Aim:
Next Generation Sequencing (NGS) technologies are more frequently used to determine HLA genotypes. Since all sequence reads originate from a single molecule, in most cases the nucleotide sequences of the paternal and maternal alleles are determined separately. During PCR amplification the DNA template is amplified, resulting in copies of the original template. However, this process is not perfect, leading to wrong bases in individual molecules. These errors can be identified in NGS, where it is considered noise. In addition, hybrid molecules are generated which represent reactions which start at one molecule, but at some point the product re-anneals to another template, and continues to elongate. Especially if a template originates from a heterozygous sample, and covers multiple positions where the maternal and paternal allele differ, these hybrid molecules can be detected by NGS. Identifying phasing is hampered by hybrid reads. In this study we quantified the hybrid molecules as hybrid reads by NGS, and applied it as a quality value for the PCR amplification.

Methods:
We created a software tool that calculates the average cross-over percentage in a dataset. This percentage is a measure for the chance of a crossover taking place between any 2 neighbouring nucleotides. This percentage can be used to predict the expected cross-over rate between nucleotides at any distance and thus to determine the maximum distance in which two heterozygous positions can be phased reliably.

HLA gene amplicons were generated with GenDx NGSgo-AmpX in independent experiments in different laboratories. NGS sequencing was done on Illumina MiSeq and Ion Torrent PGM. NGS HLA typing was performed using NGSengine. The results were used to quantify the number of hybrid reads in several different HLA gene amplification systems.

Results

Figure 1. Example of an HLA-A typing result as generated with NGSengine.

In one experiment, the cross-over rates were compared on data generated with different NGS technologies (Illumina MiSeq, Ion Torrent PGM). The same amplification protocol was performed in a second laboratory and analyzed on the Ion Torrent PGM (Table 1). Amplifications performed at one site, resulted in comparable crossover percentages, independent of the sequencing technology that was used. When the same amplification protocol was applied in a second laboratory, the rate of crossovers was twice as low, probably because a different thermocycler was used. Based on these initial experiments, the NGSgo-AmpX amplification strategy for Class I, resulted in an average cross-over rate of 0.080 – 0.183%.

These percentages were used to predict the percentage of reads that would have a crossover between two heterozygous positions (Figure 4).

Figure 2. Crossover percentages in a HLA typing experiment.

Figure 3. Distribution of crossover events in an HLA-A amplicon.

Table 1: The chance (%) of a recombination between two neighbouring base positions.

<table>
<thead>
<tr>
<th>HLA</th>
<th>MiSeq Site 1</th>
<th>IonTorrent Site 1</th>
<th>IonTorrent Site 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-A</td>
<td>0.185</td>
<td>0.189</td>
<td>0.086</td>
</tr>
<tr>
<td>HLA-B</td>
<td>0.167</td>
<td>0.169</td>
<td>0.088</td>
</tr>
<tr>
<td>HLA-C</td>
<td>0.166</td>
<td>0.144</td>
<td>0.083</td>
</tr>
<tr>
<td>Average</td>
<td>0.185</td>
<td>0.162</td>
<td>0.089</td>
</tr>
</tbody>
</table>

Conclusions:
- Hybrid reads occur
- Hybrids hamper phasing
- Hybrid reads can be used as quality control for amplification robustness
- NGSengine allows for easy investigation of hybrid reads
- Data support the random distribution of recombination.
- With recombination rates of around 0.15%, phasing will be difficult to detect if positions are further apart than 300 bp.