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 Tvedebrink, Paul Svante Enkens, Helle Smidt Mogensen, Niels Morling

Abstract

The methodology of regression and distributional analysis of the noise yielded satisfying 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 indicated that the number of drop-outs decreased by approximately 30%. Studies of different data sets supported this improvement and suggests that the methodology of the threshold determination is adequate for the noise filtering of quantitative STR data.

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. 1)
- Stutters (peaks located four basesaps before the true peak - are proposed to originate from primer mispairings)

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 method of pre-filtering.

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

<table>
<thead>
<tr>
<th>Classification</th>
<th>Noise</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>True alleles</td>
<td>379</td>
<td>5562</td>
</tr>
<tr>
<td>Stutters</td>
<td>328</td>
<td>25</td>
</tr>
<tr>
<td>Back-stutters</td>
<td>2231</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>7296</td>
<td>275</td>
</tr>
<tr>
<td>Smar position</td>
<td>13932</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>26442</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>

The remaining peaks passing the filter were all detected to be off-ladder and thus removed from the analysed data. The results were also analysed following the standard protocol of the Section of Forensic Genetics at University of Copenhagen. Using the technique recommended by the manufacturer, 262 drop-outs were observed together with 26 stutters and 14 drop-ins.

Discussion

Three times the standard deviation was also used in [4] for determining the limit of detection (LOD). However, the parameters \( \kappa \) and \( \sigma \) in [4] were based on negative controls and reagent blank samples, which implies that the parameters were computed for capillaries not containing the actual sample itself. This does not take the possible differences between capillaries 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 satisfying results in order to deduce a stochastic filter for STR DNA samples.

References