TYPELOADER 2: Automated full-length submission of novel alleles



Bianca Schöne¹, Markus Fuhrmann¹, Vineeth Surendranath¹, Julia Phieler¹, Alexander H Schmidt^{1,2}, Vinzenz Lange¹, Gerhard Schöfl¹

¹**DKMS** Life Science Lab, Dresden, Germany; ²**DKMS** gGmbH, Tübingen, Germany

Motivation

In order to use full-length sequencing for genotyping of HLA and KIR, it is essential that novel alleles are characterised and submitted in full length and that known alleles are extended to cover the complete length of the gene. However, the manual annotation and submission of fulllength sequences to the European Nucleotide Archive (ENA) and the IPD-IMGT/HLA or IPD-KIR databases is a time-consuming and error-prone task. In 2016 we presented TypeLoader, a tool for automated full-length submission of novel HLA alleles. Here, we present an update, TypeLoader2, with new features and extended capabilities.

TypeLoader (2016): semi-automated submission to ENA and IPD

- Accepts full-length sequences of novel alleles \rightarrow Automatically creates all files for submission
- Web application that needs a Linux-server backend



Features: TypeLoader workflow. a. ENA submission. To annotate a novel allele sequence, the closest known allele is identified

- ✓ Support for all HLA genes
- Upload sequences as fasta or XML
- Generates all submission files for ENA and IPD-IMGT/HLA
- \checkmark 95% less hands-on time than with manual submission
- \Rightarrow Submitted >1000 novel HLA alleles at DKMS Life Science Lab

using BLAST against the reference HLA database

- (Module 1a) and the exon/intron information is transferred to the novel allele (Module 1b). A file ready for submission to ENA is created from the sequence data together with all required metadata (Module 2).
- **IPD-IMGT/HLA submission.** A file ready for submission to IPD-IMGT/HLA is created from the sequence data together with all required metadata (Module 3). The sections marked in blue are steps that have been automated, while the steps marked in orange are steps that require manual intervention.

TypeLoader2 (2018): convenient administration of all your novel allele data

- Standalone application: no server required
 - > Just download, install, and use
- Integrated SQLite database: maintain all metadata of your novel alleles
 - \geq Keep all files in a central, organised structure
 - \geq Add, access, edit, or submit data directly



| TypeLoader | | | | | | | | | | | _ | |
|---|-------------------------|-------------------------------|----------------|---------------|----------------|-------------------------|---------------------------|---------------|--------------|-------------------------|------------------------|--|
| <u>N</u> ew <u>O</u> verviews <u>S</u> ubmit alleles Options | | | | | | | | | | | | |
| New Sequence New Project Alleles overview | Project overview Submit | t to ENA Submit t | to IPD 🔅 | | | | | | | | | |
| Projects and Samples: | Projects overv | /iew: | | | | | | | | | | |
| Open 20180907_ADMIN_mixed_test | | | | | | | | | | | | |
| 20180906_ADMIN_HLA-B_NEB4 20180824_ADMIN_HLA-B_NEB1 | | | | | | | | | | | | |
| 20180719_ADMIN_KIR2DS1_NEB1 | Show closed projects | Show closed projects! Filter: | | | | | | | | | | |
| 20180716_ADMIN_KIR3DP1_PB4 20180713_ADMIN_mixed_PB4 | Project | Name 🔻 | Project Status | Creation Date | User Name | Gene | Pool | | Title | Description | Number of Allel | |
| ID28006748 ID34174974 | 1 20180907_ADMI | N_mixed_test | Open | 07.09.2018 | Admin Istrator | mixed | test | | | | 4 | |
| ✓ ID34174974 (2) | 2 20180906_ADMI | N_HLA-B_NEB4 | Open | 06.09.2018 | Admin Istrator | HLA-B | NEB4 | | | | 0 | |
| ↓ ID84010723 ↓ ID89557039 | 3 20180824_ADMJ | N_HLA-B_NEB1 | Open | 24.08.2018 | Admin Istrator | HLA-B | NEB1 | GUI-project1 | | A GUI-generated project | 5 | |
| 20180711_ADMIN_HLA-B_NEB1 | 4 20180719_ADM | IN_KIR2DS1_NEB1 | Open | 19.07.2018 | Admin Istrator | KIR2DS1 | NEB1 | | | | 2 | |
| 20180710_ADMIN_mixed_TEST 20180710_ADMIN_mixed_NEP1 | 20180716_ADM | IN_KIR3DP1_PB4 | Open | 16.07.2018 | Admin Istrator | Add r | new Targ | et Allele | | | 2 | |
| Navigation Area. | 20180713_ADMI | N_mixed_PB4 | Open | 13.07.2018 | Admin Istrator | | Project \ | | | A GUI-generated project | 5 | |
| Data is automatically | 20180711 ADM | N_HLA-B_NEB1 | Open | 11.07.2018 | Admin Istrator | | t as curre nit Project | nt Project | | A GUI-generated project | 7 | |
| sorted into projects | 20180710_ADMI | N mixed TEST | Open | 10.07.2018 | Admin Istrator | | nit Project | | | A GUI-generated project | 6 | |
| containing samples . | 0190710 ADM | IN_mixed_NEB1 | | 10.07.2018 | Admin Istrator | Delete project if empty | | | | A GUI-generated project | 2 | |
| All samples show status | S | | | | | f all | | locto | | A Gol-generated project | 2 | |
| icons, indicating wheth | | | | ig an ove | | | | | | u datailad | - | |
| they are done (<), | | | - | | | | rviev | v of all | samples, o | r detalled | 3 | |
| pending response from | | | | oroject or | | | | | | 1. 1.1 | | |
| ENA or IPD (©), or need | | | | | | - | | | ut are not e | | | |
| user interaction (). | | | | | ng and c | an be | e acc | cessed f | rom any ov | erview or from | 1 | |
| Projects and Samples: | Sample Vie | ne navig | ation are | a. | | | | | | | Download files | |
| Open 20180907_ADMIN_mixed_test | General Inform | | | | | | | | | | | |
| 20180906_ADMIN_HLA-B_NEB4 20180824_ADMIN_HLA-B_NEB1 | | | | | Filter: | | | | | Target Allele 💌 | Filter! Remove Filter | |
| 20180719_ADMIN_KIR2DS1_NEB1 20180716_ADMIN_KIR3DP1_PB4 | Internal Donor-ID | ID34174974 | | Target Allele | | Cell Line | | Allele Status | Lab Status | Project | | |
| 20180713_ADMIN_mixed_PB4 ID28006748 | External Donor-ID | DEDKM593276 | 56 | #1 (HLA-A) | | KMS-LSL- | | ENA su | ubmitted | completed | 20180713_ADMIN_mixed_P | |
| 102000748 1034174974 | | | | #2 (HLA-A) | [| KMS-LSL- | A-67 | IPD ac | cepted | completed | 20180713 ADMIN mixed P | |

Additional new features:

- Programmatic submission to ENA via a REST API
- Automatic detection of null alleles caused by frameshift mutations
- Experimental support for KIR alleles

| ID34174974 | | | | #2 (HLA-A) | DKMS-LSL-A-67 | IPD accepted | completed | 20180713_ADMIN_mixed_PB4 | | | |
|--|-------------------|--------------------------|----------------|----------------------|-------------------|---------------------|-----------|--------------------------|--|--|--|
| ID34174974 (2) ID84010723 | Customer | DKMS | | | | | | | | | |
| ID89557039 | Details about A | Details about Allele #2: | | | | | | | | | |
| | General Origi | nal Genotyping | Lab Processing | New Genotyping ENA S | | Confirm all changes | | | | | |
| | Internal Sample | Internal Sample ID | | | | | | | | | |
| | Allele Nr. in San | Allele Nr. in Sample | | | | | | | | | |
| | Project Name | Project Name | | MIN_mixed_PB4 | | | | | | | |
| | Nr. in Project | Nr. in Project | | | | | | | | | |
| | Cell Line | Cell Line | | 67 | | | | | | | |
| | Local Name | Local Name | | _2 | | | | | | | |
| | Gene | | HLA-A | | Reset all changes | | | | | | |
| | Goal | Goal | | | | | | | | | |
| | Allele Status | Allele Status | | | | | | | | | |
| | Lab Status | Lab Status | | | | | | | | | |
| | Internal Allele N | lame | | | | | | | | | |
| | Official Allele N | lame | | | | | | | | | |
| | | | | | | | | | | | |
| Two loo dowlo Cyou bicol Hoor | Instanfa an (CLU) | | | | | | | | | | |
| TypeLoader's Graphical User | Interface (GUI) |). | | | | | | | | | |

- **Complete TypeLoader-GUI displaying the Projects Overview.** In this overview, aggregated data about all projects is shown. Closed projects are hidden by default but can be shown optionally.
- Sample View. This detailed view shows all data about the novel alleles of one sample and allows viewing, updating and editing of all metadata as well as downloading, viewing and editing of all files for this sample.

Conclusion

TypeLoader2 can easily be used on standard PCs without dependencies on other software.



Download from GitHub: https://github.com/DKMS-LSL/typeloader

With the added internal database, TypeLoader2 can be used as a lab's central information platform on their novel full-length sequences.

We hope that the increased convenience and scope of TypeLoader2 will foster the submission of more full-length sequences to the IPD-IMGT/HLA

and IPD-KIR databases, ultimately promoting the widespread use of full-length sequencing for genotyping of both HLA and KIR.



DKMS Life Science Lab Blasewitzerstr. 46 01307 Dresden, Germany





