B108 Validation Study of the TrueAllele™ Automated Data Review System

Kristy Kadash, PhD, Brian Kozlowski, MFS, Lisa Biega, MS, and Barry Duceman, PhD, New York State Police DNA Databank, New York State Police Forensic Investigation Center, 1220 Washington Avenue, Building 30, Albany, NY*

The goals of this research project are to compare an automated data review method (TrueAllele™, Cybergenetics, Inc.) to the established STR analysis software (GeneScan® and Genotyper®, ABI) for use in DNA databank laboratories.

The stated goal of the TrueAllele™ system for databank laboratories is to alleviate the shortage of skilled data reviewers by automating most of the steps in the review process. Theoretically, this would decrease the amount of time needed to analyze scores of profiles, thereby increasing efficiency. To be useful for databank laboratories, the system must handle high throughput with minimal error. The New York State (NYS) Convicted Offender DNA Databank is the first U.S. laboratory to conduct an internal validation of this software for the purpose of generating profiles from ABI 3700® capillary data for upload into the state convicted offender database.

TrueAllele™ detects, quantitates, and types allelic peaks, thereby combining the tasks of the GeneScan® and Genotyper® software (ABI) currently used by the NYS Convicted Offender DNA Databank. TrueAllele™ also prioritizes the allele calls based on several userdefined rules. As a result, the user should only review low quality data. Profiles that meet all rule requirements are automatically accepted without being examined. A sub-program within TrueAllele™, AutoValidate, checks for extraction, amplification, or pipetting errors by comparing shared loci between STR panels and multiple runs of a given sample.

The validation 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. To accomplish this, the users reviewed nearly 42,000 allele calls and made desired changes in rule firings to dictate future software behavior. Cybergenetics programmers adjusted thresholds based on the frequencies of these User rule firings. The software went through three rounds of threshold optimization, at which point it was decided that TrueAllele satisfactorily assigned low quality scores to all questionable calls and higher scores to acceptable calls. The end result was a set of parameters that the NYS Convicted Offender DNA Databank could confidently use to generate accurate and reproducible DNA profiles.

Because the TrueAllele™ expert review system operates very differently from GeneScan® and Genotyper®, an extensive concordance study was performed to ensure that the output of each analytical technique was equivalent. More than 2,000 samples were typed with ABI software and TrueAllele™. Allele designations were identical for 99.8% of samples. Concordance in sample state assignment (accept or reject) was greater than 93%. The remaining differences largely stemmed from disagreements regarding sample rejection.

TrueAllele™ differs from the widely used ABI software in significant ways. The biggest adjustment is that profiles are displayed in a locus-based format rather than in toto. The context of the whole profile has been removed from the primary user interface. Second, sample prioritization means that not all data are reviewed. Consequently, many profiles could be loaded in the state database without human evaluation. Third, editing of erroneous calls is permitted under certain conditions such as cases of dye bleed-through and the occurrence of spikes (specific to capillary electrophoresis). While possibly introducing a certain amount of subjectivity, editing serves as an important time-saver compared to the current NYS Convicted Offender DNA Databank protocol. Finally, TrueAllele™ uses a different method of size standard calculation and display than Genotyper®. Again following guidelines, users can modify the precise location of internal size standard peaks in cases of dye bleed-through and overlap that distort the peak shape.

Editing of allele designations and size standard placement under controlled circumstances will eliminate the need to re-run a significant amount of samples. Other issues will likely require different samples to be rejected, however. First, baseline activity tends to be greater in TrueAllele™ than in Genotyper®. In some cases, the presence of "unexplained peaks" might raise the suspicion of contamination unnecessarily. Second, peak height values differ between software systems. Peaks that are below the set minimum threshold in TrueAllele™ may be above the same threshold value in Genotyper®. This will necessitate the rejection of more samples due to low signal. Finally, confirmation of off-ladder alleles in TrueAllele™ is more consistent than in Genotyper®. Every off-ladder allele will be brought to the reviewer's attention for confirmation, whereas Genotyper® might use either a numerical label or the "OL allele" label. As a result, alleles with numerical labels may not be confirmed with a second run.

Overall, TrueAllele™ and Genotyper® generated comparable results. However, the reasons for rejecting

 ** Presenting Author*

samples changed slightly due to the issues mentioned above. This led to the discord regarding sample rejection. TrueAllele™ reliably brought to attention questionable calls while permitting high quality profiles to pass.

TrueAllele™ was designed to save time by focusing the review on poor data and by eliminating the need for complete re-analysis technical review. This thorough validation project, which included a large concordance study, proved TrueAllele™ to be dependable for use at the NYS Convicted Offender DNA Databank.

STR Analysis, DNA Databank, Data Review

 ** Presenting Author*