

ASSIGNTM SBT v3.6+

HLA Sequence Analysis Software

Operators Manual

Version No: 6.0
Issue Date: February 2012



CGX0036+



Conexio Genomics Pty Ltd
8/31 Pakenham St
Fremantle 6160
Western Australia
Australia



Qarad bvba
Cipalstraat 3
B-2440 Geel
Belgium

Table of Contents

About Assign™ SBT 3.6+	4
1. Compatibility.....	4
1.1. Computer Operating Systems	4
1.2. Data files supported	4
2. Overview	4
2.1. Functions and Features.....	4
2.2. Performance Characteristics	4
2.3. Intended Use	4
2.4. Limitations	5
1. Installation	5
1.1. To install:	5
1.2. Installation of Reference libraries, NMDP codes, P and G group files and CWD files.	6
2. Login and adding users	7
2.1. Login	7
2.2. Adding Users	7
3. Settings	8
3.1 General	8
3.2 Naming Conventions	8
3.3 Reference aliases	10
3.4 Variant Positions	10
4. Importing Sequences	11
4.1 Importing sequences by directory	11
4.2 Importing sequences individually	12
5. The Screen layout, Editing and Analysis	12
5.1 Sample ID Pane	12
5.2 Sequence electropherograms	14
5.3 Results pane	16
6. Navigator.....	17
6.1 Positions for Validation	17
6.2 Moving Between Validation Positions	18
6.3 Arrow functions	18
6.4 Other Navigator functions	18
6.5 Confirming or Editing Base Calls	18
7. Other Sample and Sequence Editing Functions.....	18
7.1 Resizing the EPG	18
7.2 Hiding EPG traces	18
8. View Options.....	19
8.1 Status Bar	19
8.2 Panes Functions	19
8.3 Letters and Dots	19
8.4 Nucleotides and Codons	19
8.5 View Unaligned	20
8.6 View All EPG	20
8.7 Filter Confirmed	20

9.	Data Analysis and Editing EPGs	20
9.1.	Logging On	20
9.2.	Importing Data	20
9.3.	Navigation	20
9.4.	Resolution Primer Layers	22
9.5.	Auditing	22
10.	Reporting.....	22
10.1.	Genotyping Report	23
10.2.	HARPS Report	24
10.3.	FASTA Report	26
10.4.	Quality Reports	26
Appendix 1	28
References	28
Support and Contact Details	28

About Assign™ SBT 3.6+

1. Compatibility

1.1. Computer Operating Systems

1.1.1. Assign™ is a Windows® based program that will run on Windows® XP, Windows® Vista, and Windows® 7 operating systems.

1.1.2. Microsoft® Excel® 97 or above is required for the creation of reports.

1.2. Data files supported

Assign™ requires .ab1, .abd or .scf sequence files from automated DNA sequencers. The files should be run through the Applied Biosystems™ by Life Technologies™ Sequence Analysis software or similar prior to their import into Assign™.

2. Overview

2.1. Functions and Features

2.1.1. Sequences from multiple loci can be imported into the same layout.

2.1.2. The sample identifier, sequence electropherogram and allele assignment results are all visible on one screen

2.1.3. Sequence editing includes priority analysis of positions of low quality and positions that are mismatched with closely related alleles.

2.1.4. Simultaneous analysis of sequences for the resolution of heterozygous ambiguities is possible.

2.1.5. Analysis of non-coding sequence for Class I alleles is possible.

2.1.6. Assign™ SBT 3.6+ can report CWD alleles, G groups and P groups.

CWD alleles, or common and well documented alleles, are those alleles with a calculated allele frequency in one or more populations. The user can edit the CWD file according to allele frequencies of the local population.

G groups are those groups of alleles that share the same nucleotide sequence in exons 2 and 3 for class I loci and exon 2 for class II loci.

P groups are those alleles that share the same amino acid sequence in exons 2 and 3 of class I and exon 2 of class II alleles.

NOTE: G and P groups are updated directly from the IMGT/HLA database.

These options enable laboratories to determine the probability of an ambiguous report or if an ambiguous report contains functional differences between alleles.

2.1.7. Assign™ SBT 3.6+ includes sample to sample and run to run QC analysis.

2.2. Performance Characteristics

2.2.1. **Throughput:** Assign™ SBT 3.6+ has successfully imported over 5,000 sequence electropherograms into a single project.

2.2.2. **Base Call Accuracy:** Assign™ contains a unique base caller developed to improve the accuracy of heterozygous base calls. However base call accuracy is influenced by sequence data quality. Generally, samples with a base call quality score (described in detail below) of >85 do not have incorrect base calls.

2.3. Intended Use

- To enable the operator to check and edit automated DNA sequence assignments
- To determine the HLA type from a DNA sequence following sequence editing.

2.4.Limitations

- 2.4.1.Sequence with background noise and poorly separated peaks may result in incorrect base calls and the potential for incorrect typing results. This is true for all sequence analysis software. However this limitation is offset in Assign™ SBT as poor quality data is easily identified within the software so that it can be reviewed for incorrect base calls. Such data can be excluded from analysis so that the risk of an incorrect genotype is minimised.
- 2.4.2.Assign™ SBT 3.6+ compares a sample sequence with a library of sequences from known alleles. The report lists those allele/s combinations in the library that are identical to the sample sequence. It is possible that the same sequence could be derived from alleles yet to be described and whose sequence is not yet part of the library. Therefore caution must be taken when interpreting the genotype report as a HLA type.

Getting Started and Using the Software

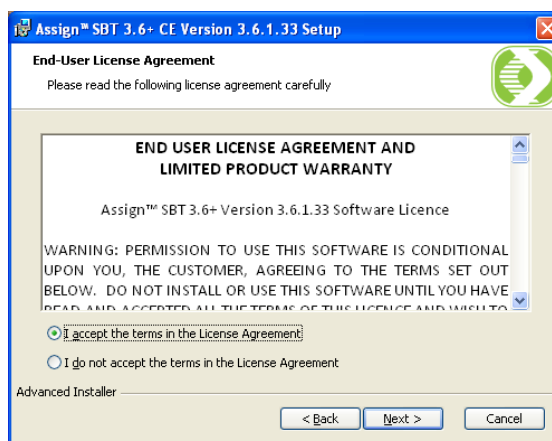
1. Installation

Assign™ SBT 3.6+ is a standalone computer software program that should be installed on the computer on which SBT analysis is performed. It is recommended that Assign™ is installed by a user with complete administrator access to the computer. The installer package can be acquired by contacting Olerup GmbH/Olerup Inc. via their website <http://www.olerup.com/>. It is also helpful if the computer has access to the internet to facilitate the system updates with new libraries and other files as needed.

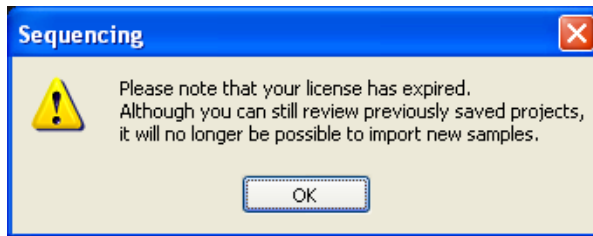
If you plan to use Assign™ SBT 3.6+ on multiple computers within your organisation, it is possible to deploy to a shared network drive. This allows user logins and settings to be shared across computers and allows license keys to be stored in a single location. Please contact Conexio Genomics Pty Ltd. for the *Assign™ Networking Setup Protocol* document.

1.1. To install:

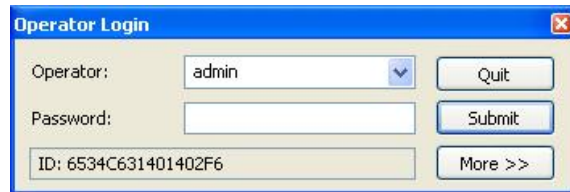
- 1.1.1.Double click on the installer file icon (.msi file) and follow the instructions for installation.
- 1.1.2.Accept the terms in the License Agreement, then click **Next**.



- 1.1.3.Select the Installation Folder location. It is best to leave the installation at the default location. Click **Next**.
- 1.1.4.Click **Install** to begin the Installation.
- 1.1.5.Once the Installation is complete, click **Finish**.
- 1.1.6.The software will not be functional without a licence key file. The licence key files are specific for the computer hardware ID.
- 1.1.6.1. To obtain this Hardware ID, launch the software by double clicking on the Assign™ SBT icon on the desktop.
- 1.1.6.2. A warning message alerting the user to expired license keys will appear, click **OK** to continue.



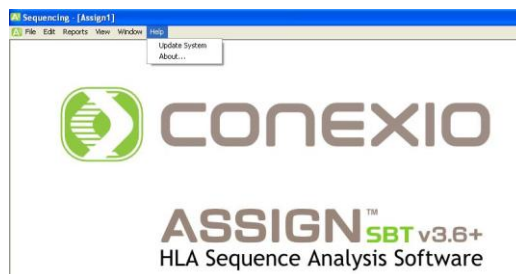
- 1.1.6.3. The Hardware ID is located at the bottom of the Login screen. Copy this ID and paste it into an email and send to keys@conexio-genomics.com to obtain the license key files.



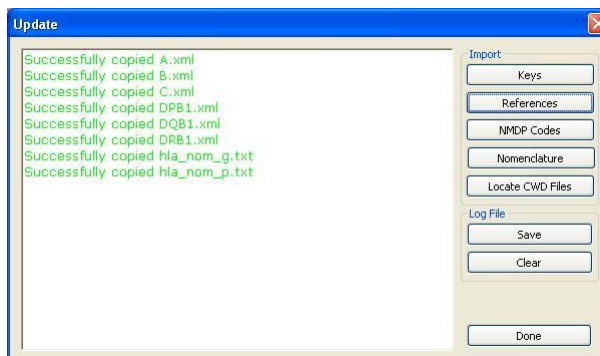
1.2. Installation of Reference libraries, NMDP codes, P and G group files and CWD files.

Before running the software the appropriate files are required to be installed. Updated reference files are available semi-annually from the Conexio Genomics website. Updated NMDP codes are also available from the Conexio Genomics website. The frequency of updating this file is dependent upon the user's needs.

- 1.2.1. Go to the Conexio Genomics website, www.conexio-genomics.com and click on the **Downloads** tab.
- 1.2.2. Click on the **Assign™ SBT 3.6+** tab. Click on the library version that you require.
- 1.2.3. A zip file will be downloaded to your computer. Unzip the file and save the **References** folder to the desktop or other convenient location. The References folder will contain the gene specific reference files, and files for the P and G Groups.
- 1.2.4. Launch the software. Use the default operator (admin) and password (cg01) login. Click on **Help | Update System** on the top menu bar.



- 1.2.5. Click on the **References** button and navigate to the unzipped References folder.
- 1.2.6. Highlight the .xml reference files to update then click **Open**. The references will be imported to the correct location within the software. Green text will confirm successful update of reference files. If Red text is present, the import failed and the process should be repeated.



1.2.7.Repeat the process with the NMDP codes by clicking on the **Assign™ SBT 3.6+** page located under the **Downloads** tab on the Conexio website. The NMDP code file is named 'numer.txt' and must not be renamed. Save the .txt file to your computer. Log into the software and click on **Help | Update System | NMDP codes**. Navigate to the numer.txt file and highlight it. Click **Open**. The code file will be imported to the proper location within the software.

1.2.8.When the licence key(s) are received, use the **Keys update** function within the software by clicking on **Help | Update System | Keys**.

1.2.9.To update the CWD file, click on the **Locate CWD Files** button. This will open Windows® Explorer; locate the CWD file. Double click to open the file in Notepad. The default CWD file is based on the report by Cano (1). The user can modify the CWD list according to the allele frequencies within their population, or leave the file as the default. If additional custom CWD files are required, they can be created by using the example CWD file provided in the Templates folder.

NOTE: Including CWD, P and G group data interpretation is optional. The CWD file may require modification with new IMGT/HLA database releases.

2. Login and adding users

2.1.Login

2.1.1.Launch the software by double clicking on the Assign™ 3.6+ icon located on the desktop.

2.1.2.The default operator is 'admin' and the default current password is 'cg01'.

NOTE: It is recommended that the admin password is never changed.

2.2.Adding Users

2.2.1.Enter the admin login and password then click **More**. Below will be a section to add additional users.

2.2.2.Type in the new operator's name in the **Edit Operator** section. Select and type a password for that user. Re-type the password. Select the **Operator Level**. Click **Add/Update** directly next to the Retype Password box.



The screenshot shows a software window titled "Operator Login". It contains several sections:

- Operator:** A dropdown menu set to "admin" and a "Quit" button.
- Password:** A text box with four dots and a "Submit" button.
- ID:** A text box containing "6534C631401402F6" and a "Less <<" button.
- Edit Users:** A section with:
 - Edit Operator:** A dropdown menu set to "admin".
 - New Password:** An empty text box.
 - Retype Password:** An empty text box and an "Add / Update" button.
 - Default settings:** A dropdown menu set to "default" and a "Remove User" button.
 - Operator Level:** A dropdown menu set to "final reviewer (with full access)".
- System File Location:** A text box containing "C:\Documents and Settings\All Users\Application Data\Conexio Ge" and "Browse..." and "Move" buttons.

2.2.3.Repeat for additional users.

2.2.4.To launch the software under a particular user, double click on the Assign™ icon. In the Operator dropdown, select the user. Type in the password, then click **Submit**.

3. Settings

The settings menu enables the user to configure the software for their requirements. Settings can be saved as different settings files to enable the software to be individualized for different users.

3.1 General

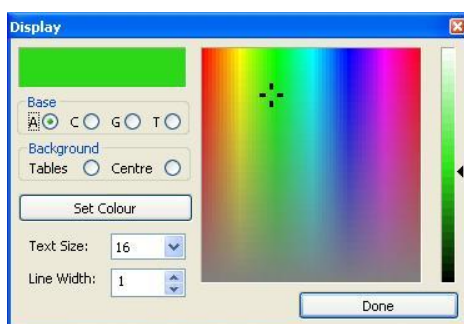
General Settings allow modifications to the interface such as changes to font, electropherograms colours and line thickness.

3.1.1. To open the default settings file, click on **Edit | Settings** on the top menu bar in the software. The default settings are located in the **General** tab.

3.1.2. Customize the Display

3.1.2.1. Click on **Display** in the General tab.

3.1.2.2. The Display options will appear.

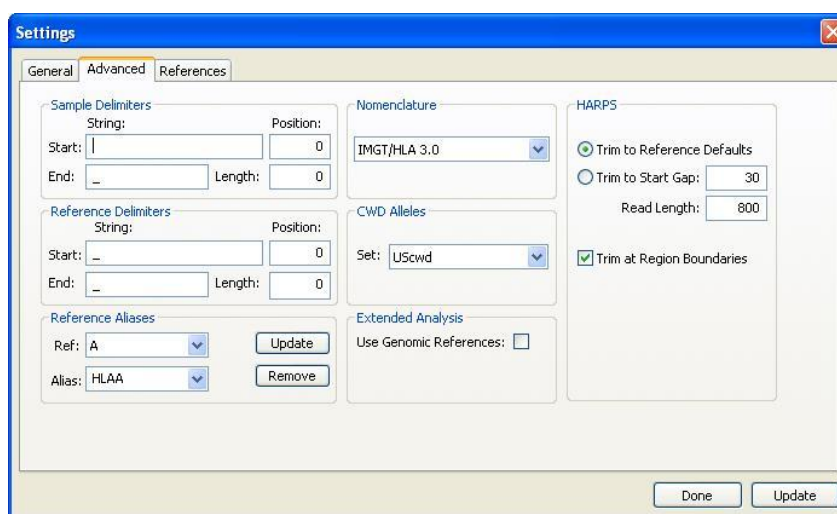


3.1.2.3. Adjust the base colours, background colours, text size, and line width (EPG tracing width). When complete, click **Done**.

3.1.2.4. To create a new settings file if desired, type in a new settings file name and proceed to the **Advanced** tab to create the naming conventions and locus alias names.

3.2. Naming Conventions

In the **Advanced** tab the user can enter the parameters that define the sample name and the locus identifier in sample sequence filenames. The sequence filename must be unique for a sample and consists of the sample name and an identifier that can be used as an alias for the locus being genotyped. If a standard system is used by the operator, analysis will occur automatically and data from different loci can be entered into the same layout in a single step.



3.2.1. **Sample Delimiters** have been used to separate components of the sequence file name.

Example

Sequence filename A01[12345_DQ2F_A01

Delimiters used to separate the components of the sequence filename:

[separates the PCR number (A01) and the sample name (12345)

_ separates the sample name and the locus (HLA-DQB1) and sequencing primer (exon 2 forward)

_ has also been used to separate the locus and well location (A01)

3.2.2. Set the Naming convention by defining the **Sample Delimiters**.

In the example above, the sample name begins with [and the sample name ends with

_. Enter [in the **Start string box**, and enter _ in the **End string box**.

3.2.3. **Alias names**

In the example above, DQ is used as an alias for HLA-DQB1.

3.2.3.1. In the **Reference Aliases** section, select the locus from the **Ref:** drop down menu.

3.2.3.2. Select the alias used (DQ) in the Alias drop down menu.

If the alias is not present, type it into the Alias box, and then click **Update** directly to the **right of the Ref box**.

3.2.3.3. Repeat this for each locus alias you will be using. After all aliases have been added, click on the **Update** button in the **lower right hand corner** of the **Settings** box.

3.2.4. **Nomenclature**

The HLA nomenclature standards changed in April 2010 from v2.0 to v3.0. In **Select the Output Naming Standard**, either IMGT/HLA3.0 or 2.0 can be selected for the naming convention of the reported alleles

3.2.5. **HARPS**

HARPs[®] are sequencing primers designed to sequence only one of the alleles in a heterozygous sample. HARPs[®] are used for the resolution of heterozygous ambiguities, by producing hemizygous sequence for one of the alleles. The sequence data from a HARP[®] is combined with the existing sequence data from the sample to produce a high resolution genotype report.

Selecting the **Trim to Reference Defaults** option will activate the default settings. The default settings are set to report HARPs[®] with a start gap of 20 bases and to limit the sequence read length to the exon to which the HARP[®] has been designed. The exceptions to this are those HARPs[®] that are designed to anneal immediately before, and in the direction of, intron 2. These HARPs[®] are designed for analysis of the entire neighbouring exon.

Alternatively, the operator may tailor the HARP[®] settings by defining their own values for the **Trim to Start Gap** and **Read Length** parameters based on their sequencing capabilities. The **Trim to Start Gap** enables only good quality sequence to be analysed by eliminating the poorly resolved data at the start of a sequence. The Start Gap will vary between laboratories according to the sequence reaction clean-up method used, the DNA Sequencer make and model, and the polymer in the sequencer capillaries. Enter the number of base pairs between the end of the primer and the first usable sequence generated in the **Start Gap**.

The **Read Length** parameter enables the operator to define the maximum number of bases which can be read on a single run. Defining the read length will ensure that the software does not report a HARP[®] which will require a longer read to achieve ambiguity resolution.

If **Trim at Regions Boundaries** is selected, the analysis will stop at the end of the exon, regardless of the read length defined in the settings.

3.2.6. CWD Alleles

The CWD alleles are stored in a text file. A default set of CWD alleles based on those described by Cano (1) are provided with the software. However laboratories may want to create their own list.

Select the **CWD** allele set you will be utilizing, if any. The installer provides a basic CWD file that can be updated by the user.

3.2.7. Extended Analysis

By selecting **Use Genomic References**, the sequence analysis will enable the comparison of the sample sequence with the genomic references provided by the IMGT/HLA database. Genomic sequence analysis enables the typing of alleles characterised by polymorphisms in non-coding regions. These alleles include expression variants such as A*01:01:01:02N, A*24:02:01:02L and others.

By leaving the **Use Genomic References** box unchecked, analysis will be performed against the standard cDNA references.

If the laboratory intends to utilize analysis of the non-coding regions, put a check in the **Use Genomic References** box.

3.3. Reference aliases

Reference aliases are used in the sequence filenames to indicate the HARP[®] being used for sequencing.

3.3.1. Click on the **References** tab of the **Settings** menu to establish alias names for the primers used.

3.3.2. Click on the **Load Reference** box at the top right of the screen. Navigate to the References folder and click on the reference needed. Click **OK**.

3.3.3. The reference information will populate the screen.

3.3.4. To establish a reference alias for the RB-TT197-F HARP[®] for example, highlight the DRB1.xml file and click **OK**.

3.3.5. In the **References** tab, in the lower left corner drop down menu, select the RB-TT197-F primer. To the right of the primer will be the alias drop down menu. Click the drop down menu to determine if your naming alias is present.

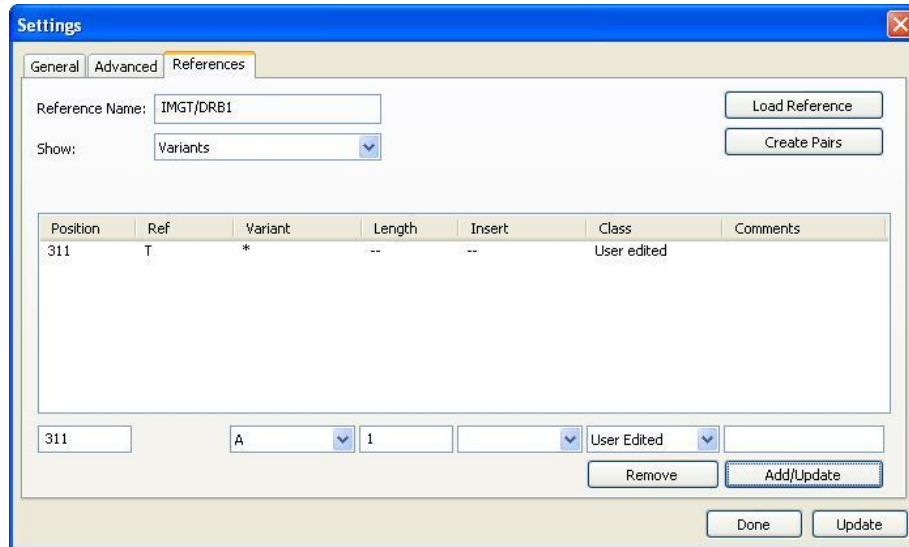
3.3.6. To add an alias, type the alias into this box. For example, TT197 as an alias for the RB-TT197-F. Click **Add Alias** directly below the alias drop down menu box. Then, click **Update** in the lower right corner. After all aliases have been added, click **Done**.

3.4. Variant Positions

Variant Positions is a tool to draw attention to sequence artefacts that may result in base call errors, or other positions within a sequence where automated base calls may be incorrect frequently. Such positions can be included as those positions within the sequence that must be validated by the user in order to generate a report.

3.4.1. Load the reference for which you want to create a variant position following the instructions above.

3.4.2. Once the reference is loaded, click on the **Show** dropdown then select **Variants**. This will open the variant position box.



3.4.3. **Create Pairs** enables the users to create a heterozygous text sequence from selected allele pairs.

3.4.4. In the lower left corner, in the blank box under **Position**, type in the position for the variant.

3.4.5. In the **Variant** drop down, select the variant base type (usually * so any call at that position is flagged).

3.4.6. Enter the **Length** of the variant in bases.

3.4.7. In the **Insert** box, enter bases if this variant includes an insertion. Leave blank if no insertion is expected.

3.4.8. Select the **Class** of variant (usually **User Edited**). Enter any **Comments** desired.

3.4.9. Click **Add/Update** to add the variant position.

3.4.10. Repeat for additional variants.

3.4.11. Click **Update** then **Done** when complete. A purple box will be displayed at each variant position indicated in the layout (described in further detail in 5.2.5.3.4 of this manual).

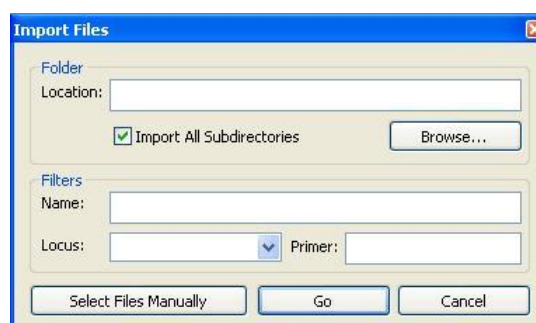
4. Importing Sequences

Once the sequence file-naming convention is defined in the software **Settings**, sequences can be imported by browsing to a directory and importing its contents or by importing the sequence files individually. Importing from a directory also allows filters to be applied so that only specific samples are imported, or those sequences from a particular locus or sequencing primer.

4.1. Importing sequences by directory

4.1.1. Open a new layout by going to **File** and selecting **New**. Ctl-N also performs this function. To import sequences into an existing layout select **File|Open** and navigate to the location of the required layout. Ctl-O also performs this function.

4.1.2. To import sequences by directory (i.e. import all electropherograms in a given folder), select **File | Import | Electropherograms** on the top menu bar.



- 4.1.3. In the pop up window, click on **Browse** and navigate to the folder that contains the sequences. Highlight the folder and click **OK**.
- 4.1.4. The **folder location** will populate the **Import Files** menu. Check the **Import All Subdirectories** box if the content of all subdirectories are to be imported. Click **Go**.
- 4.1.5. Use the **Filters** dialogue to import samples of a specific **Name**, or for all samples for a given **Locus** or for sequence files generated from a specific sequencing **Primer**.

4.2. Importing sequences individually

- 4.2.1. Open a new or existing layout, as described in 4.1.1
- 4.2.2. To import only selected electropherograms from a folder, click on **Select Files Manually**. Navigate to the folder containing the necessary sequences. Highlight all the sequences to be imported using the Ctrl or Shift key. Then click **Open**.

5. The Screen layout, Editing and Analysis

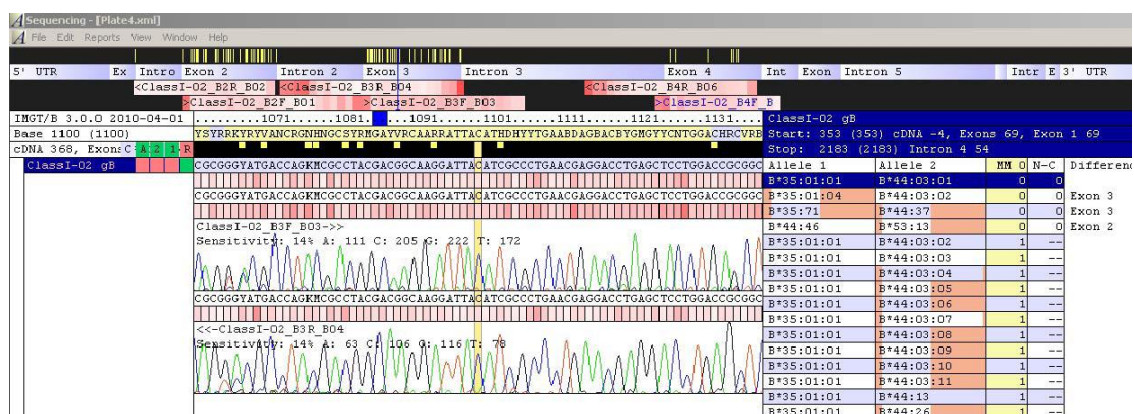
Once the sequence data has been imported, the software screen becomes populated with sample, sequence and result information. There is a predominance of white to red shading used to demonstrate sequence data quality

The use of shading to indicate sequence data quality

Sequencing based typing errors may occur if a base call error is made, and the probability of a base call error is increased if the quality of the data is poor. Assign™ SBT contains a quality scoring algorithm that assesses the quality of a sequence peak based on the peak's shape, whether or not it is well separated from neighbouring peaks and whether or not there is non-specific background.

A Base Call Score (BCS) from 0-50 is calculated for each peak and is represented in a box under the base call as a shade from red to white, where red is a BCS of 0 and white has a BCS of 50, while a BCS between 0-50 is shaded accordingly. The consensus sequence BCS is calculated from the BCS of sequences that contribute to the consensus. The BCS for positions within a sequence can be used to calculate a quality score of a sample.

The use of visual colour shading enables a sample with poor quality data and/or poor quality positions within the sequence to be readily identified and checked for possible base call errors.



5.1. Sample ID Pane

The screen layout shows information for a particular sample. It includes the sample ID, the electropherogram data, the aligned sequences for a sample and the best matched allele combinations.

The samples imported into the layout are listed on the left side of the screen. Sample names are colour coded to indicate overall data quality for the sample.

Exon	Exon 2				
>1	U318_A2F_A02_02				
<1	U318_A2R_A02_02				
IMGT/HLA A 2007-04-12					
Base 799 (871)					
Exon 4	252	C	A	2	1
U312	A				
U313	A				
U315	A				
U316	A				
U317	A				
U318	A				
U319	A				
U320	A				
U321	A				
U322	A				
U323	A				
U324	A				
U325	A				

5.1.1. The sample panel also include 5 columns of boxes (see figure above).

5.1.1.1. A light blue box in the **C** column indicates if **comments** have been made about a sample. Right clicking on the sample allows comments to be added and reviewed. These comments are included in the genotype report.

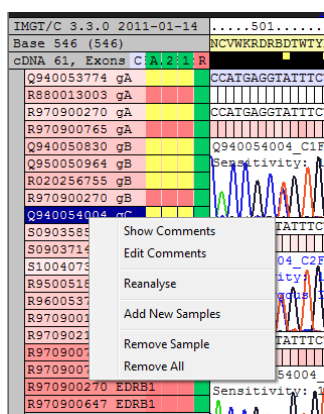
5.1.1.2. A green box in the **A** column indicates that the sample has been verified at all positions indicated in the Navigator. This box changes to green once all the positions requiring confirmation have been confirmed using the navigator. (See below for Navigator bar use.)

5.1.1.3. A green box in the **1** column indicates that the sample has undergone the **first review**. After the first reviewer has performed the analysis, the yellow box in the 1 column must be clicked to change it to green.

5.1.1.4. A green box in the **2** column indicates the sample has undergone a **second review** by another reviewer. Checking this box will lock the sample and prevent any further edits unless the box is manually unchecked.

5.1.1.5. A green box in the **R** column indicates the sample can be reported using the **Report Generator**.

5.1.2. Right click on the sample name in the Sample Pane to access sample options.



5.1.2.1. **Show Comments** will display any quality warnings or comments about a sample.

5.1.2.2. **Edit Comments** provides a text box to record any comments about a sample. These comments will appear on the report. A light blue box in the C column will indicate that a comment is present.

5.1.2.3. **Reanalyse** will remove any edits and trims that have been made, restoring the sample to the initial state following import.

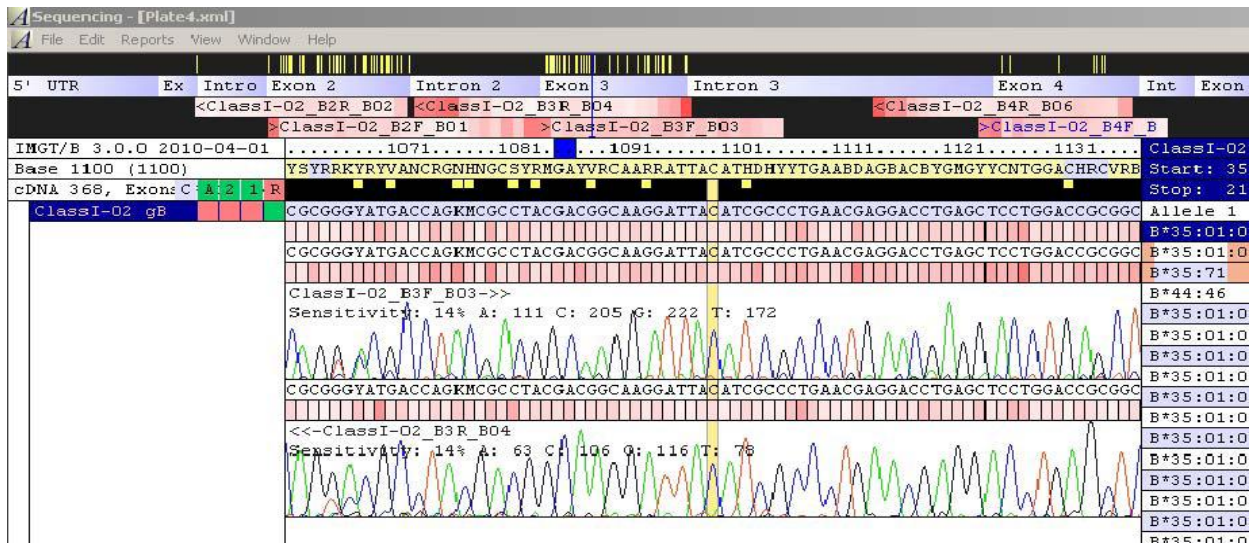
5.1.2.4. **Add New Samples** will launch the import files menu.

5.1.2.5. **Remove Sample** will remove the highlighted sample from the project.

5.1.2.6. **Remove All** will remove all samples from the project.

5.2. Sequence electropherograms

Importing the sequence electropherograms results in a display of how the sequence files are orientated according to the gene structure, the sequence electropherograms themselves, the Assign™ base calls and quality score information.



The Assign™ layout contains important information to assist with the analysis of DNA sequence data.

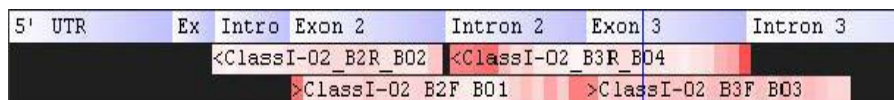
5.2.1. The Structure of the Gene being Sequenced

The blue band shown below indicates the genetic structure of the reference sequence. The yellow bars above this indicate those positions within the sequence which differ between closely related allele combinations listed in the results pane.



5.2.2. Sample Sequence Alignments

The bands shaded white to red indicate the sequence data alignments. They are shaded white to red according to sequence quality. This enables “at-a-glance” location of poor quality regions for manual review. The sequence filename and direction of sequencing (< or >) is also included.



5.2.3. The Library Consensus Sequence

Beneath the sequence alignment map is the consensus sequence of alleles within the library.

```
.....1071.....1081. ....1091.....1101.....1111.....1121.....1131.
YSYRRKYRYVANCRGNHNGCSYRMGAYVRC&ARRATTACATHDHYTGA&BDAG&B&CBYGMGYC&NTGG&CHRC
```

The sequence is shaded yellow and white to indicate exonic and intronic sequences, respectively. In addition, positions shaded light blue indicate there are alleles in the library that contain deletions at this position and dark blue regions indicate the position of insertions in some alleles.

5.2.4. The Sample Consensus Sequence

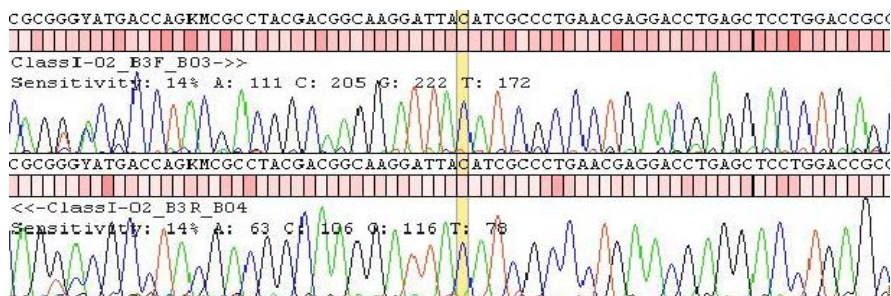
The sequence data below the library consensus sequence is the sample consensus sequence.

```
CGCGGGYATGACCAGKMC&CCT&CG&ACGG&CA&GG&ATT&CATCGCCCTGA&AC&G&G&ACCTG&AGCTCCTGG&ACCG&C
```

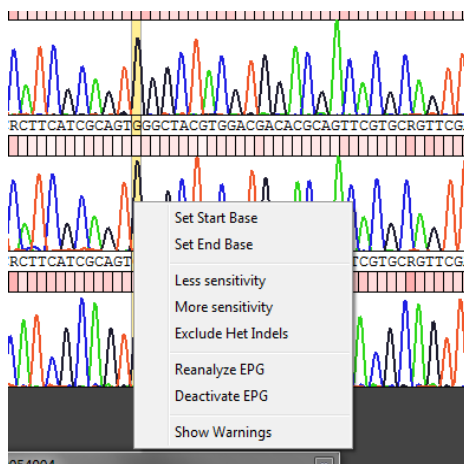
The boxes underneath the base calls are shaded white to red to indicate the quality of the consensus base call. Note: it is the sample consensus sequence that is compared to the sequence library and this is where any base call edits are incorporated into the analysis.

5.2.5. The Sample Sequence Data

Beneath the sample consensus sequence is the sample sequence electropherogram data, the software base calls and quality indicator. The electropherogram panels contain the sequence filename and the “sensitivity” of heterozygous base calls. That is, the percentage that one peak needs to be within another before a heterozygous base call is made on data with no background. The signal intensities of the 4 bases are also indicated.



5.2.5.1. Right clicking on a given electropherogram gives access to options for each electropherogram; any changes made will only be applied to the selected electropherogram.



- 5.2.5.1.1. **Set Start Base** will trim off all data to the left of the cursor.
- 5.2.5.1.2. **Set End Base** will trim off all data to the right of the cursor.
- 5.2.5.1.3. **Less sensitivity** will filter out background noise, raising the threshold by 10% to a maximum of 50%.
- 5.2.5.1.4. **More sensitivity** decreases the threshold to a minimum of 10% thus increasing sensitivity, calling more heterozygous bases.
- 5.2.5.1.5. **Exclude Het Indels** will filter out excess background in the sequence due to a co-amplified stretch of DNA. When the signal to noise is relatively low, the extra signal can be interpreted as a heterozygous Indel. Choosing "Exclude Het InDel" informs the software that the additional signal should be interpreted as background and not as a real sequence feature.
- 5.2.5.1.6. **Reanalyze EPG** will remove any user edits and trims from the electropherogram.
- 5.2.5.1.7. **Deactivate EPG** will remove the electropherogram from analysis, but does not remove it from the layout. Right clicking on the deactivated electropherogram again enable the option to **Activate EPG**.
- 5.2.5.1.8. **Show Warnings** will display any quality warnings about that particular electropherogram.

5.3.4. Selecting the additional HARP® column will result in a shortened allele list by filtering out any allele combinations with mismatches to the consensus sequence. The remaining alleles listed which are not sequenced by the HARP® will be greyed, allowing the user to clearly identify which alleles are sequenced by the HARP®.

5.3.5. The **N-C column** (see figure below) indicates the number of mismatches in the non-coding region. The **N-C** column appears if genomic references are active. The Non-Coding analysis may be activated by clicking on the N-C column, displaying the number of mismatches between the test sequence and reference alleles.

If either the N-C or IND column (see 5.3.6) is shaded pink, the data from these regions are NOT included in the analysis. To activate the N-C and IND layers to include the data in the analysis, click on the appropriate column. Clicking again will deactivate the layers. When the N-C and IND layers are activated, the data is included in the analysis and the data within these layers must be edited.

NOTE: If the **Use Genomic References** box is not checked in the **Edit/Settings/Advanced** tab, the N-C column is not created.

Allele 1	Allele 2	MM	N-C	IND	Differences
A*02:01:01:01	A*31:01:02	0			
A*02:01:01:02L	A*31:01:02	0			5' UTR
A*02:01:01:03	A*31:01:02	0			Intron 1
A*02:20:01	A*31:02	0			Exon 2
A*02:40	A*31:41	0			Exon 3
A*02:243	A*31:21	0			Exon 2
A*02:01:01:01	A*31:01:03	1			
A*02:01:01:01	A*31:01:04	1			
A*02:01:01:01	A*31:01:05	1			
A*02:01:01:01	A*31:01:06	1			

5.3.6. The **IND column** contains mismatch information in the heterozygous insertion/deletion (indel) data. The software is sensitive to runs of mixed bases since many introns have heterozygous indels.

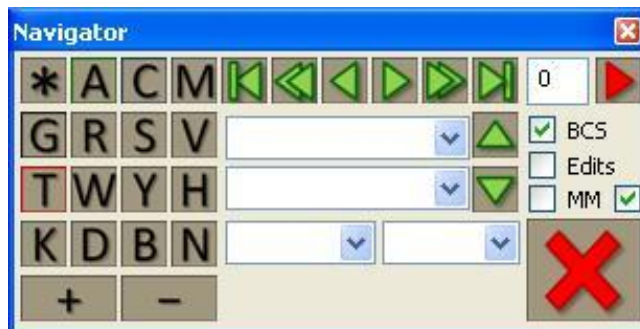
When there is background in the sequence due to a co-amplified stretch of DNA or when the signal to noise is relatively low, the extra signal can be interpreted as a heterozygous Indel. Choosing "**Exclude Het Indel**" informs the software that the additional signal should be interpreted as background and not as a real sequence feature.

NOTE: If the software does not detect heterozygous insertion/deletion data, the IND column is not created.

5.3.7. The **Differences** column indicates the regions within the reference sequence that contain the sequence differences between the ambiguous allele combinations.

6. Navigator

The navigator enables sequence editing, moving between samples and moving between positions within a sequence. Importantly the Navigator is used to validate automated base calls.





6.1. Positions for Validation



6.1.1. The user can select which positions to include for validation by checking the box for positions with a low Base Call Score (**BCS**), edited positions (**Edits**), potential Mismatch positions (**MM**), and user defined variant positions (**box to the right of the MM**). Bases with a

BCS of lower than 70 (or lower than 35 for single direction sequences) will be included for validation if the **BCS** box has been checked.



6.1.2. **Base Call Score (BCS)** is a quality measurement that is determined by peak spacing, presence of background noise, and signal strength. The BCS for each base appears above the BCS box in the navigator.



6.2. Moving Between Validation Positions



Moving to positions for validation is performed by selecting the double arrow buttons or using the  button. The  button indicates that the current position has not been validated by the operator.

Clicking the  button validates the base call and changes the  to a green tick, and the software will then proceed to the next position to be validated. Once all positions have been validated, the green tick will remain.

6.3. Arrow functions

6.3.1. Selecting either   (single arrow) button moves the EPG one position left or right.

6.3.2. Selecting either   (double arrow) button moves the EPG to the next position which requires validation.

6.3.3. Selecting either   (blocked arrow) button moves the EPG to the start or the end of the sequence.

6.3.4. Selecting the up or down arrow moves to the sample above or below in the Sample Pane.



6.4. Other Navigator functions

6.4.1. The **Master drop down menu** selects between master sequences, HARP sequences, Master-intron, or Master-indel sequences.

6.4.2. The **No Offset drop down menu** allows the user to choose the base numbering motif desired.

6.4.3. The **Codon and Base locations** are located underneath the No Offset drop down menu. To navigate to a particular base, enter the base position in the right drop down and hit enter.

6.5. Confirming or Editing Base Calls

6.5.1. Clicking on the  button will **confirm the base call** and move to the next position that meets the “to be validated” criteria. As  is clicked, a green box will appear above the base indicating that it has been confirmed. Once ALL validation positions have been confirmed, the yellow box under the **Audit** column in the Sample Pane will turn green.

6.5.2. If a **base call needs to be edited**, the call can be changed manually using the base letters on the Navigator. The raw data is not changed with edits. The consensus sequence is changed, and these changes are recorded in the saved project. When the project is opened, the changes are reapplied to the raw data.

7. Other Sample and Sequence Editing Functions

7.1. Resizing the EPG

The EPG can be resized by pressing the **Shift** key and the **up/down** or **left/right arrows** on the computer keyboard.

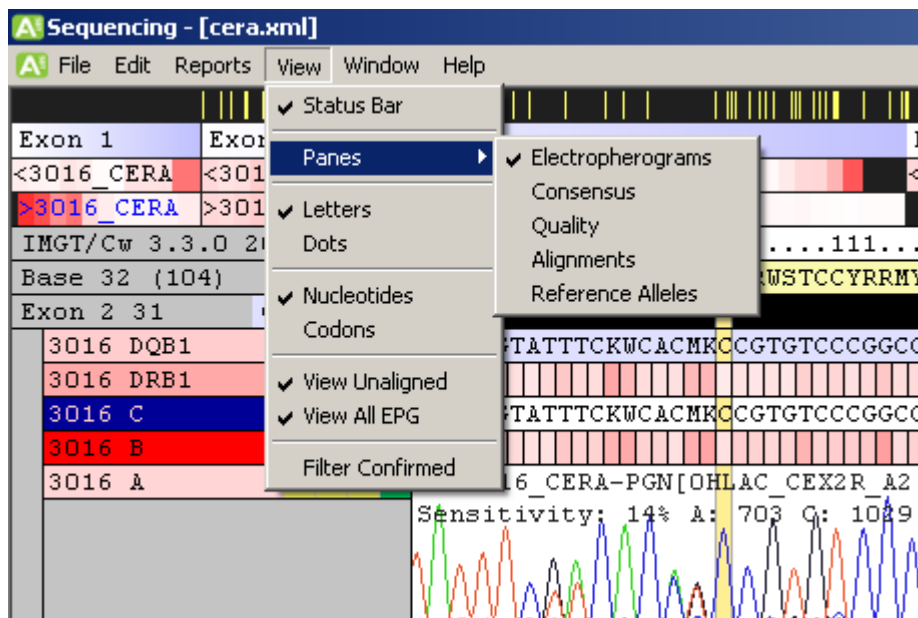
7.2. Hiding EPG traces

Pressing the **Shift** key and one of the letters representing the 4 bases (i.e. **A, C, G** or **T**) simultaneously will remove the trace of this base from the EPG. Repeating the process will return the trace.

This function is useful if heterozygous peaks are perfectly overlaid and the base call requires confirmation.

8. View Options

The **View** options enable the sample sequence data and the sequence data from the alleles in the library to be viewed in different ways.



8.1. Status Bar

The **Status Bar** at the bottom of the screen shows the status of the project.

8.2. Panes Functions

8.2.1. **Electropherogram** is the default view setting and displays the electropherogram tracings of the sample.

8.2.2. **Consensus displays** the consensus sequence for all samples in a project.

8.2.3. **Quality** displays the consensus sequence shaded according to the consensus sequence base call score for each base for every sample in the project.

8.2.4. The **Alignments** option displays the consensus sequence for each of the possible allele combinations for a given sample. Mismatches with the sample consensus appear highlighted in yellow.

8.2.5. **Reference Alleles** shows the sequence of alleles within the library compared to the sequence of the selected sample. Differences are highlighted in yellow. The alleles are shown in the Results Pane.

The user can select specific allele sequences to align together by typing the allele names into the text box at the bottom of the Navigator.

8.3. Letters and Dots

8.3.1. Selecting **Letters** will show the bases for alignments and reference as letters.

8.3.2. Selecting **Dots** will show dots at each base where the alignment or reference matches the consensus sequence. Bases that differ will be shown as letters.

8.4. Nucleotides and Codons

8.4.1. Selecting **Nucleotides** will show the base numbering.

8.4.2. Selecting **Codons** will show the codon numbering.


8.5. View Unaligned

When using the coding sequence reference, the **View Unaligned** option will include or exclude the intronic overlap of the sequences between exons.

8.6. View All EPG

View All EPG enables the HARP electropherograms to be seen with the EPG of the F and R sequences. Note that the highlighted sequence in the results pane (yellow column) is highlighted in the EPG pane. By selecting the HARPs layer, the HARP's EPGs become highlighted allowing edits/validation.

8.7. Filter Confirmed

When **Filter Confirmed** is selected, positions confirmed using the , alleles within the results pane are excluded and only those alleles with 0 mismatches with the test sequence remain. To restore the list of possible allele combinations, unclick on the **Filter Confirmed** option in the view pane.

9. Data Analysis and Editing EPGs

9.1. Logging On

Log into the Assign™ SBT 3.6+ software and select the settings file desired by clicking on **Edit** then **Settings**. Select the settings file and click **Done**.

9.2. Importing Data

9.2.1. Open a new or existing project layout by selecting **File|New** or **File|Open**. Refer to 4.1.1 for more details.

9.2.2. Import the desired data set using either the **Browse** option to import an entire folder of data or the **Select Files Manually** to selectively import files.

9.2.3. Imported samples will be displayed as a list on the left side of the screen.

9.2.4. The electropherogram data will be located in the centre of screen.

9.2.5. The allele assignments for the active (highlighted in blue) sample will be displayed at the right side of the screen.

9.2.6. Each of these panes can be sized by dragging the frame to the desired width to optimize the amount of electropherogram data displayed.

9.2.7. Resize the electropherogram peaks if desired.



9.3. Navigation

9.3.1. The Navigator box is used to navigate through the data checking the critical bases.

9.3.2. Set the desired bases to be audited in the Navigator box: BCS, Edits, MM, and variant positions. Refer to 6.1 for more details.

9.3.3. Highlight the sample to be reviewed by clicking on it in the left sample pane.

9.3.4. Navigate to the first base in the sample by clicking on the left arrow with the bar in the navigator.

9.3.5. Using the  button, navigate through the sequence verifying all desired bases. As each base is confirmed, the  will change to a green tick and the cursor will move to the next base to be verified. A green box will appear above each base pair that has been verified using this method.

9.3.5.1. There should be at least one allele pair in the Results Pane on the right that indicates no mismatches in the MM0 column by a "0".

9.3.5.2. Once the Master Layer has been reviewed, the Navigator will take the user to any resolution primer layers that are present. These layers must also be reviewed before the sample analysis is complete.

9.3.5.3. Once all priority review bases have been verified, the will change to a green tick indicating no additional bases need confirming. In the Sample Pane, the box under the A column (Audit) will turn green.

9.3.5.4. Clicking on the box under the **1** column will indicate the sample has been reviewed once.

9.3.5.5. At this point, the project should be saved to prevent any accidental loss of data review. Click on **File** then **Save As** (for new layouts) or **Save** (for existing layouts). Select a file name and location to save the project. The saved project (.xml format) indicates which electropherograms were used including their saved location, any edits and confirmations that were done, and information about the user.

9.3.5.6. If **Genomic References** were selected, the NC and IND column may be present. Refer to the differences column to determine if analysis of the non-coding regions is beneficial. Click on the **NC** column to highlight it in yellow. Navigate to the mismatch positions and make any edits necessary. Once all review and edits have been made, click back on the **MMO** column to view the allele pairs.

9.3.5.7. Repeat on the **IND** column if desired. The IND column should be reviewed only when there is a clear insertion or deletion in the sample. Poor quality or mobility shifts can trigger the Heterozygous Insertion or Deletion warning. Use the **Exclude Het Indel** option to eliminate any warnings due to background noise and not true Indels. Below is an example of a clear indel (boxed).



9.3.5.8. Heterozygous insertion/deletion data can be analysed and edited the same way as all other sequence data. The consensus sample sequence should include the sequence present in individual forward and reverse sequence files.

9.3.5.9. Once the review has been completed on the first sample, click on the second sample and repeat the process.

NOTE: Save your work often to prevent any loss of review data.

9.3.6. If a secondary review is desired, the project should be opened by the second reviewer to preserve the audit trail. The second reviewer logs onto the software and selects **Open** in the **File** menu. Navigate to the saved project (.xml file). The software will locate the raw data, import it and apply all changes and verifications that have been performed by any previous reviewers.

9.3.6.1. Select the desired priority review positions in the Navigator box. Using the left **double arrow button** to navigate to the required positions. Using the button will cancel and reapply the audited positions.

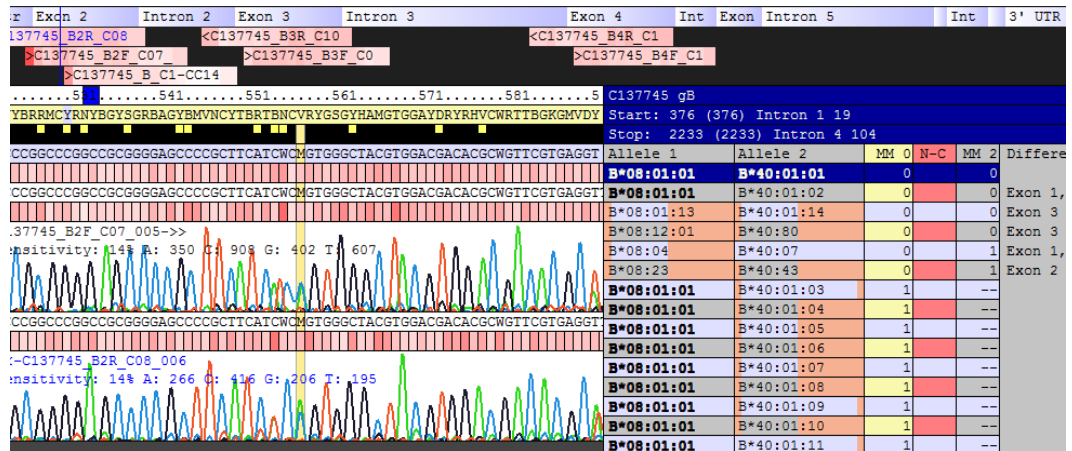
9.3.6.2. Once all positions have been reviewed, the second reviewer can click on the box in the **2** column in the sample pane.

9.3.6.3. Once the second review box is checked, the sample is 'locked' and no more edits can be made unless the box is manually un-checked.

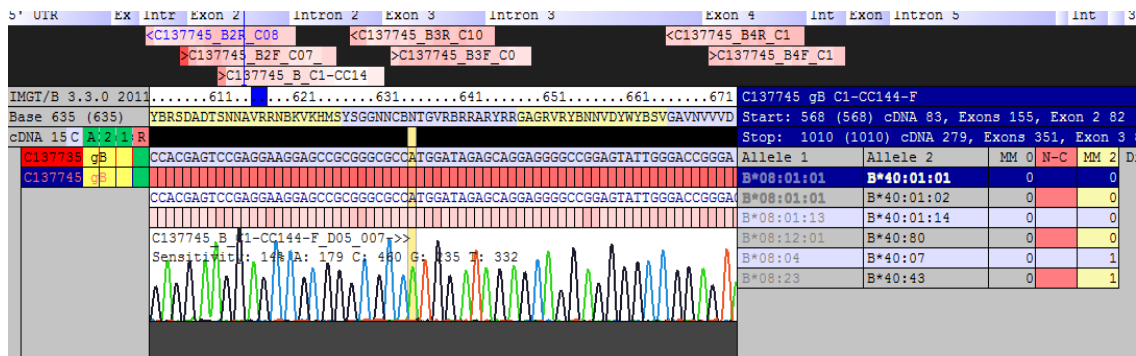
9.3.6.4. Save often to prevent any accidental loss of review data.

9.4. Resolution Primer Layers

Heterozygous Ambiguity Resolution Primers (HARPs[®]) are used to resolve heterozygous ambiguities within the regions sequenced. HARPs[®] target one of the alleles present, creating a hemizygous sequence. In the example below, multiple allele combinations match the target sequence resulting in “ambiguities” upon completion of the analysis of the master sequence, as indicated in the MM0 column highlighted in yellow.



Use of a HARP[®] that will sequence only one of the alleles in the pair will result in a hemizygous sequence in the MM1 layer (shown below). This will allow the elimination of some of the heterozygous ambiguities. Allele combinations with no mismatches in the MM0 and MM1 column will be included in the report. Alleles with mismatches in the MM1 column have been eliminated as a possible allele ambiguity by the HARP[®] sequence.



9.5. Auditing

All user interaction is logged in the audit trail. The software logs the time and date when any edits were made, when the project was saved and the user performing the action. This information can be printed on the allele report along with the sample allele assignments.

10. Reporting

The Assign[™] SBT 3.6+ reports enable a comprehensive assessment of the sequence data. A genotyping report lists the alleles best matched to the sample sequence and also enables CWD alleles to be indicated. The genotype report options also allow alleles to be reported as functional groups by reporting G groups and P groups and enable the user to structure the report specific for their requirements. The software can also report the HARPs[®] required to resolve the heterozygous ambiguities.

To access the report functions, click on **Reports** then **Report Generator** on the top menu bar.

10.1. Genotyping Report

The Genotyping report is used to report the allele combinations that have identical sequence to the sequence of the sample.

10.1.1. The **Output Filters** can be used to filter for a single sample/locus or all samples/loci in a project

10.1.2. The **Full Report Section** enables laboratories to customise their report format. The drop down menus under **Sample** enables the laboratory to include or exclude specific items from the report.

10.1.2.1. The **Sample** section contains the **Auditing, Match Summary, G Groups,** and **P Groups** reports options.

10.1.2.1.1. **Match Summary** will list all the matched allele pair combinations for all samples selected. If the CWD option is selected, CWD alleles will appear in bold type in the report.

Sample:	Q950050964	
Reference:	IMGT/B 3.3.0 2011-01-14	
Summary		
The allele pairs listed below are compatible with the consensus sequence.		
B*18:01:01	B*44:02:01:01	
B*18:01:01	B*44:02:01:02S	Intron 4
B*18:01:05	B*44:02:01:01	Exon 2
B*18:01:05	B*44:02:01:02S	Exon 2, Intron 4
B*18:01:05	B*44:27:01	Exon 2
B*18:09	B*44:09	Exon 2
B*18:12	B*44:12	Exon 2
B*18:20	B*44:51	Exon 3
B*18:43	B*44:55	Exon 2

10.1.2.1.2. The **G Group Report** option enables those Class I alleles with identical nucleotide sequence in exons 2 and 3 to be reported under the same code. Class II alleles that are identical in exon 2 are also reported using this report option. Such G groups are designated by an upper case 'G' which follows the first 3 fields of the allele designation of the lowest numbered allele in the group. Groups containing CWD Alleles will be bolded on the report.

Sample: Q950050964	
Reference: IMGT/B 3.3.0 2011-01-14	
G Group Summary	
The allele pairs listed below are compatible with the consensus sequence.	
B*18:01:01G	B*44:02:01G
B*18:09	B*44:09
B*18:12	B*44:12
B*18:20	B*44:51
B*18:43	B*44:55

10.1.2.1.3. The **P Group Report** option will list the alleles grouped according protein sequences as encoded by exons 2 and 3 for HLA Class I alleles, and exon 2 only for HLA Class II alleles. P Groups containing CWD alleles will be bolded on the report.

Sample: Q950050964	
Reference: IMGT/B 3.3.0 2011-01-14	
P Group Summary	
The allele pairs listed below are compatible with the consensus sequence.	
B*18:01P	B*44:02P
B*18:09	B*44:09
B*18:12	B*44:12
B*18:20	B*44:51
B*18:43	B*44:55

10.1.2.1.4. The **Auditing** option will include a comprehensive audit report including date, time, and identification of the operator validating the results.

10.1.2.2. The **Layers** section contains the **Edit List**, **Electropherogram List**, **Sequences**, **Mismatch List**, and **Mismatch table** options. These can be included or excluded according to the needs of the laboratory.

10.1.2.3. The **Additional Information** section can be used to add comments specific for a typing run. These comments appear at the top of the report.

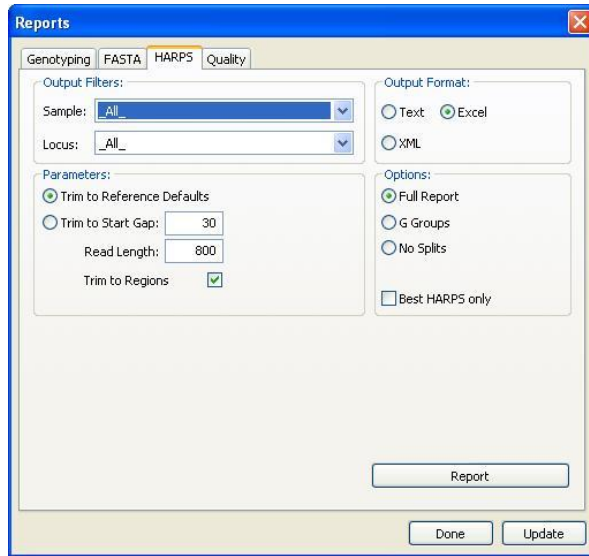
10.1.2.4. **Sort by** is used to sort the report by **Sample Name** or **Locus**.

10.1.2.5. **Summary Options** includes options to add **NMDP Codes** and **HARPS®** used to the report. The **Full+Part** option indicates which alleles are fully typed in the IMGT database, while **Differences** indicates where each allele pair differs from the others.

10.1.2.6. **Audit Options** allows the operator to choose to report all the Save events for the project by clicking the **Save** box. **Confirm** records change and priority base confirmation.

10.2. HARPS Report

10.2.1. The HARPS® report indicates to the operator which HARPs® are required to resolve heterozygous ambiguities.



10.2.2. The **Output Filters** can be used to filter for a single sample/locus or all samples/loci in a project.

10.2.3. **Parameters** enable the user to obtain maximum benefit from the HARPs[®] by customising the settings for their sequencing capabilities. The analysis parameters are defined by maximum sequence read length and the start of good quality sequence from the HARP[®].

10.2.3.1. **Trim to Reference Defaults.** The references defaults are set to report HARPs[®] with a start gap of 20 bases and to limit the sequence read length to the exon to which the HARP[®] has been designed. The exceptions to this are those HARPs[®] that are designed to anneal immediately before, and in the direction of, intron 2. These HARPs[®] are designed for analysis of the entire neighbouring exon.

10.2.3.2. Selecting **Trim to Start Gap** enables the operator to tailor the HARPs[®] report and analysis to their sequencing capabilities. If good quality sequence cannot be obtained less than 30 bases from the start of the HARP[®], then the user should set the **Trim to Start Gap** to 30 or greater. If the user does not routinely produce sequence of more than 400 base pairs the read length should be set to 400 or less. If **Trim to Regions** is selected, the analysis will stop at the end of the exon, regardless of the read length setting.

See the appendix 1 for more details regarding the HARPS report.

10.2.4. The operator can choose between a **Full Report** and a **G Groups Report** under the **Type** section. The HARPS report will list all HARPs[®] that will resolve the heterozygous ambiguity. Selecting **Best HARPs[®] only** reports only the HARP[®] with the highest HARP[®] score.

10.2.5. The **Full Report** will list the HARP resolution for each allele pair.

10.2.6. When **No Splits** is selected, only the HARPs[®] reported are shown in the report.

Sample: R020245221
Reference: IMGT/B 3.3.0 2011-01-14

Use ONE primer from group 1: C1-CT97-F (254)
 Use ONE primer from group 2: C1-GA559-R (507) C1-AC559-R (507)

Splits:

B*07:02:01	B*08:01:01
B*07:61	B*08:01:01
B*07:02:03	B*08:01:11
B*07:02:14	B*08:01:13
B*07:05:01	B*08:07
B*07:06	B*08:07

Unresolved ambiguities remain within Exon 5.

10.2.7. The **G Group Report** will group the resolution based on the broader G Grouping of the allele pairs.

NOTE: G groups that contain a CWD allele are shown in black cells.

Sample: R020245221
Reference: IMGT/B 3.3.0 2011-01-14

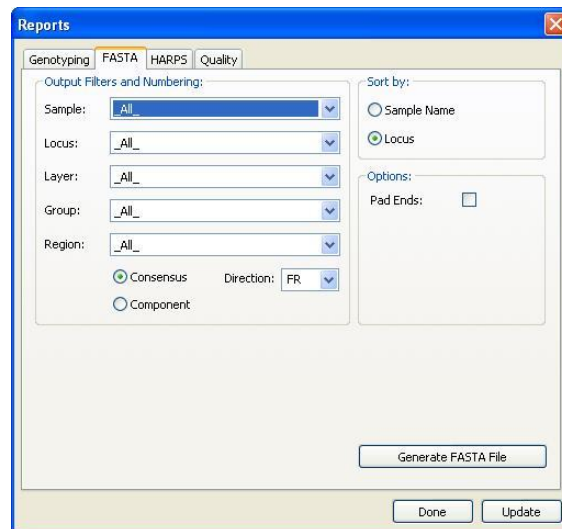
Use ONE primer from group 1: C1-CT97-F (254)
Use ONE primer from group 2: C1-GA559-R (507) C1-AC559-R (507) C1-GG539-R (507)

Splits:

B*07:02:01G	B*08:01:01G
B*07:02:03	B*08:01:11
B*07:02:14	B*08:01:13

The **Output Format** can be in a **Text** file, **Excel** (default) or **XML** format.

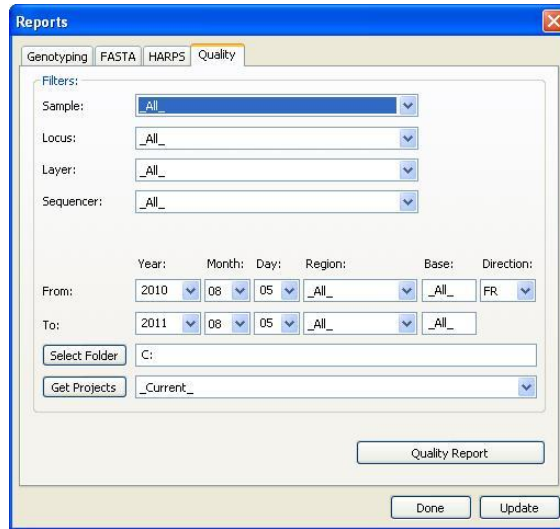
10.3. FASTA Report



The FASTA report allows the production of sequences in FASTA text format. Selecting the **sample**, **locus**, **layer**, **group** and **region** provides a detailed description of the FASTA file in the FASTA file name.

10.4. Quality Reports

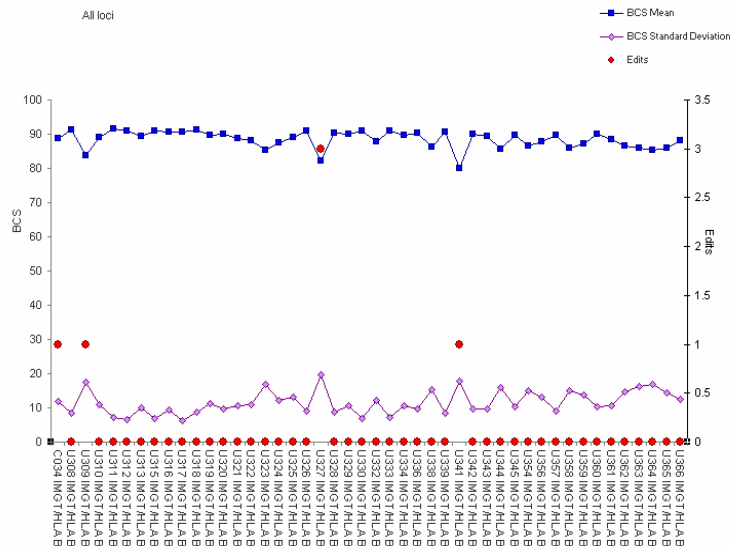
The base call score at each position is used to create the quality control information for each sample, which can then be compared between different samples to create a quality report for the assay. The principle is that if the mean and standard deviation of base call scores for a number of sequence positions can be calculated this will provide a quality value for the region of sequence from which the mean and standard deviation are calculated. This information can then be used to calculate quality information for specific sequencing primers, for different assays and different samples. The data can be used to monitor the performance of a test and set performance criteria that can be used when assessing changes, such as reagent batch changes, or DNA extraction procedures.



10.4.1. The **Quality** report dialogue enables the user to select parameters for quality analysis. In addition to the sample information, the user can select across regions (i.e. an exon) within the sequence to analyse, or a specific range of bases within a region

10.4.2. Leaving the **Get Projects** default at **_Current_** will produce a quality report in the active Assign™ SBT layout.

10.4.2.1. By Clicking **Get Projects** and then browsing to directories in the **_Current_** drop down menu, a quality report can be generated from saved layouts within the browsed directories.



The above Quality Report shows the mean BCS in blue and the standard deviation in pink for all HLA-B exon 2 samples for a single SBT run. The red dots show the number of edits made for each sample.

Appendix 1

Analysis parameters for HARPS® when the HARPS report is set to **Trim to Reference Defaults**.

The following HARPS® are designed to anneal to motifs in one exon to sequence the neighbouring exon. Note that the read length for such HARPS® must consider the intron length of approximately 250 bp for class I and the length of the exon so a minimum sequence read length in excess of 550bp may be required.

C1-GT355-R

C1-GG362-R

C1-CG319-F

C1-CG343-F

C1-CA343-F

References

1 Cano P, Klitz W, Mack SJ et al (2007) Common and well-documented HLA alleles: report of the Ad-Hoc committee of the American Society for Histocompatibility and Immunogenetics., Hum Immunol. 68(5):392-417. Epub 2007 Feb 15.

Support and Contact Details

Conexio Genomics Pty Ltd
8/31 Pakenham St
Fremantle 6160
Western Australia
Tel: +61 422863227
email: support@conexio-genomics.com
Skype: conexiocgx
Website: www.conexio-genomics.com

Or your local distributor

Conexio, Assign and HARPS are trademarks of Conexio 4 Pty Ltd. HARPS is a registered trademark in some territories.

For ordering details, please refer to the Olerup website (<http://www.olerup.com>).