Evaluation of GeneMapper® *ID-X* and GeneMarker® HID for use in Forensic DNA Analysis

Ronald Schmidt BS [1], Justin Godby MSFS [1], Valerie Bostwick MSFS[1], Theresa Caragine PhD[2]

[1] Marshall University Forensic Science Center – 1401 Forensic Science Drive, Huntington, West Virginia 25701
[2] New York City – Office of the Chief Medical Examiner – 421 East 26th Street, New York City, New York 10016

ABSTRACT

Advancement in robotic technologies has, in recent years, caused a shift in the backlog of forensic DNA from evidence processing to sample profile analysis. In order to accommodate this new need, software systems are being developed to compete in a limited market. Different system options provide opportunities to tailor software to specific agency requirements in order to reduce analyst workload and laboratory resource utilization. The studies completed here provide a direct item comparison of system features for two DNA analysis software systems: Applied Biosystems (Foster City, California) GeneMapper® *ID-X* (GM*ID-X*) and SoftGenetics LLC (State College, Pennsylvania) GeneMarker® HID (GM-HID). Simplified analysis and editing procedures along with several options applied during automated analysis prove GeneMarker® HID to be a more efficient software system. Algorithms utilized by the program to generate electropherograms also result in a gain of information and a significant reduction in the number of edits needed per sample. Along with these general advantages, GeneMarker® HID was able to solve issues specific to the New York City-Office of the Chief Medical Examiner (OCME) making it the preferred software package for future use.

INTRODUCTION

DNA analysis, while a highly discriminating identification technique, is a costly and time-consuming process. Together, time requirements and increased sample submission have contributed to a significant increase in backlogs that exist in both forensic casework and database sample processing at many agencies around the globe (1-3). Automation of laboratory methods used in DNA analysis has reduced the amount of time it takes to generate electronic data files for interpretation by trained analysts (3,4). However, the one area of this process that remains largely unaffected by automation is data analysis (2). It has been reported that this aspect of the process can constitute upwards of half of the case workload facing analysts (2,5). Though commonly employed software packages such as Applied Biosystems' (AB) GeneScan®/Genotyper® (GS/GT) and GeneMapper® *ID* (GMID) are sufficient for data file analysis, recently several new software systems have been made available to the forensic science community that are meant to ease time constraints placed on crime laboratory analysts.

GM*ID-X*, an upgrade to AB's previous GMID software, and GM-HID (SoftGenetics) are two such programs. Along with being expert systems, these systems offer a variety of additional tools meant to assist analysts with processing case data. These tool include mixture deconvolution assistants and, in the case of GM-HID, relationship testing and database searching (2,6). While the time saving effects of using these programs as expert systems are well documented, undertaking such a measure would require a time consuming validation project. These systems also contain differences in base programming that will cause them to interpret the same data in similar but different manners. Also, different analysis parameters and data input allow laboratories to customize software packages to suit specific needs.

The emergence of new forensic technologies will, in the future, require laboratories to employ newer software systems such as GM*ID-X* and GM-HID. In this study, we sought to evaluate aspects of the user interfaces of these programs in order to reduce analyst time expenditures as well as insure high fidelity data output.

MATERIALS AND METHODS

- Previously generated data files used – no new laboratory work was completed

- Samples – amplified with AB Identifiler® and run on AB 3130xL genetic analyzers

- Data was produced under both low copy and high copy protocols, and contained single source and generated mixtures profiles, as well as low and high amplification input samples

- Analysis included OCME validated stutter ratios and a 10% general filter

- GeneMapper® *ID-X* v1.0/ GeneMarker® HID v1.95 used for new data file analysis

- Data file editing was completed in accordance with OCME standard protocols

- For each study qualitative results and/or appropriate statistics were compared

- Comparison studies were completed to assess ease of use for each system, amount of necessary user interaction, possible reductions in reporting of erroneous data, and impact on amount of useful data that was reported. Studies included: Data File Analysis, Profile Editing Procedures, Saturated Data, Stutter filters, Pull-up Correction, Software System Edit Requirements, Average Peak Height Differences, and a Concordance Study with GeneScan®/Genotyper® Data

- Several studies of GeneMarker® HID's Saturated Peak Repair option were also completed in order to assess the viability of the option for use in casework. Studies included: Saturated Repair Option Height Ratio Comparison, Saturated Repair Option Pull-up Edit Movement Comparison

RESULTS

Data File Analysis

- Systems load sample files in a similar fashion
- GeneMapper® *ID-X*
  - Sample designations along with Analysis Methods and Size Standards need to be manually assigned per sample prior to analysis
- GeneMarker® HID
  - Sample designations assigned using predefined identifiers
  - Analysis Methods with Size Standards saved and assigned with one click for entire sample set

Profile Editing Procedures

- GeneMapper® *ID-X*
  - Allele label re-entered through dialog box prior to assigning edit
  - Edit code is manually assigned through a subsequent dialog box
  - While deleting allele labels edit code entry may be required
- GeneMarker® HID
  - Entry of allele label not required to assign an edit code
  - Edit codes assigned from dialog box – manual or 10 preset options
  - Deleting allele labels requires no further input

Saturated Data

- GeneMapper® *ID-X*
  - An indicator line crosses all color channels to alert user of saturated data
- GeneMarker® HID
  - No static warning of present saturated data
  - Offers a Saturated Peak Repair option for automated analysis
    - Removes pull-up from incorrect color channels
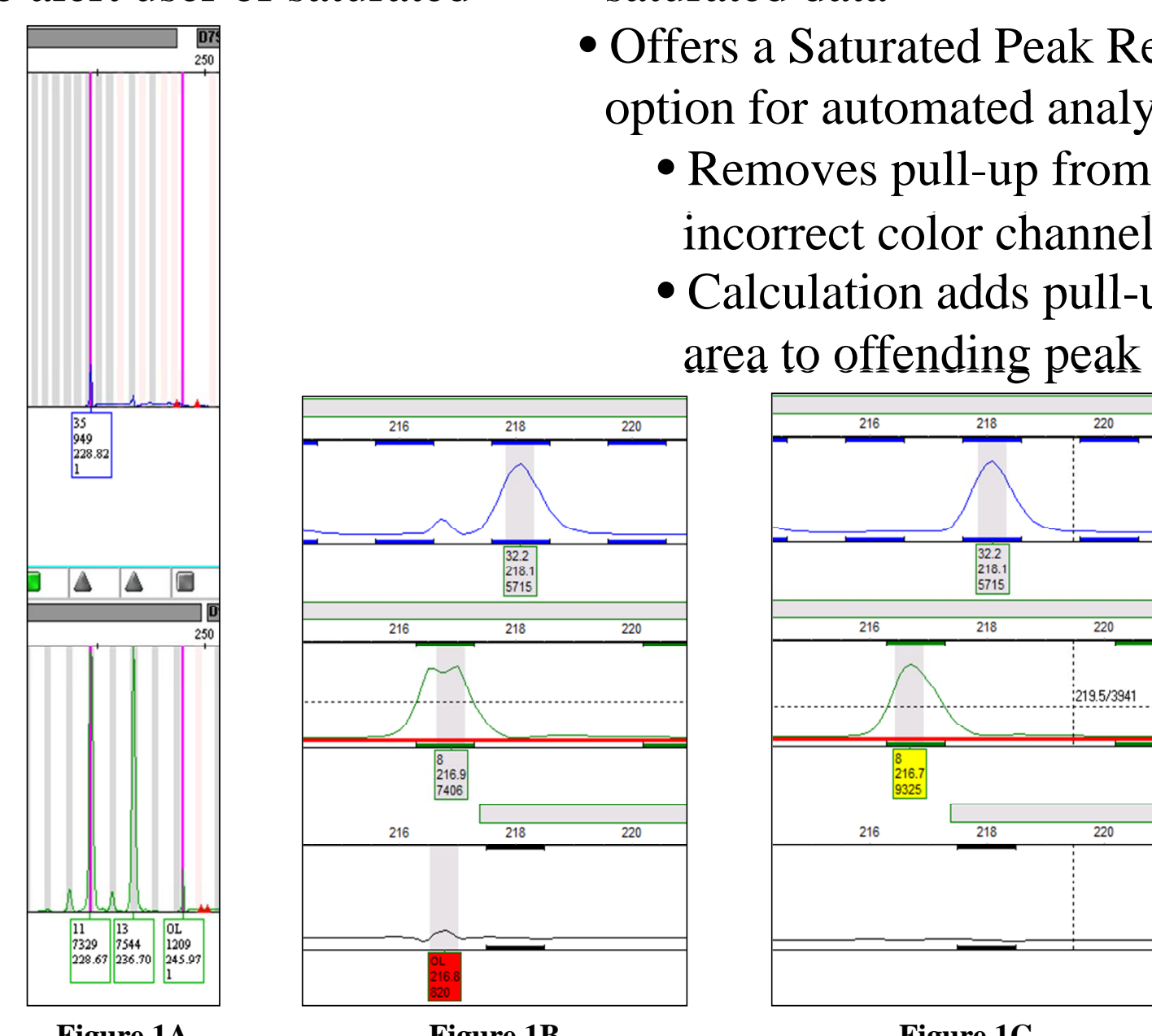   - Calculation adds pull-up peak area to offending peak

Figure 1: (A) Pink oversaturation indicator lines present in GeneMapper® *ID-X*. (B) Oversaturated data displayed in GeneMarker® HID. (C) Data from panel B reanalyzed with Saturated Repair option enabled. Peaks attributed to the offending saturated green peak are removed from color traces and added to the green peak.

Stutter Filters

- GeneMapper® *ID-X*
  - 8 filters available for each marker
  - Individual basepair location and ratios
    - 4 minus filters
    - 4 plus filters
- GeneMarker® HID
  - 3 filters available for each marker
  - N-x, N-2x, and N+x basepair locations with individual ratios
    - N − length of the fragment
    - x = length of repeat for marker

Pull-up Correction

- GeneMapper® *ID-X*
  - Some testing protocols display distorted peak shapes and extra allele calls from pull-up with kit positive samples
- GeneMarker® HID
  - Pull-up Correction option allows for the elimination of this issue by subtracting out extreme size standard pull-up
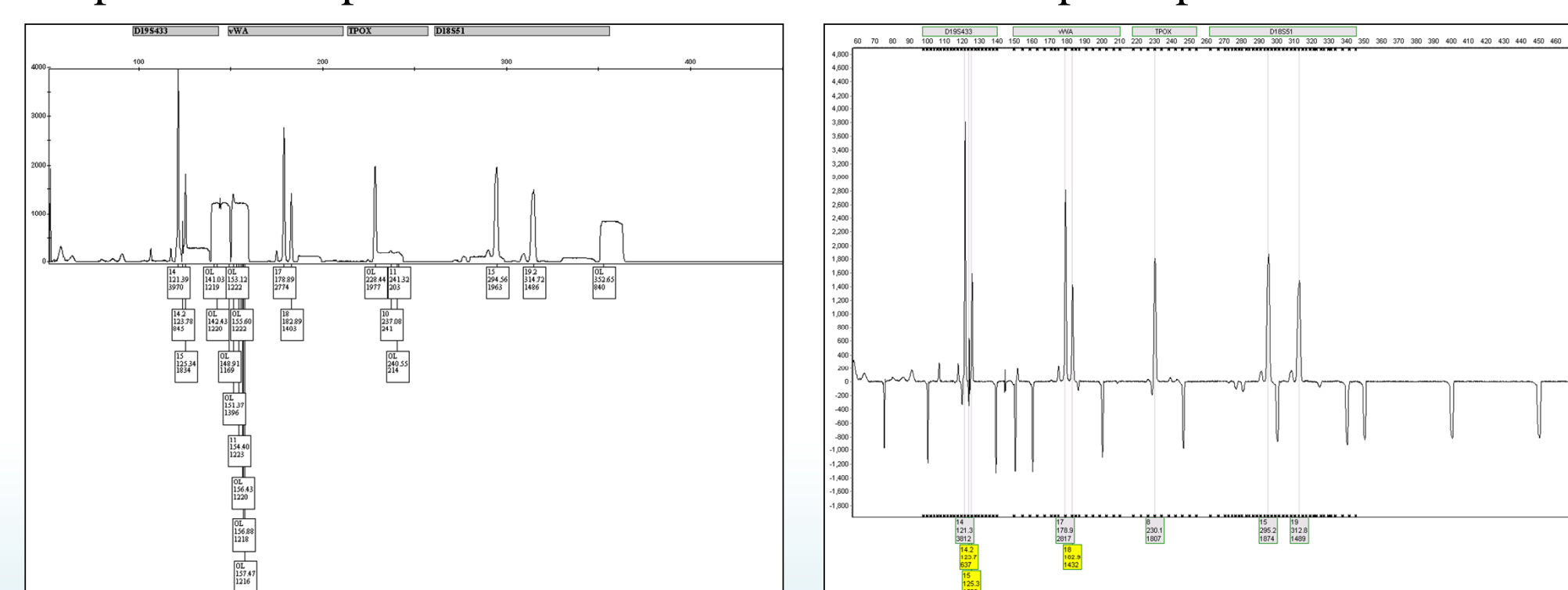

Figure 2: (A) Distorted peak shapes and extra allele calls attributed to internal size standard pull-up seen when using GeneMapper® *ID-X*. (B) Run data analyzed in GeneMarker® HID with the Pull-up Correction option enabled during analysis. Peaks under baseline mirror peaks present in GS 500 LIZ size standard.

Software System Edit Requirements

Table 1: Average edits per sample for examined software systems. Difference in relation to GeneScan®/Genotyper® shows GeneMapper® *ID-X* and GeneMarker® HID with Saturated Repair option enabled exhibited significant reductions in required numbers of edits.

50 pg input	GS/GT	GM*ID-X*	GM-HID (No Repair)	GM-HID (Repair)
Average Edits per Sample	6.1	5.4	8.8	4.4
Difference from GS/GT	--	-11%	45.2%	-27.4%

100 pg input	GS/GT	GM*ID-X*	GM-HID (No Repair)	GM-HID (Repair)
Average Edits per Sample	21.9	17.5	23.4	4.8
Difference from GS/GT	--	-20.2%	6.8%	-78.3%

Table 2: Direct comparison of positive results for systems of interest. GeneMarker® HID showed a further reduction in editing in relation to GeneMapper® *ID-X* with Saturated Repair option enabled.

50 pg input	GM*ID-X*	GM-HID (Repair)
Average Edits per Sample	5.4	4.4
Difference From GM*ID-X*	--	-18.5%
100 pg input	GM*ID-X*	GM-HID (Repair)
Average Edits per Sample	17.5	4.8
Difference From GM*ID-X*	--	-72.9%

Average Peak Height Differences

- GeneMapper® *ID-X*
  - Peak heights 2.8% lower than GeneScan®/Genotyper®
- GeneMarker® HID
  - Peak heights 0.4% lower than GeneScan®/Genotyper®

Concordance Study With GeneScan®/Genotyper Data®

- GeneMapper® *ID-X*
  - Gained one accurate allele
  - Retained inaccurate alleles
  - Retained all accurate alleles
- GeneMarker® HID
  - Gained three accurate alleles
  - Lost two inaccurate alleles
  - Lost one accurate allele

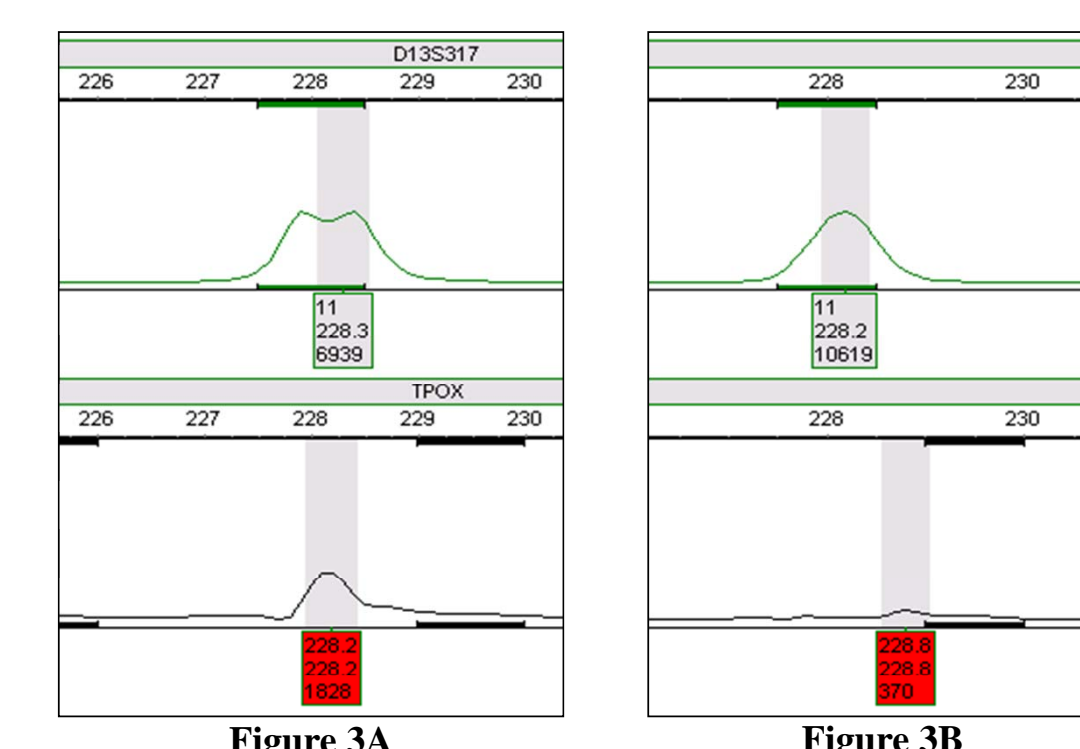### Saturated Repair Option Height Ratio Comparison

Table 3: Average difference between the peak height ratios of identical samples. Values calculated as the absolute value of the difference between ratios with and without the repair option engaged.

Peaks Repaired	Average Height Ratio Difference	Standard Deviation
None	0.03	0.05
One	0.12	0.12
Two	0.06	0.05

- 12.1% of effected loci displayed a reversal in major/minor peak relationships as compared to equivalent data unaffected by the Repair Option
- 4.8% of artificial peak height values were reduced below a 50% threshold

Saturated Repair Option Pull-up Edit Movement

- Repair option will remove normal peak shaped area
- Several peaks displayed positive side tailing not removed by Repair Option
- Residual peaks may still be called as artifacts by the software
- This shifted basepair length may limit edit code assignment and label removal
- Study included 253 repaired peaks – 13 residual peaks – 5.1% of all peaks

Figure 3: (A) Peak morphology of a pull-up peak that will leave a residual peak after repair. (B) Peak morphology of a residual peak after repair. (C) Table displays length difference data for true data peak and artifact peak with and without the Repair option enabled.

Difference Between Peak Length (bp)	
Average Before Repair	0.1
Standard Deviation	0.1
Average After Repair	0.6
Standard Deviation	0.1

DISCUSSION

- While both systems offer similar options for analysis method configuration, GM-HID offers several unique analysis options that may be applied at data analysis. GM-HID also reduces analyst interaction time by requiring less manual information input prior to automated review.

- The majority of analyst interaction with sample data occurs while editing erroneous data. GM*ID-X* utilizes an interface that requires significant analyst interaction at multiple steps during the process. GM-HID however, offers a streamlined procedure that reduces time spent editing sample profiles.

- Saturated data can complicate data analysis by misrepresenting both true peak RFU values as well as heterozygote peak height ratios. The indicator line provided by GM*ID-X* alerts analysts to such issues; however the problem itself will need to be rectified by additional laboratory work. GM-HID offers a Saturated Peak Repair option that uses a calculation to correct issues encountered during electrophoresis and reduces analyst interaction with sample data (Figure 1).

- While GM*ID-X* offers far greater flexibility with its available stutter filters settings, GM-HID did allow for more customization than currently employed software. Both programs were able to alleviate chemistry kit issues.

- Current protocols utilized at the OCME will cause severe pull-up to be displayed with kit positive samples. GM*ID-X* retains this problem as seen in previous AB software. GM-HID contains a Pull-up Correction option that automatically subtracts out severe pull-up and eliminates the need for corrective laboratory work or time-consuming editing (Figure 2).

- Both systems exhibit a reduction in edits as compared to previously used software (Tables 1,2). The GM*ID-X* reduction may partially be attributed to the flagging of some pull-up peaks as spike artifacts, removing them from results. The GM-HID reduction seen when utilizing the Saturated Repair option also can be attributed mostly to the removal of pull-up in samples. Algorithm differences may also be partially responsible for loss; however this was not quantified in the study.

- Both systems displayed a reduction in average RFU peak heights as compared to GS/GT. There was, however, a marked increase in peaks heights in GM-HID as compared to GM*ID-X*. The validity of extra data with GM-HID was not assessed in this study.

- Concordance study results appear to be independent of peak RFU heights as both systems exhibit a reduction in relation to GS/GT while exhibiting a gain in accurate data. Differences in information most likely result from different and/or improved peak recognition algorithms.

- Given these results, the viability of the Saturated Peak Repair option offered by GM-HID was assessed. There was no significant difference in peak height ratios displayed by the system when data was or was not affected by the software (Table 3). On a few occasions the PHR was reduced below 50%, the minimum requirement for association in mixture samples. However, this result was not seen in any of the examined mixed samples.

- During the viabilities study for the saturated peak repair option an issue with residual pull-up peaks that displayed a significant shift in basepair length from non-repaired data arose. However, software warnings coupled with analyst training could minimize the effect of a new artifact on data review time (Figure 3).

REFERENCES

1. Fang R, et. al. "The HID EVOlution System for Automation of DNA Quantification and Short Tandem Repeat Analysis". JALA 2010; 15: 65-73.
2. Holland MM, Parson W. "GeneMarker® HID: A Reliable Software Tool for the Analysis of Forensic STR Data." Journal of Forensic Sciences 2011; 56(1): 29-35.
3. Frappier R, Calandro L, Schade LL. "Improving Forensic DNA Laboratory Throughput: Enhanced Data Analysis and Expert Systems Capability." Forensic Magazine Feburary/March 2008.
4. Power T, Mccabe B, Harbison S. "FaSTR DNA: A New Expert System for Forensic DNA Analysis." Forensic Science International: Genetics 2008; 2: 159-165.
5. Butler JM. *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers, 2nd Ed.* Burlington, MA: Elsevier Academic, 2005.
6. SoftGenetics. *GeneMarker® HID v 1.95 User Manual*. State College, PA: SoftGenetics 2010.

ACKNOWLEDGMENTS