High-resolution HLA haplotypes inference and estimation in a panel of 304 mother-child pairs

Daniel Alvarez[1], Loes van de Pasch[1], Gideon Hönger[2,3,4], Stefan Schaub[2,3,4], Erik Rozemuller[1]

GenDx, Utrecht, The Netherlands & University Hospital Basel

Introduction

Human leukocyte antigen (HLA) genes are in linkage disequilibrium (LD), enabling recognition of conserved haplotypes. HLA allele segregation occurs following Mendelian inheritance rules, which allows for haplotype inference in nuclear families. Usually, nuclear families with pedigree-size >2 are preferred, in order to resolve ambiguities. Here we assess feasibility and potential of haplotype inference in a panel of mother-child pairs. HLA haplotype frequencies constitute a valuable resource for improved HLA typing. Our second goal was to describe a two-loci haplotype estimation method based on available NMDP haplotype frequencies, and applied to haplotypes HLA-DRB1~DRB345, HLA-DQB1~DRB1, HLA-B~C, HLA-B~DRB1 and HLA-A~B.

Material & Methods

Haplotypes in a panel of 304 mother-child pairs, HLA typed at high-resolution, were inferred for twelve HLA genes (HLA-A, B, C, G, DRB1, DRB3, DRB4, DRB5, DQA1, DQB1, DPA1, DPB1). Typing results were obtained using the NGSgo® workflow and NGSengine® software (GenDx). For each pair, three haplotypes (shared between mother and child, unshared in mother and unshared in child) are inferred. Any two-loci haplotype combinations from samples with a missing locus (due to technical reasons) were excluded. Two-loci haplotype Caucasian frequencies were obtained from NMDP, due to the population group present in the panel.

Results

In a total of 912 inferred haplotypes with 9120 loci (counting HLA-DRB3/4/5 as a single locus), we were able to unambiguously delineate alleles in 89.2 % of the cases (Figure 1). Ambiguous haplotype results due to identical heterozygous genotypes in mother and child represented only 9,8% of the cases, proving the potential of mother-child pairs for haplotype inference. In a few cases haplotype inference was discordant, due to genotype ambiguities (not taken into consideration in the method) or technical reasons (e.g. allele dropout).

For development of a haplotype estimation method making use of NMDP haplotype frequencies (limited to two-fields), typing resolution was reduced to two-fields and performed as shown in Figure 2. Validation of the method was done by comparison with the previously inferred haplotypes. The mean accuracy for all two-loci haplotypes was 93,9% in a total of 1628 estimations (Table 1). For haplotypes with the highest LD between them (HLA-B~C, HLA-DQB1~DRB1 and HLA-DRB1-DRB345), the accuracy was close to 100% in 942 estimations. In HLA-B~DRB1 and HLA-A~B, with LD, accuracy decreased to 85% in 686 estimations. The differences in the distribution of haplotype frequencies and the coverage of haplotype frequencies in the panel is shown in Figure 3. The graph shows the coverage of haplotype frequencies in the population over the number of uniquely identified haplotypes, sorted by decreasing frequency. It illustrates that for loci with a high LD, a lower number of haplotypes cover the same percentage in a population, e.g. 27 HLA-DQB1~DRB1 haplotypes (high LD) cover about 90% of the population, whereas 205 HLA-B~DRB1haplotypes (low LD) are needed to cover 90%.

Figure 1. Haplotype inference results
Percentage of each haplotype category, assigned to all 1920 loci present in all inferred haplotypes.

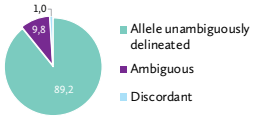

Figure 2. Haplotype estimation method
Two possible combinations of unique haplotypes are computed. Ranking and frequencies are included from NMDP haplotype frequency tables. An estimation metric P is calculated: P(Haplotype.combination) = Freq.x*Freq.y. The combination with the highest metric (shown in green/left in this example) is the most likely combination and is selected as the estimated haplotype.

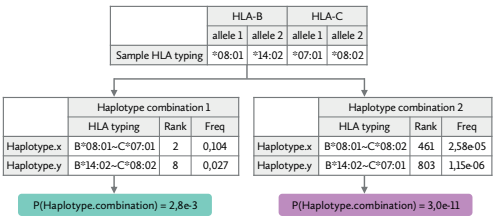

Figure 3. Coverage (%) of haplotype frequencies over number of unique haplotypes

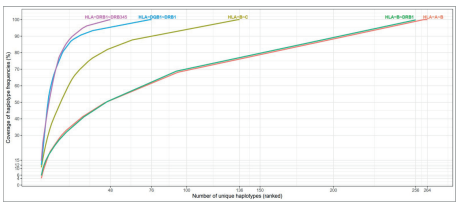

Table 1. Haplotype estimation method accuracy
Mean accuracy and individual haplotypes accuracy were obtained divided the number of correctly estimated haplotype combinations by the number of total estimations made.

Haplotype combination	Estimation accuracy	Number of estimations
HLA-A~B	85,8%	345
HLA-B~C	99,2%	374
HLA-B~DRB1	85,9%	341
HLA-DRB1~DQB1	100%	381
HLA-DRB1~DRB345	100%	187
Mean accuracy	93,9%	1628

Conclusion

Two-loci haplotypes can be unambiguously inferred from ~90% of mother-child pairs.
Two-loci haplotypes from individual samples can reliably be estimated with ~85-100% accuracy based on NMDP population haplotype frequencies.

References

1 GenDx, Utrecht, the Netherlands 2 Clinic for Transplantation Immunology and Nephrology, University Hospital Basel, Basel, Switzerland 3 Transplantation Immunology, Department of Biomedicine, University of Basel, Basel, Switzerland 4 HLA-Diagnostics and Immunogenetics, Department of Laboratory Medicine, University Hospital Basel, Basel, Switzerland