Instructions for Use

LymphoTrack® Software – MiSeq™ Version 2.4.6

This software package is provided with each LymphoTrack Assay for the Illumina® MiSeq instrument to analyze FASTQ files. The software is composed of three distinct parts: the LymphoTrack Data Analysis Application, a LymphoTrack Reporter plugin that runs automatically after analysis, and an optional LymphoTrack Visualization method consisting of four Microsoft Excel® XLSM spreadsheets. The software runs on supported Microsoft Windows® platforms.

This software package allows for clonality analysis of single or multiple target data (*IGHV* Leader [or "Leader"], *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK, TRB* or *TRG*) and additionally provides the somatic hypermutation status of samples run with the Leader or *IGH* FR1 assays.

For guidance with minimal residual disease (MRD) studies, please refer to the technical bulletin entitled *Study of MRD – Using LymphoTrack Assays – MiSeq*, **REF** M-0036. This document can be found on the LymphoTrack MRD Software (**REF** 75000008). To request the technical bulletin, emailing marketing@invivoscribe.com.

RUO This assay is for Research Use Only. Not for use in Diagnostic Procedures.

Product Compatibility

This LymphoTrack Software – MiSeq package (**REF** 75000009) was developed specifically for and is compatible with **only** the following Invivoscribe assays:

Catalog #	Description	Quantity
REF 71210059	LymphoTrack IGHV Leader Somatic Hypermutation Assay Kit A – MiSeq	8 indices – 5 reactions each
REF 71210069	LymphoTrack IGHV Leader Somatic Hypermutation Assay Panel – MiSeq	24 indices – 5 reactions each
REF 71210009	LymphoTrack IGH FR1 Assay Kit A – MiSeq	8 indices – 5 reactions each
REF 71210039	LymphoTrack IGH FR1 Assay Panel – MiSeq	24 indices – 5 reactions each
REF 71210149	LymphoTrack IGH FR1 Assay Panel B – MiSeq	24 indices – 5 reactions each
REF 71210089	LymphoTrack IGH FR2 Assay Kit A – MiSeq	8 indices – 5 reactions each
REF 71210099	LymphoTrack IGH FR2 Assay Panel – MiSeq	24 indices – 5 reactions each
REF 71210109	LymphoTrack IGH FR3 Assay Kit A – MiSeq	8 indices – 5 reactions each
REF 71210119	LymphoTrack IGH FR3 Assay Panel – MiSeq	24 indices – 5 reactions each
REF 71210129	LymphoTrack IGH FR1/2/3 Assay Kit A – MiSeq	8 indices per FR – 5 reactions each
REF 71210139	LymphoTrack IGH FR1/2/3 Assay Panel – MiSeq	24 indices per FR – 5 reactions each
REF 71220009	LymphoTrack IGK Assay Kit A – MiSeq	8 indices – 5 reactions each
REF 71220019	LymphoTrack IGK Assay Panel – MiSeq	24 indices – 5 reactions each
REF 72250009	LymphoTrack TRB Assay Kit A – MiSeq	8 indices – 5 reactions each
REF 72250019	LymphoTrack TRB Assay Panel – MiSeq	24 indices – 5 reactions each
REF 72270019	LymphoTrack TRG Assay Kit A – MiSeq	8 indices – 5 reactions each
REF 72270009	LymphoTrack TRG Assay Panel – MiSeq	24 indices – 5 reactions each

Table of Contents

1. Product Use

The RUO LymphoTrack Software – MiSeq package is intended to be used for bioinformatics data analysis and visualization of raw FASTQ files from the Illumina MiSeq instrument when used with LymphoTrack Assays.

2. Glossary and Abbreviations

2.1. Glossary

Table 1. Glossary Terms

Term	Definition
Clonal	The aggregate of genetically identical cells or organisms produced from a single progenitor cell.
Invalid	A sample or control result which does not meet the validity criteria.
Semantic versioning	A software version scheme consisting of 3 numbers (Major.Minor.Patch), aligned with the risk factor of the update.
Software	LymphoTrack Software - MiSeq
System	The entire package of software, hardware and assay (as applicable) that make up the medical device.

2.2. Abbreviations

Table 2. Abbreviations defined

Abbreviation	Definitions
CDR3	Complementarity-determining region, formed by the joining of V-(D)-J segments and establishes the diversity of the cellular immune response
CSV	Comma separated value file format
FASTQ	A text-based format for storing both a biological sequence (usually nucleotide sequence) and its corresponding quality scores
FFPE	Formalin-fixed paraffin-embedded
IFU	Instructions for use
IGH FR1	Immunoglobulin heavy chain Framework 1
IGH FR2	Immunoglobulin heavy chain Framework 2
IGH FR3	Immunoglobulin heavy chain Framework 3
IGK	Immunoglobulin kappa light chain
PCR	Polymerase chain reaction
PDF	Portable document format
SHM	Somatic hypermutation
TRB	T-cell receptor beta
TRG	T-cell receptor gamma
TSV	Tab-separated value file format
USB	Universal Serial Bus; allows data exchange and delivery of power between many various types of electronics

3.   Minimum System Requirements

- **Processor**: Intel Core 2 Duo or newer CPU recommended.
- **Hard Drive:** At least 80 GB of free disk space is required; 250 GB recommended.
- **RAM:** 4 GB required; 8 GB or more recommended.
- **Operating System:** Windows 10 Pro or Windows 11 Pro (64-bit) is required.
- A PDF reader, such as Adobe Acrobat Reader, to visualize data reports generated by the LymphoTrack Reporter.
- Microsoft Office Excel 2007, 2010 or 2013 with Macro support enabled is required for deprecated data visualization.
- A CD-ROM drive.
- The LymphoTrack Software – MiSeq v2.4.6 has been validated for use with Windows 10 Pro and Windows 11 Pro, configured for English (United States) language and default display size settings, and Microsoft Excel 2007/2010/2013 for alternate data visualization. Use of other Windows/Excel versions and/or language/display size settings may not be compatible.

4.   Warnings and Precautions

- **Default screen size**. Optimal screen resolution for LymphoTrack Software – MiSeq is 100%. Increasing the monitor default screen size above 100% may affect the display of field text in the user interface.

- **System font**. User interface is designed to use the default system font settings on a Windows computer using the English (US) language locale. Changes to the system font settings or using a language locale other than English (US) may impact the display of field text in the user interface.

- **Compatible files.** LymphoTrack Software – MiSeq is compatible with FASTQ.GZ files generated by the MiSeq Reporter Software. LymphoTrack Software – MiSeq is not compatible with FASTQ.GZ files generated by the *Bcl2fastq* software.

- **Adapter trimming.** Adapter trimming must be selected ☑ when creating MiSeq sample sheets so the resulting FASTQ.GZ files generated are adaptor trimmed.

- **Sleep or hibernate settings.** Consider disabling the sleep and/or sleep or hibernate mode features before executing the LymphoTrack Data Analysis Application. If the computer enters sleep or hibernate mode, the software may stop running.

- **Characters in pathname and file name.** 1) Avoid spaces in the pathname for the data files or software (pathnames include file folders and file names); more than one consecutive space is not permitted. 2) It is important that the filenames only contain alphanumeric and hyphen characters (A-Z, a-z, 0-9, , - [hyphen]). If the software encounters a character not within this set, it may fail. Please make sure to follow this sample naming nomenclature when setting up the Sample Sheet for the MiSeq. 3) Do not exceed 255 characters for a pathname. Sample names that cause the output filenames to approach this limit are difficult to copy to other locations.

- **Duplicate samples must have a unique sample name.** Provide a unique name and identifier for each sample when creating sample names. If samples are run in duplicate, they can be named "Sample1a" and "Sample1b". Failure to provide unique names to samples that will be run together on the same flow cell will result in only one sample being analyzed by the software during the analysis process.

5.   Installing the Software

The LymphoTrack Software – MiSeq package (**REF** 75000009) can downloaded through the invivoscribe Software Portal or copied from the provided CD.

- Install software only on a local drive and not a network drive.
- The software may not function properly when running across a network connection.

> **Note:**   Administrator permission is required to Install the software

5.1.   Load the LymphoTrack Software – MiSeq CD into the disk drive; **_OR_**

5.2.   Download the LymphoTrack Software – MiSeq package from the Invivoscribe Software Portal.

    5.2.1.   Using any web browser, navigate to https://catalog.invivoscribe.com/softwareportal/

    5.2.2.   Complete the following text fields:

        5.2.2.1.   <u>Email:</u> Enter a valid email address. A link to the software download will be sent to this address.

        5.2.2.2.   <u>Customer Account Number:</u> Enter your unique ID used when placing orders with Invivoscribe.

        5.2.2.3.   <u>Software Code:</u> Enter the software code found at the bottom of the associated sales order.

    5.2.3.   Check the *Terms and Conditions* box to proceed.

    5.2.4.   Click the **Request Software** icon.

- Following valid input text into the above fields, a link to the software download will be sent to the provided email address.

    5.2.5.   Click the link or *copy+paste* into a web browser; the software will automatically download.

5.3.   Install the LymphoTrack Software - MiSeq

    5.3.1.   Double click the **_LymphoTrackMiSeq-2.4.6.RUO.msi_** file.

        5.3.1.1.   If a message appears, stating *Windows protected your PC*, click **More info** and verify the publisher is <u>Invivoscribe, Inc.,</u> then click **Run anyway**.

- The software installer will open and allow the software to be installed to a local drive.
- The default install location is *C:\Invivoscribe\LymphoTrackMiSeq-2.4.6.RUO*.

    5.3.2.   Follow the Windows Setup Wizard installation prompts, select the directory file path to save the software and click **Install**.

- Install software only on a local drive (not a network drive); the software might not function properly if run across a network connection.

        5.3.2.1.   If the *User Account Control dialog* appears, click **Yes**.

    5.3.3.   Once installation is complete, click **Finish**.

- The software is now ready to use and can be accessed by navigating to the *Start Menu* or the directory file path selected in step 5.3.2.

6. Running the Data Analysis Software

6.1. Identify the paired-end, adapter-trimmed FASTQ.GZ files to be analyzed.

- These files are located within the MiSeq run folder and are the FASTQ.GZ files that will be used as input for the **LymphoTrackMiSeq** data analysis application.
- The *Undetermined_ *.FASTQ.GZ files are not required for analysis.

6.2. Copy the FASTQ.GZ files identified above to an empty folder on the computer. Download these files to the same local computer which has the software installed.

- The drive containing this folder requires at least 80 GB of available space.
- This folder will be the **<input_folder>** referenced below for the software.

6.3. Double-click on the ***LymphoTrackDxMiSeq-2.4.6.RUO*** application file located in the **Invivoscribe** folder created during installation.

- The default location is *Local Disk (C:) > Invivoscribe*.

6.4. Read the license agreement. To accept the terms, click the **Accept** button to proceed.

6.5. Check the desired analysis mode(s) from the *Choose analysis mode(s)* checkbox options.

- The analysis mode must correspond to the type of data in the run (Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK*, *TRB* and/or *TRG*).

6.6. Select the desired decimal mark notation from the *Choose decimal mark* options.

- This option impacts the use of the Microsoft Excel data visualization macro later and is a required input.
- The choices are to use a decimal point (".") or a decimal comma (",") for the decimal mark.

IMPORTANT! **Decimal mark.** To determine which option to choose, open a new Excel worksheet and type *=10000* into cell A1. Right-click this cell and select **Format Cells**.

- From the *Number* tab, look in the *Category:* list, and select **Number**. To the right is a *Decimal places* value which is set to **2**.
- Below this value is a *Use 1000 Separator* checkbox. Check this box and look at the sample number above. It will contain the original number entered, formatted according to the Excel defaults for a number with a decimal mark.
- If this number appears as *10,000.00*, then select the **decimal point** option. If this number appears as *10.000,00*, then select the **decimal comma** option.

6.7. Click on the **Browse** button to choose an input folder. This folder is the **<input_folder>** referenced above and contains the files to be analyzed. When the input folder is selected, it will appear empty.

IMPORTANT! **The <input_folder> selected must not contain any sample files with the same name as the input folder.** If the program was run with a sample that shared the same filename as the folder that was provided as input to the analysis functionality, the program will freeze and analysis runtime will increase significantly (*i.e,*. it will take 30 minutes to process a small sample that takes seconds to process normally.

6.8. Once the **<input_folder>** and analysis mode are selected, click the **Launch Program** button.

IMPORTANT! **The <input_folder> selected must contain only FASTQ or FASTQ.GZ files.** If neither file type nor a mix of file types are present, the program will display an error message and request a different folder to be chosen. If the FASTQ files are generated by *Bcl2fastq* (the filenames contain the barcode sequence instead of the "S1, S2, S3" *etc*… nomenclature) the program will display an error message and request the data to be re-analyzed using the MiSeq Reporter Software.

6.9. **Status of analysis**. After clicking the **Launch Program** button, the main program window will display the program's status. When the program is finished analyzing the input data files, the main program window will display the message *Analysis is complete…* and show the location of the output folder as well as provide a prompt to open the appropriate Excel Data Visualization spreadsheet.

▪ A new analysis can be started from the main program window as soon as the previous analysis is complete.

IMPORTANT! **Length of analysis.** The size and number of samples being analyzed, as well as the analysis mode and the computer's hardware will determine the amount of time the program takes to run. A typical read distribution for a 24-sample run is 200,000 to 1 million reads per sample. These numbers may change with future releases of the MiSeq platform.

Analysis of a typical single *IGHV* Leader, *IGH* FR1*, IGH* FR2, *IGH* FR3, *IGK, TRB* or *TRG* sample requires approximately 1-2 minutes to complete. A typical 24-sample Leader-only, *IGH* FR1-only, *IGH* FR2-only, *IGH* FR3-only, *IGK*-only, *TRB*-only or *TRG*-only run requires approximately 25-50 minutes to complete. Analysis of a typical 24-sample combined Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK, TRB* and/or *TRG* run multiplies the analysis time by the number of targets being analyzed.

6.10. **Output folder(s).** Depending on the analysis mode chosen, one or more output folders will be created within the **<input_folder>** containing the source FASTQ or FASTQ.GZ files.

▪ Table 3 outlines the relationship between the analysis mode and the name of the output folder(s).

Table 3. Output folder name according to analysis mode

Analysis Mode	Output folder name(s)
Leader	*LEADER_output*
IGH FR1	*IGH_FR1_output*
IGH FR2	*IGH_FR2_output*
IGH FR3	*IGH_FR3_output*
TRB	*TRB_output*
TRG	*TRG_output*
IGK	*IGK_output*
Any combination of: Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK*, *TRB* and *TRG*	Two to seven of the output folders listed above

6.11. After the program has finished running successfully, the output folder will contain several TSV files and a PDF file for each sample processed.

▪ Note that "*" in the tables below represents the sample name as entered into the sample sheet, followed by platform specific information (*e.g., samplename_S1_L001_001_combined*).

▪ For Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3, *TRB* and *IGK* analysis, please see Table 4 for descriptions of the nine files provided for each sample.

Table 4. Description of *IGH, TRB* and *IGK* output files.

Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3, *TRB* and *IGK* Analysis	
.fastq_read_summary_merged_top10_searchtop500.tsv	Same format as the *read_summary* file, but only includes the top 10 reads from the read_summary merged with reads in the top 500. A read will merge with another if they are only 1 or 2 nucleotides (nts) different, with the read with the highest count maintaining its sequence identity, and the read with the lower count added to the higher one.
.fastq_unique_reads.tsv	A FASTA file of unique sequences, sorted by count.
.fastq_CDR3_unique_reads.tsv	A FASTA file of putative CDR3 sequences found in the FASTQ_UNIQUE_READS file.

Table 4. Description of *IGH, TRB* and *IGK* output files.

Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3, *TRB* and *IGK* Analysis	
*.fastq_VJ_sequence_frequency_family.tsv	Each family is a grouping of all possible recombinations found in a particular V or J region such as a V1 family or a J4 family. Within the V1 family any number of recombinations can be found such as V1-69 and V1-8. This file contains a list of all possible V family – J family recombinations, each with a sorted list of reads (from highest to lowest %). This file is best used in a "stack" graph in Excel.
*.fastq_VJ_usage.tsv	A count matrix and % matrix of the different V-J recombination events.
*.fastq_read_summary.tsv	Contains a list of the top 200 unique sequences sorted by count (number of reads), specifying: ▪ Which V-J recombination event it best fits ▪ Length ▪ Actual % of total reads ▪ Cumulative % of total reads ▪ How well the read covers the identified V-gene targeted by the primers (%) (*IGH/IGK*) ▪ Putative CDR3 sequence Leader and *IGH* FR1 also contains: ▪ Mutation rate to partial V-gene (%) Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3 also contains: ▪ If the rearrangement is predicted to be in-frame (Y/N/n/a). If either the V or J gene segment usage cannot be predicted due to the quality of the alignment, the software will not be able to predict if the gene is in-frame and the result will be n/a. ▪ If the rearrangement is predicted to not contain a stop codon (Y/N). When an in-frame prediction cannot be made a prediction for whether a stop codon is present cannot be accurately made.
*.fastq_read_summary_family.tsv	Contains a sorted list of unique sequences, by count, specifying: ▪ Which V-J family recombination event it best fits ▪ Length ▪ Actual % of total reads ▪ Cumulative % of total reads ▪ How well the read covers the identified V-gene targeted by the primers (%) (*IGH/IGK*) ▪ Putative CDR3 sequence Leader and *IGH* FR1 also contains: ▪ Mutation rate to partial V-gene (%) Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3 also contains: ▪ If the rearrangement is predicted to be in-frame (Y/N/n/a) If the rearrangement is predicted to not contain a stop codon (Y/N)
*.fastq_VJ_usage_family.tsv	A count matrix and % matrix of the different V family – J family recombination events.
*<timestamp>.pdf	The PDF report generated by the LymphoTrack Reporter will contain similar information as the *merged read summary* file, the *read summary* file, a *V – J usage* graph, and a *V – J sequence frequency* graph. See *section 3 Data Visualization using LymphoTrack Reporter* for more information.

Note: *IGK* analysis additionally includes information for other elements in the *IGK* locus involved in recombinations (IGKINTR and IGKDEL).

Note: *TRB* analysis additionally includes information for other elements in the *TRB* locus involved in recombinations of D – J region

Note: *Putative CDR3 sequence* is determined by matching an exact definition of a CDR3 sequence. Any mutation of anchor points or disruption of the reading frame will result in a value of *not found*"

- For *TRG* analysis, please see Table 5 for descriptions of the seven files provided for each sample.

Table 5. Description of *TRG* output files.

TRG Analysis	
*.fastq_read_summary.tsv	A list of unique sequences sorted by count, specifying: - Which V-J recombination event it best fits - Length - Actual % of total reads - Cumulative % of total reads - Putative CDR3 sequence
*.fastq_read_summary_merged_top10_searchtop500.tsv	Same format as the read_summary file, but only includes the top 10 reads from the read_summary merged with reads in the top 500. A read will merge with another if they are only 1 or 2 nts different, with the read with the higher count maintaining its sequence identity, and the read with the lower count added to the higher one.
*.fastq_unique_reads.tsv	A FASTA file of unique sequences, sorted by count.
*.fastq_CDR3_unique_reads.tsv	A FASTA file of putative CDR3 sequences found in the FASTQ_UNIQUE_READS file.
*.fastq_VJ_sequence_frequency.tsv	A list of all possible V-J recombinations, each with a sorted list of reads (from highest to lowest %). The first number represents the % of total reads that the first unique read accounted for out of all reads sequenced. This file is best used in a "stack" graph in Excel.
*.fastq_VJ_usage.tsv	A count matrix and % matrix of the different V-J recombination events.
*<timestamp>.pdf	The PDF report generated by the LymphoTrack Reporter will contain similar information as the *merged read summary* file, the *read summary* file, a *VJ usage* graph, and a *VJ sequence frequency* graph. See *section 3 Data Visualization using LymphoTrack Reporter* for more information.

Note: "Putative CDR3 sequence" is determined by matching an exact definition of a CDR3 sequence. Any mutation of anchor points or disruption of the reading frame will result in a value of "not found".

- The output folder(s) contain PDF reports and TSV output data files.
- The TSV files can be used as inputs for the Data Visualization spreadsheet(s).

7. Data Visualization Using LymphoTrack Reporter (Automated, Recommended)

7.1. A PDF report for each sample is automatically generated at the completion of analysis.

- The report is named after the sample name present in the sample sheet, followed by platform specific information, and ends with a timestamp (DDMMYYhhmmss).

IMPORTANT!   **Caution when saving the PDF report:** Depending on the program used to open the PDF, one or more of the columns can be edited. Do NOT make changes to any part of the PDF. If the user suspects an edit has been made, do NOT save the PDF. If the user suspects an edit has been saved, re-run the analysis to generate a new PDF report.

7.2. The first page of the PDF report will contain an **_Index Q30_** value, calculated as the percentage of bases that have a quality score of 30 or higher, but only from reads that are assigned to an index after trimming. This value can also be used to determine if the index passed the standard Illumina specification.

- In some cases, the run Q30 score provided by the sequencer will underestimate the quality of a run because of counting undetermined sequences and untrimmed regions towards the Q30.
- Undetermined and untrimmed regions are typically lower quality and do not contribute to the downstream analysis by LymphoTrack.

7.3. The first page of the PDF report will contain a table of the **_merged read summary_** results, followed by the graphs of the **_V – J Sequence Frequency_** and **_V – J Usage_** results, and the remaining pages contain a table of the unmerged **_read summary_** results.

- The contents of these are detailed in Table 6:

Table 6. Pages included in the *PDF* output file.

PDF Output for LymphoTrack Reporter	
Merged Read Summary	Same format as the Read Summary worksheet, but only for the top 10 reads after merging with the top 500 if reads are only 1 or 2 nts different from each other. Additionally, the index Q30 score is displayed under the Total Read Count.
V – J Sequence Frequency Graph	A histogram showing the top 200 sequence frequencies of each V family – J family (for *IGH/IGK* and *TRB*) or V gene – J gene (for *TRG*) recombination.
V – J Usage Graph	A 3D bar chart showing the V family-J family usage of the top 200 sequences (for *IGH/IGK* and *TRB*) or V gene – J gene usage for all sequences (for *TRG*).
Read Summary	Contains a sorted list of the top 200 unique sequences by count, specifying: - Which V – J recombination event each sequence best fits - Sequence - Sequence length - Actual % of total reads - Cumulative % of total reads listed for top 200 reads How well the read covers the identified V-gene targeted by the primers (%). (*IGH/IGK*) - Putative CDR3 sequence Leader and *IGH* FR1 also contains: - Mutation rate to partial V-gene (%) Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3 also contains: - If the rearrangement is predicted to be in-frame (Y/N/n/a). If either the V or J gene segment usage cannot be predicted due to the quality of the alignment, the software will not be able to predict if the gene is in-frame and the result will be n/a. - If the rearrangement is predicted to not contain a stop codon (Y/N). When an in-frame prediction cannot be made a prediction for whether a stop codon is present cannot be accurately made.

Note: "Putative CDR3 sequence" is determined by matching an exact definition of a CDR3 sequence. Any mutation of anchor points or disruption of the reading frame will result in a value of "not found".

Note: *IGK* analysis additionally includes information for other elements in the *IGK* locus involved in recombinations (IGKINTR and IGKDEL).

Note: *TRB* analysis additionally includes information for other elements in the *TRB* locus involved in recombinations of D – J region

8. Data Visualization (Manual): Running the Excel Data Visualization Spreadsheets- Discommended

8.1. Check the Security settings within Microsoft Excel to make sure Macro support is properly enabled:

☐ Excel 2007: Click the **Office** button and choose **Excel Options**.
☐ Excel 2010/2013: Click the **File** menu and choose **Options**.
☐ From the *Options* window, click on **Trust Center** and then the **Trust Center Settings...** button.
☐ Click on **Macro Settings** and set the *Disable all macros with notification* option.
☐ Click **Ok** and **Ok** again.
☐ Close Excel.

IMPORTANT! Because the visualization files can be saved as a workbook (which would remove the ability of the Macro to function), copy the visualization files to a different file path before beginning analyses.

8.2. Double-click the appropriate *Data Visualization* spreadsheet file to open it.

▪ The visualization files are located in the *visualizations* sub-folder previously located in the installation folder of the software. The following visualization spreadsheet files are available:

☐ *LymphoTrackIGHVisualization.xlsm* for Leader, *IGH* FR1, *IGH* FR2 and *IGH* FR3
☐ *LymphoTrackIGKVisualization.xlsm* for *IGK*
☐ LymphoTrackTRBVisualization.xlsm for *TRB*
☐ *LymphoTrackTRGVisualization.xlsm* for *TRG*

8.3. When opening the file, the message: ***Security Warning - Macros have been disabled*** may appear*.*

☐ Excel 2007: Click the **Options...** button to the right of the warning and select *Enable this content* and click **Ok**.
☐ Excel 2010/2013: Click the **Enable Content** button.

The data visualization macro will start automatically.

8.4. The License Agreement will be displayed; this must be accepted to use the software.

▪ After accepting the License Agreement, a dialog box will open with the prompt: *Please select the Read Summary file for the sample you want.*

Note: The macro can only perform data visualization on one sample at a time, so this procedure will need to be repeated for each sample to be visualized.

8.5. Browse to the output folder created by the Data Analysis application that contains the data to be visualized and select the ***FASTQ_READ_SUMMARY*** file for the sample to be examined.

8.6. Click the **Open** button.

8.6.1. If a message regarding *Complex Formatting* appears, click **Yes**.

▪ The macro will load the data files for the chosen sample and create several worksheets, each of which will present different readable views of the data.
▪ This may take a few moments, depending on the size of the sample and the computer's hardware.

IMPORTANT! Due to a limitation in Microsoft Excel, the Visualization software is not capable of displaying more than 1,048,576 rows of data per sheet. Only the first 1,048,576 rows of data will be shown if any of the Data Analysis output files for the current sample contain more rows than this limit.

If this limit is reached, the program will display a message stating which sheet(s) are affected by this limitation. Click **OK** to proceed.

Though not all data may be displayed, the graphs are not affected and will still be accurate.

IMPORTANT!   Graphs are generated assuming a *left-to-right* display direction. In countries where the normal reading direction is *right-to-left*, the default Excel settings may be incompatible, and graphs are not generated correctly. To fix this, the Excel display option must be changed to *left-to-right*.

1) This option can be set from either **Office → Excel Options** in Excel 2007, or **File → Options** in Excel 2010/2013.
2) Select **Advanced**, then scroll down to the *Display* settings.
3) Change the *Default direction* option to *Left-to-right*, then click **OK**.
   - The next time Excel is opened, the worksheets will read from left-to-right, and the Visualization software will generate charts correctly.

   If this situation occurs and further assistance is required, please contact our Technical Support department at +1 (858) 224-6600, Monday through Friday 7:00AM – 5:00PM PST/PDT, or via email at support@invivoscribe.com.

8.7.   **Save the visualization file.** When processing is complete, a message will appear stating the program has completed. Saving the file at this point is strongly recommended.

   8.7.1.   If a warning message regarding saving as a macro-free workbook appears, click the **Yes** button.

IMPORTANT!   When saving the file, make sure to choose *Save As* and select an Excel Workbook file format (XLSX or XLS), because the default behavior of choosing **Save** or clicking the disk button will be to replace the XLSM Macro with the current workbook. This will require reinstallation of the macro with a new copy from the software CD for subsequent analyses.

8.8.   The completed workbook for Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3*, IGK, TRB* and *TRG* will include the following eight worksheets described in Table 7.

Table 7. Worksheets included in Excel output file.

Completed Workbook for Leader, *IGH* FR1*, IGH* FR2*, IGH* FR3*, IGK TRB* and *TRG*	
Read Summary	Contains a sorted list of the top 200 unique sequences by count, specifying: • Which V-J recombination event each sequence best fits • Sequence • Sequence length • Actual % of total reads • Cumulative % of total reads listed for all reads (*TRG* only) • Cumulative % of total reads listed for top 200 reads (*IGH/IGK* and *TRB*) • How well the read covers the identified V-gene targeted by the primers (%) (*IGH/IGK*) • Putative CDR3 sequence Leader and *IGH* FR1 also contains: • Mutation rate to partial V-gene (%) Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3 also contains: • If the rearrangement is predicted to be in-frame (Y/N/n/a). If either the V or J gene segment usage cannot be predicted due to the quality of the alignment, the software will not be able to predict if the gene is in-frame and the result will be n/a. If the rearrangement is predicted to not contain a stop codon (Y/N). When an in-frame prediction cannot be made a prediction for whether a stop codon is present cannot be accurately made • Putative CDR3 sequence
Merged Read Summary	Same format as the *Read Summary* worksheet, but only for the top 10 reads after merging with the top 500 if reads are only 1 or 2 nts different from each other.
VJ Sequence Frequency Graph	A histogram showing the top 200 sequence frequencies of each V family – J family (for *IGH/IGK* and *TRB*) or V gene – J gene (for *TRG*) recombination.
VJ Usage	Raw counts and frequencies of the top 200 V family - J family (for *IGH/IGK* and *TRB*) or V gene – J gene (for *TRG*) recombinations.

Table 5. Worksheets included in Excel output file.

Completed Workbook for Leader, *IGH* FR1*, IGH* FR2, *IGH* FR3, *IGK TRB* and *TRG*	
VJ Usage Percent Graph	A 3-D graph showing the top 200 sequence frequencies of each V family – J family (for *IGH/IGK* and *TRB*) or V gene – J gene (for *TRG*) recombination.
VJ Usage Raw Graph	A 3-D graph showing the top 200 raw counts of each V family – J family (for *IGH/IGK* and *TRB*) or V gene – J gene (for *TRG*) recombination.
VJ Sequence Frequency	A "stack" list of all possible V family – J family (for *IGH/IGK* and *TRB*) or V gene – J gene (for *TRG*) recombinations, each with a sorted list of the top 200 reads (from highest to lowest %).
Unique Reads	A FASTA file of unique sequences, sorted by count.

Note: *IGH* listed in the table above refers to Leader, *IGH* FR1, *IGH* FR2 and *IGH* FR3 analysis. The frequency graphs for these gene targets only specify *IGH* in the title. The data output folder specifies whether Leader or one of the framework (FR) regions was analyzed by the software and this information is included in the data file location field at the top of the *Read Summary* worksheet.

Note: *IGK* analysis additionally includes information for other elements in the *IGK* locus involved in recombinations (IGKINTR and IGKDEL).

Note: *TRB* analysis additionally includes information for other elements in the *TRB* locus involved in recombinations of the D – J region

Note: Top 200: When specified, the worksheets will show the top sequences (up to 200) for Leader, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK* and *TRB*. All the sequences are shown for *TRG* in the *Read Summary*.

Note: *Putative CDR3 sequence* is determined by matching an exact definition of a CDR3 sequence. Any mutation of anchor points or disruption of the reading frame will result in a value of *not found*.

IMPORTANT! V/J gene assignments found in the *Read Summary* can sometimes be *none*. These assignments are based on the quality of the best alignment between the read and the reference genes. If the best alignment falls below a quality threshold, the gene assignment cannot be confidently made and is assigned a *none*.

IMPORTANT! The Mutation rate to partial V-gene (%) will display "n/a" for *IGH* FR2 and FR3. There is limited V-gene coverage due to the shorter amplicon sizes of these targets which results in inaccurate mutation rate calculations.

☐ When analyzing the somatic hypermutation status of samples, focus on the top 4 reads (limit monoclonal or biclonal specimens to the top 2 or 4 reads, respectively), as the somatic hypermutation rate (SHM) and in-frame/stop-codon analyses are most informative for these reads. Due to the unbiased nature of the NGS platforms, many more reads than just the top 4 are sequenced and analyzed, most of which are low abundance rearrangements, or can be attributed to artifacts in amplification or sequencing. SHM and in-frame/stop-codon analyses are not useful in non-clonal reads in these cases. Please note that Invivoscribe uses IgBLAST (http://www.ncbi.nlm.nih.gov/igblast/showGermline.cgi) for reference sequences and results may differ slightly from the IMGT database. In addition, many pseudogenes and non-functional genes are not included in the analysis. Interpret rearrangements with somatic hypermutation statuses at or near the 2% cutoff with caution, particularly when running the *IGH* FR1 assay which only targets a portion of the *IGHV* region as the sequence upstream of the primer binding site will not be assessed.

9. Technical and Customer Service

We appreciate your business. We are happy to assist you with understanding this software, and will provide ongoing technical assistance Monday through Friday to keep the assays performing efficiently in your laboratory.

Contact Information

▄ Invivoscribe, Inc

10222 Barnes Canyon Road | Building 1 | San Diego | California 92121-2711 | USA

Phone: +1 858 224-6600 | Fax: +1 858 224-6601 | Business Hours: 7:00AM - 5:00 PM PST/PDT

Technical Service: support@invivoscribe.com | Customer Service: sales@invivoscribe.com | Website: www.invivoscribe.com

10. Symbols

The following symbols are used in Invivoscribe NGS product labeling.

REF	Catalog Number		Expiration Date
VOL	Reagent Volume		Manufacturer
LOT	Lot Number		Consult Instructions for Use
	Storage Conditions	RUO	Research Use Only

11. Legal Notice

Please visit https://invivoscribe.com/legal-notices/ for the most up-to-date legal notices.

12. Software Frequently Asked Questions

Q1: **Why does the Read Summary from the *IGH*, *IGK* or *TRB* analysis have fewer than 200 reads?**

A1: The LymphoTrack Software – MiSeq aligns the top 200 reads during analysis, but the Read Summary only reports reads that have aligned to a V or J gene found in the reference. If a read cannot align to the reference, it is left out of the Read Summary.

The reasons for a non-alignment to the reference are numerous, but usually focus on a problem with the sample, leading to a large amount of non-specific amplification to occur. This event can stem from low numbers of target cells in the sample, improper PCR conditions, or low-quality sample types, such as fragmented FFPE. To reduce the chance of these events, please follow the Instructions for Use precisely.

Q2: **Why does the V-assignment or J-assignment say "none" in the Read Summary table?**

A2: To assign a read to a specific V and/or J gene, a minimum quality threshold must be attained during the alignment. If the alignment is poor, a confident assignment cannot be made and is instead labeled *none*. Note that a read cannot be assigned *none* for both V and J genes, as these would not be reported, per the condition discussed in question 2.

Q3: **The *V – J Sequence Frequency Graph* has strange values for the y-axis. The *% total reads* is in the millions, how is this possible?**

A3: This is likely due to running the LymphoTrack program using the incorrect decimal type for your region. See section 6. *Running the Data Analysis Software*, specifically for selecting the decimal mark.

Q4: **All of the graph worksheets are empty, but the non-graph worksheets have data. Where did the graphs go?**

A4: This may be due to the default direction of the Excel program being *right-to-left* instead of *left-to-right*. If the column labels on an empty worksheet start at **A** on the right-side of the worksheet, switch the direction to *left-to-right* before using the Data Visualization macro. To change this setting, go to ***Excel options*** from the Office button in the upper left-hand corner, click on **Advanced**, and scroll down to the ***Display*** section. An option called ***Default direction*** is listed here and selecting **Left-to-right** will allow subsequent data visualization runs to work correctly.

Note: This will permanently change the direction until it is changed back. Closing Excel will not revert this change.

Q5:   **The program stopped after clicking _Launch Program_ and produced this error:**

Application terminated with error: Filename not compatible. Please use only MiSeq MSR software generated FASTQ files.

For help with this error, please contact Invivoscribe Technical Support.

A5:   The LymphoTrack software **–** Miseq expects paired-end FASTQ files generated by the MiSeq Reporter (MSR) software contained within the MiSeq machine. If the MSR software was not used to generate sequence files, the base-calling must be performed using the MSR software. Please refer to the MiSeq User Guide for more information.

Performing the base-calling using the MSR software is important because, while other programs can generate very similar results as the MSR software (for example, _Bcl2fastq_), these programs can also generate very different results, depending on the specific parameters. The LymphoTrack Assay is only valid when analyzing FASTQ files generated from the MSR software.

Q6:   **Why do the analysis results seem to include fewer reads than expected?**

A6:   If there are very few or zero total reads analyzed according to the Read Summary, make sure the correct analysis type is selected (Leader, _IGH_ FR1, _IGH_ FR2, _IGH_ FR3, _IGK, TRB_ or _TRG_) when running the LymphoTrack software **–** Miseq . If there seems to be a depressed number of total reads, compare the size of the FASTQ files being analyzed with size of the FASTQ files present on the MiSeq instrument. If the MiSeq instrument has larger files, then re-transfer the sequence data to ensure analysis of the full dataset.

The MiSeq sequencer displays different messages when processing the sequencing data. One message may state _Sequencing Complete_, and another may later state _Analysis Complete_. Transferring the FASTQ data in between these two messages may result in an incomplete LymphoTrack analysis.

Q7:   **Why does the Read Summary contain a read in the top ten that is not in the Merged Read Summary?**

A7:   The "disappearing" read may have been merged with a read ranked above it, or it is possible the alignment of this read was exceptionally poor. During the merging process, reads in the read summary are aligned with those in the unique reads file; however, reads in the read summary have been oriented to match the alignment to the reference and are possibly in the anti-sense orientation compared to the reference gene. This happens only when the alignment is so poor and when a sense or anti-sense alignment is nearly equivalent, and the anti-sense alignment was slightly better. In cases such as these, it is best to ignore the read entirely.

Q8:   **Why is the analysis is taking a long time (over an hour per sample)?**

A8:   The log file provides a breakdown of how long each step is taking. Following the line **_INFO: Decompressing files_**, there is another line that states how many files and how much time the decompression used. This number will be in milliseconds. If decompression is taking several hours, check if the target files are on a fileserver. Typically, running any I/O intensive process like decompression over a network will be much slower. If possible, copy the files over to the local computer, or decompress the files directly before running this software. Make sure that the compressed and decompressed files do not both exist in the same folder. The software will not proceed if both FASTQ and FASTQ.GZ files are in the target folder.

If the subsequent parsing and analysis steps are taking a long time, the FASTQ files may be very large. Examples of large FASTQ files would include when less than 24 indices are included on a single MiSeq run and as a result, the samples sequenced have a larger depth of coverage. A second example would be if the library is unbalanced and contains a higher proportion of a sample with one index versus the proportion of samples with other indexes. The predicted run-time stated in this IFU assumes a 24-index run with balanced libraries.

Q9: **The Run Q30 reported by the MiSeq instrument fails the Illumina specification. If the LymphoTrack analysis is run and the controls look like they performed as expected, did the run really fail?**

A9: The Run Q30 is calculated in real time during the base-calling process. At this point, no other data processing has taken place, such as de-multiplexing and adapter trimming. This means that any sequence from a poor quality index or occurring outside of the adapter sequence is contributing to the Q30 score. Sequence found outside of the adapter sequence is typically low quality, and this scenario happens often with primer dimers. If a library has a sufficient level of primer dimers, this can cause a drop in Run Q30. As a secondary check, the LymphoTrack Reporter output displays the index Q30 score as calculated from all trimmed reads in that index. If an index Q30 score also fails the Illumina specification, consider that index to have failed.

Q10: **Why are the Top 10 merged reads different when the same data is reanalyzed?**

A10: The Top 10 merged read counts may vary slightly when data is analyzed multiple times due to the sorting of sequences when there are multiple sequences with the exact same read count. In cases where multiple sequences have identical read counts, the software may sort and rank them randomly. Observed differences should be insignificant and not affect the interpretation of results.