AlloSeq Assign

Instructions for Use

IFU082
Software Version Number: 1.0.5
Document Version No: 10.0
Issue Date: October 2023

REF ASA1.0

CareDx Pty Ltd,
20 Collie Street,
Fremantle, WA 6160,
Australia

CareDx Pty Ltd

For Research Use Only. Not for use in diagnostic procedures. No claim or representation is intended to provide information for the diagnosis, prevention, or treatment of a disease.

Table of Contents

Chapter 1: Introduction

The AlloSeq Assign software (from here described as “Assign”) and AlloSeq Tx, a targeted hybrid capture NGS sequencing kit, together constitute a system for genotyping genes that are important in matching in transplantation. These genes include the HLA genes. Assign imports sequence data from fastq.gz files generated by an Illumina sequencing instrument, creates a consensus sequence per locus for a sample, enables base call review and editing, and compares the consensus sequence with a library of sequences of reference alleles. The software lists the best matched alleles to assist the user in assigning a genotype.

Assign has the following features and functionality:

- Imports sequences from multiple samples and multiple loci per sample into a user-friendly interface
- View of sample identifiers, loci headers, sequence reads, base calls, and allele assignments
- A complete analysis audit trail
- Enables analysis of exon data only or exon + non-coding
- Generates reports that include CIWD alleles, G groups, and P groups
- Phase-resolve paired-end sequence data from AlloSeq libraries sequenced on the Illumina sequencing system

Refer to IFU083 AlloSeq Tx IFU RUO for details of the associated AlloSeq Tx reagent kits

Reference Databases

Assign compares a sample sequence with a library of sequences from reference databases. Assign utilizes databases from IMGT/HLA, ISBT/ABO and NCBI/CCR5 to assign genotypes to sequences.

IMGT/HLA Database

The IMGT/HLA database includes sequences approved by the World Health Organization (WHO) Nomenclature Committee for Factors of the HLA System. The IMGT/HLA database is part of the international ImMunoGeneTics (IMGT) project (www.imgt.org).

ISBT/ABO Database

The ISBT/ABO database contains sequences provided by the International Society of Blood Transfusion (ISBT) Working Party for Red Cell Immunogenetics and Blood Group Terminology and is available from www.erythrogene.com.

NCBI/CCR5

The NCBI/CCR5 database contains sequences for CCR5 from NCBI. Refer to the Reference release notes for further information.

CWD/CIWD

The CWD/CIWD database lists Common, Intermediate and Well-Documented (CIWD) alleles. Refer to the Annotation section for further information.

Performance Characteristics

24 samples can be analysed in under 30 minutes using a computer with minimum Computing requirements. See table below for import time performance for different PC specifications. Note: These times are provided as a guide only and may vary depending on sample quality and other processes running on the PC.

PC Spec	Assay	# Samples	Average import time
64GB RAM, i7 processor, 3.4GHz	AlloSeq Tx	96	16 minutes
64GB RAM, i7 processor, 3.4GHz	AlloSeq Tx	24	7 minutes
32GB RAM	AlloSeq Tx	96	40 minutes
32GB RAM	AlloSeq Tx	24	7 minutes
16GB RAM	AlloSeq Tx	96	40 minutes

Limitations

Sequencing with short read sequencing technology from Illumina Sequencers

AlloSeq Tx has been optimised for Illumina Sequencers (See IFU). Illumina sequencing is a "short read" sequencing platform that sequences 150 bp of both ends of a DNA fragment typically 500 bp in length. In order to phase two polymorphisms, the distance between these polymorphisms must be within the length of the DNA fragments being sequenced. Polymorphisms outside this region will not be phased. The inability to phase increases the risk of a report that includes a heterozygous ambiguity. A report that lists just one genotype when the genetic sequence has not been fully phased reflects the limited number of alleles in the reference library. An alternative, yet to be described, pair of alleles may have the same consensus sequence and may, in fact be the correct answer.

Limitations of hybrid capture assays for HLA typing/MHC genetic matching

The MHC has evolved following replication and diversification of genetic sequences that include HLA and other sequences. Consequently, the MHC contains numerous homologous genes. Biotinylated probes capture sequence fragments that are up to 20% different from the target loci, so many non-specific sequences are captured. Assign filters these reads and allocates them to the correct gene. However, in some cases, reads are allocated to the wrong gene. It is our experience that this issue is rare and when it happens to such an extent that the misallocation of reads impacts analysis the match report is unable to find a perfect match between the sample consensus sequence and the reference library.

Please note that the misallocation of reads may be a symptom of reduced assay specificity caused by suboptimal assay conditions. Please contact your Technical Support team if this issue is suspected.

A transplant geneticist will understand that whilst extensive testing has been performed, the incredibly diverse nature of the MHC means that all scenarios cannot be tested and, as in all HLA typing assays, sample specific artefacts may complicate analysis and assignment of an HLA type.

Limitations of the IMGT/HLA database

As described above, Assign compares sample sequence with allele sequences from the IMGT/HLA Database and lists the best matched allele pairs. The IMGT/HLA database is updated with sequences from newly described alleles every three months and the best matched alleles are relevant only to the database used at that time. CareDx will release updated Assign reference files every six months.

An expert in the field of transplant genetics/HLA typing is required to interpret this data to provide the most likely HLA type. An expert will understand the limitation of the reference sequence database.

Not all recognized HLA alleles have been sequenced to the same extent. All classical HLA class I genes have been sequenced in exons 2+3, some have been sequenced in exon 2+3+4, some have full coding sequence, and some have full genetic sequence (as defined by the WHO nomenclature committee). Thus, an assigned HLA type from a limited reference sequence may one day be renamed as variants are identified in regions not originally sequenced. A classic example is DRB1*14:01. DRB1*14:01 was defined by sequence in exon 2 of HLA-DRB1. Researchers sequenced exon 2+3 and noticed that an exon 3 polymorphism split DRB1*14:01 to DRB1*14:01 and DRB1*14:54. The newly named allele DRB1*14:54 was determined to be the most common of the 2 alleles in all populations where DRB1*14:01 was identified. Thus, many samples originally typed as DRB1*14:01 from exon 2 sequence may now be recognized as mistyped. It is possible that donors and patients considered to be matched for DRB1*14:01 are in fact mismatched. The implications of this are not known.

The closest we can get to an accurate HLA type is to report 2 x alleles that are fully phased across the entire gene and for the reference alleles to also be sequenced across the entire gene. However, this is still limited by the arbitrary gene boundaries defined by the WHO nomenclature committee and the lack of a definition of where a gene starts and stops. It is still possible that variants exist beyond the boundaries.

Four field typing of class 2 genes

Note: The coverage of introns is incomplete for DPA1, DPB1, DQA1, DQB1, DRB1, DRB3, DRB4 and DRB5. Reporting alleles of these genes to four field resolution may produce results that are inconsistent with genotypes derived from fully sequencing the gene. In addition, alleles that do not have intronic sequence posted to the IMGT database may be promoted in the results table, while similar alleles that have been fully characterized are listed further down the table.

General Limitations

Assign is validated for use with sequence data generated by the AlloSeq Tx products sequenced on validated Illumina instruments. Assign should not be used to analyse data generated any other way. Poor quality data including consensus sequences with background noise or low depth of sequence coverage might result in incorrect consensus base calls and incorrect typing. Assign includes a simple visual interface to view read quality and depth of sequencing coverage, which enables rapid identification of poor read quality and low depth of sequence coverage.

Assign aligns the sequence reads from a sample fastq file against the consensus sequence constructed from the IMGT database for each locus concerned. To facilitate accurate alignment the alleles for the DRB1 loci have been split into different groups according to intronic similarities (for more information refer to the *Locus Consensus Sequence* and *Coverage Summary Panel* sections). Thus, a sample may have DRB1 alleles belonging to two different DRB1 groups or may have two alleles belonging to the same DRB1 group or may be homozygous thus having two copies of the same allele. Similarly, whether the sample contains alleles for HLA-DRB3, -DRB4 or -DRB5 will depend on which haplotype the sample carries, and as such the sample may contain no HLA-DRB3/DBR4/DBR5 alleles at all, may have one allele for one locus, two alleles for a single locus or two alleles for different loci. In instances where a single allele is present for a locus, if the duplicate homozygous check box is selected on the reporting options, then the software will report two copies of the same allele.

Therefore, caution is advised when interpreting the genotype report as an HLA type.

Chapter 2: Computing Requirements and Compatibility

To ensure optimal performance, use the following minimum computing requirements:

- 1 GHz or faster 64-bit Intel quad-core processor, or equivalent
- 16 GB RAM, minimum
- 16 GB available hard disk space

Sequence data files can be stored locally or on a network location. Depending on network performance, the software might experience a significant delay in processing while files are imported from a network location.

Computer Operating System And Software

Assign runs on Windows and has been validated with Windows 10 and Windows Server 2012 operating systems. Assign is not compatible with the following editions of Windows: Embedded (including Windows on the Illumina sequencing system), RT, Starter, Mobile, and Phone, or any hardware that does not support a standard keyboard, mouse, and monitor. Microsoft Excel 97, or later, is required for generating Excel reports from Assign.

Screen Resolution

The recommended screen resolution is 1920 × 1080 pixels. Use multiple monitors or increase the screen resolution on your monitor to expand the number of viewable fields for each locus.

Compatible Data File Formats

Assign is compatible with the FASTQ file format, either zipped (*.fastq.gz) or unzipped (*.fastq). The Illumina sequencing system generates these file formats. For more information about the FASTQ file format, for example, see the MiSeq Reporter Generate FASTQ Workflow Reference Guide (document # 15042322).

Backward Compatibility with Previous Settings

The installer for AlloSeq Assign v1.0.5 includes the settings file Tx17.1 v1.0.2, Tx17.1 v1.0.3 and 1.0.4 to allow users to open projects saved with previous AlloSeq Assign settings. Due to differences in the way indels are handled by the algorithm between versions, when reopening a project previously analysed with older AlloSeq Assign settings, there is a possibility that some insertions may be called as mismatches. This can be rectified by right clicking on the impacted locus and selecting reanalyse.

If a user selects Tx17.1 v1.0.1 settings or imports a project previously analysed with Tx17.1 v1.0.1 settings into AlloSeq Assign v1.0.5 and selects ‘Default’ from the fields dropdown list, non-coding layers will be displayed for all loci.

Chapter 3: Installation

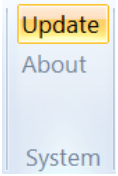
CareDx recommends administrator access to the computer before installing Assign. Ensure the computer is connected to the internet to facilitate system updates with new libraries and other files when needed.

1. Double-click the installer (*.msi) file and follow the prompts to install the software.
2. Review the License Agreement.
3. If you accept the terms in the License Agreement, click **Next** to continue.
4. Select the Installation Folder location. CareDx recommends that you accept the Default location. Click **Next**.
5. Select shortcut options, and then click **Next**.
6. Click **Install** to begin the installation.
7. When the installation is complete, click **Finish**.

Chapter 4: Getting Started

1. Double-click the Assign icon on the desktop or in the installation location.
2. In the Operator Login dialog box, select the operator from the drop-down list. The default operator is **admin**.
3. Enter the password. The default password for the admin operator is **cg01.**
4. Click **Sign In** to start the software.

Add License



1. In the System group, click **Update**.
2. Browse to the Assign license file obtained from CareDx, and then click **Open**.
3. Click **Done**, and then restart the software.
4. Licenses can also be saved in the software installation folder.
   NOTE: License keys are provided with a fixed time frame (e.g., 6 months).
   To obtain a new license key, contact CareDx Technical Support.

Add Operators

1 Click **More** to expand the Operator Login dialog box and access the Edit Users section.
2 In the **Edit Operator** field, enter a new operator name.
3 Enter a password for the new operator and retype the same password for verification.
4 From the **Default Settings** drop-down list, select the appropriate settings file for the assay used**.**
5 Select this setting for all operators analysing AlloSeq Tx data. Operators with sufficient privileges can modify settings directly in Assign.
6 From the **Operator Level** drop-down list, select from the options listed in the table below.
7 Click **Add/Update**.

Operator Level Permissions:

	First Reviewer (edit only)	***First Reviewer (with access to settings)***	***Final Reviewer (with full access)***
Can change settings	No	Yes	Yes
Can edit sequences not yet approved by a final reviewer	Yes	Yes	Yes
Can sign the final review checkbox	No	No	Yes

Updating References

Updated references are provided by CareDx twice yearly and can be downloaded from the CareDx Website. Users should update according to the frequency determined by their local regulatory requirements. To update:

1 Save the files and unzip the files on the desktop or a network location.
2 After logging in and adding operators, Select **Update** on the **System** menu in the ribbon.
3 Click References and CWD.
4 Select all the reference files and click **Open.**
5 Select all the CWD files and click **Open.**
6 After importing the new references and CWD, close the software and relaunch, select the new CWD file from the dropdown and click **Update.**

7 The new references and CWD files will then be applied.

Updating NMDP codes [Optional]

NMDP codes compatible with Assign can be downloaded from https://hml.nmdp.org/mac/files/numer.v3.zip
To update:

1 Save the files and unzip the file on the desktop or a network location.
2 Right click on the file and select rename
3 Delete the .v3 from the filename and save the file.
4 Log in to Assign and Select **Update** on the **System** menu in the ribbon.
5 Click **NMDP Codes**.
6 Select the file and click **Open**.

Import and Analyse Sequences

Importing sequences can take from minutes to hours depending on the number of files imported and the computer system performance. During import, Assign is unavailable, and the application title bar indicates that the software is not responding. The software responds after import and analysis is complete. The first import takes slightly longer to analyse after the initial installation and after a reference update due to simultaneous update of system files.

After you import sequences into Assign, analysis begins automatically. Analysis includes alignment of reads, consensus base calling, phasing and a comparison of the consensus sequence with the reference library.

Importing Errors

FASTQ sample file names typically have the following format:

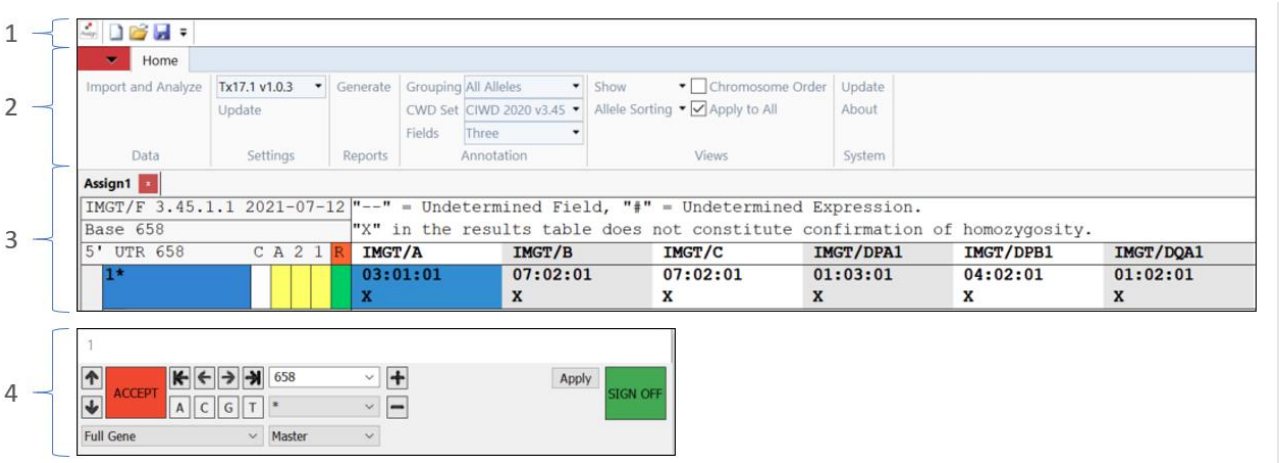
***samplename-HLA-date*_S1.FASTQ.gz**

(e.g., 00001234-HLA-04262016_S5_L001_R1_001.fastq.gz, 00001234-HLA-04262016_S5_L001_R2_001.fastq.gz)
The "-HLA-" portion of the sample name is critical for identification and analysis in Assign.

After the analysis of imported sequences is complete, any of the following warnings appear to indicate that the files were not successfully imported:

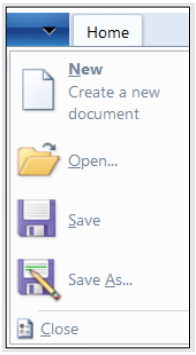
- No sample identifier/delimiter
- There are no dashes (-) in the file name as expected.
- There are no appropriate characters before the first dash to name the sample.
- No target identified/delimiter (i.e., HLA missing).
- Unable to combine 1 or more paired end files.
- Only a single read was imported for 1 or more samples.

Chapter 5: Navigating the Interface



1 **File menu**—Allows you to create new, open, and save sequences in Assign.
2 **Home tab**—Provides access to change settings and views.
3 **Sample panel**—Lists the samples in a project and tracks reviewer comments and the laboratory analysis pipeline. For more information, see *Sample Panel*.
4 **Navigator**—Helps you navigate to base positions of interest. For more information, see *Navigator*.

File Menu

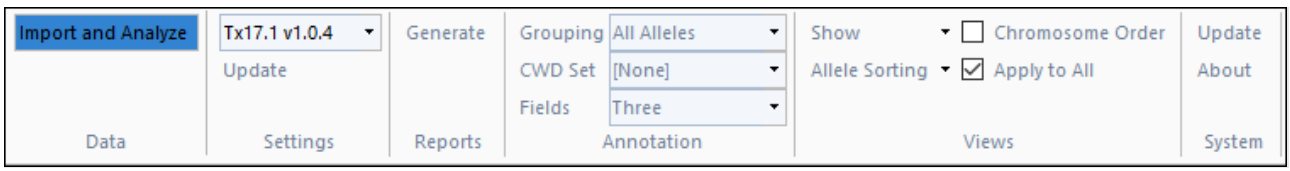


The File menu is located to the left of the Home tab. Click the down arrow to open the File menu. Use the File menu to create new, open, and save projects.

Click **Save** or **Save As** to save aligned reads, the Sample Consensus Sequence, and user edits.
It is recommended to save projects regularly as best practice.

Home Tab

The Home tab is divided into the following groups: Data, Settings, Reports, Annotation, Views and System.



Data

The Data group allows you to import and analyse sequence data.

1. Click an open document tab to choose the import destination. To create a document, click the File button and select **New** or press **Ctrl+N**. In the Data group, click **Import and Analyse.**
2. Navigate to the folder containing the FASTQ files.
3. Use the Ctrl key to select individual files or the Shift key to select a group of files that you want to import and analyse. Use Ctrl + A to select all the files in a folder. The search box at the top right of the import dialog can also be used to find a particular sample or locus for analysis.
4. Click **Open** to begin import and analysis.
   NOTE: Each sample generates a FASTQ file for Read 1 and Read 2. Make sure that you select both FASTQ files.
   For optimal analysis, both Read 1 and Read 2 FASTQ files are imported and analysed simultaneously.

Settings

Settings allows the selection of files containing different analysis and reporting parameters. This menu may not contain options for change. Clicking **Update** will save the current display options as default, including the current pane, number of fields and CWD set.

NOTE: When Settings files for versions prior to v1.0.3 are used with v1.0.3 and above the samples will be analysed against the 17 loci plus ABO and CCR5 although data for these genes is not present.

Reports

The Reports group allows you to generate:

- Genotyping and sequence data in FASTA format.
- Genotyping report file formats are text, Excel, or XML.
- Fragment analysis report.
- HML report.

For more information, see *Generating Reports.*

Annotation

The Annotation group allows you to view and report genotype matches as:

- **G Groups**—Consolidates the Results panel list into G groups.
- **P Groups**—Consolidates the Results panel list into P groups.
- **All Alleles**—Shows all allele matches in the Results panel.

The **CWD Set** shows the Common, Intermediate and Well-Documented (CIWD) alleles in bold in the Results panel. The CIWD files are generated from a combination of version 3.0.0 of the CIWD catalogue[1] and 2.0.0 of the CWD catalogue[2] and are editable to reflect CIWD alleles in your population. CIWD files are updated six monthly with the IMGT reference release to include any alleles that have been updated resulting in an allele split. CIWD files can be edited by opening the files included with each reference release with a text editor such as Notepad. The cat 1 file contains the common and intermediate alleles, the cat 1-2 file contains common, intermediate, and well-documented alleles.

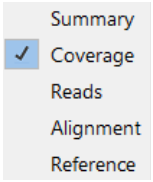
Important: Given CIWD files can be modified by the user to include population specific alleles, CareDx accepts no responsibility for the integrity of these CIWD files once they have been downloaded from the CareDx website. The **Fields** dropdown allows you to limit the view of alleles in the results list to the specified number of fields. If you select Maximum, the sequence used for the typing is automatically expanded to the maximum coverage for all loci. When the number of fields is decreased, alleles can appear several times in the results table. Identical allele pairs can occur because there are 3rd or 4th field differences with resulting mismatch differences. If you select Default, and apply to all is checked, Class I loci will be displayed to 4 fields, Class II to 3 fields and MIC to 2 fields.

[1] Hurley, CK, Kempenich, J, Wadsworth, K, et al. Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0. HLA. 2020; 95: 516– 531. https://doi.org/10.1111/tan.13811

[2] Mack et al. Common and well-documented HLA alleles: 2012 update to the CWD catalogue. *Tissue Antigens*, 81:194-203, March 20, 2013.

Views

The Views group allows you to navigate between panels to view sequence data in different ways. Use the Show drop-down list to choose the **Summary**, **Coverage**, **Reads**, **Alignment**, or **Reference** view.



Summary - Comprises 4 panels:

- Typing Summary panel—Shows the types assigned.
- Motifs
- Quality Summary panel—Shows the percentage of base calls with quality ≥ Q30.
- Coverage Summary panel—Shows the mean depth of sequencing coverage.

Coverage - Shows the mean coverage and base call composition across the amplicon.
Reads - Shows reads used in base calling.
Alignment - Comparison of the Sample Consensus Sequence and the allele pairs lists in Results panel.
Reference - Comparison of the Sample Consensus Sequence and the reference sequences for a locus.

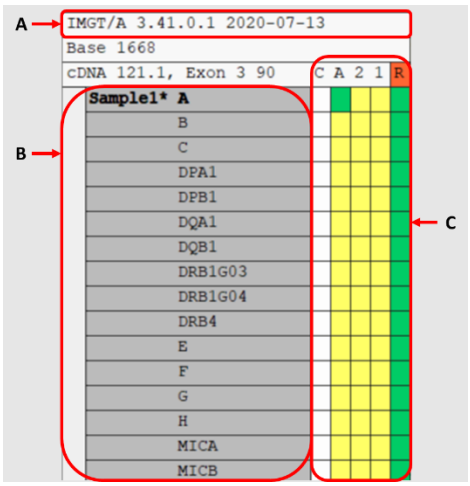
The Allele Sorting drop-down allows you to sort the Results panel by alleles with the greatest reference sequence coverage (alleles that have full gene sequenced are listed in priority over alleles that only have full or limited exons), or by the number of mismatches.

System

The System group allows you to update and view information within the Assign software. Click **Update** to open a dialog box that allows you import keys, references, NMDP codes, and CWD files. Click **About** to open a dialog box that provides the software version and licensing information.

Sample Panel

The Sample panel shows the sample names, the loci sequenced for each sample, the IMGT/HLA reference release, and the status of the review for each locus.



A IMGT/HLA reference and database version
B Sample ID and Loci
C Review hierarchy, report enabling, and locus-specific commenting

Assign References

The first row in the Sample panel shows the reference database used for assignment of allele nomenclature to the sample sequence. For more information, see *Reference Databases*. For information on the second row, see *Coordinates*. The following examples indicate specific information about the databases:

IMGT/HLA Reference

IMGT/A 3.35.0.0 2019-01-23

- **IMGT** - is the reference database
- **A** - is the gene name

- **3.35.0.0** - the first 3 fields indicate the IMGT/HLA database release version; the fourth field indicates the Assign database release version.
- **2019-01-23** - is the date of the IMGT/HLA release

ISBT/ABO Reference

ISBT/ABO 1.1.0 2017-10-23

- **ISBT** - is the reference database
- **ABO** - is the gene name
- **1.1.0** – 1.1 is the ISBT/ABO release version, .0 is the Assign database release version.
- **2017-10-23** - is the date of the ISBT/ABO release

NCBI/CCR5

NCBI/CCR5 NG_012637.1.0

- **NCBI –** is the reference database
- **CCR5 –** is the gene name
- **NG_012637.1.0** – NG_012637.1 is the reference consensus sequence accession number, .0 is the Assign database version.

Samples and Loci

Click a locus to view information for the selected locus in the Sequences and Results panels.

Review Hierarchy

The review hierarchy section of the Sample panel includes five columns, which allow for the option of multiple levels of review and comment for each sample and each locus listed. The columns are labelled C, A, 1, 2, and R. Each review level is tracked and audited.

- **Column C**—By default, the box in column C is white. Right-click the locus to add a comment related to the review. When comments are present, the box changes to light blue. Comments added in column C are included in the full genotyping report.
- **Column A**—By default, the box in column A is yellow. When the sample is verified at all positions indicated in the Navigator, the box in column A changes to green automatically.
- **Column 2**—By default, the box in column 2 is yellow. When the second review is complete, click the yellow box to change it to green, indicating the second review is complete and locking the sample. No further edits are possible unless the box is cleared manually.
- **Column 1**—By default, the box in column 1 is yellow. After the first review is complete, click the yellow box or the Sign Off button on the navigator. The box changes to green, which indicates that the first review is complete.
- **Column R**—A green box in column R indicates that the locus is included in generated reports. Click the box to turn it red to prevent reporting on the locus.

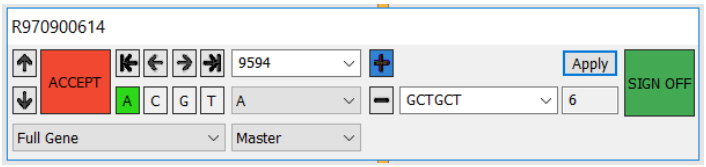
Sample Panel Options

Additional options are available for any locus listed in the Sample panel. To view options, right-click on a locus name. The following options are available:

- **Show Comments**—Shows any quality warnings or comments about a sample locus.
- **Edit Comments**—Opens a field to add or edit comments about the selected sample. These comments appear on the report. A light blue box in column C indicates that a comment is present.
- **Reanalyse**—Removes any edits and trims made to the selected locus.
- **Remove**—Removes the selected locus from the project.

Navigator

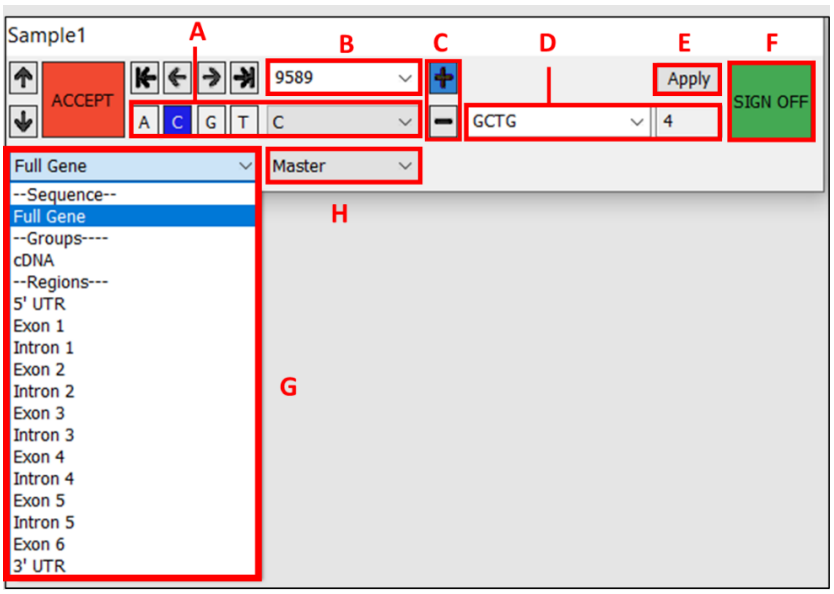
Use the Navigator to navigate to a base position of interest, confirm base calls, or make base call edits. You can drag the Navigator anywhere on the screen.



Basic Navigation

Navigation Icon	Description
	Click the up and down arrows to navigate between loci in the Sample panel.
	Click **Accept** to confirm a base call at a specific position. Click **Reject** to change a previously accepted base call.
	Use the previous and next arrows to navigate between positions to be audited.
	Use the first and last arrows to navigate to the first and last base positions.
	Click **Sign Off** to confirm that the review is complete. If there are no positions left to review this button will appear. Once Sign Off is selected by the first reviewer, the 1st review column in the review hierarchy will automatically turn green and the software automatically moves to next gene in project.
	Click Apply to navigate to a position or interest.

Advanced Navigation



A Base Selection

B Active Nucleotide Position and/or Mismatch List

C Insertion and Deletion Selection

D Insertion Details

E Apply Insertion

F Sign Off

G Nucleotide Position Field Change

H Layer selection

Base Selection

The highlighted base indicates the base call at the current position. Multiple highlighted bases indicate that more than one base was positively identified at the current position. A highlighted + indicates an insertion. A highlighted - indicates a deletion.

1. Editing base calls is performed by adding or removing a base at the active position that is consistent with the consensus base call judged by the operator. This is performed by clicking **A**, **C**, **G**, or **T, + or -**, or selecting from the base selection drop down list.
2. Click **Accept** to accept the selected base and move to the next low confidence position. See *Confidence Indicators* for low confidence positions criteria.
3. To change a previously accepted base call, click **Reject** to enable editing.

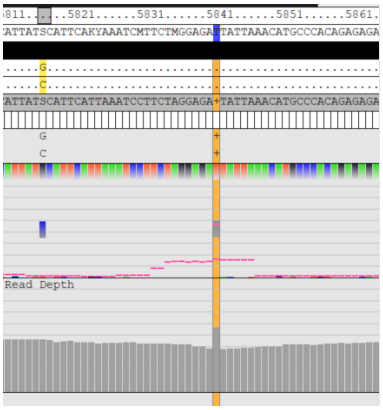
Mismatch List

The Mismatch list shows the selected position. Select the down arrow to show all mismatch positions for the selected allele pair in the active mismatch columns. To move the cursor to a selected position, enter a number in the nucleotide position field and click **Apply** or select an option from the list to move to that position.

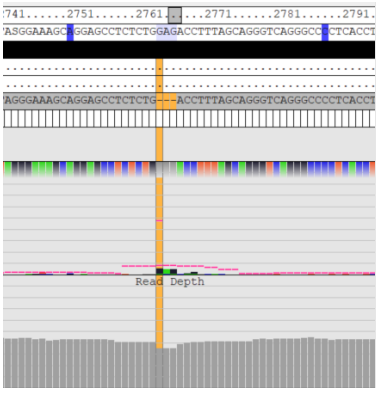
Indel Details

At a position where an insertion or deletion is present, the appropriate **+** (insertion) or **–** (deletion) box is highlighted in blue. The length of the insertion and bases included in that insertion are indicated next to the symbols. The bases inserted can be edited in the box and saved by clicking apply in the navigator.

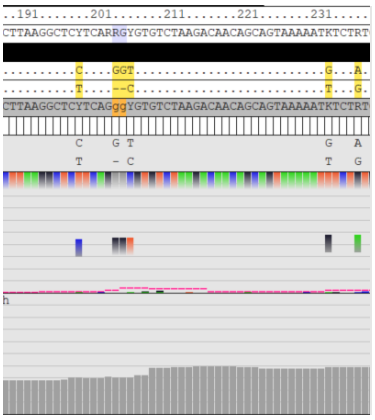
Insertions and deletions are represented in the consensus sequence in the following way:



An insertion is represented by a + in the Sample Consensus sequence. The inserted base(s) fall directly before the +. In the Navigator, the inserted sequence can be viewed in the text box. When both alleles have an insertion at the same position, use the dropdown box to view inserted sequence for both alleles.



A homozygous deletion is represented by – in the Sample Consensus sequence. The – indicates that either a homozygous allele has a deletion, or that both alleles in a heterozygous sample have the deletion. When a deletion is present in both alleles, the deletion will not be displayed in the phase layer.



A heterozygous deletion is represented by lower case letters in the Sample Consensus Sequence, indicating that one allele has the deletion, and one allele does not. Only the first position of the deletion will appear in the phase layer. Deletions will only count for one mismatch in the analysis layer.

Nucleotide Position by Regions or Groups

- The default numbering begins at the first base of the gene. Use the drop-down list to view the numbering system according to the region or group.
- **Full Gene -** Position in gene sequence based on the locus consensus sequence.
- **Regions** - For a particular position of interest, choose the region of the gene, enter the relative position in the mismatch list, and then click **Apply**. Use this feature for quick navigation.
- **Groups -** Groups define a string of regions. cDNA is a group of exons.

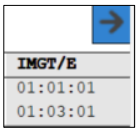
Views

Clicking **Show** allows you to choose between Summary, Coverage, Reads, Alignment and Reference panes.

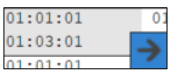
Summary

The following Summary panels are available within the Summary view. Each are accessed as unique windows with the default window being the typing summary.

- Typing Summary panel
- Sequence Motifs
- Quality Summary panel
- Coverage Summary panel



To move between Summary panels, hover over the blue box in the upper-right corner of a Summary view and click the blue arrow that appears. This arrow cycles through the Summary panels.

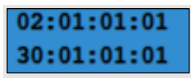


Where there are more loci than can fit on the screen, they can be viewed on the second page of the Summary panel by clicking on the blue arrow on the bottom right of the screen. Click the blue arrow on the bottom left to navigate back to the main summary panel.

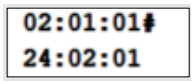
NOTE: Click Chromosome order to change the order of the loci between alphanumeric and chromosomal. Changing the order of the loci in the Summary view also changes the order in the Summary report. To retain this setting click **Update** in the Settings section of the Home tab.

Typing Summary Panel

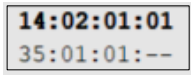
The Typing Summary panel shows the samples and the best matched alleles assigned to each locus for each sample. The Typing Summary panel uses the following markers:



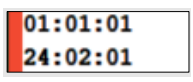
Active sample—A blue highlight indicates the active sample. Click the highlighted area to open the sample and locus in the Coverage view for further investigation. Complete fields indicate an unambiguous typing result.



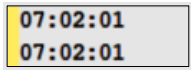
Ambiguous expression—A **#** indicates an ambiguous expression in an allele typing.



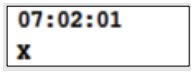
Ambiguous fields—A double dash (--) indicates an ambiguous field in the typing result. For example, --:01 indicates an ambiguity in the first field, 01:-- indicates an ambiguity in the second field, and 01:01:-- indicates an ambiguity in the third field.



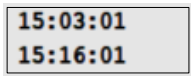
Confidence warning, red—A red box immediately to the left of an allele pair indicates a locus that might warrant further investigation. This warning can indicate insufficient read quality, or mean depth of coverage below 100.



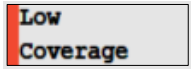
Homozygous calls duplicated, yellow— A yellow box immediately to the left of an allele pair indicates that the duplicate homozygous calls box in the reports dialog box has been checked and that the sample is homozygous at the loci in question. If the duplicate homozygous calls box is unchecked, then the second allele will be displayed as an X, as described below. See the reports section below for more information.



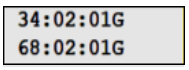
X shown for second allele—An X shown for the second allele indicates that no heterozygous positions were detected in the analysed sequence. The presence of an X in the summary screen does not constitute confirmation of homozygosity.



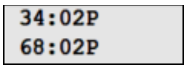
Bold text—Indicates a CIWD allele.



Low Coverage – Indicates a sample with too few reads to align to the references. Typically, this warning is displayed with negative controls or samples that fail to meet quality metrics.



G-only—Displays all alleles for all loci where IMGT have provided G-grouping as G-groups.



P-only— Displays all alleles for all loci where IMGT have provided P-grouping as P-groups.

Sequence Motifs

The sequence motifs tab within the summary display indicates the presence of motifs defined within the reference files. Motifs indicated on this tab include the presence of Bw4/Bw6 motifs in HLA-A, -B and -C, the genotype of the DPB1 SNP position rs9277534 which encodes expression variants, and other SNP positions determined to be important. A complete list of the motifs included in the references is provided in the Reference Release Notes.

Sequence Motifs.

IMGT/A	IMGT/B	IMGT/C	IMGT/DPA1	IMGT/DPB1	IMGT/DQA1	IMGT/DQB1	IMGT/DRB1	IMGT/DRB3	IMGT/DRB4
	Bw6	Bw6		rs9277534:AA					
Bw4	Bw4	Bw6		rs9277534:GG					
Bw4	Bw6	Bw6		rs9277534:AA					

Quality Summary Panel

A quality score, or Q-score, is the probability of an incorrect base call. During Illumina sequencing, each base in a read is assigned a Q-score. A higher Q-score indicates a smaller probability of error. For example, Q30, represents a 1 in 1000 chance of an incorrect call with a corresponding 99.9% call accuracy. The Quality Summary panel shows the percentage of base calls with Q30 or higher scores for each locus. A confidence warning appears for loci when the percentage of base calls with a Q30 score is 75% or less.

Percent base calls with Q30 or better.

IMGT/A	IMGT/B	IMGT/C	IMGT/DPA1	IMGT/DPB1	IMGT/DQA1	IMGT/DQB1	IMGT/DRB1	IMGT/DRB3	IMGT/DRB4
97%	96%	97%	97%	97%	97%	96%	97%	97%	

Coverage Summary Panel

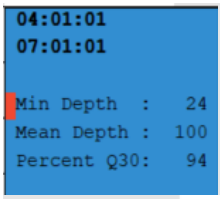
The Coverage Summary panel shows the mean depth of sequencing coverage for each locus in the project. The depth of sequencing coverage is the mean number of bases at each sequenced position in the sequence data. Warnings are present when loci do not meet specifications of 100x average coverage. If the 2 alleles of a locus are split between 2 groups (e.g., DRB1*01 and DRB1*03), a warning appears when the group does not meet specifications of 50x average coverage.

Average sample coverage.									
IMGT/A	**IMGT/B**	**IMGT/C**	**IMGT/DPA1**	**IMGT/DPB1**	**IMGT/DQA1**	**IMGT/DQB1**	**IMGT/DRB1**	**IMGT/DRB3**	**IMGT/DRB4**
224	231	221	201	171	214	202	196	201	
95	111	100	80	59	72	73	29 32		

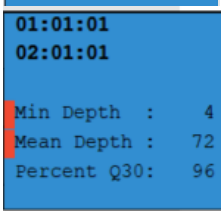
Sample with coverage warnings displayed in red.

Gene Summary

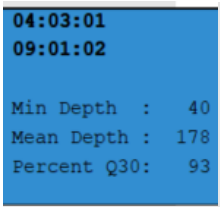
Hover the mouse cursor over an allele typing to see the following summary information.



Min Depth— Minimum sequence coverage across the regions that are covered by the probe panel. Warnings shown as red bars next to the metric are present when the min depth is below the threshold set in the references.



Mean Depth— Mean sequence coverage across the regions that are covered by the probe panel. Warnings shown as red bars next to the metric are present when the mean depth is below 100x coverage, or 50x for alleles of split loci.

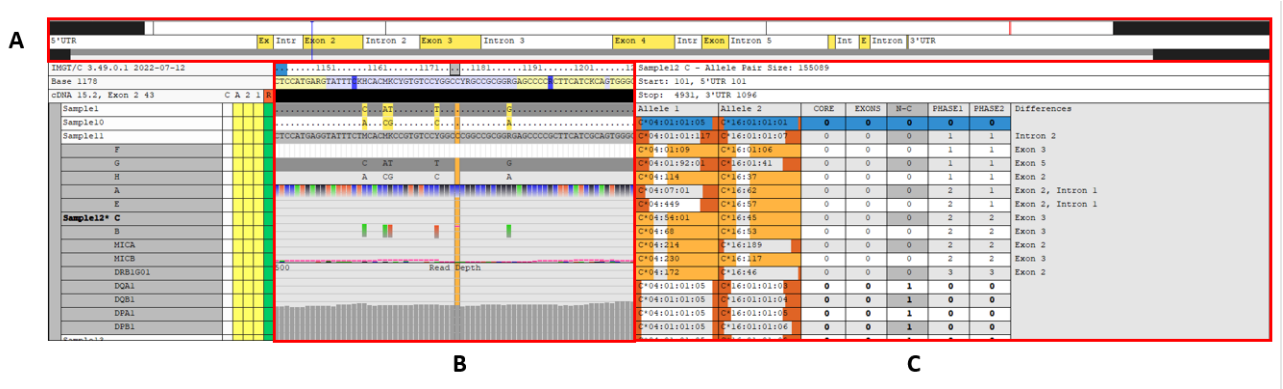


Percent Q30—Percentage of bases > Q30.

NOTE: When alleles from more than 1 group are present, the summary shows results for the first listed allele.

Coverage View

The Coverage view comprises the Confidence Plot, Locus Structure, and Phase Block Display, the Sequences panel, and the Results panel. To see the Coverage view, in the Views group, click **Show**, and then click **Coverage**.



A. **Confidence Plot, Locus Structure, and Phase Block Display**—Shows a view of the high-level locus structure, such as UTRs, introns, and exons, indicates base call confidence and position, and indicates blocks of phased sequence.

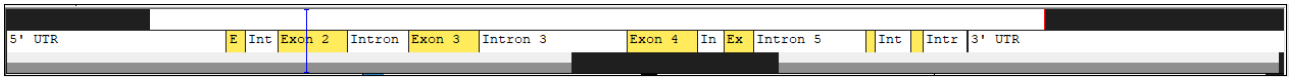
B. **Sequences Panel**—Shows consensus reference sequence, sample sequence, base calls, depth of sequencing coverage, base call quality, and alternate sequence reads.

C. **Results Panel**—Shows the allele combinations that most closely match the sample sequence and shows the mismatches between the sample sequence and the reference sequence when present.

Move the Coordinate scroll box in the Sequences panel to find positions where base call confidence is low. Use the Results panel to find mismatches with allele pairs.

Confidence Plot, Locus Structure, And Phase Block Display

Three rows span the width of the screen at the top of the Coverage view.



Top: **Confidence Plot**
Middle: **Locus Structure**
Bottom: **Phase Block Display**
Click a row to move the blue line, which indicates the region in view in the Sequences panel.

Confidence Plot

The Confidence Plot uses colours to show positions where base call confidence might warrant further investigation.

Black indicates no coverage. Common reasons for no coverage include the following:

- The region is outside area of probe coverage for the analysed locus
- The reference sequence contains an insertion or deletion that is absent in the sample

Increasing shades of red indicate any of the following conditions:

- Sequence coverage ≥ Q30 is below the minimum depth threshold for the locus.
- Mean quality score for base calls at this position is low
- A base above noise threshold is not called in consensus
- A base below noise threshold is called in consensus

White indicates complete coverage.

Locus Structure

The Locus Structure uses yellow to indicate an exon/coding sequence and white to indicate an intron/noncoding sequence.

Yellow - Exons that are in the active Mismatch Column of the Results panel.
White - Noncoding regions that are in the active Mismatch Column of the Results panel.
Grey - Regions that are not in the active Mismatch Column of the Results panel, for example, this occurs for exons not covered by the Core layer analysis when only the core layer is active.

Phase Block Display

In the Phase Block Display, regions are either grey or black.

Light Grey/dark Grey - Regions where bases are phased together. Light grey/dark grey shading corresponds to the Phasing tracks.
Black - Sections where the phase relationship between polymorphic positions can't be established.

In this example, there are 2 x blocks of phased sequence. Phasing may not be possible if the sequenced fragments or smaller than the distance between the polymorphic positions.

Sequences Panel

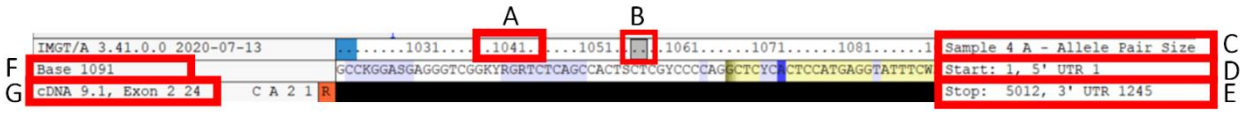
The Sequences panel on the Coverage view is composed of the Sequences section and the Base Calling section.

Sequences Section

The Sequences section of the Sequences panel includes information from comparisons of reference sequences with sample sequences. These rows are updated when you select different allele pairs in the Results panel.

1 Coordinates
2 Locus Consensus Sequence
3 Sequence Edit Indicator
4 Allele 1 Reference Sequence
5 Allele 2 Reference Sequence
6 Sample Consensus Sequence
7 Confidence Indicator
8 Phasing Track

1. Coordinates



A Gene coordinates
B Coordinate scroll box—Drag the grey box to scan along coordinates
C Sample name and locus
D Start position and location
E Stop position and location
F Highlighted base coordinate in the exon, intron, or UTR (from Sequences panel)
G Highlighted base associated codon coordinate in the gene (from Sequences panel)

2. Locus Consensus Sequence

The Locus Consensus Sequence represents a collection of common variants and motifs. Rare variants are not included.

- Yellow indicates exonic/coding sequence.
- White indicates intronic/non-coding sequence.
- Light Blue indicates deletions present in some alleles. The number of highlighted bases indicates the size of the deletion.
- Dark Blue indicates insertions present in some alleles. The inserted bases fall directly before the highlighted bases.

For HLA-DRB1, the Sample Consensus Sequence is compared with sequences of alleles that have been divided into groups with similar intronic sequence structure. Therefore, the consensus sequence represents the consensus of the best matched allele group. HLA-DRB1 alleles are split into 4 groups: DRB1G01, DRB1G03, DRB1G04, and DRB1G07.

3. Sequence Edit Indicator

The Sequence Edit Indicator row shows a color-coded edit status and acceptance status of each base in the sequence. The base edit status changes when you edit the originally called sequence using the Navigator.

	Colour Code	Edit status	Acceptance status
	Black (default)	Not edited	Not accepted
	Green	Not edited	Accepted
	Blue	Edited	Not accepted
	Blue/Green	Edited	Accepted

4. Allele 1 Reference Sequence

The Allele 1 Reference Sequence shows the IMGT/HLA reference for an allele in the highlighted allele pair selected in the Results panel. Allele 1 Reference Sequence is shaded dark grey.

- A base is displayed in this row when the allele sequence differs from the Sample Consensus Sequence for the sample, or the position is heterozygous.
- Blank positions indicate that the reference sequence is missing for the selected allele.
- A dot (.) indicates that the allele sequence is identical to the observed sequence at the selected position.
- A star (*) indicates a sequence with no intron information in the reference library.

5. Allele 2 Reference Sequence

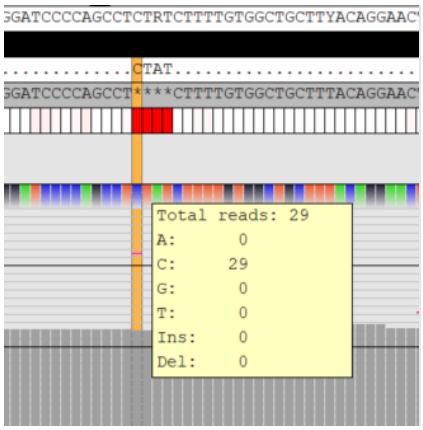
The Allele 2 Reference Sequence shows the IMGT/HLA reference for an allele in the highlighted allele pair selected in the Results panel. Allele 2 Reference Sequence is shaded light grey.

- A base is displayed in this row when the allele sequence differs from the Sample Consensus Sequence for the sample, or the position is heterozygous.
- Blank positions indicate that the reference sequence is missing for the selected allele.
- A dot (.) indicates that the allele sequence is identical to the observed sequence at the selected position.
- A star (*) indicates a sequence with no intron information in the reference library.

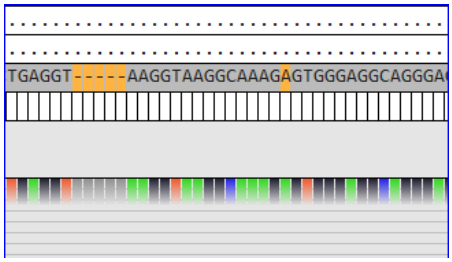
6. Sample Consensus Sequence

The Sample Consensus Sequence shows the consensus sequence of the sample sequenced with the AlloSeq Tx 17 Sequencing Panel.

Positions that are below the minimum depth threshold will be excluded from the Sample Consensus sequence. This is indicated by a star (*) as shown in the image below. For most loci the minimum depth threshold is set to 30 reads. Refer to the Reference release notes for each reference release for the thresholds.



Orange shading in the sample consensus indicates a polymorphism that is not included in the combined sequence for the gene.



7. Confidence Indicator

The Confidence Indicator is a per-base representation of the Confidence Plot. The confidence of a base call at any given position can vary based on several factors, including balance of the alleles, noise threshold, depth of coverage, and sequence quality. White in the Confidence Indicator denotes a high confidence base call. A red Confidence Indicator denotes base calls in which any of the following conditions have occurred:

1. <75% of reads have a quality score of Q30 or higher
2. A base above noise threshold is not called in consensus
3. A base below noise threshold is called in consensus
4. Edited positions

8. Phasing Track

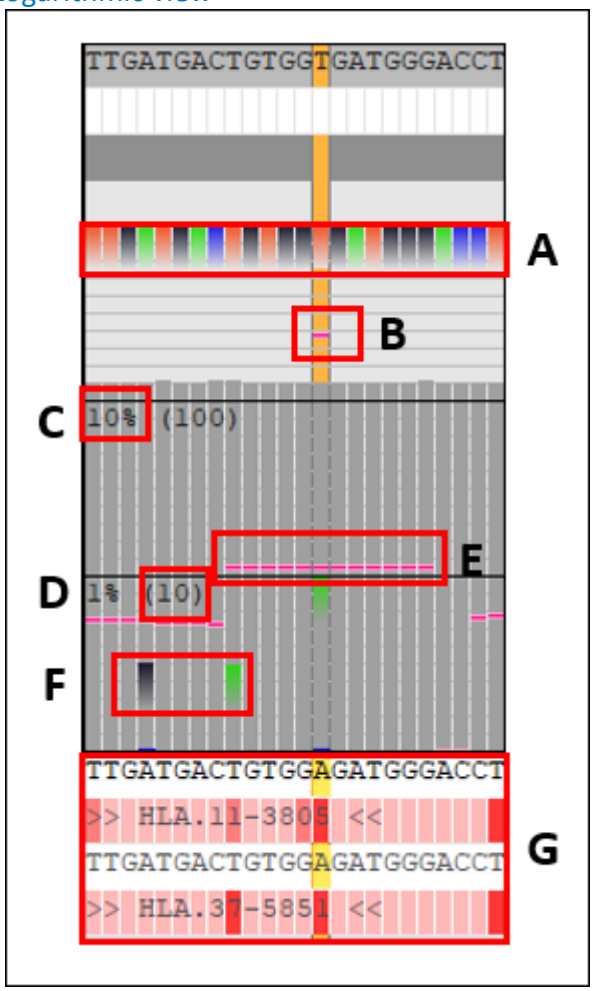
For heterozygous allele combinations, the Phasing Track rows show the phase relationship between bases connected by single reads or paired reads. A phase assignment is made only when most phasing sequences are concordant. Dark grey shading indicates the phase relating to allele 1 and light grey indicates the phase relating to phase 2.

Sequence Depth of Coverage Summary and Base Calling

The central window shows the sequence depth of coverage (DoC) information as histograms. This window summarises the sequences that contribute to the consensus. The grey bars indicate the DoC at each position and the sequence content is indicated by the coloured blocks. The data can be shown in linear or logarithmic form. The location of the coloured blocks indicates the percent contribution of a specific base to the DoC.

CTRL + L allows switching between **logarithmic** and **linear** view.

Logarithmic view



A Primary base called. The following colours indicate the most frequently occurring base call for a given position.

A—Green
C—Blue
G—Black
T—Red

B Approximate allele ratio. When a base location is highlighted, the top pink line indicates the approximate mean read depth ratio of the second allele present in the sample for heterozygous loci.

C Base call ratio. Shown using a logarithmic scale, as follows:

- Lowest section has a ratio between 0% and 1%
- Middle section has a ratio between 1% to 10%
- Highest section has a ratio between 10% to 100%

D Depth of sequencing coverage. Shown with grey bars for each base using the logarithmic scale in parentheses:

- Lowest section shows coverage depth between 0x and 10x
- Middle section shows coverage depth between 10x and 100x
- Highest section shows coverage depth between 100x and 1000x

NOTE: If the coverage is below 30, no call is shown.

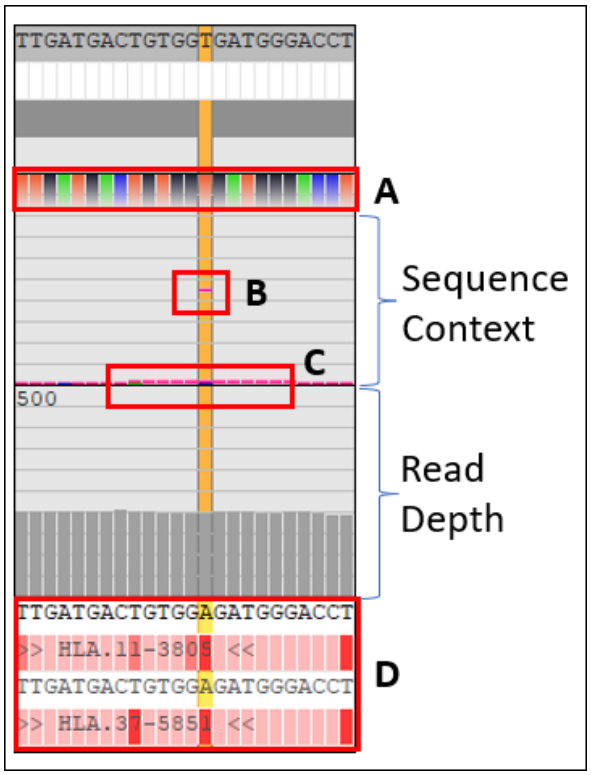
E Approximate noise threshold. Noise is a common by-product of specificity, sequencing errors, and sequence alignment. Assign dynamically sets a threshold for noise at any given base position. A pink dashed line indicates the Approximate Noise Threshold at all base locations. Typically, base calls below the noise threshold are not called.

F Other base calls. Shows the base calls that differ from the most frequently occurring base call for a given position and use the same colour indicators used in the Primary Base Called section.

G Sequence reads covering that base position that were not included in the consensus sequence for the sample

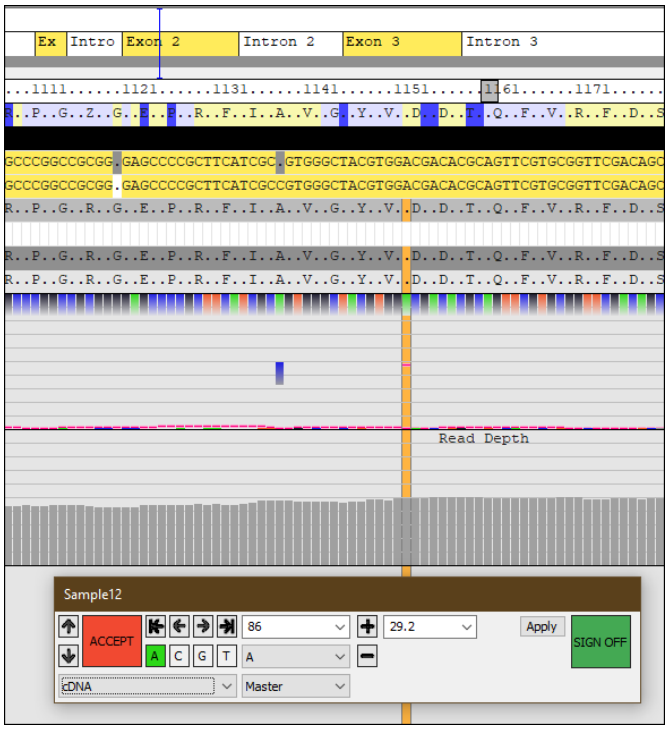
Linear View

- **A** Most frequently called base
- **B** Approximate allele ratio
- **C** Approximate noise threshold
- **D** Sequence reads covering that base position that were not included in the consensus sequence for the sample. Each line in Sequence Context is 10%, each line in Read Depth is 25 reads.



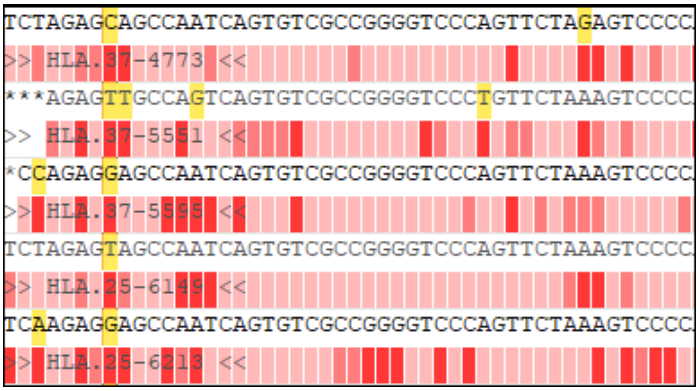
Amino acid view

To activate the amino acid view, select cDNA from the dropdown in the navigator, click in any exon and press **Ctrl + A**.



Sequence Reads

The Sequence Reads section contains calls that are not included in the Sample Consensus Sequence at the highlighted base position. All reads contributing to the alignment can be viewed in the **Reads View**.



The quality of the base call for alternate reads, as reported in the FASTQ file, is shown below the sequence in a gradient of red.

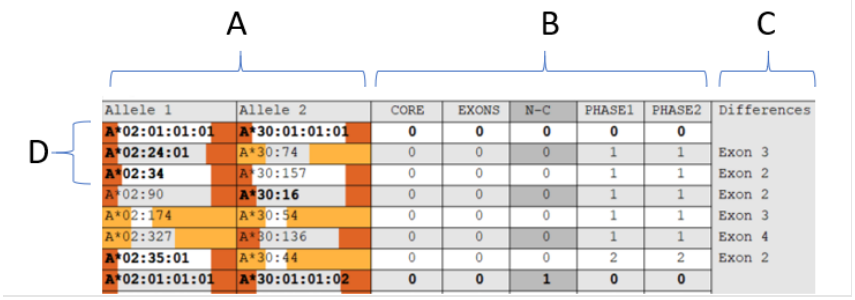
- **Dark red**—Lowest quality base call.
- **Light pink**—Highest quality base call.

Right-click on a read to open a menu.

- **Copy Sequence**—Places all the bases in the read on the clipboard
- **Copy Aligned**—Places the bases used during alignment on the clipboard
- **BLAST Sequence**—Submits the full sequence to NCBI BLAST
- **Copy Pair**—Places the sequence of a read pair on the clipboard

Results Panel

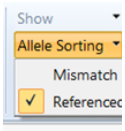
The Results panel lists all the IMGT/HLA allele pairs that are best matched to the Sample Consensus Sequence. The Results panel also provides information for each of the allele pairs listed.



Allele 1	Allele 2	CORE	EXONS	N-C	PHASE1	PHASE2	Differences
A*02:01:01:01	**A*30:01:01:01**	**0**	**0**	**0**	**0**	**0**	
A*02:24:01	A*30:74	0	0	0	1	1	Exon 3
A*02:34	A*30:157	0	0	0	1	1	Exon 2
A*02:90	**A*30:16**	0	0	0	1	1	Exon 2
A*02:174	A*30:54	0	0	0	1	1	Exon 3
A*02:327	A*30:136	0	0	0	1	1	Exon 4
A*02:35:01	A*30:44	0	0	0	2	2	Exon 2
A*02:01:01:01	**A*30:01:01:02**	**0**	**0**	**1**	**0**	**0**	

A Allele columns
B Sequence Mismatch columns
C Differences column
D CIWD alleles are bolded

Allele Columns



As a default all allele pairs appear in order based on the number of mismatches they contain when compared to the Sample Consensus Sequence. Allele pairs with no mismatches appear at the top of the columns followed by pairs with increasing numbers of mismatches. When the Duplicate Homozygous Calls box is unchecked in the reports tab, if no heterozygous positions are detected in the sequence used for the typing (default is all exons), the Allele 2 column contains an X. The presence of an X does not constitute confirmation of homozygosity. When a heterozygous position is found in the active sequence, a second allele is reported. When the results table or report is truncated to 2 or 3 fields, the second allele can appear identical. Selecting Referenced from the Allele Sorting menu sorts the allele pairs by the most completely referenced alleles.

Common, Intermediate and Well-Documented (CIWD/CWD) Alleles

In the Results panel and Summary panel, CIWD/CWD alleles are shown in bold, as described above.

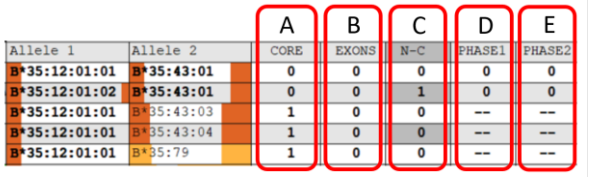
IMGT/HLA Reference Coverage

The allele pairs are banded white and grey by alternating rows for ease of viewing. Sometimes, the allele includes orange, which indicates that a part of the reference sequence is missing in the IMGT/HLA reference for that allele. Dark orange indicates the allele has genomic coverage in the IMGT database and the missing sequence is in the

non-coding region, where light orange indicates the allele has cDNA coverage only in the IMGT database and the missing sequence is in the coding region. The allele container width is directly proportional to the sequence length.

Mismatch Columns

The number of mismatches in the selected regions appear in the columns to the right of the allele pairs.



		A	B	C	D	E
Allele 1	Allele 2	CORE	EXONS	N-C	PHASE1	PHASE2
B*35:12:01:01	B*35:43:01	0	0	0	0	0
B*35:12:01:02	B*35:43:01	0	0	1	0	0
B*35:12:01:01	B*35:43:03	1	0	0	--	--
B*35:12:01:01	B*35:43:04	1	0	0	--	--
B*35:12:01:01	B*35:79	1	0	0	--	--

A Mismatches in core exons; including additional positions for non-coding expression mutations that are known. For class I alleles “Core” is exons 2-4 and exons 2-3 for class II
B Mismatches in remaining exons
C Mismatches in noncoding sequence (introns and UTRs)
D Mismatches in phasing of Allele 1
E Mismatches in phasing of Allele 2
Dashes in the phase layers indicate a mismatch to the reference sequence.

Navigating the Mismatch Columns

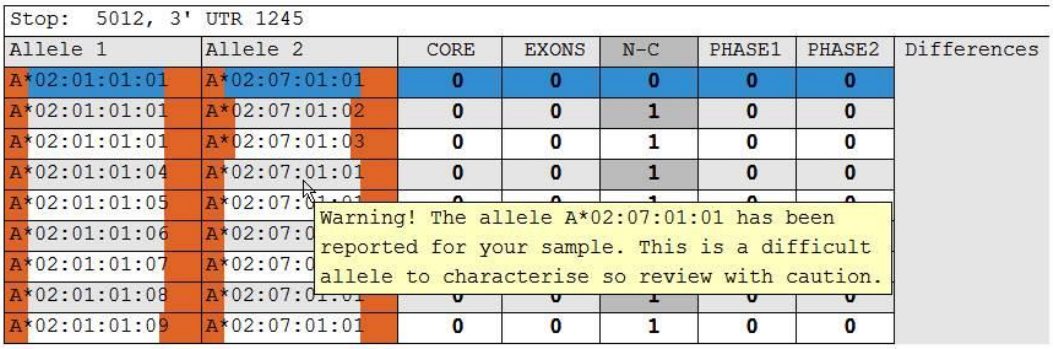
Of the 5 possible mismatch columns, the **Core and Exons** columns are activated upon import. When the default display is set to Maximum the noncoding column will be present for Class I but not Class II. Click the **Core** column header to expand or collapse the **Exons** column. Click the **Exons** column header to expand or collapse the **N-C** column. The phase mismatch columns are present only if necessary, to resolve a sequence ambiguity.

Differences Column

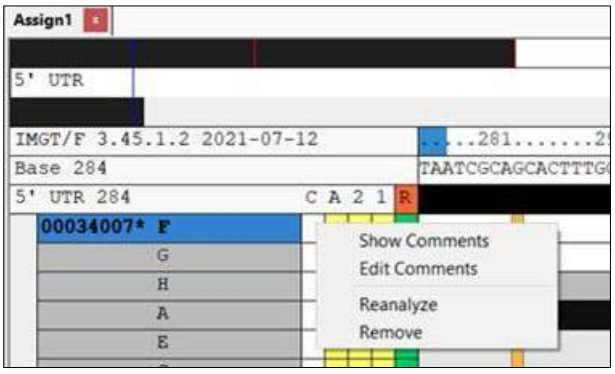
The Differences Column indicates the location of differences between the allele pairs relative to the first allele pair listed. Where ambiguities exist, the regions in which they might be resolved are indicated in this column.

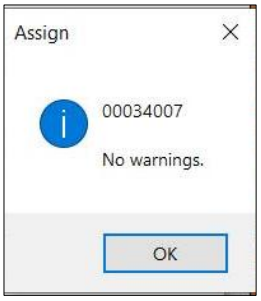
Warnings

On the screen view, when the mouse is hovered over an allele with a comment, the comment will be displayed in a popup box.



If there are no warnings another message will be displayed when entering in the comments section:





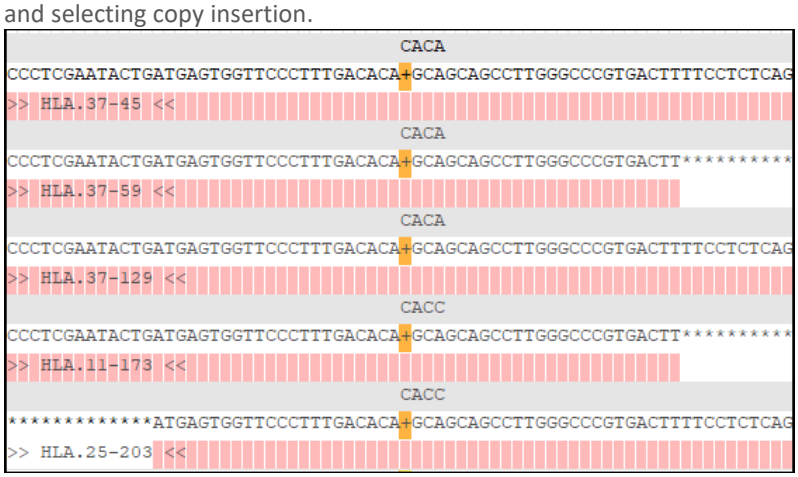
Reads View

The Reads view shows the sequence reads used in base calling for the selected position.

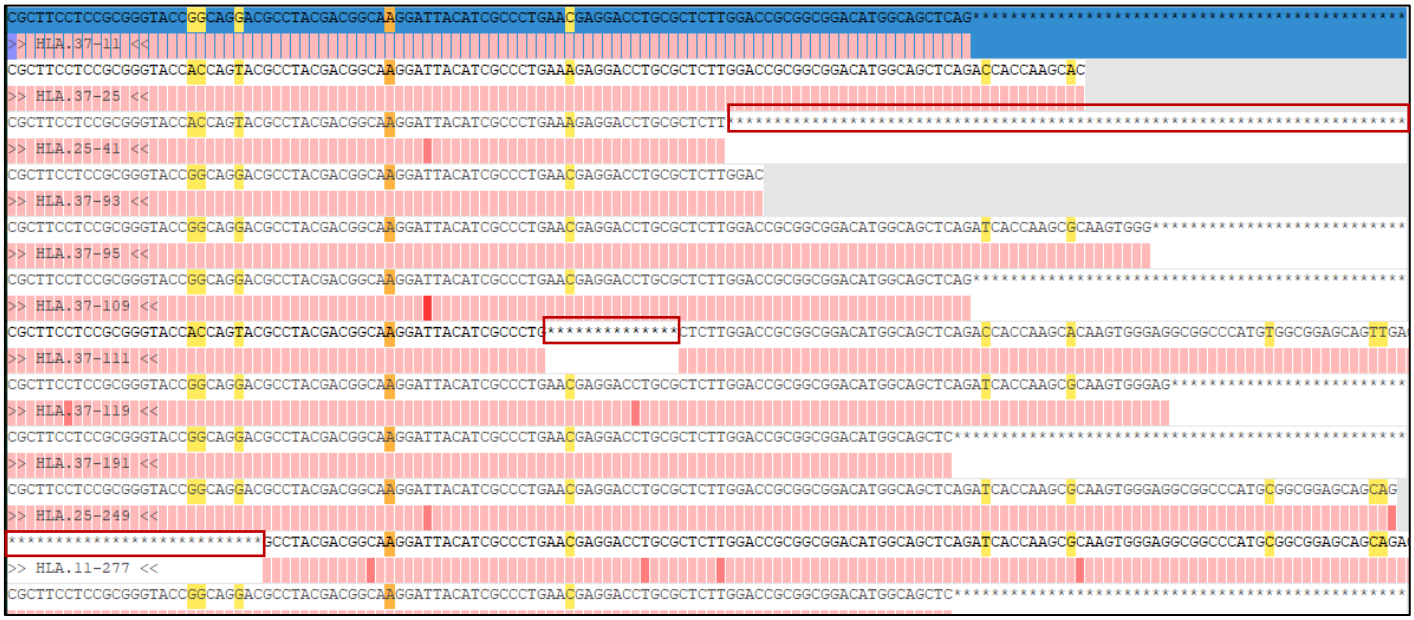
To see the Reads view, in the Views group, click **Show**, and then click **Reads**.

A Nucleotide Sequences
B Base call quality from FASTQ file—Quality is shown in light pink (highest quality) to dark red (lowest quality).
C Read scroll arrows—Use the scroll arrows to view the next batch of reads. You can also navigate by pressing Page Up and Page Down on your keyboard. To hide reads for a specific nucleotide at the base selected, press the Shift key and the nucleotide letter simultaneously. The reads reappear using the same keys. For example, press Shift + A to hide the reads calling A at the selected position. Right-click on a sequence to open a menu that includes the options to copy the sequence to the clipboard, send the sequence to BLAST for alignment, or display warnings for a sample.

Reads that contain inserted sequence when compared to the reference sequence are indicated by a '+'. The inserted sequence is displayed above the + in the reads view. The inserted sequence can be copied by right clicking on the + and selecting copy insertion.



Stars (*) in the reads view indicate the links between read pairs.



Alignment View and Reference View

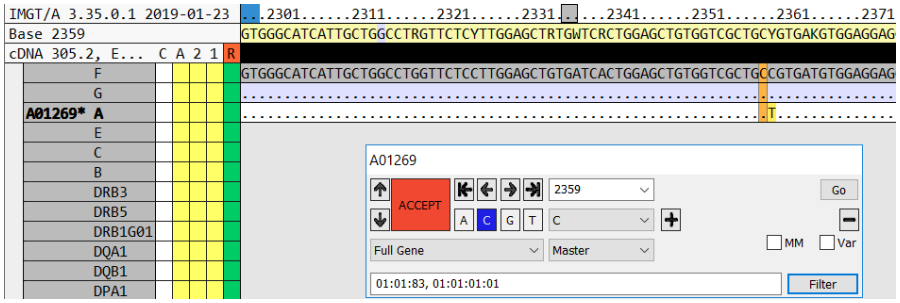
The Alignment view and Reference view provide comparisons of the Sample Consensus Sequence and your data against the IMGT sequences.

Alignment View

The Alignment view shows a comparison of the Sample Consensus Sequence and the allele pairs listed in the Results panel. Click the headings **Allele 1** or **Allele 2** to add or remove the contribution from the alleles in that column. To see the Alignment view, in the Views group, click **Show**, and then click **Alignment**.

Reference View

The Reference view shows a comparison of the Sample Consensus Sequence and the reference sequences for a locus. To see the Reference view, in the Views group, click **Show**, and then click **Reference**. You can limit the reference alleles that appear in the Reference view. Enter the reference alleles of interest into the lower field of the Navigator, and then click **Filter** to the right of the text field. Alleles that contain the text entered in the box are shown. You can enter
multiple entries, separated by commas, into the filter field.



Chapter 6: Generating Reports

Types of Reports

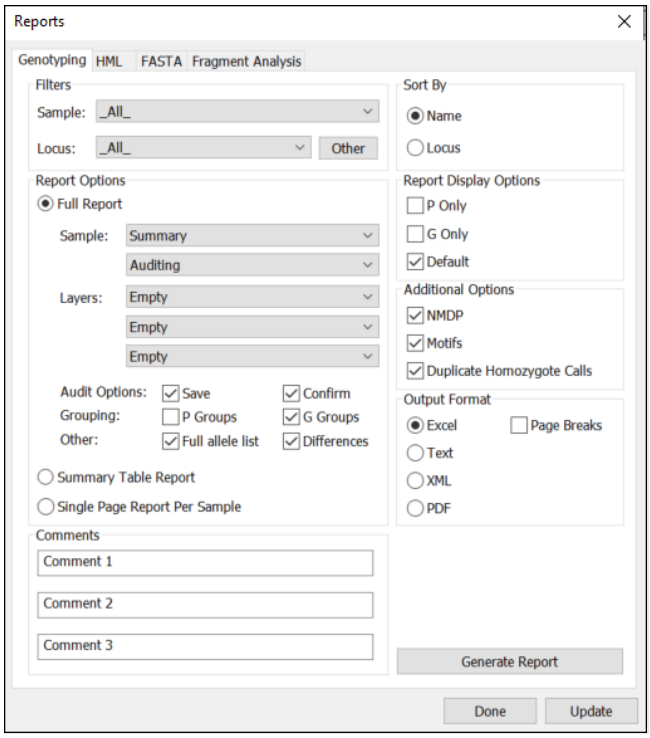
Assign generates a genotyping report, FASTA report or HML report.

- Genotyping Report—Reports on a single sample or locus or all samples and loci in the project.
- Sequence Data Report in FASTA format—Produces a fasta file of the Sample Consensus Sequence using IUPAC designations.
- Fragment Analysis—Reports for the distribution of DNA fragments clustered and read by the Illumina sequencer and imported into Assign.
- HML report- Reports on a single sample or locus or all samples and loci in the project in HML format.

Reports can be customized with a logo, page numbers, date and time, and other references about the report. For more information, see *Changing the Full Report Logo*.

Genotyping Report

Click **Generate** to launch the reporting tool.



Generating a Full Report

A full genotyping report includes a header with your preferred logo, page numbers, created date and time, sample name and references used, and the CWD set used.

1. On the Genotyping tab, in the Filters section, use the **Sample** list to select the samples to include in the report. Select **All** to include all samples in the project.
2. From the **Locus** list, with Tx17 settings, select either All, 6 Loci, 11 Loci, 17 Loci or Other, with Tx8 settings select from All, 8 Loci or Other.
   a. Select **All** to include all loci in the project in the report.
   b. Select 6 Loci to include HLA-A, -B, -C, -DRB1, -DQB1, -DPB1 in the project report.
   c. Select 11 Loci to include HLA-A, -B, -C, -DRB1, -DRB3/4/5, -DQB1, -DQA1, -DPB1, -DPA1 in the project report.
   d. Select 17 Loci to include the 11 loci listed above plus HLA-E, -F, -G, -H and MICA/MICB in the project report.
   e. Selecting Other will allow the user to select which loci to report from the list which is opened by selecting the Other button.

3 In the Sorting section, select either sample **Name** or **Locus** to sort the report.
4 Select the **Full Report** radio button.
5 In the Full Report section, use the **Sample** lists to select **Summary** or **Auditing** from the list. Select **Empty** if a selection list is not needed.

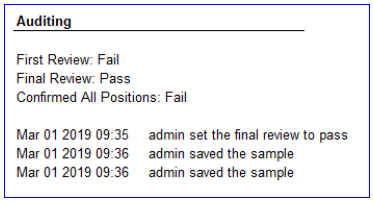
- **Summary**—Includes any warnings regarding the typing and the allele pairs that are compatible with the Sample Consensus Sequence (as edited) for each locus selected in the Filters section. Additional modifications to this section of the report are available in Summary Options.
- **Auditing**—For each Locus selected in the Filters section, the Auditing report includes the reviewer status as either Pass or Fail and whether all positions have been confirmed as either Pass or Fail. The report stamps the date, time, and user for each item passed. Additional modifications to this section of the report are available in Audit Options.

6 In the Full Report section, use the **Layers** lists to select the level of layer detail to include in the report. Select **Empty** if a selection list is not needed.
    - **Sequences**—For each locus selected in the Filters, the Sequences report prints the Sample Consensus Sequence (as edited).
    - **Edit List**—For each locus selected in the Filters, the Edit List report shows the edited positions, the edit that was made, and the user that made the edit.
    - **Mismatch List**— For each locus selected in the Filters, the Mismatch list displays the best mismatches to the sample. The mismatch limits apply to the entire gene sequence. This feature is useful for novel alleles.
7 In the Summary Options section, select the checkbox for each option to include in the report.

Summary Option	Description
Full Allele List	Includes all alleles.
P Groups	Reports ambiguous alleles to P groups, all other alleles to 2 fields. For more information, see hla.alleles.org/alleles/p_groups.html.
G Groups	Reports ambiguous alleles to G groups, all other alleles to 3 fields. For more information, see hla.alleles.org/alleles/g_groups.html.
NMDP	Provides the NMDP code corresponding to the matching allele pair for a locus.
Differences	Includes the information in the differences column of the Results panel.
Motifs	Include Bw4/Bw6 and DPB1 expression variants, and other motifs as listed in the Reference Release Notes.

NOTE: If a sample result is ambiguous, the software will automatically apply G or P group resolution to the highest level of typing possible. If it is not possible to condense the ambiguity as a G or P group, then the list of ambiguous allele combinations will be listed in the report in numerical order. In addition, the ambiguity strings will be reported on a new “Ambiguities” tab on the Summary Report.

8 In the Audit Options section, select **Save** to generate a history of save and load events. Select **Confirm** to include a history of reviewer confirmations.



9 In the Output Format section, select from the following formats:

- **Text**—Generates a report of the selected options into text format.
- **Excel**—Generates a report of the selected options into an Excel spreadsheet.
- **XML**—Generates a report of the selected options into a tagged *.xml file that is best suited for importing into an external database.
- **PDF** – Generates a report of the selected options in PDF format.
- **Page Breaks**—Adds page breaks to the Excel spreadsheet.

10 [Optional] Select **Duplicate Homozygote Calls** to print 2 alleles instead of allele 1 and X for putative homozygous samples. This box will also impact how the homozygous calls are displayed in the summary view and allele panes. To retain this preference, either check of uncheck the box as required, click update in the reports tab, and then update in the settings section of the home ribbon.
11 Click **Generate Report**. Excel reports generate and open automatically in Excel. Text or XML reports generate when you choose a save location on your computer.

Summary Table Report

A Summary Table report includes a header with your preferred logo, page numbers, created date and time, software version, references used, the CWD set used, and the Operator generating the report. The summary panes are displayed in separate tabs in the excel workbook.

1 On the Genotyping tab, in the Filters section, use the **Sample** list to select the samples to include in the report. Select **All** to include all samples in the project.
2 From the **Locus** list, with Tx17 settings select either All, 6 loci, 11 Loci, 17 Loci or Other, with Tx8 settings select from All, 8 Loci or Other.
   a. Select **All** to include all loci in the project in the report.
   b. Select 6 Loci to include HLA-A, -B, -C, -DRB1, -DQB1, -DPB1 in the project report.
   c. Select 11 Loci to include HLA-A, -B, -C, -DRB1, -DRB3/4/5, -DQB1, -DQA1, -DPB1, -DPA1 in the project report.
   d. Select 17 Loci to include the 11 loci listed above plus HLA-E, -F, -G, -H, MICA and MICB in the project report.
   e. Selecting Other will allow the user to select which loci to report from the list which is opened by selecting the Other button.
3 Select the **Summary Table Report** radio button.
4 In the Additional Options, select from NMDP and Motifs if desired.
   a. **IMPORTANT:** Selecting the **Motifs / Additional Options** will include the motif information on the summary tab of the excel report. This additional line per sample may generate a report format which is incompatible with some LIMS and database utilities. If this incompatibility occurs with your LIMS, it is recommended to leave the Motifs box unselected in the summary options (use Update option to retain settings) and utilise the Motifs tab in the excel report instead.
5 Select the desired Output Format: Text or Excel.
6 [Optional] Select **Duplicate Homozygote Calls** to print 2 alleles instead of allele 1 and X for putative homozygous samples. This box will also impact how the homozygous calls are displayed in the summary view and allele panes. To retain this preference, either check or uncheck the box as required, click update in the reports tab, and then update in the settings section of the home ribbon.
7 Click **Generate Report**. Excel reports generate and open automatically in Excel. Text reports generate when you choose a save location on your computer.

Single page report per sample

The single page report displays the alleles for each gene as well as the G groups and P groups in a single page for each sample. The gene content is also listed for each gene. Note that for genes where 4 fields have been selected, and the content for the gene is less than 98.5, the 4 field allele will not be listed.

1 On the Genotyping tab, in the Filters section, use the **Sample** list to select the samples to include in the report. Select **All** to include all samples in the project.
2 From the **Locus** list, with Tx17 settings select either All, 11 Loci, 17 Loci or Other.
   a. Select **All** to include all loci in the project in the report.
   b. Select 11 Loci to include HLA-A, -B, -C, -DRB1, -DRB3/4/5, -DQB1, -DQA1, -DPB1, -DPA1 in the project report.
   c. Select 17 Loci to include the 11 loci listed above plus HLA-E, -F, -G, -H, MICA and MICB in the project report.
   d. Selecting Other will allow the user to select which loci to report from the list which is opened by selecting the Other button.
3 Select the **Single Page Report Per Sample** radio button.
4 Select the desired Output Format: Text, Excel, XML or PDF.
5 [Optional] Select **Duplicate Homozygote Calls** to print 2 alleles instead of allele 1 and X for putative homozygous samples. This box will also impact how the homozygous calls are displayed in the summary

view and allele panes. To retain this preference, either check of uncheck the box as required, click update in the reports tab, and then update in the settings section of the home ribbon.

6 Click Report. Excel reports generate and open automatically in Excel. Text, XML and PDF reports generate when you choose a save location on your computer.

Changing the Full Report Logo

You can alter the image by directly editing the Excel template included with Assign. To change the logo, open Excel then choose the Genotyping.xlt template file. In a default installation, the template is located in C:\ProgramData\CareDx\AlloSeq v1.0.5\data\templates. For a custom installation folder, navigate to the appropriate folder and then choose data\templates\Genotyping.xlt. To replace the logo image, under Print, view Page Setup and edit the header and footer.

Changing the PDF Report Logo

You can alter the image by replacing the .png image included in Assign. To change the logo, save your logo as "CareDx-logo.png" and replace the image located in C:\ProgramData\CareDx\AlloSeq v1.0.5\data\templates. It is recommended to use an image that is approx. 513 x 219 pixels.

FASTA Report

The FASTA file format is a simple text-based format that has become a standard bioinformatics tool for representing genetic sequences. The FASTA format begins with a description line that includes a greater than symbol (>) followed by the unique identifier which could be the name of the sample/locus/entry. The next line in the FASTA is the Sample Consensus Sequence using the IUPAC designations.

1 Click **Generate** to launch the reporting tool, and select the FASTA tab.
2 In the Output Filters and Numbering section, use the **Sample** list to select an individual sample to include in the report. Select **All** to include all samples. The sample name is included automatically in the FASTA description line preceding the sequence.
3 From the **Locus** list, select an individual locus to report on the selected samples. Select **All** to include all the loci for the samples selected. Select the checkbox to insert the locus name into the FASTA file (eg, > Sample Name_IMGT/A).
4 From the **Layer** list, select a single layer to restrict output. Select the checkbox to insert the layer name into the FASTA file.
5 From the **Group** list, select a designated group of regions to restrict output.
6 From the **Region** list, select a designated region, such as an exon. Select the checkbox to insert the region name into the FASTA file.
7 In the Sort by section, select either sample **Name** or **Locus** to sort the report.
8 In the Options section, select the **Pad Ends** checkbox to add N base calls to each sequence to cover the entire amplicon.
9 Click **Generate Report**, and then choose a save location on your computer.

Fragment Analysis

The Fragment Analysis is an Excel report that provides details of the distribution of fragment sizes imported into Assign for each sample and locus.

1 Click **Reports** to launch the reporting tool.
2 On the Fragment Analysis tab, do 1 of the following:
   a. Select a single sample in the project and select a single locus or all the loci from the dropdown lists
   b. Select all the samples and select a single locus or all the loci from the dropdown lists
3 Click **Report** to generate the fragment analysis.

The Fragment Analysis automatically opens in Excel.

Customised Locus Reporting

There are two locus sets predefined in the Tx17 settings (11 loci or 17 loci).

11 Loci: HLA-A, -B, -C, -DRB1, -DRB3/4/5, -DQB1, -DQA1, -DPB1, -DPA1

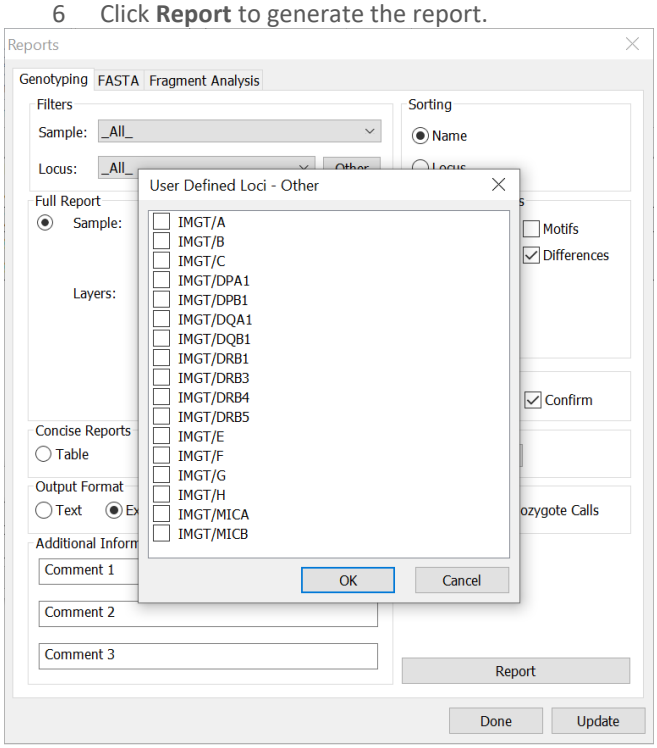
17 Loci: Those listed above plus HLA-E, -F, -G, -H and MICA/MICB.

There is one locus set predefined in the Tx8 settings (8 Loci)

8 Loci: HLA-A, -B, -C, -DRB1, -DPB1, -DQB1, ABO, CCR5

If required, a report can be generated to display specific loci:

1 Click **Reports** to launch the reporting tool.
2 In the locus box, select other.
3 Click the **Other** button.
4 Select the loci you would like to report by checking the relevant boxes and then click ok.
5 Click update in the reports tab and the settings section of the home ribbon to save this setting for future use.
6 Click **Report** to generate the report.



HML report

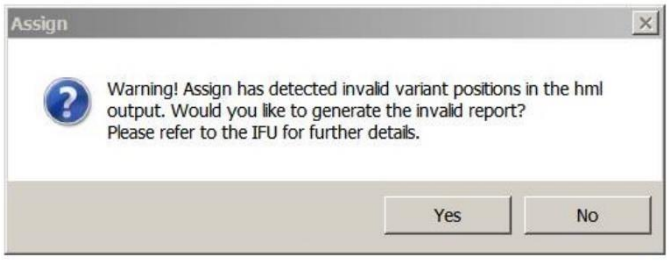
HML reporting functionality has been included in AlloSeq Assign v1.0.5 to enable submission of data to the IHIWS.

The HML report generated from AlloSeq Assign is specific to the 1.0.1 schema provided by NMDP.
For further information on HML and the schema used, refer to https://bioinformatics.bethematchclinical.org/hla-resources/hml/ and http://schemas.nmdp.org/.

To generate the HML report:

1. Click **Generate** to launch the reporting tool, and select the HML tab.
2. Use the **Sample** list to select the samples to include in the report. Select **All** to include all samples in the project.
3. Click **Modify Additional Information** and enter details. The Additional Information will save after the reports window is closed.
4. As default, Assign will generate a GLString per locus. Selecting the **Summative GLString** option will report a single GLString for each sample.

IMPORTANT When generating the HML report, AlloSeq Assign performs a validity check of the variant positions for each gene to ensure that no degenerate nucleotides are reported. If a gene fails this validity check, Assign will display the following error message:



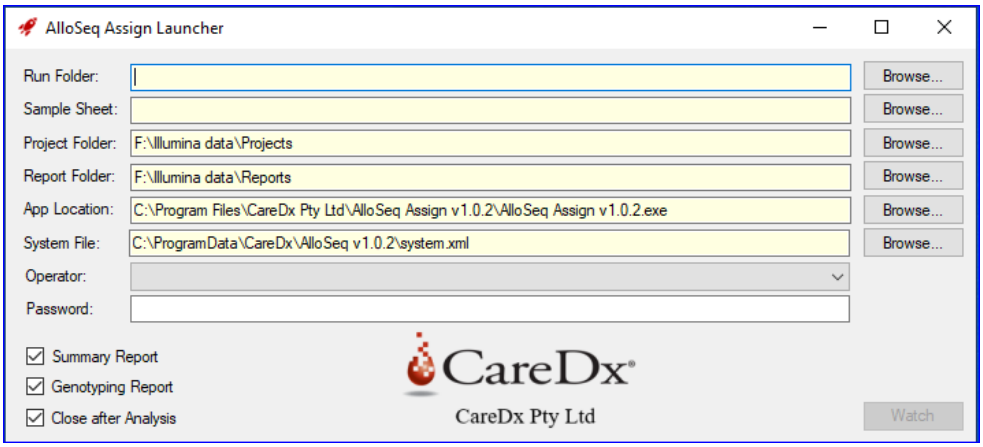
Selecting No to the message will generate a HML report that does not include the impacted gene.
To enable users to edit and submit the affected gene, selecting Yes will generate 2 reports, one without the impacted gene, and the other with the invalid variants. Please note that HML files generated with invalid variants cannot be submitted to the IHIWS. Invalid variant positions have been observed in 1/25500 genes tested, due to lack of phase where the software cannot attribute the variant position to either allele.

Chapter 7: AlloSeq Assign Launcher

AlloSeq Assign software version 1.0.2 and above is compatible with AlloSeq Assign Launcher.

AlloSeq Assign Launcher is a standalone application designed to automatically launch AlloSeq Assign once the fastq files have been created by the sequencer.

See IFU098_AlloSeq Assign Launcher for details of how to use this application.



Chapter 8: Glossary

CIWD/CWD: The Common, Intermediate and Well-Documented (CIWD) alleles identifies the subset of HLA alleles for which the frequencies are well known (common), or the alleles identified multiple times through the use of sequence-based typing methods (well-documented). For more information on the CIWD alleles see: https://www.ihiw18.org/component-immunogenetics/download-common-and-well-documented-alleles-3-0/
Or for the CWD list see: http://igdawg.org/cwd.html

G groups: HLA alleles that have identical nucleotide sequences across the exons encoding the antigen binding domains (exon 2 and 3 for HLA class I and exon 2 only for HLA class II alleles). For more information see: http://hla.alleles.org/alleles/g_groups.html

P groups: HLA alleles that have identical protein sequences across the exons encoding the antigen binding domains (encoded by exon 2 and 3 for HLA class I and exon 2 only for HLA class II alleles). For more information see: http://hla.alleles.org/alleles/p_groups.html

HLA nomenclature: Assign converts sequences into HLA nomenclature version 3.0, established in 2010, in agreement with the WHO Nomenclature Committee for Factors of the HLA System (www.imgt.org).

The HLA nomenclature uses the following format: **HLA-A*02:01:01:02L**

HLA	The HLA Prefix
-	The hyphen separates the gene name from the HLA prefix.
A	The gene name.
*****	The asterisk separates the gene name from the sequence information and indicates genetic typing.
02	Field 1—The allele group.
:	A colon separates fields.
01	Field 2—Differentiate alleles with unique protein sequence.
:	A colon separates fields
01	Field 3—Synonymous DNA substitutions within coding regions of the gene.
:	A colon separates fields.
02	Field 4—Differences in the noncoding regions of the gene.
L	This expression modifier is present regardless of the no. of fields reported. To date, the following modifiers are possible: • N denotes Null—An allele that is not expressed. • L denotes Low—An allele encoding a protein with significantly reduced or low cell surface expression. • S denotes Secreted—An allele encoding a protein that is expressed as a secreted molecule only. • Q denotes Questionable—An allele with a mutation that has previously been shown to have a significant effect on cell surface expression but is not confirmed. Therefore, its expression remains questionable.

MICA/B nomenclature:
The MICA/B nomenclature uses the following format: **MICB*002:01:01**

MIC	The MIC Prefix
B	The gene name.
*****	The asterisk separates the gene name from the sequence information and indicates genetic typing.
002	Field 1— Differentiate alleles with unique protein sequence.
:	A colon separates fields.
01	Field 2— Synonymous DNA substitutions within coding regions of the gene.
:	A colon separates fields.
01	Field 3— Differences in the noncoding regions of the gene.

ABO nomenclature:
The ABO nomenclature uses the following format: **ABO*A1.01**

ABO	The gene name.
*	The asterisk separates the gene name from the sequence information and indicates genetic typing.
A1	Field 1— Differentiate alleles with unique protein sequence.
.	A period separates fields.
01	Field 2— identifies a sequence difference.

CCR5 nomenclature:
The CCR5 nomenclature uses the following format: **CCR5*W**

CCR5	The gene name.
*	The asterisk separates the gene name from the sequence information and indicates genetic typing.
W/D	W represents the wildtype allele and D represents the Δ'32 (delta-32 deletion, rs333 number, c.554_585del)

Degenerate base designations: The consensus sequence rows in the Sequences section (rows 2 and 6) include International Union of Pure and Applied Chemistry (IUPAC) degenerate base designations.

IUPAC Code	Bases	Description
W	A T	Weak
S	C G	Strong
M	A C	Amino
K	G T	Keto
R	A G	Purine
Y	C T	Pyrimidine
B	G C T	not A
D	A G T	not C
H	A C T	not G
V	A C G	not T
N	A C G T	all bases
*		no base call

BW4/Bw6 Sequence Motifs: The sequence motif functionality in Assign reports presence of defined motifs based on nucleotide or amino acid sequences in the sequence alignment. The recognition of Bw4 and Bw6 motifs reported from Assign are based on the amino acid sequences reported by Gumperz et al.

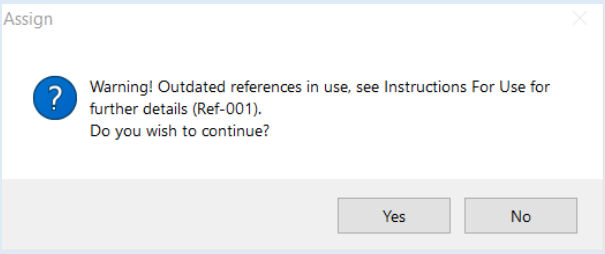
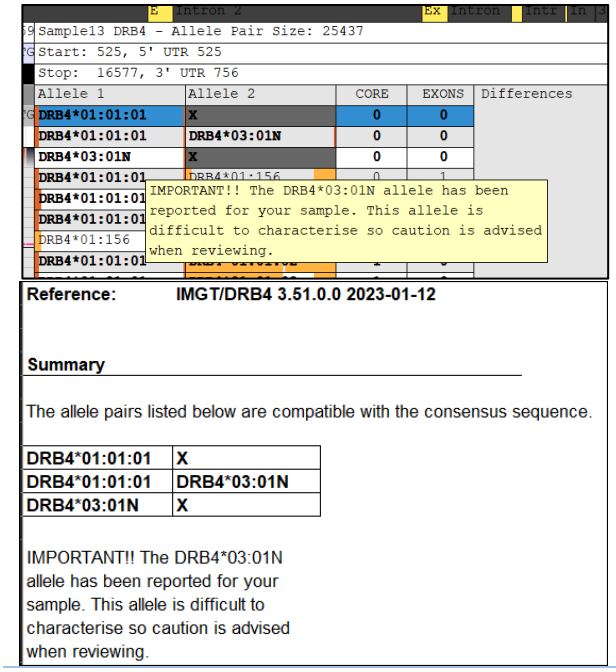
Serological epitope	Class I locus	Class I position 77	80	81	82	83
Bw4	HLA-A,B	N	I	A	L	R
	HLA-B	N	T	A	L	R
	HLA-A	S	I	A	L	R
	HLA-B	S	T	L	L	R
	HLA-B	D	T	L	L	R
Bw6	HLA-B	S	N	L	R	G
	HLA-B	G	N	L	R	G

Figure 1: Class I HLA Sequence Motifs Determining the Bw4 and Bw6 Serological Epitopes.

[1]*Gumperz, J., Litwin, V., Phillips, J., Lanier, L. and Parham, P. (1995). The Bw4 public epitope of HLA-B molecules confers reactivity with natural killer cell clones that express NKB1, a putative HLA receptor. Journal of Experimental Medicine, 181(3), pp.1133-114*

Shortcut keys

Keys	Description
Ctrl + L	Change from log/linear
Ctrl + M	Hide/unhide map area
Ctrl + F	Find sequence
Ctrl + A	Activate amino acid view
Shift + A/G/C/T	Filter by base
Right Arrow	Move one base Right
Ctrl + Right Arrow	Move to the next flagged position
Ctrl + Shift + Right Arrow	Go to the end of the consensus sequence
Left Arrow	Move one base Left
Ctrl + Left Arrow	Skip to the previous flagged position
Ctrl + Shift + Left Arrow	Go to the start of the consensus sequence
Up Arrow	Move to the previous sample
Shift + Up Arrow	Reduce the size of sequence depth of coverage display
Down Arrow	Move to the next sample
Shift + Down Arrow	Increase the size of sequence depth of coverage display
Tab	Confirm base call at current position
A/C/G/T/M/K/R/W/D/S/Y/B/V/H/N	Edit the base at the current position
Shift + I	Toggle read information

Error message	Description
Ref-001 	AlloSeq Assign version 1.0.3 and higher includes a number of changes that rely on using reference versions 3.45.1.1 and newer. Using references prior to this may result in issues with inappropriate phasing and additional ambiguities with DQB1*03.
Difficult to characterise allele warning 	This warning message will display in the coverage pane when hovering over the alleles and in the full report when alleles that are difficult to characterise are listed as a possible result. These difficult to characterise alleles may incorrectly report as the best match due to large indels.

Chapter 9: Support and Contact Details

Website: https://labproducts.caredx.com/

For ordering details, please refer to the CareDx website: https://labproducts.caredx.com/

For Technical Support please email: techsupport-global@caredx.com

Chapter 10: Revision History

Version	Date	Modification
1.0	17 Oct19	Initial issue of AlloSeq Assign IFU.
1.1	11May20	Details added to Indel Details and Confidence Indicators. Instructions added for generating a concise report. Limitation of software added. Report section updated to include PDF format.
1.2	19Jun20	Section references added to Limitations.
2.0	04Dec20	Updated Limitations
3.0	30Mar21	IFU updated and reviewed to reflect changes in v1.0.2.1270: - Added reference to AlloSeq Assign Launcher: Chapter 7: AlloSeq Assign Launcher - Added description of orange shading in the sample consensus: Chapter 5: Sequences Section - Added further description of differences column: Chapter 5: Differences Column - Added reference to Tx8 settings: Chapter 4: Add Operators Chapter 6: Generating Reports - Added ABO and CCR5: Chapter 1: Reference Databases, Chapter 5: Home Tab and Assign References Chapter 6: Customised locus reporting Chapter 8: Glossary - Updated Performance Characteristics with import times for different PCs: Chapter 1: Performance Characteristics - Added amino acid view: Chapter 5: Sequences Section - Added reference to AlloSeq Tx 8: Chapter 1: Introduction and Performance Characteristics
4.0	01Apr21	Added limited liability disclaimer for user modified CIWD files in Chapter 5: Annotation.
5.0	23Nov21	IFU updated and reviewed to reflect changes in v1.0.3: - Updated reports section with new screenshot for reporting window. - Added method for changing the pdf report logo. - Removed references to concise reports - Updated full report process based on changes to the report window. - Updated summary table report and single page per sample report methods. - Added HML reports section. - Added legacy shortcuts and error message to the appendix - Updated backward compatibility - Added four field typing of class II genes to limitations.
6.0	15Sep22	Removed reference to the AlloSeq Tx kits and updated refer to the AlloSeq Tx IFU for details of the AlloSeq Tx Kits. Performance Characteristics to be updated to reflect all SKUs for AlloSeq Tx. Add SW version number. Reissued by Hira Meraj 20-Oct-22 (DCR 2022-688)

7.0	27Feb23	Updated for v1.0.4: ▪Backward compatibility updated ▪Updated getting started screenshot ▪Removed reference to Tx17.1 ▪Updated home tab screenshot ▪Coverage view screenshots ▪Added details of grey shading in the allele reference sequences, coverage map and phasing tracks ▪Updated sequence DOC screenshots ▪Amino acid view ▪Updated Reads view screenshots, removed reference to read direction indicator, added new insertion display and stars link the read pairs. ▪Added note to single page report per sample to describe the gene content and limitation. ▪Added summative GLSTring option to HML report section. ▪Removed duplicated shortcut keys
8.0	15May23	Software versions in “Backward Compatibility with Previous Settings” section have been updated.
9.0	29Aug23	Warning section has been added in the results panel. Removed wrong specification in the revision history section for version 8.0. Updated to v1.0.5 throughout.
10.0	09Oct23	Removed Ctrl+G use instructions and instructions for use of P only and G only summary reports.