A multiplex whole-gene HLA typing strategy evaluation on various NGS platforms: both long reads (PacBio, MinION), and short reads (Illumina, Ion Torrent, MGI).

Sake van Wageningen, Pascal van der Weele, Bart Valkenburg, Helma de Bruin, Erik Rozemuller, Maarten Penning

GenDx, Utrecht, The Netherlands

Aim

To investigate which NGS platforms available on the market are suitable for high resolution HLA typing, we performed a platform comparison study (Figure 1). Different NGS Libraries were generated from the same amplicons for the following platforms: Illumina Miseq, ThermoFisher Ion S5, PacBio Sequel IIe, MGI DNBseq-G50 and ONT MinION Mk1B. All studies were performed by GenDx, independently of the NGS platform manufacturers.

Methods

HLA amplicons were generated from a 58 sample HLA diversity panel (GeT-RM, HLA58) from the Coriell institute (Figure 2). Using an 11-gene multiplex amplification, whole gene amplicons were generated for HLA-A, -B, -C, -DRB1, -DQB1, -DPA1, -DPB1, - DQA, -DRB3, -DRB5. For HLA-DRB4 a single amplicon containing exons 2 and 3 was generated (NGSgo®-MX11-3; GenDx). Amplicon pools were used as input for library preparation for the different NGS sequencing platforms. All data analysis was performed using NGSengine software (GenDx).

Results and conclusions

The Illumina Miseq (Fig 3A) and PacBio Sequel IIe (Fig 3B) generated HLA typing results that were fully concordant with the pre-type information. Moreover, the Sequel IIe data analysis in all cases yielded fully phased amplicons. ThermoFisher Ion S5 (Fig 3C) and MGI DNBseq-G50 (Fig 3D) sequencers also generated highly concordant data (98% and 97%). Analysis of some loci in some samples required extra attention and re-analysis with adjusted analysis settings. Ion S5 data quality decreased in loci with homopolymers and in certain regions of DRB loci with high homology to other DRB genes. DNBseq-G50 data showed high sequencing noise at a few specific positions. This artefact will be further investigated in the future. ONT MinION Mk1B (Fig 3E) showed high concordance for class I genes, but data analysis of class II genes was hampered in several cases by high background noise in the data. Higher noise is innate to ONT sequencing, but this effect is exacerbated by homopolymers and STR regions in intronic HLA regions. Newer ONT sequencing chemistry compared to the one used here (R10.3) will be tested in the future for further optimization.

Figure 1: Platforms used in this study
Disclaimer; All studies were performed by GenDx, independently of the NGS platform manufacturers.

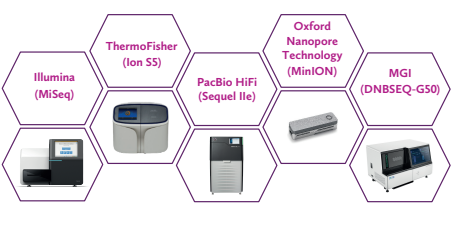
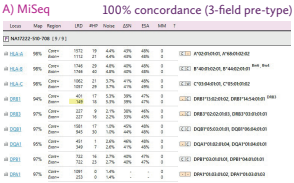


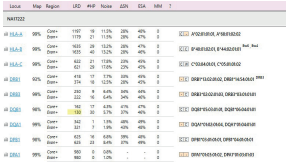
Figure 2: Setup of the Study
- HLA diverse 58 sample panel
- Multiplexed amplification; NGSgo-MX11-3

Figure 3: NGSengine data overview

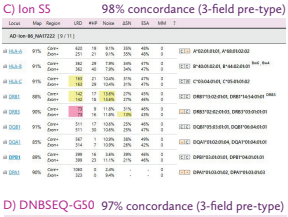
A) MiSeq 100% concordance (3-field pre-type)



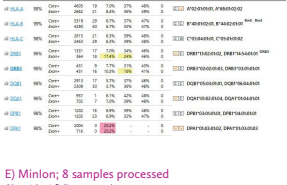
B) Sequel IIe 100% concordance (3-field pre-type)



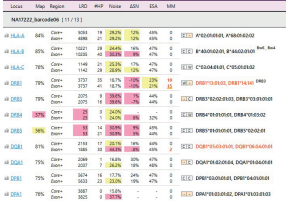
C) Ion S5 98% concordance (3-field pre-type)



D) DNBSEQ-G50 97% concordance (3-field pre-type)



E) Minlon; 8 samples processed
Class I loci fully concordant
Class II loci (DRB1, DQB1) challenging



Legend figure 3:

White (no marking): high quality
Yellow: intermediate quality
Pink: low quality

Lowest Read Depth (LRD)
Heterozygous positions (HP)
Delta Signal to Noise (ΔSN)
Estimated Second Allele% (ESA)
Mismatches (MM)

GENDX