Stochastic filtering of quantitative data from STR DNA analysis

Tvedebrink, Torben; Eriksen, Poul Svante; Mogensen, Helle Smidt; Morling, Niels

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Stochastic filtering of quantitative data from STR DNA analysis

Torben Tvedebrink1, Paul Svante Enkens2, Helle Smidt Mogensen3, Niels Morling4

1 Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Denmark
2 Department of Mathematical Sciences, Aalborg University, Denmark
3 Forensic Science Group, Department of Forensic Genetics, University of Copenhagen, Denmark
4 European Forensic Science Institute, Denmark

Introduction

The quantitative data observed from analysing STR DNA is a mixture of contributions from various sources. Apart from the true allelic peaks, the observed signal consists of at least three components resulting from the measurement technique and the PCR amplification:

- Background noise (random noise due to the apparatus used for measurements).
- Pull-up effects (more systematic increase caused by overlap in the spectrum, see right picture of Fig. 2).
- Stutters (peaks located four basepairs before the true peak - are proposed to originate from primer mispairing).

We present filtering techniques for all three technical artifacts based on statistical analysis of data from controlled experiments conducted at The Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Denmark.

The filter

In the sections below, we describe the methods used, in filtering the observed data. The samples were prepared as described in [5] and the data were analysed using a threshold of 5 rfu on peak heights and with no stutter filter or any other pre-filtering.

Figure 2 shows a flow chart of the filter, and in Section “Noise filter”, we describe how, the signal detection limit is determined. This limit is also used as threshold when deleting signals. In Sections “Pull-up filter” and “Stutter filter”, the two remaining filters based on regression for pull-ups and stutters are presented.

Results

We have used our filter on 191 two-person mixtures. In Table 3, we have summarised the performance of the overall filter. It is worth emphasising that 179 of the 191 samples were dropped out and that the stutter filter only let 32 stutters (25 stutters and 7 backstutters) slip through. In addition to the stutter peaks, another 181 (128 drop-ins, 45 pull-up effects and 8 smears) on-ladder peaks were classified as proper peaks by the filter.

Table 3. Results for the overall filter. Smear is peaks < 1.5 bp from true peak.

<table>
<thead>
<tr>
<th>Classified peak</th>
<th>Noise</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>True alleles</td>
<td>179</td>
<td>5562</td>
</tr>
<tr>
<td>Stutters</td>
<td>3287</td>
<td>25</td>
</tr>
<tr>
<td>Back-stutters</td>
<td>2221</td>
<td>7</td>
</tr>
<tr>
<td>On-ladder alleles</td>
<td>16798</td>
<td>128</td>
</tr>
<tr>
<td>Off-ladder observations</td>
<td>72961</td>
<td>275</td>
</tr>
<tr>
<td>Smax position</td>
<td>19392</td>
<td>545</td>
</tr>
<tr>
<td>On-ladder smear</td>
<td>2460</td>
<td>8</td>
</tr>
<tr>
<td>Off-ladder smear</td>
<td>26642</td>
<td>537</td>
</tr>
<tr>
<td>Pull-up peaks</td>
<td>8034</td>
<td>198</td>
</tr>
<tr>
<td>On-ladder pull-up peaks</td>
<td>1753</td>
<td>45</td>
</tr>
<tr>
<td>Off-ladder pull-up peaks</td>
<td>6282</td>
<td>153</td>
</tr>
</tbody>
</table>

Discussion

Three times the standard deviation was also used in [6] for determining the limit of detection (LOD). However, the parameters μ and σ in [6] were based on negative controls and reagent blank samples, which implies that the parameters were computed for callabilities not containing the actual sample itself. This does not take the possible differences between callabilities within a batch into account.

An advantage of the locus specific threshold is that it enables the case worker to assess the noise level of the sample. Furthermore, in cases where the distribution of the transformed peak heights deviates substantially from normality, the sample may be subject to extensive noise and/or contamination of some sort.

Conclusion

The methodology of regression and distributional analysis of the noise yielded satisfactory results in order to deduce a stochastic filter for STR DNA samples. Comparisons of the results with those based on the recommendations of the manufacturer, 262 drop-outs were observed together with 26 stutters and 14 drop-ins.

References