NIST National Institute of Standards and Technology

Email: erica.butts@nist.gov
Phone: 301-975-5107

Validation of PowerPlex 16 HS in Comparison to Identifiler Plus on the Applied Biosystems 3500 Genetic Analyzer

Erica L.R. Butts, Carolyn R. (Becky) Hill, David L. Duewer, John M. Butler, Peter M. Vallone

National Institute of Standards and Technology (NIST), 100 Bureau Drive MS 8314, Gaithersburg, MD 20899-8314

Poster # 42 at:
22th ISHI, National Harbor, MD
October 3-6, 2011

Poster available for download from STRBase:
http://www.cstl.nist.gov/biotech/strbase/pub_pres/ButtsISHI2011poster.pdf

The need to internally validate new technology is an on-going process within the forensic DNA community. Prior to being implemented into a forensic laboratory, a new product used in forensic DNA analysis must be validated internally within each laboratory to ensure the technology generates reliable and reproducible results, as suggested by the SWGDAM revised validation guidelines published in 2004 [1]. With the continuing advancement of technology, validation methods will continue to be an important part of any laboratory process. The ABI 3500 Genetic Analyzer is the newest capillary electrophoresis instrument available to the forensic DNA community. With several significant changes from the previous 31xx generation instruments, to include but not limited to, the type of laser, increased signal intensity, and RFID tracking of reagents [2]. An internal validation of the Applied Biosystems 8-capillary 3500 Genetic Analyzer was performed using two commercial short tandem repeat (STR) multiplex kits (PowerPlex 16 HS and Identifiler Plus).

Validation experiments to evaluate performance of the 3500 platform consisted of a precision study, sensitivity study, and genotype concordance. Injection parameters were also varied to identify the optimal injection time for both PowerPlex 16 HS and Identifiler Plus on an 8-capillary ABI 3500. Results from these studies were used in setting analytical and stochastic thresholds for both kits. Data are shown for both PowerPlex 16 HS and Identifiler Plus to examine the differences in performance between the two STR typing kits.

Funding
This work was funded in part through interagency agreement between the National Institute of Justice and the NIST Office of Law Enforcements and between NIST and the FBI Biometrics Center of Excellence.

References

[1] SWGDAM (2004). Revised validation guidelines. Forensic Science Communications, 6(3). Available at http://www2.fbi.gov/hq/lab/fsc/backissu/july2004/standars/2004_03_standards02.htm.
[2] Applied Biosystems. (2011). User Bulletin Applied Biosystems 3500/3500xL Genetic Analyzer. Foster City, CA
[3] Butler, J.M. (2009) *Fundamentals of Forensic DNA Typing*. Elsevier Academic Press: San Diego
[4] Gilder, J. R., Doom, T.E.; Inman, K., .et al. Run-Specific Limits of Detection and Quantitation for STR-based DNA Testing. J Forensic Sci. 52 (2007): 97-101.
[5] Armbruster, D. A., Tillman, M. D., L. M. Hubbs. Limit of detection (LQD)/limit of quantitation (LOQ): comparison of the empirical and the statistical methods exemplified with GC-MS assays of abused drugs. Clin.Chem. 40 (1994): 1233-1238.
[6] Ellison, S.L.R., Barwick, V.J., Farrant, T.J.D. (2009). Practical Statistics for the Analytical Scientist. RSC Publishing: Cambridge.

What is Validation?

From Section 1.1 of the SWGDAM Revised Validation Guidelines: Validation is the process by which the scientific community acquires the necessary information to:

(a) Assess the ability of a procedure to obtain reliable results.
(b) Determine the conditions under which such results can be obtained.
(c) Define the limitations of the procedure.

The validation process identifies aspects of a procedure that are critical and must be carefully controlled and monitored [3].

Three main objectives of validation include: Reliability, Reproducibility, and Robustness.

Reliable method – the obtained results are accurate and correctly reflect the sample being tested
Reproducible method – the same or very similar results are obtained each time a sample is tested
Robust method – successful results are obtained a high percentage of the time and few, if any, samples need to be repeated

General steps in an internal validation include:

- Review literature and learn the technique
- Determine necessary validation studies
- Collect/obtain samples, if necessary
- Perform validation studies maintaining all documentation
- Summarize the studies and submit for approval to Technical Leader
- Write-up the analytical procedure(s). Include quality assurance (controls, standards, critical reagents and equipment) and data interpretation, as applicable
- Determine required training and design training module(s)
- Design qualifying or competency test

Internal Validation Experimental Design

The experimental design of a validation is dependent on both the method being validated as well as the purpose. An internal validation is an accumulation of test data within a laboratory to demonstrate that established methods and procedures perform as expected in the laboratory. **The SWGDAM Revised Validation Guidelines suggest a total of 50 samples for internal validation testing,** in which some samples may overlap into different experimental examinations.

	Test	Types of Samples Used	Number Examined
Reliability	Injection Parameters	Samples heterozygous at 15 loci plus Amelogenin at 1.0 ng DNA input	15 (3 samples per injection)
Reliability	Size Standard Dilution	Samples heterozygous at 15 loci plus Amelogenin at 1.0 ng DNA input; Amplified Extraction Blanks	16
Reproducibility	Precision	Allelic Ladders	24
Reproducibility	Concordance	50 genomic DNA samples at 1.0 ng DNA input; SRM 2391b: 10 genomic DNA samples at 1.0 ng DNA input	60
Robustness	Sensitivity	3 dilution series of samples heterozygous at 15 loci plus Amelogenin	84 (4 replicates of each dilution series)
Robustness	Mixtures (Data Not Shown)	Mixture dilution series (1:1, 1:3, 1:5, 1:7, 1:9, 1:10, and inverse) of 2 samples heterozygous at 15 loci plus Amelogenin at 1.0 ng DNA input	28
	Total Number of Samples Examined		**227**

Sample Setup and Parameters

	PowerPlex 16 HS	Identifiler Plus
Total Reaction Volume	12.5 µL	12.5 µL
Cycle Number	30	28
Size Standard	ILS 600	LIZ 600 v2.0
Size Standard Dilution for CE	0.25 µL + 9.75 µL HiDi	0.5 µL + 8.5 µL HiDi
Injection Parameters	1.2 kV for 10 sec (Default: 1.2 kV for 15 sec)	1.2 kV for 5 sec (Default: 1.2 kV for 15 sec)

Primary Differences Between 31xx and 3500

	31xx Platforms	3500 Platforms
Laser	Argon ion (AR+) with 488/514 nm wavelength	Single-line 505 nm, solid-state, long-life laser
Power Requirement	220V	110V
File Generated	.fsa files	.hid files
Signal Normalization	None	Instrument-to-instrument available
Optimal Signal Intensity	1500-3000 RFU	~4x greater than 31xx platforms

Other Features of the 3500 Platform

- The .hid files generated require the use of GeneMapper *ID-X* v1.2 for data analysis
- Required use of pre-packaged reagents with RFID tracking technology
- No lower polymer block which results in fewer air bubbles
- Improved temperature control due to improved sealing around the oven and detector cell
- Improved peak height uniformity across capillaries, runs, and instruments
- Normalization feature only available with Applied Biosystems kits and the use of the LIZ 600 v2.0 size standard
- 6-dye detection capability

Analytical Threshold Methodology

The minimum threshold for data comparison and peak detection in the DNA typing process [3].

The analytical threshold (AT) was calculated by evaluating the limit of quantitation (LOQ) which is defined as the threshold beneath which measurements of signal strength cannot be reliably used [4]. The LOQ is commonly expressed as the average background signal plus 10 standard deviations [5]. The LOQ was determined by evaluating the baseline noise values from the sensitivity study (DNA dilution series) data. Data was analyzed with a threshold of 1 RFU set within GeneMapper *ID-X* v1.2 software. Calls for all alleles and artifacts (stutter, n+4, pull-up, etc) were removed. Remaining values were exported where the average noise was calculated per dye channel.

AT= Average Noise RFU + (10 x Standard Deviation)

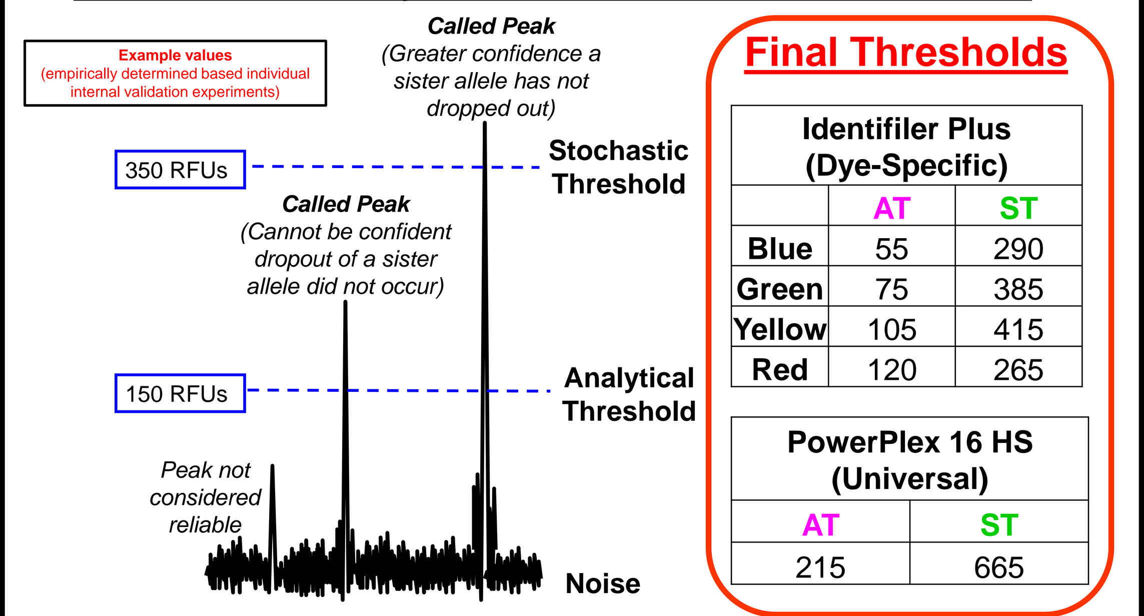
The analytical threshold for PowerPlex 16 HS and Identifiler Plus when applying one universal threshold across all dye channels was 215 RFU and 120 RFU, respectively. The statistical difference between each dye channel for both PowerPlex 16 HS and Identifiler Plus were calculated using a z-test [6] and proved to be significant enough to treat the dye channels independently for Identifiler Plus, but not significantly different for PowerPlex 16 HS.

Stochastic Threshold Methodology

The peak height value above which it is reasonable to assume that, at a given locus, allelic dropout of a sister allele has not occurred [3].

The stochastic threshold (ST) was calculated by evaluating the data from the sensitivity study (DNA dilution series). The sample amounts where dropout was observed (50 pg, 30 pg, 10 pg) were examined for stochastic effects including severe imbalance of heterozygous alleles and allelic and full locus dropout. Several methods for deriving a stochastic threshold were evaluated.

Overview of Analytical and Stochastic Thresholds

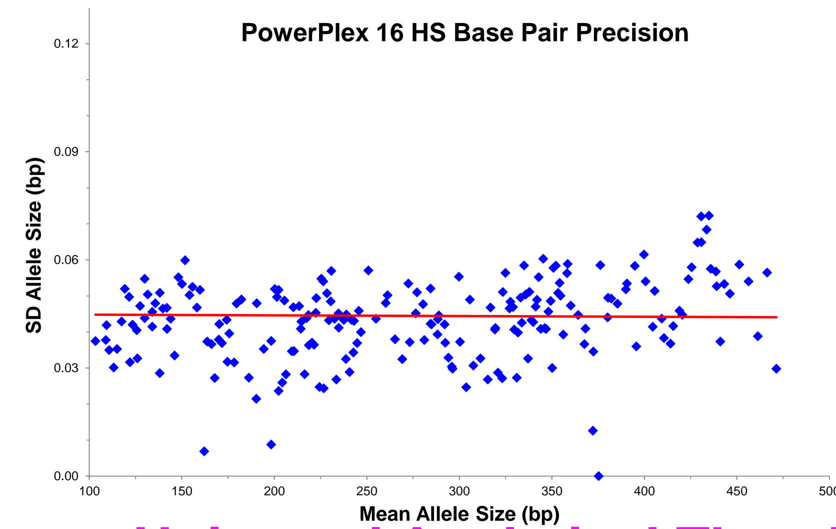


Final Thresholds

Identifiler Plus (Dye-Specific)	AT	ST
Blue	55	290
Green	75	385
Yellow	105	415
Red	120	265

PowerPlex 16 HS (Universal) AT	ST
215	665

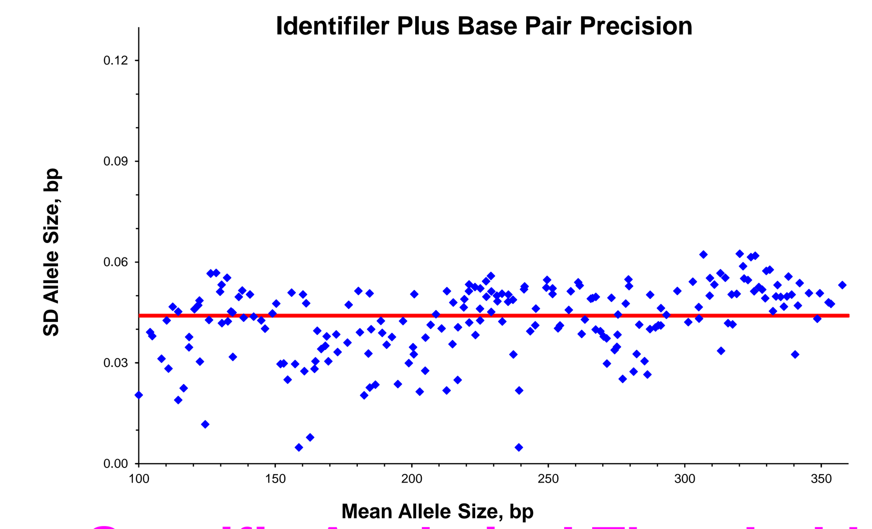
PowerPlex 16 HS Validation



Precision

Precision of **base pair sizing +/- 0.05 bp** between **allelic ladders tested** for **PowerPlex 16 HS** and **+/- 0.04 bp** for **Identifiler Plus**. The graph represents the allelic base pair precision of each allele within 24 allelic ladders. The x-axis is the average allele size in base pairs and the y-axis is the standard deviation of allele size in base pairs.

Identifiler Plus Validation

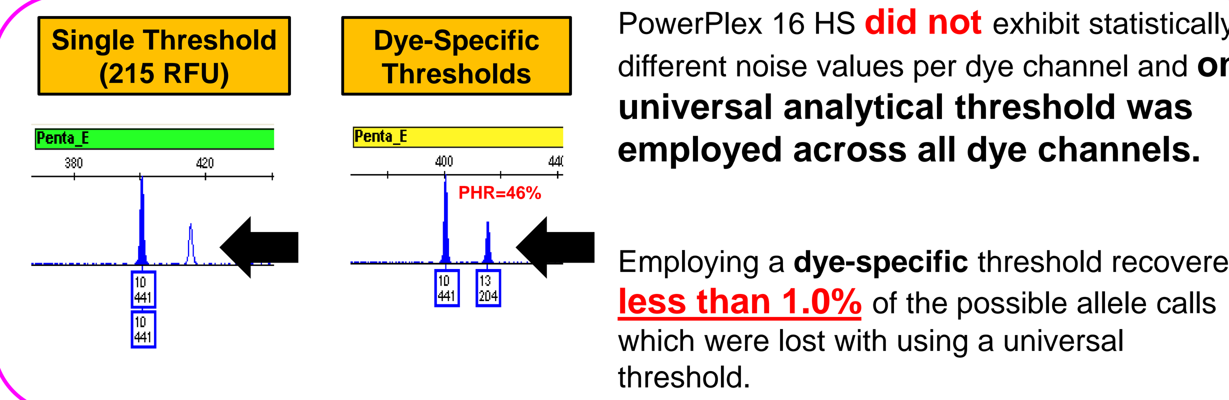


Universal Analytical Threshold

Dye Channel	Average RFU	Stdev	Min RFU	Max RFU	Calculated LOQ (RFU)	Universal Analytical Threshold
PowerPlex 16 HS						
Blue	16	17.4	2	99	190	215
Green	20	19.3	2	99	212	215
Yellow	18	19.3	2	99	211	215

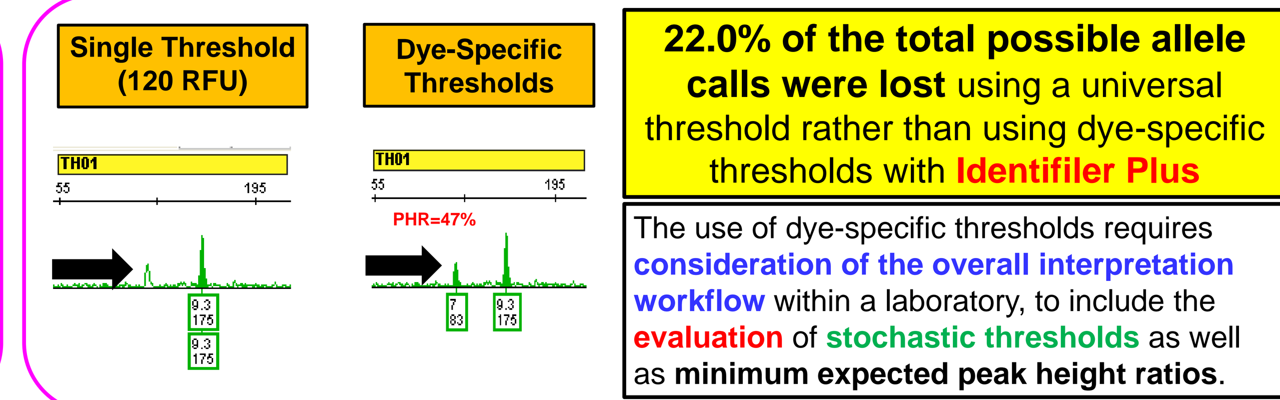
Dye-Specific Analytical Threshold

Dye Channel	Average RFU	Stdev	Min RFU	Max RFU	Calculated LOQ (RFU)	Dye-Specific Analytical Threshold
Identifiler Plus						
Blue	10	4.6	3	68	55	55
Green	16	5.6	3	78	72	75
Yellow	24	7.9	7	63	103	105
Red	31	8.9	7	81	120	120



PowerPlex 16 HS **did not** exhibit statistically different noise values per dye channel and **one universal analytical threshold was employed across all dye channels.**

Employing a **dye-specific** threshold recovered **less than 1.0%** of the possible allele calls which were lost with using a universal threshold.



22.0% of the total possible allele calls were lost using a universal threshold rather than using dye-specific thresholds with **Identifiler Plus**

The use of dye-specific thresholds requires **consideration of the overall interpretation workflow** within a laboratory, to include the **evaluation** of **stochastic thresholds** as well as **minimum expected peak height ratios.**

Stochastic Threshold

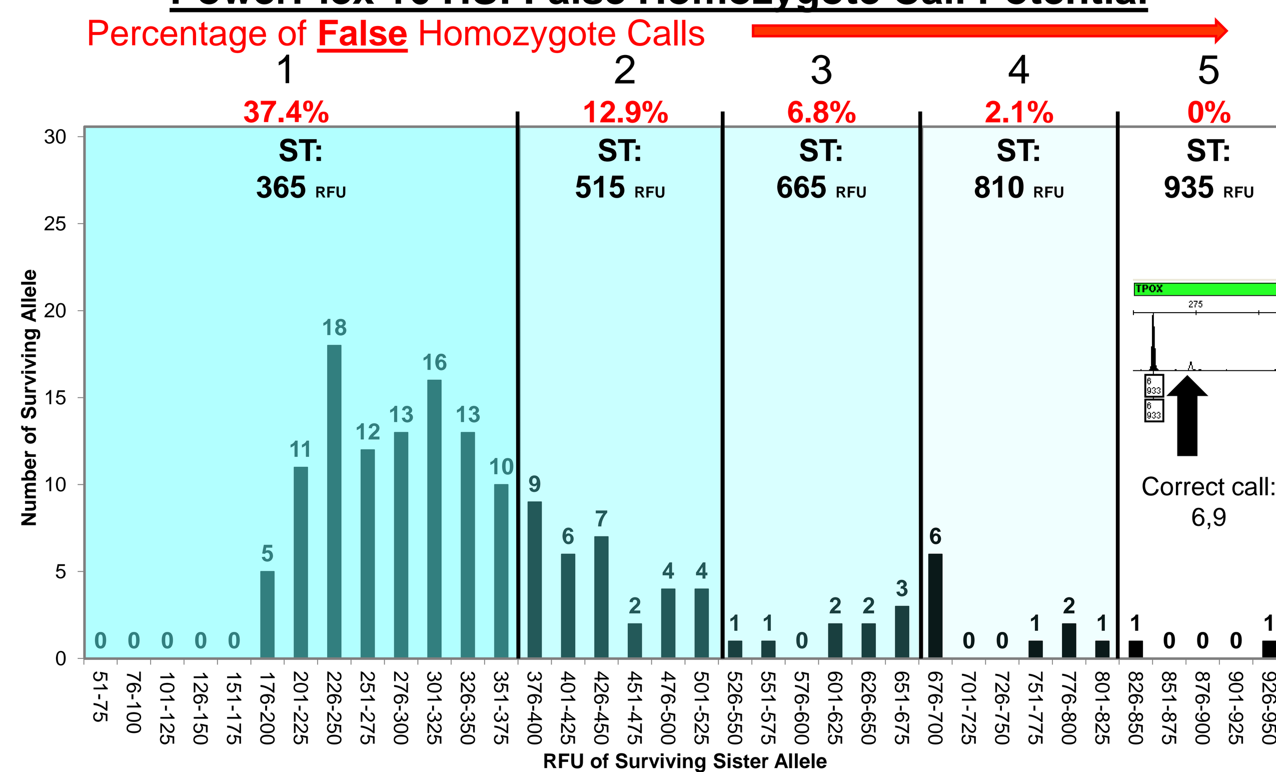
Several methods were examined to determine the stochastic threshold:

1. Average RFU of surviving false homozygote peaks
2. Average RFU of surviving false homozygous peaks + 1 x Standard Deviation
3. Average RFU of surviving false homozygous peaks + 2 x Standard Deviation
4. Average RFU of surviving false homozygous peaks + 3 x Standard Deviation
5. Maximum RFU of surviving false homozygote peak

The stochastic threshold should be evaluated to **best fit the data produced within the scope of the validation**. There **may** be occasions where data falls outside the bounds of the stochastic threshold leading to potential false homozygote calls.

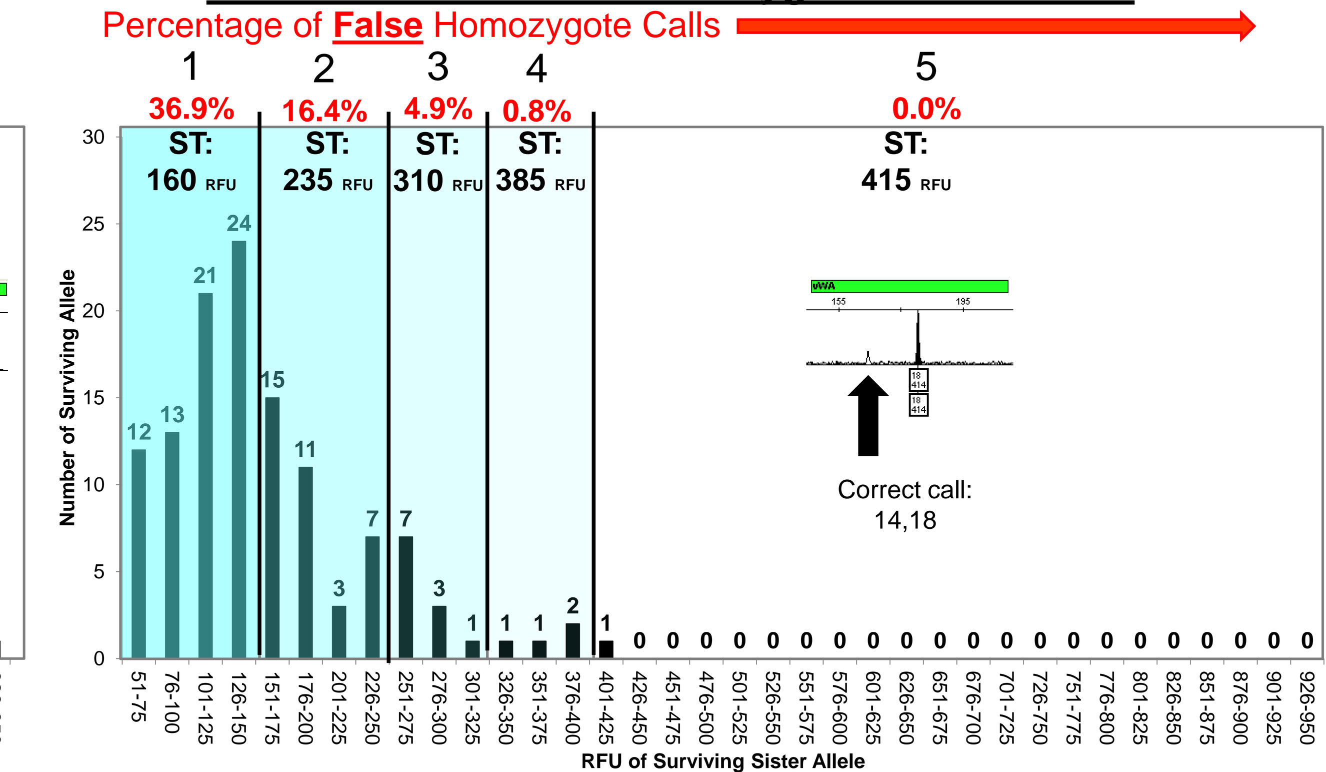
PowerPlex 16 HS: False Homozygote Call Potential

Percentage of **False** Homozygote Calls



Identifiler Plus: False Homozygote Call Potential

Percentage of **False** Homozygote Calls



Conclusions

The 3500 has proven to be reliable, reproducible and robust throughout our internal validation of both PowerPlex 16 HS and Identifiler Plus. All 60 samples tested for concordance were concordant between both PowerPlex 16 HS and Identifiler Plus as well as between testing using the 3500 and 3130xl. A total of 2809 alleles for Identifiler Plus and 2829 alleles for PowerPlex 16 HS were examined between concordance and sensitivity studies.

In testing both kits, fundamental differences were established between PCR cycle number and separation parameters on the 3500. The difference in the number of cycles between PowerPlex 16 HS and Identifiler Plus was 2 cycles (30 cycles and 28 cycles, respectively). Injection parameters on the 3500 were optimized for both PowerPlex 16 HS and Identifiler Plus (injection of 1.2 kV for 10 seconds compared to injection of 1.2 kV for 5 seconds). Both of these differences can account in part for the differences between the LOQ as well as the observed stochastic effects and relative peak heights per kit.

The analysis of setting dye-specific analytical thresholds rather than applying one universal analytical threshold resulted in less allelic and full locus dropout for Identifiler Plus. PowerPlex 16 HS did not have a statistical difference between the calculated noise values to employ the use of dye-specific analytical thresholds. For this reason a universal analytical threshold was set for all dye channels. The linked nature between analytical and stochastic thresholds suggests if the analytical threshold is adjusted, then the stochastic threshold should be reevaluated along with minimum expected peak height ratios.