Scientific Working Group on DNA Analysis Methods

Module for Autosomal Multiplex Kit Internal Validation Guidance

The following are best practice guidelines to help laboratories design internal validation studies for autosomal multiplex kits. It is recognized that laboratories may have validation experimental approaches that differ from those listed in this document. Likewise, the examples given throughout this document are informational and are not meant to dictate the types and numbers of samples every laboratory must use to satisfy each study. If the laboratory determines that a study is not applicable, the reason(s) will be documented.

Prior to beginning validation studies, the laboratory should verify the instrument operation is within manufacturer specifications to ensure optimal performance. Laboratories should include samples extracted using all methods expected to be tested with the multiplex kit. This will allow laboratories to determine if the validated extraction chemistries are compatible with the performance of the multiplex kit. It is also recommended that laboratories use samples recently quantified with current laboratory procedures. Purchased human DNA samples may be used for comparison or to supplement the laboratory processed samples.

Validation studies cannot account for all scenarios that may arise during casework examinations; however, laboratories should endeavor to cover the range of variation expected to be encountered with forensic samples. Following implementation, laboratories should review results and if necessary, conduct supplemental studies to improve workflow, thresholds and/or interpretations.

This Autosomal Multiplex Kit Internal Validation Guidance Module should be used in conjunction with the SWGDAM Validation Guidelines for Forensic DNA Analysis Methods (https://www.swgdam.org/publications). The studies in each module are not synchronized to the

FBI QAS; instead, they are presented in a suggested order to conserve resources such as time, reagents, samples and consumables and to streamline required testing.

1. Sensitivity and Stochastic Studies

1.1 *Study Purpose:*

Sensitivity studies are performed to determine the dynamic range, ideal target range, limit of detection, heterozygote balance (e.g., peak height ratio [PHR]) and the signal-to-noise ratio associated with the assay. Sensitivity studies can also be used to evaluate stochastic effects generally resulting from low quantity and/or low quality samples as well as assessing oversaturation effects created by high template samples in addition to quantification thresholds. Data obtained from the sensitivity study should be used to determine an analytical threshold for each dye channel – a threshold can be established either per dye channel or across all dye channels of the multiplex kit. However, to establish an appropriate baseline level of your process, samples which exceed or are too far below the optimal amplification target range determined by the laboratory should not be used for analytical threshold determination. Samples with too much DNA may produce baselines with excessive levels of noise whereas samples containing little to no DNA may underestimate the levels of noise.

Laboratories should characterize peak height ratio variation observed for each locus of the multiplex kit and determine expected peak height ratios over a range of DNA input quantities. The determined ratio(s) can be locus or kit specific, and data from all dye channels should be used. This information will allow the laboratory to better understand allelic imbalance, identify the possibility of allelic dropout and inform interpretation guidelines.

Laboratories should determine expected PCR stutter artifacts (e.g., n-4, n+4, n-2, etc.) for all loci in the multiplex kit. This information is expected to assist in the development of laboratory interpretation guidelines.

Laboratories should establish a stochastic threshold for any binary interpretation method. A stochastic threshold is the peak height value below which it is reasonable to assume

that, at a given locus, allelic dropout of a sister allele in a heterozygous pair may have occurred.

If a laboratory chooses to make modifications to sensitivity such as cycle numbers, injection amount or time, validation of these parameters, to include the stochastic threshold, would be necessary.

1.2 *Study Considerations:*

The following sample types are recommended for use with this study:

- single source samples of good quality
- samples that exhibit a high degree of heterozygosity
- samples that exhibit a large separation between alleles for heterozygous genotypes
- amplification negative controls
- single source samples containing a range of DNA template quantities

1.3 *Study Example:*

For example, three reference samples are utilized with a well-characterized concentration that has recently been determined (e.g., replicate quantification determinations utilizing the validated laboratory procedure). The subsequent dilutions are also quantified to ensure they represent accurate estimates of input quantities. A dilution series for each sample, spanning a range of input quantities (e.g., 2 ng, 1 ng, 0.5 ng, 0.25 ng, 0.125 ng, 0.06 ng, 0.03 ng, 0.015 ng, 0.007 ng, 0.003 ng) is prepared and amplified in triplicate. Samples with dilutions at lower template concentrations of this range are expected to produce STR profiles where allele dropout is observed. Analyzing three replicates of three unique samples using the example dilution series described would provide data from 90 individual amplifications.

1.4 *Study Outcome:*

At the completion of these experiments, the laboratory should be able to utilize the data generated to establish the system's limit of detection as well as the analytical and stochastic thresholds to be used during data interpretation. Variation observed at

different DNA template quantities may include peak height ratios, inter-locus balance or average peak height.

The laboratory can use this data to determine the upper and lower limits of reliable interpretation as well as the optimal DNA input quantity or range. Optimal DNA template targets or ranges may be represented by a DNA quantity or range between the highest and lowest concentrations tested in the study. Samples amplified at an optimal template target or range should clearly distinguish true alleles from artifacts and should typically result in complete DNA profiles. This study will allow a laboratory to define an optimal target range, but does not preclude amplification of samples outside of the optimal target range. Upper limits may be determined based upon spectral pull-up, excessive stutter, off scale signal, increased artifact detection, locus balance and/or PHR. Lower limits may be determined based upon PHR, allelic and/or locus drop-out, allelic drop-in and/or elevated stutter.

The interpretation of DNA profiles, either by binary or probabilistic genotyping methods, is impacted by the sensitivity of the analysis. The results of the sensitivity studies can be used to generate a baseline sensitivity for the system. This would be specific to the amplification volume, amplification cycle number, volume of amplicon used for the injection, injection parameters and the instrument. The baseline sensitivity can be measured in the expected number of alleles detected or average relative fluorescent units (RFU) per locus for a given DNA template amount. If a lab has multiple capillary electrophoresis (CE) instruments and would like to have one analytical threshold (AT) across all CEs, then it is important to have similar sensitivities with data obtained from each instrument.

Baseline sensitivity can be used to calibrate the sensitivity of additional instruments, check new lots of amplification kits or inform performance check protocols following future maintenance. Maintaining sensitivity over time and across instruments, based on the expected baseline range established by the validation data, can help to ensure continuing consistency in data interpretation. This data can also be merged with that

obtained from other studies to calculate expected rates of stutter, baseline signal and heterozygote balance.

When evaluating the results, data from all dye channels should be used, and dye-artifacts or other known artifacts should be characterized and removed prior to calculating the AT. Samples demonstrating off-scale data or excessive artifacts should not be used for this study.

For binary interpretation approaches, laboratories should calculate peak height ratios and corresponding standard deviation using alleles from heterozygous pairs (e.g., low RFU peak/high RFU peak, or low molecular weight peak/high molecular weight peak). Alleles in the heterozygous pair should be separated by more than one repeat to avoid contribution from stutter artifacts. Peak area ratios may be calculated instead of peak height ratios. The laboratory should establish interpretation guidelines that determine acceptable PHR variation/PHR thresholds under defined conditions such as template quantity or peak heights. Samples demonstrating broad peaks, poor separation, degradation or inhibition should not be used for this analysis.

Laboratories should characterize stutter based upon size and amplitude relative to an allelic peak and calculate average stutter per locus (or allele), standard deviation and minimum/maximum stutter per locus. The data evaluation will be performed with the stutter filters turned off, and only the alleles detected and corresponding stutter peaks included, with non-stutter artifacts removed. Samples used to evaluate stutter should include the range of template amounts encountered in casework, however loci containing off-scale peaks should not be used. Laboratories should compare stutter data obtained from the internal validation to any manufacturer's developmentally validated data to ensure consistency. The laboratory should establish stutter thresholds to be used for data interpretation. Laboratory stutter thresholds may be established using published developmental validation studies if the laboratory has confirmed that these stutter values work with their internal validation data.

2. Precision (repeatability and reproducibility)and Accuracy

2.1 *Study Purpose:*

Laboratories should demonstrate that allele sizing/allele calling are sufficient to differentiate between alleles when using the multiplex kit for repeatability and reproducibility studies. Accuracy of the system (genotype concordance) is discussed in the Known and Non-probative section.

2.2 *Study Considerations:*

Environmental factors (e.g., ambient temperature, humidity) may affect sizing precision and therefore should be taken into consideration when setting up an instrument. These factors should be monitored throughout the validation study to determine if adjustments to the instrument and/or the environment are required.

Base pair sizing precision can be determined using injections of allelic ladders in different capillaries throughout the course of the internal validation studies. Sizing precision should be performed for all capillary electrophoresis instruments individually and should also be conducted for any modification to an instrument parameter or run condition that may affect migration (i.e., voltage, polymer). All capillaries of a CE instrument should be tested for precision.

2.3 *Study Example:*

For example, to determine the precision of an eight capillary instrument, one master mix containing the appropriate volumes of size standard, formamide and allelic ladder are prepared and aliquoted to each of the 8 wells in a single column of a 96 well plate. These allelic ladder samples are injected ten consecutive times on the CE instrument for a total of 80 injections. A laboratory may also choose to repeat this study on subsequent days, or at different times of the day if environmental factors (e.g., the room temperature) vary within the laboratory.

2.4 *Study Outcome:*

Base pair sizing is measured using the base pair size for each allele in the ladder. The average base pair size and standard deviation is calculated using this data. In general, three times the standard deviation (confidence interval) provides an estimate of the precision and should be <0.5 base pairs when targeting alleles that differ by one base.

Base pair sizing variations >0.5 base pairs may result in an incorrect allele call. A laboratory may use this information to determine the frequency of ladder injections and the laboratory policy regarding off ladder (or off-bin) allele designations.

Replicates and/or re-injections of samples from the sensitivity study can also be used to address the precision of RFUs and sizing, and inform expectations for PHRs in forensic samples.

3. Mixture Study

3.1 *Study Purpose:*

Laboratories should evaluate mixed DNA samples, representative of the number of contributors (NOC), contributor ratios and template quantities that are expected to be interpreted. These studies will assist the laboratory to establish guidelines for mixture interpretation, which may include determination of the NOC to the mixture, determination of the major and minor contributor profiles and contributor ratios or proportions in relation to male:autosomal or male:female DNA quantification determination.

3.2 *Study Considerations:*

Laboratories should conduct mixture studies after the sensitivity, PHR, AT, stochastic threshold and stutter studies have been completed; furthermore, the samples should be analyzed using the thresholds established by these studies. The following should be considered when creating mixtures to be included in this study:

- Prepare mixtures with high quality, single source known samples with a well-characterized concentration that has recently been determined (e.g., replicate quantification determinations utilizing the validated laboratory procedure)
- Include a number of unique known samples to include in the study and consider the potential use of available commercial mixture samples (e.g., NIST mixed source reference sample, proficiency test sample)
- The DNA samples should be extracted and quantified with the validated laboratory procedures intended for use with the typing kit being validated.

- Assess differing amounts of allele sharing between samples (e.g., utilizing related samples or samples with significant allele sharing to samples with little to no allele sharing) and instances of challenging peak resolution (e.g., 12 major & 12.1 minor alleles) for a set of mixture samples
- Samples subjected to degradation and/or inhibitors may be included

When designing this study, select combinations of samples that exhibit allele sharing and stutter masking and that challenge the resolution and dynamic range of the CE instrument. For example, increasing the NOC, and/or decreasing the amount of DNA challenges the method's ability to readily represent the true number of contributors. Goals of this study include:

- Assessing the limitations of the laboratory's process to estimate the NOC to a mixed sample. This study should include a variety of mixture combinations (with varying numbers of contributors) to assist in developing appropriate guidelines to reasonably estimate the NOC.
- Detecting the contributors to a mixed sample with varying contributor ratios. This study should include prepared mixtures with a range of mixture contributions which reflect the types of samples processed in the laboratory.
- Detecting the contributors to a mixed sample with varying amounts of overall template. This study should include prepared mixtures which exhibit saturation and/or drop out.
- Demonstrating potential variability in amplifications. This study should include multiple amplification replicates of prepared mixtures.

3.3 *Study Example:*

For example, extract and quantify ten or more unique reference samples (combination of both male and female samples). Prepare mixed DNA samples varying the gender, number, and combination of contributors, the contributor ratio (e.g., see table below) and varying the range of template amount based on mixtures commonly encountered in casework. The DNA samples are combined in order to challenge allele sharing, stutter masking, CE resolution, dynamic range, NOC determination, and NOC male

determination. Laboratories can utilize the resulting data to set quantification thresholds and/or male quantification thresholds. Each mixture set is amplified in duplicate. Mixtures are included which incorporate degraded DNA as well as mixtures where the NOC exceeds the number expected to be interpreted by the laboratory (in this example this laboratory would be expected to interpret up to four person mixtures).

2-person	3-person	4-person	5-person
M:F F:M	F:M:M M:F:M	F:F:M:M F:M:F:M	M:M:M:F:M M:F:M:M:F
1:1	1:1:1	1:1:1:1	1:1:1:1:1
2:1	2:1:1	3:1:1:1	5:2:2:1:1
3:1	5:1:1	5:2:2:1	10:5:5:2:2
4:1	5:2:1	10:5:2:1	20:10:2:1:1
5:1	10:5:1	10:5:5:1	
10:1	10:2:1	10:5:1:1	
20:1	20:5:1		
50:1			
100:1			
150:1			
200:1			
300:1			

3.4 *Study Outcome:*

The data from this study will provide an understanding of the performance of mixed DNA samples and will form the basis of the laboratory's mixture interpretation protocol. When evaluating the data, the laboratory may determine the criteria to be used to identify a mixture (e.g., the number of alleles present, peak height imbalance, etc.), distinguish alleles from potential artifacts (e.g., stutter) and estimate the number and relative ratio of contributors. The laboratory may define parameters for discerning a major contributor(s), and may determine the lower limit where a minor contributor component can be detected (e.g., 20:1, 50:1, 100:1). The effects of pull-up and stutter on mixture interpretation when a minor contributor to the mixture is observed at similar levels and how mixture ratios can vary across the profile at high and low levels of DNA template may also be considered.

The ability to reasonably assume the NOC in a mixture is an essential component of mixture interpretation. A variety of methods may be used to evaluate the NOC including maximum allele count per locus, total allele count, mixture ratios, PHRs, maximum likelihood or software tool(s). Laboratories may define the parameters for determining the NOC considering the effects of allele sharing, allelic dropout, allelic drop-in and if data below the analytical threshold will be used in the assessment of the NOC. The study will identify the following:

- which loci of the multiplex kit are better indicators of a mixed profile (e.g., observation of six alleles at the SE33 locus is a good indicator that the NOC is 3 vs six alleles at the TPOX locus which may suggest a NOC >3),
- the limitations of using maximum allele count per locus at multiple loci to determine the NOC (e.g., a 4-person mixture with 6 alleles at each of 8 loci may be mistakenly identified as a NOC of 3),
- and how degradation affects the determination of the NOC (e.g., impact on mixture ratio when different levels of degradation are present or the impact on the NOC determination when data at high molecular weight loci are absent).

The laboratory should determine the criteria to be used to determine if a mixed DNA profile is suitable for interpretation and the maximum NOC which will be interpreted. For example, a mixture may be determined to be uninterpretable if the profile has limited data and does not demonstrate a minimum number of alleles and/or loci, or is too complex and the NOC exceeds a defined maximum.

Upon completion of this study, the laboratory may establish interpretation guidelines, such as determining the estimated NOC, and ensure their appropriateness for the range of mixtures expected to be observed in casework.

4. Contamination Study

4.1 *Study Purpose:*

Laboratories should determine the susceptibility for contamination and/or drop-in occurring from the laboratory environment, consumables, reagents, instrumentation, or

analysts. Negative controls (negative amplification and/or reagent blank controls) and known samples can be used for the detection of exogenous DNA (including allele drop-in) originating from reagents, consumables, other samples, analyst(s) and/or the laboratory environment.

4.2 *Study Considerations:*

No additional studies are necessary to assess the potential for contamination. Data generated from all samples included with the other internal validation studies should be processed through DNA typing and assessed and evaluated for any potential contamination. Allele drop-in can be documented and a frequency determined.

4.3 *Study Example:*

For example, during validation and performance checks, numerous reagent blanks and amplification negative controls were processed in addition to hundreds of known DNA reference samples of varying concentrations. All the aforementioned blanks, controls and samples were processed through DNA typing.

4.4 *Study Outcome:*

The DNA profile results are reviewed to determine if contamination is present. Care should be taken when examining unexpected peaks to determine if they are truly a contamination peak, or could be an artifact such as increased stutter, pullup, a spike, etc. If contamination is detected, an investigation should be conducted into the possible source(s) of that contamination. Depending on the extent of contamination, it might be very difficult to determine the source of the contamination and re-amplifying and/or re-extracting (as applicable) the sample and obtaining clean results may demonstrate the contamination was an isolated sporadic event (e.g., drop-in).

If there is enough genetic information in the contamination that it can be determined to have originated from a sample processed simultaneously or from an analyst of any step in the process, the sample should be re-amplified and re-extracted (as applicable) as a means to eliminate the source and cause of contamination. If the source of contamination cannot be identified within the laboratory, it may be helpful to contact the manufacturer

of consumables to search an internal employee database. Purchasing and use of "DNA-free" consumables and disposable labware is also recommended if not already implemented. If contamination is observed, a plan shall be documented in the procedures to mitigate it and/or to address its occurrence(s) during normal laboratory operations.

Laboratories may use the information obtained from these evaluations to determine laboratory policy for the following:

- procedure requirements,
- laboratory environmental controls (workflow, cleaning),
- reporting guidelines,
- quality system requirements if contamination is detected
- personal protective equipment (PPE) requirements.

If higher than typical contamination risk is inherent to the method, cautionary steps may be documented in the procedure to mitigate it. For example, a laboratory may determine based on the data, that face masks and shields minimize contamination occurrences when worn during sampling and extraction and may therefore establish this as a requirement.

5. Known and Non-probative Evidence Samples or Mock Evidence Samples Study

5.1 *Study Purpose:*

Known and non-probative samples, or mock evidence samples, are used to evaluate genotype concordance information (and therefore the degree of accuracy of the system), as well as to assess the performance of validated parameters generated from the internal validation studies. This study may be used to assess the relative performance of the multiplex kit being validated against previously validated multiplex kit(s).

5.2 *Study Considerations:*

Concordance samples should be evaluated after conditions for casework samples have been established, including: optimal target/range, thermal cycling conditions, injection

time, and voltage. These samples should not have been used previously within these validation studies. The following sample types are recommended for use with this study:

- Samples from adjudicated cases or mock samples that represent those typically encountered by the laboratory for casework
- Single source samples of varying quality and/or quantity (appropriate NIST SRM (concordance) and/or proficiency test samples with reported consensus profiles, purchased DNA samples, positive controls, laboratory generated samples, etc. with known genotypes)
- DNA mixture samples of varying quality (proficiency test samples or purchased DNA mixture samples with known genotypes, laboratory generated mixtures, etc.)
- Samples representative of items and/or stains typically encountered by the testing laboratory
- Samples representing a variety of substrates and body fluids processed with respective extraction method(s)
- Samples representing a variety of template quantities, including low template DNA samples
- Samples created in the laboratory such as artificially degraded and/or inhibited/adulterated single source and mixed DNA samples (e.g., heat-degraded, moisture, super-glue fumed, enzyme or hematin treated, etc.)
- Inhibited/adulterated samples collected from environmental settings (e.g., soil, oil, sunlight, trunk of a hot car, etc.)

5.3 *Study Example:*

For example, three NIST SRM sample(s) are amplified and run concurrently with the laboratory's internal quality assurance sample(s) (NIST-traceable samples) using the established casework methods. Fifteen additional samples are processed through amplification and typing, including archived proficiency tests of varying substrates and body fluids processed with respective extraction method(s), laboratory generated single source and mixed DNA samples that are representative of the items typically encountered

by the laboratory, and inhibited/adulterated samples (i.e. samples degraded with heat, moisture and/or enzyme treatment).

These samples are processed using the parameters established from the other internal validation studies.

5.4 *Study Outcome:*

The STR typing results should be compared to any previous results (where applicable) and to any reference DNA profiles available to ensure concordance (i.e., demonstrate agreement between the results obtained compared to those using previous methods or published data).

The multiplex kit being validated should produce robust, reproducible results for high quality/quantity samples and concordant results for challenged samples. Any discordance should be documented and, where possible, a reason given for the discordance. Discordance may not invalidate the method. Some expected reasons for discordance may include null alleles, primer site or somatic mutations, or stochastic effects for low and challenged samples. If the multiplex kit fails to produce the expected outcome, the parameters may be re-evaluated, and additional validation studies or investigation may be necessary.

References

Bregu, J., Conklin, D., Coronado, E., Terrill, M., Cotton, R. W. and Grgicak, C. M. (2013) *Analytical Thresholds and Sensitivity: Establishing RFU Thresholds for Forensic DNA Analysis*. J Forensic Sci 58: 120–129. doi:10.1111/1556-4029.12008

Butler, J. M. (2011) Quality Assurance and Validation. In: *Advanced Topics in Forensic DNA Typing: Methodology.* Elsevier.

Gilder, J. R., Doom, T. E., Inman, K. and Krane, D. E. (2007) *Run-Specific Limits of Detection and Quantitation for STR-based DNA Testing*. J Forensic Sci 52: 97–101. doi:10.1111/j.1556-4029.2006.00318.x

Promega Corporation. (2013) *Internal Validation Guide of Autosomal STR Systems for Forensic Laboratories Reference Manual*. Available at http://www.promega.com.

Recommended Minimum Criteria for the Validation of Various Aspects of the DNA Profiling Process. ENSFI 2010.

Validation information on STRBase. Available at https://strbase.nist.gov/validation.htm .