Rode Kruis
Vlaanderen

Mapping the strengths and weaknesses of a HLA next generation sequencing workflow

V. Van Sandt , A. Senev , J. Kerkhofs , M. Li , T. Liu , C. Heylen , MP. Emonds

Next generation sequencing separates alleles, making the test more sensitive, but also more prone to cross contamination. Next to that NGS can increase HLA genotyping resolution. Allelic resolution is, however not clear cut, especially for alleles that differ in the number of short tandem repeats (STRs). To map the power versus pitfalls of our NGS workflow we evaluated the MIA FORA test sensitivity and the impact of cross contamination. We assessed the value of allelic resolution looking at the quality of STR calling and compared these results with third generation SMRT sequencing (TGS, PacBio) and Sanger data.

The MIA FORA NGS test sensitivity was very high. Results with a minimum coverage of 40 were concordant with the reference genotype (allelic res.) down to 2ng/µl for HLA-A, 0,1ng/µl for DRB1345 and down to 0,05ng/µl for B, C, DQ and DP. To evaluate the impact of cross contamination we spiked homozygous samples (33ng/µl) with a DNA reference sample (0,001 to 10ng/µl). Class I data showed no impact of the cross contamination. In Class II DBR1345, DPA1 and DPB1 data a second allele was added to the homozygous result starting from a spike concentration of 1ng/µl. For DQA1 and DQB1 the impact of the cross contamination was seen even down to 0,05ng/µl. These data reflect the power and limitations of the NGS interpretation algorithms, where allelic imbalance is more tolerated in Class II alleles and especially in DQ. MIA FORA allelic resolution was reproducible, with just 1 fourth field discordance out of 270 allele calls, demonstrating the robustness of the workflow and software. Analysis of homopolymer regions in MIA FORA NGS data showed confident calls, with only limited background of reads with INDELS(5%). PacBio CCS(circular consensus corrected reads) data of the same samples showed significantly higher background in the homopolymer regions (up to 20%). STRs of more than one nucleotide showed a more diverse distribution in the number of repeats in the NGS data, rendering allelic resolution in these cases questionable. The analysis of the TGS data of these samples is still on going.