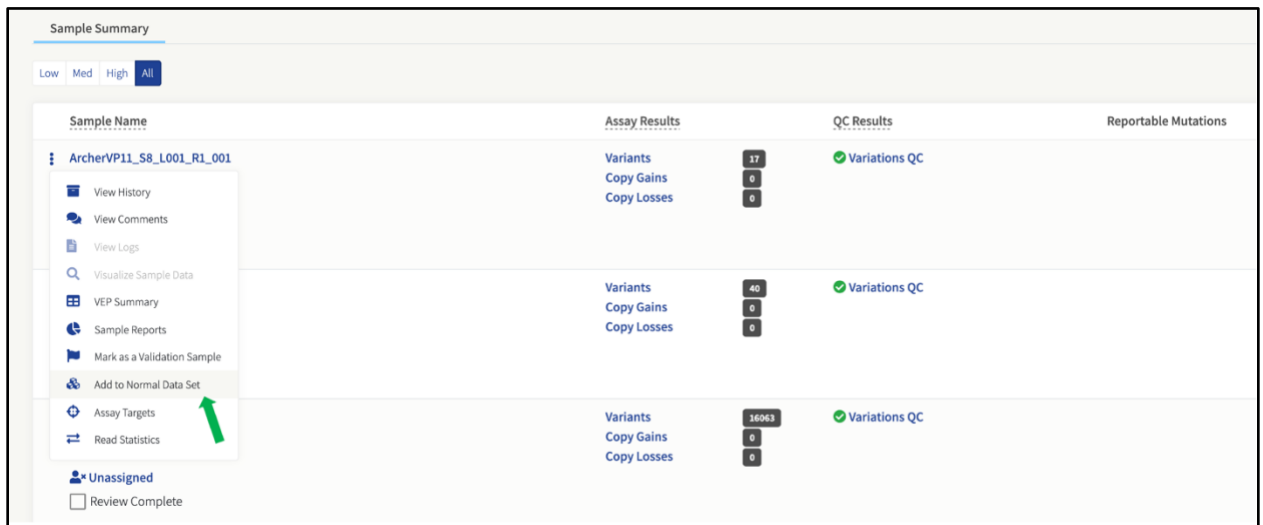
### Adding a Sample to an Normal Data Set from Job Details

A user can add a sample to a Normal Data Set from the job detail page. This can be useful for the following reasons:

- This method will display the exact reason why a sample cannot be added to a Normal Data Set.
- If you know to which job the sample belongs, it will be faster to find the sample.

This method is accessed via the job details page where each sample is listed with its overall results.

1. Click the **ellipsis drop-down** icon.
2. Click **Add to Normal Data Set**.



Sample Name	Assay Results	QC Results	Reportable Mutations
ArcherVP11_S8_L001_R1_001	Variants: 17 Copy Gains: 0 Copy Losses: 0	Variations QC: <span style="color: green;">✔</span>	
	Variants: 40 Copy Gains: 0 Copy Losses: 0	Variations QC: <span style="color: green;">✔</span>	
	Variants: 18063 Copy Gains: 0 Copy Losses: 0	Variations QC: <span style="color: green;">✔</span>	

Modal View:

**Add Sample to Normal Data Set**
✕

---

**Sample:**

**Platform:**

**Panel:**

**Normal Data Set Type:**

**Normal Data Set:**

ArcherVP11\_S8\_L001\_R1\_001

Illumina

VariantPlex Solid Tumor AK0051 v1.2

CNV
  VARIANT

Existing
  New

\* Only normal data sets that match the platform, panel, type and normal data set required settings of this job are displayed.  
 † Normal data set is not editable. Selecting this will clone the normal data set prior to adding the sample.

Include all samples from this job

Cancel
Add

If this sample cannot be added to a Normal Data Set, the exact reason will be displayed in place of the existing Normal Data Set drop down menu. Also listed is the platform and panel for the job; only Normal Data Sets defined with the matching platform and panel will be shown in the existing Normal Data Set drop down menu. Note that only Normal Data Sets whose settings match those of this job’s will be visible.

The first option is to add the sample to an existing Normal Data Set. This option has a drop-down menu allowing the user to select an applicable Normal Data Set. There might be an icon preceding the Normal Data Set name; this is explained in the legend below:

- No icon. The Normal Data Set is editable. Selecting one of these Normal Data Sets will add the sample to the Normal Data Set and redirect the user back to the job details page giving a *successful* message.
- † Normal Data Set is not editable because it has already been committed. Selecting one of these will clone the Normal Data Set prior to adding the sample and the user will be directed to the “**Create Normal Data Set**” page providing them the option to change the name, samples, and notes.
- ‡ Sample is already part of said Normal Data Set and is not selectable.

The second option in this mode is to add the sample to a new Normal Data Set. This will direct the user to the Create Normal Data Set page with the sample selected. The user will need to fill out the remaining required fields.

The final option is the **Include all Job Samples** checkbox. This will only be available if all the samples in the job can be added to Normal Data Sets. Upon selecting this option, the workflows are the same as above except all samples in the job are added.





### Existing Normal Data Sets

This is the initial page that loads when accessing Normal Data Sets from the configuration drop down link.

Create Normal Data Set						
Data Set Name	Type	Panel	Platform	Owner	Visibility	Notes
Upgrade Normal Data Set	SNP/InDel	cGSP5043 v0.1.1	Illumina			
Change Visibility	SNP/InDel	STFocus v1.0	Illumina			
Change Owner	SNP/InDel	cGSP5045 v0.3	Illumina			
Delete						

This page shows all Normal Data Sets a user is allowed to view. It demonstrates how Normal Data Sets are configured and provides a list of possible actions.

If the viewer owns the Normal Data Set, or has admin privileges they will also be able to:

-  Change the owner.
-  Change the visibility.
-  Delete the Normal Data Set, if no jobs were run with it.
-  Upgrade the Normal Data Set if it is deprecated.

It is important to understand what changing the visibility of a Normal Data Set entails. If a user tries to change it to any user that is not Owner, they are giving others access to the sample ID and names contained within that Normal Data Set.

The users who have access to the Normal Data Set will be able to run jobs against it and clone the Normal Data Set if they want to make amendments. When providing others access to the Normal Data Set, a user must also remember to check that the panel which is associated with the Normal Data Set is visible for those users. Otherwise, they will not be able to use it when running jobs.

From this page, users can also get more information on the Normal Data Set by clicking on the Normal Data Set name. The details page will include:

- Samples included in the data set.
- Required settings and their values.
- Jobs that were run against the Normal Data Set.
- Protocols using the Normal Data Set.

### Submitting a New Job with Normal Data Sets

When setting up a new job, once the Panel, Platform, Input Material, and Analysis Type (SNP/InDel, DNA CNV or both) are selected, the Normal Data Set drop down will be populated.

Only Normal Data Sets the user has permission to see and match the analysis type, platform, and panel, will be present in the drop-down menu under Variant or CNV Normal Data Set.

### 2. Pipeline

SELECT SEQUENCING PLATFORM

Illumina  Ion Torrent

SELECT INPUT MATERIAL

RNA  DNA  cfDNA

SELECT ANALYSIS TYPE

**Option Set A**

Copy Number Variation  SNP/InDel  Structural Variation  Target Coverage  MSI (coming soon)

**Option Set B**

Read Processing Only

TARGETED MUTATIONS

Select ▼

Optional

SELECT DNA SNP DETECTION

Germline Variation  Somatic Mutation

VARIANT NORMAL DATA SET

Select ▼

Optional

Normal Data Sets which are not deprecated are selectable. Ones with asterisks preceding their names are deprecated and must be upgraded prior to use.

When a user submits a job using a Normal Data Set, there is a chance the settings will not match those required by the Normal Data Set. In that case, a warning will appear showing which user settings will be overridden by the Normal Data Set's settings. The user must acknowledge their settings will be overridden prior to job submission in these cases.

## Protocols and Normal Data Sets

The process of creating a protocol with an associated Normal Data Set is virtually the same as described above except for how the settings are handled. When running an analysis, the system overrides the user settings to match those of the Normal Data Set. For protocols there is an additional constraint; protocol folder settings that are required by the Normal Data Set are not editable after the protocol is created.

## Deprecated and Upgrading Normal Data Sets

If a Normal Data Set is deprecated, that means that its samples were processed using an older version of the software that measured background noise differently. If a Normal Data Set is deprecated it will be noted as such when looking at existing Normal Data Sets. An additional action of upgrade (⬆) will also be available if the user is the owner or has admin privileges.

Create Normal Data Set

Data Set Name	Type	Panel	Platform	Owner	Visibility	Notes
10 ⚠️ <b>Deprecated</b>	SNP/InDel	cGSP5045 v0.3	Illumina			

If a Normal Data Set is deprecated:

- Any job that uses the Normal Data Set cannot be resubmitted nor cloned.
- No new jobs can use the Normal Data Set.
- Any existing protocols that use the Normal Data Set are disabled.
- No new protocols can be created that use the Normal Data Set.
- The Normal Data Set cannot be cloned.

When a user upgrades a Normal Data Set by clicking the upgrade action icon, all samples that are part of the Normal Data Set will be cloned and reprocessed. These job names will be prefixed with "Normal Data Set upgrade rerun". Once all jobs finish successfully, the Normal Data Set will be upgraded and useable on the system again.

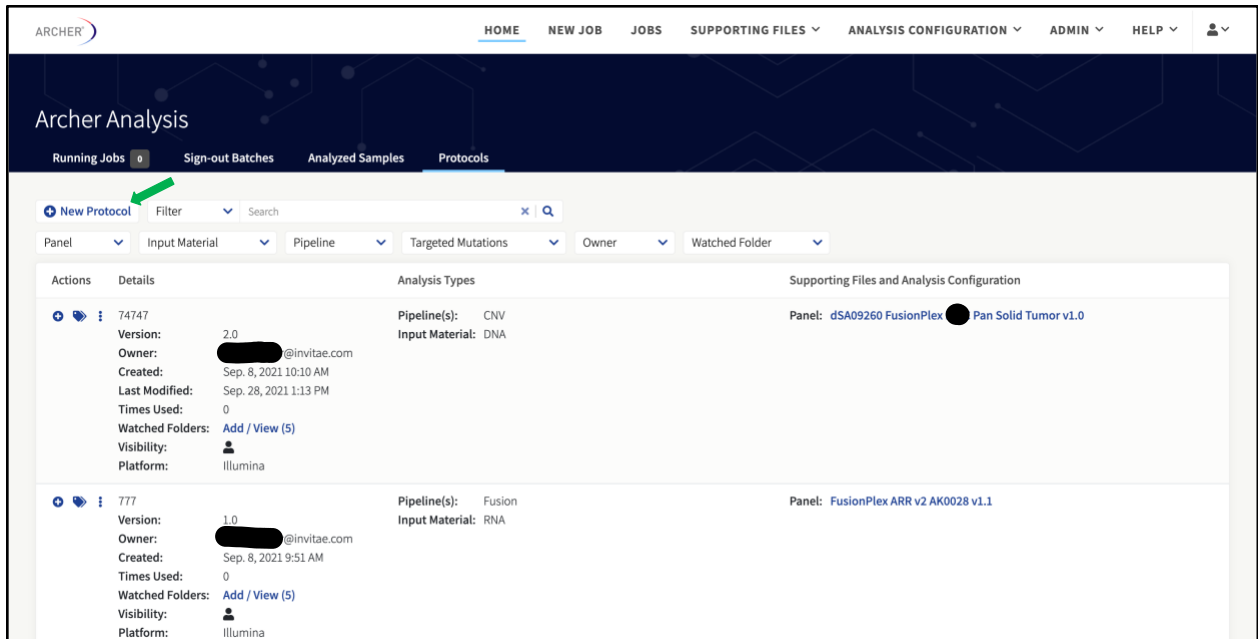
## Workflow Automation and Integrations

### Protocols and Watched Folders

The existing Watched Folders feature has been merged into the new Protocols feature. This allows users to completely configure a new analysis pipeline, save it to a library of protocols, select specific protocols to quickly submit jobs for new samples, and to specify watched folders for any protocol to monitor for and automatically process new samples.

## Create a New Protocol

To create a new analysis protocol, navigate to **Home** → **Protocols**, and click **New Protocol**.



This will open the Create New Protocol dialogue (similar to New Job), as detailed below:

1. Set Up
  - a. Enter a **name for the protocol**
  - b. Enter a **version number**
  - c. Select a **panel** from the existing list
2. Pipeline
  - a. Select the **sequencing platform** on which the samples were analyzed
  - b. Select the **type of input material** (RNA, DNA, or cfDNA)
  - c. Select **analysis pipeline(s)** specific to the input material
  - d. Select **optional settings** as needed
3. Settings
  - a. Optionally override general and pipeline-specific settings
4. Hooks
  - a. Optionally enable sample pre/post hooks and job "post" -> "pre/post" hooks

## Create New Protocol

### 1. Set Up

ENTER PROTOCOL NAME

New Protocol

PROTOCOL VERSION

1.0

SELECT PANEL

VariantPlex |



### 2. Pipeline

SELECT SEQUENCING PLATFORM

Illumina  Ion Torrent

SELECT INPUT MATERIAL

RNA  DNA  cfDNA

SELECT ANALYSIS TYPE

Option Set A

Copy Number Variation  SNP/InDel  Structural Variation  Target Coverage  MSI (coming soon)

Option Set B

Read Processing Only

TARGETED MUTATIONS

Select



Optional

SELECT DNA SNP DETECTION

Germline Variation  Somatic Mutation

VARIANT NORMAL DATA SET

Select



Optional



### 3. Settings

Use My Settings     Use System Default Settings    [Clear Changes](#)

General Analysis Settings	
Min Average Unique DNA and AMBIG Start Sites per GSP2	System Default: 50 <input type="text" value="40"/>
Min Average Unique RNA Start Sites per GSP2 Controls	<input type="text" value="10"/>
Error Correction	Off <input checked="" type="checkbox"/> On
Read Depth Normalization	<input type="text" value="3500000"/>
Barcode Length	<input type="text" value="8"/> ▾
Common Region Set	Dynamic sear ▾
Number of Threads	<input type="text" value="1"/>
Pre-/Post-Alignment Read Cleaner	Off <input type="checkbox"/> On
Merge Reads in Read Cleaner	Off <input type="checkbox"/> On
Debug	Off <input type="checkbox"/> On

### 4. Hooks - OPTIONAL ⊖

⚠ Enabled hooks were set by an administrator on this system. Changing which ones are enabled is for expert users only.

<p><b>SAMPLE PRE HOOKS</b></p> <p><input type="checkbox"/> Lower Case to Capital in FASTQ</p> <p><b>JOB POST HOOKS</b></p> <p><input type="checkbox"/> JobDoneHook</p> <p><input type="checkbox"/> python_posthook_summarize_full_results</p>	<p><b>SAMPLE POST HOOKS</b></p> <p><input type="checkbox"/> SampleDoneHook_test</p>
---	---

Click **Save** to create a new protocol.

## Start a New Job Using an Existing Protocol

To start a new job using an existing protocol, navigate to **Protocols** → **New job** icon under the **Actions** column for a saved protocol. Enter **job name**, upload sequenced libraries (FASTQ/BAM format), and click **Submit Job** to start the job.

### New Job from Protocol

#### 1. Set Up

PROTOCOL NAME  
74747

ENTER JOB NAME

#### 2. Batch Sign-Out - OPTIONAL +

#### 3. Sequenced Libraries

Upload the sequence files for the libraries to be analyzed.  
Because this job is configured for the Illumina platform, only .fastq and .fastq.gz files may be uploaded.

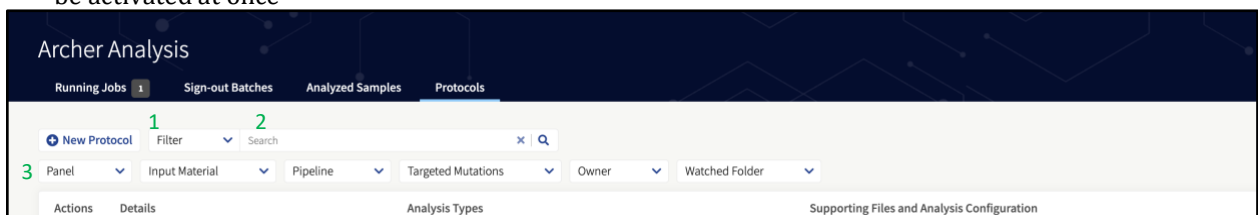
[📁 Drag and drop your FASTQ files here or click to browse.](#)

Submit Job

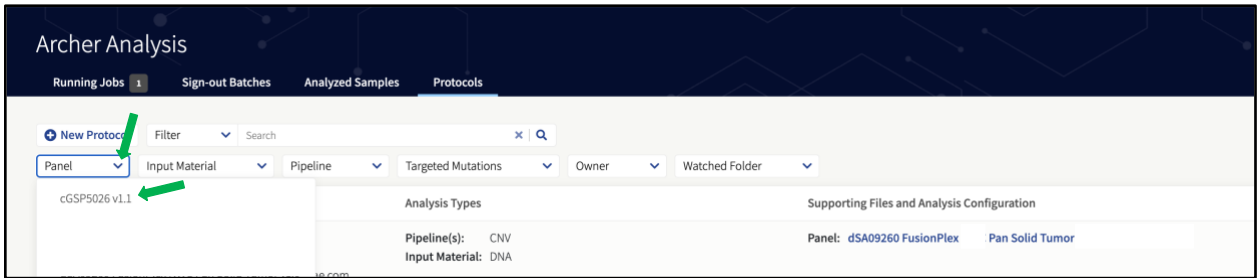
## Search and Filter Existing Protocols

Archer Protocols can be searched and filtered as follows:

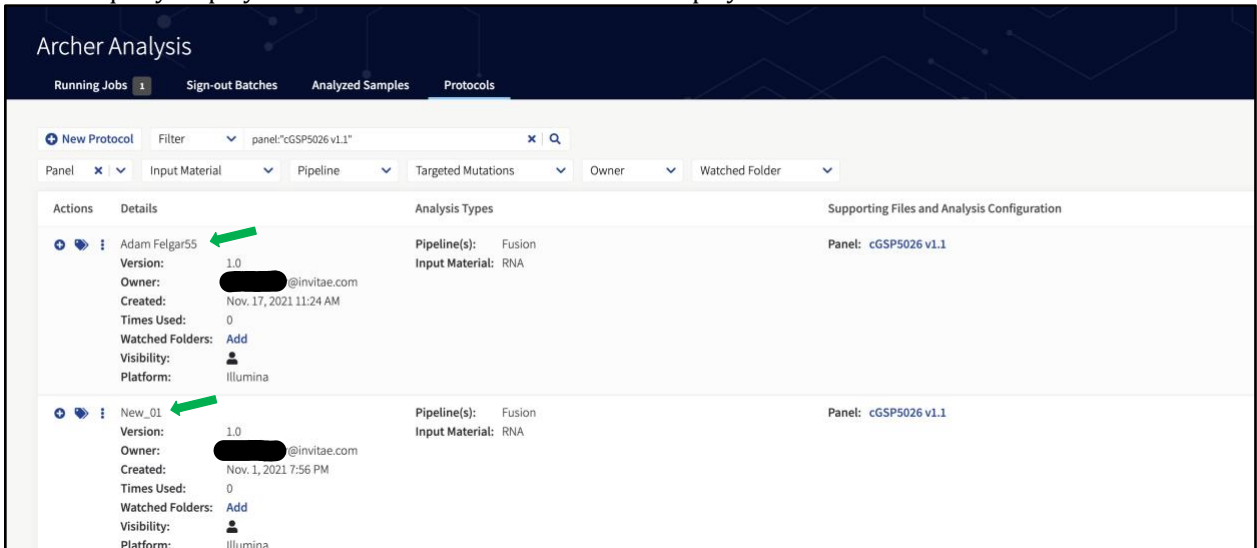
1. Category Filter – This dropdown list holds commonly used filters to help streamline a search
2. Main Search Field - Type queries here
3. Category Filters - Depending on data, one, two, or ten category filters could be available for use. Category filters display horizontally and are clickable dropdown menus. Multiple category filters may be activated at once



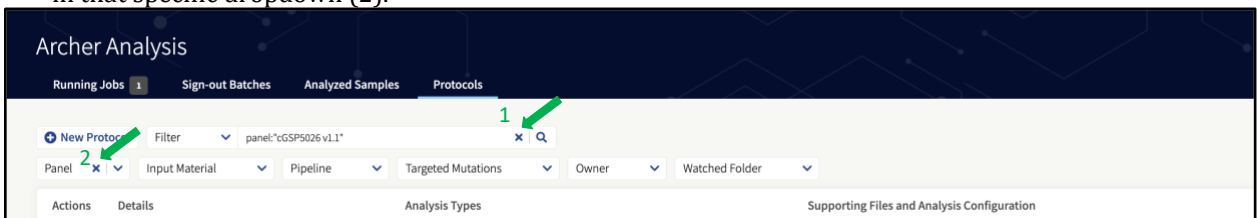
4. Click a **Category Filter** and select a **panel** to see the protocols that use that panel.



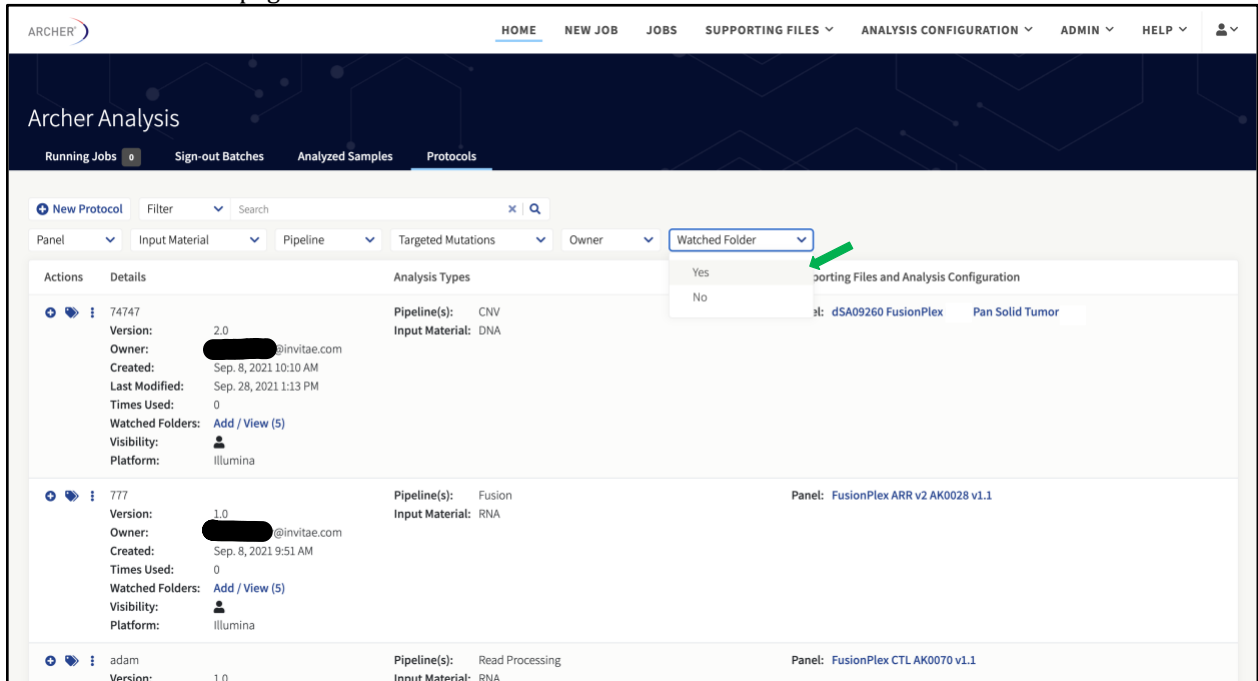
5. The query displays in the Search field and the results display in the table.



6. To clear the entire query, click the **X** next to the Search field (1). To clear only a Category Filter, click the **X** in that specific dropdown (2).



7. To quickly find a saved Protocol, use the predefined filters, search field, or specific dropdown filters at the top of the Protocols page.



The screenshot shows the Archer Analysis interface. At the top, there are navigation tabs: HOME, NEW JOB, JOBS, SUPPORTING FILES, ANALYSIS CONFIGURATION, ADMIN, and HELP. Below this is the 'Archer Analysis' header with sub-tabs: Running Jobs, Sign-out Batches, Analyzed Samples, and Protocols. The Protocols tab is active, showing a list of protocols. The first protocol (ID 74747) is selected, and its details are shown in a table. A dropdown menu is open under the 'Watched Folder' column, showing options 'Yes' and 'No'. A green arrow points to the 'Watched Folder' dropdown.

## Configure Existing Protocols

Additional options to configure existing protocols can be found under the Actions column in the Protocols tab. These include:

- View details – See the details for a Protocol and its past versions, including pipeline, input material, target panel, and settings
- Edit Protocol – Edit an existing Protocol and save as a new version
- Change Protocol Name – Change the name of an existing Protocol
- Clone Protocol – Create a copy of an existing Protocol
- Change Visibility – Configure which users can see an existing Protocol
- Change Owner – Pass ownership of a Protocol to another user based on email address
- Delete – Delete an existing Protocol

## Set up Watched Folders

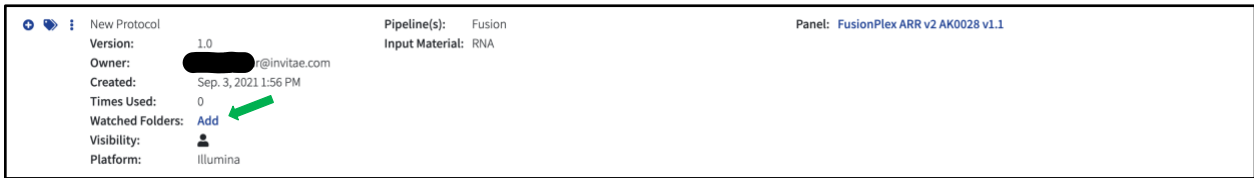
The Watched Folders mechanism allows the automatic execution of a predefined workflow whenever a set of FASTQ or BAM files are moved to a specified watched directory. There are two steps for the setup of the workflow automation:

1. Set up a new watched folder for an existing Protocol
2. Develop a script or procedure to move FASTQ or BAM files to the Watched Folder

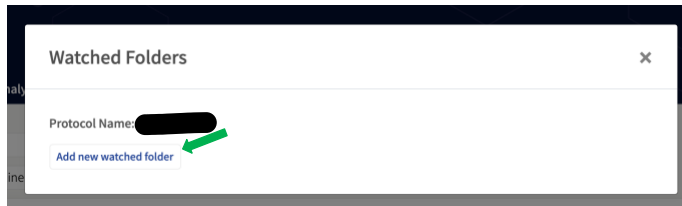
## Add a New Watched Folder

To allow Archer Analysis to automatically execute a Protocol when FASTQ or BAM files are placed in a watched folder, use the web interface to create a folder accessible to the sever.

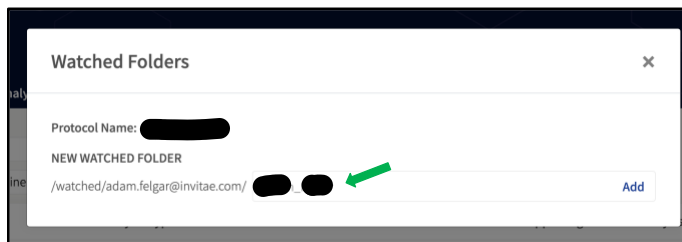
1. Navigate to **Protocols** → **Add** or **Add / View** for Watched Folders under the Details column.



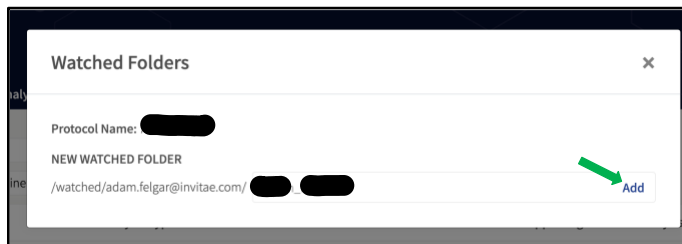
2. Click **Add new watched folder**.



3. In the **NEW WATCHED FOLDER** field, enter a desired name of the watched directory. Do not use spaces or symbols other than hyphen or underscore.




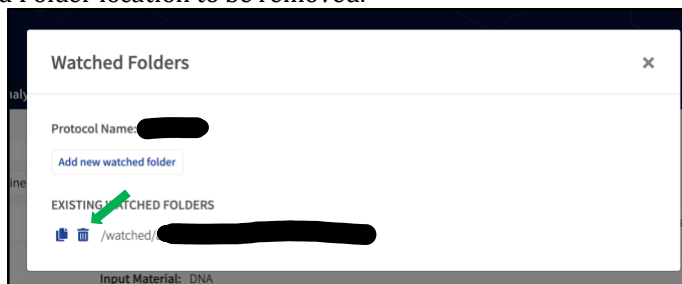
4. Click **Add** to create a new Watched Folder location.



Once complete, the new Watched Folder will appear alongside its protocol in the Protocols list.

## Remove a Watched Folder

To remove a Watched Folder, click **Add / View** for an existing Protocol, and then click the  icon next to the Watched Folder location to be removed.



Removing a watched 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 watched directory will NOT delete the folder on the server but will only remove it from monitoring.

## Remove Assets Used in a Protocol

When a Supporting File or Normal Data Set is selected for one or more Protocol, it cannot be removed until all Protocols referring to it are removed. This ensures that Protocols relying on the Supporting File or Normal Data Set can be executed correctly.

## Develop Procedure/Script for the Movement of FASTQ or BAM Files

Once the Watched 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 or BAM files to the watched folder.

For information on setting up procedures or scripts to move FASTQ or BAM files for the Archer Analysis on-premises VM, please contact [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com).

## How Archer Analysis Knows When to Execute a Workflow

Archer Analysis continuously monitors each of the Watched Folder locations and looks for either of the following:

- 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 watched 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 samples) 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 Technologies™ Ion Torrent™ system or the Illumina™ single read libraries, create a file with the structure “[SAMPLE\_NAME].completed”.

## Example for a Job Containing a Single Sample

Example of how to create a single sample for automatic execution. Content of the Watched Folder:

```
/var/www/analysis/watched_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 Watched Folder will result in a job with the name “**BC-112\_NA1473-FFPE\_S11\_L001**”



Always ensure the FASTQ or BAM files are moved/copied to the Watched Folder BEFORE the “.completed” file is created. If the FASTQ or BAM files are NOT present, the job will produce an error.

### Example for a Job Containing Multiple Samples

```
/var/www/analysis/watched_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 four samples being analyzed at the same time. Note that the file “4\_samples.completed” is located in the top level of the Watched Folder AND is created AFTER all the FASTQ files have been moved/copied/linked, to avoid errors in the job.

### Location of Files in the Watched Folder upon Analysis

Files (or links to files) that are placed in the Watched Folder will be moved to the special directory “picked\_up\_files” (in the top level of the watched directory) where they will remain.

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

```
/var/www/analysis/watched_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 Watched Folder AFTER the files are picked up for analysis:

```
/var/www/analysis/watched_folder_ARR
|
+ - picked_up_files
  |
  + - BC-112_NA1473-FFPE_S11_L001_R1_001.fastq
  + - BC-112_NA1473-FFPE_S11_L001_R2_001.fastq
```

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.

### File Name Collisions

If the FASTQ files for a sample that had been previously analyzed by Archer Analysis are placed in the Watch Folder again, the system will avoid the files from being overwritten in the “**picked\_up\_files**” directory by appending a number (incrementing for each duplicate found) to the file name in the “picked\_up\_files” directory. Here’s an example of the situation BEFORE the sample BC-112 is run again:

```
/var/www/analysis/watched_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/watched\_folder\_ARR

```

|
+ - picked_up_files
|
+ -BC-112_NA1473-FFPE_S11_L001_R1_001.fastq
+ -BC-112_NA1473-FFPE_S11_L001_R2_001.fastq
+ -BC-112_NA1473-FFPE_S11_L001_R1_001.fastq-1
+ -BC-112_NA1473-FFPE_S11_L001_R2_001.fastq-1
+ - BC-112_NA1473-FFPE_S11_L001_R1_001.fastq
+ - BC-112_NA1473-FFPE_S11_L001_R2_001.fastq

```

The job will still run, and the job will have the same name as the sample, but the FASTQ file name will be changed.

### S3 Watched Folders

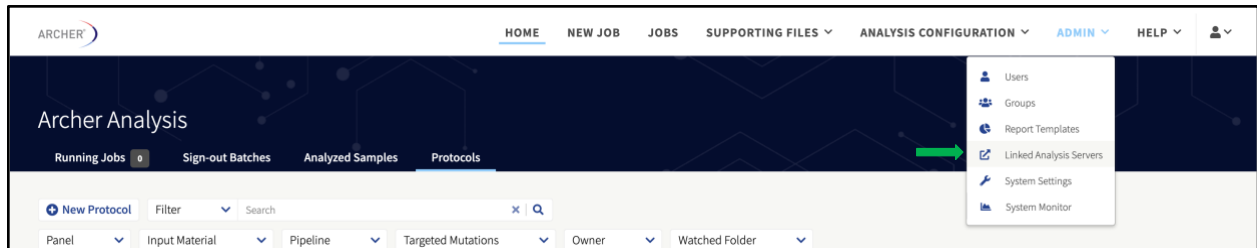
Unlimited instances have the option of using Amazon S3 as a Watched Folder location. Additional setup of the S3 Bucket is required. Please contact [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com) for additional details.



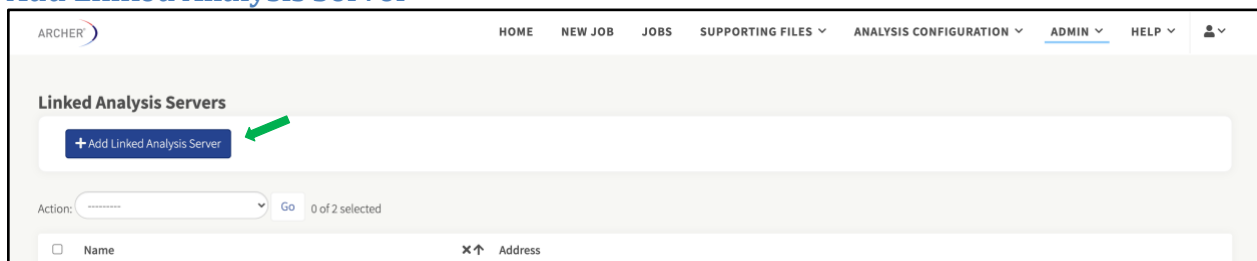
Customers can also use the REST API for job submission (SSL optionally supported) as an alternative to using Watched Folders. Please refer to <server url>/api/swagger for REST API documentation, and please contact [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com) for additional questions.

### SSL-Enabled Linked Analysis Servers

Linked Analysis servers allow users to clone jobs between different versions of instances of Archer Analysis. Admin users can access “Linked Analysis Servers” by selecting “**Admin**” and then “**Linked Analysis Servers**”. Users can reference secure servers that have HTTPS enabled via the REST API.



### Add Linked Analysis Server

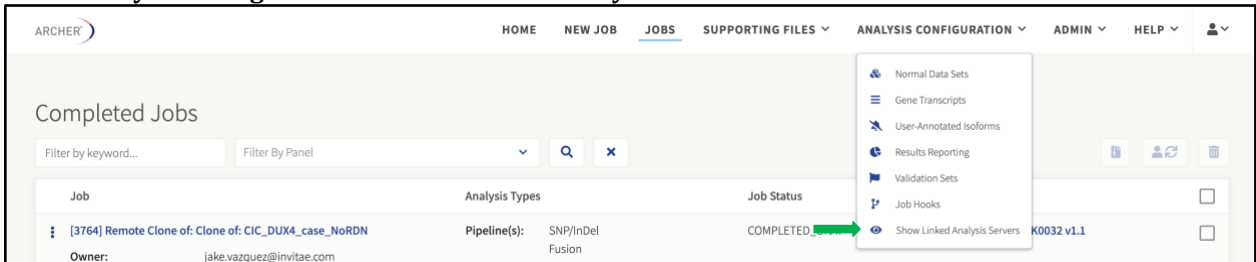





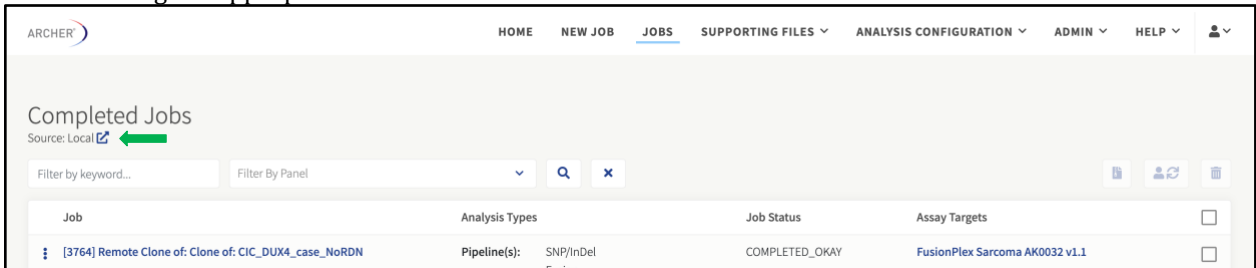
## Enable Linked Analysis Server

Admin users are given permission to link Archer Analysis instances. From the menu, select **Jobs** to access all the results that were run through the Archer Analysis pipeline.

Select **Analysis Configuration** → **Show Linked Analysis Servers**.



Under Jobs, Admin users will be able to see Source: Local  and change the server source to a different instance using the appropriate access credentials.



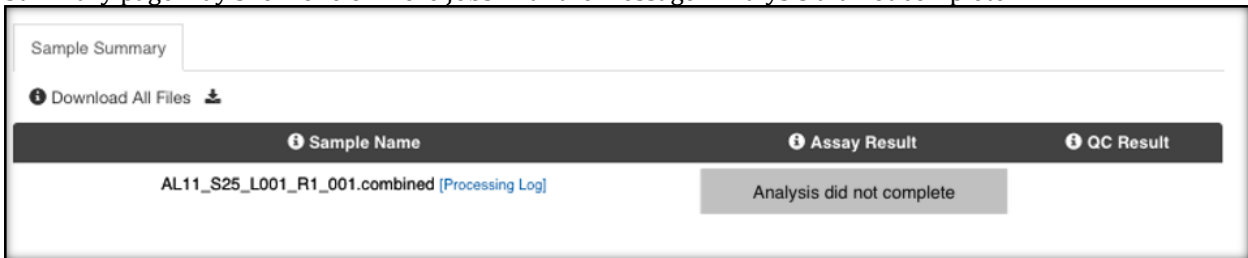
Once the source has been switched, users can search for jobs run on different instances of Archer Analysis on the Jobs page, and then clone a job from the linked server onto the local server by selecting **Clone Job** using the more-actions menu.

## Troubleshooting

This section covers problems that may arise during and immediately after the processing of jobs. For problems that arise during installation of the VM, please consult the *Archer Analysis VM Installation Guide*.

### Job finished with job status “COMPLETED\_ERROR”

This status indicates the detection of errors in one or more of the samples during processing. The Sample Summary page may show one or more jobs with the message “Analysis did not complete.”



If only one or a few of the samples finished with errors, it is likely the error occurred during the processing of the sample. Access the “**three-dot**”/”**more actions**” menu under “**View Logs**” to determine the location of the error. The easiest way to find the error is to search for the text “[ERROR].”

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] > 'an'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.
```

This may occur if the /var/www/analysis directory on the virtual machine or unlimited instance has reached its capacity. **Do not delete the files directly from the file system.** Navigate to the jobs page to delete analyses that are no longer needed, in order to free up space. Alternatively, provision the virtual machine with more disk space, when required. If you are using an Unlimited instance, contact your sales representative to request 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 files to [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com) for support.

### Job is stuck with job status “NEW”

This indicates the queue manager software (“poller”) is not running. The easiest way to restart the poller is to perform a complete restart of the virtual machine or email [adx-tech-support@invitae.com](mailto:adx-tech-support@invitae.com) when running samples on an Unlimited Instance.

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 the job will display a “COMPLETED\_ERROR” status. Re-run the job by accessing the “**three-dot**”/”**more actions**” menu and select “**Rerun Job**” to restart the job with the same settings.

### Job is stuck with job status “HOLD”

This status indicates the job is still being processed but may also be the result of a processing step crashing. A job will display a “HOLD” status when it is processing the FASTQ files after upload.

If the job continues to display a “HOLD” status (for large files this could take up to an hour), restart the VM or restart the poller as shown in *Job is stuck with job status “NEW”*.

## Appendix

### Glossary of General Terms

- GUI** – Graphical User Interface. This is accessed via any modern web browser; Google Chrome is recommended.
- VM** – Virtual Machine. Archer Analysis includes all elements of a standalone computer, which can be run on top of virtualization software; VMware is recommended.
- CLI** – Command Line Interface. This is means for directly interacting with the underlying components of Analysis (Linux OS, etc.) by means of a terminal (console) window.
- Intergenic** – Noncoding DNA regions between genes.
- Depth (DP)** – The total high quality unique molecule depth covering a variant. For SNP/InDel calling, high quality unique molecules are generally based on the per-nucleotide basequals of the involved bases. Only molecules meeting the criteria will be considered part of the support at the variant location.
- Molecular Bins** – Unique fragments/molecules from the original sample, as determined by use of molecular barcodes (MBCs) and deduplication. Molecular bins are defined as reads having the same random molecular barcode in the ligated adapter. This also requires the reads to have all the same start sites. The same MBC sequence could be seen in different start site locations.
- MBC** – Molecular Barcode in the ligated adapter.
- GSP2** – The second Gene Specific Primer (GSP) used in the nested PCR of Anchored Multiplex PCR (AMP). It is the sequence that can be found in the final read output from the sequencer.
- Fusion Candidate (FC)** – A fusion or isoform call that meets the minimum requirements specified in the User Settings.
- Alternate Observations (AO)** – The number of unique molecules supporting a variant call (alternative to the reference sequence, as defined by hg19.)
- Targeted Mutation File (TMF)** – This file contains a list of specific SNVs and InDels, in Variant Call Format (VCF), that is used with Targeted Mutation calling (Vision).

### Format of the GTF Files

The Target Region file is a GTF file containing the names and locations of the primers. Gene Transfer Format (GTF) is described at this site:

<http://genome.ucsc.edu/FAQ/FAQformat.html#format4>

The **gene\_id** field contains the name of the GSP2 and includes the name of the targeted gene.

The **transcript\_id** field contains the name of the transcripts the target is derived from (can often be empty: "").

For successful processing, the required fields in column 9 are:

- The name field containing the name of the GSP2.
- The function field containing a comma separated list of workflow pipelines that reads from this primer can be processed with.

Additional fields in column 9 that may be present but are not required and could be empty ("") include: transcript, gene\_id, exon\_number, transcript\_id, target\_ROI, and CNV\_group.

See below for a sample VariantPlex panel GTF file.

```
chr1 ArcherAssay GSP2 115251081 115251115 42 + . gene_id "NRAS_chr1_115251080_35_p_GSP2"; transcript_id "";function "CNV,SNV"; target_ROI "chr1:115251135-115251135"
chr1 ArcherAssay GSP2 115251300 115251325 42 - . gene_id "NRAS_chr1_115251299_26_n_GSP2"; transcript_id "";function "CNV,SNV"; target_ROI "chr1:115251135-115251135"
chr1 ArcherAssay GSP2 115252158 115252183 42 + . gene_id "NRAS_chr1_115252157_26_p_GSP2"; transcript_id "";function "CNV,SNV"; target_ROI "chr1:115252169-115252169"
chr1 ArcherAssay GSP2 115252249 115252276 42 - . gene_id "NRAS_chr1_115252248_28_n_GSP2"; transcript_id "";function "CNV,SNV"; target_ROI "chr1:115252169-115252169"
chr1 ArcherAssay GSP2 115256370 115256400 42 + . gene_id "NRAS_chr1_115256369_31_p_GSP2"; transcript_id "";function "CNV,SNV"; target_ROI "chr1:115256400-115256400"
chr1 ArcherAssay GSP2 115256628 115256662 42 - . gene_id "NRAS_chr1_115256627_35_n_GSP2"; transcript_id "";function "SNV"; target_ROI "chr1:115256400-115256400"
chr1 ArcherAssay GSP2 115258636 115258658 42 + . gene_id "NRAS_chr1_115258635_23_p_GSP2"; transcript_id "";function "CNV,SNV"; target_ROI "chr1:115258650-115258650"
chr1 ArcherAssay GSP2 115258791 115258819 42 - . gene_id "NRAS_chr1_115258790_29_n_GSP2"; transcript_id "";function "CNV,SNV"; target_ROI "chr1:115258650-115258650"
```

```
chr1Δ ArcherAssayΔprimer_bindΔ120458152Δ 120458177Δ 1.0Δ-Δ .Δ name "NOTCH2_chr1_120458151_26_-_A1_GSP2";
chr1Δ ArcherAssayΔprimer_bindΔ120458319Δ 120458340Δ 1.0Δ+Δ .Δ name "NOTCH2_chr1_120458318_22+_A1_GSP2";
chr1Δ ArcherAssayΔprimer_bindΔ120458391Δ 120458413Δ 1.0Δ-Δ .Δ name "NOTCH2_chr1_120458390_23_-_A1_GSP2";
chr1Δ ArcherAssayΔprimer_bindΔ120458569Δ 120458588Δ 1.0Δ+Δ .Δ name "NOTCH2_chr1_120458568_20+_A1_GSP2";
chr1Δ ArcherAssayΔprimer_bindΔ120458597Δ 120458616Δ 1.0Δ-Δ .Δ name "NOTCH2_chr1_120458596_20_-_A1_GSP2";

gene_id "NOTCH2"; exon_number "34"; transcript_id "NM_024408"; function "SNV,CNV_NO_DISPLAY"; target_ROI "chr1:
gene_id "NOTCH2"; exon_number "34"; transcript_id "NM_024408"; function "SNV,CNV_NO_DISPLAY"; target_ROI "chr1:
gene_id "NOTCH2"; exon_number "34"; transcript_id "NM_024408"; function "SNV,CNV_NO_DISPLAY"; target_ROI "chr1:
gene_id "NOTCH2"; exon_number "34"; transcript_id "NM_024408"; function "SNV,CNV"; target_ROI "chr1:120457922-1
gene_id "NOTCH2"; exon_number "34"; transcript_id "NM_024408"; function "SNV,CNV"; target_ROI "chr1:120457922-1
```

## Format of the Targeted Mutation File (TMF)

The format of the targeted mutation file is a standard Variant Call Format (VCF) file, of at least version 4.1, with some special INFO fields.

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 fields are:

- **Archer\_Gene** – The gene in which the mutation is found
- **Archer\_MutationCDS** – The Coding DNA Sequence (CDS) change in HGVS format
- **Archer\_MutationAA** – The Amino Acid change in HGVS format

An example of a TMF:

```
##fileformat=VCFv4.1
##FILTER=<ID=PASS,Description="All filters passed">
##fileDate=20170108
##reference=GRCh37
##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">
##INFO=<ID=CNT,Number=1,Type=Integer,Description="How many samples have this mutation">
##INFO=<ID=STRAND,Number=1,Type=String,Description="Gene strand">
#CHROM POS ID REF ALT QUAL FILTER INFO
chr1 11174395 COSM1185313 A T . . Archer_Gene=MTOR;STRAND=-;Archer_MutationCDS=c.7280T>A;Archer_MutationAA=p.L2427Q;CNT=4
chr1 11177096 COSM1662881 C T . . Archer_Gene=MTOR;STRAND=-;Archer_MutationCDS=c.6981G>A;Archer_MutationAA=p.M2327I;CNT=3
chr1 11177096 COSM4413485 C A . . Archer_Gene=MTOR;STRAND=-;Archer_MutationCDS=c.6981G>T;Archer_MutationAA=p.M2327I;CNT=1
chr1 11184577 COSM20417 C T . . Archer_Gene=MTOR;STRAND=-;Archer_MutationCDS=c.6644G>A;Archer_MutationAA=p.S2315Y;CNT=0
```



When creating a custom TMF, it is recommended to use a provided VCF file as a guide to ensure the proper format of the file is maintained. Otherwise, Archer Analysis will not be able to validate the file and upload will fail. For example, it is important the REF column contains the correct hg19 reference base. It is also important that the INFO header fields are included.

## Technology Licenses

© 2022 Invitae Corporation. All rights reserved. Archer<sup>®</sup>, FusionPlex<sup>®</sup>, VariantPlex<sup>®</sup>, Immunoverse<sup>™</sup>, Reveal ctDNA<sup>™</sup>, Archer Analysis<sup>™</sup> and Quiver<sup>®</sup> are trademarks of Invitae Corporation. Illumina<sup>®</sup> is a registered trademarks of Illumina, Inc., VMware vSphere<sup>®</sup>, VMware Fusion<sup>®</sup>, VMware Player<sup>®</sup> are a registered trademark or trademark of VMware, Inc. Life Technologies<sup>™</sup>, Ion Torrent<sup>™</sup>, Ion PGM<sup>™</sup> and Ion Proton<sup>™</sup> are trademarks of Thermo Fisher Scientific Inc. Mac<sup>®</sup> and Mac OS X<sup>®</sup> are registered trademarks of Apple Inc. Oracle<sup>®</sup> and VirtualBox<sup>™</sup> are registered trademarks or trademarks of Oracle Inc. Windows<sup>®</sup> and Internet Explorer<sup>®</sup> browser are registered trademarks of Microsoft Corporation. Chrome<sup>™</sup> browser is a trademark of Google Inc.

### Invitae

1400 16th Street  
San Francisco, CA 94103  
(800) 436-3037

[www.invitae.com](http://www.invitae.com)