

Instructions for entering & uploading Typing report

17th International HLA & Immunogenetics Workshop Database

*Note: To enter sample's profiles and pedigrees, see **Instructions for creating datasets, entering sample's profile & pedigrees, 17th International HLA & Immunogenetics Workshop Database.** ("Instructions to Enter IHIWS Sample's Profile")*

Contents

1. [Connecting to the 17th Workshop Database](#)
2. [Create a Typing report](#)
3. [Enter HLA typing](#)
4. [Transfer the fastq files and XML/HML reports that require conversion to IHIWS-compatible XML report to the IHIWS secure FTP site](#)
5. [Generating XML or HML IHIWS compatible output files from your analysis software](#)
6. [Convert your software analysis generated report to an IHIWS compatible XML report](#)
7. [Upload XML compatible files to the database](#)
8. [Enter HLA type by Batch upload GLstring](#)
9. [Category List](#)
10. [Typing Report Elements](#)
11. [Elements Attributes description](#)
12. [SampleID](#)

Instructions for entering & uploading Typing report

17th International HLA & Immunogenetics Workshop Database

1. *Connecting to the 17th Workshop Database*

To connect to the 17th International HLA & Immunogenetics (IHIW) Database, [visit http://ihiws.org](http://ihiws.org) and click the “Register to Participate” button.

Once connected to the database, users will be presented with the login page. To see more detail instructions on how to access the database, please see the “Registration Manual” in our website.

2. *Create a Typing report*

A Typing report can be created by any lab member associated with the Labcode assigned to the lab participating in the project

- a. Once logged in the IHIWS Database site, select the “Lab member” tab from the light blue navigation bar
- b. Select “Maintain Typing Data”
- c. Create a report ID: **Lab Defined Report ID** Is an internal identifier used by the reporting laboratory to identify a given report. Each such laboratory identifier should allow the reporting laboratory to unambiguously identify the report. To create a report ID:
 - a. Click “Create a typing Report”
 - b. Enter the Lab defined ID name
 - c. Click “Create”

3. Enter HLA Typing

The typing data may be entered one by one, or via an XML file (XML Schema maybe downloaded directly from the database) or via an HML file (Histoimmunogenetics Markup Language format).

Note: Current compatibility status of most common analysis platforms with IHIWS database allow data to be imported directly from output files after minor changes are applied using the IHIWS database tools also require the upload of fastq or htr files into the IHIWS secure FTP site. IHIWS compatible HML files generated by the software, must be parsed to IHIWS XML file before upload.

4. Transfer the fastq (raw data) files and XML/HML reports that require conversion to IHIWS-compatible XML report to the IHIWS secure FTP site:

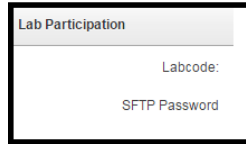
This is a required step to generate IHIWS-HML compatible reports. You may need to install software to allow secure file transfer between your local computer and the IHIWS secure FTP site, such as WinSCP for windows or FileZilla for Mac:

- File protocol = SFTP

Instructions for entering & uploading Typing report

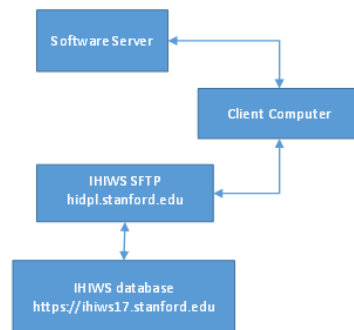
17th International HLA & Immunogenetics Workshop Database

- Host Name = hidpl.stanford.edu
- User Name = labcode (six digit ID provide to each lab when register to participate in the IHIWS)
- Password = *one per labcode. See Lab participation section of the Home Page for your Lab's SFTP password*



The screenshot shows a form titled "Lab Participation" with two input fields: "Labcode:" and "SFTP Password".

- Each lab has a folder named "upload" under which all the raw data files are to be stored. You may create subfolders under this folder to store your raw data based on how you want to organize the files, by project or by software used.
- End user will need to move files from the SFTP "upload" folder to the local computer to process files with the specific platform software. Vice-versa, if files need to be processed by the IHIWS database, they can only be accessed from the SFTP site.



5. Generating XML or HML output files that are IHIWS compatible from your analysis software if not using the SFTP process:

Depending on the software used to generate your typing data, HML or XML files may need to be processed prior to uploading onto the IHIWS database. The database allows users to upload IHIWS-compatible XML reports. To generate XML compatible reports, select the appropriate tool from the database Tools section. Select the "Lab member" tab from the light blue navigation bar to access the Tools section.

Instructions for entering & uploading Typing report

17th International HLA & Immunogenetics Workshop Database

Tools
Generate sample configuration files for Omixon
Convert GenDx xml file to IHIW xml file
Concat HML files
Parse Hml to IHIW xml
Generate GL strings from Sirona HML file
Generate GL String from Scisco csv
Job Log

Alternatively, if the report is generated by any of the software/version **as listed below**, the report can be uploaded directly onto the IHIWS SFTP and our system automatically converts the required data elements to IHIWS compatible XML report.

Note#1: *If you generate an IHIWS compatible HML file using the tools provided in the database for reports not compatible with step 6 below, the file must be parsed to IHIWS XML file before upload. Use the tool “Parse HML to IHIWS XML” to perform this process.*

Note#2: ***SampleID: When entering typing data, the Sample ID must be exactly as entered in the sample profile.** XML and HML generated reports will include the sample ID as it is entered in the run. This sample ID MUST be exactly as entered in the sample profile in the database, otherwise it will not match when uploading the HML or XML file outputs to the database.*

Note#3: *Make sure to name all files and folders without a space, use underscore if needed. System is case sensitive.*

Note #4 *Use the “Concatenate HML files” tool if you generated a report for each sample. This tool allows the user to concatenate all the reports into a single report.*

6. Convert your software analysis generated report to an IHIWS compatible XML report:

a. **SFTP generated IHIWS compatible XML reports:**

The following software have been updated to generate an HML or XML reports that can be uploaded directly onto the IHIWS SFTP and our system automatically converts the required data elements to IHIWS compatible XML report. This report can then be uploaded directly into the IHIWS database (<https://ihiws17.stanford.edu>)

- Omixon HML reports (HLA Twin™, version 2.1.0)
- GenDx XML (NGSEngine) reports version 2.4.0 (*Please contact <http://www.gendx.com/support>*)

Instructions for entering & uploading Typing report

17th International HLA & Immunogenetics Workshop Database

- ThermoFisher/One Lambda (TypeStream version 1.1) XML report. A standalone software (which can be downloaded from ThermoFisher website) allows the users to generate the report
- MIA FOR A version 3.1

To create IHIWS compatible XML reports from files generated by these software:

- To upload the XML or HML file reports from any of these software listed above to the SFTP see section in this document, or click on the link [Transfer the fastq files to the IHIWS secure FTP site](#)
 - Omixon and ThermoFisher and MIA FORA the file must be uploaded in the **upload/hml** folder in the SFTP
 - GenDx, the file must be uploaded in the **upload\gendx** folder in the SFTP
- After a short wait (~5 minutes), the IHIWS compatible XML report can be found in the **upload/ws_xml** folder in the SFTP.
- This file can then copied to your local computer so it can be uploaded directly into the IHIWS database.
- To Upload IHIWS XML compatible files to the database, see section in this document, or click on the link [Upload XML compatible files to the database](#)
- Contact your software vendor for assistance generating reports compatible with the 17th IHIWS.

b. **Illumina/Conexio report:**

Illumina/Connexio (TruSight HLA) – HML output

- The participants who use Illumina TruSight software will generate a “.cgp” file which is submitted to Illumina to convert to IHIWS.XML compatible file. Once the IHIWS XML compatible file is received by the end user, follow instructions to upload to the database. See section in this document [Upload XML compatible files to the database](#)
- To generate a “.cgp” file follow the directions below:
 - “.cgp” is generated after manually performing analysis in 4th – field and saving as “.cpg”. See section “Navigating the Interface” in the TruSight HLA Assign 2.1 RUO Software Guide (link below)
 - Intronic data should be included, to do so, make sure the “N-C” column is selected during the analysis. Click the Exons column header to expand or collapse the N-C column. For more complete instructions, See section “Navigating the Mismatch Columns” from the TruSight HLA Assign 2.1 RUO Software Guide (link below).

Instructions for entering & uploading Typing report

17th International HLA & Immunogenetics Workshop Database

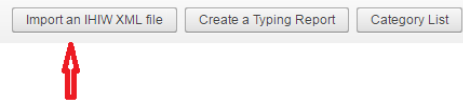
- c. It may be helpful to check final typing results using individual sample report (excel) see section “Generating Reports” in the TruSight HLA Assign 2.1 RUO Software Guide (link below)
 - d. For more specific instructions see TruSight HLA Assign 2.1 RUO Software Guide in the Illumina website
https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/assign/trusight-hla-assign-2-1-ruo-software-guide-1000000010450-01.pdf
 - iii. To submit the “.cgp” file to Illumina Contact Sandra Balladares at sballadares@illumina.com for information on your link to your specific Illumina box where the file can be deposited.
 - iv. The fastq files containing the raw data will need to be uploaded to the SFTP. See [Transfer the fastq files to the IHIWS secure FTP site](#)
 - v. The end user is responsible to upload the compatible IHIWS file to the IHIWS database.
 - c. **Immucor/Sirona (MIA FORA) HML reports**
Immucor/Sirona (MIA FORA) - HML output. Currently this HML is not compatible with the IHIWS database. GL String can be generated from the HML output using the IHIWS database Tool

Using the Tool in the IHIWS database, process the HML output file to generate GLString typing that can be uploaded by batch upload into the database (see [Enter HLA type by Batch upload GLstring](#))
 - d. **Scisco Genetics (ScisGo HLA) csv report:**
Scisco Genetics (ScisGo HLA) – GL String can be generated from the from csv output file using the IHIWS database Tool
 - i. Before using the tool, clean up the csv file:
 - Scisco uses comma (,) to represent ambiguities. User must replace (,) with (/) using notepad editor
 - Remove comments columns (Column E & F). Open the file in excel and remove any columns beyond GL string (Columns E and F)
 - ii. Use the Tool in the IHIWS database, process the csv output file to generate GLString typing. The typing can be uploaded by batch upload into the database (see [Enter HLA type by Batch upload GLstring](#))
 - e. **Histogenetics report:**
Most current HML format used to report to NMDP
7. Upload the typing using IHIWS compatible XML files in the database:

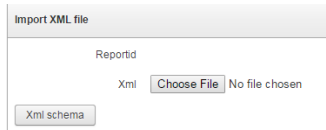
Instructions for entering & uploading Typing report

17th International HLA & Immunogenetics Workshop Database

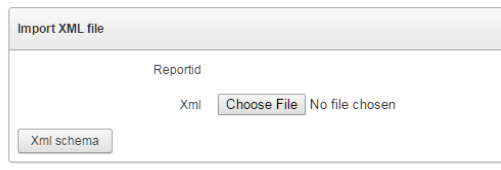
- a. Select the “Lab member” tab from the light blue navigation bar and click on Maintain Typing Data.
- b. HML reports need to be parsed to XML reports before upload (see tools section)
- c. Click on “Import an IHIWS XML file”



- d. Select the corresponding IHIWS compatible XML file



- e. Click Create
- f. Review for Invalid rows and correct the file as appropriate and try to upload again (note: Most common error is an invalid SampleID)
- g. “Parse and View” This will allow the user to make the necessary corrections to the file based on the listed errors or minor changes required to meet the IHIWS database schema. The following documents can be used as reference:
 - Click on XML schema to review IHIWS XML format
 - MIRING Validator for HML: <http://miring.b12x.org/>
 - HML schema: <https://github.com/nmdp-bioinformatics/hml/blob/master/spec/hml/1.0.1/hml-1.0.1.xsd>
 - IHIWS XML schema: Click on the Xml Schema button under Import an IHIWS XML file section to get the IHIWS XML report schema.



- h. XML Files: Once all information is entered, via XML or one sample at the time, you may select “Apply Changes and create XML”. This XML may be used for other analysis or as a template for other reports
- i. Once all information is entered, via XML or one sample at the time, you may select “Apply Changes and create XML”. This XML may be used for other analysis or as a template for other reports.
 - a. To upload the HLA types one by one or using BATCH upload
 - a. Click on the edit tool to continue

Instructions for entering & uploading Typing report

17th International HLA & Immunogenetics Workshop Database



- b. Complete the following typing report information for all the samples typed:
 - i. Software: Create Software Applied
 - ii. Hardware: Create Hardware Used
 - iii. Reagent: Create Reagent

Select or enter as appropriate each element, in some cases, multiple selections are possible

Please see sections [Typing Report Elements and Elements Attributes description](#) below for a detailed explanation.

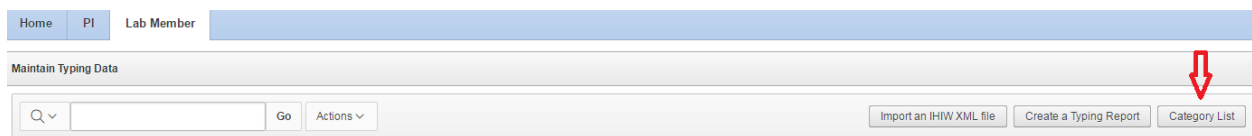
- c. Enter the HLA type for each sample, using GL string format under Create Sample section:
 - i. Click Create Sample
 - ii. Enter each sampleID, DB version, HLA type in Genotype_GL format
 - iii. Click Create
- d. To enter **GL string HLA types** using batch upload under Create Sample section:
 - i. Click batch import
 - ii. Upload file containing the columns: SampleID, DB_Version, Genotype_GL

Note: You may use the following tool to verify the GL string has the correct schema. <https://gl.nmdp.org/imgt-hla/3.23.0/explorer/>

- e. Enter the Locus Information. See Elements Attributes description for explanation of each attribute.
- f. To complete the report select Apply Changes and Create XML at the top right of the IHIW_Repot screen.

8. Category List

There are a number of attributes that require a selection from a drop down menu, use the following reference table for drop down selections, or click on the 'Category List' button under Lab Member tab to view the selection



9. Typing Report Elements and Attributes

The typing report contains the following elements specific to the report:

Instructions for entering & uploading Typing report

17th International HLA & Immunogenetics Workshop Database

- i. Software
- ii. Hardware
- iii. Reagent Protocol
- iv. Samples
- v. Locus

10. Attributes description

- i. **Software:**
 - a. Software Manufacturer: the manufacturer of the software
 - b. **Software Name** (Required Field): the name of the software applied
 - c. Software Parameters: any non-default parameter settings used in the application of the software
 - d. Software Version: the version, or other identifier, defining the software applied
 - e. **Function** (Required Field): Choose the function to which the software was applied, multiple selections are possible by holding the control key
Read processing functions may be applied to reads generated by the instrument prior to the generation of a consensus sequence.
Base Calling/Consensus generation is the creation of a consensus sequence through the analysis of a set of reads.
 - f. **Allele calling** is the identification of HLA genotypes. These functions may not be discretely distributed to individual pieces of software.
- ii. **Hardware:**
 - a. Typing method: Select from the drop down menu a Primary NGS Typing Method and a Secondary Typing Method
 - b. **Instrument name** (Required Field): the name of the instrument
 - c. Instrument Firmware: the version number, or other identifier, defining the software used on the instrument for data-analysis
 - d. Instrument Model Number: the model number, or other identifier, defining the type of instrument used for the typing
 - e. Manufacturer: the company or organization that built the instrument
- iii. **Reagent Protocol:**
 - a. **Protocol Name** (Required Field): a short description of the protocol and reagents used in performing the typing
 - b. Protocol Source: The company, laboratory or other entity that developed the protocol and reagents used to perform the typing
 - c. Protocol External Identifier: A reference to an external source that describes the protocol and reagents applied. This could be a reference to a Genetic Testing Registry (GTR) entry, a laboratory or manufacturer webpage describing the protocol and reagents, a DOI or citation of a published paper that includes the protocol.
 - d. Protocol Internal Identifier: If a protocol document is uploaded, include the file-name here.
 - e. Protocol Deviations: Any changes to the above referenced protocol should be noted here.
 - f. Specific Locus. Multiple selections are possible by holding the control key
- iv. **Samples:**

Instructions for entering & uploading Typing report

17th International HLA & Immunogenetics Workshop Database

- a. **SampleID** (required field): an alphanumeric string that allows the reporting laboratory to unambiguously identify the sample typed. This identifier should correspond to a sample already registered as part of a 17HIWS project.

Note: Once a sample profile is entered, the system generates an internal HIWS ID. To make it more convenient, the users can choose to upload the typing data using their own lab defined unique sample ID used when creating a Sample Profile in the system (as long as they comply with the requirements that require that PHI not be included as part of the ID). Users are still required to create a sample profile prior to entering the typing data.

- b. **DB version**: Select from the drop down the appropriate IMGT/HLA Database version. The column must match one of the supporting version of GL String validation web service: <https://gl.nmdp.org/> for strict mode validation; for example, 3.20.0. It could also be empty for non-strict mode validation
- c. **Genotype_GL (Required Field)**: a locus-level HLA genotype recorded using GL String format, as defined by Milius et al. 2013 (doi: 10.1111/tan.12150). Use of GL Strings allows all genotyping ambiguities to be accurately reported, along with phase information between loci. All ambiguity should be reported for a given sample. If a given locus was tested for, but found to be not present in the subject typed (e.g., a subject homozygous for DRB1*01 alleles will have no DRB3, DRB4, or DRB5 genes), report 'Absent' (e.g., DRB3*Absent) for that locus in the GL String. If polymorphisms have been identified that are not present in the reference database (novel polymorphisms), provide a GL String that describes the allele(s) that are closest in sequence to the novel allele sequence detected.

Validated using web service <https://gl.nmdp.org/>

See an example below for GL string format and the appropriate delimiters:

NOTE: database may allow user to analyze families to build haplotypes – currently in development

Order	Delimiter	Description	GLString Example
1	^	Unphased multilocus genotype	HLA-A*02:69+HLA-A*23:30 HLA-A*02:302+HLA-A*23:26/HLA-A*23:39^HLA-B*44:02:13+HLA-B*49:08
2		Genotype list	HLA-A*02:69+HLA-A*23:30 HLA-A*02:302+HLA-A*23:26/HLA-A*23:39
3	+	Genotype	HLA-A*02:302+HLA-A*23:26/HLA-A*23:39
4	~	Haplotype	HLA-A*23:26~HLA-B*44:02:13
5	/	Allele list	HLA-A*23:26/HLA-A*23:39

Note: Describe the novel sequence aspects in the NovelPolymorphism attribute of the Locus element below

- v. Locus:

Instructions for entering & uploading Typing report

17th International HLA & Immunogenetics Workshop Database

- a. **SampleID** (required field): Select from the drop down menu - should match one of the SampleID in "Create Sample" section.
- b. **HLA Typing** (required field): Select from the drop down menu - the HLA allele specified in the Genotype_GL attribute that this Locus element corresponds to. Should match one of the allele in the sample's GL String.
- c. Alignment Reference DB: Select from the drop down menu - the IMGT/HLA Database release version (e.g., IMGT/HLA Database 3.18.0) or Genome Reference Consortium release version (GRCh37) used for aligning reads for consensus generation.
- d. BaseCalling Reference DB: Select from the drop down menu - the IMGT/HLA Database release version (e.g., IMGT/HLA Database 3.18.0) used to identify the genotype from the consensus sequence.
- e. Consensus Sequence: A nucleotide sequence representing a contiguous phased region of DNA. This can correspond to a single feature, or to multiple contiguous features. If a locus is absent, this is not reported.
- f. **HLA Feature** (required field): Select from the drop down menu - (5'-UTR, 3'-UTR, Intron1, Intron2, Intron3, Intron4, Intron5, Intron6, Intron7, Exon1, Exon2, Exon3, Exon4, Exon5, Exon6, Exon7, Exon8), identify the feature, or specific set of contiguous features, to which this Locus element applies. Multiple sets of contiguous sequences may be defined per Locus element, If only individual gene sequences have been sequenced, identify each gene feature separately. It should match one of the feature names in Home->Category List.
- g. **Locus name** (required field): Select from the drop down menu - (HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5, HLA-DQA1, HLA-DQB1, HLA-DPA1, HLA-DPB1), the HLA locus for which the sequence data and metadata in a given Locus element are reported. It should match one of the HLA Locus names in Home->Category List.
- h. MeanReadDepth: the mean read depth across the length of the consensus sequence string. If a locus is Absent, this is not reported.
- i. DataFileLoc: (URL, Filepath or Filename), the URL, path or name of the file containing the primary read data generated by the instrument. Primary read data should be reported as unmapped reads with quality scores (e.g., FASTQ, htr, SFF, or unmapped BAM with quality scores).
- j. PhasingGroup: (Integer), Select from the drop down menu - an integer(indexed at 0 for each Locus_name) that identifies Locus elements that share phase. E.g., if phase is known between all Locus elements for a given Locus-name, PhasingGroup would be either 0 or 1 for a heterozygote at the locus, and 0 for a homozygote at the locus. Each element for which phase is unknown gets a new PhasingGroup number. It should be either 0 or 1 and the triplet (FEATURE, LOCUS_NAME, PHASINGGROUP) should be unique for each sample.
- k. NovelPolymorphism: describe any novel sequence polymorphisms resulting in a sequence that does not correspond to an allele in the reference database using a notation that identifies the reference database version, reference allele accession number, feature in which the novel polymorphism is found, and difference from the referenced feature. For example, IMGT/HLA|3.18.0|HLA00001|1.4|G56C identifies a G>C transversion at position 56 of exon 2 in a sequence that is otherwise identical to the exon 2 sequence of HLA-A*01:01:01:01.