Conexio Assign™ v1.0
TruSight HLA Analysis Software
User Guide

FOR RESEARCH USE ONLY

ILLUMINA PROPRIETARY
Part # 15059520 Rev. B
March 2015

illumina®

Revision History

Part #	Revision	Date	Description of Change
15059520	B	March 2015	Update the document cover.
15059520	A	February 2015	Initial release

Introduction

The Assign™ software assists with the assignment of a human leukocyte antigen (HLA) type. The software is designed to analyze data from libraries prepared with the Illumina TruSight™ HLA Sequencing Panel for DNA and then sequenced on the MiSeq® system.

Using Assign, you can import sequence data, perform base calling, edit sequences, and compare a consensus sequence with a library of sequences of HLA alleles.

Assign has the following features and functionality:

- Import sequences from multiple samples and multiple loci per sample into a user-friendly interface
- View sample identifiers, loci headers, sequence reads, base calls, and allele assignments
- Complete analysis audit trail
- Sort allele assignments based on regions of each locus, such as core exons, all exons, or entire sequences
- Generate reports that include CWD alleles, G groups, and P groups
- Perform sample-to-sample and run-to-run QC analysis
- Phase-resolve paired-end sequence data from Illumina TruSight HLA libraries sequenced on the MiSeq system

IMGT/HLA Database

Assign compares a sample sequence with a library of sequences from known alleles listed in the IMGT/HLA database, which comprises sequences of the human major histocompatibility complex, known as the human leukocyte antigen (HLA). The IMGT/HLA database includes sequences for 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).

Performance Characteristics

Assign can import sequence data from up to 24 samples generated by the TruSight HLA Sequencing Panel into a single project.

Base Call Accuracy

Assign contains a unique base caller that improves the accuracy of heterozygous base calls. However, sequence data quality and depth of sequencing coverage can influence base call accuracy.

Limitations

Poor quality data including sequences with background noise or low depth of sequencing coverage might result in incorrect 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 sequencing coverage.

Assign compares a sample sequence with a library of sequences from known alleles listed in the IMGT/HLA database. The report lists those allele combinations in the library that are identical to the sample sequence. However, the same sequence might be derived from alleles yet to be described and whose sequence is not yet part of the library. Therefore, caution is advised when interpreting the genotype report as a HLA type.

Computing Requirements and Compatibility

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

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

Assign program files require approximately 15 MB of hard disk space.

A single sample prepared with the TruSight HLA Sequencing Panel and sequenced on a MiSeq using a paired-end 250 bp run produces the following file formats and sizes:

- 85 MB in *.fastq.gz file format (zipped)
- 200 MB in *.fastq file format (unzipped)

Alternatively, store sequence data files on a network location and import into Assign over a network connection. Depending on network performance, the software might experience a significant delay in processing while files are copied from a network location.

Computer Operating System and Software

Assign runs on Windows and requires Windows Vista, Windows 7, Windows 8, Windows Server 2008, or Windows Server 2012 operating systems.

Assign is not compatible with the following editions of Windows: Embedded (including Windows on the MiSeq), 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 reports from Assign.

Compatible Data File Formats

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

Installation

Install Assign on a local computer or on a shared network drive. Installing the software on a shared network allows other users to log in, share settings across computers, and store license keys in a single location.

Local Installation

Illumina recommends administrator access to the computer before installing Assign. Make sure that 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. Accept the terms in the License Agreement, and then click **Next**.
4. Select the Installation Folder location. Illumina recommends that you accept the default location. Click **Next**.
5. Browse to the location of the Assign license files received from Illumina. Select the files, and then click **Open**.

NOTE
When you receive future license files, you can store the files in the installation folder location.

6. Click **Install** to begin the installation.
7. When the installation is complete, click **Finish**.

Shared Installation

Install Assign on a shared or networked computer using the same steps for installing the software locally.

NOTE
If network or shared drive permissions prevent installation on the shared drive, login to the shared computer as an admin to install the software.

Performing a run using MiSeq v2 chemistry requires a minimum of 25 Gb free space on the C:\ drive for a single 24-sample analysis run. For optimal performance, make sure that each computer connected to the shared computer meets or exceeds the minimum hardware and operating requirements. Assign uses the processing resources of the connected computer instead of the resources of the shared computer.

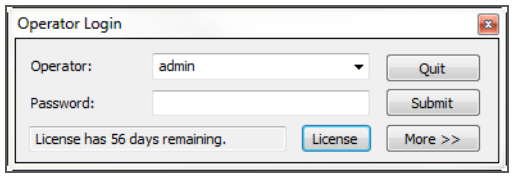
During processing, temporary files are created on the system drive of the connected computer. The temporary file sizes are approximately double the total uncompressed size of the input FASTQ files. There must be sufficient space on the C:\ drive to accommodate the temporary program files. For more information, see *Computing Requirements and Compatibility* on page 5.

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**.

   NOTE
   Illumina recommends that you do not change the admin password.
4. Click **Submit** to start the software.

Figure 1 Operator Login



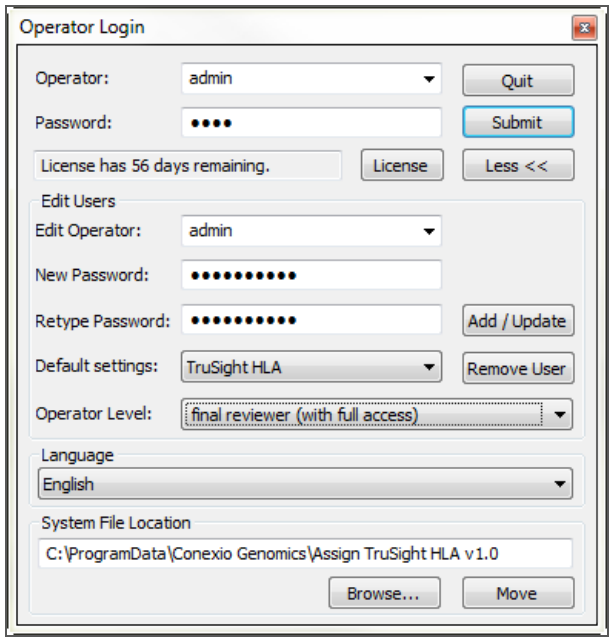
Add Operators

1. Double-click the Assign icon on the desktop or in the installation location.
2. In the Operator Login dialog box, from the **Operator** drop-down list, select an operator.
   The default operator is **admin**.
3. Enter the password.
   The default password for the admin operator is **cg01**.
4. Click **More** to expand the Operator Login dialog box and access the Edit Users section.
5. In the **Edit Operator** field, enter a new operator name.
6. Enter a password for the new operator and retype the same password for verification.
7. From the **Default Settings** drop-down list, select **TruSight HLA**.
   Select this setting for all operators analyzing TruSight HLA data. Operators with sufficient privileges can modify settings directly in Assign.
8. From the **Operator Level** drop-down list, select from the following options.

Operator Level	Permissions
First Reviewer (edit only)	Cannot change settings. Can edit sequences not yet approved by a final reviewer. Cannot sign the final review checkbox.
First Reviewer (with access to settings)	Can change settings. Can edit sequences not yet approved by a final reviewer. Cannot sign the final review checkbox.

Operator Level	Permissions
Final Reviewer (with full access)	Can change settings. Can edit sequences not yet approved by a final reviewer. Can sign the final review checkbox.

Figure 2 Add Operator



9 Click **Add/Update**.

Import and Analyze Sequences

1 Click an open document tab to choose the import destination. To create a new document, click the File button and select **New** or press Ctrl+N.

2 On the Home tab, in the Data group, click **Import and Analyze**.

3 Navigate to the folder containing the FASTA/FASTQ/GZ files.

4 Use the Ctrl or Shift keys to highlight the desired files. Click **Open** to begin import and analysis.

NOTE
Each locus generates a FASTQ file for Read 1 and Read 2. Make sure that you select both FASTQ files.

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.

TIP
To abort an import and close Assign, from the computer Task Manager, highlight the application and click **End Task**.

After you import sequences into Assign, analysis begins automatically. Analysis includes alignment of reads, base calling, IMGT/HLA reference alignment, and HLA typing.

Importing Errors

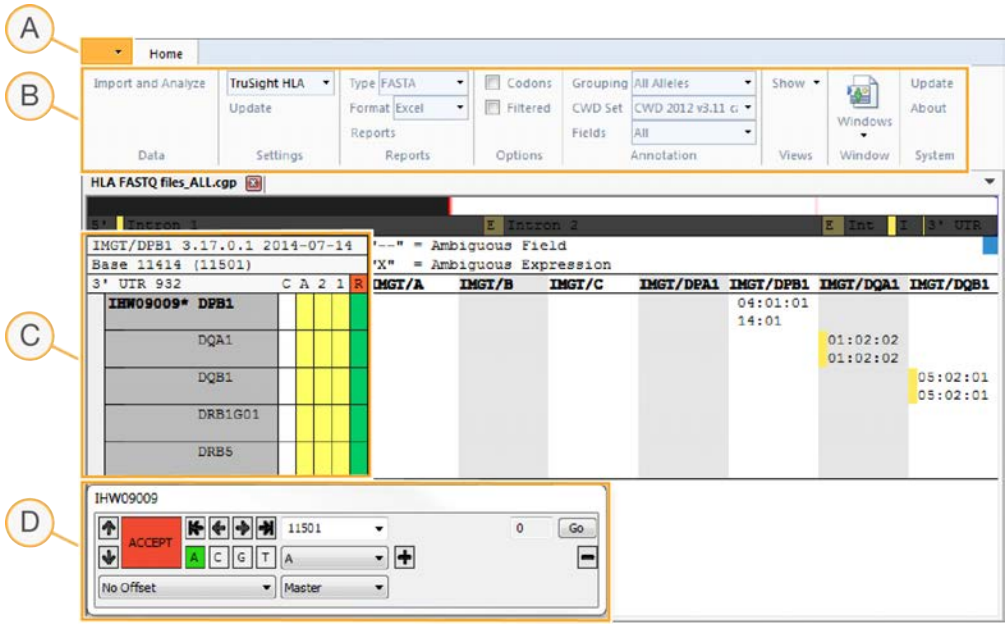
After the analysis of imported sequences is complete, either of the following warnings appear to indicate that the files were not successfully imported.

- 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
  - An appropriate gene name is missing or incorrect (eg, A, B, C, DPA1, DPB1, DQA1, DQB1, DRB1).

  For information on how to name samples properly, see *Create TruSight HLA Sample Plates and Sample Sheets with IEM (part # 15069713)*.

Navigating the Assign Interface

Figure 3 Assign Interface



A **File menu**—Allows you to create new, open, and save sequences in Assign.

B **Home tab**—Provides access to change settings and views.

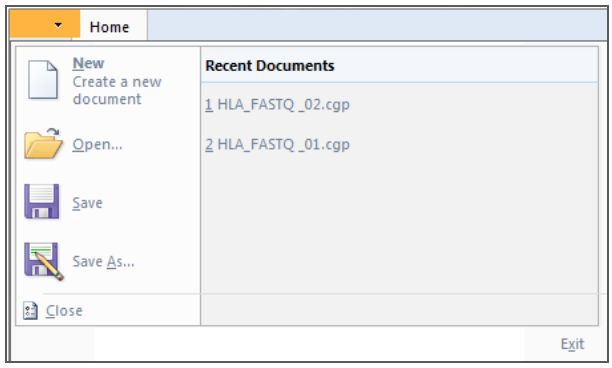
C **Sample panel**—Lists the samples in a project, expands to show each locus typed, and tracks reviewer comments and the laboratory analysis pipeline. For more information, see *Sample Panel* on page 13.

D **Navigator**—Helps you navigate to base positions of interest. For more information, see *Navigator* on page 15.

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.

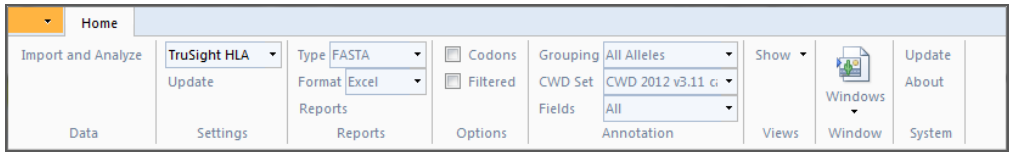
Figure 4 File Menu



Home Tab

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

Figure 5 Home Tab



Data

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

1 In the Data group, click **Import and Analyze**.

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 analyze. Use Ctrl + A to select all of 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. When searching for files, it is possible to create a project with input files from multiple folders.

4 Click **Open**.

NOTE
Each locus 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 analyzed simultaneously.

Analysis begins automatically upon import of the files, which includes alignment of the sequencing reads, assembly to form a consensus sequence, phasing, IMGT/HLA reference matching, and HLA typing.

Settings

The Settings group allows you to select the column configuration for the Results panel. For more information, see *Results Panel* on page 28.

The TruSight HLA setting is the default configuration for the Results panel. To change the default configuration, select customized settings in the Reports, Options, and Annotation groups, and then click **Update** in the Settings group.

Reports

The Reports group allows you to generate 2 types of reports in 3 file formats.

- Report types are Genotyping and FASTA.
- Report file formats are text, Excel, or XML.

For more information, see *Generating Reports* on page 33.

Options

The Options group allows you to switch between viewing options.

- **Codons**—Switches views between nucleotide and codon numbering.
- **Filtered**—Removes allele pairs from the Results panel that are not consistent with base calls that have been confirmed.

Annotation

The Annotation group allows you to consolidate annotations into the following groups:

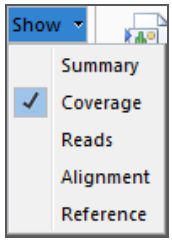
- **G Groups**—Consolidates the Results panel list into G groups.
- **P Groups**—Consolidates the Results panel list into P groups.
- **All Alleles**—Shows all allele matches in the Results panel.

The CWD Set shows a list of the Common and Well-Documented (CWD) alleles, which are indicated in bold in the Results panel.

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.

Figure 6 Views Group



- **Summary**—Comprises 3 panels.
  - Typing Summary panel—Shows the types assigned.
  - Quality Summary panel—Shows the percentage of reads with ≥ Q30.
  - Coverage Summary panel—Shows the depth of sequencing coverage.

  For more information, see *Summary View* on page 19.
- **Coverage**—Shows the read depth and base call composition at each consensus position. For more information, see *Coverage View* on page 21.
- **Reads**—Shows reads used in base calling. For more information, see *Reads View* on page 31.
- **Alignment**—Shows a comparison of the Sample Consensus Sequence and the allele pairs lists in the Results panel. For more information, see *Alignment View* on page 32.
- **Reference**—Shows a comparison of the Sample Consensus Sequence and the reference sequences for a locus. For more information, see *Reference View* on page 32.

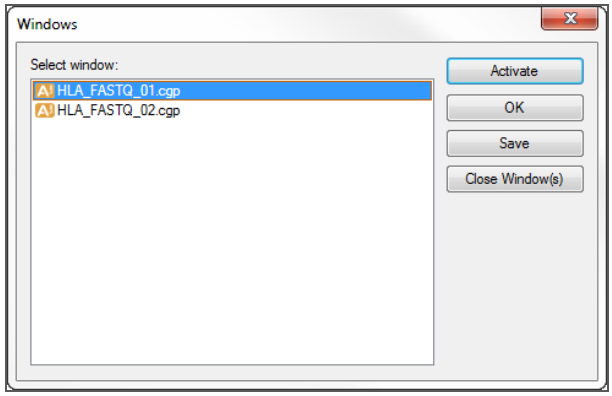
Window

In the Window group, the **Windows** list allows you to control open file windows.

Click **New Window** to duplicate the active window in a new tab. The active window file name appears in bold on the tab.

Click **Windows** to open a dialog box that allows you to activate, save, or close a currently open window.

Figure 7 Windows Dialog Box



- **Select window**—Lists the open file windows. Click a file name to highlight it.
- **Activate**—Click to activate the highlighted file window.
- **OK**—Click to close the dialog box without applying changes.
- **Save**—Click to save the highlighted file.
- **Close Windows**—Click to close the highlighted files.

System

The System group allows you to update and view information on the Assign software.

Click **Update** to open a dialog box that allows you to do the following:

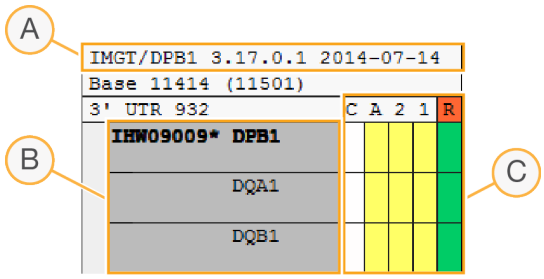
- Import keys, references, NMDP codes, and Nomenclature.
- Locate CWD Files, CWD Update.
- Save or clear a log file.

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.

Figure 8 Sample Panel



A IMGT/HLA reference
B Samples and Loci
C Review hierarchy, report enabling, and locus-specific commenting

IMGT/HLA Reference

The first row in the Sample panel shows the IMGT/HLA reference database used for assignment of HLA nomenclature to the sample sequence. For more information, see *IMGT/HLA Database* on page 4.

The following example indicates specific information about the database:

IMGT/A 3.15.0.0 2014-01-17

- IMGT is the reference database
- A is the gene name
- 3.15.0.0 is the IMGT/HLA database release
- 2014-01-17 is the date of the IMGT/HLA release

Assign converts sample sequences into HLA nomenclature version 3.0, established in 2010, in agreement with the WHO Nomenclature Committee for Factors of the HLA System.

The HLA nomenclature uses the following format:

HLA-A*02:101:01:02N

HLA	The HLA Prefix
-	The hyphen separates the gene name from the HLA prefix.
A	The gene name. For TruSight HLA, the gene name can be A, B, C, DRB1, DRB3, DRB4, DRB5, DQB1, DPB1, DQA1, or DPA1.
*	The asterisk separates the gene name from the sequence information.
02	Field 1—The allele group; alleles that encode an antigen.
:	A colon separates fields.
101	Field 2—Specific alleles that differ at the protein level from DNA substitutions and result in non-synonymous amino acid substitutions.
:	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.
N	This expression modifier is present regardless of the number of fields reported. 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.

Samples and Loci

Click a sample name to view the loci that have been identified for the selected sample. 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 5 columns, which allow for multiple levels of review and comment for each sample and each locus listed. The columns are labeled C, A, 1, 2, and R. Each review level is tracked and audited.

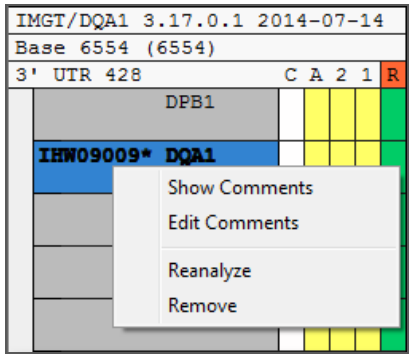
- **Column C**—By default, the box in column C is white. Right-click the sample or locus to add a comment related to the review. When comments are present, the box changes to light blue. Comments added in column C are included in the report.
- **Column A**—By default, the box in column A is yellow. When 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. The box changes to green, which indicates that the first review is complete.
- **Column R**—A green box in column R indicates that the review is complete and the sample can be reported by generating a report.

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.
- **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.
- **Reanalyze**—Removes any edits and trims made to the selected locus and restores the locus to the state following import.
- **Remove**—Removes the selected locus from the project.

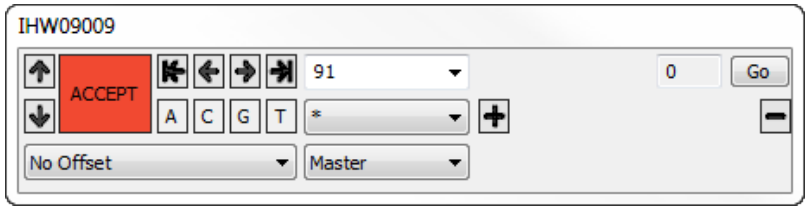
Figure 9 Sample Panel Options



Navigator

Use the Navigator to navigate to a base position of interest. You can drag the Navigator anywhere on the screen.

Figure 10 The Navigator

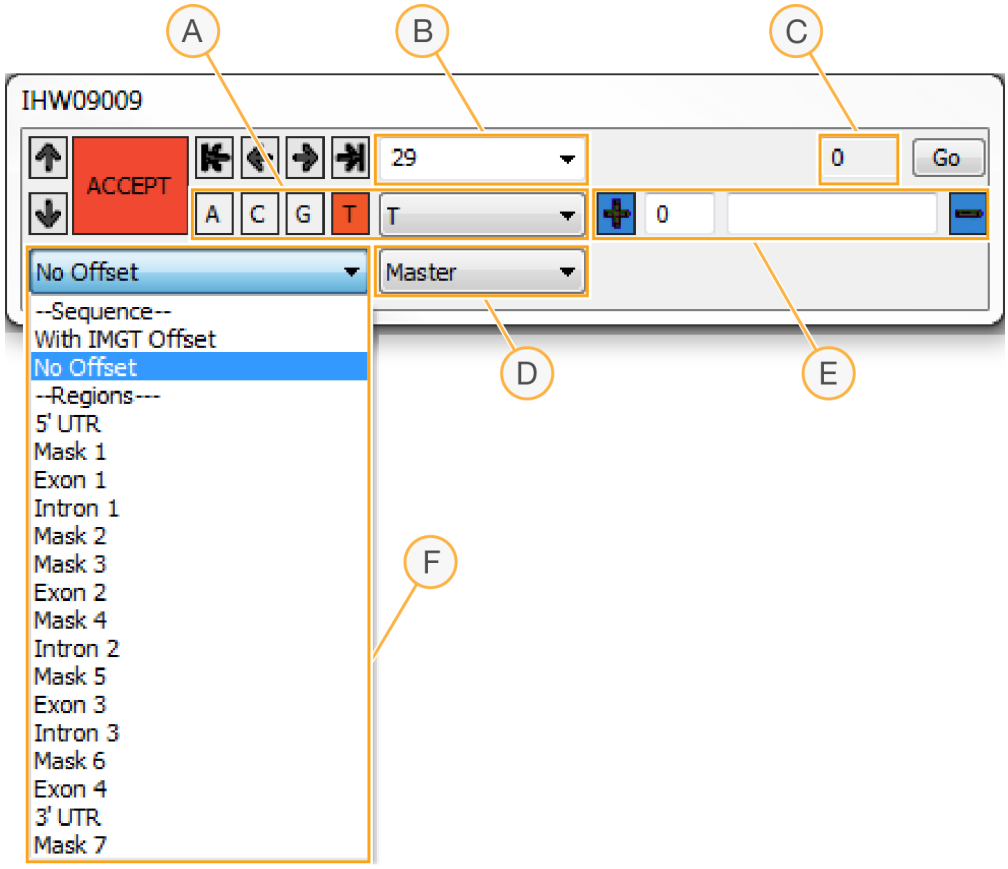


Basic Navigation

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 base positions highlighted in the Confidence Indicator. For more information, see *Confidence Indicator* on page 25.
	Use the first and last arrows to navigate to the base positions highlighted at either end in the Confidence Indicator.
	Click **Go** to make a selection.

Advanced Navigation

Figure 11 Advanced Navigation



A Base Selection
B Mismatch List
C Depth of Coverage Indicator
D Phase Tracks List
E Indel Details
F Nucleotide Position Field Change

Base Selection

The highlighted base indicates the base call at the current position. Multiple highlighted base indicates mixed bases.

A highlighted ➕ indicates an insertion. A highlighted ➖ indicates a deletion.

1 To add or remove a base at the current position, click **A**, **C**, **G**, or **T**, or select from the base selection list.

2 Click **Accept** to accept the selected base and move to the next mismatch position.

3 To change a previously accepted base call, click **Reject** to enable editing.

When you accept a base position that appears as a mismatch in an allele pair, xx/ appears in the Mismatch Column for that allele pair, which eliminates that allele pair from consideration. For more information, see *Locus Structure* on page 22.

Use with the **Filtered** option in the Options group to eliminate possible allele pairs from the Results panel. For more information, see *Options* on page 11.

Mismatch List

The Mismatch list shows the selected position and 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 **Go**. Select an option from the list to move to a position entered previously.

Depth of Coverage Indicator

The Depth of Coverage Indicator shows a numerical value from 0 to 99 for the selected base. A value of 0 indicates low confidence and a value of 99 indicates high confidence.

The Depth of Coverage Indicator value is associated with the coloring on the Confidence Indicator. Dark red indicates low confidence and white indicates high confidence. For more information, see *Confidence Indicator* on page 25.

Phase Tracks List

Use the Phase Tracks drop-down list to switch between layers (sequence and phase tracks) in a locus.

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 or deletion and the bases included in that insertion or deletion are indicated in the space between those symbols.

Nucleotide Position Field Change

The default numbering begins at the first base of the gene. Use the drop-down list to change the numbering system. You can also view the position of a base within an exon. Use offset numbering to determine the position within the coding sequence.

- **Sequence**
  - No Offset (default)—Position in gene sequence based on the locus consensus sequence.
  - With IMGT Offset—Position in gene sequence relative to the allele defined as the reference sequence by IMGT.
- **Groups**
  - cDNA—Position in cDNA. In this view, introns are numbered individually beginning at 1 in each intron.
- **Regions**— For a particular position of interest, choose the region of the gene, enter the relative position in the mismatch list, and then click **Go**. Use this feature for quick navigation.

Summary View

After imported files have completed analysis, the default view is the Summary view. To see the Summary view later, click **Show** in the Views group, then click **Summary**. Alternatively, hover the mouse cursor over the blue box in the upper-left corner of the view in the Coverage, Reads, Alignment, or Reference views, and then click the blue arrow that appears.

The following Summary panels are available within the Summary view:

- Typing Summary panel
- Quality Summary panel
- Coverage Summary panel

Navigating the Summary Panels

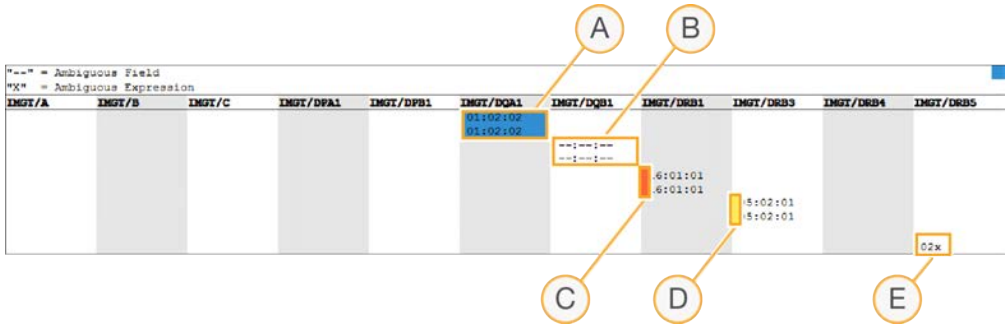
To move between Summary panels, hover over the blue box in the upper-right corner of a Summary view and click the blue arrow that appears. This arrow cycles through the Summary panels.

Typing Summary Panel

The Typing Summary panel shows the samples and types assigned to each locus for each sample. In addition to the typing results, this panel shows whether sequence or expression ambiguities exist, each of which warrant further investigation.

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

Figure 12 Typing Summary Panel



A **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.

B **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.

C **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 coverage or read quality.

D **Confidence warning, yellow**—A yellow box immediately to the left of an allele pair indicates a homozygous locus that might warrant further investigation.

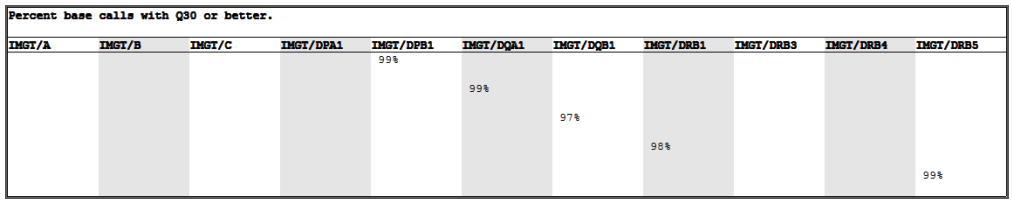
E **Ambiguous expression**—An X indicates an ambiguous expression in an allele typing.

Quality Summary Panel

A quality score, or Q-score, is a modified Phred score that measures 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, a Q-score of 30, indicated as 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 reads with Q30 or higher scores for each locus. A confidence warning appears for loci when the percentage of reads with a Q30 score is 75% or less.

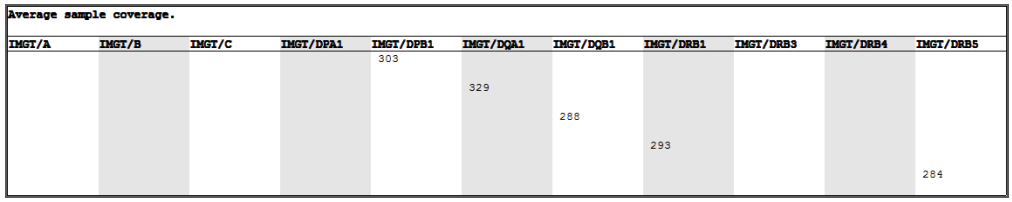
Figure 13 Quality Summary Panel



Coverage Summary Panel

The Coverage Summary panel shows the average depth of sequencing coverage for each locus in the project. The depth of sequencing coverage is the number of observations of a particular base in the sequence data. Warnings are present when loci do not meet specifications of 100x average coverage for 2 alleles, or 50x average coverage for a single allele.

Figure 14 Coverage Summary Panel



Coverage View

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

Figure 15 Coverage View

A **Confidence Plot and Locus Structure**—Shows a view of the high-level locus structure, such as UTRs, introns, and exons, and indicates base call confidence and position. For more information, see *Confidence Plot and Locus Structure* on page 21.

B **Sequences Panel**—Shows consensus reference sequence, sample sequence, base calls, depth of sequencing coverage, base call quality, and alternate sequence reads. For more information, see *Sequences Panel* on page 22.

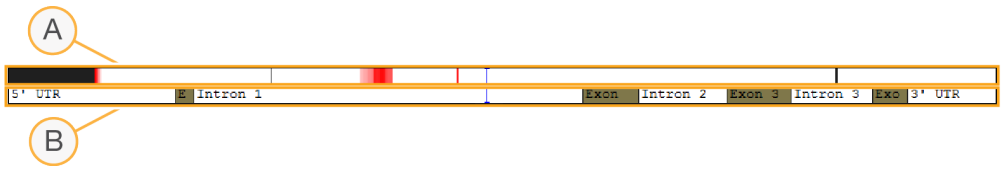
C **Results Panel**—Shows the allele combinations that most closely match the sample sequence, and shows the mismatches between the sample sequence and the reference sequence when present. For more information, see *Results Panel* on page 28.

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 and Locus Structure

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

Figure 16 Confidence Plot and Locus Structure



A **Confidence Plot**

B **Locus Structure**

Click either row to move the blue line, which indicates the region in view in the Sequences panel.

Confidence Plot

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

Figure 17 Confidence Plot colors

- Black indicates no coverage. Common reasons for no coverage include the following:
  - The amplicon does not cover the full genomic sequence for the analyzed locus
  - The reference sequence contains an insertion that is absent in the sample
- Increasing shades of red indicate any of the following conditions:
  - Sequence coverage at Q30 below 100x
  - Base calls have low mean quality
  - Base above noise threshold not called in consensus
  - Base noise below threshold called in consensus
- White indicates complete coverage.

Locus Structure

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

Figure 18 Locus Structure colors

- **Bright yellow**—Exons that are in the active Mismatch Column of the Results panel.
- **Dark yellow**—Exons that are not currently in the active Mismatch Column of the Results panel.
- **White**—Noncoding regions that are in the active Mismatch Column of the Results panel.
- **Gray**—Noncoding regions that are not in the active Mismatch Column of the Results panel.

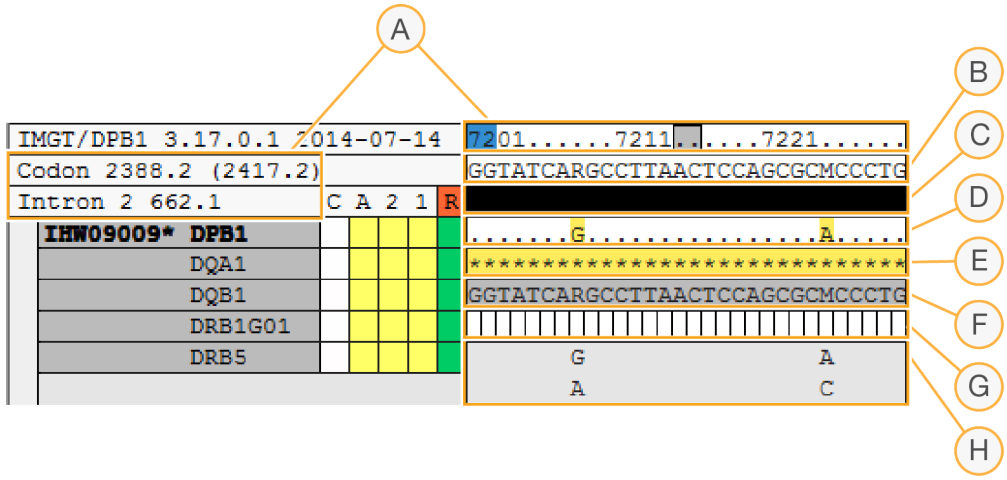
Sequences Panel

The Sequences panel on the Coverage view is comprised 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.

Figure 19 Sequences Section



- **A** Coordinates
- **B** Locus Consensus Sequence
- **C** Sequence Edit Indicator
- **D** Allele 1 Reference Sequence
- **E** Allele 2 Reference Sequence
- **F** Sample Consensus Sequence
- **G** Confidence Indicator
- **H** Phasing Track

TIP

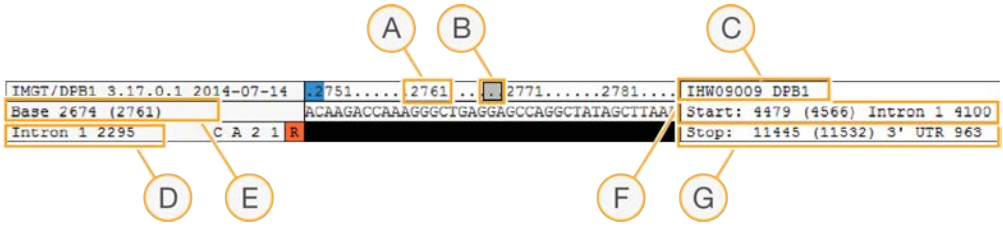
In the Sample Consensus Sequence, press CTRL+A to alternate between nucleotide and amino acid views.

The consensus sequence rows in the Sequences section (rows B and E) include International Union of Pure and Applied Chemistry (IUPAC) degenerate base designations.

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	C G T	not A
D	A G T	not C
H	A C T	not G
V	A C G	not A
N	A C G T	all bases
*		no base call

Coordinates

Figure 20 Coordinates



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

Locus Consensus Sequence

The Locus Consensus Sequence is the reference sequence that includes all known sequences for a locus, including all known insertions and sequences that might not be present in all alleles.

- Yellow indicates exonic/coding sequence
- White indicates intronic/non-coding sequence
- Blue indicates insertions present in some alleles

For HLA-DRB1 and HLA-DQB1, 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
- HLA-DQB1 alleles are split into 2 groups: DQB1 and DQB1G06

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.

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

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.

- A base is displayed in this row when the allele sequence differs from the observed 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.

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.

- 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 Sample Consensus Sequence at the selected position.

Sample Consensus Sequence

The Sample Consensus Sequence shows the consensus sequence of the sample sequenced with the TruSight HLA Sequencing Panel.

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 frequency of the alleles, noise threshold, depth of coverage, and sequence quality.

White in the Confidence Indicator denotes a high confidence base call. A bright red Confidence Indicator denotes base calls in which any of the following conditions have occurred:

1. Sequence coverage at Q30 below 100x
2. Mean quality score for base calls at this position is low
3. Base above noise threshold not called in consensus
4. Base noise below threshold called in consensus

Use the Navigator to move between red confidence flags. For more information, see *Basic Navigation* on page 16.

Phasing Track

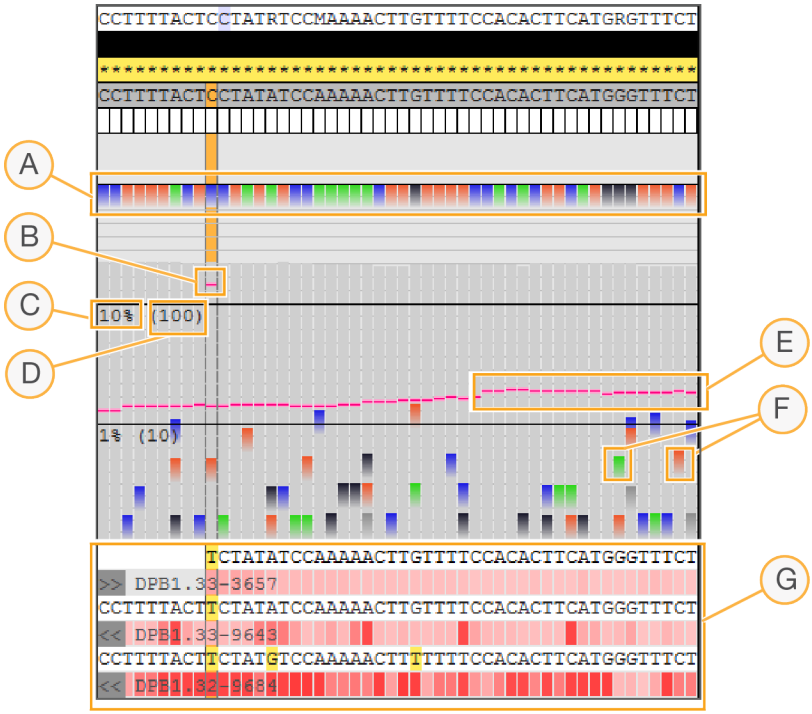
For heterozygous allele combinations, the Phasing Track rows show the phase relationship between bases connected by single reads or paired reads. A phase assignment is made only when most phasing sequences are concordant.

The top row corresponds to Allele 1/MM1 in the Results panel and the bottom row corresponds to Allele 2/MM2. If there is a large distance between heterozygous positions and the software is unable to link phase, there might be rare instances where the top row corresponds to Allele 2/MM2.

Base Calling

Base-level information appears below the Sequences section of the Sequences panel.

Figure 21 Base Calling



- **A** Primary base called
- **B** Approximate allele ratio
- **C** Base call ratio
- **D** Depth of sequencing coverage
- **E** Approximate noise threshold
- **F** Other base calls
- **G** Sequence reads covering that base position

Right-click on a read to open a menu. Select **Copy Sequence** to place all of the bases in the read on the clipboard. Select **Copy Aligned** to place the bases used during alignment on the clipboard. Select **BLAST** to submit the full sequence to NCBI BLAST.

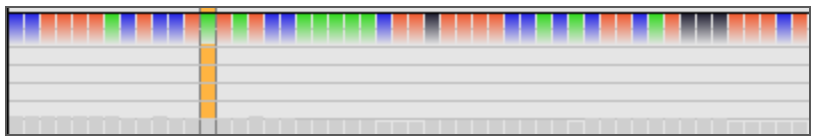
Primary Base Called

In the Primary Base Called section, the following colors indicate the most frequently occurring base call for a given position.

Figure 22 Primary Base Called Color Indicators

- **A**—Green
- **C**—Blue
- **G**—Black
- **T**—Red

Figure 23 Primary Base Called



Approximate Allele Ratio

When a base location is highlighted, a pink line indicates the approximate read depth ratio of the second allele present in the sample.

Base Call Ratio

Base calls are 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%

When there are more than 2 base calls total, the second highest frequency base call is positioned at the sum of the second, third, and fourth highest base calls at that position. This feature is intended to prevent a conflict in the rare event that a second and third base call occurs at the same frequency.

Depth of Sequencing Coverage

The depth of sequencing coverage is shown with gray 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

Approximate Noise Threshold

Noise is a common byproduct of amplification fidelity, specificity, 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.

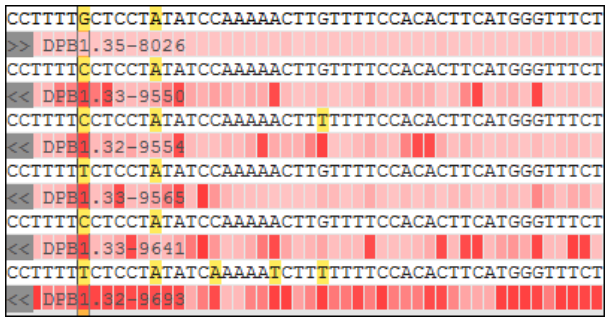
Other Base Calls

The Other Base Calls section shows the base calls that differ from the most frequently occurring base call for a given position and use the same color indicators used in the Primary Base Called section.

Sequence Reads

The Sequence Reads section contains calls that are not included in the Sample Consensus Sequence at the highlighted base position.

Figure 24 Sequence Reads



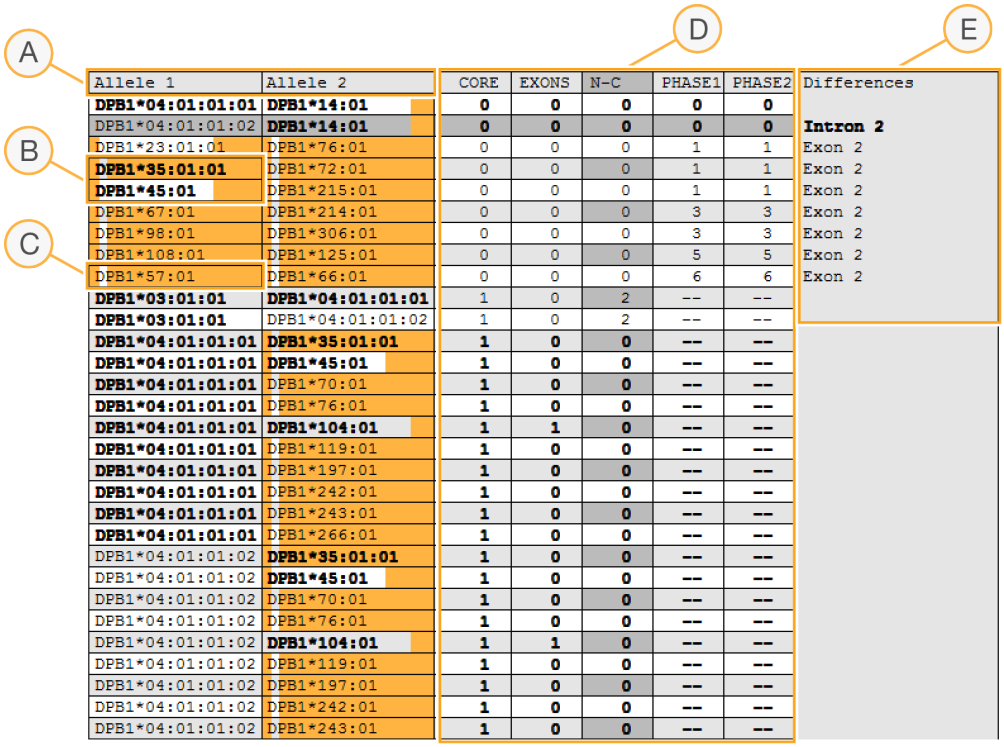
The quality of the base call for alternate reads, as reported in the FASTQ file, is shown below the sequence in a gradient of red.

- **Dark red**—Lowest quality read.
- **Light pink**—Highest quality read.

Results Panel

In the Coverage view, the Results panel lists all of the IMGT/HLA allele pairs that exactly match or closely match the Sample Consensus Sequence. The Results panel also provides information for each of the allele pairs listed.

Figure 25 Results Panel



Allele 1	Allele 2	CORE	EXONS	N-C	PHASE1	PHASE2	Differences
DPB1*04:01:01:01	**DPB1*14:01**	**0**	**0**	**0**	**0**	**0**	
DPB1*04:01:01:02	**DPB1*14:01**	**0**	**0**	**0**	**0**	**0**	**Intron 2**
DPB1*23:01:01	DPB1*76:01	0	0	0	1	1	Exon 2
DPB1*35:01:01	DPB1*72:01	0	0	0	1	1	Exon 2
DPB1*45:01	DPB1*215:01	0	0	0	1	1	Exon 2
DPB1*67:01	DPB1*214:01	0	0	0	3	3	Exon 2
DPB1*98:01	DPB1*306:01	0	0	0	3	3	Exon 2
DPB1*108:01	DPB1*125:01	0	0	0	5	5	Exon 2
DPB1*57:01	DPB1*66:01	0	0	0	6	6	Exon 2
DPB1*03:01:01	**DPB1*04:01:01:01**	1	0	2	--	--	
DPB1*03:01:01	DPB1*04:01:01:02	1	0	2	--	--	
DPB1*04:01:01:01	**DPB1*35:01:01**	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:01	**DPB1*45:01**	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:01	DPB1*70:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:01	DPB1*76:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:01	**DPB1*104:01**	**1**	**1**	**0**	**--**	**--**	
DPB1*04:01:01:01	DPB1*119:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:01	DPB1*197:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:01	DPB1*242:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:01	DPB1*243:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:01	DPB1*266:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:02	**DPB1*35:01:01**	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:02	**DPB1*45:01**	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:02	DPB1*70:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:02	DPB1*76:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:02	**DPB1*104:01**	**1**	**1**	**0**	**--**	**--**	
DPB1*04:01:01:02	DPB1*119:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:02	DPB1*197:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:02	DPB1*242:01	**1**	**0**	**0**	**--**	**--**	
DPB1*04:01:01:02	DPB1*243:01	**1**	**0**	**0**	**--**	**--**	

A Allele columns
B Common and Well-Documented (CWD) alleles (in bold)
C IMGT/HLA reference coverage
D Mismatch columns
E Differences column

Allele Columns

In the Alleles columns, all allele pairs appear in order based on the number of mismatches they contain. Allele pairs with zero mismatches appear at the top of the columns followed by pairs with increasing numbers of mismatches.

Common and Well-Documented (CWD) Alleles

In the Results panel, CWD alleles are shown in bold.

IMGT/HLA Reference Coverage

The allele pairs are banded white and gray by alternating rows for ease of viewing. In some cases, the allele includes orange, which indicates that a part of the reference

sequence is missing in the IMGT/HLA reference for that allele. The allele container width is directly proportional to the amplicon length.

In the following example of HLA-DPA1, the amplicon spans the entire length of the gene. The DPA1*01:03:01:05 allele has reference sequence for the complete gene. The DPA1*01:03:03 allele only has reference sequence available for exon 2; DPA1*02:01:01 has exon 2 and 3 only, and DPA1*01:03:01:01 has reference sequence for the region spanning exons 1–5. In this example, the shading is identical for DPA*02:01:01 and a sequence that has intron 2 sequence as well as exon 2 and 3. The shading is also identical for DPA1*01:03:01:01 and an allele with only cDNA sequence.

Figure 26 IMGT/HLA Reference

Allele 1	Allele 2
DPA1*01:03:01:05	**DPA1*02:01:01**
DPA1*01:03:03	DPA1*02:01:02
DPA1*01:06:01	DPA1*02:03
DPA1*01:03:01:02	**DPA1*02:01:01**
DPA1*01:03:01:01	**DPA1*02:01:01**
DPA1*01:03:01:03	**DPA1*02:01:01**
DPA1*01:03:01:04	**DPA1*02:01:01**
DPA1*01:03:01:01	DPA1*02:01:02
DPA1*01:03:01:01	DPA1*02:01:04
DPA1*01:03:01:01	DPA1*02:01:06
DPA1*01:03:01:01	DPA1*02:01:07

Mismatch Columns

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

The 5-column configuration shows the following information:

- The first column shows sequence mismatches in Class I exons 2, 3, and 4 and Class II exons 2 and 3.
- The second column shows the remaining exons.
- The third column shows the remainder of the amplicon.
- The fourth and fifth columns show the phase mismatches in heterozygous alleles.

Figure 27 Mismatch Columns

Allele 1	Allele 2	A CORE	B EXONS	C N-C	D PHASE1	E PHASE2
DPB1*108:01	DPB1*125:01	0	0	0	5	5
DPB1*57:01	DPB1*66:01	0	0	0	6	6
DPB1*03:01:01	**DPB1*04:01:01:01**	1	0	2	--	--
DPB1*03:01:01	DPB1*04:01:01:02	1	0	2	--	--
DPB1*04:01:01:01	**DPB1*35:01:01**	**1**	**0**	**0**	**--**	**--**

A Mismatches in exons 2, 3, and 4 (exons 2 and 3 only 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

Auto-expansion of the Mismatch Columns

Using the auto-expand feature, the mismatch columns expand to make an unambiguous typing as long as expanding the columns does not incur mismatches. The auto-expand

feature is designed to prevent biasing against alleles with complete reference sequences. Therefore, if the expansion into the next column favors an allele pair with an incomplete reference, the mismatch columns do not auto-expand.

For example, the mismatch columns do not auto-expand into the third column if the following occurs:

- The top 2 allele pairs have no mismatches in the exons (first 2 columns).
- The top pair has a complete reference.
- The second pair has an exon sequence in its reference.
- There is a single intronic mismatch for the top pair.

Navigating the Mismatch Columns

Of the 5 possible mismatch columns, the **Core** and **Exons** columns are always present. 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 needed to resolve a sequence ambiguity.

Phasing mismatches are only calculated for the pairs with the lowest number of mismatches in the first mismatch column.

Differences Column

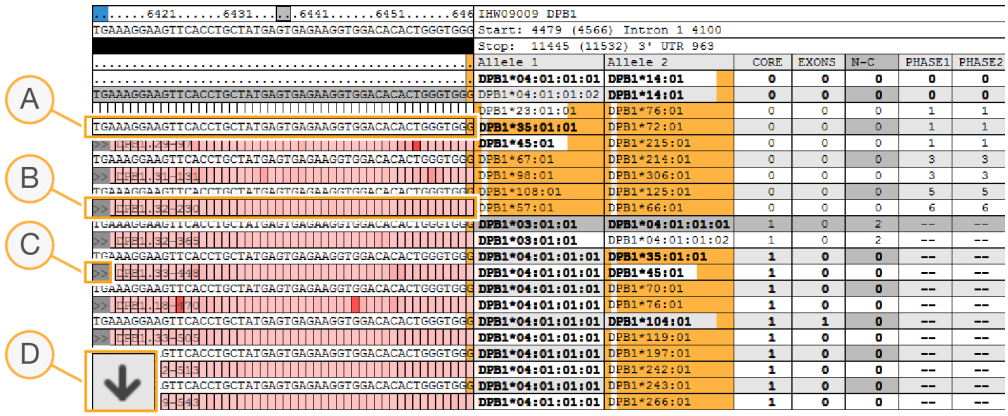
The Differences Column indicates the location of differences between the allele pairs.

Where ambiguities exist, the regions in which they might be resolved are indicated in this column.

Reads View

The Reads view shows 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**.

Figure 28 Reads View



A **Nucleotide Sequences**

B **Base call quality from FASTQ file**—Quality is shown in light pink (highest quality) to dark red (lowest quality).

C **Vertical scroll arrow**—Use the scroll arrow to look at more reads covering the selected position.

D **Read scroll arrows**—Use the scroll arrows to navigate to the beginning and end of the sequence reads. You can also navigate by pressing Page Up and Page Down on your keyboard.

To hide reads for a specific nucleotide, 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.

Alignment View and Reference View

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

Alignment View

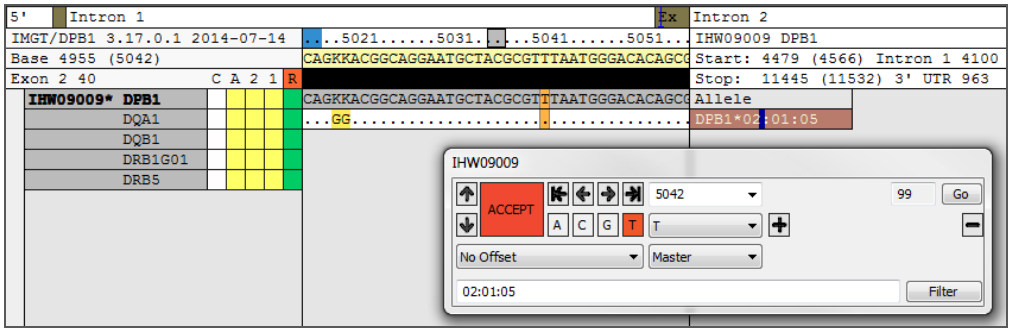
The Alignment view shows a comparison of the Sample Consensus Sequence and the allele pairs listed in the Results panel. Click the headings **Allele 1** or **Allele 2** to add or remove the contribution from the alleles in that column. To see the Alignment view, in the Views group, click **Show**, and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

Figure 29 Limiting Reference Alleles



Generating Reports

Types of Reports

Assign generates a genotyping report or a FASTA report.

- The genotyping report reports on a single sample or locus or all samples and loci in the project.
- The FASTA report reports the Sample Consensus Sequence using the IUPAC designations.

Reports can be customized with a logo, page numbers, date and time, and other references about the report.

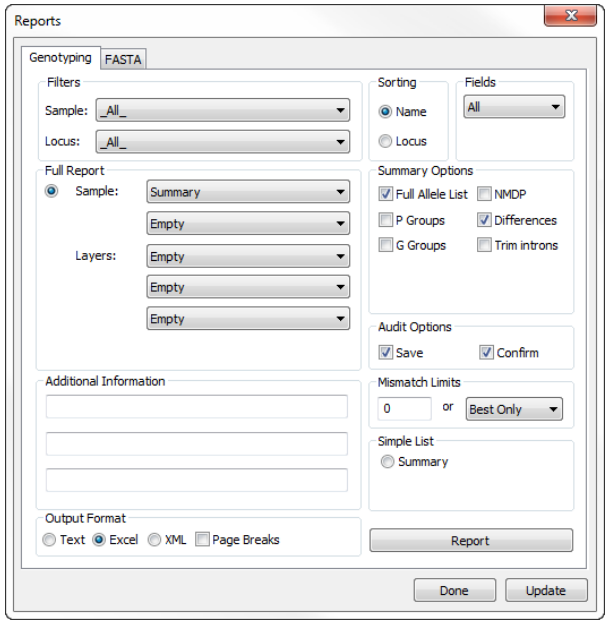
Genotyping Report

1 On the **Home** tab, under **Reports**, from the **Type** list, select **Genotyping**.

2 From the **Format** list, select your preferred file format.

3 Click **Reports** to launch the reporting tool.

Figure 30 Genotyping Report



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, select an individual locus to include in the report. Select **All** to include all loci in the project.

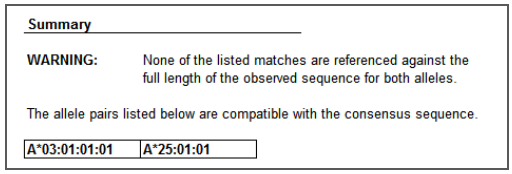
3 In the Sorting section, select either sample **Name** or **Locus** to sort the report.

4 In the Fields section, select the number of fields to report from the list.

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.

Figure 31 Warning Description



- **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. Options include the number of bases sequenced and the beginning and ending base pair positions within their respective sections of the locus. Select **Empty** if a selection list is not needed.

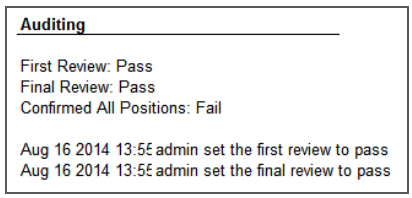
- **Sequences**—For each locus selected in the Filters, the Sequences report prints the Sample Consensus Sequence (as edited).
- **Edit List**—For each locus selected in the Filters, the Edit List report shows the edited positions, the edit that was made, and the user that made the edit.
- **Mismatch List**—When selecting the Mismatch List layer for reporting, set values in the **Mismatch Limits** section. To obtain the desired mismatch list in the report, enter the desired combination of number of mismatches and select the relation to best match from the list. 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	Includes P groups. For more information, see hla.alleles.org/alleles/p_groups.html.
G Groups	Includes G groups. 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.
Trim Introns	When the “Sequences” layer is selected for reporting, intron sequences are removed from the sequences and only cDNA sequence is provided in the 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.

Figure 32 Auditing Report



9 In the Output Format section, select from the following formats:
   a **Text**—Generates a report of the selected options into text format.
   b **Excel**—Generates a report of the selected options into an Excel spreadsheet.
   c **XML**—Generates a report of the selected options into a tagged *.xml file that is best suited for importing into an external database.
   d **Page Breaks**—Adds page breaks to the Excel spreadsheet.

10 Click **Report**. Excel reports generate and open automatically in Excel. Text or XML 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\Conexio Genomics\Assign TruSight HLA v1.0\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.

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 (>). The next line in the FASTA is the Sample Consensus Sequence using the IUPAC designations.

1 On the Home tab, in the Reports section, select **FASTA** from the **Type** list.

2 From the **Format** list, select your preferred file format.

3 Click **Reports** to launch the reporting tool.

4 On the FASTA tab, 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.

5 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 (e.g., >SampleName_IMGT/A).

6 From the **Layer** list, select a single layer to restrict output. Select the checkbox to insert the layer name into the FASTA file.

7 From the **Group** list, select a designated group of regions to restrict output.

8 From the **Region** list, select a designated region, such as an exon. Select the checkbox to insert the region name into the FASTA file.

9 Select **Consensus** to write the combined sequence for a sample to the output file.

10 Select **Component sequences** to write each individual sequence read to the file. Sequences can be filtered so that forward and reverse sequences (FR), forward only sequences (F), or reverse only sequences (R) are included in the report.

11 In the Sort by section, select either sample **Name** or **Locus** to sort the report.

12 In the Options section, select the **Pad Ends** checkbox to add N base calls to each sequence to cover the entire amplicon.

13 Click **Generate Report**, and then choose a save location on your computer.

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 1 Illumina General Contact Information

Website	www.illumina.com
Email	techsupport@illumina.com

Table 2 Illumina Customer Support Telephone Numbers

Region	Contact Number	Region	Contact Number
North America	1.800.809.4566	Italy	800.874909
Australia	1.800.775.688	Netherlands	0800.0223859
Austria	0800.296575	New Zealand	0800.451.650
Belgium	0800.81102	Norway	800.16836
Denmark	80882346	Spain	900.812168
Finland	0800.918363	Sweden	020790181
France	0800.911850	Switzerland	0800.563118
Germany	0800.180.8994	United Kingdom	0800.917.0041
Ireland	1.800.812949	Other countries	+44.1799.534000

Safety Data Sheets

Safety data sheets (SDSs) are available on the Illumina website at support.illumina.com/sds.html.

Product Documentation

Product documentation in PDF is available for download from the Illumina website. Go to support.illumina.com, select a product, then click **Documentation & Literature**.

Part # 15059520 Rev. B

Illumina
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com