illumina®

HLA Sequencing of IHWG Reference Samples with TruSight® HLA v2 Sequencing Panel

Results of repeated sequencing of 72 International Histocompatibility Working Group (IHWG) reference samples on the MiSeq® and MiniSeq™ Systems.

Introduction

The human leukocyte antigen (HLA) system is a gene region encoding the major histocompatibility complex (MHC) proteins and is responsible for alerting the immune system to infected, malfunctioning, and foreign cells. It is the most polymorphic region of the genome, meaning that there are more unique HLA sequences in the human population than in any other family of genes. This high sequence variability is important to fine-tune the adaptive immune system. The 2016 release of the IPD-IMGT/HLA Database, a specialist database for sequences of the human MHC complex, includes 14,473 unique HLA allele sequences.[1,2]

Sequencing of the HLA region is commonly used to understand transplant rejection in organ, bone marrow, and stem cell transplants. HLA mutations have also been associated with many autoimmune diseases, in which the immune system becomes self-reactive and mistakenly attacks healthy tissues. These diseases include multiple sclerosis, rheumatoid arthritis, type 1 diabetes, celiac disease, Graves' disease, lupus, Crohn's disease, psoriasis, and more.[3-6] The HLA system also has a significant role in HIV, as certain HLA alleles appear to be protective against HIV and other alleles appear to predispose to HIV.[7] Furthermore, people with certain HLA alleles have an adverse reaction to Abacavir,[8] an HIV antiretroviral drug.

Abacavir is not the only drug in which HLA type can result in adverse reactions. Carbamazepine[9] (used to treat seizures, nerve pain, and bipolar disorder), flucloxacillin[10] (antibiotic), ximelagatran[11] (anticoagulant), beta lactams[12] (antibiotics including penicillin and amoxicillin), oxicams[13] (used to treat pain), and even aspirin can cause adverse reactions in people with a particular HLA type.[14]

The TruSight HLA v2 Sequencing Panel uses Illumina next-generation sequencing (NGS) technology to generate phase-resolved sequencing results for 11 HLA loci, which cover all commonly typed HLA loci, plus those with emerging relevance. This white paper does the following:

- Presents HLA sequencing data on 72 IHWG reference samples using the TruSight HLA v2 Sequencing Panel
- Discusses in detail 5 alleles with discordance to reference typing due to no typing, requiring manual review
- Compares HLA typing results for reference samples sequenced on the MiSeq and MiniSeq Systems

Materials

The study was performed with purified DNA from 72 samples (Supplemental Table 1: IHWG Samples and References) from IHWG reference panels.[15] These samples were chosen for broad coverage of common, well-documented (CWD), rare, and null alleles, allele combinations, homozygotes, and heterozygotes. The 72 samples consisted of the following:

- 11 samples from the IHWG Consanguineous Reference Panel comprised of cell lines from 48 DNA samples from the 10th workshop indicated to be HLA homozygous by descent.
- 37 samples from the IHWG Sequence Polymorphism (SP) Reference Panel, a combination of 51 DNA samples (also includes 4 samples that overlap with the Consanguineous Reference Panel) typed using high-resolution methods available at the time of the 13th workshop.
- 13 samples from the IHWG Anthropology Reference Panel composed of 15 samples (also includes 10 samples that overlap with the SP Reference Panel) from different regions of the world.
- 1 additional sample from the 10th IHWG Workshop, 4 additional samples were selected from the 12th IHWG Workshop, 11 additional samples were selected from the 13th IHWG workshop, and 2 samples were selected from the IHWG null sample repository.

This panel of 72 samples includes alleles representing 87.5% (128 of 147) of the CWD 2.0.0[16] G groups (Table 1) and includes references for 287 unique alleles and 959 total referenced alleles (Table 2).

Table 1: Coverage of CWD 2.0.0 G Groups

Locus	Groups	Covered	Percent	Missing Groups
A	32	29	90.63%	02:22:01G, 24:05:01G, 68:01:01G
B	43	34	79.07%	13:01:01G, 15:07:01G, 15:11:01G, 15:18:01G, 15:25:01G, 27:04:01G, 27:07:01G, 35:43:01G
C	23	21	91.30%	01:03:01G, 18:01:01G
DRB1	14	11	78.57%	11:06:01G, 11:11:01G, 11:13:01G
DRB3	4	3	75.00%	02:01:01G
DRB4	1	1	100.00%	
DRB5	2	2	100.00%	
DQA1	7	7	100.00%	
DQB1	12	12	100.00%	
DPA1	1	1	100.00%	
DPB1	8	7	87.50%	23:01:01G
Total	147	128	87.50%	

The CWD alleles catalogue is periodically updated and published to identify the subset of HLA alleles with well-known frequency (common) or are well-documented. CWD 2.0.0 includes 1122 alleles that have been assigned to 147 G groups. The 72-sample IHWG panel used in this study had at least one allele from 128 of the 147 CWD G groups.

Table 2: Alleles with References by Locus

Locus	Unique Referenced Alleles	Total Referenced Alleles
A	51	121
B	70	122
C	34	119
DRB1	45	114
DRB3	7	43
DRB4	6	31
DRB5	3	10
DQA1	18	90
DQB1	19	114
DPA1	12	80
DPB1	22	115
Total	287	959

The 72-sample IHWG panel selected for this study had 287 unique HLA alleles referenced and a total of 959 alleles with references.

Methods

Library Preparation

DNA obtained from IHWG was prepared for sequencing using TruSight HLA v2 Sequencing Panel Library Preparation Kits (Figure 1). Long-range PCR was performed in 8 separate PCR reactions per sample (HLA-A, -B, -C, -DRB1/3/4/5, -DQA1, -DQB1, -DPA1, and -DPB1) for 30 cycles of 17.5 minutes per cycle for a total time of about 9 hours (Step 1). Upon completion of PCR and PCR cleanup (Step 2), samples underwent normalization (Step 3). The TruSight HLA v2 Kit uses a proprietary bead-based normalization developed by Illumina that requires no additional lab equipment.

This normalization step allows for each amplicon to be normalized en masse, rather than normalization based on averaging a few quantified amplicons. This is important because amplification will vary from sample to sample and from allele to allele, so normalizing each reaction provides consistency of DNA quantity input for library prep, which avoids dropping out of an allele or sample that is poorly represented.

The normalized amplicons were then subject to a tagmentation reaction using Illumina Nextera® technology, specifically modified for HLA sequencing, to simultaneously fragment and tag each amplicon with adapters (Step 4). The tagged DNA was then pooled into a single well per sample (Step 5). Limited-cycle PCR amplified the tagged DNA and added sequencing indexes (Step 6). The TruSight HLA v2 libraries use a dual-indexing approaching with 8-base index barcodes flanking the template DNA. The PCR products were purified (Step 7) and all individual sample libraries were pooled and loaded onto the sequencer (Step 8). Find a more detailed description of the method and protocols in the TruSight HLA v2 Sequencing Panel Reference Guide.

Samples were prepared 12 samples at a time, stored, and pooled before sequencing.

Step		Total Time	Hands-On	Schedule
1	Generate HLA Amplicons	11 hours	30 minutes	Day 1 @ 4:00pm
2	PCR Cleanup	30 minutes	30 minutes	Day 2 @ 9:00am
3	Bead-Based Normalization	1 hour	30 minutes	Day 2 @ 9:30am
4	Tagmentation	30 minutes	15 minutes	Day 2 @ 10:30am
5	Pooling and Cleanup	30 minutes	30 minutes	Day 2 @ 11:00am
6	Amplify PCR	1.5 hours	30 minutes	Day 2 @ 11:30am
7	PCR Cleanup	15 minutes	15 minutes	Day 2 @ 1:00pm
8	Library Pooling and Loading	45 minutes	45 minutes	Day 2 @ 1:15pm
	Total	16 hours	< 4 hours	Day 2 @ 2:00pm

Figure 1: TruSight HLA v2 Sequencing Panel Library Preparation Protocol—TruSight HLA v2 library preparation usually begins late in the day, providing ample time to receive samples, accession samples, and isolate DNA on the same day. The DNA is loaded onto thermal cyclers to amplify overnight. With a 12-sample run, the Day 2 library preparation will take about 5 hours so that the sequencer is loaded in the early afternoon. In total, library preparation from DNA to sequencer is ~ 16 hours with ~ 4 hours of hands-on time.

Sequencing

Libraries were run in parallel on the MiSeq and MiniSeq Systems. Paired-end 150-base pair sequencing (2 × 150 bp) was used on both systems. On the MiSeq System, 24 samples were pooled into each run using the MiSeq Reagent Kit v2 Micro (300 cycles). On the MiniSeq System, all 72 samples were pooled onto a single MiniSeq High Output Reagent Kit (300 cycles). The MiSeq System runs averaged 19 hours per run and the MiniSeq System runs averaged 24 hours per run (Table 3).

Table 3: MiSeq and MiniSeq Systems Specifications for the TruSight HLA v2 Sequencing Panel

MiSeq System (4-channel chemistry)				
Reagent Kit	**Max. Reads**	**2 × 150 Output**	**2 × 150 Run Time**	**Max. Samples**
v2 Nano	1M	300 Mb	17 hours	6
v2 Micro	4M	1.2 Gb	19 hours	24
V2 Standard	15M	4.5 Gb	24 hours	96
V3 Standard	25M	7.5 Gb	39 hours	144
MiniSeq System (2-channel chemistry)				
Reagent Kit	**Max. Reads**	**2 × 150 Output**	**2 × 150 Run Time**	**Max. Samples**
Mid Output	8M	2.4 Gb	17 hours	48
High Output	25M	7.5 Gb	24 hours	144

Data Analysis

The compressed fastq.gz files (2 per sample; Read 1 and Read 2) output from the MiSeq and MiniSeq Systems were imported directly into TruSight HLA Assign 2.0 (v2.0.0.920) Software. IMGT/HLA version 3.23 and CWD version 2.0.0 software were used for analysis and typing of the results. Assign 2.0 Software performed the alignment of individual reads to a consensus reference sequence (2 consensus reference sequences for HLA-DQB1 and 4 consensus reference sequences for HLA-DRB1).

The software then phased all heterozygous positions using base positions under individual reads and in paired reads. Often, multiple pairs of reads need to be layered to align phase directly. This method allows for phasing of heterozygous positions as far apart as the longest fragment sequenced (usually 1000–1300 bases). These phased alignments generate a consensus sequence for each locus comprising a mean depth of coverage per locus of more than 100×. Following sample multiplexing recommendations, mean depth of coverage is usually over 200× per locus. These consensus sequences were compared to the IMGT/HLA database housed within the Assign 2.0 Software to generate an HLA type for each locus.

Results: MiSeq System

Evaluation of Concordance with Reference Typing

Sequencing of the 72 IHWG samples on the MiSeq System resulted in consensus sequences for 1294 alleles, 959 of which had references for comparison. 907 (94.58%) alleles were concordant with the reference typing. The 52 typing results that did not match the reference sequence fell into 3 categories: 42 (4.38% of total referenced alleles) had incorrect or incomplete reference typings that prompted updates to the reference typings (Table 4), 5 (0.52% of total referenced alleles) had novel exonic variants, and 5 (0.52% of total referenced alleles) were not assigned a type and required manual editing.

All alleles with reference typings for which the results were concordant, alleles with reference typings that appeared to be incorrect, and novel alleles were used to calculate accuracy for the TruSight HLA v2 Sequencing Panel (Figure 2). 5 alleles were considered inaccurate and required a total of 9 base edits for concordance. These 5 alleles were not typed rather than being assigned an incorrect typing.

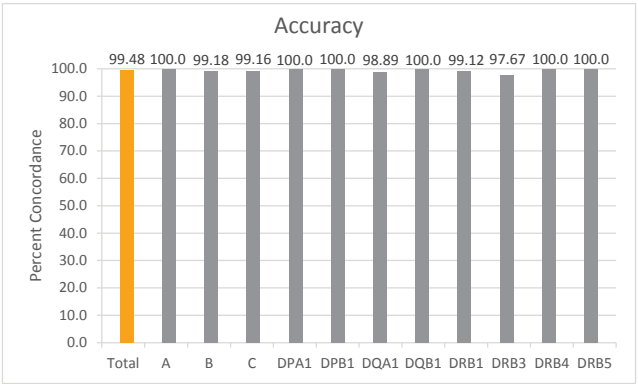

Figure 2: Accuracy of the TruSight HLA v2 Sequencing Panel on the MiSeq System—The accuracy is calculated using 959 alleles with references. A TruSight HLA v2 result was considered accurate when a result was fully concordant (907 alleles, 94.58%), when the result matched an update to the reference (42 alleles, 4.38%), and when the allele contained a novel variant (5 alleles, 0.52%). This resulted in a total of 954 alleles called accurately (99.48%).

Table 4: IWHG Reference Updates

Sample ID	Locus	Reference Typing	TruSight HLA v2 Typing	Notes
IHW01136	A	11:01	01:01:01	No evidence of the 11:01. This is consistent with findings from Wittig et al.[17] in which they also saw no evidence of the 11:01.
IHW09099	A	02:17:01	02:17:02	This is an exon 1 mismatch of a single base position (315) in which a G was called, consistent with 02:17:02. A T at this position is consistent with the 02:17:01. With over 200 reads in the analysis, there is no evidence of the T call at this position. This is consistent with Wittig et al.[17]
IHW01040	B	39:06:01	39:06:02	This is an exon 3 mismatch of a single base position (1028) in which a T was called, consistent with 39:06:02. A C at this position is consistent with the 39:06:01. With over 200 reads in the analysis, there is no evidence of the C call at this position. This is consistent with Wittig et al.[17]
IHW09388	B	40:01:01	40:01:02	Single base mismatch to 40:01:01 in exon 1 at position 356 (G>C). The C is consistent with CWD allele 40:01:02.
IHW01040	C	07:01	07:18	This is mismatched to the 07:01:01 at a single base position (2857 C>T) in exon 6. The T call at this position was made with over 100 reads of the more than 200 reads covering this position. The T at this position is consistent with the 07:18. Wittig et al. called this locus ambiguous with both 07:01/07:02 and 07:02/07:18 as possible results.
IHW01136	C	15:02	15:13	There is a single exon 4 mismatch to the 15:02:01 at position 2008 (G>C). The C is consistent with the 15:13 and is covered by more than 80 reads of the more than 200 reads covering the position. This is consistent with Wittig et al.[17]
IHW09114	C	02:02	02:10:01	There are 2 mismatches to the 02:02, one in exon 3 (1150 A>G) and one in exon 4 (1954 A>G). Both positions are covered by more than 200 reads all going to the G's. The G call at these positions is consistent with the 02:10. Furthermore, the 17:01 was not in the reference yet is clearly present.
IHW09373	C	07:01	07:18	There is a single base exon 6 mismatch to the 07:01 at position 2857 (CC>CT) with more than 80 reads covering the T and more than 200 total reads covering the position. The T at this position is consistent with 07:18. Wittig et al.[17] called this ambiguous with both the 07:01 and the 07:18 as potential results.
IHW01059	DPA1	01:04	01:03:01	There is a single exon 2 heterozygous position (4284 AA>AC) with over 200 reads and 50/50 allele balance. There are 6 other het positions called in the amplicon (4 in the 3' UTR and 2 in intron 1). These positions are consistent with coverage and allele balance to the exon 2 het position.
IHW09375	DPA1	02:02:03	02:02:02	There is a single base exon 2 mismatch to the 02:02:03 at position 4260 (G>A). The A is consistent with 02:02:02 and is covered by more than 100 reads. The 02:02:02 is consistent with Wittig et al.[17]
IHW09381	DPA1	01:01	01:03:01	There is no DPA1*01:01 currently characterized in the IMGT/HLA database.
IHW09040	DPB1	03:01	104:01	There is a single base exon 4 mismatch to 03:01 at position 10163 (G>A). The A is consistent with 104:01 and covered by more than 200 reads. The 104:01 homozygote is consistent with Wittig et al.[17]
IHW09373	DPB1	03:01	104:01	There is a single base exon 4 mismatch to 03:01 at position 10163 (G>A). The A is consistent with 104:01 and covered by more than 80 reads. Wittig et al. called this with an ambiguity string containing both the 03:01 and 104:01.
IHW09387	DPB1	13:01	133:01	There is a single base exon 2 mismatch to 13:01:01 at position 5133 (A>G). The G is consistent with 133:01 and is covered by 200 reads. The 133:01 is consistent with Wittig et al.[17]
IHW09398	DPB1	03:01	104:01	There is a single base exon 4 mismatch to 03:01 at position 10163 (G>A). The A is consistent with 104:01 and covered by more than 80 reads. Wittig et al. called this with an ambiguity string containing both the 03:01 and 104:01.
IHW09431	DPB1	03:01	104:01	There is a single base exon 4 mismatch to 03:01 at position 10163 (G>A). The A is consistent with 104:01 and covered by more than 80 reads. Wittig et al. called this with an ambiguity string containing both the 03:01 and 104:01.
IHW01040	DQA1	05:01	05:05:01	This is a frequent reference issue in which the reference calls 05:01 and the NGS result calls the more common 05:05. These differ from one another in 3 exon positions (777 exon 1, 5580 exon 3, and 6030 exon 4). Exon 2 sequence is identical between the 2. This is consistent with Wittig et al.[17]
IHW01093	DQA1	05:01	05:05:01	This is a frequent reference issue in which the reference calls 05:01 and the NGS result calls the more common 05:05. These differ from one another in 3 exon positions (777 exon 1, 5580 exon 3, and 6030 exon 4). Exon 2 sequence is identical between the 2.
IHW01141	DQA1	05:01	05:05:01	This is a frequent reference issue in which the reference calls 05:01 and the NGS result calls the more common 05:05. These differ from one another in 3 exon positions (777 exon 1, 5580 exon 3, and 6030 exon 4). Exon 2 sequence is identical between the 2.
IHW09009	DQA1	01:02:01	01:02:02	There is a single base mismatch to the 01:02:01 in exon 3 (5391 T>C). We also see no evidence of heterozygosity. This is consistent with Wittig et al.[17]
IHW09016	DQA1	05:01	05:05:01	This is a frequent reference issue in which the reference calls 05:01 and the NGS result calls the more common 05:05. These differ from one another in 3 exon positions (777 exon 1, 5580 exon 3, and 6030 exon 4). Exon 2 sequence is identical between the 2. This is consistent with Wittig et al.[17]
IHW09035	DQA1	05:01	05:05:01	This is a frequent reference issue in which the reference calls 05:01 and the NGS result calls the more common 05:05. These differ from one another in 3 exon positions (777 exon 1, 5580 exon 3, and 6030 exon 4). Exon 2 sequence is identical between the 2. This is consistent with Wittig et al.[17]

Table 4: IWHG Reference Updates, continued

Sample ID	Locus	Reference Typing	TruSight HLA v2 Typing	Notes
IHW09040	DQA1	05:01	05:05:01	This is a frequent reference issue in which the reference calls 05:01 and the NGS result calls the more common 05:05. These differ from one another in 3 exon positions (777 exon 1, 5580 exon 3, and 6030 exon 4). Exon 2 sequence is identical between the 2. This is consistent with Wittig et al.[17]
IHW09077	DQA1	03:01	03:02	There are 2 mismatches in exon 1 and 3 to the 03:01. These are perfect matches to the 03:02. Wittig et al.[17] made this same call.
IHW09114	DQA1	05:01	05:05:01	This is a frequent reference issue in which the reference calls 05:01 and the NGS result calls the more common 05:05. These differ from one another in 3 exon positions (777 exon 1, 5580 exon 3, and 6030 exon 4). Exon 2 sequence is identical between the 2.
IHW09388	DQA1	03:02	03:03:01	There is a single base mismatch to 03:02 in exon 1 at position 799 (C>T). The T is a perfect match to 03:03:01.
IHW01059	DQB1	02:01	02:02:01	There is a single base exon 3 mismatch to 02:01 at position 5548 (A>G). The G is a perfect match to 02:02:01.
IHW09040	DQB1	03:01	03:19	There is a single exon 3 mismatch to 03:01 at position 5698 (C>T). The T is a perfect match to 03:19. The 03:19 homozygous call is consistent with Wittig et al.[17] Also, the 03:19 result was independently verified by Sanger sequencing of exon 2 at Eton laboratories.
IHW09047	DQB1	02:01	02:02:01	There is a single base exon 3 mismatch to 02:01 at position 5548 (A>G). The G is a perfect match to 02:02:01. The 02:02:01 is the same call made by Wittig et al.[17] The 02:02 typing was confirmed by Sanger sequencing of exon 3 performed by the BloodCenter of Wisconsin.
IHW09114	DQB1	03:01	03:19	There is a single exon 3 mismatch to 03:01 at position 5698 (C>T). The T is a perfect match to 03:19.
IHW09273	DQB1	02:01	02:02:01	There is a single base exon 3 mismatch to 02:01 at position 5548 (A>G). The G is a perfect match to 02:02:01.
IHW09367	DQB1	05:02:01	03:03:02	There is no evidence of the 05:02:01 in our sequence data. Wittig et al.[17] also called the 03:03:02 homozygote. The 03:03 typing was confirmed by Sanger sequencing of exon 2 performed by the BloodCenter of Wisconsin.
IHW09398	DQB1	02:01	02:02:01	There is a single base exon 3 mismatch to 02:01 at position 5548 (A>G). The G is a perfect match to 02:02:01. The 02:02:01 is the same call made by Wittig et al.[17]
IHW09398	DQB1	03:01	03:19	There is a single exon 3 mismatch to 03:01 at position 5698 (C>T). The T is a perfect match to 03:19. Wittig et al. also called an 03:19. The 03:19 typing was confirmed by Sanger sequencing of exon 2 performed by the BloodCenter of Wisconsin.
IHW09056	DRB1	14:01	14:54:01	This is a frequent reference error in which the 14:01 is called in favor of the 14:54. These alleles differ at one exon 3 base position 11292 (T>C). This is a perfect match to 14:54. This is the same call made by Wittig et al.[17]
IHW09061	DRB1	14:01	14:54:01	This is a frequent reference error in which the 14:01 is called in favor of the 14:54. These alleles differ at one exon 3 base position 11292 (T>C). This is a perfect match to 14:54.
IHW09077	DRB3	03:01:01	03:01:03	There are 2 mismatches to the 03:01:01 in exons 3 (10946 A>G) and 4 (11705 T>C). Both of these changes are consistent with 03:01:03.
IHW09253	DRB3	03:01:01	03:01:03	There are 2 mismatches to the 03:01:01 in exons 3 (10946 A>G) and 4 (11705 T>C). Both of these changes are consistent with 03:01:03.
IHW09029	DRB4	01:01	01:03:01	The 01:01 and 01:03 alleles differ from one another in exon 3 at position 13110. The 01:01 calls for an A and the 01:03 calls for a G. The G is called here with more than 200 reads attributed to the G call.
IHW09114	DRB4	01:01	01:03:01	The 01:01 and 01:03 alleles differ from one another in exon 3 at position 13110. The 01:01 calls for an A and the 01:03 calls for a G. The G is called here with more than 200 reads attributed to the G call.
IHW09377	DRB4	01:01:02	01:01:01	There is no DRB4*01:01:02 allele characterized in the current IMGT/HLA database. This typing is a perfect match to 01:01:01.
IHW09388	DRB4	01:01	01:03:01	The 01:01 and 01:03 alleles differ from one another in exon 3 at position 13110. The 01:01 calls for an A and the 01:03 calls for a G. The G is called here with more than 200 reads attributed to the G call.

The majority (33 of 42) of this reference discordance is likely because outlying exons were not sequenced or investigated with conventional methods available at the time this typing was performed. These are highlighted in white.

Discovery of Novel Exonic Variants

5 alleles (0.52% of total referenced alleles) had novel exonic variants (Figure 3).

IHW09220 and IHW09378: Reference alignments for DPA1*02:01:01 and DPA1*02:02 both expect an A at position 4314 (Exon 2 position 107). The IHW09220 sample consensus sequence made up of more than 200 reads shows a G in ~ 55% of these reads and an A in ~ 45% of the reads. The A was phased and aligned to the 02:01:01 and the G was phased and aligned to the second allele. There is no match in IMGT/HLA to this sequence, including a G at this position, indicating this to be a novel variant. IHW09378 has the same novel sequence of the 02 allele (Figure 3A).

IHW09266: Reference alignments for DQB1*03:01:01 and DQB1*05:01:01 both expect a C at position 612 (Exon 1 position 81). The IHW09266 sample consensus sequence made up of more than 250 reads shows a C in ~ 50% of the reads and a T in ~ 50% of the reads. The C was phased and aligned to the 03:01:01 and the T was phased and aligned to the second allele. There is no match in IMGT/HLA to this sequence, including a T at this position, indicating this to be a novel variant (Figure 3B).

IHW09369: The reference alignment for B*15:26N expects a T at position 295 (Exon 1 position 11). The reference typing for this locus is 15:26N homozygote. The IHW09369 sample consensus sequence made up of more than 300 reads shows a C at this position in all reads. There is no match in IMGT/HLA to this sequence, including a C at this position, indicating this to be a novel variant (Figure 3C).

Manual Review of No Typing Results

5 alleles (0.52% of total referenced alleles) had no typings during analysis with TruSight HLA Assign 2.0 Software and required manual review and editing of the sequence to generate typings (Figure 4). In 4 of the 5 cases, homozygous loci contained 1 to 3 base calls at low frequency. The fifth case included a single low frequency base call in a heterozygous locus. In all cases, no typings were made and there are no examples of incorrect typings being made.

IHW01040 HLA-DRB3: The reference typing for this locus is actually 01 and 01. This seems unlikely given the DRB1 typing of 08:01 and 13:03:01 for which the 08:01 is rarely paired with a DRB3/4/5. Therefore, it is expected to see only one 01:01:02 allele. In this case, the top ranked allele is 01:01:02 and 01:01:02 with this single heterozygous base position and a mismatch to the 01:01:02. This T appears to be present in about 11% of the reads and appears to be called erroneously. The 2 options at this position are to remove the T call, resulting in a 01:01:02 homozygous typing or make no call at the position with the same 01:01:02 homozygous result.

IHW09021 HLA-C: The reference typing for this locus is 17:01. In this case, the top ranked allele pair is 17:01:01 and 17:01:01 with 3 heterozygous base positions as the mismatches to a perfect match. These calls appear to be present in about 11% of the reads and appear to be called erroneously. The 2 options at these positions are to remove the low frequency calls or make all these positions no calls, resulting in a 17:01:01 homozygous typing in both cases.

IHW09044 HLA-DQA1: The reference typing for this locus is 05:01. In this case, the top ranked allele pair is 05:01:01 and 05:05:01 with multiple mismatches to the 05:01. Apparently the actual typing is 05:05:01 homozygous with this single base position as the sole heterozygous call. The C call appears to be present in about 11% of the reads and appears to be called erroneously. The 2 options at this position are to remove the low frequency call or make no call, resulting in a 05:05:01 homozygous typing in both cases.

IHW09058 HLA-DRB1: The reference typing for this locus is 13:01. In this case, the top ranked allele pair is 13:01:01 and 13:01:01 with these 3 heterozygous base positions and mismatches to a perfect match. These calls appear to be present in about 11% of the reads and appear to be called erroneously. The 2 options at these positions are to remove the low frequency calls, resulting in a 13:01:01 homozygous typing, or make all these positions no calls, resulting in a 13:01:01/13:121 homozygous typing and ambiguity.

IHW09378 HLA-B: The reference typing for this locus is 41:02:01 and 51:01:01. A T at about 11% frequency in the reads appears to have been called erroneously. The 2 options at this position are to remove the low frequency call or make no call, both resulting in an unambiguous 41:02:01 and 51:01:01 typing result.

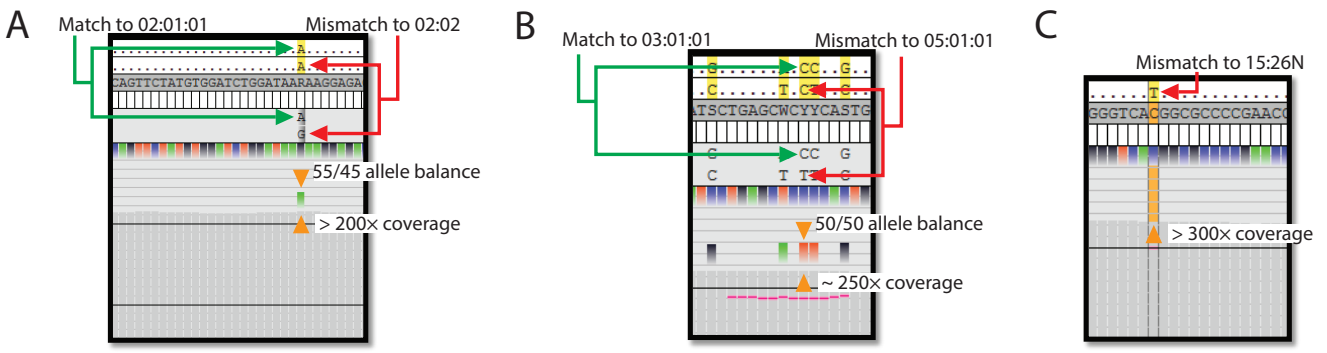

Figure 3: Novel Exonic Variants—The top 1 or 2 rows (white) in each figure are the reference alignments for that particular allele, with expected base calls at specific positions highlighted in yellow. The second or third row (gray) in each figure is the sample consensus sequence made up of hundreds of reads. **(A) IHW09220 and IHW09378 HLA-DPA1** Reference alignments for DPA1*02:01:01 (row 1) and DPA1*02:02 (row 2) both expect an A at position 4314 (Exon 2 position 107). The sample consensus sequence made up of more than 200 reads (row 3) shows a G in ~ 55% of these reads and an A in ~ 45% of the reads. The A was phased and aligned to the 02:01:01 and the G was phased and aligned to the second allele. **(B) IHW09266 HLA-DQB1** Reference alignments for DQB1*03:01:01 (row 1) and DQB1*05:01:01 (row 2) both expect a C at position 612 (Exon 1 position 81). The sample consensus sequence made up of more than 250 reads (row 3) shows a C in ~ 50% of the reads and a T in ~ 50% of the reads. The C was phased and aligned to the 03:01:01 and the T was phased and aligned to the second allele. **(C) IHW09369 HLA-B** The reference typing for this locus is 15:26N homozygote. The reference alignment for B*15:26N (row 1) expects a T at position 295 (Exon 1 position 11). The sample consensus sequence made up of more than 300 reads (row 2) shows a C at this position in all reads.

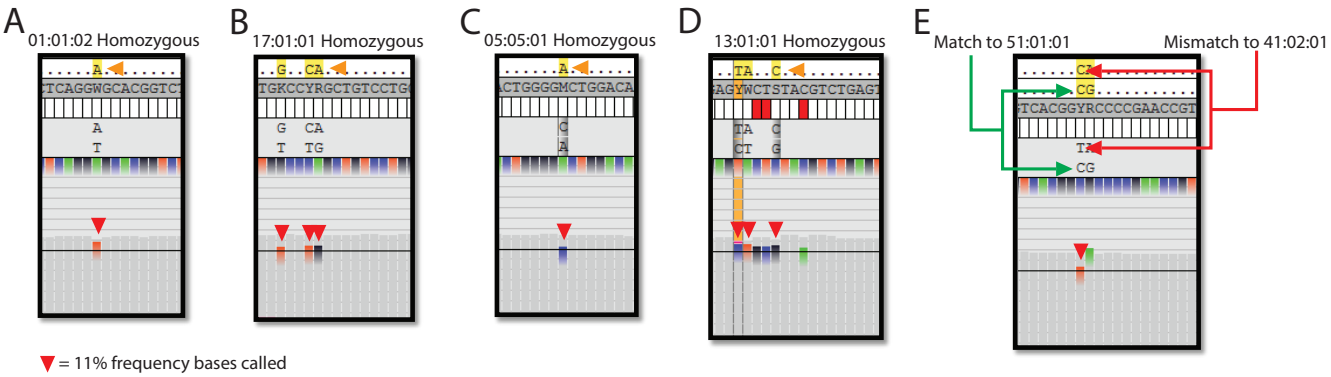

Figure 4: Manual Review of No Typing Results and Edits—5 alleles resulted in no typings during analysis with TruSight HLA Assign 2.0 Software and required manual review and editing. In all cases, the calls made appear to be present in ~ 11% of the reads (red triangles) and appear to be called erroneously. Options are to remove the low frequency calls or make all such positions no calls. **(A) IHW01040 HLA-DRB3** The reference typing for this locus is actually 01 and 01. In this case, the top ranked allele is 01:01:02 and 01:01:02 with this single heterozygous base position and a mismatch to the 01:01:02. **(B) IHW09021 HLA-C** The reference typing for this locus is 17:01. In this case, the top ranked allele pair is 17:01:01 and 17:01:01 with these 3 heterozygous base positions as the mismatches to a perfect match. **(C) IHW09044 HLA-DQA1** The reference typing for this locus is 05:01. In this case, the top ranked allele pair is 05:01:01 and 05:05:01 with multiple mismatches to the 05:01. Apparently the actual typing is 05:05:01 homozygous with this single base position as the sole heterozygous call. **(D) IHW09058 HLA-DRB1** The reference typing for this locus is 13:01. In this case, the top ranked allele pair is 13:01:01 and 13:01:01 with these 3 heterozygous base positions as the mismatches to a perfect match. **(E) IHW09378 HLA-B** The reference typing for this locus is 41:02:01 and 51:01:01, with a no typing result for the 41:02:01 allele.

Resolving HLA Typing Ambiguities

There were 41 ambiguities found in the 1292 sequenced alleles (Figure 5). There are 3 types of ambiguities present in these data (Table 5).

1. **Amplicon Ambiguities**: These ambiguities were resolved outside the amplicon and include DPB1*13:01:01/107:01 (13:01:01G group), DRB1*12:01:01/12:10 (12:01:01G group), and DRB1*08:01:01/03. All 3 of these ambiguities were resolved in exon 1 of their respective loci and exons 1 of DPB1 and DRB1 are not covered by TruSight HLA v2. Amplicon ambiguities accounted for 8 of the 41 ambiguities (19.51%). These amplicon ambiguities are highlighted in blue.
2. **Conditional Ambiguities**: These ambiguities appear based on the combination of alleles and usually occur because the homology between the alleles leads to large gaps in heterozygous base positions that are used to phase. With gaps of 1Kb or more, phase will be lost resulting in a phase ambiguity. This most commonly occurs in DPB1 in which intron 2 is over 4500 bases in length with certain combinations having het gaps of more than 4Kb (eg DPB1*04:01:01 paired with DPB1*04:02:01). These conditional ambiguities will appear every time these alleles are combined with one another, but not every time the allele is present in a different combination. Conditional ambiguities occurred in 28 of the 41 (68.29%) ambiguities and account for all the DQB1 ambiguities and all the DPB1 ambiguities except for the 13:01:01/107:01. Conditional ambiguities are highlighted in green.
3. **Phase Ambiguities**: These ambiguities rarely occur and are specific to a sequencing run. In other words, the same sample and allele combination will usually sequence unambiguously, but on occasion may not have the appropriate read structure to resolve phase. There are 3 loci (5 of 41 alleles, 12.20%) in which this occurred in these runs and these are highlighted in pink.

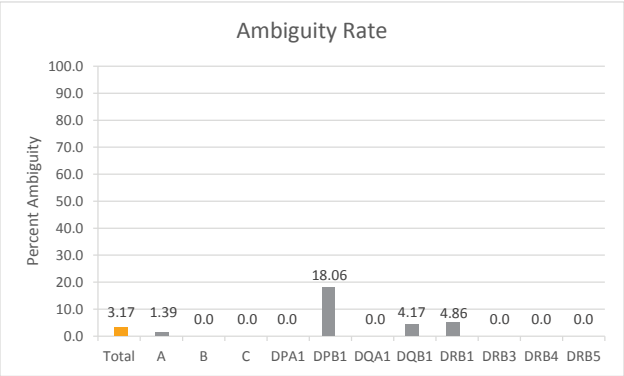

Figure 5: Ambiguity Rates of the TruSight HLA v2 Sequencing Panel on the MiSeq System—The ambiguity rate was calculated using the 1292 alleles that were sequenced and these data were calculated independently of the references. There were 41 total alleles with ambiguities.

Table 5: TruSight HLA v2 Sequencing Panel Ambiguities on the MiSeq System

Sample ID	Locus	Ambiguity
IHW01018	DRB1	08:01:01/03
IHW01040	DPB1	03:01:01/124:01
		04:01:01/350:01
IHW01040	DRB1	08:01:01/03
IHW01093	DPB1	04:01:01/105:01
		04:02:01/126:01
IHW01137	DPB1	04:01:01/105:01
		04:02:01/126:01
IHW01141	DPB1	04:01:01/105:01
		04:02:01/126:01
IHW01141	DRB1	12:01:01/12:10
IHW01174	DPB1	03:01:01/124:01
		04:01:01/350:01
IHW09044	DPB1	02:01:02/105:01
		04:02:01/416:01
IHW09045	DPB1	02:01:02/105:01
		04:02:01/416:01
IHW09045	DRB1	12:01:01/12:10
IHW09056	DPB1	13:01:01/107:01
IHW09013	DPB1	02:01:02/106:01
		19:01/414:01
IHW09114	DRB1	13:01:01/107:01
IHW09267	DQB1	06:03:01/06:39
		06:04:01/06:41
IHW09273	DRB1	12:01:01/12:10
IHW09366	DQB1	06:02:01/06:39
		06:04:01/06:84
IHW09378	DQB1	06:03:01/06:41
		06:09:01/06:88
IHW09388	DPB1	04:01:01/105:01
		04:02:01/126:01
IHW09398	DRB1	08:04:01/08:20
		13:03:01/08:40
IHW09417	DPB1	04:01:01/105:01
		04:02:01/126:01
IHW09501	DPB1	04:04:01/133:01
		13:01:01/107:01/350:01
IHW09502	A	02:05:01/02:22:01
		02:01:01/02:14
IHW09502	DPB1	03:01:01/124:01
		04:01:01/350:01

Results: MiniSeq System

Evaluation of Concordance with Reference Typing

Sequencing of the 72 IHWG samples on the MiniSeq System resulted in consensus sequences for 1292 alleles, 959 of which had references for comparison. 906 (94.47%) were concordant with the reference typing. The 52 typing results that did not match the reference sequence fell into 3 categories: 42 (4.38% of total referenced alleles) had incorrect or incomplete reference typings that prompted updates to the reference database (Table 4), 5 (0.52% of total referenced alleles) had novel exonic variants, and 6 (0.63% of total referenced alleles) required manual editing. The reference updates highlighted based on typing with the TruSight HLA v2 Sequencing Panel (Table 4) remain unchanged as all instances of incorrect or incomplete reference typing from sequencing on the MiSeq System are identical with data from the MiniSeq System.

All alleles with references for which the results were concordant, alleles with references that appear to be incorrect, and novel alleles were used to calculate accuracy for the TruSight HLA v2 Sequencing Panel (Figure 6). 6 alleles were considered inaccurate. Typing with the TruSight HLA v2 Sequencing Panel on the MiniSeq System resulted in similar accuracy compared to the MiSeq System (Figure 6).

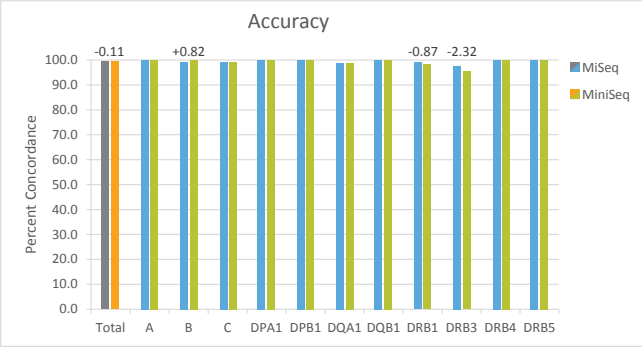

Figure 6: Accuracy of the TruSight HLA v2 Sequencing Panel on the MiniSeq System—The accuracy is calculated by using 959 alleles with references. A TruSight HLA v2 Sequencing Panel result was considered accurate when a result was fully concordant (906 alleles, 94.47%), when the result matched an update to the reference (42 alleles, 4.38%), and when the allele contained a novel variant (5 alleles, 0.52%) for a total of 953 alleles called accurately (99.37%). Therefore, there were 6 alleles that were considered inaccurate. Differences in total accuracy and accuracy for each allele between the MiniSeq (orange/green bars) and MiSeq (gray/blue bars) Systems are indicated.

HLA Typing Ambiguities on the MiniSeq System

There were 42 ambiguities found in the 1292 sequenced alleles (Figure 7). Ambiguities found with the TruSight HLA v2 Sequencing Panel on the MiniSeq System are largely identical to those occurring on the MiSeq System (Table 5). There were 8 loci that differed between the MiniSeq and MiSeq Systems (Table 6).

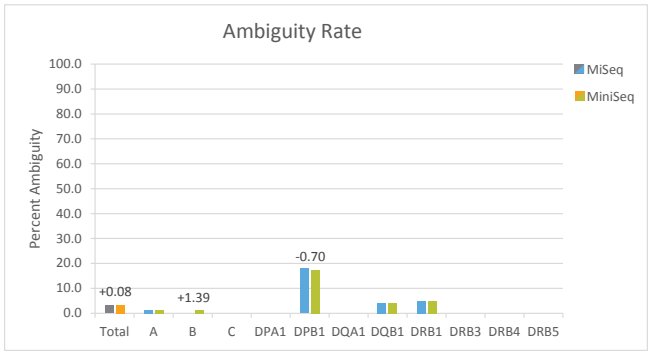

Figure 7: Ambiguity Rates of the TruSight HLA v2 Sequencing Panel on the MiniSeq System—The ambiguity rate was calculated using the 1292 alleles that were sequenced, and these data were calculated independently of the references. There were 42 total alleles with ambiguities. Differences in total ambiguity rate and ambiguity rates for each allele between the MiniSeq (orange/green bars) and MiSeq (gray/blue bars) Systems are indicated.

Table 6: TruSight HLA v2 Sequencing Panel Differences Between the MiSeq and MiniSeq Systems

Sample ID	Locus	Reference Typing	MiSeq System Typing	MiniSeq System Typing	System with Correct Typing	Notes
IHW01040	DRB1	08:01	08:01:01/03	08:01:01/03		
		13:03	13:03:01	13:55	MiSeq System	
IHW09058	DRB1	13:01:01	--:--	13:01:01	MiniSeq System	This locus required 3 base edits when sequenced on the MiSeq System.
			--:--	X		
IHW09084	DRB1	16:01	16:01:01	--:--	MiSeq System	This locus resulted in no typing being made with the MiniSeq System data.
			X	--:--		
IHW09084	DRB5		02:02	--:--	MiSeq System	This locus resulted in no typing being made with the MiniSeq System data.
			X	--:--		
IHW09378	B	41:02:01	--:--	41:02:01	MiniSeq System	This locus required 1 base edit when sequenced on the MiSeq System.
		51:01:01	51:01:01	51:01:01		
IHW09459	DRB3	02:02:01	02:02:01	--:--	MiSeq System	This locus requires 1 base edit when sequenced on the MiniSeq System.
			X	--:--		
IHW09501	DPB1		04:04:01/133:01	04:01:01	Both	This locus contained a phase ambiguity when sequenced on the MiSeq System.
			13:01:01/107:01/350:01	13:01:01/107:01		
IHW09502	A		02:01:01/02:14	02:01:01	Both	This locus contained a phase ambiguity when sequenced on the MiSeq System.
			02:05:01/02:22:01	02:05:01		

Conclusions

The TruSight HLA v2 Sequencing Panel is a fast and accurate method for high-resolution HLA typing. With this method, labs can generate HLA sequences and corresponding typing results in less than 48 hours from receipt of sample. Typing results are delivered for up to 11 loci per sample with high accuracy (99.47%) and low ambiguity rates (3.17%). This level of performance is achieved with no manual editing of the sequence. In addition, the performance of TruSight HLA v2 libraries were assessed side-by-side on the MiSeq and the MiniSeq Systems, comparing these instruments and 4-channel and 2-channel chemistries. There is not a significant difference between these systems and both the MiSeq and the MiniSeq Systems perform with high accuracy and low ambiguity rates.

Additional Resources

HLA Sequencing of International Histocompatibility Working Group (IHWG) Reference Samples with TruSight HLA v2 Sequencing Panel Supplementary Tables (available online)

TruSight HLA Assign 2.0 72-Sample Panel on MiSeq System Project File

TruSight HLA Assign 2.0 72-Sample Panel on MiniSeq System Project File

TruSight HLA v2 Sequencing Panel Reference Guide

TruSight HLA Assign 2.0 Reference Guide

Learn More

To learn more about the TruSight HLA v2 Sequencing Panel and TruSight HLA Assign 2.0 Software, visit www.illumina.com/hlaseq

References

1. Robinson J, Halliwell JA, Hayhurst JH, et al. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res*. 2015;43:D423-431.
2. Marsh SGE, Albert ED, Bodmer WF, et al. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*. 2010;75:291-455.
3. Cough SCL and Simmonds MJ. The HLA region and Autoimmune Disease: Associations and Mechanisms of Action. *Curr Genomics*. 2007;8:453-465.
4. Woodrow JC and Ilchysyn A. HLA antigens in psoriasis and psoriatic arthritis. *J Med Genetics*. 1985;22:492-495.
5. Fogdell A, Hillert J, Sachs C, and Olerup O. The multiple sclerosis- and narcolepsy-associated HLA class II haplotype includes the DRB5*0101 allele. *Tissue Antigens*. 1995;46:333-336.
6. Woodrow JC and Ilchysyn A. HLA antigens in psoriasis and psoriatic arthritis. *J Med Genetics*. 1985;22:492-495.
7. Carrington M, Nelson GW, Martin MP, et al. HLA and HIV-1: Heterozygote Advantage and B*35-Cw*04 Disadvantage. *Sci*. 1999;283:1748-1752.
8. Mallal S, Nolan D, Witt C, et al. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet*. 2002;359:727-732.
9. Ferrell PB and McLeod HL. Carbamazepine, HLA-B*1502 and risk of Stevens-Johnson syndrome and toxic epidermal necrolysis: US FDA recommendations. *Pharmacogenomics*. 2008;9:1543-1546.
10. Daly AK, Donaldson PT, Bhatnagar P, et al. HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat Genetics*. 2009;41:816-819.
11. Alfirevic A and Pirmohamed M. Predictive Genetic Testing for Drug-Induced Liver Injury: Considerations of Clinical Utility. *Clin Pharmacol Ther*. 2012;92:376-380.
12. Romano A, De Santis A, Romito A, et al. Delayed Hypersensitivity to Aminopenicillins Is Related to Major Histocompatibility Complex Genes. *Ann Allergy Asthma Immunol*. 1998;80:433-437.
13. Lonjou C, Borot N, Sekula P, et al. A European study of HLA-B in Stevens–Johnson syndrome and toxic epidermal necrolysis related to five high-risk drugs. *Pharmacogenet and Genomics*. 2008;18:99-107.
14. Dekker JW, Nizankowska E, Schmitz-Schumann M, et al. Aspirin-induced asthma and HLA-DRB1 and HLA-DPB1 genotypes. *Clin Exp Allergy*. 1997;27:574-577.
15. International Histocompatibility Working Group Website, Reference Panels. http://ihwg.org/reference/index.html. Accessed April 18, 2016.
16. Mack SJ, Cano P, Hollenbach JA, et al. Common and well-documented HLA alleles: 2012 update to the CWD catalogue. *Tissue Antigens*. 2013;81:194-203.
17. Wittig M, Anmarkrud JA, Kassens JC, et al. Development of a high-resolution NGS-based HLA-typing and analysis pipeline. *Nucleic Acids Res*. 2015;43:e70.

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

illumina®