



Omixon handbook

Omixon HLA Twin CE
Version 3.1.1

Date of issue:
10/17/2018

1 Table of Contents

1	Table of Contents	2
2	Company information.....	9
3	Extended Installation Guide	10
3.1	Windows Installation Guide.....	10
3.1.1	Installation steps.....	10
3.1.2	Uninstalling the product.....	10
3.2	Linux Installation Guide.....	10
3.2.1	Installation steps.....	11
3.2.2	Uninstalling the product.....	11
3.3	Mac OS X Installation Guide.....	11
3.3.1	Installation steps.....	11
3.3.2	Uninstalling the product.....	12
4	Genotyping dashboard	13
4.1	General overview	13
4.2	Top panel.....	13
4.2.1	Top row.....	13
4.2.2	Middle row	13
4.2.3	Bottom row	14
4.3	Genotyping and other functions	14
4.3.1	Typing and data analysis options	14
4.3.2	File browser functions	14
4.3.3	Application Settings.....	14
4.4	File browser	14
4.4.1	Navigation in the file browser	14
4.4.2	Samples and analyses	15
4.5	Traffic lights.....	15
4.6	Context menu.....	15
4.7	Simple Genotyping	15
4.8	Advanced Genotyping.....	16
4.8.1	Introduction	16
4.8.2	Protocol	16

4.8.3	Advanced Options.....	16
4.9	Genotyping Reanalysis	17
4.10	Export results	17
4.10.1	General information.....	17
4.10.2	HML export	17
	General information.....	17
	HML export wizard	18
4.10.3	PDF export.....	18
4.10.4	JSON export	19
	Export options.....	19
	Export format and content	19
4.10.5	HPRIM export.....	21
	Filter analysis results	21
	Select analysis results.....	22
	Export configuration	22
5	Genotyping analysis result	23
5.1	Introduction	23
5.1.1	Commenting.....	23
5.2	Top panel.....	23
5.2.1	Top row.....	23
5.2.2	Middle row	23
5.2.3	Bottom row	24
5.3	Interpreting Results	24
5.3.1	Troubleshooting missing results.....	24
5.4	Filtering Alleles.....	25
5.5	Traffic Lights.....	25
5.5.1	Concordance light.....	25
5.5.2	Quality control light	25
5.6	Assigning Alleles.....	26
5.7	Export table	26
5.8	Linkage disequilibrium details	27
5.8.1	Overview	27
5.8.2	Case by case guide	27
5.8.3	Notes.....	28

5.8.4	DQA1, DQB1 allele dropout warning.....	28
6	Genotyping sample result	29
6.1	Top panel.....	29
6.1.1	Top row.....	29
6.1.2	Middle row.....	29
6.1.3	Bottom row	29
6.2	Available functions	30
6.2.1	Opening the browser	30
6.2.2	Detailed genotyping information.....	30
6.2.3	Customizing displayed results	30
6.2.4	Assigning alleles.....	30
6.2.5	Commenting.....	30
6.3	Genotype	30
6.3.1	Tree view	30
6.3.2	Troubleshooting missing results.....	31
6.3.3	Adding alleles manually	31
6.3.4	Removing additional alleles.....	31
6.4	Quality Control.....	31
6.5	Data statistics.....	32
6.5.1	Overview.....	32
6.5.2	Allele imbalance.....	33
6.5.3	Fragment size.....	33
6.5.4	Read quality	33
6.5.5	Troubleshooting missing results.....	31
6.6	Genotyping mismatch result.....	34
6.7	Genotyping result novelties	34
6.8	Allele frequency information.....	34
7	Gene Browser	35
7.1	Introduction	35
7.2	Top panel.....	35
7.2.1	Top row.....	35
7.2.2	Middle row.....	36
7.2.3	Bottom row	36

7.3	Settings and Functions	36
7.3.1	Filtering the Allele Candidates	36
7.3.2	Exporting consensus sequences	36
7.3.3	Zooming	37
7.3.4	Jumping around.....	37
7.3.5	Changing short read track settings.....	37
7.3.6	Track management.....	37
7.3.7	Display setup	37
7.3.8	Exporting Gene browser data.....	37
7.3.9	Rotating the browser	37
7.4	Context menu.....	37
7.5	Metadata Panel	38
7.5.1	Metadata Panel functions in the Gene Browser:	38
7.6	Shortcuts	38
7.7	Manage tracks	39
8	Settings dashboard.....	40
8.1	General information.....	40
8.2	Sidebar	40
8.2.1	General	40
8.2.2	Database.....	40
8.2.3	Administration	40
8.2.4	Automation	40
8.2.5	Export Settings.....	40
8.2.6	Screen Settings	40
8.3	Sample and Result Folders.....	41
8.4	Analysis Protocols.....	41
8.4.1	Configure protocol.....	41
8.4.2	Create protocol	41
	Introduction	41
	General	41
	Sample data type.....	42
	Analysis options	42
8.5	Install New Database	42
8.5.1	Download problems	42

8.6	Select Active Database	43
8.7	Remove Database	43
8.8	Configure Database Extensions	43
8.9	User Management	44
8.10	Upload Licence.....	44
8.11	Typer Server Management	44
8.12	Configure Automatic Analysis	44
8.13	HPRIM Export Configuration	45
8.14	Gene Browser	45
8.14.1	Display configuration.....	45
8.14.2	Color setup	45
8.14.3	Data options	45
8.14.4	Metadata options.....	45
8.14.5	Advanced settings.....	45
8.15	Analysis Result Screens	45
9	Process manager.....	46
9.1	Colour coding.....	46
9.2	Handling Tasks.....	46
9.2.1	Abort Task	46
9.2.2	Delete Task	46
9.2.3	Pause and Resume Task.....	46
9.2.4	Find a Task.....	46
10	Detailed automation guide	47
10.1	Overview.....	47
10.2	Configuration	47
10.3	Management	48
10.4	Logging	48
11	Notifications.....	49
11.1	IMGT Database	49
11.2	License Expiry.....	49
11.3	Extended allele notification	49
11.4	Deviating from the Factory default protocol.....	49

12	HTTPS support	50
13	Command Line Interface	53
13.1	Overview	53
13.2	How does it work?	53
13.3	Command types	53
13.3.1	Initialize	53
13.3.2	License file import	53
13.3.3	HLA Database listing	54
13.3.4	HLA Database installation	54
13.3.5	Change HLA Database version	54
13.3.6	List assay versions	55
13.3.7	Set assay version	55
13.3.8	Python script execution	55
13.3.9	Status of the server tasks	56
14	Omixon Twin Admin command line interface (Admin CLI)	57
14.1	Important notes	57
14.2	Password reset	58
14.3	Cleaning licences	58
14.4	Activate disabled user	58
15	Application performance tuning	60
16	Omixon database relocation	61
17	Glossary	62
18	List of shortcuts	63
18.1	Generic shortcuts	63
18.2	Genotyping dashboard	63
18.3	Genotyping analysis result	63
18.4	Genotyping sample result	63
18.5	Gene browser	63
19	Acknowledgements	65
19.1	Collaborators	65
19.2	Third party tools and databases	65

19.3 Citations 65



2 Company information

This product is manufactured by Omixon Biocomputing Ltd.

Address:

H-1117 Budapest
Fehérvári út 50-52.
Hungary, EU

Website: <http://www.omixon.com>

Technical contact: support@omixon.com¹

Sales contact: sales@omixon.com²

¹ <mailto:support@omixon.com>

² <mailto:sales@omixon.com>



3 Extended Installation Guide

3.1 Windows Installation Guide

We provide an installer package for Windows operating systems bundled with a Java Runtime Environment (JRE).

3.1.1 Installation steps

Launch the executable.

Welcome Dialog:

- Choose "No, install to a different directory" to perform a clean installation.
- Select "Yes, update the existing installation" to upgrade a previously installed version of Omixon HLA and preserve your results.

License Agreement: Please, accept the software license agreement to proceed with the installation.

Select Destination Directory: Select the path where you would like the application to be installed.

Data Directory: Select data directory where the permanent data files will be stored. In case of an upgrade please do not change the proposed path - it has to point to the previously used database folder.

Temp Directory: Select temporary files directory.

Memory Setup: Set the maximum amount of memory to use for the application. In case of a server installation you can set the memory values for the Omixon Server and the Typer Server separately. Chosen values should fall between the minimum required and maximum available values shown on the screen.

Network Connection Setup (Typer Server installation only): Here you can set the IP address and port where the *Omixon HLA Typer Server* should connect to the *Omixon HLA Server*.

Select Start Menu Folder: Enter name for Start Menu item. Use the checkboxes to make the Start Menu folder visible for all users of the operating system or to skip this step.

Completing Setup: After clicking the Next button, the installation process starts. When it is finished, the Completing the Omixon HLA Setup dialog indicates the success of the installation.

Launching the application:

If you've decided to create a start menu folder in the previous step, *Omixon HLA* shortcut is available now. The application can be started by clicking on the startup icon or running the executable from application directory.

In Windows 8, and 8.1 there is no Start menu implemented, to run Omixon HLA application, press Windows key+s, then in the search field start typing " Omixon ", the application will pop up in results frame, then click to run it.

3.1.2 Uninstalling the product

If you want to remove the application, an uninstaller executable is available in the directory where the application was installed. If you have created Start Menu items during installation, *Uninstall Omixon HLA* icon launches the uninstaller.

If you intend to use future versions and don't want your work data to be deleted, keep the checkbox unchecked. If you wish to cleanup the working directory, then select the option 'Yes'.

Click 'Next' to proceed with the uninstallation process.

Finally, the installer informs you about the outcome of uninstalling the software.

3.2 Linux Installation Guide

We provide an installer package for Linux operating systems bundled with a Java Runtime Environment (JRE):

The install packages are single file shell scripts, suitable to install on various Linux distributions.

The downloaded installer does not have the permissions to run directly. Open a terminal window to make it executable with the following command:

```
chmod +x installer_name
```

After that, it can be started with the following command:



./installer_name

3.2.1 Installation steps

Launch the executable.

Welcome Dialog:

- Choose "No, install to a different directory" to perform a clean installation.
- Select "Yes, update the existing installation" to upgrade a previously installed version of Omixon HLA and preserve your results.

License Agreement: Please, accept the software license agreement to proceed with the installation.

Select Destination Directory: Select the path where you would like the application to be installed.

Data Directory: Select data directory where the permanent data files will be stored. In case of an upgrade please do not change the proposed path - it has to point to the previously used database folder.

Temp Directory: Select temporary files directory.

Memory Setup: Set the maximum amount of memory to use for the application. In case of a server installation you can set the memory values for the Omixon Server and the Typer Server separately. Chosen values should fall between the minimum required and maximum available values shown on the screen.

Network Connection Setup (Typer Server installation only): Here you can set the IP address and port where the *Omixon HLA Typer Server* should connect to the *Omixon HLA Server*.

Select Directory for Symlinks: Set symlink location. It is useful for users who start the application from the command line. Since the installer is creating a startup icon in the System menu, creating symlinks is optional for most users.

Completing Setup: After clicking the Next button, the installation process starts. When it is finished, the Completing the Omixon HLA Setup dialog indicates the success of the installation.

Launching the application: An application icon is placed into the System menu. By clicking on this icon, the application starts and is ready for use.

3.2.2 Uninstalling the product

If you want to remove the application, an uninstaller executable is available in the directory where the application was installed. If you have created Start Menu items during installation, *Uninstall Omixon HLA* icon launches the uninstaller.

If you intend to use future versions and don't want your work data to be deleted, keep the checkbox unchecked. If you wish to cleanup the working directory, then select the option 'Yes'.

Click 'Next' to proceed with the uninstallation process.

Finally, the installer informs you about the outcome of uninstalling the software.

3.3 Mac OS X Installation Guide

As Mac OS X is an exclusively 64-bit operating system and the Java Runtime Environment is updated regularly the installer does not contain a Java Runtime Environment (JRE).

This avoids any conflicts among the different JREs but may cause incompatibility problems. The required JRE version is: 1.8

If you don't have Java installed on your Mac, it will be automatically downloaded during installation.

3.3.1 Installation steps

Open the installer which is packed within a DMG archive. Click on the *Omixon HLA Installer.app* icon.

Welcome Dialog:

- Choose "No, install to a different directory" to perform a clean installation.
- Select "Yes, update the existing installation" to upgrade a previously installed version of Omixon HLA and preserve your results.

License Agreement: Please, accept the software license agreement to proceed with the installation.

Select Destination Directory: Please use the provided default path to install the application.

Data Directory: Please use the default location for the data directory where the permanent data - eg. results - will be stored.

Temp Directory: Select temporary files directory.



Memory Setup: Set the maximum amount of memory to use for the application. In case of a server installation you can set the memory values for the Omixon Server and the Typet Server separately. Chosen values should fall between the minimum required and maximum available values shown on the screen.

Completing Setup: After clicking the Next button in the previous step, the installation process starts. When it is finished, the Completing the Omixon HLA Setup dialog indicates the success of the installation.

Launching the application: In Application list, *Omixon HLA* is available now. By clicking the *Omixon-HLA* icon, application starts and is ready for use.

3.3.2 Uninstalling the product

If you want to remove the application, an uninstaller executable is available in the directory where the application was installed. If you have created Start Menu items during installation, *Uninstall Omixon HLA* icon launches the uninstaller.

If you intend to use future versions and don't want your work data to be deleted, keep the checkbox unchecked. If you wish to cleanup the working directory, then select the option 'Yes'.

Click 'Next' to proceed with the uninstallation process.

Finally, the installer informs you about the outcome of uninstalling the software.

4 Genotyping dashboard

4.1 General overview

This is the home dashboard of the software. All Genotyping functions are available from here.

The dashboard consists of the following subscreens:

- Top panel: this contains all the main functions and some high level information about the current folder and sample. It also shows information about the current user and active database and provides some navigation functions.
- File browser: this part of the screen can be used for navigation between all accessible folders.

4.2 Top panel

On the top of the screen you can see three rows of functions.

4.2.1 Top row

On the left, you can find the *Toggle fullscreen* button which hides the less important parts of the current screen, while on the right you can see the following things:

- the ID of the current user,
- the memory usage widget panel,
- the status panel of the process manager,
- the welcome tutorial button,
- the logout button,
- and the exit button.

In case of the desktop version, the actual memory usage can be observed at the top right corner of the screen. This displays the amount of currently used memory compared to the allowed maximum. Note, that the software might be allocating more memory from the operating system compared to the reported actual memory usage to make further computations faster by keeping some previously allocated memory pieces for later use, thus making the allocation of new memory unnecessary during further computations. By clicking on the memory report widget the software tries to free up as much memory as possible on a best effort basis: this is only an attempt, there are many technical factors which might block returning all memory to the operating system. It is also important to mention that by explicitly asking to free up memory the subsequent computations might become slightly slower, since the option of reusing certain structures allocated before becomes impossible. Only use this option when all planned analyses have been finished and you would like to use other programs more intensively in parallel while investigating the analysis results.

4.2.2 Middle row

There are four navigation buttons that are available on every screen of the tool. These general navigation buttons work very similar to navigation functions available in most widely used internet browsers:

- The *Back* button will take you to the previous screen.
- The *Forward* button will take you to the next screen.
- The *Up* button will move you one level up in the application.
- The *Home* button will take you to the starting screen (i.e. the Genotyping dashboard).

Right from the navigation button, in the information panel, you can find relevant high-level information about the current screen. On the *Genotyping dashboard*, you can see the release date and version of the active IMGT database.

On the far right hand side of the middle row, you can find the bookmark and context specific help buttons. The help tool shows relevant usage information and tips about functions and data available on the current screen of the application and can also be opened by pressing F1.

You can search the Help for any expression by using the Magnifier glass icon displayed on the popup window. Relevant hits are displayed in the left panel where you can select them and open them in the right hand reading pane. You can close the search by clicking the X displayed beside the search field.

4.2.3 Bottom row

The bottom row of the top panel contains a series of buttons which contain the main functions available on the screen.

4.3 Genotyping and other functions

4.3.1 Typing and data analysis options

- **Simple Genotyping** - This function starts the standard analysis with the Factory default protocol which is fully configured for the Omixon Holotype assay and is the only typing function available for all users.
- **Advanced Genotyping** - This function starts the Advanced Genotyping wizard which is only available for users with superuser privileges. This advanced wizard allows the setting of several advanced parameters for the Genotyping algorithms.
- **View Results** - Leads to the more detailed Genotyping analysis result screen which is suitable for working with multiple genotyping results.
- **View Details** - Opens the Genotyping sample result screen where locus level results, detailed QC metrics and data statistics can be reviewed for the selected genotyping result.
- **Export Results** - By pressing this button, the Export results wizard can be opened. Using this wizard, genotyping results can be exported in the standard HML, PDF or JSON format.
- **Re-Analyse** - This function allows the re-analysis of previously typed samples.

Note
When you create a new protocol and set it as default, after pressing the Simple Genotyping or the Analysis button, you will be warned about the deviation from the factory default protocol. The reason for the warning is to avoid clinical usage of results which are for research use only.

4.3.2 File browser functions

- **Sample Filter** - Toggles the file browser between showing and hiding samples.
- **Analysis Filter** - Toggles the file browser between showing all analyses, showing the most recent analysis per sample and hiding all analyses.
- **Analysis State Filter** - Toggles the file browser between showing all analyses or analyses that are in progress, waiting for approval or approved.
- **Filter Files** - Opens the *Filter analyses and samples* wizard which can be used for filtering the samples/analyses in the current folder using simple string matching, string matching with wildcards or regular expressions.
- **Sort Files** - Sorts samples and analyses by name.

4.3.3 Application Settings

This function opens the Settings dashboard, where tool-wide settings and administrative functions can be found.

4.4 File browser

4.4.1 Navigation in the file browser

The file browser works very similarly to a high number of well known file browsers. Navigation is single click based: folders can be entered using a single left mouse click. Moving up to the parent folder can be achieved by left clicking on the top ".." folder or by using the folder breadcrumbs at the top of the file browser.

Folders without read permission are shown in red color.

4.4.2 Samples and analyses

Only short read files (in the supported file formats) and genotyping results (with a htr extension) are shown in the file browser. Paired-end or mate-pair fastq files are automatically paired and the sample is represented by the R1 file within the file browser to avoid redundancies and make navigation easier. For each sample, the pairing status and total size of the short read files are shown. On the right hand side of each sample row you can find the *Results* and *Details* buttons which provide shortcuts for the *Genotyping analysis result* and *Genotyping sample result* views, respectively.

Those analyses which were produced using Custom analysis - typing parameters have been altered from the default protocol - are displayed with red font.

The same markup is used for those results where audit details are not available. Audit details are stored related to the Task displayed in the Process manager which represents the typing. It is possible to lose this data if the database had been deleted due to an error, damaged or if the result had been produced with a previous version of the software which allowed removing these details.

4.5 Traffic lights

Sample level results are annotated using a "traffic light" system with the following output:

- (green): all QC tests gave passed or info results and the genotyping results of the two algorithms were fully concordant.
- (red): at least one of the QC measures failed and/or the concordance between the two algorithms was under two fields.
- (yellow): in all other cases, a yellow light is shown.

4.6 Context menu

Like in most parts of the tool, a context specific menu, containing additional functions can be opened using the right button of the mouse (or touchpad). On the *Genotyping dashboard* the following functions are available in this right-click menu:

- *Toggle select all*: selects or deselects all files in the current browser.
- *Filter files*: opens the file filter wizard.
- *Reduce/Expand file name*: shortens sample and analysis names using regular expressions.
- *View selected results*: takes you to the *HLA Typing analysis* result dashboard of the selected analyses.
- *Show protocol*: shows the parameters used for the selected analysis.
- *Save as protocol*: saves the parameters used for the selected analysis
- *Delete selected items*: deletes the selected items.
- *Move selected items*: moves selected samples and analyses to a new location on the file system.
- *Rename selected result*: allows the user to rename the selected result file.
- *Copy selected result original name*: copies the full name of the analysis to the clipboard.
- *Create new folder*: creates new folder on the file system.
- *Cut/Copy/Paste*: cuts, copies or pastes selected samples and analyses.

4.7 Simple Genotyping

This function allows you to analyse data produced with the Omixon Holotype assay. All parameters are pre-set and fine-tuned for this assay. Datatype is autodetected at the beginning of the typing and the amount of data processed is adjusted accordingly. For running genotyping with other settings, please use the *Advanced Genotyping* function. You need to select files for the input. This function will allow you to *process multiple samples at the same time*, i.e. multiple pairs of fastq (or fastq.gz) files. The two input files in a file pair are assumed to have the exact matching reads, in the exact same order (there is currently no built-in check for this). A background task will be started, and once it's finished a new item will appear in the main list on the *Genotyping dashboard* where you will be able to see the result.

Note
When you create a new protocol and set is as default, after pressing the Simple Genotyping or the Analysis button, you will be warned about the deviation from the factory default protocol. The reason for the warning is to avoid clinical usage of results which are for research use only.

4.8 Advanced Genotyping

4.8.1 Introduction

This function is available from the Genotyping dashboard.

When an analysis is started with the Advanced Genotyping option a warning will be displayed to alarm the user about potential deviation from the Factory Default Protocol. The reason for the warning is to avoid clinical usage of results which are for research use only.

First you need to select one or multiple samples to analyse. This function allows you to process multiple samples at the same time, that is, multiple pairs of fastq (or fastq.gz) files. The two input files in a file pair are assumed to have the exact matching reads in the exact same order (currently there is no in-built check for this). Before the analysis you also need to select which sequencer created the short read data. After starting the analysis, a background task will be started. When the task is finished, the result will appear in the main list on the *Genotyping dashboard*.

4.8.2 Protocol

Protocols are stored sets of Genotyping analysis parameters. Advanced Genotyping allows you to select stored protocols. If a particular set of parameters work well for your samples, you can save these as a protocol and apply it to other samples later. Selecting an existing protocol can speed up the usage of the Advanced Genotyping wizard, as most of the parameters are already preset in the protocol. Protocols can be copied, edited, removed and created at any time.

4.8.3 Advanced Options

Analysis options

- *Maximum pairs processed per locus*: Twin uses locus based subsampling to optimize results. This value determines how many reads will be used for each locus. The datatype is assessed at the beginning of each typing and the amount of reads processed is adjusted accordingly. Note, that only reads longer than 75 bases are selected during subsampling (for paired data, both reads in the pair have to be over this minimum length threshold). You can experiment with this value to find the best match for your data. For samples produced with the Holotype assay and other targeted data sets the default 4000 read pairs per locus works well. In case you run into problematic samples increasing the read count can help resolve some of these issues.
- *Disable quality based subsampling*: If this checkbox is checked, subsampling will be based on the per locus read count and the number and type of targeted loci. This option is used for generating mappability statistics.
- *Ignore rare alleles*: Using this option, very rare alleles will be ignored in the analysis and will not be included or reported. If this checkbox is unchecked, rare alleles will be marked in the results. If this option is greyed out, you need to install a new database in order to import the rare alleles.
- *Alternative method execution*: You can set the alternative method execution based on the QC report status or it can be switched off. Using this option the analysis performance can be improved with high quality samples. **Always** option runs the alternative typing method always. **QC inspect/investigate/failed** option runs alternative typing method for those loci where QC reported inspect/investigate/failed status and for loci with novel alleles regardless of the QC status (this is the default configuration). **QC failed** option runs alternative typing method for those loci where QC failed and for loci with novel alleles regardless of the QC status. **Never** option turns off the alternative typing method completely.

- **Novel allele detection:** This feature detects SNPs and indels in the short read allele alignments. Based on the detected variants and the original reference sequence of the allele, novel allele sequences are generated. The name of a novel allele candidate consists of the name of the original allele plus a hashtag and a number. For example, HLA-A*01:01:01:01#1 is a novel allele generated based on the sequence of HLA-A*01:01:01:01 and one or more SNPs found in the short read alignment of this allele. Novel alleles have a double reference track in the Gene Browser, which means that both the reference of the novel allele candidate (Novel ref) and the reference sequence of the related allele (Rel ref) - i.e. the sequence of the allele the novel allele candidate originated from - are shown. Novel positions are accepted only when at least 90% of the reads and at least 10-fold coverage depth support the novel reference value. Note that in many cases multiple similar novel allele candidates are generated from different alleles. There are three versions of the novel allele detection settings:
 - not running novel allele detection
 - running novel allele detection only when exon mismatches are present (note: this function will try to resolve both exon and intron novelties)
 - running novel allele detection if any mismatches are present (whole gene novel allele detection).
- **Save read mappings:** Use this option to save the mapped reads to browse them in the Gene Browser later.

Gene targeting

You can select the targeted gene family and loci set. The predefined sets are based on the different Holotype HLA Kit configurations (5, 7, 11-loci) and there is an option to set custom loci for the genotyping, which can be useful for reanalysis purposes. When the 'Auto' option is selected, gene family and targeted loci selection is disabled since these parameters will be determined from the sample during the analysis.

4.9 Genotyping Reanalysis

Any analysis can be rerun for the same set of input files - samples - which it had been setup with. Apart from the input files, any parameters and settings can be changed. This allows you to correct any mistakes - like choosing an inappropriate parameter - and also enables examination of results with different settings or IMGT database versions. The number of free re-analyses is unlimited.

4.10 Export results

4.10.1 General information

Genotyping results can be exported in the standard HML or in PDF format using the Export Results wizard available from the Genotyping dashboard.

On the first screen of the wizard, export format and the preferred location for the exported files can be selected.

If you check the *Use LD* checkbox, results will be exported as shown on the *Genotyping analysis result* screen when LD functionality is switched on. If the option is unchecked LD information will not be included in the exports.

4.10.2 HML export

General information

The currently supported HML Report format is version 1.0.1. For reference and variant reporting the first best match allele pair of the result is used and the reference is extracted from the IMGT database. If no allele pair is available then we fall back to the GRCh38 reference: in this case the coordinates are not as accurate therefore those results/exports shouldn't be used for further processing except some special research cases. A HML report can be generated from imported results as well.



HML export wizard

Export options

All alleles or only assigned alleles can be exported by choosing the All or Assigned options, respectively.

The *Export precision* options lets you refine the precision level of the reported genotypes/alleles:

- P group: export P groups instead of alleles
- 2 fields: export 2 field allele names
- G group: export G groups instead of alleles
- 3+2 fields: export 3 field allele names for class I and 2 field allele names for class II
- 3 fields: export 3 field allele names
- 4 fields: export 4 field allele names

The *Precision reduction for novelties* option allows further automatic refinement of the precision: in case of novelties/mismatches it reduces the precision automatically to 3 field. Test Id and Test Id Source can also be specified on this screen.

Advanced export options

On this screen, values for several HML fields can be specified.

By default, one HML file is created per sample and file names are generated using analyses names. Single file export can be selected to export all selected results with the specified file name.

NMDP export options

This optional step of the wizard can be used for specifying packing list files. A default center code can also be specified in this step of the wizard.

4.10.3 PDF export

A PDF report can be generated based on the results of an analysis. The contents are highly customizable, you can select which metrics, statistics and other details should appear in the report. The alleles can also be selected by restricting the reported alleles to the Assigned alleles only. If you check the *Use LD* checkbox, results will be exported as shown on the *Genotyping analysis result* screen when LD functionality is switched on. If the option is unchecked LD information will not be included in the exports.

The organization of the report is the following:

- Header: mandatory section on top of each page - contents cannot be customized.
- Sample name: as displayed in the software.
- Analysis date: date and time when the analysis has run.
- Analyzed by: user name used to produce the analysis.
- Approval date: if the analysis results have been approved the date of the approval is displayed. In case the analysis have been approved more than once the last date is displayed.
- Approved by: if the analysis results have been approved the user name of the approved is displayed.
- Sent for Approval by: if the analysis results have been sent for approval the user's name who sent the results is displayed. If the results have not been sent for approval, the field contains the 'Not Sent' message.
- Genotyping result: contains allele candidates for each locus and the related P&G groups or the phenotype where applicable. When a result is too ambiguous to report - more than 10 allele pairs would be included in the report - the report will contain a warning about this as it is difficult to display so many alleles. You can reduce the amount of displayed alleles for example by assigning only the relevant ones and re-generating the report.
- Analysis configuration: contains the details of the protocol and a warning in case the result had not been produced with the Holotype protocol.
- Quality control summary: concordance and overall QC result for each locus
- Mismatches
- Novelties
- Alignment statistics: coverage and average coverage depth for each allele candidate on each locus
- Data Statistics: ratios of processed and skipped reads, noise and other details also displayed on the Data Statistics tab
- Comments: user comments related to samples or loci



- Locus specific content: details relevant for each locus - contents are customizable.
- Quality control details: detailed quality control metric values for each allele candidate in each locus

Selected report customization options are pertained for each user separately so you won't have to select the required contents again if you are satisfied with the setup as it was the last time you used it.

4.10.4 JSON export

Export options

You can choose to export all alleles from a typing result or assigned alleles only.

Export format and content

Results will be exported in standard JSON format.

i Sample JSON file

```
{
  "sampleName": "Sample",
  "analysisRunName": "Sample_result",
  "analysisDate": "Sep 8, 2017 12:55:14 PM",
  "executedBy": "Unknown",
  "approvedBy": "Not approved",
  "sampleComment": "Sample level comment",
  "genotypes": [
    {
      "gene": "HLA-A",
      "genotype": "01:03#1+02:40:01",
      "genotypeGGroup": "No G group+No G group",
      "genotypePGroup": "02:40P+No P group"
    },
    {
      "gene": "HLA-B",
      "genotype": "08:74#1+53:21#1",
      "genotypeGGroup": "No G group+No G group",
      "genotypePGroup": "No P group+No P group"
    }
  ],
  "geneResults": [
    {
      "gene": "HLA-A",
      "frequentQcMetrics": [
        {
          "name": "Read count",
          "value": "3464",
          "code": "READ_COUNT"
        }
      ],
      "rareQcMetrics": [
        {
          "name": "Read quality",
          "value": "36.22",
          "code": "READ_QUALITY"
        }
      ]
    },
    {
      "allele": "01:03",
      "averageCoverageDepth": 20
    },
    {
      "allele": "02:40:01",
      "averageCoverageDepth": 26
    }
  ],
  "alleleFrequencies": [
  ],
  "comment": "Locus level comment"
},
{
  "gene": "HLA-B",
  "frequentQcMetrics": [
```

```
{
  "name": "Read count",
  "value": "3498",
  "code": "READ_COUNT"
},
"rareQcMetrics": [
  {
    "name": "Read quality",
    "value": "35.97",
    "code": "READ_QUALITY"
  }
],
"alleles": [
  {
    "allele": "08:74",
    "averageCoverageDepth": 6
  },
  {
    "allele": "53:21",
    "averageCoverageDepth": 8
  }
],
"alleleFrequencies": [
],
"comment": "No comment available"
},
"configuration": {
  "sequencer": "ILLUMINA",
  "dataType": "PAIRED",
  "typingMethod": "TWIN",
  "maxReadsPerLocus": 4000,
  "disableQualityBasedSubsampling": false,
  "ignoreRareAlleles": false,
  "novelAlleleDetectionMode": "EXON_NOVELTIES_ONLY",
  "applicationVersion": "2.2.0",
  "applicationBuildId": "0ab106d522677cac281333ab6aa7e45981e0b7e9",
  "databaseVersion": "3.25.0_3(+E)",
  "defaultConfig": true
}
}
```

4.10.5 HPRIM export

If Omixon HLA is configured to have a LIMS connection you can use this wizard to launch the process of exporting analysis results into the configured LIMS system.

Filter analysis results

As first step you can filter down the analysis results you have selected for export in the *Genotyping dashboard* to ease the final selection of the results to be exported on the next page

The following filter criteria are available:



- *Approved only*: if this is turned on then only the approved results are shown (default is *on*)
- *Exported only*: if this is turned on then only the previously exported results are shown - in other words this filter can be used for re-exporting results after modification (default is *off*)
- *Show approved results since*: if the *Approved only* checkbox is on then you can set an initial date as an additional filter based on the analysis execution time (default is one week before the actual date)
- *Show exported results since*: if the *Exported only* checkbox is on then you can set an initial date as an additional filter based on the analysis result export time (default is one week before the actual date).

Select analysis results

As next step you can setup the final set of analysis results to be exported. Here only those analysis results are listed which are matching all the filter criteria set in the previous wizard step.

For each analysis result the following attributes are shown to help with the identification:

- *Analysis name*: name of the result as shown in other screens in the application
- *Execution date*: date and time of the sample analysis execution
- *Approved by*: login name of the user who has approved the result
- *Approval date*: date and time of the approval

To change the selection of multiple items in the list use the CTRL and SHIFT keys while clicking on the individual analyses.

Export configuration

As last step in the wizard the following export options can be set:

- *Export method*: *Export by sample* or *Export by batch* - the former one defines that each typing result should be placed into a separate file (this is the default setup) and in the latter case all typing results will be exported into one file.
- *Export location*: the folder where export files should be placed can be selected here (default is set to the folder which is specified on the *Settings dashboard* as a system wide configuration).

Once you have finished the wizard the export generation is executed automatically in the background. The process can be monitored just like any other background task in the process manager.

5 Genotyping analysis result

5.1 Introduction

High-level overview of the Genotyping results.

A high-level overview table of the genotyping results. This can include the summary of results for many samples. You can use Ctrl + mouse wheel to zoom in and out of the table.

Each Sample can be selected and unselected (Ctrl+click). The following options are available:

- *Sample Details* - leads to the *Genotyping sample result* dashboard which contains detailed data and statistics.
- *Browse Alignment* - opens the first pair of consensus sequences with the best matching allele candidates in the Gene browser.

The results can be exported in TXT (tab delimited text), CSV (comma separated text), XLS (Excel) or XLSX (Excel XML) format.

You can export the results with or without the flags that are displayed on the screen by selecting between the Overview and the Overview with allele flags options.

Tip - If you would like to export only the assigned allele candidates first click the **Assigned Only** button, then export the results.

5.1.1 Commenting

Comments can be added to a selected sample using the Sample Comment button. Commented samples are marked with a small red triangle in the top right corner of the sample name field on this screen. These comments are by default included in the PDF report but they can be excluded if necessary.

Comments can be removed by clicking the Sample Comment button again and removing the text.

5.2 Top panel

On the top of the screen you can see three rows of functions.

5.2.1 Top row

On the left, you can find the *Toggle fullscreen* button which hides the less important parts of the current screen, while on the right you can see the following things:

- the ID of the current user,
- the memory usage widget panel,
- the status panel of the process manager,
- the welcome tutorial button,
- the logout button,
- and the exit button.

In case of the desktop version, the actual memory usage can be observed at the top right corner of the screen. This displays the amount of currently used memory compared to the allowed maximum. Note, that the software might be allocating more memory from the operating system compared to the reported actual memory usage to make further computations faster by keeping some previously allocated memory pieces for later use, thus making the allocation of new memory unnecessary during further computations. By clicking on the memory report widget the software tries to free up as much memory as possible on a best effort basis: this is only an attempt, there are many technical factors which might block returning all memory to the operating system. It is also important to mention that by explicitly asking to free up memory the subsequent computations might become slightly slower, since the option of reusing certain structures allocated before becomes impossible. Only use this option when all planned analyses have been finished and you would like to use other programs more intensively in parallel while investigating the analysis results.

5.2.2 Middle row

There are four navigation buttons that are available on every screen of the tool. These general navigation buttons work very similar to navigation functions available in most widely used internet browsers:

- The *Back* button will take you to the previous screen.
- The *Forward* button will take you to the next screen.



- The *Up* button will move you one level up in the application.
- The *Home* button will take you to the starting screen (i.e. the Genotyping dashboard).

Right from the navigation button, in the information panel, you can find relevant high-level information about the current screen. On the *Genotyping dashboard*, you can see the release date and version of the active IMGT database.

On the far right hand side of the middle row, you can find the bookmark and context specific help buttons. The help tool shows relevant usage information and tips about functions and data available on the current screen of the application and can also be opened by pressing F1.

You can search the Help for any expression by using the Magnifier glass icon displayed on the popup window. Relevant hits are displayed in the left panel where you can select them and open them in the right hand reading pane. You can close the search by clicking the X displayed beside the search field.

5.2.3 Bottom row

The bottom row of the top panel contains a series of buttons which contain the main functions available on the screen.

5.3 Interpreting Results

By default, only the *best matching* candidates are displayed in the results table. These are the pairs of candidates with *the lowest exon and non-exon mismatch counts* compared to the pair of consensus sequences generated from the short reads. If an ambiguous best match is returned, then there will be more than two best matching candidates for the locus. Hover your mouse over any result to display a tooltip that contains further details about the results.

Novel alleles are named relative to an existing allele in the database. The base allele is selected in a way to take into account the biochemical similarity of the novel allele and the base allele as much as possible. This means that sometimes alleles with more mismatches/novelities on the non key exons are used as a base allele if this leads to a higher similarity/lower mismatch count - ideally even equivalence - on the key exons (exon 2 and exon 3 for class I genes and exon 2 for class II genes).

By pushing the *Best Matches Only* button, it is possible to see other close allele candidates.

By selecting a sample (clicking on it in the table) you will be presented with more options for drilling down into the details of the results. You can see details for a sample, you can browse allele candidates and consensus alignments in the browser, and you can see the pair(s) of candidates for each locus. More information is available within the help page of each of these functions.

A gene is homozygous if the same allele is present on both of the chromosomes and heterozygous if two different alleles of the same gene are present. In case there is only one copy of a gene is present then the locus is hemizygous. Details about the allele display can be found under the Traffic Lights, Quality control light section. Alleles from different genes are not in random recombination if the genes are located close to each in a chromosome. This phenomenon is called 'linkage disequilibrium and this information is used for determining if a particular locus is hemizygous.

5.3.1 Troubleshooting missing results

When no alleles could be reported for a targeted gene, a markup describing the possible reason for the missing allele call is shown. For additional information, hover over the info icon next to the markup and read the tooltip. The following cases can be reported:

For non DRB3/4/5 loci:

- *No data present* - No data present means that either the locus was not targeted in the sample, or it has dropped out during sequencing and should be re-sequenced.
- *Insufficient or low quality data* - There is insufficient data or the data is of low quality in the sample. Quality control results should be checked for more detail.

For DRB3/4/5:

- *Allele not expected* - There is no allele expected at this locus based on known linkage disequilibrium with HLA-DRB1 and HLA-DQB1.
- *Expected allele not found* - This markup means that based on known linkage disequilibrium information, data was expected for the locus/allele but was not found.
- *Unexpected allele found* - Data was found for a locus/allele, which was not expected based on known linkage disequilibrium information.
- *Insufficient or low quality data* - There is insufficient data or the data is of low quality in the sample. Quality control results should be checked for more detail.

When no alleles are reported for a targeted gene it is suggested to rerun the sample in question using a higher number of reads. (The number of processed reads can be set in the *Advanced Genotyping wizard*.) The reasons behind the missing allele level results can be that the

coverage does not reach the minimum threshold on the allele or on the exons, or the coverage depth is too small. Processing more reads can help making the signals that support the correct alleles stronger.

5.4 Filtering Alleles

A summary of the currently active filters is displayed to the top right of the Genotyping analysis results screen. The following filter options are available:

- **Setup Loci** - A set of visible loci can be selected.
- **Setup Filters** - Where allele frequency information is available a gray information icon is displayed by the allele. This marking can be switched on and off on this popup panel.
- **Best Matches Only** - Toggles remainder alleles.
- **Assignment State** - In combination with the Best Matches Only filter, the following combinations can be shown:
 - Assigned alleles, best matches and remainders
 - Only assigned alleles
 - Best matches and remainders
- **Workflow State** - This is a sample level filtering option that toggles between showing samples in progress, waiting for approval and only approved samples.

5.5 Traffic Lights

Results are annotated using a "Traffic light" system. Similarly to an actual traffic light, three different colours are used, all with different meanings. Unlike in a real traffic light, "mixed colours" are also available.

The different colours are based on concordance between consensus genotyping (CG) and statistical genotyping (SG) results (first light). Please note that the Concordance traffic light is only available when the Twin genotyping method is used.

5.5.1 Concordance light

The concordance traffic light system has the following outputs:

- (green): results of SG and CG are fully concordant,
- (yellow/green): results of SG and CG are concordant up to 6-digits (3rd field),
- (yellow): results of SG and CG are concordant up to 4-digits (2nd field),
- (red/yellow): results of SG and CG are concordant up to 2-digits (1st field),
- (red): results of SG and CG are discordant.




Note, that concordance traffic lights are only presented for best matching alleles.






The second traffic light is based on locus level quality control (QC) measures.

5.5.2 Quality control light

- (green) - PASSED: the locus passed all QC tests,
- (yellow/green) - INFO: one or more QC tests produced lower than average results,
- (yellow) - INSPECT: one or more QC tests produced concerning results, manual inspection of the results needed,
- (red/yellow) - INVESTIGATE: one or more QC tests showed low result quality, manual inspection and possibly reanalysis is needed,
- (red) - FAILED: one or more QC tests showed very low result quality, manual inspection is needed to determine the cause and the locus or sample likely needs re-sequencing or re-typed by alternative methods.

Some other markings can also be presented for alleles:

- Alleles displayed with the blue font are homozygous.
- Rare alleles are marked with an exclamation point icon .
- Novel alleles containing exonic (or exonic and intronic) novelties are marked with , while novel alleles containing only intronic novelties are marked with .

- The presence of additional imbalanced remainder alleles that are close to the best match in coverage, but have significantly lower coverage depth is marked with a scale icon . These imbalanced alleles are displayed with italic font.
- Alleles with extended allele sequence are marked with a *plus sign* .
- If a minor allele with well-known low amplification is present in the imbalanced minor allele list, the allele is marked with . In this case validation of the homozygous result using an alternate genotyping method (e.g. SSO) is strongly suggested.
- Hemizygous alleles are marked with . In case a locus is hemizygous only one allele is displayed and the other cell is left empty. In case the zygosity of a locus can't be determined based on the data available, the alleles are marked with .

5.6 Assigning Alleles

For each locus, one or more allele pair candidates can be assigned. Candidates can be assigned manually by clicking on the check mark before the allele's name. The following assignment options and assignment history are available via the Assignment option:

- All alleles with a certain level of overall QC result or above can be assigned.
- Assign best matches
- Assign only heterozygous results
- Assign only unambiguous results
- Assign novel alleles

All assignments can be deleted by the *Clear assignments* button in the wizard. Note, that all the 'Assign ...' options on this screen affect all the displayed samples and alleles - they don't affect alleles that are not being displayed due to filtering criteria. If you prefer to do the assignments separately for each sample/allele, you can find similar functions on the *Genotyping sample result* dashboard. Allele assignment only works for samples in "In progress" state. The function is blocked in "Ready" and "Approved" states.

5.7 Export table

Every Genotyping result table can be exported in either CSV (comma separated text), TXT (tab delimited text), XLS or XLSX format. Note, that only visible results are exported (i.e. if an allele is filtered out based on low coverage, it won't be shown in the exported file either).

There are several similar export table functions:

On the *Genotyping analysis results* dashboard, you can export all the results from an analysis. If you select the *Overview* export, only the allele names and quality control measures are exported. The *Detailed* table also contains flags and statistics for each allele candidate.

Flag abbreviations:

- R - rare,
- I - imbalanced,
- E - extended allele.

Quality control test result abbreviations:

- P - passed,
- I1 - info,
- I2 - inspect,
- I3 - investigate,
- F - failed.

Concordance test result abbreviations:

- C - fully concordant,
- C2 - concordant on a two digit level (1st field),
- C4 - concordant on a four digit level (2nd field),
- C6 - concordant on a six digit level (3rd field),
- D - discordant.

5.8 Linkage disequilibrium details

5.8.1 Overview

Based on the known genetic rules related to linkage disequilibrium (LD) and the database that is integrated into the software in case of DRB3, DRB4 and DRB5 loci additional information can be reported next to the actual genotyping results. These can be either markups describing the relation between the alleles in the genotyping result and the alleles expected based on LD or can be actual allele display changes in the result - e.g. showing up an allele only once for hemizygous results.

Note
If more than 50 best matching results can be found on a locus, LD will not be calculated. If you press the Show LD details button, no information will be displayed.

The output of matching the genotyping result with the LD based expected result is reported in the software as follows:

Matching genotyping result with linkage disequilibrium	Allele present in genotyping result	Allele <i>not</i> present in genotyping result
Allele expected based on LD	Allele displayed, no markup	Allele not displayed, bold markup says: 'Expected allele not found'
Allele <i>not</i> expected based on LD	Allele not displayed, bold markup says: 'Unexpected allele found'	Allele not displayed, markup says: 'Allele not expected'
There is no information about the allele in the LD database	Allele displayed, bold markup says: 'No linkage information available'	Allele not displayed, markup says: 'Allele not expected'

5.8.2 Case by case guide

Specific cases can be detailed as follows.

Expected allele not found

- Allele present in the LD based expected result but not present in the genotype
- DRB4*01:01 has an extra annotation: "This allele is known to drop out, please consider checking this sample with an alternative typing technology (e.g. SSO/SSP)."
- The only allele is null in the genotype of DRB4 meanwhile the LD based expected result contains DRB4*01:01 too

Unexpected allele found

- Genotyping result present for all of DRB3/4/5 but the LD based expected result contains the alleles of the other two loci only
- Genotyping result present for two of DRB3/4/5 but the LD based expected result contains the alleles of one other locus only
- Heterozygous genotype but the LD based expected result contains only one of the alleles and another locus
- Heterozygous genotype but the LD based expected result contains only one of the alleles and denotes the other allele to be not present
- Second allele of a heterozygous genotype where both alleles match the LD based expected result and the other pair matches another locus
- Second allele of a heterozygous genotype where both alleles match the LD based expected result and the other pair refers to an allele known to be not present

Allele not expected

- Allele is the only result for a locus and it matches one of the LD based expected result pairs meanwhile the other pair denotes the other allele to be not present
- Genotype not present and no LD based expected result is available

No linkage info available



- Allele not present in the LD based expected result
- Ambiguous allele meanwhile only other allele(s) match the LD based expected result

Unknown

- Multiple contradicting rules among the above ones could be applied to the result

5.8.3 Notes

- Results are assumed to be hemizygous by default
- Alleles are considered up to 2 fields precision
- If result is not present on DRB1 then we abandon the LD based result evaluation
- Novel alleles are interpreted by taking their base allele, the novelty part is not considered

5.8.4 DQA1, DQB1 allele dropout warning

As allele dropouts are very hard to distinguish from naturally homozygous results, LD information is used to detect possible DQA1 or DQB1 allele dropouts.

In case DQA1 or DQB1 is found homozygous by the algorithm, then it is checked if one of the alleles of the other locus is linked with an allele which is not the homozygous one. In case there is such allele(s) then a warning is displayed next to the homozygous result. (E.g. if DQB1 is homozygous for 06:03 and DQA1 is heterozygous (01:03+05:05): DQA1*01:03 is linked with DQB1*06:03 and DQA1*05:05 is linked with DQB1*03:01, so we display a warning about the possible DQB1*03:01 dropout.)

If both DQA1 and DQB1 are homozygous then they are checked against DRB1 the same way as described above.

6 Genotyping sample result

6.1 Top panel

On the top of the screen you can see three rows of functions.

6.1.1 Top row

On the left, you can find the *Toggle fullscreen* button which hides the less important parts of the current screen, while on the right you can see the following things:

- the ID of the current user,
- the memory usage widget panel,
- the status panel of the process manager,
- the welcome tutorial button,
- the logout button,
- and the exit button.

In case of the desktop version, the actual memory usage can be observed at the top right corner of the screen. This displays the amount of currently used memory compared to the allowed maximum. Note, that the software might be allocating more memory from the operating system compared to the reported actual memory usage to make further computations faster by keeping some previously allocated memory pieces for later use, thus making the allocation of new memory unnecessary during further computations. By clicking on the memory report widget the software tries to free up as much memory as possible on a best effort basis: this is only an attempt, there are many technical factors which might block returning all memory to the operating system. It is also important to mention that by explicitly asking to free up memory the subsequent computations might become slightly slower, since the option of reusing certain structures allocated before becomes impossible. Only use this option when all planned analyses have been finished and you would like to use other programs more intensively in parallel while investigating the analysis results.

6.1.2 Middle row

There are four navigation buttons that are available on every screen of the tool. These general navigation buttons work very similar to navigation functions available in most widely used internet browsers:

- The *Back* button will take you to the previous screen.
- The *Forward* button will take you to the next screen.
- The *Up* button will move you one level up in the application.
- The *Home* button will take you to the starting screen (i.e. the HLA Typing dashboard).

Right from the navigation button, in the information panel, you can find relevant high-level information about the current screen. On the *Genotyping dashboard*, you can see the release date and version of the active IMGT database.

On the far right hand side of the middle row, you can find the bookmark and context specific help buttons. The help tool shows relevant usage information and tips about functions and data available on the current screen of the application and can also be opened by pressing F1.

You can search the Help for any expression by using the Magnifier glass icon displayed on the popup window. Relevant hits are displayed in the left panel where you can select them and open them in the right hand reading pane. You can close the search by clicking the X displayed beside the search field.

6.1.3 Bottom row

The bottom row of the top panel contains a series of buttons which contain the main functions available on the screen.

6.2 Available functions

6.2.1 Opening the browser

Both chromosomes or a single consensus sequence and the corresponding alleles can be viewed in the Gene browser, using the *Browse Alignment*, *Browse Allele 1* or *Browse Allele 2* functions. Note that the same settings can be chosen from within the Gene browser.

6.2.2 Detailed genotyping information

Mismatch and coverage statistics and detailed mismatch and novelty lists for a selected allele pair can be viewed using the *Genotype Details*, *Show Mismatches* and *Show Novelties* functions.

6.2.3 Customizing displayed results

Loci shown in this screen can be customized using the *Setup Loci* function. Where allele frequency information is available a gray information icon is displayed by the allele. This marking can be switched on and off using the *Setup Filters* wizard. Alleles can also be filtered by status, using the *Best Matches Only* and *Assigned Only* functions. Result resolution can be changed using the *Genotype Precision* button. All assignments can be deleted by the *Unassign All* button. Note, that all the 'Assign ...' buttons on this screen affect all the displayed samples and alleles - they don't affect alleles that are not being displayed due to filtering criteria. If you prefer to do the assignments separately for each sample/allele, you can find similar functions on the *Genotyping sample result* dashboard.

6.2.4 Assigning alleles

For each locus, one or more allele pair candidates can be assigned. Candidates can be assigned manually by clicking on the check mark before the allele's name. The following assignment options and assignment history are available via the Assignment option:

- All alleles with a certain level of overall QC result or above can be assigned.
- Assign best matches
- Assign only heterozygous results
- Assign only unambiguous results
- Assign novel alleles

All assignments can be deleted by the *Clear assignments* button in the wizard.

Allele assignment only works for samples in "In progress" state. The function is blocked in "Ready" and "Approved" states.

6.2.5 Commenting

Comments can be added to a selected sample or each loci within a selected sample using the Sample Comment or the Locus comment buttons. Commented samples and loci are marked with a small red triangle in the top right corner of the sample name field or the allele field on the Genotyping Analysis result screen. These comments are by default included in the PDF report but they can be excluded if necessary. Comments can be removed by clicking the Sample Comment or Locus comment button again and removing the text.

6.3 Genotype

6.3.1 Tree view

The Genotype tab displays the results in a tree format. The different levels of the tree refer to different properties in different gene families. In case of HLA the first levels of the tree are the P group and the G group - where available. Using the *Genotype Precision* button reduces the result display level to these groups.

In case of ABO results the first level of the tree is the Phenotype information derived from the results.

6.3.2 Troubleshooting missing results

When no alleles could be reported for a targeted gene, a markup describing the possible reason for the missing allele call is shown. For additional information, hover over the info icon next to the markup and read the tooltip. The following cases can be reported:

For non DRB3/4/5 loci:

- *No data present* - No data present means that either the locus was not targeted in the sample, or it has dropped out during sequencing and should be re-sequenced.
- *Insufficient or low quality data* - There is insufficient data or the data is of low quality in the sample. Quality control results should be checked for more detail.

For DRB3/4/5:

- *Allele not expected* - There is no allele expected at this locus based on known linkage disequilibrium with HLA-DRB1 and HLA-DQB1.
- *Expected allele not found* - This markup means that based on known linkage disequilibrium information, data was expected for the locus/allele but was not found.
- *Unexpected allele found* - Data was found for a locus/allele, which was not expected based on known linkage disequilibrium information.
- *Insufficient or low quality data* - There is insufficient data or the data is of low quality in the sample. Quality control results should be checked for more detail.

When no alleles are reported for a targeted gene it is suggested to rerun the sample in question using a higher number of reads. (The number of processed reads can be set in the *Advanced Genotyping wizard*.) The reasons behind the missing allele level results can be that the coverage does not reach the minimum threshold on the allele or on the exons, or the coverage depth is too small. Processing more reads can help making the signals that support the correct alleles stronger.

6.3.3 Adding alleles manually

Right click on the Genotype panel to find the context menu which provides the "Add genotype" option to add alleles to the displayed result. A popup window will be displayed where the original result is highlighted. Browse the displayed P and G groups to locate the allele pair candidate that you wish to add. Please note that the already present result cannot be added again. Please note that only allele pair candidates can be added, single alleles cannot. You can add allele pair candidates from any P and G group and the allele pair will be displayed on the Genotype panel accordingly with a different blue markup in front of the pair.

Adding alleles manually is only possible for samples in "In progress" state. The function is not available in "Ready" and "Approved" states.

6.3.4 Removing additional alleles

The manually added allele pair candidates can be removed by selecting them and using the right mouse button to summon the context menu. Select the "Delete user added genotype" option to remove the allele pair candidate.

Removing additional alleles is only possible for samples in "In progress" state. The function is not available in "Ready" and "Approved" states.

6.4 Quality Control

Several quality control measures are calculated for every loci. Each measure for each loci is marked with a traffic light system. Thresholds for light colour assignment are presented as a tooltip in the right hand side of every cell. The QC measures are the following:

Overall

Overall QC status of the loci, calculated as the value of the worst QC measure or measures.

Primary QCs for Interpretation

- *Read count* - total number of reads mapped to a locus.
- *Noise ratio* - number of reads considered as noise relative to the total number of reads mapped to a locus.
- *Key- exon spot noise ratio* - maximum number of reads assumed to be systematic noise relative to the total number of reads at a certain position in the key exon regions (i.e. exons 2 and 3 for Class I loci and exon 2 for Class II genes in the HLA family).
- *Consensus coverage key exon minimum depth* - lowest number of reads supporting the coverage across all key exon positions.
- *Key exon allele imbalance* - average ratio of the alleles for all key exon regions. Values around 50%-50% indicate a balanced heterozygous result, while values close to 100%-0% indicate imbalance in the proportion of reads derived from the two chromosomes. Homozygous results get 50%-50% assigned. Note, that it is not always possible to distinguish between homozygous regions with a small amount of contamination or artifact reads and highly imbalanced heterozygous regions.

- *Genotype available* - a *Yes* value indicates that a genotype could be identified for a locus while *No* indicates that no best matching alleles were found.

Secondary QCs for Interpretation

- *Fragment size* - average length of the sequenced fragments identified by read pairs mapped to a locus. Fragment size is calculated the following way: alignment length for read 1, plus alignment length for read 2, plus average distance between read 1 and read 2 based on all alignments for the read pair.
- *Read quality* - average quality of all reads mapped to a locus.
- *Other exon spot noise ratio* - maximum number of reads assumed to be systematic noise relative to the total number of reads at a certain location in the non key exon regions.
- *PCR crossover artifact ratio* - number of reads assumed to be PCR crossover artifacts relative to the total number of reads mapped to the locus.
- *Key exon mismatch count* - number of variants identified between the alleles in the result pair and the consensus sequences in key exonic regions.

Warnings for Troubleshooting

- *Read length* - average length of all reads mapped to a locus.
- *Crossmapping (intergenic ambiguity)* - number of reads mapping to multiple genes relative to the total number of reads mapped to a locus.
- *Ambiguous layout (intra-genic ambiguity)* - number of reads with an ambiguous relation to other reads relative to the total number of reads mapped to a locus.
- *Non-exon spot noise ratio* - maximum number of reads assumed to be noise relative to the total number of reads at a certain location in the non exonic regions.
- *Continuous consensus* - continuous consensus sequences get a value of 1, while non-continuous consensus sequences get a value of 0.
- *Fully phased consensus* - fully phased consensus sequences get a value of 1, while not fully phased sequences get a value of 0.
- *Consensus coverage other exon minimum depth* - lowest number of reads supporting the coverage across non key exon positions.
- *Consensus coverage non-exon minimum depth* - lowest number of reads supporting the coverage across all non exon positions.
- *Other exon allele imbalance* - average ratio of the alleles for non key exon regions with consensus sequences. Values around 50%-50% indicate a balanced heterozygous result, while values close to 100%-0% indicate imbalance in the proportion of reads derived from the two chromosomes. Homozygous results get 50%-50% assigned. Note, that it is not always possible to distinguish between homozygous regions with a small amount of contamination or artifact reads and highly imbalanced heterozygous regions.
- *Non-exon allele imbalance* - average ratio of the alleles for the non-exon region with consensus sequences. Values around 50%-50% indicate a balanced heterozygous result, while values close to 100%-0% indicate imbalance in the proportion of reads derived from the two chromosomes. Homozygous results get 50%-50% assigned. Note, that it is not always possible to distinguish between homozygous regions with a small amount of contamination or artifact reads and highly imbalanced heterozygous regions.
- *Other exon mismatch count* - number of variants identified between the alleles in the result pair and the consensus sequences in non key exons.
- *Non-exon mismatch count* - number of variants identified between the alleles in the result pair and the consensus sequences in non-coding regions (introns and UTRs).
- *Novel position count* - the number of novel positions identified throughout the gene. A high number of novel positions can indicate incorrect phasing or low consensus quality.

QC results can be exported using the *Export results* button. Note, that you can find some troubleshooting tips in the Advanced user guide section of the Omixon HLA user manual.

6.5 Data statistics

6.5.1 Overview

Read counts and proportions are available for several different steps of the analysis. In the two columns on the left "remaining reads" (i.e. reads that will be used in the next step of the analysis) are shown, while on the right "filtered reads" (i.e. reads that have been removed from the analysis for some reason) are presented. The following counts are listed:

- *Processed* - total number of processed reads from the input data files.
- *Skipped* - reads that don't map to any of the targeted loci or were skipped to balance the amount of reads between targeted loci.
- *Invalid orientation* - currently, only FR orientation is accepted, read pairs with FF or RR orientation are ignored.
- *Crossmapping* - reads mapping to multiple loci, by default these are ignored.

- *Dropped during filtering* - the total number of reads that were not aligned (this includes crossmapped reads, off-target reads, reads with invalid orientation and reads that were dropped because their pair could not be aligned to the HLA region).
- *Kept after filtering* - reads that were mapped to targeted loci and were not filtered due to any of the reasons above.
- *Lost during consensus contig generation* - reads lost due to their ambiguous relation to other reads.
- *Assembled* - total number of reads in the consensus contigs.
- *Removed as noise* - reads removed as contamination (based on supporting read proportions for variants, basically consensus fragments introducing very lowly supported variants are removed).
- *PCR crossover artifact* - reads removed as chromosomal crossover artifact. Note, that currently only artifacts within a single region are removed.
- *Used for final consensus generation* - the total number of reads used for final consensus generation.

6.5.2 Allele imbalance

This figure shows the per-region allelic imbalance for all the genes. On the y-axis you can see the allelic imbalance measure (%) while on the x-axis, you can see all loci. The allele imbalance measure reflects the imbalance between reads originating from different chromosomes. The measure is calculated using heterozygous positions and presented as a pair of percentages. Percentage pairs, where both values are around 50% represent balanced heterozygous positions, while low-high percentage pairs mark imbalance. Homozygous regions are presented as 50%-50%. Note, that it is not always possible to distinguish between homozygous regions with a small amount of contamination and highly imbalanced heterozygous regions.

Overall allele imbalance: average allelic imbalance, only considering heterozygous regions.

6.5.3 Fragment size

This histogram shows the fragment size distribution of paired reads. (Fragment size is calculated as the length of both reads in a pair, plus the inner mate distance between R1 and R2, which can be negative if the reads overlap.) On the y-axis, you can see the number of reads and on the x-axis you can see the fragment length. Note, that the x-axis shows "binned" values: values with non-zero read counts are grouped by 10, therefore columns often (and especially with higher values) represent a relatively big "stretch" of fragment length. Again, by hovering over a column with the mouse, you can see the exact values on both the x and y-axis for that specific column.

6.5.4 Read quality

On this graph, the base quality per 5 bases is shown for the processed reads. Read positions are on the x-axis while on the y-axis quality values are shown. Besides the 5-base average, the 90th percentile (High quality) and 10th percentile (Low quality) quality values are shown. Note, that by hovering the mouse over a group of columns, the exact position quality values can be viewed.

The average quality and length for processed and aligned reads are shown. Note, that reads shorter than 75 bp are filtered out, so all processed reads are longer than this threshold (when paired data is used, both reads in every processed pair have to pass the length filter). After read processing, but before the alignment, quality based trimming is performed. Because of this step, aligned reads are usually somewhat shorter and have a little higher average quality than processed reads.

6.5.5 Troubleshooting missing results

When no alleles could be reported for a targeted gene, a markup describing the possible reason for the missing allele call is shown. For additional information, hover over the info icon next to the markup and read the tooltip. The following cases can be reported:

For non DRB3/4/5 loci:

- *No data present* - No data present means that either the locus was not targeted in the sample, or it has dropped out during sequencing and should be re-sequenced.
- *Insufficient or low quality data* - There is insufficient data or the data is of low quality in the sample. Quality control results should be checked for more detail.

For DRB3/4/5:

- *Allele not expected* - There is no allele expected at this locus based on known linkage disequilibrium with HLA-DRB1 and HLA-DQB1.
- *Expected allele not found* - This markup means that based on known linkage disequilibrium information, data was expected for the locus/allele but was not found.
- *Unexpected allele found* - Data was found for a locus/allele, which was not expected based on known linkage disequilibrium information.

- *Insufficient or low quality data* - There is insufficient data or the data is of low quality in the sample. Quality control results should be checked for more detail.

When no alleles are reported for a targeted gene it is suggested to rerun the sample in question using a higher number of reads. (The number of processed reads can be set in the *Advanced Genotyping wizard*.) The reasons behind the missing allele level results can be that the coverage does not reach the minimum threshold on the allele or on the exons, or the coverage depth is too small. Processing more reads can help making the signals that support the correct alleles stronger.

6.6 Genotyping mismatch result

On this screen, a detailed list of mismatches between the selected alleles and the corresponding consensus pair are available. After clicking on a specific mismatch, you can jump to the selected mismatch in the Gene browser. You can filter mismatches based on allele status (*best match* or *assigned*) or by *chromosome*. You can export this detailed list using the *Export Results* option.

6.7 Genotyping result novelties

On this screen, a detailed list of novelties (i.e. base changes that have been applied to the allele from which the novel allele was created) is available. After clicking on a specific novelty, you can jump to the selected novelty in the Gene browser. You can filter novelties based on allele status (best match or assigned) or by chromosome. You can export this detailed list using the *Export Results* option.

6.8 Allele frequency information

Information from the Allele Frequencies database (<http://www.allelefrequencies.net>) is incorporated into our downloadable IMGT databases from version 3.24. It is available for the alleles that are listed on the Allele Frequencies website and can be viewed under:

- Genotyping analysis result screen: by hovering over the information button displayed after the name of each allele candidate.
- Genotyping Sample result screen: by hovering over the information button displayed after the name of each allele candidate.
- Genotyping Sample result screen / Genotype Details pop-up: Containing the most exhaustive list on all available Allele frequency information.
- Gene browser screen: by hovering over the information button displayed under the name of each allele candidate.
- PDF report: by exporting a PDF report from the Genotyping dashboard screen with the *Export allele frequency* box ticked in the wizard.

7 Gene Browser

7.1 Introduction

The Gene Browser allows visual inspection of genomics data. Multiple allele candidates can be browsed together.

With default settings, the following tracks are available in the browser:

- *Position track* - Shows the coordinates for all visible tracks. Numbering starts from one.
- *Phasing track group*:

Phasing track - This track contains annotations for continuously phased regions (aka Phasing regions).

Variants track - Shows the number of overlapping read pairs between two consecutive heterozygous positions (i.e. two position where the two consensus sequences differ from each other). The 'Straight' label shows the number of reads for each consensus that supports the phasing shown in the browser, while the 'Cross' label shows the number of supporting reads for the other possible phasing of the two positions.

- *Consensus sequence 1* - The generated consensus sequence for one of the chromosomes.
- *Coverage depth for consensus 1* - Shows the depth of coverage for every position of the consensus sequence 1 *assembly*.
- *Consensus sequence 2* - The generated consensus sequence for the other chromosome.
- *Coverage depth for consensus 2* - Shows the depth of coverage for every position of the consensus sequence 2 *assembly*.
- *Allele 1 sequence* - Nucleotide sequence of the allele that matches the first consensus the best.
- *Region annotation for allele 1* - Annotations for exons, introns and UTRs are shown for allele 1.
- *Coverage depth track for allele 1* - Shows the depth of coverage for every position of the allele 1 *alignment*.
- *Allele 2 sequence* - Nucleotide sequence of the allele that matches the second consensus the best.
- *Region annotation for allele 2* - Annotations for exons, introns and UTRs are shown for allele 2.
- *Coverage depth track for allele 2* - Shows the depth of coverage for every position of the allele 2 *alignment*.

For novel alleles, two reference tracks are shown: the reference sequence of the novel allele (*Novel ref*) and the reference sequence of the closely related allele (*Rel ref*) from which the novel allele was derived.

Note that consensus sequences and the corresponding short reads can be viewed in the browser, even when no allele match pairs are found.

Additional tracks:

- *Noise track* - Shows systematic noise filtered out during consensus assembly. The noise consensus contains the major nucleotide for every position.
- *Amino acid track* - Shows the amino acid sequence for all allele and consensus sequences, including novel alleles, colored based on amino acid hydrophobicity.

By default, detailed coverage tracks are displayed for the allele alignments, alongside region annotations. The coverage track has a built-in base statistics visualization support: for bases in reads different from the actual consensus/reference base the corresponding coverage depth is shown with the associated nucleotide base color proportionally.

Additional modes for short read tracks

Other than the default *coverage depth mode*, the following alternative short read visualization modes are available for the short read track:

- *Short read mode* - Shows short reads displayed in a stranded fashion, so that forward strand reads (pink) and reverse strand reads (yellow) can be easily distinguished within the display.
- *Fragment mode* - Paired visualization mode that shows the corresponding forward and reverse reads in pairs in the same line. Overlapping sections between read pairs are marked with blue, while non-overlapping reads are connected with a thin line.

In both of the above modes, the short reads track can be *collapsed* which gives a summary view of the short reads (and does not allow each read to be inspected in detail).

7.2 Top panel

On the top of the screen you can see three rows of functions.

7.2.1 Top row

On the left, you can find the *Toggle fullscreen* button which hides the less important parts of the current screen, while on the right you can see the following things:

- the ID of the current user,

- the memory usage widget panel,
- the status panel of the process manager,
- the welcome tutorial button,
- the logout button,
- and the exit button.

In case of the desktop version, the actual memory usage can be observed at the top right corner of the screen. This displays the amount of currently used memory compared to the allowed maximum. Note, that the software might be allocating more memory from the operating system compared to the reported actual memory usage to make further computations faster by keeping some previously allocated memory pieces for later use, thus making the allocation of new memory unnecessary during further computations. By clicking on the memory report widget the software tries to free up as much memory as possible on a best effort basis: this is only an attempt, there are many technical factors which might block returning all memory to the operating system. It is also important to mention that by explicitly asking to free up memory the subsequent computations might become slightly slower, since the option of reusing certain structures allocated before becomes impossible. Only use this option when all planned analyses have been finished and you would like to use other programs more intensively in parallel while investigating the analysis results.

7.2.2 Middle row

There are four navigation buttons that are available on every screen of the tool. These general navigation buttons work very similar to navigation functions available in most widely used internet browsers:

- The *Back* button will take you to the previous screen.
- The *Forward* button will take you to the next screen.
- The *Up* button will move you one level up in the application.
- The *Home* button will take you to the starting screen (i.e. the Genotyping dashboard).

Right from the navigation button, in the information panel, you can find relevant high-level information about the current screen. On the *Genotyping dashboard*, you can see the release date and version of the active IMGT database.

On the far right hand side of the middle row, you can find the bookmark and context specific help buttons. The help tool shows relevant usage information and tips about functions and data available on the current screen of the application and can also be opened by pressing F1.

You can search the Help for any expression by using the Magnifier glass icon displayed on the popup window. Relevant hits are displayed in the left panel where you can select them and open them in the right hand reading pane. You can close the search by clicking the X displayed beside the search field.

7.2.3 Bottom row

The bottom row of the top panel contains a series of buttons which contain the main functions available on the screen.

7.3 Settings and Functions

7.3.1 Filtering the Allele Candidates

By clicking the *Assigned Only* button, unassigned allele candidates can be hidden. By clicking the *Best Matches Only* button, lower ranked allele candidates can be hidden. Consensuses and allele candidates can also be filtered by chromosome, using the *Displayed Allele(s)* function.

7.3.2 Exporting consensus sequences

Consensus sequences shown in the current browser session can be exported in FASTA format using the *Export Consensus* function. You can choose to export the sequences with or without alignment. Exports without alignment contain only the consensus sequences, exports with alignment contain the consensus and the reference sequences. In case of statistical only analyses the result sequences can only be exported without alignment.

7.3.3 Zooming

By default the browser starts in *Drag* mode, where the mouse wheel (or '+' and '-' keys) can be used to zoom in and out, and the display can be moved by clicking and dragging with the mouse. *Drill* mode can also be used, where the region highlighted by a mouse click and drag will be 'drilled into' once the mouse button is released.

7.3.4 Jumping around

You can jump to a position in the reference using *Jump To Position*. If you have previously selected a feature within the display (such as an annotation or short read), then you can jump back to that feature at any time by using the small arrow button at the bottom of the display.

7.3.5 Changing short read track settings

You can display more short reads by using the *collapsed* pile-up view. You can also *page* around the short reads using the small arrows at the top of the short read track. You can set the page size using the up and down arrows, and the current page by using the left and right arrows. This is useful when you have deep sequencing and want to scroll through a few hundred short reads at a time.

7.3.6 Track management

You can configure which tracks you would like to see in the display.

7.3.7 Display setup

You can hide or display strand, indels and soft-clips. The scaling of the coverage track can be changed between an absolute and a relative (global) scale. You can also set whether nucleotide and quality sequences for short reads should be showed on the bottom metadata panel or not.

7.3.8 Exporting Gene browser data

You can *Copy to Clipboard* the details of individual, selected items in the display. You can also copy the reference sequence belonging to a selected item (read or annotation), by clicking on the *Copy reference to clipboard* button.

You can also *Capture a Screenshot* of the currently visible tracks within the Gene Browser. This will create a PNG image file at the chosen location.

7.3.9 Rotating the browser

The vertical Gene Browser view is more useful for comparing multiple allele candidates. The view can be rotated to use a horizontal display, which is a bit more traditional for Genome Browsers, and is more useful when browsing a single allele candidate.

7.4 Context menu

By clicking the right mouse button, you can open the Context menu. In the Gene Browser, this menu contains the following functions:

- *Show base statistics at cursor*: Shows the number of supporting reads for each base at the position of the cursor for all visible short read tracks. Read counts for indels are also shown.
- *Toggle cursor lock*: Locks/releases the cursor.
- *Find sequence*: Searches for a specific nucleotide sequence in the reference of all visible consensus and alleles.
- *Toggle fullscreen*: Toggles between full screen (i.e. only the browser is shown) and the normal view.
- *Toggle reference masked*: Hides all the bases shared by the different allele reference sequences on the screen. This way, only the differences between the allele candidates are shown.
- *Add custom allele(s) to 1st chromosome*: Adds one or more alleles of the currently selected gene from the IMGT database to the first consensus. Reads that can be aligned to this custom allele will be shown in the browser. Note, that allele candidates that are already in the candidate list for that gene cannot be added again.

- *Add allele(s) of other result pair(s) to 1st chromosome:* Adds one or more alleles of the currently selected gene from the not-displayed result candidates to the 1st chromosome. Note, that allele candidates that are already in the candidate list for that gene cannot be added again.
- *Add allele(s) of alternative typing method result to 1st chromosome:* Adds one or more alleles of the currently selected gene from the SG result candidates to the 1st chromosome.
- *Remove custom allele(s) from 1st chromosome:* Removes one or more of the previously added custom alleles from the first consensus.
- *Add custom candidate(s) to 2nd chromosome:* Adds one or more alleles of the currently selected gene from the IMGT database to the second consensus. Reads that can be aligned to this custom allele will be shown in the browser. Note, that allele candidates that are already in the candidate list for that gene cannot be added again.
- *Add allele(s) of other result pair(s) to 2nd chromosome:* Adds one or more alleles of the currently selected gene from the not-displayed result candidates to the 2nd chromosome. Note, that allele candidates that are already in the candidate list for that gene cannot be added again.
- *Add allele(s) of alternative typing method result to 2nd chromosome:* Adds one or more alleles of the currently selected gene from the SG result candidates to the 2nd chromosome.
- *Remove custom allele(s) from 2nd chromosome:* Removes one or more of the previously added custom alleles from the second consensus
- *Remove all custom alleles:* Removes all previously added custom candidates.
- *Convert custom allele pair to result genotype:* the manually added custom alleles can be converted to genotype results which allows them to be displayed and handled as results on other screens including assignment, approval, and result export.

7.5 Metadata Panel

At the bottom of the browser screen you can see the *Metadata Panel* (a.k.a. Bottom Panel). This section of the screen shows information about the currently viewed region and features. The Metadata Panel has three subpanels:

- On the left you can see the start and end positions of the currently shown region, the current cursor position and the per-base coverage at the cursor position.
- On the right, you can see information about the read/annotation at the current cursor location.
- In the middle, you can see metadata about the currently selected read or annotation. This middle subpanel also contains a set of options for exporting information about a single read or annotation. For a detailed explanation about these functions, see the list below!

7.5.1 Metadata Panel functions in the Gene Browser:

- *Copy to clipboard:* Copies the metadata of the selected read/annotation to the clipboard.
- *Copy reference to clipboard:* Copies a region of the reference sequence (or sequences) that overlaps with the selected read or annotation.
- *Jump to selection:* Centers the browser screen on the selected read or annotation.

7.6 Shortcuts

- F2 shows base statistics at cursor position.
- UP/DOWN arrow keys can be used for scrolling alongside the reference contig.
- LEFT/RIGHT arrow keys can be used for scrolling through the reads at the current position/area.
- PAGE UP/PAGE DOWN can be used for scrolling alongside the reference contig in bigger steps.
- HOME/END jumps to the beginning/end of the current reference contig.
- +/- zooms in and out.
- SPACE locks/unlocks the cursor (i.e. green line when unlocked, red line when locked).
- BACKSPACE: Show the whole reference contig (i.e. zooms out totally) or when the cursor is locked, jumps to the locked cursor.
- Mousewheel zooms in/out (in 'Drag' mode) or movea alongside the reference (in 'Drill' mode).
- SHIFT+mousewheel scrolls through the reads at the current position/area.
- Left mouse button can be used for dragging the display (in 'Drag' mode) or for highlighting and zooming in on a region (in 'Drill' mode by clicking and releasing the left button).
- CTRL+F can be used to find a nucleotide sequence in any of the shown consensus or all sequences.
- CTRL+M toggles fullscreen.



- CTRL+D turns the reference mask on and off. (With the mask on, only differences between allele and consensus sequences are shown).
- CTRL+A can be used for adding custom alleles to the first chromosome.
- CTRL+P can be used for adding alleles from other result pairs to the first chromosome.
- CTRL+T can be used for adding results of the alternative typing algorithm to the first chromosome.
- CTRL+R removes custom candidates from the first chromosome.
- CTRL+B can be used for adding custom alleles to the second chromosome.
- CTRL+Q can be used for adding alleles from other result pairs to the second chromosome.
- CTRL+U can be used for adding results of the alternative typing algorithm to the second chromosome.
- CTRL+S removes custom candidates from the second chromosome.
- X switches the browser between horizontal and vertical modes.
- C switches the browser between collapsed and uncollapsed modes.
- T switches the browser between drag and drill navigation modes.

7.7 Manage tracks

In this wizard, you can hide and show any tracks.

The visibility of any track can be set with the *Show Track(s)* and *Hide Track(s)* functions. Hidden tracks are marked with a darker grey color in the *Manage Tracks* wizard track list.

8 Settings dashboard

Reachable from the *Genotyping dashboard* using the *Application settings* button, the *Settings dashboard* displays an overview of the settings in the tool, allows access to administration features and display configurations. Some general information about the current version of the software and the current user is also available on this dashboard.

8.1 General information

There are three blocks of information on the Settings Dashboard:

- Omixon HLA edition: this part contains the name and version of the software, the build identifier with a dedicated copy to clipboard button and some contact and copyright information.
- Omixon HLA edition: this part contains the name, version and reference number of the software, the build identifier with a dedicated copy to clipboard button and some contact and copyright information.
- User info: this part contains the login name, first and last name of the current user.
- License info: this part shows the number of available credits and the expiration date of the license.

8.2 Sidebar

The left sidebar contains the following function sets:

8.2.1 General

In this function group you can set where analysis data and result files are stored, create and manage protocols, set targeted genes for analysis and select the assay version to be used for analysis. For details about protocols see the *Analysis Protocols* help page.

8.2.2 Database

With the *Install New Database* function, you can set up one or multiple versions of the IMGT database used for genotyping. With the *Select Active Database* function, you can specify the active version of the database. Genotyping will always be initiated using the active version. In case you want to remove an installed database you can utilize the *Remove Database* function which allows you to delete any installed database that is not currently selected as active. You can set whether or not to use database extensions in the *Configure Database Extensions* menu.

8.2.3 Administration

With the *User management* option, you can create, edit and remove users. With the *Display Hardware Key* option, you can display an alphanumeric identifier for your computer which can be used for generating a license for that specific machine. The *Upload Licence* option can be used for manually importing a license file into the software.

8.2.4 Automation

This function group allows you to configure automatic analysis on server-client configurations.

8.2.5 Export Settings

Here you can configure LIMS export.

8.2.6 Screen Settings

In this function group you can change the display configurations for the Gene Browser and result screens. Note, that these changes will modify the default behavior and appearance of the software. If you only want to temporarily change browser settings the *Display settings* option on the browser screen should be used. You can also modify the default filters of the Genotyping result screens. Be aware,



that if you unselect a locus in this wizard, results for that locus won't be shown, regardless of the typing results. For both configuration sets, you can set all the parameters back to the default values using the *Restore defaults* function. For details about these settings, please see the following help pages: *Analysis Result Screens* and *Gene Browser*.

8.3 Sample and Result Folders

Users can optionally set where their sample files are in the file system and where the analysis result files should be stored. If the user select a sample for analysis that can be found in the subfolder of the configured base directory then the result file will be created in the subfolder of the base result directory but the same folder hierarchy.

For example, if the configured values are

- base sample folder: /data/samples and
- base result folder: /data/results

and user select the */data/samples/folder1/SampleA_R1.fastq.gz* for analysis. In this case the result will be created in the */data/results/folder1* directory.

Optionally user can set that the result files must be generated in the timestamped subfolders (i.e. */data/results/folder1/20151101_091102*).

If the selected sample is outside of the base directory then the result will be created in the same folder as the sample is.

Note: the timestamped subfolder name format is yyyyMMdd_HHmm.

8.4 Analysis Protocols

Selecting the *Show protocol* option in the right-click context menu of a sample displays the *Protocol configuration* popup. The protocol configuration is the set of parameters which was applied in the selected genotyping run. The protocol of a specific analysis can be saved using the *Save as protocol* option in the *Genotyping dashboard* context menu.

8.4.1 Configure protocol

Selecting the *Show protocol* option in the right-click context menu of an analysis result on the Genotyping dashboard displays the *Protocol configuration* popup. The protocol configuration is the set of parameters which was applied in the selected genotyping analysis run.

8.4.2 Create protocol

Introduction

This function is available from the Analysis Protocols dashboard. You can view, edit and use protocols from here.

Protocols are stored sets of genotyping analysis parameters. Using an existing protocol can speed up the genotyping wizard, as most of the parameters will be set to those stored in the protocol. If you settle on a protocol that works well for your samples, you can save this protocol from any successful run and reuse it again later. You can also edit existing protocols, and create new ones at any time.

The Factory Default protocol is readily available after installation and is recommended to be used for genotyping any samples that have been produced with the Hologate HLA kits. Modification of this protocol is disabled for this reason - to preserve the recommended settings, so you cannot accidentally lose them.

The 'Create New Protocol' option is available anytime to design your own protocols. It's also possible to set any protocol as default.

General

You can name your new protocol here and you can mark it as a 'Default Protocol'.

IMPORTANT

The protocol marked as 'Default protocol' on the Protocol configuration popup will be used for genotyping 'by default' meaning that it will be utilized for every occasion when the 'Analyse' button or the 'Simple Genotyping' button is used. These options don't allow further configuration therefore please be considerate when setting a Protocol - other than the Factory Default - as a default.



Sample data type

You can select the data type (Single data, Paired data, Single data split to multiple files, Paired data split to multiple files. When using paired reads, the two input files are assumed to have the exact matching reads, in the exact same order (there is currently no built-in check for this).

Analysis options

These options dictate where the data has come from and effect how the underlying algorithms deal with the data. Some data sources (e.g. whole genome and whole exome) are more 'noisy', and the algorithms can help to filter out this extra noise.

- *Process all reads*: For whole exome sets, it is recommended to process all the reads (or pairs) because of the lower coverage that is usually found in these kind of data sets.
- *Maximum pairs processed per locus*: Twin uses locus based subsampling to optimize results. This value determines how many reads will be used for each locus. Note, that only reads longer than 75 bases are selected during subsampling (for paired data, both reads in the pair have to be over this minimum length threshold). You can experiment with this value to find the best match for your data. For samples produced with the Holotype assay and other targeted data sets the default 4000 read pairs per locus work well. In case you run into problematic samples increasing the read count can help resolve the issues.
- *Ignore rare alleles*: This option allows you to ignore very rare alleles in the analysis - they will not be included or reported in this case. If this option is not chosen, then these rare alleles will be marked in the results. If this option is greyed out, you will need to run *Install New database* again, in order to import the rare alleles database.
- *Novel allele detection*: This feature detects SNPs and indels in the short read allele alignments. Based on the detected variants and the original reference sequence of the specific allele, novel allele sequences are generated. The names of novel allele candidates contain the allele name of the original allele plus a hashtag and a number. For example, HLA-A*01:01:01:01#1 is a novel allele that has been generated based on the allele sequence of HLA-A*01:01:01:01 and one or more SNPs found in the short read alignment of this allele. Novel alleles have a double reference track in the HLA Genome Browser, both the reference of the novel allele candidate (Novel ref) and the reference sequence of the related allele (Rel ref) (i.e. the sequence of the allele the novel allele candidate originated from) are shown. Novel positions are accepted only when at least 90% of the reads and at least 10-fold coverage depth support the novel reference value. Note, that in many cases multiple similar novel allele candidates are generated from different alleles. There are three versions of the novel allele detection settings: you can choose not to run novel allele detection, you can choose to run novel allele detection only when exon mismatches are present (note, this will try to resolve both exon and intron novelties) or you can choose to run novel allele detection if any mismatches are present (i.e. whole gene novel allele detection).
- *Save read mappings*: Use this setting to save the mapped reads if you would like to browse them in the HLA genome browser.

8.5 Install New Database

This only needs to be run once for every IMGT database version. On the first startup an initial import process automatically imports the latest IMGT database. Additional database files can be downloaded from the Internet and installed by selecting them from the provided list. This should only take a few seconds.

For most of the databases a database extension (i.e. additional sequence information for some of the alleles) is available. Database extensions can be configured using the *Configure Database Extensions* option on the Settings dashboard.

8.5.1 Download problems

If for some reason you cannot download the database archive from within the tool (e.g. due to network security settings) you can download the archives manually and use the *Select local file* option instead of the *Download file* option in the wizard.

The following database versions can be downloaded:

- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.10.0.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.14.0.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.15.0.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.16.0.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.17.0.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.19.0.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.20.0.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.21.0.oxdb.gz



- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.24.0_1.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.24.0_2.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.25.0_3.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.26.0_3.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.27.0_3.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.28.0_4.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.29.0.1_5.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.30.0_5.oxdb.gz
- http://omixon-download.s3.amazonaws.com/dbs/omixon_3.31.0_5.oxdb.gz³

8.6 Select Active Database


With this function, you can select the active IMGT database version. This active version will be used for new genotyping analyses. For existing analyses, the database version used during the previous analysis must be installed but doesn't have to be selected as active, therefore previous results can be viewed, regardless of the active IMGT database version.

If a database extension (i.e. additional sequence information for some of the alleles) is available for the selected database version and extensions are not turned on, a wizard is shown after activating the database, so usage of allele extensions can be turned on or off. Database extensions can also be configured at any time using the *Configure Database Extensions* option on the *Settings dashboard*.

8.7 Remove Database

With this function, obsolete or corrupt versions of the IMGT database can be removed from the software.

8.8 Configure Database Extensions

The database extension contains additional partial sequences for some of the alleles in the IMGT database. These additional sequences are not yet part of the official IMGT database most likely due to technical reasons (e.g. a full allele sequence is not available), but the information they contain have been published and is well known within the HLA community. If the database extension is turned on (i.e. "Use genotype reference extension database?" is checked), the extended sequences are added to the official IMGT database and are used during the genotyping workflow. On the graphical user interface, extended alleles are clearly marked with a *plus sign* .

List of extended alleles in the different databases:

3.24.0_1 and 3.24.0_2:

- DRB4*01:03:01:02N (intron 1 sequence)

3.25.0_3

- C*06:116N (intron 3 sequence)
- DPB1*105:01 (intron 2 sequence)
- DRB1*09:01:02 (intron 2 and intron 3 sequences)
- DRB4*01:03:01:02N (intron 1 sequence)

3.28.0_4

- C*06:116N (intron 3 sequence)
- DRB1*09:01:02 (intron 2 and intron 3 sequences)
- DRB4*01:03:01:02N (intron 1 sequence)

3.29.0.1_5, 3.30.0_5 and 3.31.0_5

- C*06:116N (intron 3 sequence)
- DRB4*01:03:01:02N (intron 1 sequence)

³ http://omixon-download.s3.amazonaws.com/dbs/omixon_3.30.0_5.oxdb.gz



8.9 User Management

User Management allows User accounts to be added, edited and deactivated. There must be at least one 'Super User' at all times. The first registered user automatically becomes the Super User, this user cannot be deactivated.

The following properties must be set for all users:

- User (login name)
- Password
- User role

The following properties are optional:

- First name
- Last name
- Department

There are currently two user roles within the application:

- Analyst: this role is set up for standard activities related to sample analysis including Holotype HLA typing, View/Use Protocol, Export functions.
- SuperUser: authorised to use all functions in the application. On top of the Analyst role's permissions it also includes Administrative functions, Advanced HLA typing, Import function and Create/Save Protocol.

8.10 Upload Licence

Use this function to upload a new license which extends the number of credits or time available in the application. Please make sure that you do not modify the license file during download. Simply save it to your machine then browse and select the file for uploading then click Finish to complete the operation. A pop up message informs you if the upload was successful. Afterwards the new credits and available time will be immediately displayed on the Settings screen. In case the upload is not successful an error message informs you about the possible causes. If the license has accidentally been modified you can simply download it again from the provided link and repeat the above steps. If the hardware key does not seem to match please check that the key under "Settings / Display Hardware Key" matches the one previously sent to the Omixon support team.

If the correct key has been provided or you get a different error message or have any questions regarding the described process please contact support@omixon.com⁴. For detailed information about licensing and pricing, please contact sales@omixon.com⁵.

8.11 Typer Server Management

Only applicable in server-client environment. In this menu you can view connection port, ID and availability of the *Typer Servers* connected to the *Omixon Server*.

8.12 Configure Automatic Analysis

The folder structure generated by the Illumina machines is the following:

```
/root/data/miseq/Run1/Data/Intenstities/BaseCalls/*.fastq.gz
```

When Automation service is set up it watches the configured analysis folders that contain the files generated by the Illumina Local Run Manager software.

By default automation periodically checks the folders and watches the *CompletedJobInfo.xml* files that are generated after sequencing runs are completed. These files are located in the root level of the run folder. To override the default behavior you have the option to supply up to three hour values separated by commas. When doing so automation only checks the folders once at each given time.

You can configure Automation from the Application Settings *Configure Automatic Analysis* menu.

Based on the example folder structure above the configuration fields should be set like this:

Parent directory for all run folders: /root/data/miseq/

⁴ <mailto:support@omixon.com>

⁵ <mailto:sales@omixon.com>



Parent directory is the root folder where the analysis results are generated to. Please make sure you enter your own parent directory as it is specific to your computer.

Sample location inside each run folder: Data/Intenstities/BaseCalls

Sample location folder is the default subfolder defined by the Illumina's Local Run Manager software.

Set the *Autostart service* flag if you want Automation to start automatically when restarting the Omixon Server.

On the *Select automation protocol* screen you can select the protocol to be used by the Automation. If no protocol was selected the default protocol is used.

8.13 HPRIM Export Configuration

If Omixon HLA is configured to have a LIMS connection you can use this wizard to setup the default export folder for the LIMS export function.

8.14 Gene Browser

You can define the starting (i.e. used when the software is freshly started) graphical and navigation settings for the browser here. These settings can be saved and will still be in effect after restart. The browser settings are user specific.

8.14.1 Display configuration

On this screen the orientation of the browser screen (horizontal or vertical), the navigation mode (*Drag & Zoom* or *Drill & Scroll*) and the short read pileup visualisation settings can be changed.

8.14.2 Color setup

On this screen, the color codes for nucleotides used in the browser can be changed.

8.14.3 Data options

On this screen you can select whether inserts or deletions should get an extra mark in zoomed out views of the browser. Soft clips and strand information can be shown or hidden as well. The default scale for the coverage tracks can also be changed on this screen.

8.14.4 Metadata options

With these settings, you can show/hide read sequences and quality information on the metadata panels in the browser (i.e. the middle and right subpanels in the bottom browser panel).

8.14.5 Advanced settings

The zooming resolution of the Gene browser can also be set on this screen by changing the *Overview factor*. This is a the maximum number of nucleotides displayed per pixel. The default setting is 20. The minimum number of pixels for collapsed read width can be set as well.

8.15 Analysis Result Screens

In this wizard you can set the default values for the genotyping result screens. You can select the genes that should be shown by default. You can also set which allele candidate categories should be visible on the result screens by default.

9 Process manager

The Process manager always displays the actual status of the currently running Task in the top right corner of the screen. The Process manager screen can be opened by clicking on the panel at any time.

Tasks represent analyses that have been setup and are - or were - under processing. Tasks are shown as a list with the latest being on top by default. The information on these tabs is available as long as it's not deleted manually.

9.1 Colour coding

- Running and paused tasks are displayed in yellow.
- Submitted tasks which are being queued to be processed are displayed in blue.
- Failed tasks are red.
- Completed tasks are green.
- Aborted tasks are also blue.

9.2 Handling Tasks

Click on the tick available on the left side to select a Task. Multiple items can be selected.

9.2.1 Abort Task

After selection any Submitted currently running Task can be aborted.

9.2.2 Delete Task

Tasks related to analyses store the relevant audit details therefore removal of Tasks have been disabled to provide compliance with audit requirements.

9.2.3 Pause and Resume Task

Running tasks can be temporarily stopped and later resumed. This function is not available in client-server environment.

9.2.4 Find a Task

Click on the Task Filtering button to reduce the number of displayed Tasks by searching for the required items based on Date or Status.


Show task properties

Click on a task to show it's properties. Use the *Copy to clipboard* button to copy the contents of the selected tab page to the clipboard.

10 Detailed automation guide

10.1 Overview

Omixon HLA Twin provides an integrated automated genotyping solution. Automation can be configured and managed from the user interface.

 Automation is only available in client-server mode. You need to have admin permission if you would like to configure and manage it.

10.2 Configuration

The folder structure generated by the Illumina machines is the following:
`/root/data/miseq/Run1/Data/Intenstities/BaseCalls/*.fastq.gz`

When Automation service is set up it watches the configured analysis folders that contain the files generated by the Illumina Local Run Manager software.

By default automation periodically checks the folders and watches the *CompletedJobInfo.xml* files that are generated after sequencing runs are completed. These files are located in the root level of the run folder. To override the default behavior you have the option to supply up to three hour values separated by commas. When doing so automation only checks the folders once at each given time.

You can configure Automation from the Application Settings *Configure Automatic Analysis* menu.

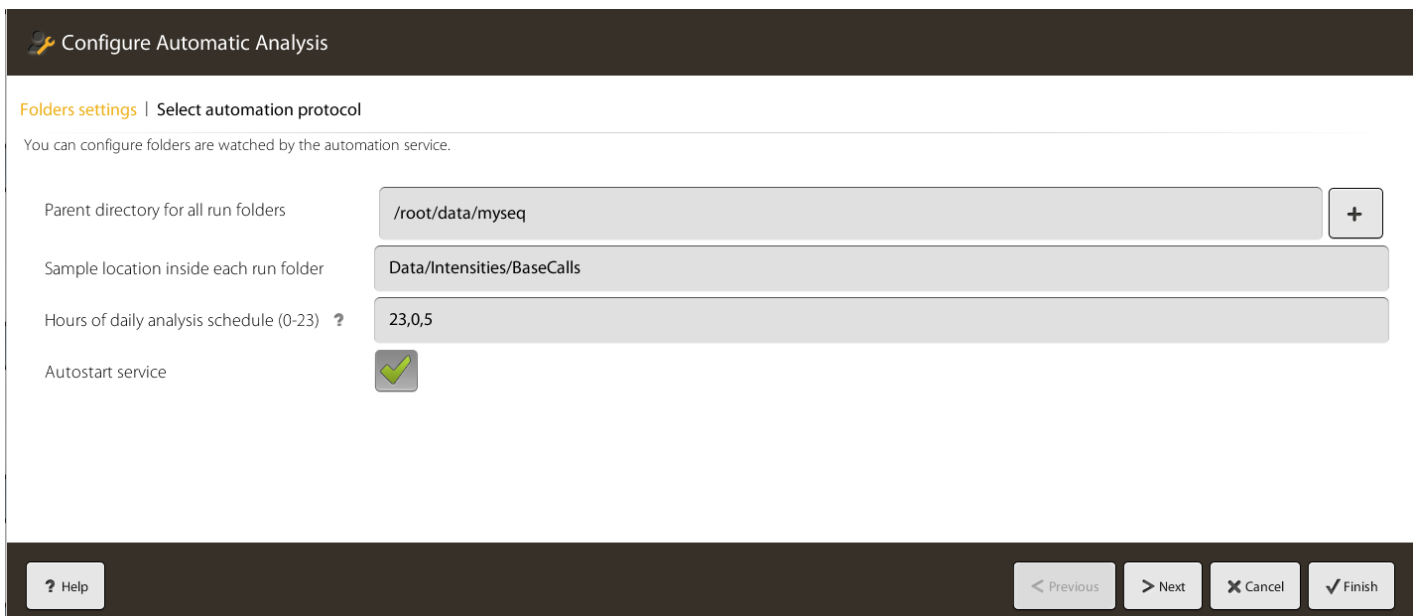
Based on the example folder structure above the configuration fields should be set like this:

Parent directory for all run folders: `/root/data/miseq/`

Parent directory is the root folder where the analysis results are generated to. Please make sure you enter your own parent directory as it is specific to your computer.

Sample location inside each run folder: `Data/Intenstities/BaseCalls`

Sample location folder is the default subfolder defined by the Illumina's Local Run Manager software.



If you want Automation to start automatically when you restart the Omixon Server then set the *Autostart service* flag.

On the *Select automation protocol* screen you can select the protocol that should be used by the Automation. If no selection then the default protocol is used.

Configure Automatic Analysis

Folders settings | Select automation protocol

You can configure folders are watched by the automation service.

Parent directory for all run folders	<input type="text"/>	<input data-bbox="1412 454 1465 506" type="button" value="+"/>
Sample location inside each run folder	<input type="text" value="Data/Intensities/BaseCalls"/>	
Hours of daily analysis schedule (0-23) ?	<input type="text"/>	
Autostart service	<input type="checkbox"/>	

You can optionally delay the automation if your analysis files are written out in undetermined order - for example the *CompletedJobInfo.xml* marker file is written out before the last sample file is. In this case maybe not all of your sample batch is being analyzed. Setting the - *Domixon.automation.delay=120* parameter in the *omixon-hla-twin-server.vmoptions* file is used for delaying the automation by 120 sec.

10.3 Management

Automation service can be started or stopped manually from the Application Settings *Manage Automation Service* menu.

Automation service

You can start or stop automation service

<input data-bbox="137 1415 352 1525" type="button" value=" Start service"/>	<input data-bbox="932 1415 1114 1525" type="button" value=" X Close"/>
---	--

10.4 Logging

Automation generates log entries into the *OMIXON_TWIN_INSTALL/logs/automation.log* file.



11 Notifications

11.1 IMGT Database

In accordance with ASHI guidelines Omixon HLA displays a warning message if the IMGT Database used for genotyping is older than 1 year. In order to switch to a more up-to-date database you can use the database management functions available on the Application Settings Dashboard.

11.2 License Expiry

In order to help you keep your software running warnings are provided as the license approaches its expiry date. A notification will be displayed one month to license expiry and after every login within this time period. When the notification is shown, you can choose to hide this message and it will not be shown again or you can choose to proceed without turning the notification off.

11.3 Extended allele notification

When a new database is installed in which extended allele sequences are available a notification is shown for allowing the usage of extended allele sequences. Note, that the usage of extended alleles is suggested as they can help resolving common ambiguities. You can find the list of extended alleles in the Configure Database Extensions section.

11.4 Deviating from the Factory default protocol

A notification is shown when not the Factory default is used. Please be aware, that the use of the Factory Default protocol is strongly suggested for analyzing Holotype data as all the parameters are fine tuned for this data type.



12 HTTPS support

In client-server Twin installations HTTPS communication can be enabled instead of the default HTTP.

In order to use secure communication, public and private key and certificates must be generated for the server as well as for the client(s). The simplest way to generate keys and certificates is to use the `keytool` application that comes with the JDK, as it generates keys and certificates directly into the keystore.

Using the Linux script below you can generate all the required files:

Keystore generation on Linux

```
#!/bin/sh
# script for re-generating the keystores and truststores needed for Omixon Twin HTTPS support

JAVA_HOME=$HOME/OmixonHlaTwin/jre
PATH=$JAVA_HOME/bin:$PATH
# Create server keystore - file omixon-server.keystore
keytool -genkey -v -alias omixon-server -keyalg RSA -keysize 1024 -keystore omixon-
server.keystore -validity 365 -keypass changeit -storepass changeit -dname "CN=Omixon,
OU=Omixon"

# Export Server's Public Key - file omixon-server.cer
keytool -export -keystore omixon-server.keystore -alias omixon-server -file omixon-server.cer
-keypass changeit -storepass changeit

# Export Client Key Store - file omixon-client.keystore
keytool -genkey -v -alias omixon-client -keyalg RSA -keysize 1024 -keystore omixon-
client.keystore -validity 365 -keypass changeit -storepass changeit -dname "CN=Omixon,
OU=Omixon"

# Exporting Client's Public Key - file omixon-client.cer
keytool -export -keystore omixon-client.keystore -alias omixon-client -file omixon-client.cer
-keypass changeit -storepass changeit

# Importing Client's Public key into server's truststore
keytool -import -v -trustcacerts -alias omixon-client -file omixon-client.cer -keystore
omixon-server.keystore -keypass changeit -storepass changeit -noprompt

# Importing Server's Public key into client's truststore
keytool -import -v -trustcacerts -alias omixon-server -file omixon-server.cer -keystore
omixon-client.keystore -keypass changeit -storepass changeit -noprompt
```

This is the corresponding Windows generator script:



Keystore generation on Windows

```
@echo off
SET JAVA_HOME=%HOMEDRIVE%%HOMEPATH%\OmixonHlaTwin\jre
SET PATH=%JAVA_HOME%\bin;%PATH%

REM Create server keystore - file omixon-server.keystore
keytool -genkey -v -alias omixon-server -keyalg RSA -keysize 1024 -keystore omixon-
server.keystore -validity 365 -keypass changeit -storepass changeit -dname "CN=Omixon,
OU=Omixon"

REM # Export Server's Public Key - file omixon-server.cer
keytool -export -keystore omixon-server.keystore -alias omixon-server -file omixon-server.cer
-keypass changeit -storepass changeit

REM # Export Client Key Store - file omixon-client.keystore
keytool -genkey -v -alias omixon-client -keyalg RSA -keysize 1024 -keystore omixon-
client.keystore -validity 365 -keypass changeit -storepass changeit -dname "CN=Omixon,
OU=Omixon"

REM # Exporting Client's Public Key - file omixon-client.cer
keytool -export -keystore omixon-client.keystore -alias omixon-client -file omixon-client.cer
-keypass changeit -storepass changeit

REM # Importing Client's Public key into server's truststore
keytool -import -v -trustcacerts -alias omixon-client -file omixon-client.cer -keystore
omixon-server.keystore -keypass changeit -storepass changeit -noprompt

REM # Importing Server's Public key into client's truststore
keytool -import -v -trustcacerts -alias omixon-server -file omixon-server.cer -keystore
omixon-client.keystore -keypass changeit -storepass changeit -noprompt
```

IMPORTANT

Please make sure you update the *OmixonHlaTwin* parameter in the beginning of the script with the actual name of the installation folder in case it is different from the default name.

To use the above scripts substitute the *changeit* value in the *-keypass* attribute with your actual password. Validity time can be configured with the *-storepass* attribute - default validity time is 365 days

In the *vmoption* file of the Omixon Twin server you have to set the following parameters:

Server settings

```
-Domixon.keystore=/path/omixon-server.keystore
-Domixon.keystore.password=changeit
-Domixon.truststore=/path/omixon-client.keystore
-Domixon.truststore.password=changeit
```

where *path* refers to the folder where the server keystore files are.



On the client side you have to set the following details in the client *vmoptions* file:

Client settings

```
-Domixon.truststore=/path/omixon-client.keystore  
-Domixon.truststore.password=changeit
```

13 Command Line Interface

13.1 Overview

The application's Command Line Interface (CLI) allows you to execute some functions by calling them remotely. We support Python script execution using Jython that is an implementation of the Python programming language designed to run on the Java platform.

13.2 How does it work?

CLI can be used only in client-server mode, in the desktop version network layer of the application is not used. CLI commands can only be executed when the server is running and network connection is available.

Each command type has its own parameters that must be provided by the user. If one of the mandatory parameter is missing, the command is not executed and an error is reported in the console. Business functions can be called by authorized users, the authentication parameters are part of the CLI command (see the *Command types* section).

There are two main components in the CLI: *ScriptingTool* and *OmixonTools*. The latter component defines the API that can be called from Python, the former one is used when system related commands should be executed directly (i.e. HLA Database installation).

The *ScriptingTool* redirects the command execution to the proper command wrapper implementation. The given command wrapper contains all business logic - i.e. parsing the parameters, calling server side functions in the right sequence, etc.

The *OmixonTools* implementation can be called from Python scripts directly that are interpreted and executed by Jython.

Phases of the command execution are logged on the console, this way user is informed continuously about the status of the calling.

The preferred console type can be set as a command parameter:

- color (COLOR) - default
- black-and-white (BW)
- logger

In the last case, the application's messages are redirected to a log file. When the command execution is finished, the user is notified.

13.3 Command types

The following section describes the supported command types.

13.3.1 Initialize

Can be used for creating the super user if there is no available already registered user.

Initialize command syntax

```
initialize -host localhost -port 4380 -username admin -password admin -console COLOR
```

where

- *host* and *port* define address of the application server
- *username* and *password* define the authentication parameters of the super user to be created
- *console* defines the console type (optional).

13.3.2 License file import

Supports importing a licence file from command line.

License command syntax

```
license -host localhost -port 4380 -username admin -password admin -license /path/omixon-server.lic -console COLOR
```

where

- *username* and *password* define the authentication parameters of the given user who tries to execute action
- *license* defines the license file path.

13.3.3 HLA Database listing

Just run the `hladb` command without `-install` or `-change` argument. e.g.:

HLA Database listing

```
hladb -username admin -password admin -host localhost -port 4380

* INFO: Available databases: 3.10.0, 3.14.0, 3.15.0, 3.16.0, 3.17.0, 3.19.0, 3.20.0, 3.21.0,
3.24.0_1,
* INFO: Installed databases: 3.10.0, 3.21.0, 3.24.0_1,
```

13.3.4 HLA Database installation

IMGT databases can be installed by version number - in this case public network should be accessible by the server - alternatively, database bundles can be imported from local files.

HLA Database install by version command syntax

```
hladb -host localhost -port 4380 -username admin -password admin -install file /home/user/db/omixon_3.29.0.1_5.oxdb.gz
or
hladb -host localhost -port 4380 -username admin -password admin -install internet 3.29.0.1_5
```

where

- *username* and *password* define the authentication parameters of the given user who tries to execute action
- *install file omixon_3.29.0.1_5.oxdb.gz* is the operation and its argument to install a database from the specified file

or

- *install internet 3.29.0.1_5* is the operation and its argument for downloading and installing the specified database version from the internet.

13.3.5 Change HLA Database version

Changes the active database version. This version of the IMGT database bundle will be used for all analyses.

Change HLA Database version

```
hladb -username admin -password admin -host localhost -port 4380 -change 3.21.0

* INFO: Successfully connected to server localhost on port 4380
* INFO: hla.dat.date: 31/05/2016, hla.dat.version: 3.24.0_1
* INFO: Changing db version from 3.24.0_1 to 3.21.0
```

13.3.6 List assay versions

Lists the available assay versions and shows which one is currently selected for analysis.

List assay versions

```
assay_version -username admin -password admin -host localhost -port 4380 -list

* INFO: Available assay versions:
* INFO: Display name      Version code      Selected
* INFO: Holotype HLA v1 kit      0
* INFO: Holotype HLA v2.1 kit    1                X
```

13.3.7 Set assay version

Sets the assay version for the analysis. The version code can be obtained from the output of the *assay_version -list* command.

Select assay version

```
assay_version -username admin -password admin -host localhost -port 4380 -version 1

* INFO: Successfully connected to server localhost on port 4380
* INFO: hla.dat.date: 09/05/2017, hla.dat.version: 3.28.0_4
* INFO: Assay version has been set.
```

13.3.8 Python script execution

The following API of the *OmixonTools* can be called from Python scripts:

- disconnect from the server
- register an observer that will be notified about the server events
- call files - FASTQ or BAM - filtering
 - using single or paired FASTQ files
 - by custom configuration
 - by supplying a sample folder path - with this option batch filtering will be scheduled for all sample files in the folder
- starting paired and single FASTQ files typing
 - using single or paired FASTQ files
 - by custom configuration

- by supplying a sample folder path - with this option batch typing will be scheduled for all sample files in the folder
- schedule big data analysis
 - using single or paired FASTQ files
 - by custom configuration

Python scripts can be executed using the following syntax.

Python script execution syntax

```
run example.py -host localhost -port 4380 -username admin -password admin -args directory=/home/username/fastq_data_folder/ file1=sample_file1.fastq.gz file2=sample_file2.fastq.gz config=/home/username/hla.conf
```

where

- *run* is followed by the python script that should be executed
- *args* are arguments for the script; eg: -args arg1=human arg2=1024 arg3=13,71,49 (access in the script with Arguments.get('arg1',')) etc.

Supported arguments

- *directory*: where the sample files are located
- *file1* and or *file2*: names of the sample files
- *config*: in case of custom filtering or genotyping a configuration file must be specified

13.3.9 Status of the server tasks

Users can fetch the status of the scheduled and executed/aborted tasks using the following syntax.

Server status command syntax

```
status -taskstatus ALL -detail DETAIL -host localhost -port 4380 -username admin -password admin -console COLOR
```

where

- *taskstatus* is comma separated task status [ALL, CREATED, SUBMITTED, SUCCEEDED, FAILED, ABORTED]
- *detail* defines the level of detail [DETAIL, SUMMARY, BOTH].



14 Omixon Twin Admin command line interface (Admin CLI)

In some cases system or user related data should be changed or cleaned in the Omixon Twin/Explore application. For example a user forgets his password that should be reset or application database is relocated to a different host and there is a license issue that should be fixed. These issues can be fixed using the Admin CLI that is part of the installed application.

14.1 Important notes

- Before you execute any commands make sure that the Omixon Desktop or Server application is up and running on the given host.
- Also make sure that any other installations are not running. Execute the commands from the installation directory of the running instance.
- In case the database you are trying to access has been created with a non-default name, make sure you provide the database name that you are trying to work with.

The following command types are supported:

- list-users
- reset-pwd
- clean-lic
- activate-user

Running the script with no parameters will list the supported commands:

```
username@machine ~/OmixonHlaTwin $ ./omixon-hla-twin-admin-cli
```

Usage: <main class> [options] [command] [command options]

Commands:

```
list-users    List registered users
Usage: list-users [options]
Options:
  -dbname
      Omixon database name
      Default: .omixon-hla-twin
  -dbpassword
      Omixon database password
      Default: omixon
  -dbuser
      Omixon database user
      Default: omixon

reset-pwd     Reset user password
Usage: reset-pwd [options]
Options:
  -dbname
      Omixon database name
      Default: .omixon-hla-twin
  -dbpassword
      Omixon database password
      Default: omixon
  -dbuser
      Omixon database user
      Default: omixon
* -password
      New password
```



```
* -user
    Username that password should be set

clean-lic    Clean uploaded licences
Usage: clean-lic [options]
Options:
  -dbname
    Omixon database name
    Default: .omixon-hla-twin
  -dbpassword
    Omixon database password
    Default: omixon
  -dbuser
    Omixon database user
    Default: omixon

activate-user Activate disabled user
Usage: activate-user [options]
Options:
  -dbname
    Omixon database name
    Default: .omixon-hla-twin
  -dbpassword
    Omixon database password
    Default: omixon
  -dbuser
    Omixon database user
    Default: omixon

* -user
    Username that should be activated
```

where *dbname*, *dbpassword*, *dbuser* are optional, but the mandatory parameters are marked with star (*).

14.2 Password reset

The application's users credentials are stored in the OrientDB that is run in the desktop or in client-server installation in the server. User passwords can be reset by executing the following command:

```
username@machine ~/OmixonHlaTwin $ ./omixon-hla-twin-admin-cli reset-pwd -
password blankit -user test
```

14.3 Cleaning licences

Installed licenses can be removed from the system by executing the following command:

```
username@machine ~/OmixonHlaTwin $ ./omixon-hla-twin-admin-cli clean-lic
```

You can check on the *Settings dashboard* screen of the application where no available license credits should be shown after license removal.

14.4 Activate disabled user

Users that have been disabled are not displayed in the *User Management* menu in the software.

Users can be enabled by executing the following command:



```
username@machine ~/OmixonHlaTwin $ ./omixon-hla-twin-admin-cli activate-user -  
user test
```



15 Application performance tuning

The Omixon HLA application is implemented in the Java language. The installed Omixon HLA application contains the optimal configuration for all use cases but there are some tweaks that can be useful in case the application does not perform as well as expected. Sometimes application tuning requires only changing JVM options but there are other options. This document provides some tips and tricks that can help you to improve application performance.

Run fewer programs at the same time

Genotyping is a CPU intensive task and requires more memory than a "normal" application.

Close unused programs while Omixon HLA application runs to free up CPU and memory.

Use SSD instead of HDD

Sample and analysis result files are larger than a "normal" file, so it is crucial to decrease the I/O operations on a physical disk.

Use Solid-state drives (SSD) that are inherently faster than a hard disk drive (HDD).

- File opening speed: SSD up to ~30% faster than HDD
- File copy / write speed: SSD generally above 200 MB/s and up to 550 MB/s for cutting edge drives, and in case of HDD the range can be anywhere from 50 – 120MB/s.

Local disk vs network storage

The Omixon HLA application massively reads and writes mounted physical disks where your samples and result files are stored. Storages have inherent *latency* that refers to the increasing amount of time it takes to access a given file on the hard disk drives but in case of Network-Attached Storages (NAS) we have to calculate the distances between the files and the applications. These distances can cause delays over a wide-area-network (WAN) because the data is stored far away from where it is being used.

If possible, use local disks instead of NAS for minimizing this latency.

Too many files in one folder

When you open a folder, the Omixon HLA application parses the files - sample and result ones - before generating the file list for the user. If the folder contains too many files, the execution time is longer.

If possible, store a maximum of around 100 samples and 100 result files in one folder.

You can use file filters to get better response time, because less information is generated and sent by the application to the clients.

Distribute load on multiple CPUs

Genotyping is a CPU intensive task and by default all available CPUs are calculated and roughly 25% of available threads are kept free to provide sufficient resources for the Omixon HLA reporting functions while the genotyping process is running. Optionally, you can set the `hla.general.threadLimit` parameter in the `vmoption` file of the application to change this default behavior. If this value was set to 0, all CPUs are bound to the genotyper. If the value is set below 0 thread limit it is dynamically calculated by the application i.e. at least 2 CPU cores should not be used by the genotyper.

Set up targeted loci

You can decrease the execution time of the genotyping process if you set the targeted loci manually to match the set of loci that was sequenced. In this case filtering and analysis will only be executed for the configured loci. The advanced typing wizard can be used for specifying the list of targeted loci for a single analysis task.

Set up per locus read count

Using the default read count for Holotype samples is strongly recommended as these read counts were determined based on the characteristics of the assay and targeted loci. Running with a higher read count is only suggested for problematic loci. Note, that higher read counts can increase the memory requirements significantly.



16 Omixon database relocation

Omixon HLA application metadata is stored in the Omixon database that is created by the application on the fly when it is being started at very first time and updated automatically in case a new version is installed.

The Omixon database folder must be set during installation but it can be relocated later - i.e. because of a new disk should be used for storage.

Database relocation steps are:

- stop the Omixon HLA application - desktop or server
- move the database to the new folder
- change the *-Dogve.data.dir* application parameter value to the new one in the *vmoption* file
- start the Omixon HLA application.

! The Omixon application database should be located on the same host where the Omixon HLA application (desktop or server) is running.
If you move the database to a new host - i.e. the Omixon HLA application is installed on a new machine - the database can be relocated, but a new license should be installed. Before you upload the new license you should clean the previous one using the Admin CLI.

17 Glossary

- **alignment score:** a quality score that is calculated for every alignment of a short read. The alignment score is basically based on the mismatches between the read and reference sequences.
- **allele:** a sequence version of a gene.
- **allele search:** the selection of allele pairs based on the generated consensus sequence and the IMGT/HLA database (Robinson et al. 2015).
- **ambiguity:** multiple alleles or allele pairs are called. For the statistical genotyping method these are supported by a nearly equal amount of reads, while in the consensus genotyping results multiple allele pairs can have the equal amount of mismatches due to e.g. off-target differences between the alleles or unresolved phase in the consensus.
- **assembly:** aligning and merging the fragments or short reads produced from a longer sequence in order to reconstruct the original sequence.
- **CG:** consensus based HLA genotyping method.
- **contig:** contiguous sequence, the result of de novo short read assembly.
- **continuous alignment:** a short read alignment that is generated for continuous allele region groups or consensus contigs.
- **de novo assembly:** short read assembly that doesn't use any information provided by a reference sequence (or sequences).
- **genotype:** one (or more) allele pair(s) for an HLA locus in a single sample.
- **interloci crossmapping:** reads map to multiple HLA loci (e.g. to HLA-DRB1 and HLA-DRB3).
- **locus (plural: loci):** the location of a gene.
- **long range PCR:** long range polymerase chain reaction is a type of PCR that was optimized for amplifying long (~2 to ~15 kb) DNA fragments.
- **mismatch:** difference between the consensus and reference allele sequences.
- **MSA:** multiple sequence alignment. An alignment of three or more biological sequences.
- **novelty:** a mismatch that haven't been previously described in the allele definitions of the IMGT/HLA database (Robinson et al. 2015).
- **PCR crossover artifact:** reads generated from a hybrid PCR product that contains information from multiple PCR amplicons often derived from different copies of the chromosome (e.g. see Holcomb et al. 2014).
- **phasing:** separation of signals coming from the two copies of the chromosomes (in this case chr6).
- **phase break:** a position between two fully phased subsequences between which phasing cannot be resolved (e.g. due to a long homozygous stretch or PCR crossover artifacts).
- **pooled (sample):** after the long range PCR step of the protocol, the amplicons of the different loci are mixed together. The result is a single pair of fastq files that contain reads from all targeted loci.
- **QC measures:** quality control measures that can indicate sequencing, data and result quality issues.
- **QC alignment:** short read alignment for the alleles that were selected based on the consensus.
- **random noise:** randomly distributed sequencing errors in the short reads.
- **region:** gene region: exon, intron or UTR.
- **SBT:** sequence based typing. Usually provides low resolution HLA types, as in most protocols, only exon 2 and exon 3 are sequenced using Sanger sequencing by capillary electrophoresis.
- **segmented alignment:** a non-continuous short read alignment that is generated independently for each region.
- **serological typing:** antibody based HLA typing method. Only allows low resolution typing.
- **SG:** statistics based HLA genotyping method.
- **splicing:** is a modification of the pre-mRNA in which introns are removed and exons are joined.
- **splicing ambiguity:** in some cases based solely on the consensus sequence and the allele definitions in the IMGT/HLA database the exact gene region coordinates for the consensus cannot be determined.
- **SSO:** sequence specific oligonucleotide based HLA typing. Uses hybridization probes. Only provides low resolution typing.
- **SSP:** sequence specific primer based HLA typing. PCR based technique that uses a list of primer pairs specific for alleles or allele groups. Only provides low resolution HLA typing.
- **strand bias:** an imbalance between the number of reads mapping to the forward and reverse strand.
- **systematic noise/artifact:** a group of reads that contain a systematic error that differs from both consensus sequences.
- **targeted locus (plural: targeted loci):** a gene or a set of genes for which sequencing was attempted.
- **unpooled (sample):** sequencing data generated using the long range PCR amplicons of a single locus. The result is a single pair of fastq files that contain reads from a single targeted locus.
- **variant support:** the number of reads supporting the reference and alternate bases at a heterozygous position.

18 List of shortcuts

18.1 Generic shortcuts

- F1 key opens the Help for the current page or wizard.
- F8 key closes the window and exits the application
- F11 key switches to fullscreen display mode to utilize as much space as possible for visualization - note that certain platforms and window managers are not working properly together with this function, the first time you use you will get a warning message about this.
- ALT+F4 closes the application.
- CTRL + C/V/X copies/pastes/cuts texts or files.
- SHIFT/CTRL + left click can be used for multiple selection/deselection
- Left/Right/Up/Down arrow buttons can be used for scrolling on the screen if necessary.
- DEL can be used for deleting the selected item(s).
- F9 or CTRL-PLUS: scale up window contents
- F10 or CTRL-MINUS: scale down window contents
- F12 or CTRL-ZERO: reset window contents scaling

18.2 Genotyping dashboard

- CTRL+A can be used for selecting all samples and analyses.
- CTRL+F opens the file filter wizard.
- CTRL+R expands/reduces sample and analysis names.
- CTRL+H shows/hides previous analyses.
- DEL deletes the selected items.

18.3 Genotyping analysis result

- CTRL+C, after the selection of a sample, this shortcut copies the selected genotype to the clipboard.
- F2 key puts the cursor to the next sample with a failed locus.
- F3 key puts the cursor to the next sample with a discordant locus.
- F4 key puts the cursor to the next sample with a QC failed locus.
- CTRL+mousewheel can be used for zooming in and out.

18.4 Genotyping sample result

- CTRL+C, after the selection of a genotype, this shortcut copies the selected genotype to the clipboard.

18.5 Gene browser

- F5 jumps to the same locus of the previous sample.
- F6 jumps to the same locus of the next sample.
- F2 shows base statistics at cursor position.
- UP/DOWN arrow keys can be used for scrolling alongside the reference contig.
- LEFT/RIGHT arrow keys can be used for scrolling through the reads at the current position/area.
- PAGE UP/PAGE DOWN can be used for scrolling alongside the reference contig in bigger steps.
- HOME/END jumps to the beginning/end of the current reference contig.
- +/- zooms in and out.
- SPACE locks/unlocks the cursor (i.e. green line when unlocked, red line when locked).



- BACKSPACE: Show the whole reference contig (i.e. zooms out totally) or when the cursor is locked, jumps to the locked cursor.
- Mousewheel zooms in/out (in 'Drag' mode) or move alongside the reference (in 'Drill' mode).
- SHIFT+mousewheel scrolls through the reads at the current position/area.
- Left mouse button can be used for dragging the display (in 'Drag' mode) or for highlighting and zooming in on a region (in 'Drill' mode by clicking and releasing the left button).

- CTRL+F can be used to find a nucleotide sequence in any of the shown consensus or all sequences.
- CTRL+M toggles fullscreen.
- CTRL+D turns the reference mask on and off. (With the mask on, only differences between allele and consensus sequences are shown).
- CTRL+A can be used for adding custom alleles to the first chromosome.
- CTRL+P can be used for adding alleles from other result pairs to the first chromosome.
- CTRL+T can be used for adding results of the alternative typing algorithm to the first chromosome.
- CTRL+R removes custom candidates from the first chromosome.
- CTRL+B can be used for adding custom alleles to the second chromosome.
- CTRL+Q can be used for adding alleles from other result pairs to the second chromosome.
- CTRL+U can be used for adding results of the alternative typing algorithm to the second chromosome.
- CTRL+S removes custom candidates from the second chromosome.
- X switches the browser between horizontal and vertical modes.
- C switches the browser between collapsed and uncollapsed modes.
- T switches the browser between drag and drill navigation modes.

19 Acknowledgements

19.1 Collaborators

We would particularly like to thank Dr. Derek Middleton for providing access to the Allele Frequency Net database.

19.2 Third party tools and databases

The IMGT/HLA database is used for HLA genotyping (1, 2).

Linkage disequilibrium data used in the software was derived from NMDP Registry Haplotype Frequencies database (3) and from Wikiversity (4).

Allele frequency information was derived from the Allele*Frequency Net database (5).

A number of tools from SamTools Picard are used within the software for handling and manipulating SAM and BAM files (6, 7).

19.3 Citations

1. Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SGE: The IMGT/HLA Database. *Nucleic Acids Research* (2011), **39 Suppl 1**:D1171-1176.
2. Robinson J, Malik A, Parham P, Bodmer JG, Marsh SGE: IMGT/HLA - a sequence database for the human major histocompatibility complex. *Tissue Antigens* (2000), **55**:280-287.
3. Gragert L, Madbouly A, Freeman J, Maiers M: Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Human Immunology* (2013), **74(10)**: 1313-1320. <http://dx.doi.org/10.1016/j.humimm.2013.06.025>.
4. https://en.wikiversity.org/wiki/Genetics/Human_Leukocyte_Antigen/Linkage_Disequilibrium/DR-DQ_Blocks - Data exported: 2016.04.21
5. Gonzalez-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, Silva AL, Silva AL, Ghattaoraya GS, Alfirevic A, Jones AR and Middleton D: Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acid Research* (2015), **28**:D784-788.
6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Processing Subgroup: The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* (2009), **25**:2078-2079.
7. Li H: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* (2011), **27**:2987-2993.