Human Immunology 75 (2014) 1040–1046

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Next generation sequencing to determine HLA class II genotypes in a cohort of hematopoietic cell transplant patients and donors

Anajane G. Smith [a], Chul-Woo Pyo [a], Wyatt Nelson [a], Edward Gow [a], Ruihan Wang [a], Shu Shen [a], Maggie Sprague [b], Shalini E. Pereira [a,b], Daniel E. Geraghty [a,*], John A. Hansen [a,b,*]

[a] Fred Hutchinson Cancer Research Center, Seattle, WA 98109, United States
[b] Clinical Immunogenetics Laboratory, Seattle Cancer Care Alliance, Seattle, WA 98109, United States

ARTICLE INFO

Article history:
Received 15 May 2014
Accepted 18 August 2014
Available online 27 August 2014

Keywords:
Next generation sequencing
HLA-class II
Genotyping
High resolution
Multiplex reaction PCR

ABSTRACT

Current high-resolution HLA typing technologies frequently produce ambiguous results that mandate extended testing prior to reporting. Through multiplex sequencing of individual amplicons from many individuals at multiple loci, next generation sequencing (NGS) promises to eliminate heterozygote ambiguities and extend the breadth of genetic data acquired with little additional effort. We report here on assessment of a novel NGS HLA genotyping system for resequencing exons 2 and 3 of DRB1/B3/B4/B5, DQA1 and DQB1 and exon 2 of DPA1 and DPB1 on the MiSeq platform. In a cohort of 2605 hematopoietic cell transplant recipients and donors, NGS achieved 99.6% accuracy for DRB1 allele assignments and 99.5% for DQB1, compared to legacy genotypes generated pretransplant. NGS provided at least single 4-digit allele resolution for 97% of genotypes at DRB1 and 100% at DQB1. Overall, NGS typing identified 166 class II alleles, including 9 novel sequences with greater than 99% accuracy for DRB1 and DQB1 genotypes and elimination of diploid ambiguities through in-phase sequencing demonstrated the robust reliability of the NGS HLA genotyping reagents and analysis software employed in this study.

1. Introduction

The human leukocyte antigen complex (HLA) comprises a 4 Mb region on chromosome 6p21.3 with many genes involved in immune function. The HLA class I and class II genes encode molecules that bind a vast array of endogenous and foreign peptide fragments for presentation to T lymphocytes and, thus, play fundamental roles in the body's recognition of self and non-self, defense against infectious disease, susceptibility to autoimmune disease, and in the induction of alloimmune responses in the context of transfusion and transplantation. Molecular typing has revealed an extraordinary level of variation among HLA-A, -B, -C, DR, DQ and DP genes, with extensive sharing of polymorphic sequence cassettes between HLA alleles. Without separating the alleles in heterozygous individuals, the inability to assign the cis–trans phase of polymorphisms frequently generates ambiguous typing results. Laboratories that require high resolution typing for selecting donors for hematopoietic cell transplantation (HCT) have developed complex, time-consuming and expensive strategies for separating the HLA alleles in heterozygous specimens [1–5]. Another challenge has been the identification of alleles with polymorphism outside of the antigen recognition site (ARS), encoded by exon 2 in HLA-class II genes. Current genotyping strategies in clinical laboratories focus largely on polymorphism within the HLA ARS which plays a critical function in T cell mediated adaptive immunity [4,6].

Next generation sequencing (NGS) technology minimizes challenges inherent in current HLA typing technologies [7,8]. NGS produces fully phased HLA typing in a single pass by parallel sequencing of individual DNA transcripts, eliminating heterozygote phase ambiguity within a region targeted for amplification, and simplifying data analysis. NGS also accommodates multiplexing of specimens and simultaneous analysis of multiple genes, including regions outside the ARS, at little additional cost. The NGS system evaluated in this study involved sequencing of locus specific PCR amplicons using the MiSeq platform (Illumina, San Diego, CA). The system included software for extracting the HLA genotypes of individual samples from the multiplexed raw NGS

* Corresponding authors. Address: Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, D4-100, Seattle, WA 98109-1024, United States (D.E. Geraghty). Address: Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, D2-100, Seattle, WA 98109-1024, United States (J.A. Hansen).
E-mail addresses: geraghty@fhcrc.org (D.E. Geraghty), jhansen@fhcrc.org (J.A. Hansen).

http://dx.doi.org/10.1016/j.humimm.2014.08.206
0198-8859/

data from exons 2 and 3 of DRB1, DRB3, DRB4, DRB5, DQA1, and DQB1 and exon 2 of DPA1 and DPB1. This approach was tested on a cohort of 2605 HCT donors and recipients for whom legacy DRB1 and DQB1 data were available.

2. Materials and methods

2.1. Study population

The study population comprised 2605 HCT patients and donors transplanted at the Fred Hutchinson Cancer Research Center and the Seattle Cancer Care Alliance between 1992 and 2004 [9]. The cohort included 486 HLA identical sibling pairs, 111 HLA mismatched related pairs, 502 unrelated donor-recipient pairs, and 407 recipient and donor singletons. Patients were self-described as Caucasian (86%) and "other", including African Americans, Asians, Hispanics and Native Americans. All recipient and donor samples were collected before HCT according to Institutional Review Board approved human subjects research protocols.

2.2. DNA specimens

Genomic DNA, at a target concentration of 60 ng/μl, was distributed into 384 well plates together with 83 "no DNA" negative controls. Subsequent sample handling for PCR set up was performed with a robot (Biomek, Beckman Coulter, Brea, CA).

2.3. NGS-based HLA typing system

Genotyping was carried out using reagents and software obtained from Scisco Genetics Inc. (Seattle, WA http://www.sciscogenetics.com/services/integrated-genotyping-system/). Eleven locus specific primer pairs were used for initial PCR amplification of exons 2 and 3 from DQA1, DQB1, DRB1, DRB3, DRB4 and DRB5 and exon 2 from DPA1 and DPB1, including four primer pairs specific for DRB genes, 3 for DQB genes, 2 for DQA, and 1 each for DPA and DPB genes (Fig. 1). Initial PCR used 2 μl of the 60 ng/μl archived DNA for a nominal quantity of 120 ng per amplification. All PCR amplicons of individual samples were pooled, purified through Exonuclease I and Alkaline Phosphatase treatment, combined with adaptor sequences, and tagged with unique index sequences [10], in a separate PCR reaction (adaptor and index sequences are provided in Supplementary Table S1 NGS approach). Subsequently, all the samples in a 384 well plate batch, each uniquely identified or barcoded, were pooled and multiplex sequencing was carried out using the MiSeq Sequencing System (Illumina, San Diego, CA).

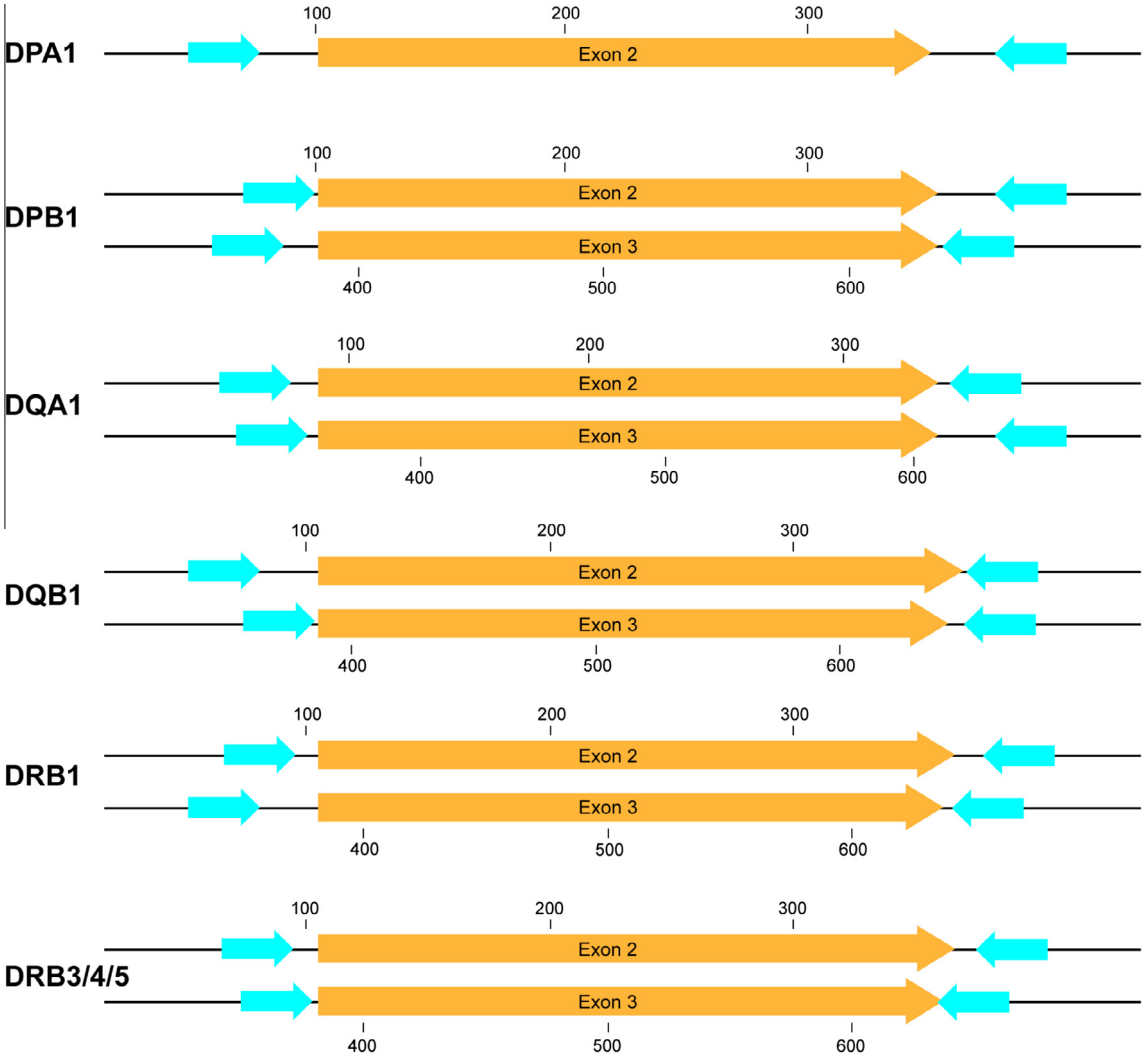

Fig. 1. Amplification strategy for NGS HLA class II typing. PCR primer locations are indicated by arrows flanking exon 2 of DPA1 and exons 2 and 3 of DPB1, DQA1, DQB1, DRB1 and DRB3/4/5. All primers lie within the flanking intron sequence with the exception of HLA-DPB1 exon 2 forward which overlaps by 1 base with the exon sequence. Although only DPB1 exon 2 was typed by NGS in this study, primers for DPB1 exon 3 are now available, as indicated above.

The analysis software compiled the MiSeq paired end read data for each barcoded sample and compared the sequences with ".fasta" data from the IMGT HLA database (http://www.ebi.ac.uk/ipd/imgt/hla/download.htm) to identify the most likely allele match. Statistics for the likelihood of each allele match include (1) the number of paired ends with a perfect match; (2) the number of paired ends with a single mismatch; (3) the number of unique pairs for a given type; (4) the number of positions with >30% mismatching; (5) consistency between independently derived exon data; and (6) aggregate quality values of base calls at each position. Phase between exons was determined by coincidence with established databases.

The HLA analysis software is a Java application and can be run as a standalone analysis program or as a component of the Genetics Management System (GeMS) a platform designed to support workflow management, data analysis, and record keeping. GeMS (http://www.sciscogenetics.com/wp-content/uploads/2013/11/IGS-flyer.pdf) consists of a desktop application deployed in either a 2-tier client–server configuration or as a stand-alone application with an embedded database. The application is written in Java and based on the Eclipse RCP application framework. The desktop portion of GeMS is pure Java and can be run on Windows, Mac, or Linux desktop or laptop computers. When run in the client–server configuration, a platform specific PostgreSQL database is used.

The DQA1, DQB1 and DRB1/3/4/5 genotypes identified by NGS were reported as single alleles or allele groups with polymorphism outside of exons 2 and 3 (indicated with ‡ in tables and text). DPA1 and DPB1 results not discriminated to single alleles were reported with the "G" designations for groups that differ outside of exon 2 [11] (http://www.ebi.ac.uk/ipd/imgt/hla/ambig.html).

NGS HLA genotyping and data analysis was carried out in a blinded fashion with no knowledge of Legacy typing results or the relationship between specimens. Allele assignments were reported without review of results from the other class II loci or reference to known allele frequencies.

2.4. Legacy typing

The original HLA results generated for the study cohort used technologies such as SSOP, SSP and SBT that evolved from 1992 through 2004 [12–14]. Reagents in early assays were designed to detect the limited exon 2 polymorphism that defined DRB1 and DQB1 alleles known at the time of the typing. Legacy genotypes were designated at the four digit level and alleles with silent or synonymous nucleotide substitutions in exon 2 were not captured in the Legacy data. For example, DRB1*11:01:01 and DRB1*11:01:02, reported in legacy as DRB1*11:01, differ by a silent nucleotide substitution at codon 90 where both ACA (*11:01:01) and ACG (*11:01:02) code for threonine.

2.5. Discrepancy resolution

Discordant results between NGS and legacy were resolved by reviewing the original typing records and the NGS raw data. NGS allele assignments were compared in related donor and recipient samples. Four specimens, with previously unknown DQB1 sequences, were analyzed by traditional Sanger sequencing (AlleleSEQR, Celera, Alameda, CA) with Assign software (Conexio, Freemantle, Au).

3. Results

NGS HLA allele assignments for DRB1 and DQB1 were compared with the legacy four digit allele assignments, while DQA1 genotypes were reviewed in the context of common, expected DRB1–DQA1–DQB1 associations (Ref. [25], Begovich et al.). 430 specimens (17%) with a discordant allele assignment were flagged for review. Of those, 316 discordant DRB1 ($n = 106$) or DQB1 alleles ($n = 210$) involved exon 3 polymorphism not assessed by Legacy typing. Thirteen specimens were excluded from analysis because of sample identity issues and NGS results of nine specimens were excluded due to pipetting errors during the manual transfer of the archived DNA to the 384 well plates used in NGS procedures. Among the remaining 2583 specimens, a minority had too few sequence reads to allow allele assignments at one or more loci. Nevertheless, NGS allele assignments were available for 2523 specimens (98%) at DRB1, 2443 (95%) at DQB1, 2498 (97%) at DQA1, 2527 (98%) at DPA1 and 2533 (98%) at DPB1, and across all loci including DRB3, DRB4 and DRB5 for over 90% of the specimens.

Legacy typing was discordant for 112 allele assignments at DRB1 (Table 1), of which 6 (0.12%) involved exon 2 polymorphism: 2 discordant DRB1*15 alleles, 2 discordant DRB1*11 alleles, and one each of DRB1*04 and DRB1*03 alleles. The other 106 discordant assignments involved exon 3 polymorphism, with Legacy

Table 1
Legacy DRB1 and DQB1 discordant data.

DRB1 Legacy	DRB1 NGS	Exon	N[#]	DQB1 Legacy	DQB1 NGS	Exon	N[+]
*04:08[+]	*04:04	2	1	*06:04[+]	*06:09	2	9
*11:02[+]	*11:03	2	1	*03:01/*03:02[+]	*03:04	2	5
*03:01[+]	*03:15	2	1	*03:01[+]	*03:02	2	3
*11:01[+]	*11:39	2	1	DQ7	*03:02	2	2
*15:01[#]	*15:22	2	1	*03:02[+]	*03:01	2	1
*15XX[#]	*15:27	2	1	*03:05[+]	*03:02	2	1
*14:01[^]	*14:54[^]	3	106	*06:02[+]	*06:10	2	2
				*06:02[+]	*06:11	2	2
				*02[+]	*02new	2	1
				*06XX[#]	*06:17	2	2
				*03:01[+]	*03:19[^]	3	22
				*02:01[+]	*02:02[^]	3	183
				*03:01[+], *04:02[+], *05:01[+]	*03[n], 04[n], 05[n]	3	5
Subtotal			112 (2.2%)	Subtotal			238 (4.9%)

Exon indicates where the discordant alleles differ, exon 2 versus exon 3.
N[1], Number of discordant DRB1 alleles among a total of 5046 DRB1 assignments.
N[2], Number of discordant DQB1 alleles among a total of 4886 DQB1 assignments.
[#] Clinical reports included comments describing identification of a previously unknown allele sequence (DRB1 = 2; DQB1 = 1).
[+] Legacy typing discordant by PCR-SSOP technology (DRB1 = 4; DQB1 = 24).
[^] Alleles differ in exon 3, not analyzed by Legacy typing (DRB1 = 106; DQB1 = 210).
[n] New DQB1 allele sequence with polymorphism in exon 3, not analyzed by Legacy typing.

typing of DRB1*14:01 versus DRB1*14:54:01 by NGS. Only seven of DRB1*14:01 in the Legacy data were confirmed by NGS as DRB1*14:01:01.

NGS typing at DRB1 was discordant for 21 of 5046 (0.4%) allele assignments, all of which resulted in false homozygosity (Table 2). In one case, DRB1*14:08 was not reported due to a "rule set" error in the HLA allele calling script. Twenty specimens with low sequence read numbers gave concordant assignments for one allele but failed to assign a second specificity. DRB1*07:01 was missed in 17 of 651 (2.6%) DRB1*07:01 positive cases and one each of DRB1*01:01, *11:01 or *13:02 were not assigned.

At DQB1, Legacy typing was discordant for 238 allele assignments (Table 1), of which 28 (0.6%) involved exon 2 polymorphism. Legacy assignments for DQB1*06 alleles were most frequently discordant with 9 of DQB1*06:04 found to be DQB1*06:09 by NGS, 4 of DQB1*06:02 typed by NGS as DQB1*06:10 ($n = 2$) or DQB1*06:11 ($n = 2$), and 2 of legacy DQB1*06XX identified by NGS as DQB1*06:17. Another 12 discordant alleles involved Legacy typing of DQB1*03 alleles, with miscalls for DQB1*03:04 ($n = 5$) most frequent. One case involved a novel DQB1*02 sequence with exon 2 polymorphism identified by NGS. The 210 discordant alleles with exon 3 polymorphism included 22 of DQB1*03:19 (NGS) versus DQB1*03:01 (legacy) and 183 examples of DQB1*02:01 in legacy versus DQB1*02:02 by NGS. The other 5 cases involved 3 previously unknown exon 3 sequences detected by NGS: DQB1*03:01new ($n = 2$), DQB1*04:02new ($n = 2$) and DQB1*05:01new ($n = 1$).

NGS typing at DQB1 was discordant for 26 (0.5%) allele assignments (Table 2). Low sequence reads in 7 specimens resulted in 4 of DQB1*02:02 reported as DQB1*02:01 and 3 cases with one concordant allele, but a second specificity was not assigned. Script "rule set" errors caused 4 cases of false homozygosity when DQB1*05:04 was not assigned in heterozygous specimens. NGS allele calling script errors resulted in 2 cases of DQB1*06:02, 06:04 reported by NGS as DQB1*06:04, 06:39 and false positive allele calls in 13 homozygous specimens with DQB1*06:02, 06:02 reported as DQB1*06:02, 06:47 ($n = 11$) and DQB1*06:03, 06:03 reported as DQB1*06:03, 06:44 ($n = 2$).

At DQA1, common DRB1–DQA1–DQB1 associations suggested false homozygosity in four specimens with low sequence reads. In three cases, DQA1*05:05 was not identified in the presence of DQA1*05:01 and in one DQA1*01:02 was not detected in the presence of DQA1*01:01.

At DPA1 and DPB1, accuracy of NGS data was assessed by review of DP matching among 475 sibling pairs allele matched for HLA-A/B/C/DRB1/DQA1/DQB1 and with NGS DP results available for both recipient and donor. DP mismatched pairs were flagged for review of the NGS raw sequencing data, but no results were changed. 460 (97%) pairs were reported as DPB1 allele matched and DPB1 matched pairs were also DPA1 matched. 14 of 15 pairs were mismatched for one DPB1 allele and one was mismatched for both DPB1 alleles suggesting 16 DQB1–DPB1 recombination events among the 1900 patient/donor haplotypes giving an inferred recombination rate of $\geqslant 0.008$.

All together, NGS typing identified 166 HLA class II genotypes (Fig. 2). DRB1 was most polymorphic with 56 alleles, followed by 38 of DPB1 alleles, 29 of DQB1, 15 of DQA1, 12 of DPA1, 7 of DRB3, 5 of DRB4 and 4 of DRB5 alleles. These included 9 new allele sequences detected at DRB3 ($n = 2$), DRB4 ($n = 1$), DQB1 ($n = 4$) and DPB1 ($n = 2$), each of which appears to be the result of a single nucleotide polymorphism in a previously known allele. Supplementary Table S2 provides a full listing of the 166 genotypes identified by NGS and Supplementary Table S3 includes descriptions of the new allele sequences.

At DRB1, the DRB1*04 and DRB1*14 allele families were most polymorphic with ten and eight alleles, respectively. NGS identified six alleles each of DRB1*08, DRB1*11, DRB1*13 and DRB1*15, four of DRB1*01, three of DRB1*03, two alleles each of DRB1*12 and DRB1*16 and one each of DRB1*07, DRB1*09 and DRB1*10. NGS genotyping achieved single allele resolution for 49 of the 56 DRB1 (88%) alleles, with only two genotypes differing at the 4 digit level: DRB1*04:07/*04:92 with polymorphism in exon 4 and DRB1*12:01/*12:10 which differ in the DRB1 signal peptide. Five other DRB1 allele sets, DRB1*03:01‡, 07:01‡, *08:01‡, *15:01‡ and *15:03G, exhibit silent substitutions outside of exons 2 and 3.

Sixteen alleles were identified at the HLA-DRB3 ($n = 7$), -DRB4 ($n = 5$) and -DRB5 ($n = 4$) loci, including 3 novel sequences: DRB3*01new, DRB3*02new, and DRB4*01new. Three genotypes were not resolved to the allele level: DRB3*01:01‡ and DRB3*02:02‡ carried silent substitutions outside exons 2 and 3; and DRB4*01:03̂ included DRB4*01:03:01:02N. At DRB5, preliminary NG analysis of a related donor/recipient pair was ambiguous for DRB5*01:02/*01:08N. Examination of the raw NGS data revealed the exon 3 deletion characteristic of DRB5*01:08N.

The twenty-nine DQB1 alleles included nine each of DQB1*03 and DQB1*06, five of DQB1*05, and three each of DQB1*02 and DQB1*04. Four novel DQB1 sequences were identified: DQB1*02-new, DQB1*05new, DQB1*03new, and DQB1*04new. Subsequent to this study, the same DQB1*02 sequence, submitted to IMGT by another laboratory, was named DQB1*02:14 in March 2013 (http://www.ebi.ac.uk/ipd/imgt/hla/) and recent Sanger sequencing confirmed our example of DQB1*02:14. In Supplementary

Table 2
NGS DRB1 and DQB1 discordant data.

DRB1 NGS	N†	DQB1 NGS	N‡
*14:08 not assigned†	1	*05:04 not assigned†	4
*07:01 not assigned^	17	*03:03 not assigned^	1
*01:01 not assigned^	1	*06:02 not assigned^	1
*11:01 not assigned^	1	*06:03 not assigned^	1
*13:02 not assigned^	1	*02:02 reported as *02:01^	4
		*06:02 reported as 06:47‡	11
		*06:03 reported 06:44‡	2
		*06:02 reported as *06:39‡	2
Subtotal	21 (0.4%)	Subtotal	26 (0.5%)

N[1], Number of discordant DRB1 alleles among a total of 5046 DRB1 assignments.
N[2], Number of discordant DQB1 alleles among a total of 4886 DQB1 assignments.
† NGS software rule set errors (DRB1 = 1; DQB1 = 4).
‡ NGS software allele calling script errors (DRB1 = 0; DQB1 = 15).
^ NGS low sequence read errors (DRB1 = 20; DQB1 = 7).

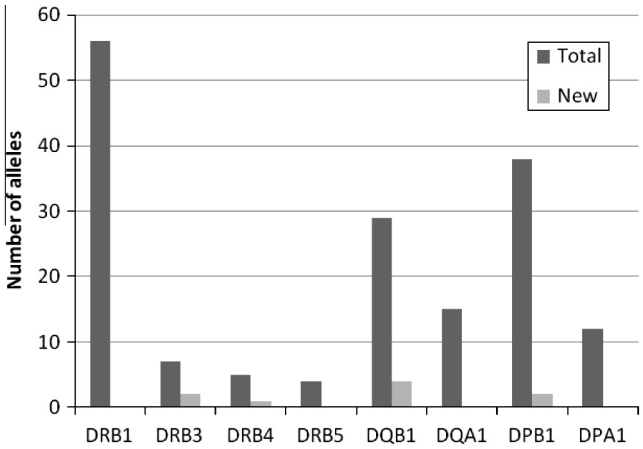

Fig. 2. 166 HLA class II alleles identified by NGS. Total: the total number of genotypes at each locus. DRB1 = 56, DRB3 = 7, DRB4 = 5, DRB5 = 4, DQB1 = 29, DQA1 = 15, DPB1 = 38, DPA1 = 12. New: previously unknown sequences found by NGS. DRB3 = 2, DRB4 = 1, DQB1 = 4, DPB1 = 2.

material, Table S4 and Fig. S1 show the NGS analysis data of the specimen with DQB1*05new and Fig. S2 shows recent Sanger sequencing of the same specimen. Interestingly, the N494 G to A polymorphism is also found in DQB1*05:14, *06:38, and *06:108. NGS achieved single allele resolution for 25 of the 29 DQB1 alleles (86%) and no genotypes differed at the 4 digit level, although DQB1*03:01‡, *03:03‡, *05:01‡, and *05:03‡ have silent substitutions outside of exons 2 and 3.

Next generation sequencing at DQA1 identified 15 genotypes, including five of DQA1*01, three each of DQA1*04 and *05, two of DQA1*03 and one each of DQA1*02 and *06. Four allele groups had silent substitutions outside exons 2 and 3: DQA1*01:02:01‡, *01:02:02‡, *01:03‡, and *05:01‡. Four genotypes were not resolved to the four digit level due to polymorphism in exon 1 and/or exon 4: DQA1*01:01/01:04:01/01:05; DQA1*03:02/03:03:01; DQA1*05:03/05:07; and DQA1*05:05/05:09/05:11. Low NGS reads for exon 3 sequences resulted in 184 DQA1 genotypes that were reported as allele groups with polymorphism outside of exon 2.

At HLA-DP, 12 DPA1 alleles and 38 DPB1 alleles were identified, including two novel DPB1 sequences identified in single unrelated individuals: DPB1*02new and DPB1*04new.

4. Discussion

A novel and very powerful feature of NGS HLA genotyping is the ability to produce fully phased HLA sequences in a single pass, thus eliminating heterozygote phase ambiguity and simplifying data analysis. The multiplex capability, through incorporation of sample specific markers, provides simultaneous sequencing and parallel processing of multiple gene transcripts from many individuals, allowing HLA-class II genotyping to include not only DRB1 and DQB1, but also DQA1, DRB3, DRB4, DRB5, DPA1 and DPB1, as well as regions outside the exon 2 encoded ARS. The increased capacity is achieved with little or no increase in reagent cost, input DNA or time required for machine operation and data analysis, providing single allele resolution in the majority of individuals.

The first well-established NGS protocol, using the 454 platform (Roche Molecular Systems, Pleasanton, CA) was reported in 2009 [15]. Recent improvements in the 454 sequencing workflow and software have simplified bench procedures and data analysis [16]. Life Technologies (Carlsbad, CA) has introduced a novel NGS HLA system using Ion Torrent technology. Approaches to achieve complete genomic sequencing of HLA, including exon and intron sequences extending 2–4 kb, have also been described [17,18]. In these, amplicons from long-range PCR of multiple exons are sheared in a shotgun sequencing approach with computational phasing and assembly of the complete genomic sequences allowing 8-digit allele identification. An NGS approach targeting exons 2 and 3 of HLA-A, B, C, DRB1, DQB1 and DPB1, together with microfluidic chip technology to facilitate routine high volume throughput, has been reported [19]. The NGS HLA typing system (Scisco Genetics, Inc., Seattle, WA) evaluated in our study uses the MiSeq platform (Illumina, San Diego, CA) for sequencing of locus specific PCR amplicons. Novel software was provided to extract the HLA genotypes of individual samples from the multiplexed raw NGS data of exons 2 and 3 for DRB1, DRB3, DRB4, DRB5, DQA1, and DQB1 and exon 2 for DPA1 and DPB1.

The existing legacy HLA data used for comparative analysis in this study employed a succession of molecular methods for analysis of the exon 2 polymorphism exhibited by DRB1 and DQB1 alleles. In the legacy versus NGS comparison, the few legacy miscalls (0.1% of DRB1 and 0.6% of DQB1 allele assignments) involving antigen recognition site polymorphism is reassuring documentation of the overall accuracy (>99.5%) of the historic clinical typing for our HCT cohort. A few miscalls may have been the result of false positive or false negative probe reactions in analyses performed in the early 1990's. However, the majority of the legacy miscalls were likely due to early SSOP assays designed to discriminate the alleles known at the time of the typing. For example, DRB1*03:15 and DRB1*11:39, reported to IMGT in 1999 and 2000, respectively, were not known when early SSOP assigned DRB1*03:01 and DRB1*11:01. In addition, comments in the original clinical reports for DRB1*15:22, DRB1*15:27, and DQB1*06:17 indicated detection of novel sequences. Among the 6 discordant DRB1 allele assignments, only DRB1*03:15 and DRB1*11:39 changed the DRB1 match status from matched to mismatched. Among the 28 discordant DQB1 allele assignments, 3 changed from matched to mismatched and 3 changed from mismatched to matched status.

The Legacy discordant results involving exon 3 polymorphism revealed specific associations previously published or reported to IMGT-HLA. DQB1*02:01:01 and DQB1*02:02 were primarily associated with DRB1*03:01 and DRB1*07:01, respectively [20] and DQB1*03:19 was found with DRB1*11:02 [21]. The more frequent DRB1*14:54:01 was found in association with DRB3*02:02, while the less frequent DRB1*14:01:01 associated with DRB3*02:24 [22–24]. While analysis of exons outside the ARS may not be critical for HCT matching, the higher level of resolution improves the quality of the reported typing by reducing ambiguity in HLA allele assignments.

In addition, it is remarkable that 460 of 475 patient-sibling pairs, with NGS results available for both individuals, were found to be identical for their DPB1 alleles. The inferred DQB1–DPB1 recombination rate of 0.008 (0.8%) is virtually identical to the rate previously reported [25].

The NGS detection of the exon 3 deletion of DRB5*01:08N indicates that this technology should detect other known or novel deletions or insertions that often generate in an HLA allele not expressed on the cell surface. Although DRB4*01:03:01:02N was not discriminated in this study, recent adjustments in the analysis software allow detection of the intron 1, splice site polymorphism of this common DRB4 null allele. Although this HCT cohort was predominantly Caucasian, many of the DRB1 alleles identified by NGS are frequent in minority populations [26], suggesting that the next generation sequencing reagents and software used in this study will provide accurate HLA genotyping in diverse human populations.

The discordant NGS allele assignments (0.4% at DRB1 and 0.5% at DQB1) were attributable to low read sequence data or software problems, including allele calling script errors and allele "rule set" errors. Low reads were largely the result of insufficient archived DNA available for initial PCR. While this study used approximately 120 ng of DNA per amplification, the NGS system described here has successfully used as little as 5 ng per PCR, as well as DNA from cheek swabs, blood spots and phi29 amplification. Defined limits of acceptable sequence read numbers, critical for genotyping accuracy, can be monitored and outlier data flagged with the NGS analysis software. Script errors involved incorrect reporting of a secondary allele when generating output for reporting, but after data processing. The rule set is the panel of known alleles used to type the data. In this analysis, DRB1*14:08 and DQB1*05:04 were inadvertently left out of the rule set and assigned in error as novel sequences, but then not reported in the processed data. Feedback from this study facilitated improvements in the NGS HLA analysis software.

The value of sequence-based high resolution genotyping in the selection of unrelated HCT donors matched for HLA-A, B, C and DRB1 has been well established [1–5]. Other studies have also demonstrated the impact of disparity for HLA-DP, DQ and DRB3/4/5 [27–29] and, recently, shown that mismatching for certain key peptide-binding residues in class I HLA-A, B and C molecules

was associated with increased risk of acute and chronic GVHD and death [30]. Umbilical cord blood has become an important option for HCT partly because of less stringent requirements for HLA matching. [31]. However, a recent publication has shown that higher levels of HLA matching improve survival after umbilical cord blood HCT [32]. These insights into HLA class I and class II mediated allorecognition underscore the value of both the breadth and the depth of data provided by NGS HLA genotyping for the efficient and precise assessment of genetic variation that impacts HCT outcome.

In summary, the NGS typing of this study cohort provided the opportunity to verify the historic legacy DRB1 and DQB1 allele assignments and correct the HLA genotypes for the few specimens with historic allele miscalls. The greater than 99% accuracy of NGS for DRB1 and DQB1 allele assignments, concordance of NGS data among blinded specimens of related donor/recipient pairs, and the identification of novel sequences demonstrate the robust reliability of the NGS-based HLA typing reagents and analysis software employed in this study. It is important to note that the in-phase sequencing, inherent in the next generation technology, provides unambiguous results in a single pass for the HLA class II alleles with polymorphism in exons 2 and 3. Thus, NGS eliminates the complex, time-consuming and expensive extended testing often required by current Sanger sequencing in heterozygous individuals. In addition, the simultaneous sequencing of multiple HLA genes is more cost effective and efficient than current technologies that require independent analysis of each locus. Although this study was limited to HLA class II, the NGS technology applied here has been extended to HLA-class I exons 2, 3, and 4, allowing simultaneous, in-phase analysis of HLA-A, B, C, DRB1, DRB3/4/5, DQB1, DQA1, DPB1 and DPA1 loci.

The extraordinary polymorphism of HLA genes has compelled development of increasingly accurate, sensitive and efficient genotyping methods. Next generation sequencing technology places HLA typing laboratories at the cusp of yet another transformational change that promises to eliminate current confounding phase ambiguities and provide simultaneous analysis of numerous specimens across multiple HLA loci in a single pass of testing with little increase in expenditures. The growing array of NGS HLA typing systems reported in use or under evaluation suggests that several different NGS technologies may become available for routine HLA typing. Each system reported to date is unique in the number of HLA loci analyzed, the amplification and sequencing schemes, and the equipment and software utilized. While current NGS systems are geared primarily toward high throughput laboratories, further innovation should provide applications suitable for all situations and availability of a diversity of systems ensures that both clinical and research laboratories can chose the best system for their particular needs.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.humimm.2014.08.206.

References

[1] Morishima Y, Sasazuki T, Inoko H, Juji T, Akaza T, Yamamoto K, et al. The clinical significance of human leukocyte antigen (HLA) allele compatibility in patients receiving a marrow transplant from serologically HLA-A, HLA-B, and HLA-DR matched unrelated donors. Blood 2002;99:4200–6.

[2] Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, et al. High-resolution donor–recipient HLA matching contributes to the success of unrelated donor marrow transplantation. Blood 2007;110:4576–83.

[3] Woolfrey A, Klein JP, Haagenson M, Spellman S, Petersdorf E, Oudshoorn M, et al. HLA-C antigen mismatch is associated with worse outcome in unrelated donor peripheral blood stem cell transplantation. Biol Blood Marrow Transplant 2011;17:885–92.

[4] Spellman SR, Eapen M, Logan BR, Mueller C, Rubinstein P, Setterholm MI, et al. A perspective on the selection of unrelated donors and cord blood units for transplantation. Blood 2012;120:259–65.

[5] Furst D, Muller C, Vucinic V, Bunjes D, Herr W, Gramatzki M, et al. High-resolution HLA matching in hematopoietic stem cell transplantation: a retrospective collaborative analysis. Blood 2013;122:3220–9.

[6] Horn PA, Albis-Camps M, Verboom M, Bunce M, Yousaf K, Williams S, et al. The nature of diversity of HLA-DRB1 exon 3. Tissue Antigens 2007;70:335–7.

[7] Lind C, Ferriola D, Mackiewicz K, Heron S, Rogers M, Slavich L, et al. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. Hum Immunol 2010;71:1033–42.

[8] Erlich H. HLA DNA typing: past, present, and future (Review). Tissue Antigens 2012;80:1–11.

[9] Chien JW, Zhang XC, Fan W, Wang H, Zhao LP, Martin PJ, et al. Evaluation of published single nucleotide polymorphisms associated with acute graft versus host disease. Blood 2012;119:5311–9.

[10] Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol 2010;11:R119.

[11] Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, et al. Nomenclature for factors of the HLA system, 2010 (Review). Tissue Antigens 2010;75:291–455.

[12] Petersdorf EW, Smith AG, Haase AM, Martin PJ, Hansen JA. Polymorphism of HLA-DRw52-associated DRB1 genes as defined by sequence-specific oligonucleotide probe hybridization and sequencing. Tissue Antigens 1991;38:169–77.

[13] Mickelson E, Smith A, McKinney S, Anderson G, Hansen JA. A comparative study of HLA-DRB1 typing by standard serology and hybridization of non-radioactive sequence-specific oligonucleotide probes to PCR-amplified DNA. Tissue Antigens 1993;41:86–93.

[14] Smith AG, Nelson JL, Regen L, Guthrie LA, Donadi E, Mickelson EM, et al. Six new DR52-associated DRB1 alleles, three of DR8, two of DR11, and one of DR6, reflect a variety of mechanisms which generate polymorphism in the MHC. Tissue Antigens 1996;48:118–26.

[15] Bentley G, Higuchi R, Hoglund B, Goodridge D, Sayer D, Trachtenberg EA, et al. High-resolution, high-throughput HLA genotyping by next-generation sequencing. Tissue Antigens 2009;74:393–403.

[16] Moonsamy PV, Williams T, Bonella P, Holcomb CL, Hoglund BN, Hillman G, et al. High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array™ System for simplified amplicon library preparation. Tissue Antigens 2013;81:141–9.

[17] Wang C, Krishnakumar S, Wilhelmy J, Babrzadeh F, Stepanyan L, Su LF, et al. High-throughput, high-fidelity HLA genotyping with deep sequencing. Proc Natl Acad Sci U S A 2012;109:8676–81.

[18] Shiina T, Suzuki S, Ozaki Y, Taira H, Kikkawa E, Shigenari A, et al. Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers. Tissue Antigens 2012;80:305–16.

[19] Lange V, Bohme I, Hofmann J, Lang K, Sauter J, Schone B, et al. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. BMC Genomics 2014;15:63.

[20] Bodmer JG, Marsh SG, Albert ED, Bodmer WF, Dupont B, Erlich HA, et al. Nomenclature for factors of the HLA system, 1994. Tissue Antigens 1994;44:1–18.

[21] Witter K, Mautner J, Albert T, Zahn R, Kauke T. HLA-DQB1*0319, a novel HLA-DQB1 allele, shows strong haplotype association to HLA-DRB1*1102. Tissue Antigens 2007;70:73–5.

[22] Furst D, Solgi G, Schrezenmeier H, Mytilineos J. The frequency of DRB1*1454 in South German Caucasians. Tissue Antigens 2010;76:57–9.

[23] Lee KW, Jung YA. Additional sequence analysis outside exon 2 clarifies DRB1*12 and DRB1*14 allelic frequencies in Koreans. Hum Immunol 2009;70:464–7.

[24] Xiao Y, Lazaro AM, Masaberg C, Haagenson M, Vierra-Green C, Spellman S, et al. Evaluating the potential impact of mismatches outside the antigen recognition site in unrelated hematopoietic stem cell transplantation: HLA-DRB1*1454 and DRB1*140101. Tissue Antigens 2009;73:595–8.

[25] Begovich AB, McClure GR, Suraj VC, Helmuth RC, Fildes N, Bugawan TL, et al. Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. J Immunol 1992;148:249–58.

[26] Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. Hum Immunol 2007;68:779–88.

[27] Shaw BE, Gooley TA, Malkki M, Madrigal JA, Begovich AB, Horowitz MM, et al. The importance of HLA-DPB1 in unrelated donor haematopoietic cell transplantation. Blood 2007;110:4560–6.

[28] Fleischhauer K, Shaw BE, Gooley T, Malkki M, Bardy P, Bignon JD, et al. Effect of T-cell-epitope matching at HLA-DPB1 in recipients of unrelated-donor haemopoietic-cell transplantation: a retrospective study [Erratum appears in Lancet Oncol. 2012 Apr, 13(4), pp. e134–5]. Lancet Oncol 2012;13:366–74.

[29] Fernandez-Vina MA, Klein JP, Haagenson M, Spellman SR, Anasetti C, Noreen H, et al. Multiple mismatches at the low expression HLA loci DP, DQ, and DRB3/4/5 associate with adverse outcomes in hematopoietic stem cell transplantation. Blood 2013;121:4603–10.

[30] Pidala J, Wang T, Haagenson M, Spellman SR, Askar M, Battiwalla M, et al. Amino acid substitution at peptide-binding pockets of HLA class I molecules increases risk of severe acute GVHD and mortality. Blood 2013;122:3651–8.
[31] Kamani N, Spellman S, Hurley CK, Barker JN, Smith FO, Oudshoorn M, et al. State of the art review: HLA matching and outcome of unrelated donor umbilical cord blood transplants (Review). Biol Blood Marrow Transplant 2008;14:1–6.
[32] Eapen M, Klein JP, Ruggeri A, Spellman S, Lee SJ, Anasetti C, et al. Impact of allele-level HLA matching on outcomes after myeloablative single unit umbilical cord blood transplantation for hematologic malignancy. Blood 2013 [Epub ahead of print; PMID 24141369].