COMPARING HLA TYPES OF TARGETED AMPLICON SAMPLES BY ANALYZING SANGER AND NEXT GENERATION SEQUENCING DATA

Attila Tordai[1], Dora Inotai[1], Endre Major[2], Attila Bérces[2], Szilveszter Juhos[2], Tim Hague[2]
Laboratory of Transplantation Immunogenetics – Omixon Biocomputing (www.omixon.com)

Introduction

Since their first commercialization next generation sequencing (NGS) technologies have spread rapidly to molecular genetic laboratories, in many cases overshadowing traditional Sanger sequencing due to their relative low cost and fast sample turnaround times. One of the driving forces of these large-scale sequencing studies has been hunting for mutations that can be associated to genetic diseases (Kilpinen, 2013). In particular among these mutations, different types of the human leukocyte antigen (HLA) gene are showing significant correlation to susceptibility for – or protection from – certain conditions. Still, using NGS for HLA typing is lagging behind although there are pioneering NGS based HLA typing studies using 454 and Illumina reads (Erlich et al. 2011, Wang et al.).

We present the results of eight different validation samples for parallel high resolution HLA-typing by sequencing (SBT) performed by both IonTorrent and Illumina MiSeq next generation sequencing (NGS) technologies analyzed by Omixon Target HLA and Sanger capillary sequencing analyzed by the SBT Engine software.
Results following the NGS workflow managed to have 100% percent concordance to HLA types determined by Sanger SBT, presenting a fast processing alternative to classical sequence based HLA typing.

Table 1: HLA types using Sanger SBT

Sample	HLA-A	HLA-A	HLA-B	HLA-B	HLA-C	HLA-C	HLA-DRB1	HLA-DRB1
B9143	A*02:01	A*26:01	B*38:01:01	B*41:01	C*12:03	C*17:01	DRB1*03:01	DRB1*14:54
B9170	A*03:01:01	A*26:01:01	B*07:02	B*38:01	C*07:02	C*12:03	DRB1*11:04:01	DRB1*15:01
B9063	A*01:01	A*68:01	B*08:01:01	B*44:03	C*04:01	C*07:01	DRB1*03:01	DRB1*07:01
B9243	A*02:01	A*32:01:01	B*40:01	B*51:01	C*03:04:01	C*14:02:01	DRB1*07:01:01	DRB1*11:01
B9226	A*01:01	A*02:01	B*38:01	B*40:01	C*12:03	C*17:01	DRB1*01:02:01	DRB1*13:02:01
B9189	A*02:05:01	A*26:01	B*35:01	B*50:01:01	C*04:01	C*06:02	DRB1*12:01	DRB1*13:01
B9153	A*29:02:01	A*68:01:01	B*40:01	B*44:02:01:01	C*03:04:01	C*16:01:01	DRB1*04:01:01	DRB1*13:02:01
B9146	A*01:01	A*02:01	A*08:01:01	A*51:01:01	C*07:01	C*15:02	DRB1*03:01	DRB1*11:01

Methods

The amplicons were designed to target the HLA-A,B,C genes, exons 1,2,3 and exons 4,5,6,7,8 separately, as well as the HLA-DRB1 gene exon 2. The samples were fragmented by sonication and 600 bp fragments were selected for NGS followed by barcoding each sample before mixing them and preparing two sequencer-specific libraries. Illumina sequencing followed the 150 bp paired-end protocol while Ion Torrent sequencing followed 250 bp single-end protocol.

For NGS data analysis, a purpose-built computational algorithm implemented in Omixon Target software was used to determine HLA types for all loci that can be found in the IMGT/HLA Database. The algorithm selects the HLA type candidates based on alignment errors and coverage statistics. We aligned the reads only to exons in the IMGT/HLA reference allele sequences that allowed us to achieve at least six digits resolution. After alignment, for each reference allele's coding sequences we are storing the coverage depth and coverage width values.
Coverage width is the extent of the exon covered by reads. In the next step we filtered using all this allele coverage data again, and left only those candidates that are worth processing further. This filter discards putative alleles having too few reads covering the reference and/or having long parts or whole exons not covered at all. Finally, having narrowed down to only few dozens of possible individual alleles, we are searching for allele pairs in a way that we are optimizing for both coverage depth and coverage width, reporting allele pairs that containing both a high number of mapped reads and are have adequate coverage of exons for both alleles at each locus.

Table 2: HLA types using NGS SBT (with reported ambiguities)

Sample	HLA-A	HLA-A	HLA-B	HLA-B	HLA-C	HLA-C	DRB1	DRB1
B9143	02:01:01	26:01:01	38:01:01	41:01	12:03:01	17:01:01	03:01:08,03:01:01	14:01:01,14:54:01
B9170	26:01:01	03:01:01	38:01:01	07:02:01	12:03:01	07:02:01	11:04:01	15:01:17,15:01:01
B9063	68:01:01	01:01:01	08:01:01	44:03:01	04:01:01	07:01:01	03:01:01,03:01:08	07:01:01
B9243	32:01:01	02:01:01	40:01:02	51:01:01	03:04:01	14:02:01	11:01:01,11:01:08,11:97	07:01:01
B9226	02:01:01	01:01:01	40:01:02	38:01:01	12:03:01	17:01:01	01:02:01	13:02:01
B9189	02:05:01	26:01:01	35:01:01	50:01:01	04:01:01	06:02:01	13:117,13:01:01	12:01:01,12:10,12:06,12:17
B9153	68:01:01	29:02:01	40:01:02	44:02:01	03:04:01	16:01:01	04:01:01	13:02:01
B9146	02:01:01	01:01:01	08:01:01	51:01:01	15:02:01	07:01:01	03:01:08,03:01:01	11:01:01,11:01:08,11:97

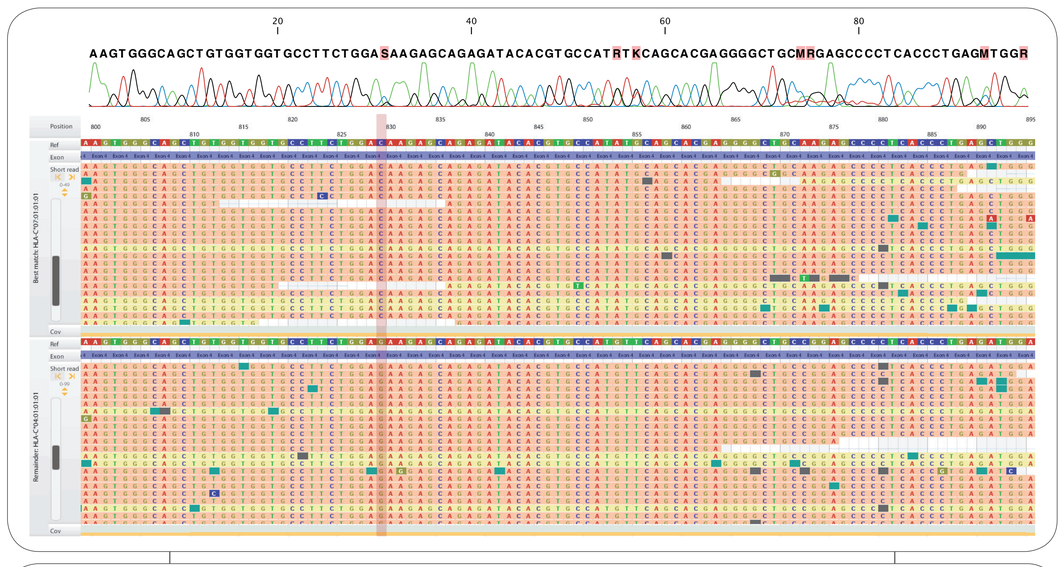

Fig. 3 – HLA-C:
Sanger trace and IonTorrent reads for sample B9063 HLA-C (types HLA-C*04:01 and HLA-C*07:01) on exon 4 showing 7 differences in the Sanger trace. Although HLA types are differing from each other most of the time on exon 2 and 3, using NGS typing it is possible to have much higher accuracy because it is easier to process other exons as well.

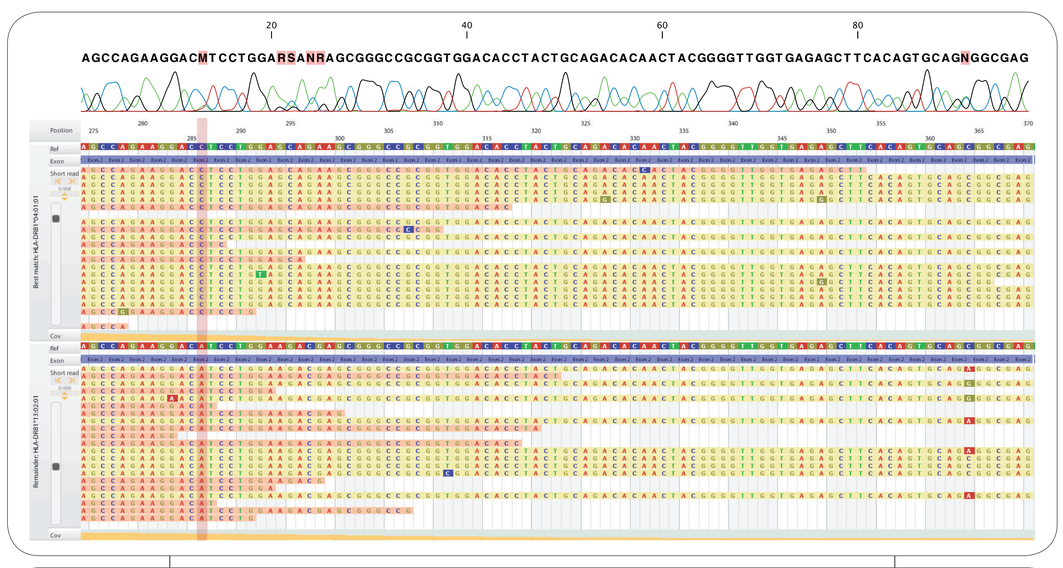

Fig. 4 – HLA-DRB1:
The HLA-DRB1 traces and Illumina reads for sample B9153 with types HLA-DRB1*04:01:01 and HLA-DRB1*13:02:01. The method is capable to distinguish DBRB1 alleles as well, but the resolution is lower since there was only one exon targetet for sequencing.

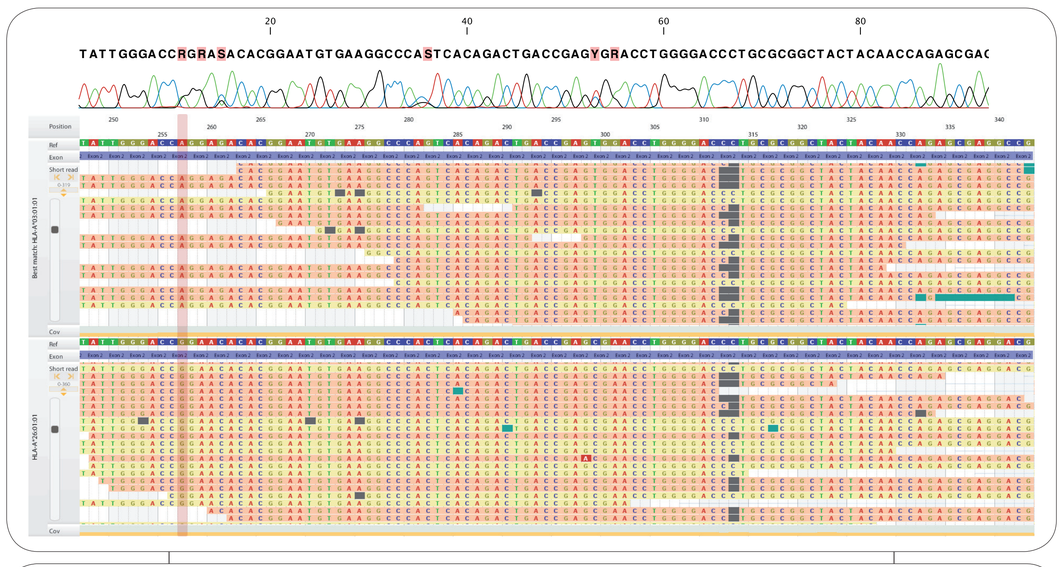

Fig. 1 – HLA-A:
Detailed view of a complex area from sample B9170 having types HLA-A*03:01:01and HLA-A*26:01:01. The aligned reference sequences are showing seven differences also present in the Sanger trace. Reads from NGS sequencers can be separated without explicit phasing. The software was able to find the correct allele in spite of the inherent homopolymer errors of IonTorrent technology.

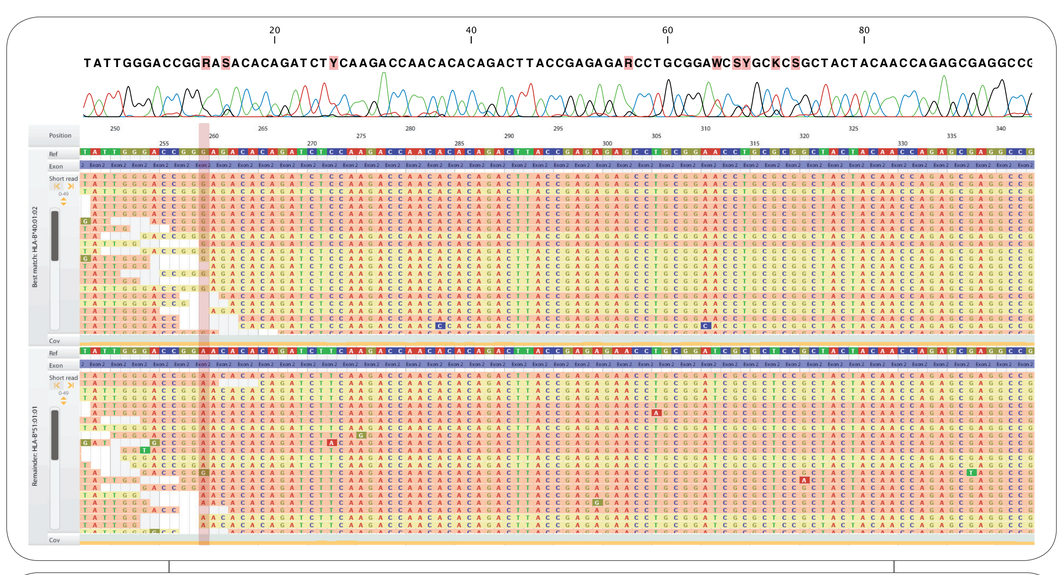

Fig. 2 – HLA-B:
For HLA-B the same precision can be achieved (showing a complex region on exon 2 Illumina reads): either the Illumina or the IonTorrent reads can be separated without phasing and the correct tpye can be assigned.

Conclusions

For both Illumina and Ion Torrent platforms, we reached 100% concordance with Sanger sequencing based typing for HLA-A, B, and C and DRB1. Processing NGS data by the computer algorithm is quick, all the samples were finished in ~90 minutes for IonTorrent and ~35 mins for Illumina on a commodity PC.

Ambiguities of the DRB1 loci were the result of limiting the amplification to exon 2. Results from IonTorrent technology tend to show ambiguities due to the homopolymer errors present in their reads. Our findings confirm that with correctly targeted genes it is possible to reach accurate high resolution HLA typing with either Illumina or IonTorrent reads and these results are comparable to those that are achievable by Sanger based SBT.

References:

1, Kilpinen H, Barrett JC.: How next-generation sequencing is transforming complex disease genetics. Trends Genet. 2013 Jan;29(1):23-30. doi: 10.1016/j.tig.2012.10.001. Epub 2012 Oct 25.

2, Erlich RL, Jia X, Anderson S, Banks E, Gao X, Carrington M, Gupta N, DePristo MA, Henn MR, Lennon NJ, de Bakker PI.: Next-generation sequencing for HLA typing of class I loci. BMC Genomics. 2011 Jan 18;12:42. doi: 10.1186/1471-2164-12-42.

3, Wang C, Krishnakumar S, Wilhelmy J, Babrzadeh F, Stepanyan L, Su LF, Levinson D, Fernandez-Viña MA, Davis RW, Davis MM, Mindrinos M.: High-throughput, high-fidelity HLA genotyping with deep sequencing. Proc Natl Acad Sci U S A. 2012 May 29;109(22):8676-81. doi: 10.1073/pnas.1206614109. Epub 2012 May 15.

Contact:

1 Laboratory of Transplantation Immunogenetics Hungarian National Blood Transfusion Service, Budapest, Hungary
2 Omixon Biocomputing Ltd, Budapest, Hungary
Corresponding author: szilveszter.juhos@omixon.com