J Forensic Sci, January 2007, Vol. 52, No. 1
doi:10.1111/j.1556-4029.2006.00318.x
Available online at: www.blackwell-synergy.com

TECHNICAL NOTE

Jason R. Gilder,[1] *M.S.; Travis E. Doom,*[2] *Ph.D.; Keith Inman,*[3] *M. Crim.; and Dan E. Krane,*[4] *Ph.D.*

Run-Specific Limits of Detection and Quantitation for STR-based DNA Testing

ABSTRACT: STR-based DNA profiling is an exceptionally sensitive analytical technique that is often used to obtain results at the very limits of its sensitivity. The challenge of reliably distinguishing between signal and noise in such situations is one that has been rigorously addressed in numerous other analytical disciplines. However, an inability to determine accurately the height of electropherogram baselines has caused forensic DNA profiling laboratories to utilize alternative approaches. Minimum thresholds established during laboratory validation studies have become the *de facto* standard for distinguishing between reliable signal and noise/technical artifacts. These minimum peak height thresholds generally fail to consider variability in the sensitivity of instruments, reagents, and the skill of human analysts involved in the DNA profiling process over the course of time. Software (BatchExtract) made publicly available by the National Center for Biotechnology Information now provides an alternative means of establishing limits of detection and quantitation that is more consistent with those used in other analytical disciplines. We have used that software to determine the height of each data collection point for each dye along a control sample's electropherogram trace. These values were then used to determine a limit of detection (the average amount of background noise plus three standard deviations) and a limit of quantitation (the average amount of background noise plus 10 standard deviations) for each control sample. Analyses of the electropherogram data associated with the positive, negative, and reagent blank controls included in 50 different capillary electrophoresis runs validate that this approach could be used to determine run-specific thresholds objectively for use in forensic DNA casework.

KEYWORDS: forensic science, DNA typing, threshold, minimum peak height, limit of detection, limit of quantitation, bioinformatics

STR-based DNA profiling methodology is effectively at the theoretical detection limit in that typable results can be generated from as little starting material as a single cell (1,2). However, one of the most challenging aspects of forensic DNA analysis is the interpretation of low-level testing results where it is difficult to distinguish reliably between noise and signal from template DNA that is associated with an evidence sample (3,4). This difficulty with minimal samples is often compounded by the consumptive nature of PCR-based DNA testing (5,6) when material is unavailable for replicate testing. Forensic DNA testing laboratories typically endeavor to minimize the effect of baseline noise and stochastic artifacts by relying upon very conservative minimum peak height thresholds (commonly fixed in the range of 50–200 relative fluorescent units [RFUs]) that are established during the course of their validation processes (7–10). However, the conservative nature of these commonly used thresholds can also arbitrarily remove from consideration legitimate signal from trace and secondary contributors to an evidentiary sample—matters of critical importance in many criminal investigations.

Any measurement made with a light-detecting instrument, such as a genetic analyzer, is subject to at least some level of background noise (11)—defined here as a signal not associated with amplified DNA. Instrument-related factors that may contribute to background noise in DNA-testing experiments are typically run-specific and include (but are not necessarily limited to) the age and condition of the polymer and capillary being used; dirty capillary windows; and dirty pump blocks (12). Background noise may also differ between instruments due to differences in charged couple device detectors, laser effectiveness and alignment, and cleanliness and alignment of the optical components (10). Many amplification-related factors that contribute to background noise (such as analyst skill and stocks of chemicals) are also run-specific and might be reasonably expected to have varying impacts over time.

Many analytical disciplines aside from forensic DNA profiling have needed to account rigorously for background noise mixed with low levels of signal (13,14). In the uncommon circumstances where background noise occurs at a constant level, it can simply be subtracted from an analyzed signal to obtain true measurements of the tested material (11). It is much more common, however, for background noise, such as that associated with DNA testing results, to not be constant. In these instances, it is commonly assumed that noise magnitude is independent of analyte signal and that noise levels are distributed in a Gaussian fashion that can be effectively characterized with a mean and a standard deviation (11,13–15). Two different signal-to-noise thresholds can be readily derived from the mean (μ) and standard deviation (σ) of the noise levels from a particular test and instrument: a limit of detection (LOD) and a limit of quantitation (LOQ) (11,13–15). The LOD is the smallest quantity of analyte that the analytical process can reliably detect. LOD is expressed as a statistical

[1]Forensic Bioinformatics Inc., 2850 Presidential Drive, Suite 150, Fairborn, OH 45324.
[2]Department of Computer Science and Engineering, Wright State University, Dayton, OH 45435-0001.
[3]Forensic Analytical Specialties Inc., 3777 Depot Rd., Suite 409, Hayward, CA 94545.
[4]Biological Sciences, Wright State University, Dayton, OH 45435-0001.

Received 23 April 2006; and in revised form 14 July 2006; accepted 5 Aug. 2006; published 8 Dec. 2006.

confidence limit of noise error, usually 99.7% (i.e., three standard deviations) or

$$LOD = \mu_b + 3\sigma_b \quad (1)$$

where μ_b is the average amount of background noise and σ_b is the standard deviation associated with that value (11,13–15). The LOQ represents the threshold beneath which measurements of signal strength cannot be reliably used to determine the relative quantity of detected analyte (e.g., because such measurements may include an appreciable amount of signal arising from background noise). LOQ is commonly expressed as the average background signal plus 10 standard deviations (11,13–15), or

$$LOQ = \mu_b + 10\sigma_b \quad (2)$$

Forensic DNA testing laboratories routinely test a positive control, negative control, and reagent blank with every DNA analysis run (7–9). While these controls are utilized primarily as sentinels for gross failures of the DNA-testing processes, such as cross-contamination of samples, as well as contamination or inappropriate activity of reagents, they also contain an abundance of subtle but important information about the running environment of the DNA-testing system—particularly as it pertains to background noise. In this technical note, we describe a methodology that invokes generally accepted practices from other analytical disciplines and uses information associated with those ubiquitous controls to establish objective run-specific electropherogram peak height thresholds.

Materials and Methods

Baseline Noise Determination

Data for this study were obtained from 50 STR-based DNA testing runs generated by four analysts working at Forensic Analytical Specialties Inc. (Hayward, CA) using the laboratory's validated standard protocols (e.g., no additional rounds of amplification were used as might be the case for low-copy-number analyses). All DNA profiles were generated with the Profiler Plus® commercial testing kit during the course of actual casework associated with approximately 150 cases conducted between 2004 and 2006. Each run was performed on the same Applied Biosystems 310 Genetic Analyzer and contained a positive control; a negative control; and a reagent blank. A positive control consisted of template DNA from the 9947A immortal lymphoid cell line (16). This positive control DNA is provided by the manufacturer of the test kit and its STR genotype is well characterized. Negative controls begin at the amplification step and contain all of the reagents used for amplification (but no template DNA). A reagent blank is a sample that contains all of the reagents used from the beginning of the extraction of a sample through amplification and typing, but again containing no template DNA. When a single run contained more than one injection of a given control, the last injection was used. No other information associated with a run (e.g., that associated with reference or evidentiary samples) was used. Electronic data files associated with these control samples (with any case-specific information removed) are available on the Internet at: www.bioforensics.com/baseline/baseline.zip.

The National Center for Biotechnology Information's (NCBI) BatchExtract software (17) was used to obtain the trace and peak data from Applied Biosystem's GeneScan® sample files. BatchExtract provides the height (in RFUs) of each data-collection point (DCP) for each dye along a sample's electropherogram trace. BatchExtract also provides additional information associated with labeled peaks, including the DCPs where GeneScan® considered peaks to begin and end. DCP regions containing an ROX size standard peak were excluded (masked) from consideration in all dye colors to avoid any complications from spectral overlap artifacts (i.e., pull-up) (3,4). A total of 296,592 DCPs associated with the 50 negative controls ($\mu = 5932$ DCP/run, $\sigma = 131$ DCP) and 297,315 DCPs associated with the 50 reagent blank controls ($\mu = 5946$ DCP/run, $\sigma = 87$ DCP) remained for inclusion in subsequent analyses after masking was completed. Similarly, DCP regions (+ and − 55 DCPs to account conservatively for potential stutter artifacts) associated with the expected alleles for the 9947A immortal lymphoid cell line (16) were also masked in all dye colors for positive control samples. 120,762 DCPs associated with the 50 positive controls ($\mu = 2415$ DCP per run, $\sigma = 198$ DCP) remained for inclusion in subsequent analyses after masking was completed. Shareware that performs these analyses (including masking) on the output of BatchExtract is available at www.bioforensics.com (18).

Test Mixture

A two-person mixture was created by combining the genomic DNA of two unrelated individuals with known genotypes in a ratio of approximately 10 to 1. The major contributor was known to be a female with the following STR-DNA profile: D3S1358 18, 18; vWA 16, 19; FGA 20, 21; D8S1179 13, 15; D21S11 32.2, 32.2; D18S51 15, 17; D5S818 11, 12; D13S317 11, 11; and D7S820 8, 10. The secondary contributor was known to be a male with an STR-DNA profile of: D3S1358 13, 17; vWA 17, 18; FGA 22, 24; D8S1179 11, 11; D21S11 28, 30; D18S51 12, 19; D5S818 11, 13; D13S317 10, 11; and D7S820 11, 12. The electropherograms for the mixed sample were generated with the same Applied Biosystems 310 Genetic Analyzer and protocols as those used to generate the control samples described above.

Results

The distribution of baseline RFU level at each nonmasked DCP was generally Gaussian for each of the 50 analyzed negative, reagent blank, and positive controls (Fig. 1). Histograms displaying the distribution of all three controls for all 50 runs included in this analysis can be found online at www.bioforensics.com/baseline/baseline.zip. Differences in the average baseline levels within each of the 50 analyzed runs were small between negative and positive control samples (with an average difference of the averages of only 0.60 RFUs). Differences in the average baseline levels within each of the 50 analyzed runs were similarly small between negative and reagent blank controls (with an average difference of μ_b values of 0.41 RFUs) and between positive and reagent blank samples (with an average difference of μ_b values of 0.46 RFUs). While the inferred LOQ thresholds for all three controls were very similar within runs, average background noise values (μ_b) and standard deviations (σ_b) varied substantially between runs (Table 1), such that $\mu_b + 10\sigma_b$ (LOQ thresholds) derived from positive controls, negative controls, and reagent blank controls ranged from 27.7 to 75.7; 30.0 to 145.4; and 30.0 to 116.5 RFUs, respectively.

All of the combined average LOD and LOQ fall below 100 RFUs. Baseline values were found to be generally homogeneous in that the minimum and average LOD and LOQ were within three standard deviations of each other for each of the 150 analyzed controls. The maximum values for μ_b were generally similar in

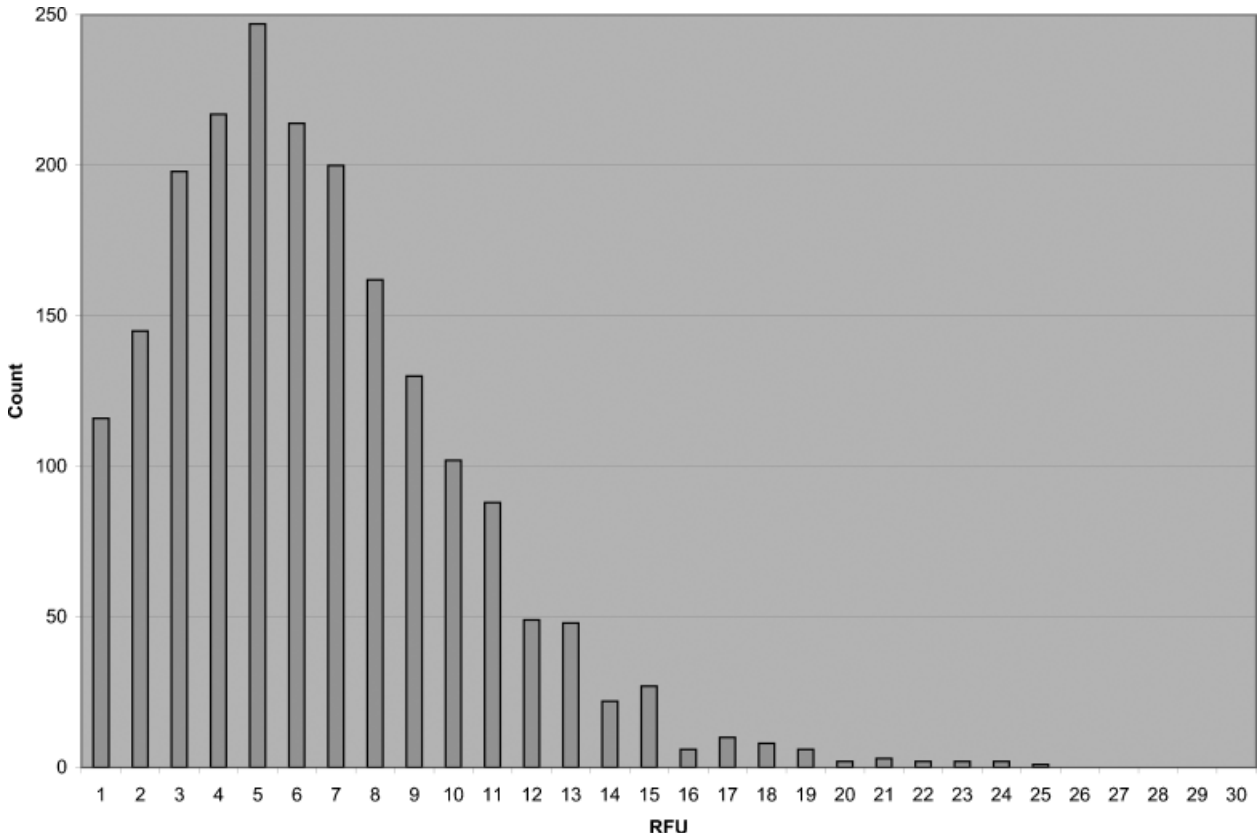

FIG. 1—*Representative histogram taken from the distribution of measured relative fluorescent unit (RFU) levels at all nonmasked data-collection points in the first of 50 negative control samples after masking. This distribution is from a blue channel and exhibits an average baseline approximately equal to that of the population's average baseline signal (5.5 RFUs).*

each of the three different control types, with a maximum-observed difference within a run of only 8.8 RFUs (between a negative control and positive control).

Single averages and standard deviations for each of the 50 analyzed runs were also generated by considering all DCP values for a run together (i.e., independent of which of the three different controls they came from). Standard deviations for these larger data sets were generally smaller than those observed when each of the three controls was considered separately although the calculated LOD and LOQ values were very similar to those obtained by considering the three controls for runs separately (Table 1).

When considering the dye channels separately, the green channel, on average, exhibited the highest amount of baseline signal and the yellow channel exhibited the least (Table 2). The negative control containing the highest average baseline in the green channel exhibited a uniformly elevated baseline and was responsible for the single highest-observed LOQ (Table 1). The LOQ determined for this sample with information from all three color channels (145.4 RFUs) was found to be more conservative than the LOQ determined from the green channel alone (89.2 RFUs) such that no noise in the green channel would have been confused with signal.

TABLE 1—*Maximum, minimum, and average baseline levels observed in the set of reagent blanks, negative controls, and positive controls (determined from controls in 50 different runs).*

	μ_b	σ_b	$\mu_b+3\sigma_b$	$\mu_b+10\sigma_b$
Positive Control				
Maximum	6.7	6.9	27.4	75.7
Average	5.0	3.7	16.1	42.0
Minimum	3.7	2.4	10.9	27.7
Negative Control				
Maximum	13.4	13.2	53.0	145.4
Average	5.4	3.9	17.1	44.4
Minimum	4.0	2.6	11.8	30.0
Reagent Blank				
Maximum	6.5	11.0	39.5	116.5
Average	5.3	4.0	17.3	45.3
Minimum	4.0	2.6	11.8	30.0
All three controls averaged				
Maximum	7.1	7.3	29.0	80.1
Average	5.2	3.9	16.9	44.2
Minimum	3.9	2.5	11.4	28.9

All values are in RFUs.

TABLE 2—*Maximum, minimum, and average baseline levels observed in each of three color channels for reagent blanks, negative controls, and positive controls.*

	Minimum	Average	Maximum
Positive Control			
Blue	3.7	5.2	9.7
Green	4.3	5.8	7.4
Yellow	3.0	4.1	6.4
Negative Control			
Blue	4.0	5.3	8.0
Green	4.6	6.7	31.2
Yellow	3.0	4.0	6.4
Reagent Blank			
Blue	3.7	5.4	8.5
Green	4.8	6.2	8.6
Yellow	3.4	4.3	6.1

All values are in RFUs.

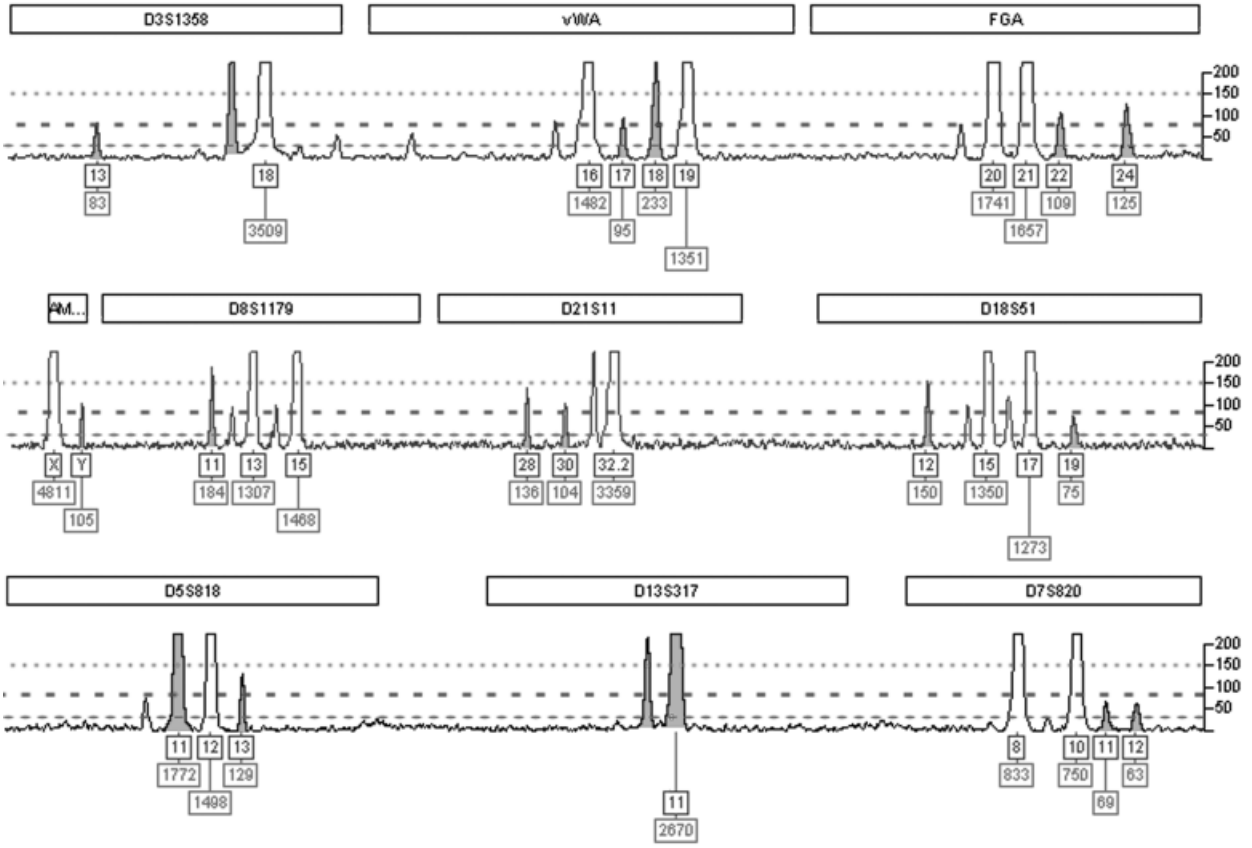

FIG. 2—*Electropherograms from an approximately 10:1 mixture of two reference samples. Three different thresholds are shown: a minimum peak height threshold at 150 relative fluorescent unit (RFU) (dotted line); a limit of quantitation (LOQ) threshold determined to be at 77 RFUs from the negative control for this electrophoresis run (dashed line); and a limit of detection (LOD) threshold determined to be at 29 RFUs for this electrophoresis run (small dashed line). Genotyper® assigned allele calls (with ABI stutter filters in place) are shown in boxes immediately below the electropherogram peaks while peak heights (in RFUs) are shown in boxes below those labels for all peaks with heights greater than the LOD. Peaks consistent with the known profile of the minor contributor are shaded.*

A known-mixed DNA profile from two unrelated individuals of an approximately 10:1 ratio was also examined using this methodology (Fig. 2). The negative control tested in the same analysis run as the mixture yielded a LOD of 29 RFUs and a LOQ of 77 RFUs. Eleven alleles (including the Y allele at the amelogenin locus) associated with the known DNA profile of the minor contributor were not labeled for this mixed sample when the GeneScan® default threshold of 150 RFUs was used. Eight alleles (including the Y allele at the amelogenin locus) associated with the male secondary contributor fall between the LOQ and the commonly used 150 RFU threshold. Similarly, three additional alleles associated with the secondary contributor fall between the LOD and the LOQ thresholds. The 17 allele (347 RFUs) at the D3 locus (which is in a stutter position relative to the major contributor's 3509 RFU 18 allele at that locus) and the 10 allele (210 RFUs) at the D13 locus (which is in a stutter position relative to the major contributor's 2670 RFU 11 allele at that locus) are the only alleles of the secondary contributor that are not labeled by Genotyper® when the threshold is set to the LOD inferred from the negative control (29 RFUs; Fig. 2). Two peaks with heights greater than the LOD that were observed in the blue channel were associated with pull-up from the green channel and were not considered.

Discussion

The similarity of the baseline levels of samples that were expected to have a high signal amplitude arising from analyte (template DNA in the positive controls) and those expected to contain little or no analyte (the negative and reagent blank controls) indicates that noise magnitude in STR-based DNA testing is independent of the analyte signal. Baseline levels for each of the three different standard controls included in each DNA profiling electrophoresis run were also very similar within runs, but differed widely between runs. These observations suggest that the baseline noise associated with capillary electrophoresis of DNA profiles is comparable with that encountered in other analytical endeavors and that generally accepted means of determining LOD and LOQ can be applied.

The samples analyzed in this study were primarily positive, negative, or reagent blank controls. It should be possible to evaluate evidentiary or reference samples included in the same capillary electrophoresis run with the LOD and LOQ values inferred from these controls. Any peaks in evidentiary or reference samples that exceed these thresholds (such as those associated with the secondary contributor in the mixture containing DNA of two unrelated individuals with known STR-DNA profiles; Fig. 2) are unlikely to be due to baseline noise. All peaks above the threshold would then require evaluation to ascertain whether they were signal from amplified genomic DNA, or whether they may have originated from technical artifacts such as pull-up, voltage spikes, or stutter.

It is worth noting that the maximum range of LOD thresholds (10.9–53.0 RFUs; Table 1) determined with this method in these 50 runs associated with casework performed by Forensic Analytical Specialties, Inc. is substantially below the minimum peak height threshold of 100 RFUs established by the laboratory during the course of their validation studies. Disregarding information

associated with electropherogram peaks well above an analytical threshold of detection (and even above an analytical threshold of quantitation) might be considered abundantly conservative under some circumstances, given that DNA testing is a very sensitive process subject to a variety of technical artifacts such as pull-up, voltage spikes, and stutter. However, in this abundance of caution, valid information about the presence of real DNA peaks is being discarded or ignored. In the instance of the mixture of two individuals with known STR-DNA profiles (Fig. 2), the lower levels of the LOQ and LOD allowed reliable recognition of alleles arising from the genomic DNA of a secondary contributor while the commonly used 150 RFU minimum peak height threshold did not. In some investigations (e.g., a mixture of a victim and perpetrator that was small enough to require consumption of the entire sample), the observation of alleles associated with a secondary contributor using the LOD threshold methodology described here could constitute critically important information that would have not been available if only conservative minimum peak height thresholds were used.

LOD and LOQ thresholds can be used to distinguish reliably between noise and legitimate DNA signal. Two approaches can be adopted with data gathered from intra-laboratory collection of baseline data. The first is to use average LOD and LOQ thresholds derived from both validation and current casework samples. These values could be constantly updated. A second approach would involve the determination of LOD and LOQ values for every run for use with the other samples within that run. In either case, empirical statistically derived values provide a more rigorous discrimination between data contributed by noise and data derived from human DNA. Thorough analyses of the data pertaining to baseline noise in control samples with software such as NCBI's BatchExtract may help draw the attention of analysts to other important issues as well. For instance, if one of the three control samples for a given run exhibits a larger average and/or standard deviation of baseline levels than the others, it may be an indication that that sample (and, perhaps the run with which it is associated) should be evaluated with greater care. Similarly, controls with elevated average and standard deviations of baseline activity might indicate the need for maintenance or replacement of reagent stocks. BatchExtract is a freely available program (17) and its output can be used with Forensic Bioinformatics' free baseline analysis program to determine the LOD and LOQ for any control sample (18).

References

1. Findlay I, Taylor A, Quirke P, Frazier R, Urquhart A. DNA fingerprinting from single cells. Nature 1997;389:555–6.
2. Oorschot RAV, Jones MK. DNA fingerprints from fingerprints. Nature 1997;387:767.
3. Thompson WC, Ford S, Doom T, Raymer M, Krane DE. Evaluating forensic DNA evidence: essential elements of a competent defense review Part 1. Champion 2003;27(3):16–25.
4. Thompson WC, Ford S, Doom T, Raymer M, Krane DE. Evaluating forensic DNA evidence: essential elements of a competent defense review Part 2. Champion 2003;27(4):24–8.
5. Leclair B, Sgueglia JB, Wojtowicz PC, Juston AC, Frégeau CJ, Fourney RM. STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes. J Forensic Sci 2003; 48(5):1001–13.
6. Frégeau CJ, Bowen KL, Leclair B, Trudel I, Bishop L, Fourney RM. AmpFLSTR® Profiler Plus™ short tandem repeat DNA analysis of casework samples, mixture samples, and non-human DNA samples amplified under reduced PCR volume conditions (25 μL). J Forensic Sci 2003;48(5):1014–34.
7. DNA Advisory Board Quality assurance standards for forensic DNA testing laboratories. Forensic Sci Comm 2000;2(3).
8. Federal Bureau of Investigation (FBI) Laboratory. National DNA Index System (NDIS) data acceptance standards. May 2005. See website http://forensics.marshall.edu/NEST/Nest%20PDFs/Documents/AppendB-NDIS-0505.pdf.
9. Scientific Working Group on DNA Analysis Methods (SWGDAM). Short Tandem Repeat (STR) interpretation guidelines. Forensic Sci Comm 2000;2(3).
10. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Brown AL, Budowle B. Validation of STR typing by capillary electrophoresis. J Forensic Sci 2001;46(3):661–76.
11. Rubinson KA, Rubinson JF. Sample size and major, minor, trace, and ultratrace components. Contemporary instrumental analysis. Upper Saddle River: Prentice Hall, 2000:150–8.
12. Applied Biosystems, Inc. (ABI). Chemistry reference for the ABI Prism® 310 genetic analyzer. Foster City: Applied Biosystems Inc., 2000.
13. Anderson N. Determination of the lower limit of detection [Letter]. Clin Chem 1989;35:2152–3.
14. Thomsen V, Schatzlein D, Mercuro D. Limits of detection in spectroscopy. Spectroscopy 2003;18(12):112–4.
15. Arinbruster DA, Tillman MD, Hubbs LM. Limit of detection (LOD)/limit of quantitation (LOQ): comparison of the empirical and the statistical methods exemplified with GC-MS assays of abused drugs. Clin Chem 1994;40:1233–8.
16. Frégeau CJ, Aubin RA, Elliott JC, Gill SS, Fourney RM. Characterization of human lymphoid cell lines GM9947 and GM9948 as intra- and inter-laboratory reference standards for DNA typing. Genomics 1995;28: 184–97.
17. National Center for Biotechnology Information (NCBI). BatchExtract. 2006. See website ftp://ftp.ncbi.nlm.nih.gov/pub/forensics/.
18. Forensic Bioinformatics Inc. LOQ: baseline analysis software. 2006. See website www.bioforensics.com/baseline/loq.zip.

Additional information and reprint requests:
Dan E. Krane, Ph.D.
Biological Sciences
3640 Colonel Glenn Highway
Dayton, OH 45435-0001
E-mail: dan.krane@wright.edu