Human Immunology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

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

Very high resolution single pass HLA genotyping using amplicon sequencing on the 454 next generation DNA sequencers: Comparison with Sanger sequencing

F. Yamamoto [a,1], B. Höglund [b,1], M. Fernandez-Vina [a], D. Tyan [a], M. Rastrou [b], T. Williams [b], P. Moonsamy [b], D. Goodridge [c], M. Anderson [a,2], H.A. Erlich [b,*], C.L. Holcomb [b]

[a] Stanford University, Palo Alto, CA, USA
[b] Roche Molecular Systems (RMS), Pleasanton, CA, USA
[c] Conexio Genomics, Perth, Australia

ARTICLE INFO

Article history:
Received 13 February 2015
Revised 19 May 2015
Accepted 21 May 2015
Available online xxxx

Keywords:
DNA sequencing
454 GS FLX
HLA genotyping
NGS

ABSTRACT

Compared to Sanger sequencing, next-generation sequencing offers advantages for high resolution HLA genotyping including increased throughput, lower cost, and reduced genotype ambiguity. Here we describe an enhancement of the Roche 454 GS GType HLA genotyping assay to provide very high resolution (VHR) typing, by the addition of 8 primer pairs to the original 14, to genotype 11 HLA loci. These additional amplicons help resolve common and well-documented alleles and exclude commonly found null alleles in genotype ambiguity strings. Simplification of workflow to reduce the initial preparation effort using early pooling of amplicons or the Fluidigm Access Array™ is also described. Performance of the VHR assay was evaluated on 28 well characterized cell lines using Conexio Assign MPS software which uses genomic, rather than cDNA, reference sequence. Concordance was 98.4%; 1.6% had no genotype assignment. Of concordant calls, 53% were unambiguous. To further assess the assay, 59 clinical samples were genotyped and results compared to unambiguous allele assignments obtained by prior sequence-based typing supplemented with SSO and/or SSP. Concordance was 98.7% with 58.2% as unambiguous calls; 1.3% could not be assigned. Our results show that the amplicon-based VHR assay is robust and can replace current Sanger methodology. Together with software enhancements, it has the potential to provide even higher resolution HLA typing.

1. Introduction

In recent years, various next generation sequencing platforms, based on massively parallel clonal sequencing, have been used to develop high resolution and high throughput HLA typing systems [1–12]. In general, two basic strategies have been used: (1) an amplicon sequencing approach focused on the highly polymorphic regions (primarily exons) of HLA class I and class II genes [1–6], or (2) long-range PCR of full or partial length individual HLA gene loci, followed by fragmentation, shot-gun sequencing and assembly [7–11]. A third strategy, namely sequencing the long range PCR product directly with the long reads made possible by the PacBio SMRT system, is now available [12]. Different applications of HLA sequencing may be best realized with one of these various strategies. Sequencing of long-range PCR products is useful for adding to HLA sequencing databases. The amplicon sequencing strategy of

Abbreviations: gPCR, genomic PCR; CWD, common and well-documented; HLA, human leukocyte antigen; HSCT, hematopoietic stem cell transplantation; HR, high resolution; MID, multiplex identifier; MM, mismatches; NGS, next generation sequencing; PTP, PicoTiter Plate; RMS, Roche Molecular Systems; SSO, sequence-specific oligonucleotide; SSP, Sequence-specific primer; VHR, very high resolution.

* Corresponding author at: Children's Hospital & Research Center, Oakland, CA, 5700 Martin Luther King Jr. Way, USA. Tel.: +1 510 450 7932.

E-mail addresses: fionay@stanford.edu (F. Yamamoto), bryan.hoglund@roche.com (B. Höglund), marcelof@stanford.edu (M. Fernandez-Vina), dtyan@stanford.edu (D. Tyan), melinda.rastrou@roche.com (M. Rastrou), priscilla.moonsamy@roche.com (P. Moonsamy), damian@conexio-genomics.com (D. Goodridge), matthew.anderson@BCW.edu (M. Anderson), herlich@chori.org (H.A. Erlich), cherie.holcomb@roche.com (C.L. Holcomb).

[1] First two authors contributed equally.
[2] Permanent address: Blood Center of Wisconsin, WI, USA.

http://dx.doi.org/10.1016/j.humimm.2015.05.002
0198-8859/

targeting the most informative exons and intron regions, which we have adopted, is advantageous for matching stem cell donor and recipient at high resolution, for research studies (population genetics and disease associations) and for donor registry typing, because the genomic PCR conveniently creates the library without the need for ligation of adaptor sequences and, typically, requires lower read coverage and fewer total reads per sample than that required for assembly of shotgun sequences derived from the whole gene. Amplicon sequencing strategies can be independent of the sequencing platform, provided that the reads are sufficiently long to cover the amplicons, preferably in both directions, and allow for high throughput and low cost per sample HLA typing. The limitation of the selected amplicon (partial gene sequence) strategy, however, is that, for some samples, the level of resolution will be lower and the genotyping ambiguity greater than with other approaches. The challenge for all of these systems based on PCR is the design of primers that amplify individual HLA loci with high specificity yet amplify all alleles at those loci with comparable efficiency. This sometimes results in the placement of primers extending into the exons which limits important sequence data and potentially masks polymorphic sites.

Here we report the development of a "very high resolution" (VHR) HLA genotyping system, which builds upon the previously reported "high resolution" (HR) amplicon sequencing system, [1,2,4] through the resolution of the majority of both the "common and well-documented alleles" (CWD) [13] and many of the null alleles commonly encountered in ambiguity strings. We added 8 amplicons (22 primer pairs in total) to the 14 primer pairs of the Roche 454 GS GType HLA High Resolution (HR) Primer Set, available as two plates, to yield DNA sequence for more exonic and intronic sequence of class I loci, and an additional exon (exon 3) for DRB and exon 2 for DPA1. The VHR-22 primer set amplifies HLA-A exon 1 with intron 1, exon 2, 3, and exons 4 and 5 with the intervening intron; HLA-B and HLA-C exon 1, 2, 3, 4, and 5; HLA-C exon 6 and 7 with intervening intron; DPA1, DPB1, DQA1 exon 2; DQB1 and the DRB loci exons 2 and 3.When generated with "fusion" primers (MID tags and adaptor sequences at the 5′ end of the target specific primers), these 8 amplicons could be generated in one additional genomic PCR plate, for a total of 3 plates. To facilitate high throughput, these 22 amplicons were also generated using the 4-primer system with the Fluidigm Access Array™[4]. A new Conexio Genomics genotyping software, MPS 1.0, was developed to allow use of genomic DNA references and enable the analysis of non-coding regions as well as the exons used previously by the Conexio ATF genotyping software.

The initial VHR-22 workflow was analogous to the previously reported workflow [1,2]. It provided reliable and very high resolution (limited ambiguity) genotyping but required extensive handling of individual amplicons. We also observed that, in a very limited number of cases (exon 2 for DRB1*04 alleles and DQB1*06:01), some alleles were not amplified as well as others in the same sample (preferential allele amplification). Preferential amplification creates a limitation for throughput since a read depth sufficient to obtain accurate genotyping of samples containing alleles that are amplified less efficiently must be obtained (i.e., more reads required). We also noted that, especially for the DRB loci for which generic primers are used, *in vitro* PCR artifacts ("crossover products" between DRB1 and DRB3/4/5) were occasionally detected [14]. In addition, we observed that genomic amplification and emulsion PCR, and sequence read recovery was less robust for the longer amplicons (HLA-C exon 3 and HLA-A exons 4–5).

To address these issues, we made several modifications in the previously published protocol (2). First, we modified the genomic PCR (gPCR) to prevent preferential allele amplification, reduce the formation of crossover products, and increase the robustness of amplification for the longest amplicons. Second, to further increase the number of sequence reads obtained for the longest amplicons (HLA-A exons 4–5 and HLA-C exon 3), we modified the emulsion PCR conditions and added more of these amplicons ("spiking") to the overall amplicon pool library. Third, we simplified the workflow by reducing the handling of amplicons during library preparation.

Finally, a collaborative study with the Stanford Blood Center was carried out in which 59 samples were analyzed to compare the VHR-22 3 plate fusion primer system and the Conexio MPS software to the Sanger-based HLA typing system currently used in most HLA labs.

2. Materials and methods

2.1. Samples

Initial genotyping at RMS was carried out on DNA extracted from cell lines of known genotype. DNA extraction was by use of the Gentra® Puregene® Kit (Qiagen, Valencia, CA). Genomic DNA samples tested by the Stanford Blood Center HLA laboratory (IRB # 24968) were extracted from peripheral blood leukocytes using either the BioRobot EZ1 or the QIAsymphony SP according to manufacturer's protocols. The samples had been previously typed for HLA-A, -B, -C, -DRB1, and DQB1 by AlleleSEQR HLA SBT and HARP (Abbott, Abbott Park, IL) performed with the 3130xl Genetic Analyzer (Life Technologies, Carlsbad, CA). The sequence data were analyzed by Assign software (Conexio Genomics, Perth, Australia). Luminex xMAP solid-phase testing with LABType® SSO typing kits and Fusion software (One Lambda, Canoga Park, CA) was also performed for HLA-A, -B, -C, -DRB1, -DRB3/4/5, -DQA1, -DQB1, -DPA1 and -DPB1when ambiguity resolution was necessary following sequencing.

2.2. Primer designs

2.2.1. Fusion primers

The HR-14 fusion primers (Roche Life Sciences GS GType HLA Assay, Branford, CT) were as described in Holcomb et al. 2011 [2] with the exception that 10 base MIDs and adapter sequences were incorporated to be compatible with the GS FLX and GS Junior Titanium chemistry. It should be noted that the amplicon for HLA-A exon 4 included complete sequence coverage of intron 4 and exon 5; this primer pair was exactly the same as that provided in the 454 Life Sciences GS GType HLA Assay, but the Conexio software was modified to allow sequence analysis of the entire amplicon for this study. As previously described [2], some of the HR-14 primers extended into the exons, and resulted in some residual genotyping ambiguity. These designs were a "trade-off" to maximize allelic coverage while maintaining specificity for a given locus. Note that specificity was less of a consideration for DRB primers in order to allow sequencing of DRB3/4/5 as well as DRB1. The primers for the eight additional amplicons of the VHR set were designed to allow sequencing of exon and intron regions of particular interest in ambiguity resolution. All of these primers gave complete coverage of the targeted exons and full or partial coverage of the targeted introns. Coverage by all primers is summarized in Fig. 1S. We observed some preferential allele amplification for two of the additional amplicons in the VHR set. The DRB primers of exon 3 amplified all alleles of the DRB1 locus but amplified exon 3 of DRB5 more efficiently (as evidenced by more recovered sequence reads) than DRB4 and DRB3. The HLA-A exon 1 primers amplified HLA-A*11:01 less efficiently than all other HLA-A alleles.

Please cite this article in press as: Yamamoto F et al. Very high resolution single pass HLA genotyping using amplicon sequencing on the 454 next generation DNA sequencers: Comparison with Sanger sequencing. Hum Immunol (2015), http://dx.doi.org/10.1016/j.humimm.2015.05.002