GENDX

personalizing diagnostics

Whole-Gene NGS-Based High-Resolution Genotyping Strategy for HLA-DPB1

HLA typing requires high-resolution typing of mainly HLA-A, -B, -C, -DRB1 and -DQB1. Application of HLA-DPB1 for matching donor and recipient is rising. In IMGT/HLA database release 3.34.0, 1097 HLA-DPB1 exon 2 sequences are registered, however only ~350 alleles have sequences registered outside of exon 2 and 3. This results in highly ambiguous typing results at the 2nd field level. DPB1 typing accuracy would clearly benefit from more full-length DPB1 sequences in the database. To supplement the database, a whole-gene amplification strategy for HLA-DPB1 from 5'UTR to 3'UTR is therefore a necessity.

The current NGSgo® amplification strategy includes all exons, however phasing issues arise due to coverage with two separate amplicons. Here, we present data on whole-gene HLA-DPB1 amplification as one amplicon (resulting in an amplicon of ~11 kb in size).

Materials & Methods

Amplification was performed on 33 IHWG and 36 UCLA gDNA samples (n=69) using the GenDx-LongMix enzyme mastermix (GenDx). The IHWG samples were selected from the DQ/DP reference panel. Successful amplification was confirmed by checking the amplicons by agarose gel electrophoresis. Representative amplicons for the whole-gene amplification are shown in Figure 1 for 7 IHWG samples (A) and for 7 UCLA samples (B). A strong band of the correct size can be observed for all samples as well as a blank negative control.

Library preparation was executed with 250 ng input per sample according to the NGSgo® workflow compatible with Illumina MiSeq (GenDx). After sequencing on an Illumina MiSeq platform applying the 300 cycle v2 kit, analysis was performed with NGSengine® software version 2.12 (GenDx).

Results

Analysis resulted in high-quality data with an average mappability of 87%, noise maximum of 1.9% and delta signal-to-noise of 41%. Figure 2 shows a base variation plot of a representative sample. In heterozygous samples the allele ratio was always within [30-70]%, with one exception where the ratio was [17-83]%. However, the distance between signal and noise was 13% and therefore the second allele could easily be typed by the software.

Concordance of the obtained typing results in comparison to the pre-typing information was found to be 84%. However, the pre-typing had been determined by exon 2 analysis only. It was observed that in all 8 samples with a discordant typing result, re-analysis with exon 2 only resulted in a concordant typing (Table 1). Therefore 100% of the typing results obtained by whole-gene amplification were concordant.

An unambiguous 4th field allelic typing result could be obtained for 60 out of 69 samples (87%). The ambiguities of the remaining 9 samples are summarized in Table 2. In 4 samples the 4th field could not be uniquely determined due to a SNP in a repetitive area that was excluded from analysis, or due to a SNP outside of the amplified region. For 5 samples only one genotype ambiguity remained. This was in all cases caused by a large distance between the discriminating SNPs, which could not be covered with the paired-end sequencing strategy of Illumina*.

Conclusion

This new strategy allows for a robust amplification from 5'UTR to 3'UTR of HLA-DPB1 as one amplicon, covering all exons, and will contribute significantly to the level of information available in the IMGT/HLA database. It will also enable more accurate genotyping of HLA-DPB1 in the future with a reduced number of ambiguities.

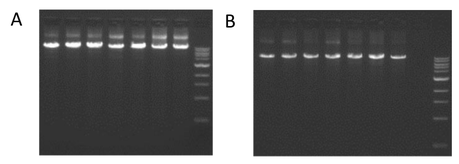

Figure 1. Agarose gel electrophoresis of whole-gene HLA-DPB1 amplicons, including a 1 Kb DNA ladder (NEB). Representative amplicons of 7 IHWG samples (IHWG9060, -9064, -9076, -9095, -9107, -9138 and -9151) (A) and 7 UCLA samples (C2-116 till C2-122) including a nega-tive no-template control on the right (B).

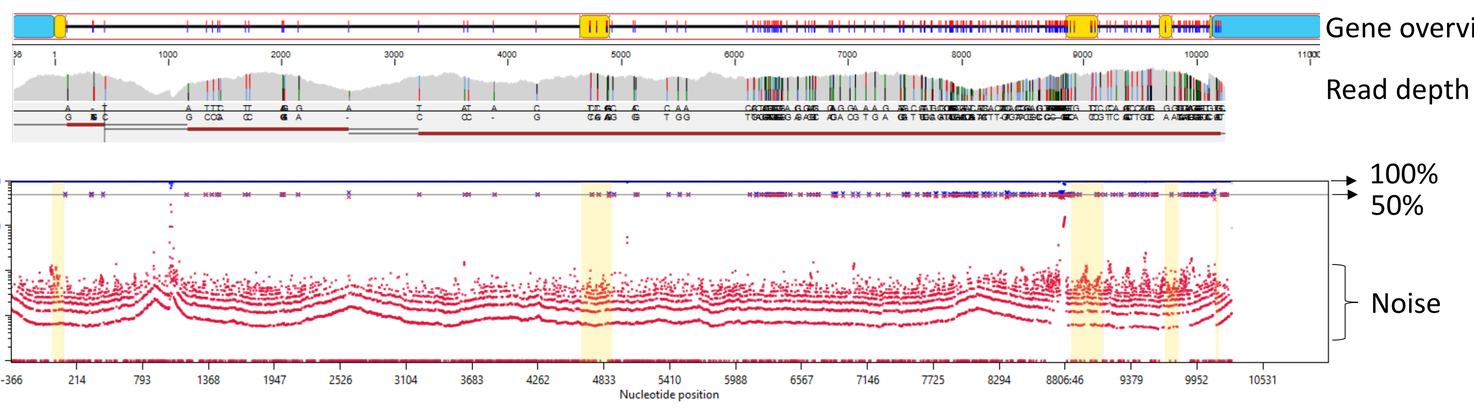

Figure 2. Overview, read depth, and base variation plot of representative sample IHWG9375 (HLA-DPB1*04:01:01:04/05:01:01:01).

Sample name	Pre-typing based on exon 2	Typing obtained by whole gene analysis
C2-104	03:01/04:01	04:01:01:14/124:01:02
C2-114	05:01/13:01	135:01/519:01
IHWG9076	13:01/13:01	107:01/107:01
IHWG9151	02:01/03:01	02:01:02:06/104:01:01:01
IHWG9189	13:01/28:01	13:01:01:04/296:01
IHWG9193	16:01/22:01	22:01:01:02/652:01
IHWG9366	02:01/19:01	02:01:02:18/106:01
IHWG9373	03:01/04:02	04:02:01:04/104:01:01:02

Table 1. Summary of the 8 samples for which a difference was found between the pre-typing based on exon 2 and the obtained typing result based on a whole-gene analysis.

Sample name	Ambiguity	Cause
C1-120	04:01:01:01/04:02:01:02 105:01:01:02/126:01:01:01	Distance between discriminating SNPs cannot be phased*
C1-212	02:01:02:01/04:02:01:02 105:01:01:08/416:01:01:02	
C1-220	04:01:01:01/04:02:01:02 105:01:01:02/126:01:01:01	
IHWG9314	04:01:01:02/46:01:01 04:01:01:05/46:01:01	
IHWG9381	02:02:01:01/04:02:01:02 105:01:01:05/721:01	
C1-219	01:01:01:01/04:02:01:02 01:01:01:03/04:02:01:02	Discriminating SNP in 3'UTR outside of amplified region
C2-108	01:01:01:01/04:02:01:02 01:01:01:03/04:02:01:02	
IHWG9267	04:01:01:01/30:01:01:01 04:01:01:06/30:01:01:01	Discriminating SNP in repetitive area of intron 2 that is excluded from analysis
IHWG9376	02:01:02:04/14:01:01:01 02:01:02:09/14:01:01:01	

Table 2. Summary of ambiguities found in 9 samples. In the remaining 60 samples, an unam-biguous 4th field allelic typing result was obtained.

Bouwmans EE, Welleweerd R, van Heck MR, Penning MT, Rozemuller EH.
GenDx, Utrecht, The Netherlands

* These ambiguities could be solved by sequencing the amplicons on a long-read sequencing platform. More information on this strategy can be found on Poster 182.

GenDx.com