Analytical thresholds for Yfiler® Plus

Mikkel Meyer Andersen[1,*], Helle Smidt Mogensen[2], Poul Svante Eriksen[1] and Niels Morling[2]

1 Dept. of Mathematical Sciences, Aalborg University, Denmark
2 Sect. of Forensic Genetics, Dept. of Forensic Medicine, University of Copenhagen, Denmark
* mikl@math.aau.dk

Introduction

- Analytical thresholds affect the balance between drop-outs of true alleles and drop-ins of false alleles
- Most softwares do not allow for sample dependent thresholds, hence average thresholds for all loci are commonly used
- AB3500/AB3500xl Genetic Analyzers (Life Technologies): Need for dye dependent thresholds since the signal intensities vary between the fluorescent dyes

Data

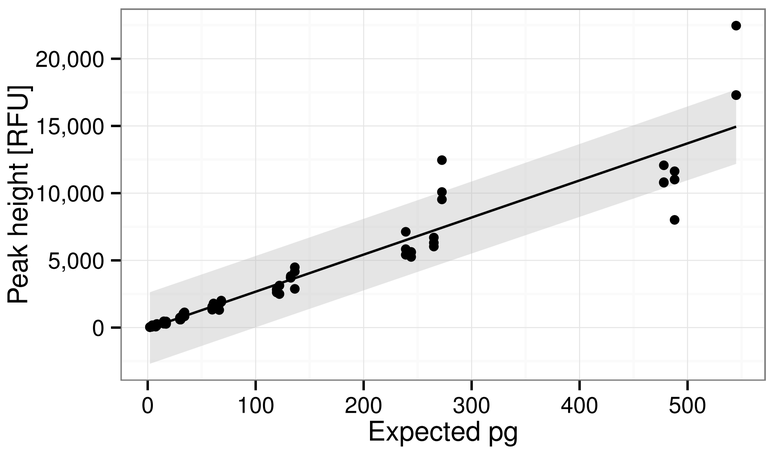
Dilution series: DNA from 4 males in 8 dilutions (F_1-F_8) with each dilution run in triplicates (n = 96 samples).

DNA was extracted from blood samples using EZ1 Investigator (Qiagen), quantified using Quantifiler® Hum (Life Technologies), amplified for 30 cycles using Yfiler®Plus (Life Technologies) and electrophoresed on an AB3500xl (Life Technologies). Results were analyzed with GeneMapper IDX 1.4 with 5 RFU threshold.

Quantified DNA in the non-diluted samples (F_1): Male A: 265 pg; Male B: 488 pg; Male C: 545 pg; Male D: 478 pg

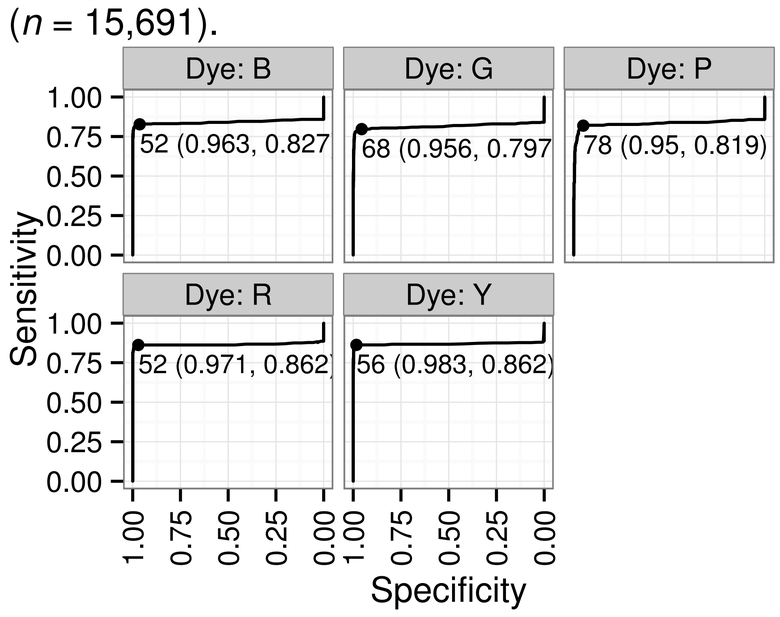
DNA amount vs peak height

- Relationship between expected pg (dilution factor, $1/2^{d-1}$ for dilution F_d, times quantified DNA in the non-diluted samples) and sample mean allele peak height
- The hatched area is a 95% prediction interval
- The adjusted R^2 values of the linear regression model is 0.816



ROC

We know profiles of the males, hence all peaks can get a label: 'Signal' (n = 2,306) or 'Noise' (n = 15,691).



- Receiver operating characteristic (ROC) curve on separating signal (allele) from noise (specificity and sensitivity have same weight)
- **Sensitivity** is true positive rate (the allele dropped out, how good are we to detect this?)
- **Specificity** is true negative rate (the allele did not drop out, how good are we to detect this?)
- The labels are thresholds followed by the coordinates for that particular point (specificity, sensitivity)

Dye	Threshold	Threshold rounded[1]	Spec[2]	Sens[3]
B	52 RFU	50 RFU	0.963	0.827
G	68 RFU	75 RFU	0.956	0.797
P	78 RFU	75 RFU	0.950	0.819
R	52 RFU	50 RFU	0.971	0.862
Y	56 RFU	50 RFU	0.983	0.862

[1] Rounded to nearest 25, [2] Spec = Specificity, [3] Sens = Sensitivity

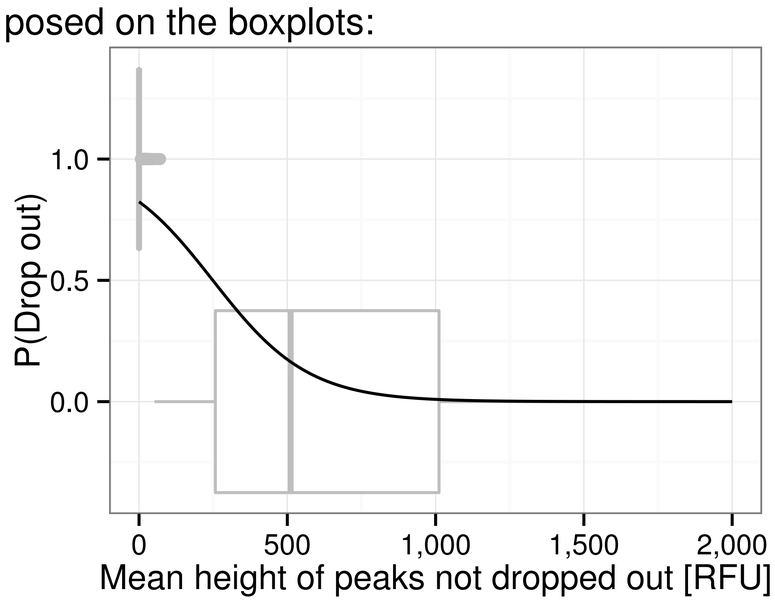
Bootstrap

- 10,000 random samples of all peaks with replacement (gives a new sample of same size)
- For each, we determined the dye thresholds with ROC (and rounded to nearest 25)
- Almost same as before except B (50 vs 75), but both 50 and 75 have large probabilities (43% and 53%, resp.)

Drop-out

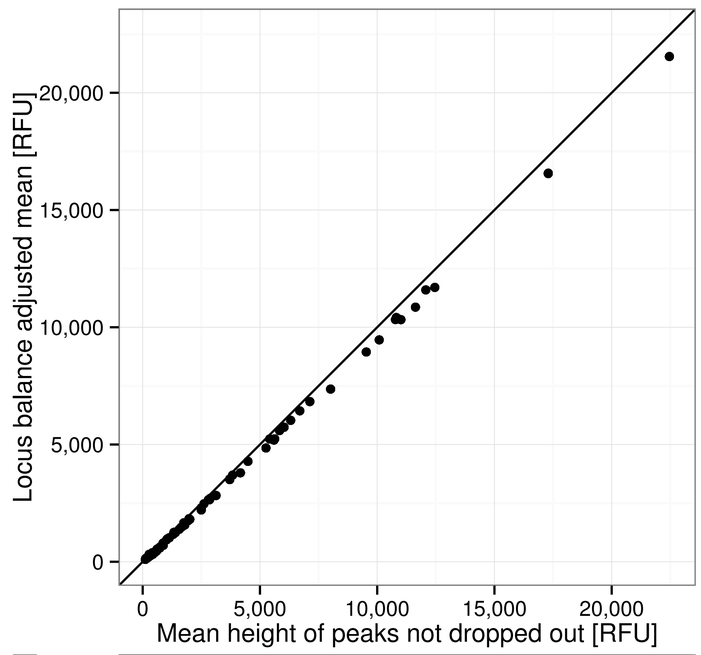
- Drop-out: When a peak has a height less than the dye specific threshold
- Fitted logistic regression model
- Full model with main effects and all first order (pairwise) interactions between explanatory variables: locus, dye and mean peak height (of the peaks not dropped out in each profile)
- Full model was reduced according to Akaikes Information Criterion (AIC)
- Final model consisted only of mean peak height of allele peaks as explanatory variable (no locus nor dye)

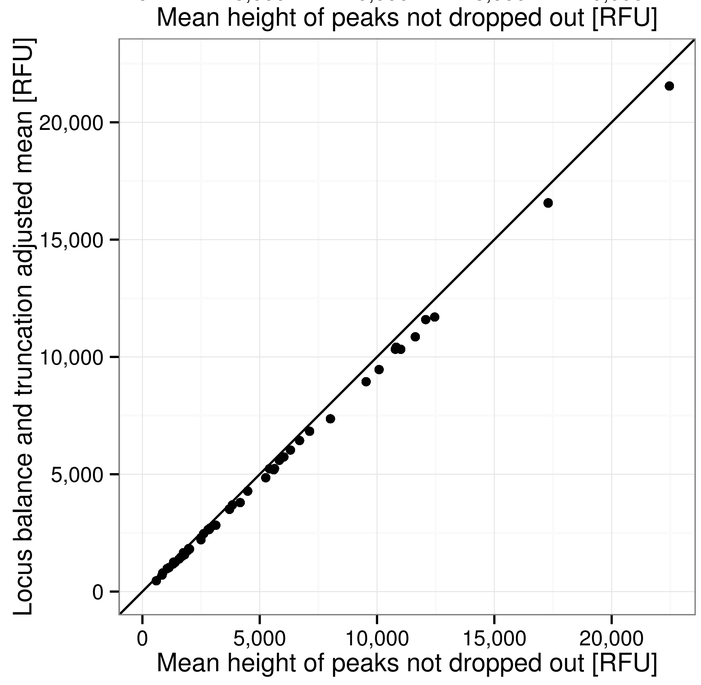
The logistic regression predictions are superimposed on the boxplots:



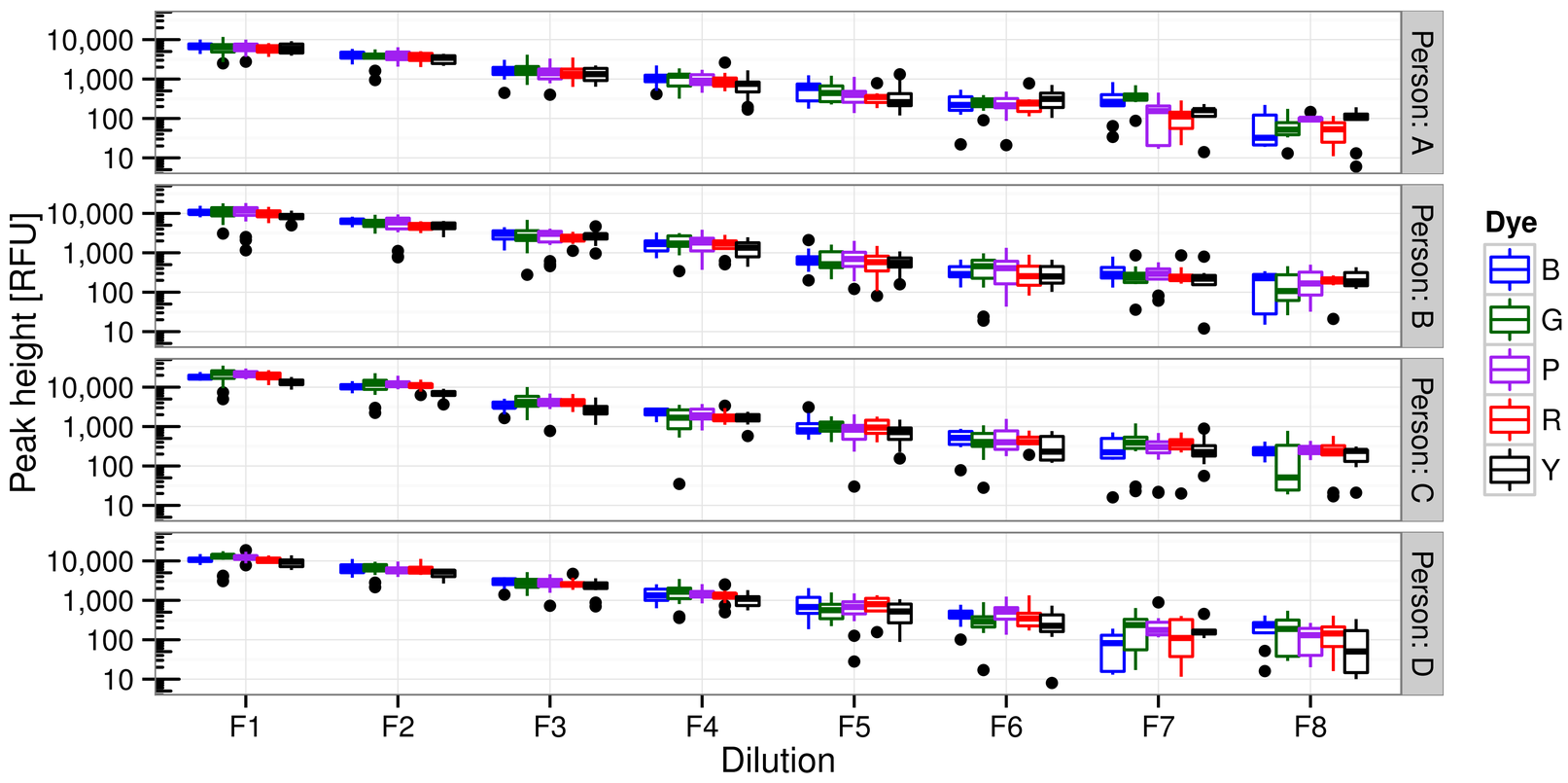
Locus balances

Correcting for locus balances as described in [1] did not make a big difference on signal strength estimates:





Allele peak height distribution for each male, dilution and dye



Each boxplot shows all allele peaks from all loci and all three replicates for the particular male, dilution and dye

References

[1] Andersen, M. M., Mogensen, H. S., Eriksen, P. S., Olofsson, J. K., Asplund, M. and Morling, N. (2013) Estimating Y-STR allelic drop-out rates and adjusting for interlocus balances. *Forensic Science International: Genetics*, 7, 327–336