# Scientific Working Group on DNA Analysis Methods

# Guidelines for the Validation of Probabilistic Genotyping Systems

## SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems

The Scientific Working Group on DNA Analysis Methods, better known by its acronym of SWGDAM, is a group of approximately 50 scientists representing Federal, State, and Local forensic DNA laboratories in the United States and Canada. During meetings, which are held twice a year, Committees discuss topics of interest to the forensic DNA community and often develop documents to provide direction and guidance for the community. In some instances, an Ad Hoc Working Group may be empanelled to address a particular topic outside of the routine SWGDAM January/July meeting schedule. These Guidelines, drafted by the SWGDAM Ad Hoc Working Group on Probabilistic Genotyping, were approved by the SWGDAM Executive Board for public comment in March 2015. Following the public comment period, the Ad Hoc Working Group forwarded the Final Guidelines to the SWGDAM Executive Board and they were approved for posting on the SWGDAM web site on June 15, 2015.

Guidance is provided herein for the validation of probabilistic genotyping software used for the analysis of autosomal short tandem repeat (STR) typing results. These guidelines are not

intended to be applied retroactively. It is anticipated that they will evolve with future developments in probabilistic genotyping systems.

**Introduction**

Probabilistic genotyping refers to the use of biological modeling, statistical theory, computer algorithms, and probability distributions to calculate likelihood ratios (LRs) and/or infer genotypes for the DNA typing results of forensic samples ("forensic DNA typing results"). Human interpretation and review is required for the interpretation of forensic DNA typing results in accordance with the FBI Director's Quality Assurance Standards for Forensic DNA Testing Laboratories[1]. Probabilistic genotyping is a tool to assist the DNA analyst in the interpretation of forensic DNA typing results. Probabilistic genotyping is not intended to replace the human evaluation of the forensic DNA typing results or the human review of the output prior to reporting.

A probabilistic genotyping system is comprised of software, or software and hardware, with analytical and statistical functions that entail complex formulae and algorithms. Particularly useful for low-level DNA samples (i.e., those in which the quantity of DNA for individuals is such that stochastic effects may be observed) and complex mixtures(i.e., multi-contributor samples, particularly those exhibiting allele sharing and/or stochastic effects), probabilistic genotyping approaches can reduce subjectivity in the analysis of DNA typing results.Historical methods of mixture interpretation consider all interpreted genotype combinations to be equally probable, whereas probabilistic approaches provide a statistical weighting to the different genotype combinations. Probabilistic genotyping does not utilize a stochastic threshold. Instead, it incorporates a probability of alleles dropping out or in.In making use of more genotyping information when performing statistical calculations and evaluating potential DNA contributors, probabilistic genotyping enhances the ability to distinguish true contributors and non-contributors.A higher LR is typically obtained when evaluating a person of interest (POI) who is a true contributor to the evidence profile, and a lower LR is typically obtained when the POI is not a true contributor. While the absence of an allele or the presence of additional allele(s)

---

[1] Probabilistic genotyping is to be distinguished from an Expert System.An Expert System, if NDIS approved and properly validated in accordance with the QAS, may only be used by a laboratory on database, known or casework reference samples to replace the manual review in accordance with the QAS and NDIS Operational Procedures. Expert Systems are not approved for use on forensic or forensic mixture DNA samples.

relative to a reference sample may support an exclusion, probabilistic genotyping approaches allow inclusion and exclusion hypotheses to be consideredby calculating a LR in which allele drop-out and drop-in may be incorporated.

The use of a likelihood ratio as a reporting statistic for probabilistic genotypingdiffers substantially from binary statistics such as the combined probability of exclusion.  Prior to validating a probabilistic genotyping system, the laboratory should ensure that it possesses the appropriate foundational knowledge in the calculation and interpretation of likelihood ratios.  Laboratories should also be aware of the features and limitations of various probabilistic genotyping programs and the impact that those items will have on the validation process.  Depending on the performance characteristics of the software, prerequisite studies may be required to, for example, establish parameters forallele drop-out and drop-in, stutter expectations, peak height variation, and the number of contributors to a mixture.Each laboratory seeking to evaluate a probabilistic genotyping system must determine which validation studies are relevant to the methodology, in the context of its application, to demonstrate the reliability of the system and any potential limitations.  The laboratory must determine the number of samples required to satisfy each guideline and may determine that a study is not necessary.  Some studies described herein may also be suitable for evaluating material modifications to existing procedures.

**Background**

Please refer to the SWGDAM Validation Guidelines for DNA Analysis Methods andthe FBI Quality Assurance Standards for Forensic DNA Testing Laboratories and for DNA Databasing Laboratories (QAS) for general background information regarding validation and definition of terms.

Probabilistic genotyping may generate a number of possible genotype combinations for a given profile, where some genotypes may be assigned more weight than others.  Allele drop-in and drop-out probabilities may be used in the determination of the weights associated with each of the possible genotypes.  There are two main approaches to probabilistic genotyping: the semi-continuous method and fully continuous method.  The semi-continuous method focuses only on the alleles present in the profile and addresses all possible genotype combinations of the

observed alleles in conjunction with a probability of drop-out and drop-in. Analysis parameters such as peak height variation, mixture ratios and stutter percentages are not typically utilized by semi-continuous software systems, although these elements may be considered during the initial manual evaluation of the data. The fully continuous method generally utilizesmore of the biological information in the profile, such as peak heights, stutter percentages and mixture ratios. The weighting of genotype combinations as more or less probable may be inferred from the data through methods such as Markov Chain Monte Carlo (MCMC) samplings from probability distributions.

The analyst will need to employ some level of interpretation before using the software to perform the calculations and should visually interpret allelic and non-allelic peaks and other characteristics of the DNA typing results, as necessitated by the software. For example, the analyst may be required to estimate and use a specific number of contributors in a statistical calculation when interpreting a DNA mixture, or to assess whether typing results should be interpreted or not based on quality.

Forensic DNA typing results interpreted by a DNA analyst using probabilistic genotyping software may be eligible for CODIS entry and upload to NDIS in accordance with the NDIS Operational Procedures if the probabilistic genotyping software has been properly validated pursuant to the QAS and these Guidelines.

1. **Validation of Probabilistic Genotyping Systems**
    1.1. The laboratory shall validate a probabilistic genotyping system prior to usage for forensic applications.
    1.2. The laboratory shall document all validation studies in accordance with the FBI Quality Assurance Standards for Forensic DNA Testing Laboratories.
    1.3. The laboratory should document or have access to documentation that explains how the software performs its operations and activities, to include the methods of analysis and statistical formulae, the data to be entered in the system, the operations performed by each portion of the user interface, the workflow of the system, and the system reports or

other outputs.  This information enables the laboratory to identify aspects of the system that should be evaluated through validation studies.

2. **System control**
   2.1. The laboratory should verify that the software is installed on computers suited to run the software, that the system has been properly installed, and that the configurations are correct.
   2.2. The laboratory should, where possible, ensure the following system control measures are in effect:

      2.2.1. Every software release should have a unique version number.  This version number should be referenced in any validation documentation or published results.
      2.2.2. Appropriate security protection to ensure only authorized users can access the software and data.
      2.2.3. Audit trails to track changes to system data and/or verification of system settings in place each time a calculation is run.
      2.2.4. User-level security to ensure that system users only perform authorized actions.

3. **Developmental Validation**
   Developmental validation of a probabilistic genotyping system is the acquisition of test data to verify the functionality of the system, the accuracy of statistical calculations and other results, the appropriateness of analytical and statistical parameters, and the determination of limitations. Developmental validation may be conducted by the manufacturer/developer of the application or the testing laboratory.  Developmental validation should also demonstrate any known or potential limitations of the system.
   3.1. The underlying scientific principle(s) of the probabilistic genotyping methods and characteristics of the software should be published in a peer-reviewed scientific journal. The underlying scientific principles of probabilistic genotyping include, but are not limited to, modeling of stutter, allelic drop-in and drop-out, Bayesian prior assumptions such as allele probabilities, and statistical formulae used in the calculation and algorithms.

3.2. Developmental validation should address, where applicable, the following:

3.2.1. Sensitivity – Studies should assess the ability of the system to reliably determine the presence of a contributor's(s') DNA over a broad variety of evidentiary typing results (to include mixtures and low-level DNA quantities). This should be evaluated using various sample types (e.g., different numbers of contributors, mixture proportions, and template quantities).

3.2.1.1. Sensitivity studies should demonstrate the potential for Type I errors (i.e., incorrect rejection of a true hypothesis), in which, for example, a contributor fails to yield a LR greater than 1 and thus his/her presence in the mixture is not supported.

3.2.1.2. Sensitivity studies should demonstrate the range of LR values that can be expected for contributors.

3.2.2. Specificity – Studies should evaluate the ability of the system to provide reliable results for non-contributors over a broad variety of evidentiary typing results (to include mixtures and low-level DNA quantities). This should be evaluated using various sample types (e.g., different numbers of contributors, mixture proportions, and template quantities).

3.2.2.1. Specificity studies should demonstrate the potential for Type II errors (i.e., failure to reject a false hypothesis), in which, for example, a non-contributor yields a LR greater than 1 and thus his/her presence in the mixture is supported.

3.2.2.2. Specificity studies should demonstrate the range of LR values that can be expected for non-contributors.

3.2.3. Precision – Studies should evaluate the variation in Likelihood Ratios calculated from repeated software analyses of the same input data. This should be evaluated using various sample types (e.g., different numbers of contributors, mixture proportions, and template quantities).

3.2.3.1. Some probabilistic genotyping approaches may not produce the same LR from repeat analyses. Where applicable, these studies should therefore demonstrate the range of LR values that can be expected from

multiple analyses of the same data and are the basis for establishing an acceptable amount of variation in LRs.

    3.2.3.2. Any parameter settings (e.g., iterations of the MCMC)that can reduce variability should be evaluated. For example, for some complex mixtures (e.g., partial profiles with more than three contributors), increasing the number of MCMC iterations can reduce variation in the likelihood ratio.

3.2.4. Case-type Samples – Studies should assess a range of data types exhibiting features that are representative of those typically encountered by testing laboratories. These features include those derived from mixtures and single-source samples, such as stutter, masked/shared alleles, differential and preferential amplification, degradation and inhibition.

    3.2.4.1. These studies should demonstrate sample and/or data types that can be reliably evaluated using the probabilistic genotyping system.

3.2.5. Control Samples – If the software is designed to assess controls, studies should evaluate whether correct results are obtained with control samples.

3.2.6. Accuracy – Studies should assess the accuracy of the calculations performed by the system, as well as allele designation functions, where applicable.

    3.2.6.1. These studies should include the comparison of the results produced by the probabilistic genotyping software to manual calculations, or results produced with an alternate software program or application, to aid in assessing accuracy of results generated by the probabilistic genotyping system. Calculations of some profiles (e.g., complex mixtures), however, may not be replicable outside of the probabilistic genotyping system.

    3.2.6.2. If the software uses raw data files from a genetic analyzer as input data, the peak calling, sizing and allele designation functions should be compared to the results of another software system to assess accuracy. Allele designations should also be compared to known genotypes where available.

4. **Internal Validation**

   Internal validation of a probabilistic genotyping software system is the accumulation of test data within the laboratory to demonstrate that the established parameters, software settings, formulae, algorithms and functions perform as expected. In accordance with the QAS, internal validation data may be shared by all locations in a multi-laboratory system.

   Depending on the features and capabilities of the probabilistic genotyping system, some DNA typing results may or may not be determined to be suitable for such analysis. To identify data features (e.g., minimum quality requirements, number of contributors) that render a profile appropriate or inappropriate for probabilistic genotyping, the laboratory should test data across a range of characteristics that are representative of those typically encountered by the testing laboratory. Data should be selected to test the system's capabilities and to identify its limitations. In particular, complex mixtures and low-level contributors should be evaluated thoroughly during internal validation, as the data from such samples generally help to define the software's limitations, as well as sample and/or data types which may potentially not be suitable for computer analysis. In addition, some exclusions may be evident without the aid of probabilistic software.

   If conducted within the same laboratory, developmental validation studies may satisfy some of the elements of the internal validation guidelines.

   4.1. The laboratory should test the system using representative data generated in-house with the amplification kit, detection instrumentation and analysis software used for casework. Additionally, some studies may be conducted by using artificially created or altered input files to further assess the capabilities and limitations of the software. Internal validation should address, where applicable to the software being evaluated:

      4.1.1. Specimens with known contributors, as well as case-type specimens that may include unknown contributors.

4.1.2.   Hypothesis testing with contributors and non-contributors

    4.1.2.1.   The laboratory should evaluate more than one set of hypotheses for individual evidentiary profiles to aid in the development of policies regarding the formulation of hypotheses.  For example, if there are two persons of interest, they may be evaluated as co-contributors and, alternatively, as each contributing with an unknown individual.  The hypotheses used for evaluation of casework profiles can have a significant impact on the results obtained.

4.1.3.   Variable DNA typing conditions (e.g., any variations in the amplification and/or electrophoresis parameters used by the laboratory to increase or decrease the detection of alleles and/or artifacts)

4.1.4.   Allelic peak height, to include off-scale peaks

4.1.5.   Single-source specimens

4.1.6.   Mixed specimens

    4.1.6.1.   Various contributor ratios (e.g., 1:1 through 1:20, 2:2:1, 4:2:1, 3:1:1, etc.)

    4.1.6.2.   Various total DNA template quantities

    4.1.6.3.   Various numbers of contributors.  The number of contributors evaluated should be based on the laboratory'sintended use of the software.A range of contributor numbers should be evaluated in order to define the limitations of the software.

    4.1.6.4.   If the number of contributors is input by the analyst, both correct and incorrect values (i.e., over- and under-estimating) should be tested.

    4.1.6.5.   Sharing of alleles among contributors

4.1.7.   Partial profiles, to include the following:

    4.1.7.1.   Allele and locus drop-out

    4.1.7.2.   DNA degradation

    4.1.7.3.   Inhibition

4.1.8.   Allele drop-in

4.1.9.   Forward and reverse stutter

4.1.10. Intra-locus peak height variation

4.1.11. Inter-locus peak height variation

4.1.12. For probabilistic genotyping systems that require in-house parameters to be established, the internal validation tests should be performed using those same parameters. The data set used to establish the parameters should be different from the data set used to validate the software using those parameters.

4.1.13. Sensitivity, specificity and precision, as described for Developmental Validation

4.1.14. Additional challenge testing (e.g., the inclusion of non-allelic peaks such as bleed-through and spikes in the typing results)

4.2. Laboratories with existing interpretation procedures should compare the results of probabilistic genotyping and of manual interpretation of the same data, notwithstanding the fact that probabilistic genotyping is inherently different from and not directly comparable to binary interpretation.The weights of evidence that are generated by these two approaches are based on different assumptions, thresholds and formulae. However, such a comparison should be conducted and evaluated for general consistency.

4.2.1. The laboratory should determine whether the results produced by the probabilistic genotyping software are intuitive and consistent with expectations based on non-probabilistic mixture analysis methods.

4.2.1.1. Generally, known specimens that are included based on non-probabilistic analyses would be expected to also be included based on probabilistic genotyping.

4.2.1.2. For single-source specimens with high quality results, genotypes derived from non-probabilistic analyses of profiles above the stochastic threshold should be in complete concordance with the results of probabilistic methods.

4.2.1.3. Generally, as the analyst's ability to deconvolute a complex mixture decreases, so do the weightings of individual genotypes within a set determined by the software.

5.  **Modification to Software**

    Modification to probabilistic genotyping software shall be addressed in accordance with the QAS.

    5.1.  Modification to the system such as a hardware or software upgrade that does not impact interpretation or analysis of the typing results or the statistical analysis shall require a performance check prior to implementation.

    5.2.  A significant change(s) to the software, defined as that which may impact interpretation or the analytical process, shall require validation prior to implementation.

    5.3.  Data used during the initial validation may be re-evaluated as a performance check or for subsequent validation assessment. The laboratory must determine the number and type of samples required to establish acceptable performance in consideration of the software modification.

**References and Suggested Readings**

Federal Bureau of Investigation, (2015) *NDIS Operational Procedures Manual*, available at http://www.fbi.gov/about-us/lab/biometric-analysis/codis/ndis-procedures-manual.

Federal Bureau of Investigation (2011) *Quality Assurance Standards for Forensic DNA Testing Laboratories,* available at http://www.fbi.gov/about-us/lab/codis/qas-standards-for-forensic-dna-testing-laboratories-effective-9-1-2011

Gill, P. et al. (2012)*DNA Commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods*, Forensic Science International Genetics 6(6): 679-688.

Kelly. H. et al. (2014)*A comparison of statistical models for the analysis of complex forensic DNA profiles*, Science &Justice 54(1): 66-70.

Scientific Working Group on DNA Analysis Methods (2010) *Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories*, available at http://www.swgdam.org/Interpretation_Guidelines_January_2010.pdf.

Scientific Working Group on DNA Analysis Methods (2012) *Validation Guidelines for DNA Analysis Methods*, available at http://www.swgdam.org/SWGDAM_Validation_Guidelines_APPROVED_Dec_2012.pdf .

Scientific Working Group on DNA Analysis Methods (2014) *Guidelines for STR Enhanced Detection Methods*, available at http://swgdam.org/SWGDAM%20Guidelines%20for%20STR%20Enhanced%20Detection%20Methods%20FINAL%20100614.pdf.

Steele, C. D. and Balding, D. J. (2014) *Statistical Evaluation of Forensic DNA Profile Evidence,* Annu. Rev. Stat. Appl. 1:361-384.