Increasing the power in paternity and relationship testing utilizing MPS for the analysis of a large SNP panel

Ida Grandell¹, Andreas Tillmar¹,²

¹ Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden
² Department of Clinical and Experimental Medicine, Faculty of Health Sciences, Linköping University, Linköping, Sweden

Our motivation for a NGS/MPS SNP panel

• Still a need for supplementary markers
  • Handle around 3000-5000 relationship cases each year
    • Inconclusive paternity/maternity cases.
    • Complex/distant relationship cases.

• Implement NGS/MPS technology for use in routine casework
  • ”Easy” to interpret.
  • Account for things not alwyss considered in a reseach project
    • QA aspects (over time).
    • Educate staff.
Design & set up
140 SNP marker panel currently know as “QIAseq Investigator SNP ID”

• 140 SNPs
  • 52 SNPforID (Sanchez et al., 2006)
  • 88 II SNPs (Pakstis et al., 2010)
• Each SNP covered by 2 forward and 2 reverse primers
  • 4 amplicons per locus

Lab workflow
(details i Grandell et al., 2016)
Result - Coverage
(N=54, DNA extracted from blood samples)

Result – Allele balance (ARF; [# reads ref allele/# total reads])
Brief summary of the technical validation  
(140 SNPs and 29 autosomal STRs)

- Removed 3 SNPs (rs1360288, rs2399332, rs4530059) due to technical reasons
- Accuracy, repeatability, sensitivity etc have been shown
  - details in Grandell et al., 2016
- Set thresholds criterias for genotyp calling:
  - Cov>200x,
  - ARF: 0.9-1 or 0-0.1 (homozygous)
  - ARF: 0.4-0.6 (heterozygous)
- FTA and high quality DNA extractions
Validation of the biostatistic features of the panel
(137 SNPs and 29 autosomal STRs)

The results indicate significant dependencies!

- **Linkage**
  - Linkage is a consequence of the biological phenomenon recombination causing closely markers to be inherited as a unit from parent to child in a higher degree than for independent markers.
  - Estimate recombination rates from multi generation family pedigree or via genetic maps.
  - The "product rule" cannot be applied in all cases due to marker dependencies!

- **Linkage disequilibrium (allelic association)**
  - Exists when alleles at different loci occur together, at a population level, more (or less) often than expected by chance.
  - If LD, haplotype frequencies should be used rather than allele frequencies.

Genotype data/ (haplotype data) from 49 Swedish individuals.
Linkage analysis based on data from HapMap 3
  - Estimation of recombination rates.
LD analysis based on 49 swedish individuals AND 1000 Genomes project
  - Exact test.
  - SNAP (http://www.broadinstitute.org/mpg/snap)
  - ”Expected” LRs for different case scenarios
    - Simulations

Validation of the biostatistic features of the panel
(137 SNPs and 29 autosomal STRs)
Validation of the biostatistic features of the panel
(131 SNPs and 29 autosomal STRs)

- 6 SNPs removed due to sign of LD
- A genetic map was created

"Expected" LRs

<table>
<thead>
<tr>
<th>Case scenario</th>
<th>$H_1^{true}$ (29 STR)</th>
<th>$H_2^{true}$ (29 STR)</th>
<th>$H_1^{true}$ (29 STR+ 131 SNP)</th>
<th>$H_2^{true}$ (29 STR+ 131 SNP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paternity (trio) vs Unrelated</td>
<td>7.0e+016</td>
<td>1.8e+032</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Paternity/Maternity (duo) vs</td>
<td>3.7e+011</td>
<td>2.6e+021</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unrelated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full siblings vs Unrelated</td>
<td>4.2e+009</td>
<td>2.3e+018</td>
<td>4.0e-008</td>
<td>2.0e-017</td>
</tr>
<tr>
<td>Full siblings vs Half siblings</td>
<td>9.4e+002</td>
<td>1.2e+006</td>
<td>0.004637</td>
<td>3.4e-006</td>
</tr>
<tr>
<td>Paternity vs Uncle</td>
<td>6.9e+003</td>
<td>5.9e+007</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Workflow, case-by-case basis

NGS-data (e.i. reads) → Bioinformatics (genotype calling) → A/T T/T C/T Familias → Likelihood ratio (LR)

Case hypotheses → Population frequencies → Recombination rates

Results – Cases tested so far

(Apart from 40 parent-child tested during validation)

<table>
<thead>
<tr>
<th>Outcome (posterior prob)</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conclusive (&gt;99%)</td>
<td>13</td>
</tr>
<tr>
<td>Still inconclusive (5%-95%)</td>
<td>2</td>
</tr>
</tbody>
</table>
Results – Real cases

Case 1
Woman + Child
Question: Is the woman the mother, aunt, full sibling or unrelated to the child?

<table>
<thead>
<tr>
<th></th>
<th>Mother</th>
<th>Aunt</th>
<th>Full sibling</th>
<th>Unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td>29 A-STR</td>
<td>93%</td>
<td>0.02%</td>
<td>7%</td>
<td>&lt;0.001%</td>
</tr>
<tr>
<td>29 A-STR + 131</td>
<td>99.999%</td>
<td>&lt;0.001%</td>
<td>0.001%</td>
<td>&lt;0.001%</td>
</tr>
</tbody>
</table>

Results – Real cases

Case 2
Mother + Child + Man
Question: Is the man a half-uncle or unrelated to the child?

<table>
<thead>
<tr>
<th></th>
<th>Half uncle</th>
<th>Unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td>29 A-STR</td>
<td>87%</td>
<td>13%</td>
</tr>
<tr>
<td>29 A-STR + 131</td>
<td>99.2%</td>
<td>0.8%</td>
</tr>
</tbody>
</table>
Results – Real cases

Case 3
Woman + woman
Question: Are they half siblings?

<table>
<thead>
<tr>
<th></th>
<th>Half siblings</th>
<th>Unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td>29 A-STR</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>29 A-STR + 131 SNP</td>
<td>0.6%</td>
<td>99.4%</td>
</tr>
</tbody>
</table>

Posterior probability (equal priors)

Summary

• Validation
  • NGS typing methodology
  • Biostatistical workflow
• Implemented in may 2016.
• Have solved earlier ”unsolved” cases.
• Aiming for accreditation (17025) next year.