Complementing the IPD-IMGT/HLA database by Complete Genomic Characterization of Non-Rare HLA-DRB1 Alleles

Viviane Albrecht[1], Christin Paech[1], Kathrin Putke[1], Madlen Pahlke[1], Steffen Klasberg[1], Gerhard Schöfl[1], Bianca Schöne[1], Alexander H. Schmidt[1,2], Vinzenz Lange[1], Anja Klussmeier[1]

[1]DKMS Life Science Lab, St. Petersburger Str. 2, 01069 Dresden, Germany

[2]DKMS, Kressbach 1, 72072 Tübingen, Germany

Introduction

The IPD-IMGT/HLA database (release 3.41.0) currently comprises 2737 HLA-DRB1 alleles. However, for most alleles (96%) only exons 2 and 3 have been reported. Furthermore, the 118 full-gene characterized alleles are not the most common ones but often represent recently discovered rare variants. This lack of complete sequence information for many common or well defined HLA-DRB1 alleles disrupts the advantage of NGS for HLA genotyping. Therefore, we initiated a project aiming for the full-gene characterization of all HLA-DRB1 alleles down to a frequency of 0.0001 in European populations.

Sample Selection & PCR

Full-gene HLA-DRB1 characterization is challenging because of the varying length of the gene, 11 to 16 kb depending on the allele, as well as long homopolymer stretches and repetitive sequences. These obstacles were overcome by the development of an robust allele-length specific amplification strategy (EFI 2019: https://dkms-lab.de/dateien/EFI2019_DRB1FullLength.pdf).

To extend the sequences of the most frequent partially known alleles in the IPD/IMGT-HLA database, we selected 624 samples harboring 125 distinct HLA-DRB1 alleles with the following criteria:

- DNA concentration > 20 ng/µl
- Allele genotyped at least 50 times in the high throughput workflow between 2016 and 2019
- availability of at least 5 independent samples

Long-Range PCR Group	Alleles	Allele Length	PCR Failure Rates	
			Short read	Long read
S	DRB1*01/15/16	11 – 12 kb	20%	13%
M	DRB1*03/08/10/11/12/13/14	13 – 14 kb	7%	6%
L	DRB1*04/07/09	15 – 16 kb	44%	39%

Table 1: PCR failure rates of grouped known HLA-DRB1 alleles. Pretyped samples are assigned to their respective group(s) using first field information.

IPD-IMGT/HLA database

In total, 315 sequences were submitted to IPD-IMGT/HLA using TypeLoader2[2]. This includes the characterization of 112 unique alleles that had been selected for full gene characterization. Additionally, 51 novel alleles and 152 allele confirmations were submitted.

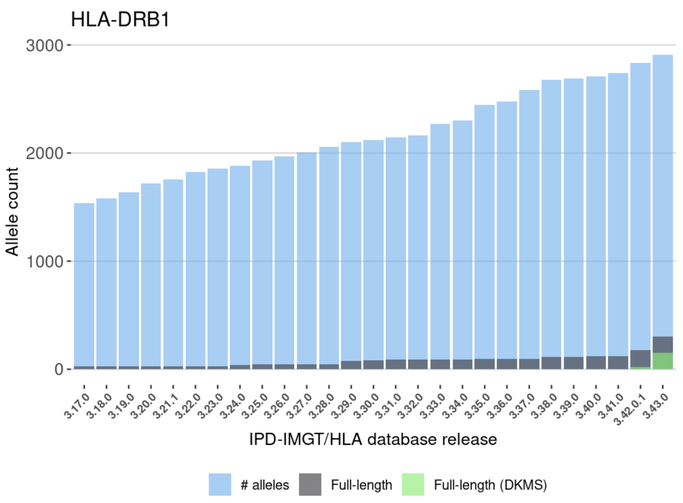

Figure 2: Growth of the IPD-IMGT/HLA database for HLA-DRB1 alleles from July 2014 to January 2021. The submissions described in this study are colored in green, submissions from other laboratories are colored in grey.

Dual Redundant Sequencing Strategy

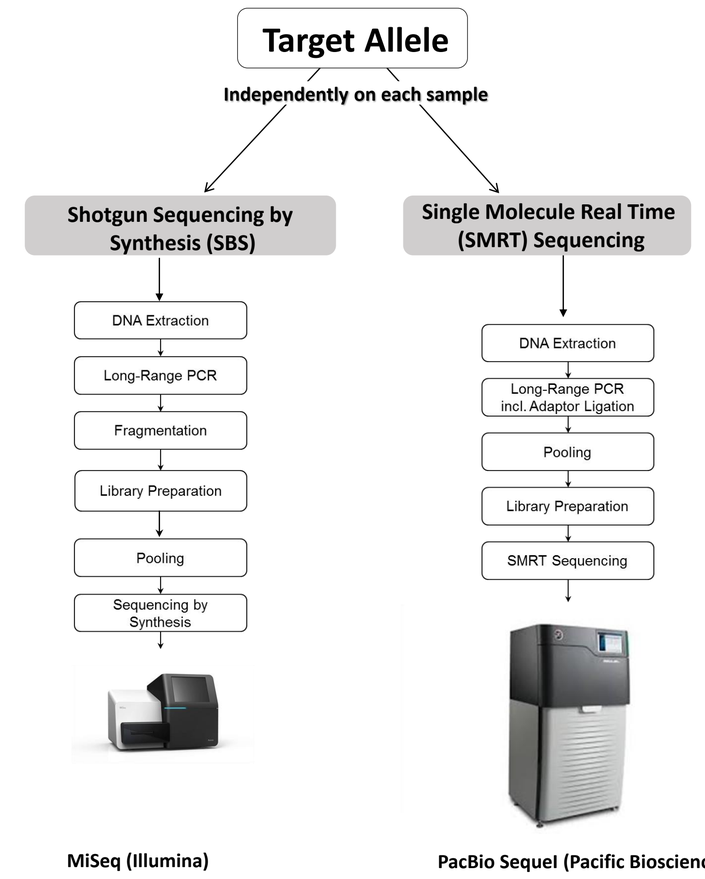

Figure 1: Workflow for the characterization of targeted HLA-DRB1 alleles

Using our dual redundant reference sequencing (DR2S)[1] pipeline, we combined short-read Illumina sequences with long-read PacBio Sequel sequences to generate highest quality error corrected consensus sequences.

Results

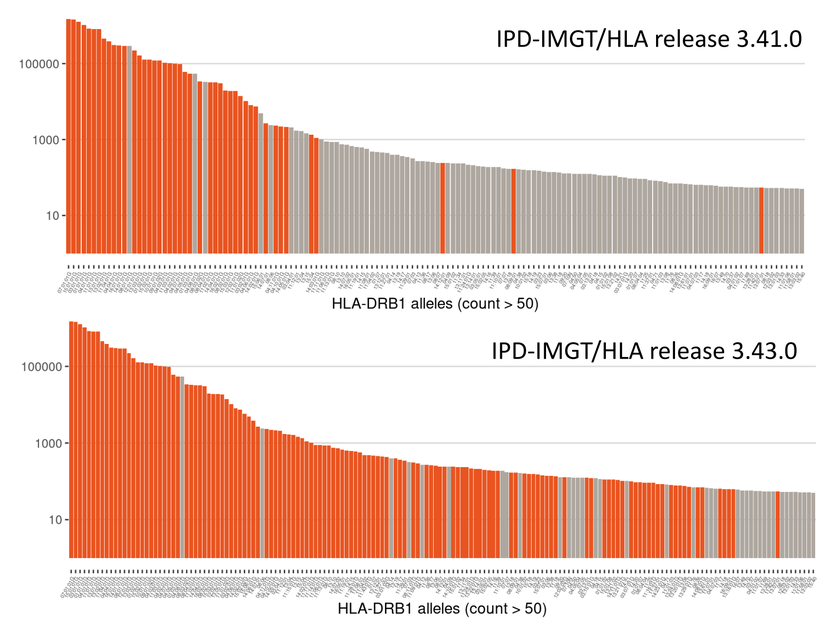

Figure 3: Full-gene characterized alleles (orange) and partially known alleles (grey) before (release 3.41) and after (release 3.43) this characterization effort sorted by number of observations in the DKMS Life Science Lab. The cumulative allele frequency of fully characterized HLA-DRB1 alleles has been increased from 93% to 99%, thereby strongly improving the basis for NGS based HLA-DRB1 genotyping.

Conclusion

We characterized and submitted 315 full-length HLA-DRB1 alleles to the IPD-IMGT/HLA database, thereby extending the sequence of 112 common or well defined alleles. But even though we raised the proportion of full-length known HLA-DRB1 alleles from 4% to 10%, about 90% of HLA-DRB1 alleles are still lacking a complete genomic sequence. Therefore, we encourage other HLA genotyping laboratories to use the described allele-length specific amplification strategy to submit HLA-DRB1 full-gene sequences in the future.

References

[1] Klasberg S, Schmidt AH, Lange V, Schöfl G. DR2S: An Integrated Algorithm Providing Reference-Grade Haplotype Sequences from Heterozygous Samples. *bioRxiv*. Published online November 10, 2020:2020.11.09.374140. doi:10.1101/2020.11.09.374140

[2] Schöne B, Fuhrmann M, Surendranath V, Schmidt AH, Lange V, Schöfl G. TypeLoader2: Automated submission of novel HLA and killer-cell immunoglobulin-like receptor alleles in full length. HLA. 2019;93(4):195-202. doi:10.1111/tan.13508

DKMS Life Science Lab

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

DKMS Life Science Lab

St. Petersburger Str. 2 • 01069 Dresden • www.dkms-lab.de