Human Immunology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

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

Development and validation of a sample sparing strategy for HLA typing utilizing next generation sequencing

Denise M. McKinney [a], Zheng Fu [b], Lucas Le [a], Jason A. Greenbaum [b], Bjoern Peters [a], Alessandro Sette [a,*]

[a] Department of Vaccine Development, La Jolla Institute for Allergy and Immunology, 9820 Athena Circle, La Jolla, CA 92037, USA
[b] Bioinformatics Core Facility, La Jolla Institute for Allergy and Immunology, 9820 Athena Circle, La Jolla, CA 92037, USA

ARTICLE INFO

Article history:
Received 20 November 2014
Revised 17 April 2015
Accepted 30 April 2015
Available online xxxx

Keywords:
HLA typing
Next generation sequencing
Sample sparing
Pipeline

ABSTRACT

We report the development of a general methodology to genotype HLA class I and class II loci. A Whole Genome Amplification (WGA) step was used as a sample sparing methodology. HLA typing data could be obtained with as few as 300 cells, underlining the usefulness of the methodology for studies for which limited cells are available. The next generation sequencing platform was validated using a panel of cell lines from the International Histocompatibility Working Group (IHWG) for HLA-A, -B, and -C. Concordance with the known, previously determined HLA types was 99%. We next developed a panel of primers to allow HLA typing of alpha and beta chains of the HLA DQ and DP loci and the beta chain of the DRB1 locus. For the beta chain genes, we employed a novel strategy using primers in the intron regions surrounding exon 2, and the introns surrounding exons 3 through 4 (DRB1) or 5 (DQB1 and DPB1). Concordance with previously determined HLA Class II types was also 99%. To increase throughput and decrease cost, we developed strategies combining multiple loci from each donor. Multiplexing of 96 samples per run resulted in increases in throughput of approximately 8-fold. The pipeline developed for this analysis (HLATyphon) is available for download at https://github.com/LJI-Bioinformatics/HLATyphon.

1. Introduction

The Human Leukocyte Antigen (HLA) complex on chromosome 6 is one of the most polymorphic regions in the human genome. The HLA genes play key function in immune activation and regulation, from infectious disease, transplantation, autoimmunity and cancer immunology [1]. Recent years have seen an explosive growth in the development of assays using Next Generation Sequencing (NGS). An obvious application for NGS is for HLA typing. However, surprisingly few groups or companies have published results from HLA typing using NGS [2,3] that are validated on large reference panels.

The IMGT/HLA database currently defines nearly 9000 alleles for HLA Class I genes (HLA-A, -B, and -C), and nearly 3000 HLA Class II genes (HLA-DRB, -DQA/DQB, and -DPA/DPB) [4]. With the explosion of whole genome sequencing projects, the number of novel HLA alleles is increasing at a phenomenal rate. Typically HLA has been typed in the clinic using sequence specific oligonucleotide (SSO) methods, sequence specific primer (SSP) methods, or Sanger sequencing [5]. These methods are reliable, but relatively costly and slow. For the research laboratory, HLA typing of research subjects is generally technically outside the expertise of the group, and most groups have outsourced their typing to commercial laboratories at considerable expense, thus limiting the number of subjects that can be typed. With the improvements and general availability of NGS, a HLA typing assay would allow for most laboratories to type subjects at a reasonable cost.

Recent improvements in NGS technologies have allowed for read lengths that are necessary for distinguishing between the two, often different, alleles expressed in an individual. Compared with the traditional Sanger method, NGS technologies produce many short reads, making the HLA genotyping process much cheaper and faster, as well as reducing ambiguous typing results of heterozygous samples in diploid genomes [6]. To date the hindrance for high-resolution HLA genotyping via NGS-based approach has been the allelic polymorphism of HLA genes. To overcome this issue a few purely computational means were

Abbreviations: WGA, Whole genome amplification; IHWG, International Histocompatibility Working Group; TAB, total aligned bases.

* Corresponding author. Tel.: +1 858 752 6916; fax: +1 858 752 6987.
E-mail address: alex@liai.org (A. Sette).

http://dx.doi.org/10.1016/j.humimm.2015.04.007
0198-8859/

developed: (1) *De novo* assembling the NGS short reads and mapping the resulting contigs against the HLA references, scoring each HLA allele on the basis of contigs depth of coverage, length and percent sequence identity [7,8]; (2) Tree-based top down greedy algorithm by using hierarchical read weighting [9]; (3) Assuming the correct reference(s) should have more mapped reads than the incorrect one(s) and performing count reads maximization [10–13]. Although these methods could yield acceptable concordance rates, their schemes were using arbitrary and empirical criteria that might introduce bias in HLA genotype calling. Thus in this study we developed and validated an in-house HLA-typing pipeline on the basis of counting read depth rather than number of mapped reads with leveraging the NGS data of longer reads.

An important consideration in designing HLA-typing assays for practical clinical research studies is the limited sample availability. In cases where donors are affected by severe pathologies, or in longitudinal studies following donors after vaccination or in clinical studies, only a few million cells are available in total. Accordingly, sample sparing technologies are required, and as few PBMC as possible should be utilized for HLA typing determination, while still preserving accuracy, experimental convenience and contained costs. Whole Genome Amplification (WGA) is a technique developed around 20 years ago to allow for amplification of the subject's genome in the case of limiting sample. Over the years, several types of WGA assays have been developed, generally either PCR based applications or isothermal amplification methods such as Multiple Displacement Amplification (MDA) [14]. Commercial products are available for both WGA methods, including REPLI-g (Qiagen) and PicoPlex WGA (Rubicon Genomics). In particular, MDA methods have been shown in many studies to be accurate and unbiased using limited samples [14,15].

Accordingly, we sought here to validate HLA typing using limited PBMC amounts. Herein we report our results in developing a sample sparing HLA typing assay including an analysis pipeline (HLATyphon) using a MiSeq (Illumina).

2. Material and methods

2.1. Sample preparation

DNA samples from cell lines with known HLA types were obtained from the International Histocompatibility Working Group (IHWG) using the SP Reference Panel, a combination of 51 DNA samples typed using the highest frequency sequence specific methods possible at the time of the 13th workshop (2007).

Alternatively, genomic DNA isolated from PBMC of study subjects by standard techniques (QIAmp; Qiagen, Valencia, CA) was used for HLA typing. All studies using human PBMCs were performed using anonymized samples following approved protocols from the relevant Research Ethics Committees, for which informed consent was obtained from all individual donors. These samples included four studies in San Diego, CA (University of CA and LJI). The cell lines used for the study include AMAI (IHW09010), AMALA (IHW09064), KAS011 (IHW09009), KOSE (IHW09056), KT17 (IHW09024), MGar (IHW09014), RML (IHW09016), RSH (IHW09021), DBB (IHW09052), M7 (IHW09215), and Priess (IHW09301).

WGA (REPLI-g; Qiagen, Valencia, CA) was used to allow typing of limiting amounts of samples. Input cell numbers ranged from 30 to 30,000 cells.

2.2. PCR amplification

Amplicons specific for HLA Class I genes were generated using primers described by Wang et al. [11]. The primers, located in exons 1 and 7, generated a PCR product of approximately 2700 bp. For Class II, amplicons specific for exons 2 and 3 from the appropriate HLA Class II genes were generated by PCR using locus-specific primers designed to amplify most polymorphic HLA genes. The primers were designed using Primer3 (v.0.4.0) [16,17], and compared to the consensus sequences present in the IMGT/HLA database [4]. For the beta chain genes (DRB1, DQB1, and DPB1), two primer sets are used to generate two amplicons spanning exon 2 and exons 3 through 4 (DRB1) or exons 3 though 5 (DQB1 and DPB1). For the alpha chain genes (DPA and DQA), a single set of primers was used to generate an amplicon spanning exons 2 through 4. A primer mix was prepared for each primer set. Each primer in the set was present in the mix at a concentration of 10 μM.

A 46 μl PCR master mix was prepared for each sample containing 10 μl of 5× Crimson LongAmp Taq Reaction Buffer (Crimson LongAmp Taq kit; New England BioLabs, Ipswich, MA), 1.5 μl of 10 mM dNTPs (New England Biolabs, Ispswich, MA), 2 μl (5 units) Crimson LongAmp Taq, 4 μl primer mix (final concentration 0.8 μM for each primer), and 28 μl nuclease-free water (Qiagen). Four μl of DNA (Qiagen purified or REPLI-G material) were added to 46 μl of the PCR master mix in a 96 well PCR tray. After mixing, the PCR was performed on a thermocycler (BioRad or Applied Biosystems). The thermal profile was 94 °C for 2 m, followed by 40 cycles at 94 °C for 30 s, 63 °C for 30 s, and 68 °C for 3 m. A finishing step of 68 °C for 7 m completed the reaction. The PCR reaction products were analyzed for the correct size fragment on a 1% agarose gel electrophoresed at 125 V for 30 min.

Amplicons of the correct size were then purified using Zymo DNA Clean-up Kit, according to the manufacturer's instructions. The purified amplicons were quantitated by fluorescence using Qubit 2.0 (Life Technologies, Carlsbad, CA). The two amplicons for each beta chain gene were combined, and a single library was prepared for the set.

2.3. Sequencing libraries

Sequencing libraries are prepared using Nextera XT reagents (Illumina, San Diego, CA). The Nextera XT kit utilizes an engineered transposome to simultaneously fragment, and tagment input DNA, adding unique adapter sequences and bar codes in the process. Briefly, 96 libraries were prepared in parallel, using 1 ng of input DNA amplicon(s) for each library. The two amplicons for each of the beta chains were combined, and one ng of the combined material was used. The protocol provided by the manufacturer was used for the tagmentation, PCR amplification and PCR clean-up steps. The libraries were purified using AMPure XP (Beckman Coulter, Brea, CA) with a ratio of 0.5:1 beads to DNA (volume:volume).

The libraries were individually normalized by fluorometric quantitation (Qubit), and pooled in equimolar amounts. The combined libraries were quantified and fragment size was determined using TapeStation (Agilent). Libraries with an average fragment size of at least 600 bp (range 300–1000 bp) and concentration of 2000 pmol/l were considered to pass QC. The pooled library was loaded at 5.4 pM on one MiSeq flowcell with 1% phiX spiked in (MiSeq Reagent Kit v3). Paired end sequencing was performed with 300 cycles in each direction.

2.4. Computational HLA genotyping pipeline

For making HLA typing calls, a protocol similar to that of Wang et al. [11] was employed, with several modifications described herein. All scripts were developed in Python 2.7 for execution on a Linux cluster through the Torque submission system. The MiSeq-generated reads were first trimmed to remove all bases with a low quality call (Q<=28) as well as those downstream. For

Please cite this article in press as: McKinney DM et al. Development and validation of a sample sparing strategy for HLA typing utilizing next generation sequencing. Hum Immunol (2015), http://dx.doi.org/10.1016/j.humimm.2015.04.007