Benchmark of tools for in silico prediction of MHC class I and class II genotypes from NGS data

Arne Claeys[1], Jasper Staut[1], Peter Merseburger[1], Kathleen Marchal[2, 3] and Jimmy Van den Eynden[1*]

1. Department of Human Structure and Repair, Anatomy and Embryology Unit, Ghent University, Ghent, Belgium.
2. Department of Information Technology, IDLab, Ghent University, Ghent, Belgium.
3. Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium.

* Author to whom correspondence should be addressed (jimmy.vandeneynden@ugent.be).

Abstract

The HLA genes are a group of highly polymorphic genes that are located in the MHC region on chromosome 6. The HLA genotype affects the presentability of tumour antigens to the immune system. Therefore, methods to accurately type HLA alleles are critical to study differences in immune response between cancer patients. PCR-based methods are the current gold standard, but large-scale datasets with PCR-based HLA genotypes are rarely available. A variety of methods for *in silico* NGS-based HLA genotyping have been developed, bypassing the need to determine these genotypes with separate experiments. However, there is currently no consensus on the best performing tool. Here, we compiled a list of 13 HLA callers and evaluated their accuracy on three different datasets. Based on these results, best-practice guidelines were constructed, and consensus HLA allele predictions were made for DNA and RNA samples from The Cancer Genome Atlas (TCGA).

Keywords

HLA genotyping, benchmark, tumour-immune interaction, The Cancer Genome Atlas

Introduction

The human Major Histocompatibility Complex (MHC) is a gene complex located on the p-arm of chromosome 6 that contains two large clusters of genes with antigen processing and presentation functions: the MHC class I and MHC class II regions [1–3]. MHC class I molecules are involved in the presentation of endogenous antigens to cytotoxic T-cells and consist of a heavy chain encoded by the *HLA-A*, *HLA-B* or *HLA-C* genes, and a light β_2 microglobulin chain [4–6]. Their role in tumour immunity has been established for a long time. Indeed, they can present neoantigens, small mutated peptides, to CD8+ T cells, resulting in an immune response and cancer cell death.

The most frequently studied MHC class II genes include *HLA-DPA1, HLA-DPB1*, *HLA-DQA1, HLA-DQB1*, *HLA-DRA1* and *HLA-DRB1*. They encode alpha and beta heterodimers that form the MHC class II complex. The role of these genes in anti-tumour immunity is emerging [7–9]. MHC class II mediated tumour-immune interaction occurs either via an indirect or a direct mechanism. First, cancer cells can secrete neoantigens that are subsequently taken up and presented on the MHC-II of professional antigen presenting cells infiltrating the tumour [7,10]. Additionally, some tumours express MHC-II themselves and can directly interact with CD4+ T-cells [7,11].

The peptide-binding region of HLA molecules is highly polymorphic and specific HLA alleles determine neoantigen binding and presentation to the immune system. This genetic diversity could lead to differential responses to immunotherapy, as illustrated by the association that has been described between MHC-I genotypes (e.g., *HLA-B62*) and survival in immune checkpoint blockade (ICB)-treated advanced melanoma patients [11]. It is currently unclear whether MHC-II genotypes also determine responses to immunotherapy. Such association studies require knowledge of the HLA genotype. PCR methods are currently the gold standard for this genotyping but datasets with PCR-based HLA calls are rarely available [12–14]. HLA genotyping can also be performed on Next Generation Sequencing (NGS) data. A plethora of tools has been developed for this task. *Polysolver* and *Optitype* are often recommended as the best performing tools for MHC-I genotyping [15]. For MHC-II genotyping there is currently no consensus about the best method. Several benchmarks have been performed previously [13,15–18], but these were either not applied to MHC class II or did not include some recently published tools.

In this study, we compiled a list of 13 tools that predict HLA genotypes from NGS data and benchmarked their performance on both the 1000 genomes dataset and on an independent cell line dataset. Subsequently we indirectly assessed the performance of these tools on 9162 DNA and 9761 RNA sequencing files from The Cancer Genome Atlas (TCGA). Based on these findings, we gave recommendations on which tool to use for a given data type and how the outputs of multiple tools can be combined into a superior consensus prediction.

Results

Selection of 13 HLA genotyping tools with variable computational resource requirements

We identified 22 available HLA genotyping tools from literature (Table 1). Thirteen tools that were free for academic use, applicable on DNA and/or RNA NGS data and ran on our Linux system were included in this study: *arcasHLA*, *HLA-HD*, *HLA-VBSeq*, *HLA*LA*, *HLAforest*, *HLAminer*, *HLAscan*, *Kourami*, *Optitype*, *PHLAT*, *Polysolver*, *seq2HLA* and *xHLA*. All 13 tools can make allele predictions for the three MHC class I genes (*HLA-A*, *HLA-B* and *HLA-C*) and 9 tools support additional calling of the MHC class II genes *HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DRB1*. Two methods support only a subset of the MHC class II genes: *xHLA* does not support calling *HLA-DPA1* and *HLA-DQA1*, while *PHLAT* does not support *HLA-DPA1* and *HLA-DPB1*.

Firstly, the time and memory consumption of the thirteen selected tools were measured on a random subset of 10 DNA and 10 RNA sequencing files from the TCGA project (Figure 1). Among the 10 DNA-based methods *HLAScan* (median 7.2 hours per file), *Optitype* (median 2.45 hours) and *HLA*LA* (median 1.7 hours) are the most time-consuming tools. The remaining DNA tools take less than 1 hour per file, with *HLAminer, Kourami* and *PHLAT* being the fastest (34s, 150s and 262s respectively). Apart from being time-consuming, *HLA*LA* is also the most memory demanding DNA tool (median 36.4 GiB per file). Other DNA tools with a median memory consumption higher than 10 GiB are *xHLA* (median 24.3 GiB), *HLA-HD* (median 15.0 GiB) and *Kourami* (median 9.18 GiB). The relatively low memory consumption of *Polysolver* and *HLAScan* makes it feasible to compensate for their long running time by processing multiple samples in parallel.

Among the 7 RNA-based methods, *HLA-HD* takes the longest time to analyse a sample (median 14.9 hours). Most methods that are compatible with both DNA and RNA data (*HLAminer*, *PHLAT* and *HLA-HD*) take a longer time on RNA data. At the other end of the spectrum, the pseudoalignment based tool *arcasHLA* takes only 40s per file. The most memory intensive tool is *HLA-HD* as well (median

82 memory peaks of 101 GiB), followed by *Optitype* (median 34.1 GiB). The other RNA tools have a
83 memory consumption lower than 10 GiB.

84 *HLA*LA* and *HLA-HD* are the best performing MHC class II genotyping tools on DNA data

85 The 10 selected algorithms that are compatible with DNA sequencing data were benchmarked using
86 WES data from the *1000 Genomes project* [19] (Figure 2). Predictions were made for *HLA-A* (n = 1012),
87 *HLA-B* (n = 1011), *HLA-C* (n = 1010), *HLA-DQB1* (n = 1008), *HLA-DRB1* (n = 1000) and *HLA-DQA1* (n =
88 68). No gold standard calls for *HLA-DPA1* and *HLA-DPB1* were available for this dataset.

89 For MHC-I genes (*HLA-A*, *HLA-B*, *HLA-C*), the best accuracy was obtained with *Optitype* (98.0%),
90 followed by *Polysolver* and *HLA*LA* (94.9% and 94.4% respectively). For MHC-II genes (*HLA-DQA1*,
91 *HLA-DQB1* and *HLA-DRB1*), the best allele predictions were made using *HLA-HD* and *HLA*LA* (96.2%
92 and 95.7% accuracy respectively). These were the only two methods to reach an accuracy of 90% on
93 all tested MHC-II genes. *HLAScan* (74.2%), *HLA-VBSeq* (60.2%) and *HLAminer* (52.8%) performed
94 considerably worse than the other tools.

95 We observed large variabilities in calling accuracies between MHC class II genes. Overall, *HLA-DQB1*
96 was the hardest MHC-II gene to call. Except for *PHLAT*, all tools obtained their worst MHC-II call
97 accuracy on this gene. *HLA-DQA1*, on the other hand, was the gene with the highest calling accuracy
98 for all tools that support it, except for *HLAminer* and *Kourami*.

99 Lower prediction accuracies are either caused by wrong allele calls or a failure to make an allele call.
100 When *HLAScan* and *Kourami* were able to make a call, their predictions were most often reliable
101 (Figure S1), but these tools regularly produced no output at all (Figure S2). *HLA-VBSeq* and *HLAminer*
102 had both a high rate of incorrect and failed calls (Figures S1-S2).

103 Subsequently, we performed an independent benchmark using the smaller NCI-60 cell line dataset
104 (n=58), which largely confirmed our results (Figure S3). Additionally, this analysis indicated that the
105 best performing MHC class II supporting tools also performed well on *HLA-DPB1*, a gene for which no
106 gold standard calls were present in the 1000 genomes dataset.

107 *HLA-HD*, *PHLAT* and *arcasHLA* are the best performing MHC class II genotyping tools on
108 RNA data

109 We then evaluated the 7 selected methods that support HLA calling on RNA sequencing data from the
110 *1000 genomes project* [20] (Figure 2). Predictions were made for *HLA-A* (n = 373), *HLA-B* (n = 372),
111 *HLA-C* (n = 372), *HLA-DQB1* (n = 371), *HLA-DRB1* (n = 362) and *HLA-DQA1* (n = 53). Again, no gold
112 standard calls for *HLA-DPA1* and *HLA-DPB1* were available for this dataset.

113 A*rcasHLA* and *Optitype* had the best MHC-I allele predictions (99.4% and 99.2% accuracy,
114 respectively), followed by *HLA-HD* (98.0%), *seq2HLA* (95.7%) and *PHLAT* (95.4%). Similar accuracies
115 were found for MHC-II allele predictions, with *HLA-HD*, *PHLAT* and *arcasHLA* performing the best
116 (99.4%, 98.9% and 98.1%, respectively). Contrary to its good prediction of MHC class I alleles, *seq2HLA*
117 has a lower accuracy for MHC class II (87.9%).

118 The high MHC-I accuracies of *arcasHLA* and *Optitype* were confirmed on the independent NCI-60
119 dataset (91.8% and 90.0%, respectively; n=58). The accuracy of *HLA-HD*, *PHLAT* and *seq2HLA* was
120 worse on the cell lines than in the benchmark on the 1000 genomes data (86.6%, 83.3% and 82.3%,
121 respectively). As MHC-II is generally not expressed in cell lines, this benchmark was not performed for
122 those genes.

A correlation analysis between observed and expected allele frequencies confirms the benchmarking results.

Being one of the few large sequencing datasets for which gold standard HLA genotypes for both MHC classes are available, many algorithms included in our benchmark are optimized and validated on files from the *1000 genomes* project, introducing a potential bias. Additionally, no gold standard HLA calls are available for *HLA-DPA1* and *HLA-DPB1*. Therefore, we performed an indirect and independent evaluation by comparing the observed allele frequencies for each tool with the expected population frequencies. (Figure 3).

We applied each of the tools on DNA and RNA sequencing data from The Cancer Genome Atlas (TCGA, n=9162 and n=9761 for DNA and RNA respectively) and calculated how often each of the alleles was predicted by a certain tool to obtain an observed allele frequency, stratifying for Caucasian American (n= 7935) and African American (n=938) ethnicities. By comparing these frequencies to the expected allele frequencies, as derived from *Allele Frequency Net* [21], strong correlations (i.e., Pearson's r higher than 0.95 for all genes and both populations) were found for the DNA-based tools *HLA-HD*, *HLA*LA*, *Optitype*, *Polysolver* and *xHLA* and for the RNA-based tools *Optitype*, *arcasHLA* and *PHLAT*. The correlations were considerably worse for *HLA-VBSeq*, *HLAminer* and *HLAforest* than for the other tools (Figure 3). These findings largely confirm the results of the benchmark on the 1000 genomes data.

HLA predictions made by the most accurate tools are complementary

We then examined the complementarity of the tools' predictions. Using a hierarchical clustering approach on the correctness of each tool and sample, we noted that only for a very small fraction of the samples the genotypes are wrongly typed by all tools simultaneously (median 0.79% for DNA and 0.68% for RNA), and that predictions are often complementary (Figures S4-S5).

We then calculated for each pair of tools how often their predictions are concordant (Figures S6-S9). Tools that performed poorly in the previous analyses (e.g., *HLAminer*, *HLA-VBSeq* and *HLAforest*) consistently have a low concordance with all other tools. In contrary, tools that scored high in the previous analyses (such as *Optitype, HLA-LA, arcasHLA* and *HLA-HD*) made predictions that are consistent with each other. Noteworthy, this is also the case for *HLA-DPA1* and *HLA-DPB1*, two genes for which no gold standard data was available, suggesting that predictions for these genes are reliable as well.

A consensus metaclassifier improves HLA predictions for DNA data

The complementarity of the tools' allele predictions opens the possibility to combine predictions of different HLA callers into a consensus prediction.

We applied a majority voting algorithm to the output of all tools, with the predicted allele pair being the one with most votes. On the DNA data, this approach outperforms the predictions of each individual tool for all genes. This is best illustrated by the *HLA-DQB1* gene, where the accuracies increased from 93.2% with the best performing tool (*HLA-LA*) to 96.5% when the voting metaclassifier was used. On RNA data, where the best tools already attain accuracies over 99% by themselves, only minor improvements were made by combining the results (Figure S10).

We then determined the minimal number of tools that must be included in the DNA-based metaclassifier to produce reliable results (Figure 4). For the DNA data, including 4 tools in the model led to a considerable improvement for all genes for both MHC classes. The best accuracies using 4 tools were observed when *Optitype*, *HLA*LA*, *Kourami* and *Polysolver* were combined for MHC-I predictions (99.0% accuracy) and with *HLA*LA*, *HLA-HD*, *PHLAT* and *xHLA* for MHC-II predictions

167 (98.1% accuracy). Raising the number of tools further only resulted in marginal gains. Strikingly, the
168 accuracy of the *HLA-DQB1* allele predictions even decreases when more tools were included in the
169 model. Therefore, we suggest combining the output of 4 tools for both MHC classes.

170 To indirectly evaluate whether the good performance of this approach is generalizable to other
171 datasets, we assessed the correlation between the expected allele frequencies and the allele
172 frequencies observed using the 4-tool DNA consensus predictions on the TCGA dataset and compared
173 the results with our previous findings. The allele frequencies predicted by the metaclassifier correlated
174 better with the expected allele frequencies (Figure 3) than was the case for the individual tools that
175 supported all genes of interest.

176 Discussion

177 Rapid technological advancements in NGS have resulted in the generation of numerous publicly
178 available WES and RNA-Seq datasets. These data have been critical for understanding the genomic
179 basis of human carcinogenesis. In the field of immuno-oncology, the availability of these genomic data
180 with corresponding clinical data opens new possibilities for studying differences in immune response
181 between cancer patients. However, this requires that the HLA genotypes for each subject can be
182 accurately determined. An ever-increasing number of NGS-based HLA typing software applications
183 have been developed. In this study, we benchmarked the performance of 13 publicly available tools.

184 First, we evaluated the tools by comparing their output to genotypes derived from a PCR-based
185 approach. While PCR methods are the gold standard for HLA typing, they have limitations that could
186 lead to ambiguous typing results [22]. Furthermore, inconsistencies have been reported across PCR-
187 based HLA typing datasets that are available for the 1000 genomes samples [23] which could have
188 affected our benchmarking results. Therefore, we also used 2 other, indirect approaches to assess the
189 performance of the different tools.

190 Both a concordance analysis between the tools' predictions and a correlation analysis between
191 predicted and expected allele frequencies confirmed our benchmarking results. To avoid biasing the
192 results of this correlation analysis, we disabled population-specific allele frequencies for the
193 algorithms that support this (i.e., *arcasHLA* and *Polysolver*). However, in the case of *arcasHLA*, when
194 no specific population is specified, it uses prior frequencies that depend on the prevalence of the
195 alleles in the entire human population. A*rcasHLA*'s usage of prior frequencies might hinder its ability
196 to call alleles that are uncommon in the specified population. The correlation analysis between
197 observed and expected allele frequencies revealed that this tool overestimated the frequency of *HLA-*
198 *DRB1*14:02* in a population where this allele was rare. Additionally, we note that *xHLA* and *Polysolver*
199 both required that the BAM data were realigned to a version of the GRCh38 reference genome that
200 does not contain alternative scaffolds. Similarly, realigning the BAM files to GRCh37 was required for
201 use with *HLA-VBSeq*. This requirement for an additional alignment step may have resulted in the loss
202 of relevant reads and hence have negatively impacted the results of these tools.

203 The optimal strategy for HLA genotyping depends on a few factors: the availability of either DNA or
204 RNA data, the size of the dataset that needs to be analysed and the available computational resources.
205 In general, DNA-based methods tend to faster and less resource demanding, while more accurate
206 predictions were made using RNA-based pipelines. For DNA data, *Optitype* and *HLA-HD* are the best
207 performing individual tools for MHC class I and MHC class II typing, respectively. For RNA data, the
208 same tools are recommended when sufficient resources are available. However, the large resource
209 and time consumption of *HLA-HD* makes its usage rather impractical on large datasets. As an

210 alternative, *arcasHLA* can be used, which is both the fastest and one of the most accurate tools for
211 RNA.

212 Finally, we have demonstrated that the accuracy of the DNA-based HLA genotype predictions can be
213 improved further by combining the output of *Optitype*, *HLA*LA*, *Kourami* and *Polysolver* for MHC-I
214 tying and combining *HLA*LA*, *HLA-HD*, *PHLAT* and *xHLA* for MHC-II typing using a majority voting rule.
215 For RNA data a similar approach did not lead to a further improvement of the prediction accuracies.

216

Materials and methods

Selection of tools

A list of existing HLA genotyping tools for NGS data was compiled from literature between October and December 2020. The tools that fulfilled the following criteria were selected for further analysis: the tool should be free for academic use, support either DNA or RNA sequencing data, should not require enrichment of the HLA region before sequencing and should be a Linux command line tool that we could successfully run on our system. When the authors provided instructions on how to update the IPD-IMGT/HLA database used by their tool, this database was updated to version 3.43. This was the case for three tools: *HLA-HD*, *HLAminer*, and *Kourami*.

Next-generation sequencing datasets for benchmark

Slices of the 1012 CRAM files of WES data from the *1000 Genomes on GRCh38* dataset [19] that were used for the benchmark on DNA data were obtained from the *International Genome Sample Resource* using the *samtools view* command. The following contigs were included in the download: the MHC region on the primary assembly (chr6:28509970-33480727), all 525 contigs starting with "HLA-" and all unmapped reads. The sliced BAM files for the RNA benchmark were obtained from the *Geuvadis* [20] RNA-Seq dataset (part of the *1000 genomes* project) via *ArrayExpress* (accession number *E-GEUV-1*). All reads mapped to the MHC region and the unmapped reads were included in the download.

Sequencing data from NCI-60 cell lines [24] were obtained from the *Sequence Read Archive* with accession numbers *SRP150855* (WES) [24] and *SRP133178* (RNA) [25]. The NCI-60 sequencing data was realigned according to the same alignment pipeline used by the *1000 Genomes on GRCh38* dataset [19]: reads were aligned to the complete GRCh38 reference genome, including ALT contigs and HLA sequences, using an alternative scaffold-aware version of BWA-MEM. As done in the same 1000 genomes alignment pipeline, PCR-introduced duplicates were marked using the *markduplicates* function in BioBamBam.

Aligned sequences of Whole Exome Sequencing (WES) and RNA sequencing experiments from *The Cancer Genome Atlas (TCGA)* were downloaded in BAM format from the *Genomic Data Commons (GDC)* portal. All 9162 available BAM files of Blood Derived normal WES samples were selected. For RNA-Seq, all 9762 RNA-Seq samples that were derived from primary tumours and were aligned using the "STAR 2-Pass" workflow, were selected. Reads mapped to the MHC region of chromosome 6 (chr6:28509970-33480727) and unmapped reads were extracted from the BAM files at download time following the instructions that are described in the GDC API. For the RNA-Seq samples one file failed to download after multiple attempts. The resulting dataset consists of 9162 blood-derived normal WES samples and 9761 primary tumour RNA-seq samples from 33 available cancer types.

The most resource intensive RNA tools were applied on a subset of the TCGA dataset. Optitype was applied on 2226 RNA files and HLAforest on 2900 files. HLA-HD was not applied on the TCGA data.

Gold standard HLA typing data

Gold standard PCR-based HLA calls for the samples from the *1000 genomes on GRCh38* dataset were provided by three earlier studies [26–28]. The HLA genotypes from these datasets were merged. Where the calls did not agree, the PCR-SBT based calls by Gourraud et al. [26] were preferred. For the NCI-60 cell lines, PCR-SBT derived HLA genotypes were provided in a study by Adams et al [29].

For both reference datasets alleles were mapped to the corresponding G-groups, as defined by IPD-IMGT (http://hla.alleles.org/alleles/g_groups.html), and trimmed to the second-field resolution.

HLA allele predictions

All 13 selected tools were run on the sliced BAM files following the guidelines of the authors.

For tools that allowed to specify a list of loci that should be called: *HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1* and *HLA-DRB1* were chosen.

Kourami was run with the *-a* (additional loci) parameter to call the *HLA-DPA1* and *HLA-DPB1* genes. In rare cases, this led to a crash of the tool and *Kourami* was run again without the "-a" parameter. For *HLAminer* only the HPRA mode was evaluated. *xHLA*, *Polysolver* and *HLA-VBSeq* were not compatible with BAM files that are aligned to a reference genome build that includes alternative contigs. For these tools, an additional realignment step was performed before the tool was executed. Input data for *xHLA* and *Polysolver* were realigned to a GRCh38 build that excludes alternative contigs. The input data for *HLA-VBSeq* was realigned to GRCh37.

All allele predictions were mapped to the corresponding G-groups and trimmed at second-field resolution.

Measuring the resource consumption

The running time and memory consumption required by the tools were measured for a subset of 10 DNA and 10 RNA sequencing files from the TCGA project. Each tool was executed in a separate Docker container that was allocated a single CPU core. When the package provided a parameter to specify the number of threads, this was set to 1. Per file, the memory usage of the Docker container was monitored using the *docker stats* command. The running time was calculated as the time interval between the start and the end of the tool, excluding the time to start the Docker container. Pre-processing steps related to re-alignment to a different genome build (as required for *xHLA*, *Polysolver* and *HLA-VBSeq*) were not included in the resource consumption assessment.

Performance metric

For each sample, two allele predictions were made. An allele prediction was labelled "correct" when it was listed as one of the two alleles in the gold standard for that patient. When a tool made a homozygous prediction, while the gold standard was heterozygous, at most one of the two predictions was labelled "correct" for that sample. The accuracy of the predictions is then defined as the proportion of all correctly predicted alleles divided by twice the number of samples. Samples where the gold standard was missing for a particular gene were ignored for that gene.

Population frequency data

Lists of expected HLA allele frequencies for an African American and for a Caucasian American population were constructed based on 18 different studies in the *Allele Frequency Net* [21] database (Table S1). The studies were selected based on the following criteria. First, we required that the study was conducted on a *Black* or *Caucasoid* population from the United States. This was not possible for *HLA-DPA1* where no HLA allele frequencies were available for these ethnicities. As a substitute, the allele frequencies of three European populations (French, Swedish and Basques) were used to approximate the allele frequencies for this gene in Caucasian Americans. As a second requirement, the HLA calls should be determined by a PCR-based method. Thirdly, the Allele Frequency Net database should have assigned a gold label to the study for the gene of interest (). Lastly, it was required that the subjects included in the selected studies were healthy subjects (i.e., selected for an anthropological study, blood donors, bone marrow registry or controls for a disease study).

Allele frequencies from different studies were combined by taking the average frequency, weighted according to the study's sample size. All alleles were mapped to the corresponding G-groups and trimmed at second-field resolution.

Correlation between expected and observed allele frequencies

For all tools and for each supported data type, the number of times that each allele was called was counted. This count was divided by the total number of samples to obtain the "observed allele frequency". The Pearson correlation was calculated between observed allele frequencies and the allele frequencies that were expected based on the *Allele Frequency Net* database.

Concordance of predictions among different tools

Per gene, the concordance of the predictions between each pair of tools was assessed by counting the number of genotype predictions made by the first tool that were also made by the second tool (for the same sample and gene). Samples where one of both tools did not make a prediction were not considered. This analysis was performed on the 1000 genomes and TCGA dataset.

Concordance of predictions between DNA and RNA samples

The concordance of the predictions made on DNA and on RNA data by the tools that support both data types (*HLAminer*, *PHLAT*, *HLA-HD* and *Optitype*) was calculated analogously as the concordance between different tools. For each gene and each tool, the number of genotype predictions made on the DNA data that were also made on the RNA data were counted. Samples where the tool did not make a prediction on either DNA or RNA data were not considered.

Consensus HLA predictions

A majority voting rule was used to determine the most likely HLA genotype for each sample. For each gene of interest, we selected the pair of alleles that has been predicted the most frequently for that sample (i.e., outputted by the highest number of tools). When multiple allele pairs had equal numbers of predictions, priority was given to the allele pair that was predicted by the tool with the best individual performance for that gene.

Selecting a minimum number of tools to make consensus HLA predictions

The minimal set of tools that must be included in the majority voting scheme to make reliable consensus predictions was determined using an iterative procedure. Initially, two tools were selected for the model: the tool that performed the best in the benchmark on the 1000 genomes data and the one that best complements that tool. The latter tool was defined as the tool that most often made a correct prediction (for both alleles) on the samples that were wrongly predicted by the best performing tool. Tools were added one by one to this initial model. At each step, the consensus predictions were made and evaluated using the gold standard HLA calls. The tool that led to the greatest increase in accuracy was added to the model. This procedure was repeated until all tools were selected.

Hardware and software environment

Analyses were performed on Ubuntu 20.04 on a Dell EMC PowerEdge R940 server with 72 physical CPU cores and 376 GiB RAM installed.

Data processing and statistical analysis

Data processing and statistical analyses were performed using R (version 4.0).

Funding

This work was supported by the Ghent University Special Research Fund Starting Grant (JVdE, BOF.STG.2019.0073.01, https://www.ugent.be/en/research/funding/bof). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgement

The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Author contributions

- Conceptualization: J.V.d.E.
- Methodology: A.C. and J.V.d.E.
- Software: A.C., J.S. and P.M.
- Validation: A.C. and J.V.d.E.
- Formal Analysis: A.C., J.S., P.M. and J.V.d.E.
- Investigation: A.C., J.S., P.M. and J.V.d.E.
- Resources: A.C., J.S., P.M. and J.V.d.E.
- Data Curation: A.C., J.S., P.M. and J.V.d.E.
- Writing – Original Draft Preparation: A.C. and J.V.d.E.
- Writing – Review & Editing: A.C., J.S., P.M., K.M. and J.V.d.E.
- Visualization: A.C. and J.V.d.E.
- Supervision: J.V.d.E.
- Project Administration: J.V.d.E.
- Funding Acquisition: J.V.d.E.

Code availability

The code used to produce the results described in this manuscript will be made available on GitHub.

Conflict of Interest

The authors declare to have no conflicts of interest.

366 ## References

367 1. Trowsdale, J. Genomic Structure and Function in the MHC. *Trends in Genetics* **1993**, *9*, 117–
368 122, doi:10.1016/0168-9525(93)90205-V.

369 2. Beck, S.; Geraghty, D.; Inoko, H.; Rowen, L.; Aguado, B.; Bahram, S.; Campbell, R.D.; Forbes,
370 S.A.; Guillaudeux, T.; Hood, L.; et al. Complete Sequence and Gene Map of a Human Major
371 Histocompatibility Complex. *Nature 1999 401:6756* **1999**, *401*, 921–923, doi:10.1038/44853.

372 3. Horton, R.; Wilming, L.; Rand, V.; Lovering, R.C.; Bruford, E.A.; Khodiyar, V.K.; Lush, M.J.; Povey,
373 S.; Talbot, C.C.; Wright, M.W.; et al. Gene Map of the Extended Human MHC. *Nature Reviews*
374 *Genetics 2004 5:12* **2004**, *5*, 889–899, doi:10.1038/nrg1489.

375 4. Halenius, A.; Gerke, C.; Hengel, H. Classical and Non-Classical MHC I Molecule Manipulation by
376 Human Cytomegalovirus: So Many Targets—but How Many Arrows in the Quiver? *Cellular &*
377 *Molecular Immunology 2015 12:2* **2014**, *12*, 139–153, doi:10.1038/cmi.2014.105.

378 5. Allen, R.L.; Hogan, L. Non-Classical MHC Class I Molecules (MHC-Ib). *eLS* **2013**,
379 doi:10.1002/9780470015902.A0024246.

380 6. Hewitt, E.W. The MHC Class I Antigen Presentation Pathway: Strategies for Viral Immune
381 Evasion. *Immunology* **2003**, *110*, 163, doi:10.1046/J.1365-2567.2003.01738.X.

382 7. Axelrod, M.L.; Cook, R.S.; Johnson, D.B.; Balko, J.M. Biological Consequences of MHC-II
383 Expression by Tumor Cells in Cancer. *Clinical Cancer Research* 2019, *25*, 2392–2402.

384 8. Alspach, E.; Lussier, D.M.; Miceli, A.P.; Kizhvatov, I.; DuPage, M.; Luoma, A.M.; Meng, W.; Lichti,
385 C.F.; Esaulova, E.; Vomund, A.N.; et al. MHC-II Neoantigens Shape Tumour Immunity and
386 Response to Immunotherapy. *Nature* **2019**, *574*, 696–701, doi:10.1038/s41586-019-1671-8.

387 9. Sun, Z.; Chen, F.; Meng, F.; Wei, J.; Liu, B. MHC Class II Restricted Neoantigen: A Promising
388 Target in Tumor Immunotherapy. *Cancer Letters* **2017**, *392*, 17–25,
389 doi:10.1016/J.CANLET.2016.12.039.

390 10. Corthay, A.; Skovseth, D.K.; Lundin, K.U.; Røsjø, E.; Omholt, H.; Hofgaard, P.O.; Haraldsen, G.;
391 Bogen, B. Primary Antitumor Immune Response Mediated by CD4+ T Cells. *Immunity* **2005**, *22*,
392 371–383, doi:10.1016/J.IMMUNI.2005.02.003.

393 11. Audun, O.; Haabeth, W.; Fauskanger, M.; Manzke, M.; Lundin, K.U.; Corthay, A.; Bogen, B.;
394 Tveita, A.A.; Haabeth, O.A.W.; Fauskanger, M.; et al. CD4+ T-Cell–Mediated Rejection of MHC
395 Class II–Positive Tumor Cells Is Dependent on Antigen Secretion and Indirect Presentation on
396 Host APCs. *Cancer Research* **2018**, *78*, 4573–4585, doi:10.1158/0008-5472.CAN-17-2426.

397 12. Szolek, A.; Schubert, B.; Mohr, C.; Sturm, M.; Feldhahn, M.; Kohlbacher, O. OptiType: Precision
398 HLA Typing from next-Generation Sequencing Data. *Bioinformatics* **2014**, *30*, 3310–3316,
399 doi:10.1093/bioinformatics/btu548.

400 13. Bauer, D.C.; Zadoorian, A.; Wilson, L.O.W.; Alliance, M.G.H.; Thorne, N.P. Evaluation of
401 Computational Programs to Predict HLA Genotypes from Genomic Sequencing Data. *Briefings*
402 *in Bioinformatics* **2018**, *19*, 179–187, doi:10.1093/BIB/BBW097.

403 14. Orenbuch, R.; Filip, I.; Comito, D.; Shaman, J.; Pe'Er, I.; Rabadan, R. ArcasHLA: High-Resolution
404 HLA Typing from RNAseq. *Bioinformatics* **2020**, *36*, 33–40,
405 doi:10.1093/BIOINFORMATICS/BTZ474.

406 15. Matey-Hernandez, M.L.; Brunak, S.; Izarzugaza, J.M.G. Benchmarking the HLA Typing
407 Performance of Polysolver and Optitype in 50 Danish Parental Trios. *BMC Bioinformatics* **2018**,
408 *19*, 1–12, doi:10.1186/S12859-018-2239-6.

409 16. Lee, M.; Seo, J.H.; Song, S.; Song, I.H.; Kim, S.Y.; Kim, Y.A.; Gong, G.; Kim, J.E.; Lee, H.J. A New
410 Human Leukocyte Antigen Typing Algorithm Combined With Currently Available Genotyping
411 Tools Based on Next-Generation Sequencing Data and Guidelines to Select the Most Likely
412 Human Leukocyte Antigen Genotype. *Frontiers in Immunology* **2021**, *12*, 4080,
413 doi:10.3389/FIMMU.2021.688183.

414 17. Li, X.; Zhou, C.; Chen, K.; Huang, B.; Liu, Q.; Ye, H. Benchmarking HLA Genotyping and Clarifying
415 HLA Impact on Survival in Tumor Immunotherapy. *Molecular Oncology* **2021**, *15*, 1764–1782,
416 doi:10.1002/1878-0261.12895.

417 18. Chen, J.; Madireddi, S.; Nagarkar, D.; Migdal, M.; vander Heiden, J.; Chang, D.; Mukhyala, K.;
418 Selvaraj, S.; Kadel, E.E.; Brauer, M.J.; et al. In Silico Tools for Accurate HLA and KIR Inference
419 from Clinical Sequencing Data Empower Immunogenetics on Individual-Patient and Population
420 Scales. *Briefings in Bioinformatics* **2021**, *22*, 1–11, doi:10.1093/BIB/BBAA223.

421 19. Zheng-Bradley, X.; Streeter, I.; Fairley, S.; Richardson, D.; Clarke, L.; Flicek, P.; Consortium, the
422 1000 G.P. Alignment of 1000 Genomes Project Reads to Reference Assembly GRCh38.
423 *Gigascience* **2017**, *6*, 1, doi:10.1093/GIGASCIENCE/GIX038.

424 20. Lappalainen, T.; Sammeth, M.; Friedländer, M.R.; ’T Hoen, P.A.C.; Monlong, J.; Rivas, M.A.;
425 Gonzàlez-Porta, M.; Kurbatova, N.; Griebel, T.; Ferreira, P.G.; et al. Transcriptome and Genome
426 Sequencing Uncovers Functional Variation in Humans. *Nature 2013 501:7468* **2013**, *501*, 506–
427 511, doi:10.1038/nature12531.

428 21. Gonzalez-Galarza, F.F.; McCabe, A.; Santos, E.J.M. dos; Jones, J.; Takeshita, L.; Ortega-Rivera,
429 N.D.; Cid-Pavon, G.M.D.; Ramsbottom, K.; Ghattaoraya, G.; Alfirevic, A.; et al. Allele Frequency
430 Net Database (AFND) 2020 Update: Gold-Standard Data Classification, Open Access Genotype
431 Data and New Query Tools. *Nucleic Acids Res* **2020**, *48*, D783–D788,
432 doi:10.1093/NAR/GKZ1029.

433 22. Adams, S.D.; Barracchini, K.C.; Chen, D.; Robbins, F.; Wang, L.; Larsen, P.; Luhm, R.; Stroncek,
434 D.F. Ambiguous Allele Combinations in HLA Class I and Class II Sequence-Based Typing: When
435 Precise Nucleotide Sequencing Leads to Imprecise Allele Identification. **2004**,
436 doi:10.1186/1479-5876-2-30.

437 23. Bauer-Mehren, A.; Bundschus, M.; Rautschka, M.; Mayer, M.A.; Sanz, F.; Furlong, L.I. Gene-
438 Disease Network Analysis Reveals Functional Modules in Mendelian, Complex and
439 Environmental Diseases. *PLoS ONE* **2011**, *6*, e20284, doi:10.1371/journal.pone.0020284.

440 24. Abaan, O.D.; Polley, E.C.; Davis, S.R.; Zhu, Y.J.; Bilke, S.; Walker, R.L.; Pineda, M.; Gindin, Y.;
441 Jiang, Y.; Reinhold, W.C.; et al. The Exomes of the NCI-60 Panel: A Genomic Resource for Cancer
442 Biology and Systems Pharmacology. *Cancer Res* **2013**, *73*, 4372–4382, doi:10.1158/0008-
443 5472.CAN-12-3342.

444 25. Reinhold, W.C.; Varma, S.; Sunshine, M.; Elloumi, F.; Ofori-Atta, K.; Lee, S.; Trepel, J.B.; Meltzer,
445 P.S.; Doroshow, J.H.; Pommier, Y. RNA Sequencing of the NCI-60: Integration into CellMiner
446 and CellMiner CDB. *Cancer Res* **2019**, *79*, 3514–3524, doi:10.1158/0008-5472.CAN-18-2047.

447 26. Gourraud, P.A.; Khankhanian, P.; Cereb, N.; Yang, S.Y.; Feolo, M.; Maiers, M.; Rioux, J.D.;
448 Hauser, S.; Oksenberg, J. HLA Diversity in the 1000 Genomes Dataset. *PLOS ONE* **2014**, *9*,
449 e97282, doi:10.1371/JOURNAL.PONE.0097282.

450 27. Boegel, S.; Löwer, M.; Schäfer, M.; Bukur, T.; de Graaf, J.; Boisguérin, V.; Türeci, Ö.; Diken, M.;
451 Castle, J.C.; Sahin, U. HLA Typing from RNA-Seq Sequence Reads. *Genome Medicine* **2012**, *4*, 1–
452 12, doi:10.1186/GM403.

453 28. Huang, Y.; Yang, J.; Ying, D.; Zhang, Y.; Shotelersuk, V.; Hirankarn, N.; Sham, P.C.; Lau, Y.L.; Yang,
454 W. HLAreporter: A Tool for HLA Typing from next Generation Sequencing Data. *Genome*
455 *Medicine* **2015**, *7*, 1–12, doi:10.1186/S13073-015-0145-3.

456 29. Adams, S.; Robbins, F.M.; Chen, D.; Wagage, D.; Holbeck, S.L.; Morse, H.C.; Stroncek, D.;
457 Marincola, F.M. HLA Class I and II Genotype of the NCI-60 Cell Lines. *Journal of Translational*
458 *Medicine* **2005**, *3*, 11, doi:10.1186/1479-5876-3-11.

459 30. Kawaguchi, S.; Higasa, K.; Shimizu, M.; Yamada, R.; Matsuda, F. HLA-HD: An Accurate HLA
460 Typing Algorithm for next-Generation Sequencing Data. *Human Mutation* **2017**, *38*, 788–797,
461 doi:10.1002/HUMU.23230.

462 31. Nariai, N.; Kojima, K.; Saito, S.; Mimori, T.; Sato, Y.; Kawai, Y.; Yamaguchi-Kabata, Y.; Yasuda, J.;
463 Nagasaki, M. HLA-VBSeq: Accurate HLA Typing at Full Resolution from Whole-Genome
464 Sequencing Data. *BMC Genomics* **2015**, *16*, 1–6, doi:10.1186/1471-2164-16-S2-S7.

465 32. Dilthey, A.T.; Mentzer, A.J.; Carapito, R.; Cutland, C.; Cereb, N.; Madhi, S.A.; Rhie, A.; Koren, S.;
466 Bahram, S.; McVean, G.; et al. HLA*LA—HLA Typing from Linearly Projected Graph Alignments.
467 *Bioinformatics* **2019**, *35*, 4394–4396, doi:10.1093/BIOINFORMATICS/BTZ235.

468 33. Kim, H.J.; Pourmand, N. HLA Haplotyping from RNA-Seq Data Using Hierarchical Read
469 Weighting. *PLOS ONE* **2013**, *8*, e67885, doi:10.1371/JOURNAL.PONE.0067885.

470 34. Warren, R.L.; Choe, G.; Freeman, D.J.; Castellarin, M.; Munro, S.; Moore, R.; Holt, R.A.
471 Derivation of HLA Types from Shotgun Sequence Datasets. *Genome Medicine* **2012**, *4*, 1–8,
472 doi:10.1186/GM396.

473 35. Ka, S.; Lee, S.; Hong, J.; Cho, Y.; Sung, J.; Kim, H.N.; Kim, H.L.; Jung, J. HLAscan: Genotyping of
474 the HLA Region Using next-Generation Sequencing Data. *BMC Bioinformatics* **2017**, *18*, 1–11,
475 doi:10.1186/S12859-017-1671-3.

476 36. Lee, H.; Kingsford, C. Kourami: Graph-Guided Assembly for Novel Human Leukocyte Antigen
477 Allele Discovery. *Genome Biology* **2018**, *19*, 1–16, doi:10.1186/S13059-018-1388-2.

478 37. Bai, Y.; Wang, D.; Fury, W. PHLAT: Inference of High-Resolution HLA Types from RNA and Whole
479 Exome Sequencing. *Methods in Molecular Biology* **2018**, *1802*, 193–201, doi:10.1007/978-1-
480 4939-8546-3_13.

481 38. Rooney, M.S.; Shukla, S.A.; Wu, C.J.; Getz, G.; Hacohen, N. Molecular and Genetic Properties of
482 Tumors Associated with Local Immune Cytolytic Activity. *Cell* **2015**, *160*, 48–61,
483 doi:10.1016/j.cell.2014.12.033.

484 39. Xie, C.; Yeo, Z.X.; Wong, M.; Piper, J.; Long, T.; Kirkness, E.F.; Biggs, W.H.; Bloom, K.; Spellman,
485 S.; Vierra-Green, C.; et al. Fast and Accurate HLA Typing from Short-Read next-Generation
486 Sequence Data with XHLA. *Proceedings of the National Academy of Sciences* **2017**, *114*, 8059–
487 8064, doi:10.1073/PNAS.1707945114.

40. Hayashi, S.; Moriyama, T.; Yamaguchi, R.; Mizuno, S.; Komura, M.; Miyano, S.; Nakagawa, H.; Imoto, S. ALPHLARD-NT: Bayesian Method for Human Leukocyte Antigen Genotyping and Mutation Calling through Simultaneous Analysis of Normal and Tumor Whole-Genome Sequence Data. *Journal of Computational Biology* **2019**, *26*, 923–937, doi:10.1089/cmb.2018.0224.

41. Liu, C.; Yang, X.; Duffy, B.; Mohanakumar, T.; Mitra, R.D.; Zody, M.C.; Pfeifer, J.D. ATHLATES: Accurate Typing of Human Leukocyte Antigen through Exome Sequencing. *Nucleic Acids Research* **2013**, *41*, e142–e142, doi:10.1093/NAR/GKT481.

42. Buchkovich, M.L.; Brown, C.C.; Robasky, K.; Chai, S.; Westfall, S.; Vincent, B.G.; Weimer, E.T.; Powers, J.G. HLAProfiler Utilizes K-Mer Profiles to Improve HLA Calling Accuracy for Rare and Common Alleles in RNA-Seq Data. *Genome Medicine* **2017**, *9*, 1–15, doi:10.1186/S13073-017-0473-6.

43. Huang, Y.; Yang, J.; Ying, D.; Zhang, Y.; Shotelersuk, V.; Hirankarn, N.; Sham, P.C.; Lau, Y.L.; Yang, W. HLAreporter: A Tool for HLA Typing from next Generation Sequencing Data. *Genome Medicine* **2015**, *7*, 1–12, doi:10.1186/S13073-015-0145-3.

44. Wittig, M.; Anmarkrud, J.A.; Kässens, J.C.; Koch, S.; Forster, M.; Ellinghaus, E.; Hov, J.R.; Sauer, S.; Schimmler, M.; Ziemann, M.; et al. Development of a High-Resolution NGS-Based HLA-Typing and Analysis Pipeline. *Nucleic Acids Research* **2015**, *43*, e70, doi:10.1093/NAR/GKV184.

45. Sverchkova, A.; Anzar, I.; Stratford, R.; Clancy, T. Improved HLA Typing of Class I and Class II Alleles from Next-Generation Sequencing Data. *HLA* **2019**, *94*, 504–513, doi:10.1111/TAN.13685.

46. Jia, X.; Han, B.; Onengut-Gumuscu, S.; Chen, W.M.; Concannon, P.J.; Rich, S.S.; Raychaudhuri, S.; de Bakker, P.I.W. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLOS ONE* **2013**, *8*, e64683, doi:10.1371/JOURNAL.PONE.0064683.

47. Cao, H.; Wu, J.; Wang, Y.; Jiang, H.; Zhang, T.; Liu, X.; Xu, Y.; Liang, D.; Gao, P.; Sun, Y.; et al. An Integrated Tool to Study MHC Region: Accurate SNV Detection and HLA Genes Typing in Human MHC Region Using Targeted High-Throughput Sequencing. *PLoS ONE* **2013**, *8*, doi:10.1371/JOURNAL.PONE.0069388.

518 # Captions of Figures and Tables

519 ## Table 1: Overview of evaluated tools for HLA genotyping.

520 List of tools for NGS-based HLA genotyping. Checkmarks and crosses indicate which input files are
521 supported and for which genes predictions can be made. The tools in the upper part of the table are
522 benchmarked in this study. Tools in the lower part of the table did not fulfil our inclusion criteria and
523 were not further considered *(Methods)*.

524 ## Figure 1: Computational resource consumption of the 13 selected tools

525 (**A-B**) Boxplots compare the amount of resources needed by the different tools to analyse one
526 sequencing file on a system with a single CPU core. Each tool was applied on DNA and/or RNA
527 sequencing files *(n=10)*, as indicated at the top of the figure. Different tools are represented with a
528 different colour of the boxplot, as indicated in the legend on the right. (**A**) Time consumption per
529 sample. (**B**) Maximal memory consumption per sample.

530 ## Figure 2: HLA allele prediction accuracies

531 Radar plots of HLA allele prediction accuracies on samples from the 1000 Genomes Project. Coloured
532 lines represent different genes, as indicated in the legend below the plots. Corners of the radar plots
533 correspond to the tools that were evaluated for that data type. The *Meta* tools correspond to the 4-
534 tool consensus metaclassifiers.

535 ## Figure 3: Correlations between observed and expected allele frequencies

536 Heatmap of correlations between observed allele frequencies and frequencies expected in an African
537 American and in a Caucasian American population. Vertical axis indicates the tools, with different
538 colours representing the data type (DNA or RNA) on which the tool was applied. Rows were sorted
539 according to the mean correlation of the tool. Size of the circles indicates the p-value of the correlation
540 test as indicated in legend. Absent circles indicate that the tool could not be evaluated on that gene.

541 ## Figure 4: Accuracies of meta-prediction models with an increasing number of included
542 tools

543 Tools were added one by one to the consensus metaclassifier model. At each step, the prediction
544 accuracies of the best performing metaclassifier model for a given number of tools were plotted at
545 the top of the figure. Unfilled markers are placed at the smallest number of tools where the maximal
546 accuracy was obtained for that gene. Black lines indicate the average accuracy of the consensus
547 predictions for the two MHC classes (averaged over all genes of that class). The table below the plot
548 indicates which tools were selected in each model for a given number of tools.

549

Supplementary materials

Supplementary tables

- Table S1: Motivation for exclusion of 9 tools from the study
- Table S2: Overview of AFN datasets used for expected allele frequencies

Supplementary figures

- Figure S1: Fraction of correct allele predictions (1000 genomes)
- Figure S2: Fraction of successful allele predictions (1000 genomes)
- Figure S3: HLA allele prediction accuracies on NCI-60 cell lines
- Figure S4: Correctness of predictions on DNA data
- Figure S5: Correctness of predictions on RNA data
- Figure S6: Concordance of HLA calls between each pair of tools on DNA data (1000 genomes)
- Figure S7: Concordance of HLA calls between each pair of tools on RNA data (1000 genomes)
- Figure S8: Concordance of HLA calls between each pair of tools on DNA data (TCGA)
- Figure S9: Concordance of HLA calls between each pair of tools on RNA data (TCGA)
- Figure S10: Comparison of accuracies of all-tool metaclassifier with best performing individual tool per gene

		Data type		Input filetype		Supported HLA loci								
		DNA	RNA	BAM	FASTQ	A	B	C	DPA1	DPB1	DQA1	DQB1	DRB1	Version
Included	*arcasHLA* [14]	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.2.0
	HLA-HD [30]	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.3.0
	HLA-VBSeq [31]	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	2
	*HLA*LA* [32]	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	1.0.1
	HLAforest [33]	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	1
	HLAminer [34]	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	1.4
	HLAscan [35]	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2.1.4
	Kourami [36]	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	0.9.6
	Optitype [12]	✓	✓	✗	✓	✓	✓	✓	✗	✗	✗	✗	✗	1.3.5
	PHLAT [37]	✓	✓	✗	✓	✓	✓	✓	✗	✗	✓	✓	✓	1.1
	Polysolver [38]	✓	✗	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗	4
	seq2HLA [27]	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	2.3
	xHLA [39]	✓	✗	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	1.2
Not included	ALPHLARD(-NT) [40]	✓	✗	?	?	✓	✓	✓	✓	✓	✓	✓	✓	
	ATHLATES [41]	✓	✗	✗	✓	✓	✓	✓	✗	✗	✗	✓	✓	
	HLAProfiler [42]	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	HLAreporter [43]	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	HLAssign [44]	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	OncoHLA [45]	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	PolyPheMe	✓	✗	✗	✓	✓	✓	✓	✗	✗	✗	✓	✓	
	SNP2HLA [46]	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	
	SOAP-HLA [47]	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	

Table 1

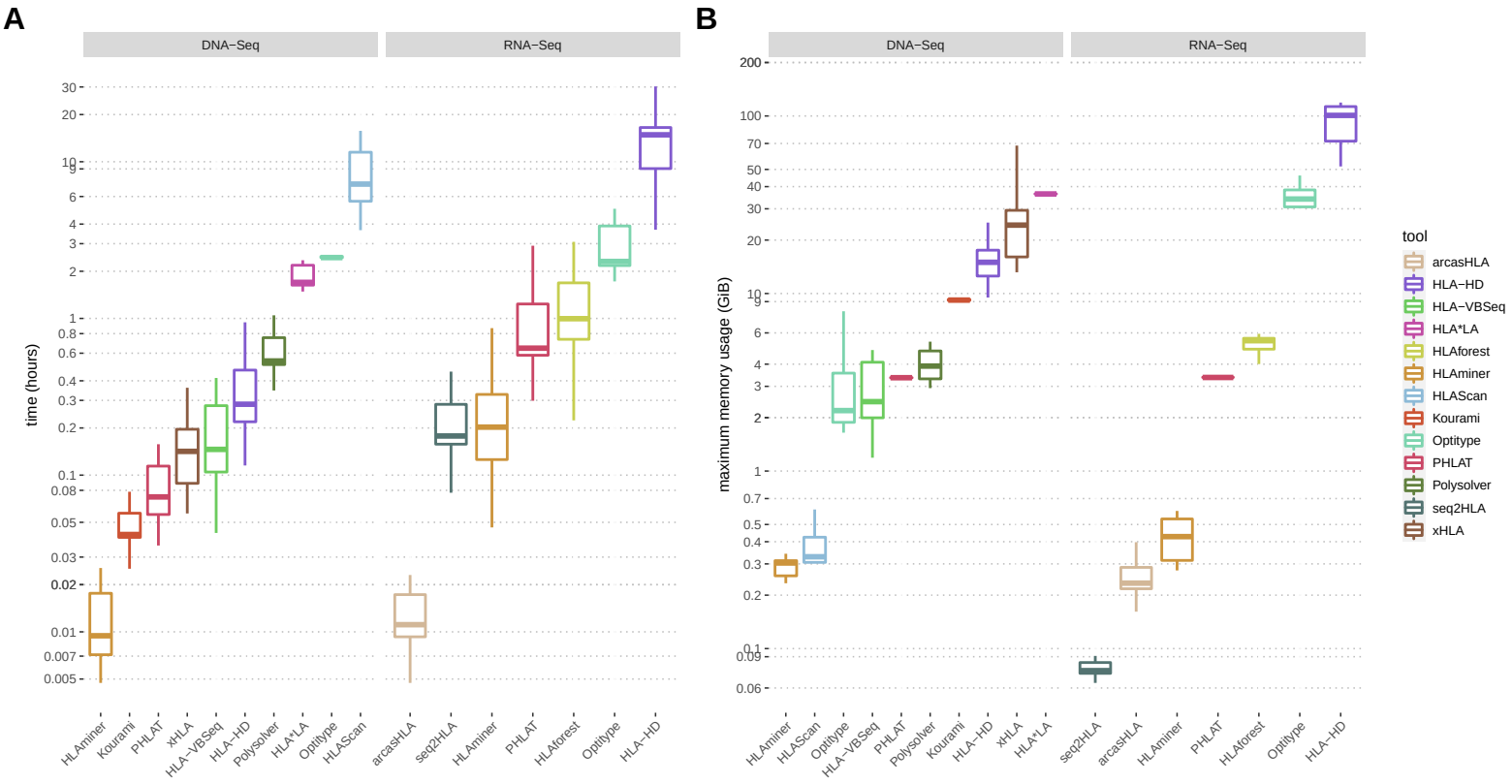
A
DNA−Seq
RNA−Seq
time (hours)
HLAminer
Kourami
PHLAT
xHLA
HLA−VBSeq
HLA−HD
Polysolver
HLA*LA
Optitype
HLAScan
arcasHLA
seq2HLA
HLAminer
PHLAT
HLAforest
Optitype
HLA−HD
B
DNA−Seq
RNA−Seq
maximum memory usage (GiB)
HLAminer
HLAScan
Optitype
HLA−VBSeq
PHLAT
Polysolver
Kourami
HLA−HD
xHLA
HLA*LA
seq2HLA
arcasHLA
HLAminer
PHLAT
HLAforest
Optitype
HLA−HD
tool
arcasHLA
HLA−HD
HLA−VBSeq
HLA*LA
HLAforest
HLAminer
HLAScan
Kourami
Optitype
PHLAT
Polysolver
seq2HLA
xHLA

Figure 1

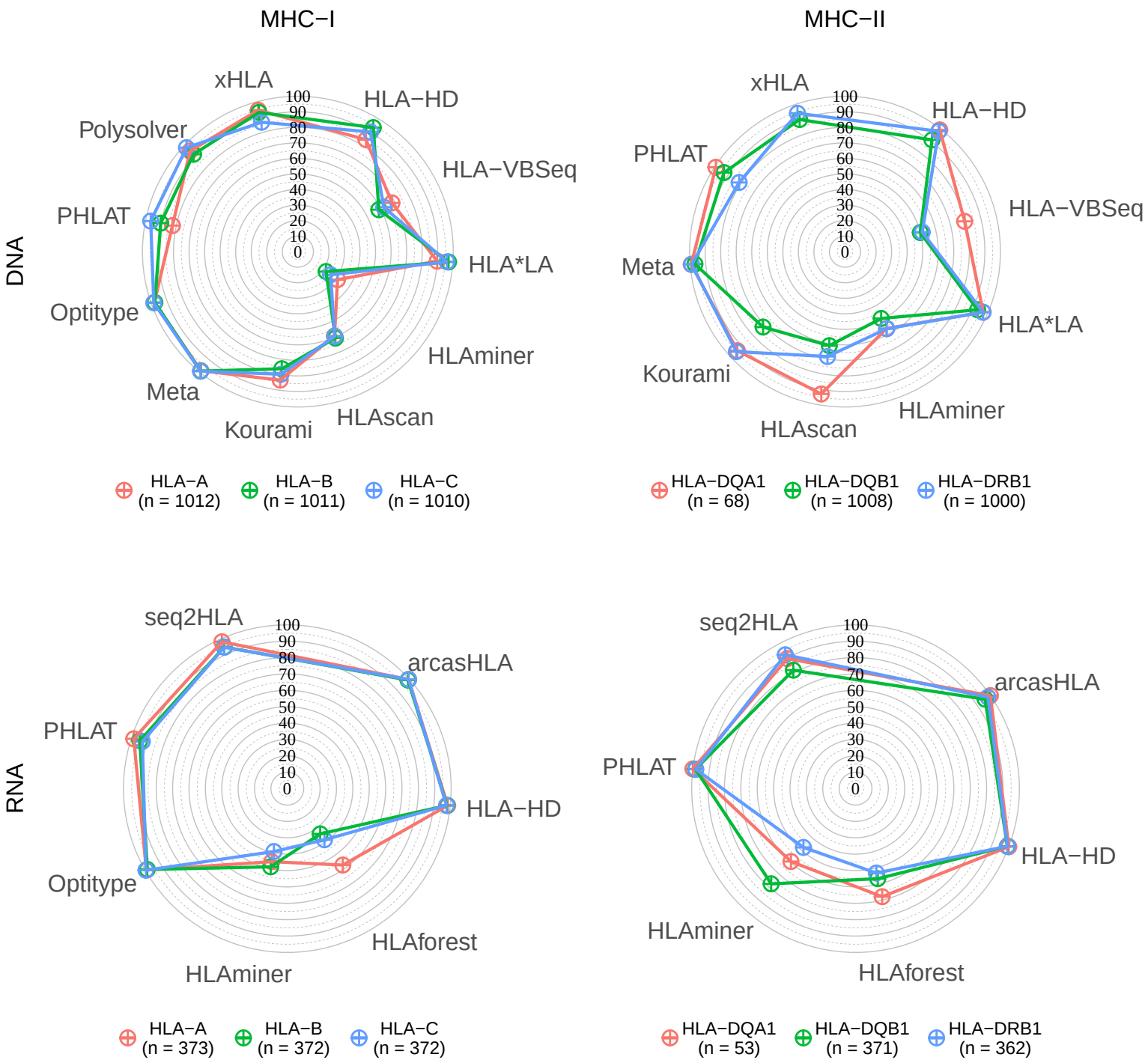
MHC−I
MHC−II
DNA
RNA
xHLA
HLA−HD
HLA−VBSeq
HLA*LA
HLAminer
HLAscan
Kourami
Meta
Optitype
PHLAT
Polysolver
seq2HLA
arcasHLA
HLAforest
100
90
80
70
60
50
40
30
20
10
0
HLA−A (n = 1012)
HLA−B (n = 1011)
HLA−C (n = 1010)
HLA−DQA1 (n = 68)
HLA−DQB1 (n = 1008)
HLA−DRB1 (n = 1000)
HLA−A (n = 373)
HLA−B (n = 372)
HLA−C (n = 372)
HLA−DQA1 (n = 53)
HLA−DQB1 (n = 371)
HLA−DRB1 (n = 362)

Figure 2

African American
Caucasian American
xHLA
Optitype
Meta
HLA−HD
HLA*LA
Optitype
Polysolver
arcasHLA
PHLAT
Kourami
seq2HLA
PHLAT
HLA−VBSeq
HLAscan
HLAminer
HLAminer
HLAforest
HLA−A
HLA−B
HLA−C
HLA−DPB1
HLA−DQA1
HLA−DQB1
HLA−DRB1
HLA−A
HLA−B
HLA−C
HLA−DPA1
HLA−DPB1
HLA−DQA1
HLA−DQB1
HLA−DRB1
ρ
1.00
0.95
0.90
0.50
data type
DNA
RNA
p−value
≤ 10−10
10−7
10−4
10−1

Figure 3

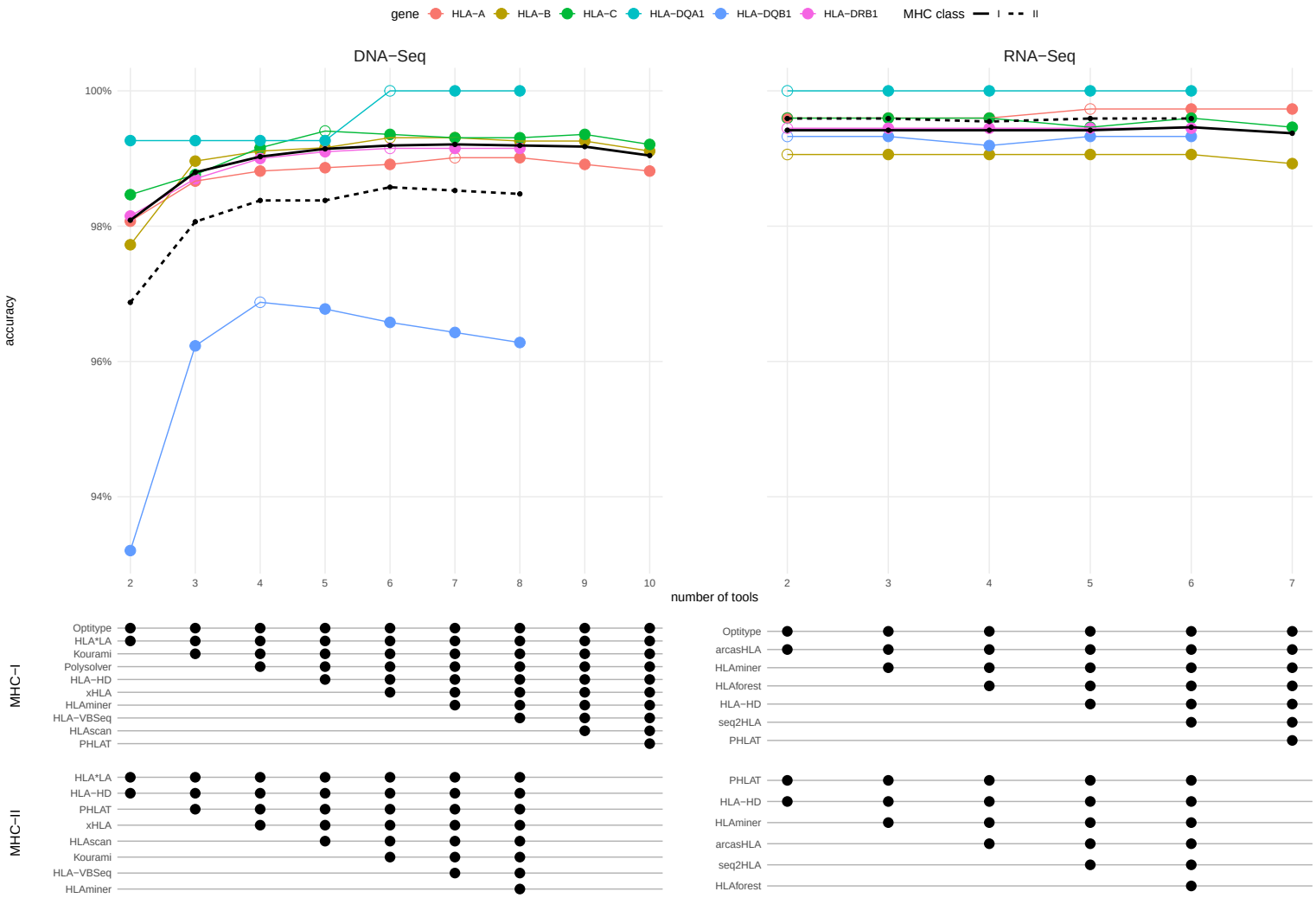
gene
HLA−A
HLA−B
HLA−C
HLA−DQA1
HLA−DQB1
HLA−DRB1
MHC class
I
II
DNA−Seq
RNA−Seq
accuracy
100%
98%
96%
94%
number of tools
MHC−I
Optitype
HLA*LA
Kourami
Polysolver
HLA−HD
xHLA
HLAminer
HLA−VBSeq
HLAscan
PHLAT
arcasHLA
seq2HLA
MHC−II

Figure 4