*J Forensic Sci,* July 2004, Vol. 49, No. 4
Paper ID JFS2003336
Available online at: www.astm.org

*Kristy Kadash,*[1] *Ph.D., D-ABC; Brian E. Kozlowski,*[1] *M.F.S; Lisa A. Biega,*[2] *M.S.; and Barry W. Duceman,*[3] *Ph.D.*

# Validation Study of the TrueAllele® Automated Data Review System*

**ABSTRACT:** The New York State Convicted Offender DNA Databank is the first U.S. lab to complete an internal validation of the TrueAllele® expert data review system. TrueAllele® is designed to assess short tandem repeat (STR) DNA data based on several key features such as peak height, shape, area, and position relative to a standard ladder and use this information to make accurate allele calls. The software then prioritizes the allele calls based on several user-defined rules. As a result, the user need only review low-quality data. The validation of this system consisted of an extensive optimization phase and a large concordance phase. During optimization, the rule settings were tailored to minimize the amount of high-quality data viewed by the user. In the concordance phase, a large dataset was typed in parallel with the ABI software Gene Scan® and Genotyper® (manual review) and TrueAllele® (automated review) for comparison of allele calls and sample state assignment. Only one significant difference was discovered out of 2048 samples in the concordance study. In this case, TrueAllele® revealed a spike in the profile that was interpreted as a DNA peak by the analyst in Genotyper®. TrueAllele® was designed to focus the review on poor data and to eliminate the need for complete reanalysis technical review. This validation project proved TrueAllele® to be dependable for use at the NYS Convicted Offender DNA Databank.

**KEYWORDS:** forensic science, STR analysis, DNA databank, automated data review

The goal of the TrueAllele® system for convicted offender DNA databanking laboratories is to alleviate the shortage of skilled data reviewers by automating steps in the review process (1). Theoretically, this would decrease the amount of time needed to analyze scores of DNA profiles, thereby increasing efficiency. To be useful, the system must handle high throughput with minimal error.

The Forensic Science Service in Great Britain currently uses the system to process all convicted offender DNA profiles (2). The New York State Convicted Offender DNA Databank is the first U.S. lab to conduct an internal validation of this software for the purpose of using it to generate profiles from databank samples on the ABI 3700® platform for upload into the Combined DNA Index System (CODIS). This paper does not address the application of the TrueAllele® expert review system to the analysis of forensic samples.

The automated process begins with the creation of a datadisk that consists of the gel or capillary data, rule thresholds used for quality assessment, and user preferences including short tandem repeat (STR) panel, choice of size standard, nomenclature of extraction and amplification controls, and ladder preference. Each sample will be typed either with the closest ladder (capillary data) or the loading ladder (gel data). During the software subroutine Image Call, the size standards for each capillary are tracked and labeled. In Cap View, the user scans the entire run to verify that all the size standards have been defined correctly. Modifications of peak assignments can be made at this step if necessary. In Control Check, the user can subsequently use a number of built-in features to review the quality of the run. At this step, an entire run may be rejected if there is a problem with the positive and negative controls or reference ladders. During Allele Call, TrueAllele® identifies and quantitates peaks and then prioritizes the samples (on a per-locus basis) according to a set of user-defined rules. Table 1 gives a brief description of each rule and the optimized parameters used by the NYS Convicted Offender DNA Databank. The user is prompted to accept, edit, or reject potentially problematic calls in Allele View without reviewing calls that are given a high-quality score. Allele View consists of several panels that contain all the necessary information to enable the user to evaluate the allele calls, including the TrueAllele® designated genotypes, the electropherogram, and peak quantitative data. At this stage, the user can also decide how much of the data to review by selecting a maximum quality score threshold. When review is complete, TrueAllele® generates a number of reports containing information on accepted, edited, and rejected samples in formats compatible with CODIS or a Laboratory Information Management System. The development of this automated genotyping procedure has been published previously (2–6).

To maximize the performance of TrueAllele®, it must be adapted to the laboratory's sample processing system while adhering to established quality-control guidelines. This is accomplished with an "optimization" period. The end result should be a set of rule parameters that can be used confidently by the laboratory to generate accurate and reproducible DNA profiles. These parameters are then compared with Gene Scan® and Genotyper® in an extensive concordance study to ensure that the outcomes of each analytical technique are equivalent before adopting TrueAllele®.

[1] Forensic Scientist, New York State Convicted Offender DNA Databank, New York State Police Forensic Investigation Center, 1220 Washington Ave., Albany, NY.

[2] Supervisor of DNA Services, New York State Convicted Offender DNA Databank, New York State Police Forensic Investigation Center, 1220 Washington Ave., Albany, NY.

[3] Director of Biological Science, New York State Police Forensic Investigation Center, 1220 Washington Ave., Albany, NY.

* This work has been presented orally at the following meetings: 28th Annual Meeting of the Northeast Association of Forensic Scientists, Atlantic City, NJ, Nov. 2002; SWGDAM Annual Meeting at FBI Headquarters, Quantico, VA, Jan. 2003; and at the 55th Annual Meeting of the American Academy of Forensic Scientists, Chicago, IL, Feb. 2003.

Received 1 Oct. 2003; and in revised form 18 Dec. 2003; accepted 1 Feb. 2004; published 26 May 2004.

Copyright © 2004 by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959.

TABLE 1—*Rule descriptions and optimized parameters.*

| Rule Name | Rule Fires When: | Action | Threshold |
|---|---|---|---|
| Amelo | Allele designations at this locus are other than XX or XY. | ON | n/a |
| Check control | Allele calls for positive control sample do not match expected designations. | ON | n/a |
| Conflict | TrueAllele's two allele calling algorithms (SVD and enumeration) do not agree on the designations. | ON | n/a |
| Crossover | An allele for one locus may have crossed over into an adjacent locus window of the same color. | ON | n/a |
| Dispersion | The designated peaks contain less than a certain percentage of the total signal in the allelic window. | ON | 60% |
| Dye to dye | Extra peaks in window may be caused by other dyes at that location (aka pull-up or bleed-through). | OFF | n/a |
| Extra allele | There are more than two significant peaks within the allelic window. | ON | 20% |
| High n peak | The height of the n (−A) peak is too high relative to the n + 1 (+A) peak. | ON | 30% |
| High signal | The height of one or more peaks exceeds the user-defined limit. | ON | 20,000 RFU |
| Lane to lane | The peaks from a neighboring lane/capillary are visible in the current window. | OFF | |
| Low homozygote | The height of the single designated peak in the window is too low. | ON | 250 RFU |
| Low signal | The height of one or more peaks is less than the user-defined limit. | ON | 150 RFU |
| Negative | Peaks were detected in the designated negative control lanes. | ON | n/a |
| New allele | One or more designated alleles are novel and have not been seen before. Add to the list of known alleles. | ON | n/a |
| Noise | No peaks were detected in the sample at this locus. | ON | n/a |
| Off ladder | One or more designated alleles are too far away from the physical or virtual ladder alleles. | ON | 0.45 bp |
| Off physical ladder | One or more designated alleles do not fall on the physical ladder (do not consider virtual ladder alleles). | ON | n/a |
| Overlap | Two alleles from different loci in different dyes are nearly the same size. Check if true alleles or bleed-through. | ON | n/a |
| Peak morphology | The designated peaks have peak fit quality less than a user-defined threshold. | ON | 20% |
| Rare | One or more peaks are designated as X.1 or X.3. | ON | n/a |
| Relative area | The area of the smaller peak is less than the expected amount relative to the larger peak. | ON | 50% |
| Relative height | The height of the smaller peak is less than the expected amount relative to the larger peak. | ON | 20% |
| Stutter | The area of the stutter peak is too high relative to the peak that is one repeat larger. | ON | 15% |
| Third peak | There is an extra peak visible just outside of the allelic window. | ON | 15% |
| Uncorrelated | The two designated peaks appear to migrate in opposite directions relative to the ladder. | ON | 0.4 bp |
| Unexpected | One or more peaks were found outside of the allelic windows (similar to Extra Allele and Third Peak rules). | ON | 20% |

## Materials and Methods

### *Optimization Phase*

The validation consisted of an extensive optimization phase followed by a large concordance phase. Prior to the initiation of TrueAllele® optimization, preliminary ABI 3700® runs were sent to Cybergenetics, Inc. Their review of these data yielded a set of preliminary rule parameters in which thresholds were set intentionally low. For the first round of optimization, a set of ABI 3700® data was processed through TrueAllele® using this template. This ensured that no sample of even slightly poor quality was given a Quality Score >0.000 (i.e., fire no rules). A user reviewed every allele call in the datadisk and made necessary changes to reflect the desired future outcome by using the User column of the Rule Analysis window. The example in Fig. 1 shows a locus with raised baseline in the Electropherogram window of Allele View (Fig. 1*a*; ladder peaks are in the background). There is a main peak at Allele 16 with several smaller "potential" peaks. TrueAllele® fired three rules for this sample, as shown in the TA column of the Rule Analysis window (Fig. 1*b*). The same rules are initially checked in the User column as well. The analyst determined that the raised baseline did not interfere with accurate allele calling at this locus. Consequently, they opted to "unfire" the rules by clicking in the User column to indicate that samples like this can be given higher-quality scores in the future. In certain circumstances, the user may want an additional rule to fire. Again, they would communicate this by checking that rule box in the User column. Optimization involved a feedback loop between NYS Convicted Offender DNA Databank and Cybergenetics, Inc. When review was complete, Cybergenetics, Inc. correlated the User rule firings to relevant peak information such as height, shape, area, and position. They adjusted the rule thresholds to better fit the user's preference and then provided the laboratory with the improved template. The original dataset and a new dataset were processed with the new template using the same review guidelines. This loop continued until a satisfactory balance was reached where high-quality data were automatically accepted while no substandard sample escaped user scrutiny.

### *Concordance Phase*

A dataset of 2048 convicted offender profiles comprised the concordance study. A sample size this large was chosen to guarantee adequate testing of not only the allele calling ability but also the sample management support files generated by TrueAllele® (information on accepted, edited, and rejected samples) in comparison with the present protocol using Gene Scan® and Genotyper®. We considered two types of concordance. First, allele calls were compared to determine the consistency between systems. Second, we compared the reasons for sample rejection based on the information on sample quality provided by each system.

The concordance study involved a combination of samples that had already completed primary analysis and technical review with ABI software ("OldBatches") and samples that were processed in both systems concurrently ("NewBatches"). There were five old batches, each containing all ABI 3700® runs for two separate extraction plates, for a total of ten plates. There were also five new batches, each containing all ABI 3700® runs for three extraction plates, for a total of 15 plates. TrueAllele® processing was performed on every batch. In Allele View, data with a Quality Score <0.300 were reviewed and accepted, edited, or rejected. All data above this score were automatically accepted.

Once all data were processed through both TrueAllele® and Genotyper®, the profiles were compared with the help of a subprogram within TrueAllele®. The AutoValidate program was provided by the Cybergenetics, Inc. group to assist the NYS Convicted Offender DNA Databank in the automated comparison of profiles generated by both systems. AutoValidate is a multistep program that first checks for profile completeness and consistency

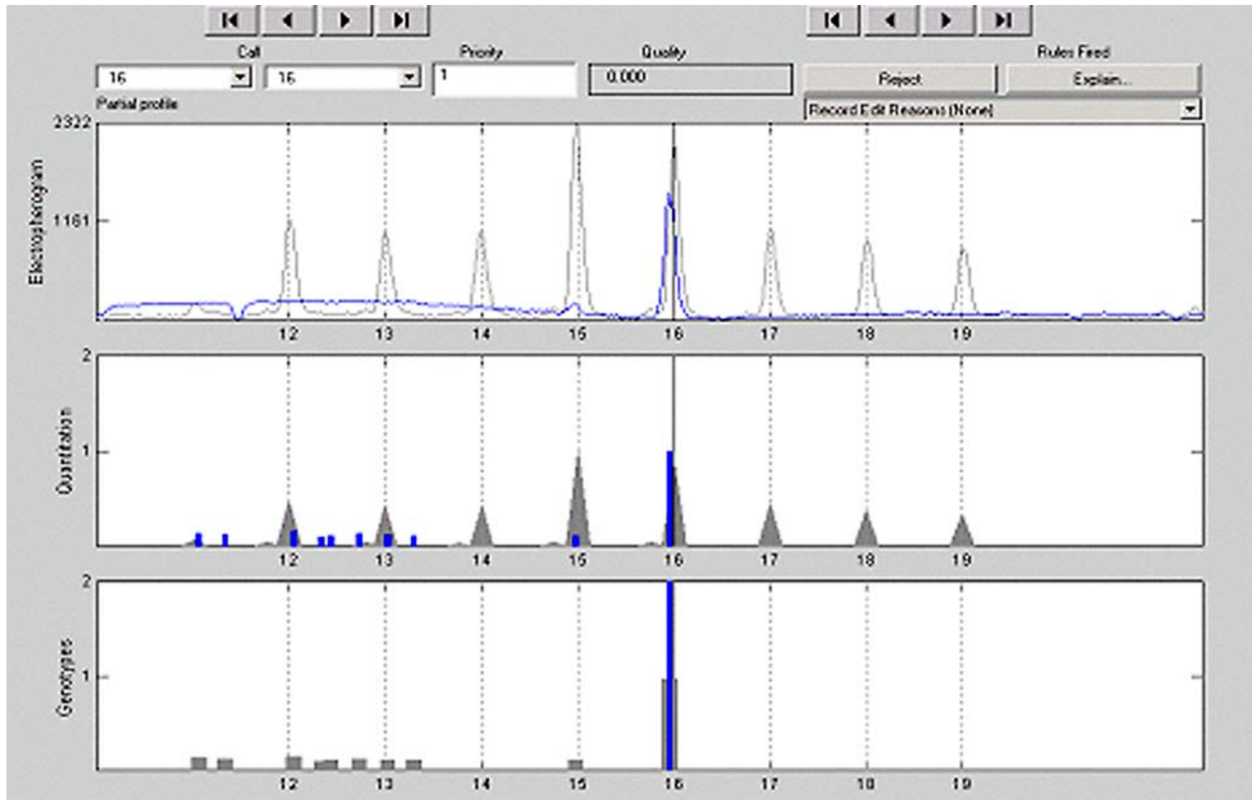


FIG. 1*a—Example of the Allele View window. TrueAllele® detected several peaks in the Electropherogram window (*top*). The ladder peaks can be seen in the background. All quantitated peaks are visible in the Quantitation window (*middle*). TrueAllele® determined the call to be 16, 16 as shown in the Genotypes window (*bottom*) and the Call text boxes at the upper left.*

within each data type. For example, TrueAllele® results files from Cofiler and Profiler Plus datadisks are placed in the TRuns folder (T = TrueAllele®). AutoValidate checks for results at all 17 loci and compares the three loci shared between the Cofiler and Profiler Plus kits. An "Inconsistent" file is generated if there are any disagreements in the shared loci or at other loci between multiple runs of a sample. All complete and accepted TrueAllele® profiles are then compared against the Genotyper® profiles in the URuns folder (U = User) that have passed an identical review process. If any dissimilarity is found in the profiles provided by the two data types, TrueAllele® writes a "Differences" file that records the inconsistent alleles. If TrueAllele® obtained a full profile for a given sample while Genotyper® produced only a partial profile or vice versa, AutoValidate generates "TUnmatch" and "UUnmatch" files, respectively. Any sample that yielded an incomplete profile in either system (listed in these two files) would not be compared automatically. Therefore, these profiles were compared manually. In addition, the reasons for rejecting samples in either one or both systems were investigated by the users. Concordance of allele calling and sample rejection between the two data analysis systems can be adequately measured from these AutoValidate reports.

## Results

A total of three rounds of optimization was required to optimize performance of TrueAllele® for the NYS Convicted Offender DNA Databank. Throughout the optimization period, nearly 42,000 allele calls were reviewed. The first round consisted of ten ABI 3700® runs. Approximately 7100 allele calls were reviewed during this round. Thresholds for six of the rules were then adjusted in Template 2. See Table 1 for a description of the rules. The *Dispersion* percentage was decreased from 80% to 60%. *Dispersion* refers to the amount of signal found within the allelic peaks, thus more signal would be tolerated outside of the designated peaks by lowering the threshold. *Extra Allele* and *Unexpected* fire when a defined amount of signal occurs in nonallelic peaks. *Extra Allele* is restricted to peaks within the ladder region while *Unexpected* applies to areas outside of the ladder region. Both indicate the possibility of contamination or bleed-through. As with the *Dispersion* adjustment, increasing the threshold percentage from 15% to 20% for both rules allowed more signal to appear outside of the designated peaks. The *High Signal* rule was raised from 10,000 RFU to 15,000 RFU because signal from the ABI 3700® used in this study tends to run high without error. The *Low Homozygote* threshold was determined to be more effective when lowered from 500 to 250 RFU. The *Low Signal* threshold was lowered from 300 to 150 RFU to meet our current guidelines for minimum peak heights for ABI 3700® data. Samples with peaks under this height must be rejected and rerun at a higher loading amount.

For the second round, the original ten ABI 3700® runs were reprocessed with Template 2 along with another ten runs. Approximately 15,000 allele calls were reviewed during Optimization Round 2. Only two changes were made to the rule parameters for Template 3. The *High Signal* rule was raised again to 20,000 RFU. The *Off-Ladder* rule was lowered to 0.45 bp, requiring that allelic peaks be closer to the ladder in order to be properly designated.

The third round consisted of the 20 previously reviewed runs plus 11 more ABI 3700® runs for an estimated total of 20,000 allele calls. Most of the edits and rejects made by the reviewer reflected sample quality rather than suboptimal rule settings. As a result,

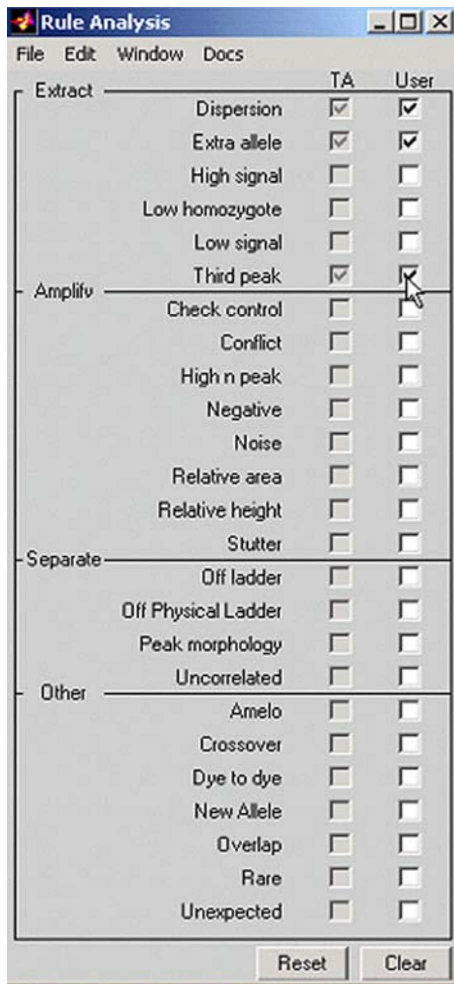


FIG. 1*b*—*The Rule Analysis window has two columns to show which rules TrueAllele® fired. Initially, both the TA and User columns match. During optimization, the user can modify rule firings to reflect how they want TrueAllele® to treat similar circumstances in the future by checking and unchecking the appropriate boxes in the User column.*

TABLE 2—*Number of rules fired for each STR panel using each template.*

| | STR Panel | Template 1 | Template 2 | Template 3 |
|---|---|---|---|---|
| Dataset 1 | | | | |
| (∼7100 allele calls) | Cofiler | 968 | 510 | 406 |
| | Profiler Plus | 1359 | 756 | 631 |
| Dataset 2 | | | | |
| (∼7900 allele calls) | Cofiler | . . . | 967 | 764 |
| | Profiler Plus | . . . | 1142 | 906 |
| Dataset 3 | | | | |
| (∼4700 allele calls) | Cofiler | . . . | . . . | 644 |
| | Profiler Plus | . . . | . . . | 806 |

Template 3 was adopted as the working set of rule parameters, and it was decided that no further optimization was necessary. The final optimized thresholds are listed in Table 1.

Rule statistics were generated for each datadisk, and comparisons were made between templates. Table 2 illustrates the improvements in rule firing seen with each new template. While the numbers seem high, the total number of allele calls and possible rule firings must be considered. Dataset 1 consists of 7100 calls. Dataset 2 and 3 contained 7900 and 4700 calls, respectively. In addition, each call could have fired up to 24 rules. The adjustments made in Template 2 reduced significantly the number of rule firings. Further improvement occurred with Template 3. By Optimization Round 3, the rule firings were deemed to be necessary and appropriate based on sample quality. Further modifications to the thresholds may have erroneously allowed lower-quality samples to be accepted without review.

TABLE 3—*Results of concordance study allele designations verified with TrueAllele® AutoValidate program.*

| Dataset | No. Samples | Matches | Nonmatches |
|---|---|---|---|
| OldBatch1 | 174 | 174 | 0 |
| OldBatch2 | 142 | 141 | 1 |
| OldBatch3 | 174 | 174 | 0 |
| OldBatch4 | 174 | 174 | 0 |
| OldBatch5 | 139 | 137 | 2 |
| Total | 803 | 800 | 3 |
| Percentages | | 99.63% | 0.37% |
| NewBatch1 | 257 | 257 | 0 |
| NewBatch2 | 261 | 260 | 1 |
| NewBatch3 | 205 | 205 | 0 |
| NewBatch4 | 261 | 261 | 0 |
| NewBatch5 | 261 | 261 | 0 |
| Total | 1245 | 1244 | 1 |
| Percentages | | 99.92% | 0.08% |
| Total % | 2048 | 99.80% | 0.02% |

Template 3 was used in the next phase of the validation. More than 2000 convicted offender samples were processed through both genotyping systems to test concordance between TrueAllele® and Genotyper®. Table 3 illustrates the comparison of allele designations between TrueAllele® and Genotyper®. Matching and nonmatching samples are enumerated for each batch of data. The OldBatches and NewBatches are counted separately, and the total percentages are calculated at the bottom of the table. All but four allele designations (99.8%) matched between the two systems. Table 4 contains information about the four samples that had allele calls that did not agree. In the first case, the difference occurred at an allele just outside the ladder at D5S818. TrueAllele® binned the allele as >16 whereas Genotyper® labeled it as 17. TrueAllele® always bins alleles that are outside of the ladder range, as it did in this case. The Genotyper® macro will not only put labels on alleles present in the ladders but also on "virtual" alleles that are just outside of the ladder range, again as it did in this case. This type of benign discrepancy occurred in less than 1% of the samples in this concordance study. In addition, CODIS permits only binned alleles to be uploaded. Therefore, automatic binning by TrueAllele® is favorable. In the second case, there was not an actual disagreement but rather an omission. All three designations for a triple allele call at TPOX were included in the TrueAllele® results file. The Genotyper® table did not contain a field for the third allele, however. The third disagreement involved a spike in one profile at D8S1179. On the Genotyper® plots, the spike resembled a DNA peak to the analyst (Fig. 2*a*), while the spike was more apparent in TrueAllele® (Figs. 2*b*, 2*c*). The shape was much sharper than a DNA peak, and the height was about half of what it appeared to be in Genotyper®. TrueAllele® did call the spike, but the user recognized it as an artifact and edited the call to reflect the real genotype. When the AutoValidate program discovered the discrepancy, the Gene Scan® and Genotyper® information on the sample were more closely inspected. Spikes exhibit certain features that distinguish them from true DNA peaks, specifically narrow peak width and their occurrence in all dyes in the electropherogram. Based on

TABLE 4—*Description of the four samples with allele differences.*

| Nonconcordant Sample | Batch | Locus | Type of Inconsistency: TrueAllele | Type of Inconsistency: Genotyper |
|---|---|---|---|---|
| 1 | OldBatch2 | D5S818 | Allele binned as >16 | Allele labeled as 17 |
| 2 | OldBatch5 | TPOX | All three alleles of triplet recorded | Third allele in triplet not recorded |
| 3 | OldBatch5 | D8S1179 | Spike readily recognized | Peak not recognized as spike |
| 4 | NewBatch2 | FGA | Marker bin not set wide enough | No error |

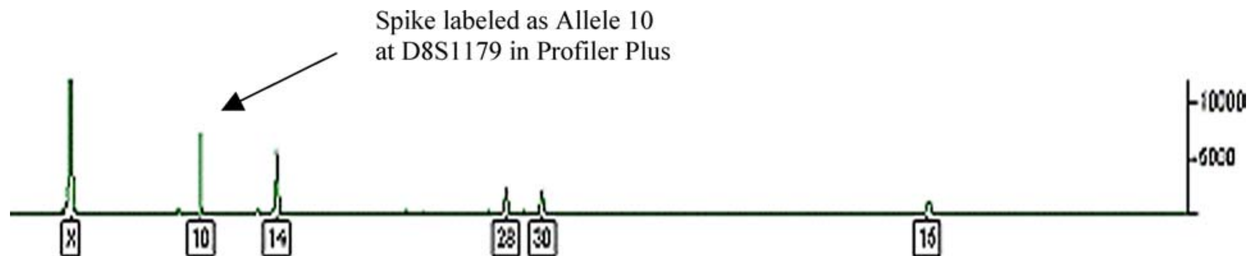


FIG. 2*a—Zoom View of Genotyper® window showing profile of green Profiler Plus loci (Amelogenin, D8S1179, D21S11, and D18S51) of problem sample. The D8S1179 locus indicates a heterozygous call of 10,14. The arrow points to the spike labeled as allele 10. This call successfully passed primary analysis and technical review. Note relative height of the two labeled peaks at D8S1179.*

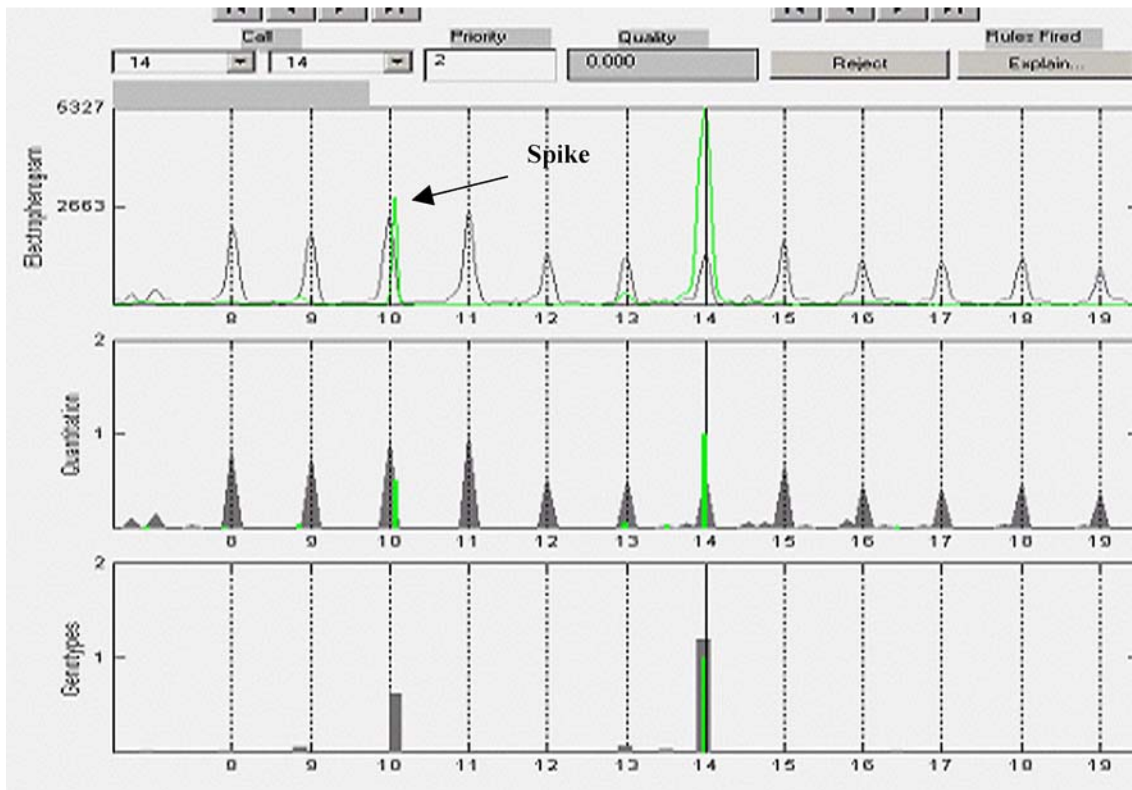


FIG. 2*b—Allele View Electropherogram window of TrueAllele® showing only D8S1179 of problem sample. The arrow points to the spike. TrueAllele® originally called this locus 10,14 as did Genotyper®. However, the user immediately recognized one peak as a spike based on shape and relative size.*

peak width, this spike should have been caught by the analyst in Genotyper®. The spike was simply more obvious to the user in the TrueAllele® display. The fourth case turned out to be a TrueAllele® user error in setting marker bins. As a result, an off-ladder allele was missed. The default marker bin settings in TrueAllele® have since been expanded to include all potential off-ladder alleles at every locus. In summary, allele call disagreements between TrueAllele® and Genotyper® stemmed from differences in their output files and human oversight.

Ultimately, samples are either accepted or rejected. Table 5 displays the results of the sample state assignments made by TrueAllele® and Genotyper®. Agreements between the two systems are classified into "Accept" and "Reject." Disagreements are divided according to which system accepted the sample. "Full in TA only" implies that the sample was rejected in Genotyper®. "Full in GT only" indicates that the sample failed in TrueAllele®. For the majority of cases, both systems agreed to accept or reject a sample (93.59%). Less than 6.5% yielded a complete profile with only

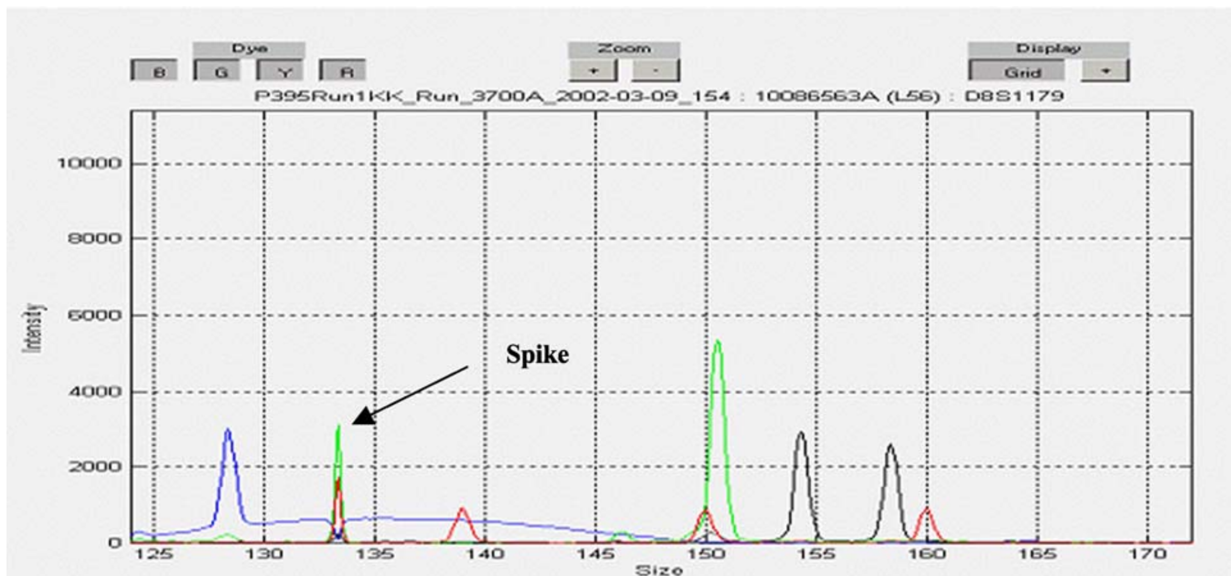


FIG. 2c—*Allele View Electropherogram window of TrueAllele® showing all dye colors in that region. The arrow points to the spike. Unlike in Genotyper® (not shown), the spike occurs in all colors.*

TABLE 5—*Results of concordance study sample state assignments.*

| Dataset | No. Samples Assigned | Agreements | | Disagreements | |
|---|---|---|---|---|---|
| | | Accept | Reject | Full in TA Only | Full in GT Only |
| OldBatch1 | 174 | 162 | 5 | 0 | 7 |
| OldBatch2 | 141 | 134 | 2 | 5 | 0 |
| OldBatch3 | 174 | 151 | 11 | 6 | 6 |
| OldBatch4 | 174 | 165 | 6 | 3 | 0 |
| OldBatch5 | 137 | 124 | 3 | 3 | 7 |
| Total | 800 | 736 | 27 | 17 | 20 |
| Percentages | | 95.37% | | 4.63% | |
| NewBatch1 | 257 | 241 | 7 | 1 | 8 |
| NewBatch2 | 260 | 226 | 7 | 8 | 19 |
| NewBatch3 | 205 | 169 | 20 | 9 | 7 |
| NewBatch4 | 261 | 244 | 3 | 6 | 8 |
| NewBatch5 | 261 | 227 | 6 | 17 | 11 |
| Total | 1244 | 1107 | 43 | 41 | 53 |
| Percentages | | 92.44% | | 7.56% | |
| Total % | 2044* | 93.59% | | 6.41% | |

* The four profile mismatches were removed from the total number of samples compared in this table. Hence the total number of samples is 2044.

TrueAllele® or Genotyper®. High signal was the primary reason that TrueAllele® accepted samples when Genotyper® did not. Too much pull-up or inverted peaks in Genotyper® necessitate sample rejection according to our current protocol. We avoided this problem in TrueAllele® because of editing permission. There were a number of explanations for samples that were rejected in TrueAllele® but not in Genotyper®, some of which stemmed from the current NYS Convicted Offender DNA Databank protocol. Differences included the inequality of the Relative Fluorescence Units (RFU) scale between TrueAllele® and Genotyper®, activity at the baseline visible in TrueAllele®, methods of size standard tracking, and consistent identification of every microvariant allele. Overall, the observed differences in sample rejection stemmed from a combination of protocol restraints and sample information display.

The majority of these rejected samples had allele calls below the 150 RFU threshold in TrueAllele®. The current NYS Convicted Offender DNA Databank protocol requires that all peaks generated by the ABI 3700® must be greater than 150 RFU to be called. Apparently, the Relative Fluorescence Units (RFU) scale in TrueAllele® does not equal the Genotyper® scale. This is likely due to the fact that TrueAllele® uses the signal data as they come off the ABI 3700® instrument whereas Gene Scan® applies peak smoothing during analysis. Smoothing is known to affect peak height. In general, it seems that peaks are slightly smaller in TrueAllele®. Therefore, more samples were rejected because they did not meet the current guideline for peak height threshold.

A smaller number of samples exhibited questionable baselines in TrueAllele® but not in Genotyper®. In general, the TrueAllele® interface shows a noisier baseline than Genotyper® plots. There are two explanations for this difference. First, unlike Gene Scan®, TrueAllele® does not apply peak smoothing measures to the data but rather it uses the raw data as mentioned above. Second, the TrueAllele® display focuses on a smaller range, a single locus rather than the whole profile. Despite these explanations, the users decided

to remain conservative and reject samples with questionable baselines during this study.

A few samples were rejected because of failed size standards. TrueAllele® uses a different method of size standard tracking than does Gene Scan®. Gene Scan® detects peaks for the selected dye color and matches them to a previously defined size standard. In Cap View, TrueAllele® tracks the peaks in each capillary and assigns size labels according to the expected distance between peaks. The user then scans each lane to make sure all size standard peaks were properly defined. On a few occasions, some peaks failed to be identified. Therefore, some size standards that passed review in Genotyper® did not pass in TrueAllele®.

Some samples were rejected in TrueAllele® because of microvariant alleles that needed to be confirmed in a second run. Our protocol states that only alleles labeled as "OL allele" must be confirmed with a second run. Some alleles were not confirmed in Genotyper® because the software labeled them using "virtual" ladder alleles. As a result, no confirmatory run was available for TrueAllele® processing. Because TrueAllele® has no "OL allele" label, we rely on the rule firings to indicate all microvariants that need to be confirmed.

One sample was rejected in TrueAllele® because of a spike interfering with an allele call. TrueAllele® tends to raise the baseline around spikes and areas of high pull-up peaks specifically in ABI 3700® data, which undergo spectral correction during the run and are therefore not entirely "raw." The raised baseline made the allele call questionable. The spike was not labeled in Genotyper®, however, and did not interfere with any other peaks. Therefore, the sample successfully passed review in Genotyper®.

In one case, TrueAllele® failed to recognize a DNA peak. Although it was greater than the detectable threshold, TrueAllele® did not quantitate this peak, possibly because its area relative to the other DNA peak in the window was too low. Because TrueAllele® did not detect the peak, the user opted to reject rather than edit the call.

## Discussion

Using Template 3 in the optimization phase, the user did not have to alter the rule firings as determined by TrueAllele®. In addition, no samples considered "low quality" under the current NYS Convicted Offender DNA Databank protocol would have escaped user review. The edits and rejections accurately reflected sample quality. Therefore, optimization ended after three rounds. At that point, 10%–15% of the allele calls had quality scores <0.500, which translates into an acceptable amount of data to review. This does not imply that only 10%–15% of samples are reviewed, however. In reality, the user will examine all of the samples at some step in the TrueAllele® process, such as Cap View and Control Check.

Concordance between the widely accepted software Gene Scan® and Genotyper® and the new automated data review program TrueAllele® was measured by comparing the profiles of more than 2000 convicted offender samples generated independently by both systems. The AutoValidate program created by Cybergenetics, Inc. facilitated this comparison. Results from each system were not identical, but the differences were trivial. TrueAllele® and Genotyper® called the same allele designations for 99.8% of samples. Four out of 2048 samples revealed differences at a single locus each. Two of these stemmed from differences in the software output format, specifically binning all alleles outside of ladder range and the lack of a third allele field in Genotyper® tables. The other two differences actually resulted from human oversight in spike recognition and setting ladder bins.

TrueAllele® and Genotyper® exhibited conflict in sample state assignment ("accept" or "reject") for 6.41% of 2048 samples. Editing ability accounts for the majority of samples accepted only in TrueAllele®. Other cases involved modifications to size standard assignments. In this study, a greater number of samples were accepted only in Genotyper®. Four issues arose that may have caused this increase in the number of rejected samples in TrueAllele®: confirmation of every off-ladder allele rather than only those labeled "OL allele," incompletely defined size standards, higher number of peaks under 150 RFU, and visibility of activity at the baseline. Each is related directly to the requirements currently stated in the NYS Convicted Offender DNA Databank protocol.

## Conclusions

TrueAllele® proved to be a reliable method of allele calling for convicted offender samples processed by the NYS Convicted Offender DNA Databank. Through the process of in-house Optimization and Concordance, the TrueAllele® system was customized to the data quality requirements of the NYS Convicted Offender DNA Databank. The optimization phase yielded guidelines with which TrueAllele® could perform a quality data assessment as well as a human reviewer. The guidelines were tested in the concordance phase with convincing results. Therefore, TrueAllele® performed, as designed, as an automated allele caller with stringent quality checks that focus human review on questionable data. Furthermore, TrueAllele® removes much of the subjectivity of human review by scrutinizing every locus in a consistent manner. In summary, no sample should pass TrueAllele® review that would fail human review given that a thorough internal validation has been conducted.

TrueAllele® is designed to operate independently of other allele calling systems. We propose to use TrueAllele® as the primary system for review of STR DNA data in our convicted offender laboratory. We plan an incremental introduction of the software into production mode. Initially, a single analyst will perform a first review with the expert data review system. A second analyst will subsequently perform a technical review of the edited calls made by the primary analyst. As experience with the software increases, the model can be shifted to greater dependence on the automated process. Our goal ultimately is to replace the current two-person review process with a single reviewer using TrueAllele®. This paradigm should be expected to reduce data quality assessment time from twofold, under the initial implementation process, to greater than fivefold using the fully implemented system.

Finally, it should be noted that supplemental validation studies are warranted for other electrophoresis platforms, STR kits, and sample types. We make no implications for the use of this system on forensic casework samples.

*Acknowledgments*

We would like to extend our appreciation to Peter Wistort, a supervisor in the NYSP DNA Databank, for his thoughts and ideas on the most effective use of TrueAllele® in our laboratory.

## References

1. Palsson B, Palsson F, Perlin M, Gubjartsson H, Stefansson K, Gulcher J. Using quality measures to facilitate allele calling in high-throughput genotyping. Genome Res 1999;9(10):1002–12. [PubMed]
2. Perlin MW, Coffman D, Crouse CA, Konotrop F, Ban JD. Automated STR data analysis: validation studies. Proceedings of the Twelfth International Symposium on Human Identification; 9–12 Oct. 2001;

Biloxi, MS: Promega, 2002. (http://www.promega.com/geneticidproc/ussymp12proc/contents/perlin.pdf)
3. Perlin MW, Lancia G, Ng S-K. Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. Ame Hum Genet 1995;57(5):1199–10.
4. Ng S-K, Mukhopadhyay N, Rayman J, Ghosh S, Perlin MW. Fully automated microsatellite genotyping: computer software (Abstract). Am J Hum Genet 1996;59(4 Suppl):A229.
5. Richards DR, Perlin MW. Quantitative analysis of gel electrophoresis data for automated genotyping applications (Abstract). Amer J Hum Genet 1995;57(4 Suppl):A26.
6. Stoughton R, Bumgarner R, Frederick III WJ, Mc Indoe RA. Data-adaptive algorithms for calling alleles in repeat polymorphisms. Electrophoresis 1997;18:1–5. [PubMed]

Additional information and reprint requests:
Kristy Kadash, Ph.D., D-ABC
Forensic Scientist
New York State Police Forensic Investigation Center
1220 Washington Ave.
Albany, NY 12226
E-mail: kkadash@troopers.state.ny.us