

# Archer™ Analysis v7.2

Supports versions 7.0, 7.1

# Table of Contents

Chapter 1—Introduction .....	12
Before you start .....	13
Resources required to operate Archer Analysis .....	13
Preparing your data for Archer Analysis .....	13
Demultiplexed FASTQ or BAM files (Required) .....	14
Panels file (Required) .....	14
Targeted Mutations file (Optional) .....	14
Regions of Interest file (Optional) .....	14
Glossary .....	15
In this guide .....	16
Additional resources .....	16
Technical support .....	16
Overview .....	17
Intended use .....	17
Test principle .....	17
Access the software .....	17
Chapter 2—Features and workflows .....	18
Introduction .....	19
Guided tutorials .....	19
Supported assays .....	20
FusionPlex® (RNA) Assays .....	20
VariantPlex® (DNA) Assays .....	20
LiquidPlex® (ctDNA™) Assays .....	20
Immunoverse™ (RNA) Assays .....	20
Outputs, system architecture, algorithms .....	21
Navigation .....	22
Main header / main menu .....	22

Dropdown menus .....	22
Footer .....	22
Info pop-ups .....	23
Log into Archer Analysis .....	24
Create an account .....	25
The landing, or "home" page .....	26
Log out of Archer Analysis .....	26
Manage users and groups .....	27
Admin users .....	27
Users .....	28
Add or edit users .....	28
Delete users .....	30
Groups .....	30
Add and edit groups .....	31
Delete groups .....	31
Manage supporting file types .....	33
Add and manage files .....	33
File actions—View, download, visibility and ownership .....	34
Delete files .....	34
Existing panels .....	34
Existing targeted mutations .....	35
Existing regions of interest .....	35
Change visibility .....	35
Change ownership .....	36
User settings .....	37
View and manage user settings .....	37
General Archer Analysis settings .....	37
Min average unique DNA and AMBIG start sites per GSP2 .....	38
Default threshold .....	38

Min average unique RNA start sizes per GSP2 controls .....	38
<b>Default threshold .....</b>	<b>39</b>
Error correction .....	39
Read depth normalization .....	39
Barcode length .....	39
Common region set .....	39
Pre- / Post-alignment read cleaner .....	40
Merge reads in read cleaner .....	40
Debug .....	40
<b>Structural variant analysis settings .....</b>	<b>41</b>
Display intronic structural variant .....	41
Min reads for valid structural variant .....	41
Cross-contamination single panel .....	42
Cross-contamination fusion confidence threshold .....	42
Expression imbalance .....	42
Min unique start sites for valid structural variant .....	42
Structural variant percent of GSP2 reads .....	42
Expression imbalance alpha .....	42
Call transcriptional read through events weak .....	42
RNA expression visualization .....	43
Min RNA relative expression value .....	43
Max RNA relative expression value .....	43
RNA expression outliers .....	43
Min reportable RNA WT percent of reads .....	43
Require DNA structural variant to have the same chrom/strand .....	43
<b>Variations-Mutations settings .....</b>	<b>44</b>
Min depth for variant call .....	44
MAPQ threshold for variant call .....	44
<b>Default value .....</b>	<b>44</b>

Min basequal for variant call .....	45
Min allele fraction for variant call .....	45
Min phred qual score for variant call .....	45
Variant downstream ROI size .....	45
Outlier detection .....	45
Min DAO for outlier detection calculation .....	45
Deep shallow threshold .....	46
<b>CNV pipeline settings .....</b>	<b>47</b>
CNV Strong Amplification Threshold .....	47
CNV Strong Deletion Threshold .....	47
CNV P-Value Threshold .....	47
<b>MSI pipeline settings .....</b>	<b>48</b>
Minimum number of loci for MSI .....	48
Minimum depth per sample for MSI .....	48
Stable threshold for MSI .....	48
High threshold for MSI .....	48
MSI alpha .....	49
Minimum depth per locus .....	49
<b>TMB pipeline settings .....</b>	<b>50</b>
TMB Min Allele Fraction for Variant .....	50
TMB Min Depth for Variant Call .....	50
TMB Max Percent Bases Less Than or Equal to 200x Coverage .....	50
Minimum Unique Fragment Filtered On-Target Ratio .....	50
High Threshold for TMB .....	51
Low Threshold for TMB .....	51
TMB (DNA only) .....	51
Output files .....	52
<b>Miscellaneous system settings .....</b>	<b>54</b>
Items per page .....	54

Landing "Home" page .....	54
Automatically expand aggregate isoform visualizations .....	54
<b>Save new user settings or restore defaults .....</b>	<b>55</b>
<b>Identify non-default settings .....</b>	<b>56</b>
<b>Job setup and management .....</b>	<b>57</b>
Starting a job: New job page .....	57
Section 1: Set Up .....	57
Choosing the pipeline(s) .....	57
Section 2: Pipeline .....	58
Common analysis types .....	58
RNA analysis types .....	58
	60
DNA and cfDNA analysis types .....	60
Section 3: Settings .....	62
Section 4: Hooks .....	63
Section 5: Batch sign-out .....	63
Section 6: Sequenced libraries .....	64
Selecting normalization parameters for CNV .....	64
Running jobs .....	66
Progress bar and status updates .....	66
Analyzing multiple samples and jobs .....	67
Stopping a running or queued job .....	67
Jobs that complete with errors .....	67
<b>Job hooks .....</b>	<b>70</b>
Add a job hook .....	70
Script arguments .....	71
Managing existing hooks .....	72
Editing existing hooks .....	72
Enabling hooks on a per job basis .....	73

Deleting existing job hooks .....	73
Analytical validation sets .....	74
Mark samples for analytical validation set .....	74
Create an analytical validation set .....	75
Delete an analytical validation set .....	76
Run an analytical validation set .....	76
View analytical validation samples .....	76
Analytical Validation Report .....	76
Run history .....	78
Chapter 3—Accessing and interpreting results .....	79
Key assay concepts .....	80
Genome reference .....	80
Raw Reads, Unique Fragments, and Unique Start Sites .....	80
RNA vs. DNA vs. Ambiguous Reads .....	81
SEX_ID .....	82
SNP_ID .....	82
Control Genes (FusionPlex Only) .....	82
Reviewing results and interpretation workflow .....	83
Sign-out batches .....	83
Analyzed samples .....	83
Minimum number of loci for MSI .....	84
Minimum depth per sample for MSI .....	84
Stable threshold for MSI .....	84
High threshold for MSI .....	84
MSI alpha .....	84
Minimum depth per locus .....	85
Sample summary .....	85
QC results .....	86
Sample QC .....	87

Anomaly QC .....	87
Fusions QC .....	87
Variations QC .....	87
Results review and reporting .....	88
Jobs .....	89
Configuration .....	89
Modal view: .....	90
View logs .....	91
Clone job and rerun job .....	92
Job reports .....	93
Download results .....	93
Job Details page (Sample Summary) .....	93
Download files .....	94
Visualize sample data .....	95
Sample Details page .....	96
Read statistics .....	96
Assay Targets .....	97
Pipeline-specific links .....	98
Pipeline-specific information .....	98
Fusions (RNA only) .....	98
Key metrics for strong and low confidence fusion/isoform candidates .....	99
Strong fusions & oncogenic isoforms and low confidence fusions .....	99
Negative evidence criteria .....	101
User-annotated false and true positives .....	102
Visualization of fusion candidate-supporting reads in IGV.js .....	102
Tracks in the IGV.js view for fusions .....	103
Protein translations .....	103
RNA expression pipeline outputs .....	104
Outputs of the RNA expression pipeline .....	104



Per primer pair RNA expression normalization .....	104
RNA expression normalization primer pair selection .....	105
Qualitative analysis .....	105
Quantitative analyses .....	106
SNP/InDel variants (RNA and DNA) .....	107
Targeted vs. non-targeted calling .....	107
CNVs (DNA only) .....	108
CNV gene/group plots .....	108
Showing CNV baselining primers .....	109
MSI (DNA Only) .....	109
MSI plots and grid .....	110
Output files .....	111
Overall Aggregate Sample Results .....	111
Individual sample per locus results .....	111
Outlier detection .....	112
Enabling/disabling outlier detection .....	112
Using outlier detection in the variant grid .....	113
Key outlier detection metrics .....	113
Lymphoma Cell of Origin (Experimental) .....	114
Immune Repertoire (RNA Only) .....	116
Job-level results .....	117
Sample-level results .....	117
Sample Details page (Read Statistics tab) .....	118
Sample Details page (Clonotype Summary tab) .....	118
Per sample plots .....	119
Post-processing after Archer Analysis .....	120
Filtering results .....	120
Editing and saving grid columns .....	121

Normal Data Sets .....	122
Normal Data Sets terminology .....	122
Guidelines for Normal Data Sets .....	123
User settings for Normal Data Sets .....	123
Creating a Normal Data Set .....	123
Sample visibility .....	124
Adding a Sample to a Normal Data Set from Job Details .....	124
Modal view: .....	125
Existing Normal Data Sets .....	126
Submitting a new job with Normal Data Sets .....	127
Protocols and Normal Data Sets .....	128
Deprecating and upgrading Normal Data Sets .....	128
Workflow automation and integrations .....	129
Protocols and Watched Folders .....	129
Create a new protocol .....	129
Start a new job using an existing protocol .....	132
Search and filter existing protocols .....	132
Configure existing protocols .....	133
Set up Watched Folders .....	134
Add a new Watched Folder .....	134
Remove a Watched Folder .....	134
Remove assets used in a protocol .....	135
Develop a script for the movement of FASTQ or BAM Files .....	135
How Archer Analysis knows when to execute a workflow .....	135
Example for a job containing a single sample .....	136
Example of a job containing multiple samples .....	136
Location of files in the Watched Folder upon analysis .....	137
File name collisions .....	138
S3 Watched Folders .....	138

SSL-Enabled Linked Analysis Servers .....	139
Troubleshooting .....	141
Common problems .....	141
Appendix A .....	143
Format of the GTF file .....	143
Appendix B .....	144
Format of the Targeted Mutation File (TMF) .....	144
Technology licenses .....	145
Technical support .....	145

# Chapter 1—Introduction



# Before you start

## Resources required to operate Archer Analysis

While interacting with Archer Analysis on a standard web browser, data is processed, stored, and served via the Analysis Server package. This server *must be* accessible to your web browser and running one of these three options:

1. Your local machine.
2. Other hardware on the local intranet.
3. A remotely hosted location on the internet.



**Tip:** 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.



**Note:** 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.



**Note:** 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, contact our [Technical Support](#).

## 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 or BAM files (Required)

For each sample to be processed in a job, the corresponding FASTQ file (for Illumina data) or BAM file (for Ion Torrent data) must be uploaded to Archer Analysis.



**Tip:** Both R1 and R2 files are required for samples that are paired-end sequenced.

Understand 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
- If the FASTQs are separated into different files corresponding to lanes on the sequencer for one sample, these files must be concatenated into two FASTQs, one file each for R1 and R2, *before* uploading
- 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 samples 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 "Add and manage files" on page 33. Consult the product literature to determine which file should be used and where to source it. Contact [Technical Support](#) 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.



**Tip:** 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.

# Glossary

Listed here are some common terms used within the Archer Analysis bioinformatics platform.

Term	Definition
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	The 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 base equals 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 Specicic 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	A fusion or isoform call that meets the minimum requirements specified in the User Settings.
Alternate Observations	The number of unique molecules supporting a variant call (alternative to the reference sequence, as defined by hg19).
Targeted Mutation File	This file contains a list of speciCic SNVs and InDels, in Variant Call Format (VCF), that is used with Targeted Mutation calling (Vision).

## In this guide

This guide will explain how to use Archer Analysis, a comprehensive bioinformatics suite designed to use in conjunction with Archer library preparation kits and [gene enrichment panels](#) for targeted sequencing of select genes and regions of interest (ROIs) using next-generation sequencing (NGS).

## Additional resources

View videos and additional resources for Archer products [here](#).

## Technical support

Visit our [support center](#) for a list of helpful answers to frequently asked questions, or [contact us](#) directly.



# Overview

## Intended use

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) as well as variant annotation. 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 the Archer Analysis Unlimited service. A free demo of this service can be accessed [here](#).

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 you start" on page 13 for additional information.

Visit our online platform [portal](#), contact your field team, or our [Technical Support](#) for further details and assistance in accessing any of the options listed here.

# Chapter 2—Features and workflows



# Introduction

Archer® Analysis is a comprehensive bioinformatics suite that includes pipelines for all of Archer's targeted NGS assays, both catalog and custom. The software covers:

- All secondary analysis (e.g., read trimming, QC, alignment, and mutation calling)
- The initial steps of tertiary analysis (e.g., annotations from [Ensembl Variant Effect Predictor](#) (VEP) and [ClinVar](#))

Once optimized, the Archer Analysis software is an excellent tool for routine use by both expert and novice users. Technical specialists are also available by [contacting us](#) to assist in this process.

## 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.

# Supported assays

The pipelines supported by the software are described here.



**Note:** The pipelines in any given product may 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
- **DNA Target Coverage** – to assess panel performance and unique coverage depth across the intended targets (the target region of interest (ROI))

## 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

## Immuniverse™ (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

# Outputs, system architecture, algorithms

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 user guide focuses on the GUI. View **Reports and other output files** and **Output from download all files** for more information about these features.



**Important!** This guide 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 more explanation is needed for security, privacy, or other concerns of your lab, additional documentation is available on these particular topics:

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

Contact your field team or our [Technical Support](#) for more in-depth documentation, or assistance.

# Navigation

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

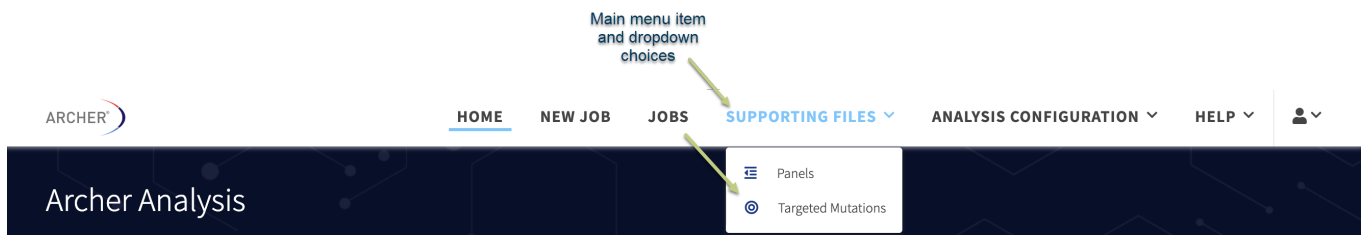
## Main header / main menu

The top of every page features a header with these persistent 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

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, and more.



For Research Use Only.  
Not for use in diagnostic procedures.

## Info pop-ups

Throughout Archer Analysis, you can click or hover over a metric, term, or 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.



**Tip:** Make use of these tool tips for a quick reference for definitions of terms, fields, or explanations of how parameters are being used in the software.

Hover over the icon and dotted line to read tooltips

Job: C	Reads	Reads
7787		59
557	62.17	chr10:61665880,chr10:43612032
355	82.94	chr10:51582939,chr10:43612032

Search for a keyword that is contained within any column. Spaces between keywords are considered an OR statement. To search for a space as part of a keyword, put quotes around the entire keyword phrase (e.g., "a / b").

Reads  
The number of unique reads supporting the event.

# Log into Archer Analysis

Archer Analysis is accessible via most web browsers except for Internet Explorer® which is incompatible and unsupported. Google Chrome™ is the recommended web browser. For Archer Analysis Unlimited instances, enter the URL that was chosen when establishing the instance (or <https://analysis.archerdx.com> for the demo site).

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.



**Important!** The Archer Analysis Software is a fully contained and secure environment, allowing you to run your analyses under your own account.

The login screen allows you to sign into your account, register for a new account, and reset your password.

ARCHER

HELP ▾

ARCHER Analysis

Log in to enter the site.

EMAIL ADDRESS

Enter email

PASSWORD

Enter password

Log in

Register new account

Create an account  
Forgot your password?

Reset 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.



**Important!** There is a limit of five (5) login attempts for incorrect passwords, after which the account will be de-activated. You 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 (case sensitive) as the login, create a password, then specify two security questions/answers. For security reasons, passwords require the following complexity:
  - Cannot be similar to personal information
  - Must contain at least nine (9) characters
  - Cannot be a commonly used 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.



**Note:** Accounts on the demo site must be approved by the Archer support team and your email verified before they become active.

Account Information	
Email	<input type="text"/>
Password	<input type="password"/>
Repeat Password	<input type="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	<input type="text" value="-----"/>
Security Answer	<input type="text"/>
Security Question 2	<input type="text" value="-----"/>
Security Answer	<input type="text"/>

Software Terms of Service

[Click here](#) to view the Software Terms of Service.

I agree to the Software Terms of Service.

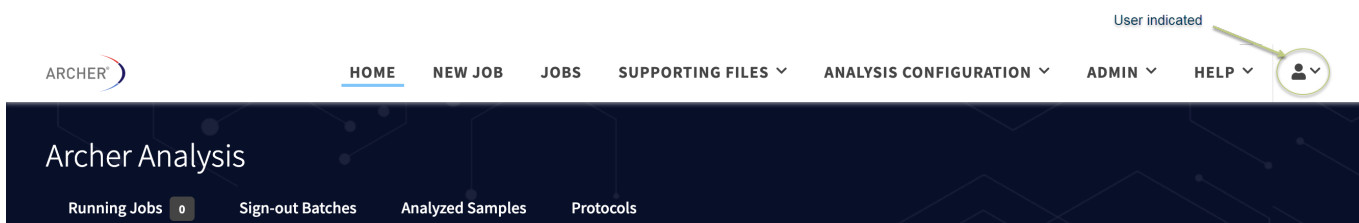
[Create Account](#)

## The landing, or "home" page

After successfully logging in, the Archer Analysis home page displays.



**Note:** If you have been assigned Admin privileges, or assigned to a User Group with jobs sharing enabled, you may already notice jobs started by other users.



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

### Log out of Archer Analysis

Click the **user icon**→**Logout**.



**Note:** Users will be automatically logged out after **thirty minutes of inactivity**.

# Manage users and groups

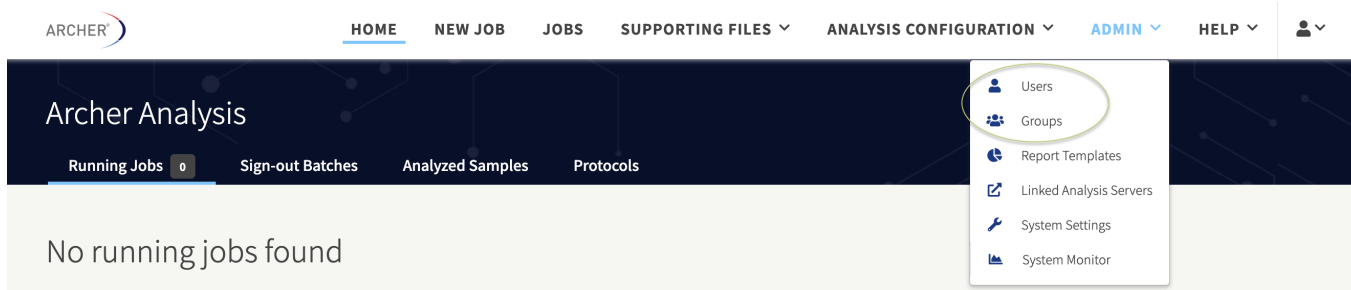
Archer Analysis utilizes user groups to control the visibility of job and sample data, as well as access to global system settings. There are two user types: **admins** and **basic users**. Additional permissions and data access granularity can be defined via custom user groups. Software permissions can be applied on a per-user and a per-group basis.



**Tip:** A user 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.

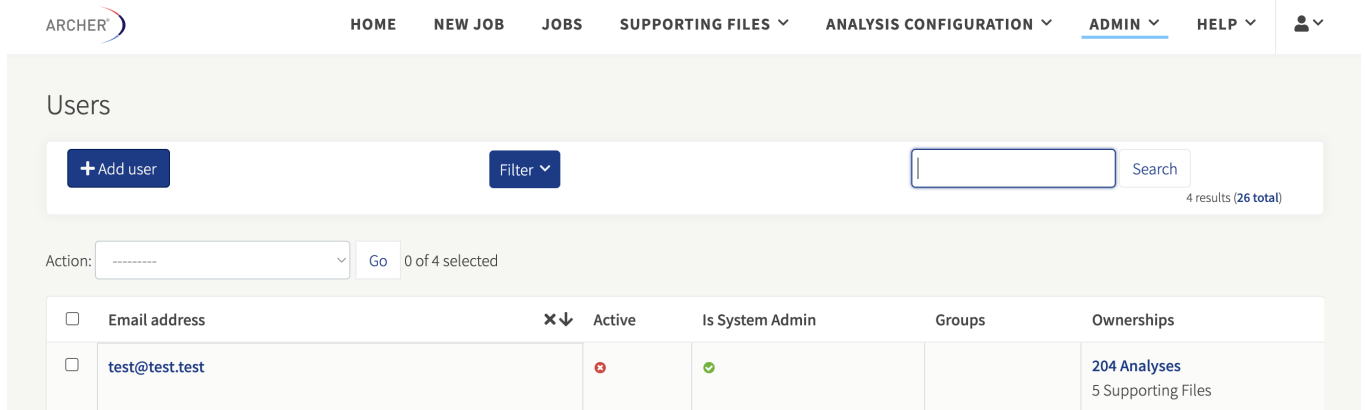
One of the primary steps in installing Archer Analysis as a virtual machine is to set up the first admin account via the command line interface (CLI). Refer to the *Archer Analysis Virtual Machine Installation Guide* for further details. For Archer Analysis Unlimited (AAU) instances, admin user(s) will be added by Archer 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:

- User and group management
- Access to experimental features
- Ability to view all jobs in the system
- Ability to change ownership of jobs to any user
- Management of Targeted Mutations (VCF) and Assay Targets (GTF) files

# Users

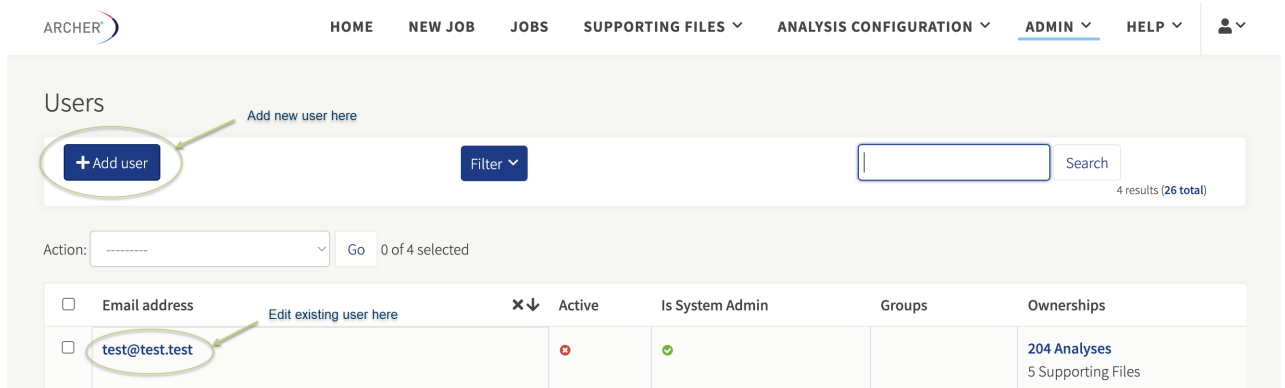
Selecting **ADMIN**→**Users** will navigate to a table containing a list of all the users in the system. The columns indicate the status of each user, whether the account is active, whether they are a designated system admin, the groups to which they belong, and how many jobs and supporting files they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



Multiple users can be selected by using the checkboxes in the left-most column, at which point they can be activated, deactivated, or deleted. You can search for users by entering the entire or part of the email address in the **Search** field.

## Add or edit users

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



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.

## Users &gt; Add user

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

Email address:

Password:

Your password can't be too similar to your other personal information.  
Your password must contain at least 9 characters.  
Your password can't be a commonly used password.  
Your password must contain 1 uppercase letter, 1 lowercase letter, 1 digit, 1 special character.

Password confirmation:

Enter the same password as before, for verification.

Add relevant information in text fields,  
then click Save.

Save

- 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.

Sample Credits: 48 Sample credits are usually purchased, but you can manually [add](#) and [subtract](#) them. [View all transactions](#)

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.

5 Inherited permissions:

Permissions inherited by the user via group permissions

6 Additional permissions:

Available Permissions

Filter

users | user | Can add protocol

users | user | Cancel jobs

users | user | Can change gene transcript priority order

users | user | Change owner of all jobs

Choose all

Chosen Permissions

Remove all

Additional permissions assigned to a user that are not inherited via group permissions

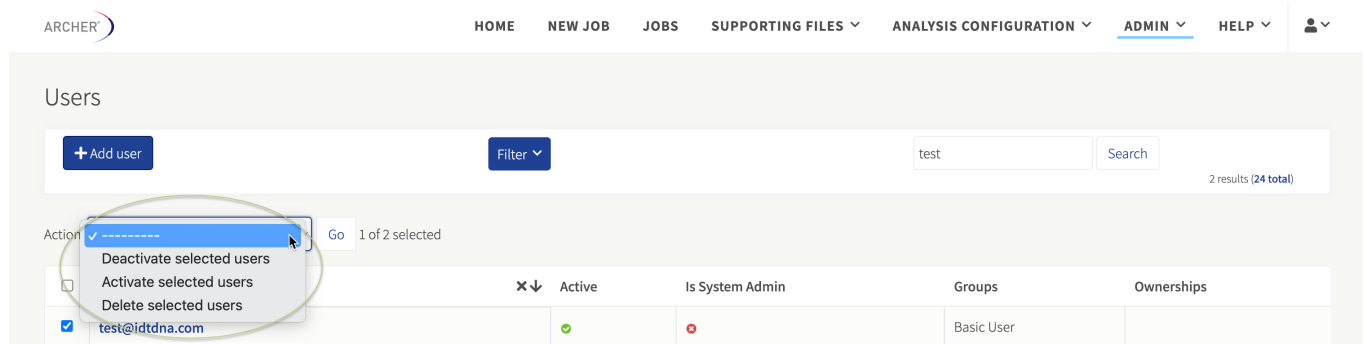
- A new temporary password can be assigned to the user.
- Users can be deactivated by unchecking the **Active** checkbox.
- Users can be granted the ability to manage users and groups by checking the **Is System Admin** checkbox.
- Group membership is controlled by choosing Available Groups from the Group Selection List and moving them to the **Chosen Groups** list.
- 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**, then using the **Action** box.



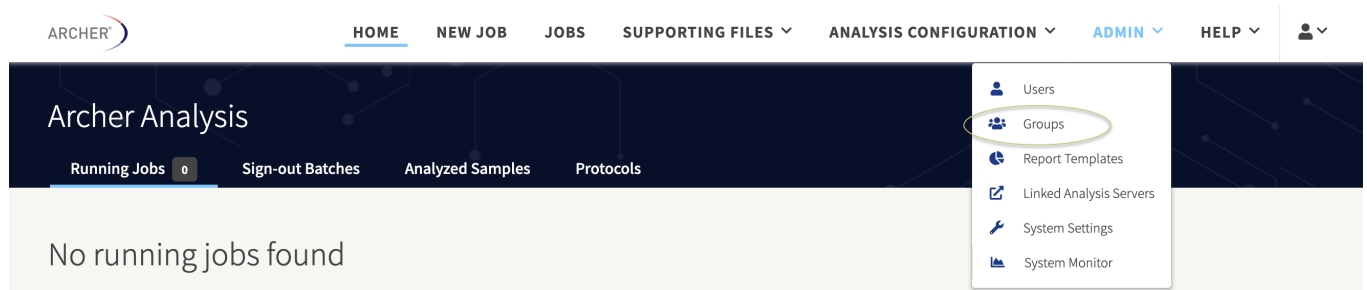
The screenshot shows the Archer Analysis interface. At the top, there is a navigation bar with links for HOME, NEW JOB, JOBS, SUPPORTING FILES, ANALYSIS CONFIGURATION, ADMIN, and HELP. Below this is the 'Users' page header. A search bar contains the text 'test' and shows '2 results (24 total)'. Below the search bar, there is an 'Action' dropdown menu with a blue background and a white border. The menu is open, showing three options: 'Deactivate selected users', 'Activate selected users', and 'Delete selected users'. The 'Delete selected users' option is highlighted. To the right of the menu, it says 'Go 1 of 2 selected'. Below the menu is a table with columns for 'Active', 'Is System Admin', 'Groups', and 'Ownerships'. The first row of the table has a checkbox checked, a green checkmark in the 'Active' column, a red 'X' in the 'Is System Admin' column, and 'Basic User' in the 'Groups' column. The username 'test@ltdna.com' is visible in the first column of the table.

## Groups

Groups serve a couple of purposes:

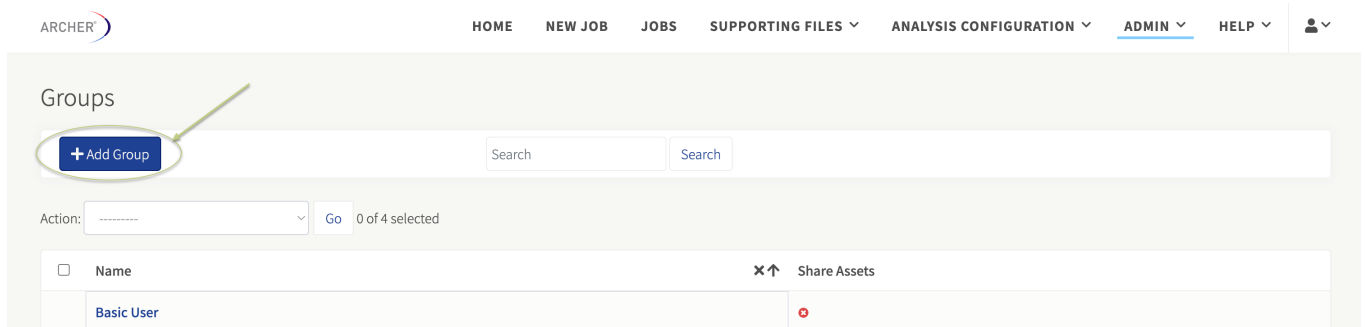
- Managing groups so many individual permissions do not have to be manually assigned to individual users
- Provide visibility into Shared Assets

Access from the main menu bar, click **Admin**→**Groups** to manage all groups.

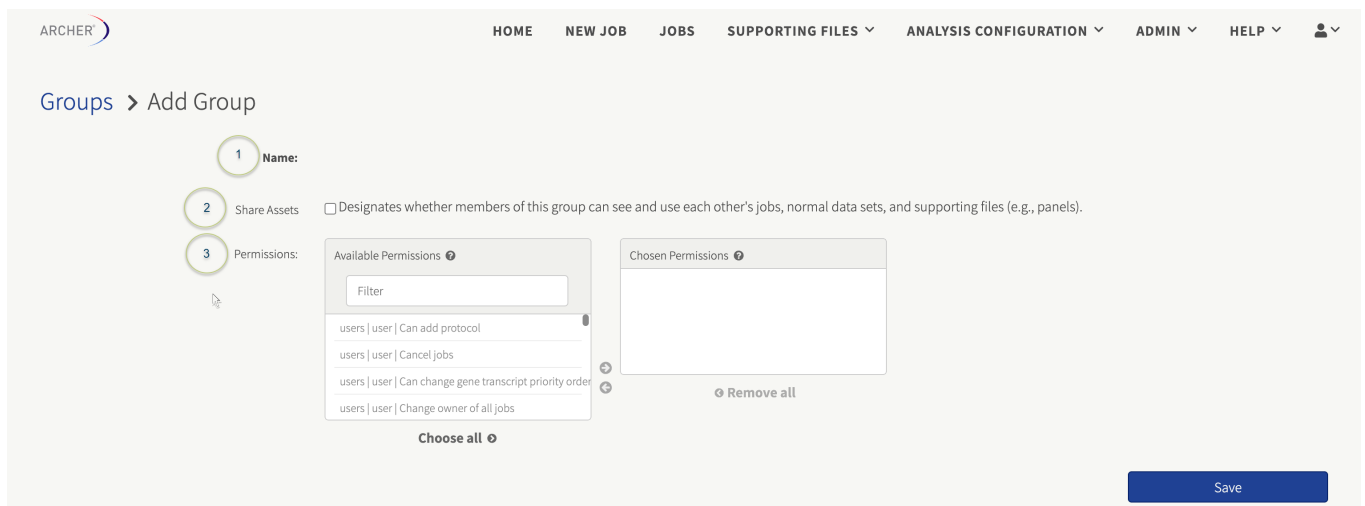


The screenshot shows the Archer Analysis main menu bar. The navigation bar at the top has links for HOME, NEW JOB, JOBS, SUPPORTING FILES, ANALYSIS CONFIGURATION, ADMIN, and HELP. Below this is a dark blue header with the text 'Archer Analysis' and several tabs: 'Running Jobs 0', 'Sign-out Batches', 'Analyzed Samples', and 'Protocols'. Below the header, there is a message 'No running jobs found'. On the right side of the header, there is a dropdown menu with several options: 'Users', 'Groups', 'Report Templates', 'Linked Analysis Servers', 'System Settings', and 'System Monitor'. The 'Groups' option is highlighted with a green circle.

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.

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



**Important!** 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 assign users to multiple groups only 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 (above).

## Delete groups

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

## Groups

+ Add Group

Search

Search

Action

✓ -----  
Delete selected Groups

Go

1 of 4 selected

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

4 Groups



# Manage supporting file types

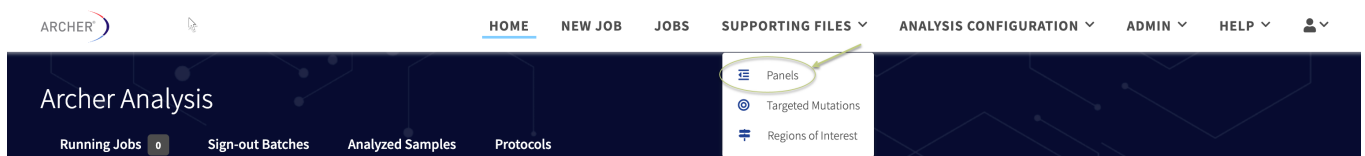
There are three types of files specified during the set-up of an Archer Analysis job:

- 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.

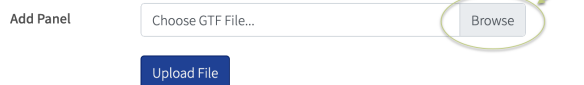
## Add and manage files

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

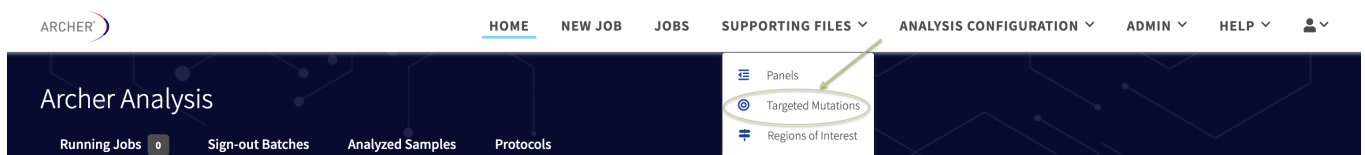


Click **Browse** to add a new GTF file. The file will be validated upon upload.

### Manage Panels

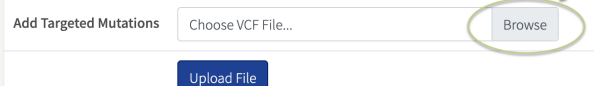


For VCF files, select **Supporting Files**→**Targeted Mutations**.

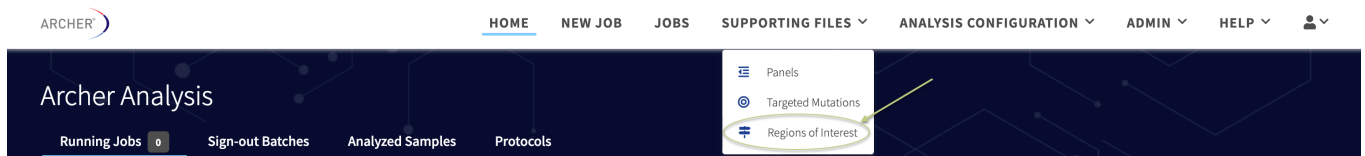


Click **Browse** to add a new VCF file. The file will be validated upon upload.

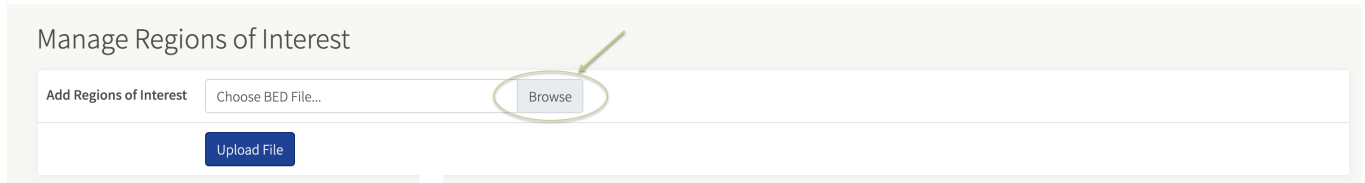
### Manage Targeted Mutations



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



Click **Browse** to add a new BED file. The file will be validated upon upload.



## 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 do one of the following:

- View the file within the browser
- Download a copy of the file
- Edit the visibility
- Change ownership

### Delete files

If a file does not have any jobs or protocols with which it's associated, it can be deleted. Rather, click the **Delete** option in the corresponding menu.

In the tables shown here, you can view existing file names, visibility, owner, etc. Usage, as the example shows below, indicates how many jobs have used the file and whether it is part of a protocol workflow.

### Existing panels

**Existing Panels**

lung

Name	Visibility	Usage	Owner	Group Sharing
<input type="button" value="Hand"/> FusionPlex Expanded Lung 18090 v1.3 <input type="button" value="List"/> View Panel us 5006 v1.2 <input type="button" value="Download"/> Download Panel 133 v1.2 <input type="button" value="Eye"/> Change Visibility 8090 v1.3 <input type="button" value="X"/> Delete	<input type="button" value="Globe"/>	2 jobs	System	N/A
	<input type="button" value="Globe"/>		System	N/A
	<input type="button" value="Globe"/>		System	N/A
	<input type="button" value="Globe"/>		System	N/A

## Existing targeted mutations

**Existing Targeted Mutations**

core

Name	Visibility	Usage	Owner	Group Sharing
FusionPlex Core Solid Tumor 22408 Targeted Mutations v1.0			System	N/A

- View Targeted Mutations
- Download Targeted Mutations
- Change Visibility
- Delete

## Existing regions of interest

**Existing Regions of Interest**

variantplex

Name	Visibility	Usage	Owner	Group Sharing
VariantPlex AML Focus TargetROI v1.0			System	N/A
VariantPlex AML Focus TargetROI v1.0			System	N/A
VariantPlex AML Focus TargetROI v1.0			System	N/A
VariantPlex AML Focus TargetROI v1.0			System	N/A

- View Regions of Interest
- Download Regions of Interest
- Change Visibility
- Delete

## Change visibility

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

Select	Symbol	Name	Description
<input type="radio"/>		Owner	Only the owner will see the file.
<input type="radio"/>		Group	All users who belong to any of the owner's groups that have sharing enabled will see the file.
<input type="radio"/>		Global	All users on the system will be able to see the file.
<input type="radio"/>		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.
<input type="radio"/>		Hidden	No one (including the owner) will be able to see the file when performing new analyses.

## Change ownership



**Tip:** Any user with the 'Manage all Supporting Files' permission may change a supporting file's ownership. See "View and manage user settings" on the next page to learn how.

Selecting **Change Owner** 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.

### Change File Ownership



New owner:

Cancel

Change owner

# User settings



**Note:** User settings are applied across all jobs, that means the user settings will apply to every future job started by a user.

These settings control processes that run for all samples (i.e., read cleaning, normalization), as well as for individual processing pipelines (structural variant, SNP/InDel, etc.).



**Important!** The default values have been carefully selected and should not be changed without extensive experience with the software or consulting a Field Bioinformatics Specialist. For further assistance, contact your field team, or [Technical Support](#).

## View and manage user settings

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




**Note:** 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 Archer Analysis settings

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

## User Settings

General Analysis Settings	
Min Average Unique DNA and AMBIG Start Sites per GSP2	<input type="text" value="50"/>
Min Average Unique RNA Start Sites per GSP2 Controls	<input type="text" value="10"/>
 Error Correction	<input type="text" value="On"/>
Read Depth Normalization	<input type="text" value="3500000"/>
Barcode Length	<input type="text" value="8"/>
Common Region Set	<input type="text" value="Dynamic search for correct set"/>
Pre-/Post-Alignment Read Cleaner	<input type="text" value="Off"/>
Merge Reads in Read Cleaner	<input type="text" value="Off"/>
Debug	<input type="text" value="Off"/>
Enable Unified VCF	<input type="text" value="On"/>

### 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.

#### Default threshold

**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.



**Important!** 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 sizes 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 starts 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 encompass only reads identified as RNA reads. View "Key assay concepts" on page 80 for further explanation of unique start sites, control GSP2s, and how reads are classified as RNA/DNA/ambiguous.

## Default threshold

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.



**Important!** 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 your lab's specific requirements.

## 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.



**Tip:** 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 this setting. "0" represents NO subsampling; all reads in the FASTQ file will be analyzed. The default setting is 3500000 for most pipelines, however for cfDNA input material, the default setting is 10000000. **Note:** setting this value to 0 may increase processing time if there are many reads in the FASTQs.

## Barcode length

This setting refers to 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 1
- Illumina Expanded Set Option 2
- Legacy Ion Torrent Set

## 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 as seen here:

Structural Variant Analysis Settings	
Display Intronic Structural Variants	On
Min Reads for Valid Structural Variant	5
Cross-Contamination Single Panel	On
Cross-Contamination Fusion Confidence Threshold	0.05
Expression Imbalance	On
Min Unique Start Sites for Valid Structural Variant	3
Structural Variant 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
Require DNA Structural Variant to have the same Chrom/Strand	On

## 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 enables the detection of cross-contamination between samples run with the same panel 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 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 enables 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 (x).

## 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 (x) icon), regardless of any other criteria, the fusion candidate will be placed in the Weak Evidence bin. Transcriptional read-through 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.

### **Require DNA structural variant to have the same chrom/strand**

This setting, when ON, configures the pipeline to only report DNA fusions that have the same chromosome and strand for all legs of the fusion. When OFF, all candidates will be reported.

# Variations-Mutations settings

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

Variations/Mutations Settings	
Min Depth for Variant Call	<input type="text" value="4"/>
 MAPQ Threshold for Variant Call	<input type="text" value="0"/>
 Min Basequal for Variant Call	<input type="text" value="20"/>
Min Allele Fraction for Variant Call	<input type="text" value="0.001"/>
Min Phred Qual Score for Variant Call	<input type="text" value="1"/>
 Variant Downstream ROI Size	<input type="text" value="400"/>
Outlier Detection	<input type="text" value="On"/>
Min DAO for Outlier Detection Calculation	<input type="text" value="0"/>
 Deep Shallow Threshold	<input type="text" value="5"/>
VEP Report Everything	<input type="text" value="Off"/>

## 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.

### Default value

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.

## Deep shallow threshold

The depth at which a molecular bin will be considered a 'deep' bin. Deep bins are considered higher confidence 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 fold increase in the 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 fold decrease in the 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 MSI pipeline in Analysis.

MSI Settings	
Baseline	Off
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 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.



**Tip:** 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.

# TMB pipeline settings

This section of the *User Settings* page contains parameters that are preset in the in the TMB pipeline and can be modified within the ranges specified here.

TMB Overrides	
TMB Min Allele Fraction for Variant Call	0.001
TMB Min Depth for Variant Call	4

## TMB Min Allele Fraction for Variant

Require at least this fraction of observations (AO/DP) supporting an alternate allele in order to evaluate the position in SNPs/InDel analysis for a given sample.

## TMB Min Depth for Variant Call

Require at least this number of alternate observations (AO) supporting an allele in order to evaluate the position.

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

TMB Settings	
TMB Max Percent Bases Less Than Equal 200 Times Coverage	0.3
Minimum Unique Fragment Filtered On-Target Ratio	410
High Threshold for TMB	20
Low Threshold for TMB	5

## TMB Max Percent Bases Less Than or Equal to 200x Coverage

Maximum percentage of bases with less than or equal 200X coverage.

## Minimum Unique Fragment Filtered On-Target Ratio

Minimum unique fragment filtered on target ratio for the sample to pass TMB QC, where unique fragment filtered on target value/gsp2 count of panel should be greater than or equal to this value.

## High Threshold for TMB

The minimum score to classify a result as being TMB-High. Scores below this value but above `TMB_LOW_THRESHOLD` will be classified as TMB-Intermediate. For VariantPlex Pan Solid Tumor and VariantPlex Complete Solid Tumor panels we recommend defaults of 5 and 6, respectively.

## Low Threshold for TMB

The maximum score to classify a result as being TMB-Low. Scores above this value but below `TMB_HIGH_THRESHOLD` will be classified as TMB-Intermediate. For VariantPlex Pan Solid Tumor and VariantPlex Complete Solid Tumor panels we recommend defaults of 20 and 17.5, respectively.

## TMB (DNA only)

TMB detection can be selected in addition to other assay types if the user has a panel that is designed for TMB analysis. The TMB algorithm calculates the TMB score based on the number of variants (nonsynonymous and synonymous) that are predicted to be tumor-specific (somatic). A linear model then uses the somatic variants to predict a TMB score based on the panel used.

Due to variability in panel-based TMB calculations Archer Analysis uses an intermediate range between low and high TMB, which is configurable by the user:

- TMB low: Score below **TMB Low Threshold**
- TMB intermediate: Score above **TMB Low Threshold** and below **TMB High Threshold**
- TMB high: Score above **TMB High Threshold**

Two TMB specific quality control checks are made to ensure consistency in TMB calling. The first is the number of unique on target reads per gsp2 in the panel (Minimum Unique Fragment Filtered On-Target Ratio), which ensures the analysis is getting enough unique reads. The second is the percentage of bases under 200X coverage (Maximum Percent Bases Less than or Equal to 200X Coverage), which ensures that there is well distributed coverage of enough depth for variant calling.



**Note:** The default thresholds for the TMB QC checks and TMB calling can be configured to user-specified values. Refer to the TMB pipeline settings above.

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

<p>GAB319_R1</p> <p>Low Med High</p> <p>Unassigned</p> <p><input type="checkbox"/> Review Complete</p>	<p>431 Variants</p> <p>TMB-Intermediate TMB</p>	<p>Sample QC</p> <p>Variations QC</p> <p>TMB QC</p>
<p>GAB320_R1</p> <p>Low Med High</p> <p>Unassigned</p> <p><input type="checkbox"/> Review Complete</p>	<p>434 Variants</p> <p>TMB-High TMB</p>	<p>Sample QC</p> <p>Variations QC</p> <p>TMB QC</p>
<p>GAB321_R1</p> <p>Low Med High</p> <p>Unassigned</p> <p><input type="checkbox"/> Review Complete</p>	<p>407 Variants</p> <p>TMB-Low TMB</p>	<p>Sample QC</p> <p>Variations QC</p> <p>TMB QC</p>



**Important!** If a sample fails TMB QC the results will not be shown. However, if you acknowledges the QC fail, you will then be able to see the results, such as the sample below.

<p>Variants</p> <p>TMB</p> <p>Assay Targets</p> <p>Read Statistics</p>	<p>&gt; Sample: GAB351_S31_R1_001</p> <p>TMB Status: TMB-High</p> <p>TMB Score: 21.8 mutations/Mb</p>
--	---

## Output files

There is an output directory named “TMB” that contains all the TMB files. There are two sample specific files (`{sample_name}_somatic_variants.csv` & `{sample_name}_TMB.tsv`) and one overall results file (`overall_results.tsv`). The sample somatic variants file contains the variants that were predicted to be tumor specific. The sample TMB file contains the following columns:

- **Name:** The name of the sample as gleaned from the R1 fastq.
- **Score:** The TMB score for the sample.
- **qc\_BASES\_LESS\_THAN\_EQUAL\_200x\_COVERAGE:** This column starts with a pass/fail description, followed by the fraction of bases less than 200X coverage and finally the QC setting for this metric.
- **qc\_UNIQUE\_FRAGMENT\_FILTERED\_ON\_TARGET:** This column starts with a pass/fail description, followed by the number of unique fragments filtered on target and finally the QC setting for this metric. Note that in the TMB job settings this is scaled by the number of gsp2s in the panel (to make it panel agnostic), and here it is shown as the total (panel specific).

The overall results file contains the following columns for each sample:

- **Name:** The name of the sample as gleaned from the R1 fastq.
- **Score:** The TMB score for the sample.
- **QC:** quality control pass/fail for the sample and if the sample failed which metrics it failed.

# Miscellaneous system settings

This section of the User Settings page contains parameters that either a) do not affect data analysis, or b) do not fit in the other page sections described earlier.

Misc Settings	
Items Per Page	25
Landing Page	Running Jobs
Automatically Expand Aggregate Isoform Visualizations	Off

## Items per page

This setting specifies the number of items that will display per page (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**:



# Identify non-default settings

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

Min Basequal for Variant Call		20
Min Allele Fraction for Variant Call	(default: 0.001)	0.008
Min Phred Qual Score for Variant Call		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 “**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 the section “Add and manage files” on page 33.



**Important!** 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 in “Sample Details page” on page 96.

New Job

1. Set Up

ENTER JOB NAME

Enter job name here

SELECT PANEL

Select

Next

## 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 sequencing platform and 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     cfDNA

### Common analysis types

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

**Read Processing Only**—Select this only to 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 "Before you start" on page 13 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 Immunoverse assays. No other pipelines can be run in the same job. This analysis type is only available for samples sequenced on the Illumina platform.



**Note:** The RNA Expression pipeline will run for all FusionPlex jobs for which the **User > Settings "RNA Expression Visualization"** is set to **"On"**.



**Tip:** 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.

## 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

### VARIANT NORMAL DATA SET

Select

Optional



**Note:** 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 DNA and cfDNA Analysis Types section below for more details on the specific SNP/InDel Detection Types.

## 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—see *Select DNA SNP/InDel Detection* 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 "Before you start" on page 13 of this guide for more information on TMFs. An optional field will display to specify a Variant Normal Data Set to be used as an alternative to intra-job normalization. See "Normal Data Sets" on page 122 for more information.
- **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 *Selecting Normalization Parameters for CNV*. An optional field will display to specify a Variant Normal Data Set to be used as an alternative to intra-job normalization. See "Normal Data Sets" on page 122 for more information.
- **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. The Target Coverage pipeline is unsupported with cfDNA inputs.

## 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  TMB  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

## 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  TMB  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

## Section 3: Settings

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



**Note:** Available settings will change based on the specific kind of job you are setting up.

Partial view of the Settings section after the **View / edit settings** button has been clicked.

### 3. Settings



Use My Settings    Use System Default Settings

General Analysis Settings	
Min.Average.Unique.DNA.and.AMBIG.Start.Sites.per.GSP2	<input type="text" value="50"/>
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="350000"/>
Barcode.Length	<input type="text" value="8"/> ▾
Common.Region.Set	Dynamic sear ▾
Pre-/Post-Alignment.Read.Cleaner	Off <input type="checkbox"/> On
Merge Reads in Read.Cleaner	Off <input type="checkbox"/> On

## 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 "Section 4: Hooks" above for more information.

### 4. Hooks - OPTIONAL



## Section 5: Batch sign-out

To facilitate downstream analysis, the laboratory/sequencing Batch ID can be added as well as the Results Review and Reporting can be assigned to any current user in the system. A Batch due date for review/reporting can be specified, too. These actions are optional.



**Tip:** 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



## Section 6: Sequenced libraries

This last section shows where the sequence files are uploaded:

- If the Illumina platform is chosen during set up, then FASTQ files for all samples to be analyzed must be uploaded. Note that both R1 and R2 files must be selected for each sample.
- If the Ion Torrent 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.](#)



**Tip:** FASTQ or BAM files to be analyzed in a job may derive from different folders on your own device. You 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 get copied to (and permanently saved) in the system. Also, FASTQ files can be 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



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
MyRNA_S20_L001_R1_001	MyRNA_S20_L001_R1_001.fastq.gz	81.77MB
	MyRNA_S20_L001_R2_001.fastq.gz	87.75MB



### 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
MyRNA_S20_L001_R1_001	<input checked="" type="checkbox"/> Tumor	1	MyRNA_S20_L001_R1_001.fastq.gz	81.77MB
	<input type="checkbox"/> Normal		MyRNA_S20_L001_R2_001.fastq.gz	87.75MB



**Tip:** 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. Upon clicking this command, the **Submit Job** function begins. A progress bar, as shown below, indicates the upload status.

### Uploading files ×

MyRNA_S20_L001_R1_001.fastq.gz	0%
MyRNA_S20_L001_R2_001.fastq.gz	0%



**Note:** 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 [this one](#).

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, you 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 **Homepage**—to change this location, refer to "Miscellaneous system settings" on page 54).

Job ID and Name: [4772] MyJobMay19

Owner: cgeaney@idtdna.com

Analysis Version: 7.2.0-rc-9

Submitted: May 19, 2023 02:42 PM

Started: May 19, 2023 02:48 PM

Pipeline(s): Fusion SNP/InDel

Input Material: RNA

STATUS: RUNNING

FusionPlex Expanded Lung 18090 v1.3

[1] MyRNA\_S20\_L001\_R1\_001 [2/14] Deduplication -> 50000 molecular bins deduplicated. 13%

Progress bar

Functionality via the dropdown menu

Job ID and Name: [4772] MyJobMay19

Owner: cgeaney@idtdna.com

Analysis Version: 7.2.0-rc-9

Submitted: May 19, 2023 02:42 PM

Started: May 19, 2023 02:48 PM

Pipeline(s): Fusion SNP/InDel

Input Material: RNA

STATUS: RUNNING

FusionPlex Expanded Lung 18090 v1.3

[1] MyRNA\_S20\_L001\_R1\_001 [3/14] Alignment -> Performing RNA Alignment 20%

Configuration

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 "Running jobs" above for more detail on reanalyzing jobs.

**Caution:** 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:

Click Samples to view sample status

Current step of the analysis workflow

[1] MyRNA\_S20\_L001\_R1\_001 [3/14] Alignment -> Performing RNA Alignment 20%



**Note:** Below each sample's progress bar, the steps of the analysis workflow are indicated. In the example above, "[3/14]" indicates that step 3 of a total of 14 steps is being executed for Sample [1].

## Analyzing multiple samples and 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.



**Note:** The VM can be configured to process multiple samples simultaneously if the VM is running on a host with sufficient resources. See the *Archer Analysis Virtual Machine Installation Guide* for further information (Contact our [Technical Support](#) for questions, and to obtain this separate document).

## Stopping a running or queued job

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

There are a few options when you elect to cancel a currently running job:

- **Continue**—continues the job and stops the cancel command
- **Delete**—stops the job and removes all the data and results from the server
- **Stop**—stops the job but leaves the sequencing files on the server and allow the samples to be rerun in the future

### Cancel Job



Are you sure you want to cancel this job?

Continue

Delete

 Stop

## Jobs that complete with errors

It is possible that a job will finish with errors. If a job fails during a run, the samples 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, or contact [Technical Support](#) for assistance.



**Tip:** In some cases, the error prevents the sample from being processed completely. This is indicated with the message **“This analysis did not complete”** in the Job Details page, as shown. To troubleshoot these errors, you will need to view the Logs. Click the ellipsis menu icon (⋮) to access the **View Logs** link.

Sample Summary

Low Med High All

Click to access the View Logs link

Sample Name	Assay Results	QC Results
S20518_S3_L001_R1_001 ⋮ This analysis did not complete.		

Click to access the #.log.stderr.txt file

Sample	Job	Assay Results	QC Results
[10206] Sample1_S1_R1_001 View History View Comments View Logs Visualize Sample Data VEP Summary	[7781] Pan Solid Pipeline(s): SNP/InDel Input Material: DNA Include Non-Targeted Mutations: Yes Mutation Detection Type: Somatic Mutation Panel: VariantPlex Pan Solid Tumor 20196 v1.3 Status: COMPLETED_OKAY	146 Variants	✔ Variations QC ✔ Sample QC

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

# Job Logs: 7781



Primary Workflow

Summarization and Staging

Workflow Configs

Content of #.log.stderr.txt file shown below

1.log.stderr.txt



SHOWING THE LAST 40 LINES

[Download Raw File](#)

```
[2023-02-28 18:53:11,714] base_call_iterator.py [INFO] > Processing bed entry 1600 of 3384...
[2023-02-28 18:53:26,185] base_call_iterator.py [INFO] > Processing bed entry 1700 of 3384...
[2023-02-28 18:53:45,263] base_call_iterator.py [INFO] > Processing bed entry 1800 of 3384...
[2023-02-28 18:54:06,433] base_call_iterator.py [INFO] > Processing bed entry 1900 of 3384...
[2023-02-28 18:54:21,177] base_call_iterator.py [INFO] > Processing bed entry 2000 of 3384...
[2023-02-28 18:54:45,282] base_call_iterator.py [INFO] > Processing bed entry 2100 of 3384...
[2023-02-28 18:55:12,716] base_call_iterator.py [INFO] > Processing bed entry 2200 of 3384...
[2023-02-28 18:55:36,590] base_call_iterator.py [INFO] > Processing bed entry 2300 of 3384...
[2023-02-28 18:56:00,385] base_call_iterator.py [INFO] > Processing bed entry 2400 of 3384...
[2023-02-28 18:56:21,323] base_call_iterator.py [INFO] > Processing bed entry 2500 of 3384...
[2023-02-28 18:56:51,062] base_call_iterator.py [INFO] > Processing bed entry 2600 of 3384...
[2023-02-28 18:57:06,351] base_call_iterator.py [INFO] > Processing bed entry 2700 of 3384...
[2023-02-28 18:57:37,822] base_call_iterator.py [INFO] > Processing bed entry 2800 of 3384...
[2023-02-28 18:58:10,159] base_call_iterator.py [INFO] > Processing bed entry 2900 of 3384...
[2023-02-28 18:58:34,435] base_call_iterator.py [INFO] > Processing bed entry 3000 of 3384...
[2023-02-28 18:58:58,994] base_call_iterator.py [INFO] > Processing bed entry 3100 of 3384...
[2023-02-28 18:59:23,257] base_call_iterator.py [INFO] > Processing bed entry 3200 of 3384...
[2023-02-28 18:59:51,535] base_call_iterator.py [INFO] > Processing bed entry 3300 of 3384...
[2023-02-28 19:00:10,740] master_workflow.py [CHECKPOINT] > Archer Analysis Pipeline
[2023-02-28 19:00:10,909] variations_timing.py [INFO] > VARIATIONS_TIMING: time spent in alignment, lofreq, and freebayes
is 780.44s; full time is 13063.52s which is 16.739x; FAIL
[2023-02-28 19:00:10,921] workflow.py [CHECKPOINT] > Sample Summary Workflow
[2023-02-28 19:00:10,922] workflow.py [STATUS] > Starting Workflow
[2023-02-28 19:00:10,922] summarize_sample.py [STATUS] > Compressing and indexing alignment files
[2023-02-28 19:00:10,922] summarize_sample.py [INFO] > Processing /tmp/archer-
temp/7781/output.1/Sample1_S1_R1_001.molbar.trimmed.deduped.sam
[2023-02-28 19:00:10,974] summarize_sample.py [INFO] > Processing /tmp/archer-
```

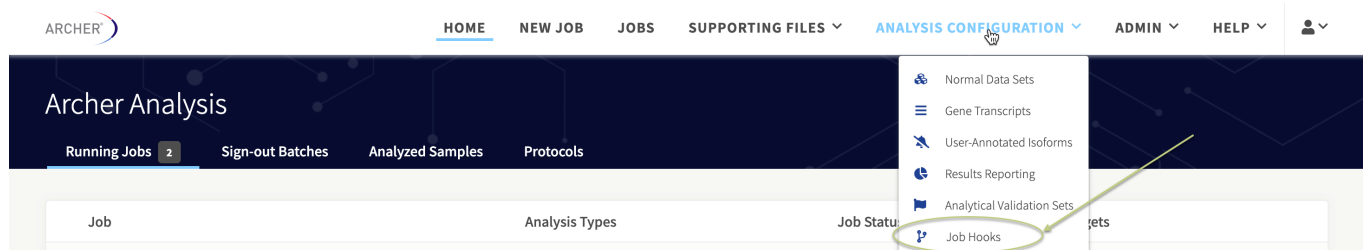
# Job hooks

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



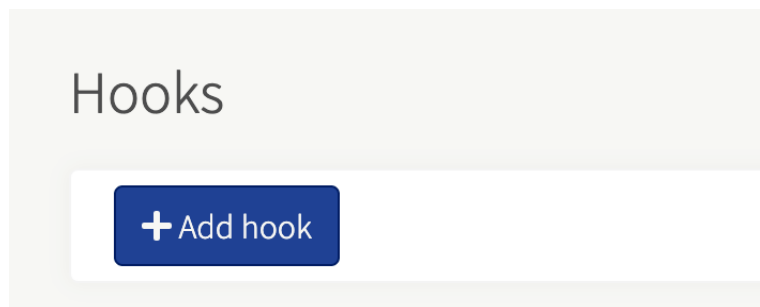
**Note** You must have Admin privileges to create job hooks. However, standard users may access and utilize job hooks once they have been created.

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



## Add a job hook

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



From the Add Hook page, you will create a new job hook by providing 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}

[Required args for all hooks](#)  
\${HOOK\_OUTPUT\_DIR}

**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.

**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 **Save**.

## 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 the desired hooks to enable/disable.

### 4. Hooks - OPTIONAL



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

#### JOB POST HOOKS

GetAverageCoveragePerTarget

Once job setup is complete, click **Submit Analysis** to begin job processing. See "Job setup and management" on page 57 for more detail.

## Deleting existing job 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 2 selected

<input type="checkbox"/>	Name	Hook type	Script file	Active	Move
<input checked="" type="checkbox"/>	GetAverageCoveragePerTarget	Job Post Hook	hooks/jobHook_aggTargetCov_incl_bugFix.py		

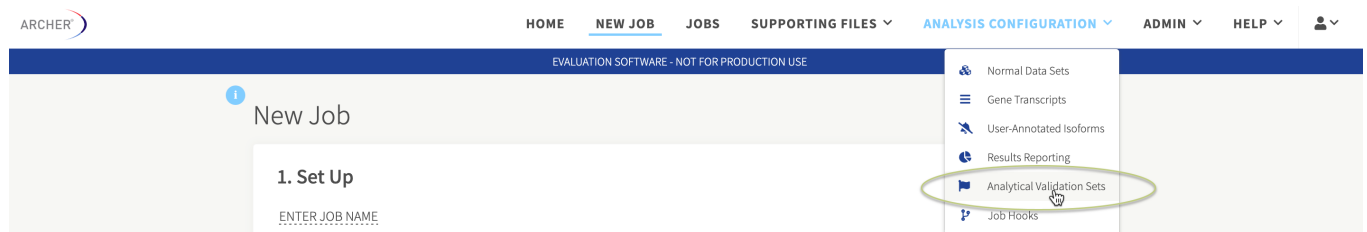
# Analytical validation sets

The use of analytical validation sets is important when you want to validate the version of Archer Analysis you are running. You can select any of your 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 you when an analytical validation set passes/fails but also provides a detailed downloadable PDF report.




**Note:** Only Admin users can run analytical validation sets.

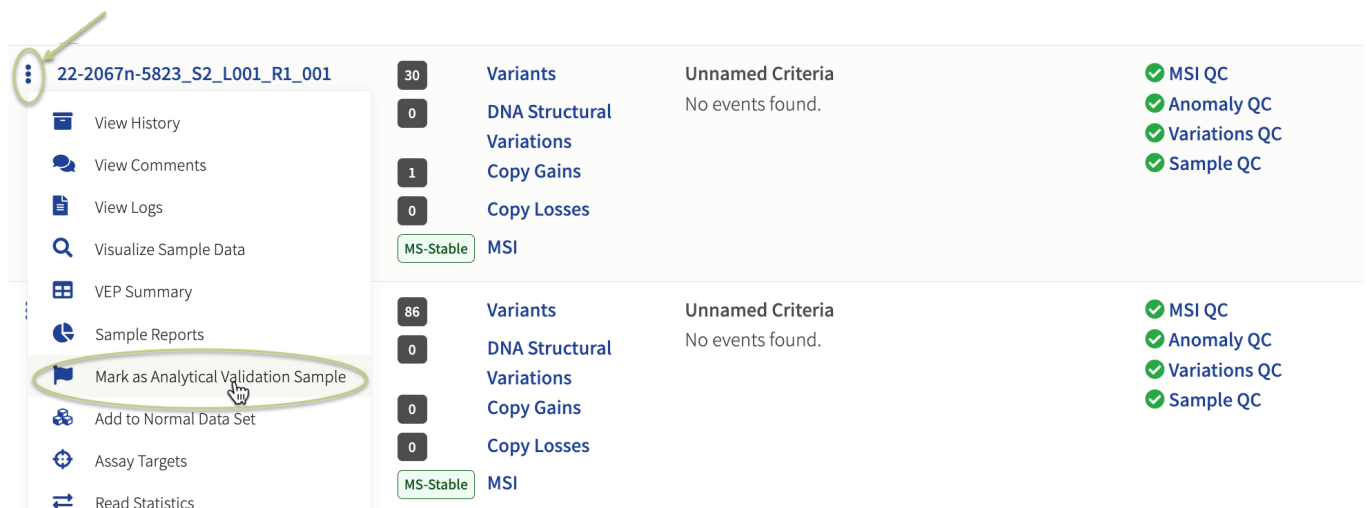
To access Analytical Validation Sets, select the **Analysis Configuration** → **Analytical Validation Sets** menu to add/manage Analytical Validation Sets.



## 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 or samples to incorporate on the Job Details page.

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



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
Archer Default	aweinstein@idtdna.com	4	10	May 22, 2023	Pass

From the Analytical Validation Sets page, click Create New Analytical Validation Set. The Analytical Validation Samples page will appear.

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

Analytical Validation Samples


Create new

+ Create New Analytical Validation Set Remove Analytical Validation Samples

Select job

Job	Analysis Types	Assay Targets	Sample Count
<input type="checkbox"/> [4636] CNV Validation Analysis Version: 7.0.0-12 Completed: April 14, 2023 12:27 PM	Pipeline(s): CNV Input Material: DNA	CNV Validation v2	3
	ARR_Fusion_V2_ROS1_0.5x_R1		Comment: Default analytical validation set
	ARR_Fusion_V2_ROS1_2x_R1		Comment: Default analytical validation set
	ARR_Fusion_V2_normals_R1		Comment: Default analytical validation set

Name the analytical validation set, then click **Ok**. The Analytical Validations Sets page appears with the newly created analytical validation set.

You 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 you to select the specific variants to include for analytical validation testing.

### Reportable Items




<input type="checkbox"/> Gene	Copy Number	Standard Deviation of Copy Number	P Value
<input checked="" type="checkbox"/> ROS1	0.50	0.00	0.00e+00

Close








Save

## Delete an analytical validation set

To delete an analytical validation set or sets from the Analytical Validation Sets page, click the  icon to the left of the analytical validation set and select **Delete**.



**Note:** 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
   Archer Default 	aweinstein@idtdna.com	4	10	May 22, 2023	Pass
   Delete	aweinstein@idtdna.com	4	10	April 14, 2023	Pass

## 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.

## View analytical validation samples

You 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.

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

## Analytical Validation Report

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

Analytical Validation Set	Archer Default
Overall Status	PASS
Started	May 22, 2023 11:18 AM
Completed	May 22, 2023 11:35 AM
Software Version	7.2.0-rc-9
Sample Count	10

### HOOK3\_11\_RET\_12\_R1

Sample Links	<a href="#">observed</a> / <a href="#">expected</a>	Job Type	<b>Pipeline(s):</b> Fusion <b>Input Material:</b> RNA
QC	Sample QC: PASS Fusion QC: PASS	Job Status	COMPLETED_OKAY
Overall Result	PASS		
Fusions Found	1 (1 expected)		
<b>Details</b>			

### Weak\_LTK\_10\_ALK\_20\_R1

Sample Links	<a href="#">observed</a> / <a href="#">expected</a>	Job Type	<b>Pipeline(s):</b> Fusion <b>Input Material:</b> RNA
QC	Sample QC: FAIL Fusion QC: NUMBER OF UNIQUE RNA GSP2 CONTROL START SITES LOW	Job Status	COMPLETED_OKAY
Overall Result	PASS		
Isoforms	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 the test was run.

## Analytical Validation Sets

[Create New Analytical Validation Set](#)

Analytical Validation Set Name	Created By	Job Count	Sample Count	Last Run	Result
Archer Default	aweinstein@idtdna.com	4	10	May 22, 2023	Pass
<a href="#">View Run</a> <a href="#">Download PDF Report</a>	<b>Started By:</b> aweinstein@idtdna.com <b>Software Version:</b> 7.2.0-rc-9	<b>Started:</b> May 22, 2023, 11:18 a.m. <b>Completed:</b> May 22, 2023, 11:35 a.m. <b>Result:</b> Pass			
<a href="#">View Run</a> <a href="#">Download PDF Report</a>	<b>Started By:</b> aweinstein@idtdna.com <b>Software Version:</b> 7.2.0-rc-9	<b>Started:</b> May 16, 2023, 2:14 p.m. <b>Completed:</b> May 16, 2023, 2:59 p.m. <b>Result:</b> 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.

# Chapter 3—Accessing and interpreting results



# Key assay concepts

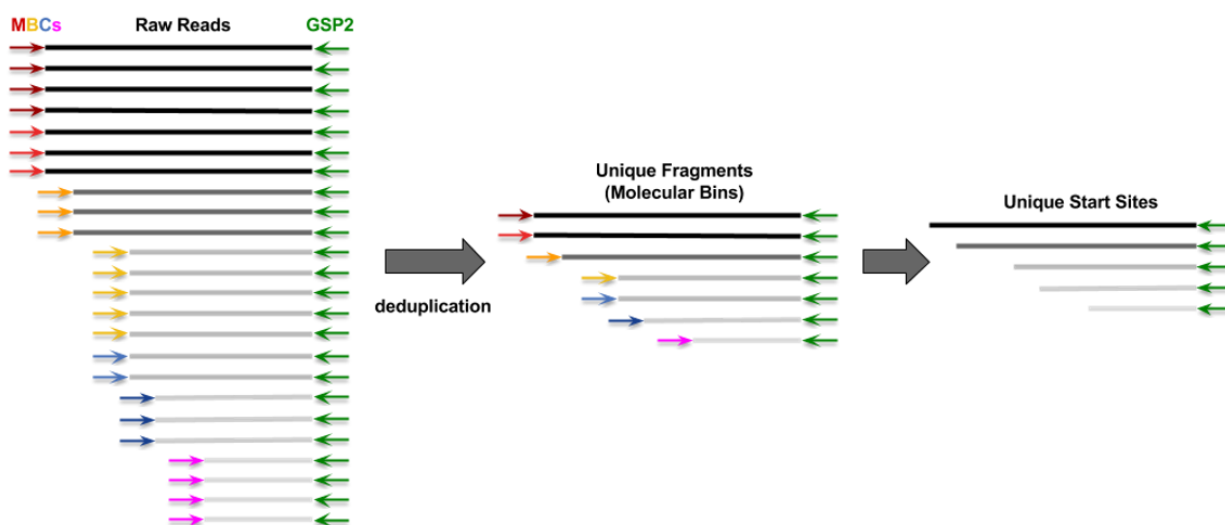
There are several key concepts that should be understood when working with Anchored Multiplex PCR (AMP) chemistry, in order to understand many of the terms and metrics in this section of the guide, as well as to properly interpret results from all Archer assays. These concepts are reviewed here.

## 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** (aka, Total Fragments)—the total number of reads (or read pairs) reported from the sequencer.

**Unique Fragments** (aka, Molecular Bins)—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 that low quality raw reads are filtered out before this step—see "Read statistics" on page 96).

**Unique Start Sites**—a subset of Unique Fragments; 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.





**Tip:** 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 more detail) 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**—reads 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.



**Tip:** 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, but instead utilizes a third-party tool, MiXCR. See "Immune Repertoire (RNA Only)" on page 116 for more information.

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

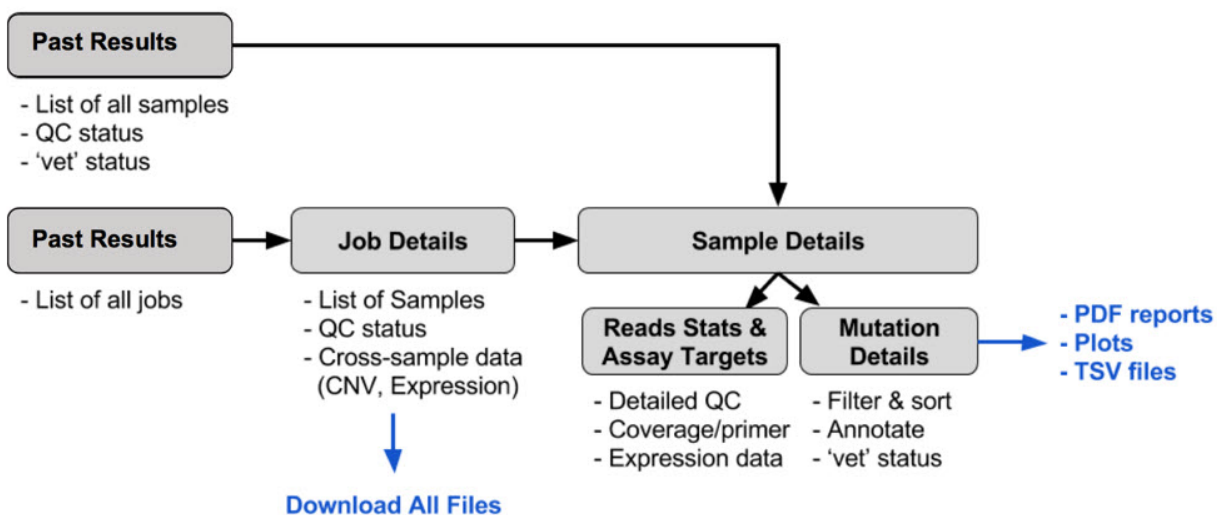
Some Archer panels contain SNP\_ID primers that can be used to report genotype across a series of highly polymorphic genomic positions in mostly intronic and intergenic regions. If a panel containing SNP\_ID primers and the appropriate function flag in the GTF is used, there will be an output file that 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 "QC results" on page 86 for further details on these points.

# Reviewing results and 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" on page 54 for further instruction):

## 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 the "MSI pipeline settings" on page 48 section which contains parameters that only affect the CNV pipeline in Analysis).

MSI Settings	
Baseline	Off
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).



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

Sample	Job	Assay Results	QC Results
[12432] 22-2075-042423-5823_S6_L001_R1_001 <small>Sample ID      Sample name</small>	[8470] Clone of: 051923 Expanded ST Pipeline(s): SNP/InDel CNV Structural Variation MSI Input Material: DNA Mutation Detection Type: Somatic Mutation Instrument: NovaSeq Panel: VariantPlex Expanded Solid Tumor 20123 v1.2 Status: COMPLETED_OKAY <small>Archer panel</small>	51 Variants 0 DNA Structural Variations 1 Copy Gains 0 Copy Losses MS-Stable MSI	MSI QC Anomaly QC Variations QC Sample QC

## 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:

Running Jobs 0 Sign-out Batches Analyzed Samples Protocols



Search Filter By Panel

Sample	Job	Assay Results	QC Results
[12432] 22-2075-042423-5823_S6_L001_R1_001	[8470] Clone of: 051923 Expanded ST Pipeline(s): SNP/InDel, CNV, Structural Variation, MSI Input Material: DNA Mutation Detection Type: Somatic Mutation Instrument: NovaSeq Panel: VariantPlex Expanded Solid Tumor 20123 v1.2 Status: COMPLETED_OKAY	51 Variants 0 DNA Structural Variations 1 Copy Gains 0 Copy Losses MS-Stable MSI	MSI QC Anomaly QC Variations QC Sample QC
[12432] 22-2075-042423-5823_S6_L001_R1_001	[8470] Clone of: 051923 Expanded ST Pipeline(s): SNP/InDel, CNV, Structural Variation, MSI Input Material: DNA Mutation Detection Type: Somatic Mutation Instrument: NovaSeq Panel: VariantPlex Expanded Solid Tumor 20123 v1.2 Status: COMPLETED_OKAY	30 Variants 0 DNA Structural Variations 1 Copy Gains 0 Copy Losses MS-Stable MSI	MSI QC Anomaly QC Variations QC Sample QC

Click icon to access available actions

- View History
- View Comments
- View Logs
- Visualize Sample Data
- VEP Summary
- Sample Reports
- Mark as Analytical Validation Sample
- Add to Normal Data Set
- Assay Targets
- Read Statistics

Click to view/edit QC status

1. Click the sample ID/name under the **Sample** column to jump to the "Sample Details page" on page 96 and get an in-depth view of all results for the sample (see Sample Details).
2. Click the job ID/name under the **Job** column to jump to the "Job Details page (Sample Summary)" on page 93 of the job the sample was run in (see Job 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" on page 96. See "Pipeline-specific information" on page 98 for more detail.
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 the "QC results" below section for more detail.

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 does 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" on page 80 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" on page 80 for explanations of unique start sites, RNA reads, and control GSP2s).



**Important!** 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.



**Tip:** Since the PreSeq™ RNA QC Assay is correlated to this Analysis QC metric, it provides a means of rescuing samples early in library preparation from further processing. Consult the [FusionPlex protocol](#) for your assay for more detail 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" on page 80 for explanations of unique start sites and DNA/ambiguous reads).



**Important!** 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.

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** pop-up, 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 Archer Analysis settings" on page 37 for more detail). You can use this pop-up to manually override the pass/fail call, and add comments about the sample; all actions are recorded and viewable in the **History** section at the top:

### QC Metrics ✕

---

#### History


History of comments & Pass/Fail alterations

Reviewer	Date	Type	Action	Comment
No history found.				

#### Metrics

From User Settings

MSI QC
Anomaly QC
Variations QC
Sample QC

Metric	Value	Threshold	Pass
Average Unique DNA Start Sites per GSP2	67.97	>= 50.0	

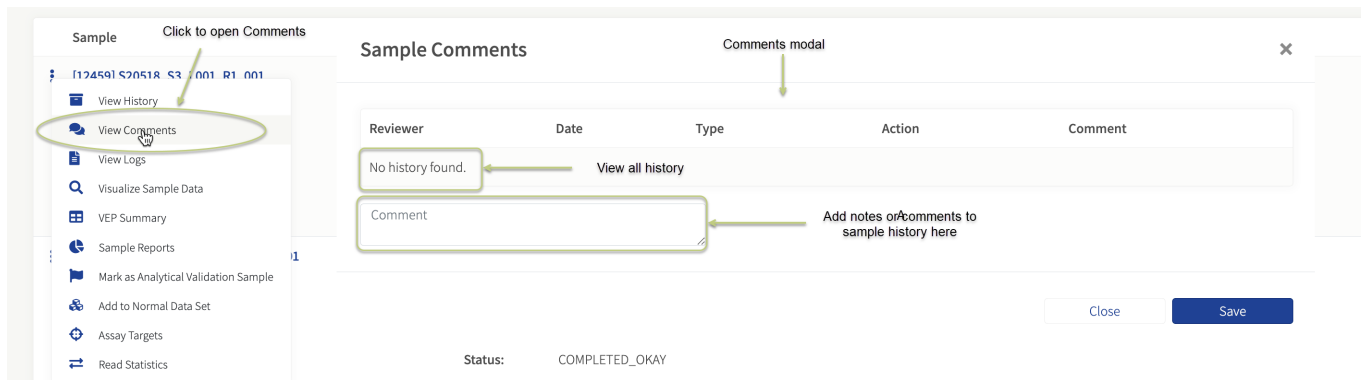
Overall Anomaly QC:  
 ▼

Comment

## 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:





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 on the main menu, under the Jobs tab.

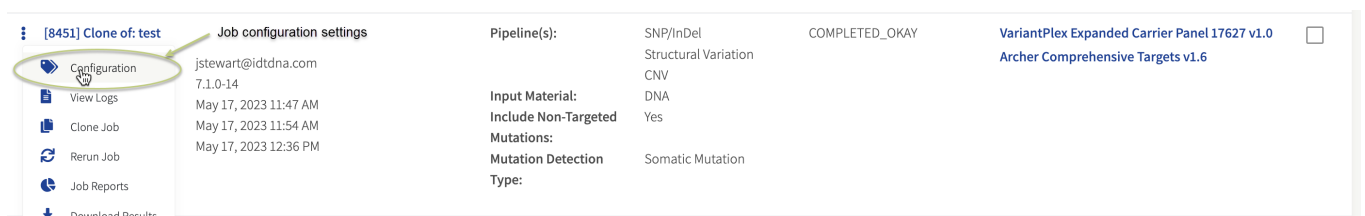
The *Jobs* page also shows a list of all completed jobs, including any jobs run by other users that the current user has permission to view (see "Manage users and groups" on page 27 for further details) Specific jobs can be located by typing a query into the search box, or by filtering by any GTF files currently loaded into the system that have associated jobs:

The job list includes descriptive information about each job, as follows:

- **Job**—includes Job ID, job name, owner, analysis version, and timestamps for when each job was submitted, started, and completed
- **Analysis Types**—The pipelines 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 the job was set up; if a Targeted Mutations file was chosen or Normal Data Sets were used, they will show 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

---

**Job Name:** [8451] Clone of: test

## DETAILS

---

**Status:**  COMPLETED\_OKAY  
**Owner:** jstewart@idtdna.com  
**Submitted:** May 17, 2023 11:47 AM  
**Started:** May 17, 2023 11:54 AM  
**Completed:** May 17, 2023 12:36 PM  
**Analysis Version:** 7.1.0-14  
**Platform:** Illumina

## ANALYSIS TYPES

---

**Pipeline(s):** SNP/InDel  
CNV  
Structural Variation  
**Input Material:** DNA  
**Include Non-Targeted Mutations:** Yes  
**Mutation Detection Type:** Somatic Mutation

## SUPPORTING FILES AND ANALYSIS CONFIGURATION

---

**Panel:** VariantPlex Expanded Carrier Panel 17627 v1.0  
**Targeted Mutation:** Archer Comprehensive Targets v1.6  
**File:**


### [Modal view:](#)

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" on page 56 for more information):

## ANALYSIS SETTINGS

General Analysis Settings	
<u>Min Average Unique DNA and AMBIG Start Sites per GSP2</u>	50
<u>Min Average Unique RNA Start Sites per GSP2 Controls</u>	10
<u>Error Correction</u>	On
<u>Read Depth Normalization</u>	System Default: 3500000 0
<u>Barcode Length</u>	8
<u>Common Region Set</u>	Dynamic search for correct set
<u>Pre-/Post-Alignment Read Cleaner</u>	Off

## 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	View job logs	Analysis Types	Job Status	Assay Targets
[8450] Clone of: ARCHER_TRAINING_1_S5		Pipeline(s): Fusion Input Material: RNA	COMPLETED_OKAY	FusionPlex MAML3 Supplement 22729 concatenated with FusionPlex Lung v2 18090 v1.3 v1.0 <input type="checkbox"/>
	<ul style="list-style-type: none"><li>Configuration canobile@idtdna.com 7.1.0-14</li><li><b>View Logs</b> May 17, 2023 10:56 AM</li><li>Clone Job May 17, 2023 11:04 AM</li><li>Rerun Job May 17, 2023 11:54 AM</li><li>Job Reports</li><li>Download Results</li></ul>	Pipeline(s): Read Processing Input Material: DNA	COMPLETED_OKAY	VariantPlex Expanded Carrier Panel 17627 v1.0 <input type="checkbox"/>



**Tip:** 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, 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:

Download main job log file

Primary Workflow

Summarization and Staging

Workflow Configs

1.log.stderr.txt



SHOWING THE LAST 40 LINES

Download Raw File

```
[2023-05-17 17:52:00,669] fusion_metrics.py [INFO] > Using FGFR1_chr8_38274839_27_-_A1_GSP2 to check mispriming for candidate 7910
[2023-05-17 17:52:02,455] fusion_metrics.py [INFO] > Using PIK3CA_chr3_178936984_32_-_A1_GSP2 to check mispriming for candidate 4608
[2023-05-17 17:52:04,290] fusion_metrics.py [INFO] > Using FGFR2_chr10_123325175_27+_A1_GSP2 to check mispriming for candidate 1711
[2023-05-17 17:52:06,005] fusion_metrics.py [INFO] > Using ALK_chr2_29446473_24+_A1_GSP2 to check mispriming for candidate 4015
[2023-05-17 17:52:06,005] off_target_by_homology_check.py [INFO] > Mispriming due to lack of exonic bases on target arm
[2023-05-17 17:52:07,710] fusion_metrics.py [INFO] > Using ROS1_chr6_117730777_22+_A1_GSP2 to check mispriming for candidate 5806
[2023-05-17 17:52:07,710] off_target_by_homology_check.py [INFO] > Mispriming due to lack of exonic bases on target arm
[2023-05-17 17:52:09,371] fusion_metrics.py [INFO] > Using NTRK2_chr9_87475960_34_-_A1_GSP2 to check mispriming for candidate 8450
[2023-05-17 17:52:11,207] fusion_metrics.py [INFO] > Using NTRK1_chr1_156845948_20+_A1_GSP2 to check mispriming for candidate 885
```




Contact [Technical Support](#) 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.




**Note:** 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” () under the  menu:


Job	Analysis Types	Job Status	Assay Targets
[8450] Clone of: ARCHER_TRAINING_1_S5 Configuration: canobile@idtdna.com, 7.1.0-14 View Logs: May 17, 2023 10:56 AM Clone Job: May 17, 2023 11:04 AM Rerun Job: May 17, 2023 11:54 AM Job Reports Download Results	Pipeline(s): Fusion Input Material: RNA	COMPLETED_OKAY	FusionPlex MAML3 Supplement 22729 concatenated with FusionPlex Lung v2 18090 v1.3 v1.0
	Pipeline(s): Read Processing Input Material: DNA	COMPLETED_OKAY	VariantPlex Expanded Carrier Panel 17627 v1.0

## 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 more detail.

Job	Analysis Types	Job Status	Assay Targets
[8450] Clone of: ARCHER_TRAINING_1_S5 Configuration: canobile@idtdna.com, 7.1.0-14 View Logs: May 17, 2023 10:56 AM Clone Job: May 17, 2023 11:04 AM Rerun Job: May 17, 2023 11:54 AM Job Reports Download Results	Pipeline(s): Fusion Input Material: RNA	COMPLETED_OKAY	FusionPlex MAML3 Supplement 22729 concatenated with FusionPlex Lung v2 18090 v1.3 v1.0
	Pipeline(s): Read Processing Input Material: DNA	COMPLETED_OKAY	VariantPlex Expanded Carrier Panel 17627 v1.0

## Download results

To download analysis results for the job, select "Download Results" (  ). See Output from "Download All Files" for more detail.

Job	Analysis Types	Job Status	Assay Targets
[8450] Clone of: ARCHER_TRAINING_1_S5 Configuration: canobile@idtdna.com, 7.1.0-14 View Logs: May 17, 2023 10:56 AM Clone Job: May 17, 2023 11:04 AM Rerun Job: May 17, 2023 11:54 AM Job Reports Download Results	Pipeline(s): Fusion Input Material: RNA	COMPLETED_OKAY	FusionPlex MAML3 Supplement 22729 concatenated with FusionPlex Lung v2 18090 v1.3 v1.0
	Pipeline(s): Read Processing Input Material: DNA	COMPLETED_OKAY	VariantPlex Expanded Carrier Panel 17627 v1.0

## 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.



**Tip:** The specific "Assay Result" categories that are shown depends on how the job was set up; see "Starting a job: New job page" on page 57 for more information.

Job: Clone of: ARCHER\_TRAINING\_1\_S5 Job Options ▾


Sample Summary    RNA Expression

Low Med High All

Sample Name	Assay Results	Fusion Results	QC Results	Reportable Mutations
18T120_S5_L001_R1_001 Low Med High * Unassigned <input type="checkbox"/> Review Complete	2 Fusions	Strong Fusions & Oncogenic Isoforms WSCD1 → BRAF <input checked="" type="checkbox"/> PGM2 → MAML3 <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> Fusion QC <input checked="" type="checkbox"/> Sample QC	

## Download files

To download files for the job, select “Download Results” (  ). This opens 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 for you to download by selecting  .

Job: Clone of: ARCHER\_TRAINING\_1\_S5 Job Options ▾

Sample Summary    RNA Expression    **Download results**

Low Med High All

Sample Name	Assay Results	Fusion Results	QC Results	Reportable Mutations
18T120_S5_L001_R1_001 Low Med High * Unassigned <input type="checkbox"/> Review Complete	2 Fusions	Strong Fusions & Oncogenic Isoforms WSCD1 → BRAF <input checked="" type="checkbox"/> PGM2 → MAML3 <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> Fusion QC <input checked="" type="checkbox"/> Sample QC	

Configuration

View Logs

Clone Job

Rerun Job

Job Reports

**Download Results**



**Tip:** 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 Output Files

Supporting Files

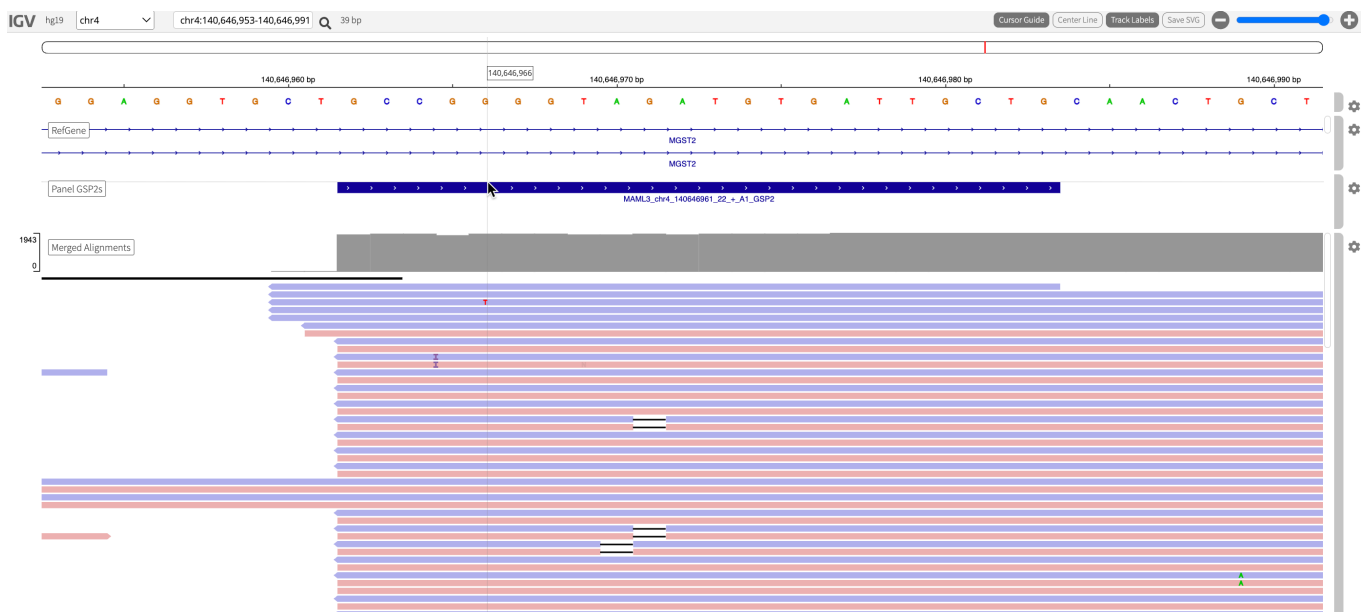
- Select All
- 18T120\_S5\_L001\_R1\_001\_full\_results.txt
- summaries/Summary.deduped.counts.txt
- summaries/Summary.deduped.counts\_pct.txt
- summaries/Summary.expression\_summary\_table.txt
- summaries/Summary.non\_deduped\_counts.txt
- summaries/Summary.non\_deduped\_counts\_pct.txt
- summaries/Summary.per\_sample\_columns\_expression\_summary\_table.txt
- 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 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, 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	
1,659,498	1,574,167 (94.9%)		1,021,694	
<a href="#">Export Data (tsv)</a>				
Fragment Type	Total Fragments	Mapped	Pass Alignment Filter	On Target %
All Fragments	1,021,694 (100.0%)	1,019,764 (99.8%)	99.8%	98.1%
Unique Fragments	80,564 (7.9%)	79,555 (98.7%)	98.7%	89.1%
<a href="#">Export Data (tsv)</a>				
DNA/RNA Statistics				
Type	DNA Reads	RNA Reads	Ambiguous Reads	
All Fragments	610,820 (61.1%)	90,017 (9.0%)	299,606 (29.9%)	
Unique Fragments	42,581 (60.1%)	6,052 (8.5%)	22,233 (31.4%)	
Average Unique Fragments per GSP2	127.84	18.17	66.77	
Unique Start Sites	9,578 (60.4%)	1,607 (10.1%)	4,800 (30.2%)	
Average Unique Start Sites per GSP2	29.2	4.93	15.29	
Average Unique Start Sites per GSP2 Control	37.42	10.5	4.25	
<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
44.5	10.5

## Miscellaneous Statistics

On Target Deduplication Ratio
14.12:1

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

## DNA/RNA Fragment Lengths

DNA Fragment Median Length (bp)	DNA Fragment Mean Length (bp)	RNA Fragment Median Length (bp)	RNA Fragment Mean Length (bp)
85.0	91.5	87.0	90.3

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



**Tip:** For your benefit, most of the metrics contain tool tips (mouse-over context sensitive Help).

## 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:

Assay Targets					
Unique Fragments	Unique Start Sites	DNA	RNA	Ambiguous	Total
Target		Fragments	RNA Fragments	DNA Fragments	Ambiguous Fragments
ALK_chr2_29430108_25+_A1_GSP2		206 (0.291%)	4 (0.066%)	202 (0.474%)	0 (0.000%)
ALK_chr2_29432692_21-_A1_GSP2		198 (0.279%)	0 (0.000%)	109 (0.256%)	89 (0.400%)
ALK_chr2_29443578_18+_A1_GSP2		258 (0.364%)	2 (0.033%)	35 (0.082%)	221 (0.994%)
ALK_chr2_29443634_21-_A1_GSP2		140 (0.198%)	0 (0.000%)	41 (0.096%)	99 (0.445%)
ALK_chr2_29443665_25+_A1_GSP2		242 (0.341%)	2 (0.033%)	203 (0.477%)	37 (0.166%)
ALK_chr2_29445228_26+_A1_GSP2		142 (0.200%)	3 (0.050%)	82 (0.193%)	57 (0.256%)
ALK_chr2_29445431_22+_A1_GSP2		193 (0.272%)	3 (0.050%)	172 (0.404%)	18 (0.081%)
ALK_chr2_29446286_24+_A1_GSP2		642 (0.906%)	0 (0.000%)	0 (0.000%)	642 (2.888%)
ALK_chr2_29446355_20+_A1_GSP2		221 (0.312%)	0 (0.000%)	0 (0.000%)	221 (0.994%)
ALK_chr2_29446432_22+_A1_GSP2		372 (0.525%)	22 (0.364%)	0 (0.000%)	350 (1.574%)
ALK_chr2_29446473_24+_A1_GSP2		286 (0.404%)	23 (0.380%)	0 (0.000%)	263 (1.183%)
ALK_chr2_29448389_20+_A1_GSP2		139 (0.196%)	1 (0.017%)	109 (0.256%)	29 (0.130%)
ALK_chr2_29449887_37+_A1_GSP2		39 (0.055%)	0 (0.000%)	22 (0.052%)	17 (0.076%)
ALK_chr2_29450497_24+_A1_GSP2		134 (0.189%)	6 (0.099%)	113 (0.265%)	15 (0.067%)
ALK_chr2_29451882_22+_A1_GSP2		157 (0.222%)	8 (0.132%)	102 (0.240%)	47 (0.211%)
ALK_chr2_29456485_24+_A1_GSP2		157 (0.222%)	13 (0.215%)	39 (0.092%)	105 (0.472%)
ALK_chr2_29474058_21+_A1_GSP2		77 (0.109%)	0 (0.000%)	45 (0.106%)	32 (0.144%)
ALK_chr2_29498337_19+_A1_GSP2		266 (0.375%)	9 (0.149%)	253 (0.594%)	4 (0.018%)

The *Assay Targets* page is itself divided into seven different sections. The top section of the page is dedicated to Controls (for RNA panels only, see "Control Genes (FusionPlex Only)" on page 82 to learn more). The bottom section is for the actual Assay Targets. Across all sections described here:

- **Unique Molecular Bins** – defined as the total unique reads (i.e., de-duplicated reads). Data is further sorted into Fragments, RNA Fragments, DNA Fragments, and Ambiguous Fragments (see "RNA vs. DNA vs. Ambiguous Reads" on page 81)
- **Unique Start Sites** – defined as the number of start sites occupied by at least one read. Data is further sorted into Fragments, RNA Fragments, DNA Fragments, and Ambiguous Fragments (see "RNA vs. DNA vs. Ambiguous Reads" on page 81)
- **DNA** – represents the data for DNA reads only and further separated by DNA reads that have a unique molecular bin and those DNA reads that have unique start sites
- **RNA** – represents the data for RNA reads only and further separated by normalized expression, RNA reads that have a unique molecular bin, and those RNA reads that have unique start sites
- **Ambiguous** – represents the data for ambiguous reads only and further separated by ambiguous reads that have a unique molecular bin and those ambiguous reads that have unique start sites
- **Total** – represents the data for all reads combined (RNA vs. DNA vs. Ambiguous Reads) and further separated by reads that have a unique molecular bin and those reads that have unique start sites

## Pipeline-specific links

The additional links found in the "Sample Details page" on page 96 show information for each specific pipeline selected at job setup.

## Pipeline-specific information

### Fusions (RNA only)

The RNA Fusion pipeline will detect gene fusion events by annotating the de-novo assembled RNA reads with BLAST. Archer-specific filtering logic is used to reduce misalignments and false positives.

The Fusion Grid displays results of candidates of any Fusions or Isoforms from this pipeline:

The screenshot shows the Fusion Grid interface. At the top, there are tabs for "Strong Fusions & Oncogenic Isoforms", "Duplications of Interest", "Low Confidence Fusions", and "All Results". Below the tabs is a table with columns: Actions, Classification, Report, Artifact, Genes, SS, Reads, %Reads, Break, Cat, Type, InFrame, TO, Rept, Art, Known Partners, Strong, User-Annotated True Positive, and User-Annotated False. The first row shows a fusion event: WSCD1 → BRAF, with 3 start sites, 14 reads, and 26.92% of reads. Below the table, there are two visualizations. The first is labeled "Aggregate Variant - 1 Component Variant in total" and shows a genomic map with exons 8 and 9. The second is labeled "1st Component Variant - Strongest" and shows a similar genomic map. Both visualizations highlight the fusion breakpoint between exon 9 of WSCD1 and exon 8 of BRAF.

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 primers that support the fusion.



**Tip:** 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. In order to call a fusion, the assembled read or contig must extend at least 20bp into the fusion partner. For reads to be considered supporting the fusion they must span the fusion breakpoint by 5bp on either side of the breakpoint.
- **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.





**Note:** 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 Archer Analysis settings” on page 37.
2. **Presence of fusion in Quiver** – If the fusion is found in the database of known fusions, called [Quiver](#), 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.



**Important!** 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 reads** – Percent reads is the proportion of breakpoint spanning reads that support the candidate relative to the total number of RNA reads spanning the breakpoint. Percent reads need 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 reads 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.

Change the “Min Unique Start Sites for Valid Structural Variation” parameter under “Structural Variation Settings” to adjust this cutoff.









**Note:** 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 below 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`

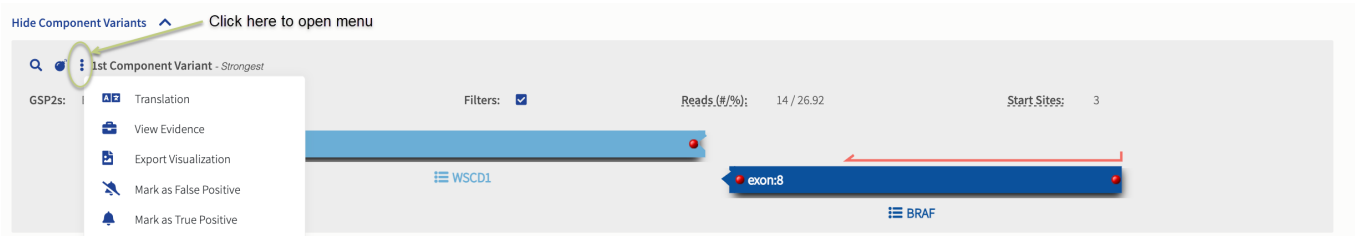
under the “Structural Variation Analysis Settings.”

- 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

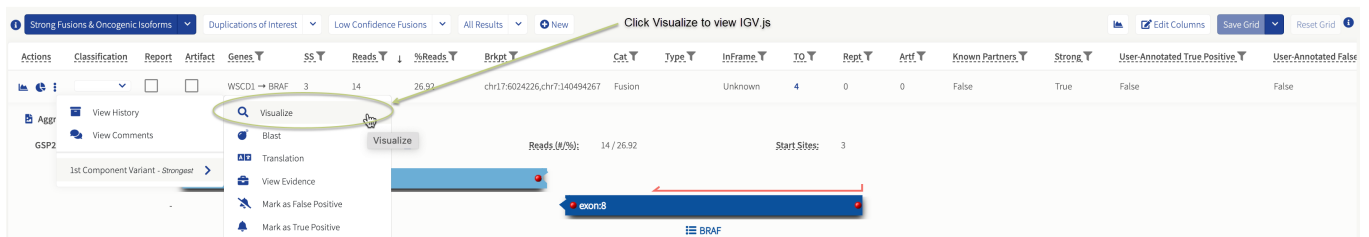
It is possible to override the positive and negative evidence criteria by annotating specific fusions (according to breakpoint). Such annotation will be recorded and used in a manner similar to Quiver entries, and will trump all other criteria:



The screenshot shows a web interface for fusion analysis. A dropdown menu is open for the '1st Component Variant - Strongest' entry. The menu options are: Translation, View Evidence, Export Visualization, Mark as False Positive, and Mark as True Positive. In the background, a genomic track shows a fusion between WSCD1 and BRAF, with a red arrow indicating the fusion junction. The BRAF gene is shown with its exon 8 highlighted in blue. The interface also displays 'Reads (#/%)' as 14 / 26.92 and 'Start Sites' as 3.

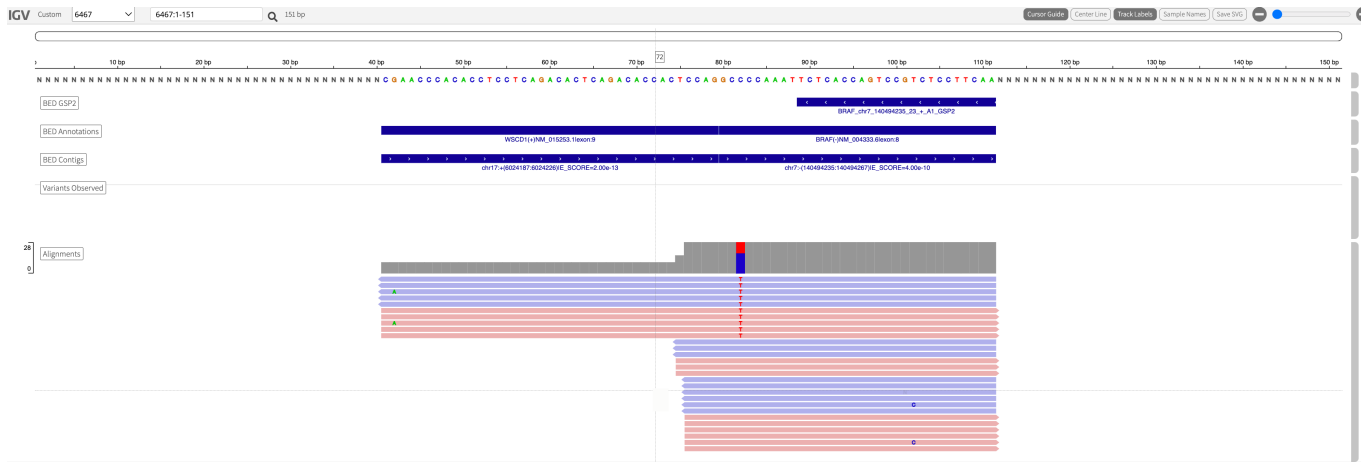
## Visualization of fusion candidate-supporting reads in IGV.js

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



The screenshot shows a table of fusion candidates. The table has columns for Actions, Classification, Report, Artifact, Genes, SS, Reads, %Reads, Brkpt, Cat, Type, InFrame, IO, Rept, Artf, Known Partners, Strong, User-Annotated True Positive, and User-Annotated False. The first row shows a fusion between WSCD1 and BRAF with 3 reads and 26.92% of reads. A 'Visualize' button is highlighted in the table. Below the table, a genomic track shows the fusion between WSCD1 and BRAF, with a red arrow indicating the fusion junction. The BRAF gene is shown with its exon 8 highlighted in blue. The interface also displays 'Reads (#/%)' as 14 / 26.92 and 'Start Sites' as 3.

The IGV.js genome browser opens 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.  
**Important!** 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  box provides a new window with the prediction results and a link to the BLAST application.

<u>In Frame</u>	<u>5' Gene</u>	<u>5' Transcripts</u>	<u>3' Gene</u>	<u>3' Transcripts</u>
Unknown	WSCD1	XM_005256573.1 XM_005256572.2 NM_015253.1 XM_024450682.1	BRAF	XM_017012558.1 NM_001354609.2 NM_004333.6 XM_017012559.1
<a href="#">BLAST Sequence</a> <a href="#">Query Pfam</a> MAKPFRLQKFLRRTQFLFFLTAAYLMTGSLLLQVRVALPQGPRAPGPLQTLPVAAVALGVLLDSRALHDPVSPPELLLGVDMLQSPLTRPRPGPRWLRSRN:				
Unknown	WSCD1	XM_011523777.2	BRAF	XM_017012558.1 NM_001354609.2 NM_004333.6 XM_017012559.1
<a href="#">BLAST Sequence</a> <a href="#">Query Pfam</a> MAKPFRLQKFLRRTQFLFFLTAAYLMTGSLLLQVRVALPQGPRAPGPLQTLPVAAVALGVLLDSRALHDPVSPPELLLGVDMLQSPLTRPRPGPRWLRSRN:				

Close

## 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.

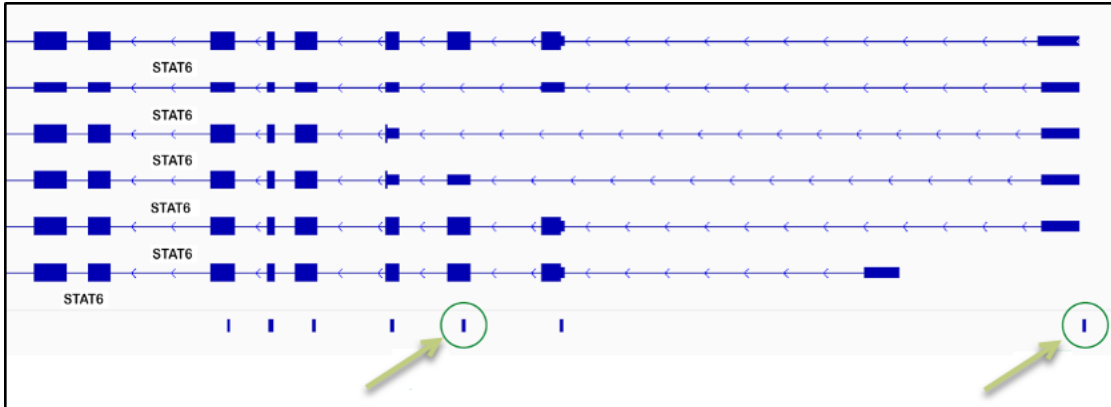
The normalized per GSP2 values from this calculation are written out to the `rna_expression_visualization.tsv` data frame. These values are log transformed (`logshift_rna_expression_visualization.tsv`) and binned to generate the heatmap visualizations.



## 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.

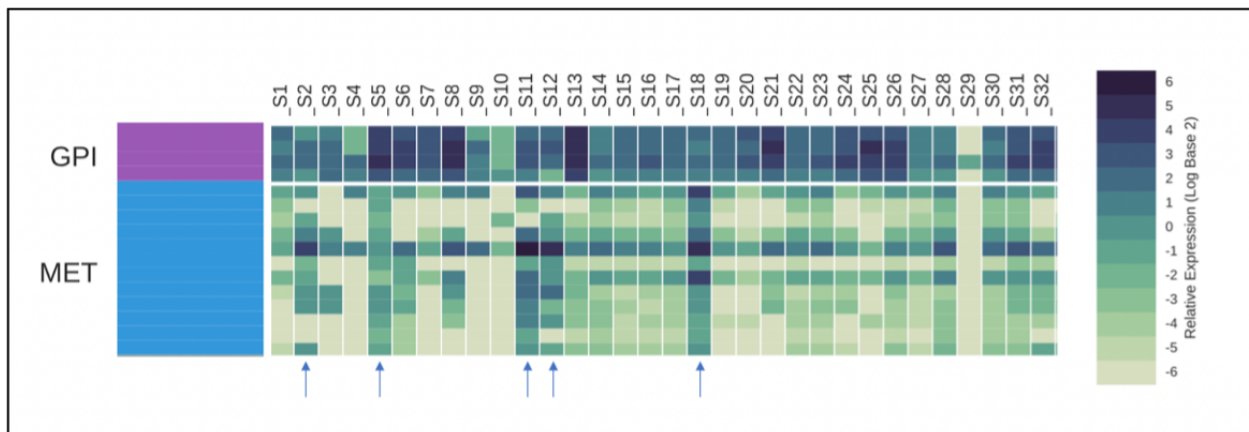


**Note:** This screenshot shows all 5 STAT6 RefSeq isoforms. In green circles are two primers that target a subset of the isoforms. The primer in the second green circle (on the right) will always generate 0 (or very few) RNA reads because it amplifies into the 5' UTR (its amplicon does not cross an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

## Qualitative analysis

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)





**Note:** This Heatmap represents the Log2 transformed, normalized RNA expression values.

## Quantitative analyses

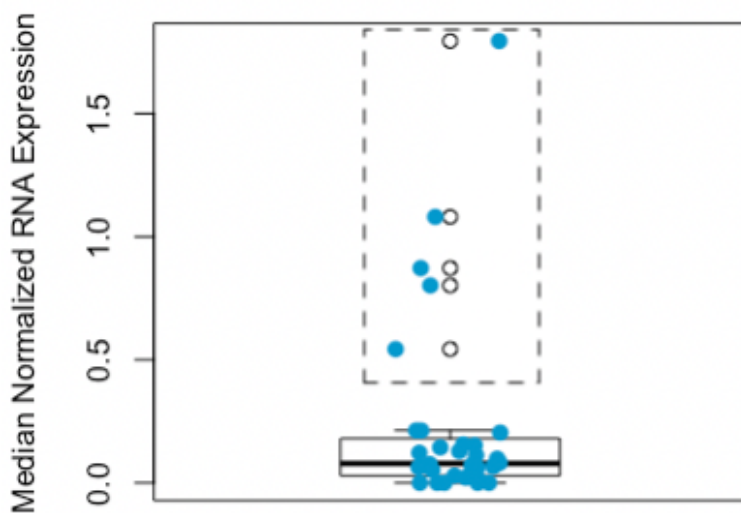
Customers and internal teams exploring custom RNA expression analysis pipelines often begin with the `rna_expression_visualization.tsv` data frame. This output contains the normalized per GSP2 expression values for each primer, and each sample.

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
MET	MET_chr7_116339157_20_-_A1_GSP2	0.00316822	1.41217073	2.67248273	2.07484592	3.43656349	0.35385702
MET	MET_chr7_116340225_29+_A1_GSP2	0	0.21497343	0.55780284	0.31993089	0.81179537	0.01538509
MET	MET_chr7_116380008_24_-_A1_GSP2	0	0.29635623	1.08463928	0.42343794	1.80854276	0.01153882
MET	MET_chr7_116380910_24_-_A1_GSP2	0	0.63570716	1.6677042	0.95172719	2.68568046	0.11154189
MET	MET_chr7_116395414_23_-_A1_GSP2	0.00950467	1.51095614	2.91939911	1.98478135	3.97812957	0.66540504
MET	MET_chr7_116411556_24_-_A1_GSP2	0	0.07114597	0.27014119	0.09409732	0.40460561	0.04615526
MET	MET_chr7_116411645_28+_A1_GSP2	0	0.29123782	1.06671136	0.56794454	1.09088464	0.11923443
MET	MET_chr7_116411907_25_-_A1_GSP2	0	0.07984728	0.58184256	0.11090041	0.7748788	0
MET	MET_chr7_116414939_29_-_A1_GSP2	0.00316822	0.19296425	0.72526593	0.31657027	0.79813624	0.01538509
MET	MET_chr7_116417450_29_-_A1_GSP2	0	0.19091688	1.04878344	0.31858664	0.86975439	0.01153882
MET	MET_chr7_116418834_27_-_A1_GSP2	0	0.08240648	0.40582296	0.14316235	0.32449665	0.01538509
MET	MET_chr7_116435945_25_-_A1_GSP2	0	0.28407204	1.28103152	0.52627287	1.49106026	0.03846272
		Sample_1	Sample_2	Sample_3	Sample_4	Sample_5	Sample_6
	MET_MEDIAN	0	0.24952274	1.0577474	0.37168442	0.98031951	0.0269239

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 ( $=\min(\max(x), Q_3 + 1.5 * IQR)$ ).



- Samples that display expression above this threshold shown in the dashed-line box above (and indicated with arrows in the Heatmap image in Qualitative analysis) display RNA expression consistent with MET amplification (in this example) and were investigated further.



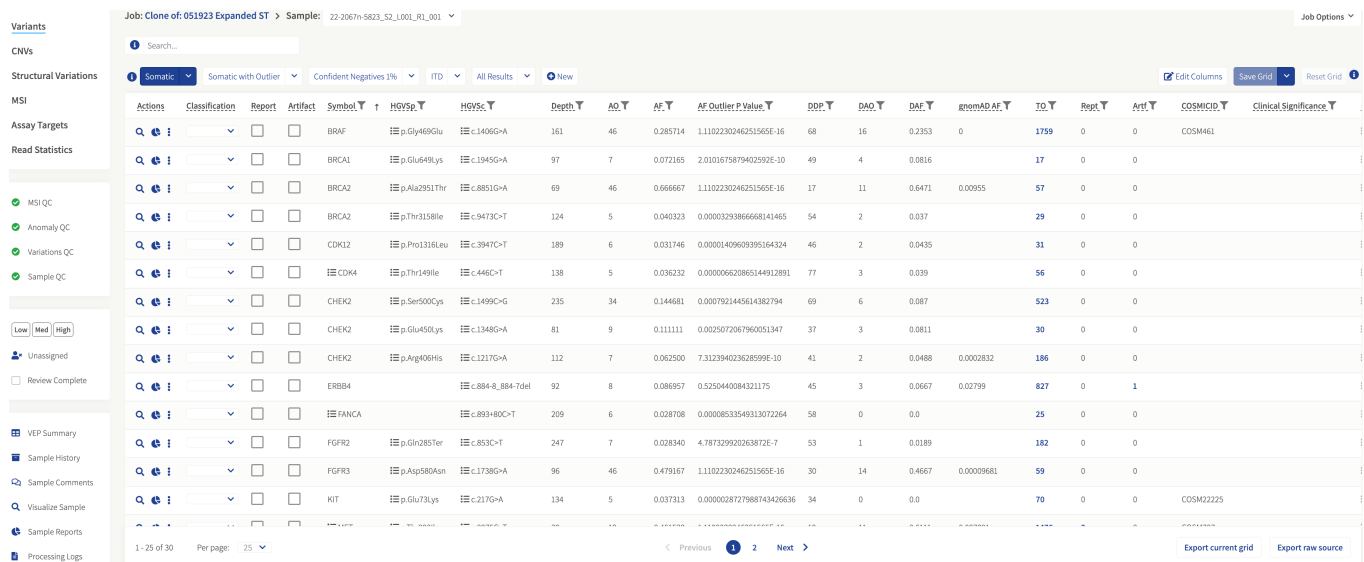
**Note:** Comparator data was not used to refine the threshold for this example set; however, this will likely be required to balance sensitivity/specificity for expression-based analyses.

## 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

Analysis can be configured to detect targeted variants of interest, de novo variants, or both. To detect targeted variants, a Targeted Mutation File (TMF) containing a list of the variants of interest must be selected when setting up the new job. To also include de novo variant calling, select “Include Non-targeted variants” = YES. With no TMF, de novo variants will be detected by default. See *Starting a Job: New Job Page* for more information. If a TMF is used, variants found in the TMF will be identified by green font; de novo variants will be in black font. Results for all variants in the TMF that fall within the Region of Interest (ROI) of the panel can be found in the All Results tab. Those that meet the criteria of a filter set will also be included in that filter set’s tab.



Actions	Classification	Report	Artifact	Symbol	HGVS	Depth	AF	AF.T	AF.Outlier.P.Value	SDP	PAF	DAF	AF.MAD.AF	TO	RefL	Arst	COSMIC	Clinical Significance
				BRAF	p.Gly469Glu	161	46	0.285714	1.1102230246251565E-16	68	16	0.2353	0	1759	0	0	COSM461	
				BRCA1	p.Glu649Lys	97	7	0.072165	2.0101675879402592E-10	49	4	0.0816		17	0	0		
				BRCA2	p.Ala295LThr	69	46	0.666667	1.1102230246251565E-16	17	11	0.6471	0.00955	57	0	0		
				BRCA2	p Thr318Ile	124	5	0.040323	0.00003293866668141465	54	2	0.037		29	0	0		
				CDK12	p.Pro131Leu	189	6	0.031746	0.00001409609395164324	46	2	0.0435		31	0	0		
				COX4	p Thr149Ile	138	5	0.030232	0.000006620865144912891	77	3	0.039		56	0	0		
				CHEK2	p.Ser500Cys	235	34	0.144681	0.0007921445614382794	69	6	0.087		523	0	0		
				CHEK2	p.Glu450Lys	81	9	0.111111	0.0025072067960051347	37	3	0.0811		30	0	0		
				CHEK2	p.Arg406His	112	7	0.062500	7.312394023628599E-10	41	2	0.0488	0.0002832	186	0	0		
				ERBB4	c.884_8_884-7del	92	8	0.080957	0.5250440084521175	45	3	0.0667	0.02799	827	0	1		
				FANCA	c.893+80C>T	209	6	0.028708	0.00008533549313072264	58	0	0.0		25	0	0		
				FGFR2	p.Gln285Ter	247	7	0.028340	4.787329920263872E-7	53	1	0.0189		182	0	0		
				FGFR3	p.Asp580Asn	96	46	0.479167	1.1102230246251565E-16	30	14	0.4667	0.00009681	59	0	0		
				KIT	p.Glu731Lys	134	5	0.037313	0.0000028727988743426636	34	0	0.0		70	0	0	COSM22225	

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://ensembl.org). The default variant transcript annotation is applied by sorting the VEP annotations according to the order of the Archer Analysis Gene Transcripts list, without consideration for the exact version numbers at the ends of the transcripts.

## 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 Details page (Sample Summary)” on page 93 will list the results of the CNV analysis by highlighting the genes that show significant copy gains or losses:

[12428] 22-2067n-5823\_S2\_L001\_R1\_001

[8470] Clone of: 051923 Expanded ST

Pipeline(s):  
 SNP/InDel  
 CNV  
 Structural Variation  
 MSI

Input Material: DNA

Mutation Detection Type: Somatic Mutation

Instrument: NovaSeq

Panel: VariantPlex Expanded Solid Tumor 20123 v1.2

Status: COMPLETED\_OKAY

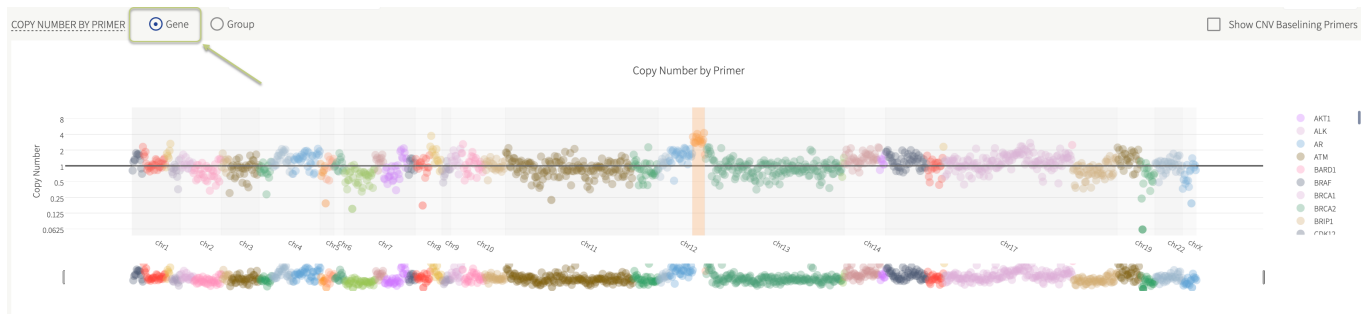
Legend:

- 30 Variants
- 8 DNA Structural Variations
- 1 Copy Gains
- 6 Copy Losses
- MS-Stable MSI
- MSI QC
- Anomaly QC
- Variations QC
- Sample QC

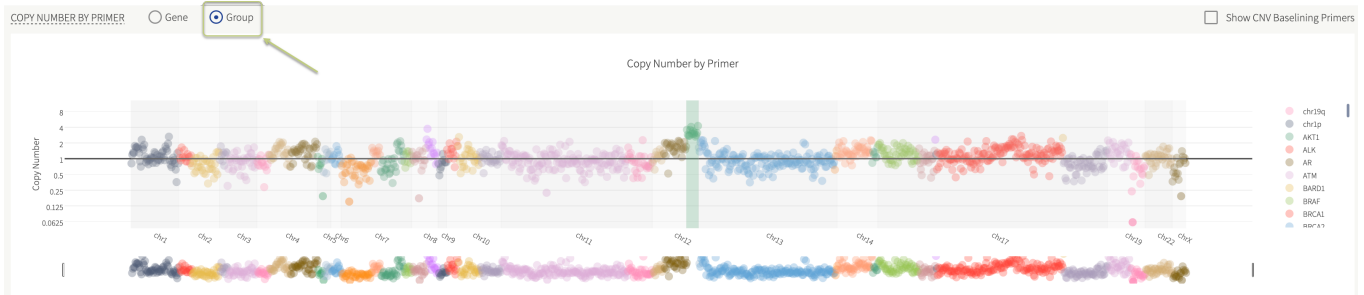
## CNV gene/group plots



**Note:** 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.



**Tip:** 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.

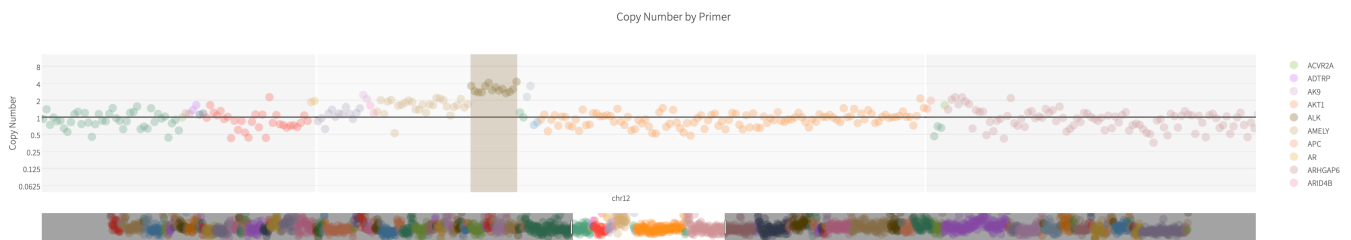


The CNV grid contains the summarized CNV results for both genes and groups regardless of the toggle for the plot.

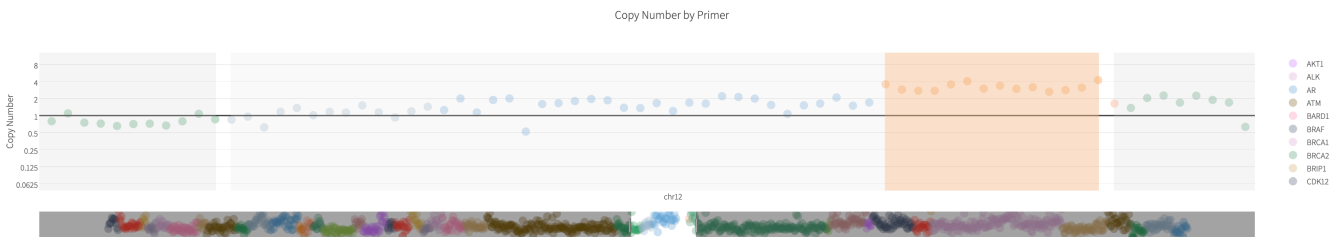
### Showing CNV baselining primers

For the primer-by-primer copy number plot there is an additional "Show CNV Baselining 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 Baselining Primers" option.

Whether 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 Baselining 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" on page 48 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.

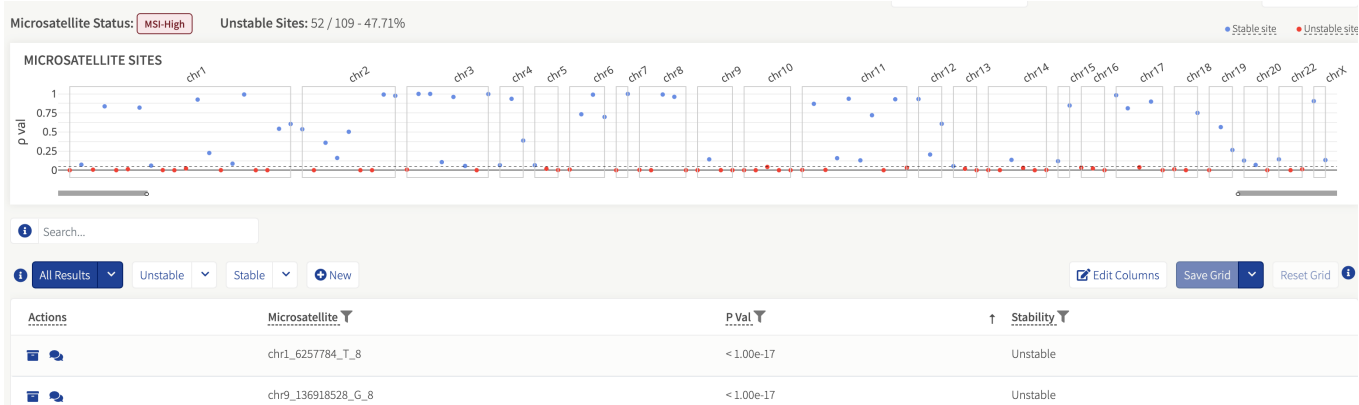
The *Job Detail* page will display the MSI call highlighted in green (MS-Stable), red (MSI-High), yellow (MSI-Intermediate), and the MSI specific QC status.

<p>⋮ GAB367_S1_R1_001</p> <p>Low Med High</p> <p>👤 Unassigned</p> <p><input type="checkbox"/> Review Complete</p>	<p>364 Variants</p> <p>MS-Stable MSI</p>	<p>✔ Sample QC</p> <p>✔ Variations QC</p> <p>✔ MSI QC</p>
<p>⋮ GAB368_S2_R1_001</p> <p>Low Med High</p> <p>👤 Unassigned</p> <p><input type="checkbox"/> Review Complete</p>	<p>437 Variants</p> <p>MSI-High MSI</p>	<p>✔ Sample QC</p> <p>✔ Variations QC</p> <p>✔ MSI QC</p>
<p>⋮ GAB369_S3_R1_001</p> <p>Low Med High</p> <p>👤 Unassigned</p> <p><input type="checkbox"/> Review Complete</p>	<p>417 Variants</p> <p>MS-Stable MSI</p>	<p>✔ Sample QC</p> <p>✔ Variations QC</p> <p>✔ MSI QC</p>

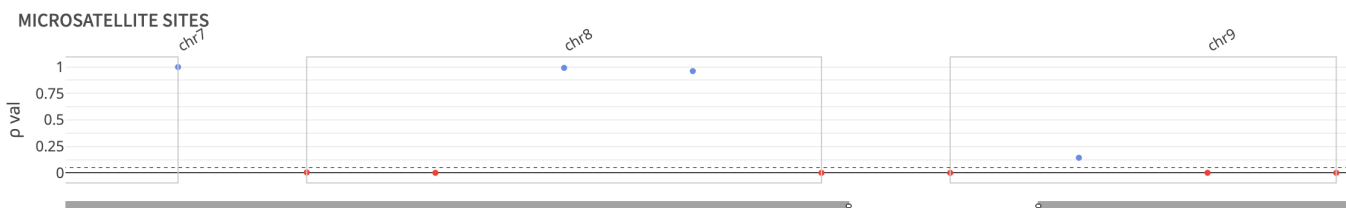
## MSI plots and grid

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.



**Tip:** You can zoom in on the MSI loci grid for enhanced viewing.



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

## 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 "MSI pipeline settings" on page 48 for more detail 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

The "Outlier Detection" algorithm 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 through 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.



**Important!** 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" on page 37 page. Scroll down to the "Variations-Mutations settings" on page 44 section where the toggle for **OUTLIER\_DETECTION** can be found.

Variations/Mutations Settings	
Min Depth for Variant Call	4
MAPQ Threshold for Variant Call	0
Min Basequal for Variant Call	20
Min Allele Fraction for Variant Call	0.001
Min Phred Qual Score for Variant Call	1
Variant Downstream ROI Size	400
Outlier Detection	<div style="border: 1px solid #ccc; padding: 2px;"> <input checked="" type="checkbox"/> On             <input type="checkbox"/> Off         </div>
Min DAO for Outlier Detection Calculation	0
Deep Shallow Threshold	5
VEP Report Everything	Off





**Tip:** 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" on page 37.

## 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	AF Outlier P Value	DDP
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ASXL1	p.Trp583Ter	c.1748G>A	2716	1213	0.446613	1.1102230246251565E-16	225
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	BCORL1	p.Thr371Pro	c.1111A>C	1092	36	0.032967	0.0003325689326165726	642
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	BCORL1	p.Leu372Pro	c.1115T>C	968	129	0.133264	1.1102230246251565E-16	558
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	BCORL1	p.Ala375Pro	c.1123G>C	920	91	0.098913	1.1102230246251565E-16	529
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CEBPA	p.Pro97ValfsTer64	c.289_290delinsGTGG	2068	1639	0.792553	1.1102230246251565E-16	942
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	EZH2	p.Gln331Ter	c.991C>T	1078	911	0.845083	1.1102230246251565E-16	4



**Note:** Outlier Detection will not be beneficial to run on Germline panels. To save analysis runtime, manually disable the feature in "User settings" on page 37.



**Tip:** 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.

Sample Name	Assay Results	Fusion Results	QC Results	Reportable Mutations
ABC_S4_L001_R1_001 [Low] [Med] [High] * Unassigned <input type="checkbox"/> Review Complete	0 Fusions ABC Lymphoma 0 Variants	Unnamed Criteria No events found.	<input checked="" type="checkbox"/> Fusion QC <input checked="" type="checkbox"/> Variations QC <input checked="" type="checkbox"/> Sample QC	

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” on page 96.

Fusions

Variants

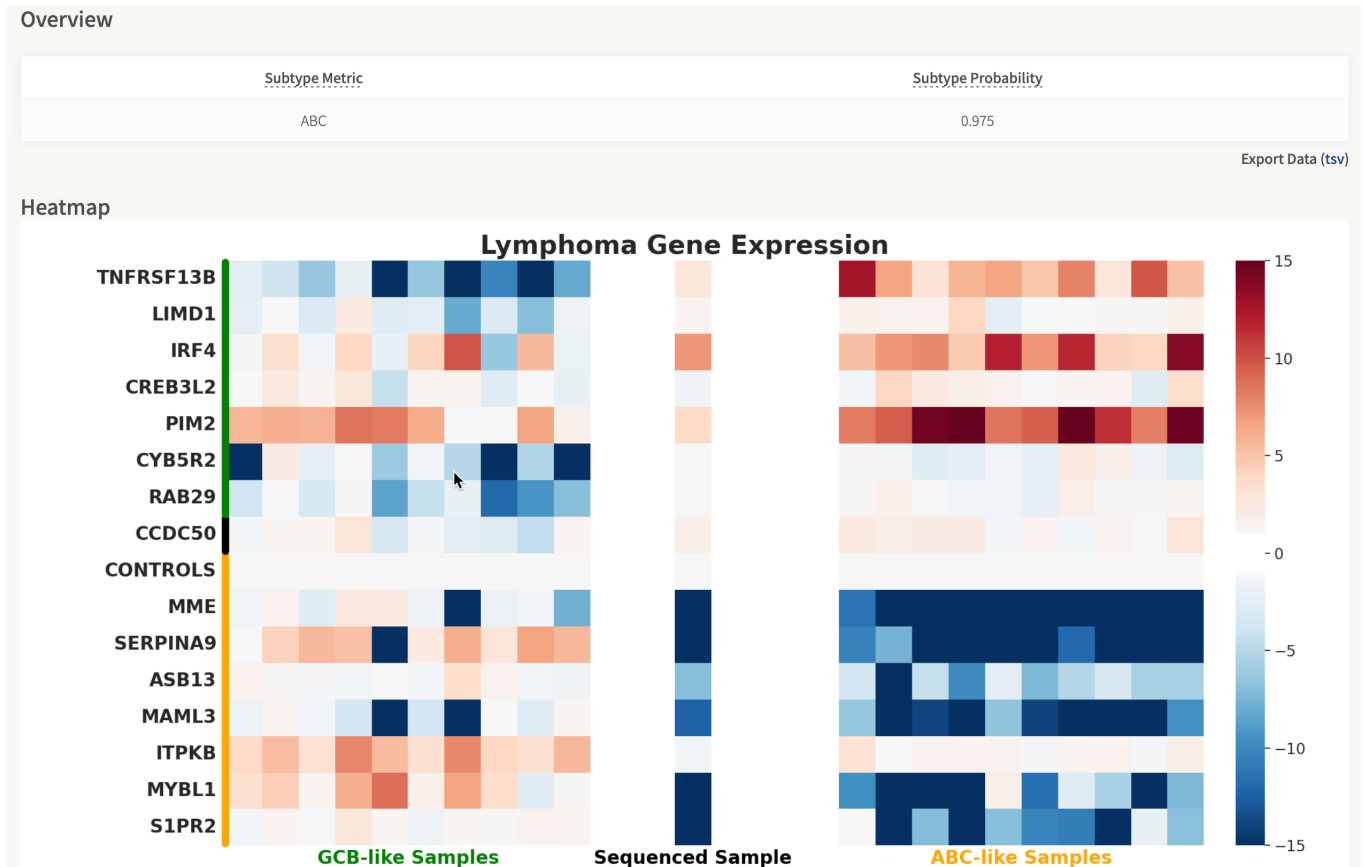
Lymphoma

Assay Targets

Read Statistics

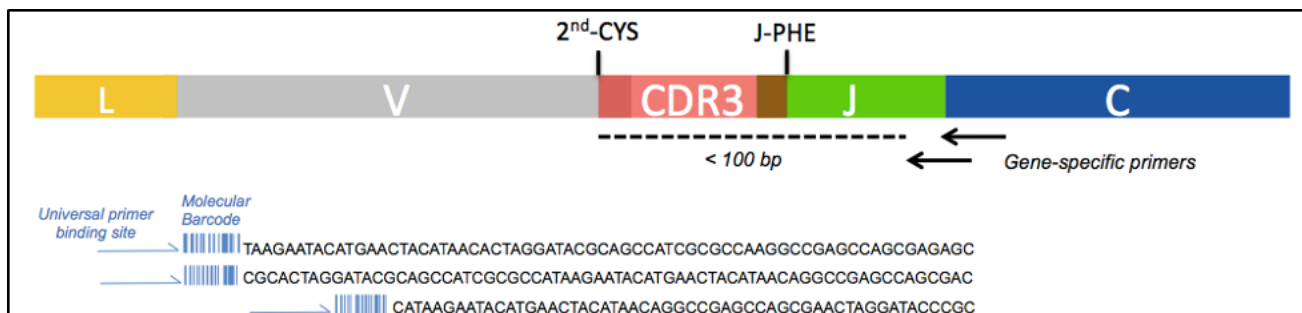
Lymphoma COO Metrics:

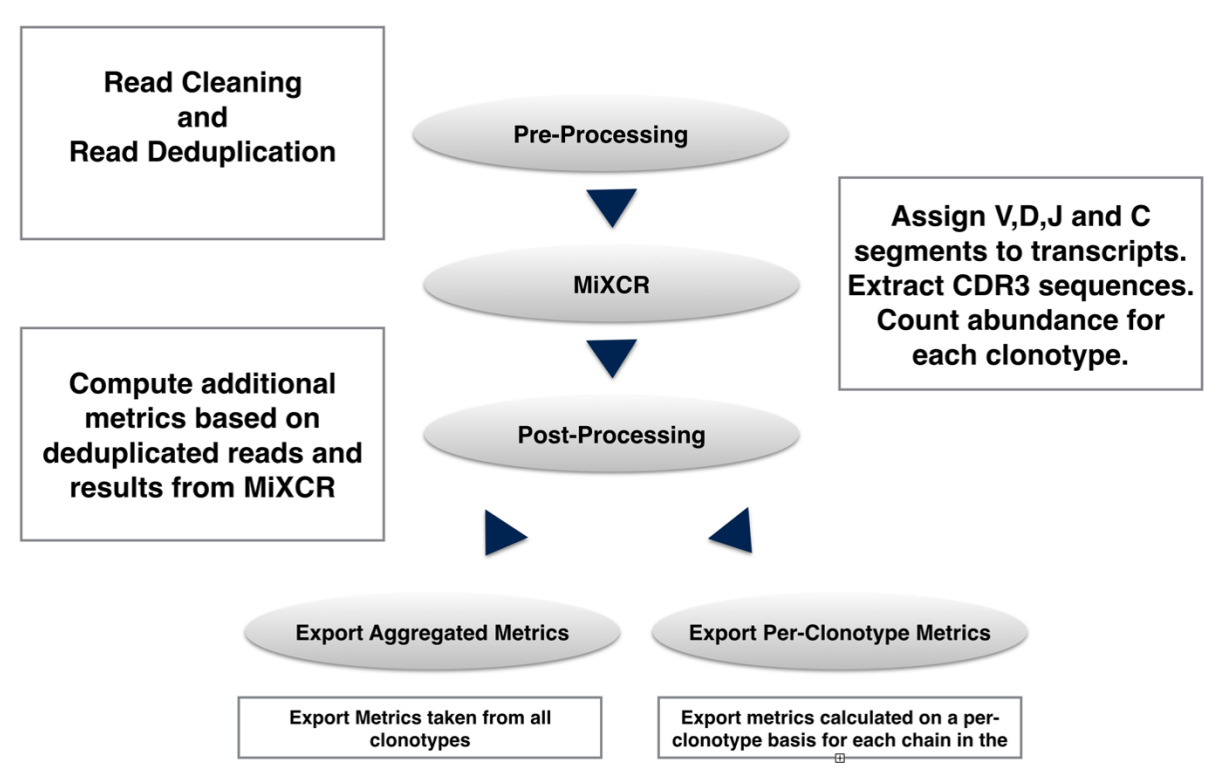
- **Subtype Metric** – The reported lymphoma cell-of-origin subtype (GCB, ABC or Unclassified).
- **Subtype Probability** – Estimated probability of subtype classification.
- **Lymphoma Gene Expression Heat map** – This is a heat map of the gene expression considered in the COO algorithm. The results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



## 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 Immunoverse 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 page (Sample Summary)" on page 93 and "Sample Details page" on page 96:

View summary metrics for all samples in the job

Job: Clone of: Clone of: Remote Clone of: Immuniverse TCR B/G Demo Data

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

Low Med High All

Download all metrics for job via modal

Sample Name	Assay Results	QC Results
NM3813_S8_R1_001 Low Med High Unassigned <input type="checkbox"/> Review Complete	0 TRA Chain 246 TRB Chain 4 TRG Chain 0 TRD Chain	Sample QC
SM1347_S36_R1_001 Low Med High Unassigned <input type="checkbox"/> Review Complete	2 TRA Chain 65149 TRB Chain 4306 TRG Chain 0 TRD Chain	Sample QC

View metrics per sample (Sample Details page)

Total clonotypes per chain

Job Options

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

## 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:

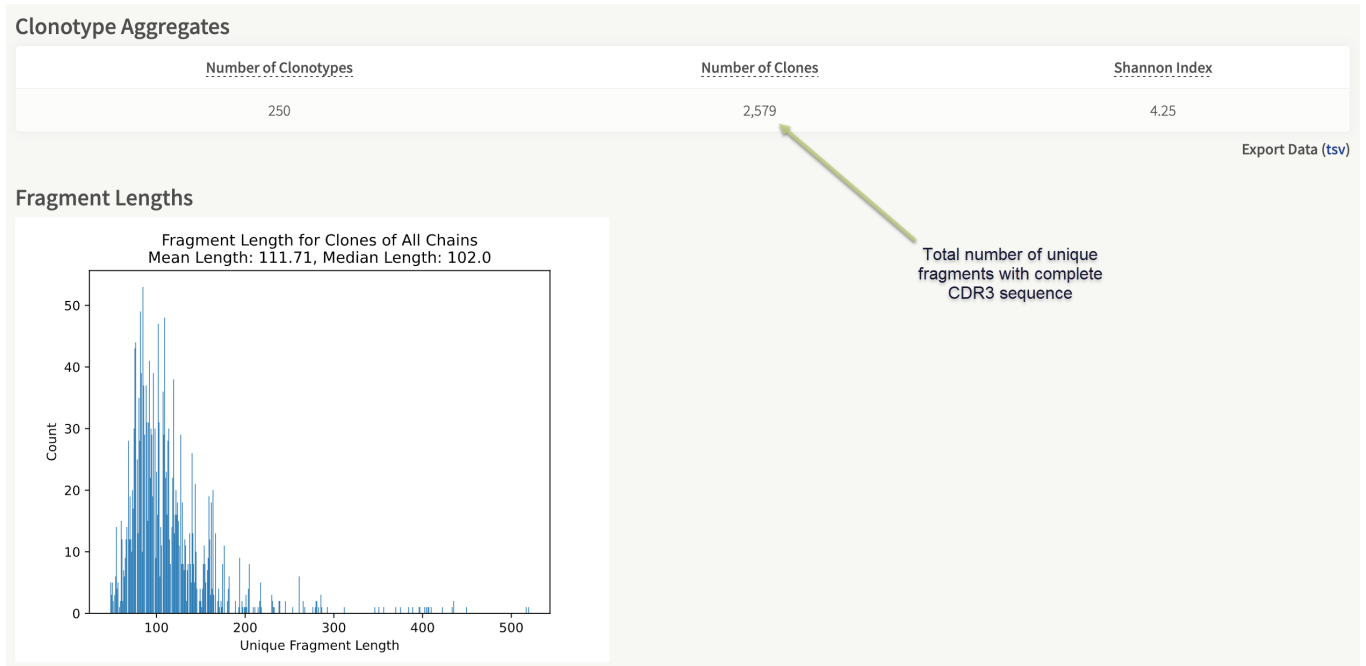
- **Read Statistics** – overviews key metrics before and after processing read data
- **Clonotypes** – provides an in depth look at clones and clonotypes

[Sample Details page \(Read Statistics tab\)](#)

The top sections of the **Read Statistics** tab provide metrics similar to other Analysis pipelines.



**Tip:** More unique information for this pipeline is provided in the 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).



**Tip:** Besides 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:

Job: Clone of: Clone of: Remote Clone of: Immunoverse TCR B/G Demo Data > Sample: NM3813\_SB\_R1\_001 Job Options

TRA TRB TRG TRD Metrics are reported on a per chain basis

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

Search...

Default Hypermutation Default + New Filtering options Edit Columns Save Grid Reset Grid

Actions	CDR3 Sequence	Clone Frequency	Clone Abundance	Raw Fragment Abundance	Predicted V Region	Predicted D Region
	TGCAGCGTCTGGACAGGGCTAAGTTCTTC	0.15525291828793775	399	97610	TRBV29-1	TRBD1
	TGTGCCAGCAGATGGACTCATGCAAGCACAGATACGGCAGTATTTT	0.07354085603112841	189	25603	TRBV6-5	TRBD2
	TGTGCCAGCAGCTTCAGGAGGGTGGCAATGAGCAGTTCTTC	0.0688715953307393	177	38000	TRBV5-6	TRBD2
	TGTGCCAGCAGCATACTAGCGGGCAGACCAAGAGACCCAGTACTTC	0.06653696498054475	171	19255	TRBV2	TRBD2
	TGTGCCAGCAGCCTCTTAGCGGGGACCTAGGGAGCAGTTCTTC	0.04824902723735409	124	23279	TRBV7-8	TRBD2
	TGTGCCAGCAGAACCGGGACAGGGAACCTCTACGAGCAGTACTTC	0.026459143968871595	68	8659	TRBV27	TRBD1
	TGTGCCAGCAGCACACCGGGACAGTACCGGGGGAATCAGCCCCAGCATTTT	0.01867704280155642	48	11037	TRBV21-1	TRBD1
	TGTGCCAGCAGCTTTAGCACCGGGGAGCTGTTTTTT	0.01089494163424245	28	4128	AMBIGUOUS	INDETERMINATE
	TGTGCCAGCAGCTAGTCTAGAACGACTGAAGCTTTCTTT	0.00933852140077821	24	3239	TRBV7-9	INDETERMINATE
	TGTGCCAGCAGCCCCAGTGGACTAGCGGGGGCAAGGATACGCAGTATTTT	0.0089494163424245	23	2023	TRBV9	TRBD2
	TGCGCCAGCAGTAAATCGACAGGGGGCAATGAGCAGTTCTTC	0.007782101167315175	20	5047	TRBV23-1	TRBD1

1 - 25 of 246 Per page: 25 < Previous **1** 2 3 4 5 ... 10 Next > Export current grid Export raw source

## Per sample plots

Here we detail the per-sample plots that can be accessed via the “Clonotypes” section:

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

- Data Summary:** Contains the grid with metrics available on a per-clonotype basis for the selected chain.
- Frequencies per Clonotype:** Contains visual displays of segment usage frequencies of each clonotype:
  - Predicted V-Region Frequency: frequency among all V-regions found per-clonotype.
  - Predicted D-Region Frequency: frequency among all D-regions found per-clonotype.
  - Predicted J-Region Frequency: frequency among all J-regions found per-clonotype.
  - Predicted C-Region Frequency: frequency among all C-regions found per-clonotype.
  - CDR3 Translation Frequency: Frequency of CDR3 translations amongst all clonotypes of the specified chain
  - Productive CDR3 Translation Frequency: Frequency of CDR3 translations amongst all clonotypes of the specified chain of only productive CDR3 translations.
  - CDR3 Sequence Frequency: Frequency of CDR3 sequences amongst all clonotypes of the specified chain
- Frequencies per Clone:** Contains visual displays of segment usage frequencies of each clonotype the plots that can be selected are:
  - Predicted V-Region Frequency: frequency among all V-regions found per-clone.
  - 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 **Job Options > Download Results** on the "Job Details page (Sample Summary)" on page 93. 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



**Note:** 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.



## 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 **Save Grid**) or multiple samples within a job (by clicking on the dropdown and selecting **Save to All Samples**).

Search...

Default Hypermutation Default New

Clone Frequency

Edit Columns Save Grid (Grid Changed) Reset Grid

Actions	CDR3 Sequence	Clone Abundance	Clone Frequency	Raw Fragment Abundance	Predicted V Region	Predicted D Region
	TGCAGCGTCTGGACAGGGTCTAAGTTCTTC	399	0.15525291828793775	97610	TRBV29-1	TRBD1
	TGTGCCAGCAGATGGACTCATGCAAGCACAGATACGCAGTATTTT	189	0.07354085603112841	25603	TRBV6-5	TRBD2
	TGTGCCAGCAGCTTCAGGAGGGGTGCCAATGAGCAGTTCTTC	177	0.0688715953307393	38000	TRBV5-6	TRBD2

Search...

Default Hypermutation Default New

Clone Frequency Clone Abundance Raw Fragment Abundance Predicted V Region Predicted D Region

Edit Columns Save Grid (Grid Changed) Save to All Samples Reset Grid

Actions	CDR3 Sequence	Clone Frequency	Clone Abundance	Raw Fragment Abundance	Predicted V Region	Predicted D Region
	TGCAGCGTCTGGACAGGGTCTAAGTTCTTC	0.15525291828793775	399	97610	TRBV29-1	TRBD1
	TGTGCCAGCAGATGGACTCATGCAAGCACAGATACGCAGTATTTT	0.07354085603112841	189	25603	TRBV6-5	TRBD2

# Normal Data Sets

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.

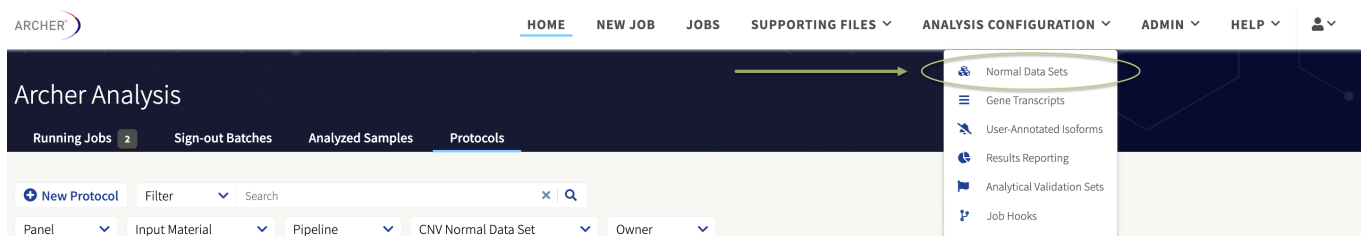
When designating a Normal Data Set of samples, it is critical to ensure those samples do NOT contain the variations to measure in the assay. If they are present, the ability to accurately estimate the background noise of those variations will be diminished — subsequently lowering the sensitivity of those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



**Important!** For SNP/InDel Normal Data Sets, it is vital 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 you want to add/subtract samples from a committed Normal Data Set, you 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.



**Tip:** The Normal Data Set can be upgraded (new for v7.2) by re-running the samples in the Normal Data Set. See the "Deprecating and upgrading Normal Data Sets" on page 128 section for more on how to upgrade.

## Guidelines for 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
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



**Note:** Because it can be difficult to acquire tumor/normal paired samples, it is acceptable to utilize other normal control samples such as Genome in a Bottle ([GIAB](#)) when creating a Normal Data Set. Advantageously, you can reuse the same Normal Data Set in subsequent analyses.

Settings that must be the same are defined in this section. To help you remember what these settings are, they are indicated within "User settings" on page 37 with a Normal Data Set icon.

User 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

## 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**.

## Normal Data Sets

Create Normal Data Set

When creating a Normal Data Set, you 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 you want to create a Normal Data Set for both analysis types, you 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

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.



**Tip:** Samples will be grouped together based on the values of the concordant settings listed above. 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.

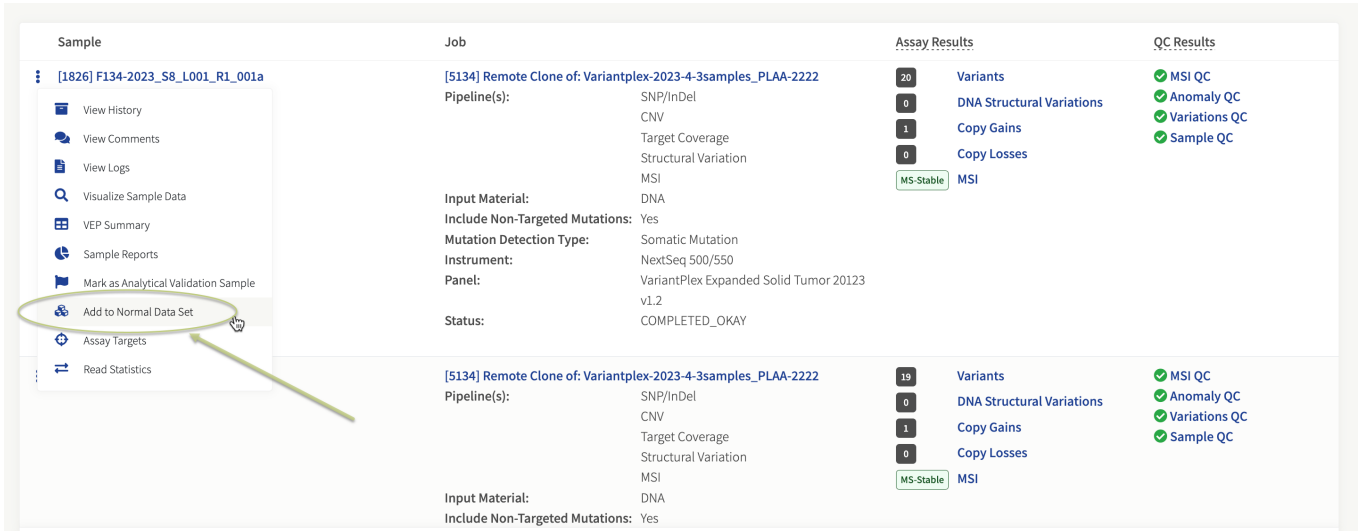
## Adding a Sample to a Normal Data Set from Job Details

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  icon.
2. Click **Add to Normal Data Set**.



Sample	Job	Assay Results	QC Results
[1826] F134-2023_S8_L001_R1_001a	[5134] Remote Clone of: Variantplex-2023-4-3samples_PLAA-2222	20 Variants 0 DNA Structural Variations 1 Copy Gains 0 Copy Losses MSI (MS-Stable)	MSI QC Anomaly QC Variations QC Sample QC
[5134] Remote Clone of: Variantplex-2023-4-3samples_PLAA-2222	[5134] Remote Clone of: Variantplex-2023-4-3samples_PLAA-2222	19 Variants 0 DNA Structural Variations 1 Copy Gains 0 Copy Losses MSI (MS-Stable)	MSI QC Anomaly QC Variations QC Sample QC

## Modal view:

Add Sample to Normal Data Set
✕

---

**Sample:** F134-2023\_S8\_L001\_R1\_001a

**Platform:** Illumina

**Panel:** VariantPlex Expanded Solid Tumor 20123 v1.2

**Normal Data Set Type:**  CNV  SNP/InDel

**Normal Data Set:**  Existing  New

Please select ... ▼

Only Normal Data Sets that are compatible with this sample's platform, panel, pipeline type, and job settings are shown as options.

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:

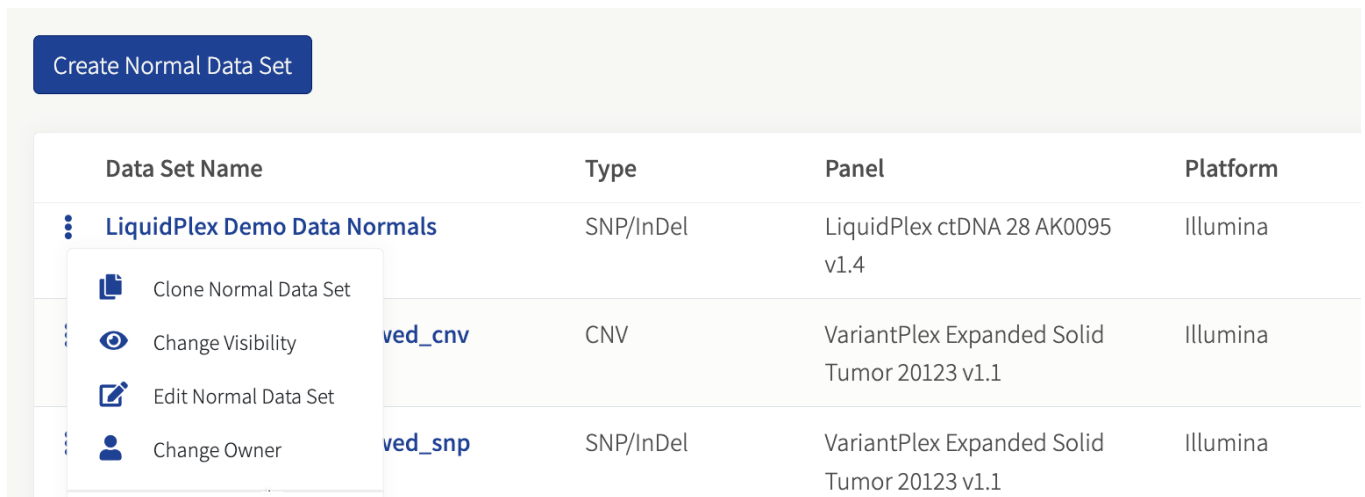
- **No icon** The Normal Data Set is editable. Selecting one of these Normal Data Sets will add the sample to the Normal Data Set and redirect the user back to the job details page giving a successful message.
- **†** Normal Data Set is not editable because it has already been committed. Selecting one of these will clone the Normal Data Set before adding the sample and you 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 you to the *Create Normal Data Set* page with the sample selected. Then, you 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.



Data Set Name	Type	Panel	Platform
LiquidPlex Demo Data Normals	SNP/InDel	LiquidPlex ctDNA 28 AK0095 v1.4	Illumina
ved_cnv	CNV	VariantPlex Expanded Solid Tumor 20123 v1.1	Illumina
ved_snp	SNP/InDel	VariantPlex Expanded Solid Tumor 20123 v1.1	Illumina

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, you 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**

SNP/InDel    Copy Number Variation    Structural Variation    MSI    TMB    Target Coverage

**Option Set B**

Read Processing Only

TARGETED MUTATIONS

Select ▼

Optional

SELECT DNA SNP DETECTION

Somatic Mutation    Germline Variation

VARIANT NORMAL DATA SET

Select ▼

Optional




**Tip:** Normal Data Sets which are not deprecated are selectable; ones with asterisks preceding their names are deprecated and must be upgraded before 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.

## Deprecating 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
 Test ND-CNV  <b>Deprecated</b>	CNV	VariantPlex Comprehensive Solid Tumor v1 18224 v1	Illumina	hel-batal@idtdna.com		

If a Normal Data Set is deprecated, the following are true:

- Any job that uses the Normal Data Set cannot be resubmitted or 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



**Note:** 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.



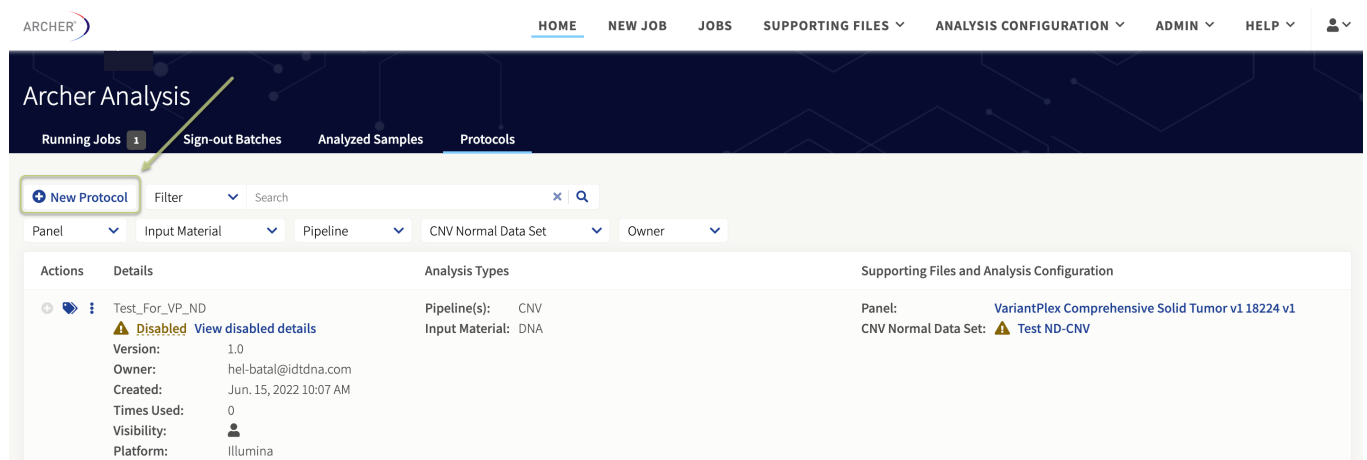
# Workflow automation and integrations

## 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**, then click **New Protocol**.



This will open the **Create New Protocol** dialog (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 any **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

Enter Name

PROTOCOL VERSION

1.0

SELECT PANEL

VariantPlex Myeloid 8401 v1.3

## 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  TMB  Target Coverage

Option Set B

Read Processing Only

TARGETED MUTATIONS

Select

Optional

SELECT DNA SNP DETECTION

Somatic Mutation  Germline Variation

VARIANT NORMAL DATA SET

Select

Optional

CNV NORMAL DATA SET

Select

### 3. Settings



Use My Settings  Use System Default Settings

#### General Analysis Settings

Min Average Unique DNA and AMBIG Start Sites per GSP2

Min Average Unique RNA Start Sites per GSP2 Controls

Error Correction Off  On

Read Depth Normalization

Randomly subsamples all samples to the provided read depth. 0 represents no normalization.

Barcode Length

Common Region Set

Pre-/Post-Alignment Read Cleaner Off  On

Merge Reads in Read Cleaner Off  On

Debug Off  On

Enable Unified VCF Off  On

#### Structural Variant Analysis Settings

Display Intronic Structural Variants Off  On

Min Reads for Valid Structural Variant

### 4. Hooks - OPTIONAL



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

#### JOB PRE HOOKS

Test

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

## Start a new job using an existing protocol

To start a new job using an existing protocol, navigate to **Protocols**→**New job** under the **Actions** column for a saved protocol. Enter a **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

Test\_For\_VP\_SC\_Protocol

ENTER JOB NAME

My Job May 23

#### 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.

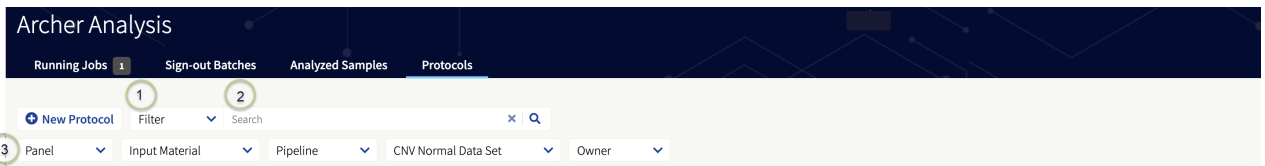
 Drag and drop your FASTQ files here or click to browse.

Submit Job

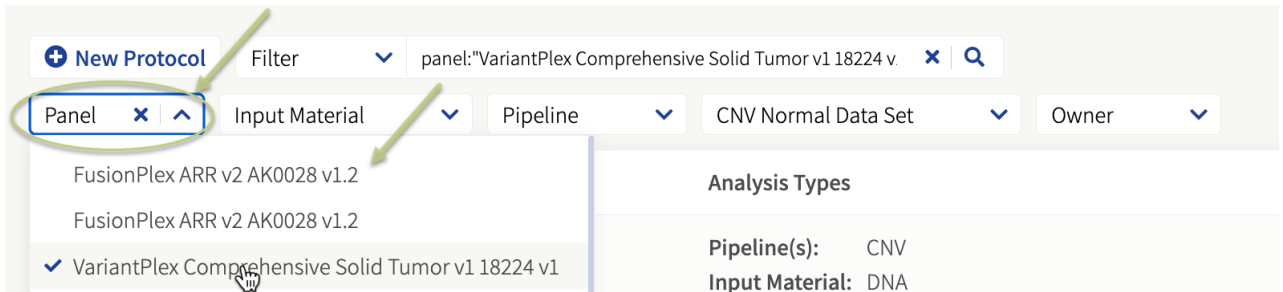
## Search and filter existing protocols

Archer protocols can be searched and filtered as follows:

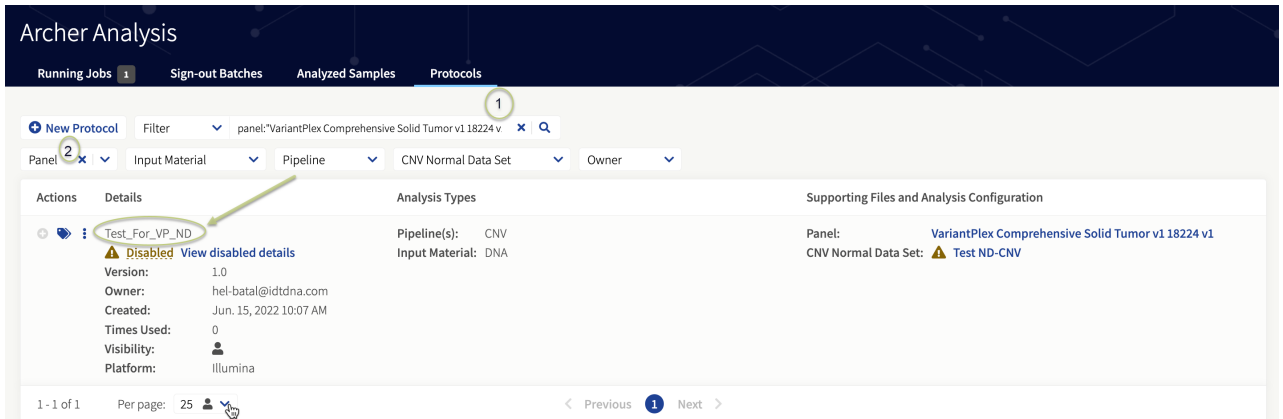
1. **Category Filter** – This dropdown list holds commonly used filters to help streamline a search
2. **Main Search Field** - Type queries here
3. **Category Filters** - Depending on data, one, two, or ten category filters could be available for use. Category filters display horizontally and are clickable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

## 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.

Actions	Details	Analysis Types	Supporting Files and Analysis Configuration
	<b>My Protocol May 23</b> Version: 1.0 Owner: cgeaney@idtdna.com Created: May 23, 2023 6:03 PM Times Used: 0 Watched Folders: <b>Add</b> Visibility: Platform: Illumina	<b>Pipeline(s):</b> SNP/InDel Structural Variation CNV MSI <b>Input Material:</b> DNA <b>Mutation Detection Type:</b> Somatic Mutation <b>Instrument:</b> NovaSeq	<b>Panel:</b> VariantPlex Expanded Solid Tumor 20123 v1.2

## 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**.

### Watched Folders



Protocol Name: My Protocol May 23

**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.

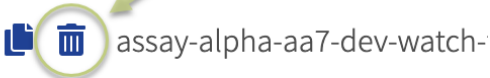
# Watched Folders

---

**Protocol Name:** My Protocol May 23

[Add new watched folder](#)

## EXISTING WATCHED FOLDERS



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 a 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 [Technical Support](#).

## 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.



**Tip:** For single read technologies such as 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**”



**Important!** 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 of 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:** 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
```



**Important!!** Do not remove the files in the “**picked\_up\_files**” folder 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
```



**Note:** Even with a file name collision, the job will still run, and the job will have the same name as the sample however, 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 [Technical Support](#) for additional details.



**Tip:** You can use the REST API for job submission (SSL optionally supported) as an alternative to using Watched Folders. Refer to `<server url>/api/swagger` for REST API documentation or contact [Technical Support](#) for additional questions.

## SSL-Enabled Linked Analysis Servers

Linked Analysis servers allow you 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.

The screenshot shows the Archer Analysis web interface. The top navigation bar includes: ARCHER, HOME, NEW JOB, JOBS, SUPPORTING FILES, ANALYSIS CONFIGURATION, ADMIN, and HELP. The ADMIN dropdown menu is open, showing options: Users, Groups, Report Templates, **Linked Analysis Servers** (circled in green), System Settings, and System Monitor. Below the menu, the 'Linked Analysis Servers' page is displayed, featuring a '+ Add Linked Analysis Server' button (circled in green) and an 'Action:' dropdown menu.

### 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**.

The screenshot shows the Archer Analysis web interface with the 'Completed Jobs' page. The top navigation bar includes: ARCHER, HOME, NEW JOB, **JOBS**, SUPPORTING FILES, ANALYSIS CONFIGURATION, ADMIN, and HELP. The ANALYSIS CONFIGURATION dropdown menu is open, showing options: Normal Data Sets, Gene Transcripts, User-Annotated Isoforms, Results Reporting, Analytical Validation Sets, Job Hooks, and **Show Linked Analysis Servers** (circled in green). Below the menu, the 'Completed Jobs' table is visible, showing a job with ID [5143] DNA\_SV\_SameChrOFF, owner canobile@idtdna.com, and analysis version 7.2.0-rc-9.

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.

## Completed Jobs

Source Local

Filter by keyword...

Filter By Panel




Job	Analysis Types	Job Status	Assay Targets	
[5143] DNA_SV_SameChrOFF	Pipeline(s): Structural Variation	COMPLETED_OKAY	dGSP10668 v1.1	<input type="checkbox"/>

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

# Troubleshooting

## Common problems

This table covers problems and warnings that commonly arise during and immediately after the processing of jobs. For problems occurring during installation of the VM specifically, please consult the *Archer Analysis VM Installation Guide*.

Job status warning	Description	Possible solution
Job finished with a job status of "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.	<p>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]."</p>  <p>A common reason for failure of an individual job is running out of disk space:</p> <pre>[0] 2024-01-15 10:00:00 [INFO] - Checking for required files. [0] 2024-01-15 10:00:00 [INFO] - All files are present. [0] 2024-01-15 10:00:00 [INFO] - All files are present. [0] 2024-01-15 10:00:00 [INFO] - All files are present. [0] 2024-01-15 10:00:00 [INFO] - All files are present. [0] 2024-01-15 10:00:00 [INFO] - All files are present. [0] 2024-01-15 10:00:00 [INFO] - All files are present. [0] 2024-01-15 10:00:00 [INFO] - All files are present. [0] 2024-01-15 10:00:00 [INFO] - All files are present. [0] 2024-01-15 10:00:00 [INFO] - All files are present. [0] 2024-01-15 10:00:00 [INFO] - All files are present.</pre> <p>This may occur if the <code>/var/www/analysis</code> 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 and delete analyses that are no longer needed 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.</p> <p>If the error is not pointing to an obvious solution, please</p>

Job status warning	Description	Possible solution
		<p>send the log files to <a href="#">tech support</a> for further assistance.</p>
<p>Job is stuck with a job status of "NEW"</p>	<p>This indicates the queue manager software ("poller") is not running.</p>	<p>The easiest way to restart the poller is to perform a complete restart of the virtual machine or email <a href="#">tech support</a> when running samples on an Unlimited Instance.</p> <p>If the poller is still not running, check the poller log file (stderr.poller) for more information.</p> <p>If the VM was restarted, it is likely the job will display a "COMPLETED_ERROR" status. Re-run the job by accessing the <b>"three-dot"</b> / <b>"more actions"</b> menu and select <b>"Rerun Job"</b> to restart the job with the same settings.</p>
<p>Job is stuck with a job status of "HOLD"</p>	<p>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.</p>	<p>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 <i>Job is stuck with job status "NEW"</i>.</p>

# Appendix A

## Format of the GTF file

The Target Region or panel file is a GTF file containing the names and locations of the primers. Gene Transfer Format (GTF) is described [here](#).

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:

- Name field containing the name of the GSP2
- 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.

Here is a sample VariantPlex panel GTF file:

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

# Appendix B

## 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 this [web page](#).

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 11184573 COSM20417 C T . . Archer_Gene=MTOR;STRAND=-;Archer_MutationCDS=c.6644C>A;Archer_MutationAA=p.S2315Y;CNT=0
```



**IMPORTANT!** 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

© 2023 Archer. All rights reserved. Illumina® is a registered trademarks of Illumina, Inc., VMware vSphere®, VMware Fusion®, VMware Player® are a registered trademark or trademark of VMware, Inc. Life Technologies™, Ion Torrent™, Ion PGM™ and Ion Proton™ are trademarks of Thermo Fisher Scientific Inc. Mac® and Mac OS X® are registered trademarks of Apple Inc. Oracle® and VirtualBox™ are registered trademarks or trademarks of Oracle Inc. Windows® and Internet Explorer® browser are registered trademarks of Microsoft Corporation. Chrome™ browser is a trademark of Google Inc.

## Technical support

Visit our [support center](#) for a list of helpful answers to frequently asked questions, or contact us directly at [archer-tech@idtdna.com](mailto:archer-tech@idtdna.com).

For more than 30 years, IDT's innovative tools and solutions for genomics applications have been driving advances that inspire scientists to dream big and achieve their next breakthroughs. IDT develops, manufactures, and markets nucleic acid products that support the life sciences industry in the areas of academic and commercial research, agriculture, medical diagnostics, and pharmaceutical development. We have a global reach with personalized customer service.

SEE WHAT MORE WE CAN DO FOR YOU AT [WWW.IDTDNA.COM](http://WWW.IDTDNA.COM).

**For research use only. Not for use in diagnostic procedures.** Unless otherwise agreed to in writing, IDT does not intend these products to be used in clinical applications and does not warrant their fitness or suitability for any clinical diagnostic use. Purchaser is solely responsible for all decisions regarding the use of these products and any associated regulatory or legal obligations. Doc ID: RUO23-2168\_001