



# Accuracy of Programs for the Determination of Human Leukocyte Antigen Alleles from Next-Generation Sequencing Data

Antti Larjo<sup>1</sup>, Robert Eveleigh<sup>2</sup>, Elina Kilpeläinen<sup>1</sup>, Tony Kwan<sup>2</sup>, Tomi Pastinen<sup>2</sup>, Satu Koskela<sup>1</sup> and Jukka Partanen<sup>1\*</sup>

<sup>1</sup>Finnish Red Cross Blood Service, Helsinki, Finland, <sup>2</sup>McGill University and Genome Quebec Innovation Centre, Montreal, QC, Canada

## OPEN ACCESS

### Edited by:

Alok Srivastava,  
Christian Medical College &  
Hospital, India

### Reviewed by:

Anat R. Tambur,  
Northwestern University,  
United States  
Alexandre Walencik,  
Établissement Français du Sang,  
France

### \*Correspondence:

Jukka Partanen  
jukka.partanen@veripalvelu.fi

### Specialty section:

This article was submitted to  
Alloimmunity and Transplantation,  
a section of the journal  
Frontiers in Immunology

Received: 11 July 2017

Accepted: 01 December 2017

Published: 13 December 2017

### Citation:

Larjo A, Eveleigh R, Kilpeläinen E,  
Kwan T, Pastinen T, Koskela S and  
Partanen J (2017) Accuracy of  
Programs for the Determination of  
Human Leukocyte Antigen Alleles  
from Next-Generation  
Sequencing Data.  
*Front. Immunol.* 8:1815.  
doi: 10.3389/fimmu.2017.01815

The human leukocyte antigen (HLA) genes code for proteins that play a central role in the function of the immune system by presenting peptide antigens to T cells. As HLA genes show extremely high genetic polymorphism, HLA typing at the allele level is demanding and is based on DNA sequencing. Determination of HLA alleles is warranted as HLA alleles are major genetic risk factors in autoimmune diseases and are matched in transplantation. Here, we compared the accuracy of several published HLA-typing algorithms that are based on next-generation sequencing (NGS) data. As genome sequencing is becoming increasingly common in research, we wanted to test how well HLA alleles can be deduced from genome data produced in studies with objectives other than HLA typing and in platforms not especially designed for HLA typing. The accuracies were assessed using datasets consisting of NGS data produced using an in-house sequencing platform, including the full 4 Mbp HLA segment, from 94 stem cell transplantation patients and exome sequences from 63 samples of the 1000 Genomes collection. In the patient dataset, none of the software gave perfect results for all the samples and genes when programs were used with the default settings. However, we found that ensemble prediction of the results or modifications of the settings could be used to improve accuracy. For the exome-only data, most of the algorithms did not perform very well. The results indicate that the use of these algorithms for accurate HLA allele determination is not straightforward when based on NGS data not especially targeted to the HLA typing and their accurate use requires HLA expertise.

**Keywords:** genetic variation, histocompatibility, human leukocyte antigen alleles, genome sequence, transplantation

## INTRODUCTION

Successful transplantation of solid organs and stem cells requires good immunogenetic matching. The matching of alleles in human leukocyte antigen (HLA) genes is particularly crucial (1, 2). All HLA genes are located in the MHC gene complex on chromosome 6p21.3 (3). The most important transplantation HLA molecules, HLA-A, -B, -C, -DR, -DQ, and -DP, are integral membrane proteins, which bind short peptides that are recognized by T cells of the cellular immune system. HLA molecules are the starting point of the adaptive immune response, making them interesting not only

in the normal immune response but also in the susceptibility to autoimmune diseases and tissue and cell transplantation matching. The hallmark of the HLA genes is their very extensive genetic polymorphism. More than 17,000 alleles have been identified in the HLA genes in the IPD-IMGT/HLA database (4). Their determination is challenging due to the extremely high number of alleles, closely related gene sequences between HLA genes, and similar or even identical gene segments shared by some alleles or genes.

Currently, PCR-based molecular techniques, including Sanger sequencing, are the gold standard for HLA allele determination (“HLA typing”). The sequencing using commercial HLA typing kits usually includes the whole exons 2–4 for the class I genes HLA-A, -B, and -C, most of exon 2 for the class II HLA gene DRB1, the whole exon 2 for HLA-DPB1, and exons 2 and 3 for HLA-DQB1. Methods based on next-generation sequencing (NGS) techniques are also emerging. For the NGS-HLA typing, a subset of exons, the whole HLA genes or the whole MHC region may be targeted (5–10). Several commercial HLA kits based on different NGS techniques are already available. These targeted panels specifically designed for HLA allele determination together with long range PCR enable high accuracy HLA typing at the allele resolution level for clinical transplantation use.

Genomic sequencing projects may encompass the HLA gene region, hence producing data of which HLA alleles may be deduced. The data may originate from different genomic sources depending on the targeting method utilized. For example, a typical target may be the whole exome or a set of genes. NGS can also be performed without targeting, as is done in whole-genome or transcriptome sequencing. Even though HLA allele determination in many research projects is not the main objective, the projects certainly would benefit from a valid interpretation of HLA alleles based on the sequence produced.

Most of the widely used high-throughput NGS instruments produce only short reads, from approximately 100 to a few hundred nucleotides per read. This, together with the lack of information on which of the chromosomes or highly similar genes a particular read originates from, causes problems in the bioinformatics analyses of NGS data. These problems are particularly substantial in the case of the highly polymorphic HLA genes, which can differ from each other only by a single nucleotide. In addition, different genes may share highly homologous sequences, leading to problems in the short-read alignment. NGS technologies producing longer read lengths have the potential to solve many of these problems.

As algorithms or computer programs are a key factor in the utilization of NGS data, we selected a set of publicly available programs for the determination of HLA alleles based on NGS data and tested their accuracies using two datasets with different characteristics. One dataset included the whole-genomic MHC region as part of an in-house sequencing platform (11) in 94 patient samples, while the other, a set of samples from the 1000 Genomes catalog (12), targeted primarily exons. The in-house sequencing platform primarily focuses on identification of rare variants in immunologically relevant regulatory areas but also includes the full genomic sequencing of the MHC region, which, however, was not optimized for HLA allele determination. We

selected both assembly based and alignment-based methods, which increases methodological diversity. In addition, we used an ensemble approach (13) to test whether accuracy could be increased by combining results from different programs. The classifiers that are part of an ensemble should be diverse enough so that they do not all produce the same erroneous result. Hence, it was essential to select programs that applied different approaches, either assembly based or alignment based. We also describe some modifications to the default instructions to achieve more reliable results with one of the programs. While none of the programs, when used alone or according to the standard default instructions gave perfect allele assignments, our results show that the ensemble approach produced better results.

## MATERIALS AND METHODS

### Finnish Red Cross Blood Service (FRCBS) Dataset

A total of 94 samples from patients who were HLA typed for possible hematopoietic stem cell transplantation formed the FRCBS dataset. This study was carried out in accordance with the recommendations of the Ethical Review Board of Helsinki University Hospital with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethical Review Board of Helsinki University Hospital.

Clinical HLA typing was performed from 2003 to 2008 in the HLA laboratory of the FRCBS, using methods accredited by the European Federation for Immunogenetics. Four different techniques were used. For both low- and medium-resolution HLA typing, LIPA (Innogenetics Group, Gent, Belgium), rSSO-Luminex Technology (Labtype, One Lambda Inc., Canoga Park, CA, USA), and PCR-SSP (Micro SSP™ Generic HLA Class I/II DNA Typing Trays, One Lambda Inc., Canoga Park, CA, USA; Olerup SSP® genotyping, Olerup SSP AB, Stockholm, Sweden) were used. The results were analyzed with the appropriate software provided by the manufacturers. Sequence-based typing for determining the medium-resolution HLA alleles was performed with AlleleSEQR PCR/Sequencing kits (Atria Genetics, Hayward, CA, USA), using the ABI 3130xl genetic analyzer (Applied Biosystems, Thermo Fisher Scientific, MA, USA), and the results were analyzed with the Assign 3.5+ software (Conexio Genomics Pty Ltd., Fremantle, Australia) according to the supplier's instructions. For the high-resolution HLA typing, combinations of techniques described earlier were used.

Next-generation sequencing was performed using Roche SeqCap EZ Human MHC Design capture, which captures approximately 95% of the MHC/HLA region (11). The sequencing was performed with Illumina HiSeq 2000, yielding 100 bp paired-end reads and a median on-target coverage of 27.5× per sample. The data were quality checked using FastQC, and adapters were trimmed using Cutadapt (14).

### 1000 Genomes Dataset

We used data of 63 samples from the 1000 Genomes project (12). The samples represented three different ethnic groups: 33

Finnish, 15 Puerto Ricans, and 15 Yoruba. HLA typing had previously been carried out at the intermediate-resolution level using Sanger sequencing for the HLA-A, -B, -C, -DRB1, and -DQB1 genes (15), hence, these genes were included in the test. The samples were not selected based on the date of sequencing, depth of sequencing, or any other quality measure as our purpose was to test the methods on data produced using the standard, exome sequencing.

## Programs for Determination of HLA Alleles from NGS Data

The programs can be categorized into read-mapping and assembly based approaches. Instead of using a single reference, as is usually done for NGS data, the read-mapping methods are often based on the alignment of NGS reads with a reference sequence set consisting of all known HLA alleles. Then, the HLA alleles of the sample are predicted based on the properties of the alignments, often summarized as a likelihood or probability score formed by, for example, the number of mapped reads and the overall quality of the alignment. The assembly based methods first construct or “assemble” the NGS reads into larger contigs, which are then queried against a reference database containing all known HLA allele sequences. The reference sequences and allele names are derived from the IPD-IMGT/HLA database (4).

Below, we briefly describe the methods selected for comparison in this study, but refer the reader to the references for further information on the programs.

All the methods were installed and used as instructed in their manuals.

### ATHLATES

ATHLATES (16), version 1.0, is an assembly based method developed for use with exome sequencing data. ATHLATES first filters the sequencing reads by aligning them to HLA allele sequences (either gDNA or cDNA) obtained from the IPD-IMGT/HLA database, allowing for soft clipping to include intron–exon spanning reads. Reads mapping to more than one HLA gene (e.g., HLA-A and non-HLA-A) is excluded. Paired-end reads are merged and reads with potential sequencing errors (low frequency *k*-mers) are discarded. Contig assembly is initiated by using each paired-end read as a contig. Contigs are then merged by considering those sharing an *l*-mer and *l* is decreased iteratively from the full read length until a fixed threshold is met. We kept track of the frequency at each base all the time. Contigs sharing longer and more high-frequency substrings are prioritized in comparisons and merging.

The exons of each HLA allele are then matched with the assembled contigs and the overall difference (Hamming distance) is calculated as the sum of the differences of each exon in the allele. Only alleles with no more than two mismatches and adequate coverage (20× read coverage, minimum of 85% of exon sequence covered by best-hit contigs, more than 70% of cDNA length captured by summed exon lengths) are considered further.

A list of candidate alleles is then formed by selecting those with no missed exons and no more than one mismatch. When the alleles correspond to multiple protein coding sequences, no

mismatches are allowed. Pairs of alleles are formed from this list, scores are calculated for each pair (using a scoring scheme based on multiple sequence alignment), and, finally, the pair(s) with the best score is reported.

ATHLATES was installed and used as instructed in the ATHLATES User Manual 1.0. Novoalign V3.03.00 was used for alignment using the parameters recommended in the ATHLATES manual and for defining the fragment lengths according to the input data.

From the typing results, we selected only allele pairs listed under “Inferred Allelic Pairs.” As ATHLATES often reports more than one possible allele pair, the list was reduced to arrive at only one allele pair to make the comparison with other typing methods more unbiased. The selection was carried out by allowing the listed alleles to vote and then selecting the highest shared level, e.g., alleles A\*11:01:01:01 and A\*11:01:01:02 would result in A\*11:01:01. The call was considered to be empty if low-resolution assignments failed to achieve a higher vote than the higher ones. If the files had no calls at all, they were interpreted as having a missing allele.

ATHLATES does not report only a single allele pair, but a list of the best allele pairs. To keep the comparison fair to other programs, a voting scheme similar to the one described below was used to select the most likely alleles. Selection of the most likely allele pair could also be performed, for example, based on population frequencies of alleles, but this approach was not utilized here.

### HLAssign

HLAssign was developed (mainly) for HLA typing using specific targeted capturing with baits that was designed to consider the highly polymorphic sequences in the HLA region (16). The Windows 64-bit version 14.04 with database IPD-IMGT/HLA v 3.21.0 was used. HLAssign works by mapping all the reads to cDNA sequences from the IPD-IMGT/HLA database (4). It then discards those that are not completely covered or only have coverage on a small, central portion of the read. Several different parameters/statistics are then calculated for all the allele pairs formed from the remaining alleles. The allele pair with the highest weighted harmonic mean of the scaled parameters is selected as the most likely allele pair. This is the allele pair we used for comparisons.

### OptiType

OptiType (17), version 1.3.1, is a mapping-based method and can produce typings from both DNA and RNA sequencing data. It also aims at utilizing multi-mapping reads, whereas, in many other programs, these are discarded. OptiType gives results only for major class I genes (A, B, and C). Even though in principle it should be possible to modify the program to also type class II genes, we did not carry out such modifications.

OptiType has its own HLA allele reference database, that is, constructed from exons 2 and 3 of class I HLA alleles from the IPD-IMGT/HLA database constituting an intermediate-resolution reference data (4). It also includes the intervening intronic sequences (using special phylogenetic-based imputation if intron sequences are missing) for use with DNA sequencing

data. Non-classical HLA class I genes HLA G, H, and J are also included. Sequencing reads are first aligned against the database, allowing multiple matches per read. A binary matrix is then created, indicating which alleles best align to each read. Using this matrix and integer linear programming, the best allele combination is selected by maximizing the number of reads mapping to each gene.

### HLAreporter

HLAreporter version 1.03 (18) begins by filtering the NGS reads, which is done by mapping them to a reference sequence panel. This reference set is at intermediate level of resolution consisting of exons 2–4 for class I HLA genes and exons 2 and 3 for class II HLA genes. The sequences are obtained from the IPD-IMGT/HLA database (4). The panel is further modified by appending 50 bp of intronic sequence to both ends of each exon, including non-classical HLA class I genes. NGS reads that receive no mapping results or map perfectly to more than one gene are filtered away. The remaining set of reads for each gene is then used to assemble contigs. Then, to match the assembled contigs, two more databases are used: one with only exons 2 and 3 for class I and exon 2 for class II genes, and the other one with sequences for less polymorphic exons (exon 4 for class I and exon 3 for class II genes). Matching to the first database allows identification of the alleles, but often only at the G-group level, while the second database is used to further break down the result. Contig-HLA allele matching is performed by calculating scores for each contig (as a product of contig size, average depth of coverage, and percentage of exonic sequence) as well as counting the score for an allele by summing the scores of the contigs supporting this allele. The program only accepts perfect matches between the assembled contigs and the candidate HLA alleles or allele groups. The best scoring of those are reported as the result.

A list of possible alleles may be produced, but there can be several possible alleles listed for both allele pairs. In the same way as for ATHLATES we used voting to reduce the lists to a single allele pair. If the voting gave no consensus, the gene call was marked as missing.

### Omixon Target

This program, version 1.93, was the only commercially available program. Only a limited amount of information appears to be public and can be found on the www site of the producer www.omixon.com, where a reference to Ref. (19) is given. The algorithm (19) works by aligning reads against the IPD-IMGT/HLA sequences with certain constraints. It then scores the alleles and reports the most likely allele pair. However, the exact functioning of the program might differ from that described in the publication. As the Omixon Target program was no longer available by the end of year 2016 we also tested its updated version Omixon Explore (Version 1.0).

### Ensemble Prediction

We used majority voting to combine the results from different typing programs into an ensemble prediction. Because of the multilevel nature of HLA alleles, there are situations where a gene is voted as two alleles, for example, 01:01:01 and 01:01:02. In this

case, the resulting majority vote would be 01:01, i.e., the most detailed level up to which the majority of the voting methods agree. To find such majority-voted alleles, we used a tree to capture all the alleles and their votes. Let this majority voting tree be called *V*. The root node of *V* is an empty node, meaning no typing result (which can occur if there is no consensus on any of the alleles).

For each voted allele, a temporary voting tree is formed so that the level of detail in typing increases toward the child nodes. For example, allele 01:01:02 would result in a tree 01 → 01 → 02. Each node also tracks the given vote (by default one). Such temporary trees are then added under the root node of *V* so that new nodes are appended and, for the existing nodes, the vote count is incremented. Each node of *V* also keeps track of the alleles that have contributed to its votes.

For each sample and gene, all typing methods/programs are given two votes each, since the programs (typically) return an allele pair. However, for some genes there are typing results from only a subset of the programs (e.g., OptiType includes HLA-A, -B, and -C genes only). Therefore, there is not always the same number of voting programs for each gene. If a typing program returns an ambiguous result (for both or either allele), such results are skipped and not used in voting. Some programs might also return missing calls meaning that the missing typing results are considered as “evidence” for the allele being absent. Thus, they contribute a vote for a missing allele, even though such results might also be indicative of various typing problems (related to sequencing or algorithms).

Selection of the majority-voted alleles is done by traversing *V* from the root node toward child nodes, if there is at least one child node whose vote reaches the required threshold (i.e., more than half of all the votes). If more than one child node has the same (highest) number of votes, the child can be selected randomly. When it is not possible to go any deeper (due to going below the voting threshold or being at a leaf node), the allele group represented by this node is selected, and the votes of all alleles that contributed to this node are subtracted from *V*. This selection is done as long as there are nodes that have received more than half of all the possible votes. If there are no nodes with sufficient votes to start with, the result is deemed ambiguous.

Depending on the threshold, the voting result can range from no alleles to several alleles. Varying the threshold can be used to control the number of returned alleles. However, it is not always possible to get, for example, just a pair of alleles with unique protein sequence as the result. This is due to possible ties in voting. In this case, either one or three alleles may be given as the result for any gene. In our comparisons, we used a threshold of 0.5. In the event of more than two options per gene, the genes were investigated to see if there were alleles that could be combined at the level of the accuracy of the reference typing. As the voting does not necessarily return only one pair per gene, varying numbers of allele comparisons may have to be performed. For example, if for a given sample the gene voting resulted in three alleles or allele groups, one of which could not be matched to the reference typing, the unmatched one is counted as an error and thus the result is three allele comparisons.

## G Groups

Some programs, such as the HLAreporter, do not return allele pairs, but instead report G groups. A G group for a gene consists of the alleles with identical nucleotide sequences in exons 2 and 3 (for HLA class I genes) or exon 2 (for HLA class II genes). A single G group can contain tens of different alternative alleles, reflecting uncertainty in the typing. For such results, a temporary G-group tree is formed consisting of all the alleles in the G group and each node (also parent nodes) have weights of one. For example, for G group A\*02:11:01G both alleles 02:11:01 and 02:69 would get one vote. The low-resolution assignment A\*02 gets only one vote as well even though it is shared by both alleles in the G group. The G-group tree is then added to the actual voting tree V. When selecting a majority-voted allele, the whole G-group tree is again subtracted from the voting tree if the majority-voted allele was voted by this G group.

## RESULTS

We tested four publicly available programs (HLAssign, HLAreporter, ATHLATES, and Optitype) and one commercial (Omixon Target) program package for their accuracy in the interpretation of HLA alleles based on genomic NGS sequence data. Two datasets were tested. First, the FRCBS dataset ( $N = 94$  samples of apparent Finnish origin) was produced using a sequencing platform (11) that has been developed to screen rare DNA variants in immunologically relevant gene regions and has not been optimized for HLA typing. The platform uses the Roche SeqCap EZ Human MHC Design capture and basically produces the full genomic sequence of the 4 Mbp HLA segment, with nearly 100% coverage of exons of the HLA genes. Sequencing depths of HLA genes, as calculated by ATHLATES, were as follows: HLA-A 14-138x; HLA-B 13-99x; HLA-C 14-165x; HLA-DRB1 22-216x;

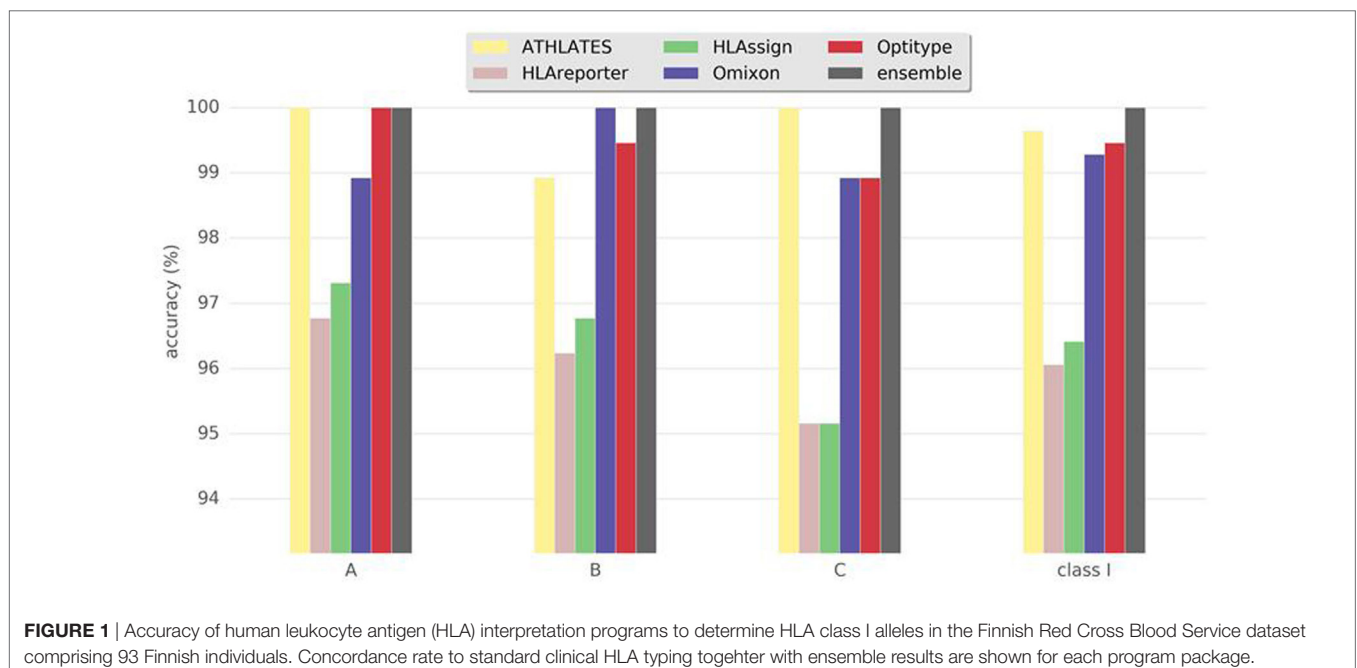
HLA-DQA1 14-191x; HLA-DQB1 15-165x; and HLA-DPB1 15-152x. The FRCBS dataset had been previously typed using standard EFI-accredited clinical methods. Second, to test the exome-only data, we selected 63 samples from the 1000 Genome dataset; the samples were of Finnish ( $N = 33$  samples), Puerto Rican ( $N = 15$  samples), and Yoruba ( $N = 15$  samples) origins. The 1000 Genome dataset covered only the exons of HLA genes, rather than the entire genomic sequence of MHC.

## FRCBS Data

The sequencing datasets were first trimmed for adapter sequences and nucleotides of poor quality using Cutadapt (14). All typing programs were run as described in Section Materials and Methods. As one of the 94 samples obviously showed a loss of heterozygosity (LOH) (see below for details for this sample), the comparisons were done with only 93 samples. The HLA results assigned by each software tested and the clinical HLA typing are shown in Table S1 in Supplementary Material.

Not all the programs were intended to be used for the typing of all HLA genes. Hence, the comparisons were divided into HLA class I and II genes. All the programs were able to give typing results for HLA-A, -B, and -C genes. We calculated the concordance between the results from each program and the reference, clinical HLA result (Figure 1; Table 1) for each of the HLA class I genes separately, as well as for class I combined. Overall, the best performing programs were ATHLATES, Omixon, and OptiType, which all gave the same alleles as the reference typing in 99–100% of cases. The other two programs gave concordant results in 96–97% of the cases. The ensemble prediction was beneficial as 100% concordance with the reference typing was achieved with this method for all the three HLA class I genes.

Figure 2 and Table 1 show the results for class II genes. It is noteworthy that OptiType did not return any results for this



set of genes; therefore, OptiType was not included in the final comparison. Even though HLAreporter was not among the best performing programs for class I genes, it performed best for class II genes, with a 98% concordance with the reference typing. Omixon also achieved an excellent result with a concordance rate of over 96%. It is noteworthy that ATHLATES gave a poor performance for the HLA-DPB1 gene. The program is clearly not designed for DPB1 typing using the default settings and instructions. However, as shown later, its performance can be enhanced significantly with certain modifications. HLA-DQB1 seemed to present a challenge for HLAssign, with a concordance rate of only approximately 80%. We did not investigate this further. The ensemble result for class II gave an excellent outcome with a 99% concordance rate. Four HLA-DRB3-5 alleles were not

concordant. An HLA-DRB5\*02:06; 02:02:01 heterozygosity was suggested in a sample with only one DRB5-associated class II haplotype, and an extra DRB4\*03:01N was proposed in three samples with HLA-DRB1\*01:15 haplotypes. It should be noted that a discordant result is not to be interpreted as incorrect, as it only implies discordance to the reference result.

Since not all clinical HLA typings were performed to the same resolution level, we next tested whether the concordance rates would change if only those cases with a unique amino acid sequence assignments were included in the analyses. For HLA class I, the accuracies were in general slightly lower, except for ATHLATES, which improved its accuracy. The differences seen in class II gene results were negligible (data not shown).

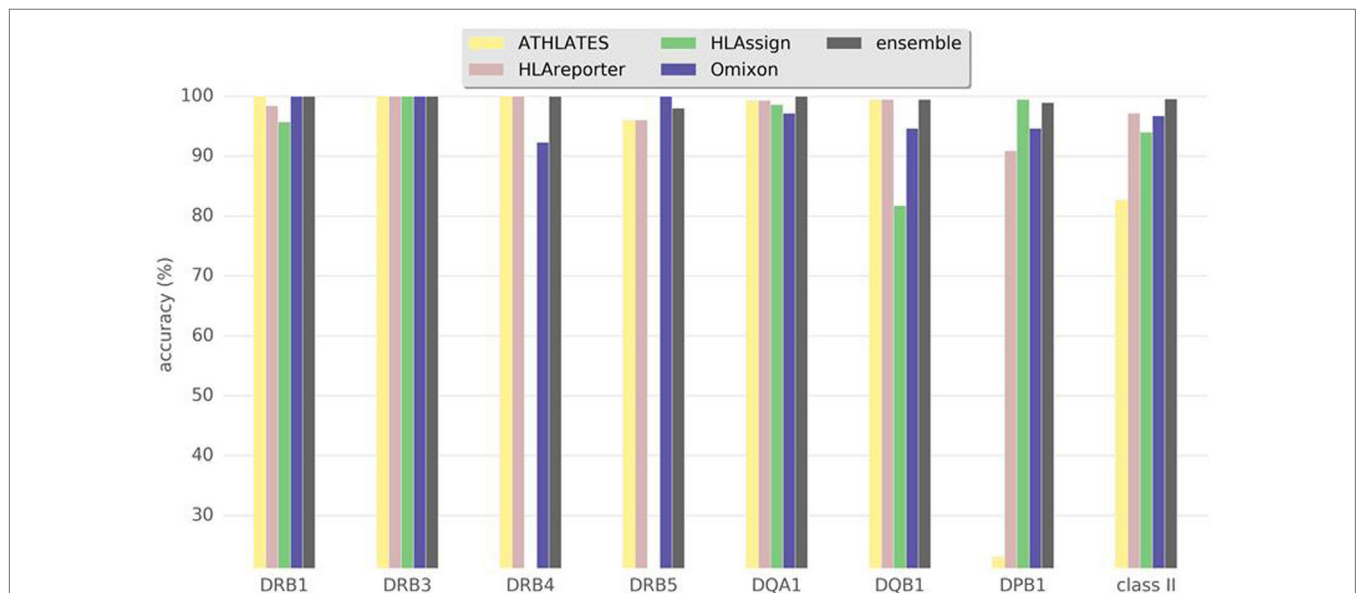
The discordant results varied between the different software. However, some samples seemed to be problematic in particular loci as the majority vote gave more than two possible assignments or the consensus could be drawn only at the one-field resolution level (i.e., low-resolution assignment). As an example, there were three samples with a deviant majority vote result and a discordant HLA assignment with Omixon Target (Table 2). The discordant assignment could not be explained by low coverage or sequencing depth, as both alleles had an average coverage range of 97–100 and average sequencing depth range of 118–322. Furthermore, two out of three discrepancies disappeared after analyzing the data with an updated version of the Omixon software, Omixon Explore (samples FRC13 and FRC37). In the only remaining discrepant sample (FRC36), the average sequencing depth of HLA-DQB1\*03:22 was lower than that of HLA-DQB1\*03:01:01 (118 versus 173, respectively). The full sequence of DQB1\*03:22 allele is not known.

Among the 94 Finnish samples, one was found to have an LOH in the original clinical typings. Hence, it was not included in the comparison. In the clinical typings, three separate samples from

**TABLE 1** | Summary of accuracies of the HLA interpretation programs and the ensemble result based on Finnish Red Cross Blood Service dataset comprising 93 samples.

	Program	Concordant result	Different	Accuracy %
Class I	HLAssign	538	20	96.42
	HLAreporter	533	25	95.52
	ATHLATES	556	2	99.64
	Optitype	555	3	99.46
	Omixon Target	554	3	99.28
	Ensemble	558	0	100
Class II	HLAssign	703	45	93.98
	HLAreporter	836	15	98.24
	ATHLATES	703	147	82.24
	Optitype	na	na	na
	Omixon Target	822	28	96.71
	Ensemble	848	4	99.53

Concordance rates to standard clinical human leukocyte antigen (HLA) typing are shown.



**FIGURE 2** | Accuracy of human leukocyte antigen (HLA) interpretation programs to type HLA class II alleles in the Finnish Red Cross Blood Service dataset comprising 93 Finnish individuals. Concordance rate to standard clinical HLA typing together with ensemble results are shown for each program package.

**TABLE 2** | Sequencing depth and coverage in three samples with discrepant results in human leukocyte antigen (HLA) interpretation by the Omixon Explore software.

	Exons covered	Depth range	Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	Exon 7	Omixon Target	Omixon Explore	Reference method
<b>All samples HLA-DQB1</b>	5	3–279	37–305	3–105	51–446	40–307	155–413	na	na			
<b>FRC13</b>												
HLA-DQB1 allele 1	4	Average 134	167	89	148	178	–	na	na	02:01	02:01	02:01
HLA-DQB1 allele 2	4	Average 138	167	89	148	178	–	na	na	02:06	02:02	02:02
<b>FRC36</b>												
HLA-DQB1 allele 1	4	Average 173	189	77	245	205	–	na	na	03:01	03:01	03:01
HLA-DQB1 allele 2	2	Average 118	–	77	158	–	–	na	na	03:22	03:22	03:01
<b>All samples HLA-A</b>	7	16–322	4–206	5–239	4–305	4–463	12–426	1–389	1–366			
<b>FRC37</b>												
HLA-A allele 1	6	Average 322	206	239	305	463	426	–	1	02:01	02:01	02:01
HLA-A allele 2	6	Average 322	206	239	305	463	426	–	1	02:197	02:01	02:01

Samples FRC13 and FRC36 had an HLA-DQB1 discrepancy and sample FRC37 had an HLA-A discrepancy.

the individual were required to confirm the heterozygosity in the HLA-C locus. Obviously, higher proportions of malignant cells with an LOH hampered HLA determination. The NGS programs suggested homozygous HLA-C assignment, which was also suggested by the majority vote. It is noteworthy that Omixon correctly suggested an unbalanced result in HLA-C locus, indicating a minor fraction of reads from another putative allele. We have recently reported (20) a systematic screening of LOH in patients waiting for stem cell transplantation due to hematological malignancies. The sample from this study was not included in the study.

## Modified ATHLATES

Improvements in the ATHLATES results could be achieved by using the Mosaik v2.2.3 (21) aligner as described in the ATHLATES user manual 1.0 with the addition of the `-om` option to redirect multiple mapped reads to a separate bam file. When these reads were excluded from the HLA typing phase substantially better results were obtained (details not shown), demonstrating that minor modifications to the programs may result in more accurate performance.

## 1000 Genomes Data

As exome sequencing is probably the most widely used method for genomic sequencing, we tested the ability of the programs to determine HLA alleles from such non-MHC or non-HLA targeted data. We used 63 samples from the 1000 Genomes Project for which HLA alleles had been determined using Sanger sequencing (15). The selected samples were from three different ethnic groups (33 Finnish, 15 Puerto Ricans, and 15 Yoruba), and for each sample, the HLA types for five genes (HLA-A, -B, -C, -DQB1, and -DRB1) were available. Samples were not selected based on the date of sequencing, depth of sequencing, or any other quality measure as our purpose was to test the methods on standard, exome sequenced samples.

HLAssign was tested on several exome datasets, but it failed to produce calls, except for just a few genes in some samples. Thus, HLAssign was excluded from further analyses with the

1000 Genomes Project data. In addition, we did not test the commercial Omixon Target program because it was no longer available but was replaced by the new software, Omixon Explore. Its performance with 1000 Genome samples has been published (19).

For both HLAreporter and ATHLATES, we used voting to get a single best allele or allele group pair for each gene. If this did not produce a consensus, the allele call was marked as “no call” and was counted as a typing error. If voting gave an HLA assignment at the one-field resolution level, it was also marked as “no call.” Missing calls were also counted as errors unless the reference typings indicated a missing allele as well, although there were no missing calls for any of the typed genes in the 1000 Genomes set.

Results for the comparisons are shown in **Table 3**. To estimate the relative accuracy between the programs for the cases where data quality and/or coverage were not a limiting factor, we identified the intersection of genes for which the different programs returned results and counted the accuracies for the resulting set of alleles. Both HLAreporter and ATHLATES missed a large fraction of the alleles, whereas OptiType gave calls for all the alleles, with an accuracy of over 98%. This difference may be related to the fact that HLAreporter requires higher read coverage than was available. In addition, the assembly based methods cannot cope with sequence gaps, whereas gaps present no serious problem for the mapping-based OptiType. One factor explaining the HLAreporter discrepancy between the accuracies presented here and those achieved by Huang et al. (18) may be that we did not select only the good quality samples from the 1000 Genomes collection.

## DISCUSSION

A major finding in this article was that none of the programs alone gave perfect assignments for all HLA genes based on sequencing platform that was not specifically designed for HLA typing but for other research purposes. In fact, some of the programs were clearly unsuitable for use with certain genes, particularly when

**TABLE 3** | Summary of accuracies of the human leukocyte antigen (HLA) interpretation programs for 1000 Genomes samples.

Program	Number of alleles	Correct (all alleles)	Accuracy (all alleles) (%)	Missing alleles (%)	Correct (intersection $N = 164$ )	Accuracy (intersection) (%)
HLAreporter	630	202	32.06	190 (30.16)	91	55.49
ATHLATES	630	468	74.29	78 (12.38)	137	83.54
Optitype	378	372	98.41	0 (0.00)	161	98.17

Missing allele refers to either a completely missing call or one for which no consensus could be formed based on a list of detected alleles. Intersection denotes the set of alleles for all sample-gene pairs that had a non-missing typing in all tested typing programs. No results could be achieved using HLAssign.

tested using the 1000 Genome data containing only exomes. This may be related to the relatively low sequencing depth of the 1000 Genome samples as compared with the depth needed for clinical HLA typing. Second, none of the programs consistently outperformed the other programs in all the cases and genes. Therefore, it was not possible to conclude the best performing program. Finally, with the ensemble prediction method, the accuracy improved, reaching almost 100% concordance with the reference. However, it is noteworthy that the ensemble approach is not very practical in clinical settings, but it may be used for research purposes to get the best possible HLA determination. In cases with discrepancy in typing results, we could not conclude which of the allelic alternatives was the correct one, as it was no longer possible for us to have genuine control alleles by performing clinical grade NGS sequencing of these samples. Reliable HLA results can hence be achieved by using several programs and by applying the ensemble approach in the research context where the genome data are obtained without HLA genes being the primary analysis target.

The ensemble approach, however, has its drawbacks. All methods returning the same allele may lull us into a false sense of security regarding the allele call. It should be noted that even a unanimous result is not necessarily the right one, but can, for example, be caused by problems in sequencing or in the reference allele database. Use of sequencing methods that are based on short sequencing reads results in sequence alignment problems when applied to allele interpretation of the highly homologous HLA alleles and genes. Also the data with short reads and low read depth are very challenging for the highly polymorphic genes, such as HLA genes, resulting in ambiguous HLA assignments. This has been solved in the commercial HLA typing kits where read lengths are several hundred nucleotides and the sequencing depth is usually several thousands. We also noted that some of the problems with accuracy were associated with the properties of the software. A good example is shown in **Table 2** in which the updated version of the Omixon Explore program could resolve two of the three discrepancies found using the older version Omixon Target. Furthermore, this example shows the problems related to partial sequences known of some alleles: only exons 2 and 3 sequences of DQB1\*03:22 are known. We were also able to show that the ATHLATES program could be adjusted to give much more reliable results by using modifiers as described by Lee et al. (21). We assume that similar modifications could be done to other programs as well.

The results presented here for some of the programs were likely worse than the results that might be achieved by an experienced

HLA professional looking at the list of reported alleles and making interpretations based on the population allele or haplotype frequencies. Instead, we used a simple minded approach and compared the programs as instructed by their manuals and applied default parameters. In addition, in cases where a program gave multiple alternatives with equal scores, we did not utilize the reference typing or population frequencies to determine the apparently correct alternative. In the real research project, it certainly is wise to do some interpretation based on prior population data.

Some differences in the relative performances of the programs might be due to the various versions of the IPD-IMGT/HLA database they use. However, we did not try to unify the versions since the programs use their own, slightly modified format of the database and because the programs were all released within a short period. The IPD-IMGT/HLA database is an obvious choice for reference. However, the sequences therein have been found to contain some errors (8), and we do not know whether some of the programs utilize the corrected sequences.

Determination of HLA class I alleles from the 1000 Genomes exome NGS data has been reported earlier (19). They reported an accuracy of over 90% between the NGS and Sanger data. In the case of the 1000 Genomes samples, it is notable that in our study most of the tested programs returned a significant number of missing results. A missing result may indicate that the gene was missing or that the quality of the data was insufficient for the program. The latter is a known problem because the targeting baits for exome sequencing are usually designed for the human reference genome, ignoring the very high variability in the HLA region (8). However, OptiType retained a very high accuracy even for exome-only data. Therefore, exons 2 and 3 for HLA class I genes were likely well captured in the exome sequences of the 1000 Genomes Project. Alternatively, even though the exons might not be fully captured, OptiType is still able to make the most of the data, whereas some other programs had filtering criteria that could not be met when the exons were sequenced only partly. One such program is HLAssign where the cDNA allele sequences need to be fully covered in the alignment.

The results of this study clearly indicate that selecting a program for HLA allele determination based on NGS data that were not designed for HLA typing purposes is not simple. This process requires a good understanding of the type of NGS data produced and the HLA allele frequencies in the study population. Some modifications in the programs, the adoption of an ensemble approach or testing with multiple programs may be needed for accurate performance.



## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Ethical Review Board of Helsinki University Hospital with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethical Review Board of Helsinki University Hospital.

## AUTHOR CONTRIBUTIONS

AL, RE, EK, TK, and TP produced the genomic NGS data and performed the bioinformatics analyses; AL, SK, and JP planned the study and wrote the manuscript.

## REFERENCES

- Süsal C, Opelz G. Current role of human leukocyte antigen matching in kidney transplantation. *Curr Opin Organ Transplant* (2013) 18:438–44. doi:10.1097/MOT.0b013e3283636ddf
- Charron D, Petersdorf E. The HLA system in hematopoietic stem cell transplantation. 1st ed. In: Socié G, Blazar B, editors. *Immune Biology of Allogeneic Hematopoietic Stem Cell Transplantation*. London: Elsevier (2013). p. 19–38.
- The MHC Sequencing Consortium. Complete sequence and gene map of a human major histocompatibility complex. *Nature* (1999) 401:921–3. doi:10.1038/44853
- Robinson J, Halliwell JA, McWilliam H, Lopez R, Parham P, Marsh SGE. The IMGT/HLA database. *Nucleic Acids Res* (2013) 41:D1222–7. doi:10.1093/nar/gks949
- Erlich HA. HLA typing using next generation sequencing: an overview. *Hum Immunol* (2015) 76:887–905. doi:10.1016/j.humimm.2015.03.001
- Wang C, Krishnakumar S, Wilhelm J, Babrzadeh F, Stepanyan L, Su LF, et al. High-throughput, high-fidelity HLA genotyping with deep sequencing. *Proc Natl Acad Sci U S A* (2014) 109:8676–81. doi:10.1073/pnas.1206614109
- Gabriel C, Fürst D, Fae I, Wenda S, Zollikofer C, Mytilineos J, et al. HLA typing by next-generation sequencing – getting closer to reality. *Tissue Antigens* (2014) 83:65–75. doi:10.1111/tan.12298
- Wittig M, Anmarkrud JA, Kässens JC, Koch S, Forster M, Ellinghaus E, et al. Development of a high-resolution NGS-based HLA-typing and analysis pipeline. *Nucleic Acids Res* (2015) 43:e70. doi:10.1093/nar/gkv184
- Hosomichi K, Shiina T, Tajima A, Inoue I. The impact of next-generation sequencing technologies on HLA research. *J Hum Genet* (2015) 60:665–73. doi:10.1038/jhg.2015.102
- 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) 14:1. doi:10.1186/1471-2164-14-355
- Morin A, Kwan T, Ge B, Letourneau L, Ban M, Tandre K, et al. Immunoseq: the identification of functionally relevant variants through targeted capture and sequencing of active regulatory regions in human immune cells. *BMC Med Genomics* (2016) 9:59. doi:10.1186/s12920-016-0220-7
- Genomes Project Consortium; Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature* (2010) 467:1061–73. doi:10.1038/nature09534

## ACKNOWLEDGMENTS

This study was funded by Tekes, the Finnish Funding Agency for Innovation, as part of the SalWe GetItDone program (grant ID 3982/31/2013) and by Finnish government VTR funding for medical research. The authors want to note that Omixon Ltd gave their commercial Omixon Target program free of charge for this study. Omixon has not participated in the interpretation of the results and has not seen them before publication.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/articles/10.3389/fimmu.2017.01815/full#supplementary-material>.

- Polikar R. Ensemble based systems in decision making. *IEEE Circuits Syst Mag* (2006) 6(3):21–45. doi:10.1109/MCAS.2006.1688199
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* (2011) 17:10–2. doi:10.14806/ej.17.1.200
- Gourraud PA, Khankhanian P, Cereb N, Yang SY, Feolo M, Maiers M, et al. HLA diversity in the genomes dataset. *PLoS One* (2014) 9:e97282. doi:10.1371/journal.pone.0097282
- Liu C, Yang X, Duffy B, Mohanakumar T, Mitra RD, Zody MC, et al. ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res* (2013) 41:e142. doi:10.1093/nar/gkt481
- Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* (2014) 30:3310–6. doi:10.1093/bioinformatics/btu548
- Huang Y, Yang J, Ying D, Zhang Y, Shotelersuk V, Hirankarn N, et al. HLAReporter: a tool for HLA typing from next generation sequencing data. *Genome Med* (2015) 7:25. doi:10.1186/s13073-015-0145-3
- Major E, Rigó K, Hague T, Bérces A, Juhos S. HLA typing from genomes whole genome and whole exome illumina data. *PLoS One* (2013) 8:e78410. doi:10.1371/journal.pone.0078410
- Linjama T, Impola U, Niittyvuopio R, Kuitinen O, Kaare A, Rimpiläinen J, et al. Conflicting HLA assignment by three different methods due to the apparent loss of heterozygosity in the MHC region. *HLA* (2016) 87:350–5. doi:10.1111/tan.12770
- Lee WP, Stromberg MP, Ward A, Steward C, Garrison EP, Marth GT. MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS One* (2014) 9:e90581. doi:10.1371/journal.pone.0090581

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Larjo, Eveleigh, Kilpeläinen, Kwan, Pastinen, Koskela and Partanen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.