



ASSIGNTM ATF v1.5

Variant Detection and Genotyping Sequence Analysis Software

User Manual

Version No: 4.0

Issue Date: January 2014

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

REF

CGX00150



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

Table of Contents

| | |
|---|----|
| Table of Contents..... | 2 |
| About Assign™ ATF 1.5..... | 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..... | 6 |
| 2.3 Limitations | 6 |
| Quick Guide | 7 |
| 3. Brief Description of Assign™ ATF 1.5 Functions - Sufficient Information for BASIC Analysis ... | 7 |
| 3.1 Installation – Full Details in Section 4 | 7 |
| 3.2 Login – Full Details in Section 5..... | 8 |
| 3.3 Defining Analysis Settings – Full Details in Section 6.2..... | 8 |
| 3.4 Create Reference Sequence – Full Details - Section 6.3 | 9 |
| 3.5 Entering File Naming Convention – Full Details in Section 6.4 | 13 |
| 3.6 Import your Sequences – Full Details in Section 7 | 14 |
| 3.7 Sequence Analysis and Editing – Full Details in Section 8..... | 15 |
| 3.8 Producing a Report – Full Details in Section 9 | 16 |
| 3.9 Saving and Opening Projects – Full Details in Section 10..... | 16 |
| Detailed Guide - Getting Started and Using the Software | 17 |
| 4. Installation | 17 |
| 4.1 To Install..... | 17 |
| 5. Login and User Management..... | 19 |
| 5.1 Login..... | 19 |
| 5.2 Adding Users..... | 19 |
| 5.3 Changing a Password | 20 |
| 5.4 System File Location..... | 20 |
| 6. Setting up Assign™ ATF 1.5 for Analysis | 21 |
| 6.1 General Settings Tab | 21 |
| 6.2 Data Analysis Settings (Engine Tab) | 22 |
| 6.3 Creating Reference Sequences (Reference Tab)..... | 25 |
| 6.4 Naming Conventions (Naming Tab)..... | 33 |
| 7. Importing Sequences for Analysis | 36 |
| 7.1 Importing Sequences by Directory | 36 |
| 7.2 Importing Sequences Individually | 37 |
| 8. The Screen Layout, Editing, and Analysis | 38 |
| 8.1 Sequence Data Quality is indicated by Shading | 38 |
| 8.2 Sample ID Pane | 38 |

| | | |
|---|--|----|
| 8.3 | Sequence Pane | 40 |
| 8.4 | View Options | 44 |
| 8.5 | Results Pane | 46 |
| 8.6 | Navigator | 47 |
| 8.7 | Priority Editing and Confirming Base Calls | 48 |
| 8.8 | Reviewing the Analysis | 49 |
| 8.9 | Audit Trail | 49 |
| 9. | Reporting | 50 |
| 9.1 | Variants Reports | 50 |
| 9.2 | Genotyping Reports | 51 |
| 9.3 | FASTA Reports | 53 |
| 9.4 | Quality Reports | 54 |
| 10. | Opening, Saving, and Printing Projects | 56 |
| 10.1 | Select the Desired Settings File | 56 |
| 10.2 | Opening Projects | 56 |
| 10.3 | Saving Projects | 56 |
| 10.4 | Printing Projects | 56 |
| Appendix 1: Frequently Asked Questions | | 57 |
| Installation FAQs | | 57 |
| I Am Using a Previous Version of Assign™ ATF 1.5 – How Do I Install v1.5? | | 57 |
| Can I install Assign™ ATF 1.5 on a network drive, rather than on individual PCs? | | 57 |
| Analysis FAQs | | 57 |
| All sequences from a sample are not appearing together as part of the same active sample? | | 57 |
| The sequence data has cut off base calling but I can still see good quality sequence? | | 57 |
| How do I undo a mistakenly trimmed sequence? | | 57 |
| What Keyboard Shortcuts (Hotkeys) are available in Assign™ ATF 1.5? | | 58 |
| The EPG indicates the presence of an indel but this has not been reported in the report. | | 59 |
| Newly loaded sequences shown as already reviewed. | | 59 |
| The EPG contains the correct base call at a variant site but this has been changed in the consensus sequence. | | 59 |
| Appendix 2: Why Should I Use Reference Manager? | | 60 |
| Assign™ ATF 1.5 Analysis WITHOUT using Reference Manager | | 60 |
| Assign™ ATF 1.5 Analysis WITH using Reference Manager | | 60 |
| Appendix 3: Manual Editing of the Reference Sequence | | 61 |
| Appendix 4: Analysis Settings and Sequencing Applications | | 63 |
| Appendix 5: Heterozygous Insertion/Deletion and Poor Quality Sequence Warnings | | 64 |
| Recommended Approaches for Analysing Heterozygous Indel Sequences | | 65 |
| Support and Contact Details | | 67 |

About Assign™ ATF 1.5

Assign™ ATF 1.5 is a sophisticated, automated software program that is used for the rapid, high throughput, quality controlled analysis of DNA sequence electropherograms (EPGs) from automated DNA sequencers. It has been developed by laboratory scientists and expert computer programmers with extensive experience in DNA sequencing.

Assign™ ATF 1.5 can be used for an extensive range of sequencing applications as well as producing quality control information in a unique and informative manner.

1. Compatibility

1.1 Computer Operating Systems

- 1.1.1 Assign™ ATF 1.5 is a Windows® based program that will run on Windows® XP, Windows® 7 and Windows® 8 operating systems. Performance cannot be guaranteed for Windows® Vista.
- 1.1.2 Microsoft® Excel® 97 or above is required for the creation of reports.
- 1.1.3 Installations on 32 or 64 bit Intel (or compatible) hardware are also supported, with the exception of virtualised and server environments. Conexio Genomics can provide limited support for installations where the program and/or data files are stored on a network.

Note: Assign™ software has not been designed to be compatible with RT, Embedded, Starter, Mobile, and Phone editions of Windows as well as any hardware which do not support a standard keyboard, mouse and screen.

1.2 Data Files Supported

- 1.2.1 Assign™ ATF 1.5 requires **.ab1**, **.abd**, or **.scf** sequence files from automated DNA sequencers or **.fasta** sequences. The **.ab1**, **.abd**, and **.scf** files should be run through the Applied Biosystems™ by Life Technologies™ Sequence Analysis software before they are imported into Assign™ ATF. Automated DNA sequencers from Applied Biosystems™ by Life Technologies™, Beckman and Amersham have been used successfully with Assign™ ATF 1.5.

2. Overview

2.1 Functions and Features

2.1.1 Tailored to an Extensive Range of Sequencing Applications

- 2.1.1.1 Variant/Mutation Detection;
- 2.1.1.2 Quality Control;
- 2.1.1.3 Sequence Alignment;
- 2.1.1.4 Genotyping.

2.1.2 Efficient and Rapid Setup

- 2.1.2.1 Create reference sequences directly from GenBank and be ready for analysis in minutes.
- 2.1.2.2 Improve analysis speed and quality by tailoring analysis according to your PCR strategy, to remove poor quality and unwanted sequences.

2.1.3 User Friendly

Straight-forward user interface and easy to interpret layout.

- 2.1.4 **High Throughput**
 - 2.1.4.1 Import thousands of sequences, from multiple genes into a single project for high throughput analysis.
 - 2.1.4.2 Display electropherogram data or consensus text sequences as mismatches with the reference sequence
- 2.1.5 **Accurate Base Calling**

Assign™ ATF 1.5 has a unique base caller, optimized for heterozygous base calling even in the presence of poor quality data. Assign™ ATF 1.5 performs a dynamic assessment of background noise and compensates for this in order to perform accurate base calling.
- 2.1.6 **Highly Sensitive and Accurate Mutation Detection**
 - 2.1.6.1 Assign™ ATF 1.5 uses proprietary base calling and alignment algorithms. The base caller has been designed specifically for accurate detection of mixed bases/heterozygotes.
 - 2.1.6.2 Assign™ ATF 1.5 includes an optional, patented approach to EPG analysis that normalises the data and enables the quantitative nature of DNA sequencing to be exploited. Nicknamed 'Picket Fence' analysis, this approach further improves heterozygous base calling to deliver accurate detection of low level mutants.
- 2.1.7 **Rapid Analysis - Priority Review and Editing**
 - 2.1.7.1 Removes data analysis as a bottleneck for high throughput sequencing-based applications.
 - 2.1.7.2 Sequence review and editing is streamlined by facilitating priority analysis of positions of importance (e.g. bases with low quality, bases mismatched with the reference sequence, user-defined variant positions, and/or user-edited positions). Assign™ ATF 1.5 navigates you directly to those positions matching your set criteria, substantially accelerating analysis time.
- 2.1.8 **Quality Driven Analysis with a Quality Control Focus**
 - 2.1.8.1 Strong focus on data quality – generates visible quality indicators based on critical quality parameters including peak symmetry, background noise and separation from neighbouring peaks.
 - 2.1.8.2 Allows performance criteria establishment for the acceptance or rejection of base calls, entire EPGs, and/or entire samples.
 - 2.1.8.3 Enables automatic generation of longitudinal quality control reports, allowing for run to run analysis of quality. This facilitates the assessment of the effect of changes (such as reagent changes) on sequence data quality.
 - 2.1.8.4 Different levels of user access are available to allow only selected users to perform final reviews (Quality Control Checks), prior to final release.
- 2.1.9 **Audit Trail**

An analysis audit trail is recorded and can be reported. This allows reference to, and reporting of, which user has performed which steps. This further improves quality control by providing traceability in sequence analysis and reporting.
- 2.1.10 **'Distinctive' Features**
 - 2.1.10.1 Enables analysis of heterozygous insertions and deletions. Calculates the number of inserted or deleted bases for reporting, enabling all sequence variants to be reported.

2.1.10.2 Allows the ability to include overlapping genes and coding sequences in a single project, maximising analysis of the impact of sequence variants.

2.2 Performance Characteristics

2.2.1 **High Throughput:** Assign™ ATF 1.5 has successfully imported over 5,000 sequence EPGs into a single project.

2.2.2 **Base Call Accuracy:** Assign™ ATF 1.5 contains a unique base caller developed specifically for heterozygous base call accuracy.

2.3 Limitations

2.3.1 Sequence base call accuracy is dependent on the quality of sequence data.

2.3.2 A minimum availability of 4GB of memory is recommended for optimal performance of the software.

Quick Guide

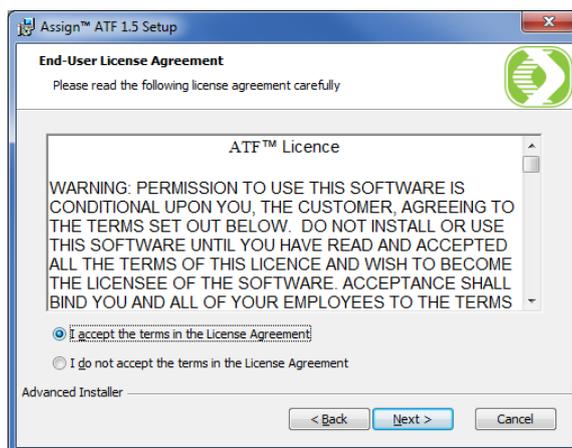
This Quick Guide can be used for basic setup and analysis. If you prefer a more comprehensive guide please skip to the Detailed Guide on page 17. *It is recommended that the Detailed Guide is read to ensure optimal operation and full utilisation of Assign™ ATF 1.5 features.*

3. Brief Description of Assign™ ATF 1.5 Functions - Sufficient Information for BASIC Analysis

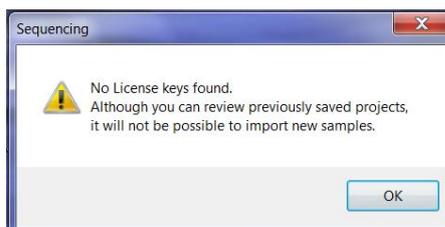
- Installation.
- Logging in.
- Defining analysis settings for automated analysis of data.
- Creating a reference sequence to which sample sequences are compared.
- Defining naming conventions to enable automatic importing and alignment of sequences.
- Importing sequence data.
- Sequence data analysis, editing and confirmation.
- Report generation.
- Saving and opening projects.

3.1 Installation – Full Details in Section 4

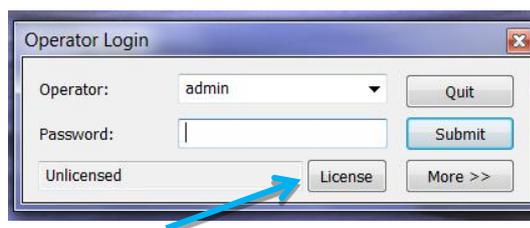
- 3.1.1 Download the installer from the link provided and save the installer to your hard drive.
- 3.1.2 Right click on .zip file and **Extract all**.
- 3.1.3 Double click on the installer file icon and follow the instructions for installation.
- 3.1.4 Read and accept the terms in the License Agreement, then click **Next**.



- 3.1.5 Click **Install** to begin the Installation.
- 3.1.6 Once the Installation is complete, click **Finish**.
- 3.1.7 To obtain the licence key, launch the software by double clicking on the **Assign™ ATF 1.5** icon located on the desktop.
- 3.1.8 A warning message alerting the user to no license key will appear, click **OK** to continue.

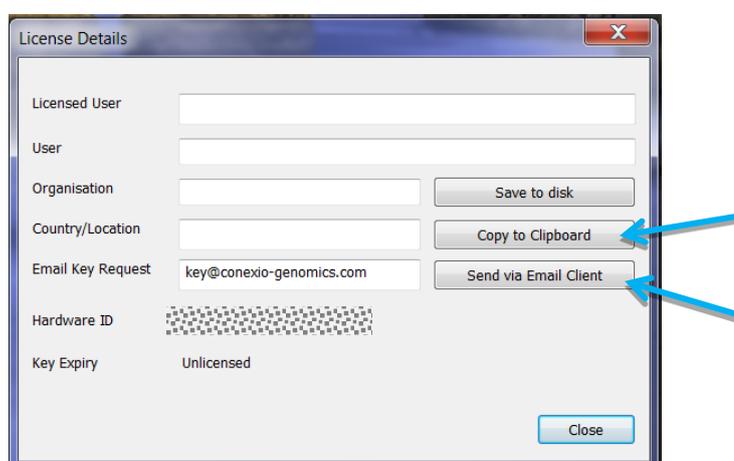


3.1.9 Click **License** in the **Operator login** box.



3.1.10 In the Licensed User box complete all fields. *Note: enter a valid email address in **User**.*

Click **Send via Email Client** to request your licence key file. Alternatively, click **Copy to Clipboard**. Paste into an email and send licence key request to keys@conexio-genomics.com



3.1.11 Once your keys are emailed to you ensure you save them in the same folder as the software (e.g. **C:\ProgramDataConexio Genomics\ATF 150**).

*If you cannot locate the folder double click on the **Assign™ ATF 1.5** desktop shortcut, use the Operator **admin** and enter the default password **cg01** and click **More>>**. The **System File Location** is listed at the bottom of the screen. This is the folder in which you will need to save your key(s).*

Ensure you save only the key file(s) and NOT the entire folder that you were emailed (e.g. FFFFFFFF401602A7) or the keys will not be functional.

3.2 Login – Full Details in Section 5

3.2.1 Launch Assign™ ATF 1.5 by clicking on the **Assign™ ATF 1.5** icon on the desktop.

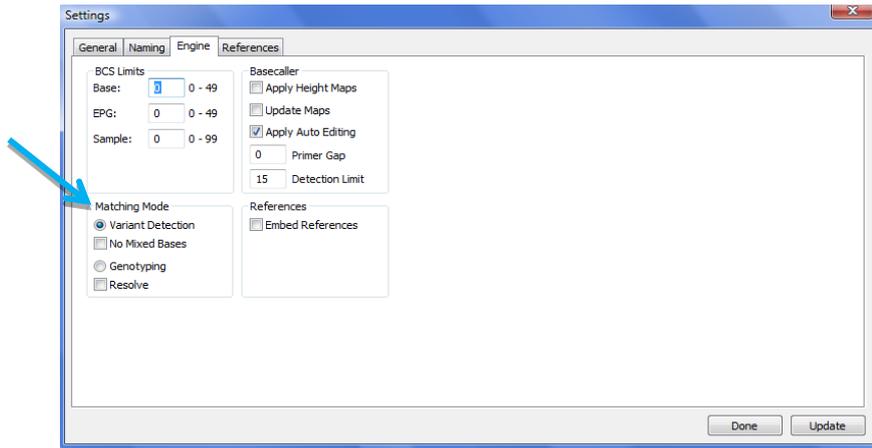
3.2.2 Login using the Operator **admin** and the default password **cg01**.

Additional users can be added with unique passwords and varying levels of access.

3.3 Defining Analysis Settings – Full Details in Section 6.2

Data analysis settings must be defined to ensure that Assign™ ATF 1.5 performs the correct analysis for your application. The most important of which is to select the matching mode.

3.3.1 Go to **Edit > Settings > Engine**.



3.3.2 Select the appropriate Matching Mode:

- Variant Detection** when comparing a sequence with a single reference sequence.
- Genotyping** when comparing a sequence against a reference library where heterozygous sequence is expected.

3.3.3 Ensure **Auto Editing** (*automatic base calling based on historic data*) is **NOT** checked for variant detection.

3.3.4 Click **Update** to save changes.

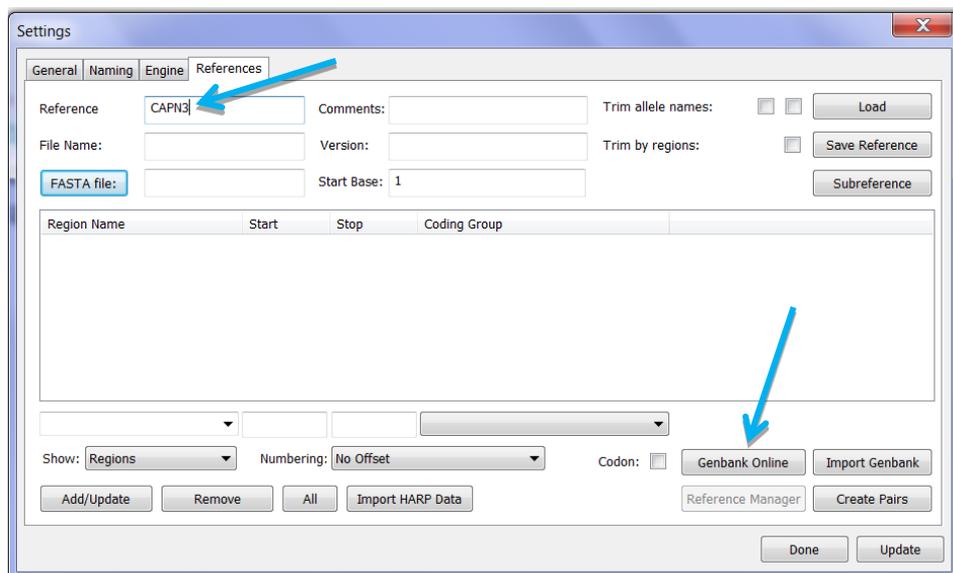
*Other settings available in **Engine** include setting limits to exclude poor quality data (**BCS Limits**), and Base Calling settings such as the use of the Picket Fence algorithm to normalise peak heights for more accurate detection of heterozygotes (**Apply Height Maps/Update Maps**), and the cutoff limit for heterozygous detection (**Detection Limits**).*

3.4 Create Reference Sequence – Full Details - Section 6.3

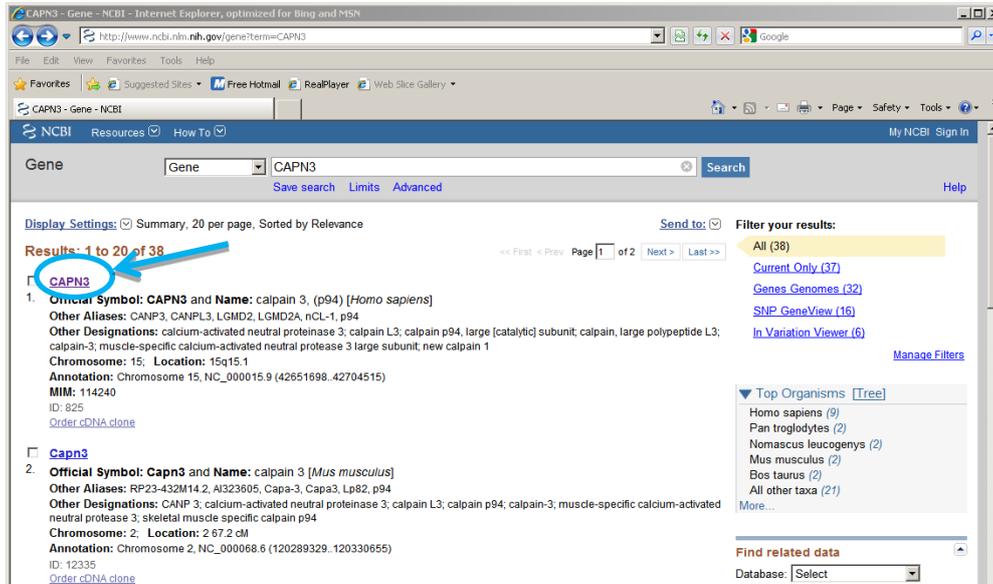
The reference sequence is the sequence to which test sequences are aligned.

3.4.1 Creating a Reference from Genbank File – Full Details in Section 6.3.1

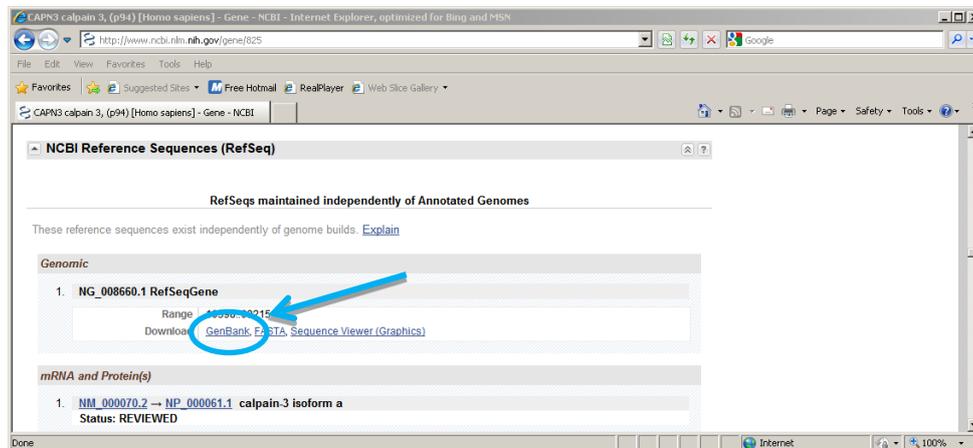
- 3.4.1.1 In Assign ATF 1.5, click **Edit > Settings** and open the **Reference** tab. Enter the appropriate gene name in **Reference** and click **Genbank Online**. The internet browser will navigate to the 'Gene' database at NCBI (<https://ncbi.nlm.nih.gov/gene/>)



3.4.1.2 Select the appropriate entry:

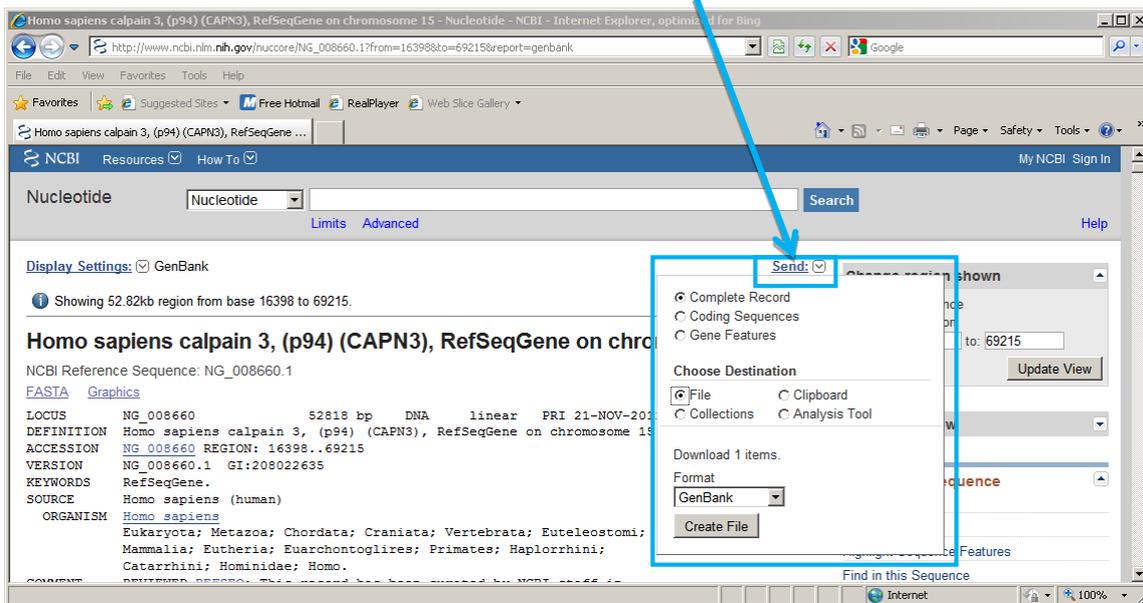


3.4.1.3 Scroll through the page until you reach the NCBI Reference Sequences (RefSeq) section and in the Genomic subsection click GenBank.



3.4.1.4 Confirm that the Genbank entry is the gene you wish to sequence and then perform the following:

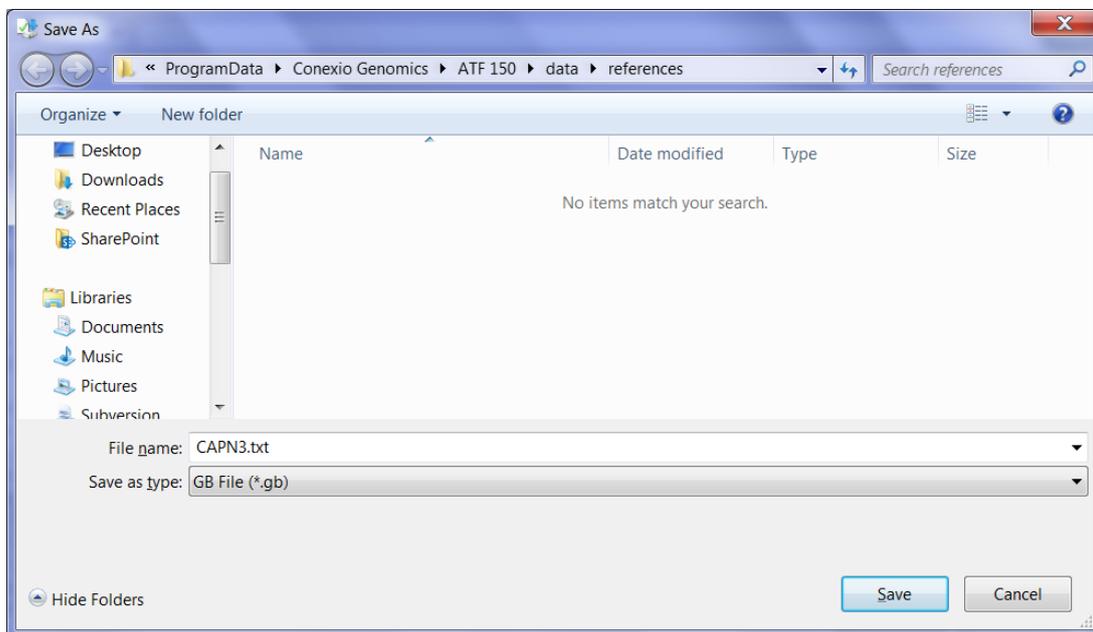
- ◆ Click **Send**.
- ◆ Select **Complete Record**.
- ◆ Select File in **Choose Destination**.
- ◆ Select **GenBank** in the **Format** drop-down menu.
- ◆ Click **Create File**.



3.4.1.5 Your navigator's download window will appear. Select **Save as**.

3.4.1.6 Save the file as either a GenBank (.gb) file or as a text file (text file option allows manual editing if necessary). To save as a text file, navigate to the **ATF 150 > data > references** folder and change the file extension from '.gb' to '.txt' (as shown in the image below). To save as a GenBank file do not change the extension. Also, change the file name from 'sequence' to the name of the gene.

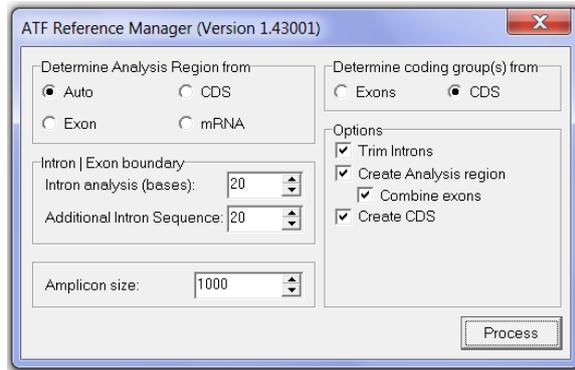
Assign ATF may NOT support file names containing characters other than alpha-numerical.



*If you cannot locate the **references** folder; within the Assign™ ATF 1.5 software, go **Help > Locate Data Directory**. The reference folder is within the **data** folder.*

3.4.1.7 Within Assign™ ATF 1.5 go to **Edit > Settings > References**.

- 3.4.2.1 Click **Reference Manager**. Click **Yes** to the question “Do you wish to save the current reference before launching?” The Reference Manager window will open.



- 3.4.2.2 Select your desired settings. For a straight forward analysis use the default settings and also check **Trim Introns**. Refer to section 6.3.2 for specific details of each setting.

- 3.4.2.3 Click **Process**.

- 3.4.2.4 A Windows Explorer window will open. Select the **.txt** or **.gb** file created from the GenBank entry and click **Open**.

- 3.4.2.5 A window will open with details of the reference and stating that it has been saved. Select **OK**.

Assign™ ATF 1.5 will create an annotated reference sequence based on the information in GenBank.

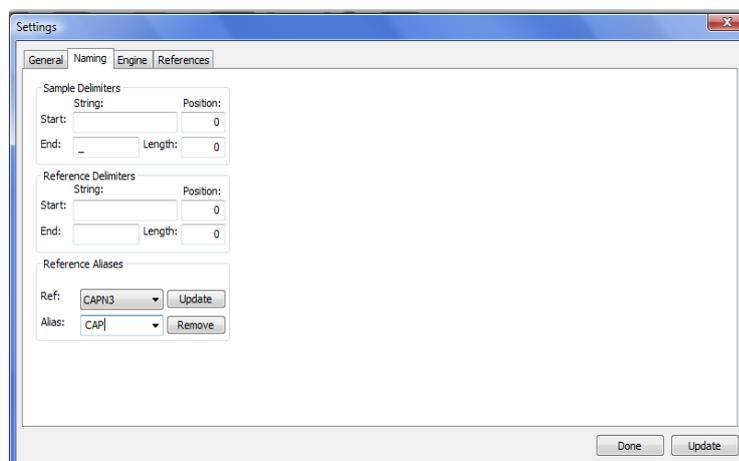
Additional information, such as known variants can be added manually as per section 6.3.3. Additional reference sequence functions and details to manually edit reference sequences are available in sections 6.3.4 and 6.3.5 respectively.

Reference sequences can also be created from FASTA files as per section 6.3.6.

3.5 Entering File Naming Convention – Full Details in Section 6.4

Your sequence file naming convention must be defined to allow automatic grouping of different sequence files for a sample and analysis of the consensus sequence against the appropriate reference sequence.

- 3.5.1 Go to **Edit > Settings > Naming**.



- 3.5.2 Define the location of the Sample ID and Reference within the sequence filename by either:

- Defining how the name relates to delimiters within the filename (if delimiters are routinely used to distinguish the various components of the sequence-file name)

Example: Sequence filename: **A01[12345_CAPN_ex2F**

Sample name starts with [= Enter [in the **String > Start** field

Sample name ends with _ = Enter _ in the **String > End** field.

OR

- Defining the position of the start of the filename and the number of characters used (ONLY if the same number of characters are routinely used to distinguish the various components in the sequence-file name AND the name of each component always starts in the same position)

Sample name starts at position 5 = Enter 5 in the **Position > Start** field

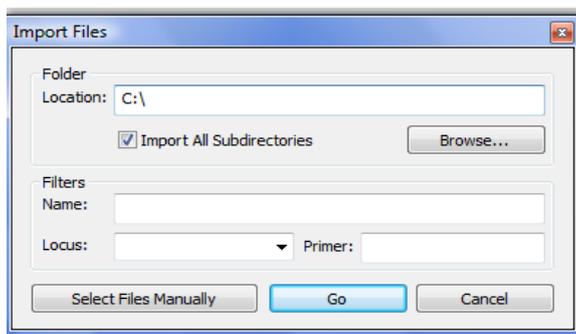
Sample name is 5 characters long = Enter 5 in the **Position > Length** field

- 3.5.3 In **Reference Aliases** select the reference sequence (from the **Ref:** drop down menu) and enter the alias used in the sequence filename that defines the reference (from the **Alias** drop-down menu). *In the example above CAP is used as an alias for CAPN3.*
- 3.5.4 Click **Update** to save changes (directly to the right of the **Ref** field).
- 3.5.5 Repeat this for each reference alias. Click on the **Update** button in the lower right hand corner of the **Settings** window to save the changes.

3.6 Import your Sequences – Full Details in Section 7

*Please take care when commencing your work to ensure that your desired settings file is loaded in **Edit > Settings > General** (refer to section 6.1.1). This will not be relevant if you only have a single 'default' settings file.*

- 3.6.1 Logging into Assign™ ATF 1.5 automatically opens a new project. Multiple projects can be created by selecting **File > New**.
- 3.6.2 To open an existing project select **File > Open** and navigate to the location of the required project to open it.
- 3.6.3 Go to **File > Import > Electropherograms** to import sequence data.



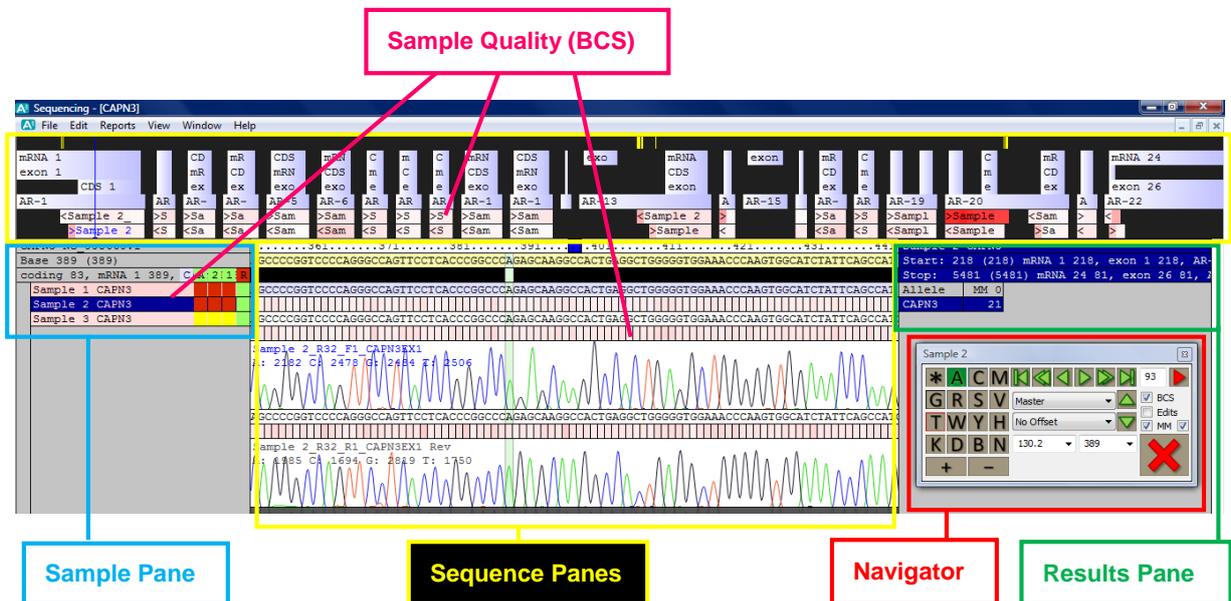
- 3.6.4 To import all of the sequences in a directory click **Browse**, navigate to the directory and select **Go**.
- 3.6.5 To import individual or specific files click **Select Files Manually**, navigate to the files and click **Open**.
- 3.6.6 If asked to select a reference, select the relevant reference from the drop down menu. Check **Use for Subsequent Samples** if all samples are to be compared to the same reference.

If you accurately set up a reference alias in the naming convention (section 3.5.3-Quick Guide or 6.4.3-Detailed Guide) the reference will be automatically selected and you will not need to perform this step.

Text files in FASTA format can also be imported.

3.7 Sequence Analysis and Editing – Full Details in Section 8

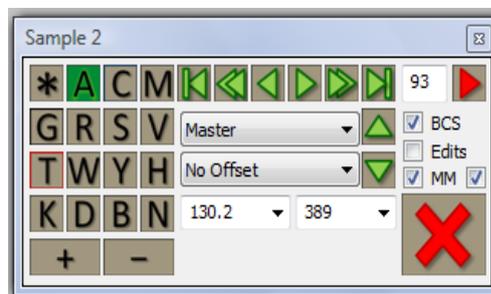
Once the sequence data has been imported, the software screen becomes populated with sample, sequence and result information. Refer to section 8 for a detailed explanation of the screen layout.



The screen shows the gene and reference structures, sample ID, the EPG data, and the aligned consensus sequences for a sample against the reference.

Assign™ ATF 1.5 contains a quality scoring algorithm (Base Call Score, BCS) that assesses the data quality. Throughout the various sections of the screen white to red shading is used to indicate sequence data quality for each base, EPG, and sample (red = poor quality). This enables easy identification of regions of poor quality where base call errors are more likely.

The **Navigator** enables sequence editing, moving between samples and moving between positions within a sequence. It is used to confirm base calls at user-defined positions such as where there is poor quality (BCS), where user edits have been made, where there are mismatches between the sample and the reference sequence(s), and/or where there are variant positions. The user can define which of these positions they wish to easily locate and use the **Navigator** arrows to skip to these positions. This allows 'priority editing' and eliminates the need to review all positions in the sequence.



3.7.1 Selecting either  (single arrow) button moves one position left or right.

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

- 3.7.3 Selecting the up or down arrow moves to the sample above or below in the **Sample Pane**.
- 3.7.4 Perform Priority Editing and Base Call Confirmation using the **Navigator** (refer to section 8.7 for more details). The base calls included in the priority edits can be selected using the **BCS/Edits/MM** boxes to choose positions with a low quality score, positions mismatched with the reference sequence; user defined variant positions and edited positions.
- 3.7.5 Click either   button to navigate to the next priority position.
- 3.7.6 Confirm a base call by selecting the **X** button. Base call changes can be made using the key pad on the **Navigator**. Continue until all base calls are confirmed and a  remains. The 'A' column in the **Sample ID** pane will also turn green to indicate that analysis is complete.
- 3.7.7 Additional editing functions are available by right clicking with the cursor on the EPG. EPGs can be trimmed (**Set Start Base/Set End Base**), de-activated and removed from the analysis, or removed from the project.

Checking the BCS box on the Navigator ensures positions of low quality are assessed.

Base call editing changes the consensus sequence only. Mismatches compared to the reference sequence are indicated in the results pane to the right of the EPGs.

*Assign™ ATF 1.5 allows the review of analysis (quality control checks). Two levels of review are possible (review '1' and '2' as displayed in the **Sample Pane**). Only authorised users may perform final reviews. Refer to section 8.8 for details.*

Refer to section 8 for a comprehensive description of all analysis features and functions.

3.8 Producing a Report – Full Details in Section 9

Once base calls have been confirmed, Assign™ ATF 1.5 enables comprehensive assessment of the sequence data in the form of reports. The standard report formats are:

- **Variants Report:** *Used for applications where test sequence is compared with a single reference sequence.*
- **Genotype Report:** *Used for genotyping applications when matching a sample sequence against a library of known sequences.*
- **FASTA Report:** *Provides a FASTA file of sequences from all samples in the Assign™ ATF 1.5 project.*
- **Quality Report:** *(BCS Report) Enables a quality control analysis of samples within the Assign™ ATF 1.5 project or for all projects within a specific directory.*

- 3.8.1 Reports are created in **Reports > Report Generator**.
- 3.8.2 Select the appropriate report format.
- 3.8.3 Tailor the report to your requirements by selecting the appropriate report functions and click **Generate Report**.

3.9 Saving and Opening Projects – Full Details in Section 10

- 3.9.1 Save by going to **File > Save** or **File > Save As**, as appropriate.
 - 3.9.2 Saved projects must be opened by **File > Open** and browsing to the saved project.
- Assign™ ATF 1.5 saves the project information including edits and links to EPGs as an .xml file. Assign™ ATF 1.5 projects cannot be opened by clicking on the project xml file. Projects (including EPGs) can also be printed.*

Detailed Guide - Getting Started and Using the Software

4. Installation

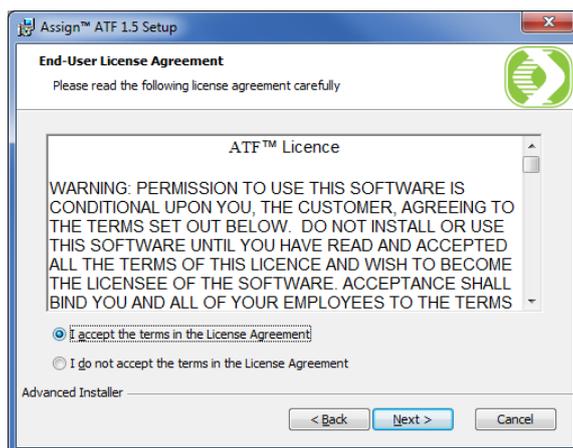
Assign™ ATF 1.5 is a standalone computer software program that should be installed on the computer on which sequence analysis is to be performed. It is recommended that Assign™ ATF 1.5 is installed by a user with complete administrator access to the computer. It is also helpful if the computer has access to the internet to facilitate the system updates and other files as needed.

The installer package can be acquired by contacting Conexio Genomics Pty Ltd at support@conexio-genomics.com

If you plan to use Assign™ ATF 1.5 on multiple computers within your organisation, it is possible to deploy the software 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 further details.

4.1 To Install

- 4.1.1 Download the installer from the link provided and save the installer to a location of your choice on your hard drive.
- 4.1.2 Right click on .zip file and **Extract all**.
- 4.1.3 Double click on the installer file icon and follow the instructions for installation.
- 4.1.4 Read and accept the terms in the License Agreement, then click **Next**.



- 4.1.5 Click **Install** to begin the Installation.
- 4.1.6 Once the Installation is complete, click **Finish**.

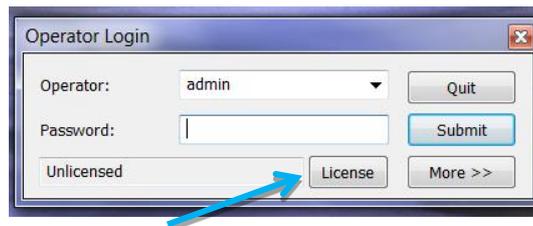
New data cannot be analysed in Assign™ ATF without a licence key file. The licence key files are specific for the computer hardware ID and are available after installation:

- 4.1.7 To obtain the licence key, launch the software by double clicking on the **Assign™ ATF 1.5** icon located on the desktop.

- 4.1.8 A warning message alerting the user to 'no license key' will appear, click **OK** to continue.

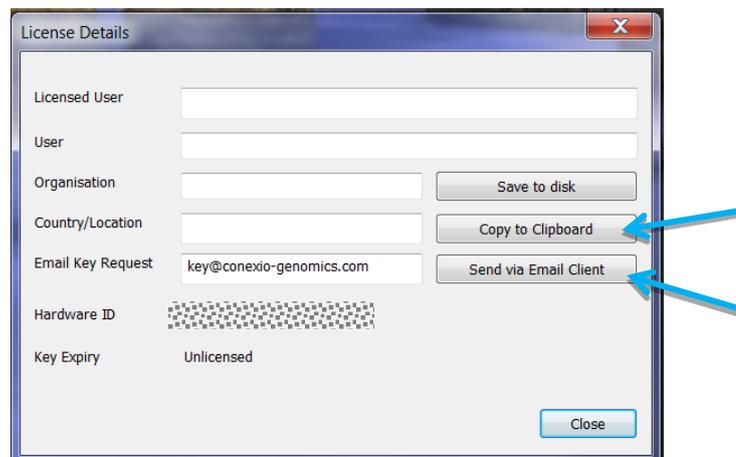


- 4.1.9 Click **Licence** in the **Operator login** box.



- 4.1.10 In the Licensed User box complete all fields. *Note: enter a valid email address in **User**.*

Click **Send via Email Client** to request your licence key file. Alternatively, click **Copy to Clipboard**. Paste into an email and send licence key request to keys@conexio-genomics.com



- 4.1.11 Once your keys are emailed to you ensure you save them in the same folder as the software (e.g. **C:\ProgramDataConexio Genomics\ATF 150**).

*If you cannot locate the folder double click on the **Assign™ ATF** desktop shortcut, use the Operator **admin** and enter the default password **cg01** and click **More>>**. The **System File Location** is listed at the bottom of the screen. This is the folder in which you will need to save your key(s).*

*Ensure you save only the key file(s) and NOT the entire folder that you were emailed (e.g. **FFFFFFFF401602A7**) or the keys will not be functional.*

5. Login and User Management

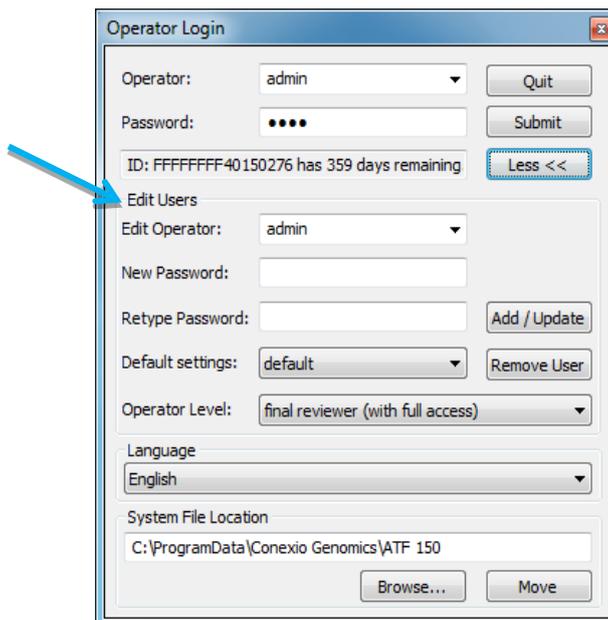
5.1 Login

5.1.1 Ensure you have obtained and saved your licence key. Launch the software by double clicking on the **Assign™ ATF 1.5** icon located on the desktop.

5.1.2 The default Operator is **admin** and the default current password is **cg01**.

5.2 Adding Users

5.2.1 Enter the admin login and password and click **More>>**. A section will open in which you can add additional users.



5.2.2 Type the new user's name in the **Edit Operator** section. Type a password for the new user in **New Password** and retype it in **Retype Password**.

5.2.3 Generally the **Default settings** will be left as 'default'.

5.2.4 Select the **Operator Level**. This refers to the user's level of access based on authority. Restricting access ensures reports are not created without the appropriate level of approval. The user levels include:

◆ **First reviewer (edit only):**

- Can NOT change settings.
- Can edit sequences that have NOT been signed off by a final reviewer.
- Can sign on or off first check box.
- Can NOT sign on or off second (final) check box.
- Can NOT edit sequences that have been signed off by a final reviewer.

◆ **First reviewer (with access to settings):**

- Can change settings.
- Can edit sequences that have NOT been signed off by a final reviewer.
- Can sign on or off first check box.
- Can NOT sign on or off second (final) check box.
- Can NOT edit sequences that have been signed off by a final reviewer.

◆ **Final reviewer (with full access):**

- Can change settings.
- Can edit sequences that have NOT been signed off by a final reviewer.
- Can sign on or off first check box.
- Can sign on or off second (final) check box.
- Can NOT edit sequences that have been signed off by a final reviewer (unless the reviewer clicks to sign off the second review).

Note: Signing off means an editor is satisfied with a result. If a sample is signed off by a “Final Reviewer” it can no longer be edited (unless a “Final Reviewer” signs off the second review). All changes in status are recorded.

5.2.5 Click **Add/Update** (directly next to the **Retype Password** field).

5.2.6 Repeat for additional users.

5.2.7 Once the users have been added they can login by selecting their name from the **Operator** dropdown, typing their password and clicking **Submit**.

5.3 Changing a Password

5.3.1 Login using the user name and password of the account you wish to change and click **More>>**. A section will open in which you can change the password.

5.3.2 Type the new password in **New Password** and retype it in **Retype Password** to confirm.

5.3.3 Click **Add/Update** (directly next to the **Retype Password** field).

5.4 System File Location

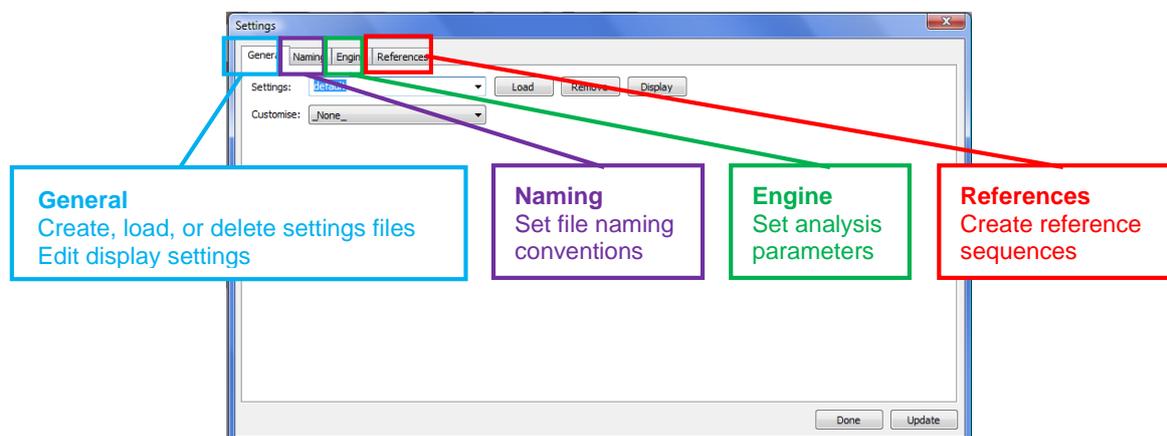
5.4.1 If the location of the system files is moved for any reason, ensure the location is updated in this **Operator Login** screen. If this is not updated Assign™ ATF 1.5 will not be functional.

6. Setting up Assign™ ATF 1.5 for Analysis

Assign™ ATF 1.5 enables the user to configure the sequence analysis parameters for their requirements. Either the “default” settings can be edited or individualised settings can be saved as “settings” files for different applications or different users.

Note that customising the settings is an important step. For example it will not be possible to import any sequence data until references are created.

Open the **Settings** window by selecting the **Edit** drop down menu and clicking **Settings**. The following functions are available:



6.1 General Settings Tab

Customise is a specialised setting and is to be used under the guidance of Conexio Genomics.

6.1.1 Working With “Default” or Customised Setting Files

Unique settings can be created for users or projects. NOTE: It is common for most labs for ALL analyses and ALL users to use the same standard settings.

Creating new settings files is performed in the **General** tab:

6.1.1.1 To edit “default” settings, leave “default” in the **Settings** drop down menu and then perform the appropriate edits as described in the following sections.

6.1.1.2 To create a new setting file, type the new name in **Settings** and then perform the appropriate edits as described in the following sections. Ensure you click **Update** after any changes are made.

- Example 1: Create a file according to the gene being analysed with the required settings and simply load this file every time you want to perform this application (rather than having to change the settings each time).
- Example 2: Create a settings file called “John” if this user routinely uses different settings.

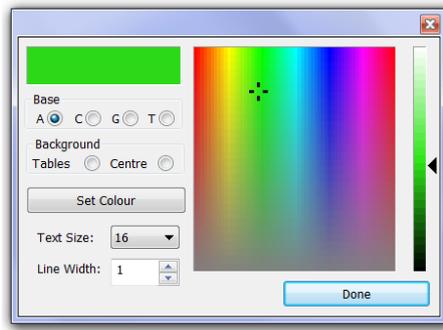
6.1.1.3 Once settings files have been created they can be opened by selecting them in the main **Settings** drop down menu and clicking **Load**.

If you create new settings files, please take care when commencing work on a project to ensure that the correct settings file is loaded.

6.1.2 Editing Display Settings

Modifications can be made to the interface such as changes to font, EPG colours and line thickness.

6.1.2.1 Within the **General** tab click on **Display** to open the Display window.

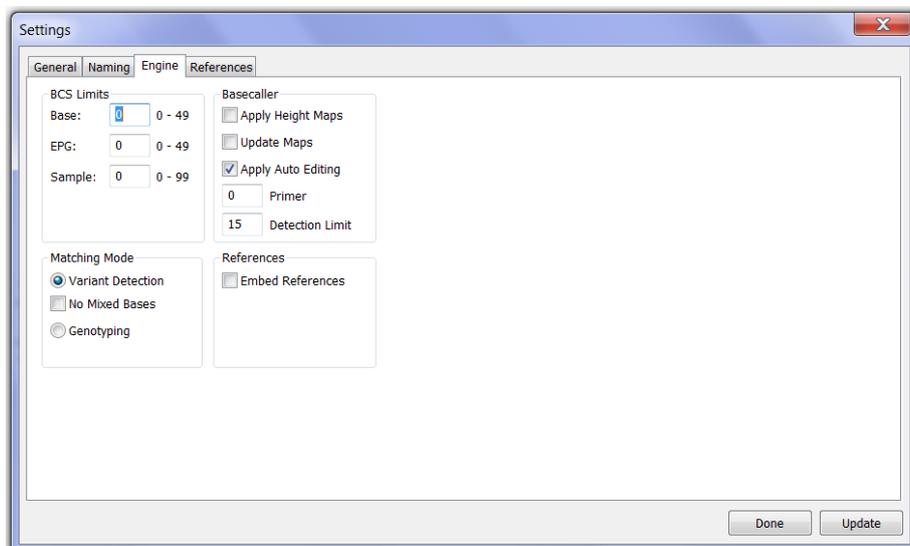


6.1.2.2 Adjust the base colours, background colours, text size, and line width (EPG tracing width). Click **Set Colour** for each colour change to save. When complete, click **Done**.

6.1.2.3 In the main **Settings** window ensure you click **Update** to save any changes.

6.2 Data Analysis Settings (Engine Tab)

*Important analysis settings should be selected to optimise sequence analysis based on the application that Assign™ ATF 1.5 is being used for. This is performed in the **Engine** tab. These settings include setting limits to exclude poor quality data (**BCS Limits**), the matching mode (either **Variant Detection** or **Genotyping**), and Basecalling settings such as the use of the Picket Fence algorithm to normalise peak heights for more accurate detection of heterozygotes (**Apply Height Maps/Update Maps**), automatic base calling based on historic data (**Auto Editing**), and the cutoff limit for heterozygous detection (**Detection Limits**).*



6.2.1 Base Call Score (BCS) Limits

The Base Call Score is the basic unit of Assign™ ATF 1.5's quality assessment system. The BCS reflects the integrity of the peak shape, the background and the separation from neighbouring peaks. The higher the number the better the sequence quality and the more confidence that a given base call is correct. A perfect peak will have a BCS of 50.

The BCS of a consensus sequence is calculated from the BCS of individual sequences that constitute the consensus sequence. The maximum BCS for a consensus sequence is 100. The BCS does not discriminate against heterozygous base calls and as a result the mean BCS at positions across a region provides a

quantitative quality score for the sequence region. This principle can be able to determine a quality score for a sample. The Quality Reports enable a comprehensive application of BCS for quality control purposes.

The following BCS are calculated:

- **Base BCS** is the quality assessment score for a single sequence peak. This will be a BCS between 0-50 with a perfect peak having a BCS of 50.
- **EPG BCS** is the quality assessment score for the EPG. This is a mean BCS for all positions within an EPG. This will be a BCS between 0-50.
- **Sample BCS** is the quality assessment score for the sample. If there is more than one EPG covering a position (e.g. a sample is sequenced in both directions) it is based on the consensus sequence. This will be a BCS between 0-100.

The **BCS Limits** section filters base calling depending on quality. Positions within a sequence, an EPG, or a sample will not be analysed unless they have a value above that entered in the **BCS Limits** fields.

Note: the default values of 0 are sufficient to enable analysis and are the recommended settings. Entering BCS limits speeds up the analysis of poor quality data.

6.2.1.1 Enter the desired number in to each field. Using the default of 0 will result in all data being included.

- Setting a **Base** limit value will result in all bases with a score lower than this value not being called and will be assigned a *
- Setting an **EPG** limit value will result in the exclusion of an EPG if the mean BCS of all positions falls below the value used.
- Setting a **Sample** limit will result in the exclusion of a sample if the mean BCS of the sample falls below the value used.

6.2.1.2 After entering the appropriate values click **Update** to save the changes.

6.2.2 Matching Mode

Two of the main functions of Assign™ ATF 1.5 are **Variant Detection** (e.g. BRCA testing) and **Genotyping** (e.g. HepC genotyping or highly polymorphic genetic systems).

Specifying which of these two applications you are using will optimise the analysis.

6.2.2.1 **Variant Detection:** Select this when identifying sequence differences between a test sequence and a single reference sequence.

6.2.2.2 **Genotyping:** Select this when comparing the test sequence against a library of sequences of alleles to determine to which alleles or combination of alleles the test sequence is best matched.

6.2.2.3 **No Mixed Bases:** This ensures the base call of a single peak – either A, C, G or T (rather than heterozygote). This is useful for base calling poor quality hemizygous data.

6.2.2.4 After making your desired selections click **Update** to save the changes.

6.2.3 Basecaller

Assign™ ATF 1.5 has a unique base calling function. The **Basecaller** section enables activation of the **Picket Fences** algorithm and **Auto Editing**.

Picket Fences is a normalisation algorithm that improves base call accuracy for re-sequencing projects. Ideally, homozygous peak heights would be the same height as each other and heterozygous peaks would be 50% of homozygous peak heights. However, this is not the case due to the variable incorporation rates of di-deoxynucleotide nucleotides.

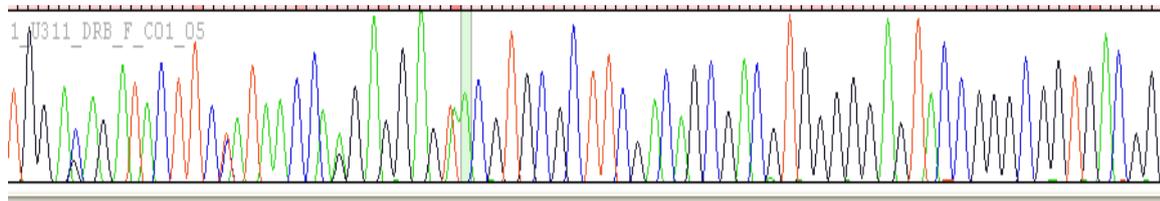
Despite the variable di-deoxynucleotide incorporation rates between positions within a sequence, the incorporation rate at any one position within a sequence is highly reproducible between different samples. As a result, a homozygous base at any particular position within a sequence has a predicted peak height. Picket Fence analysis presents the sequence peak heights of an EPG relative to the expected homozygous peak height. As a result, homozygous peaks are usually the same height and heterozygous peak heights are 50% of homozygous peaks.

Base calling is then performed on this data resulting in an increase in heterozygous base calling accuracy.

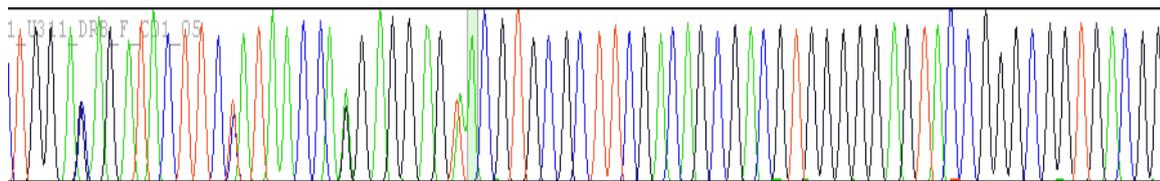
Auto Editing is an intuitive base call algorithm that is applied when the quality of a sequence peak is poor. Assign™ ATF 1.5 uses prior base calling information at this position as a guide to the most likely base.

Auto Editing should **NOT** be used for variant detection, SNP discovery applications, or for other applications where the sample sequence is compared with a single sequence.

Conventional EPG Analysis



Picket Fence Analysis



Suggested Applications for Picket Fence Analysis:

- High throughput genotyping on optimized data.
- Comparing SNP frequencies on pooled DNA.
- Accurate detection of low level mutations.
- Quality Control of reagents - ensuring equivalent amplification of alleles.
- Genotyping of alleles defined by insertion/deletion polymorphisms.

Suggested Applications for NON-Picket Fence Analysis:

- High throughput SNP screening on non-optimized data, or data of variable quality.
- Non re-sequencing applications.
- Contig assembly from cloned data.

6.2.3.1 **Apply Height Maps:** Checking this box turns on **Picked Fence** analysis.

6.2.3.2 **Update Maps:** Picket Fence analysis uses normalisation maps taken from sequence data. Checking **Update Maps** allows Assign™ ATF 1.5 to use the data from imported samples to update the normalisation maps and thereby improve the Picket Fence function.

6.2.3.3 **Apply Auto Editing:** Checking this box instructs the software to perform Auto Editing. *Auto Editing should **NOT** be used for variant detection, SNP discovery applications, or for other applications where the sample sequence is compared with a single sequence.*

6.2.3.4 **Detection Limits:** Set the cut off limit for heterozygous detection. A limit between 11 - 15 is recommended for ideal sensitivity and specificity.

*Note that **Primer** is a HLA specific application and is not commonly used. Please contact Conexio Genomics if you require further information.*

Suggested Applications for Auto Editing:

- Sequence based genotyping when comparing an unknown sequence against a sequence library.

6.2.3.5 After making your desired selections click **Update** to save the changes.

*For a summary of **Matching Mode** and **Basecaller** analysis settings to be used for particular sequencing applications please refer to Appendix 4: Analysis Settings and Sequencing Applications.*

6.3 Creating Reference Sequences (Reference Tab)

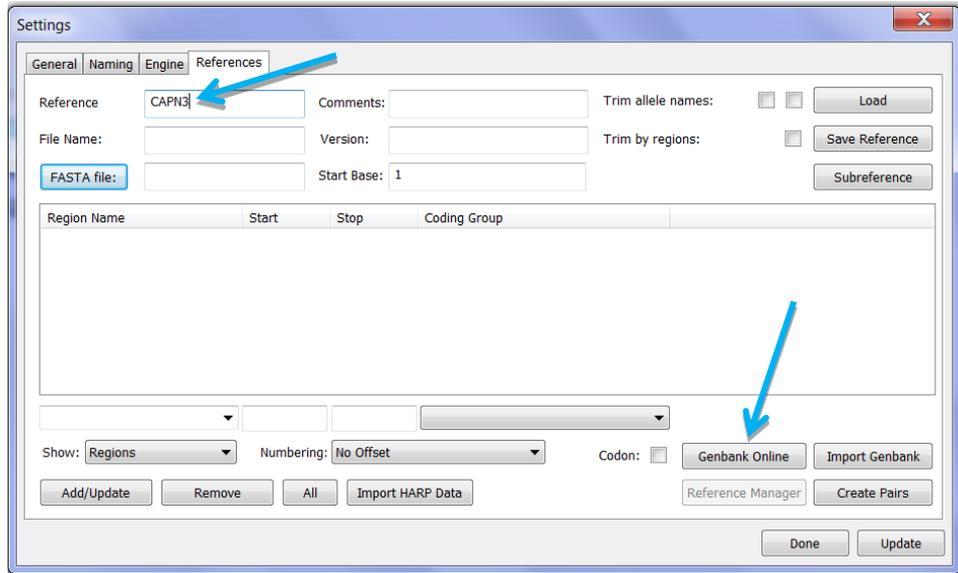
The reference sequence is the sequence to which test sequences are aligned. Reference sequences in Assign™ ATF 1.5 can be made by importing GenBank information or using text sequence in FASTA format. The reference sequence can be annotated with the location of various genetic structures such as exons and un-translated regions etc. Annotated GenBank files will automatically result in an annotated reference sequence in Assign™ ATF 1.5. Reference sequences created from FASTA files require manual annotation.

Reference sequences can be set up and created within minutes if the automated Reference Manager tool is used.

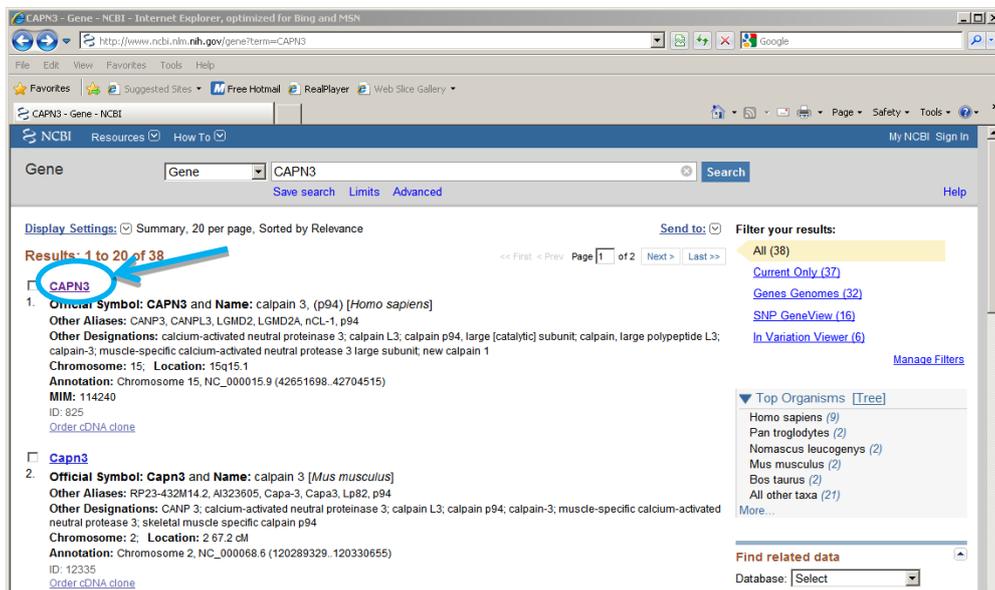
6.3.1 Creating a Reference From a Genbank File

In this example we are creating a reference for the gene **CAPN3** using GenBank.

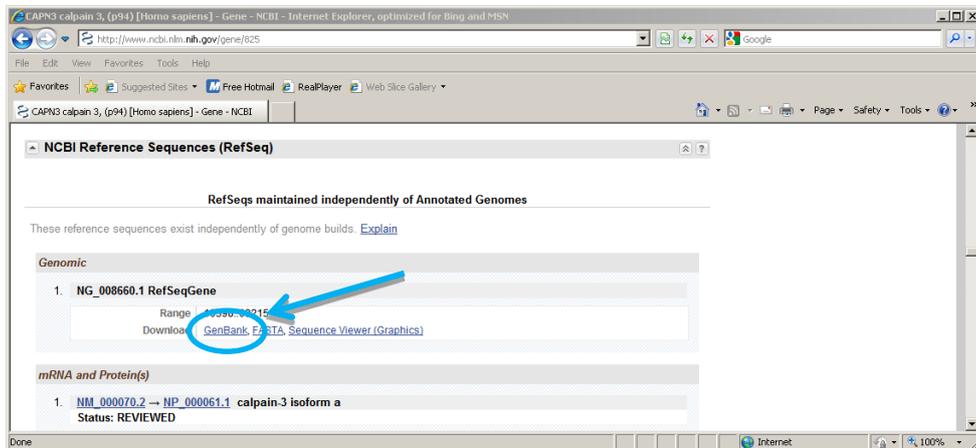
- 6.3.1.1 In Assign™ ATF 1.5, click **Edit > Settings** and open the **Reference** tab. Enter the appropriate gene name in **Reference** and click **Genbank Online**. The internet browser will navigate to the Genbank website.



6.3.1.2 Select the appropriate entry:

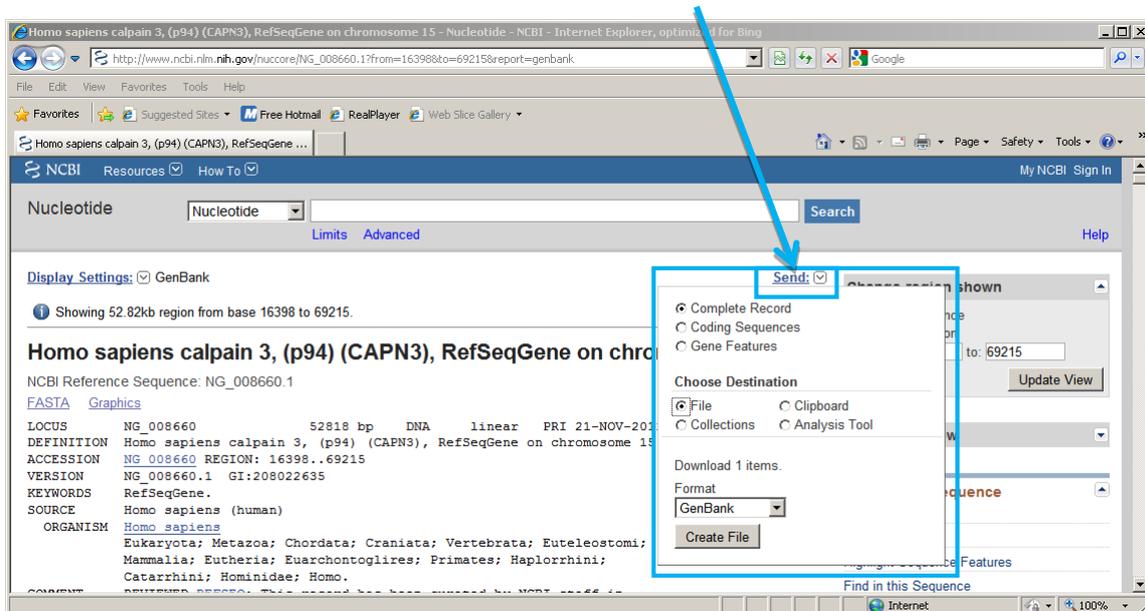


6.3.1.3 Scroll through the page until you reach the **NCBI Reference Sequences (RefSeq)** section and in the **Genomic** subsection click **GenBank**.



6.3.1.4 Confirm that the Genbank entry is the gene you wish to sequence and then perform the following:

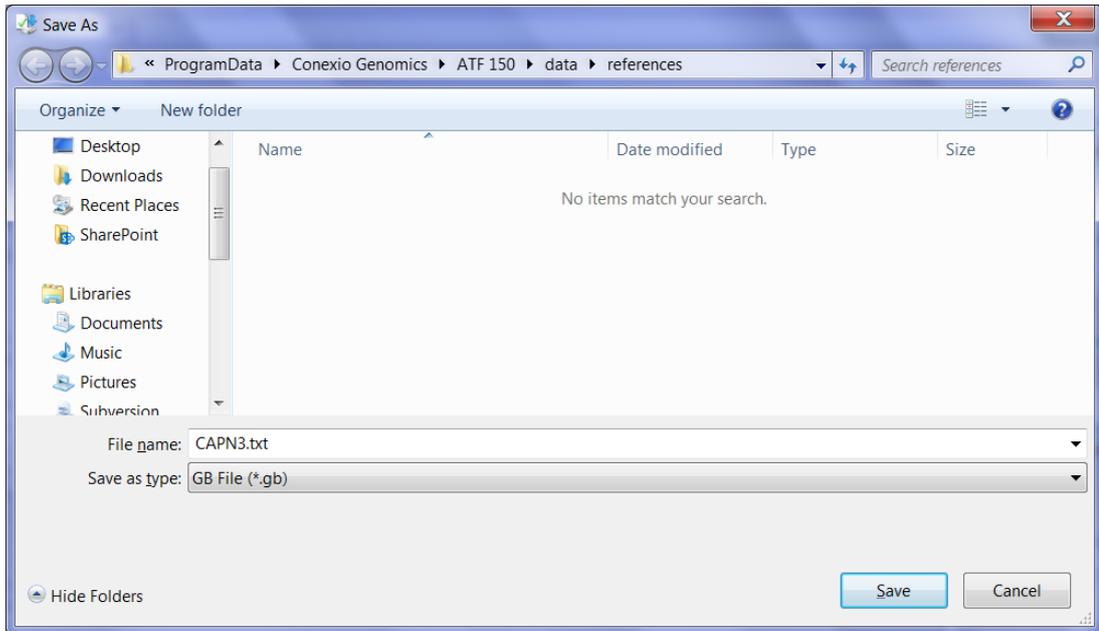
- ◆ Click **Send**.
- ◆ Select **Complete Record**.
- ◆ Select File in **Choose Destination**.
- ◆ Select **GenBank** in the **Format** drop-down menu.
- ◆ Click **Create File**.



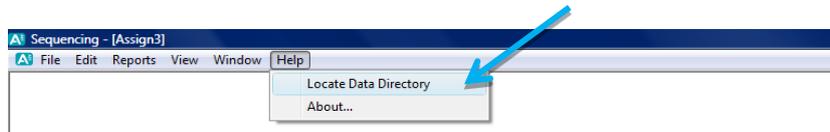
6.3.1.5 Your navigator's download window will appear. Select **Save as**.

6.3.1.6 Save the file as either a GenBank (.gb) file or as a text file (text file option allows manual editing if necessary). To save as a text file, navigate to the **ATF 150 > data > references** folder and change the file extension from '.gb' to '.txt' (as shown in the image below). To save as a GenBank file do not change the extension. Also, change the file name from 'sequence' to the name of the gene.

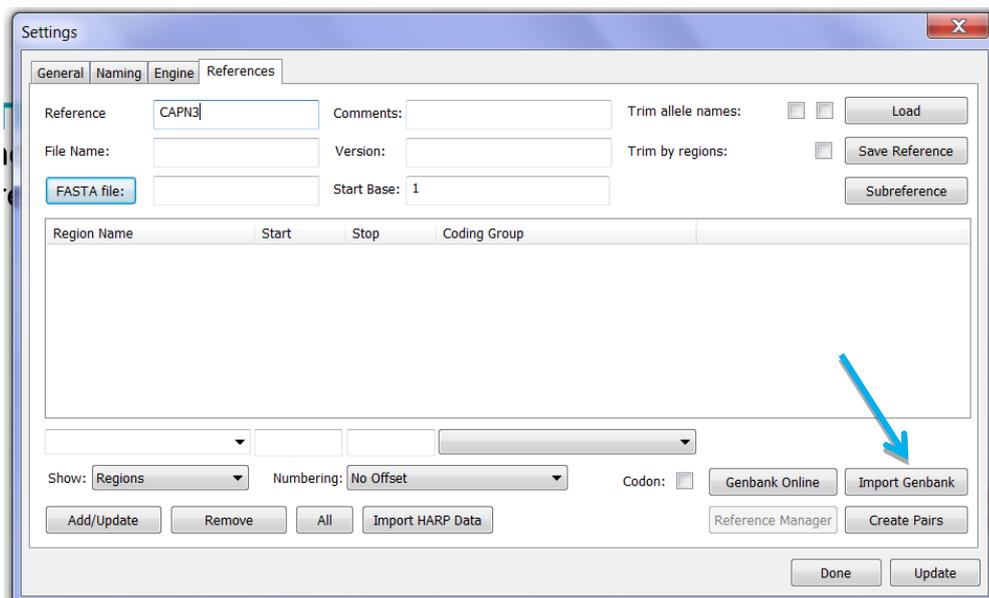
Assign ATF may NOT support file names containing characters other than alpha-numerical. See Limitations 2.3.2.



*The **ATF references** folder is usually located at **C:\ProgramData\Conexio Genomics\ATF 150\data\references**. However, it may be located elsewhere depending on your operating system. To locate the data folder; within the Assign™ ATF 1.5 software, go **Help > Locate Data Directory**. The **reference** folder is within the **data** folder. Save the reference within that reference folder.*



- 6.3.1.7 Once saved, the GenBank file is now ready for use in Assign™ ATF 1.5 for the creation of a reference sequence.
- 6.3.1.8 Within Assign™ ATF 1.5, go to **Edit > Settings > References**.
- 6.3.1.9 Click **Import GenBank** and a Windows Explorer window will open. Browse to the **ATF 150 > data > references** folder, select the GenBank file you just created and click **Open**.

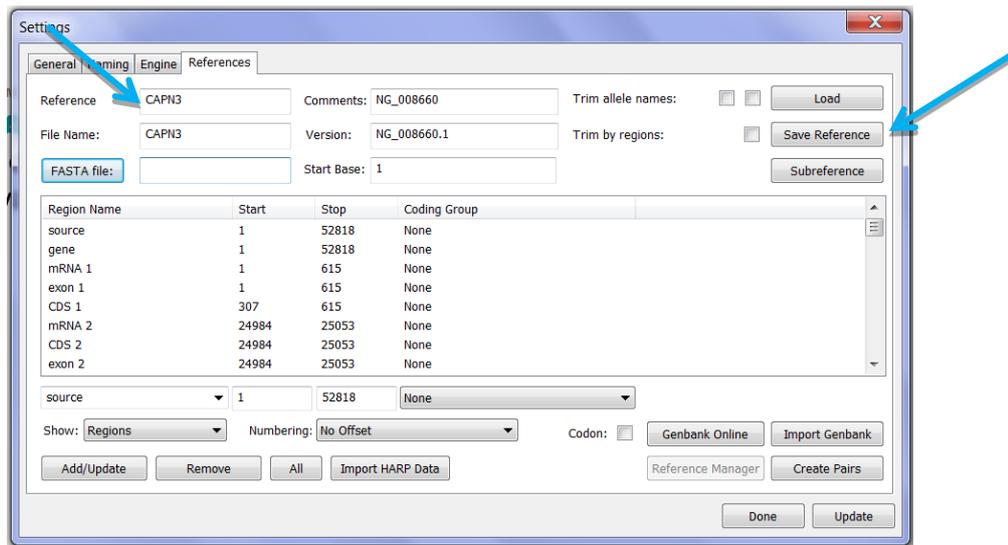


6.3.1.10 The **References** window will be populated with the information contained in the GenBank file.

6.3.1.11 Typically, a user would change the **Reference Name** and **File Name** to reflect the name of the gene.

In the example both the **Reference Name** and **File Name** have been changed from the default to CAPN3.

6.3.1.12 Click **Save Reference**.



6.3.2 Automatic Creation of Coding Group and Analysis Region in Reference Manager

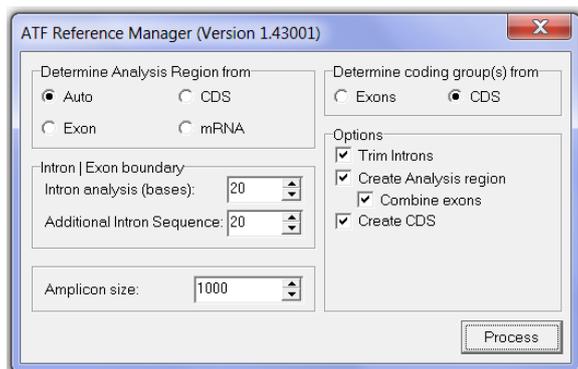
Reference sequences are split into **Regions** and **Groups**. Typically, in a genomic reference the Regions will be exons, CDS and mRNA. Groups consist of a set of Regions. Regions (e.g. CDS) can be concatenated to create a coding **Group**, enabling a sequence change to be interpreted in the context of a codon, or amino acid.

Assign™ ATF 1.5 has a function that enables sequence data to be trimmed at Region boundaries. Therefore, in addition to the functional regions (CDS, exons and mRNA) **Analysis Regions** can be created. Typically Analysis Regions would be set around the regions of the gene that are PCR amplified and sequenced.

The user can set the Analysis Regions relative to an exon. For example, if the testing strategy includes the amplification of an exon plus 20bp of flanking intron, these parameters can be set automatically so that when sequence data is imported, only the exon +20bp of intron is analysed.

References imported from GenBank already contain information about the various **Regions**. The Assign™ ATF 1.5 **Reference Manager** can be used to automatically create a **Coding Group** and **Analysis Regions** that reflect the PCR amplification strategy. For more information refer to Appendix 2: Why Should I Use Reference Manager?

- 6.3.2.1 Click **Reference Manager**. Click **Yes** to the question “Do you wish to save the current reference before launching?” The **Reference Manager** window will open:



Choose your required settings as appropriate:

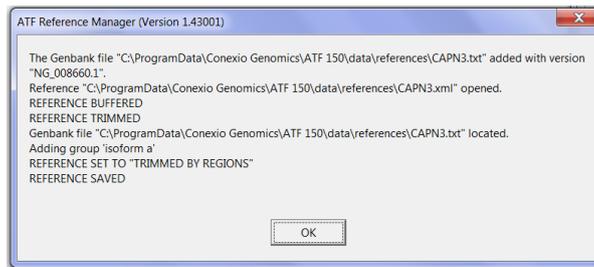
- 6.3.2.2 **Determine Analysis Regions from:** This allows the user to create the analysis region dependent on the PCR amplification strategy; whether it is based on regions defined by CDS, Exons or mRNA.
- Auto** is the default and is required for a typical analysis. **Auto** ensures that an Analysis Region is created at any position where there is a CDS, mRNA or Exon.
- 6.3.2.3 **Intron | Exon Boundary:** This is the amount of intron sequence bordering the Analysis Region (exon, CDS and/or mRNA) that you wish to include in your analysis:
- Intron analysis (bases):** Enter the number of bases adjacent to the Analysis Region that you wish to include in the analysis. The same Buffer Length will be set for all Analysis Regions, but can be changed manually in the References window as described in **Appendix 3: Manual Editing of the Reference Sequence, section A3.1.3.**
 - Additional Intron Sequence:** Enter an amount of sequence, in addition to the intron sequence to be included in the analysis. This sequence is not automatically included in the analysis but enables subsequent inclusion, should the Analysis region be considered too small.
- 6.3.2.4 **Amplicon size:** Enter the PCR fragment size. This enables multiple exons amplified in the same amplicon to be included in the same analysis region.
- 6.3.2.5 **Determine Coding Groups from:** This function enables you to create your **Coding Group(s)** from Exons or the CDS, depending on how the coding sequence is defined in the reference.
- 6.3.2.6 **Trim Introns:** This allows the removal of data from the reference sequence that isn't to be included in the analysis.
- 6.3.2.7 **Create Analysis Regions:** Generally this will be left checked. If unchecked, Analysis Regions will not be created. If the sequencing strategy for your gene is not based around exons and includes non-coding regions, there is no need to create Analysis Regions. In this case, Reference Manager is used only to automatically create coding groups.
- 6.3.2.8 **Combine exons:** If checked, regions will be combined into one Analysis Region if the regions are included in one amplicon. If unchecked, Analysis Regions will be created for all individual region.

6.3.2.9 **Create CDS:** Generally this will be left checked. If checked, it will create coding groups for as many different genes or partial genes that are present within the reference sequence. If unchecked, coding groups will not be created.

6.3.2.10 Once your settings have been selected click **Process** to run the **Reference Manager**.

6.3.2.11 A Windows Explorer window will open. Select the relevant **.txt** or **.gb** file for your gene and click **Open**.

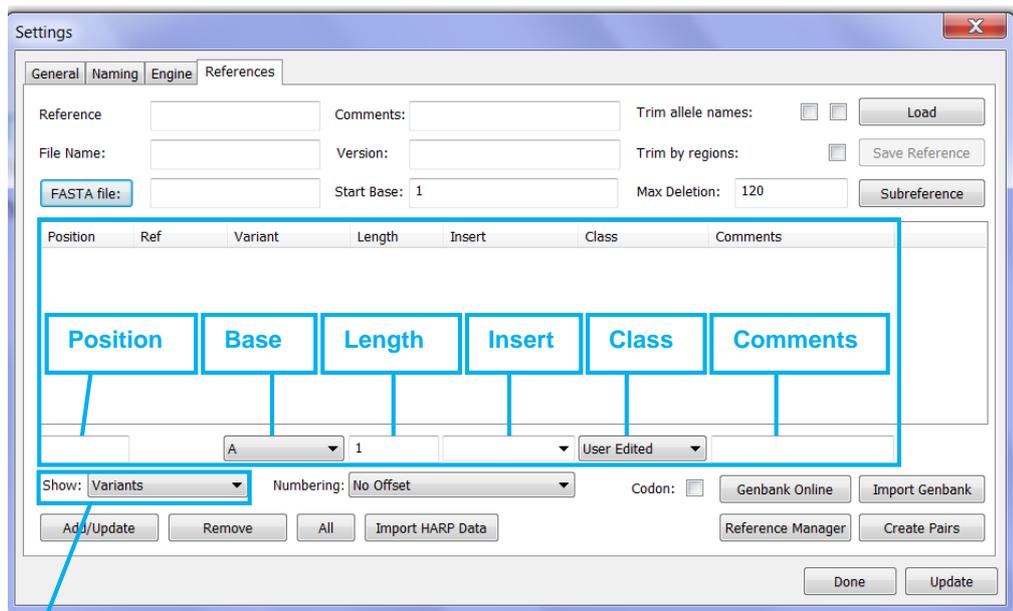
6.3.2.12 A window will open with details of the reference and stating that it has been saved. Select **OK**.



6.3.3 Setting Variants in Reference Sequence

Known variant positions can be entered into the reference sequence. When analysing sequences the user is then able to select variant positions for Priority Editing (refer to section 8.7 for details), enabling quick checking of base calling at these variant positions. Variants are indicated by purple boxes above the sample consensus sequence (refer to section 8.3.5) and can be displayed on the report. Variants can also be added from the analysis screen (refer to section 8.3.6.5).

6.3.3.1 Within the **Edit > Settings >References** window open the reference file you wish to edit by clicking **Load**. Browse to the appropriate reference (**.xml**) file in the **ATF 150 > data > references** folder, select it and click **Open**.



Select **Variants** in **Show** drop down menu

6.3.3.2 Select **Variants** from the **Show** drop down menu.

6.3.3.3 Enter the **Position** (lower left corner) of the variant in the reference sequence.

Further edits that can be made for setting variants in the reference sequence are detailed below. Please note these are optional if applicable to the user but are not required for successful analysis:

6.3.3.4 If required, enter the **Variant** nucleotide from the drop down menu under **Variant** (usually * so any call at that position is flagged).

6.3.3.5 If required, enter the **Length** of the variant in bases (if insertion or deletion variants are >1).

6.3.3.6 If required, in the **Insert** field enter the sequence of bases for a known insertion. Otherwise leave blank.

6.3.3.7 If required, enter the **Class** of the variant (User Edited, Known, Novel, Pathogenic).

6.3.3.8 If required, enter any **Comments**.

6.3.3.9 Click **Add/Update**.

6.3.3.10 Repeat for additional variants as required.

6.3.3.11 Click **Update** (bottom right hand corner) and click **Save Reference** to save changes.

6.3.4 **Additional Reference Sequence Functions**

6.3.4.1 **Subreference** can be used to create a new reference sequence from a region within the current reference sequence. For example, if an exon is nominated as a region within the existing sequence, a subreference of that single exon can be created.

6.3.4.2 **Trim by Regions** only allows analysis of EPG in defined regions. This function is usually applied to Genomic reference sequences where Ref Manager has been used to create analysis regions and also for cDNA references when genomic sequences are amplified.

6.3.5 **Manual Editing of the Reference Sequence**

The Reference Manager is used to automatically create Coding Group and Analysis Regions and results in standard settings for each region. However, Reference information can be edited manually. Refer to Appendix 3: Manual Editing of the Reference Sequence for details.

6.3.6 **Creating a New Reference from Text Files in FASTA Format**

A reference sequence can be created from a single sequence or multiple sequence variants from of the same gene in FASTA format. FASTA format is characterised by a ">" sign followed by the sequence name on the first line and the sequence on the next line.

The following sequence will be used as an example:

```
>Sequence 1
ACGTCGATCAGTACAGCTTTCTGACGATCCAGTTAGGGATCACCCAGACCC.....
>Sequence 2
ACGTCGATCCGTACAGCTTTCTGACGATCCAGTTAGGGATCACCCAGACCC.....etc
```

If you have sequences for multiple variants and you wish to compare test sequences against the sequences of the variants (genotyping), ensure all sequences are in a single file and all sequences in this file are in FASTA format.

- 6.3.6.1 Within **Edit >Settings > References** enter the name of the reference sequence in the **Reference** field.

This is usually descriptive and can contain detailed information about the reference sequence.

- 6.3.6.2 Enter the name of the file that you wish to save this reference sequence in the **File Name** field.

This is usually a short name. The Reference name and the Filename can be the same.

- 6.3.6.3 Click the **FASTA file:** button. This will launch a file search dialogue. Browse to the FASTA file that contains your reference sequence.

- 6.3.6.4 Additional information regarding the reference sequence can be entered in the **Comments** field and the **Version** field can be used to distinguish between multiple versions of a reference sequence or allele library.

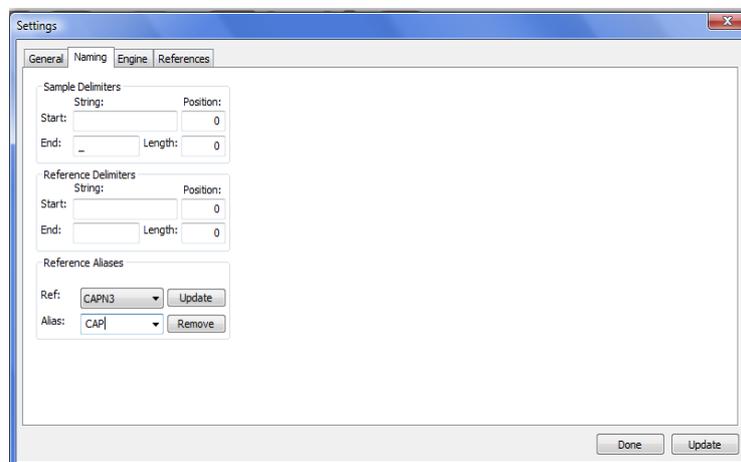
- 6.3.6.5 Once imported Click **Update** (bottom right hand corner) to save.

6.4 Naming Conventions (Naming Tab)

Using a standard sequence file naming convention enables Assign™ ATF 1.5 to link all EPG for a sample and to analyse the test sequence(s) against the appropriate reference sequence.

*In the **Naming** tab the user can enter the parameters that define the routine sample name and the locus identifier in sample sequence filenames. The sequence filename must be unique for a sample and consist of the sample name and an identifier that can be used as an alias for the locus being analysed. Either delimiter symbols or a standard number of characters can be used to define the location of the sample and reference name.*

Assign™ ATF 1.5 may not support file names containing characters other than alpha-numerical.



To assist in explanation, the following example sequence filename will be used for each of the following sections:

Example: Sequence filename: **A01[12345_CAPN_ex2F**

Delimiters have been used to separate the components of the filename:

[Separates the PCR number (**A01**) and the sample name (**12345**);

_ Separates the sample name and the locus (**CAPN**); and

_ Separates the locus and the primer name (**ex2F**).

6.4.1 Defining Sample Delimiters

6.4.1.1 Using delimiters to define the sample name:

If delimiters are routinely used to distinguish the various components of the sequence filename (such as in the example above) you can use the delimiters to define the sample name.

Using the example above:

Sample name starts with [= Enter [in the **String > Start** field
Sample name ends with _ = Enter _ in the **String > End** field.

OR

6.4.1.2 Using character position to define the sample name:

ONLY if the same number of characters is routinely used to distinguish the various components in the sequence filename AND the name of each component always starts in the same position, the position can be used to define the sample name.

Using the example above:

Sample name starts at position 5 = Enter 5 in the **Position > Start** field
Sample name is 5 characters long = Enter 5 in the **Position > Length** field

6.4.1.3 Click on the **Update** button in the lower right hand corner of the **Settings** window.

6.4.2 Defining Reference Delimiters

6.4.2.1 Using delimiters to define the reference name:

If delimiters are routinely used to distinguish the various components of the name (such as in the example above) you can use the delimiters to define the reference name.

Using the example above:

Reference name starts with _ = Enter _ in the **String > Start** field
Reference name ends with _ = Enter _ in the **String > End** field.

OR

6.4.2.2 Using character position to define the reference name:

ONLY if the same number of characters is routinely used to distinguish the various components in the sequence filename AND the name of each component always starts in the same position, the position can be used to define the reference name.

Using the example above:

Reference name starts at position 11 = Enter 11 in the **Position > Start** field
Reference name is 4 characters long = Enter 4 in the **Position > Length** field

6.4.2.3 Click **Update** to save the changes.

6.4.3 Defining Reference Aliases

*For each reference available it is possible to enter a **Reference Alias**. This will be a series of characters that will be used in the sequence filename for all samples associated with that particular reference. In the example above CAP is used as an alias for CAPN3. Creating an alias for each reference allows Assign™ ATF 1.5 to automatically select that particular reference whenever that alias name is included in a sample's sequence-file name. This eliminates the need to manually select the reference when importing the sample sequences.*

6.4.3.1 In the **Reference Aliases** section, select the reference from the **Ref:** drop down menu.

- 6.4.3.2 Select the alias from the **Alias** drop-down menu. **CAP** in this example. If the alias is not present, type it into the **Alias** field, and then click **Update** (directly to the right of the **Ref** field).
- 6.4.3.3 Repeat this for each reference 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** window to save the changes.

7. Importing Sequences for Analysis

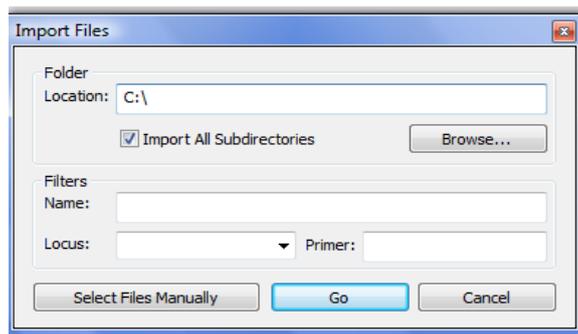
Once parameters have been established in **Settings** (including the creation of applicable references) Electropherograms (EPG) or text sequence can be imported into Assign™ ATF 1.5 for analysis. Sequences can be imported as individual files or by directory, including subdirectories. Importing sequences by directories enables high throughput analysis. It also allows filters to be applied so that only specific samples are imported, or those sequences from a particular locus e.g. all sequences with the same sample name within the selected folder can be imported. This is useful for comparing sequences from the same patient over time or for importing sequences from different loci for the same patient.

Sequences can be imported into a new project or an existing project.

7.1 Importing Sequences by Directory

Please take care when commencing your work to ensure that your desired settings file is loaded in **Edit > Settings > General** (refer to section 6.1.1). This will not be relevant if you only have a single 'default' settings file.

- 7.1.1 Logging into Assign™ ATF 1.5 automatically opens a new project. Multiple projects can be created by selecting **File > New** (Ctl-N also performs this function).
- 7.1.2 To open an existing project select **File > Open** (Ctl-O also performs this function). Navigate to the location of the required project to open it.
- 7.1.3 To import sequences by directory (i.e. import all EPGs in a given folder), select **File > Import > Electropherograms** on the top menu bar.



- 7.1.4 In the pop up window, click on **Browse** and navigate to the folder that contains the sequences. Highlight the folder and click **OK**.
- 7.1.5 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**.
- 7.1.6 If asked to select a reference, select the relevant reference from the drop down menu. Check **Use for Subsequent Samples** if all samples are to be compared to the same reference. *If you accurately set up a reference alias in the naming convention (section 6.4.3) the reference will be automatically selected and you will not need to perform this step.*
- 7.1.7 Use the **Filters** dialogue to import specific samples filtered by:
 - 7.1.7.1 **Name:** All sequence files with a specific sample name. The file names must be identical in the sample identifier region as defined in the naming settings (see section 6.4.1).
 - 7.1.7.2 **Locus:** All sequence files from a specific locus. The file names must be identical in the reference identifier region as defined in the naming settings (see section 6.4.2).

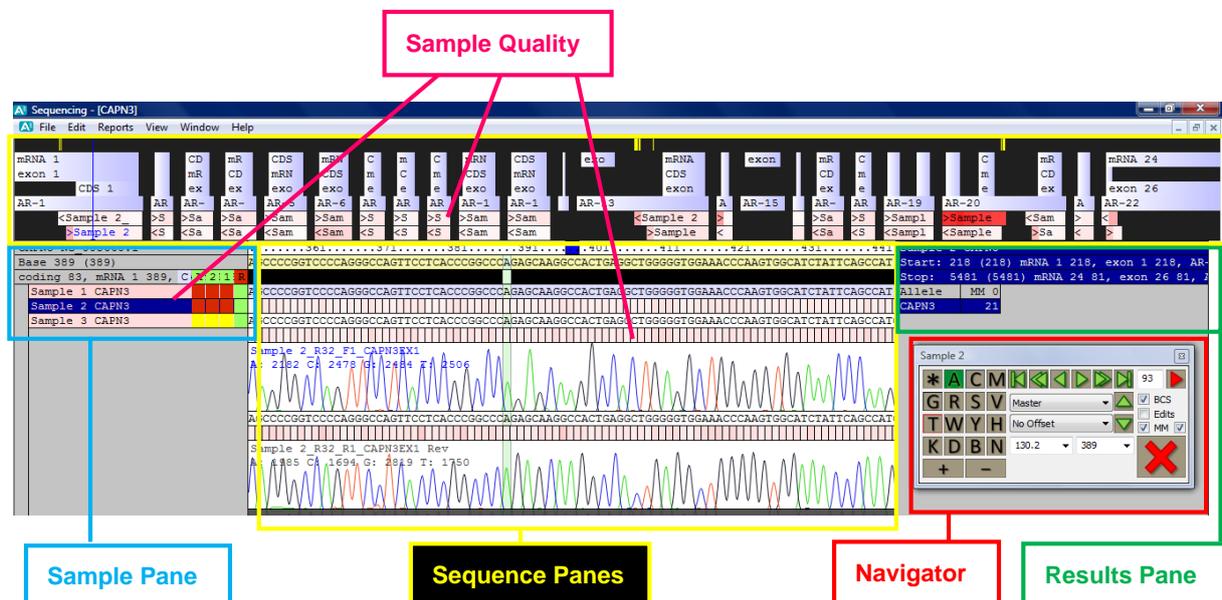
7.2 Importing Sequences Individually

- 7.2.1 Open a new or existing project, as described above.
- 7.2.2 To import only selected EPGs 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. Click **Open**.

8. The Screen Layout, Editing, and Analysis

Once the sequence data has been imported, the software screen becomes populated with sample, sequence, and result information. Sequences can be viewed and edited. Edits result in real time updates of the results screen.

There are several major panes; the **Sample Pane**, **Sequence Panes**, the **Results Pane**, and the **Navigator**.



The screen shows information for a particular sample. It includes the sample ID, the EPG data, and the aligned consensus sequences for a sample against the reference.

8.1 Sequence Data Quality is indicated by Shading

Throughout the various sections of the screen white to red shading is used to demonstrate sequence data quality. The probability of a base call error is increased if the quality of the data is poor. Assign™ ATF 1.5 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.

8.2 Sample ID Pane

All of the samples imported into the project are listed on the left side of the screen in the **Sample ID Pane**:

Sample Quality
Sample names are colour coded from white to red to indicate overall data quality for the sample. Red

Report Status
A green box in the R column indicates the sample can be reported using the Report Generator. Clicking the green box will remove the sample from the report and turn the box red.

Displayed Sample
The sample that is currently displayed on the screen (i.e. EPGs, consensus sequence etc) will be highlighted dark blue. A different sample can be displayed by selecting it.

Comments
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 report.

Analysis Status
A green box in the A column indicates that all priority positions requiring confirmation have been confirmed using the Navigator. This automatically turns green once all priority positions have been confirmed.

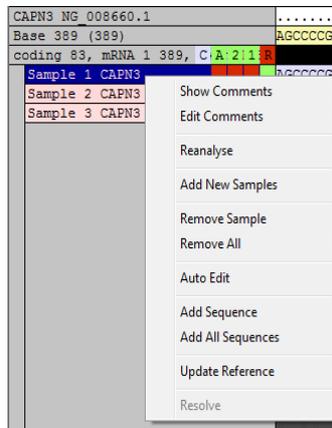
2nd Review Status
A green box in the 2 column indicates the sample has undergone a **second review** by reviewer with high authority. Checking this box will lock the sample and prevent any further edits unless the box is manually unchecked.

1st Review Status
A green box in the 1 column indicates that the sample has undergone the **first review**. After the first reviewer has completed analysis, the yellow/red box in the 1 column must be clicked to change it to green.

| Sample | C | A | 2 | 1 | R |
|----------------|---|---|---|---|---|
| Sample 1 CAPN3 | | | | | |
| Sample 2 CAPN3 | | | | | |
| Sample 3 CAPN3 | | | | | |

Red boxes in the A, 1, and 2 columns indicate a QC warning. **Yellow boxes** have no QC warning.

8.2.1 Right click on the sample name in the **Sample Pane** to access sample options:



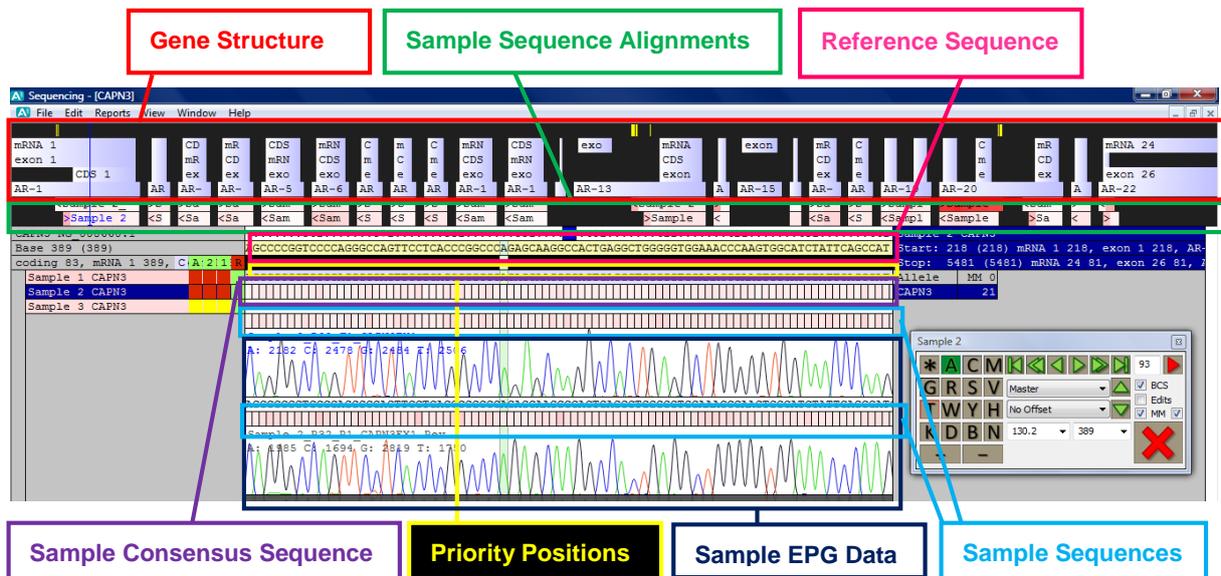
- 8.2.1.1 **Show Comments** will display any quality warnings or comments about a sample.
- 8.2.1.2 **Edit Comments** provides a text box to record any comments about a sample. These comments will appear on the report. A light blue box in the C column will indicate that a comment is present.
- 8.2.1.3 **Reanalyse** will remove any edits and trims that have been made, restoring the sample to the initial state following import. Can be used if the analysis settings have been edited after the sequences have been imported.
- 8.2.1.4 **Add New Samples** will launch the **Import Files** window to import new samples.
- 8.2.1.5 **Remove Sample** will remove the highlighted sample from the project.
- 8.2.1.6 **Remove All** will remove all samples from the project.

- 8.2.1.7 **Auto Edit** will run the auto edit function if this hasn't already been selected in settings (section 6.2.3.3).
- 8.2.1.8 **Add Sequence** will add the sequence information of the active sample to the reference to create a Library. Subsequent analysis will list sequence differences between active samples and the additional sequences in the reference library.
- 8.2.1.9 Similar to 8.2.1.8 **Add All Sequences** will add all sequences in the project to the references.
- 8.2.1.10 **Update Reference** will set the sequence of the active sample as the reference sequence against which all samples within the project will be compared. All sequences in the project will be reanalysed against this new reference.

Note that **Resolve** is a HLA specific application and is not commonly used. Please contact Conexio Genomics if you require further information.

8.3 Sequence Pane

Importing the sequence EPGs results in a display of how the sequence files are orientated according to the gene structure, the sequence EPGs themselves, the Assign™ ATF 1.5 base calls and quality score information.



8.3.1 Gene Structure

The light blue bands shown below indicate the genetic structure of the reference sequence. The yellow bars above this indicate those positions within the sequence which differ between the sample and the reference.



8.3.2 Sample Sequence Alignments

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



8.3.3 The Reference Sequence

Beneath the sequence alignment map is the reference sequence. If using a library (rather than a single reference) the sequence displayed is a consensus. The sequence is shaded yellow to indicate exons and white to introns. In addition, positions shaded light blue indicate the reference sequence contains deletions at this position and dark blue regions indicate the position of insertions.

```

.....361.....371.....381.....391......401......411......421......431......441
AGCCCCGGTCCCCAGGGCCAGTTCCTCACCCGGGCCAGAGCAAGGCCACTGAGGCTGGGGGTGGAAACCCCAAGTGGCATCTATTCAGCCAT
  
```

8.3.4 The Sample Consensus Sequence

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

```

AGCCCCGGTCCCCAGGGCCAGTTCCTCACCCGGGCCAGAGCAAGGCCACTGAGGCTGGGGGTGGAAACCCCAAGTGGCATCTATTCAGCCAT
  
```

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

8.3.5 Priority Positions

Coloured boxes in the black row below the reference sequence (and above the sample consensus sequence) highlight specific sites:

```

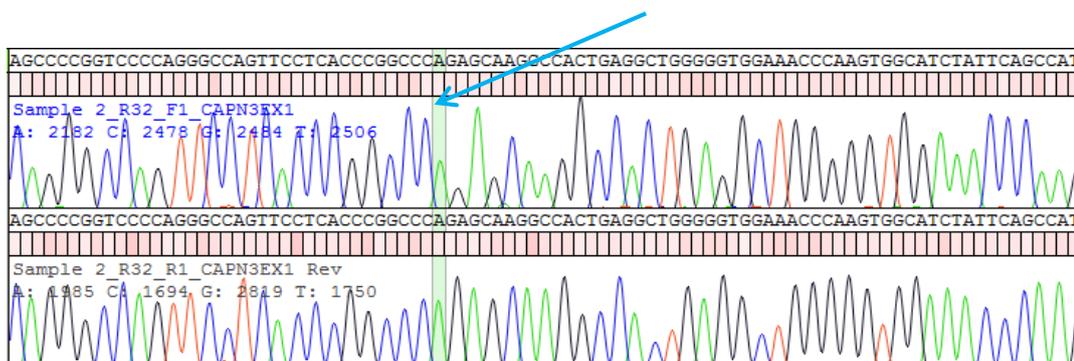
.....301......311......321......331
SRVDVVRBNBDEHVNSRTGRACRMYBWNTGSAKRYRHAR
  
```

- 8.3.5.1 **Yellow boxes** indicate positions within the sequence that differ from the reference sequence.
- 8.3.5.2 **Green boxes** appear when the base call has been confirmed using the Navigator (refer to 8.7.5).
- 8.3.5.3 **Orange boxes** indicate auto-edited positions.
- 8.3.5.4 **Blue boxes** appear when a position has been manually edited.
- 8.3.5.5 **Purple boxes** indicate a user defined variant position (refer to section 6.3.3 and 8.3.6.5).

8.3.6 The Sample EPG Data

Beneath the sample consensus sequence is the sample sequence EPG data, the software base calls and quality indicator. The EPG panels contain the sequence filename and the signal intensities of the 4 bases.

The position you are at will be indicated by a bar through the sequence and will be the colour of the base at that position.



8.3.6.1 **Migration through the sequence:** Can be performed in three ways:

- ◇ Clicking on the EPG pane and using the arrows on the computer keyboard.
- ◇ Using the **Navigator** as per section (8.6).
- ◇ Clicking on the highlighted bar and using the mouse to drag it.

8.3.6.2 **Zooming the EPG:** The height and width of the EPG can be resized by pressing the computer keyboard's **Shift** key and **up/down** or **left/right arrows** respectively.

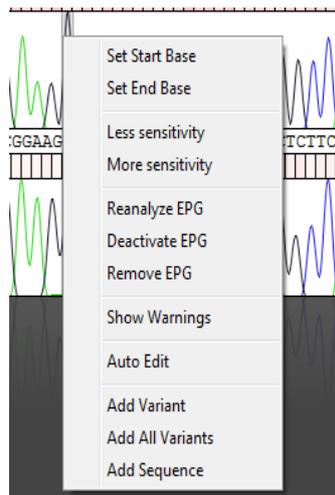
8.3.6.3 **Hiding EPG Traces:** Simultaneously pressing the computer keyboard's **Shift** key and one of the letters representing the 4 bases (**A, C, G** or **T**) will remove the trace of this base from the EPG. Repeating the process will return the trace.

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

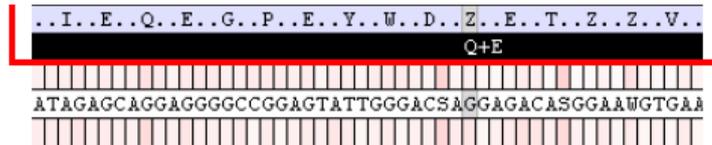
It is also useful for confirming low level mutations

8.3.6.4 **Expanding the EPG Window:** Clicking the boundaries of the panes whilst holding the mouse key enables the expansion or contraction of the EPG window.

8.3.6.5 **Electropherogram Options:** Right clicking on a given EPG gives access to options for each EPG. Any changes made will only be applied to the selected EPG.



- ◇ **Set Start Base** will trim off all data to the left of the cursor. Hold **Ctrl** on your computer keyboard simultaneously to apply to all samples.
 - ◇ **Set End Base** will trim off all data to the right of the cursor. Hold **Ctrl** on your computer keyboard simultaneously to apply to all samples.
 - ◇ **Less sensitivity** results in a reanalysis of an EPG after reducing the heterozygous detection limit. It will filter out background noise, raising the detection threshold by 10% to a maximum of 50%. This function is very effective for improving base call accuracy of data with high background.
 - ◇ **More sensitivity** results in a reanalysis of an EPG after increasing the heterozygous detection limit. It decreases the threshold of detection to a minimum of 10% thus increasing sensitivity and calling more heterozygous bases. This function is useful for accurately detecting low level mutations in data free of background.
 - ◇ **Reanalyze EPG** will remove any user edits and trims from the EPG.
 - ◇ **Deactivate EPG** will remove the EPG from analysis, but does not remove it from the project. Right clicking on the deactivated EPG again enables the option to **Activate EPG**.
 - ◇ **Remove EPG** will completely remove the EPG from the project.
 - ◇ **Show Warnings** will display any quality warnings about that particular EPG.
 - ◇ **Auto Edit** will run the auto edit function if this hasn't already been selected in settings (section 6.2.3.3).
 - ◇ **Add Variant** will add to the reference sequence, the current position as a variant.
 - ◇ **Add All Variants** will add to the reference sequence, all positions that vary from the to the reference sequence as a variant.
 - ◇ **Add Sequence** will add the sequence of the active sample as a reference. It will display in the **Results Pane**.
- 8.3.6.6 **Amino Acid Sequence View:** The amino acid sequence can be viewed instead of base pair sequence by performing the following:
- ◇ Pressing **Ctrl+A** in the EPG pane shows the amino acid sequence of the reference sequence and the consensus sequence of the sample.

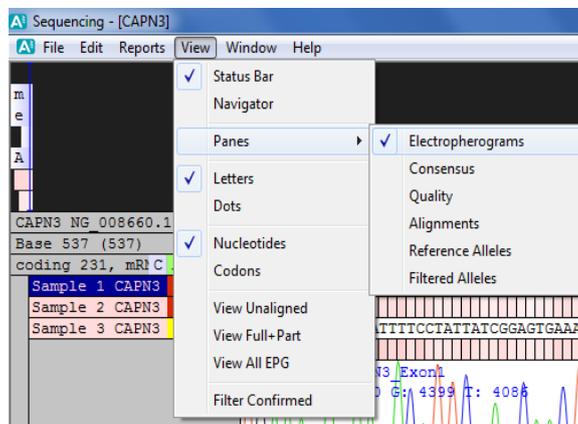


“Z” is used to represent heterozygous positions. Clicking on the “Z” or scrolling to the position within the sequence will show the amino acids at these positions in a new black line. This will assist with understanding the consequence of novel alleles.

◇ Re-press **Ctrl+A** in the EPG pane to re-display bases.

8.4 View Options

The **View** options enable the sample sequence data and the reference sequence data to be viewed in different ways. Select the **View** options from the main menu. Assign™ ATF 1.5’s ability to import thousands of sequences, it’s accurate and novel approach to base calling and the simple switch between EPG and sample text sequence simplifies high throughput SNP screening.



8.4.1 Status Bar

Click to display or un-display the **Status Bar** at the bottom of the screen. This shows the status of the software (e.g. ‘Ready’ indicates it is ready for the next action).

8.4.2 Navigator

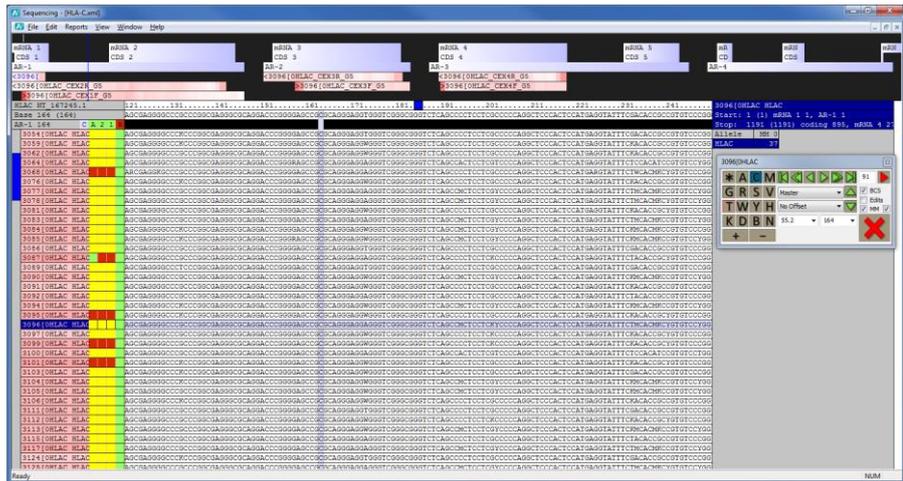
Click to re-display the **Navigator** if it has been closed.

8.4.3 Panes

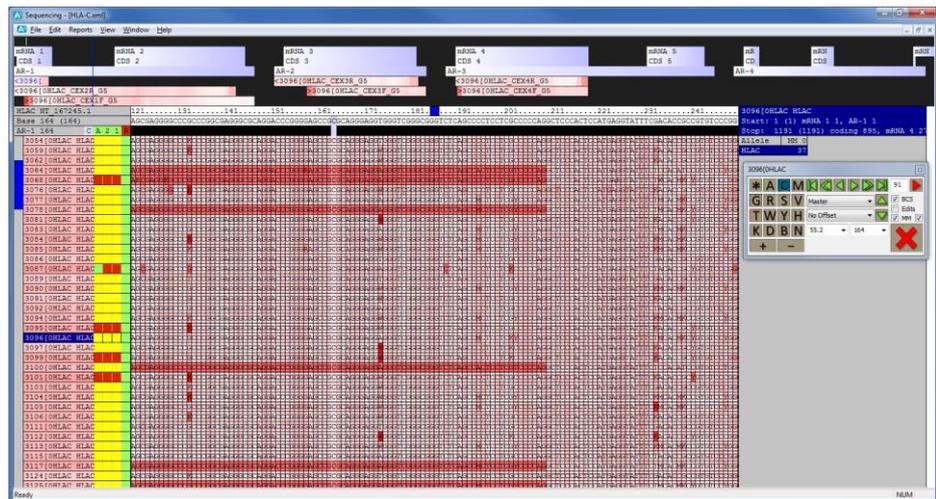
8.4.3.1 **Electropherogram** is the default view setting and displays the EPG tracings of the sample.

8.4.3.2 **Consensus** displays the text consensus sequence for all samples in a project allowing high throughput analysis.

*This can be used with **Dots** (see below) to enable easy location of positions that differ from the reference. Dots are shown at each base where the sample sequence matches the reference sequence. Bases that differ will be shown as letters.*



8.4.3.3 **Quality** displays the text consensus sequence shaded according to the consensus sequence base call score for each base for every sample in the project. *This too can be used with **Dots** (see below) to enable easy location of positions that differ from the reference. Dots are shown at each base where the sample sequence matches the reference sequence. Bases that differ will be shown as letters.*



8.4.3.4 **Alignments** is used only in Genotyping mode. It displays the text consensus sequence for each of the possible allele combinations for a given sample with the test sample at the top of the list. Mismatches with the sample consensus appear highlighted in yellow.

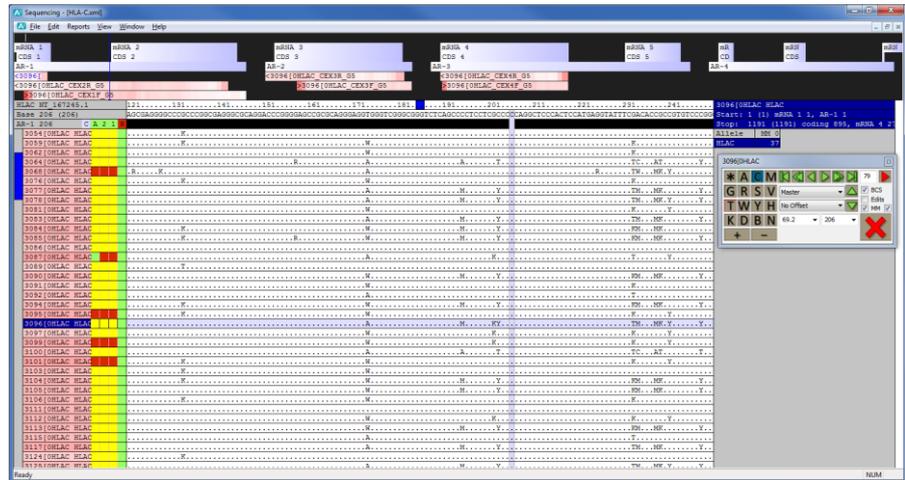
Reference Alleles is used only in Genotyping mode. It shows the sequence of alleles within the reference compared to the sequence of the selected sample. Differences are highlighted in yellow. The alleles are shown in the **Results Pane**.

Filtered Alleles is used only in Genotyping mode. As base calls are confirmed the alleles differing from the sample are automatically removed from the **Results Pane**. Selecting **Filtered Alleles** will redisplay these removed alleles.

8.4.4 Letters or Dots

8.4.4.1 **Letters** will show the bases for consensus and reference sequences as letters.

8.4.4.2 **Dots** will show dots at each base where the sample sequence matches the reference sequence. Bases that differ will be shown as letters. This enables easy location of positions that differ from the reference.



8.4.5 Nucleotides or Codons

8.4.5.1 **Nucleotides** will show the sequence positions as bases.

8.4.5.2 **Codons** will show the sequence positions as codons.

8.4.6 View Unaligned

By default **View Unaligned** is off - if using only coding sequence in a reference, any intronic overlap of the sequences between exons is excluded. Select **View Unaligned** to include the intronic overlap. To re-exclude intronic overlap, unclick the **View Unaligned** option.

8.4.7 Genotyping Specific Functions

View Full+Part, **Filter confirmed** and **View All EPG** are HLA typing specific functions. Please contact Conexio Genomics for more information.

8.5 Results Pane

The Results Pane (below) includes the sample name, the start and stop positions of the sample sequence, and lists the number of mismatches (**MMO**) with the reference sequence (**Allele**).

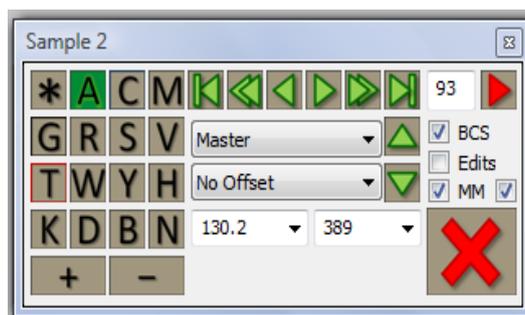
If the software detects the likelihood of indel data there will be a warning in this pane: **Please check the sequence for heterozygous indels or poor quality data**. Refer to Appendix 5: Heterozygous Insertion/Deletion and Poor Quality Sequence Warnings for the procedure to follow in the case of this warning.

| | |
|--|------|
| Sample 2 CAPN3 | |
| Start: 218 (218) mRNA 1 218, exon 1 218, AR- | |
| Stop: 5481 (5481) mRNA 24 81, exon 26 81, I | |
| Allele | MM 0 |
| CAPN3 | 21 |

If genotyping and comparing the sample sequence to a library of reference sequences an **Allele 1** and **Allele 2** column will display listing the allele pairs which are best matched to the sample sequence.

8.6 Navigator

The **Navigator** enables sequence editing, moving between samples and moving between positions within a sequence. Importantly, the **Navigator** is used to validate base calls at user-defined validation positions, such as those where there is a low BCS, where user edits have been made, where there are mismatches between the sample and the reference sequence(s), and/or where there are variant positions. The user can define which of these positions they wish to easily locate and use the **Navigator** arrows to skip to these positions. This allows 'priority editing' and eliminates the need to review all positions in the sequence. Refer to section 8.7 for a detailed description of the use of the **Navigator** for Priority Editing and Confirmation of Base Calls.



8.6.1 Arrow Functions

- 8.6.1.1 Selecting either   (single arrow) button moves one position left or right.
- 8.6.1.2 Selecting either   (double arrow) button moves to the next position requiring user confirmation (refer to section 8.6.2).
- 8.6.1.3 Selecting either   (blocked arrow) button moves to the start or the end of the sequence.
- 8.6.1.4 Selecting the up or down arrow moves to the sample above or below in the **Sample Pane**.

8.6.2 Setting Positions for User Confirmation (Priority Editing)

- 8.6.2.1 The user can select which positions to include for validation by checking or un-checking the following checkboxes:
 - ◇ **BCS** to include positions with a low Base Call Score.
 - ◆ The BCS (quality assessment number) for each base appears above the BCS box in the navigator.
 - ◆ Bases with a BCS of lower than 70 (or lower than 35 for single direction sequences) will be included for validation if the **BCS** checkbox has been checked.
 - ◇ **Edits** to include positions edited by the user.
 - ◇ **MM** to include potential Mismatch positions (positions for which the test consensus sequence differs from the reference).
 - ◇ **User Defined Variant Positions** selected by checking the checkbox to the right of MM.
- 8.6.2.2 If you wish to save the priority position settings, within **Edit > Settings** ensure the Settings folder you wish to save to is selected and click **Update**.

8.6.3 Other Navigator Functions

8.6.3.1 **Master** drop down menu selects between layers.

*Note that **Layer** is a HLA specific application and not commonly used. Please contact Conexio Genomics if you require further information.*

8.6.3.2 **No Offset** drop down menu allows the user to choose a particular region within the sequence.

8.6.3.3 **Codon** and **Base** locations are located underneath the **No Offset** drop down menu, from left to right respectively.

◇ To navigate to a particular codon type a codon and position number e.g. **130.2** to move to position **2** of codon **130**.

◇ To navigate to a particular base, enter the base position in the right drop down and hit the keyboard's **Enter** key.

8.7 Priority Editing and Confirming Base Calls

*Please take care when commencing your work to ensure that your desired settings file is loaded in **Edit > Settings > General** (refer to section 6.1.1). This will not be relevant if you only have a single 'default' settings file.*

8.7.1 Ensure the sample to be reviewed is highlighted by clicking on it in the left **Sample Pane**.

8.7.2 Navigate to the first base in the sample by clicking on the  button in the **Navigator** and then navigate to the first position requiring user validation by selecting the  button.

8.7.3 The **X** button indicates that the current position has not yet been validated by the user.

8.7.4 **If a base call needs to be edited:**

8.7.4.1 The call can be changed manually by selecting or unselecting the base letters on the **Navigator**. Mixed bases are automatically updated (e.g. if the base call was **M** but the user decides the call should be **A**, unselecting **C** will leave only **A**).

8.7.4.2 The **+** and **-** buttons can be used to add or remove insertions (+) or deletions (-).

The raw sequence data is not changed with edits, only the consensus sequence is changed.

8.7.5 **Confirm a base call** by clicking on the **X** button.

8.7.5.1 As **X** is clicked, the **X** changes to a  and a green box will appear above the base in the sample consensus sequence, indicating that the position has been confirmed.

8.7.5.2 Assign™ ATF 1.5 will then proceed to the next position to be validated. Using the **X** button, continue to navigate through the sequence verifying all desired bases.

8.7.6 Once all priority review bases have been verified, the **X** will change to a  indicating no additional bases need confirming. In the **Sample Pane**, the box under the **A** column (Audit) will turn green. This indicates that the sample is ready for review.

8.8 Reviewing the Analysis

Assign™ ATF 1.5 allows the review of analysis (quality control checks). Two levels of review are possible (review '1' and '2'). Only authorised users may perform final reviews ('2').

| CAPN3 NG 008660.1 | | | | | | |
|-------------------|-------|---|---|---|---|---|
| Base 3224 (3224) | | | | | | |
| coding 1554, mRNA | | C | A | 2 | 1 | R |
| Sample 1 | CAPN3 | | | | | |
| Sample 2 | CAPN3 | | | | | |
| Sample 3 | CAPN3 | | | | | |

8.8.1 First Review

- 8.8.1.1 Once the ✓ is displayed in the Navigator and the box under the **A** column (Audit) in the **Sample Pane** is green, the sample is ready for review.
- 8.8.1.2 Clicking on the box for the sample under the **1** column in the **Sample Pane** will indicate the sample has been reviewed once. The box will turn from red to green.
- 8.8.1.3 At this point, the project should be saved to prevent any accidental loss of data review as per section 10.3.
- 8.8.1.4 Once the review has been completed on the first sample, click on the second sample and repeat the process.

Save your work often to prevent any loss of data.

8.8.2 Second (Final) Review

- 8.8.2.1 If a secondary review is desired, the project should be opened by the second reviewer to preserve the audit trail. The second reviewer is to log on to the software and select **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.
- 8.8.2.2 Select the desired priority review positions in the **Navigator**. Using the left **double arrow button** to navigate to the required positions. Using the **X** button will cancel and reapply the audited positions.
- 8.8.2.3 Once all positions have been reviewed, the second reviewer can click on the box in the **2** column in the sample pane. The box will turn from red to green.
- 8.8.2.4 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.

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

8.9 Audit Trail

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 report.

9. Reporting

The Assign™ ATF 1.5 reports enable a comprehensive assessment of the sequence data. The standard report formats are:

- **Variants Report:** Used for applications where test sequence is compared with a single reference sequence.
- **Genotype Report:** Used for genotyping applications when matching a sample sequence against a library of known sequences.
- **FASTA Report:** Provides a FASTA file of sequences from all samples in the Assign™ ATF 1.5 project.
- **Quality Report:** (BCS Report) Enables a quality control analysis of samples within the Assign™ ATF 1.5 project or for all projects within a specific directory.

Note that **HARPS Report** is a HLA specific application. Please contact Conexio Genomics if you require further information.

If you have unique reporting requirements please contact Conexio Genomics at support@conexio-genomics.com for advice about customised reports.

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

9.1 Variants Reports

The screenshot shows a 'Reports' dialog box with the following settings:

- Output Filters and Numbering:** Sample: _All_, Locus: (empty), Layer: _All_, Group: (empty), Region: _All_, Nuc:
- Options:** BCS: , Audit: , User: , HGVS: , Observed: , All Alleles: , Indel Bases:
- Output Type:** Vertical (selected), Horizontal, Table 256 columns
- Output Format:** Text, Excel (selected), XML

The Variants report reports differences between the test sequence and the reference.

9.1.1 **Output Filters and Numbering** can be used to filter based on the following:

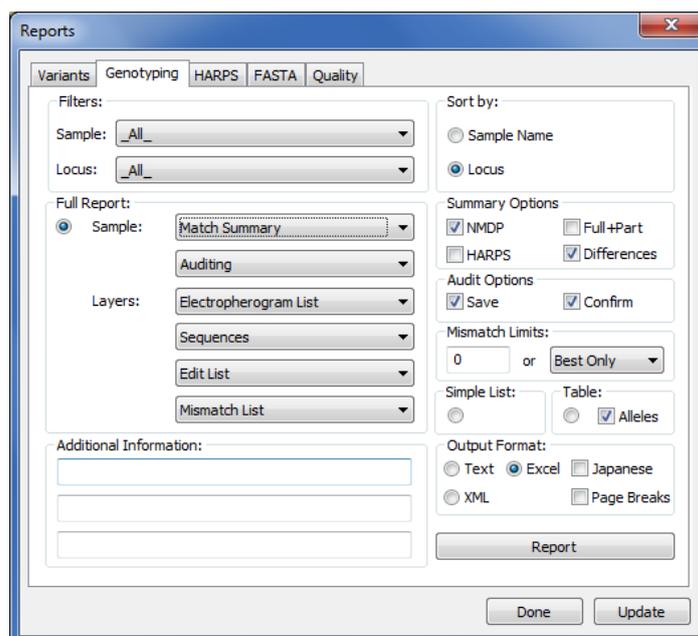
- 9.1.1.1 **Sample:** 'All' samples or an individual sample.
- 9.1.1.2 **Locus:** 'All' loci or individual loci.
- 9.1.1.3 **Group:** 'All' groups or individual groups.
- 9.1.1.4 **Region:** 'All' regions or individual regions.
- 9.1.1.5 **Nuc:** Check the box under **Nuc** for variants to be reported as nucleotides. Leave the box unchecked for variants to be reported as codons.

Note that **Layer** is a HLA specific application and is not commonly used. Please contact Conexio Genomics if you require further information.

- 9.1.2 **Options** allows the following to be displayed on the report:
 - 9.1.2.1 **BCS**: Checking this will display the base call score quality values on the report.
 - 9.1.2.2 **Audit**: Checking this will display an audit report including edits and the identification of the user who performed them.
- 9.1.3 **Variants** allows the selection of the variants to be displayed on the report:
 - 9.1.3.1 **User Defined**: Reports sequence at positions defined by the user.
 - 9.1.3.2 **HGVS**: Produces a report according to the nomenclature proposed by HGVS (<http://www.hgvs.org/mutnomen/>). Unchecking this box allows the user to select individual regions to be reported.
 - 9.1.3.3 **Observed**: Reports any sequence differences between the sample sequence and the reference sequence.
 - 9.1.3.4 **All alleles**: Reports the variants between the sample sequence and the sequence of all alleles in the database.
 - 9.1.3.5 **Indel Bases**: Reports the actual inserted or deleted sequence as opposed to reporting the number of bases.
- 9.1.4 **Output Type** enables customisation of how variants are listed on the report:
 - 9.1.4.1 **Vertical**: Each sample is listed separately with a vertical list of variants for that sample against the reference.
 - 9.1.4.2 **Horizontal**: Each sample is listed one directly under the other with a horizontal list of variants for that sample against the reference.
 - 9.1.4.3 **Tabular**: Individual regions can be reported in table format with the reference amino acid sequence listed horizontally (one amino acid per cell). Each sample is aligned underneath, stating the amino acid at each position. It is possible to choose the number of columns listed.
- 9.1.5 **Output Formats** allows the selection of the report format:
 - 9.1.5.1 **Excel**: Produces a report in an excel worksheet.
 - 9.1.5.2 **XML**: Produces a report in xml format
 - 9.1.5.3 **Text**: Produces a text file report.
- 9.1.6 **Generate Report** is to be clicked to produce the report once your desired settings have been selected.

9.2 Genotyping Reports

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



9.2.1 **Filters** can be used to filter based on the following:

9.2.1.1 **Sample:** 'All' samples or individual samples in the project.

9.2.1.2 **Locus:** 'All' loci or individual loci in the project.

9.2.2 **Full Report** enables customisation of the report format: The drop down menus enable the inclusion or exclusion of specific items from the report:

9.2.2.1 **Sample** options:

◇ **Match Summary** will list all the matched allele pair combinations for all samples selected.

◇ **Auditing** will include a comprehensive audit report including date, time, and identification of the user validating the results.

9.2.2.2 **Layers** options:

◇ **Electropherogram List:** Lists the EPG sequence files analysed.

◇ **Sequences:** Produces a list of the sample consensus sequences.

◇ **Edit List:** Lists manual and auto edits performed during the analysis.

◇ **Mismatch List:** Lists the mismatched nucleotide information of the closest matched sequences within the libraries.

9.2.2.3 **Additional Information** can be used to add comments. These comments appear at the top of the report.

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

9.2.2.5 **Summary Options** includes options to include on the report:

*Note that **NMDP Codes** and **HARPS** are designed for HLA typing*

◇ **Full+Part:** Indicates which alleles are fully or partially typed with the reference database.

◇ **Differences:** Will report the location of differences between the best matched genotype and those closest matched.

9.2.2.6 **Audit Options** (if the **Auditing** option is selected, as described above) allows reporting of:

◇ **Save:** All the Save events for the project.

◇ **Confirm:** All changed and priority base confirmations.

9.2.2.7 **Mismatch Limits** allows reporting of a nominated number of mismatches between the sample sequence and the library sequence, or the best (1 or 2) matches.

9.2.2.8 **Simple List or Table:**

◇ **Simple List:** Lists each allele as a string of text (with no summary information).

◇ **Table:** Lists each allele in a table (with no summary information).

9.2.3 **Output Formats** allows the selection of the report format:

9.2.3.1 **Text:** Produces a text file report.

9.2.3.2 **XML:** Produces a report in xml format.

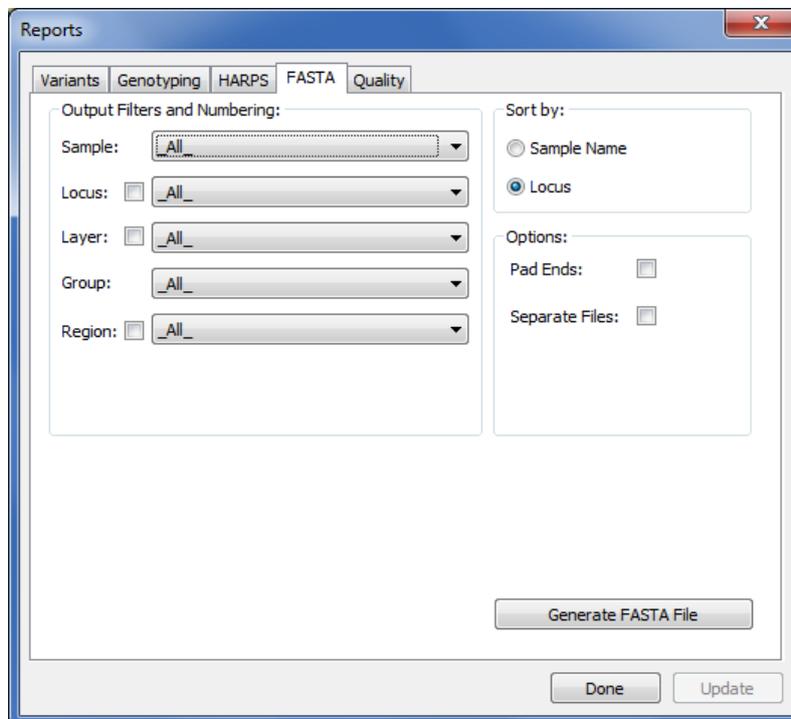
9.2.3.3 **Excel:** Produces a report in an excel worksheet.

9.2.3.4 **Page Breaks:** Adds a page break between samples.

9.2.4 **Report** is to be clicked to produce the report once your desired settings have been selected.

9.3 FASTA Reports

The FASTA report allows the production of sequences in FASTA text format.



9.3.1 **Output Filters and Numbering** can be used to filter based on the following:

9.3.1.1 **Sample:** 'All' samples or an individual sample.

9.3.1.2 **Locus:** 'All' loci or individual loci.

9.3.1.3 **Group:** 'All' groups or individual groups.

9.3.1.4 **Region:** 'All regions or individual regions.

*Note that **Layer** is a HLA specific application and is not commonly used. Please contact Conexio Genomics if you require further information.*

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

9.3.3 **Options** allows the following to be displayed on the report:

9.3.3.1 **Pad Ends:** Includes dashes (-) at the end of a sequence to enable all sequences to be the same length.

9.3.3.2 **Separate Files:** Makes a separate file for each sample.

9.3.4 **Generate FASTA File** is to be clicked to produce the report once your desired settings have been selected.

9.4 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.

The screenshot shows a software window titled "Reports" with a "Quality" tab selected. The window contains several filter options: "Sample" (dropdown menu set to "All"), "Locus" (dropdown menu set to "_All_"), "Layer" (dropdown menu set to "_All_"), and "Sequencer" (dropdown menu set to "_All_"). Below these are date and time filters: "From" (Year: 2011, Month: 03, Day: 07, Region: "_All_", Base: "_All_", Direction: "FR") and "To" (Year: 2012, Month: 03, Day: 07, Region: "_All_", Base: "_All_"). There are also fields for "Select Folder" (set to "C:") and "Get Projects" (dropdown menu set to "_Current_"). A "Quality Report" button is located at the bottom right of the dialog box, along with "Done" and "Update" buttons.

9.4.1 **Filters** can be used to filter based on the following:

9.4.1.1 **Sample:** 'All' samples or individual samples.

9.4.1.2 **Locus:** 'All' loci or individual loci.

9.4.1.3 **Sequencer:** 'All' sequencers or an individual sequencer.

Note that **Layer** is a HLA specific application. Please contact *Conexio Genomics* if you require further information.

9.4.2 **From and To** can be used to filter by date or region:

9.4.2.1 **Year Month Day**: Enter a **From** and **To** date to compare quality over time.

9.4.2.2 **Region**: 'All' regions or select a **From** and **To** region to select an individual or multiple consecutive regions for which to compare the quality.

9.4.2.3 **Base**: 'All' bases or select a **From** and **To** base to select an individual or a section of bases for which to compare the quality.

9.4.2.4 **Direction**: Select **FR** to analysis forward and reverse sequences, **F** for just forward and **R** for just reverse.

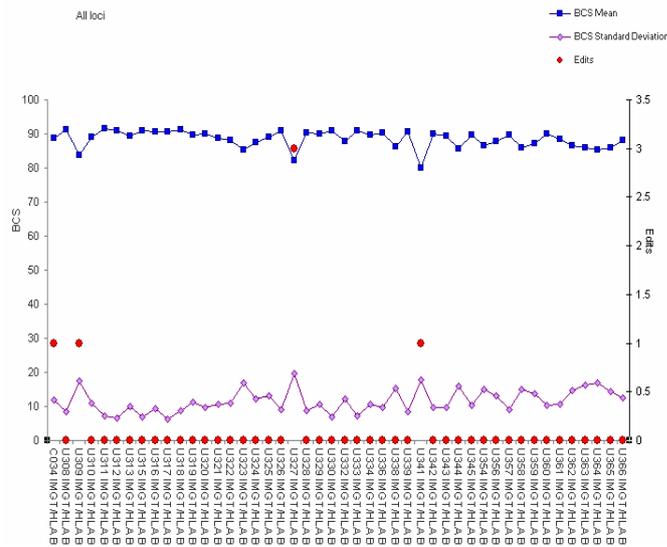
9.4.3 **Select Folder**: Click to select a specific folder within which you would like to compare quality data. **C:** is the default.

9.4.4 **Get Projects** can be used to generate a report from either a selected project or all projects saved within the directory:

9.4.4.1 The default of **_Current_** will produce a quality report of the active Assign™ ATF 1.5 project.

9.4.4.2 By Clicking **Get Projects** you can choose **_All_** or browse the drop down menu to select a specific project.

9.4.5 Once the desired settings have been selected click **Quality Report** to produce the report.



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

9.4.6 The Excel report has various worksheets:

9.4.6.1 **Data**: A spreadsheet with the quality information from which the quality graphs are produced.

9.4.6.2 **BCS Distribution**: A plot of the frequency of BCS from all positions for the consensus sequence from all samples within the project. This also contains the frequency of edits for each BCS.

9.4.6.3 **BCS Means**: A plot of the mean BCS for each sample. This also contains the frequency of edits for each sample.

9.4.6.4 **Signal Strength**: A plot of signal strengths for each sample within the project plotted with the mean BCS.

10. Opening, Saving, and Printing Projects

Assign™ ATF 1.5 projects are saved with links to the EPG. EPGs are not saved as part of the project so that the file sizes are kept as small as possible. EPGs are imported back into the project when projects are reopened.

10.1 Select the Desired Settings File

- 10.1.1 When logging back into the Assign™ ATF 1.5 software, ensure that the correct settings file is opened by clicking **Edit > Settings**. Select the desired settings file and click **Done** (refer to section 6.1.1). This will not be applicable if a single 'default' settings file is used.

10.2 Opening Projects

- 10.2.1 Open a new project by selecting **File > New** and import EPG as per section 7.
- 10.2.2 Existing projects must be opened by selecting **File > Open** and browsing to the saved project.

*Projects CANNOT be opened by clicking on the file name in a folder or in the recent items in the **File** drop down menu.*

10.3 Saving Projects

- 10.3.1 New Assign™ ATF 1.5 projects can be saved using **File > Save As**. Select a file name and location to save the project.
- 10.3.2 Current Assign™ ATF 1.5 projects can be saved using **File > Save**.
- 10.3.3 The saved project (.xml format) contains information about which EPGs were used (including their saved location), any edits and confirmations that were performed, and an audit trail (including details about the user).

10.4 Printing Projects

- 10.4.1 Assign™ ATF 1.5 projects (including the EPGs) can be printed using **File > Print**.

Appendix 1: Frequently Asked Questions

Installation FAQs

I Am Using a Previous Version of Assign™ ATF 1.5 – How Do I Install v1.5?

Current users of Assign™ ATF 1.5 can simply email support@conexio-genomics.com to be supplied with the ATF 1.5 installer.

Once downloaded, double click on the installer file icon and follow the instructions for installation.

Your current version will not be replaced.

You will need to copy your licence key file(s) from the **ATF xxx** folder (where xxx = the old version) into the newly created **ATF 150** folder.

In order to continue to use existing references and settings files, copy the contents of the **ATF xxx\data\references** and **ATF xxx\data\settings** folders into the **ATF 150\data references** and **ATF 150\data\settings** folders respectively.

Once installed the **ATF 150** folder is usually located at **C:\ProgramData\Conexio Genomics\ATF 150**. However, it may be located elsewhere depending on your operating system. To locate the data folders for both the old ATF version and the new ATF 1.5:

- Double click on the relevant **Assign™ ATF 1.5** desktop shortcut, leave the default Operator of **admin** and enter the default password of **cg01**. Click **More>>**.
- The **System File Location** is listed at the bottom of the screen.

Can I install Assign™ ATF 1.5 on a network drive, rather than on individual PCs?

Assign™ ATF 1.5 can be installed on a network drive. Please email support@conexio-genomics.com for detailed instructions. It is recommended that users seek assistance from their IT division when performing a network installation.

Analysis FAQs

All sequences from a sample are not appearing together as part of the same active sample?

This is usually because Assign™ ATF 1.5 has not been set up so that it can uniquely identify the sample name within the sequence filename. Please review **Naming Conventions** in section 6.4.

The sequence data has cut off base calling but I can still see good quality sequence?

The software has trimmed the base calling region based on quality or your chosen settings. To force base calling at trimmed positions go to **View** and select **View Unaligned**. Then move the mouse to the EPG field and right click at a position that includes the region required to be analyzed and select **Trim Right** (if the mouse is to the right of the unanalyzed sequence) or **Trim Left** (if the mouse is to the left of the unanalyzed sequence).

How do I undo a mistakenly trimmed sequence?

Right click on the EPG and select **Reanalyse**. Note that the EPG is re-imported and all edits will be lost. Note that the **Reanalyse** function can also be performed on all EPGs or all samples within a project. This is usually performed if the settings are changed after sequences have been imported.

What Keyboard Shortcuts (Hotkeys) are available in Assign™ ATF 1.5?

| Key | Function |
|---------------------------------|---|
| Right Arrow | Move one base Right |
| Shift+Right Arrow | increase (zoom in) peak width in the Electropherogram view |
| Ctrl+Right Arrow | Move to the next position with a low BCS, a mismatch and/or an edit (As toggled in the navigation window) |
| Ctrl+Shift+Right Arrow | Go to the end of the consensus sequence |
| Left Arrow | Move one base left |
| Shift+Left Arrow | Reduce (zoom out) peak width in the Electropherogram view |
| Ctrl+Left Arrow | Skip to the previous position with a low BCS, a mismatch and/or an edit (As toggled in the navigation window) |
| Ctrl+Shift+Left Arrow | Go to the start of the consensus sequence |
| Up Arrow | Move to previous sample |
| Shift+Up Arrow | Reduce the size of each Electropherogram view |
| Ctrl+Up Arrow | Increase the draw height for peaks in the current Electropherogram |
| Ctrl+Shift+Up Arrow | Increase the draw height for peaks for all Electropherograms |
| Down Arrow | Move to next sample |
| Shift+Down Arrow | Increase the size of each Electropherogram view |
| Ctrl+Down Arrow | Reduce the draw height for peaks in the current Electropherogram |
| Ctrl+Shift+Up Down | Reduce the draw height for peaks for all Electropherograms |
| Tab | Confirm base call at current position |
| A/C/G/T/M/K/R/W/D/S/Y/B/V/H/N/X | Edit the base at the current position |
| Shift+A Shift+C Shift+G Shift+T | Toggle the trace for that base on or off in all Electropherograms |
| Ctrl+A | Toggle Base/Ammino Acid view |
| Ctrl+F | Find Sequence |
| Shift+l | Toggle electropherogram information |
| Ctrl+M | Toggle Map |
| - | Shift the BCS Baseline |
| Ctrl+[Ctrl+] | Cycle sample pane view between Consensus, Quality, Alignments, Alleles and Electropherograms |

The EPG indicates the presence of an indel but this has not been reported in the report.

Assign™ ATF 1.5 can calculate the deleted sequence within an indel. However, the distinction between indel and poor quality sequence is difficult for the software to calculate. This is particularly true if the sequence quality is poor. If Assign™ ATF displays the warning ***Please check the sequence for heterozygous indels or poor quality data*** perform the steps in Appendix 5: Heterozygous Insertion/Deletion and Poor Quality Sequence Warnings.

Newly loaded sequences shown as already reviewed.

It is possible for newly imported samples to show that they have already been reviewed (green box in **A** column). The most common cause is that the **MM** and/or **BCS** boxes are NOT checked in the **Navigator**. In this case it is likely that there are no Priority Positions to be confirmed and as such the sample is marked as analysed. It is possible that if only **BCS** is checked and the sequence quality is good then there may also be no positions to validate. To resolve this, check the **Navigator** to ensure that the **MM** and **BCS** boxes are set as appropriate for your application. Refer to **Setting Positions for User Confirmation (Priority Editing)** in section 8.6.2.

The EPG contains the correct base call at a variant site but this has been changed in the consensus sequence.

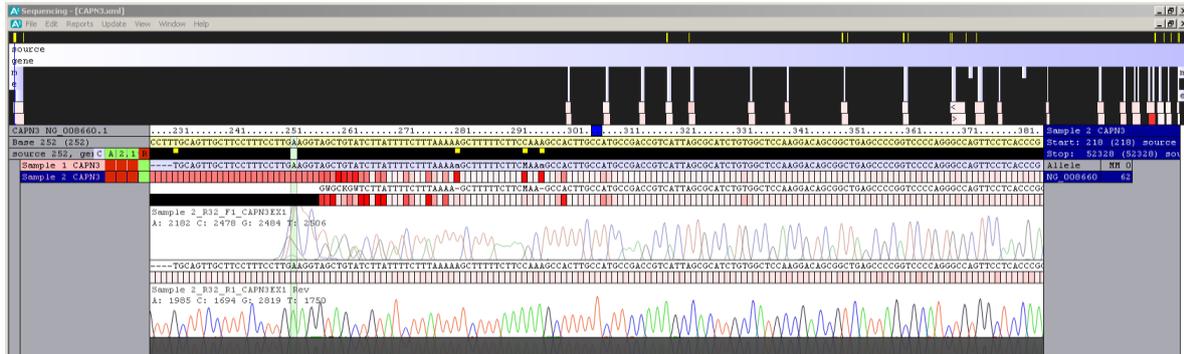
It is likely that the Assign™ ATF 1.5 **Auto Edit** function is on. In this case Assign™ ATF 1.5 will perform a base call with a significant bias to what is “expected”. The Auto edit function greatly assists base calling when Assign™ ATF 1.5 is used for genotyping highly polymorphic genes. The Auto edit function can be turned off in **Settings > Engine**. Refer to section 6.2.3.

Auto Edit is NOT recommended for Variant detection

For more information please contact Conexio Genomics at support@conexio-genomics.com.

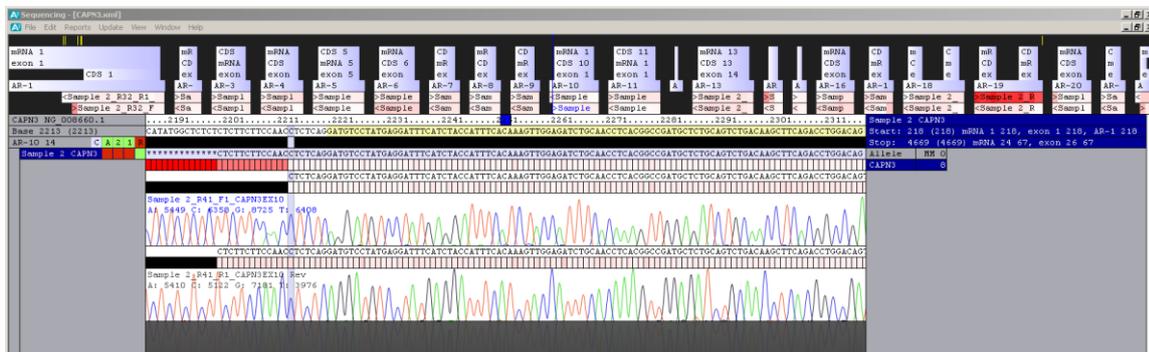
Appendix 2: Why Should I Use Reference Manager?

Assign™ ATF 1.5 Analysis WITHOUT using Reference Manager



- The coding group has to be created **manually** which can be time consuming (*not evident from the figure*).
- Within the gene structure panel regions are minimised and valuable quality information is not visible.
- The Allele name in the **Results Pane** is taken as the GenBank Accession number.
- High number of sequence mismatches with the reference due to the inclusion of poor quality sequence where Forward and Reverse sequence EPGs overlap.

Assign™ ATF 1.5 Analysis WITH using Reference Manager



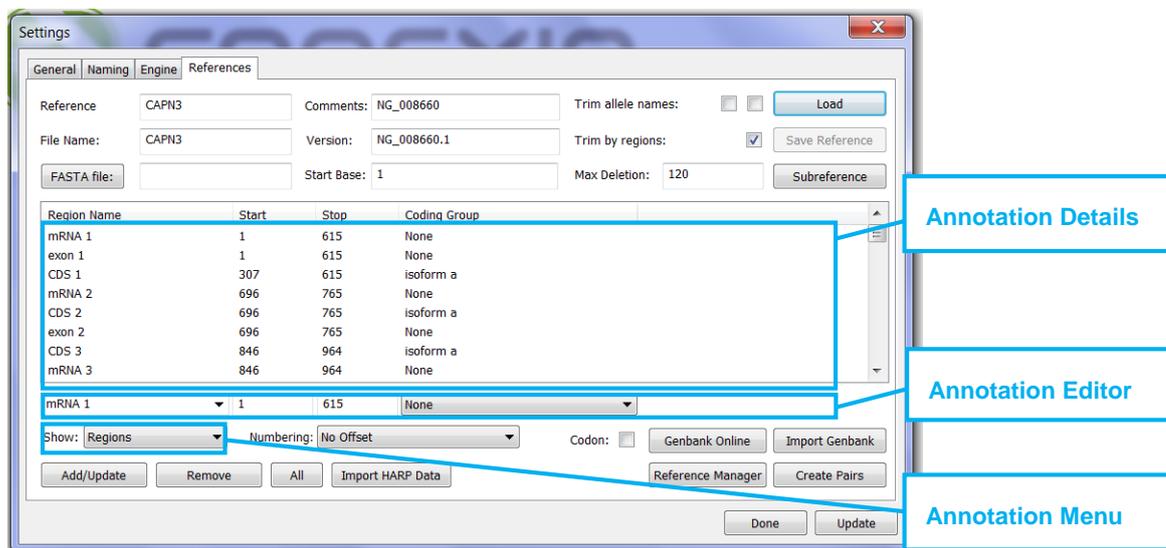
- The coding group has been created **automatically** which saves considerable time (*not evident from the figure*).
- Unused intron information is removed.
- EPG quality information can be interpreted.
- A lower number of mismatches are present and their locations are easily found.
- The Allele Name in the Results Pane has the name of the gene.

The CAPN3 reference in the figure can be further improved by adjusting the AR-1 region to remove the poor quality sequence resulting in mismatches in the 5 UTR. This can be done under **Edit > Settings > References**.

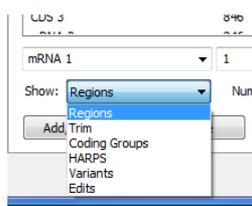
Appendix 3: Manual Editing of the Reference Sequence

The Reference Manager is used to automatically create Coding Group and Analysis Regions and results in standard settings for each region. Manual editing can be used to edit a specific region, or if for some reason you chose not to use the Reference Manager. *Manual editing is generally not required if Reference Manager is used.*

- A3.1 Within the **Edit > Settings >References** window open the reference file you wish to edit by clicking **Load**. Browse to the appropriate reference (.xml) file in the **ATF > References** folder, select it and click **Open**.



- A3.1.1 The large window contains the sequence **Annotation Details** for the item selected in the **Annotation Menu**.
- A3.1.2 Reference sequence details can be edited or entered in the **Annotation Editor**.
- A3.1.3 Use the **Annotation Menu** (**Show** drop-down menu) to select the items to be displayed in the **Annotation Window** and the **Annotation Editor**.



- A3.1.3.1 **Regions** enables the different regions within the reference sequence to be annotated. These can be overlapping.
- ◆ To add regions, choose **Regions** from the **Show** drop down menu.
 - ◆ Enter the name of the region in the box above the **Show** drop down menu e.g. 5UTR.
 - ◆ Enter the Region Start position in **Start** box and the Region end position in the **End** box. (Number the regions so that base 1 is the first base of the reference sequence).
 - ◆ Click on **Add/Update**. Perform this process for all regions.
 - ◆ Importing GenBank entries may result in many redundant and un-required regions. Several regions can be removed by typing the first few letters of the coding regions to be deleted in the left hand box in the Annotation

Editor and clicking **All** (located in the bottom line). A window will pop up indicating how many regions were removed.

- ◆ Once all regions have been edited click **Update** (bottom right hand corner) to save the changes.
- ◆ To annotate the 5' UTR as minus numbers before the start codon, enter the appropriate **Start Base** and **Update** (bottom right hand corner). To view the alternative numbering systems select between **With Offset** and **No Offset in Numbering**.

A3.1.3.2 **Trim** enables sequencing or PCR primer locations at the beginning of Regions to be excluded from the analysis (i.e. in the case where a primer site is within an exon). The user can choose not to analyse sequences at amplification primer sites if these sequences are included in the reference sequence. The **Trim** function allows the removal of sequence at the beginning or ends of regions.

E.g. the 5' PCR amplification primer is 23 bases in length and is located at the beginning of the 5'UTR. This region can be excluded from analysis by **Trimming** the length of the PCR amplification primer region.

- ◆ Select **Trim** from the **Show** drop down menu.
- ◆ Select the required region to be **Trimmed** in the **Trim region** drop down menu (**5UTR**).
- ◆ Enter the number of bases required to be **Trimmed** from the **Start** (**23**).
- ◆ Click **Add/Update** to register the changes.
- ◆ Click **Update** (bottom right hand corner) to save the changes.

A3.1.3.3 **Coding Groups** enables coding regions to be annotated. A Coding Group can be a single region or consist of several linked regions. Once the regions (e.g. exons) have been annotated in the reference sequence, common regions can be grouped to create a continuous string of sequence. For example exons can be grouped to form the coding sequence. This information is incorporated into variant reports to identify if variants result in amino acid changes.

- ◆ Select **Coding Groups** from the **Show** drop down menu.
- ◆ Enter the name of a new coding group and enter the start base.
- ◆ Select the regions to be added from the **Members** drop down menu.
- ◆ If this is a coding region, select **Yes** from the **Coding** drop down menu.
- ◆ If the coding region is in the 3'-5' (reverse) orientation of the sequence select **Yes** from the **Reverse** drop down menu.
- ◆ Click **Add/Update** to register the changes.
- ◆ To add more regions to the **Coding Group**. Select it from the drop down menu, select the next region from the **Members** drop down menu.
- ◆ Click **Add/Update**. Repeat until all members have been added.
- ◆ Click **Update** (bottom right hand corner) to save the changes.

A3.1.3.4 **Variants** enables known sequence variants to be added to the reference. Refer to section 6.3.3 for details.

A3.1.3.5 **Edits** enables editing of positions in different frames of the reference. This can be performed with nucleotide or codon numbering.

*Note that **Layer** is a HLA specific application and is not commonly used. Please contact Conexio Genomics if you require further information.*

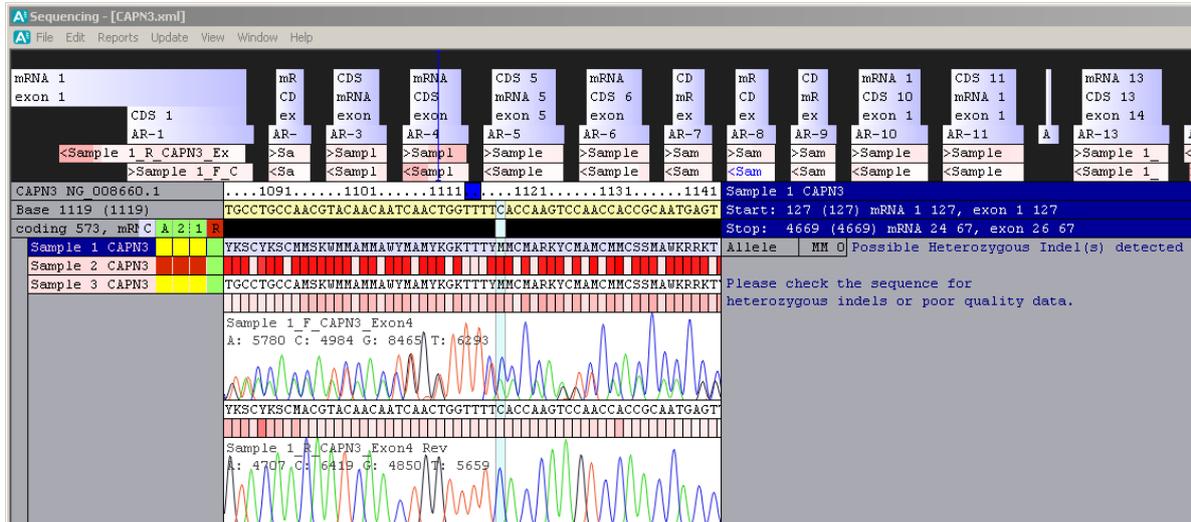
Appendix 4: Analysis Settings and Sequencing Applications

| Assign™ ATF Analysis Parameters | | Application* | | | |
|------------------------------------|-----------------------|--------------|----------------------|-------|-------------------------|
| | | Genotyping | Variant Detection | Clone | Anonymous Sequencing |
| Matching Mode | Variant Detection | No | Yes | Yes | Yes |
| | Genotyping | Yes | No | No | No |
| | No Mixed Bases | No | No | Yes | No |
| Base Caller | Apply Height Maps | Yes | Yes | No | No |
| | Update Height Maps | Yes | Yes | No | No |
| | Apply Auto Editing | Yes | No | No | No |

*Application Definitions

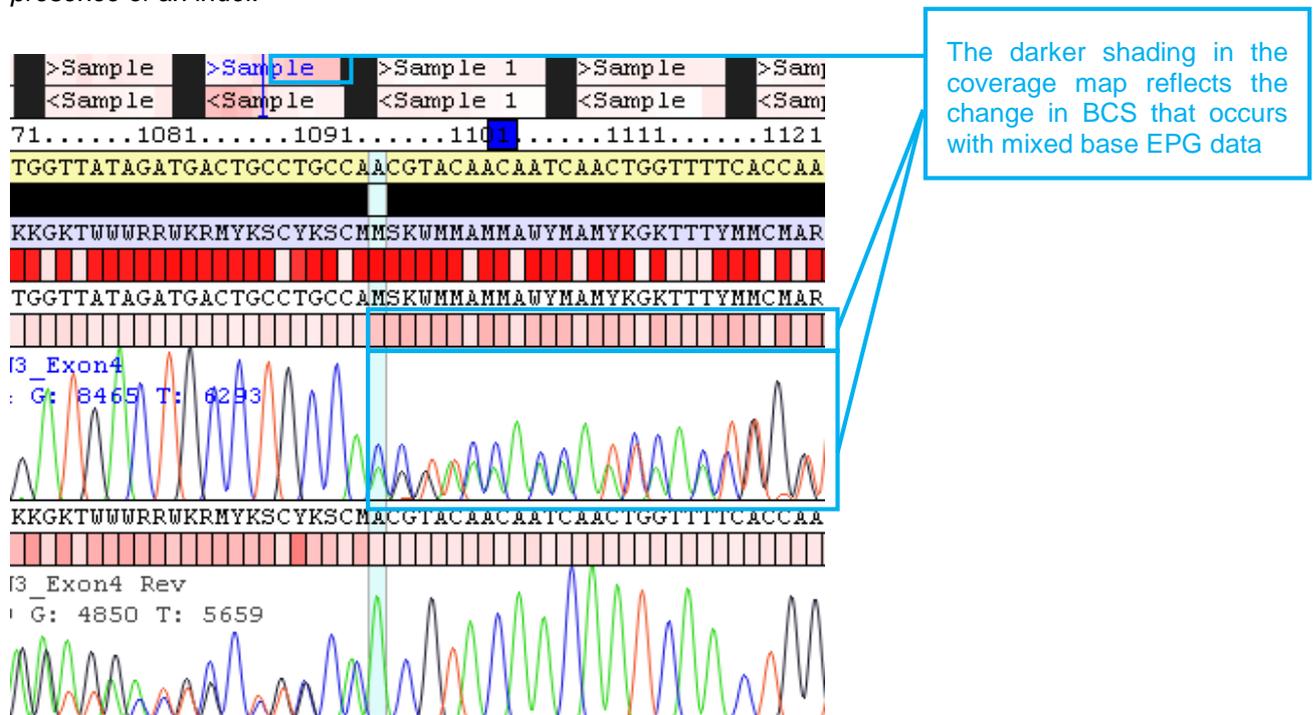
- **Genotyping:** Applications including the comparison of a test sequence with a library of sequences of variants (alleles) for the locus being sequenced.
- **Variant Detection:** Applications including SNP discovery, variants in genes associated with genetic disorders, and viral variants associated with drug resistance.
- **Clone:** Clone sequencing and contig assembly.
- **Anonymous Sequencing:** Can include sequencing clones or PCR products where a reference sequence does not exist.

Appendix 5: Heterozygous Insertion/Deletion and Poor Quality Sequence Warnings



Assign™ ATF is able to identify and analyse heterozygous insertion / deletion (indel) data and, in many cases, correctly report the inserted or deleted sequence.

Assign™ ATF contains warnings in the Results Pane that alerts the user that an indel may be present. The location of the indel can usually be located by reviewing the sequence coverage data. A defined change in shade (indicating a change from good to poor quality sequence) usually indicates the the presence of an indel.



Recommended Approaches for Analysing Heterozygous Indel Sequences

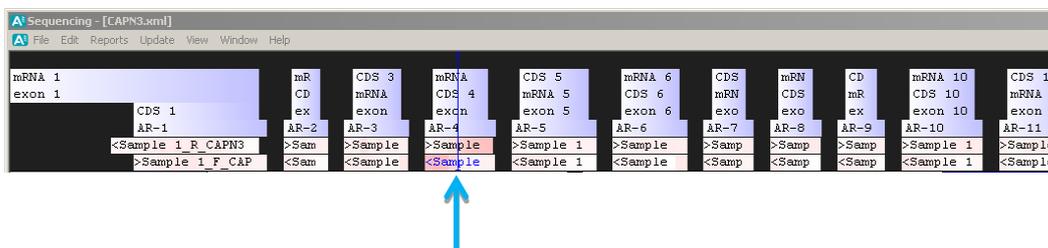
Poor quality sequence data and heterozygous indel data may give similar profiles and distinguishing between the two is difficult for ALL sequence analysis software.

Typically, when an indel is present Assign™ ATF issues a warning in the Results Pane and the number of mismatches will NOT be reported.

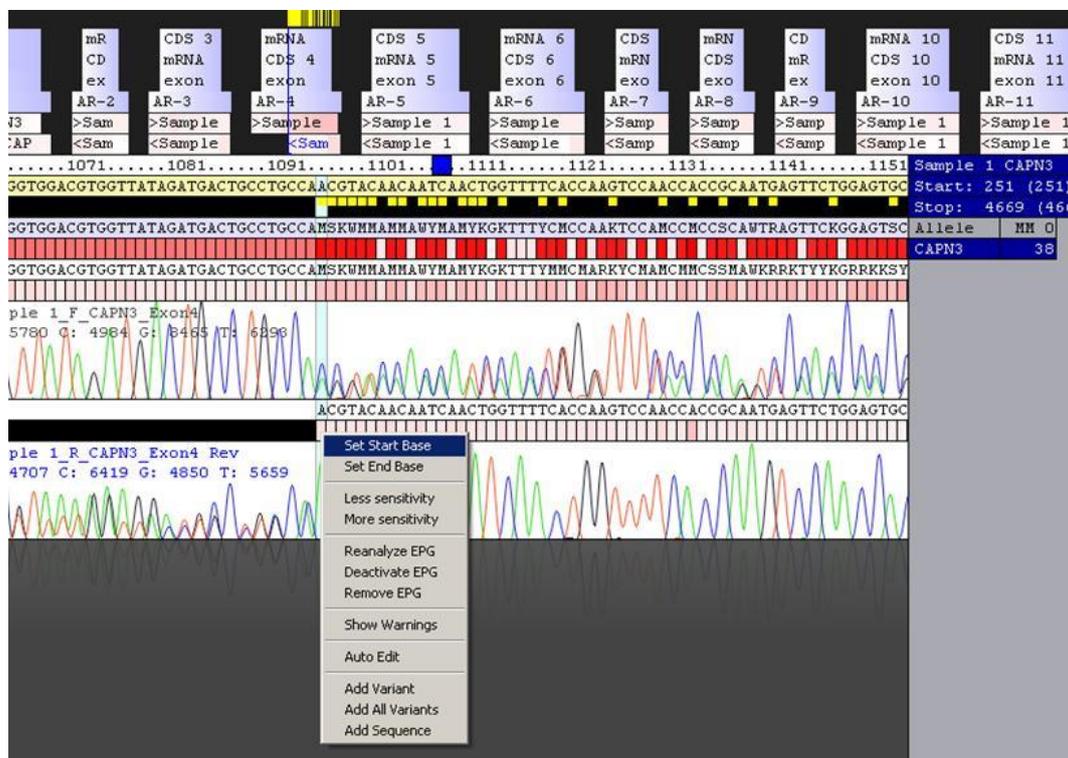
```
Sample 1 CAPN3
Start: 251 (251) mRNA 1 251, exon 1 251
Stop: 4669 (4669) mRNA 24 67, exon 26 67
Allele MM 0 Possible Heterozygous Indel(s) detected

Please check the sequence for
heterozygous indels or poor quality data.
```

- Review the sequence coverage map to determine the likely position of the indel.



- Review the sequence EPG data at this position.
- On the reverse strand sequence data, click to highlight the base at the start of the indel data, right click the mouse to reveal the menu and select **Set Start Base**.



In the example above, the software registers mismatches in both the Results Pane and above the sample consensus sequence. This indicates that base calling of the complex mixed bases is not completely correct and that the software has not yet registered the data as a heterozygous indel.

- If the user considers that the data is typical of a heterozygous indel, the quickest way to get the software to register the indel data is to trim the forward sequence by highlighting the relevant base, right clicking to reveal the menu and selecting **Set End Base**. Repeat this, if necessary, until the yellow mismatch markers above the indel sequence disappear. This indicates that the software has stopped seeing the nucleotide sequence mismatches as individual mismatches and an indel must be present.

The screenshot displays a sequence alignment interface. At the top, there are labels for mRNA and CDS regions (4-11) and AR regions (AR-4 to AR-11). Below these are sequence alignment tracks for Sample 1. A yellow highlight is placed over a region of the sequence: `GGTGGACGTGGTTATAGATGACTGCCTGCCA` in the sample and `GGTGGACGTGGTTATAGATGACTGCCTGCCAMSKWMMAMMAWYMAACTGGTTTTCCACCAAGTCCAACC` in the reference. A context menu is open over the sample sequence, with the option **Set End Base** highlighted. Other menu options include 'Less sensitivity', 'More sensitivity', 'Reanalyze EPG', 'Deactivate EPG', 'Remove EPG', 'Show Warnings', 'Auto Edit', 'Add Variant', 'Add All Variants', and 'Add Sequence'. Below the sequence is an EPG (Electropherogram) plot showing signal intensity for each base. A peak is labeled at position 6293. The variant report on the right shows 'Sample 1 CAPN3' with 'Start: 239 (239)', 'Stop: 4669 (4669)', 'Allele: CAPN3', and 'MM: 3'.

The Variant report will now include the indel.

Support and Contact Details

Conexio Genomics Pty Ltd

8/31 Pakenham St
Fremantle 6160
Western Australia

Tel: +61 422 863 227

email: support@conexio-genomics.com

Skype: conexiocgx

Website: www.conexio-genomics.com

Conexio™ and Assign™ are trademarks of Conexio 4 Pty Ltd.

For ordering details, please refer to the Conexio website <http://www.conexio-genomics.com> or email support@conexio-genomics.com