RUO PiVAT Platform v1 User Guide

PiVAT Platform v1.x

Released July 2025

Legal Notice

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

Table of Contents

Introduction

This user manual describes how to use Pillar Bioscience's (Pillar) PiVAT software. PiVAT is Pillar's genome sequence data software.

Software Requirements

The PiVAT software requires a supported browser to access the user interface. The current versions of the following browsers are supported:

- Google Chrome
- Mozilla Firefox

Symbols Used in this User Manual:

This symbol is a reminder to pay attention to details that could affect either proper installation or performance**.**

This symbol is an indication of a critical detail not to be overlooked.

Connect to PiVAT

1. Login to PiVAT using the username and password provided by Pillar Biosciences.

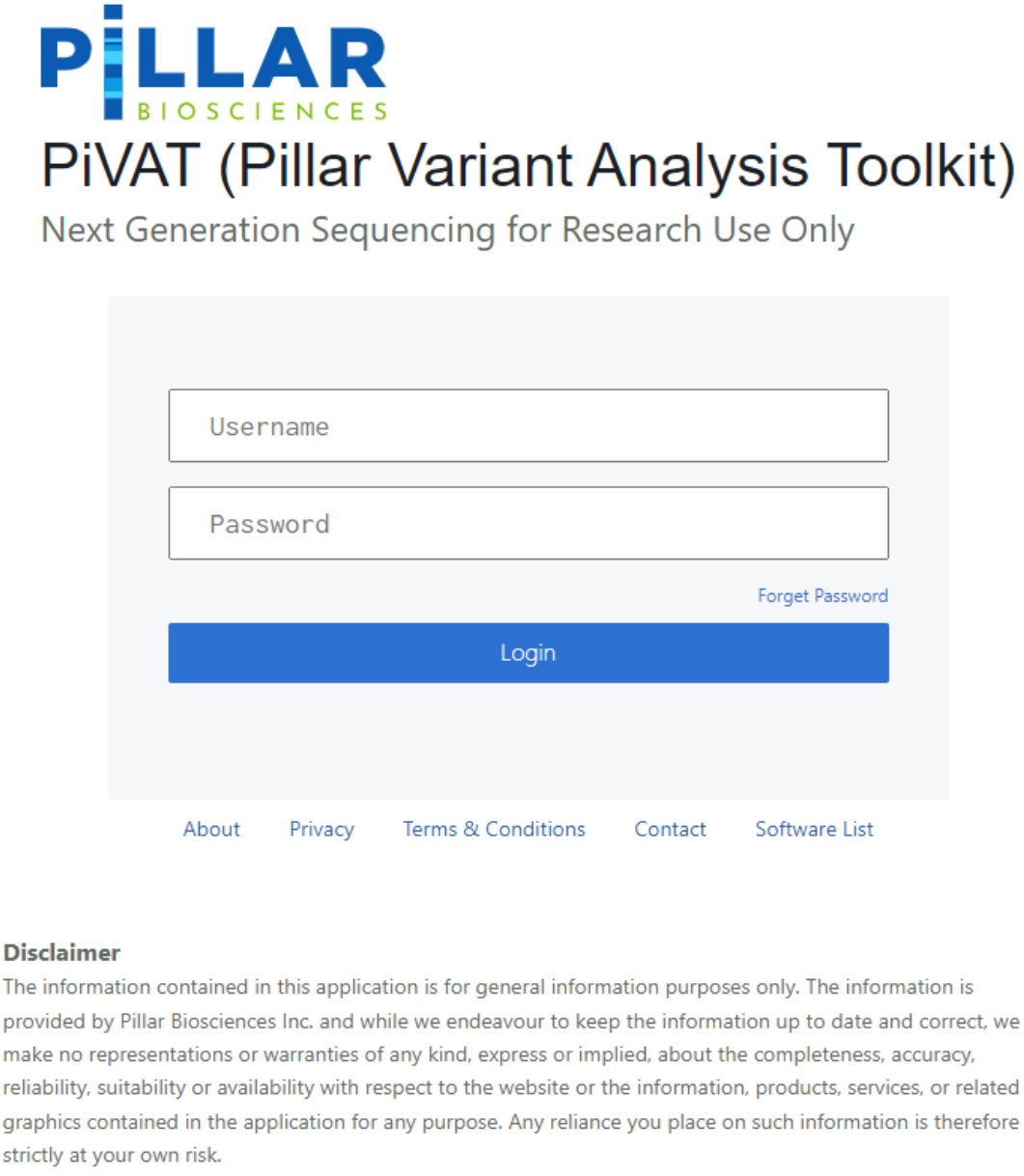


Figure 1 PiVAT login page

Accept Terms and Conditions

1. The following screen is displayed upon user's first login.
2. Read the terms and conditions and click the **Agree** button to continue.

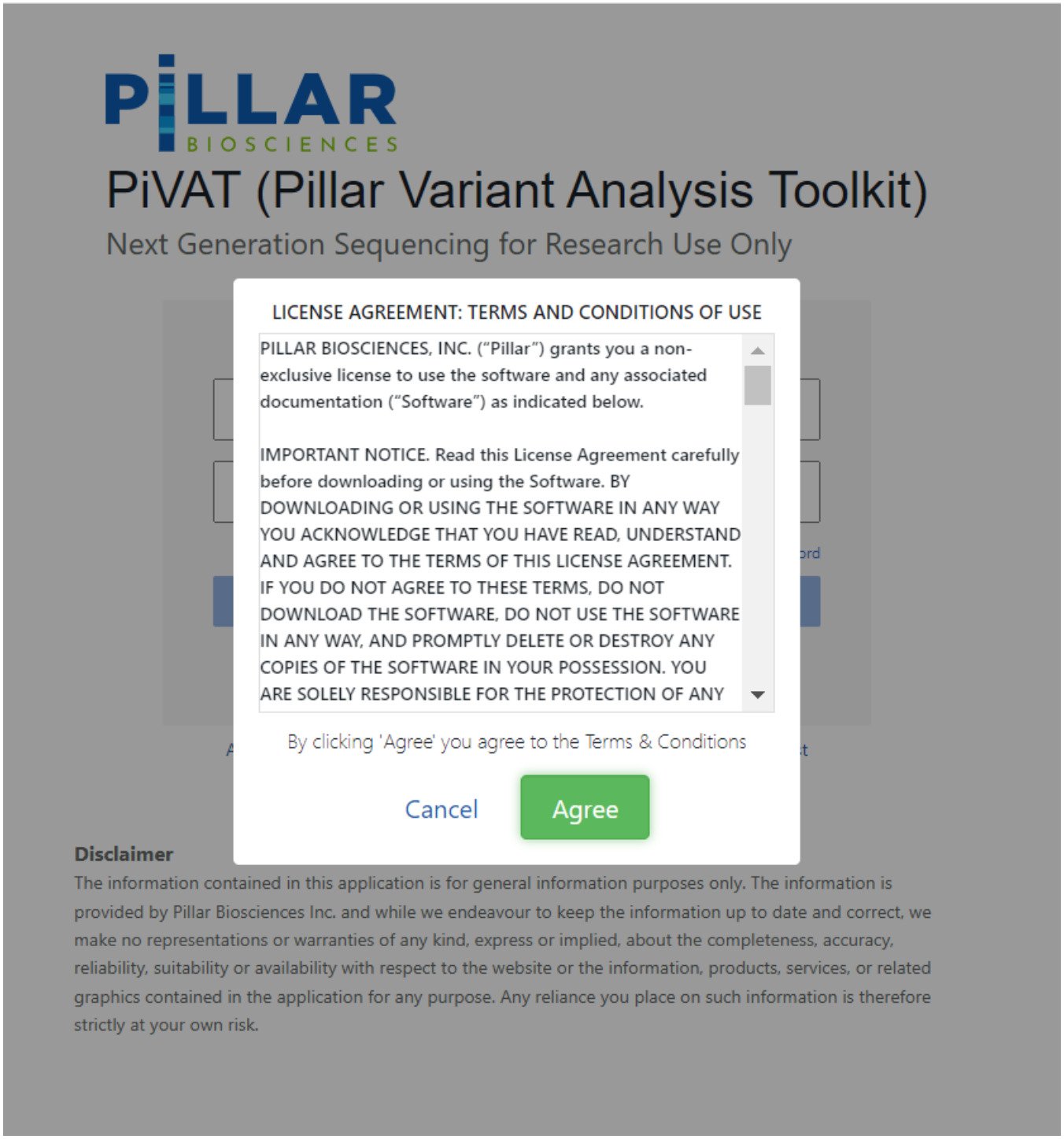


Figure 2 PiVAT License Agreement

Using the PiVAT Software

Dashboard

1. User may navigate through the application using the Dashboard screen.

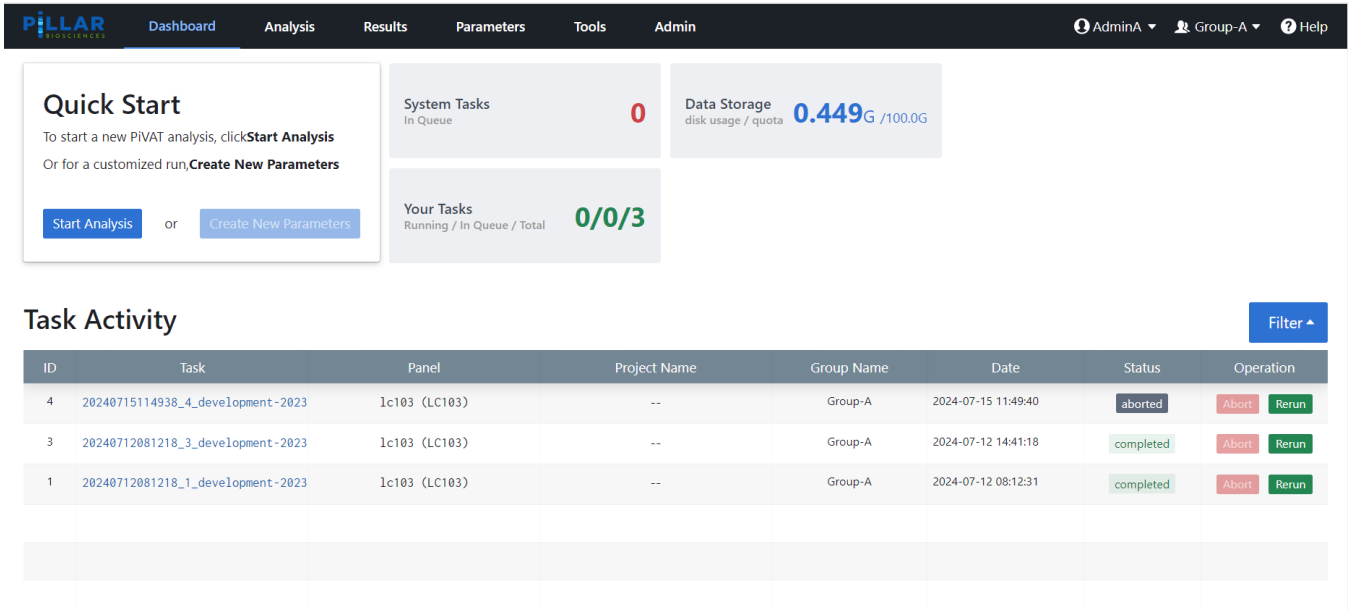


Figure 3 PiVAT Dashboard page

Navigation Bar

1. The following Navigation Bar is displayed on the top of PiVAT Dashboard page. Additional items are displayed for admin users and detailed in the PiVAT Admin Guide document.
2. The page being displayed will be highlighted on the Navigation Bar.
3. Logged in user is indicated next to icon.
4. User Group is indicated next to icon.

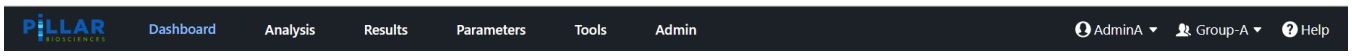


Figure 4 Navigation Bar on PiVAT Dashboard

Dashboard: Returns the user to the Dashboard screen

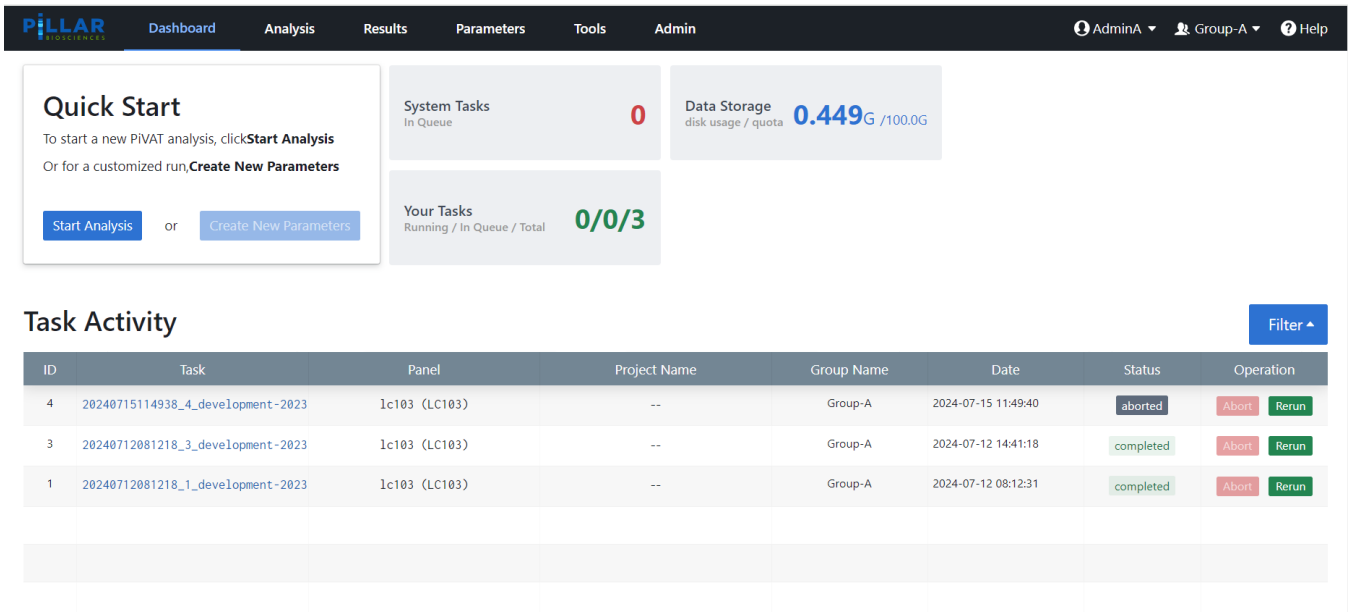


Figure 5: PiVAT Dashboard page

Analysis: Displays the screen for creating, starting, and viewing analysis tasks on the left menu bar include:

- **Project View**: Shows input files organized by project. If toggled off, will show file tree-based input navigation.
- **Results**: Shows queued and completed analysis tasks from user's User Group.
- **Parameters**: Allows Admin users to create custom parameters for a given panel (optional).

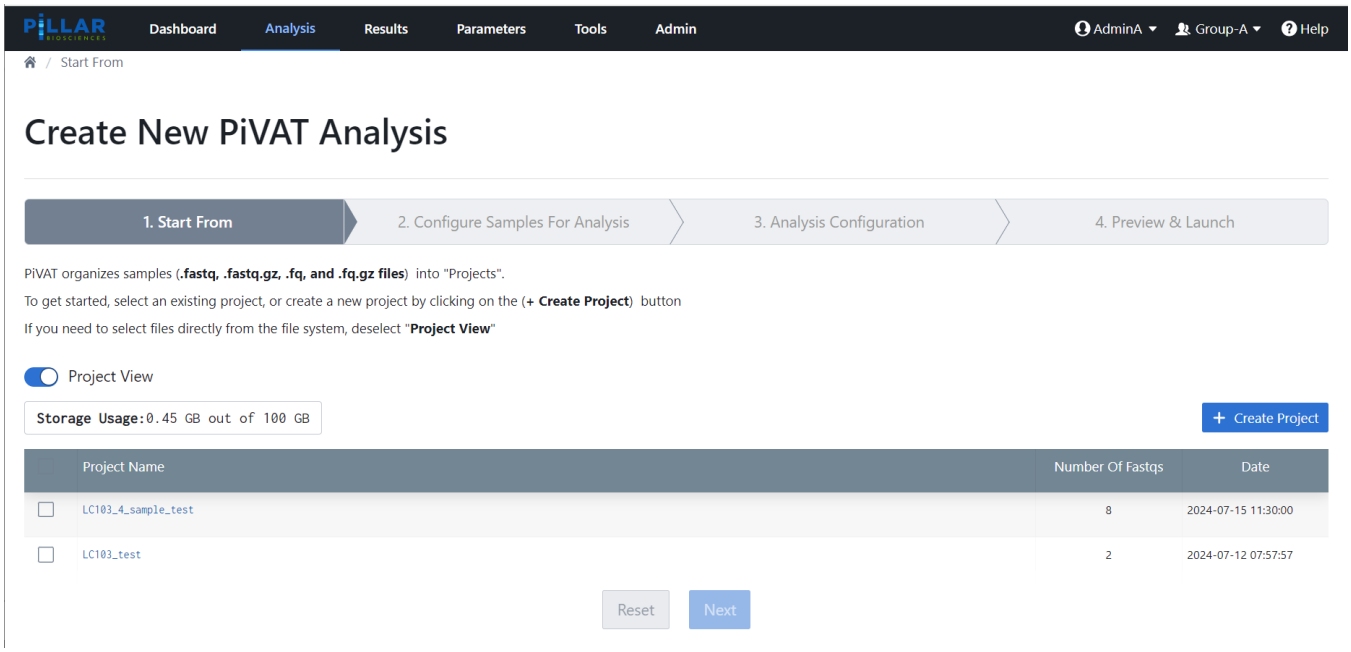


Figure 6: PiVAT Analysis page

Tools: Displays the screens for Integrative Genomics Viewer (IGV), Data Management and Download on the left menu bar.

- **IGV**: Choose *bam* and *bai* files from task server, and enter the locus you want to focus on.
    - Bam/bai files are located in Home/user/output/{panel name}/{run name}/{sample name}/analysis

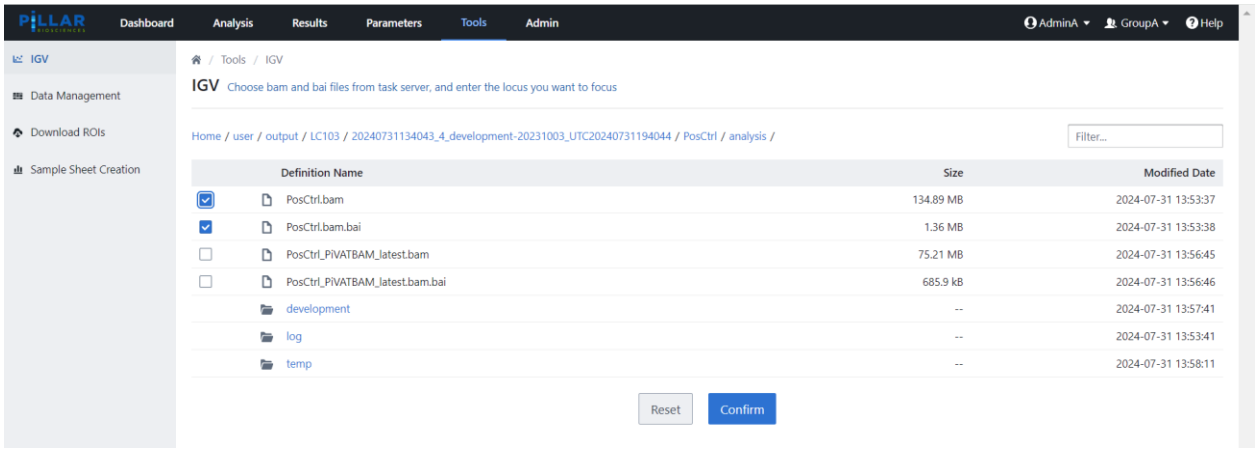


Figure 7: PiVAT Tools/IGV page

- **Data Management**: Sample data files can be uploaded to this section for use as inputs to the analysis tasks. The Project Name corresponds to a folder with the same name in the uploads directory. Sample files used for a task may be uploaded to a Project and be used as a bundled input for the Analysis Task. Project data, including both input and output files may be deleted using the Delete button
    - Note only System Admin and Group Admin users may delete project data

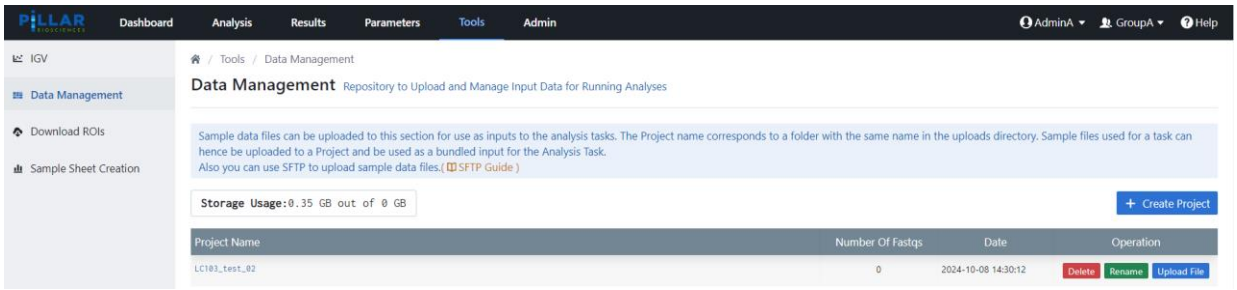


Figure 8 PiVAT Tools/Data Management page

- **Download ROIs**: Select and download panel and analysis parameters in bed format.
    - **Note**: only available to system Admin and Group Admin users

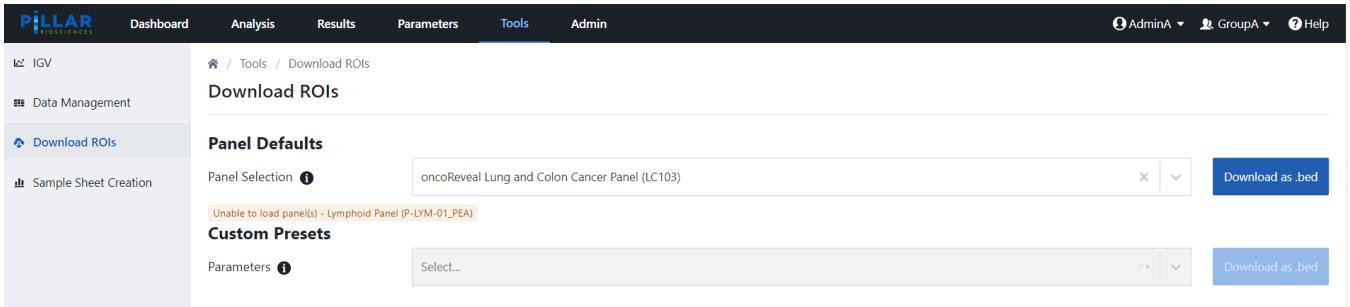


Figure 9 Download ROIs page

- **Sample Sheet Creation**: Allows user to create and download an Illumina v1 sample sheet compatible with MiSeq and NextSeq 550 sequencers

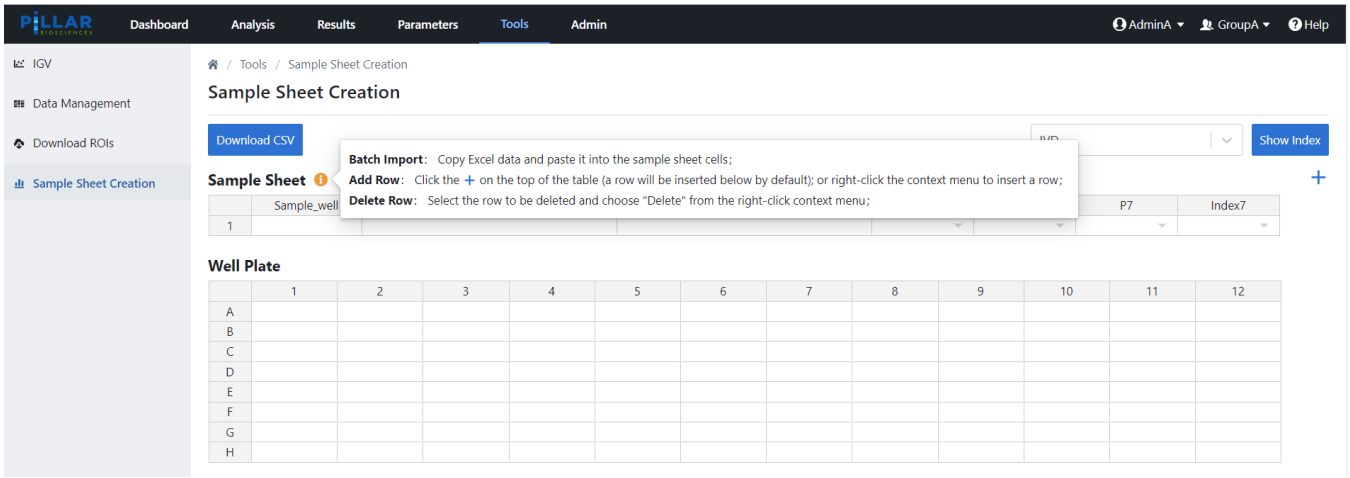


Figure 10 Sample Sheet creation page

Help: Displays product information, instructions for SFTP, and a downloadable User Manual in PDF format.

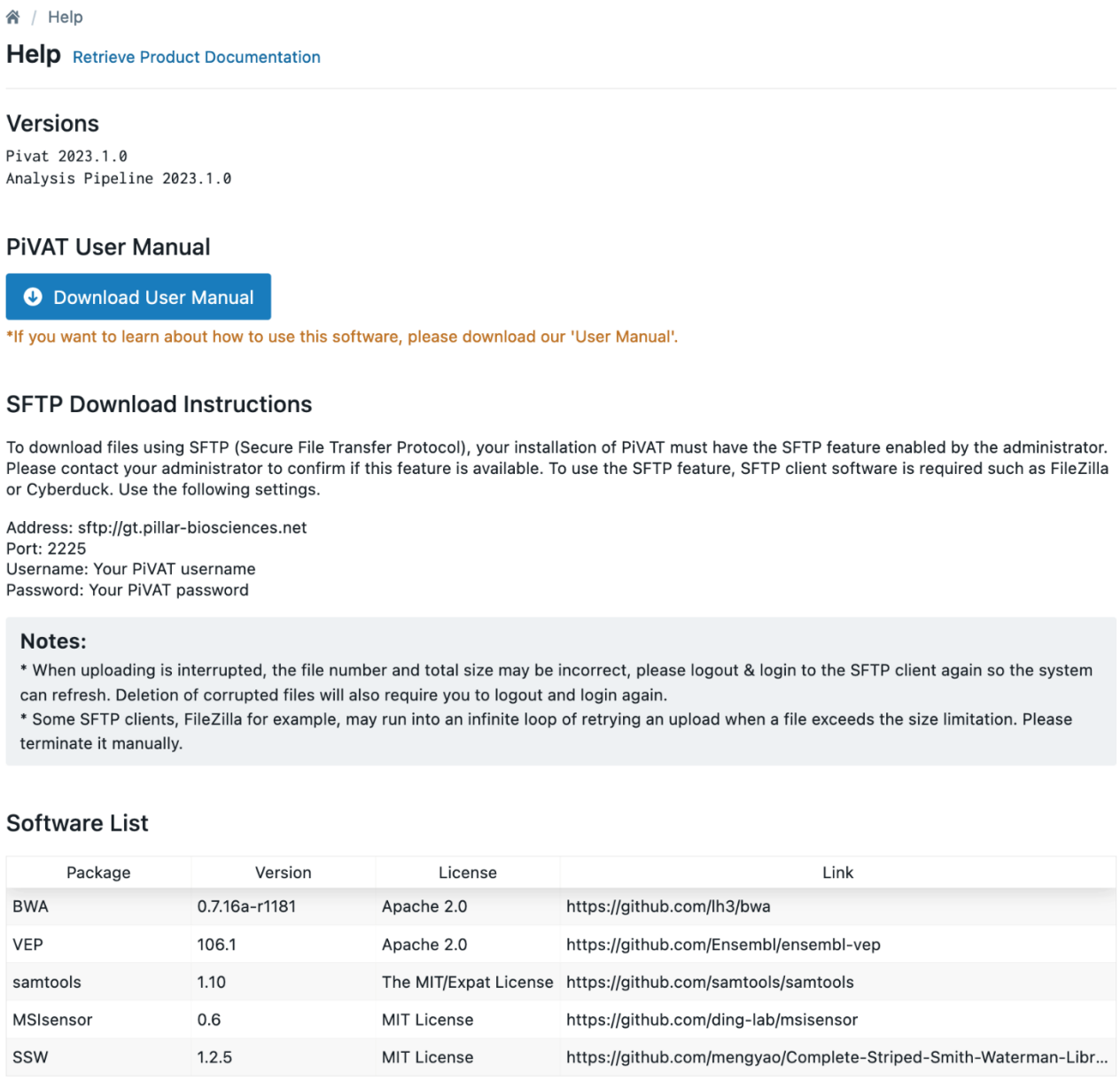


Figure 11: Help page

Setup an Analysis

Creating Analysis Parameters

1. Begin with selecting **Parameters** on the Navigation Bar or selecting **Create New Parameters** from the quickstart menu to display the following Create New PiVAT Analysis page.

The Custom parameter setting described in this Creating Analysis Parameters section is not required to start a run.

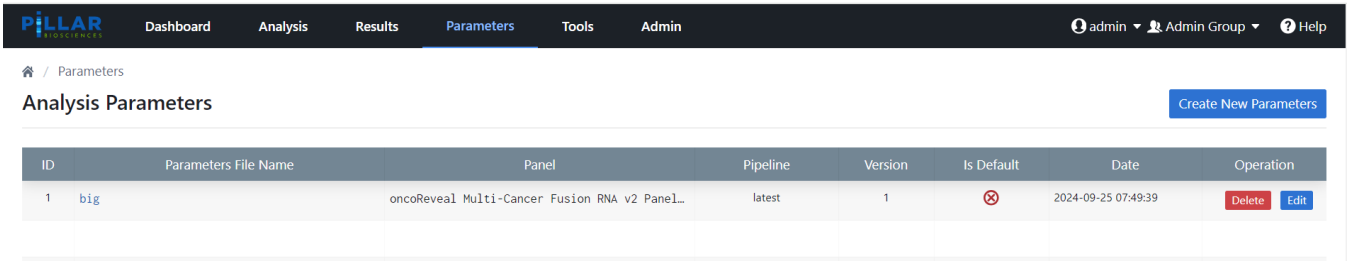


Figure 12: Analysis Parameters page

2. Select **Create New Parameters** to display the following Analysis Parameters page. This enables the user to customize the analysis parameters.

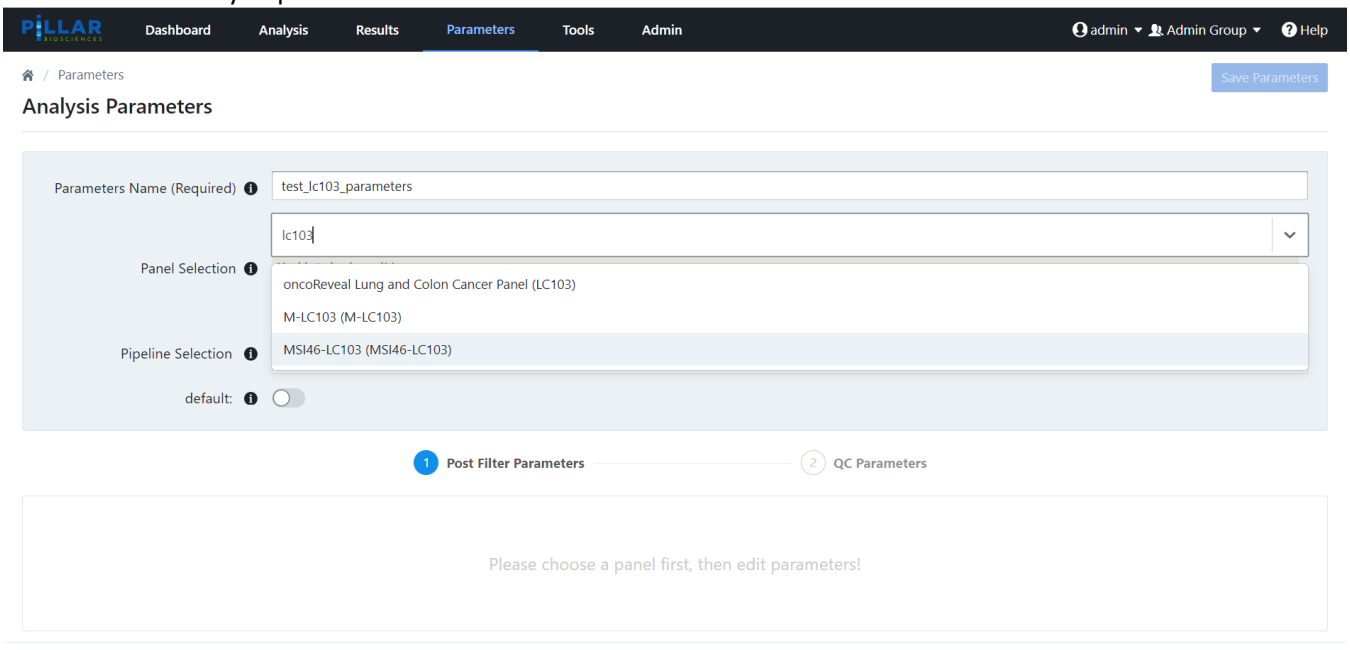


Figure 13: Create New Parameters page

- **Parameters Name:** Enter desired parameters file name. The user defined parameters will be saved with this name for future use.

- **Panel Selection:** Select desired panel using drop-down menu.
- **Pipeline Selection:** The PiVAT Platform can support multiple different analysis pipelines. Different pipelines may have different parameter options. Use Pipeline Selection to choose which pipeline you would like to create new parameters for.
- **Default:** Toggling this on will make the new parameters the default parameters for the selected panel on the chosen pipeline for users within that User Group.

 While System Admin and Group Admin users may choose which custom parameters they would like to use for a given run, Analyses executed by Regular users will always use the Default Parameters for a given panel/pipeline/User Group, or no custom parameters if Default Parameters have not been set for that panel + pipeline.

- Select ① **Create New Parameters** and navigate across the tabs to customize settings for, Variant filtering, CNV filtering, Fusion filtering and QC

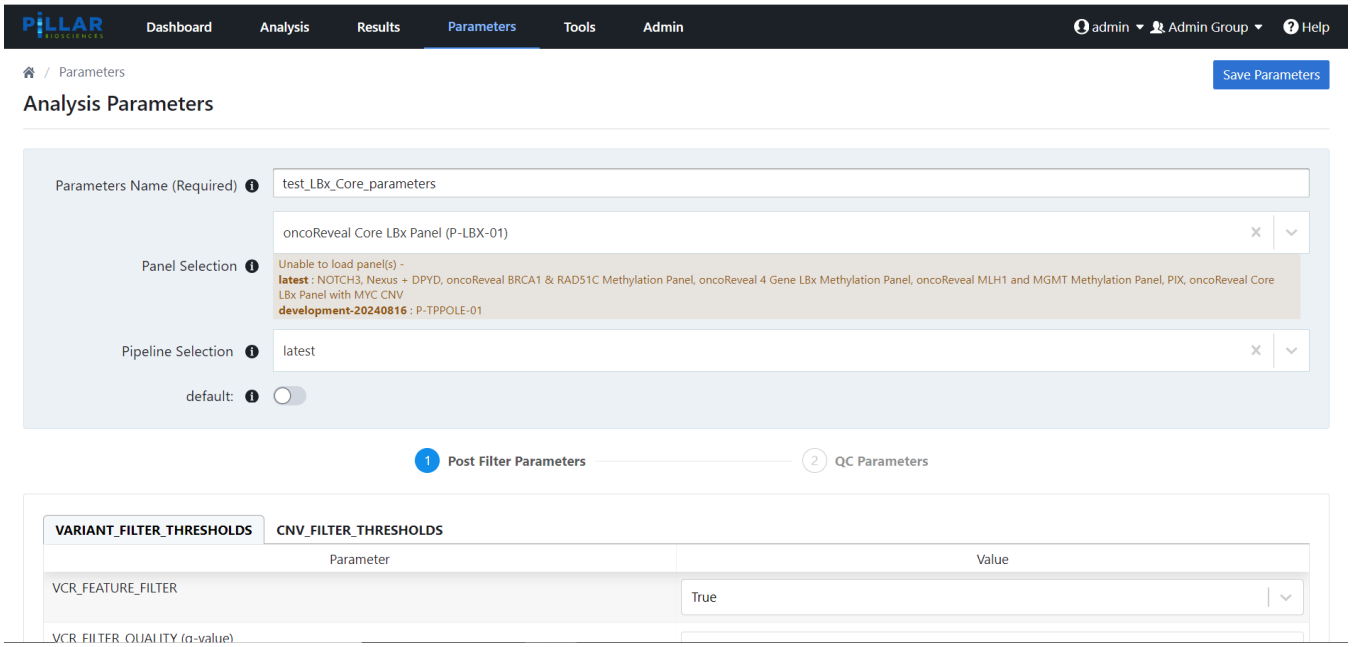


Figure 14: Analysis Parameters: ① Post Filter Parameters page

3. Users may also customize settings for **Grouping** on the Post Filter Parameters page

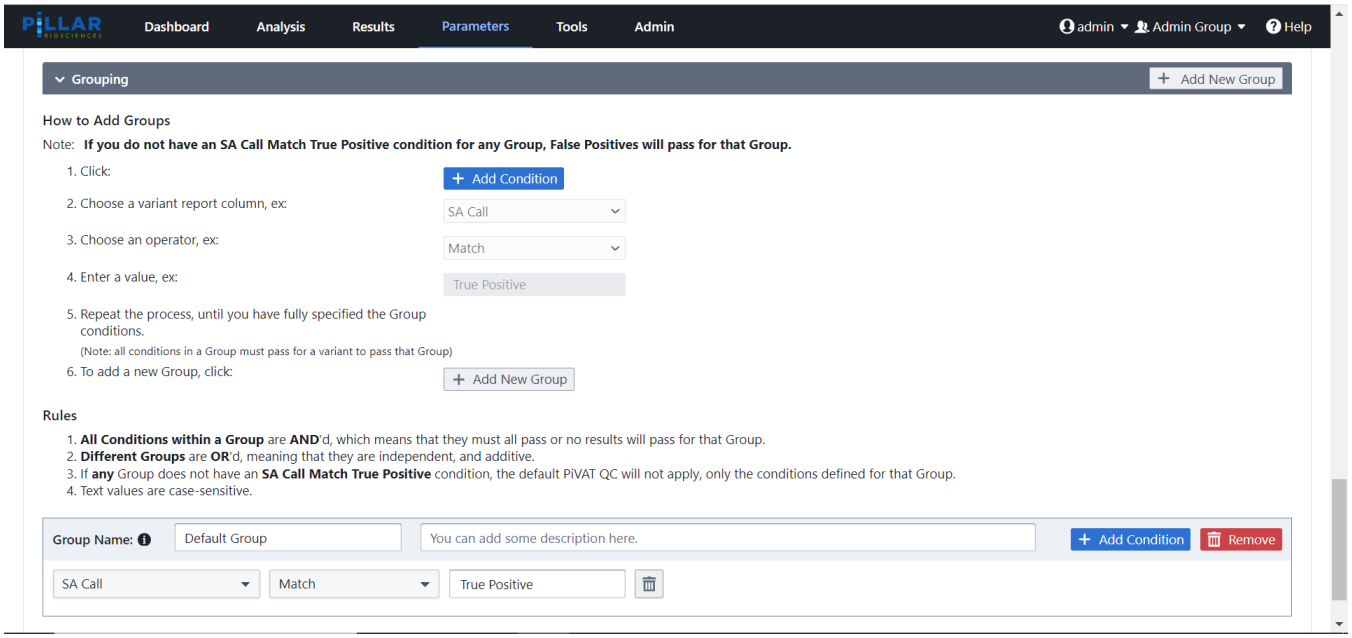


Figure 15: Grouping

4. In ② **Post Filter Parameters,** use the CNV Threshold Grid to customize thresholds for filtering CNV calls. The CNV Threshold Grid will only appear if the selected panel is a CNV panel.

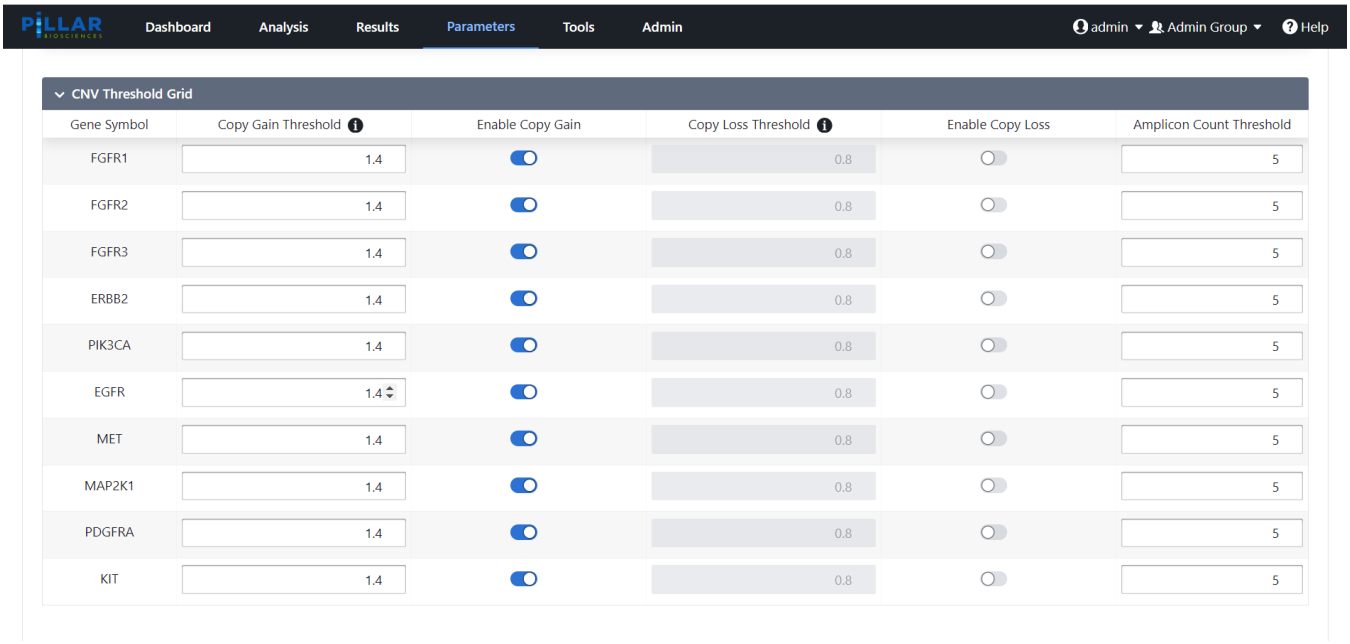


Figure 15: Post Filter Parameters page CNV Threshold Grid

Note: When enabling Copy Loss for panels using the cfDNA CNV Caller (see Appendix C, currently only P-LBX-01): the **CNV_MAX_COPY_NUMBER** parameter will need to be adjusted to ensure that calls below the Copy Loss Threshold are reported. See Appendix C.

5. Select ③ **QC Parameters** to customize QC parameters for **Sample** and controls (**NTC**, **PosCtrl**, **NegCtrl**).

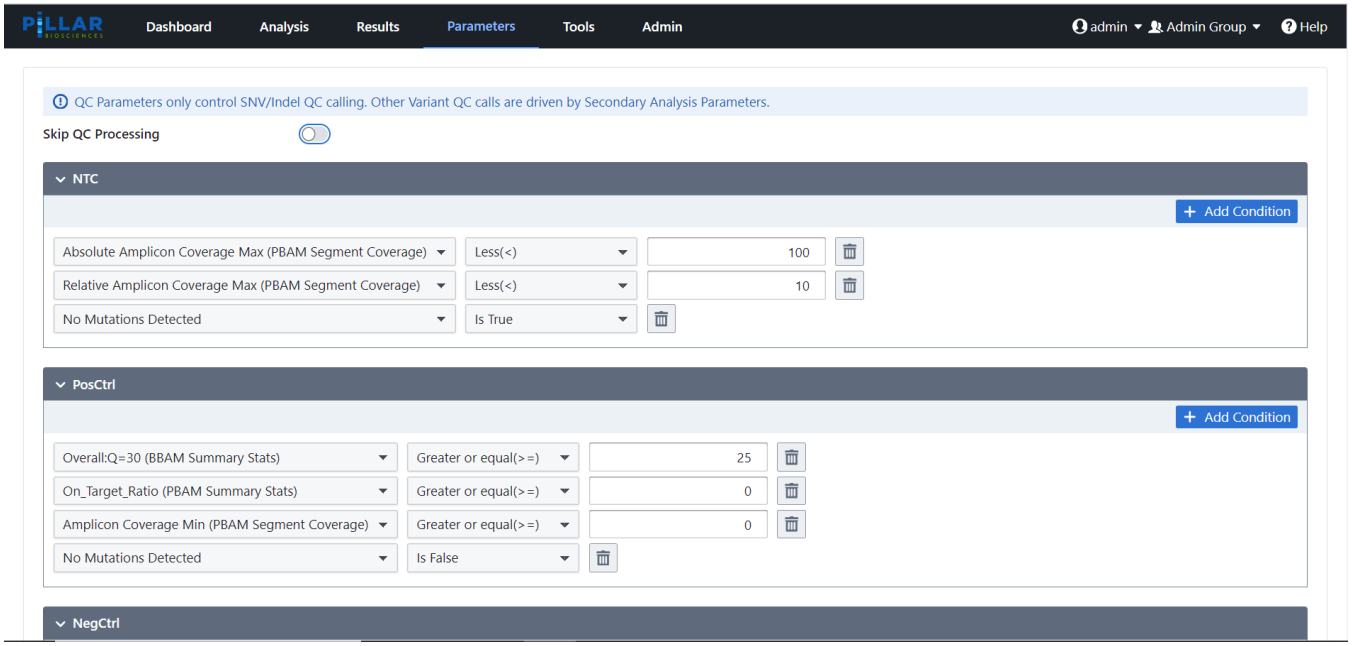


Figure 16: QC Parameters page

QC Parameters only control SNV/Indel QC calling. Other Variant QC calls are driven by Secondary Analysis Parameters.

Start Analysis

Please refer to the following appendices for caller-specific instructions.

- Appendix A: SNV/Indels Caller
- Appendix B: Microsatellite Instability
- Appendix C: Somatic CNV Caller
- Appendix D: Thalassemia CNV Caller
- Appendix D: Thalassemia CNV Caller
- Appendix D: Thalassemia CNV Caller

1. Select **Start Analysis** on the left menu bar to display the following Start Analysis: 1. Start From page.
   - **Select Panel:** Select desired panel from the drop-down menu.
   - **Parameters File**: Parameters file selection is optional. Refer to Creating Analysis Parameters section for instructions to create custom parameters file. Default parameters will be applied if parameters file is NOT selected.
   - **Select Samples**: Select desired FASTQ format files. Refer to Data Management section for instructions to import data.

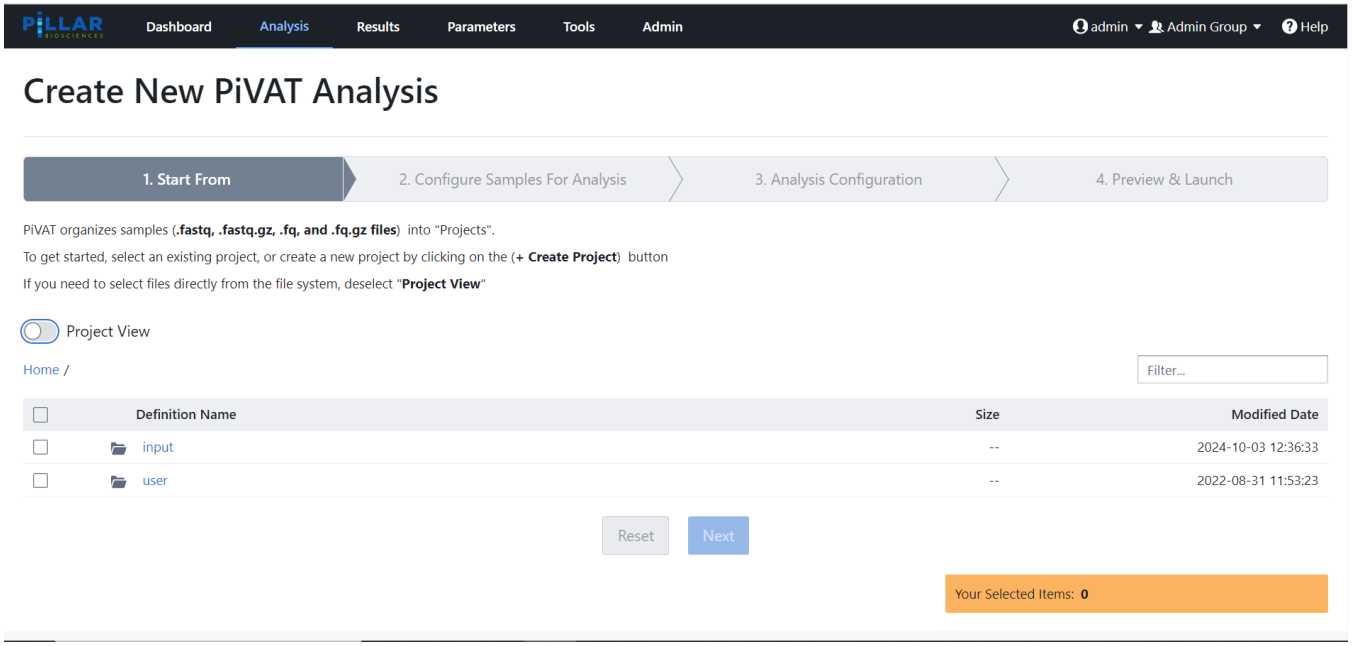


Figure 17: Start Analysis page

2. Select **Next** to display the following window: 2. Configure Samples for Analysis page.

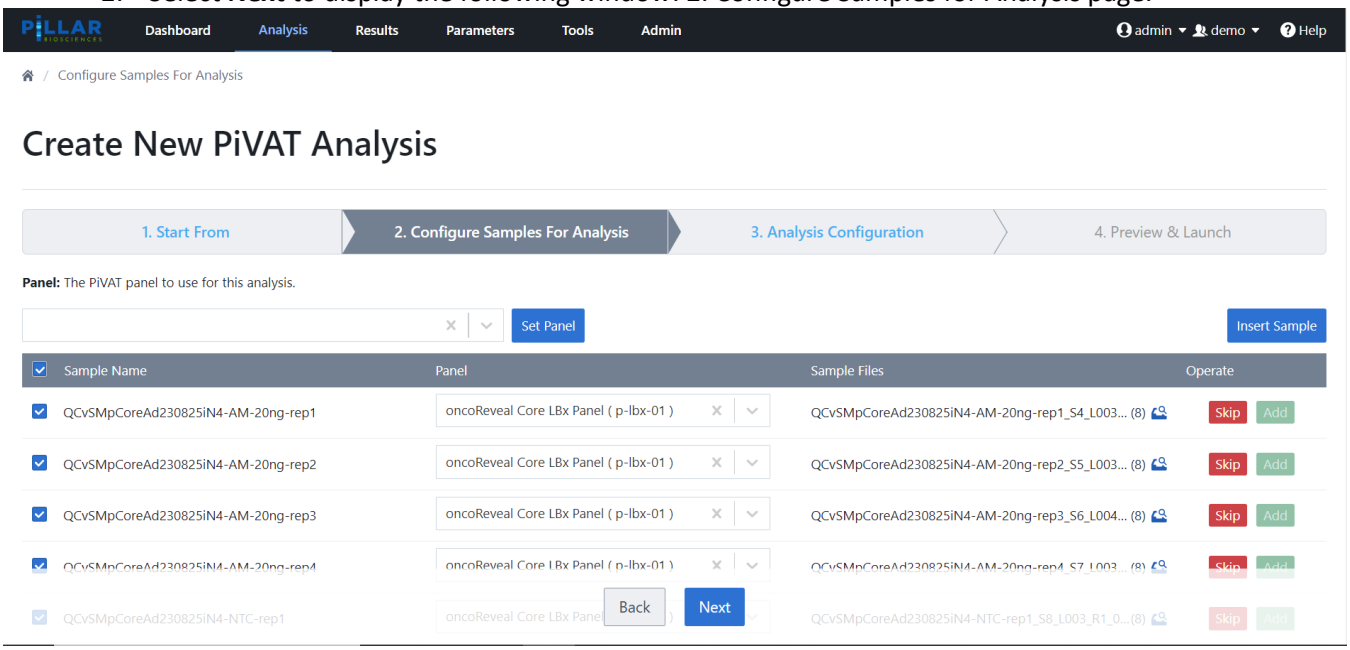


Figure 18: Configure Samples for Analysis page

- Select the panel to be used to analyze each sample
- Panel may be changed for multiple samples by using the check boxes and Set Panel button

3. Select **Next** to display the following window: 3. Analysis Configuration page.

At the Analysis Configuration page, users may rename individual samples.

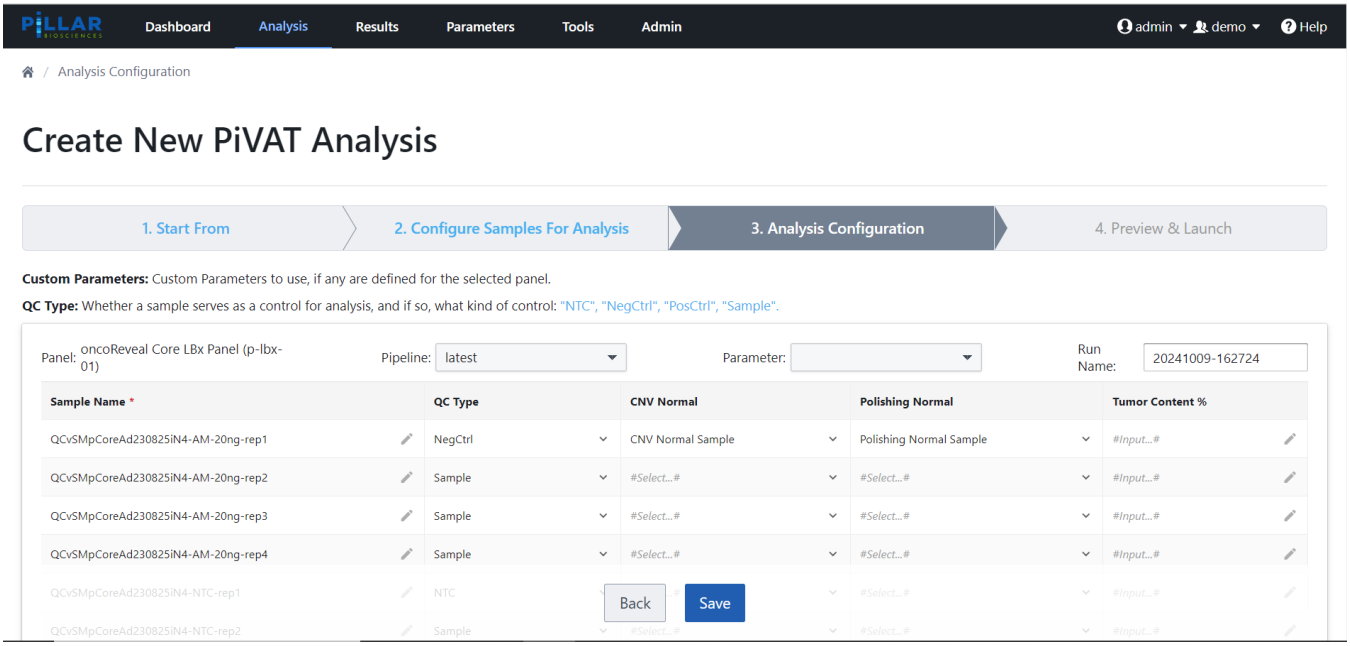


Figure 19: Analysis Configuration page

- **Pipeline:** The analysis pipeline to be used for the run. If multiple PiVAT Analysis Pipelines are installed on the PiVAT Platform, they may be selected here
- **Parameter**: Parameters file selection is optional. Refer to Creating Analysis Parameters section for instructions to create custom parameters file. Default parameters will be applied if parameters file is NOT selected
- **Run Name:** Enter desired name for the analysis to uniquely identify the run.

4. Users may edit sample information such as **Sample Name**, **QC Type** as well as edit or skip **Sample Files**.
5. Users may also select samples using check boxes to define **Tumor-Normal Paired** and **CNV Normal Samples**.

Tumor-Normal Paired definition is required for performing Paired MSI panels, refer to Appendix B: Microsatellite Instability (MSI) Caller for details.

CNV Normal Samples definition is recommended for CNV analysis, refer to Appendix C: Somatic CNV Caller for details.

CNV Normal Samples definition is required for Thalassemia CNV analysis, refer to Appendix D: Thalassemia CNV Caller for details.

P-LBX-01 requires at least 1 CNV Normal to be specified (using Define CNV Normal Samples button as seen in the image below). If sample tumor content is known, it should be specified in the **Tumor Content %** column (as a percentage) for improved CNV calling performance. However, calls will still be done in terms of sample copy number, tumor content solely functions to improve the signal to noise tuning.

6. Select Next to display the following window: Preview & Launch page. Once user confirms the analysis setup, select **Launch Analysis**.

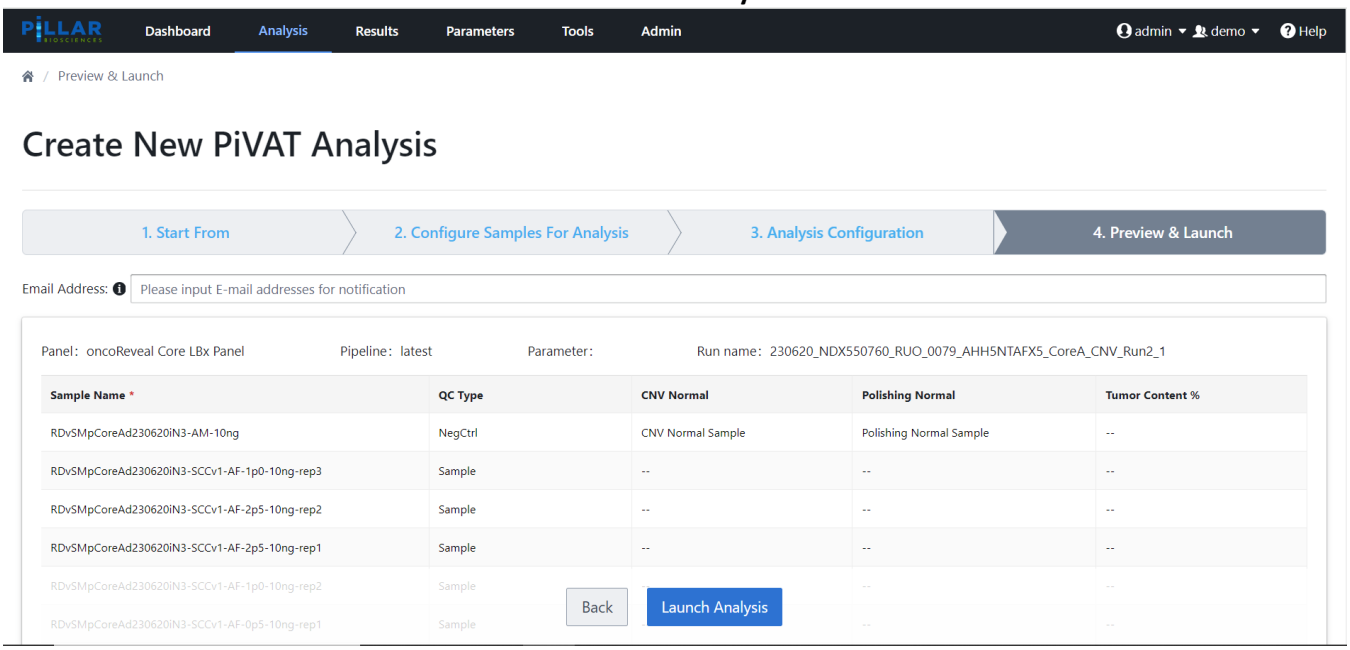


Figure 20: Preview & Launch

7. A pop-up message will appear. User may select:

- **Start Another Analysis** to redirect to **Start Analysis** page, or
- **Go to Analysis Results** to redirect to **Results** page.

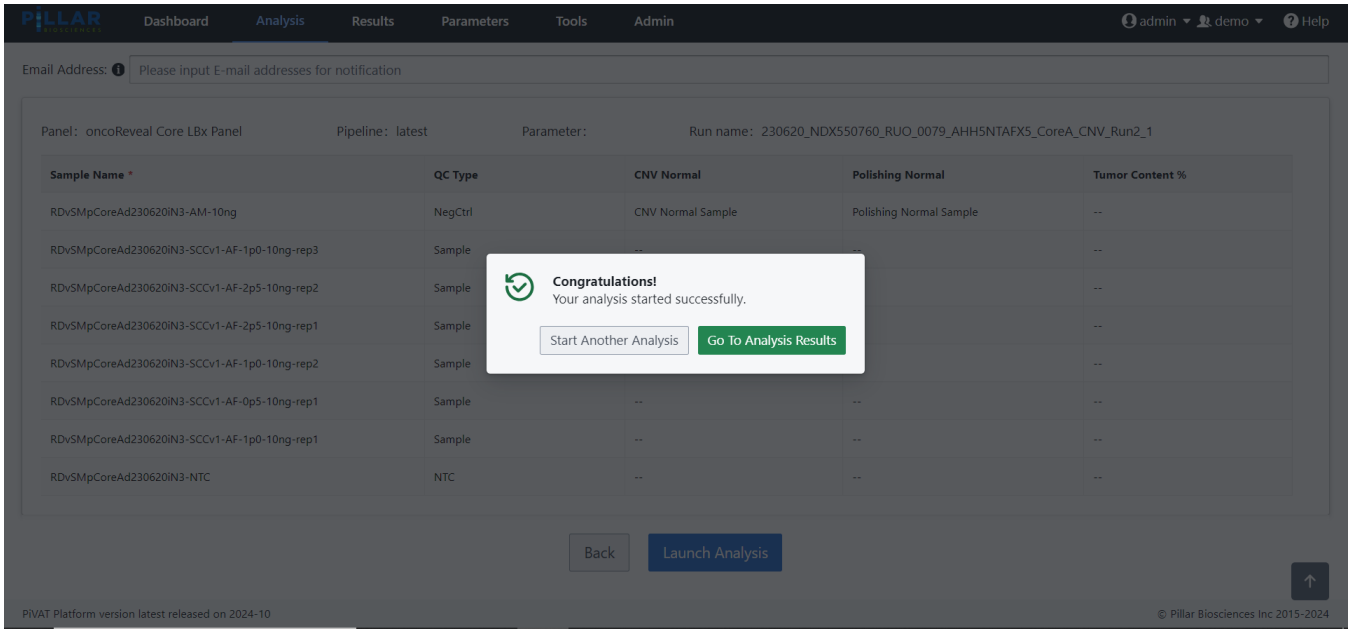


Figure 21 Start Analysis: 3. Preview & Launch page

Monitoring the Task Status

1. The PiVAT Dashboard will display a summary of analysis tasks, data storage and **Task Activity** of the User Group of the user currently logged in.

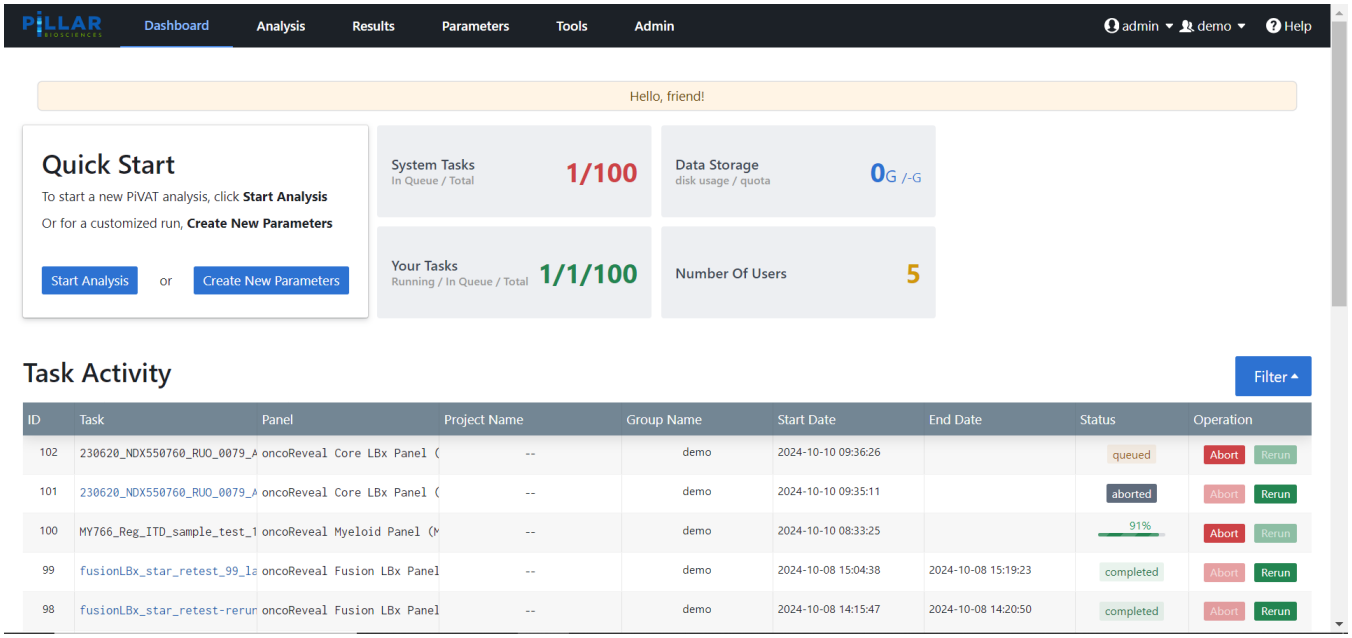


Figure 22: Dashboard page

2. The following information is provided for each task row:

- **ID:** Unique task ID number for current database.
- **Task:** Analysis name entered by user during analysis setup.
- **Panel:** Panel selected for the analysis.
- **Project Name:** Name of the Project associated with the analysis.
- **Group Name:** Name of the User Group of the user that initiated the analysis.
- **Start Date:** Date when the analysis task was created.
- **End Date:** Date when the analysis task completed or was aborted
- **Status**: Analysis task status
  - ✓ **queued**: There are other task(s) running. This task is queued until preceding tasks are finished.
  - ✓ **running**: This analysis is currently running.
  - ✓ **completed**: The analysis has completed successfully.
  - ✓ **failed**: An error occurred, and the analysis did not complete.
  - ✓ **aborted**: This task was aborted by a user.
- **Operation**: User may select **Abort** to terminate an analysis task in running state or **Rerun** to repeat an analysis as-is or with edits to analysis setup.

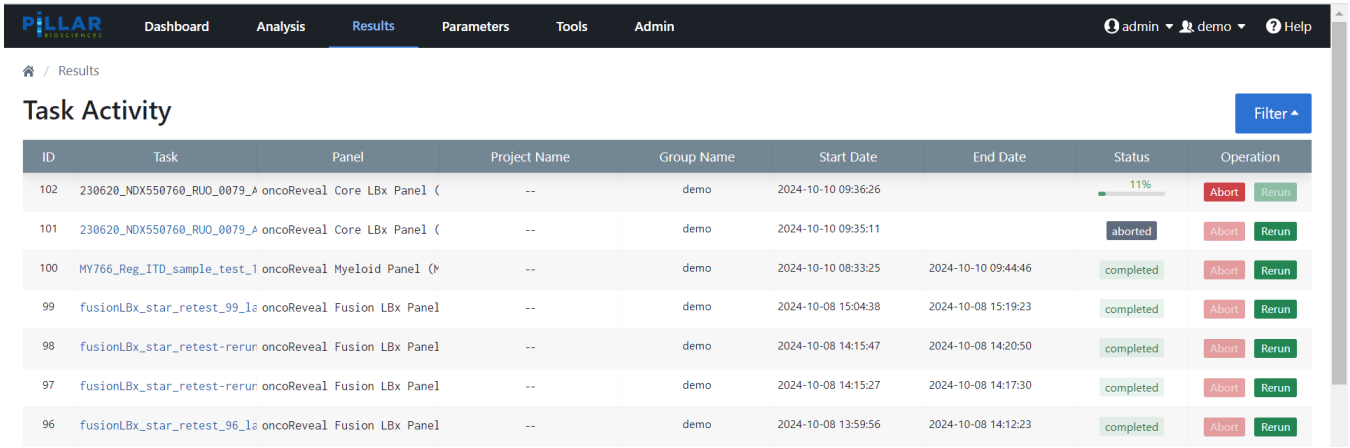


Figure 23: Analysis Results Task Activity page

View Summary of Task Details

1. Hover the mouse over the task name to view a pop-up summary of analysis.

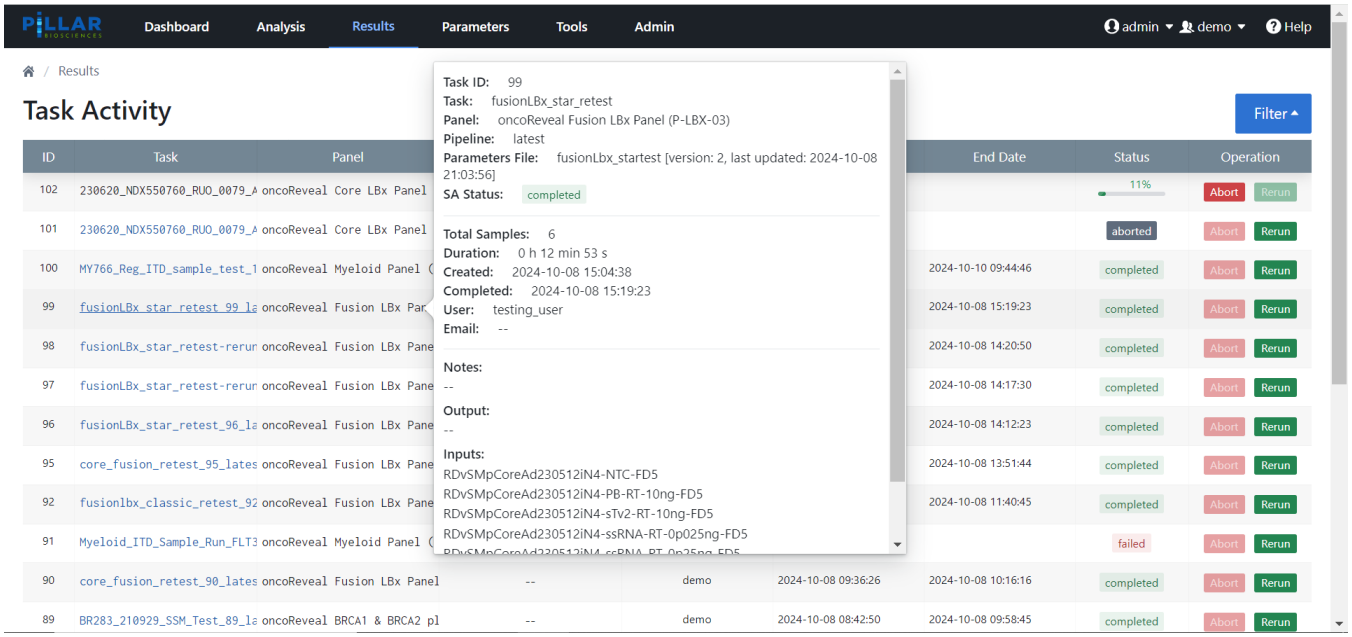


Figure 24: Analysis Task detail.

2. The summary information includes:

- Task ID
- Task name
- Panel
- Pipeline
- Parameters file name
- Task status
- Total number of Samples
- Duration
- Created and Completed date/time stamps. Note that the time between the Created and Completed time stamps may be longer than the Duration if the task was queued before starting
- User
- Email address (if configured)
- Output folder location
- Input files
- Arguments

Download Analysis Results

1. Select task hyperlink from list of completed tasks to display the Analysis Task page. Users may download analysis results, view log file and task information or rerun an analysis task.

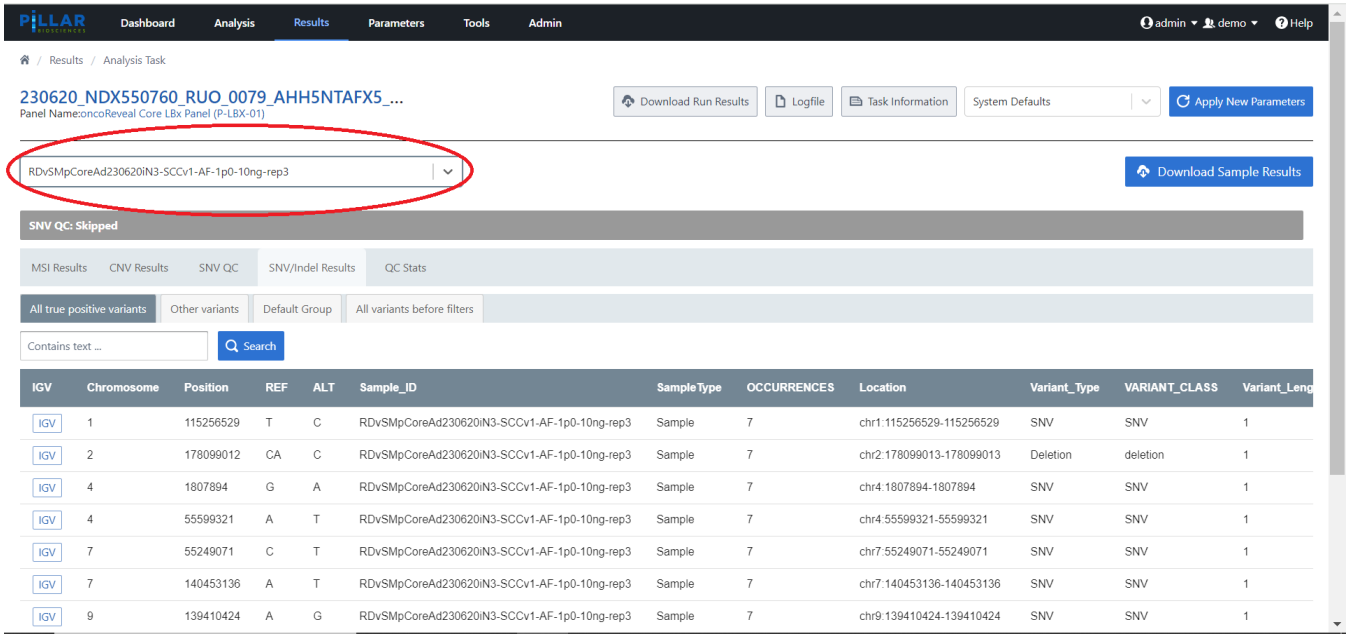


Figure 25: Analysis Task page with sample selection circled

2. Click the **Download Sample Results** button to download a zip formatted file with the sample results files. Click the **Download Run Results** button to download a zip file containing all sample and run results.

Results are shown for the sample selected only. Sample selection drop down is above

File download is unavailable if the task is in the **queued** or **deleted** status.

Users may delete the zip file from this screen. Please verify that analysis data has been downloaded and stored in accordance with your organization's IT policies *before* deleting data from the PiVAT system.

3. Analysis outputs may be accessed individually on the GUI via the Browse Downloadable Results page, accessible by clicking the "Task Information" button on the Results/Analysis Task page, then clicking on "Browse Downloadable Result Data" on the following Task Information page as shown in the following images:

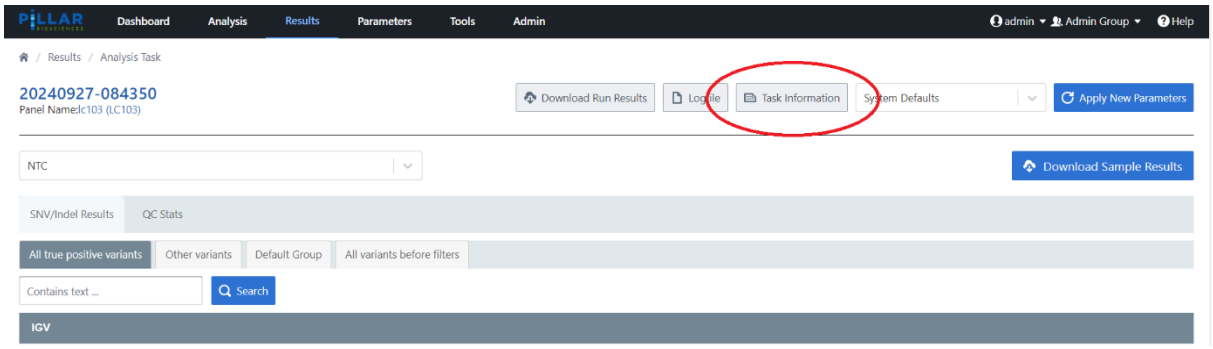
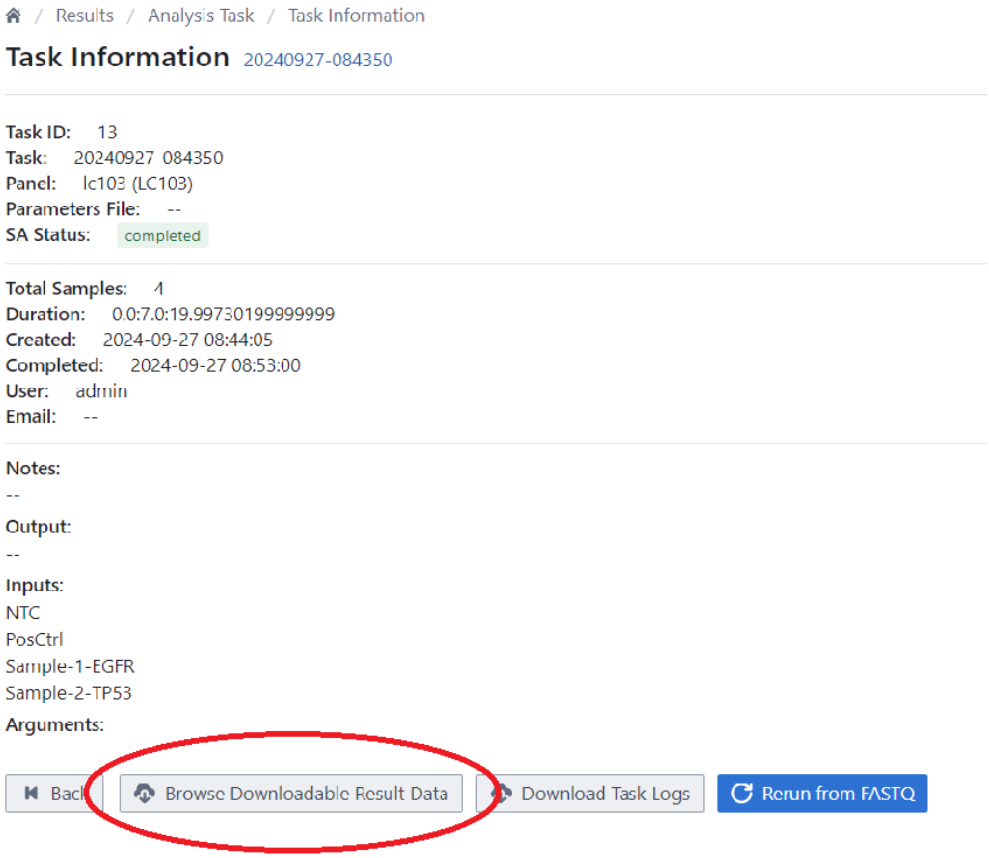


Figure 26: Analysis Task page with Task Information circled

3.1. Click the **Task Information** button to display the summary of analysis, this is the same information you get by hovering mouse over the task name on Analysis Result page.



3.2.

Figure 27: Task Information page with Browse Downloadable Data circled

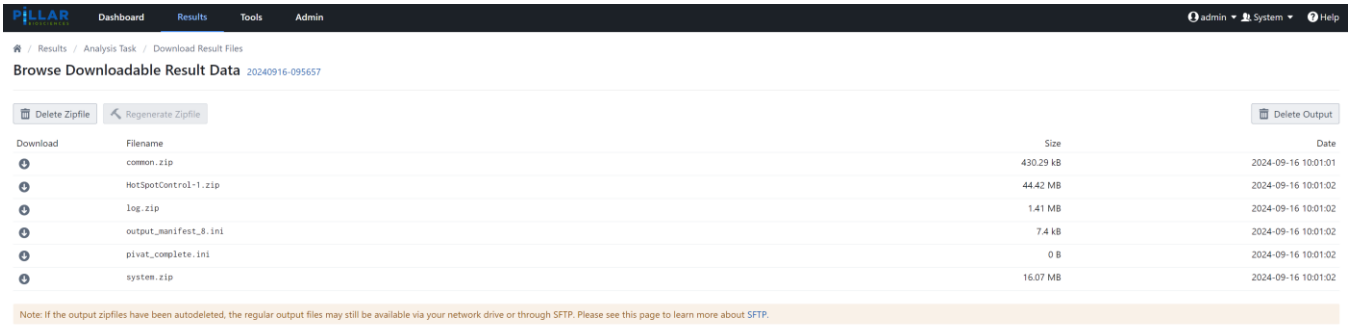


Figure 28: Browse Downloadable Result Data page

4. A deleted zipfile may be recovered by selecting the **Regenerate Zipfile** button. This selection is only available if the files are still present on the system. If the output zipfiles have been deleted, the regular output files may still be available via your network drive or through SFTP. Refer to the Help page to learn more about SFTP.
5. Click the **Back** button or Analysis / Analysis Results / **Analysis Task** to return to Analysis Task page.
6. Click the **Download Task Logs** button from Task Analysis page to download task logs. The information is not available when the task is in the running, queued or deleted status.
7. Click the **Back** button or Analysis / Analysis Results / **Analysis Task** to return to Analysis Task page.
8. Click the **Rerun** button for the option to repeat the analysis as-is or with edits to analysis setup.

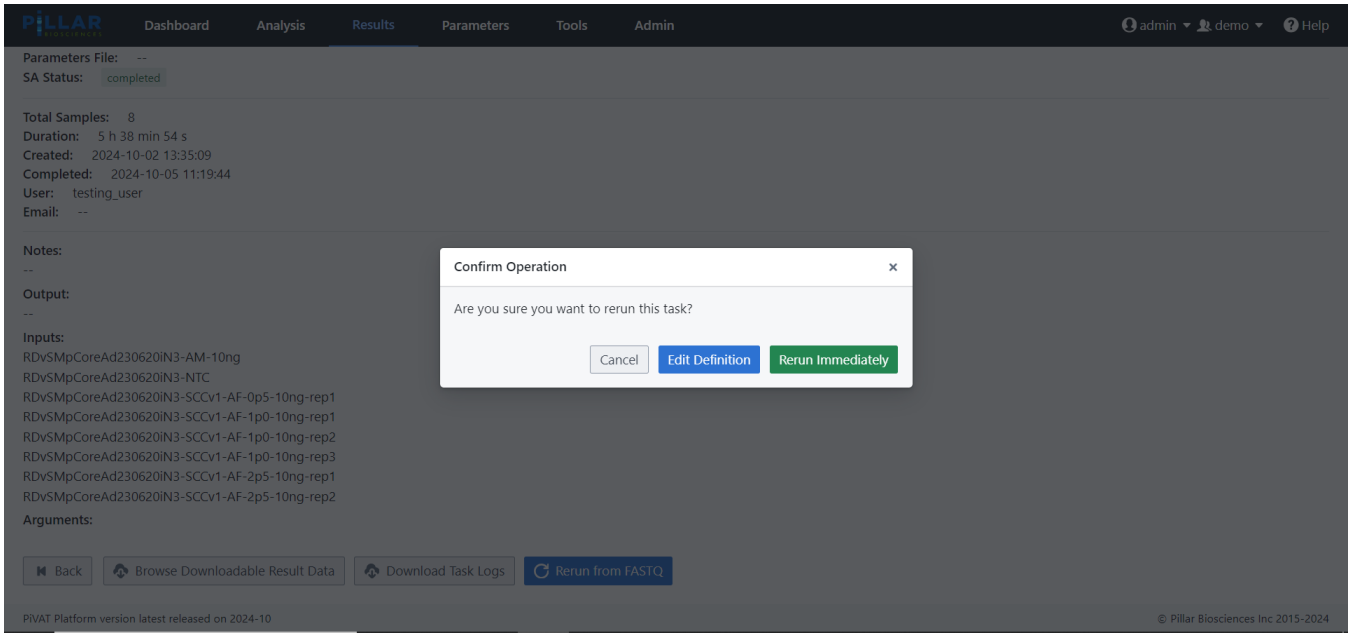


Figure 29: Rerun confirmation message

View Analysis Results

1. Users may navigate across the tabs to view MSI Results, CNV Results, SNV QC, SNV/InDel Results, or QC Stats. For fusion results, the user may view Fusion Calls, Alternative Splicing, or Expression Controls.
2. Click the **SNV QC** tab to display summary of Sample Type and Status for a sample.

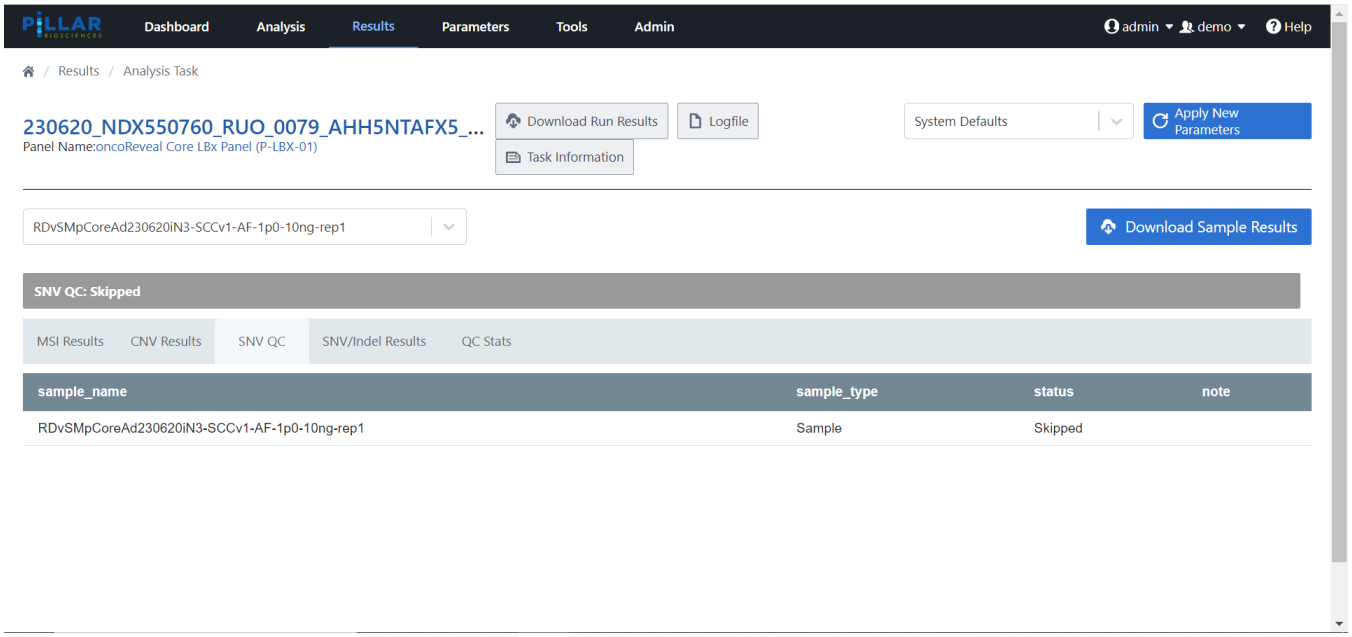


Figure 30: Analysis Task: QC Summary tab

3. Click the **QC Stat** tab to display and navigate sub-tabs to view Summary, Coverage Depth, Coverage Uniformity, and Segment Stats.

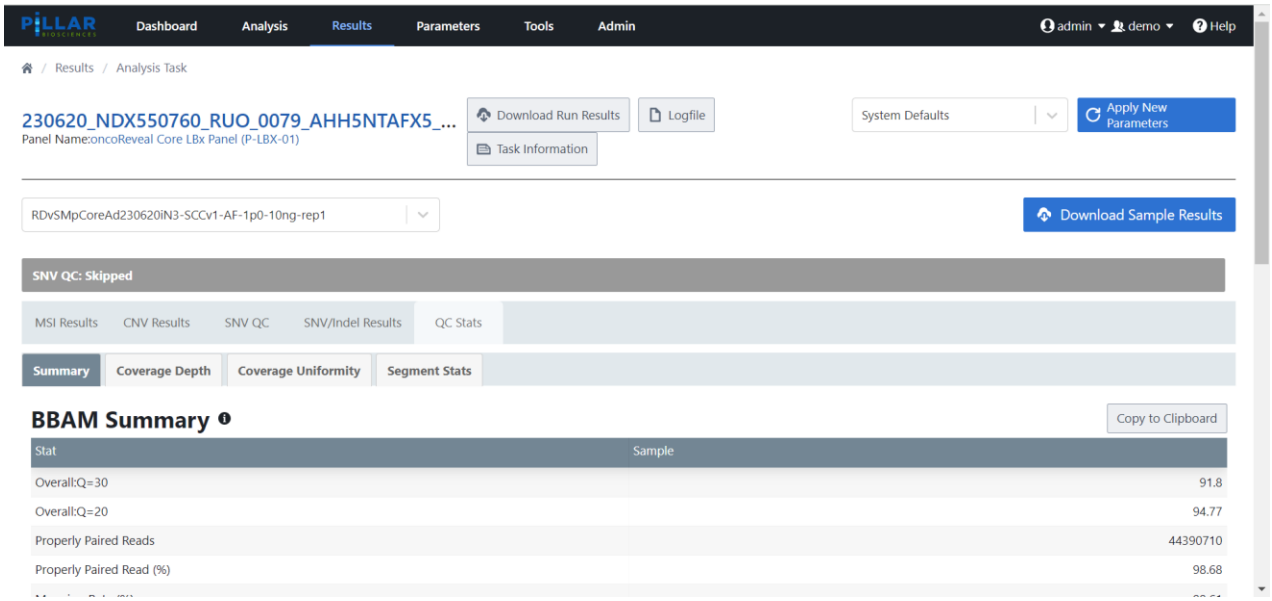


Figure 31: QC Stats tab

4. Click the **SNV/InDel Results** tab to display and navigate sub-tabs to view SNV/InDel results. SNV/InDel results grouping will be displayed according to the Parameters File applied.

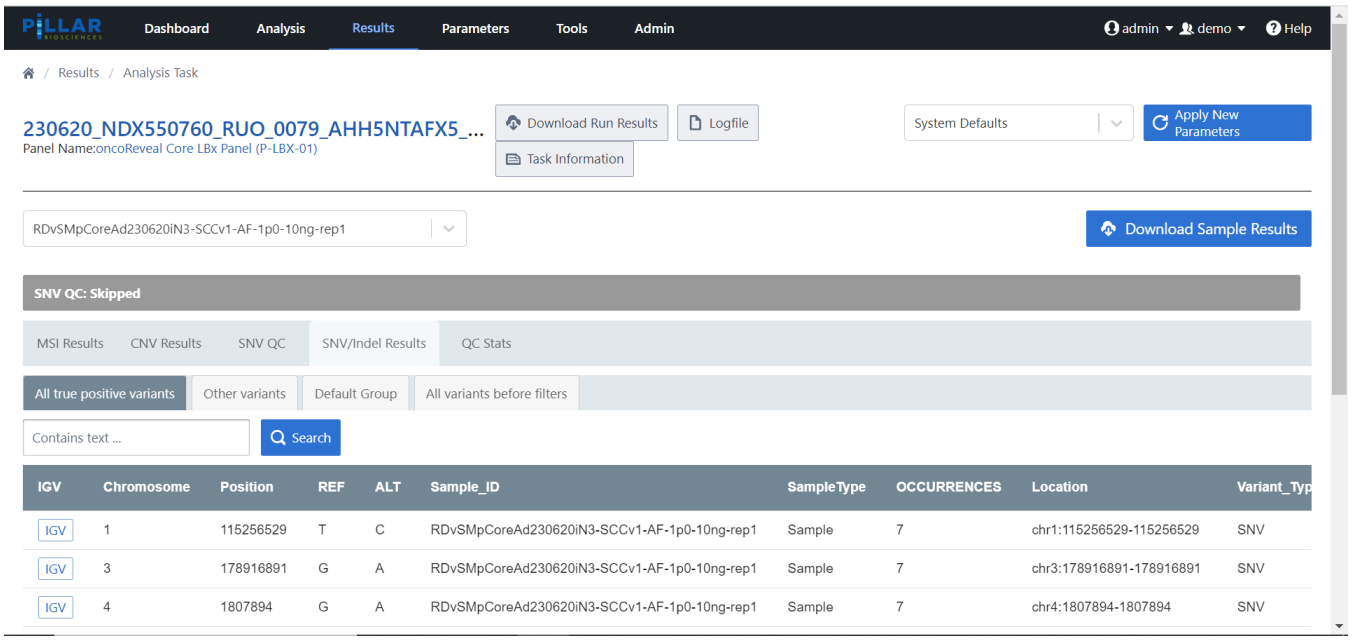


Figure 32: SNV/InDel Results tab

5. Click the **Download Run Results** button download all sample and run results, or click Download Sample Results to download results for the selected sample.

PiVAT Output Content

Figure 33: PiVAT output displayed on the Browse Downloadable Data page

A successfully completed PiVAT run will create a run output folder which contains the following files/folders:

- individual sample folder(s)
    - shown as “HotSpotControl – 1.zip” in the image above
    - contains: bam files, samtools log files
- log folder
    - shown as “log/” in the image above
    - contains all analysis pipeline run logs
- output manifest file
    - named “output_manifest_{PiVAT task ID}”
    - shown as “output_manifest_8” in the image above
    - this file contains, in json format, a record of all the final variant calling output and log files
- completion semaphore file
    - always named “pivat_complete.ini”
    - file is empty
    - used to indicate the PIVAT run has completed successfully
- system folder
    - always named “system/”
    - contains:
        - all data exchange (pickle) files
        - filter_result sub directory, containing data exchange files and filter and QC processing logs.
            - Note, re-run from “apply new parameters” button will create a new filter_result subdirectory for the rerun with new parameters
        - final_report subdirectory, containing all final report files (**pdf, xlsx, and vcf results files)**, additional data exchange files, report log files
            - Note, re-run from “apply new parameters” button will create a new final_report subdirectory for the rerun with new parameters
- zips folder

- always named “zips/”
- contains:
  - all output listed above (sample folders, log folder, systems folder, output manifest file, and completion semaphore file).
  - Note that output folders exist as zip files within the zips folder.

User password

1. A user can change their own password by selecting the **Profile** option from the drop-down menu for the username. The following page is displayed:

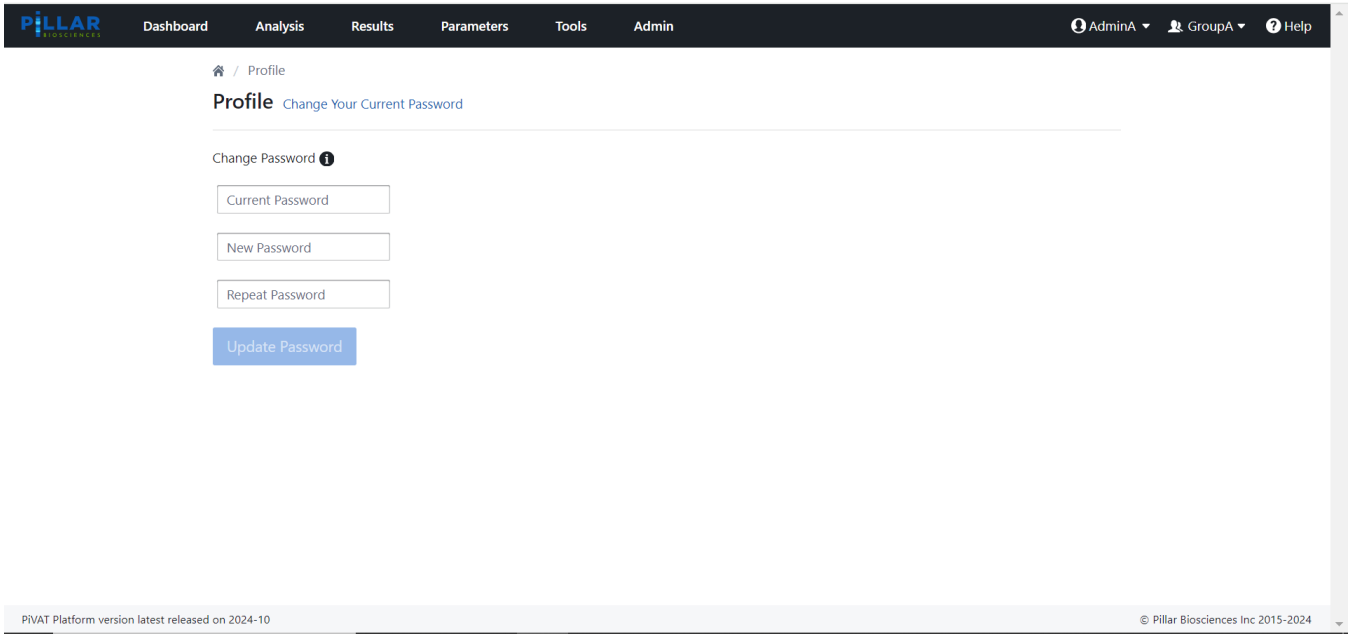


Figure 34: Profile page

2. Create new password and then click **Update Password** to confirm.
3. To exit without changing your password, click or any option from the Navigation Bar.

Record the new password in accordance with your organization’s IT policies.

Data Management

1. Begin by clicking **Tools** on the Navigation Bar and **Data Management** on the left menu bar to display the following Data Management page.
2. The Data Management page is used to upload input data (FASTQ) to the PiVAT system for analysis.
3. Storage data is displayed for verification of sufficient free storage before attempting to upload data.

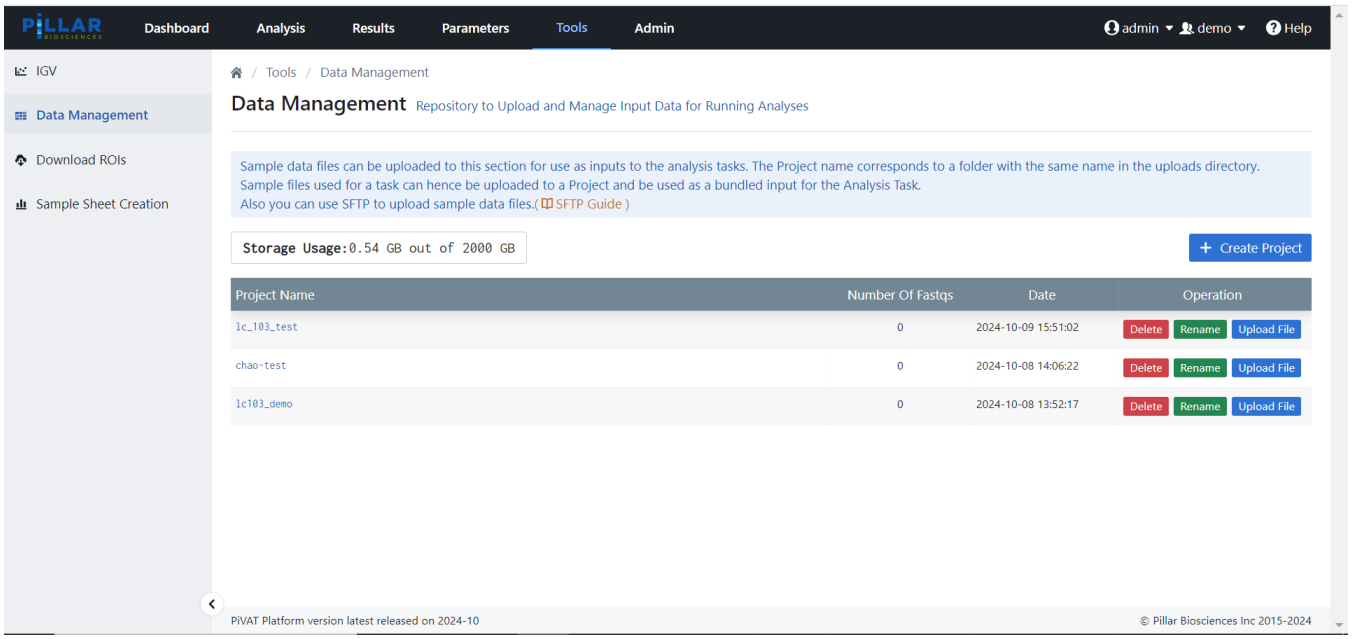


Figure 35: Data Management page

4. Input data can be uploaded to a Project listed on the page. To create a new project, click **+ Create Project**.
5. All projects associated with the current user's User Group are listed on this page. Click the Project Name to which you wish to upload data and the following modal will be displayed:

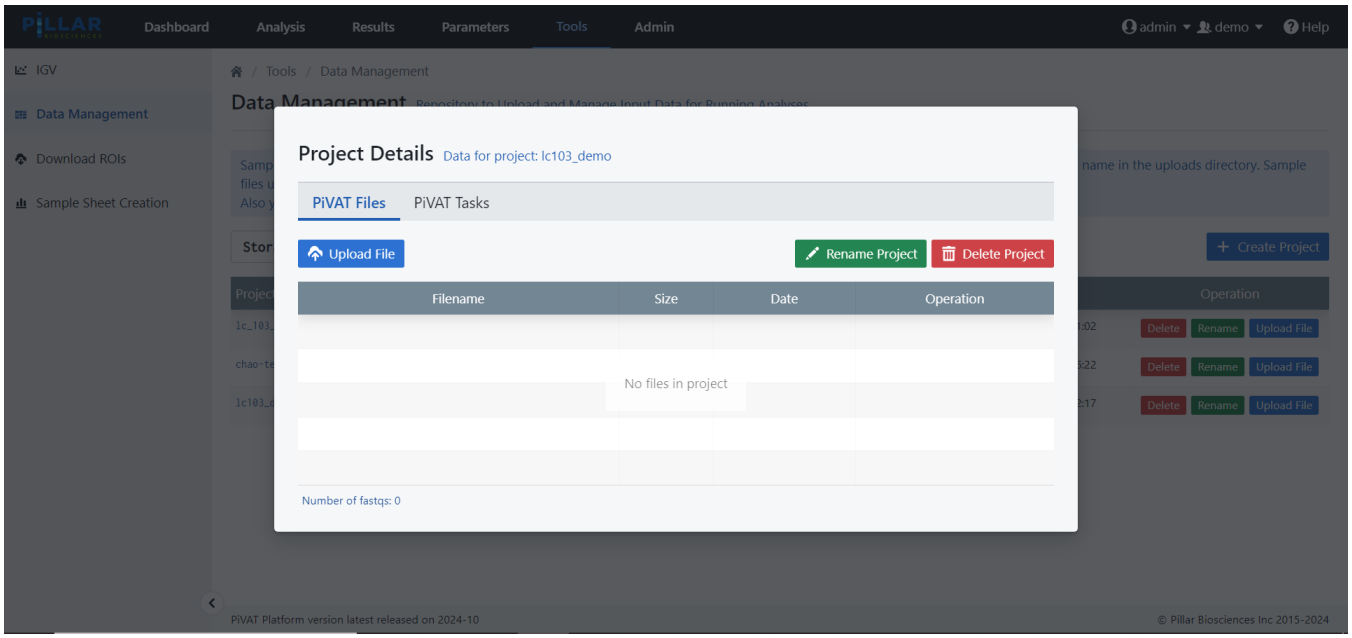


Figure 36: Project Details modal

6. Click the **Upload File** button to display the Upload Input Files page. From here users may drag & drop files or select the files for upload. Once your files are displayed on the page, select **Start Upload**.

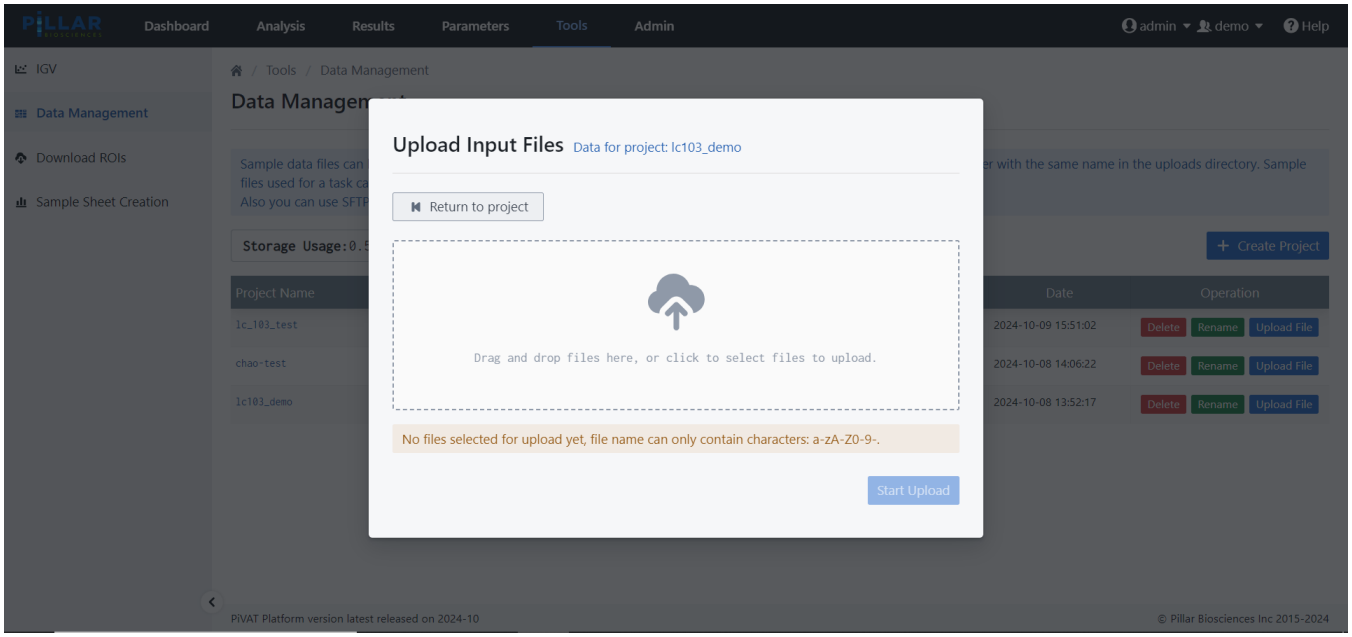


Figure 37: Upload Input Files page

7. Users may also select **Rename** or **Delete** the project from the Project Details page.

- Take caution when deleting a project, as it will result in all input and output data associated with the project to be deleted.
- Selection of directories for upload is only supported by the Drag and Drop option.

SFTP Upload/Download

To download files using SFTP (Secure File Transfer Protocol), your installation of PiVAT must have the SFTP feature enabled by the administrator. Please contact your administrator to confirm if this feature is available. To use the SFTP feature, SFTP client software is required such as FileZilla or Cyberduck. Use the following settings.

Address: sftp://*[IP address of your PiVAT system]*
Port: 2225
Username: Your PiVAT username
Password: Your PiVAT password

When uploading is interrupted, the file number and total size may be incorrect, please logout & login to the SFTP client again so the system can refresh. Deletion of corrupted files will also require you to logout and login again.

Some SFTP clients, FileZilla for example, may run into an infinite loop of retrying an upload when a file exceeds the size limitation. Please terminate it manually.

SFTP Use Example

1. Download FileZilla: https://filezilla-project.org
2. Enter sftp://, followed by the url of the PiVAT system, enter the username and password, as seen below. The first time you log in to that server, you will receive an “Unknown host key” warning. Select “Always trust this host, add this key to the cache” and click “OK”

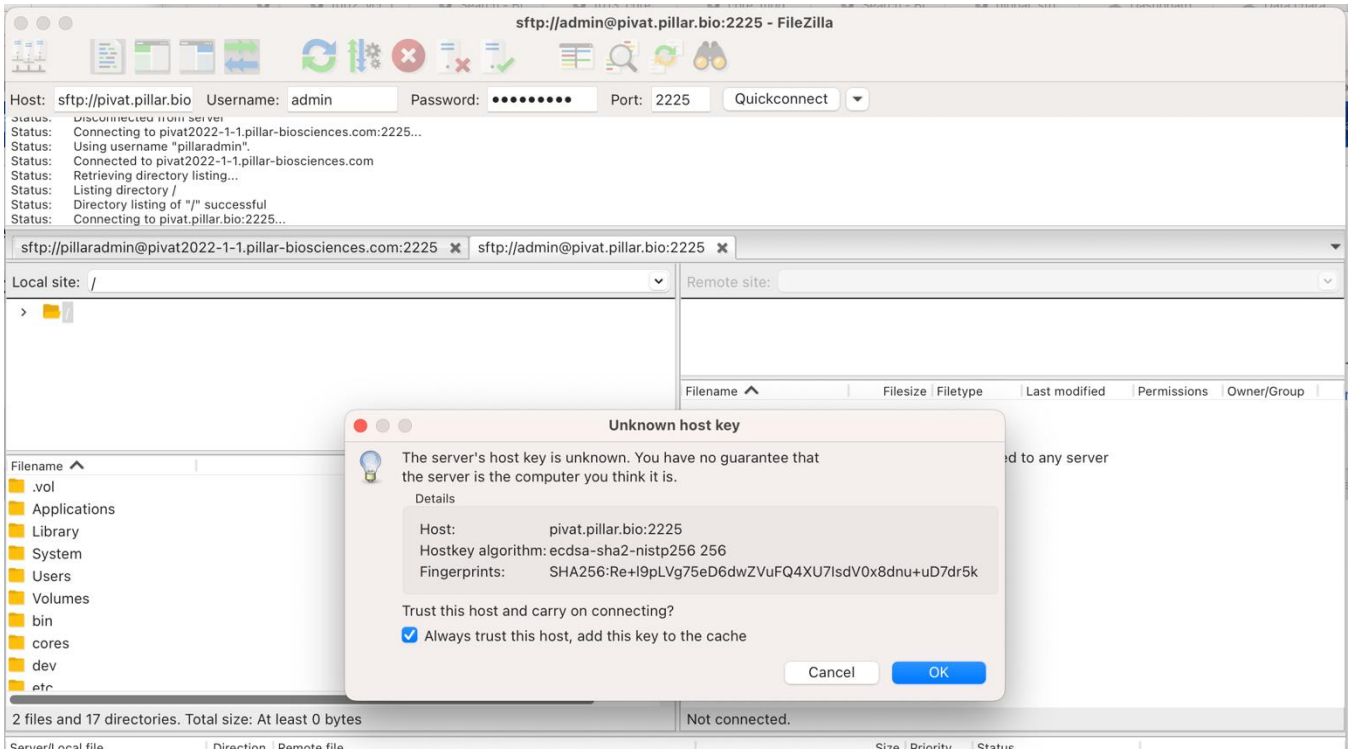


Figure 38: FileZilla example with unknown host key warning

3. To upload a file or folder, drag the file or folder from “Local site” (left side of screen), to a folder on the “Remote site” (right side of screen). For downloads, drag a file or folder from the “Remote site” to a folder on the “local site”.

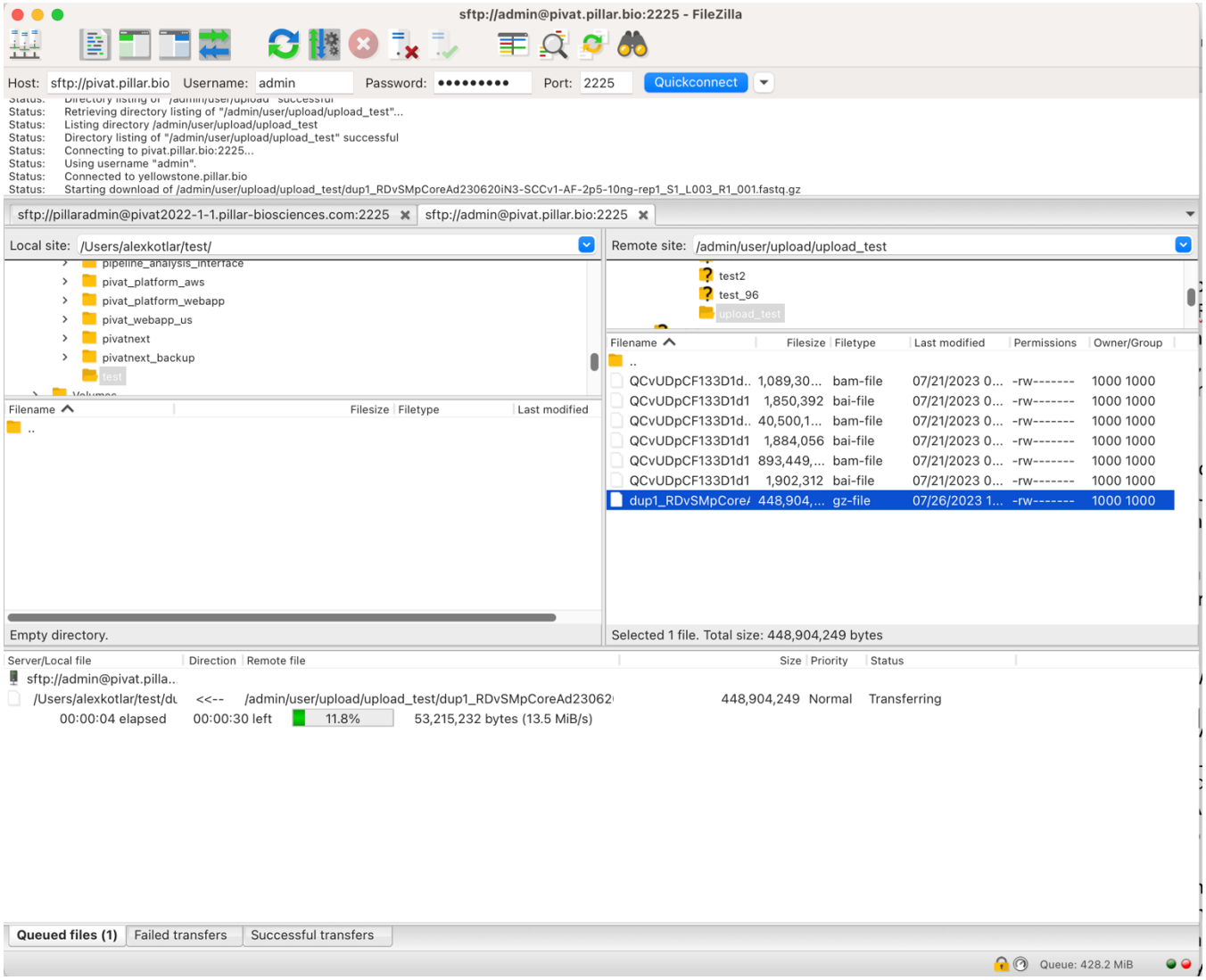


Figure 39: Filezilla file transfer

Troubleshooting

If the system requires a restart, the user should check that all running and queued analysis tasks have completed before restarting the system.

If a task fails due to insufficient disk space, an administrator user can free up space on the system. There must be at least 100 GB of free space on the system for a task to start.

To avoid performance issues, do not run other applications on the workstation while PiVAT is running.

If the PiVAT application will not load, contact Pillar Biosciences at (800) 514-9307 or support@pillarbiosci.com.

If PiVAT is nonresponsive, it is possible to restart PiVAT using the following command:

- Open a terminal window
- cd /pillar/docker_files
- sudo docker-compose down
    - Superuser password required
- sudo docker ps
    - Verify that no docker containers are active
- sudo docker-compose up -d
- sudo docker ps
    - Verify that the required containers are now running.

Legal Notices

Company Information

Pillar Biosciences, Inc.
9 Strathmore Road
Natick, MA 01760

(800) 514-9307
support@pillarbiosci.com
https://pillarbiosci.com

Appendix A: SNV/Indels Caller

The SNV/Indels analysis reports all SNV and indel variant calls that have passed PiVAT's filtering criteria based on parameters set. The output will display true positive SNV and indel variants and annotations.

SNV/Indels Analysis Parameters Setup

Please refer to Creating Analysis Parameters section of this document for instructions on how to create and save analysis parameters.

1. Navigate to ① **Post Filter Parameters** page to access the following list of adjustable parameters and default values in [ ] by tab.

QC Parameters [default]
VCR_FILTER_QUALITY (q-value) Minimum variant quality acceptance. Default value may change depending on the chosen panel for the analysis.
VCR_FEATURE_FILTER [True] Filters variants by the Feature column with NM identifiers. Default is set to True.
VCR_FILTER_QUALITY (q-value) Minimum variant quality acceptance. Default value may change depending on the chosen panel for the analysis.
VCR_FILTER_FREQUENCY (pct) Minimum variant frequency. Default value may change depending on the chosen panel for the analysis.
VCR_FILTER_REPEAT_FREQUENCY (pct) Minimum variant frequency for repeat regions. Default value may change depending on the chosen panel for the analysis.
VCR_FILTER_COVERAGE (count) [50] Minimum total coverage. Default is set to 50.
VCR_FILTER_MIN_RATIO [-2] Minimum variant read direction ratio. -2 implies variant is on the reverse read while reference on the forward read. 0 means no preference for forward or reverse read of variant with respective to the reference. Default is set to -2.
VCR_FILTER_MAX_RATIO [2] Maximum variant read direction ratio. Setting to 2 implies variant is on the reverse read while reference on the forward read. Setting to 0 means no preference for forward or reverse read of variant with respective to the reference. Default is set to 2.
VCR_FILTER_VARIANT_COVERAGE (count) Minimum variant coverage. (Total Coverage * Variant Read Frequency)

QC Parameters [default]
VCR_REPEAT_REGION_QUALITY_FILTER (q-value) [30] Minimum variant quality within a repeat region. Default is set to 30.

PiVAT Output: FINAL_REPORT_CUSTOMER_RESULTS file

The FINAL_RESPORT_CUSTOMER_RESULTS.xlsx file contains SNP/InDel results for all samples within the run. Definition and/or description of result columns reported in CUSTOMER_RESULTS file sheet tabs are provided below.

Sheet Tab(s)	Column Name	Definition/Description
All Variant Report, True Positive Variants	Sample_ID	User defined library ID, derived from FASTQ file name
	SampleType	Type of sample used for analysis. i.e., Sample, NTC, PosCtrl, or NegCtrl
	Chromosome	Chromosome number in human genome where variant occurs
	Position	Base position on given chromosome where variant occurs (hg19)
	REF	Sequence found in human reference genome (hg19)
	ALT	Variant sequence found in sample
	Location	Genomic coordinates of variant (hg19)
	Variant_Type	SNV = single nucleotide variant; Deletion = one or more base deletion; Insertion=one or more base insertion; Delins = "deletion-insertion", variant characterized by both a deletion of one or more bases AND an insertion of one of more bases on the same allele. i.e., AGACTA ---> ATTCTA resulting from deletion of GA and insertion of TT or AGACTA ---> ATCTA resulting from deletion of GA and insertion of T
	VARIANT_CLASS	Sequence Ontology (SO) terms to identify variant type
	Variant_Length	Number of affected bases
	Variant_Net_Length	The net length between the difference of a variant between the length of the reference sequence and the alternate sequence. Variant_Net_Length = (length of alternate sequence – length of Reference sequence)
	Amplicon_ID	Amplicon in assay in which variant occurs.
	Variant_Read_Frequency_(%)	Occurrence of variant in sequencing data, as a percentage of total sequenced reads in the given segment. (Variant_Coverage / Total_Coverage)*100

Sheet Tab(s)	Column Name	Definition/Description
	Variant_Coverage	Number of reads that contain variant in segment.
	Total_Coverage	Total number of reads in segment.
	Variant_Quality	Measure of variant call quality. Score is from 0 (low quality) to 40 (highest quality)
	Variant_Read_Direction_Ratio	Measure of strand bias in variant calling. Value of 0 indicates variant was found on both the positive strand and negative strand at equal rates (ideal). Values greater than zero or less than zero indicate percentage of bias towards the negative strand (negative values) or positive strand (positive values)
	Zygosity	Germline mutation panels only. HETEROZYGOUS = one out of two alleles contain variant; HOMOZYGOUS = both alleles contain variant
	Consequence	Type of mutation. Examples: frameshift_variant = variant produces an mRNA transcript that is out of the normal reading frame. synonymous_variant = variant does not affect amino acid sequence of final protein due to redundancy of codons. missense_variant = variant leads to change in amino acid sequence. inframe_deletion = deletions of 3, or multiples of 3, which remove whole codon sequences. intron_variant = variant occurs in intron. splice_region_variant = variant in splice site, may lead to splicing error in final mRNA transcript. stop_gained = variant results in premature stop codon in mRNA transcript. 3_prime_UTR_variant = variant occurs in 3" untranslated region (UTR)
	Impact	Impact of variant on normal gene function
	Gene_Symbol	Gene name
	Gene_ID	NCBI gene ID number
	Feature	NCBI transcript identifier
	All_Features	List of all NCBI transcript identifiers associated with the variant
	HGVSC	HGVSC variant name. Describes position and variant change in gene (nucleic acid) sequence in Feature
	HGVSP	HGVSP variant name. Describes position and variant change in protein (amino acid) sequence in Feature
	Transcript	The VEP annotated transcript name (e.g., NM_001127500.3)
	c_dot	VEP HGVSc annotation with transcript name removed (e.g., `c.3912C>T`)

Sheet Tab(s)	Column Name	Definition/Description
	p_dot	VEP HGVSp annotation with transcript removed (e.g., `p.Asp1304=`)
	Exon	Affected exon number in Feature out of total number of exons. Applies to variants within exon regions only. (affected exon # / total # of exons)
	Intron	Affected intron number in Feature out of total number of introns. Applies to variants within intron regions only. (affected intron # / total # of introns)
	CDS_Position	cDNA sequence position in Feature
	Protein_Position	Affected amino acid number in Feature
	Amino_Acids	Amino acid resulting from variant (reference amino acid / variant amino acid)
	Codons	Position in codon where variant occurs and resulting codon with variant
	Co-located_Known_Variation	Variants that occur at the same position in publicly available databases
	Strand	Transcribed strand. -1 = negative strand, 1 = positive strand
	Repeat	Number of repeat bases, including repeat sequence (if applicable). i.e. 14G indicates a repeat of 14 Gs
	HGVS_Offset	Correction factor to synchronize genomic base position for indels in repeat regions in negative strand genes. Read alignment uses left-align paradigm (left most base assumed to be indel; first genomic position) while HGVS uses right-align paradigm (right most base assumed to be indel; first cDNA position)
	SIFT	Output from SIFT (v5.2.2). PREDICTION (SIFT score). Predicts impact of variant on protein function. Prediction possibilities: tolerated, deleterious. SIFT score is a range from 0 - 1. Scores > 0.05 are tolerated, scores of < 0.05 are deleterious
	POLYPHEN	Output from PolyPhen (v2.2.2). PREDICTION (PolyPhen score). Predicts impact of variant on protein function. Prediction possibilities: benign, possibly damaging, probably damaging. PolyPhen score is a range from 0 (benign) - 1 (probably damaging) and is the probability of variant leading to a damaging mutation. Benign = variant unlikely to cause damage to protein function. Possibly damaging = variant likely to cause damage to protein function but prediction is with low confidence. Probably damaging = variant likely to cause damage to protein function and prediction is with high confidence.

Sheet Tab(s)	Column Name	Definition/Description
	AF	Allele frequency of variant found in global population (1000 Genomes version Phase 3)
	AFR_AF	Allele frequency of variant found in African populations (1000 Genomes version Phase 3)
	AMR_AF	Allele frequency of variant found in American populations (1000 Genomes version Phase 3)
	EAS_AF	Allele frequency of variant found in East Asian populations (1000 Genomes version Phase 3)
	EUR_AF	Allele frequency of variant found in European populations (1000 Genomes version Phase 3)
	SAS_AF	Allele frequency of variant found in South Asian populations (1000 Genomes version Phase 3)
	gnomAD_AF	Allele frequency of variant found in global population (gnomAD version r2.1)
	gnomAD_AFR_AF	Allele frequency of variant found in African/African American populations (gnomAD version r2.1)
	gnomAD_AMR_AF	Allele frequency of variant found in Latino/Admixed American populations (gnomAD version r2.1)
	gnomAD_ASJ_AF	Allele frequency of variant found in Ashkenazi Jewish populations (gnomAD version r2.1)
	gnomAD_EAS_AF	Allele frequency of variant found in East Asian populations (gnomAD version r2.1)
	gnomAD_FIN_AF	Allele frequency of variant found in European (Finnish) populations (gnomAD version r2.1)
	gnomAD_NFE_AF	Allele frequency of variant found in European (non-Finnish) populations (gnomAD version r2.1)
	gnomAD_OTH_AF	Allele frequency of variant found in Other populations (gnomAD version r2.1)
	gnomAD_SAS_AF	Allele frequency of variant found in South Asian populations (gnomAD version r2.1)
	CLIN_SIG	Clinical significance (ClinVar v201912)
	PiVAT_ClinSig	PiVAT clinical significance based on CLIN_SIG and IMPACT
	Coverage_Mean	Mean base coverage of all bases within the defined ROI; with paired end sequencing, merged paired-end reads (forward and reverse) create a coverage of 1x

Sheet Tab(s)	Column Name	Definition/Description
PBAM Stats Summary	STDEV	Standard deviation of mean base coverage
	Coverage_Median	Median base coverage of all bases within the defined ROI; with paired end sequencing, merged paired-end reads (forward and reverse) create a coverage of 1x
	Coverage_Max	Maximum base coverage of all bases within the defined ROI
	Coverage_Min	Minimum base coverage of all bases within the defined ROI
	Total_Number_Of_Reads	Total number of reads on target for all amplicons
	Total_Valid_Reads	Total number of reads contributing to paired end assembly after filtering
	On_Target_Ratio	The ratio between Total_Valid_Reads and Total_Number_Of_Reads. On_Target_Ratio = (Total_Valid_Reads / Total_Number_Of_Reads)

Appendix B: Microsatellite Instability (MSI) Caller

Microsatellite instability (MSI) is measured using MSIsensor (https://github.com/ding-lab/msisensor) in conjunction with Pillar's custom MSI46 panel. This software compares the distribution of read lengths at 53 microsatellite (MS) sites in the tumor genome with the same sites in a matched normal reference. MSI status (instable—MSI-high, or stable—MSS) is determined by the number of sites which differ.

Pooled vs un-pooled MSI Sample Setup

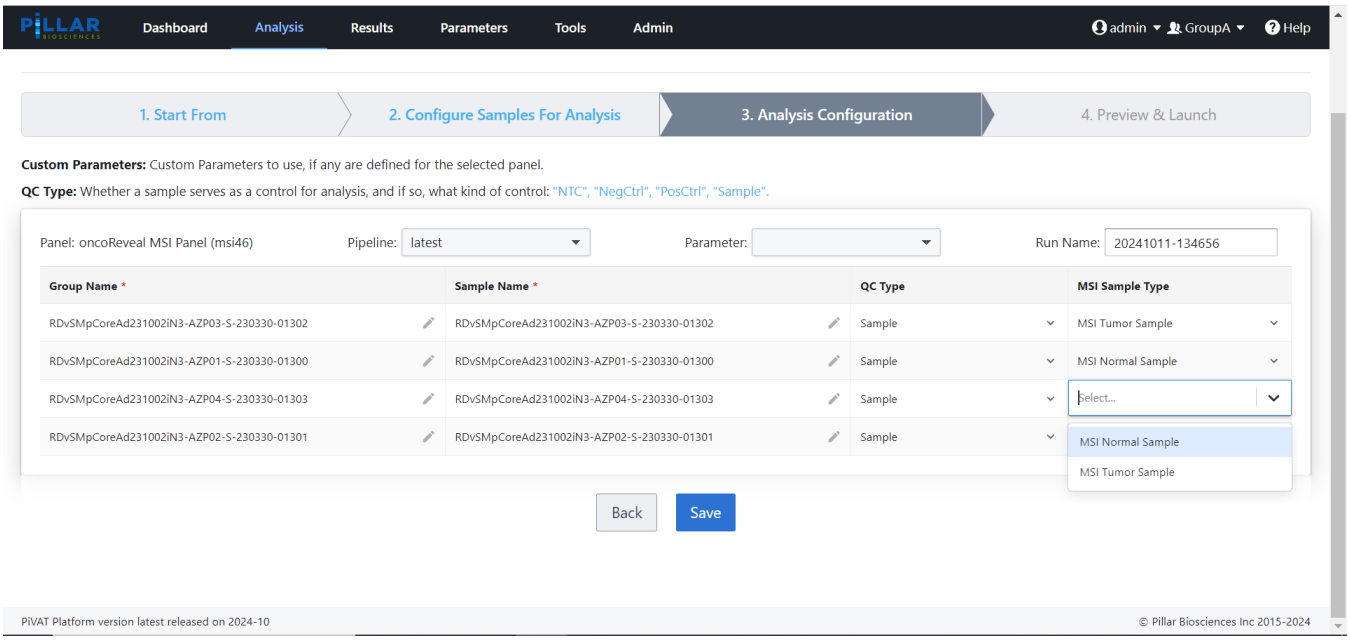
For panels which utilize MSI pooled normal (eg: oncoReveal Core LBx Panel) the user does not need to designate tumor/normal pairing. For panels which use paired tumor/normal MSI calling (eg: oncoReveal MSI Panel) The user must provide a matched normal sample for use as a reference in determining each microsatellite site's baseline read length distribution. Ideally this sample should be prepared with the same methodology and sequenced in parallel with its corresponding tumor sample.

Paired MSI Analysis Setup

This section provides specific instructions to start a paired MSI analysis. Please refer to Start Analysis section on page 16 for general instructions on how to start a PiVAT analysis.

1. Select **Start Analysis** from the Quickstart menu or select **Analysis** from the navigation bar.
2. From the Start Analysis: 1. Start From page select your desired samples.
3. Select Next to proceed.
4. From the Start Analysis 2: Configure Samples for Analysis page, select the **MSI46** panel (or other panel that includes paired MSI analysis) from the drop-down menu and select the desired samples to include in analysis.
5. Select **NEXT** to proceed.
6. From the Start Analysis: 3. Analysis Configuration page enter desired **Run Name**.

7. For each tumor/normal sample select the appropriate MSI Sample Type from the drop down menu.



8.

Figure 40: Start Analysis: 2. Edit Definition page – define MSI tumor, MSI normal samples.

9. Once all the samples have been paired select **NEXT**.

10. Review the analysis input and select Launch Analysis to begin the run. The task status may now be monitored on the PiVAT Dashboard.

PiVAT Output: MSI Results

MSI results are included in the CUSTOMER_REPORTS and SAMPLE_REPORTS xlsx files

Definition and/or description of result columns reported in MSI_RESULTS file sheet tabs are provided below.

Sheet Tab	Column Name	Definition/Description
MSI Call Report	**Tumor_Sample**	Unique Sample ID for each MSI tumor sample
	Normal_Sample	Unique Sample ID for each MSI normal sample (in a matched pair for un-pooled MSI).
	Tumor_Sample_Type	The sample type associated with the MSI tumor sample.
	Normal_Sample_Type	The sample type associated with the normal sample from a matched pair
	MSI_Call	Whether the tumor sample exhibits microsatellite instability (True = MSI-High, False = MSS)
	Total_Number_of_Sites	The number of MS sites assessed for a tumor/normal pair
	Number_of_Somatic_Sites	The number of MS sites that are different between the tumor and the matched normal sample

Sheet Tab	Column Name	Definition/Description
	Somatic_Sites_Ratio	The ratio of differing MS sites to total MS sites assessed
Read Count Distribution	**Tumor_Sample**	Unique Sample ID for each tumor sample in a matched pair
	Normal_Sample	Unique Sample ID for each normal sample in a matched pair
	Tumor_Sample_Type	The sample type associated with the tumor sample from a matched pair
	Normal_Sample_Type	The sample type associated with the normal sample from a matched pair
	Target_Name	The amplicon ID covering the MS site
	Chromosome	The chromosome the MS site is located on
	MSI_Start	The location of the MS start site (hg19)
	Left_Flank	The 5-bp sequence to the left (upstream) of the MS site
	Repeat_Unit_Bases	The reference number of repeats and the [repeating sequence unit] comprising the MS site
	Right_Flank	The 5-bp sequence to the right (downstream) of the MS site
	Number_of_Repeat_Units	The number of possible repeating units (1-100)
	Tumor_Read_Count	The number of reads in the tumor sample with a repeat length specified by **Number_of_Repeat_Units** field
Somatic Sites	**Normal_Read_Count**	The number of reads in the normal sample with a repeat length specified by **Number_of_Repeat_Units** field
	Tumor_Sample	Unique Sample ID for each tumor sample in a matched pair
	Normal_Sample	Unique Sample ID for each normal sample in a matched pair
	Tumor_Sample_Type	The sample type associated with the tumor sample from a matched pair
	Normal_Sample_Type	The sample type associated with the normal sample from a matched pair
	Target_Name	The amplicon ID covering the MS site
	Chromosome	The chromosome the MS site is located on
	MSI_Start	The location of the MS start site (hg19)
	Left_Flank	The 5-bp sequence to the left (upstream) of the MS site
	Repeat_Times	The reference number of repeats of the repeating sequence unit
	Repeat_Unit_Bases	The bases comprising the repeating unit
	Right_Flank	The 5-bp sequence to the right (downstream) of the MS site
	Difference	The difference score (0-1) assigned to the MS site between tumor and normal

Sheet Tab	Column Name	Definition/Description
	P_Value	The p-value associated with the difference score
	FDR	The FDR-adjusted p-value associated with the difference score
	Rank	The rank, by p-value, of the MS site
Germline Sites	**Tumor_Sample**	Unique Sample ID for each tumor sample in a matched pair
	Normal_Sample	Unique Sample ID for each normal sample in a matched pair
	Tumor_Sample_Type	The sample type associated with the tumor sample from a matched pair
	Normal_Sample_Type	The sample type associated with the normal sample from a matched pair
	Target_Name	The amplicon ID covering the MS site
	Chromosome	The chromosome the MS site is located on
	MSI_Start	The location of the MS start site (hg19)
	Left_Flank	The 5-bp sequence to the left (upstream) of the MS site
	Repeat_Times	The reference number of repeats of the repeating sequence unit
	Repeat_Unit_Bases	The bases comprising the repeating unit
	Right_Flank	The 5-bp sequence to the right (downstream) of the MS site
	Genotype	The repeat lengths of alleles in the Normal sample

Appendix C: Somatic CNV Caller

There are two copy number variation (CNV) callers implemented within PiVAT, one for cell-free DNA (cfDNA) panels and the other for non-cfDNA panels, i.e., formalin-fixed paraffin-embedded (FFPE) and tissue. **Exon-level calls are only made for the BRCA1/BRCA2 genes in non-FFPE tissue samples**; for the rest of the sample types, calls are restricted to gene-level only. Both variant calling algorithms work on detecting deviations in the log copy ratios of amplicon coverage when compared with a normal sample(s) as a baseline. A Dirichlet process-based algorithm is used for cfDNA samples, and a state space model is used for non-cfDNA samples.

CNV Sample Setup

FFPE/Tissue CNV caller: For the best FFPE/tissue CNV caller performance, **3 to 5 (minimum 2) normal samples**, which are used to normalize calls, are recommended. These normal samples should have similar sample conditions and preparation processes to the test sample. See the section below for the recommended setting for CNV analysis.

cfDNA CNV caller: The cfDNA CNV caller requires **at least one normal sample** for call normalization. The quality of the calls improves with increasing numbers of normal samples. **Please note** that the clinical normal samples work best as normal reference samples. In the absence of clinical normal samples, reference standards may be used. However, they tend to lower CNA calling sensitivity and specificity.

For both FFPE/Tissue and cfDNA CNV callers, normals should closely match the background of the putative non-normal samples under test. **Please refer to your panel user manual for panel-specific configuration and usage recommendations.**

CNV Analysis Setup

This section provides specific instructions to start a CNV analysis. Please refer to Start Analysis section on page 16 for general instructions on how to start a PiVAT analysis.

1. Select Start Analysis on the left menu bar.
2. From the Start Analysis: 1. Start From page, select your samples
3. Select Next to proceed.
4. From the Start Analysis: 2. Configure Samples for Analysis page, select a CNV panel from the drop-down menu and select the desired samples to include in the analysis.
5. Select Next to proceed.
6. From the Start Analysis: 3. Analysis Configuration page, enter desired **Analysis Name**.
7. The samples you have selected will appear in the table as shown below.
8. To define normal samples select “CNV Normal Sample” from the CNV Normal drop down menu.

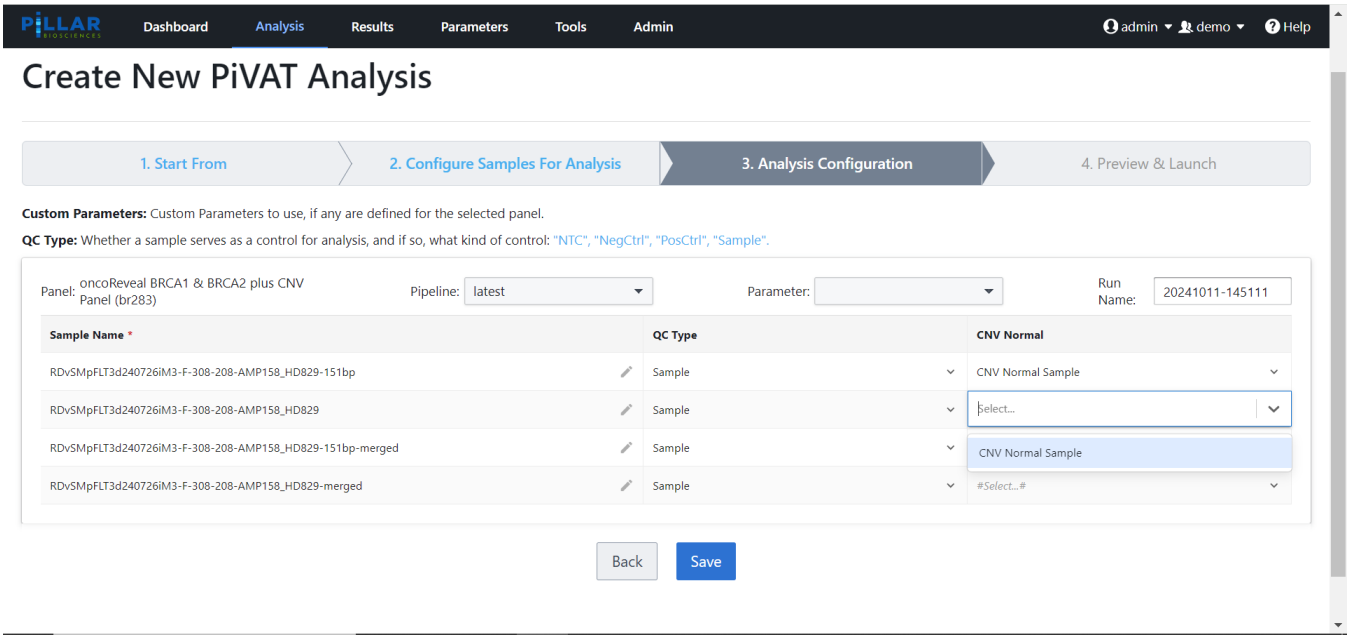


Figure 41: Analysis Configuration page: Set CNV Normal samples

9. If the option is present, **Tumor Content %** should be specified if known to improve performance. In the **Tumor Content %** column enter a value from 1e-8 to 100, representing the tumor content as a percentage (%) for the sample on that row. If this is left empty the tumor content will be given a default based on the panel.

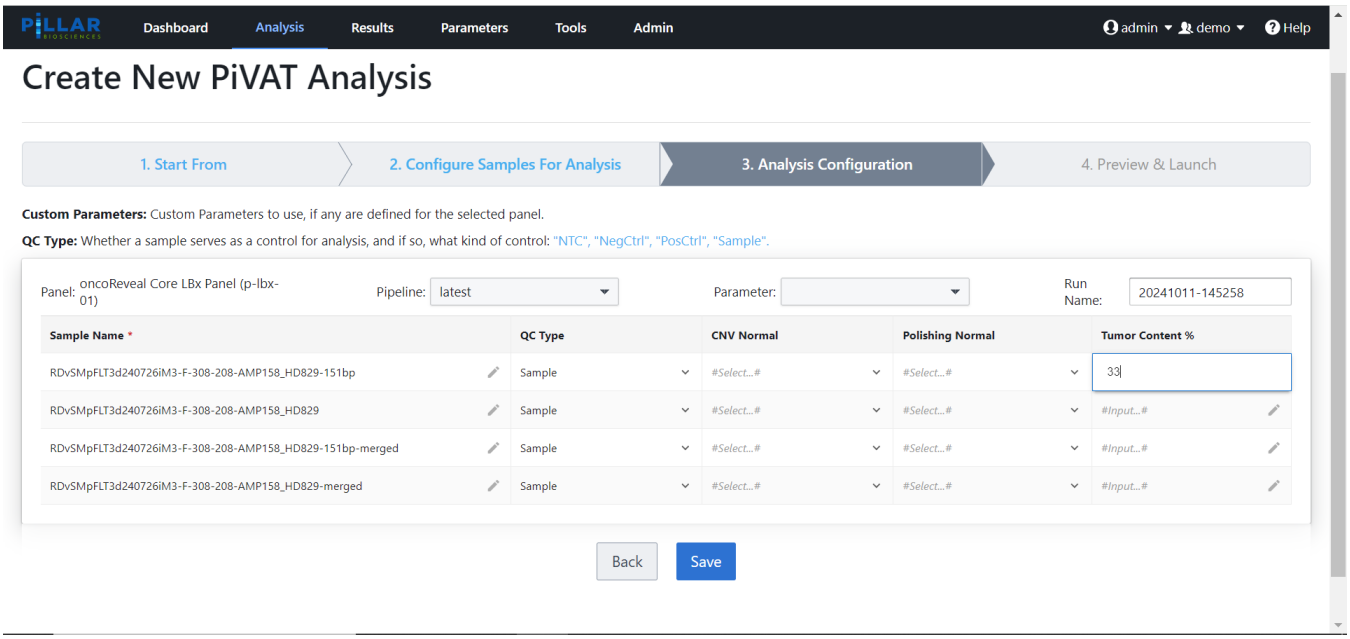


Figure 72: Analysis Configuration page – Set Tumor content

10. **For FFPE and Tissue panels:** If no normal negative reference samples are provided, at least 10 potentially positive samples are required to run CNV analysis. **Note that samples that fail CNV QC will not be included**

toward the 10 sample requirement. In this setting, the CNV analysis will run with a median percentile normalization algorithm. The sample setting configuration will be automatically detected by PiVAT. No user parameter adjustment is needed. In addition, the number of samples with the same CNV type cannot exceed 30% of all samples. We recommend pooling samples with different CNVs (CNVs with different exons/genes/lengths) in this setting. **Please refer to the panel user manual for specific recommendations for that panel.**

11. The normalized copy number ratio can be found in 'CNV Calls' sheet in run and sample level RESULTS excel file. Note that the "Copy Number Ratio" in PiVAT is defined as the copy number ratio of a potentially positive sample to that of the negative reference samples (diploid with 2 copy number).
12. Select Launch Analysis on the Preview and Launch page after confirming your analysis settings.

CNV parameters

The following is a list of user adjustable QC parameters and default values for CNV analysis. Please note these are the **PiVAT** default values, rather than the default values for your **panel**. Please consult your panel user manual for panel-specific defaults and recommendations, before defining custom parameters.

1. **CNV_P_VALUE_THRESHOLD**

   A potential CNV must have a P-Value lower than this threshold to be called. Values are numbers from 0 to 1.

CNV Threshold Grid

The CNV Threshold grid allows the user to adjust the parameters around which CNVs are called. The adjustable parameters are:

1. **Copy Gain Threshold**
   A potential Copy Gain must have a copy number greater than the Copy Gain Threshold to be called.
2. **Enable Copy Gain**
   Enable Copy Gain must be toggled on to call copy gains.
3. **Copy Loss Threshold**
   A potential Copy Loss must have a copy number less than the Copy Loss Threshold to be called.
4. **Enable Copy Loss**
   Enable Copy Loss must be toggled on to call copy losses.

5. **Amplicon Count Threshold**
   Copy gain/loss must be observed in a number of amplicons greater or equal to this threshold.

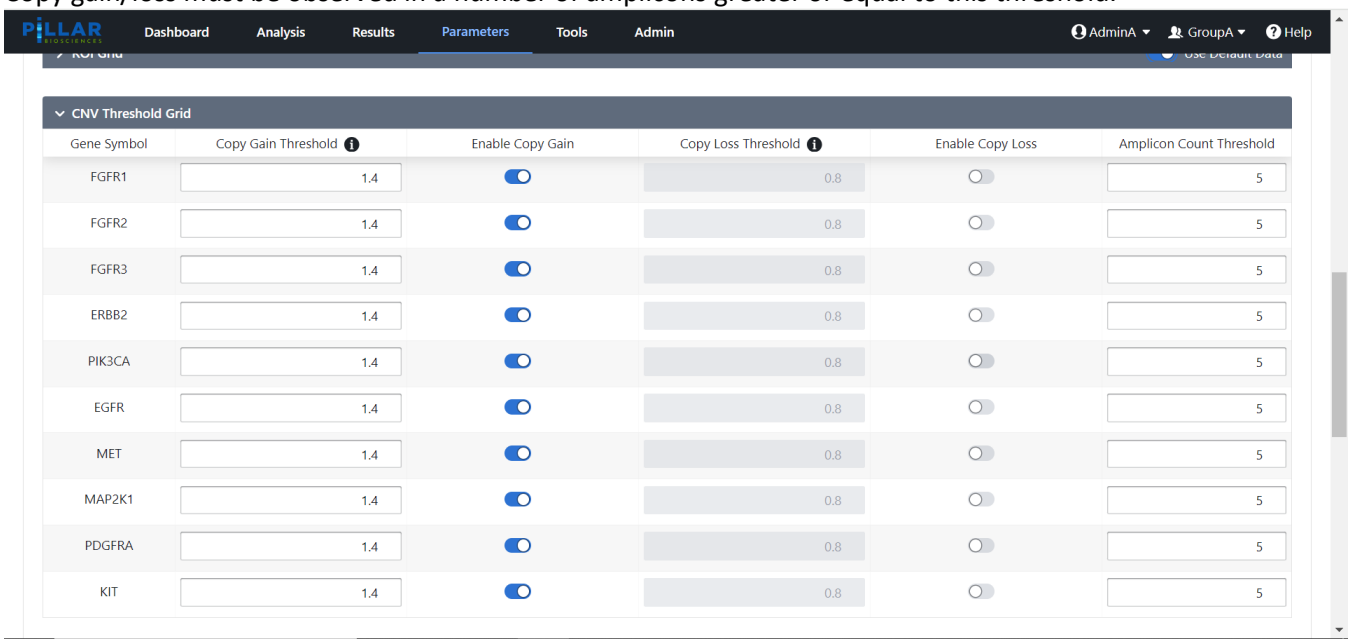


Figure 43: CNV Threshold Grid

PiVAT Output: CNV_RESULTS file

Definition and/or description of result columns reported in CNV Calls tab of FINAL_REPORT xlsx files are provided below.

Sheet Tab	Column Name	Definition/Description
CNV Calls	**Sample ID**	Unique Sample ID for each sample.
	Gene ID	Simplified gene name for the called CNV segment.
	Copy Number Ratio	The ratio of the copy number of called CNV segment to the copy number (2) of a negative normal sample.
	Relative Gain/Loss	Relative Gain or Loss label of the called CNV segment. If the copy number ratio is more than the copy gain threshold, it is labeled as "Gain". If it is less than the copy loss threshold it is labeled as "Loss".
	Std Dev	Standard deviation calculated from the copy number ratios of the amplicons in the called CNV segment.
	P-Value	The probability of the CNV segment being CNV neutral.
	Amplicon Count	Count of amplicons in the called CNV segment.

Sheet Tab	Column Name	Definition/Description
	CNV Start	The estimated start location of the called CNV segment.
	CNV End	The estimated end location of the called CNV segment.
	Final_Call	The CNV call (either Copy Gain, Copy Loss, or Neutral)
Filtered CNV Samples	**Sample ID**	Unique Sample ID for each sample
	Filter Reason	The reason that the sample is filtered out from the CNV analysis.
	Sample Type	Indicate whether the reported sample is “Sample” or “Negative Reference”. This is pre-defined by the user before starting the SA analysis.
CNV Segment Coverages	This sheet reports the raw segment coverages of each amplicon for all the samples initially added to the CNV analysis. The column headers are Amplicon Names and the row indices are Sample IDs.	

CNV Calls Table

The CNV Calls table contains the CNV calls for the sample selected**.**

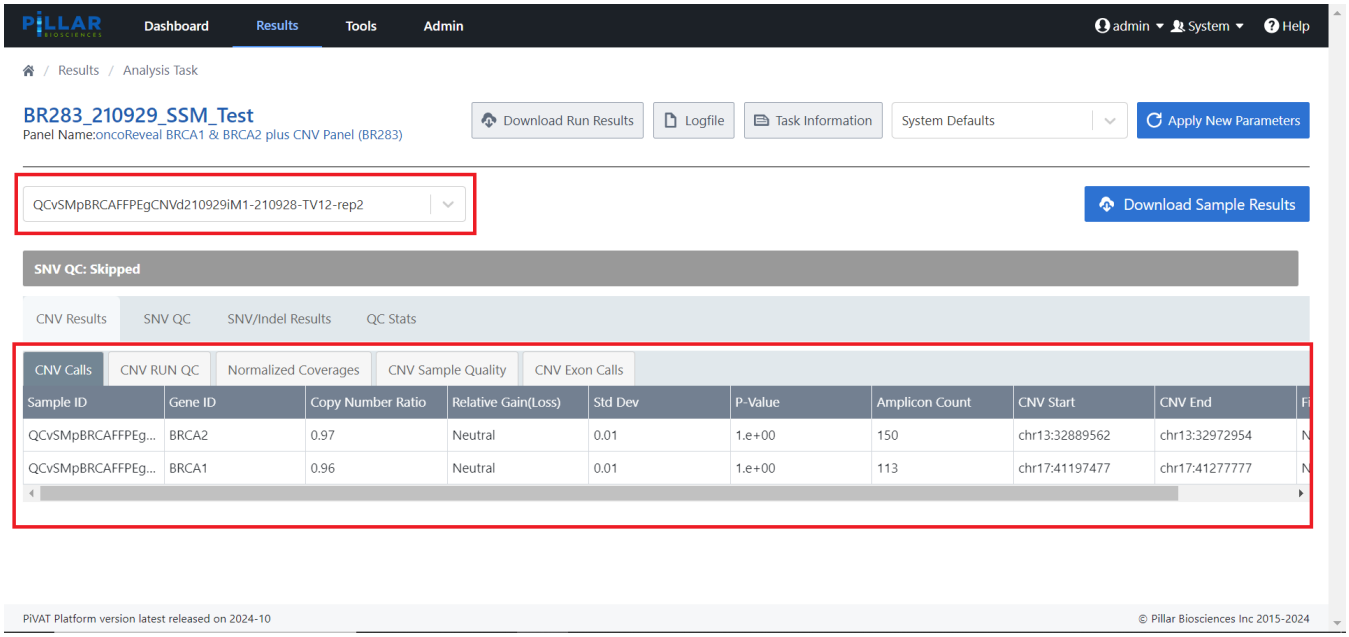


Figure 44 Analysis Results: CNV Calls Table

Definition and/or description of result columns reported in the CNV Calls table are provided below.

Column Name	Definition/Description
Sample ID	Unique Sample ID for each sample.
Gene ID	Simplified gene name for the called CNV segment.
Copy Number Ratio	The ratio of the copy number of called CNV segment to the copy number (2) of a negative normal sample.
Relative Gain/Loss	Relative Gain or Loss label of the called CNV segment. If the copy number ratio is more than 1, it is labeled as “Gain”. Otherwise, it is labeled as “Loss”.
Std Dev	Standard deviation calculated from the copy number ratios of the amplicons in the called CNV segment.
P-Value	**For FFPE / Tissue CNV caller:** P-value of the called CNV segment. **For cfDNA caller**: 0 if cfDNA CNV Caller considers the copy number to be non-neutral, or 1 otherwise.
Amplicon Count	Count of amplicons in the called CNV segment.
CNV Start	The estimated start location of the called CNV segment.
CNV End	The estimated end location of the called CNV segment.
Exon List	The list of the exons in the called CNV segment (only available for BRCA1/2 genes)

CNV Plot

A scatter plot will be displayed for each selected CNV sample under the Normalized Coverages tab. Each point represents the copy number ratio of a target.

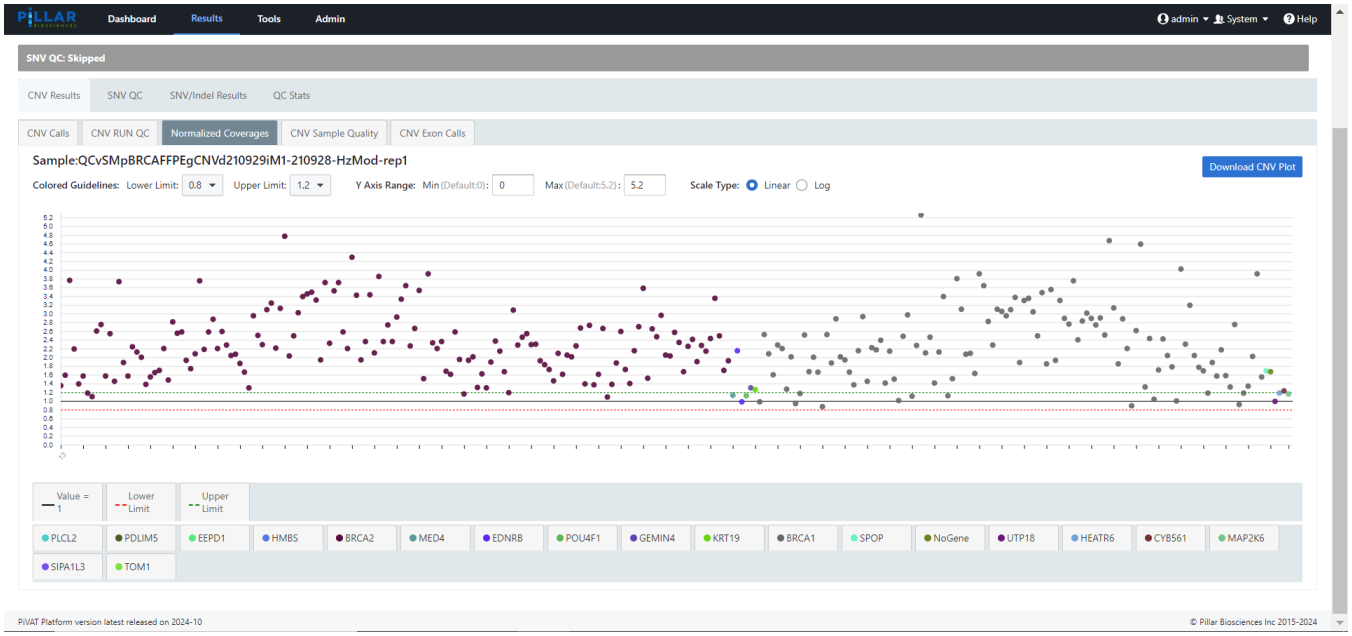


Figure 45: Analysis Results CNV Plot

Genes can be toggled on/off by clicking the name of the gene in the legend. In the example below, the **BRCA1** gene has been hidden.

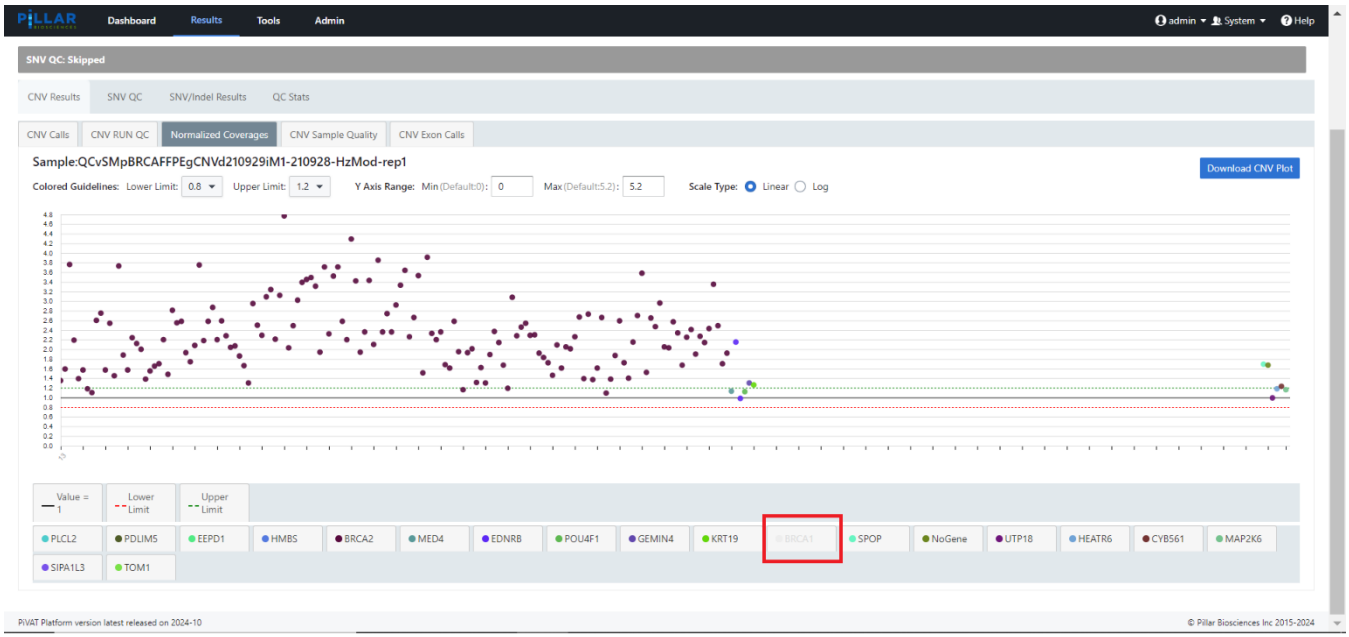


Figure 46: Analysis Results CNV Plot – Toggle Gene

The plot can be saved as a PNG by clicking the blue **Download CNV Plot** button. Any change made to the plot will be reflected in the saved image.

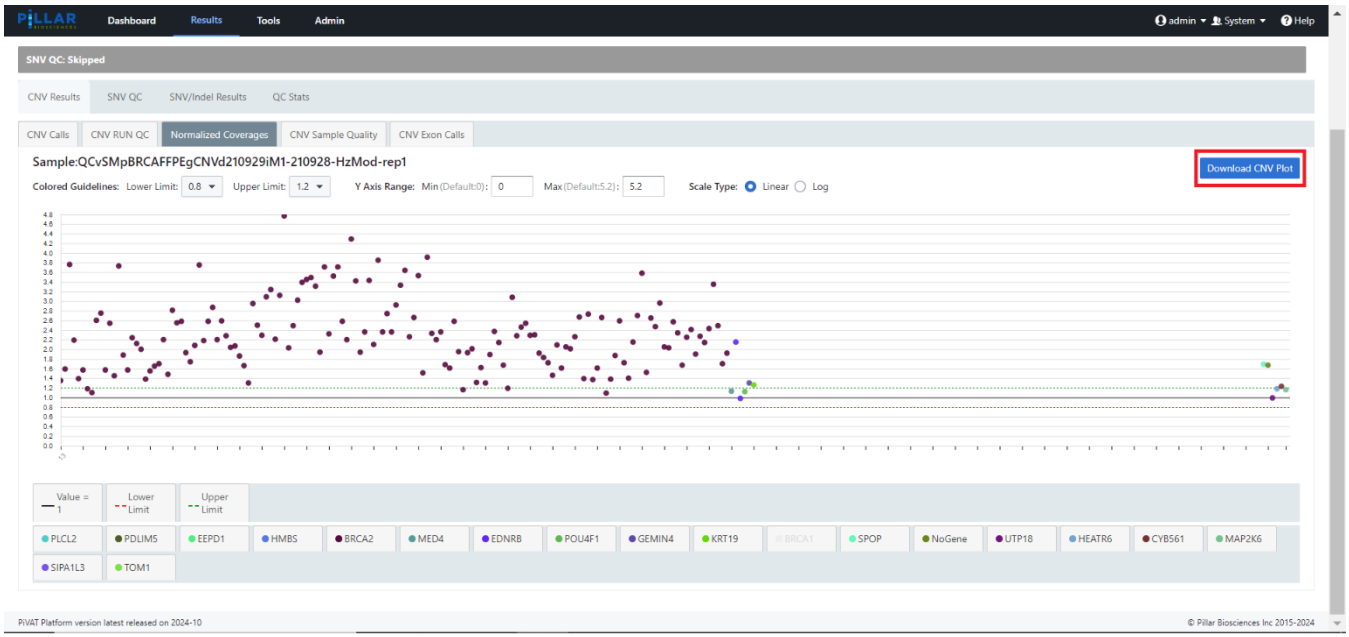


Figure 47: Analysis Results CNV Sample Plot – Save Plot as PNG

Note that the plot shows the copy number ratios of all targets in the genes of interest for the panel. The data points on the plot do not get filtered out by the selected preset filter.

Appendix D: Thalassemia CNV Caller

The Thalassemia analysis involves comparing the sample to one or more negative reference samples. A state space model is used to estimate the log copy ratios at each amplicon, which helps reduce noise in the raw empirical data. To determine the Thalassemia type, the estimated copy ratios are then compared to a list of known Thalassemia types and the best candidate is chosen using statistical decision theory..

Thalassemia CNV Sample Setup

Please refer to you Thalassemia panel user manual for details on the number of normals for call normalization to use.

See **Error! Reference source not found.** section in Appendix C: Somatic CNV Caller for instructions to define CNV normal samples.

Thalassemia CNV Analysis Setup

1. See CNV Analysis Setup section in Appendix C: Somatic CNV Caller for instructions to setup a Thalassemia CNV analysis.
2. The THAL_RESULTS excel file is output upon completion of the analysis.
3. Detailed Thalassemia call information is in the "Thalassemia Calls" sheet of the THAL_RESULTS excel file.
4. The fully normalized copy number ratios can be found in "Fully Normalized" sheet in THAL_RESULTS excel file. Note that the copy number ratio in PiVAT is defined as the copy number ratio of a potentially positive sample to that of the negative reference samples (diploid with 2 copy number).

Thalassemia CNV QC parameters

See CNV parameters section in Appendix C: Somatic CNV Caller for a list of user adjustable QC parameters.

PiVAT Output: THAL_RESULTS file

Thalassemia results are found in the **CNV Sample Diagnosis** tab of the run level FINAL_REPORT excel file.

Definition and/or description of result columns reported in FINAL_REPORT excel file at the run level. Sample level results include only the "CNV Calls" tab.

Sheet Tab	Column Name	Definition/Description
CNV Sample Diagnosis	**Sample ID**	Unique Sample ID for each sample.
	Diagnosis	Inferred Thalassemia diagnosis
	Copy Number Ratio	The ratio of the copy number of called CNV segment to the copy number (2) of a negative normal sample.
	Probability	Indicate whether a sample is defined as "Sample" or "Negative_Reference" by the user.
	Log Evidence	Unnormalized posterior probability of the **Diagnosis.** Also known as Bayseian evidence or normalizing constant. Useful for debugging.

Sheet Tab	Column Name	Definition/Description
	CNV Start	Genomic position of start of observed CNV
	CNV End	Genomic position of end of observed CNV
CNV Segment Coverage	This sheet reports the raw segment coverages of each amplicon for all the samples initially added to the Thalassemia analysis. The column headers are Amplicon Names and the row indices are Sample IDs.	
Filtered CNV Samples	**Sample ID**	Unique Sample ID for each sample
	Filter Reason	The reason that the sample is filtered out from the CNV analysis.
	Sample Type	Indicate whether the reported sample is "Sample" or "Negative Reference". This is pre-defined by the user before starting the SA analysis.

Thalassemia Type Calls Table

Thalassemia Type is shown under the CNV Results; CNV Sample Diagnosis tab in the **Diagnosis** column. Potential alpha Thalassemia types include:

Alpha Diagnosis Name	Thalassemia Profile
a_-_3.7del	Alpha 3.7 deletion
a_-_4.2del	Alpha 4.2 deletion
a_-_3.7dup	Alpha 3.7 duplication
a_-_4.2dup	Alpha 4.2 duplication
Alpha Deletion (2.7k or 2.4k or 1.2k)	Alpha deletion of 2.7, 2.4, or 1.2
Alpha Deletion (MED or SEA or FIL or THAI or 61.9k or 44.6k or 30.8k or 27.6k or 20.5k)	Alpha subtypes of phillipines, southeast asian, Thai, or other listed ddeletions
Alpha Duplication (186.8k or 69.4k)	Alpha duplication, subtypes 186.8 or 69.4
Alpha Duplication (134.8k or 121.2k)	Alpha duplication, subtypes 134.8 or 121.2

We also allow for specific combinations of two types to be present. Such combinations will be denoted with a / (i.e. a_-_3.7del/a_-4.2dup would be an alpha 3.7 deletion and alpha 4.2 duplication). The allowed combinations are alpha 3.7 del/dup and alpha 4.2 del/dup with any of the remaining subtypes.

Beta subtypes include

BetaDiagnosis Name	Thalassemia Profile
b_FilipenoDel	Beta deletion, phillipines subtype
b__118kDel	Beta deletion, 118k subtype
b_204Dup	Beta duplication, 204k subtype

B_ChineseDel	Beta deletion, Chinese subtype
b_67.8Dup	Beta deletion, 67.8k subtype
b_GermanDel	Beta deletion, Germansubtype
b_SEA-HPFHDel	Beta deletion, South East Asian subtype
b_BlackDel	Beta deletion, Black subtype
b_IndianDel	Beta deletion, Indiansubtype
b_TurkishDel	Beta deletion, Turkishsubtype
b__21kDel	Beta deletion, 21.3k subtype
b-TaiwaneseDel	Beta deletion, Taiwanese subtype
b__7.3kDel	Beta deletion, 7.3k subtype

These are only supported as heterozygous variations.

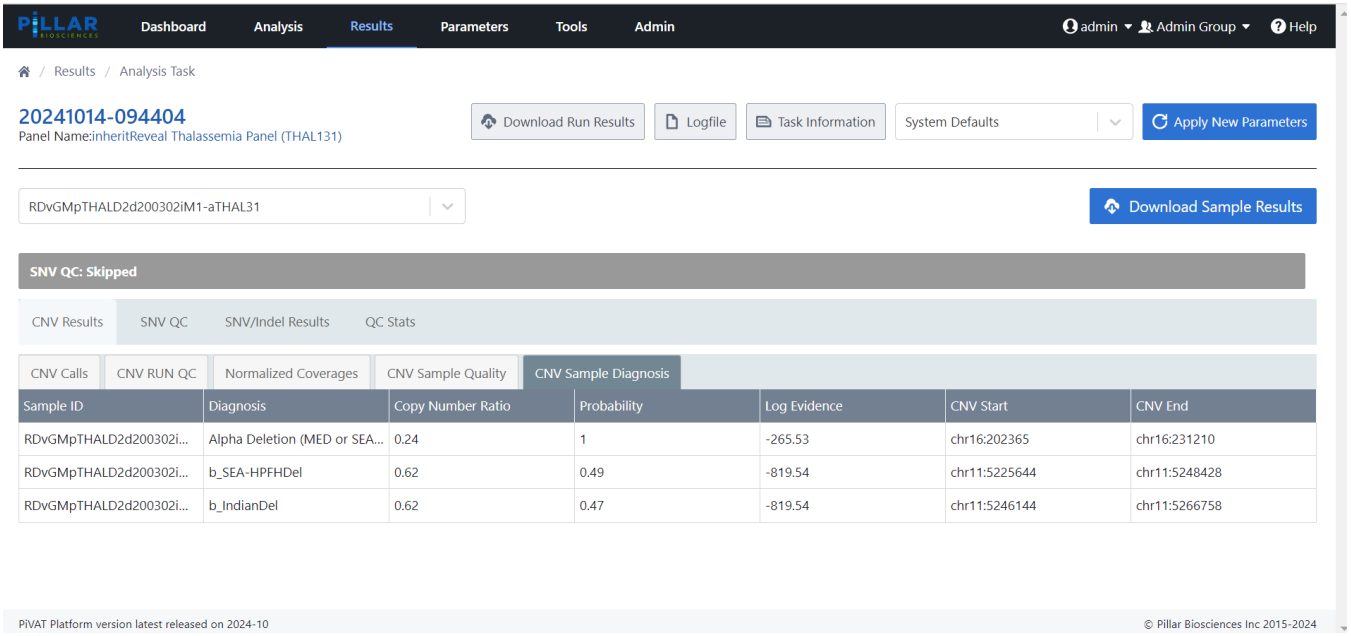


Figure 48 Analysis Results: Thalassemia Type Calls Table

Definition and/or description of result columns reported in the CNV Sample Diagnosis table are provided below.

Column Name	**Definition/Description**
Sample ID	Unique Sample ID for each sample.
Diagnosis	The Thalassemia type detected by the “Thal Type Caller” for a sample.
Copy Number Ratio	The ratio of the copy number of called Thalassemia CNV to the copy number (2) of a negative normal sample.

P-Value	P-value of the called Thalassemia CNV.
Relative Gain/Loss	Relative Gain or Loss label of the called CNV segment. If the copy number ratio is more than 1, it is labeled as “Gain”. Otherwise, it is labeled as “Loss”.
Chromosome	The chromosome the Thalassemia type call is located on.
CNV Start Min	The minimum start position of the detected thalassemia.
CNV Start Max	The maximum start position of the detected thalassemia.
CNV End Min	The minimum end position of the detected thalassemia.
CNV End Max	The maximum end position of the detected thalassemia.

Thalassemia Plot

See CNV Plot section in Appendix C: Somatic CNV Caller for a description of CNV/Thalassemia sample plots, and how to save plots as PNG files.

Note that for the Thalassemia plot, the copy number ratios are grouped by chromosome and not by the target genes. The chromosomes can be toggled to show or hide the data points, similar to the CNV plot.

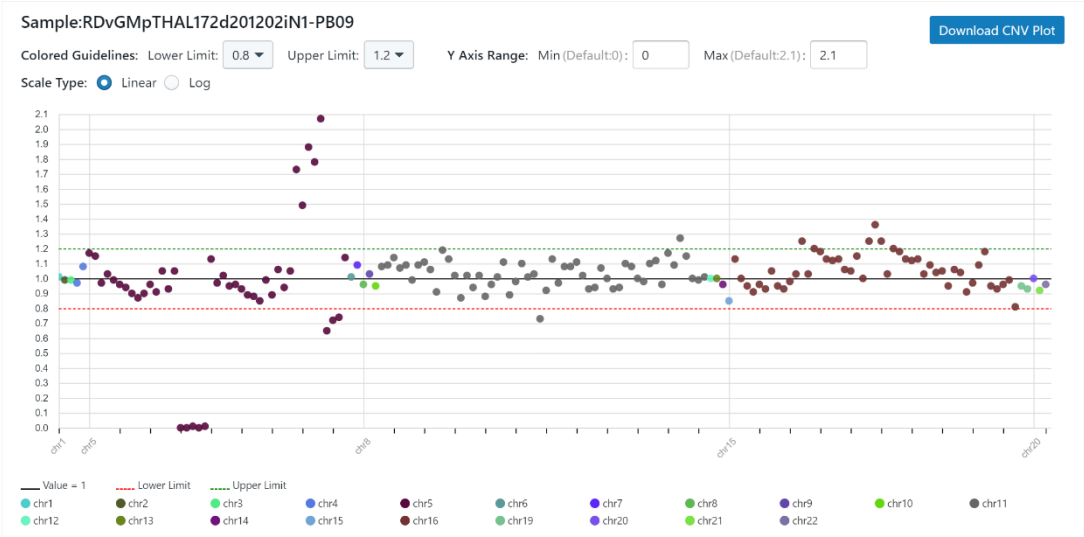


Figure 49: Thalassemia plot

Appendix E: SMA Caller

The SMA analysis is based on the double normalization method. The normalization baseline is calculated from negative reference samples.

SMA Sample Setup

Please refer to you SMA panel user manual for details on the number of normals for call normalization to use.

See **Error! Reference source not found.** section in Appendix C: Somatic CNV Caller for instructions to define normal samples to be used as negative references.

SMA Analysis Setup

1. See CNV Analysis Setup section in Appendix C: Somatic CNV Caller for instructions to setup a SMA Thalassemia analysis.
2. The SMA_RESULTS excel file is output upon completion of the analysis.
3. Detailed SMA call information is in the “SMA Call Report” sheet of the SMA_RESULTS excel file.
4. Each sample’s detailed SMA Report can be downloaded as a PDF file in the “Report” section of the results page.
5. The fully normalized copy number ratios can be found in “Fully Normalized” sheet in SMA_RESULTS excel file. Note that the copy number ratio in PiVAT is defined as the copy number ratio of a potentially positive sample to that of the negative reference samples (diploid with 2 copy number).

SMA QC parameters

See CNV parameters section in Appendix C: Somatic CNV Caller for a list of user adjustable QC parameters.

PiVAT Output: FINAL_REPORT file

Definition and/or description of result columns reported in SMA Call Report tab of FINAL_REPORT xlsx file are provided below.

Sheet Tab	Column Name	Definition/Description
SMA Call Report	**Sample ID**	Unique Sample ID for each sample.
	Gene-Exon	Name of the gene/exon pair, or if a control target, name of the control amplicon.
	Location	Genomic coordinates of exon (hg19).
	Copy Number	Copy number of the gene-exon or of the control amplicon
	Copy Number Ratio	The copy number ratio of the gene-exon or control-amplicon to the copy number (2) of a negative normal sample.

SMA Calls Table

SMA Calls table contains the copy number ratios for the sample(s) selected in **Select Sample.**

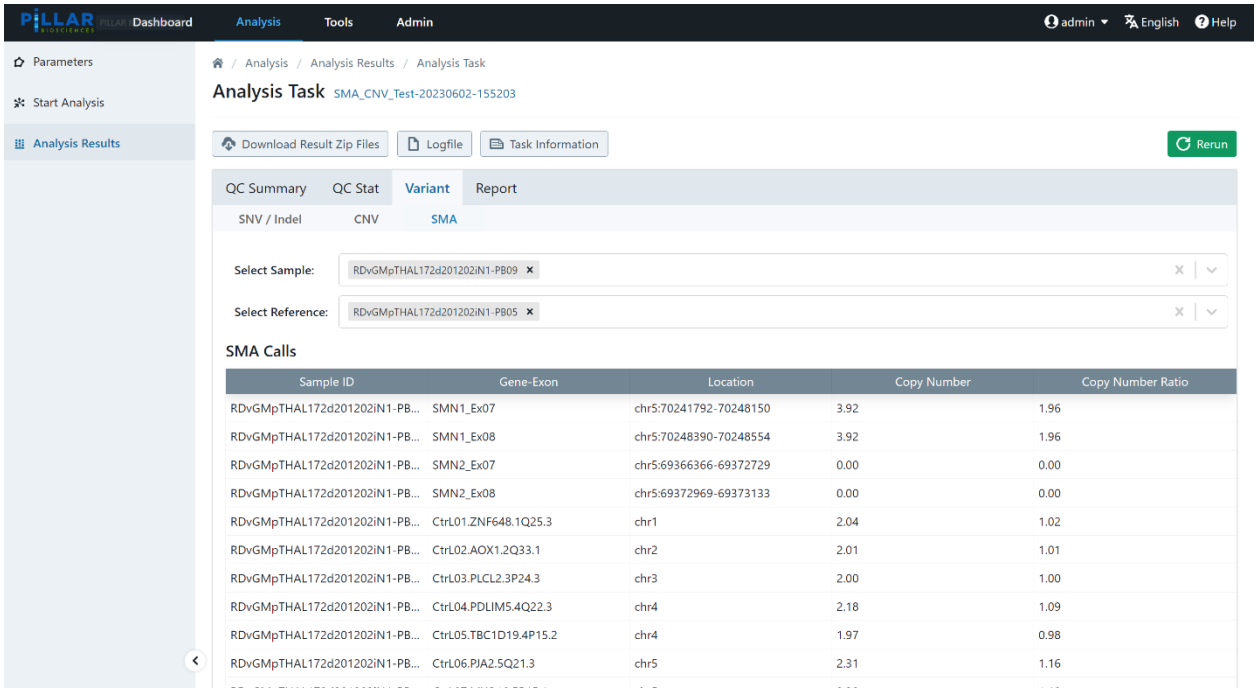


Figure 50 Analysis Results: SMA Calls table

Definition and/or description of result columns reported in the SMA Calls table are provided below.

Column Name	Definition/Description

Sample ID	Unique Sample ID for each sample.
Gene-Exon	Name of the gene/exon pair, or if a control target, name of the control amplicon.
Location	Genomic coordinates of exon (hg19).
Copy Number	Copy number of the gene-exon or of the control amplicon
Copy Number Ratio	The ratio of the copy number of called Thalassemia CNV to the copy number (2) of a negative normal sample.

SMA Plot

A SMA plot for each sample will be below the SMA Results tab. Each boxplot represents the copy number ratios of all CNV Normal Samples specified in the run, and the orange data points represent the copy number ratios of the potentially positive sample. The x-axis represents gene-exon labels, and the y-axis represents copy number ratio.

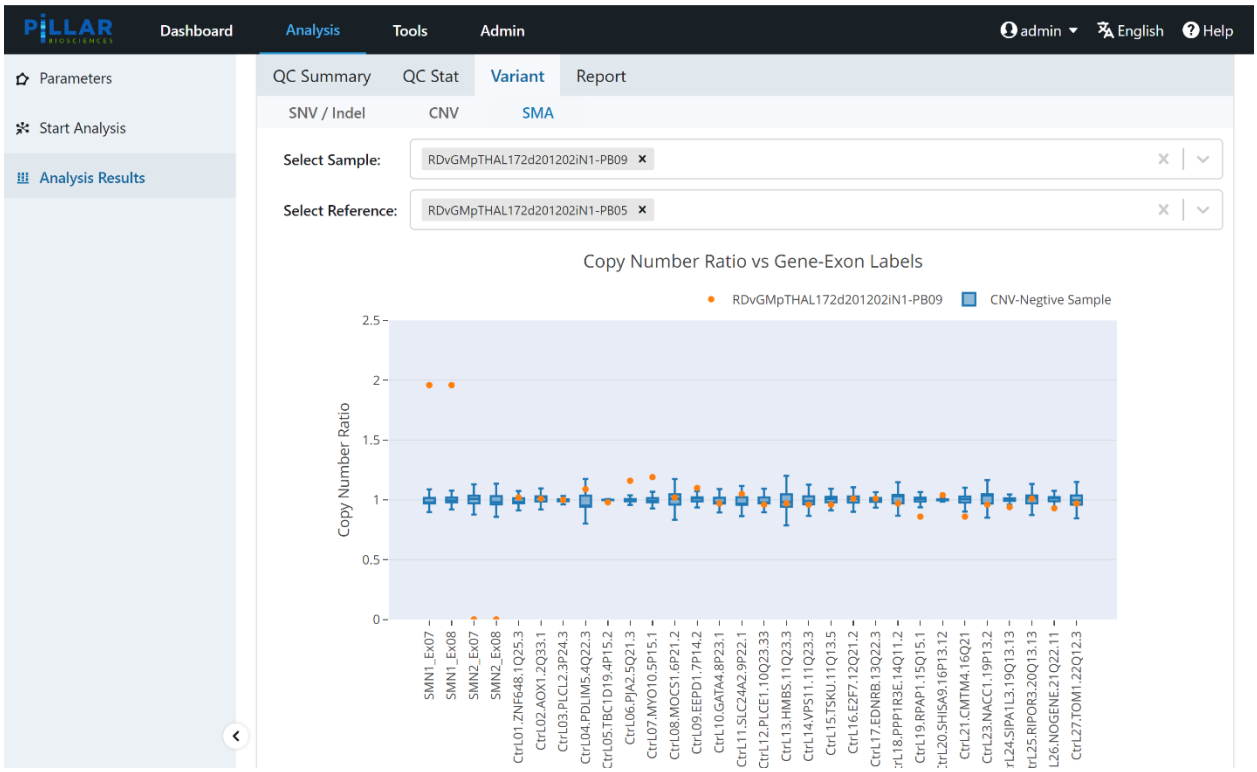


Figure 51: Analysis Results: SMA Sample Box Plot

To save the box plot graph, hover over the graph, and click the "**Download plot as a png**" icon as with CNV results..

Appendix F: Fusion Caller

Pillar Fusion panels detect common fusion transcripts in a simple, multiplex reaction. The output provides likely fusion transcripts by gene and exon pairs.

Fusion Sample Setup

Control samples are not required for fusion analysis.

Fusion Analysis Parameters Setup

1. Navigate to ① **Post Filter Parameters** page to access the following list of adjustable parameters and default values in ( ) by tab. **Please note that these are PiVAT defaults, rather than panel defaults. Please refer to your panel user manual for panel-specific defaults and recommendations.**

Tab	QC Parameters [default]
STAR_FUSION_FILTER_THRESHOLDS	**STAR_FUSION_COUNT_THRESHOLD (20)** Minimum number of reads in which a fusion is observed required for fusion to be called as a True Positive
	STAR_FUSION_FFPM_THRESHOLD (200) Minimum number of reads in which a fusion is observed per million total reads required for fusion to be called as a True Positive

PiVAT Output: STAR_FUSION_RESULTS xlsx file

Definition and/or description of result columns reported in STAR_FUSION_RESULTS file sheets are provided below.

Sheet Tab	Column/Row Name	Definition/Description
Fusion Calls	**SAMPLE_NAME**	Name of sample
	FusionName	FusionName = LeftGene(LeftExon):RightGene(RightExon)
	HGVS	Fusion ID in HGVS Nomenclature
	JunctionReadCount	the number of RNA-Seq fragments containing a read that aligns as a split read at the site of the putative fusion junction.
	SpanningFragCount	the number of RNA-Seq fragments that encompass the fusion junction such that one read of the pair aligns to a different gene than the other paired-end read of that fragment
	Est_J	the estimated junction read counts that have been corrected for multiple mappings
	Est_S	the estimated split read counts that have been corrected for multiple mappings
	SpliceType	indicates whether the proposed breakpoint occurs at reference exon junctions as provided by the reference transcript structure annotations (ex. gencode)
	FFPM	fusion fragments per million total reads
	PROT_FUSION_TYPE	Denotes whether the translated fusion protein is in frame (INFRAME), frameshifted (FRAMESHIFT), or no result (.)
	LeftBreakpoint	Genomic position of the left fusion breakpoint
	LeftExon	Exon immediately preceding the fusion breakpoint
	RightBreakpoint	Genomic position of the right fusion breakpoint
	RightExon	Exon immediately following the fusion breakpoint
	FusionBreakpointSeq	Genomic sequence around the fusion breakpoint
	JunctionReadCount_filter_result	PASS or FAIL boolean result of Junction Read Count filter

Sheet Tab	Column/Row Name	Definition/Description
	FFPM_filter_result	PASS or FAIL boolean result of FFPM filter
	Call_Type	PiVAT variant call type. “True Positive” if the fusion passes all STAR Fusion filter thresholds
Expression Controls	**Sample Name**	Name of Sample
	Type	Type of target (“Internal Control” or "3'/5' expression imbalance")
	Gene	Gene used as Expression Control target for this sample
	AllTargets	Exons targeted within Expression Control Gene for this sample
	Total_counts	Total read counts over Expression Control Gene targets for this sample
Overall Stats	**Before filtering: total reads**	Total reads before filtering for a given sample
	After filtering: total reads	Total reads after filtering for a given sample
	Overall: q30(%)	Percent of total reads (after filtering)with quality scores of at least 30
	Overall q20(%)	Percent of total reads with quality scores of at least 20
	Filtering result: low quality reads	Number of reads filtered based on insufficient quality
	Filtering result: passed filter reads	Number off reads passing filters
	Filtering result: too many n reads	Number of reads filtered due to too many unresolved bases
	Filtering result: too short reads	Number of reads filtered due to insufficient length
	Filtering result: too long reads	Number of reads filtered for excessive length
	Mapped reads	Number of (post filter) mapped reads
	Mapped reads (%)	Percent of total reads (post filter) successfully mapped
	Average input read length	Average input read length
	Average mapped length	Average mapped length
	Fusion candidate reads	
	Fusion positive reads	

Sheet Tab	Column/Row Name	Definition/Description
	Control segment reads	Number of reads aligned to expression Control regions
	Total targeted transcript reads	
	% total targeted transcript reads	

Appendix G: FASTQ File Name Format

A full description of the Illumina FASTQ Name Format can be found under the "Naming" section here:

https://support.illumina.com/help/BaseSpace_OLH_009008/Content/Source/Informatics/BS/NamingConvention_FASTQ-files-swBS.htm

- It is not necessary to name FASTQ files following the Illumina naming convention to analyze them through PiVAT.
- The following terms should be avoided as sample names: “*common*” and “*logs*”. Sample names should not vary by case only.
- Sample names are case insensitive when working in Windows or MacOS environments but, case sensitive when working in Linux. This could result in files being hidden, or overwritten.
- Do not use any patient identifiable information in any file names.