Chapter 5

In Phase HLA Genotyping by Next Generation Sequencing — A Comparison Between Two Massively Parallel Sequencing Bench-Top Systems, the Roche GS Junior and Ion Torrent PGM

Jerzy K. Kulski, Shingo Suzuki, Yuki Ozaki, Shigeki Mitsunaga, Hidetoshi Inoko and Takashi Shiina

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/57556

1. Introduction

Human Leukocyte Antigen (HLA) is the major histocompatibility complex in humans and it is critically involved in the rejection of hematopoietic stem cell or organ transplants [1-3] and in the pathogenesis of numerous autoimmune diseases [4]. Rejection and autoimmunity is believed to occur because the HLA presents aberrant histocompatibility antigen or public epitopes that play a key role in self-nonself recognition via various mechanisms including molecular mimicry and antibody mediated rejection [5-9]. The HLA genomic region on chromosome 6p21 encodes more than 200 genes including nine classical HLA genes, HLA-A, -B, -C in the class I region and HLA–DPA1, -DPB1, -DQA1, -DQB1, -DRA, -DRB1 in the class II region, that are the most polymorphic in the human genome contributing to over 7000 alleles and numerous HLA haplotypes implicated in disease resistance or susceptibility [10]. There are also a variety of non-classical HLA genes, such as HLA-DO and HLA-DM [11], HLA-E, -F, -G, MICA and MICB, which have received less attention in clinical medical research than the classical HLA genes [12,13].

The determination of the classical HLA alleles by DNA typing techniques over the past thirty years has assisted with the efficient and rapid HLA matching in transplantation therapy [2,14,15], research into autoimmunity and HLA related diseases [4,10], population diversity studies [16-18] and in forensic and paternity testing [19]. HLA gene alleles are also a target for

pharmacogenomics research into drug hypersensitivity [20], such as the association of HLA-B*57:01 with hypersensitivity to the antiviral agent abacavir [21] and the association of HLA-B*15:02 with anticonvulsants in Asian populations [20].

Many variations of the conventional PCR typing method have been used for HLA DNA typing, such as incorporating restriction fragment polymorphisms [22], single strand conformation polymorphism [23], sequence specific oligonucleotides (SSOs) [24], sequence specific primers (SSPs) [25] and sequence based typing (SBT), like the Sanger method [26]. The HLA DNA typing methods mainly applied today are PCR-SSO, such as the Luminex commercial methodology [27,28], and SBT by the Sanger method employing capillary sequencing of cloned chain-termination reactions [26,29]. However, the current, conventional DNA typing methods cannot readily distinguish between polymorphisms on the same chromosome (*cis*) or different chromosomes (*trans*), thereby creating ambiguities that are difficult, time-consuming and expensive to resolve. Because we cannot assign a single HLA allele to a particular chromosome when using these conventional methods in routine HLA-typing, we have to predict the most probable HLA allele assignment based on the information of allele frequency of a population or by computation using complex statistical likelihoods. Therefore, in most cases, the HLA gene haplotype that is the cluster of SNPs within a gene sequence inherited from a single parent is a statistical prediction rather than a proper empirical determination [30].

The next generation sequencing (NGS) technologies, which are also referred to as second-generation sequence technologies, are new sequencing developments [31-33] that have followed on from the first generation sequencing technologies of the Sanger-Coulson sequencing method using chain-termination dideoxynucleotide sequencing of single-stranded DNA [34] and chemical cleavage of double-stranded DNA using the Maxam-Gilbert method [35]. Recent advances in technology and cost effectiveness of NGS point the way for its implementation and wider use in HLA research and diagnostic settings with the generation of haplotype (in-phase) sequencing and a massive level of parallelism producing millions of sequencing reads for analysis [36-43]. There are a variety of commercial NGS platforms currently available for HLA gene amplicon resequencing, such as those from Roche-454 (454 GS-FLX-Titanium, GS Junior), Applied Biosystems-Agencourt (3730XL, SOLiD 3), Illumina-Solexa (Genome Analyzer GAII, MiSeq, HiSeq models), Life Technologies (Ion Torrent Proton with different high density sequencing Chips), HeliScope (Helicos) and PacBio RS II (PacBio); including three small benchtop sequencers, the Roche Genome Sequencer (GS) Junior, Ion Personal Genome Machine (PGM) sequencer and the Illumina MiSeq, that are much cheaper than their larger counterparts and that provide faster turnover rates, but a more limited data throughput [31,44-46].

Recently, the performances of the three commercially available benchtop sequencers were compared against each other directly by different investigators comparing the sequencing of bacterial genomes [45-47]. There were large differences obtained from the three platforms in cost, sequence capacity and in performance outcomes of genome depth, stability of coverage and read lengths and quality, due in part to their different sequencing methods. The Roche GS Junior was of intermediate price with the lowest sequencing capacity of >6 Mb per run and a run time of 8 h, but the highest cost per run because of expensive reagents. Its sequencing is

http://dx.doi.org/10.5772/57556

dependent on using the pyrosequencing technique, measuring the fluorescent light emitted when fluorophore-labeled dNTPs are added to sample DNA templates during the polymerase reaction [31,32]. The MiSeq was the most expensive instrument with the highest throughput at 1.5-2 Gb and a run time of 27 h. The running cost was 60 times cheaper than the Roche GS Junior. The MiSeq uses reversible terminator chemistry for sequencing in a cyclic method that involves fluorophore-labeled nucleotide incorporation, fluorescence imaging and cleavage [31,48]. The Ion PGM was the cheapest instrument with a throughput of 100 Mb using the 316 Chip, and 1 Gb using the 318 Chip, and a run time of 2-3 h. The running cost was 8 to 50 times cheaper than the Roche GS Junior depending on which Chip was purchased. The Ion PGM uses semi-conductor technology and ion-sensitive transistors to sequence DNA using only DNA polymerase and natural nucleotides in a sequencing-by-synthesis approach, with each polymerisation event resulting in pH and voltage changes identified by electronic sensors [33]. Therefore, the Ion PGM is the only "non-light" sequencer currently available in the market place.

Most pre-sequencing workflows for the benchtop sequencers and the other machines require DNA template fragmentation and library preparation where the fragments are labeled in vitro with oligonucleotide tags and adapters using commercial kits such as NimbleGen, Sure Select and other systems in order to be captured by beads or probes in preparation for clonal amplification of single stranded DNA fragments [31]. The labeling of DNA libraries with barcode sample tags, such as the multiplex identifier (MID) for Roche/454 sequencing, allows the libraries to be pooled to maximize the sequence output as a multiplex amplicon sequencing step for each sequencing run [49,50]. After construction of the template libraries, the DNA fragments are clonally amplified by emulsion PCR [51,52] or by solid phase PCR using primers attached to a solid surface [53] in order to generate sufficient single stranded DNA molecules and detectable signal for generating sequencing data [31]. Apart from selecting a suitable sequencing machine, NGS also provides challenges in analyzing and interpreting complex HLA genomic data from the millions of sequencing reads generated from the next-generation sequencers, which are different to those generated from traditional sequencers, and other HLA DNA typing methods and platforms.

We have described a new HLA DNA typing method, called Super high-resolution Single molecule - Sequence Based Typing (SS-SBT), that employs NGS and the Roche GS Junior [54] and the Ion PGM [55] massively parallel sequencing bench-top platforms (Figure 1). The SS-SBT method allows sequencing of the entire HLA gene region (promoter/enhancer, 5'UT, exons, introns, 3'UT) to detect new alleles and null alleles and solves the problem of phase ambiguity by using bioinformatics and computing for accurate phase alignments, after long-range PCR and sequencing clonally amplified single DNA molecules [42, 56].

This chapter updates our progress with the SS-SBT method [42] and describes some of the tasks required to identify the polymorphisms and other variants generated by NGS for SNP haplotyping from different classical HLA loci of individuals such as tissue donors and recipients in a relatively simple and economical way.

1.1. Aims of chapter

Here we describe and examine our latest modifications to the SS-SBT method for six-classical class I and class II HLA loci (HLA-A, -B, -C, -DPB1, -DQB1 and -DRB1) at a super-high resolution (formerly known as the 8-digit level of resolution) and three non classical class II HLA loci HLA-DRB3, -DRB4 and –DRB5 at high resolution using single DNA molecules to solve the ambiguity problem by undertaking:

1. Specific amplification of the entire gene sequence.
2. Comparisons between two different NGS platforms, the Roche GS Junior and Ion PGM.
3. Comparisons between different DNA typing software for in phase sequence analysis and validation of the huge amounts of NGS data.
4. Comparison of workflow differences in the construction of single gene and multiplex gene sequencing template libraries for 11 HLA class I and class II loci.
5. Application of the SS-SBT method including PCR multiplexing for the construction of template libraries for more efficient sequencing runs.

2. Materials and methods

2.1. HLA allele nomenclature and definition of super-high resolution

HLA genotyped alleles can be assigned at different levels of detail according to a recommended, standardized nomenclature [57]. Designations begin with HLA- as the prefix for an HLA gene and the locus name, followed by a separator * and then one or more two-digit numbers separated by colons (field separators) that specify the allele. The first two digits specify a group of alleles (first order, field one or a low resolution level). The third and fourth digits specify a non- synonymous allele (second order, field two or an intermediate resolution level). Digits five and six denote any synonymous mutations within the coding frame of the gene (third order, field three or a high resolution level). The seventh and eighth digits distinguish mutations outside the coding region (fourth order, field four or a super high resolution level). A ninth digit usually specifies an expression level or other non-genomic data and it is designated by a letter such as A (‘Aberrant’ expression), C (present in the 'Cytoplasm' but not on the cell surface), L (‘Low’ cell surface expression), N (‘Null’ alleles), Q (‘Questionable’ expression) or S (expressed as a soluble 'Secreted' molecule but is not present on the cell surface). Thus, a completely described allele may be up to 9 digits long. An example of an eight-digit or a super high resolution of an HLA-A allele that includes sequence variation within the introns or 5′/3′ extremities of the gene is HLA-A*01:01:01:01.

2.2. DNA samples

One hundred genomic DNA samples were obtained from Japanese subjects for this study on PCR amplification and NGS of HLA genes. We obtained written consent from each individual

http://dx.doi.org/10.5772/57556

and ethical approval from Tokai University School of Medicine where the research was performed. The DNA samples were isolated from the peripheral white blood cells using the QIAamp DNA Blood Mini Kit (QIAGEN, Germany). The HLA class I and class II gene alleles for nine samples (TU1 to TU8 and TU10) had been previously assigned to the field 2 allelic level of resolution (formerly known as 4-digit typing) [57] by the Luminex method [27,58] and LABType SSO kits (One Lambda, CA).

2.3. DNA extraction

Genomic DNA samples were isolated from the peripheral white blood cells using the QIAamp DNA Blood Mini Kit (Qiagen). The purity of the genomic DNA for each sample was determined by measuring the absorbance at 260 and 280 nm, with the A260/A280 values being in the range of 1.5-1.9, and the concentration of the DNA was adjusted to 10-20 ng/μl using PicoGreen assay (Life Technologies, CA).

Figure 1. The benchtop sequencers, Ion PGM from Ion Torrent and GS Junior from Roche/454

2.4. HLA DNA typing of genomic DNA samples by the Luminex method

The samples were typed at HLA-A, -B, -C, -DRB1, -DQB1 and -DPB1 using the Luminex method [27,28,58] and reagents supplied by the LABType SSO kits (One Lambda). An outline of the workflow for the Luminex method is shown in Figure 2.

2.5. HLA DNA typing of genomic DNA samples by SBT and the Sanger sequencing method

The DNA samples were typed at HLA-A, -B and -C (exons 2, 3 and 4) and -DRB1 (exon 2) using AlleleSEQR HLA-SBT Reagents (Abbott Laboratories, Abbott Park, IL). To confirm that the HLA alleles from both chromosomes were amplified with a 1:1 ratio and to validate newly discovered SNPs and indels, the nucleotide sequences of the PCR products were also directly sequenced by using the Sanger method and the ABI3130 genetic analyzer (Life Technologies, CA) in accordance with the protocol of the Big Dye terminator method (Life Technologies). The sequence-generated chromatogram data was analyzed by the Sequencher ver.4.10 DNA

sequence assembly software (Gene Code Co., MI). Sequence data were also analyzed using the assign-atf software (Conexio Genomics, Australia, [59]). An outline of the workflow for the Sanger sequencing method is shown in Figure 2.

2.6. SS-SBT by NGS

We used two different benchtop NGS platforms (Figure 1); massively parallel pyrosequencing with the Roche GS Junior Bench Top system [54] and massively parallel semiconductor sequencing with the Ion Personal Genome Machine (PGM) system [55]. A general review of the NGS workflows used for the Roche 450 system is presented by Margulies et al. [51], Metzger [31] and Erlich [41] and for the Ions PGM/Torrents system by Rothberg et al. [33].

For SS-SBT, we used a five-step workflow for HLA DNA genotyping and haplotyping for both the Roche GS Junior Bench Top system and the Ions PGM (Figure 2). These steps were LR-PCR, amplicon library construction, template preparation by emulsion (em) PCR, NGS run and HLA DNA data analysis. The first and last steps were essentially the same for both platforms, whereas steps 2 to 4 were platform specific. Two different sequencing runs, a single gene-sequencing run (SGSR) or a multiplex gene sequencing run (MGSR) were performed for the different HLA gene loci. SGSR was performed for only a single gene locus per run, whereas MGSR was performed at the same time for all of the HLA gene loci, whereby all of the LR-PCR amplicons were pooled together to construct the template libraries required for the sequencing platforms.

2.6.1. Step 1: LR-PCR

Single LR-PCR

We developed and used LR-PCR primers for eleven HLA loci, A, B, C, DRB1, DRB3, DRB4, DRB5 (DRB3/4/5), DQA1, DQB1, DPA1 and DPB1 [42, 56]. For amplification of the HLA-DRB1, group-specific primers were employed to prevent concomitant amplification of DRB1-like genes, such as DRB3, DRB4, DRB5 and DRB pseudogenes including DRB2, DRB6, DRB8 and DRB9. Others have reported different LR-PCR primer sets for some of the same HLA genes [40,43].

Pooling the LR-PCR amplicons

Amplicons from all class I and class II loci for each individual were pooled in equal amounts to yield 2-3 μg per pool of a single sample.

2.6.2. Steps 2 to 4: Amplicon library construction, emPCR and NGS

Both NGS platforms require the preparation of a fragmented template library of the PCR products and further clonal amplification of the templates by emPCR for sequencing of singled stranded DNA and detection of base-called signals (Figure 2).

http://dx.doi.org/10.5772/57556

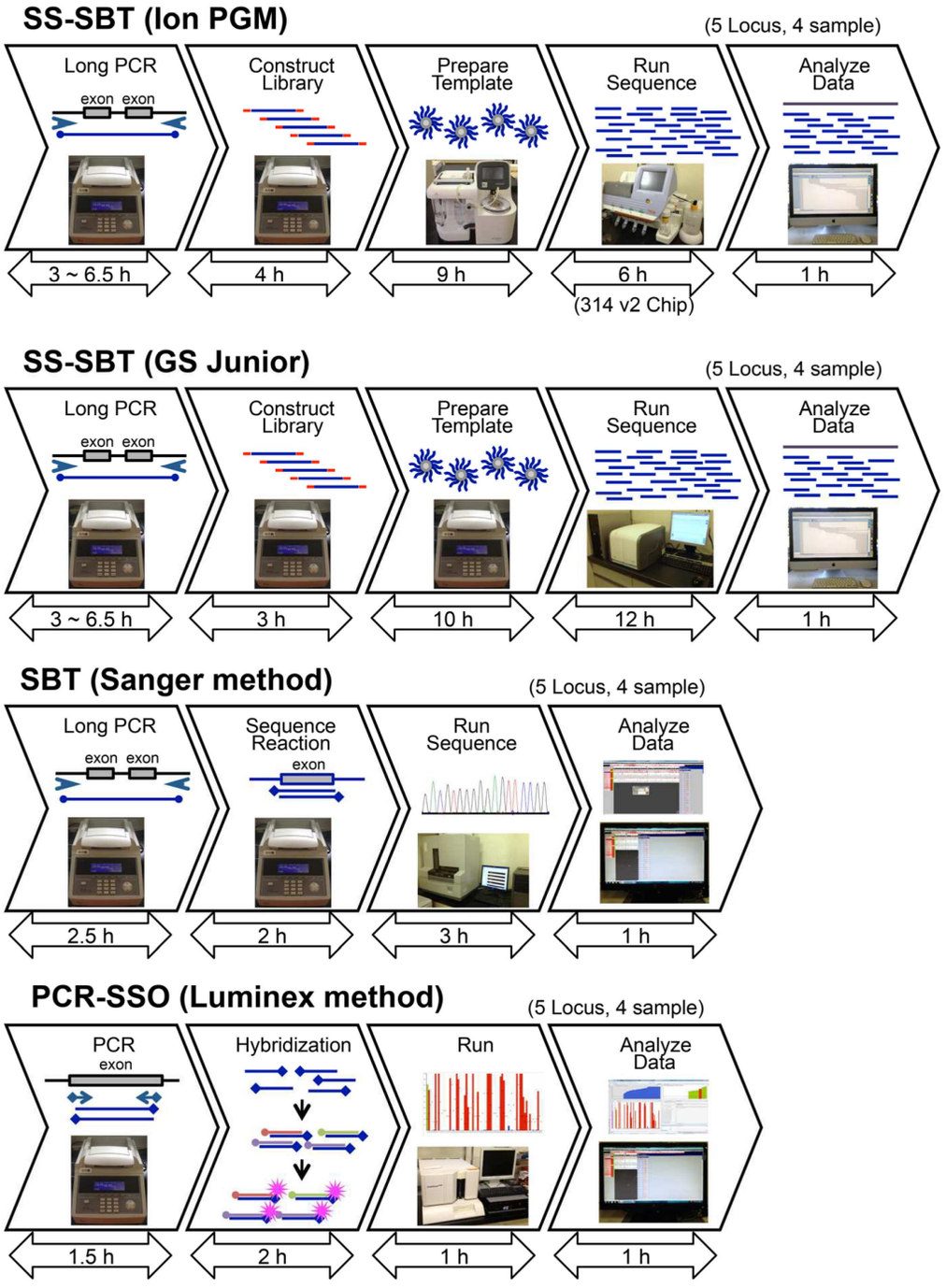


Figure 2. Workflow and time (hours) taken for each HLA genotyping step using the SS-SBT method by NGS with the Ion PGM or Roche GS Junior sequencers, by SBT using the Sanger sequencing method or by PCR-SSO using the Luminex method. The times are for typing five HLA loci using four DNA samples

2.6.2.1. Workflow for Roche GS Junior Bench Top system

Titanium libraries of single-stranded template DNA fragments were prepared for emPCR and sequencing using a GS FLX Titanium Rapid Library Preparation kit (cat no.

05608228001, 454 Life Sciences, Schweiz, CA) together with the GS Titanium Rapid Library MID Adaptors Kit (cat no. 05619211001, 454 Life Sciences) that permits the preparation of up to 12 individually MID labeled libraries per kit [60]. The workflow with the preparation kit involved fragmenting the LR-PCR amplicons by nebulization using disposable nebulisers, repairing fragment ends in a thermocycler reaction with T4 polymerase and Taq polymerase, adding adaptors such as the fusion A and B capture/sequence adapters and MID tags to the fragmented DNA by ligation with ligase and then linking the fragments of 100 bp or larger to the AMPure paramagnetic beads (Beckman Coulter Genomics, MA) with a selective binding buffer. Small fragments, primers and nucleotides were removed from the magnetic beads by washing them with 70% ethanol. Each single-strand DNA library was eluted from the beads with water or a Tris-acetate buffer, pH 8.

A MID adaptor was substituted for the adaptor provided in the GS FLX Titanium Rapid Library Preparation kit during the preparation of a library from a PCR DNA sample. The MID adaptor was ligated to each fragmented sequence of a library to provide a recognizable sequence tag at the beginning of each read. In this way, multiple libraries prepared with different MIDs can be sequenced together allowing the data analysis software to assign each read to the correct library and therefore a particular sample.

The libraries were quantified in 96-well format using the Fluoroskan Ascent CF Ver. 2.6 software (Thermo Fisher Scientific, DE) to detect the fluorescence in a PicoGreen assay (Life Technologies). The library size for each sample was measured by an Agilent 2100 Bioanalyzer using an Agilent High Sensitivity DNA Kit (Agilent Technologies, CA) and pooled with the other libraries at equimolar concentrations to create a multiplexed library.

The multiplexed libraries were clonally amplified by emPCR in order to produce hundreds of thousands or millions of copies of the same template sequence so that sufficient signal would be generated to be easily detected and recorded by the sequencing system. For preparation of emulsion beads, the single-stranded DNA fragmented libraries were added to the capture beads A and B, emulsified with oil and then amplified for at least 6 hours with an enzyme and reaction mix provided in the GS Junior Titanium emPCR Kit (Lib-L). An emulsion of PCR reagents in microreactors was prepared by adding beads A and B to the PCR reaction mix (1X amplification mix, Amplification Primers, 0.15U/ml Platinum Taq (Life Technologies), and emulsion oil and mixing vigorously using a Tissue Lyser (Qiagen, Germany).

The emulsion PCR of the library templates was carried out in an ABI 9700 (Life Technologies) thermocycler using the following cycling conditions: hotstart activation for 4 minutes at 94°C, 50 cycles of 94°C for 30 seconds, 58°C for 4.5 minutes, 68°C for 30 seconds. Sequencing primers were added to the mixture of beads and annealing buffer and annealed to the template using the following thermocycler conditions: 65°C for 5 minutes, and kept on ice for 2 minutes..

After emPCR, the emulsion was broken by isopropanol, the beads carrying the single-stranded DNA templates were enriched with primers A and B, counted, and 0.5 million

beads deposited along with enzymes and buffer into a single loading port of a GS Junior Titanium PicoTiterPlate (PTP) (Cat. No 05 996 619 001, 454 Life Sciences, Carlsbad, CA) to obtain sequence reads by pyrosequencing using the Roche GS Junior system. Packing beads, sample beads and enzyme beads were applied to the PTP as per manufacturer instructions. The GS Junior Titanium PicoTiterPlate permits only a single sequencing run per PTP. During sequencing, nucleotides flowed across the PicoTiterPlate in a fixed order and a cooled CCD camera recorded the amount of light generated by the pyrosequencing reactions. The images from the CCD camera were then converted to sequence data by the instrument's software. Bi-directional sequencing for the fragments was achieved because the single strands captured by the A and B beads allow both forward and reverse reads to be identified.

After the DNA sequencing run by the Roche GS Junior Bench Top system, data analysis such as image processing, signal correction and base-calling, binning, trimming and mapping of the sequence reads and assignment of HLA alleles were carried out in accordance with our previous publication [42]. Image processing, signal correction and base-calling were performed by the GS Run Processor Ver. 2.5 (454 Life Sciences) with full processing for shotgun or filter analysis. Quality-filter sequence reads that were passed by the assembler software (single sff file) were binned on the basis of the MID labels into 10 separate sequence sff files using the sfffile software (454 Life Sciences). These files were further quality trimmed to remove poor sequence at the end of the reads with quality values (QVs) of less than 20.

2.6.2.2. Workflow for Ion PGM system

Barcoded-library DNAs were prepared from PCR amplicons with an Ion Xpress Plus Fragment Library Kit and Ion Xpress Barcode Adaptors 1-16 Kit according to the manufacture's protocol for 200 base-read sequencing (Life Technologies). One hundred nanograms of the HLA pooled amplicon products from four individuals were used for the preparation of each DNA library. Each DNA library was clonally amplified by eight cycles of PCR. The DNA size for each library was measured by an Agilent 2100 Expert Bioanalyzer using the Agilent High Sensitivity DNA Kit (Agilent Technologies) and the concentration of each library was measured with the Ion Library Quantitation Kit using the 7500 Real-Time PCR System (Life Technologies). Each barcoded-library was mixed at equimolar concentrations then diluted according to the manufacture's recommendation. emPCR was performed using the resulting multiplexed library with the Ion Xpress Template 200 Kit on an ABI 9700 (Life Technologies) with the following cycling parameters: primary denaturation 94°C/6 min, followed by 40 cycles for 94°C/30 s, 58°C/30 s and 72°C/90 s, and 10 cycles for 94°C/30 s and 68°C/6 min. After the emulsion was broken with 1-butanol, a magnetic-bead-based process according to the manufacture's recommendation enriched the beads carrying the single-stranded DNA templates. Sequencing was performed using the Ion Sequencing 200 Kit and Ion 316 chip with a flow number of 440 for 200 base-read [61].

Ion Torrent uses a pH change (release and detection of a proton) to detect the incorporation of a base into the growing DNA strand rather than a flash of light, as used in the 454 Sequencing System [33].

In the case of the Ion PGM DNA sequencing system, the raw data processing and base-calling, trimming and output of quality-filter reads that were binned on the basis of the Ion Xpress Barcodes into 10 separate sequence sff files were all performed by the Torrent Suite 1.5.1 (Life Technologies) with full processing for shotgun analysis. These files were further quality trimmed to remove poor sequence at the end of the reads with QVs of less than 10.

2.6.2.3. Sequencing parameters, variables and definitions

The NGS methodology generates a number of sequencing parameters, variables and errors that need to be defined and noted.

A sequencing read is a contiguous length of nucleotide bases that is generated using a sequencing machine. The full set of aligned reads (coverage) reveals the entire genomic sequence in the DNA sample selected for sequencing.

Both the NGS methods described here are designed to generate hundreds of thousands to millions of short (100 to 800 bp) contiguous reads of low to high quality. Intrinsic quality measures or filters are automatically built into the sequencing analysis software to recognize and measure or statistically predict quality values (QV or Q) and remove incorrect repetitive sequence, indels, low-quality sequence at the beginning and end of runs (end clipping) and recognize accurate consensus sequences [62,63].

Generally, the QVs or Q's are based on estimated Phred quality scores [64] so that a quality score of 10 represents a 1 in 10 probability of an incorrect base call or 90% base call accuracy, 20 represents a 1 in 100 probability of an incorrect base call or 99% base call accuracy, 30 represents a 1 in 1000 probability of an incorrect base call or 99.9% base call accuracy, 40 represents a 1 in 10000 probability of an incorrect base call or 99.99% base call accuracy, and so on.

QV scores are clearly valuable because they can reveal how much of the data from a given run is usable in a resequencing or assembly experiment. Sequencing data with lower quality scores can result in a significant portion of the reads being unusable, resulting in wasted time and expense. A QV of ≥ 20 is usually considered a satisfactory cut-off score for base calling accuracy. The QV and Q reads from the GS Junior and the Ion PGM are slightly different from each other and not directly comparable, with some studies showing the GS Junior overestimates the base-calling accuracy while the Ion PGM underestimates the base-call accuracy [45,65-67].

Read depth is the number of individual sequence reads that align to a particular nucleotide position [40]. The average read depth (coverage, redundancy or depth) is the number of average reads representing or covering a given nucleotide in the reconstructed sequence, often calculated as N x L/G where G is the original length of the genome sequence, N is the number of reads and L is the average read length. Thus, for a 1000 bp sequence G reconstructed from 10 reads N with and average length of 200 nucleotides per read L, there is a 2x coverage or redundancy. However, reads are not distributed evenly over an entire genome and many bases

will be covered by fewer reads than the average coverage, while other bases will be covered by more reads than average [68].

Another important variable to consider, particularly for HLA typing with NGS is the allelic balance [40] or the ratio of the average depth [42], which is the relative number of reads originating from each of the two alleles of each locus. This variable is important to monitor in order to prevent allele drop-out. Allelic imbalance is usually not a sequencing problem, but most often is due to poor genomic PCR primer design where a set of primers favours the amplification of one haplotype over the other.

Sequence accuracy to estimate the quality of a sequencing run was estimated by including internal control DNA beads of known sequence with each run.

2.6.3. Step 5: DNA sequence data analysis and HLA assignments

The trimmed and MID or barcoded binned sequences were mapped as perfectly matched parameters using the GS Reference Mapper Ver. 2.5 (Life Technologies). Reference sequences used for mapping of the sequence reads were selected by nucleotide similarity searches with HLA allele sequences in the IMGT-HLA database using the BLAT program [69]. If a reference sequence covering the entire gene region was not available, we converted the sff files to ace files and constructed a new virtual sequence by *de novo* assembly using the sequencher Ver. 4.10 DNA sequence assembly software (Gene Code Co, Ann Arbor, MI), and used it as a reference sequence.

Three different automated NGS data processing systems, the Sequence Alignment Based Assigning Software (SEABASS) (an in house development of Tokai University, Isehara), Omixon Target (Omixon, [70]) and Assign MPS v1 (Conexio, [59]) were also assessed and compared for in phase sequence alignment of HLA genes and for allele assignment at the 8-digit level. The three programs were designed to provide software or a suite of software tools for analyzing targeted sequencing data from the next generation sequencing platforms, Roche 454, Ion Torrent and Illumina. The HLA edition of the Omixon Target can be obtained for use with MacOSX or Windows as a credit-based or annual based license fee. Conexio software for Windows can be purchased outright, whereas SEABASS is an in house development for Linux and is not available commercially at this time.

3. Results

3.1. HLA typing by Sanger SBT, the Luminex method or SS-SBT by NGS

Ten genomic samples (TU1-TU10) were genotyped at six HLA loci (HLA-A, -B, -C, -DRB1, -DQB1 and –DPB1) by three different typing methods, the Luminex, the Sanger SBT and the NGS SS-SBT method. At least one or more pairs of unresolved alleles were detected in the samples by the SBT and/or Luminex methods, whereas all the samples were easily resolved at least to the 4-digit level by the SS-SBT typing method (see the Tables and Supplementary Tables in the paper by Shiina et al [42]). Table 1 shows an example of the DNA typing results obtained for the HLA-C locus by the Luminex, SBT or SS-SBT methods as previously reported.

HLA-C typing by Luminex method				
DNA Sample ID	Allele 1	Allele 2	Number of ambiguities	
			Allele 1	Allele 2
TU1	C*01:02/07/11/+	C*03:03/11/12/+	18	17
TU2	C*01:02/07/11/+	C*03:04/06/09/+	18	23
TU3	C*03:03/04/06/+	C*07:02/10/13/+	40	29
TU4	C*03:03/04/06/+	C*14:03/10	41	2
TU5	C*03:02/04/05/+	C*07:02/10/13/+	40	29
TU6	C*03:03/04/06/+	C*14:03/10	41	2
TU7	C*01:02/07/11/+	C*07:02/32/38/+	17	25
TU8	C*08:03/06/14	**C*14:03**	3	0
TU9	C*08:01/08/20/+	C*15:02/03/10/+	7	5
TU10	C*07:02/10/19/+	C*15:02/03/07/+	27	7
HLA-C typing by SBT method				
	Allele 1	**Allele 2**	**Ambiguities (Allele 1 and Allele 2)**	
TU1	C*01:02/22/51	C*03:03/04/20/+	9	
TU2	C*01:02/22/31/+	C*03:04/28/64/+	8	
TU3	C*03:03/04/13/+	C*07:02/10/50/+	18	
TU4	C*03:03/04/20	C*14:03/10	4	
TU5	C*03:04/32/35/+	C*07:02/10/29/+	15	
TU6	C*03:03/04/20	C*14:03/10	4	
TU7	C*01:02/17/22	C*07:02/37/39/+	10	
TU8	**C*08:03:01**	**C*14:03**	0	
TU9	C*08:01/10/22	C*15:02/17	3	
TU10	C*07:02/19/39/+	C*15:02/03/07/+	8	
HLA-C typing by SS-SBT using Roche GS Junior				
	Allele 1	**Allele 2**	**Ambiguities**	
TU1	C*01:02:01	C*03:03:01	0	
TU2	C*01:02:01	C*03:04:01:02	0	
TU3	C*03:03:01	C*07:02:01:03	0	
TU4	C*03:03:01	C*14:03	0	
TU5	C*03:04:01:02	**C*07:02:01:(04)**	0	
TU6	C*03:03:01	C*14:03	0	
TU7	C*01:02:01	**C*07:02:01:(05)**	0	
TU8	C*08:03:01	C*14:03	0	
TU9	C*08:01:01	C*15:02:01	0	
TU10	C*07:02:01:01	C*15:02:01	0	

Table 1. Results of HLA DNA typing for the HLA-C locus by the Luminex, SBT or SS-SBT methods. The / is possible ambiguity, + is more than the possible ambiguities indicated by /, Parenthesis and bold letters indicate tentative allele names, not yet officially approved by the WHO Nomenclature Committee.

http://dx.doi.org/10.5772/57556

Figure 3 shows how the SS-SBT NGS typing method resolves the phase ambiguity at the HLA-DRB1 locus of a sample that the Luminex beads and other conventional typing methods were not able to resolve. Essentially, the Luminex bead method analyzed the mixture of the PCR products amplified from the paternal and maternal chromosomes and could not distinguish between HLA-DRB1*15:01/DRB1*04:05 on Sample "A" and HLA-DRB1*15:02/DRB1*04:10 on Sample "B", and therefore, resulting in ambiguity. On the other hand, in the SS-SBT method the sequenced PCR products were each amplified from a single DNA molecule (clonal amplification method by emulsion PCR), which helps to determine whether each polymorphism is paternal or maternal, thus resolving phase ambiguity. In the case shown in Figure 3, the sample was typed as DRB1*15:02/DRB1*04:10 on Sample "B".

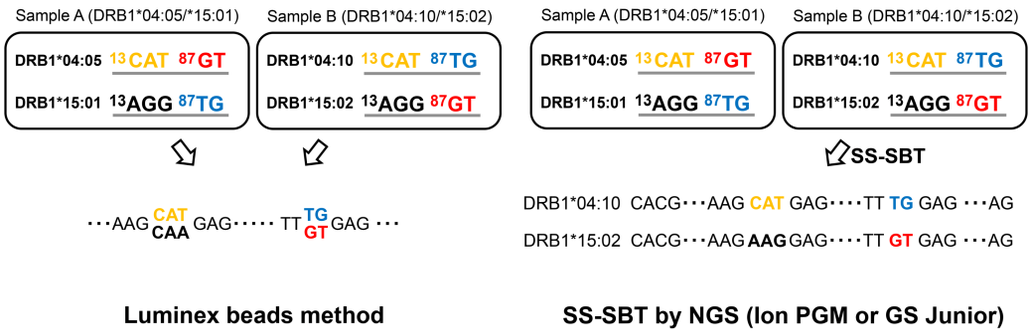


Figure 3. SS-SBT resolves phase ambiguity, which is an inherent problem of the Luminex beads method and other conventional methods

3.2. HLA typing by SS-SBT using two different NGS benchtop platforms

3.2.1. LR-PCR

3.2.1.1. Single LR-PCR

Long-range PCR (LR-PCR) was developed to amplify eleven HLA genes (HLA-A, -C, -B, -DRB1, -DQA1, -DQB1, -DPA1, -DPB1, -DRB3, -DRB4 and –DRB5) that are known to be highly polymorphic. PCR primers were designed to avoid annealing to any known polymorphic sites and to amplify regions spanning from the 5′ promoter to the 3′UTR region of the HLA genes in either a single amplification reaction or as two separately divided amplifications that could be easily merged when sequenced.

The PCR primer sets for the HLA-A, -B and -C genes were designed to amplify their sequences from the 5′-enhancer-promoter region to the 3′UTR with an amplicon size of about 5 kb. PCR primer sets were designed for HLA-DQA1, -DQB1, and -DPA1 to amplify sequences from the 5′- enhancer-promoter region to the 3′UTR with amplicon sizes of about 7 to 10 kb bases. Because the gene sizes for HLA-DRB1 and -DPB1 were too long to successfully amplify the whole gene in a single reaction, we divided the amplification of these genes into two parts. One PCR primer pair was designed to amplify the 5′-enhancer-promoter region to exon 2 and

the other primer pair was designed to amplify exon 2 to the 3'UTR, resulting in amplicon sizes of about 6 to 11 kb and 5 to 7 kb, respectively (Figure 4). The DRB1 locus produced different amplicon sizes because of allelic differences in the length of its introns.

LR-PCR methods were also designed and tested to amplify HLA-DRB3, -DRB4 and -DRB5 (*DRB3/4/5)* from intron 1 to exon 6 (3'UTR) with the product size of 5.6 kb for *DRB3*, 5.1 kb for *DRB4* and 4.7 kb for *DRB5*. The *DRB3/4/5* specific primer sets successfully amplified the HLA-DRB3, -DRB4 and -DRB5 genes from 19 positive genomic DNA samples (TU1 to TU8, TU10 to TU17 and TU20 to TU22) with PCR products varying in size between 4.7 kb to 5.6 kb (Figure 4, [56]).

Although some LR-PCR methods are known to produce low frequency extra amplification of pseudogenes, as previously reported [43], our LR-PCR methods generally produced specific amplicons of targeted genes with little or no evidence of extra sequences (Figure 4). In addition, allelic imbalance for all the LR-PCR methods was minimal as judged by the sequencing results with allelic ratios ranging on average between 0.6 and 1.6 in heterozygous samples.

3.2.1.2. Pooling LR-PCR products for library construction

In order to expedite the number of samples and HLA loci required for NGS, the single LR-PCR products were pooled in equimolar amounts to produce a single multiplex (up to 13 PCR products) for the construction of a multiplexed, barcoded sample library. Figure 5 shows an example of an electrophoretogram of the pooled PCR amplicons from the HLA-A, -B, -C and -DRB1 genes.

3.2.2. Roche GS Junior NGS system

3.2.2.1. Library preparations using different gene loci and sample numbers

Initially, we applied an individual tagging per gene amplicon method for SGSR by preparing a tagged library for each of the ten individual samples with the aim of sequencing one HLA locus at a time [42]. Thus, ten libraries were constructed from each PCR amplicon of a particular HLA locus, with each library representing an HLA gene with ten tagged individual samples. The ten individual MID-labelled ssDNA libraries were then combined into a single tube for clonal emPCR and sequencing. In this way, we performed a single sequencing run for each HLA locus represented by ten MID labelled individuals, one locus at a time. For example, Run 1 was HLA-A with samples one to ten; Run 2 was HLA-B with samples one to ten, Run 3 was HLA-B with samples one to ten, and so on until we finished Run 13, our last run, Run 13, representing the 13th LR-PCR amplicon we had amplified from the 11 HLA gene loci that we had chosen to analyze.

Later, we changed from SGSR to an individual tagging and gene-pooling method for MGSR in order to markedly reduce the number of required sequencing runs. In this approach, we first pooled the PCR amplicons obtained from the eleven separate HLA genes (HLA-A, -B, -C, -DQA1, -DQB1, -DPA1, -DRB1 (part a and b), -DPB1 (part a and b), -DRB3, -DRB4

http://dx.doi.org/10.5772/57556

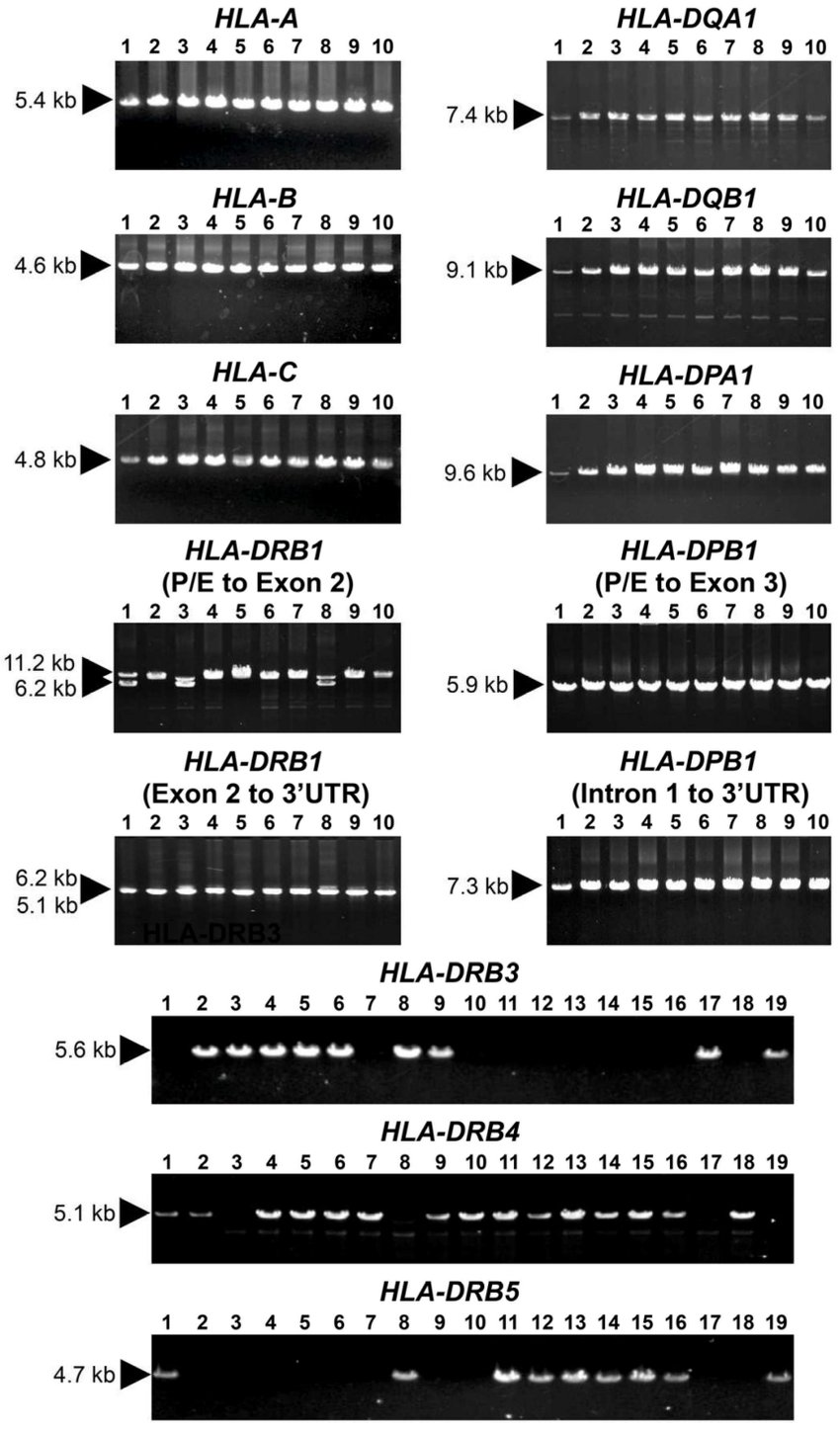


Figure 4. Agarose gel electrophoresis of LR-PCR products obtained from 10 to 19 unrelated genomic DNA samples using locus-specific primers for 11 HLA genes. The individual LR-PCR amplicons were used for SGSR or were combined into a single pooled mix for MGSR. Figure adapted from [42] and [56]

and -DRB5 of each individual at equimolar concentrations and then prepared the MID-labeled NGS libraries using the pooled amplicons from each individual. In this way, a single

sequencing run was performed for all eleven HLA loci and for a multiple number of MID-labelled individuals or DNA samples. This was possible because the sequencing adapters specifically identified and grouped the gene loci, whereas the MID tags identified the individual samples. Figure 5 shows an electrophoretogram of pooled LR-PCR amplicons amplified from samples with known DR sub-types in parenthesis such that lane 1 is a heterozygous sample of DR6 and DR9, lane 2 is heterozygote of DR2 and DR6, lane 3 is a heterozygote of DR2 and DR4, lane 4 is a heterozygote of DR3 and DR8 and lane 5 is a heterozygote of DR1 and DR7. Consequently, the PCR product from the DRB1 gene varies in size depending on the DR sub-type and associated HLA-DRB1 allele so that it is 5.1 kb (orange) and 5.2 kb (blue) in lane 1, 5.2 kb (blue) and 5.6 kb (green) in lane 2, 5.6 kb (green) and 6.2 kb (purple) in lane 3, 5.2 kb (blue) in lane 4 and 5.1 kb (orange) and 5.2 kb (blue) in lane 5.

3.2.2.2. Sequencing multiplex libraries prepared for SGSR

Table 2 shows the number of draft sequences or reads (sequences passing a quality control criteria after base calling), generated by the Roche GS Junior Bench Top system for HLA-C in ten samples using multiplex libraries prepared for SGSR. The results for the sequence information about HLA-C were reported previously by Shiina et al [42] and Ozaki et al [56] as part of the data obtained in their NGS analysis of the eleven HLA loci.

DNA Sample ID	MID type	Draft Read Numbers	Draft Read Bases (bp)	Avg Read Length (bp)	Edited Reads (% of Draft Reads)	No. Mapping Bases (bp) post edit
TU1	MID1	7,873	3,213,949	408.2	5,075 (64.5%)	1,640,715
TU2	MID2	17,926	7,472,332	416.8	12,002 (67.0%)	4,003,792
TU3	MID3	8,211	3,326,703	405.2	5,553 (67.6%)	1,678,658
TU4	MID4	18,617	7,782,001	422.5	12,600 (67.7%)	4,223,267
TU5	MID5	12,432	5,185,865	417.1	8,318 (66.9%)	2,617,187
TU6	MID6	8,292	3,468,588	418.3	5,784 (69.8%)	1,982,456
TU7	MID7	10,900	4,503,252	413.1	7,042 (64.6%)	2,181,140
TU8	MID8	11,507	4,793,240	416.5	7,801 (67.8%)	2,554,949
TU9	MID9	12,082	4,995,156	413.4	8,101 (67.1%)	2,709,149
TU10	MID10	17,674	7,415,292	419.6	12,929 (73.2%)	3,968,608
Total		125,514	52,156,378	416.2	85,205 (67.9%)	27,559,921

Table 2. Sequence read information obtained by the Roche GS Junior for HLA-C.

Draft read numbers for eight genes, HLA-A, -C, -B, -DRB1, -DQA1, -DQB1, -DPA1 and -DPB1, were 1,015,526 sequence reads in total for 10 samples per locus with a range of reads between 49,048 in HLA-DQA1 and 153,610 sequence reads in exon 2 to 3'UTR of HLA-DRB1 [42]. These were high quality reads with QV ≥ 20 and an average QV of 30.4 ranging between 28.6 and 32.4. The draft read bases for eight genes, HLA-A, -C, -B, -DRB1, -DQA1, -DQB1, -DPA1, -

http://dx.doi.org/10.5772/57556

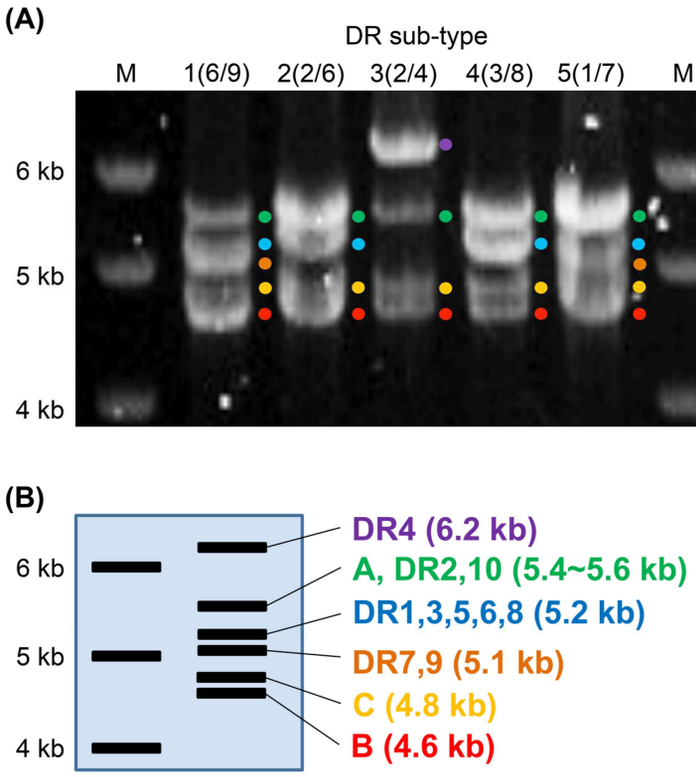


Figure 5. (A) Agarose gel electrophoresis of pooled LR-PCR amplicons of HLA-A (green), -B (red), -C (yellow) and -DRB1 (purple, green, blue and orange) used for MGRS. The lanes labeled M are bands of the 1 kb DNA size marker ladder. (B) Schematic of the expected PCR product band patterns and sizes seen in the PCR multiplex amplified from samples with different DR subtypes

DPB1, were 412,956,301 bp in total with a range between 18,077,833 bp in HLA-DQA1 and 65,384,530 bp in exon 2 to 3'UTR of HLA-DRB1 [42]. There was an overall average sequence length of 414.2 bp, ranging from 321.8 bp in intron 1 to 3'UTR of HLA-DPB1 to 441bp in HLA-DPA1. Of the draft reads, excluding error reads such as artificial insertion or deletion (indel) and nucleotide substitutions, 49.7%–75.4% matched perfectly with HLA allele sequences released from the IMGT-HLA database or our newly constructed virtual reference sequences. Therefore, the sequence reads had high quality and sufficient sequence volume for further HLA DNA genotyping and haplotyping analysis.

The HLA-DRB3, -DRB4 and -DRB5 (DRB3/4/5) gene loci were each sequenced separately by the Roche GS Junior Bench Top system using 19 samples and two separate multiplex libraries. After the sequencing run for each gene, a total of 147,269 sequence reads were provided for mapping and allele assignment of DRB3/4/5. The sequence reads were high quality reads (QV≥20) with an average QV of 30.9 ranging between 28.6 and 32.4. The draft read bases in total were 62.3 Mb in DRB3/4/5, with an overall average sequence length of 423.2 bp, ranging from 361 bp to 445.6 bp. This data suggested that the sequence reads had sufficient high quality and sequence volume for further HLA DNA genotyping and haplotyping analysis.

3.2.2.3. HLA SS-SBT DNA typing

HLA typing was applied to 10 genomic DNA samples (TU1-TU10) all of which gave more than one pair of unresolved alleles when defined by the SBT and/or Luminex methods (Table 1 and Table 3). In comparison, the pyrosequencing reads using Roche GS Junior system matched perfectly with the HLA allele sequences previously deposited in the IMGT-HLA database or against the virtual allele sequences constructed by *de novo* assembly as reference sequences. HLA allele sequences were determined at the 8-digit level in both phases in all samples of HLA-C (Table 1) and –DRB1 (Table 3) as well as HLA-A, -B, and -DQB1. Ambiguous alleles were resolved by in phase (haplotype) sequencing through the heterozygous positions thereby eliminating any doubt about possible alternative allele combinations (Figure 3). Cloning and Sanger sequencing validated the accuracy of the sequencing results by NGS (data not shown).

DNA sample ID	Sanger-SBT	
	Allele 1	**Allele 2**
TU1	DRB1*09:01/06	DRB1*15:01/02
TU2	DRB1*09:01/06	DRB1*14:05/44
TU3	DRB1*01:01/17/20	DRB1*14:05/44/100
TU4	DRB1*04:10/11	DRB1*14:01/32/54
TU5	DRB1*09:01:02	DRB1*13:02:01
TU6	DRB1*04:05/28/43/+	DRB1*13:02/36/65/+
TU7	DRB1*04:03/07/11/+	DRB1*08:03/10/29/+
TU8	DRB1*13:02/29/36	DRB1*16:02/05/10
TU9	DRB1*14:05:01	-
TU10	DRB1*09:01:02	DRB1*12:01/06/10/+
DNA sample ID	**PCR-SSOP**	
	Allele 1	**Allele 2**
TU1	DRB1*09:01/04/05	DRB1*15:01/13/16/+
TU2	DRB1*09:01/04/05/+	DRB1*14:05/23/43/+
TU3	DRB1*01:01/05/07/+	DRB1*14:05/23/45/+
TU4	DRB1*04:10	DRB1*14:01/26/54/+
TU5	DRB1*09:01/04/05	DRB1*13:02/96
TU6	DRB1*04:05/29/45/+	DRB1*13:02/73/96
TU7	DRB1*04:03/39/41/+	DRB1*08:03/23/27/+
TU8	DRB1*13:02/73/96	DRB1*16:02
TU9	DRB1*14:05/23/43/+	-
TU10	DRB1*09:01/04/05/+	DRB1*12:01/06/08/+
DNA sample ID	**SS-SBT**	
	Allele 1	**Allele 2**
TU1	DRB1*09:01:02:(01)	DRB1*15:01:01:01/02/03
TU2	DRB1*09:01:02:(02)	DRB1*14:05:01:(02)

http://dx.doi.org/10.5772/57556

TU3	DRB1*01:01:01	DRB1*14:05:01:(02)
TU4	DRB1*04:10:(03):(01)	DRB1*14:54:01:(02)
TU5	DRB1*09:01:02:(01)	DRB1*13:02:01:(02)
TU6	DRB1*04:05:01:(01)	DRB1*13:02:01:(02)
TU7	DRB1*04:03:01:(02)	DRB1*08:03:02:(02)
TU8	DRB1*13:02:01:(02)	DRB1*16:02:01:(02)
TU9	DRB1*14:05:01:(02)	-
TU10	DRB1*09:01:02:(01)	DRB1*12:01:01:(02)

Table 3. Results of HLA DNA typing for the HLA-DRB1 locus by the Sanger SBT, PCR-SSOP and NGS SS-SBT methods. The / is possible ambiguity, + is more than the possible ambiguities indicated by /, parenthesis and bold letters indicate tentative allele names, not yet officially approved by the WHO Nomenclature Committee. Blue background indicates HLA alleles with ambiguous allele.

An average sequencing depth was 387.9 in total, ranging between a depth of 53.5 in TU7 of HLA-DQB1 and 924.0 in TU4 of HLA-C. An example of the minimum and maximum depth and ratio of depth between two allele sequence reads for HLA-C in five DNA samples using the Roche GS Junior is shown in Table 4. The average depth ratio between haplotypes or alleles of HLA-A, -B, -C, -DRB1 (mix 2), -DQB1 ranged from 0.6 to 1.6, suggesting a satisfactory allelic balance was achieved with the PCR reactions. In comparison, the ratio of depth between two alleles in the HLA-DRB1 (mix 1) sequences was more variable exceeding a ratio delta of 1.0, which may be due to the marked difference in the sizes of PCR products among DRB1 groups (6-11 kb). However, this ratio of depth between two alleles did not affect sequence determination from the enhancer–promoter region to the exon 2 region of the DRB1 gene when the number of draft reads was increased.

			Depth			
Allele	**Sample ID**	**Read Num.**	**Min**	**Max**	**Ave**	**Ratio***
	TU1	4,766	87	628	339.9	0.9
	TU2	11,296	246	1,433	847.4	1.0
Allele 1	TU3	3,648	91	595	290.9	1.1
	TU4	12,232	245	1,446	924.0	1.1
	TU5	5,339	171	843	433.5	1.0
	TU1	4,876	106	634	371.1	0.9
	TU2	11,465	234	1,425	883.5	1.0
Allele 2	TU3	3,392	85	616	262.6	1.1
	TU4	12,103	230	1,447	870.6	1.1
	TU5	5,265	156	868	427.5	1.0

Table 4. Minimum and maximum depth and ratio of depth between two-allele sequence reads for HLA-C from five DNA samples using the Roche GS Junior. *Ratio is the average depth of allele 1/average depth of allele 2.

Overall, in this analysis, 21 HLA allele sequences were newly determined, 17 of them were newly identified alleles and four were alleles extended from 2-digit to 8-digit level sequences. Most of the newly identified SNPs and/or indels for 16 alleles were observed in the introns, whereas a synonymous SNP was identified in one allele of DRB1*04:10:(03):(01)) (parenthesis indicates tentative allele name, not yet officially approved by the WHO Nomenclature Committee). A comparison of the alleles detected for HLA-DRB1 by Sanger-SBT and SS-SBT in ten samples is shown in Table 3.

The HLA-DQA1, -DPA1 and -DPB1 alleles were assigned only at the 6-digit level of exon 2 with no novel alleles discovered in the 10 genomic DNA samples. This lower level assignment was likely due in part to low SNP densities preventing the haplotype genomic segments to be properly aligned and the phases separated from each other.

DRB3/4/5 sequences were identified to the field 4 level in both phases and validated by Sanger sequencing [56]. An average depth of DRB3/4/5 was 277 in total (107 in TU1 to 561.7 in TU11. The number of newly determined allele sequences at the field 4 level of resolution was two for DRB3 (DRB3*01:01:02:(02) and DRB3*02:02:01:(03)), three for DRB4 (DRB4*01:03:01:(04), DRB4*01:03:01:(05) and DRB4*01:03:01:(06)) and one for DRB5 (DRB5*01:01:01:(02)) (the allele names at field 4 are tentative as they have not yet been officially approved by the WHO Nomenclature Committee). Six allele sequences were extended from a field 2 or 3 level to a field 4 level with one for DRB3 (DRB3*03:01:01:(01)), one for DRB4 (DRB4*01:03:02:(01)) and two for DRB5 (DRB5*01:02:(01):(01) and DRB5*02:02:(01):(01)). Field 4 level haplotype structure of DRB1-DRB3/4/5 were also determined by SS-SBT and reported by [56].

3.2.2.4. Sequencing multiplex LR-PCR generated libraries prepared for MGSR

Using the libraries prepared from pooled LR-PCR for MGSRs (Figure 5), we sequenced 81 Japanese and Caucasian genomic samples by the Roche GS Junior instrument and assigned the aligned alleles across the entire gene regions of the eleven HLA gene loci without any ambiguities at least to the field 3 assignment level. Of the 164 alleles detected at the field 4 assignment level, 78 (47.6%) were newly detected alleles. We were unable to determine the allele sequences of HLA-DQA1, -DPA1 and -DPB1 at field 4 due to a lack of tag SNPs available to identify and separate the two phases within some of the noncoding regions. However, all the HLA exonic alleles were assigned without ambiguity at the field 3 level.

3.2.3. Sequence read information from the Ion PGM system

The amplicons amplified from eight HLA loci (A, B, C, DRB1, DQA1, DQB1, DPA1 and DPB1), representing a total of 60 kb of the entire gene regions (from the enhancer-promoter region to the 3'UT region), were prepared and pooled for the construction of barcoded DNA libraries from four DNA samples (TU1, TU3, TU5 and TU8) that had been previously sequenced by the Roche Junior system (Table 1). The four-barcoded samples were sequenced as a multiplex in a single run on an Ion 316™ chip. Figure 6 shows the workflow of the SS-SBT method performed by the Ion PGM™ sequencer that completed the processes from sample preparation to sequence analysis within three days. Previously, four days were needed for the manual

http://dx.doi.org/10.5772/57556

sample preparation and library construction from the time of PCR amplification to HLA allele definition [42]. Now, with the use of automated procedures, such as the AB Library Builder™ system and Ion OneTouch™ system from Life Technologies, we have shortened the workflow significantly to just two to three days.

The number of reads, number of bases, and median read-length of each sample are shown in Table 5. A total of 522 Mb of sequence data was produced (about 600K to 720K reads per sample) with a range between 123.3 Mb for TU1 and 141.6 Mb for TU5, with an overall average read length of 197.3 bp. The median read length was about 214 bp, which was sufficient to resolve the phase ambiguity in all of the samples that were tested. Draft read numbers in total were 2,646,446 reads with a range of reads from 605,501 for TU1 to 721,499 reads for TU5 that were high quality reads with QV ≥ 10 at an average QV of 19.2 in the high quality reads.

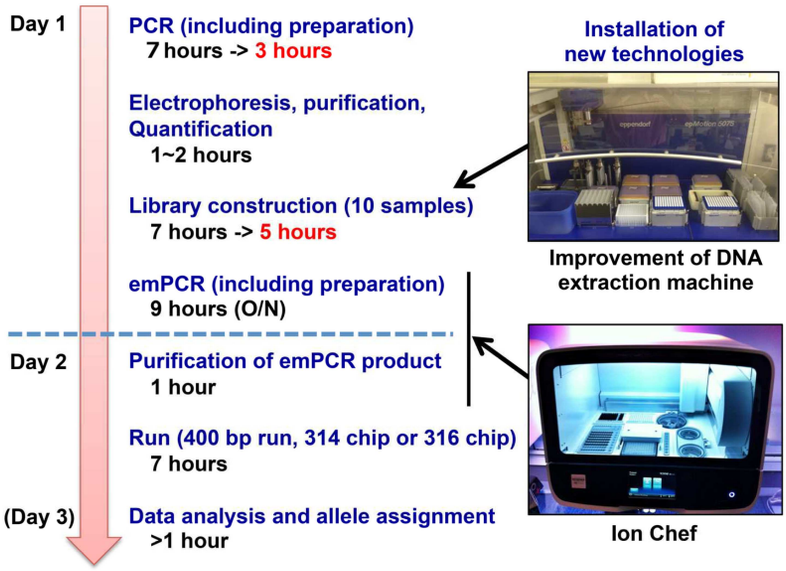


Figure 6. Improved workflow and time reduction to simplify the SS-SBT method using the Ion PGM

Sample ID	Number of Reads	Number of Bases (Mbp)	Median Read Length (bp)
TU01	605,501	123.3	220
TU03	644,314	127.2	214
TU05	721,499	141.6	214
TU08	675,132	130.2	214
Total	2,646,446	Total 522.2	Ave. 214

Table 5. Number of reads, number of bases, and median read length of each of four samples using the Ion PGM sequencer.

Alleles of HLA class I genes obtained from the typing method using Ion PGM™ are shown in Table 6. The 1,292,006 draft sequence reads (48.8% of the passed assembly reads) when compared with the reference sequences using the GS Reference Mapper (Ver. 2.5) matched consistently with the HLA alleles that were assigned by Roche GS Junior sequencing. The average depth between two-allele sequence reads spanned from 581 for *HLA-DRB1* in the TU1 sample to 2177 for *HLA-B* in the TU5 sample. The ratio between the average depth of allele 1 and average depth of allele 2 was 0.87 to 2.03. These typing results using the Ion PGM were consistent with the results obtained by Roche Junior system at an 8-digit level with no phase ambiguity. Also, for HLA class II genes, the Ion PGM typing results for HLA-DRB1, -DQA1, -DQB1, -DPA1 and -DPB1 by the SS-SBT methods was an exact match with the results obtained by Roche Junior system demonstrating that complete and correct HLA typing was carried out efficiently by both sequencing systems [42].

A comparison between the Ion PGM and the Roche GS Junior for the depth distribution obtained for HLA-B is shown in Figure 7.

The depth distribution of nucleotide calls across the HLA-B DNA sequence in different samples occurred with numerous peaks and valleys. For HLA-B the depth of nucleotide calls was lowest at a depth of 38 for allele 2 in sample TU01 where a string of polyG was sequenced in intron 7 by the Ion PGM sequencer. However, a depth of > 30 appears to be sufficient for accurate calls and a sequence run of more than 15 identical nucleotides seemed to be accurate enough both by the Roche Junior and Ion PGM sequencers in this study. Overall, the trend in the variation of depth distributions tended to be similar between the samples for both the Roche GS Junior and Ion PGM suggesting that read depth was probably more dependent on the gene sequence and grouping of nucleotides than other factors such as read length, fragmentation of PCR products for the library construction, fragment size selection, emPCR variability and efficiency of sequencing primer locations.

Locus	Sample ID	Allele 1		Allele 2	
		Allele name	Depth Ave.	Allele name	Depth Ave.
HLA-A	TU01	A*02:06:01:01	1281	A*11:01:01:01	1297
	TU03	A*24:02:01:01	1171	A*31:01:02:01	1194
	TU05	A*26:01:01:01	1922	A*31:01:02:01	1923
	TU08	A*24:02:01:01	1280	A*332:03:01:01	1267
HLA-B	TU01	B*40:02:01:01	1509	B*55:02:01(:02)*	1419
	TU03	B*02:06:01:01	1337	B*35:01:01:02	1330
	TU05	B*02:06:01:01	2167	B*35:01:01:02	2177
	TU08	B*02:06:01:01	1478	B*48:01:01:01	1428
HLA-C	TU01	C*02:06:01:01	1612	C*03:03:01:01	1576
	TU03	C*02:06:01:01	1328	C*07:02:01:03	1296
	TU05	C*02:06:01:01	1831	C*07:02:01:04	1808
	TU08	C*02:06:01:01	1661	C*14:03:01:01	1642

Table 6. Class I alleles for each sample identified by typing with Ion PGM.

http://dx.doi.org/10.5772/57556

3.3. HLA assignment software comparisons

We developed an in-house software system (SEABASS) for HLA sequence analysis, allele and in phase haplotyping and compared it to two commercial software systems, the Conexio Genomics genotyping software, Assign-MPS v1 [59], and the Omixon Target HLA Typing software program [70]. The user interface for the Assign-MPS v1 is shown in Figure 8 for (A) accurately genotyping the HLA-B*56:01:01(02) sequence and (B) not detecting DQB1*03:03:02 because of too many low quality sequence reads.

All three software systems imported and converted the.fna files (FASTA format sequence files for each read) generated by the GS Junior or the Ion PGM software into the appropriate files for HLA genotype assignments. All the software programs analysed the various sequence reads, sorted them according to genomic primer and MID tag or barcodes, compared the sequences to the IMGT/HLA sequence database and generated a consensus sequence with allele assignment taking into account the exonic and intronic sequence and phase relationship. The numbers of different reverse and forward reads for each amplicon were indicated, phased and automatically assigned a genotype only if the aligned sequences were perfectly matched with the alleles listed in the HLA sequence database (Figure 9). In cases where the aligned sequence reads were mismatched at one or more bases with the database, manual editing allowed for further investigation and assignment of a genotype.

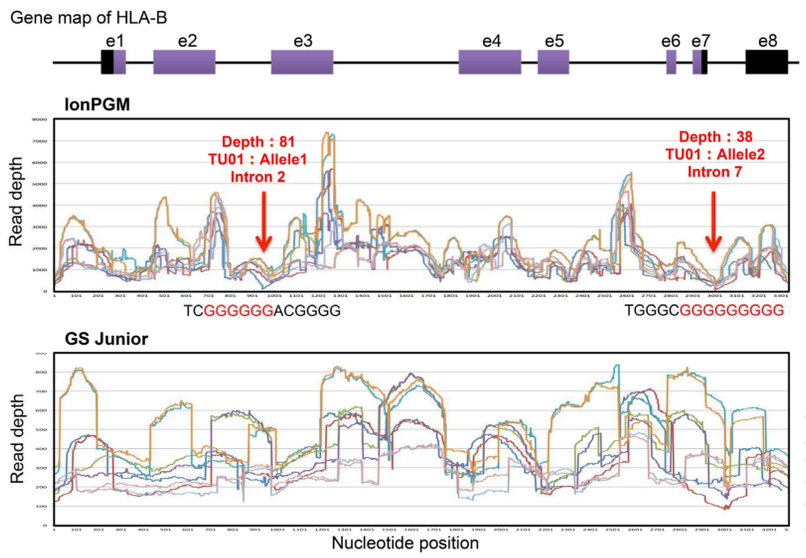


Figure 7. Distribution of depth of base calls for all samples at each nucleotide position of the HLA-B gene using Ion PGM and Roche GS Junior sequencers

A comparison of the three software systems for allele assignment in terms of platform, convenience, analysis speed, detection of new alleles, field 2 and field 4 level of typing assignment are shown in Table 7. All three software systems performed excellently for HLA typing at the field 2 level, but the SEABASS method was better than Assign and Omixon for HLA typing at the field 4 level. In addition, the SEABASS and Assign software could detect new alleles whereas the Omixon did not provide this possibility. However, Omixon was best

for analysis speed and convenience of use. Overall, there was marginal difference in the efficiency, performance, analysis and outputs between the three output systems, although we favored our in house system over the commercially available Conexio Genomics software on a cost benefits basis.

	SEABASS	Omixon 1.6.0	Assign MPS
Platform	Linux	Mac/Windows	Windows
Convenience	*	***	*
Analysis speed	*	***	**
Detection of NEW allele	YES	NO	YES
Field 2 level typing	***	**	**
Field 4 level typing	***	*	*

Table 7. Comparison of allele assignment software systems

4. Discussion

Our NGS study focused on developing a suitable SS-BT method for in phase HLA genotyping for research and diagnostic laboratories using two of the commercially available low to medium throughput capacity NGS systems, the Roche GS Junior and the Ion PGM [42,56]. So far, in our best-case scenario, we were able to sequence 11 HLA loci for 5 individuals in a single sequencing run by Roche GS Junior or the Ion PGM. We have not used the full capacity of the Ion Torrents sequencing chip and, therefore, there is a potential to use a greater number of loci or individual samples (at least 57 samples) than we have already used for a single sequencing run. Moreover, both platforms provided high-resolution or super high resolution HLA typing without ambiguities, depending on the LR-PCR design to amplify the HLA gene loci.

High-throughput HLA genotyping methodologies were previously developed using massively parallel sequencing strategies, such as Roche/454 [36,37,40-42,56,71] and Illumina MiSeq sequencing [43,72]. Most of these high-throughput HLA-genotyping studies amplified a few individual exons (usually exons two, three and four) in an exon based strategy and sequenced in a multiplexed manner. LR-PCR was used by a few of the investigators to amplify large genomic regions of each gene including introns and all or most of the exons in a single PCR [40,42,43,56,72]. We also chose to combine LR-PCR with NGS because LR-PCR requires only one or two primer sets and eliminates the need to validate multiple sets of primers to amplify all alleles in the exon-based strategy. In addition, the error rates of the polymerase enzymes used in LR-PCR, because of error repair, are typically two- to six-fold lower than that of Taq polymerase that is used in conventional PCR [73]. We amplified and sequenced from the 5′ promoter to the 3′ UTR including exons 1-7 for HLA class I and class II genes in order to substantially improve the allele resolution for genotyping in comparison with the previous conventional genotyping methods, such as SSOP and SBT, with which allele calling of

http://dx.doi.org/10.5772/57556

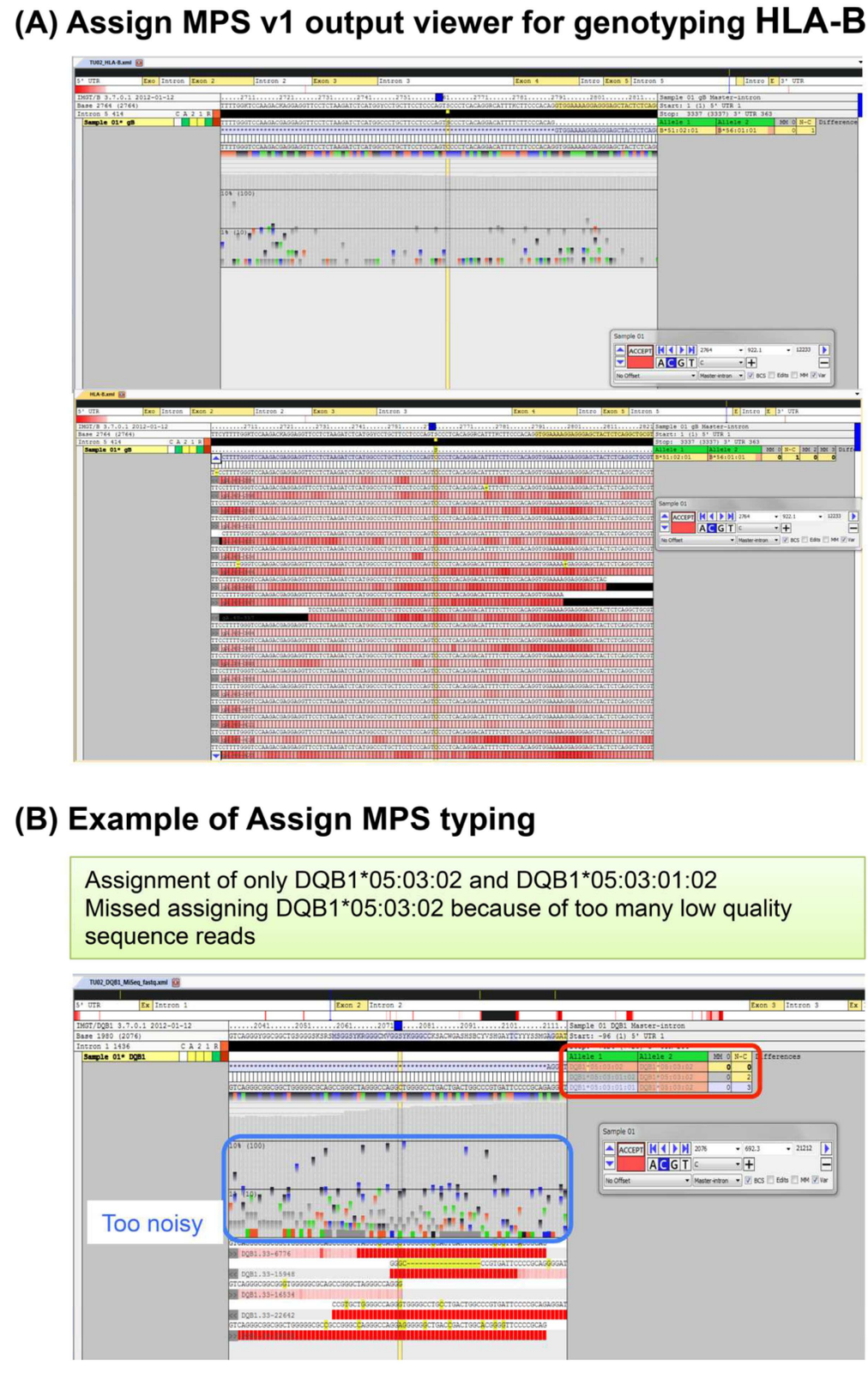
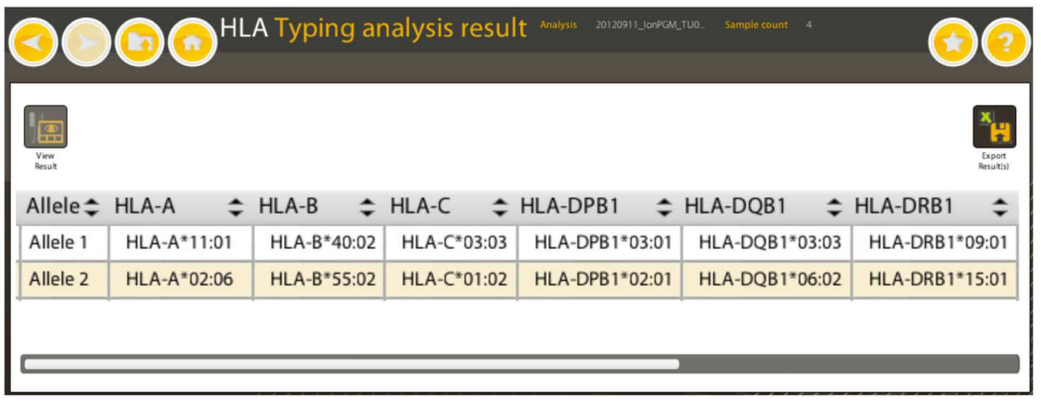


Figure 8. Assign-MPS v1 user interface (A) accurately genotyping the HLA-B*56:01:01(02) sequence, and (B) not detecting DQB1*03:03:02 because of too many low quality sequence reads

sequences was largely limited to exons 2 and 3 for HLA class I genes and exon 2 for HLA-DRB1.



Allele	HLA-A	HLA-B	HLA-C	HLA-DPB1	HLA-DQB1	HLA-DRB1
Allele 1	HLA-A*11:01	HLA-B*40:02	HLA-C*03:03	HLA-DPB1*03:01	HLA-DQB1*03:03	HLA-DRB1*09:01
Allele 2	HLA-A*02:06	HLA-B*55:02	HLA-C*01:02	HLA-DPB1*02:01	HLA-DQB1*06:02	HLA-DRB1*15:01

Figure 9. Omixon software system output viewer of HLA typing analysis results for HLA-A, -B, -C, -DPB1, -DQB1 and -DRB1 at the field 2 level of allele resolution

We developed and tested LR-PCR for the following 11 class I and class II HLA genes, HLA-A, -B. -C, -DRB1, -DRB3, -DRB4, -DRB5, -DQA1, -DQB1, -DPA1 and -DPB1 using 13 separate LR-PCRs. The entire gene sequence from the enhancer/promoter region to the 3'UTR-region was amplified for all of the HLA gene loci, except for HLA-DRB1, -DPB1 and -DRB3/4/5. PCR of HLA-DRB1 and -DPB1 was divided into two parts with enhancer/promoter to exon 2 and exon 2 to 3'UTR for HLA-DRB1, and enhancer/promoter to intron 2 and intron 1 to 3'UTR for DRB1*15:01:01:03 because of the large size of intron 1 (~ 8kb) and/or the complexity of nucleotide repeat (microsatellite) polymorphisms in intron 2 (Figure 7). The three DRB3/4/5 specific primer sets amplified the gene regions from intron 1 across to exon 6 and into the 3'UTR. The LR-PCR of these HLA genes revealed intronic as well as the exonic polymorphic sites, which extended the HLA allelic resolution phase, but also provided important phasing information to assist with the resolution of combination ambiguities and identifying previously unknown alleles outside the regions of exon 2 and 3. Although the IMGT/HLA database has sequences mostly of exons 2 and 3 [74,75], the promoter/enhancer, intron, and 3'UTR variants should not be ignored for more comprehensive HLA typing now and into the future. The genetic variants in a regulatory element such as promoter or even introns need to be extensively analyzed because autoimmune [76] and infectious diseases [77] have been associated with the differential expression levels of the HLA genes in different haplotypes [78]. In addition, null alleles resulting from intronic polymorphisms warrant investigation and resolution to better understand their functional effects [79-82].

After PCR amplicon production, there are four main steps leading to sequencing and HLA data analysis; amplicon library preparation, emulsion PCR, NGS and HLA data analysis (Figure 2). The preparation of the template library is an important step because the PCR amplicons are used either singly or pooled together into a multiplex and then labeled with sequence tags (indexes, barcodes or MIDs) during library construction to facilitate sample multiplexing prior to emulsion PCR and sequencing. The addition of different barcodes to

http://dx.doi.org/10.5772/57556

different sample libraries enables the independent detection of sequences in a mixture of different samples. Computing software is used to accurately parse the tagged sample files during the analysis of sequencing data. However, a possible disadvantage of using too many multiple barcoding tags is to lose sequencing depth because when samples are pooled for multiplexing the amount of input DNA for each sample, due to the increased sample number, is reduced.

In previous studies, most investigators using either the Roche or Illumina systems pooled their PCR amplicons at equimolar concentrations for even gene distribution before constructing barcoded libraries [36,37,40,41,43,56,71,72]. The ligation of barcodes to the fragmented DNA templates during library construction allowed a varying number of different samples to be sequenced during a single sequencing run, depending on read lengths, the capacity of the sequencing platform and the number of pooled amplicons that were used. For example, Erlich et al. [71] reported sequencing a maximum of 760 samples per run after loading a multiplex of 95-96 barcoded samples into a single lane of an 8-region PicoTiter Plate using the 454 Genome Sequencer FLX instrument. Initially, we used only a single gene-sequencing run (SGSR) by sequencing a single PCR amplicon and up to ten barcoded samples per sequencing run with the Roche GS Junior and chose to use only up to four barcoded samples per sequencing run with the Ion PGM [42]. As described in this chapter, we now have changed from SGSR to a multiplex gene-sequencing run (MGSR), for which we pooled all 13 LR-PCR amplicons together into a single sample (Figure 5) prior to constructing libraries for five barcoded samples per sequencing run. This change, from SGSR to MGSR, reduced the number of sequencing runs from 13 to 1, thereby greatly reducing the workload, the cost per sample and the overall cost per individual. Currently, five barcoded samples using 13 LR-PCR amplicons in a single sample is close to Roche GS Junior's maximum sequencing capacity of 80 Mb of sequence reads per run (assuming an average 100-bp read length) without compromising depth of coverage and increasing sequencing or genotyping errors. A statistically adequate depth of sequence coverage is essential to prevent alignment errors and minimize genotyping errors. In comparison, the Ion PGM 318 sequencing chip has far greater sequencing capacity of 1000 Mb ((assuming an average 200 bp read length) and potentially we could use the Ion PGM for genotyping up to 57 barcoded samples using the 13 LR-PCR amplicons in a single sample, if necessary. Although the sample number per sequencing run is low for the Roche GS Junior compared to high capacity NGS platforms, often a small HLA typing laboratory for transplantation matching would require the typing results from only a small number of samples on a weekly basis. If sequencing is required for a very large number of samples, such as for association studies or population diversity studies, our workflow for the Roche Junior can be easily adapted to the larger capacity platforms such as the Roche 454 Genome Sequencer FLX instrument.

Apart from different sequencing capacities, the Roche Junior and Ion PGM use different sequencing principles and procedures. Roche uses a pyrosequencing fluorescence technology with a light output detected by camera scanning [31, 32], whereas the Ion PGM uses Ion Torrent semiconductor sequencing technology with simple sequencing flow chemistry and no light [33]. Ion PGM is the first commercial sequencing machine that does not require fluorescence

and camera scanning, resulting in higher speed, lower cost, and smaller instrument size. Currently, it enables 200 bp reads in 3 hours and the sample preparation time is less than 13 hours for 8 samples in parallel. Because every new sequencing technology introduces unique errors and biases into the resulting DNA sequences, a proper understanding of the NGS specific characteristics that are used to identify and interpret reads is crucial in assessing the accuracy and applications for this new technology. Both manufacturers provide their own unique software to processes the raw acquisition data and produce read files that contain high quality consensus reads and draft assemblies. Roche has the GS analysis software and Ion Torrent has a web browser driving the Torrent Suite Software on computers attached to their respective sequencing instruments. For both platforms, we used the manufacturers software to preprocess the DNA sequencing raw data before transferring the edited data onto our HLA genotyping software. Therefore, we have not directly addressed what errors are introduced into the raw reads by the NGS apparatus. However, others have shown that the major sequencing errors with both sequencing instruments prior to pre-processing are largely related to high frequency indel polymorphisms, homopolymeric regions, replicate bias, and substitution errors that mostly increase in rate with distance from the read start [45,65-67]. In addition, the quality scores that are Phred-based can be only used to detect inserted and substituted bases for both the Roche and Ion Torrent platforms. While the Ion PGM quality scores underestimate the base accuracy, the Roche 454 quality scores tend to over estimate the base accuracy [45,65-67]. Although there are no key studies of raw error reads directly comparing the Roche GS Junior with Ion PGM outputs, a recent statistical study has concluded that the accuracy of the Ion PGM is poorer than that of light-based technologies [65]. On the other hand, direct comparisons of the Roche light-based technologies and Ion PGM of pre-edited sequencing data of bacteria has generally concluded that there is little difference in the accuracy of the sequencing data and that most errors arise with indel polymorphisms and homopolymeric regions [45-47]. Currently, we are conducting our NGS sequencing experiments with 100 DNA samples using the Roche Junior and 200 DNA samples with the Ion PGM, so, at this stage, we have not performed any accurate or meaningful statistical comparisons for the various sequencing variables produced by the two different platforms. In general, however, the indels and homopolymers are only a minor problem for SS-SBT HLA genotyping by NGS using either of the two platforms.

After pre-editing the sequences generated by the two NGS platforms using the manufacturers software, we transferred the edited data to HLA typing software. We compared a software program, called the Suzuki method, which was developed in-house, against two commercial programs, Omixon Target (Omixon) and Assign MPS (Conexio). These stand-alone software programs were assessed and compared for in phase sequence alignment of HLA genes and for allele assignment at different levels including the 8-digit level. Most genotype-calling algorithms select the HLA type candidates based on optimized alignment to cDNA references from the IMGT-HLA database due to the lack of genomic reference sequences. All three software programs consolidated and assessed the various sequence reads, sorted them according to genomic primer and barcodes, compared the sequences to the IMGT/HLA sequence database references and generated a consensus sequence with allele assignment taking into account the exonic and intronic sequence and phase relationship. The barcodes or

http://dx.doi.org/10.5772/57556

MID tags identified by the software were used to reveal the reads of the various samples. Scores for sequence measures and quality assurances were provided including sequence depth, allele depth and allelic balance. Reads were aligned to the various loci and regions based on 100% matching between the read sequence and the reference library. Finally, a consensus sequence was generated and allele assignment made taking into account exonic and intronic sequence as well as phase relationship. The numbers of different reverse and forward reads for each amplicon were indicated, phased and automatically assigned a genotype only if the aligned sequences were perfectly matched with the alleles of genomic references or constructed references. A mapping procedure was applied for each candidate allele to verify the accuracy of the HLA typing and to detect novel alleles. In cases of less than 100% match, even at 99%, mis-mapping can occur among the HLA loci. Also, at least an average 30-fold depth was necessary to identify genetic variants with high sensitivity and resolution. When sequencing read numbers were too few, then we could not make exact assignments. If aligned sequence reads were mismatched at one or more bases with the database, manual editing allowed for further investigation and assignment of a genotype. Overall, there was little difference in the efficiency, performance, analysis and outputs between the three software systems, although we favored our in house program over the commercially available software on a cost benefits basis. Because the IMGT/HLA sequence database has relatively few genomic sequences, at less than 6% of the database entries, a major task remains to continue building a suitable reference library for all the known polymorphic HLA genes.

Our main aim in using the NGS technology was to eliminate the ambiguities currently associated with the conventional HLA genotyping methods. So far, we found that the SS-SBT method is superior to other HLA DNA typing methods, especially to efficiently detect new HLA alleles and null alleles at the 8-digit level of DNA typing without ambiguity. Although, at most, only 100 Japanese and Caucasian genomic DNA samples were used in this study, we unequivocally defined the HLA-A, -B, -C, -DRB1 and -DQB1 loci to single HLA alleles at the 8-digit level without any ambiguity. In addition, 17 DRB allele sequences, seven in DRB1, three in DRB3, four in DRB4 and three in DRB5, were newly determined to the field 4 level of allele resolution without phase ambiguity by SS-SBT. However, achieving a complete depth of correct sequence information in most samples of DRB1 and in some samples of DRB3/4/5, such as TU20 in DRB3, was compromised by the presence of microsatellites in a limited number of intronic sites. Major sequence instabilities were encountered in our study with $T^{5\sim17}$ and $T^{2\sim27}$ mono-stretch sequences and $GT^{7\sim28}$ and $GA^{3\sim23}$ microsatellite repeats in intron 1 of DRB1, intron 5 of DRB1 and intron 2 of DRB1 and DRB3/4/5 alleles that were obtained in our study and from the IMGT-HLA database. These instabilities within the microsatellite repeats are probably compounded by PCR and sequencing errors [65], but could be solved by Sanger sequencing of PCR products using high fidelity DNA polymerase.

All the allele sequences, excluding DRB1*09:21, were perfectly matched to at least the previously reported field 1 or field 2 level of allele information [83]. Therefore, the newly designed primers and PCR conditions for HLA-A, -B, -C, -DRB1, -DQA1, -DQB1, -DPA1 and -DPB1 [42] and -DRB3/4/5 [56] are efficient for DNA typing by the SS-SBT method. Of the 164 alleles detected at the field 4 assignment level in 81 Japanese and Caucasian samples, 78 (47.6%) were

newly detected alleles. Therefore, simply increasing the sample size in future analyses of HLA polymorphisms by the SS-SBT method could identify new non-synonymous substitutions along with indels that generate a null allele. A new allele DRB1*09:21 that we identified in only one sample so far, in comparison to the 129 DRB1*09:01 positive genomic DNA samples typed by SS-SBT in the Japanese population, suggests that the new allele has extremely low frequency in the Japanese.

We encountered problems with some microsatellite sequences, especially for DRB1*15:01:01:03 with the complexity of microsatellite polymorphisms in intron 2. In addition, the 8-digit level HLA alleles could not be assigned in HLA-DQA1, -DPA1 and -DPB1 because the SNP and indel densities to separate both of the phases were much lower than in the other HLA loci. Resolution was difficult for phase ambiguities in the conserved sequences at HLA loci, such as exon 3 in the HLA class II genes. Therefore, most HLA allele sequences are still unknown at the 8-digit level. In this respect, collection of the 8-digit level reference sequences using HLA homozygous DNA samples, haplotype extraction methods [84] or third generation sequencers such as one molecule real time DNA sequencer PacBio RS [85,86] (Pacific Biosciences, Menlo Park, CA) that provide a 3 kb read length on average are necessary for solving the current genotyping problems to improve the SS-SBT method.

The collection of HLA allele sequences at the 8-digit level and the development of HLA allele assignment programs are necessary to improve the SS-SBT method. The average depth of HLA-A, -B and -C was stable with a ratio of ~1:1. However, in comparison, the average depth of HLA-DRB1 for LR-PCR mix 2 and -DQB1 was less stable with ratios varying between 1:1 and 1: <2 [42]. Monitoring the depth ratio and potential allele dropout is important to detect PCR bias due to unexpected variations in the primer sites. In cases where the depth ratio is drastically changed such as 1:5 or more, the primer sets and/or PCR condition may have to be modified. Nevertheless, the newly developed HLA DNA typing method SS-SBT [42] is potentially applicable to the diagnostic laboratory once some of the minor problems described above are solved in future.

Our study is the first direct comparison between the GS Junior and Ion PGM sequencing platforms for HLA genotyping. Although we have as yet to maximize the potential capacity of the Ion PGM to meet our theoretical expectation of sequencing the pooled PCR amplicons for 11 gene loci using 57 barcoded samples, the Ion PGM seems to perform as well as or better than Roche Junior on a number of fronts. However, it is difficult to choose between the two platforms at this stage and further work and comparative analysis will need to be performed before drawing a definite conclusion. Whereas there is little difference in overall performance between the two platforms at this time, the new Ion Torrent 318 chip offers greater capacity for more samples and more accurate reads with read lengths of 400 bp, for which we are presently testing 300 samples. In addition, automation of library preparations and emPCR amplification steps prior to sequencing has improved the overall turn around time from library preparation to allele readouts from four days to two/three days with the sequencing step performed overnight (Figure 2).

Development of better technologies to reduce the complication of the process and running costs is also necessary before the SS-SBT method is introduced into the diagnostic laboratory.

http://dx.doi.org/10.5772/57556

In a previous simulated test of the Ion Torrent PGM system in a clinical laboratory setting, we needed four days for the manual sample preparation and library construction from the time of PCR amplification to HLA allele definition [42]. However, with the use of automated procedures such as the AB Library Builder™ system and Ion OneTouch™ system from Life Technologies we have shortened the workflow significantly to just two days with a running cost of US $17 per locus per sample. In comparison, the workflow for the Roche Junior remains at four days from the time of PCR amplification to HLA allele definition at a running cost of US $40 per locus per sample. Also, a new protocol using the new Ion 318 chip with 32 barcodes for 400 bp-read high sequencing quality has increased sequencing capacity and enabled sequencing reads up to and beyond 400 bp [87]. Although we are at this moment testing the new Ion Torrent protocols with the Ion 318 chip it looks likely to lead to further improvements and further cost reductions. Therefore, a decrease in the running costs of NGS is expected to soon be substantially better than those of the conventional HLA DNA typing methods.

Whereas we have tested the Roche Junior and Ion PGM NGS compact systems, the Illumina MiSeq is another commercially available compact NGS [45]. The Illumina MiSeq sequencer appears to compare well with Roche Junior and Ion PGM for HLA typing [43,72] and sequencing other regions and targets of interest [45]. Other laboratories have favored the Illumina MiSeq system and have published the workflows and results for HLA typing and sequencing haplotypes to high resolution levels [43,72]. The Illumina MiSeq offers some advantages over the Roche and Ion PGM such as low DNA input amount, the pair-end reads (ie., fragments sequenced in both directions) to determine in phase alleles and possibly, a better and more reliable resolution of substitutions and indels [46]. However, the Illumina MiSeq may not perform as well for the middle reads of a 150 bp/200 bp sequence and the generated reads may exhibit rapidly increasing error rates as the read length increases, resulting in lower quality contig assemblies [46]. Also, the HLA DNA data analysis step appears to be a more difficult process [43,72] than that for the Roche Junior and Ion PGM (unpublished data). In cases where the phase of an allele is unresolved by the NGS system and HLA software other approaches may help to resolve the allele ambiguity. For example, one promising approach is to use DNA haplotype-specific extraction of the gene from the sample using a commercially available extraction kit, such as the solid-phase, capture-based EZ1 HaploPrep kit from Qiagen [88], and then resequencing the unresolved haplotype. The DNA haplotype-specific extraction procedure can extract genomic regions of up to 50 kb of contiguous sequence without amplification or concentration of the extracted DNA, and it has been used successfully to genotype and haplotype HLA class I and class II genes [84, 89-93] including adjoining genes such as HLA-B, MICA and HLA-C [89].

HLA in-phase genotyping is important for a variety of applications including for infectious disease studies, transplantation, pharmacogenomics, autoimmune diseases, population diversity and human evolution, and treatment of cancer pathology by vaccination. While we used the genome from whole blood or from peripheral blood mononuclear cells for HLA in phase-genotyping the analysis of HLA cDNA from cells and tissue (fresh or formalin fixed) would also be of value [38,39]. In humans, the HLA-A, -B, and -C locus-specific gene expression patterns were reported in the peripheral blood leukocytes, colon mucosa and larynx mucosa

by real-time PCR [94]. Recently, we investigated the relationship between haplotypes and gene expression levels of the class I genes by sequencing the genomic DNA of pigs using the Roche Genome Sequencer 454 FLX and found that the sequence read numbers closely reflected the gene expression levels in white blood cells [95]. The use of the NGS sequencing method in human studies, similar to our MHC locus-specific expression analysis in pigs, could also provide informative data in various biomedical studies on HLA gene expression, such as the detection of expression levels among inter- and intra-populations, and among different tissues, both before and after vaccination against pathogens.

High-resolution donor-recipient HLA matching contributes to the success of unrelated donor organ and marrow transplantation [2,96]. HLA typing also plays critical roles in donor and recipient matching for embryonic (ES) or induced pluripotent stem (iPS) cell transplantation therapy. Currently, an iPS cell bank project is under way, led by Kyoto University, with the participation of one of the co-authors (HI) of this chapter. It was reported that when thirty iPS cell lines that have different combinations of homozygous HLA-A, -B, and -DR are generated, these iPS cells will have matched the three loci in 82.2% of the Japanese population. For fifty iPS cell lines, the chance of a match increases to 90.7% in the Japanese population [97].

In addition to finding new polymorphisms within exons, the SS-SBT method can analyze polymorphisms in introns, promoter, enhancer, and 3′-UTR regions that have largely remained unexplored until now. By analyzing polymorphisms in the entire region of HLA genes, the functional influence of those polymorphisms can be revealed in transplantation, diseases, and adverse effects of medications. For example, it is expected that about one in a thousand people have a null allele (deficient mutant that influences function or expression) in the HLA region. Since null alleles have a profound influence on GVHD or transplant establishment, especially in hematopoietic stem cell transplantation, the detection of null alleles is considered to be of great clinical importance.

The SS-SBT method has been developed to obtain massive and accurate sequencing data easily and cost effectively. Recently, the authors of a critical review of HLA typing by NGS [98] pointed out that the Roche company will discontinue product support for the 454 sequencing systems in 2016, implying that the Roche Junior sequencing platform will be phased out commercially in the near future. However, almost all the materials and reagents required for the Ion PGM™ system are available as kit products, and one Gb of sequencing data can be obtained within 5 hours of running the protocols and data analysis. Up to fifty-seven samples can be multiplexed and eight HLA gene loci analysed at an eight-digit level in a single run. Compared to the Luminex beads method or the SBT method, the SS-SBT is more cost effective. For these reasons, the Ion PGM™ sequencer is potentially the perfect sequencer for HLA genotyping using the SS-SBT method. With an expected increase in throughput and the development of an automated system in the future, we hope that Ion PGM™ will be even easier to use for routine HLA genotyping.

In conclusion, we have developed procedures for massively parallel sequencing of multiplex products that can be used for several benchtop sequencing platforms. We have obtained sequences of sufficient high quality to permit accurate HLA in phase genotypes across the full-length gene from the 5′ promoter/enhancer region to the 3′ UTR for most of the classical class

I and class II genes. The use of sample tags or barcodes allows for optimization of second generation sequencing technologies by pooling samples and sequencing multiple samples in parallel for time- and cost-efficient workflows. We are currently working towards optimizing the Ion PGM/SS-SBT method for HLA in phase genotyping for both the clinical diagnostic and research laboratories.

Erratum

The software Assign MPS 1.0 used in this study was a beta version and still under development at the time of publishing the chapter and not available commercially from Conexio.

Author details

Jerzy K. Kulski[2*], Shingo Suzuki[1], Yuki Ozaki[1], Shigeki Mitsunaga[1], Hidetoshi Inoko[1] and Takashi Shiina[1]

*Address all correspondence to: kulski@me.com

1 Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Isehara, Kanagawa, Japan

2 Centre for Forensic Science, The University of Western Australia, Nedlands, WA, Australia

References

[1] Zinkernagel RM, Doherty PC. The discovery of MHC restriction. Immunology Today 1997;18(1): 14-7. DOI: 10.1016/S0167-5699(97)80008-4

[2] Sheldon S, Poulton K. HLA typing and its influence on organ transplantation. Methods in Molecular Biology 2006;333: 157-74. DOI: 10.1385/1-59745-049-9:157

[3] Park M, Seo JJ. Role of HLA in hematopoietic stem cell transplantation. Bone Marrow Research 2012;2012: 680841. DOI:10.1155/2012/680841

[4] Shiina T, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations: 2004.Tissue Antigens 2004;64(6): 631-49. DOI: 10.1111/j.1399-0039.2004.00327.x

[5] Bolton EM, Bradley JA. Transplantation Immunology. In: Eremin O, Sewell H. (ed.) Essential Immunology for Surgeons. Oxford University Press; 2011. Chpt 3, p199-235.

[6] Duquesnoy RJ. Antibody-reactive epitope determination with HLAMatchmaker and its clinical applications. Tissue Antigens 2011;77(6): 525-34. DOI: 10.1111/j.1399-0039.2011.01646.x

[7] Ponticelli C. The mechanisms of acute transplant rejection revisited. Journal of Nephrology 2012;25(2): 150-8. DOI: 10.5301/jn.5000048

[8] Garcia MAA, Yebra BG, Flores ALL, Guerra EG. The Major Histocompatibility Complex in transplantation. Journal of Transplantation. 2012;2012: 842141. DOI: 10.1155/2012/842141

[9] Holoshitz J. The quest for better understanding of HLA-disease association: Scenes from a road less travelled by. Discovery Medicine 2013;16(87): 93-101. www.discoverymedicine.com/Joseph-Holoshitz/2013/08/26/the-quest-for-better-understanding-of-hla-disease-association-scenes-from-a-road-less-travelled-by/ (accessed 18 December 2013).

[10] Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. Journal of Human Genetics 2009;54(1): 15-39. DOI: 10.1038/jhg.2008.5

[11] Kropshofer H, Hammerling GJ, Vogt AB. The impact of the non-classical MHC proteins HLA-DM and HLA-DO on loading of MHC class II molecules. Immunological Reviews 1999;172: 267-78. DOI: 10.1111/j.1600-065X.1999.tb01371.x

[12] Dawkins R, Leelayuwat C, Gaudieri S, Tay G, Hui J, Cattley S, Martinez P, Kulski J. Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. Immunological Reviews 1999;167: 275-304. DOI: 10.1111/j.1600-065X.1999.tb01399.x

[13] Naves EM, Cuadrado JFP, Perez Rosada A, Gomez del Moral M. Structure and function of 'non-classical" HLA class I molecules. Immunologia 2001;20(4): 207-15. http://revista.inmunologia.org/Upload/Articles/5/4/545.pdf (accessed 18 December 2013).

[14] Erlich HA, Opelz G, Hansen J. HLA DNA typing and transplantation. Immunity 2001;14(4): 347-56. DOI: 10.1016/S1074-7613(01)00115-7

[15] Mahdi B M. A glow of HLA typing in organ transplantation. Clinical and Transplantation Medicine 2013;2(1): 6. DOI: 10.1186/2001-1326-2-6

[16] Mack SJ, Sanchez-Mazas A, Single RM, Meyer D, Hill J, Dron HA, Jani AJ, Thomson G, Erlich HA. Population samples and genotyping technology. Tissue Antigens 2007;69(Suppl s1): 188-91. DOI: 10.1111/j.1399-0039.2006.00768.x

[17] Vina MA, Hollenbach JA, Lyke KE et al. Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. Philosophical Transactions of the Royal Society of London B: Biological Sciences 2012;367(1590): 820-9. DOI: 10.1098/rstb.2011.0320

[18] Nakaoka H, Mitsunaga S, Hosomichi K et al. Detection of ancestry informative HLA alleles confirms the admixed origins of Japanese population. PLoS One 2013;8(4): e60793. DOI: 10.1371/journal.pone.0060793

[19] Grubic Z, Stingl K, Martinez N, Palfi B, Brkljacic-Kerhin, V, Kastelan A. STR and HLA analysis in paternity testing. International Congress Series 2004;1261: 535-7. DOI: 10.1016/S0531-5131(03)01654-6

[20] Alfirevic A, Pirmohamed M. Drug induced hypersensitivity and the HLA complex. Pharmaceuticals 2011;4(1): 69-90. DOI: 10.3390/ph4010069

[21] 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(9308): 727-32. DOI: 10.1016/S0140-6736(02)07873-X

[22] Ota M, Fukushima H, Kulski JK, Inoko H. Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism. Nature Protocols 2007;2(11): 2857-64. DOI: 10.1038/nprot.2007.407

[23] Argu"ello JR, Madrigal JA. HLA typing by reference strand mediated conformation analysis (RSCA). Reviews in Immunogenetics 1999;1(2): 209-19.

[24] Saiki R, Walsh PS, Levenson CH, Erlich HA. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proceedings of the National Academy of Sciences of the United States of America 1989;86(16): 6230-4. www.pnas.org/content/86/16/6230 (accessed 18 December 2013)

[25] Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor–recipient matching in cadaveric transplantation. Tissue Antigens 1992;39(5): 225-35. DOI: 10.1111/j.1399-0039.1992.tb01940.x

[26] Santamaria P, Lindstrom AL, Boyce-Jacino MT, Mystera SH, Barbosab JJ, Farasa AJ, Richa SS. HLA class I sequence-based typing. Human Immunology 1993;37(1): 39-50. DOI: 10.1016/0198-8859(93)90141-M

[27] Itoh Y, Mizuki N, Shimada T, Azuma F, Itakura M, Kashiwase K, Kikkawa E, Kulski JK, Satake M, Inoko H. High-throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. Immunogenetics 2005;57(10): 717-29. DOI: 10.1007/s00251-005-0048-3

[28] Itoh Y, Inoko H, Kulski JK, Sasaki S, Meguro A, Takiyama N, Nishida T, Yuasa T, Ohno S, Mizuki N. Four-digit allele genotyping of the HLA-A and HLA-B genes in Japanese patients with Behcet's disease by a PCR-SSOP-Luminex method. Tissue Antigens 2006;67(5): 390-4. DOI: 10.1111/j.1399-0039.2006.00586.x

[29] Hutchison CA III. DNA sequencing: bench to bedside and beyond. Nucleic Acids Research 2007;35(18): 6227-37. DOI: 10.1093/nar/gkm688

[30] Leslie S, Donnelly P, McVean G. A statistical method for predicting classical HLA alleles from SNP data. The American Journal of Human Genetics 2008;82(1): 48-56. DOI: 10.1016/j.ajhg.2007.09.001

[31] Metzger ML. Sequencing technologies – the next generation. Nature Reviews in Genetics 2010;11(1): 31-46. DOI: 10.1038/nrg2626

[32] Rothberg JM, Leamon JH. The development and impact of 454 sequencing. Nature Biotechnology 2008;26(10): 1117-24. DOI: 10.1038/nbt1485

[33] Rothberg JM, Hinz W, Rearick TM et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature 2011;475(7356): 348-52. DOI: 10.1038/nature10242

[34] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 1977;74(12): 5463-7. www.pnas.org/content/74/12/5463 (accessed 18 December 2013).

[35] Maxam A, Gilbert W. A new method of sequencing DNA. Proceedings of the National Academy of Sciences, USA 1977;74(2): 560-4. www.pnas.org/content/74/2/560 (accessed 18 December 2013).

[36] Gabriel C, Danzer M, Hackl C, Kopal G, Hufnagl P, Hofer K, Polin H, Stabentheiner S, Proll J. Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification. Human Immunology 2009;70(11): 960-4. DOI: 10.1016/j.humimm.2009.08.009

[37] Holcomb CL, Hoglund B, Anderson MW et al. A multi-site study using high-resolution HLA genotyping by next generation sequencing. Tissue Antigens 2011;77(3): 206-17. DOI: 10.1111/j.1399-0039.2010.01606.x

[38] Lank SM, Golbach BA, Creager HM, Wiseman RW, Keskin DB, Reinherz EL, Brusic V, O'Connor DH. Ultra-high resolution HLA genotyping and allele discovery by highly multiplexed cDNA amplicon pyrosequencing. BMC Genomics 2012;13: 378. DOI: 10.1186/1471-2164-13-378

[39] Lank SM, Wiseman RW, Dudley DM, O'Connor DH. A novel single cDNA amplicon pyrosequencing method for high-throughput, cost-effective sequence-based HLA class I genotyping. Human Immunology 2010;71(10): 1011-7. DOI: 10.1016/j.humimm.2010.07.012

[40] Lind C, Ferriola D, Mackiewicz K et al. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. Human Immunology 2010;71(10): 1033-42. DOI: 10.1016/j.humimm.2010.06.016

[41] Erlich H. HLA DNA typing: past, present, and future. Tissue Antigens 2012;80(1): 1-11. DOI: 10.1111/j.1399-0039.2012.01881.x

[42] Shiina T, Suzuki S, Ozaki Y 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(4): 305-16. DOI: 10.1111/j.1399-0039.2012.01941.x

[43] Hosomichi K, Jinam TA, Mitsunaga S, Nakaoka H, Inoue I. Phase-defined complete sequencing of the HLA genes by next-generation sequencing. BMC Genomics 2013;14: 355. DOI: 10.1186/1471-2164-14-355

[44] Glenn TC. Field guide to next-generation DNA sequencers. Molecular Ecology Resources 2011;11(5): 759-69. DOI: 10.1111/j.1755-0998.2011.03024.x

[45] Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of next-generation sequencing systems. Journal of Biomedicine and Biotechnology 2012,2012: 251364. DOI: 10.1155/2012/251364

[46] Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. Performance comparison of benchtop high-throughput sequencing platforms. Nature Biotechnology 2012;30(5): 434-9. DOI: 10.1038/nbt.2198

[47] Quail M, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. BMC Genomics 2012;13: 341. DOI: 10.1186/1471-2164-13-341

[48] Bentley DR, Balasubramanian S, Swerdlow HP et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456(7218): 53-9. DOI: 10.1038/nature07517

[49] Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M, Fire AZ. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. Nucleic Acids Research 2007;35(19): e130. DOI: 10.1093/nar/gkm760

[50] Lennon NJ, Lintner RE, Anderson S et al. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. Genome Biology 2010;11(2): R15. DOI: 10.1186/gb-2010-11-2-r15

[51] Margulies M, Egholm M, Altman WE et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005;437(7057): 376-80. DOI: 10.1038/nature03959

[52] Shendure JA, Porreca GJ, Church GM. Overview of DNA Sequencing Strategies. In: Current Protocols in Molecular Biology. Hoboken: John Wiley and Sons; 2008. Vol 8, Ch 7. DOI: 10.1002/0471142727.mb0701s81

[53] Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Research 2006;34(3): e22. DOI: 10.1093/nar/gnj023

[54] GS Junior Bench Top System. Roche 454 Sequencing. http://www.gsjunior.com (accessed 17 October 2013).

[55] Ion Personal Genome Machine (PGM) Sequencer. Specification sheet. http://www3.appliedbiosystems.com/cms/groups/applied_markets_marketing/documents/generaldocuments/cms_094139.pdf (accessed 17 October 2013).

[56] Ozaki Y, Suzuki S, Shigenari A, Okudaira Y, Kikkawa E, Oka A, Ota M, Mitsunaga, S, Kulski JK, Inoko H, Shiina T. HLA-DRB1, -DRB3, -DRB4 and -DRB5 genotyping at a super-high resolution level by long range PCR and high-throughput sequencing. Tissue Antigens. 2014 Jan;83(1):10-6. doi: 10.1111/tan.12258. Epub 2013 Nov 30. PubMed PMID: 24355003.

[57] HLA nomenclature. http://hla.alleles.org/nomenclature/naming.html (accessed 17 October 2013).

[58] Ando A, Shigenari A, Ota M et al. SLA-DRB1 and -DQB1 genotyping by the PCR-SSOP-Luminex method. Tissue Antigens 2011;78(1): 49-55. DOI: 10.1111/j.1399-0039.2011.01669.x

[59] Conexio. http://www.conexio-genomics.com (accessed 17 October 2013).

[60] 454 Sequencing system. Guidelines for amplicon experimental design. June 2013. Roche. http://dna.uga.edu/docs/454SeqSys_AmpliconDesignGuide_Jun2013.pdf (accessed 17 October 2013).

[61] Application note. Ion Torrent amplicon sequencing. http://www3.appliedbiosystems.com/cms/groups/applied_markets_marketing/documents/generaldocuments/cms_094273.pdf (accessed 17 October 2013).

[62] Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Research 1998;8(3): 186-94. genome.cshlp.org/content/8/3/186.long (accessed 18 December 2013)

[63] Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research 1998;8(3): 175-85. DOI: 10.1101/gr.8.3.175

[64] Technical note: sequencing. Quality scores for next-generation sequencing. Illumina. http://res.illumina.com/documents/products/technotes/technote_q-scores.pdf (accessed 17 October 2013).

[65] Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW. Shining a light on dark sequencing: Characterising errors in Ion Torrent PGM data. PLoS Computational Biology 2013;9(4): e1003031. DOI: 10.1371/journal.pcbi.1003031

[66] Gilles A, Meglecz E, Pech N, Ferreira S, Malausa T, Martin JF. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics 2011;12: 245. DOI: 10.1186/1471-2164-12-245

[67] Huse SM, Huber JA, Morrison HG, Sogin ML, Mark Welch D. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biology 2007;8(7): R143. DOI: 10.1186/gb-2007-8-7-r143

[68] Technical note: sequencing. Estimating sequence coverage. Illumina. http://res.illumina.com/documents/products/technotes/technote_coverage_calculation.pdf (accessed 17 October 2013).

[69] UCSC genome bioinformatics. http://genome.ucsc.edu/ (accessed 17 October 2013).

[70] Omixon. Targeted NGS data analysis. http://www.omixon.com (accessed 17 October 2013).

[71] Erlich RL, Jia X, Anderson S et al. Next-generation sequencing for HLA typing of class I loci. BMC Genomics 2011;12: 42. DOI: 10.1186/1471-2164-12-42

[72] Wang C, Krishnakumar S, Wilhelmy J, Babrzadeh F, Stepanyan L et al. High-throughput, high-fidelity HLA genotyping with deep sequencing. Proceedings of the National Academy of Sciences of the United States of America 2012;109(22): 8676-81. DOI: 10.1073/pnas.1206614109

[73] Cline J, Braman JC, Hogrefe HH. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Research 1996;24(18): 3546-51. DOI: 10.1093/nar/24.18.3546

[74] Robinson J, Halliwell JA, McWilliam H, Lopez R, Parham P, Marsh SGE. The IMGT/HLA database. Nucleic Acids Research 2013;41(D1): D1222-7. DOI: 10.1093/nar/gks949

[75] Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SGE. The IMGT/HLA database. Nucleic Acids Research 2011;39 (suppl 1): D1171-6. DOI: 10.1093/nar/gkq998

[76] Cocco E, Meloni A, Murru MR et al. Vitamin D responsive elements within the HLA-DRB1 promoter region in Sardinian multiple sclerosis associated allele. PLoS One 2012;7(7): e41678. DOI: 10.1371/journal.pone.0041678

[77] Thomas R, Apps R, Qi Y et al. HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. Nature Genetics 2009;41(12): 1290-4. DOI: 10.1038/ng.486

[78] Vandiedonck C, Taylor MS, Lockstone HE, Plant K, Taylor JM, Durrant C, Broxholme J, Fairfax BP, Knight JC. Pervasive haplotypic variation in the spliceo-transcriptome of the human major histocompatibility complex. Genome Research 2011;21(7): 1042-54. DOI: 10.1101/gr.116681.110

[79] Elsner HA, Bernard G, Eiz-Vesper B, de Matteis M, Bernard A, Blasczyk R. Non-expression of HLA-A*2901102 N is caused by a nucleotide exchange in the mRNA

splicing site at the beginning of intron 4. Tissue Antigens 2002;59(2): 139-41. DOI: 10.1034/j.1399-0039.2002.590212.x

[80] Curran MD, Williams F, Little AM, Rima BK, Madrigal JA, Middleton D. Aberrant splicing of intron 1 creates a novel null HLA-B*1501 allele. Tissue Antigens 1999;53(3): 244-52. DOI: 10.1034/j.1399-0039.1999.530304.x

[81] Tamouza R, El Kassar N, Schaeffer V et al. A novel HLA-B*39 allele (HLA-B*3916) due to a rare mutation causing cryptic splice site activation. Human Immunology 2000;61(5): 467-73. DOI: 10.1016/S0198-8859(00)00108-7

[82] Dubois V, Tiercy JM, Labonne MP, Dormoy A, Gebuhrer L. A new HLA-B44 allele (B*44020102S) with a splicing mutation leading to a complete deletion of exon 5. Tissue Antigens 2004;63(2): 173-80. DOI: 10.1111/j.1399-0039.2004.00134.x

[83] Bettinotti MP, Mitsuishi Y, Bibee K, Lau M, Terasaki PI. Comprehensive method for the typing of HLA-A, B, and C alleles by direct sequencing of PCR products obtained from genomic DNA. Journal of Immunotherapy 1997;20(6): 425-30. journals.lww.com/immunotherapy-journal/Abstract/1997/11000/Comprehensive_Method_for_the_Typing_of_HLA_A,_B,.1.aspx (accessed 18 December 2013)

[84] Dapprich J, Cleary MA, Gabel HW, Akkapeddi A, Iglehart B, Turino C, Beaudet L, Lian J, Murphy NB. A Rapid, Automatable Method For Molecular Haplotyping. In: J.A. Hansen, (ed.) Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference. Seattle, WA: IHWG Press; 2006. Volume 2, p93-96.

[85] Eid J, Fehr A, Gray J et al. Real-time DNA sequencing from single polymerase molecules. Science 2009;323(5910): 133-8. DOI: 10.1126/science.1162986

[86] Pacific Biosciences. http://www.pacificbiosciences.com/products/ (accessed 18 October 2013).

[87] Ion 318™ Chip Kit v2 (Ion Torrent™). http://www.lifetechnologies.com/order/catalog/product/4484355 (accessed 18 October 2013).

[88] EZ1 HaploPrep Handbook - Qiagen. http://www.qiagen.com/search.aspx?q=ez1%2520haploprep&c={7B5D5E07-20AE-4E4D-B233-FEFA27C84B5B}#&&p=1 (accessed 22 October 2013).

[89] Dapprich J, Ferriola D, Magira EE, Kunkel M, Monos D. SNP-specific extraction of haplotype-resolved targeted genomic regions. Nucleic Acids Research 2008;36(15): e94. DOI: 10.1093/nar/gkn345

[90] Dapprich J, Magira E, Samonte MA, Rosenman K, Monos D. Identification of a novel HLA-DPB1 allele (DPB1*1902) by haplotype-specific extraction and nucleotide sequencing. Tissue Antigens 2007;69(3): 282-3. DOI: 10.1111/j.1399-0039.2006.00752.x

[91] Dapprich J, Witter K, Gabel H, Murphy NB, Albert ED. Identification of a new HLA-B (B*1576) by haplotype Specific Extraction. Human Immunology 2007;68(5): 418-21. DOI: 10.1016/j.humimm.2007.01.015

[92] Guo Z, Hood L, Malkki M, Petersdorf EW. Long-range multilocus haplotype phasing of the MHC. Proceedings of the National Academy of Sciences of the United States of America 2006;103(18): 6964-9. DOI: 10.1073/pnas.0602286103

[93] Nagy M, Entz P, Otremba P, Schoenemann C, Murphy N, Dapprich J. Haplotype-specific extraction: a universal method to resolve ambiguous genotypes and detect new alleles – demonstrated on HLA-B. Tissue Antigens 2007; 69(2): 176-80. DOI: 10.1111/j.1399-0039.2006.00741.x

[94] García-Ruano AB, Méndez R, Romero JM, Cabrera T, Ruiz-Cabello F, Garrido F. Analysis of HLA-ABC locus-specific transcription in normal tissues. Immunogenetics 2010;62(11-12): 711-9. DOI: 10.1007/s00251-010-0470-z

[95] Kita YF, Ando A, Tanaka K, Suzuki S, Ozaki Y, Uenishi H, Inoko H, Kulski JK, Shiina T. Application of high-resolution, massively parallel pyrosequencing for estimation of haplotypes and gene expression levels of swine leukocyte antigen (SLA) class I genes. Immunogenetics 2012;64(3): 187-99. DOI: 10.1007/s00251-011-0572-2

[96] Lee SJ, Klein J, Haagenson M et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. Blood 2007;110(13): 4576-83. DOI: 10.1182/blood-2007-06-097386

[97] Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nature Biotechnology 2008;26(1): 101-6. DOI:10.1038/nbt1374

[98] Gabriel C, Fürst D, Faé I, Wenda S, Zollikofer C, Mytilineos J, Fischer GF. HLA typing by next-generation sequencing - getting closer to reality. Tissue Antigens. 2014;83(2):65-75. DOI:10.1111/tan.12298

IntechOpen