J Forensic Sci, May 2004, Vol. 49, No. 3
Paper ID JFS2003196
Available online at: www.astm.org

John H. Ryan,[1] Ph.D.; Jeffrey K. Barrus,[1] B.S.; Bruce Budowle,[2] Ph.D.; Cynthia M. Shannon,[3] B.S.; Victor W. Thompson,[1] B.S.; and Brian E. Ward,[1] Ph.D.

The Application of an Automated Allele Concordance Analysis System (CompareCalls^SM) to Ensure the Accuracy of Single-Source STR DNA Profiles

ABSTRACT: A powerful method for validating a scientific result is to confirm specific results utilizing independent methodologies and processing pathways. Thus, we have designed, developed and validated an automated allele concordance analysis system (CompareCalls^SM, patent pending) that performs comparisons between two independent DNA analysis platforms to ensure the highest accuracy for allele calls. Application of this system in a quality assurance role has shown the potential to eliminate greater than 90% of the STR analysis required of a DNA data analyst. While this system is broadly applicable for use with any two independent STR analysis programs, either prior to or following human data review, we are presenting its application to data generated with the ABI Prism™ Genotyper™ software system versus data generated with the SurelockID^SM system. With the automated allele concordance analysis system, the GeneScan™ DNA fragment data generated from an ABI 377™ gel image are analyzed in two independent pathways. In one analysis pathway, the GeneScan™ data are imported into Genotyper™ software where STR labels are assigned to the fragment data based upon the criteria of the Kazam 20% macro. The "Kazam" macro provided with the Genotyper™ program works by labeling all peaks in a category (or locus) and then filtering (or removing) the labels from peaks, such as those in stutter positions, that meet predefined criteria. In the second pathway, the GeneScan™ data are imported into the SurelockID^SM analysis platform where STR labels and error messages are assigned to the fragment data based upon hard-coded allele calling criteria and quality parameters. The resulting STR allele calls for each analysis platform are then compared, utilizing the automated allele concordance analysis system. Any differences in the STR allele calls between the two systems are flagged in a discordance report for further review by a qualified DNA data analyst. The automated allele concordance analysis system guides the DNA data analyst to the discordant data generated by either analysis platform. Additionally, the analyst is also directed to data that are of less than pristine quality which may have an increased potential for errors in interpretation by either analysis platform or by a human DNA data analyst. Implementation of an automated allele concordance analysis system will yield high-quality data for CODIS and free the human DNA data analyst to perform other critical duties within the laboratory.

KEYWORDS: forensic science, STR profile, STR allele, CODIS, automated allele concordance analysis system, gel image, quality assurance, Kazam 20% macro

In a highly mobile society, where crime, particularly violent crime, is a concern, DNA technology is being utilized for the public good. The use of various DNA databases has been pivotal in solving crimes where there is either no known suspect or in linking multiple cases committed by the same individual (1,2). The implementation of CODIS (Combined DNA Index System) has enabled law enforcement to solve many cases (i.e., those without suspects) that previously could not be addressed (3). CODIS is a database system that enables United States federal, state and local law enforcement crime laboratories to exchange and compare DNA profiles to assist the police in solving crimes by quickly identifying the perpetrator from different crimes or to identify potential suspects that may have committed the crime(s). There are two indexes in CODIS that are primarily accessed to attempt to solve crimes; these are the Convicted Felon and the Forensic Indexes. The Convicted Felons Index contains DNA profiles of persons convicted of crimes which, under state statutes, may be included in CODIS. The Forensic Index contains DNA profiles attributed to unknown individuals derived from lawfully collected specimens obtained during the course of a criminal investigation (i.e., from cases without a suspect). In order for the use of CODIS to be effective, both of these indexes must be populated with data, and resources must be made available to process, type and enter reliable data (3).

The DNA Backlog Elimination Act of 2000 (4) was designed in part to help address the backlog of convicted felon DNA samples awaiting analysis in state and local laboratories. To rapidly process the DNA backlog, the National Institute of Justice (NIJ) awarded funds authorizing state laboratories to outsource sample processing (5). To date, over one million STR DNA profiles from convicted felons have been generated for inclusion in CODIS (3). This dramatic success has led to the expansion of collection mandates in multiple states (6), resulting in an exponential growth of available DNA samples for processing. In addition, the potential expansion of state and federal laws to collect and process DNA samples from all felons will create a potential workload of over 1.5 million DNA samples annually (7).

To accommodate such substantial numbers of convicted felon samples, high throughput analysis systems have been developed and implemented (8–12). If not properly reviewed, any STR analysis system has the potential for errors or poor quality data. To ensure the accuracy and quality of the STR profiles generated by

[1] Myriad Genetic Laboratories, 320 Wakara Way, Salt Lake City, Utah 84108.
[2] Federal Bureau of Investigation, 935 Pennsylvania Avenue, Washington, DC 20535.
[3] Ohio Bureau of Criminal Identification and Investigation, 1560 Route 56, PO Box 365, London, Ohio 43140.
Received 4 June 2003; and in revised form 28 Sept. 2003, 4 Dec. 2003, 19 Jan. 2004; accepted 19 Jan. 2004; published 7 April 2004.

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

outsourcing, the DNA Advisory Board (DAB) Standards for Forensic DNA and Convicted Offender DNA Databasing Laboratories (13) mandated that multiple quality checks be implemented. The DAB Standards for Forensic DNA and Convicted Offender DNA Databasing Laboratories require laboratories to undergo accreditation, yearly audits, on-site visits and process blind resubmissions of DNA samples. In addition, regarding STR DNA analysis, standard 17.1.1 states:

> The laboratory will establish and use appropriate review procedures to verify the integrity of the data received from the subcontractor including, but not limited to, the following: Visual inspection and evaluation of results/data.

This is currently interpreted to require the public laboratories perform a "100% technical review" of the vendor-generated STR data. Because the number of outsourced STR profiles generated by laboratories has increased, the burden of implementing standard 17.1.1 by the state laboratories has concurrently increased. This paper describes an automated allele concordance analysis system (CompareCalls^{SM}, patent pending), that meets the requirements of standard 17.1.1, while reducing the human analytical burden by greater than 98%.

The task of STR data analysis for pristine database samples is relatively straightforward when compared to the issues inherent to the data analysis of casework samples such as mixture interpretation (14) and potential stochastic effects due to damaged or low copy number DNA (15). The automated analysis of STR profiles generated from pristine database samples has been presented to the forensic community as a potential solution to the burden of human data analysis (8–12). However, there is still a possibility that any single automated analysis software platform could assign an incorrect allele call and such issues of errors in software interpretation of STR data have been documented in the forensic analysis literature (16). Historically, similar concerns have been raised in the forensic community regarding the potential for errors in *human* interpretation of DNA data. These concerns have been addressed by utilizing one of the hallmarks of the Scientific Method: "independent replication of experimental results" (17). For the human interpretation of DNA data, this can be partially achieved by utilizing a second independent human review. In a similar manner, any current concerns of the forensic community regarding the potential for errors in data interpretation of STR allele calls generated by an automated analysis platform can be addressed by designing redundancy and independence into the automated software interpretation of STR profiles. This concept of validating a scientific result through independent methodologies and processing pathways has been demonstrated to be an excellent tool for identifying incorrect results or interpretations (17) and has been recently recommended as an approach to reduce the rate of false positives in testing for infectious pathogens (18). The replication of experimental results is especially powerful if identical results are achieved by truly independent methodologies.

In this paper, we describe the design, development and validation of an automated allele concordance analysis system (19–21) that performs comparisons between two independent DNA analysis platforms to ensure the highest accuracy for allele calls (Fig. 1). This analytical tool compares STR data from one STR analysis system to that of another STR analysis system and identifies all concordant and discordant allele calls. In addition, the resulting discordant STR allele calls directs the human data analyst to rapidly identify any areas of potential concern in STR data analysis. The application of this system in a quality assurance role has shown the potential to eliminate approximately 90% of the human STR analysis required in technical review of vendor-generated data (Table 1). This system is broadly applicable for use with any two-allele calling systems, and the present paper presents its application to data generated with Genotyper^{TM} versus data generated with Myriad Genetic Laboratories SurelockID^{SM} analysis platform. This analytical tool was assessed in three phases. In Phase I, within a test database, artificial errors were introduced into STR data allele calls for one of the analysis platforms. These manufactured errors were designed to reflect common areas of data interpretation or quality issues (Figs. 2 and 3). The automated allele concordance analysis system was then applied to the data set containing the manufactured errors. In Phase II, STR allele calls were made by human data review in the SurelockID^{SM} analysis platform; the automated allele concordance analysis system was applied to STR data generated from 50 ABI 377-96 lane gels. The automated allele concordance analysis system was used to compare the unreviewed data generated by Genotyper^{TM} with human reviewed data derived from SurelockID^{SM} analysis platform. The resulting automated allele concordance analysis report was assessed by a human reviewer analyzing the same STR data in Genotyper^{TM} (Fig. 4). Finally in Phase III, the automated allele concordance analysis system was applied to a data set of 290,676 CODIS STR markers where initial allele calls were made by qualified DNA data analysts in one previously validated analysis platform and then subjected to both an external and an internal 100% technical review by qualified CODIS DNA data analysts who made allele calls in an independent analysis platform (Fig. 5).

TABLE 1—*Current performance of the CompareCalls^{SM} application on ABI 377 STR data generated from CODIS samples: The performance of the automated allele concordance analysis process was measured for four different sample sets. These sample sets represented a range of samples types and quality of STR data produced. The percentage of lanes highlighted for further review represents the number of lanes in the discordance report versus the total number of lanes analyzed. The percentage of STR markers highlighted for further review represents the number of STR markers in the discordance report versus the total number of STR markers analyzed.*

Sample Set	Lanes Highlighted for Further Review	STR Markers Highlighted for Further Review
A	11.6%	1.7%
B	12.6%	2.2%
C	7.4%	1.0%
D	6.1%	1.0%
All current sample sets	10.5%	1.6%

Materials and Methods

DNA Isolation

DNA was extracted from dried blood spots on FTA, dried blood on filter paper, dried epithelial cells on buccal swabs, and from peripheral blood mononuclear cells with a commercial kit (QIAamp blood kit; Qiagen, Chatsworth, CA) following the manufacturer's recommendations.

STR Genotyping

Multiplex genotyping was performed with AmpFℓSTR® Profiler Plus^{TM} and AmpFℓSTR® COfiler^{TM}(Applied Biosystems, Foster City, CA) following the manufacturer's recommendations with slight validated modifications.

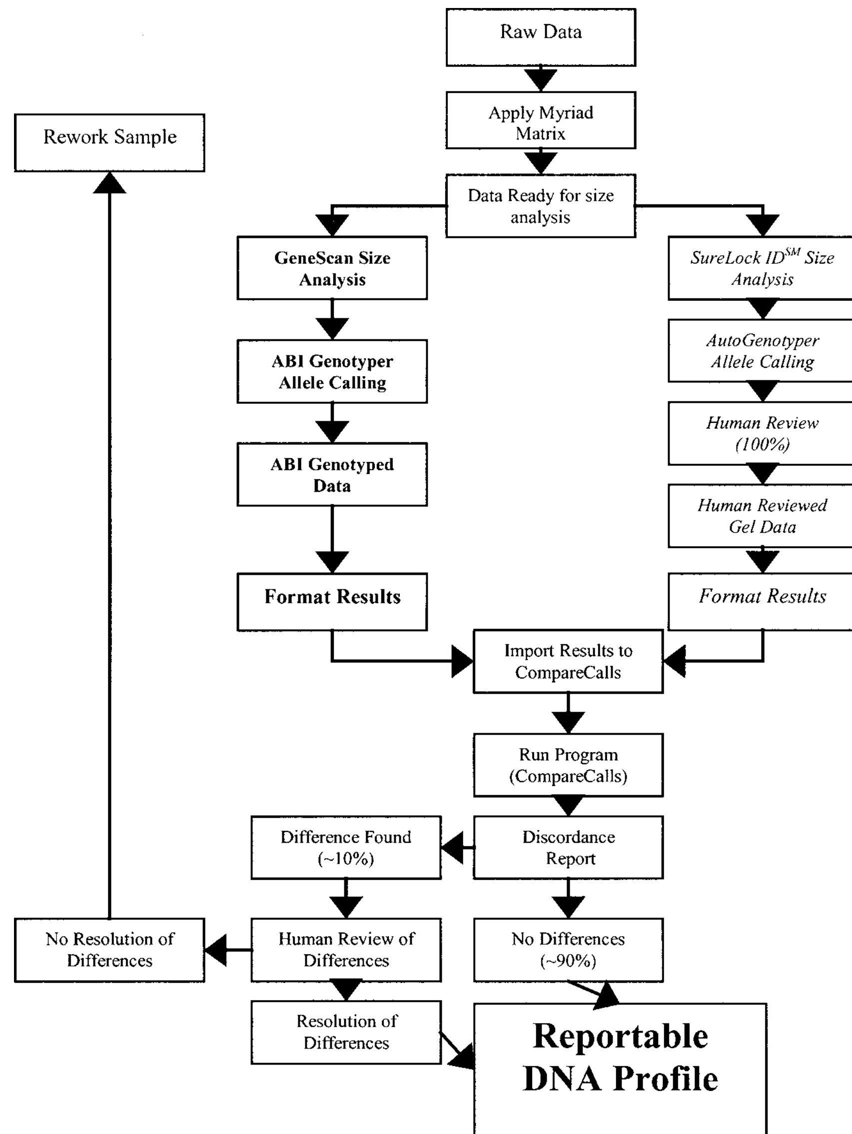

FIG. 1—*The automated allele concordance analysis process (CompareCalls^{SM}: ABI 377 DNA Sequencer Gel Images are tracked and extracted (raw data). A matrix is applied to correct for spectral overlap of fluorescent emission spectra. The resulting Genescan^{TM} trace files are then processed via two separate pathways. In Pathway 1 (bold text), the trace files are analyzed in Genescan^{TM} using a 50 RFU minimum peak amplitude for peak detection of all colors, light smoothing, and local southern method for fragment sizing. The analyzed Genescan^{TM} traces are then imported into a blank Genotyper^{TM} template file. Fragments are labeled using the "Kazam" 20% macro. The results table from the Genotyper^{TM} analysis are then exported as a text file for comparison. In Pathway 2 (italicized text), the trace files are imported into the SurelockID^{SM} system where an auto analysis is performed. Two CODIS DNA data analysts independently review the results of the auto analysis and the approved results are exported as a text file for comparison, in the same format as the Genotyper^{TM} analysis. The final STR calls for each sample are then compared using the two text files. The comparison is made on a gel lane basis and compares each allele as well as certain quality metrics (low signal, saturation, more than two alleles). Any differences between the two text files are highlighted in a Discordance Report. All differences are reviewed by a human DNA data analyst and are either resolved or the sample in question is submitted for additional processing.*

Electrophoresis

Electrophoresis was performed on an ABI 377 DNA Sequencer^{TM} (Applied Biosystems, Foster City, CA) using an acrylamide based gel for fragment separation in both 96-lane and 32-lane configurations, following the manufacturer's recommendations with slight validated modifications.

Analysis

Fragment sizing and analysis were performed with GeneScan^{TM} V3.1.2 and Genotyper^{TM} V2.5.0 (Applied Biosystems, Foster City, CA) in parallel with Myriad Genetic Laboratories SurelockID^{SM} analysis platform V9.0.

Additional Software

Inspection of the final profiles and supporting data was aided by the use of Microsoft Excel V8.0 for Office 98 on Macintosh and Microsoft Access V9.0 and Excel V9.0 for Office 2000 on PC.

Automated Allele Concordance Analysis System Pathway

ABI 377 DNA Sequencer Gel Images were tracked and extracted, following standard operating procedures. The resulting Genescan trace files were processed via two separate pathways (Fig. 1).

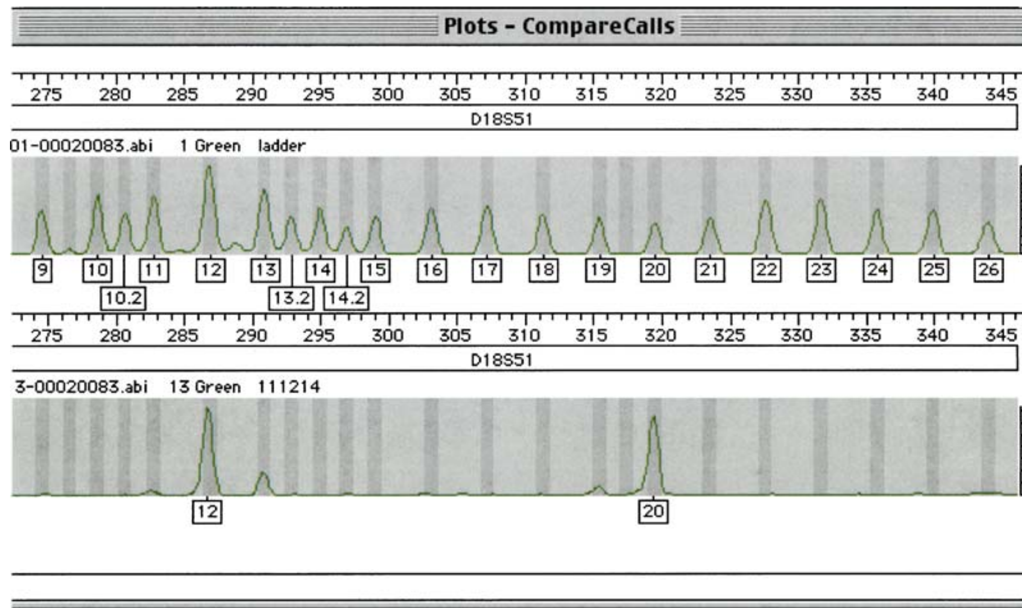

FIG. 2A—*Example of a manufactured artificial error, in which a minor 3rd allele peak has been deselected in the D18S51 marker of Gel 00020083 lane 13. The results of this analysis were then processed through the CompareCalls^{SM} pathway.*

Gel ID	Lane ID	Marker ID
00020083	13	D18S51

FIG. 2B—*The resulting CompareCalls^{SM} discordance report for the D18S51 marker of Gel 00020083 lane 13.*

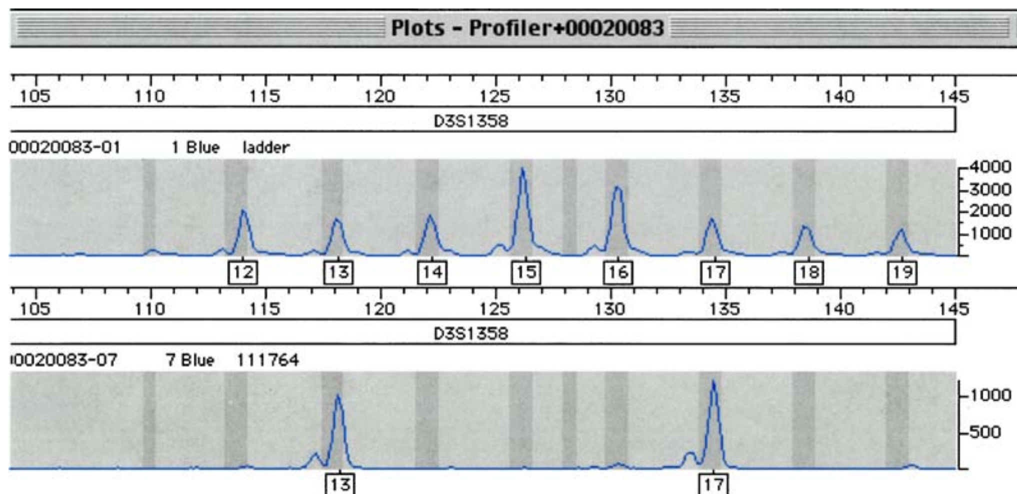

FIG. 3A—*Example of a manufactured artificial error, in which the −A peaks have been deselected in the D3S1358 marker of Gel 00020083 lane 7. The results of this analysis were then processed through the CompareCalls^{SM} pathway.*

Gel ID	Lane ID	Marker ID
00020083	07	D3S1358

FIG. 3B—*The resulting CompareCalls^{SM} discordance report for D3S1358 marker of Gel 00020083 lane 7.*

Pathway 1 (Fig. 1 bold text)—The trace files were analyzed in Genescan™ using a 50 RFU minimum peak amplitude for peak detection of all colors, light smoothing, and local southern method for fragment sizing. The analyzed Genescan™ traces were imported into a blank Genotyper template file. Fragments were labeled using the "Kazam" 20% macro. The "Kazam" macro provided with the Genotyper™ program works by labeling all peaks in a category (or locus) and then filtering (or removing) the labels from peaks, such as those in stutter positions, that meet predefined criteria (16). The results table from the Genotyper™ analysis was exported as a text file for comparison.

Pathway 2 (Fig. 1, italicized text)—The trace files were imported into the SurelockID^{SM} system where STR labels and error messages are assigned to the fragment data based upon hard-coded allele

TABLE 2—*The CompareCalls^{SM} 96-Lane Gel-based discordance report: The performance of the automated allele concordance analysis process is shown for eight 96-Lane 377 Gels. The percentage of STR markers highlighted for further review represents the number of STR markers in the discordance report versus the total number of STR markers analyzed in a 96-Lane Gel (651 STR markers in a Cofiler Gel and 930 STR markers in a Profiler Plus Gel).*

Gel ID	Further Investigation
00016331	09/651 (1.4%)
00016840	27/930 (2.9%)
00016842	16/651 (2.5%)
00016853	17/930 (1.8%)
00016857	11/651 (1.7%)
00016868	24/930 (2.6%)
00016870	20/930 (2.2%)
00016886	16/651 (2.5%)

calling criteria and quality parameters (9,10), such as a low signal threshold of 150 RFU and an allowable peak imbalance of 60%. Two CODIS DNA data analysts independently reviewed the results of the auto analysis and the approved results were exported to a text file in the same format as the Genotyper™ analysis. The sample results that were approved were based upon the analysis of the two CODIS DNA data analysts. Any sample results that were not approved were subject to additional processing to produce data quality sufficient to meet reporting criteria.

Comparison—After analysis, the final calls for each sample were compared using the two text files. Only the final calls, for samples approved by the two CODIS DNA data analysts, were subject to comparison. The comparison was made on a gel lane basis and compared each allele as well as certain quality metrics (low signal, saturation, more than two alleles.) The results of this comparison were displayed showing concordance and discordance on an ABI 377 gel basis (Table 2) and on a lane/marker basis (Table 3).

Validation

An extensive validation of the automated allele concordance analysis system, which included analysis of STR data via two independent analysis platforms, was performed. The validation was divided into three phases to test various aspects of the process.

Phase I consisted of manufacturing artificial errors in a limited data set to verify that the software detected all possible discrepancies arising from differences in the analysis platforms. Errors were introduced into the final saved calls in the SurelockID^{SM} system for the test set (Figs. 2A and 3A), and these calls were exported to a text file. The automated allele comparison process was then initiated. The trace data were reanalyzed in the GeneScan™ Genotyper ™ system and the results were exported to a text file. These files were introduced into the automated allele concordance analysis system (CompareCalls^{SM}) and the discrepancy report was examined and compared to the supporting data files (Figs. 2B and 3B).

Phase II consisted of manually analyzing 50 ABI 377-96 lane production gels in parallel with the automated allele concordance analysis process (Fig. 4). A DNA technician analyzed the 50 gels using Genotyper™ 2.5 searching for areas of concern. Any noted issues were recorded on a lane/marker basis and compared against the automated allele comparison system discrepancy report to ensure that CompareCalls^{SM} highlighted all areas of concern.

Phase III consisted of a process similar to Phase II, but was more extensive (Fig. 5). Human DNA data analysts from a public laboratory, i.e., external to Myriad Genetics, performed a 100%

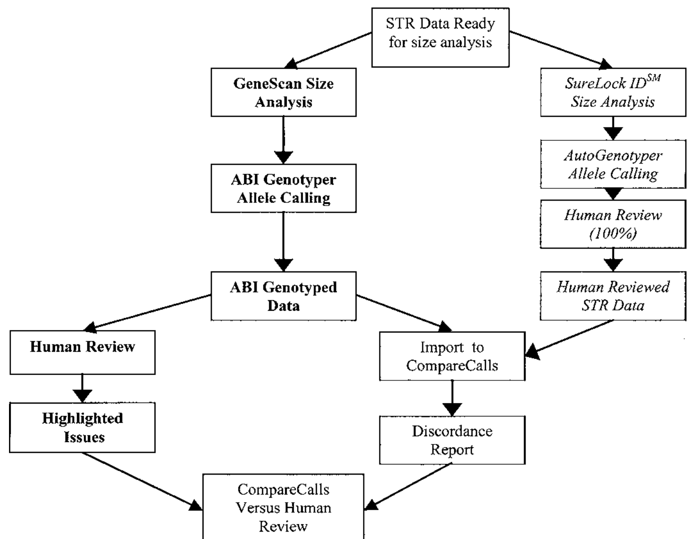

FIG. 4—*The CompareCalls^{SM} process versus human review: 50 ABI 377-96 lane production gels were manually analyzed (bold text) in parallel with the automated allele concordance analysis process outlined in Fig. 1. Any issues noted during manual analysis were recorded on a lane/marker basis and compared against the final discordance report to ensure that CompareCalls^{SM} had highlighted all areas of concern.*

TABLE 3—*The CompareCalls^{SM} STR lane/marker based discordance report: The STR markers highlighted for further review in Gel 00016840 are displayed showing the Gel Lane and the STR marker in question.*

Gel ID	Lane ID	Marker ID
00016840	17	D21S11
00016840	17	D5S818
00016840	19	D18S51
00016840	22	D8S1179
00016840	27	D3S1358
00016840	32	D3S1358
00016840	65	D3S1358

technical review of the data. In addition, a 100% technical review was performed by human DNA data analysts at Myriad Genetics. The reviewers both internally and externally analyzed ABI 377 gel data using Genotyper™ 2.5 and recorded areas of concern on lane/marker basis (bold text). These results were compared against the automated allele concordance analysis system discordance report to ensure that CompareCalls^{SM} highlighted all areas of concern.

Results and Discussion

An automated STR concordance analysis system (CompareCalls^{SM}) has been developed and evaluated to determine whether or not it is as accurate as the current state laboratory 100% technical review of STR data generated from outsourced DNA samples. The automated STR concordance analysis system is based on using two software analysis packages, where the concordant and discordant STR allele calls are identified. The concordant STR allele calls require no further human review, while the discordant STR allele calls require further human review. As long as the discordant STR allele calls are a small portion of the overall data, the automated STR concordance analysis system will alleviate much of the human STR profile interpretation currently utilized. The automated STR concordance analysis system described in this

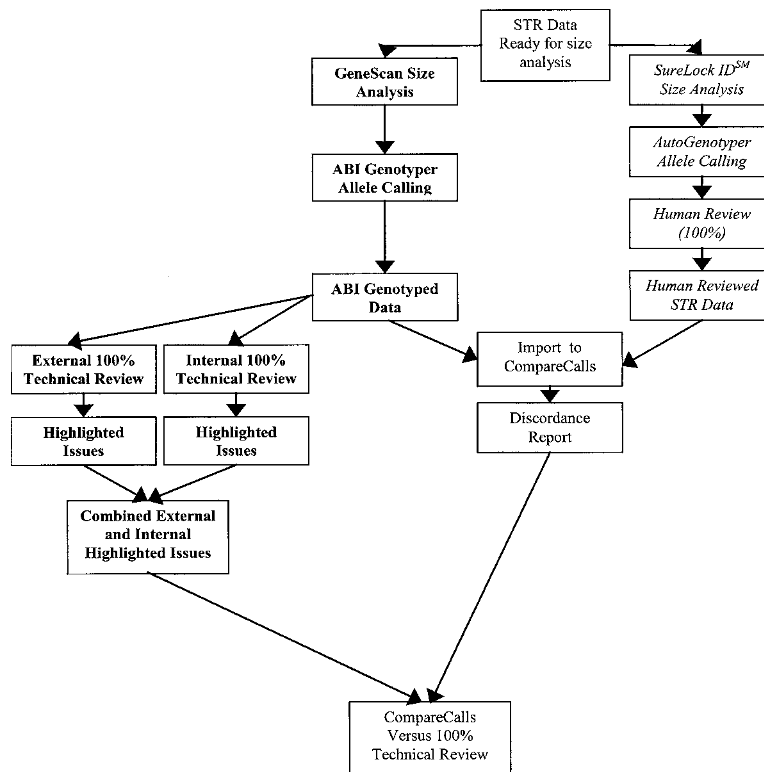

FIG. 5—*The CompareCalls^{SM} process versus 100% technical review: External and internal 100% technical review were performed (bold text) in parallel with the automated allele concordance analysis process outlined in Fig. 1. Human DNA data analysts from a public laboratory, i.e., external to Myriad Genetics, performed a 100% technical review of the data. In addition, a 100% technical review was performed on the same data by human DNA data analysts at Myriad Genetics. The reviewers both internally and externally analyzed ABI 377 gel data using Genotyper^{TM} 2.5 and recorded areas of concern on lane/marker basis (bold text). Any issues noted during the technical review were compared against the final discordance report to ensure that CompareCalls^{SM} had highlighted all areas of concern.*

study meets these requirements and can alleviate the majority of the human time required to review STR data generated from outsourced CODIS samples.

In Phase I of the validation, deliberate manufactured errors were introduced into the SurelockID^{SM} final allele calls for separate 24 STR markers (Figs. 2 and 3). The automated allele concordance analysis system was run on on the data set and every one of the manufactured errors was highlighted (data not shown). In addition, the automated allele comparison system highlighted areas for additional human review (data not shown).

In Phase II of the validation, a DNA technician analyzed 50 ABI 377-96 lane production gels using Genotyper^{TM} 2.5 searching for areas of concern (Fig. 4, bold text). Any noted issues were recorded on a lane/marker basis. The same 50 ABI 377-96 lane production gels were processed in the normal SurelockID^{SM} analysis pathway (Fig. 4, italicized text). The results of the normal SurelockID^{SM} analysis pathway were then exported to the automated allele concordance pathway. The results of the DNA technician analysis were then compared against the CompareCalls^{SM} discordance report to ensure that the automated allele concordance analysis system highlighted all areas of concern. The CompareCalls^{SM} system highlighted all issues identified by the DNA technician (data not shown). In addition, the automated allele comparison system highlighted areas for additional human review (data not shown).

In Phase III of the validation (Fig. 5, bold text), External and Internal CODIS certified DNA data analysts independently reviewed the results for 290,676 CODIS STR markers in Genotyper^{TM} 2.5. Any noted issues were recorded on a lane/marker basis. The same STR data were processed through the automated allele comparison

TABLE 4A—*Performance of CompareCalls^{SM} in External Validation: External and Internal CODIS certified DNA data analysts independently reviewed the results for 290,676 CODIS STR markers in Genotyper^{TM} 2.5. Any noted issues were recorded on a lane/marker basis. The same STR data were processed through the automated allele comparison system and all highlighted issues were recorded. Of the 34,071 total Gel Lanes analyzed, 4812 Gel Lanes were highlighted for further review (14.1%). Of the 290,676 total STR markers analyzed, 6512 STR markers were highlighted for further review (2.2%).*

ABI Gel Lanes Processed	ABI Gel Lanes Highlighted	STR Markers Processed	STR Markers Highlighted
34,071	4812 (14.1%)	290,676	6512 (2.2%)

TABLE 4B—*Performance of Human-Based Analysis: The combined results of the external and internal DNA data analyst noted issues were compared against the automated allele comparison discordance report produced for the same lanes of STR data. Of the 290,676 total STR markers analyzed, 37 potential issues noted by CODIS-certified DNA data analysts were not highlighted in the automated allele comparison discordance report. Each of these 37 potential issues was investigated by a CODIS-certified DNA data analyst. All 37 issues were found to be the result of human-based error and should not have been highlighted. It should be strongly noted that all of these issues would have been resolved during a normal 100% technical review.*

ABI Gel Lanes Analyzed	Total STR Markers Analyzed	Total Human Based Errors (% of Total)	Total Compare Call^{SM} Based Errors
34,071	290,676	37 (0.1%)	0 (0%)

system and all highlighted issues were recorded (Table 4A). The combined results of the external and internal DNA data analyst noted issues were compared against the automated allele comparison discordance report produced for the same lanes of STR data. Thirty-seven potential issues noted by CODIS certified DNA data analysts were not highlighted in the automated allele comparison discordance report (Table 4B). Each of these potential issues was investigated by a CODIS certified DNA data analyst. Upon further investigation all 37 potential issues were found to be due to one of several human-based errors. (It should be strongly noted that the human-based errors would have been detected in the course of finalizing the 100% technical review.) In 24 instances, a human reviewer had misidentified the STR marker in question. For example, a reviewer was looking at a vWA marker and documented a problem for the FGA marker. In four instances, a human reviewer miscalculated the relative peak heights of the marker alleles. For example, a reviewer calculated that the second D3 allele was 58% of the height of the first D3 allele. However, upon further investigation it was found that the second D3 allele was actually 68% of the height of the first D3 allele. In nine instances, the human reviewer indicated a potential issue such as a peak imbalance that was not found to be present upon further investigation by a CODIS certified DNA data analyst. All calls considered concordant by the automated STR analysis system were in agreement with the human technical review except for the 37 instances of human-based error (0.01% of the total STR allele calls analyzed) (Table 4B).

Since the completion of Phase III, the automated allele concordance analysis system has been applied in an internal quality assurance role in the Myriad Genetics production environment. Over 600,000 ABI 377 lanes of STR data have now been evaluated by human and automated STR concordance analysis. Currently in the Myriad Genetics production environment, the automated allele comparison system highlights approximately 10% of the lanes and 2% of the STR markers for further review (Table 1). It should be noted that the actual percentage of lanes or markers highlighted for further review by the automated allele concordance analysis system is dependent upon the quality of the STR data analyzed.

The validation studies presented in this paper demonstrate that an automated allele concordance analysis system (CompareCalls^{SM}) is at least as accurate as "100% human technical review" of STR DNA profiles. In addition, we have found that an automated allele concordance analysis system utilizing two independent DNA analysis platforms can be used to reduce the DNA data analyst time necessary to perform a "100% Technical Review" of STR profiles by approximately 90% on a lane-by-lane basis and by approximately 98% on a marker-by-marker basis (Table 1).

By utilizing the concordant STR results generated from two independent analysis pathways and platforms, the automated allele concordance analysis system presented in this paper has a significantly reduced probability of errors in STR data interpretation compared to any single STR software analysis platform. This same principle of "independent replication of experimental results" is a central tenet of the forensic sciences and has now been successfully applied to the human analysis of STR DNA profiles.

While the validation study focused on the ABI 377 platform, the principle of an automated allele concordance analysis system is broadly applicable to any commercially available STR processing platform and any two independent STR analysis programs. In addition, an automated allele concordance analysis system has the potential to be applied to DNA data in a myriad of fashions. Studies are currently underway to validate the CompareCalls^{SM} system for application to STR DNA data produced from capillary electrophoresis platforms, to STR DNA data prior to any human review, and to the STR DNA data produced from crime scene samples. Results of these studies will be presented in future reports.

In conclusion, an automated allele concordance analysis system has been developed, validated, and implemented for the STR analysis of single-source pristine samples. When concordant data are generated by the automated STR concordance system, no further human review is needed. As was demonstrated in this rigorous review, with good-quality DNA profiles, only a small percentage of the STR markers require technical review (approximately 2.0%). These findings have been supported by the review of over 600,000 lanes of STR data. Because automated allele concordance analysis systems based on independent allele calling algorithms are less prone to error than the current technical review process, the NDIS board of CODIS is justified in setting criteria for implementing expert systems. Implementation of such a validated expert system will yield high-quality data for CODIS and free the human DNA data analyst to perform other critical duties within the laboratory.

References

1. Budowle B, Moretti TR, Niezgoda SJ, Brown BL. CODIS and PCR based short tandem repeat loci: Law enforcement tools. Second European Symposium on Human Identification, 10–12 June 1998, Innsbruck, Austria.
2. Parson W, Steinlechner M, Scheithauer R, Schneider PM. National DNA intelligence databases in Europ-Report on the current situation. Proceedings of the 9th Symposium on Human Identification. 7–10 October 1998, Orlando, FL.
3. Federal Bureau of Investigation, Combined DNA Index System, National DNA Index System statistics. www.fbi.gov/hq/lab/codis/national.htm.
4. DNA Analysis Backlog Elimination Act of 2000. H.R. 4640. One Hundred and Sixth Congress of the United States of America, Second Session. Public Law 106-546, 19 December 2000.
5. National Institute of Justice Fiscal Year 2000 Budget. Crime Fighting Technologies Program. DNA backlog reduction program. 15 August 2000. www.ojp.usdoj.gov.
6. Smith AL. Forensic DNA databases. DNARESOURCE.COM.
7. United States Department of Justice, Office of Justice Programs Fiscal Year 2000 Annual Report to Congress. www.ojp.usdoj.gov.
8. Perlin MW, Coffman D, Crouse C, Konotop F, Ban J. Automated STR data analysis: Validation studies. Proceedings Twelfth International Symposium on Human Identification, 29 November 2001, Biloxi, MS.
9. Ward BE, Duceman B, Fraser A, Brown R, Wistort P, Hulick M. Cross validation of allele calling software for STR analysis in felon data basing applications "Allele Caller" versus Genotyper^{TM}. Proceedings Eleventh International Symposium on Human Identification, 10–13 October 2000, Biloxi, MS.
10. Gutin S, Scholl T, Thomas A. Advancement in allele calling accuracy and throughput. Proceedings Tenth International Symposium on Human Identification, November 1999.
11. Perlin MW. An expert system for scoring DNA database profiles. Proceedings Eleventh International Symposium on Human Identification, 10–13 October 2000, Biloxi, MS.
12. Perlin MW. Computer automation of STR scoring for forensic databases. First International Conference on Forensic Human Identification in the Millennium, 25–26 October 1999, Queen Elizabeth II Conference Centre, London.
13. DNA Advisory Board quality assurance standards for convicted offender DNA databasing laboratories. Forensic Sci Communi 2000;2(3).
14. Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KAJ, Buckleton J. Interpreting DNA mixtures. J Forensic Sci 1997;42(2):212–22. [PubMed]
15. Budowle B, Hobson D, Smerick J, Smith J. Low copy number–consideration and caution. Proceedings Twelfth International Symposium on Human Identification, 29 November 2001, Biloxi, MS.

16. Kinsey PT, Horman S. Modification of the stutter position label-filtering macro in the PE Biosystems Genotyper Version 2.5 software package: resolution of stutter-filter back talk. Forensic Sci Commun, July 2000; (2)3.
17. "Scientific Theory." Encyclopædia Britannica 2003.
18. Making The Nation Safer. The role of science and technology in countering terrorism. National Research Council of the National Academies, The National Academy Press, Washington, DC, 2002, pp.72.
19. Ryan JH. Ensuring the accuracy of STR allele calls for database samples via automated cross comparisons between two independent allele-calling platforms. 13th International Symposium on Human Identification. 7–10 October 2002, Phoenix, AZ.
20. Ryan JH, Leclair B. The SurelockID^{SM} Platform. The 8th CODIS Users Group Meeting. Washington, D.C., 5–8 November 2002.
21. Ryan JH, Barrus JK. Automated STR concordance analysis. SWGDAM, Washington, D.C., 14 January 2003.

Additional information and reprint requests:
John H. Ryan, Ph.D.
Myriad Genetic Laboratories
320 Wakara Way
Salt Lake City, Utah 84108